

Scavenger hunt: Selection of obscured active galactic nuclei combining multiband optical variability and colors[★]

D. De Cicco^{1,2,★★}, S. Cavuoti^{2,3}, M. Paolillo^{1,2,3}, V. Petrecca², Y. Maruccia², and P. Sánchez-Sáez⁴

¹ Department of Physics, University of Napoli “Federico II”, Via Cinthia 9, 80126 Napoli, Italy

² INAF – Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy

³ INFN – Sezione di Napoli, Via Cinthia 9, 80126 Napoli, Italy

⁴ European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching bei München, Germany

Received 31 July 2025 / Accepted 17 January 2026

ABSTRACT

Context. As wide-field optical surveys such as Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST) begin operations, time-domain astronomy is about to face a data revolution, paving the road for new and expanded variability studies.

Aims. This work leverages the complementary power of optical variability and color selection to identify active galactic nuclei (AGN), with a particular emphasis on optimizing the identification of obscured AGN, typically more challenging to distinguish from inactive galaxies based on optical variability alone. The analysis is designed to provide valuable insights in the context of performance forecasting for the LSST, albeit using a scaled-down version of the LSST dataset.

Methods. We present the first combined AGN selection based on *g*-, *r*-, and *i*-band light curves from the VST-COSMOS survey, spanning a 3.3 yr baseline. We identify AGN candidates independently in each band using a random forest (RF) classifier trained on features primarily related to optical variability, along with six optical/infrared colors and a morphology indicator. We subsequently merge the resulting band-specific samples in order to enhance selection purity and reliability. We then focus on defining a subset of features that significantly improve the identification of obscured AGN.

Results. The RF classifiers yield a consistent performance across the three bands, highlighting the critical role played by contamination. Using the combined three-band plus color selection we successfully recover $58_{-8}^{+9}\%$ of all AGN and $69_{-8}^{+10}\%$ of the known obscured AGN that have been independently confirmed in all three bands. When requiring confirmation in at least two out of the three bands, these fractions increase to $69_{-8}^{+10}\%$ and $80_{-9}^{+10}\%$, respectively. Moreover, we demonstrate that, while combining variability features with colors is crucial to improve obscured AGN selection, relying solely on color features results in a markedly higher contamination rate.

Key words. methods: statistical – surveys – galaxies: active

1. Introduction

Active galaxies are among the most luminous and variable extragalactic sources, powered by accretion onto supermassive black holes (SMBHs) hosted in their centers. Optical variability, generally detected on timescales from days to years, has proven to be a powerful tool for the identification of unobscured (or Type I) active galactic nuclei (AGN) in time-domain photometric surveys. Variations are aperiodic and trace instabilities in the accretion flow, probing the physical mechanisms at play in the surroundings of the SMBH.

However, the situation is markedly different for obscured (or Type II) AGN. In the scenario of the unified model (Antonucci 1993; Urry & Padovani 1995), the lack of broad optical emission lines in this class of sources is explained by the presence of a dusty torus preventing us from directly observing the accretion disk and broad-line region. As a result, the central engine is largely hidden in the optical, and what we detect is mostly emission from the narrow-line region and the host galaxy, with minimal contribution from the variable nucleus. Optical variability is therefore expected to be much weaker or absent (López-Navas et al. 2023), posing significant chal-

lenges for the selection of obscured AGN based on such a property. Despite theoretical predictions, a growing number of obscured AGN have been found to exhibit optical variability (e.g., López-Navas et al. 2022; Bernal et al. 2025), but the effectiveness of optical variability as a selection tool for obscured AGN remains a matter of active debate (see for instance the recent review by Paolillo & Papadakis 2025). This is intriguing, especially in the framework of upcoming large-scale multivisit surveys, such as the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST; Ivezić et al. 2019), expected to deliver vast amounts of high-cadence photometric data. Understanding the mechanisms behind optical variability of obscured AGN is therefore critical to improve AGN selection techniques, as well as to gain a deeper insight into AGN general structure and evolution.

This work is part of a series focused on AGN selection in the COSMOS field, based on the analysis of optical variability from time series obtained via the VLT Survey Telescope (VST; Capaccioli & Schipani 2011): this is a 2.6-meter optical telescope, with a field of view of 1 sq. deg. and a pixel scale of 0.214”; it is located in Chile, at the Paranal Observatory, and is dedicated to wide-field imaging surveys of the southern sky (see, e.g., Cappellaro et al. 2015; Falocco et al. 2015; Botticella et al. 2017; Fu et al. 2018; Liu et al. 2018, 2020; Poulain et al. 2020). Our dataset covers a 3.3 yr baseline and

[★] Observations were provided by the ESO programs 088.D-4013, 092.D-0370, and 094.D-0417 (PI G. Pignata).

^{★★} Corresponding author: demetra.decicco@unina.it

consists of observations in the g , r , and i bands; while we made extensive use of r -band data in De Cicco et al. (2015, 2019, 2021, 2022, 2025), Cavuoti et al. (2024), Maruccia et al. (2025), Kaviraj et al. (2026), the other two datasets are here used as a whole for the first time for this kind of analysis, though part of the g -band data were used in Lira et al. (2024), De Cicco et al. (2025).

The VST-COSMOS dataset can be regarded as a scaled-down analog of what the LSST is expected to deliver. While the analysis presented here is limited to a 3.3 yr baseline, our current monitoring extends beyond 11 yr, i.e., longer than the 10 yr baseline planned for LSST. Moreover, the single-visit depth of VST and LSST images are comparable. On the other hand, the VST-COSMOS area covers roughly one tenth of the area the LSST survey will cover per deep drilling field (DDF), and our analysis is limited to three optical bands, versus the six that LSST will provide. Given these considerations, one of the goals of our series of works has been to provide lower-limit expectations for the AGN selection performance of the LSST.

The primary goal of this work is to independently exploit the three optical bands available from the VST-COSMOS dataset and integrate each with static colors and morphological information from the literature, thereby combining the diagnostic power of optical variability and color selection to enhance AGN identification and expand the census of obscured AGN beyond that achieved in earlier studies. In particular, this work builds upon the approach of De Cicco et al. (2025), who made use of a random forest (RF; Breiman 2001) algorithm to identify AGN by testing different sets of features, and also focused on optimizing the selection of obscured AGN combining r - and g -band features together, identifying a feature set that returned a $(68.1 \pm 1.2)\%$ completeness for this class of sources. While this number is significantly lower than the near-100% completeness typically obtained for bright, unobscured AGN, it still represents a great improvement compared to earlier results trying to detect the bulk of the obscured AGN population (see, e.g., De Cicco et al. 2019). Here we make use of the same algorithm, but extend the analysis to three bands and expand the AGN labeled set (LS) used in the past encompassing sources with a wider diversity of properties. We propose a different approach, identifying AGN based on an independent selection across the three available bands. We then adopt the feature-optimization approach of De Cicco et al. (2025) to identify the most effective predictors to unearth obscured AGN, but this time we do it for each band separately in order to obtain independent results. We then investigate the advantages of combining the obtained samples to enhance the purity of our AGN selection.

The structure of this paper is as follows. Section 2 presents the dataset, the derived source catalogs, and the LS. Section 3 outlines the classification procedure carried out in the three bands, using two different LSs. Section 4 is dedicated to models using variability features or colors separately. Section 5 addresses the optimization of obscured AGN selection. Finally, Section 6 summarizes our main findings. We note that throughout this work we generally use the terms “obscured” for AGN that were classified as Type II via optical spectroscopy as they are characterized by peculiar optical variability, as is detailed in Section 5.

2. The VST-COSMOS gri dataset

As was mentioned above, our dataset consists of g -, r -, and i -band data. The r -band dataset consists of 54 visits, while the g - and i -band datasets have a lower observing cadence and

count 25 and 23 visits, respectively; this is due to differences in the original cadence planning of the SUDARE-VOICE surveys (Cappellaro et al. 2015; Vaccari et al. 2016; Botticella et al. 2017), designed for supernova studies, where these data come from. The single-visit depth of the g -, r -, and i -band data is of 25.0, 24.6, and 24.0 mag, respectively, at a confidence level of $\sim 5\sigma$. Each image originally covered an area of 1 sq. deg., but we masked $\approx 17\%$ of the field in order to exclude saturated stars, defected areas, and the noisiest regions on the edges of each image.

Table A.1 in Appendix A reports information about the three-band datasets. For general information about the data structure and the reduction process, we refer the reader to Section 2 of De Cicco et al. (2015) and Section 2 of De Cicco et al. (2019).

2.1. The gri catalogs

The catalog extraction and aperture correction processes follow what was done for the r -band data in De Cicco et al. (2019), described in detail in Section 3 of that work. In short, we used SExtractor (Bertin & Arnouts 1996) to obtain from each visit in each band a catalog of sources. Consistent with De Cicco et al. (2015, 2019), we chose to work with magnitudes measured within a $2''$ -diameter aperture, which typically encloses $\approx 70\%$ of the flux from a point-like source. In order to correct for the effect of having different seeing values from visit to visit and to compensate for potential calibration offsets, we identified a reference visit for each band and normalized all the others to this. Specifically, the reference visits have the lowest possible seeing compatible with the absence of significant defects; we chose images #48¹, 48, and 29 from Table A.1 for the r , g , and i band, respectively (in bold in the table).

