

# J-PLUS: Bayesian object classification with a strum of BANNJOS

A. del Pino<sup>1</sup>, C. López-Sanjuan<sup>1</sup>, A. Hernán-Caballero<sup>1</sup>, H. Domínguez-Sánchez<sup>1</sup>, R. von Marttens<sup>2,3</sup>, J. A. Fernández-Ontiveros<sup>1</sup>, P. R. T. Coelho<sup>4</sup>, A. Lumbreras-Calle<sup>1</sup>, J. Vega-Ferrero<sup>1</sup>, F. Jimenez-Esteban<sup>5</sup>, P. Cruz<sup>5</sup>, V. Marra<sup>6,7,8</sup>, M. Quartín<sup>9,10,3</sup>, C. A. Galarza<sup>11</sup>, R. E. Angulo<sup>12,13</sup>, A. J. Cenarro<sup>1</sup>, D. Cristóbal-Hornillos<sup>18</sup>, R. A. Dupke<sup>11,14,15</sup>, A. Ederoclite<sup>1</sup>, C. Hernández-Monteagudo<sup>16,17</sup>, A. Marín-Franch<sup>1</sup>, M. Moles<sup>18</sup>, L. Sodr e Jr.<sup>4</sup>, J. Varela<sup>18</sup>, and H. Vázquez Ramio<sup>1</sup>

- <sup>1</sup> Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Unidad Asociada al CSIC, Plaza San Juan 1, 44001 Teruel, Spain  
e-mail: andresdelpinomolina@gmail.com
- <sup>2</sup> Instituto de Física, Universidade Federal da Bahia, 40170-155, Salvador, BA, Brazil
- <sup>3</sup> PPGCosmo, Universidade Federal do Espírito Santo, 29075-910, Vitória, ES, Brazil
- <sup>4</sup> Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, 05508-090 São Paulo, Brazil
- <sup>5</sup> Centro de Astrobiología, CSIC-INTA, Camino bajo del castillo s/n, E-28692, Villanueva de la Cañada, Madrid, Spain
- <sup>6</sup> Departamento de Física, Universidade Federal do Espírito Santo, 29075-910, Vitória, ES, Brazil
- <sup>7</sup> INAF, Osservatorio Astronomico di Trieste, via Tiepolo 11, 34131 Trieste, Italy
- <sup>8</sup> IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151, Trieste, Italy
- <sup>9</sup> Instituto de Física, Universidade Federal do Rio de Janeiro, 21941-972, Rio de Janeiro, RJ, Brazil
- <sup>10</sup> Observatório do Valongo, Universidade Federal do Rio de Janeiro, 20080-090, Rio de Janeiro, RJ, Brazil
- <sup>11</sup> Observatório Nacional - MCTI (ON), Rua Gal. José Cristino 77, São Cristóvão, 20921-400 Rio de Janeiro, Brazil
- <sup>12</sup> Donostia International Physics Centre (DIPC), Paseo Manuel de Lardizabal 4, 20018 Donostia-San Sebastián, Spain
- <sup>13</sup> IKERBASQUE, Basque Foundation for Science, 48013, Bilbao, Spain
- <sup>14</sup> University of Michigan, Department of Astronomy, 1085 South University Ave., Ann Arbor, MI 48109, USA
- <sup>15</sup> University of Alabama, Department of Physics and Astronomy, Gallalee Hall, Tuscaloosa, AL 35401, USA
- <sup>16</sup> Instituto de Astrofísica de Canarias, La Laguna, 38205, Tenerife, Spain
- <sup>17</sup> Departamento de Astrofísica, Universidad de La Laguna, 38206, Tenerife, Spain
- <sup>18</sup> Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan 1, 44001 Teruel, Spain

Received XX XX 2024; accepted XX XX XX

## ABSTRACT

**Context.** With its 12 optical filters, the Javalambre-Photometric Local Universe Survey (J-PLUS) provides an unprecedented multicolor view of the local Universe. The third data release (DR3) covers  $3,192 \text{ deg}^2$  and contains 47.4 million objects. However, the classification algorithms currently implemented in the J-PLUS pipeline are deterministic and based solely on the sources morphology.

**Aims.** Our goal is to classify the sources identified in the J-PLUS DR3 images into stars, quasi-stellar objects (QSOs), and galaxies. For this task, we present BANNJOS, a machine learning pipeline that utilizes Bayesian neural networks to provide the full probability distribution function (PDF) of the classification.

**Methods.** BANNJOS is trained on photometric, astrometric, and morphological data from J-PLUS DR3, *Gaia* DR3, and CatWISE2020, using over 1.2 million objects with spectroscopic classification from SDSS DR18, LAMOST DR9, DESI Early Data Release, and *Gaia* DR3. Results are validated on a test set of about  $1.4 \times 10^5$  objects and cross-checked against theoretical model predictions.

**Results.** BANNJOS outperforms all previous classifiers in terms of accuracy, precision, and completeness across the entire magnitude range. It delivers over 95% accuracy for objects brighter than  $r = 21.5 \text{ mag}$ , and  $\sim 90\%$  accuracy for those up to  $r = 22 \text{ mag}$ , where J-PLUS completeness is  $\lesssim 25\%$ . BANNJOS is also the first object classifier to provide the full probability distribution function (PDF) of the classification, enabling precise object selection for high purity or completeness, and for identifying objects with complex features, like active galactic nuclei with resolved host galaxies.

**Conclusions.** BANNJOS has effectively classified J-PLUS sources into around 20 million galaxies, 1 million QSOs, and 26 million stars, with full PDFs for each, which allow for later refinement of the sample. The upcoming J-PAS survey, with its 56 color bands, will further enhance BANNJOS's ability to detail each source's nature.