We obtained a corrective factor for each visit by computing the average magnitude difference with respect to the reference image for all the sources with a magnitude in the range 16–21 mag. We chose this magnitude range as it minimizes the influence of very bright or very faint objects. In our previous works we masked part of the r -band images (saturated stars, halos, defected areas, and the image edges) making use of the masks obtained via the *Pulecenella* code (Huang et al. 2011). In De Cicco et al. (2015, 2019) we used the masked catalog obtained from a stacked image² as a reference, and matched all the single-visit catalogs with that, obtaining a master catalog of 22 927 sources, which we required being detected in at least half of the 54 r -band visits. Given the higher observing cadence of our r band with respect to the other two, here we used that same r -band master catalog as a starting point for cross-matching with the g - and i -band catalogs. As a consequence, we did not need to apply masks to the g - and i -band images, as we simply cross-matched the corresponding master catalogs with the r -band one. Consistent with what we did for the r -band master catalog, we required that the g - and i -band master catalogs include sources detected in at least half the visits. We obtained 23 990 and 26 682 sources detected in at least 13 and 12 visits

¹ We note that the reference visit for the r band is the same we used in De Cicco et al. (2019). As in that work we excluded one visit but kept track of its ID number, the reference image was there identified as #49. Here we did not keep track of that in order not to have missing numbers in the list. Indeed, the whole r -band dataset is a subset of the one we used in De Cicco et al. (2019).

² The stacked image was produced as the median of all the exposures having a seeing full width at half-maximum (FWHM) $< 0.80''$; it has a total exposure time of 19 800 s, and the limiting magnitude at 5σ above the background rms is $r(\text{AB}) \approx 26$ mag.

in the g and i band, respectively. The intersection of the three master catalogs returned a catalog of 16 910 sources.

2.2. The feature set for AGN classification

This work is based on the use of the same features used in De Cicco et al. (2019), which we report in Table 1 for the sake of convenience. The list includes 29 variability features, one morphology feature, and six colors. The variability features have been widely used in the literature (e.g., Ponti et al. 2012; Kim et al. 2016; Cabrera-Vives & Förster 2020; De Cicco et al. 2021; Sánchez-Sáez et al. 2021) for time series analysis and are computed from the source light curves; we computed these features independently for each band. Given the different sampling in each band, the maximum number of points per light curve will differ from band to band. Nevertheless, since we used these features independently in each band, this is not an issue. The morphology feature comes from the COSMOS Advanced Camera for Surveys (ACS) catalog (Koekemoer et al. 2007; Scoville et al. 2007b) from the *Hubble* Space Telescope, based on F814W imaging. This feature is a neural network-based classifier that distinguishes stars from galaxies and separates extended from point-like sources based on their morphology. In light of our performance forecasting objectives, we chose colors derived from the $uBrizy$ filters, to mirror those that will be available from the LSST³. We also included a mid-infrared (MIR) color that existing or forthcoming ancillary datasets provide or will provide in the very near future. This color proved to be essential in distinguishing AGN from inactive galaxies, as is shown, for example, in De Cicco et al. (2021). The static color features were obtained from the corresponding magnitudes reported in the COSMOS2015 catalog (Laigle et al. 2016) and thus they were computed only once, and the same values for each source were used for classification in each band. Additional information about these features can be found in Sects. 2.2, 2.3, and 2.4 of De Cicco et al. (2021).

2.3. Labeled set

As was mentioned in the introduction, this work builds on De Cicco et al. (2021, 2025). Specifically, both works made use of a LS consisting of AGN with X-ray emission and selected via optical spectroscopy from the *Chandra*-COSMOS Legacy Catalog⁴ (Marchesi et al. 2016) or AGN selected via their MIR properties based on the criterion by Donley et al. (2012), and of non-AGN which can be either stars or inactive galaxies, i.e., galaxies not showing any sign of nuclear activity. There are small differences between the LSs in De Cicco et al. (2021) and De Cicco et al. (2025) due to constraints originating from specific requirements.

The present work tests the expansion of the AGN LS used in our previous works, and hence begins with two parts where we made use of two different LSs: the first one (hereafter, the spec-MIR LS) takes the cue from the one used in the two works mentioned above, but it turns out to be smaller as we require each source in the LS to be present in the g , r , and i master catalogs. It includes 2154 sources, specifically:

- 358 AGN (versus the 380 AGN used in De Cicco et al. 2025), of which 209 spectroscopically classified as Type I and 96 spectroscopically classified as Type II; 203 of these 358 sources are classified as AGN via their MIR properties following Donley et al. (2012), and 53 of these 203 sources do not have spectroscopic classification (MIR-only AGN subsample);
- 1796 non-AGN (versus the 2163 non-AGN used in De Cicco et al. 2025).

In the second part of the work we include in the LS another sub-class of AGN, i.e., 162 additional AGN selected only on the basis of their X-ray-to-optical flux ratio X/O (following Maccacaro et al. 1988, but also Hornschemeier et al. 2001; Xue et al. 2011; Civano et al. 2012, we require them to have $X/O > -1$). Hereafter, we will label this subsample of 162 sources as X-ray-only AGN. The total LS in this second part (hereafter, the main LS) therefore consists of 2316 sources in total – 1796 non-AGN and 520 AGN – corresponding to an increase of nearly 50% with respect to the AGN in the spec-MIR LS. The inclusion of X-ray-selected AGN introduces greater diversity, including also some low-luminosity AGN. We caution that the spec-MIR LS does include AGN also confirmed via X-ray properties, but we did not select them based on this property; hence they are always classified as AGN via at least one additional property, contrary to the X-ray-only AGN subsample that we added to the main LS.

We stress that, consistent with previous works in this series, our classifiers are always binary, which means that we did not deal with sub-classification within the two classes of AGN and non-AGN, though we used that information when discussing the selection of obscured AGN. We also point out that color and morphological information is available for all the sources in our LS. Additional details about how we selected the samples of sources making up our LS are reported in Section 2.5 of De Cicco et al. (2021) and Section 2.2 of De Cicco et al. (2025); hence we do not include them here for the sake of conciseness.

Table 2 reports the number of sources belonging to AGN and non-AGN as well as to the various sub-samples of sources they consist of. We point out that spec-MIR LS is a subsample of the main LS, and also note that a source can be classified as AGN based on more than one diagnostic, which means that the sum of the numbers reported for the AGN class will exceed the corresponding number of AGN.

3. Classification via a random forest algorithm

This work aims at combining AGN classification in three different bands, and starts with two separate experiments based on the use of two distinct LSs, as is detailed in Table 2. In both parts we focused on the classification of the sources in our LS via a RF algorithm. The classification algorithm was used on the g , r , and i bands independently, then we also cross-matched the obtained results. Color features and a morphology indicator are always added to the list of variability-based features used for each band.

Our code is based on the use of the Python *scikit-learn* library (Pedregosa et al. 2011). In what follows we essentially adopted the same approach described in Section 4 of De Cicco et al. (2025), which here we briefly summarize. Since our LS is unbalanced, in the algorithm we set `class_weight = balanced_subsample`: in this way, every time a bootstrap sample is used to extract a subset of features to build a tree, the algorithm adjusts the weights of each class dynamically based on the class distribution in that specific bootstrap sample used to train the corresponding tree.

³ Since g -band magnitudes are not provided in the chosen reference catalog (see further), we replaced them with the Subaru B band, which closely overlaps the g -band wavelength coverage.

⁴ We stress that the AGN from the *Chandra*-COSMOS Legacy Catalog have, by construction, X-ray detection, but were classified as Type I or Type II via optical spectroscopy.

Table 1. List of variability, morphology, and color features used in this work.

	Feature	Description	Reference
classic var. features	A_{SF}	rms magnitude difference of the SF, computed over a 1 yr timescale	Schmidt et al. (2010)
	γ_{SF}	Logarithmic gradient of the mean change in magnitude	Schmidt et al. (2010)
	GP_DRW_ τ	Relaxation time τ (i.e., time necessary for the time series to become uncorrelated), from a DRW model for the light curve	Graham et al. (2017)
	GP_DRW_ σ	Variability of the time series at short timescales, from a DRW model for the light curve	Graham et al. (2017)
	ExcessVar	Measure of the intrinsic light curve variance	Nandra et al. (1997); Allevato et al. (2013)
	P_{var}	Probability that the source is intrinsically variable	McLaughlin et al. (1996)
	IAR $_{\phi}$	Level of autocorrelation using a discrete-time representation of a DRW model	Eyheramendy et al. (2018)
FATS features	Amplitude	Half of the difference between the median of the maximum 5% and of the minimum 5% magnitudes	Richards et al. (2011)
	AndersonDarling	Test of whether a sample of data comes from a population with a specific distribution	Nun et al. (2015)
	Autocor_length	Lag value where the autocorrelation function becomes smaller than η^e	Kim et al. (2011)
	Beyond1Std	Percentage of points with photometric mag that lie beyond 1σ from the mean	Richards et al. (2011)
	η^e	Ratio of the mean of the squares of successive mag differences to the variance of the light curve	Kim et al. (2014)
	Gskew	Median-based measure of the skew	–
	LinearTrend	Slope of a linear fit to the light curve	Richards et al. (2011)
	MaxSlope	Maximum absolute magnitude slope between two consecutive observations	Richards et al. (2011)
	Meanvariance	Ratio of the standard deviation to the mean magnitude	Nun et al. (2015)
	MedianAbsDev	Median discrepancy of the data from the median data	Richards et al. (2011)
	MedianBRP	Fraction of photometric points within amplitude/10 of the median mag	Richards et al. (2011)
	MHAOV_Period	Period obtained via the Multi-Harmonic Analysis Of Variability periodogram	Huijse et al. (2018)
	PairSlopeTrend	Fraction of increasing first differences minus fraction of decreasing first differences over the last 30 time-sorted mag measures	Richards et al. (2011)
	PercentAmplitude	Largest percentage difference between either max or min mag and median mag	Richards et al. (2011)
	Q31	Difference between the 3 rd and the 1 st quartile of the light curve	Kim et al. (2014)
	Period_fit	False-alarm probability of the largest periodogram value obtained with Lomb-Scargle	Kim et al. (2011)
	Ψ_{CS}	Range of a cumulative sum applied to the phase-folded light curve	Kim et al. (2011)
	Ψ_{η}	η^e index calculated from the folded light curve	Kim et al. (2014)
	R_{CS}	Range of a cumulative sum	Kim et al. (2011)
	Skew	Skewness measure	Richards et al. (2011)
Std	Standard deviation of the light curve	Nun et al. (2015)	
StetsonK	Robust kurtosis measure	Kim et al. (2011)	
morph.	class_star	<i>HST</i> stellarity index	Koekemoer et al. (2007), Scoville et al. (2007a)
colors	u-B	CFHT <i>u</i> mag – Subaru <i>B</i> mag	Laigle et al. (2016)
	B-r	Subaru SuprimeCam <i>B</i> mag – Subaru SuprimeCam <i>r</i> + mag	Laigle et al. (2016)
	r-i	Subaru SuprimeCam <i>r</i> + mag – Subaru SuprimeCam <i>i</i> + mag	Laigle et al. (2016)
	i-z	Subaru SuprimeCam <i>i</i> + mag – Subaru SuprimeCam <i>z</i> ++ mag	Laigle et al. (2016)
	z-y	Subaru SuprimeCam <i>z</i> ++ mag – Subaru Hyper-SuprimeCam <i>y</i> mag	Laigle et al. (2016)
	ch21	<i>Spitzer</i> 4.5 μm (<i>channel2</i>) mag – 3.6 μm (<i>channel1</i>) mag	Laigle et al. (2016)

Notes. The first two sections of the table list the variability features used; `class_star`, the only morphology feature used, follows; the bottom section reports the color features used: all of them are optical/near-infrared colors except for `ch21`, which is a MIR color. This table corresponds to Table 1 in De Cicco et al. (2021).

We optimized the performance of each RF classifier, operating on the two possible LSs in each band, testing a set of possible combinations of the main hyperparameters involved in the RF building process⁵. This was achieved via a grid search, which explores all the combinations of the selected hyperparameter values using cross-validation and predefined scoring metrics to identify the optimal set. Given the potentially infinite number of possible combinations and the high computational cost, it was necessary to constrain the set of values

explored for each hyperparameter. We tested 216 combinations of hyperparameter per classifier, focusing on the following five hyperparameters: `n_estimators`, `min_samples_split`, `max_depth`, `min_samples_leaf`, `max_features`. In Appendix B we provide a short description of these five hyperparameters, together with details about the tested values and the obtained best sets. We evaluated model performance via the balanced accuracy, which averages the recall of both classes and accounts for the class imbalance in our LS.

A common approach in machine learning is to divide the LS into two disjoint subsets: a training set (typically 70–75%) for model fitting, and a validation set (30–25%) for performance evaluation during training. However, due to the heterogeneous nature of our AGN sample – selected on the basis of different properties – this strategy would make the results highly

⁵ In machine learning, the settings you choose before training an algorithm are called hyperparameters, and they define how the algorithm will work. For a RF algorithm they will establish, for instance, the number of trees to be used in the forest, the maximum depth of each tree, the number of features to consider when splitting a node.

Table 2. Information about the number of sources in the LS and its various parts.

	AGN		non-AGN			
	Type I	Type II	MIR	X-ray	Stars	Gal
spec-MIR LS (358 AGN)	209	96	203	–	841	955
^c			(53 MIR-only)			
main LS (520 AGN)	209	96	203	477	841	955
			(53 MIR-only)	(162 X-ray only)		

Notes. The table shows the composition of the LS for the two parts of this work: the top line refers to the LS where we did not include AGN selected via their *X/O*, while these sources are included in the bottom line. The rest of the LS is unchanged and, in particular, the number of non-AGN in each LS is always 1796. The various columns indicate (left to right): AGN classified via optical spectroscopy as Type I or Type II, AGN selected via their MIR properties, AGN selected via their *X/O*, plus stars and inactive galaxies which make up the non-AGN part of the LS.

sensitive to the specific AGN types included in the training set. Following De Cicco et al. (2021, 2025), we therefore adopted the leave-one-out cross-validation (LOOCV; Sammut & Webb 2010), which treats each source in the LS as a single-instance validation set, while the remainder is used for training. This process yields a prediction for every object in the LS, effectively using the full sample for both training and validation without overlap.

The first step in our selection process is to select all the sources classified as AGN in the LS, for each of the tested models, and evaluate the performance of the classifier. We report some classic evaluation metrics for these models in Sections I and II of Table 3. The table includes two of the usual classes for a confusion matrix, namely, true positive rate (TPR), true negative rate (TNR)⁶; it also reports accuracy, precision, F1, and recall⁷, the last one also being computed for various sub-classes of AGN, defined on the basis of their selection method. Section I of the table refers to the three tests on the spec-MIR LS, which does not include AGN confirmed only via their X-ray properties, while the main LS, used in the three tests in Section II of the table, includes these sources. Each percentage reported in the table is obtained as the average of ten experiments.

The tests on the spec-MIR LS do not show any significant differences among the three bands, except for a higher value for the overall recall in the *g* band, as well as for its various sub-classes. This might be ascribed to the faster variability of AGN in the *g* band.

The tests on the main LS mostly show the same trend for all the metrics but the recall, the *r* band mostly returning slightly better performance. For what concerns recall, we note that the overall and most of the sub-class values are consistent within their uncertainties, while we obtained slightly higher values in the *g* band for obscured AGN and in the *i* band for MIR AGN. We also note that the X-ray AGN recall is basically independent on the band.

⁶ True positives (TPs): known AGN correctly classified; true negatives (TNs): known non-AGN correctly classified. A confusion matrix also reports the number of false positives (FPs): known non-AGN incorrectly classified, and the number of false negatives (FNs): known AGN incorrectly classified. For each quantity we can obtain the corresponding ratio with respect to the total number of AGN or non-AGN.

⁷ Accuracy $A = \frac{TPs+TNs}{Tot. Sample}$; precision $P = \frac{TPs}{TPs+FPs}$, $F1 = 2 \times \frac{P \times R}{P+R}$, recall $R = \frac{TPs}{TPs+FNs}$.

If we compare the two upper sections of the table in order to assess the effect of the inclusion of the X-ray only AGN in the LS, we notice that, essentially, they increase the heterogeneity of the LS, which worsens the TPR as well as the TNR. This leads, as expected, to a general worsening of all the metrics. While the recall of unobscured AGN and MIR AGN does not change dramatically, we do observe a decreased X-ray recall. We stress that, when using the spec-MIR LS, the sub-sample of AGN that are confirmed via X-ray properties is dominated by spectroscopically confirmed AGN, as they make up 305 out of the 315 sources in the X-ray LS, while the remaining 10 are AGN confirmed via their MIR properties. When we use the main LS and hence include X-ray-only AGN, on the other hand, we include potentially different sources, their only known property as AGN being their *X/O*; as a consequence, we interpret this decreased X-ray recall as another effect of heterogeneity. The most striking consequence of expanding the LS is, nevertheless, the increase in the recall of obscured AGN, which is around 50% when using the spec-MIR LS and rises to >70% values when resorting to the main LS. In spite of the generally lower performance of the classifiers based on the main LS, we stress that a diversified LS represents the most interesting case scenario, as it is what we commonly have to deal with if we aim at including as many source types as possible in our analysis. Hence, the main LS is the one we will focus our further work on in the next sections.

4. Models using optical variability or colors alone

As also mentioned in the introduction to this work, it is well known that in obscured AGN the nuclear continuum and broad-line region are hidden by circumnuclear material, so that the observed optical emission is dominated by the host galaxy. As a result, the intrinsic variability signal is heavily diluted or completely suppressed, making obscured AGN much harder to recover through optical variability-based selection alone. In previous works in this series (De Cicco et al. 2021, 2025) we discussed the primary role of color features for improving the identification of obscured AGN. This naturally raises the question of whether it would be possible to rely solely on color features for classification, thereby avoiding the complexity of time-domain data, light curves, and all the associated challenges. In what follows, we report on two complementary tests in which we constructed different RF classifiers using, respectively, variability features alone in each available band, or color features alone. The comparison with the previous tests in this work assesses the distinct advantages and shortcomings of these approaches in the context of AGN selection.

The first set of experiments in this section uses only the 29 variability features reported in Table 1. The RF classifier is trained on the main LS. Once again, we followed our standard hyperparameter optimization procedure. All the experiments limited to variability features only yielded remarkably worse results. Section III of Table 3 presents the metrics obtained when using the full set of 29 variability features from Table 1. Compared to Section II, where the same LS and the full set of features were used, a slight enhancement is observed only in the TNR, but this comes at the cost of marked declines in the TPR, F1 score, and all the recall values. In particular, the recall for obscured AGN is notably lower than in any other experiments presented in this work. We also observe that, overall, the recall in the *r* band is broadly consistent with the findings of De Cicco et al. (2021). However, we caution that the LS used in that study was less heterogeneous than the one adopted here

Table 3. Confusion matrix values for the various RF classifiers tested in the *gri* bands.