**Key words.** methods: data analysis – machine-learning, techniques: photometric, surveys

## 1. Introduction

Public photometric large digital sky surveys are revolutionizing our view of the Universe. Covering large areas of the sky ( $\gtrsim 5000 \text{ deg}^2$ ), surveys such as POSS-II (Second Palomar Observatory Sky Survey, three optical *gri* broad bands; Gal et al. 2004), SDSS (Sloan Digital Sky Survey, five optical *ugriz* broad bands; Abazajian et al. 2009), and VHS (VISTA Hemisphere

Survey, three near-infrared *HJK<sub>s</sub>* bands; McMahon et al. 2013) have provided crucial information about the large-scale structures that dominate the Universe. This allows astronomers to better understand the nature of celestial objects and to study a wide range of phenomena, from star formation to the evolution of galaxy groups. The next generation of surveys, such as DES (Dark Energy Survey, five optical *ugriz* broad bands; Flaugher 2012), UHS (UKIRT Hemisphere Survey, two near-infrared *JK<sub>s</sub>*

**Table 1.** J-PLUS photometric system and limiting magnitudes.

Filter ( $\mathcal{X}$ )	Central wavelength [nm]	FWHM [nm]	$m_{\text{lim}}^{\text{DR3}}$ [mag] <sup>a</sup>
<i>u</i>	348.5	50.8	20.8
J0378	378.5	16.8	20.8
J0395	395.0	10.0	20.8
J0410	410.0	20.0	21.0
J0430	430.0	20.0	21.0
<i>g</i>	480.3	140.9	21.8
J0515	515.0	20.0	21.0
<i>r</i>	625.4	138.8	21.8
J0660	660.0	13.8	21.0
<i>i</i>	766.8	153.5	21.3
J0861	861.0	40.0	20.4
<i>z</i>	911.4	140.9	20.5

**Notes.** <sup>(a)</sup> Limiting magnitude ( $5\sigma$ , 3 arcsec diameter aperture) of J-PLUS DR3 (López-Sanjuan et al. 2023).

bands; Dye et al. 2018), *Euclid* (three near-infrared *YJH* broad bands; Laureijs et al. 2011), and LSST (Large Synoptic Survey Telescope, six optical *ugrizY* broad bands; Ivezić et al. 2019), will push the current sensitivity limits and open the door to new fields such as time-domain astronomy.

Among the recent additions to the family of public photometric surveys is the Javalambre Photometric Local Universe Survey<sup>1</sup> (J-PLUS; Cenarro et al. 2019), carried out at the Observatorio Astrofísico de Javalambre (OAJ, Teruel, Spain; Cenarro et al. 2014) using the 83 cm Javalambre Auxiliary Survey Telescope (JAST80) and T80Cam, a panoramic camera of 9.2k × 9.2k pixels that provides a 2 deg<sup>2</sup> field of view (FoV) with a pixel scale of 0.55 arcsec pix<sup>-1</sup> (Marín-Franch et al. 2015). J-PLUS aims to cover 8500 deg<sup>2</sup> of the northern sky hemisphere with an unprecedented set of 12 filters: five broad *ugriz* bands plus seven narrow optical bands (refer to Table 1). The vast dataset produced by J-PLUS has broad astrophysical applications, enhancing our understanding of the Universe. The third data release (DR3) spans 3192 deg<sup>2</sup> (2881 deg<sup>2</sup> after masking) and catalogs 47.4 million objects with improved photometric calibration (López-Sanjuan et al. 2023), offering an unprecedented multicolor view of the local Universe.<sup>2</sup>

Due to its photometric, flux-limited nature, J-PLUS images all astronomical sources down to its limiting magnitude without pre-selection. Identifying and classifying the observed objects within its footprint, such as stars and galaxies, is crucial. As with any photometric survey, this is one of the first steps in creating science-ready data products from J-PLUS data. Reliable object classification enables the study of specific astronomical sources and aids in the discovery of uncommon or new types of objects. Classification algorithms typically employ two complementary approaches: color-based and morphological. Color classifiers leverage the distinct positions of stars, galaxies, and quasi-stellar objects (QSOs) in color-color diagrams (e.g., Huang et al. 1997; Elston et al. 2006; Baldry et al. 2010; Saglia et al. 2012; Małek et al. 2013), while morphological classifiers distinguish between point-like and extended sources based on isophotal concentration (e.g., Kron 1980; Reid et al. 1996; Odewahn et al. 2004; Vasconcellos et al. 2011). J-PLUS utilizes the CLASS\_STAR morphological classifier from the SExtractor photometry pack-

age (Bertin & Arnouts 1996). However, this algorithm, which considers only object elongation, extension, and peak brightness, simplistically categorizes objects as "star" or "not star." Such a method is prone to misclassification, particularly when distinguishing between compact sources like stars, distant active galactic nuclei, or compact galaxies. Refined classification through manual inspection is possible but requires significant time and resources.

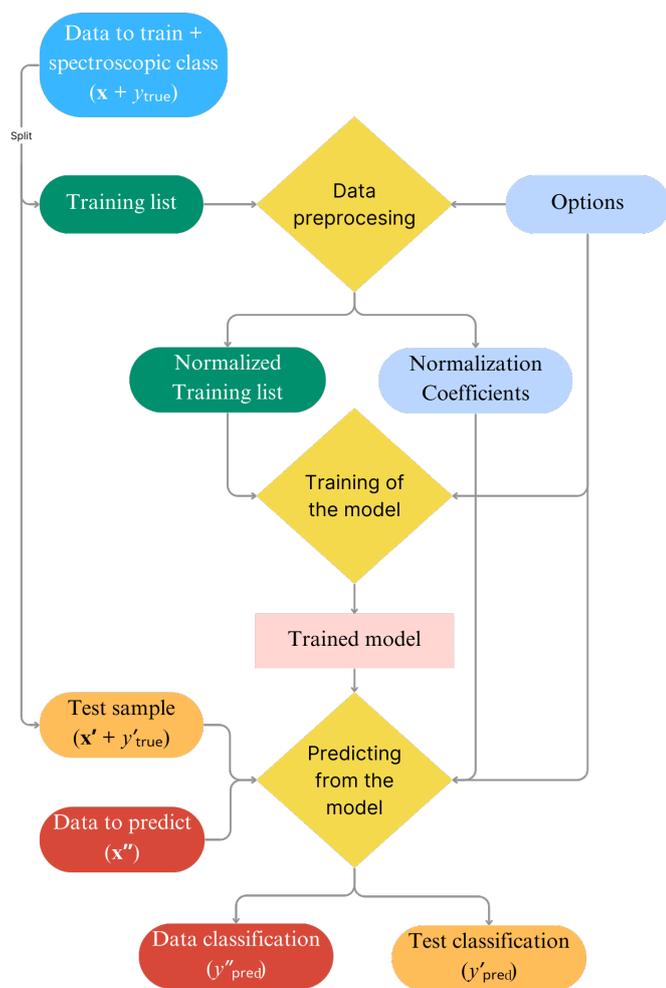
Some of these problems can be eased by incorporating prior information within a Bayesian framework (e.g., Sebk 1979; Scranton et al. 2002; Henrion et al. 2011; Molino et al. 2014). In López-Sanjuan et al. (2019), the authors introduced the `sglc_prob_star` classification, which imposes priors based on concentration, broad-band colors, object counts, and distance to the Galactic plane to achieve more reliable results, especially for sources with low signal-to-noise (S/N) ratios. However, this method does not utilize the valuable multi-filter color information available in J-PLUS, and its output remains bimodal, distinguishing only between "compact" and "extended" sources. To achieve a reliable classification based on an object's nature, not just its morphology, it is essential to utilize all photometric information in J-PLUS. Prior works have attempted this using various classification methods; for instance, Wang et al. (2022) employed the 12 photometric bands to include the QSO class. Yet, their method was limited to high S/N sources, classifying only about 3.5 million sources out of the ~13 million in J-PLUS's first data release.