Band	LS	features	TPR (recall)	TNR	accuracy	precision	F1	unobscured AGN recall	obscured AGN recall	X-ray recall	MIR recall	
I	<i>g</i>	spec-MIR	all	81.6 ± 0.2	98.18 ± 0.16	95.42 ± 0.14	89.9 ± 0.8	85.6 ± 0.4	99.7 ± 0.2	53.2 ± 0.6	85.7 ± 0.3	88.2 ± 0.4
	<i>r</i>	spec-MIR	all	79.3 ± 0.8	98.74 ± 0.15	95.52 ± 0.15	92.6 ± 0.8	85.5 ± 0.5	99.00 ± 0.15	50.1 ± 1.5	84.2 ± 0.5	86.4 ± 1.0
	<i>i</i>	spec-MIR	all	79.2 ± 0.6	98.03 ± 0.10	94.90 ± 0.16	88.9 ± 0.5	83.8 ± 0.5	99.2 ± 0.2	46.4 ± 1.5	83.2 ± 0.6	87.9 ± 0.7
II	<i>g</i>	main	all	74.8 ± 0.4	95.72 ± 0.15	91.03 ± 0.11	83.5 ± 0.4	78.7 ± 0.2	99.5 ± 0.0	73.5 ± 1.8	76.3 ± 0.4	89.2 ± 0.3
	<i>r</i>	main	all	74.3 ± 0.5	96.4 ± 0.2	91.4 ± 0.2	85.7 ± 0.7	79.4 ± 0.5	99.6 ± 0.2	71 ± 2	76.2 ± 0.5	89.5 ± 0.6
	<i>i</i>	main	all	74.9 ± 0.9	95.0 ± 0.2	90.5 ± 0.2	81.4 ± 0.6	77.8 ± 0.6	99.6 ± 0.2	72.8 ± 1.9	76.3 ± 0.9	91.5 ± 0.6
III	<i>g</i>	main	var. features only	47.69 ± 0.18	98.28 ± 0.14	86.92 ± 0.13	88.9 ± 0.8	62.1 ± 0.3	92.30 ± 0.15	29.2 ± 0.0	51.6 ± 0.2	66.45 ± 0.16
	<i>r</i>	main	var. features only	51.6 ± 0.2	99.31 ± 0.07	88.60 ± 0.07	95.6 ± 0.4	67.0 ± 0.2	97.0 ± 0.2	37.4 ± 0.3	55.6 ± 0.2	70.9 ± 0.0
	<i>i</i>	main	var. features only	47.0 ± 0.5	95.5 ± 0.3	84.6 ± 0.3	75.3 ± 1.3	57.8 ± 0.6	88.3 ± 0.7	24.5 ± 1.1	49.5 ± 0.5	69.5 ± 1.0
IV	–	main	<i>uBrizy</i> -derived colors only	73.8 ± 0.3	89.1 ± 0.3	85.7 ± 0.2	66.3 ± 0.6	69.9 ± 0.4	96.0 ± 0.3	64.6 ± 1.0	75.9 ± 0.4	82.7 ± 0.6
	–	main	all colors only	76.0 ± 0.8	93.8 ± 0.2	89.8 ± 0.3	78.1 ± 0.7	77.0 ± 0.7	99.3 ± 0.3	68 ± 3	76.6 ± 0.9	93.3 ± 0.5
V	<i>g</i>	main	13, optimized for obscured AGN sel.	77.8 ± 0.8	94.76 ± 0.14	90.96 ± 0.15	81.1 ± 0.4	79.5 ± 0.4	99.95 ± 0.15	81 ± 2	79.2 ± 0.7	91.8 ± 0.6
	<i>r</i>	main	14, optimized for obscured AGN sel.	76.7 ± 0.6	95.7 ± 0.2	91.4 ± 0.2	83.7 ± 0.7	80.0 ± 0.5	99.5 ± 0.0	79.0 ± 1.8	78. ± 0.7	90.0 ± 0.3
	<i>i</i>	main	10, optimized for obscured AGN sel.	80.1 ± 0.2	94.72 ± 0.17	91.42 ± 0.14	81.4 ± 0.5	80.7 ± 0.3	99.5 ± 0.0	84.3 ± 0.9	81.1 ± 0.2	93.3 ± 0.2

Notes. The table reports information about the band, LS, and features used for each test, as well as true positive rate (TPR), true negative rate (TNR), accuracy, overall precision values, F1, and recall for unobscured and obscured AGN, X-ray AGN, and MIR AGN. The overall value of the recall is by definition the same as the TPR. Sections I and II report the results obtained using the two different LSs adopted in this work. Sections III and IV report the results from the tests where only variability features or only colors are used, respectively, which will be discussed in Section 4. Section V is focused on the selection of obscured AGN (see also Table 5), which will be discussed in Section 5. All values are to be read as percent values. The percentage errors represent the standard deviation from the mean value derived from a set of ten experiments per classifier. The Roman numerals in the leftmost column help identify the various sections of the table throughout this work.

(as is discussed in Section 2.3). A further noteworthy outcome from this analysis is that the *r* band consistently outperforms the other two bands, which can very likely be attributed to its higher observing cadence. In particular, the *r*-band precision and the TNR from this model are the highest across the whole table, suggesting that the classifier did an excellent job in identifying non-AGN via their optical variability features alone. By contrast, compared to the other two bands, the *i* band exhibits a larger scatter in its performance, and its metrics are the lowest across all the tests in the table. Nevertheless, this band still confirms useful in identifying MIR AGN.

We then tested a RF classifier trained on the main LS, using only the color features expected to be available from the LSST – namely, *u–B*, *B–r*, *r–i*, *i–z*, and *z–y* – with no use of variability features. We followed the same hyperparameter optimization procedure adopted throughout this work and combined the results from ten experiments as usual. We reported the outcomes in the first line of Section IV of Table 3. As clearly shown in the table, this purely color-based approach results in an overall significant decline in classifier performance compared to Section II of the table. In particular, we observe that the precision sinks, being 15–17% lower, which means the classification will suffer from substantial contamination. This is a typical issue when using only colors, and excluding the variability features typically significantly hampers the classification accuracy and leads to critically contaminated AGN samples (e.g., De Cicco et al. 2021). We also note an average 6% drop in the TNR across the three bands, which means that, on average, based on the size of our LS we are misclassifying >100 more non-AGN compared to previous experiments. The recall for obscured AGN is also lower by ≈8% on average across the three bands.

We also tested the use of all six color features from Table 1, adding the MIR color feature *ch21* to the previous experiments since, as was mentioned above, this is well known to be crucial for the identification of obscured AGN. The obtained results are reported in the second line of Section IV of Table 3. In this case the results are much better than in the previous line in the table but, overall, the exclusion of the variability features worsens the performance compared to Section II (especially the *g* and *r*-band-related lines), pointing toward a higher contamina-

tion. Only the TPR and, as expected, the recall for MIR AGN are higher than in Section II.

The results analyzed in this section aim at confirming that neither optical variability features alone nor color features alone are as effective in selecting AGN as their combination is; and, in particular, if we focus on obscured AGN we find that so far the highest recall has been returned by the full set of features from Section II of Table 1. This said, in the next section we will focus on optimizing the selection of obscured AGN.

5. Selection of obscured AGN

In previous works in this series (De Cicco et al. 2021, 2025) we extensively discussed the issues affecting the selection of obscured AGN via optical variability and their generally much lower completeness compared to unobscured AGN. De Cicco et al. (2025) also showed how to identify an optimized set of features to select obscured AGN and obtained a more complete sample than previous works in the series, though still affected by a large contamination. Specifically, that previous analysis explored whether combining the *g*- and *r*-band features could improve classifier performance. In order to do so, the analysis was limited to 33 visits common to the two bands, where some gaps in either band had been filled via data imputation. The approach, however, did not result in any significant gain for the selection of obscured AGN, as the best-performing feature subsample ultimately included, among others, variability features from one band only, this being the *r*. Here we therefore adopted a different strategy, using the data from the three available bands – *g*, *r*, and *i* – independently to select AGN in each of them. This means that here we use all the available visits, without introducing artificial data to account for missing visits. This approach will be beneficial for LSST data where different bands are likely to be sampled with different cadences, although these will be higher than VST cadences. We finally merged these independent results in order to obtain a cleaner and hence more robust sample.

We have seen that our reference sample of obscured AGN consists of 96 sources detected in the X-rays and classified as Type II via optical spectroscopy. We point out that these sources

were also used in De Cicco et al. (2022), where the structure function of various subsamples of AGN selected via different properties and methods was analyzed. In that case, the sample of Type II AGN consisted of 104 sources, and thus included eight additional sources that we had to exclude from the present work based on the adopted cuts, as is detailed in Section 2.3. Nevertheless, 92% of the sample of Type II AGN used in De Cicco et al. (2022) consisted of the same Type II AGN here used, and that work clearly shows (e.g., in their Fig. 5) that this subsample of sources appears to be less variable than all the other subsamples of AGN there analyzed. This suggests that, while the Type II classification of these sources may be debated based on the low signal-to-noise ratio of the original spectra and the fact that they were observed at a different time with respect to the VST light curves used here, they are still a peculiar subsample of objects for what concerns their optical variability, less variable and harder to identify, and are therefore worth being investigated separately.