To classify the entire catalog, methods capable of handling missing data are necessary. Techniques based on machine learning, successfully implemented in other surveys, offer classification into two (e.g., Ball et al. 2006; Miller et al. 2017) or three categories (e.g., Małek et al. 2013). In the particular case of J-PLUS, in von Martens et al. (2024) authors applied eXtreme Gradient Boosting (XGBoost, Chen & Guestrin 2016) to classify objects into stars, galaxies and QSOs, being the first classifier to approach a three-class classification in this survey. Unlike Bayesian approaches, these methods provide a deterministic classification for each object. In the best of the cases, classifiers would offer deterministic probabilities for an object's class membership (e.g., star, QSO, galaxy) without uncertainty or correlation between the classes probabilities. However, models and training data carry inherent uncertainties, and the distinctions between classes are sometimes not clear, as in the case of partially resolved active galaxies blurring the lines between QSOs and galaxies. Reliable classification across the entire catalog requires models that accommodate these uncertainties, leveraging morphological, photometric, and external information to provide confidence intervals and correlations in their predictions. The inclusion of uncertainty and degeneration among classes is also critical in order to control the purity and completeness of object selections as well as biases in samples containing objects of mixed characteristics.

In this paper, we introduce BANNJOS, a pipeline utilizing Bayesian Artificial Neural Networks (BANN) to classify J-PLUS objects into three categories: stars, QSOs, and galaxies. We evaluate the classification's quality and suggest methods to enhance the purity of the selected object samples. The paper is structured as follows: Section 2 outlines the BANNJOS pipeline; Section 3 discusses its application to J-PLUS, highlighting specific execution details; Section 4 presents classification results and model validation; Section 5 present examples of selection criteria; Section 6 compares our method with three existing classifications for J-PLUS; Section 7 present the

<sup>1</sup> www.j-plus.es

<sup>2</sup> www.j-plus.es/datareleases/data\_release\_dr3



**Fig. 1.** Flowchart illustrating the flow of processing of BANNJOS. Central components are represented by yellow rhombuses, while inputs and outputs are indicated by rounded rectangles. The procedure commences with the training data (depicted in blue, located at the top-left) and the options (illustrated in light blue, positioned at the top-right), which govern the behavior of BANNJOS. Throughout the chart, colors show the data type: light blue for options or variables controlling the process, green for training data, orange for the test sample, and red for the data on which predictions are intended to be made.

results for the entire J-PLUS catalog and extra statistical validation tests; Section 8 explain known caveats, and finally, Section 9 summarizes our findings and conclusions.

## 2. BANNJOS

The Bayesian Artificial Neural Networks for the Javalambre Observatory Surveys, or BANNJOS, is a publicly available<sup>3</sup> machine learning pipeline designed to derive any desired property of an astronomical object using supervised learning techniques. It is coded in Python using the Tensorflow and Tensorflow probability libraries (Abadi et al. 2015). BANNJOS works as a general-purpose regressor and is designed to operate at a variety of user input levels. It can run nearly

<sup>3</sup> <https://github.com/AndresdPM/BANNJOS>,  
<https://gitlab.cefca.es/adelpino/bannjos>

automatically once provided with the minimum input files, but allowing for extensive customization through optional keywords, giving users greater control over the process.

In essence, BANNJOS trains a model,  $f(\mathbf{x})$ , based on the relation between a dependent variable  $y_{\text{true}}$  and a set of independent variables,  $\mathbf{x}$ , using a training sample. It then predicts this variable,  $y'_{\text{pred}}$ , in any given  $\mathbf{x}'$ . Its most critical components include:

- Reading and preprocessing the data,
- Training the model,
- Computing  $y_{\text{pred}}$ .

A flowchart showcasing the working flow of BANNJOS can be found in Figure 1. In the following we explain the details of the preprocessing of the data and the different models available.

### 2.1. Preprocessing of the Data

The training sample consists of a table containing  $\mathbf{x}$  and  $y_{\text{true}}$ . The nature of these variable sets may vary; for instance,  $y_{\text{true}}$  can include categorical, continuous, or both types of information, whereas  $\mathbf{x}$  may encompass diverse types of data, such as photometric magnitudes or the positions of sources on the detector’s focal plane. Adapting all available information for the training process is a critical step with a notable impact on the final results. Initially, BANNJOS shuffles the training set and constructs  $\mathbf{x}$  and  $y_{\text{true}}$  based on user-defined options and criteria. The data are then divided into training and test sets, with the latter containing a subset of the training data that will not be exposed to the model during its training. This test sample, denoted as  $\mathbf{x}'$  and  $y'_{\text{true}}$  in Figure 1, will serve as a validation sample to assess the quality of the results.

Next, BANNJOS normalizes  $\mathbf{x}$ , an essential step prior to training a neural network. Outliers can artificially widen the dynamic range of the training sample, thereby compressing the range containing actual information. To mitigate this, normalization is performed individually to each variable using the 0.005 and 0.995 quantiles of  $\mathbf{x}$  as the minimum and maximum, respectively, and then clipping the normalized set to the range  $[0, 1]$ . Any missing data in  $\mathbf{x}$  are assigned a default value of  $-0.1$ . This approach allows BANNJOS to handle missing data within the training sample. Lastly, BANNJOS converts any categorical data into continuous numerical values, enabling it to treat classification problems as regression problems by assigning probabilities to each class.

### 2.2. Dealing with Uncertainties with Different Models

BANNJOS provides users the option to choose between six different regression models: three deterministic and three probabilistic. The deterministic models include a  $k$ -neighbors (KNN), a random forest (RFR), and a multilayer feed-forward artificial neural network (ANN). The probabilistic models consist of an ANN with a Gaussian posterior and two Bayesian ANNs (BANNs): a variational inference ANN and a dropout ANN.

Each model has its own set of advantages and disadvantages. Simple models like  $k$ -neighbors and random forest regressors are highly scalable and perform quickly on large, well-distributed samples, with few hyperparameters requiring tuning. However, they may not be the best choice for highly complex problems and are incapable of extrapolating data.

ANNs based regressors may provide better results for complex problems but require more hyperparameters to be fine-tuned and demand greater computational resources.

Measuring reliable uncertainties associated with each model’s predictions is crucial for most scientific cases. BANNJOS computes these uncertainties,  $\sigma(y_{\text{pred}})$ , in different ways depending on the model’s capabilities.

### 2.2.1. Deterministic Models

Three from the six models implemented in BANNJOS are unable of predicting the uncertainty associated with a prediction’s nominal value. These models include KNR, RFR, and the basic ANN. In such cases, BANNJOS calculates the expected uncertainty using a  $k$ -fold cross-validation scheme. In brief, the training sample is randomly divided into  $k$  equal-sized subsamples. A nominal model is then trained on  $k - 1$  subsamples and used to predict values in the remaining subsample, which acts as the validation data. This process is repeated  $k$  times<sup>4</sup>, with each of the  $k$  subsamples used once as validation data. After completing the  $k$ -fold cross-validation, the results are used by BANNJOS to train a second model (the variance model) based on the relation between  $|y_{\text{true}} - y_{\text{pred}}|^2$  and  $\mathbf{x}$ . This variance model predicts the uncertainties for the nominal model’s predictions. In our trials, this method proved to be faster than using  $k$  models trained during the  $k$ -fold cross-validation to predict uncertainty, specially when predicting in large data sets, while providing very similar results.

### 2.2.2. Probabilistic Models

BANNJOS includes three probabilistic neural network models, two of which operate as Bayesian approximations. The first model, an ANN with a Gaussian posterior, fits and predicts both the nominal result and its aleatoric uncertainty, i.e., the uncertainty associated with randomness observed in the training sample. The variational inference BANN and dropout BANN models, however, can estimate both aleatoric and epistemic uncertainties, the latter arising from insufficient information in the training sample.

If a probabilistic model is used,  $y_{\text{pred}}$  and its associated uncertainty are derived by sampling the model’s posterior multiple times ( $N$ ). This process is equivalent to sampling the PDF of the prediction for each instance (astronomical object),  $y_{\text{pred},i}$ . BANNJOS computes several useful statistics from these PDFs such as the mean and several percentiles (see Section 3.3 and Appendices C, and F).

## 3. Using BANNJOS on J-PLUS for Object Classification

In this work, we have utilized BANNJOS for object classification in J-PLUS. We adopted a systematic approach to construct the training sample, select the optimal model and its corresponding hyperparameters, and derive the probability distribution functions, PDFs, for each object’s likelihood of belonging to a specific class.

### 3.1. Training Sample

We compiled an extensive training sample consisting of a selection of objects with J-PLUS photometry and available spectroscopic classification. We first downloaded the entire J-PLUS catalog, with over 47 million sources, taking information from several of the main scientific tables available at CEFCA’s archive<sup>5</sup>. We combined the resulting table with data from ongoing and past all-sky surveys that contain additional information about the observed sources. These data include astrometric and photometric information from *Gaia* DR3 (Gaia Collaboration et al. 2016, 2023) and CatWISE2020 (Marocco et al. 2021), as well as reddening information from dust maps of Schlafly & Finkbeiner (2011). In order combine all the data, we took advantage of the catalogues matches already available at the J-PLUS archive. Details on the specific J-PLUS archive query and the variables utilized during the training are available in Appendix A. Broadly, the training list encompasses data on:

- J-PLUS photometry, i.e., fluxes measured across eight apertures for the 12 bands using forced photometry (SExtractor dual mode);
- J-PLUS position on the CCD and morphology, i.e., ellipticity, effective radius;
- J-PLUS photometry and masking flags;
- J-PLUS Tile observation details, i.e., seeing, zero-point, noise, etc;
- CatWISE2020 photometry, i.e., `W1mpro_pm` and `W2mpro_pm` bands;
- *Gaia* and CatWISE2020 astrometry, i.e., parallaxes and the absolute value of the one-dimensional proper motions<sup>6</sup>.