As we mentioned earlier, each of the values reported in Table 3 is the average outcome of a series of ten experiments. As a consequence, each of these experiments will provide an independent classification for each source in the LS. If we focus on obscured AGN, we can combine the results from the ten experiments corresponding to each tested model, and obtain a final classification as AGN or non-AGN for each source. In order to do so, we need to set a classification threshold. In the top and middle sections of Table 4 we report the results obtained on the basis of three different thresholds, for the first two series of ten experiments testing two different LSs: first, we considered a source to be an AGN only if all the ten experiments corresponding to a given model classify it as an AGN (10/10 threshold); then we relaxed this criterion and lowered the threshold to eight out of ten positive outcomes (8/10 threshold), and finally we tested an even less conservative approach, adopting a 6/10 threshold. Once we obtained the final classification for each source per band, we cross-matched the results and obtained a final sample of sources classified as AGN in all three bands, for each of the tested thresholds. We also reported the number of sources classified as AGN in two out of the three bands, as well as the union of the results, i.e., the number of sources classified as AGN in at least one band. This last number is to be considered as an upper limit to the completeness that we can achieve based on the used data and feature sets and we stress that, although some sources are classified as AGN only in one band, this classification is still the result of multiple (10, 8 or 6) tests.

From the table we can see that, for the tests based on the spec-MIR LS, the final results intersecting the samples obtained for each band are not very different from each other, regardless of the adopted threshold; on the contrary, the threshold seems to matter more in tests using the main LS. Based on these results, we decided to build our further analysis on the test performed using the main LS, and with 8/10 as a threshold; indeed, this LS returns a much improved recall for obscured AGN, and the threshold is a representative average benchmark for the following discussion, since especially for the union the results do not change much.

Since MIR information is available from the COSMOS2015 catalog (Laigle et al. 2016) for the whole sample of 96 obscured AGN used in this work, we reported them on a diagram in Fig. 1, comparing two MIR colors obtained from the four IRAC filters corresponding to 3.6 μm , 4.5 μm , 5.8 μm , and 8.0 μm . The figure includes lines delimiting typical AGN loci based on Lacy et al. (2007) and Donley et al. (2012). It is apparent that a significant fraction of the obscured AGN in our sample do not lie in these

Table 4. Number of obscured AGN confirmed from each tested model, in each band.

spec-MIR LS	<i>g</i>	<i>r</i>	<i>i</i>	and (\cap)	2 out of 3 bands	or (\cup)
10/10	49	44	40	33 (34 $^{+7}_{-6}$ %)	46 (48 $^{+8}_{-8}$ %)	55 (57 $^{+10}_{-9}$ %)
8/10	51	46	43	36 (38 $^{+7}_{-6}$ %)	48 (50 $^{+8}_{-8}$ %)	57 (59 $^{+9}_{-8}$ %)
6/10	51	47	44	36 (38 $^{+7}_{-6}$ %)	49 (51 $^{+8}_{-7}$ %)	59 (61 $^{+9}_{-8}$ %)
<hr/>						
main LS						
10/10	65	62	59	49 (51 $^{+8}_{-8}$ %)	61 (64 $^{+9}_{-8}$ %)	80 (83 $^{+10}_{-9}$ %)
8/10	68	65	65	56 (58 $^{+9}_{-8}$ %)	66 (69 $^{+10}_{-8}$ %)	79 (82 $^{+10}_{-9}$ %)
6/10	71	67	68	60 (62 $^{+9}_{-8}$ %)	68 (71 $^{+10}_{-9}$ %)	79 (82 $^{+10}_{-9}$ %)
<hr/>						
main LS, obscured AGN-focused						
10/10	69	69	78	60 (62 $^{+9}_{-8}$ %)	74 (77 $^{+10}_{-9}$ %)	89 (93 $^{+11}_{-10}$ %)
8/10	74	72	81	66 (69 $^{+10}_{-8}$ %)	77 (80 $^{+10}_{-9}$ %)	88 (92 $^{+11}_{-10}$ %)
6/10	79	76	81	72 (75 $^{+10}_{-9}$ %)	79 (82 $^{+10}_{-9}$ %)	87 (91 $^{+11}_{-10}$ %)

Notes. The top section refers to tests using the spec-MIR LS, i.e., not including AGN confirmed only via their X-ray properties. The middle section refers to tests using the main LS, which includes these AGN. The bottom section refers to additional tests discussed later in this section, where we used the main LS and a specific selection of features for each band, aimed at optimizing the selection of obscured AGN. The various columns indicate: the adopted threshold (1); the three bands (2–4), where the final number for each model tested was obtained combining the outcomes of the ten independent experiments, according to the threshold indicated by each line; the intersection of the results (5), i.e., the number (and percentage) of obscured AGN confirmed in all three bands; the number (and percentage) of obscured AGN confirmed in two out of three bands; and the union of the results (6), i.e., the number (and percentage) of sources classified as obscured AGN in at least one band. Error bars on the reported percentages were estimated following the approach of Gehrels et al. (2015), who provided approximate formulae for 95% confidence intervals under the assumptions of Poisson and binomial statistics. Percentages were computed with respect to the total number of obscured AGN, these being 96.

regions, which is both intriguing and challenging as it shows how different AGN selection techniques are complementary: indeed, in this case it shows how selection based mostly on optical variability properties combined with colors is able to unearth AGN that would be otherwise missed via MIR-based selection, as they lie in the area where star-forming galaxies are typically found. Indeed, Donley et al. (2012) discuss how in this region we can find obscured AGN that will be mostly missed via X-ray or MIR selection. Another point is that, since the MIR data we used are not contemporary to our VST data, the state of at least some of these sources might have changed; if so, they would show different properties in optical and MIR wavelengths.

One of the primary challenges in identifying obscured AGN lies in distinguishing them from inactive galaxies. Following the approach outlined in De Cicco et al. (2025), we analyzed the feature sets in each band to assess which ones are most effective in separating these two classes of sources. For each feature, we therefore compared the distributions corresponding to obscured AGN and inactive galaxies, applying a Kolmogorov-Smirnov (K-S) test to identify the features whose distributions differ significantly and for which the probability of obtaining by chance a larger difference is low. Based on the K-S test results in each band, we defined the corresponding thresholds and thus identified a subset of discriminative features to be used in the subsequent analysis. We stress that, while the procedure closely follows the approach adopted in De Cicco et al. (2025), there are some key differences here: first of all, since here we are using the three bands independently, the light curve-derived features for

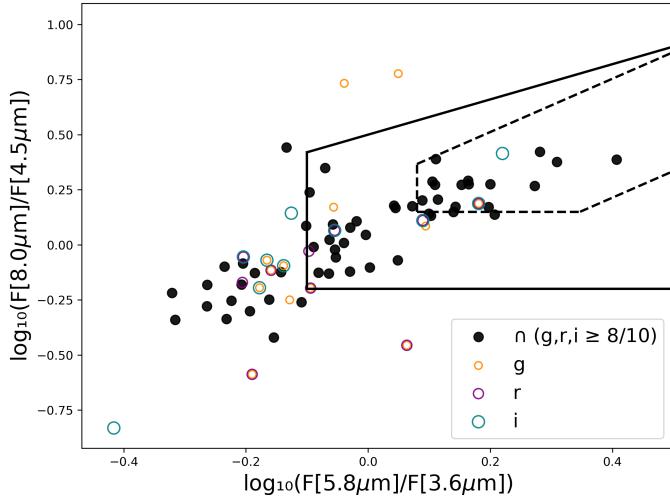


Fig. 1. Mid-infrared color-color diagram for our sample of 96 obscured AGN. The solid and dashed lines define the region where AGN are typically found after Lacy et al. (2007) and Donley et al. (2012), respectively. Filled black dots indicate sources that were classified as AGN in all three bands with an 8/10 threshold, while empty circles indicate sources classified as AGN in one or two bands only (using the same 8/10 threshold in each band), according to the legend in the figure.

each band are used separately. This allows us to make full use of the entire dataset, exploiting all the available visits, without the need for any artificial adjustments. Conversely, De Cicco et al. (2025) used r , g , and also bivariate features together, which made it necessary to resort to data imputation for the light curves in the two bands so that all the features used together had the same weight. Another difference is that here we focus on an AGN LS that is ≈ 1.4 times larger than the previous one (520 versus 380 AGN) and also more diversified, better reflecting a realistic scenario.

Essentially, for each band we used the selected subset of features and repeated the hyperparameter optimization procedure described in Section 3. We then trained a classifier using only these features and performed ten independent experiments, as usual. It is important to note that each classifier was trained on the whole main LS and not only on obscured AGN and inactive galaxies. From each set of ten experiments, we derived a feature importance ranking and subsequently removed the least informative feature, repeating the entire process iteratively. Feature importance was evaluated as the average impurity decrease across all the trees where the feature was used and involved in a split. We assessed the performance of each classifier using the same metrics reported in Table 3, and halted the iteration when a further reduction in the number of features no longer yielded any performance gains. This procedure, applied independently to the g , r , and i bands, revealed that the best-performing classifiers rely on 13, 14, and 10 features in the g , r , and i bands, respectively. We listed the selected features for each band in Table 5, while we reported the corresponding classifier performance metrics in Section V of Table 3.