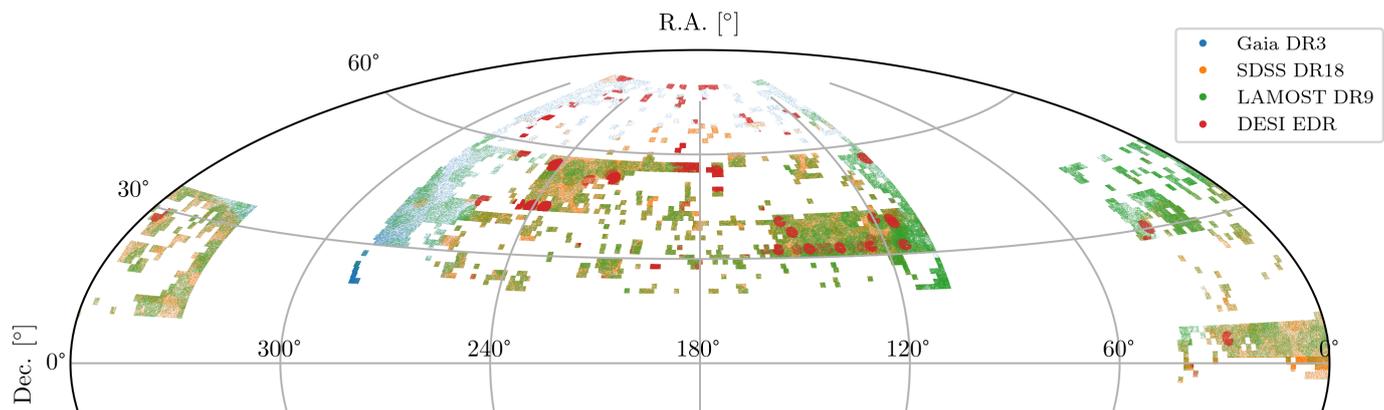
Flux measurements across different apertures provide additional insights into the source’s morphology. Its position in the detector can also be relevant, as it may help identify geometric distortions or other image-affecting cosmetic effects. Furthermore, details regarding the tile image’s quality, such as average seeing during acquisition, can be crucial for determining whether the source is extended. In total, we included 445 variables in our analysis.

In order to obtain the spectroscopic classification, we performed a cross-match of the resultant J-PLUS+*Gaia*+CatWISE2020 table with the SpecObj table from SDSS DR18 (Almeida et al. 2023), the Low Resolution spectroscopy catalog from LAMOST DR9 (Large Sky Area Multi-Object Fiber Spectroscopic Telescope; Cui et al. 2012), and the Early Data Release (EDR) from DESI (Dark Energy Spectroscopic Instrument; DESI Collaboration et al. 2023). The match is based on the sky position of sources across the catalogs, considering a source identical if its registered positions differ between catalogs by less than a specified angular separation. This maximum separation was determined in two steps. In the first step, a very large radius of 2 arcsec was used, ensuring all possible matches were included. In the second step, the maximum separation was determined based on finding 99% of matches within it. For J-PLUS and SDSS, the maximum allowed separation was found to be 0.6 arcseconds, while for J-PLUS with LAMOST and DESI, it was set

<sup>4</sup> The  $k$  value is set by the user. Larger  $k$  values will reduce bias in the prediction, but at larger computational cost.

<sup>5</sup> <https://archive.cefca.es/catalogues/jplus-dr3>

<sup>6</sup> Two-dimensional proper motions could potentially leak positional information (RA, Dec) to the model due to the Solar Galactic Reflex Velocity.



**Fig. 2.** Aitoff projection of the training list sources. The positions of the sources are represented by small dots, color-coded according to their originating survey. These are sources identified in J-PLUS with available spectroscopic or photospectroscopic classification. The small squares with 1.4-degree sides reflect the J-PLUS tiles. For a reference on the number of sources, see Table 2.

to 0.65 arcseconds. Lastly, we included a set of unequivocally classified *Gaia* sources, where the probability of belonging to a particular class is 1 and 0 for the other two classes. The resulting catalog underwent cleaning and consistency checks. Initially, we filtered out poorly measured sources by applying recommended criteria from SDSS:

- $0.864 < \text{RCHI2} < 1.496$ ,
- $\text{Z\_ERR} > 0$ ,
- $\text{Z\_ERR}/(1+\text{Z}) < 3 \times 10^{-4}$ ,
- $\text{PLATEQUALITY}$  not "bad",
- $(\text{SPECPRIMARY} = 1)$  or  $(\text{SPECLEGACY} = 1)$ ,
- $\text{SN\_MEDIAN\_ALL} \geq 2$ ,
- $\text{ZWARNING\_NOQSO} = 0$ .

The clipping levels for RCHI2 were set to the 0.1 and 0.9 quantiles of its distribution, respectively. For sources from LAMOST, the following criteria were applied:

- $\text{FIBERMASK} = 0$ ,
- $\text{Z\_ERR} > 0$ ,
- $\text{Z\_ERR}/(1+\text{Z}) < 3 \times 10^{-4}$ ,
- $\text{SN\_MEDIAN\_ALL} \geq 3$ .

For the reliable sources classified in DESI EDR, we selected those with:

- $\text{TARGETID} > 0$ ,
- $\text{SV\_PRIMARY} = 1$ ,
- $\text{ZWARN} = 0$ .

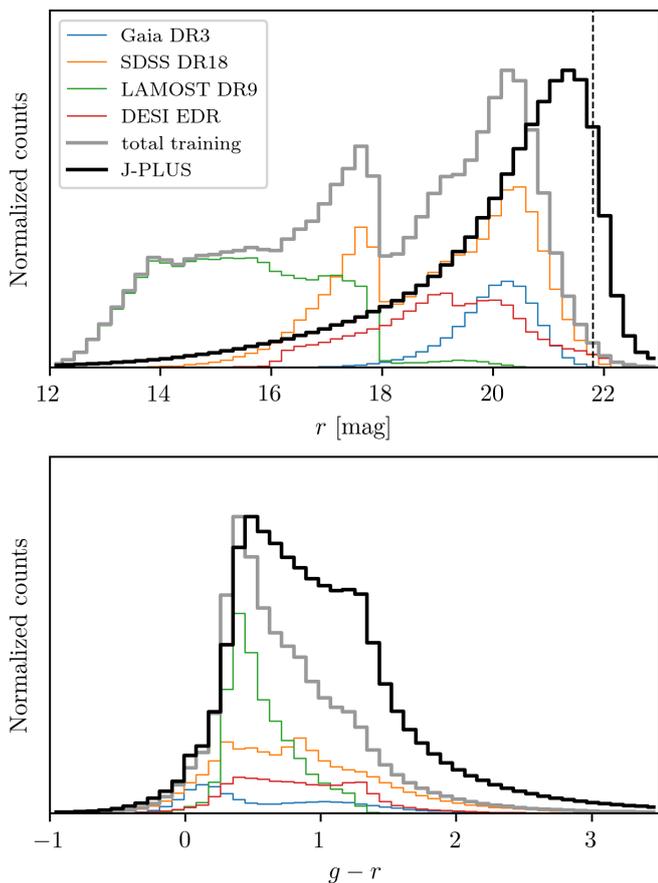
Lastly, *Gaia* sources were chosen based solely on their renormalized unit weight error, with  $\text{RUWE} \leq 1.3$ . The applied quality filters resulted in 586,641 sources from SDSS, 508,432 from LAMOST, 256,532 from DESI, and 152,566 from *Gaia*, totaling 1,504,171 sources.