Table 5 shows that eight features consistently appear in the selected lists across the three bands: these are the variability features GP_DRW_ τ and η^e , the colors ch21, u-B, r-i, i-z, z-y (i.e., all the color features considered except for B-r), and the morphology indicator class_star. Moreover, four additional variability features – ExcessVar, IAR $_\phi$, Autocor_length, R $_{cs}$ – are present in the selected feature sets for two out of

the three bands. These results suggest that, although the exact feature set always depends on the specific band and on the LS used, certain features consistently play a central role in the classification process. Interestingly, we note that several variability features reported in Table 5 (e.g., GP_DRW_ τ , IAR $_\phi$, Autocor_length, R $_{cs}$) are related to the timescale of the variations and are sensitive to the number of visits, and hence to the light curve sampling.

In De Cicco et al. (2025) we applied the same optimization procedure starting from 25 features (see their Fig. 3) and identified a final subset of eight features (see their Fig. 4) that proved most effective in separating obscured AGN and inactive galaxies. As was mentioned above, although the initial set of 25 features included both r - and g -band features, the final subset of eight contained only r -band features, together with five colors and the morphology indicator. For comparison, all the 14 features selected as most effective in the present work are included in the initial selection of 25 features used in De Cicco et al. (2025), and all eight of their final features are also found among our final subset of 14, with half of them also occupying the same rank position. Despite such a consistency, the fact that we had started from a different feature sample and had trained the algorithm on a different LS naturally led to a different set of features in the end.

Section V of Table 3 indicates that, in most cases, the results obtained across the three bands via the corresponding subsamples of features are mutually consistent. We note that the performance of the i band is generally slightly better than the other two, and this is interesting especially if we focus on the results obtained for the subsamples of obscured AGN and MIR AGN (where the latter potentially include a significant fraction of obscured AGN): indeed, this supports the well-established notion that redder bands are more effective in the identification of obscured sources. However, the recall values across the various subclasses of sources do not differ dramatically from each other.

A comparison with Section II of the table – where the same LS was used, but with the entire initial set of features – reveals that reducing the number of features generally leads to an improved performance in each band. Nonetheless, we consistently notice slightly lower values for TNR and precision across all three bands. We stress that, given the size and composition of our LS, the average $<1\%$ decrease in TNR observed translates into <20 additional misclassified sources. An increase in contamination is typically the cost of achieving a higher completeness, which is indeed reflected in our results, as apparent from the various recall values in the table. In particular, in this case the $<1\%$ higher contamination is the price to pay to obtain a recall for obscured AGN increased by $\approx 8\%$ in all three bands. We note that the most complete and most contaminated performance is associated with the i -band classifier. This could, at least in part, result from the less efficient cadence in this band, which is likely to weaken the variability features (see, e.g., Fig. 10 of De Cicco et al. 2019). We therefore expect this contamination to be reduced when using LSST data. We stress how, in this set of experiments focused on obscured AGN, a more diverse and realistic LS led to a generally worse performance but substantially improved the recall of obscured AGN.

In line with the approach adopted earlier in this work, we set a probability threshold of eight out of ten experiments as our reference, which yields single-band samples of obscured AGN larger than any other experiment here performed so far. Specifically, we were able to retrieve 74 (g band), 72 (r band), and 81 (i band) out of the 96 known obscured AGN in our LS, which

Table 5. List of features yielding the best-performing classifier in each band.

	<i>g</i>	<i>r</i>	<i>i</i>
1	ch21	ch21	ch21
2	u-B	u-B	u-B
3	class_star	class_star	class_star
4	r-i	GP_DRW_τ	B-r
5	GP_DRW_τ	r-i	r-i
6	η ^e	ExcessVar	i-z
7	IAR _φ	P _{var}	GP_DRW_τ
8	i-z	i-z	IAR _φ
9	R _{cs}	R _{cs}	z-y
10	z-y	GP_DRW_σ	η ^e
11	ExcessVar	z-y	–
12	Meanvariance	Period_fit	–
13	Autocor_length	η ^e	–
14	–	Autocor_length	–

Notes. The classifiers are evaluated using the standard performance metrics adopted in this work, with particular emphasis on the selection of obscured AGN. Within each band, features are ranked in order of decreasing importance.

translate into 77⁺¹⁰₋₉%, 75⁺¹⁰₋₉%, and 84⁺¹⁰₋₉%, respectively. We also combined the results of the ten experiments performed for each band with the corresponding selected set of features. These results are presented in the bottom section of Table 4, including the intersection and union of the retrieved samples as well as the number of sources confirmed in two out of three bands. We can see that a sample of 66 out of 96 (69⁺¹⁰₋₈%) obscured AGN are correctly classified in all three bands, while 77 AGN (80⁺⁹₋₇%) are correctly classified in two out of three bands.

In Fig. 2 we present two color-color diagrams, namely $B - r$ versus $u - B$ and $r - i$ versus $B - r$, where the sources of the main LS are shown. The left diagram suggests that the obscured AGN that we consistently failed to identify in the three bands are redder than most of the other obscured AGN in the LS, while the right diagram shows that these sources occupy the lower galaxy sequence, having redder $B - r$ but bluer $r - i$ colors. Nevertheless, based on these features alone they do not seem to constitute a separate population with respect to the “other AGN” subsample in our LS, that is to say all the AGN in the LS that are not labeled as obscured; hence these diagrams alone are not sufficient to disentangle them. We did not include other color-color diagrams based on other filters as they do not add relevant information.

Similarly to colors, the variability features alone generally fail to disentangle the obscured AGN that we are not able to confirm in any of the three bands from inactive galaxies. This can be clearly seen in Fig. 3, where we show the distribution of the feature `Autocor_length`, computed from the r -band light curves, for the AGN and non-AGN in our LS, distinguishing the various classes of sources. It is indeed apparent that the distribution for obscured AGN identified in all three bands largely overlaps the distribution for all the other AGN, reaching higher values for the feature at issue compared to non-AGN. Things change when a source is not classified as an AGN in the r band or is an obscured AGN that our classifier failed to identify in any of the bands, as the corresponding `Autocor_length` values are the lowest possible. While this feature is the least important in the ranking obtained when optimizing selection in the r band (see

Table 5), it provides a very effective visualization of the limitations of optical variability in finding some obscured AGN.

6. Summary and conclusions

This work combined the strength of multiband optical variability and color-based selection to identify AGN and, in particular, to mitigate the well-known challenges in detecting obscured sources. The analysis was based on the g -, r -, and i -band light curves extracted from the VST-COSMOS dataset, and focused on the independent identification of AGN in each band. A RF classifier was trained on an initial set of 36 features, of which 29 derived from optical variability, complemented by six optical/infrared colors and a morphology indicator. Since our goal is to provide meaningful insights and practical guidance for future large-scale surveys, the selected features were chosen to reflect those that will be readily accessible for the DDFs from the LSST dataset or from ancillary datasets (for example, the SWIRE/SERVS/DeepDrill surveys will provide magnitudes in the filters necessary to obtain our `ch21` color feature; see, e.g., Lonsdale et al. 2003; Mauduit et al. 2012).

We selected AGN candidates in each band independently, making use of a RF algorithm and testing two different LSs, the difference between the two lying in the AGN populations they include. The so-called spec-MIR LS is composed of spectroscopically confirmed AGN selected from the *Chandra*-COSMOS Legacy Catalog, which is made up of X-ray-detected sources. The objects we selected from this catalog therefore have X-ray information and a spectroscopic classification as Type I or II. The spec-MIR LS also includes MIR-confirmed AGN selected on the basis of the criterion by Donley et al. (2012). The main LS further includes sources classified as AGN in the *Chandra*-COSMOS Legacy Catalog based solely on their X-ray properties, but with no spectroscopic classification available. Our comparison of the performance of the various classifiers (see Sections I and II of Table 3) shows that the main LS, being more diversified than the spec-MIR LS, leads to an overall decline in performance, with higher contamination rates. Yet, the recall for obscured AGN increases significantly – by $\approx 20\%$ across all bands – when using the main LS, reflecting the advantage of a more diverse LS in capturing a wider variety of AGN types. Based on the adopted threshold, we are able to identify 58⁺⁹₋₈% of the known obscured AGN. Although this fraction is lower than the (68.1 ± 1.2)% reported in De Cicco et al. (2025), we note that it is more robust, as the selected sources are consistently classified as AGN in all three bands. We also note (see Fig. 1) that these sources would not be classified as AGN based solely on their MIR colors, and this highlights the effectiveness of variability as a complementary selection method to color-based criteria.

We demonstrated that integrating optical variability with color features enhances classifier performance relative to using either set of features separately. Specifically, the classifiers using only optical variability features returned high purity but low completeness, while those using only color features exhibit the opposite trend, achieving higher completeness at the expense of purity.