The spatial distribution of these sources is illustrated in Figure 2. Histograms showing the magnitude and color ranges covered by each of the surveys in the training list are displayed in Figure 3. Although the training list spans the entire J-PLUS footprint, the levels of completeness and color coverage vary, inherent to each contributing survey. For instance, *Gaia* (depicted in blue) provides uniform sampling across the footprint but is limited to relatively blue sources ( $g - r \lesssim 1.6$ )

and does not reach the depths of its spectroscopic counterparts. Conversely, surveys such as SDSS or DESI delve deeper and sample the less populated galactic poles, offering insights into distant galaxies and QSOs. Incorporating various catalogs into the training list maximizes the coverage across the parameter space spanned by J-PLUS, minimizing instances where the model must extrapolate predictions, which, in theory, should enhance results. However, due to the disparate coverage in terms of object types, photometric properties, and spatial distribution among the surveys, it is crucial to exclude any positional information from the training list to prevent the model from learning the inherent biases in the selection and properties of the sources inherent to spectroscopic surveys.

A significant number of sources have multiple measurements in two or more of the four classification catalogs. Confusion matrices for the classification of such objects, shown in Figure 4, illustrate the consistency (or lack thereof) between the surveys, indicating the relative number of objects classified identically or differently between pairs of surveys. These matrices were constructed using all objects shared between each survey, with SDSS serving as the reference. Although inter-survey consistency is generally good, it is not without discrepancies. For example, LAMOST tends to classify fewer objects as stars compared to SDSS (91%), and *Gaia* often classifies many of SDSS's galaxies as QSOs (14%). Despite these differences, the overall agreement is satisfactory, especially considering that *Gaia* contributes only around 9% of the galaxies and QSOs to the final list.

We removed all but one instance from repeated objects with consistent spectroscopic classes across catalogs (1160 in total), retaining duplicates where classifications differed depending on the catalog. We identified a small subset of objects (747) with varying classifications across two or more input lists. After visually inspecting the spectra of some of these objects, we found they were primarily galaxies misidentified as stars and vice versa (233 cases), and distant active galaxies classified either as QSOs or galaxies (467 cases). We refined this last group further by excluding nearby sources (stars) with parallax  $\varpi/\sigma(\varpi) > 1$ , meaning objects with parallax different from zero at a  $1\sigma$  confidence level. Since the dichotomy between galaxies and QSOs is not always clear, we decided to keep those objects (426), but removed those with



**Fig. 3.** Distribution of the  $r$  magnitude (top panel) and  $g - r$  color (bottom panel) for the sources in J-PLUS and the training list. The distribution of sources from each contributing survey is depicted by thin lines in various colors (as in Figure 2). The aggregate training list is represented by a thick gray line, while the entirety of J-PLUS sources, 47,751,865 objects, is shown by a thick black line. Histograms related to the training list have been normalized to the peak of the total training list (gray line), whereas the J-PLUS histograms have been normalized to their own peak. A vertical thin dashed line shows the limiting magnitude of J-PLUS DR3 ( $r = 21.8$  mag, see Table 1) The extended tail of objects with  $g - r \gtrsim 1.5$  is mostly composed by low S/N sources with poorly determined  $g$  magnitude.

implausible class combinations, such as Galaxy-Star or QSO-Star. Notably, we found no objects classified differently across three or more catalogs, meaning no objects were assigned all three possible classes.

The final list comprises 1,365,700 objects (1,365,274 unique), distributed into 480,267 galaxies, 127,633 QSOs, and 757,800 stars. Each object's record in the catalog includes photometric, astrometric, and morphological data from J-PLUS, *Gaia* DR3, and CatWISE2020, alongside spectroscopic classification from SDSS DR18 (585,336 objects in common), LAMOST DR8 (507,737 objects in common), DESI EDR (254,586 objects in common), and spectrophotometric classifications from *Gaia* DR3 (151,974 objects in common). The details about the composition of the training set can be found in Table 2. We reserved 10% of these sources to compose the 'test' set, which will be utilized later to validate and assess our results.

**Table 2.** Composition of the training set.

Survey	Galaxies	QSOs	Stars	Total
SDSS DR18	354,849	93,905	137,887	586,641
LAMOST DR8	28,289	11,219	468,924	508,432
DESI EDR	140,604	15,029	100,899	256,532
<i>Gaia</i> DR3	1,014	74,391	77,161	152,566
Total	524,756	194,544	784,871	1,504,171
Total non-repeat <sup>a</sup>	480,267	127,633	757,800	1,365,700
Training no balan.	432,315	114,774	682,041	1,229,130
Balan. downsamp. <sup>b</sup>	114,774	114,774	114,774	344,322
Balan. supersamp. <sup>c</sup>	682,041	682,041	682,041	2,046,123