In the final stage of this study we turned our attention on optimizing the selection of obscured AGN. For each band, we determined the most discriminative features relative to inactive galaxies. We used these feature sets to build a series of classifiers for each band in a recursive way, removing the least informative feature at each step, and stopping when no further gains in the performance were observed. This process yielded a set of

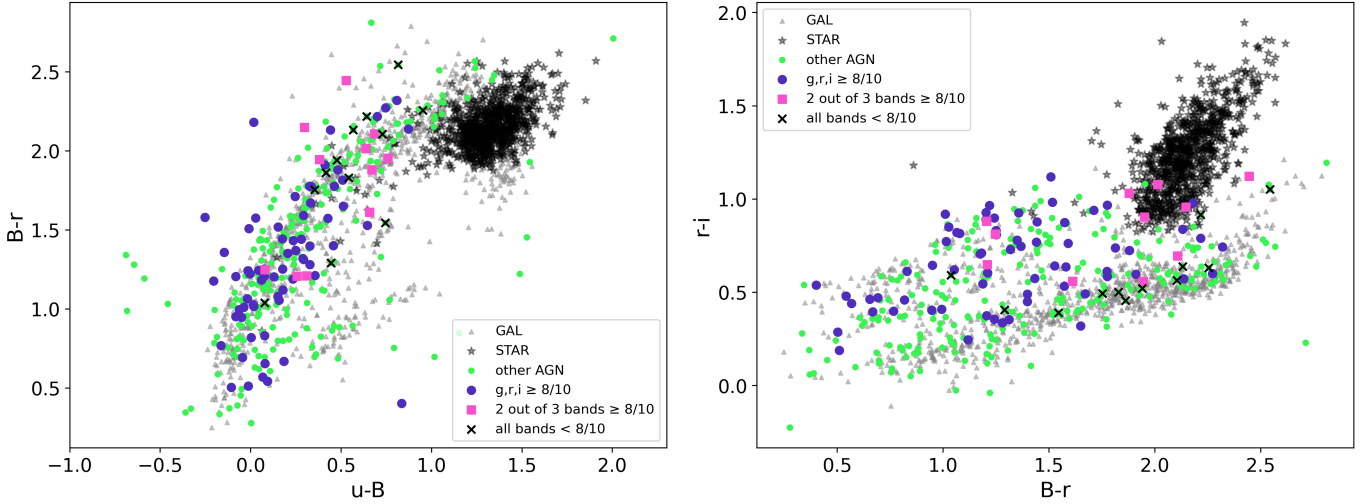


Fig. 2. Color-color diagrams where the obscured AGN that were not identified in any of the three bands (black crosses) tend to place themselves on specific loci. In both panels, the gray triangles represent inactive galaxies from the LS, while black stars represent the stars in the LS. The large violet dots indicate the obscured AGN confirmed in all three bands by at least eight out of ten experiments, while the bright pink squares indicate the ones confirmed in two out of the three bands based on the same threshold. The small green dots stand for all the AGN in the LS that are not labeled as obscured.

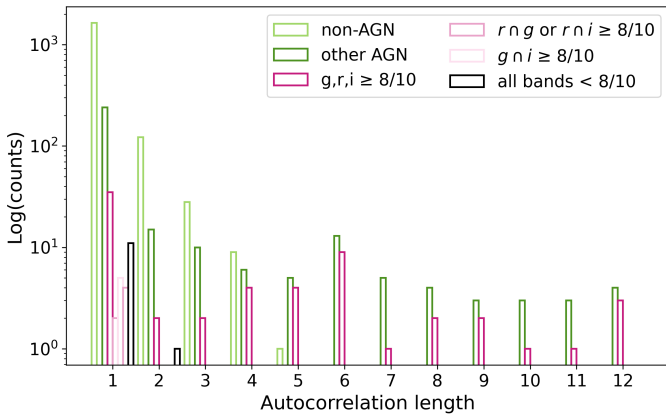


Fig. 3. Distribution of the values obtained from the r -band light curves for the `Autocor_length` feature. The AGN in the LS are split into obscured AGN confirmed in all three bands (deep magenta), in the r band plus either g or i (light magenta), or in the g and i bands, but not in the r (light pink). Dark green indicates all the other AGN, meaning all the AGN in the LS that are not labeled as obscured, while the non-AGN are shown in light green. Black represents the obscured AGN that each classifier consistently failed to identify (that is to say, they were classified as AGN in $<8/10$ experiments per band).

13, 14, and 10 features for the g , r , and i bands, respectively (see Table 5). From these results, we derived a sample of AGN candidates in each band, achieving single-band recall values for obscured AGN that exceed those obtained here and in earlier studies in this series. We subsequently combined the three samples to obtain a more robust final sample. We decided to classify a source as an AGN if it was classified as such in at least eight out of the ten experiments run for each classifier (see Table 4). This strategy resulted in a recall of $69^{+10}_{-8}\%$ for the obscured AGN identified in all three bands. While this is numerically consistent with the findings of De Cicco et al. (2025), in this work we extended the result to a more comprehensive and diversified sample of AGN classified on the basis of spectra, MIR colors, and X-ray emission. If we adopt a looser criterion, demanding

AGN to be confirmed in at least two out of three bands, we obtain an $80^{+10}_{-9}\%$ recall. The obscured AGN samples obtained combining two or three bands should be regarded as purer and more reliable than those derived from single bands. The variability of AGN across different optical bands is known to be correlated, reflecting the physical connection among emission regions located at different distances from the central SMBH. One would therefore expect that a source varying in one optical band also varies in the others, so that optical multiband variability detection alone would not necessarily provide additional information, except reducing contamination as it would remove spurious variability. In our case, however, the classification does not rely on variability alone, but on its combination with color-based properties. Since each classifier uses both variability and color features, their correlations are not trivial. The selection obtained in that specific band, combined with colors, can identify obscured AGN. As a consequence, detections in more than one band provide complementary information, leading to a more reliable and robust sample of obscured AGN. Overall, this study proves that a careful selection of features tailored to the identification of obscured AGN, combined with a multiband approach, results in a markedly enhanced recovery rate of known sources. A crucial aspect of this result is that it was achieved by training the classifier via a more diversified LS than in our previous studies. As such, it is more representative of the true AGN population, including sources with a broader range of properties than before.

In the future we plan to explore the sources that were excluded from the final sample, in order to verify whether we excluded potentially interesting obscured targets. Indeed, these sources may display variability properties distinct not only from unobscured AGN but also from the selected sample, and these properties could, in principle, vary from band to band.

Testing this method on the other DDFs, beginning with VST data – and ultimately extending the analysis to the LSST DDF data, and hence to more bands – will validate the robustness of the selection technique. An additional key step would be the spectroscopic follow-up of our obscured AGN sample, which

could offer valuable insights into both the nature and the physical state of these sources.

Acknowledgements. DD acknowledges PON R&I 2021, CUP E65F21002880003, and Fondi di Ricerca di Ateneo (FRA), linea C, progetto TORNADO. DD, MP and VP also acknowledge the financial contribution from PRIN-MIUR 2022 and from the Timedomes grant within the “INAF 2023 Finanziamento della Ricerca Fondamentale”.

References

- Allevato, V., Paolillo, M., Papadakis, I., & Pinto, C. 2013, *ApJ*, 771, 9
- Antonucci, R. 1993, *ARA&A*, 31, 473
- Bernal, S., Sánchez-Sáez, P., Arévalo, P., et al. 2025, *A&A*, 694, A127
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Botticella, M. T., Cappellaro, E., Greggio, L., et al. 2017, *A&A*, 598, A50
- Breiman, L. 2001, *Mach. Learn.*, 45, 5
- Cabrera-Vives, G., Förster, F., et al. 2020, *Astron. Comput.*, 30, 100334
- Capaccioli, M., & Schipani, P. 2011, *Messenger*, 146, 2
- Cappellaro, E., Botticella, M. T., Pignata, G., et al. 2015, *A&A*, 584, A62
- Cavuoti, S., De Cicco, D., Doorenbos, L., et al. 2024, *A&A*, 687, A246
- Civano, F., Elvis, M., Brusa, M., et al. 2012, *ApJS*, 201, 30
- De Cicco, D., Paolillo, M., Covone, G., et al. 2015, *A&A*, 574, A112
- De Cicco, D., Paolillo, M., Falocco, S., et al. 2019, *A&A*, 627, A33
- De Cicco, D., Bauer, F. E., Paolillo, M., et al. 2021, *A&A*, 645, A103
- De Cicco, D., Bauer, F. E., Paolillo, M., et al. 2022, *A&A*, 664, A117
- De Cicco, D., Zazzaro, G., Cavuoti, S., et al. 2025, *A&A*, 697, A204
- Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, *ApJ*, 748, 142
- Eyheramendy, S., Elorrieta, F., & Palma, W. 2018, *MNRAS*, 481, 4311
- Falocco, S., Paolillo, M., Covone, G., et al. 2015, *A&A*, 579, A115
- Fu, L., Liu, D., Radovich, M., et al. 2018, *MNRAS*, 479, 3858
- Gehrels, N., Spergel, D., & WFIRST SDT Project 2015, *J. Phys. Conf. Ser.*, 610, 012007
- Graham, M. J., Djorgovski, S. G., Drake, A. J., et al. 2017, *MNRAS*, 470, 4112
- Hornschemeier, A. E., Brandt, W. N., Garmire, G. P., et al. 2001, *ApJ*, 554, 742
- Huang, Z., Radovich, M., Grado, A., et al. 2011, *A&A*, 529, A93
- Huijse, P., Estévez, P. A., Förster, F., et al. 2018, *ApJS*, 236, 12
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- Kaviraj, S., De Cicco, D., Lazar, I., et al. 2026, *MNRAS*, 545, staf1918
- Kim, D.-W., et al. 2016, *MNRAS*, 456, 552
- Kim, D.-W., Protopapas, P., Byun, Y.-I., et al. 2011, *ApJ*, 735, 68
- Kim, D.-W., Protopapas, P., Bailer-Jones, C. A. L., et al. 2014, *A&A*, 566, A43
- Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196
- Lacy, M., Petric, A. O., Sajina, A., et al. 2007, *AJ*, 133, 186
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
- Lira, P., Sánchez-Sáez, P., Arévalo, P., et al. 2024, *MNRAS*, 531, 3310
- Liu, D., Fu, L., Liu, X., et al. 2018, *MNRAS*, 478, 2388
- Liu, D., Deng, W., Fan, Z., et al. 2020, *MNRAS*, 493, 3825
- Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al. 2003, *PASP*, 115, 897
- López-Navas, E., Martínez-Aldama, M. L., Bernal, S., et al. 2022, *MNRAS*, 513, L57
- López-Navas, E., Arévalo, P., Bernal, S., et al. 2023, *MNRAS*, 518, 1531
- Maccacaro, T., Gioia, I. M., Wolter, A., Zamorani, G., & Stocke, J. T. 1988, *ApJ*, 326, 680
- Marchesi, S., Civano, F., Elvis, M., et al. 2016, *ApJ*, 817, 34
- Maruccia, Y., De Cicco, D., Cavuoti, S., et al. 2025, *A&A*, 699, A303
- Mauduit, J. C., Lacy, M., Farrah, D., et al. 2012, *PASP*, 124, 714
- McLaughlin, M. A., Mattox, J. R., Cordes, J. M., & Thompson, D. J. 1996, *ApJ*, 473, 763
- Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, *ApJ*, 476, 70
- Nun, I., Protopapas, P., Sim, B., et al. 2015, ArXiv e-prints [arXiv:1506.00010]
- Paolillo, M., & Papadakis, I. 2025, *Riv. Nuovo Cim.*, 48, 537
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825
- Ponti, G., Papadakis, I., Bianchi, S., et al. 2012, *A&A*, 542, A83
- Poulain, M., Paolillo, M., De Cicco, D., et al. 2020, *A&A*, 634, A50
- Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, *ApJ*, 733, 10
- Sammur, C., & Webb, G. I. 2010, *Leave-One-Out Cross-Validation* (Springer, US: Boston, MA), 600
- Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021, *AJ*, 161, 141
- Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010, *ApJ*, 714, 1194
- Scoville, N., Abraham, R. G., Aussel, H., et al. 2007a, *ApJS*, 172, 38
- Scoville, N., Aussel, H., Brusa, M., et al. 2007b, *ApJS*, 172, 1
- Urry, C. M., & Padovani, P. 1995, *PASP*, 107, 803
- Vaccari, M., Covone, G., Radovich, M., et al. 2016, *The 4th Annual Conference on High Energy Astrophysics in Southern Africa (HEASA 2016)*, 26
- Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, *ApJS*, 195, 10

Appendix A: Dataset

Here we report detailed information about the three-band datasets used throughout this work. We refer the reader to Section 2 of De Cicco et al. (2015) and Section 2 of De Cicco et al. (2019) for an overview of the data structure and a description of the reduction process.

Table A.1. VST-COSMOS *gri* dataset.

ID	obs. date	time (days)	<i>r</i> band seeing	<i>g</i> band seeing	<i>i</i> band seeing	ID	obs. date	time (days)	<i>r</i> band seeing	<i>g</i> band seeing	<i>i</i> band seeing
1	2011-Dec-18	0	0.64	-	-	37	2014-Feb-19	794	0.89	0.91	-
2	2011-Dec-22	4	0.94	-	-	38	2014-Feb-21	796	0.93	-	0.83
3	2011-Dec-27	9	1.04	1.13	0.93	39	2014-Feb-23	798	0.81	-	-
4	2011-Dec-31	13	1.15	-	-	40	2014-Feb-26	801	0.81	0.82	-
5	2012-Jan-02	15	0.67	-	-	41	2014-Feb-28	803	0.77	-	0.91
6	2012-Jan-06	19	0.58	-	-	42	2014-Mar-04	807	-	1.24	-
7	2012-Jan-18	31	0.62	-	-	43	2014-Mar-08	811	0.91	-	0.89
8	2012-Jan-20	33	0.88	-	-	44	2014-Mar-21	824	0.96	1.04	-
9	2012-Jan-22	35	0.81	1.11	0.95	45	2014-Mar-23	826	0.92	-	0.98
10	2012-Jan-24	37	0.67	-	-	46	2014-Mar-25	828	0.66	-	-
11	2012-Jan-27	40	0.98	-	-	47	2014-Mar-29	832	0.89	0.86	-
12	2012-Jan-29	42	0.86	-	-	48	2014-Apr-04	838	0.58	0.72	-
13	2012-Feb-02	46	0.86	0.88	1.05	49	2014-Apr-07	841	0.61	-	-
14	2012-Feb-16	60	0.50	0.61	0.60	50	2014-Dec-03	1081	1.00	-	-
15	2012-Feb-19	63	0.99	-	-	50b	2014-Dec-05	1083	-	-	1.42
16	2012-Feb-21	65	0.79	-	-	50c	2014-Dec-15	1093	-	-	0.69
17	2012-Feb-23	67	0.73	-	-	51	2014-Dec-16	1094	-	0.80	-
18	2012-Feb-26	70	0.83	1.04	0.88	52	2014-Dec-25	1103	-	0.80	-
19	2012-Feb-29	73	0.90	-	-	53	2015-Jan-03	1112	-	1.09	-
20	2012-Mar-03	76	0.97	-	-	54	2015-Jan-10	1119	0.71	-	-
21	2012-Mar-13	86	0.70	-	-	55	2015-Jan-15	1124	-	0.88	-
22	2012-Mar-15	88	1.08	-	-	55b	2015-Jan-21	1130	-	-	0.78
23	2012-Mar-17	90	0.91	1.04	0.53	56	2015-Jan-23	1132	-	1.01	-
24	2012-May-08	142	0.74	-	-	57	2015-Jan-28	1137	0.90	-	-
25	2012-May-09	143	-	0.78	0.62	57b	2015-Jan-29	1138	-	-	0.84
26	2012-May-11	145	0.85	-	-	58	2015-Jan-30	1139	-	0.96	-
27	2012-May-17	151	0.77	-	-	59	2015-Jan-31	1140	0.73	-	-
28	2013-Dec-25	728	-	0.93	-	59b	2015-Feb-13	1153	-	-	0.91
29	2013-Dec-27	740	0.72	-	0.70	60	2015-Feb-14	1154	-	1.00	-
30	2013-Dec-30	743	1.00	-	-	61	2015-Feb-15	1155	0.70	-	-
31	2014-Jan-03	747	0.86	0.96	-	62	2015-Mar-10	1178	0.80	-	-
32	2014-Jan-05	749	0.81	-	0.77	63	2015-Mar-13	1181	-	1.15	-
33	2014-Jan-12	756	0.73	-	0.73	64	2015-Mar-14	1182	0.84	-	-
34	2014-Jan-21	765	1.18	1.13	-	65	2015-Mar-19	1187	1.00	-	-
35	2014-Jan-24	768	0.80	-	0.73	65b	2015-Mar-20	1188	-	-	1.00
35b	2014-Feb-02	777	-	-	0.80	66	2015-Mar-22	1190	-	1.29	-
36	2014-Feb-09	784	1.28	-	0.98						

Notes. The table reports the visit ID, date, time in days from the first observation, presence (via the corresponding seeing value) or absence (-) of a visit in the *r*, *g*, and *i* bands. The three values in bold identify the selected reference images for each band.

Appendix B: Best hyperparameters obtained per classifier

Our fine-tuning process focused on five hyperparameters, shortly described in what follows. $n_estimators$ sets the number of decision trees making up the forest, aiming to balance classification accuracy and computational time. $min_samples_split$ defines the minimum number of objects required to split a node. max_depth sets a limit on tree depth, aiming to prevent underfitting or overfitting. $min_samples_leaf$ sets the minimum number of objects for a node to survive, helping control tree size. $max_features$ determines the number of features to take into account when splitting a node.

Table B.1. Set of best values obtained from a grid search-based optimization of the five selected hyperparameters, for each RF classifier tested in this work.

RF classifier	$n_estimators$	$min_samples_split$	min_sample_leaf	max_depth	$max_features$
g , spec-MIR LS	300	10	4	10	sqrt
r , spec-MIR LS	100	2	4	10	sqrt
i , spec-MIR LS	500	10	4	10	sqrt
g , main LS	300	10	4	10	sqrt
r , main LS	300	10	4	10	sqrt
i , main LS	100	10	4	10	sqrt
g , var. features only	300	2	4	10	log2
r , var. features only	500	10	2	30	sqrt
i , var. features only	100	10	2	10	log2
$uBrizy$ -derived colors only	300	10	4	10	sqrt
all colors only	100	10	2	20	sqrt
g , 13 features	100	10	4	20	sqrt
r , 14 features	100	10	4	20	sqrt
i , 10 features	500	2	4	10	sqrt

Notes. The chosen hyperparameters to optimize (see Section 3) are: $n_estimators$ (tested values: 100, 300, 500), $min_samples_split$ (tested values: 2, 5, 10), $min_samples_leaf$ (tested values: 1, 2, 4), max_depth (tested values: 10, 20, 30, None), and $max_features$ (tested options: *sqrt* and *log2*). The first column lists the classifiers evaluated in this work, with their performance metrics shown in Table 3.