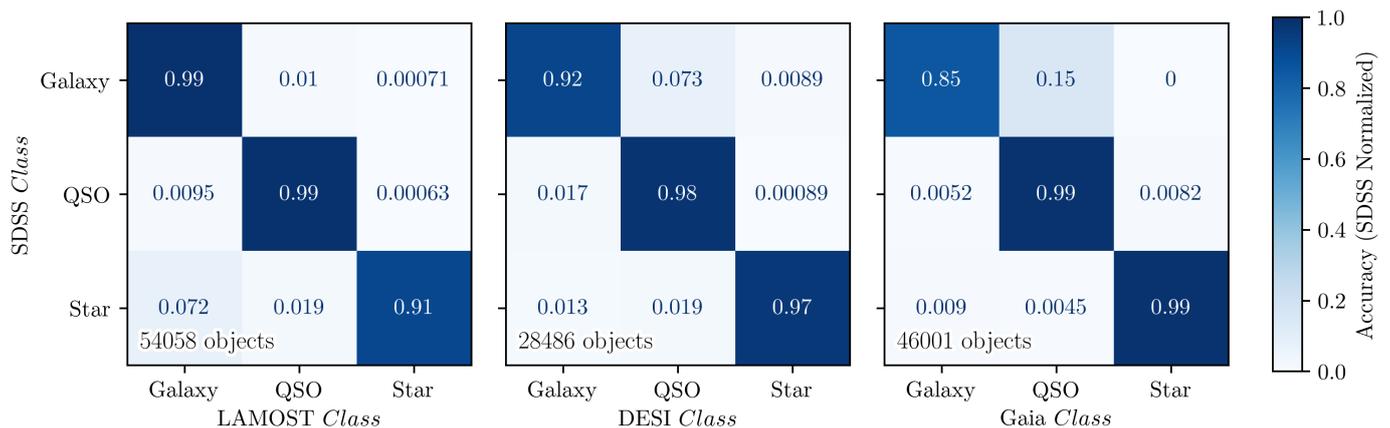
**Notes.** <sup>(a)</sup> Total non-repeated sources. Counts after removing repeated sources present in one or more surveys. <sup>(b)</sup> Balanced training list down-sampled to the least common class. <sup>(c)</sup> Balanced training list super-sampled to the most common class.

The compiled training list is disproportionately biased towards the "star" class, with galaxies being the second most prevalent. Indeed, within the J-PLUS footprint, stars are the most numerous due to the survey's depth limitations, but the class ratios in the training list might be artificially skewed by the design of the contributing surveys. To evaluate and mitigate potential biases, we tested our models using three different versions of the training list. The first version retained the original composition after removing 10% for the test sample, resulting in 1,229,130 sources. The second, a downsampled version, randomly reduced the numbers of the two most abundant classes to match those of the least populated class, the QSOs, resulting in a list of 344,322 equally distributed sources among the "Star", "QSO", and "Galaxy" classes. The third, a balanced version of the original list, employed oversampling for the less populated classes using a Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al. 2002), which uses a k-neighbor interpolator to generate new instances (sources) based on the averages of neighboring properties. This "augmented" list, balanced among classes, contains 2,046,123 sources.

### 3.2. Model Selection and Hyperparameter Tuning

To achieve optimal classification, we evaluated the models described in Section 2.2, identifying the model and hyperparameter configuration yielding the best results. Additionally, we assessed the three variations of our training sample outlined in Section 3.1. The performance of each configuration was measured on the test sample through cross-validation, comparing the spectroscopic/spectrophotometric class of an object,  $y_{\text{true}}$ , against the predicted one,  $y_{\text{pred}}$ . For BANNs, which provide the PDF for each class, we designated the class corresponding to the highest median value (quantile 0.5) from the three PDFs as the predicted class  $class = \max[PC_{\text{Galaxy}}(50), PC_{\text{QSO}}(50), PC_{\text{Star}}(50)]$ . We started evaluating the models in a coarse grid of hyperparameters in order to gain a general idea about their performance. All the ANN based models were trained using a validation sample of 40%<sup>7</sup> In general, ANNs significantly surpassed RFR and KNR in performance, with a deep dropout BANN emerging as the superior model. This result is somewhat expected, as dropout

<sup>7</sup> The model sets apart this fraction of the training list, not training on it, and evaluates the loss metrics on this data at the end of each epoch (training iteration).



**Fig. 4.** Confusion matrices between the four surveys used to compile the training list. The proportion of objects in each bin relative to the total number of objects in SDSS is indicated by varying shades of blue and is presented as a real number  $\in [0, 1]$ . The confusion matrices also specify the total number of objects used in their computation.

BANNs are less susceptible to overfitting compared to traditional and variational inference ANNs, especially when the model architecture is sufficiently deep. With the exception of RFR, the augmented training list consistently yielded slightly better results across all models. After selecting the dropout BANN as the best overall model, we evaluated its performance across an extensive grid of potential hyperparameter configurations:

- Number of hidden layers: [8, 7, 6, 5, 4, 3, 2]
- Batch size: [32, 64, 128, 256, 512]
- $L_n$ : [1600, 1300, 1000, 700, 500, 300, 200, 100, 50, 5]
- Dropout ratio at  $L_{1-8}$ : [0.1, 0.2, 0.3, 0.4, 0.5]
- Dropout ratio at  $L_0$ : [0.0, 0.1, 0.2]
- Loss function: MSE, Huber
- Initial learning rate: [ $10^{-3}$ ,  $5 \times 10^{-4}$ ,  $10^{-4}$ ,  $5 \times 10^{-5}$ ]
- Step decay in learning rate: [10, 20, 30, 50,  $\infty$ ]

Here, Number of hidden layers denotes the quantity of hidden layers between the input and output. Batch size refers to the number of samples processed before updating the model. The term  $L_n$  indicates the number of neurons in the  $n$ -th layer for  $1 \leq n \leq 8$ . The Dropout ratios represent the fraction of randomly dropped neurons in each epoch at hidden layers ( $n > 0$ ) or at the input layer ( $n = 0$ ). The Loss function is the metric minimized during the fitting process. Initial learning rate denotes the step size at each epoch during optimization, and Step decay in learning rate specifies the epoch count before the learning rate updates to half of its previous value, with  $\infty$  indicating a constant learning rate.

Given the vast number of possible hyperparameter combinations, exceeding  $10^{11}$ , testing all configurations was computationally unfeasible. Therefore, we employed a Random Search Cross-Validation approach, randomly selecting and testing 800 configurations against the test sample. We limited each model’s training to 2000 epochs and incorporated an early stopping mechanism that halts training if the loss does not improve over 50 epochs<sup>8</sup>. To make the predictions,

<sup>8</sup> Improvement is defined within a precision of  $10^{-5}$  in the loss function.

we sampled the posterior only 128 times. This allows us to keep a reasonable computational cost for all the optimization tests, while realistically sampling the hyperparameter space. Upon completing cross-validation for the 800 models, we utilized a Histogram-based Gradient Boosting Regression Tree to identify the optimal model architecture and training hyperparameters. Model evaluation was based on three metrics: balanced accuracy, average precision, and their quadratic sum. The found optimal hyperparameters for the dropout BANN are:

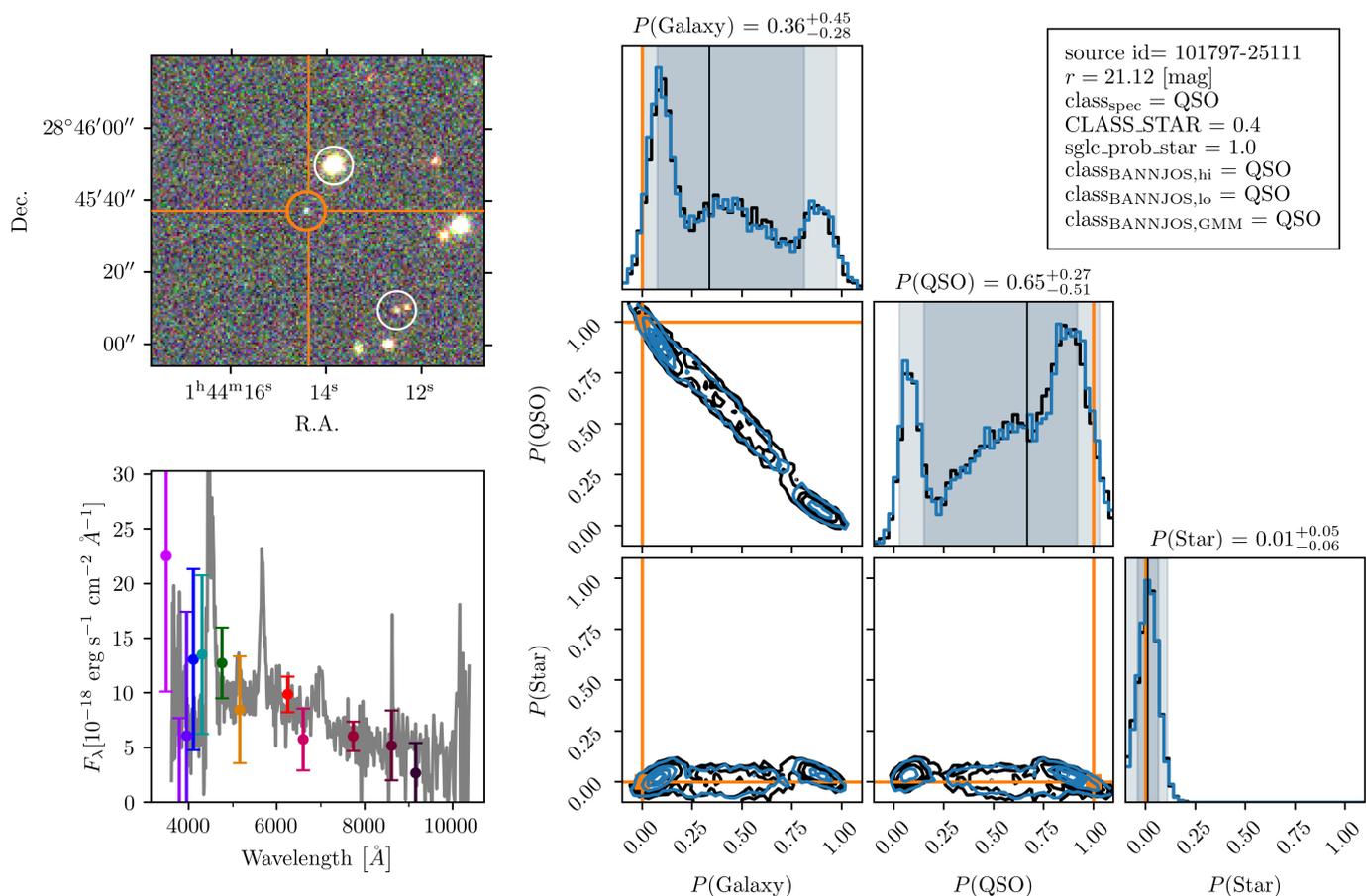
- Number of hidden layers: 4
- Batch size: 256
- $L_{1-4}$  = [700, 1300, 500, 300]
- Dropout ratio at  $L_{1-4}$ : 0.4
- Dropout ratio at  $L_0$ : 0
- Loss function: MSE
- Initial learning rate:  $10^{-4}$
- Step decay in learning rate: 30

Where now  $L_{1-4}$  is the number of neurons in the hidden layers 1 to 4. Despite this being the best configuration, our testing indicates that model performance is largely invariant to hyperparameter configuration, provided the values are reasonable. For instance, the accuracies of our top 10 models differ by less than 0.1%. More details on the impact on the accuracy that the model hyperparameters have, the correlation between some hyperparameters, and the Individual Conditional Expectation (ICE) curves, can be found in Appendix B.

### 3.3. Training the Model and Sampling the Posterior

We used the hyperparameters derived in the previous Section 3.2 to build our model. To maximize results, we extended the training stopping criteria to 10,000 epochs and updated the early stop function to 500 epochs. The model converged after 8,963 epochs with a loss function precision of  $10^{-7}$ .

A significant advantage of using a BANN model is its capability to provide the full PDF for  $y_{\text{pred}}$ . Here, the total PDF for a specific object is the sum of the individual class PDFs, defined within the 3-dimensional space delineated by orthogonal axes  $P(\text{class} = \text{Galaxy})$ ,  $P(\text{class} = \text{QSO})$ ,

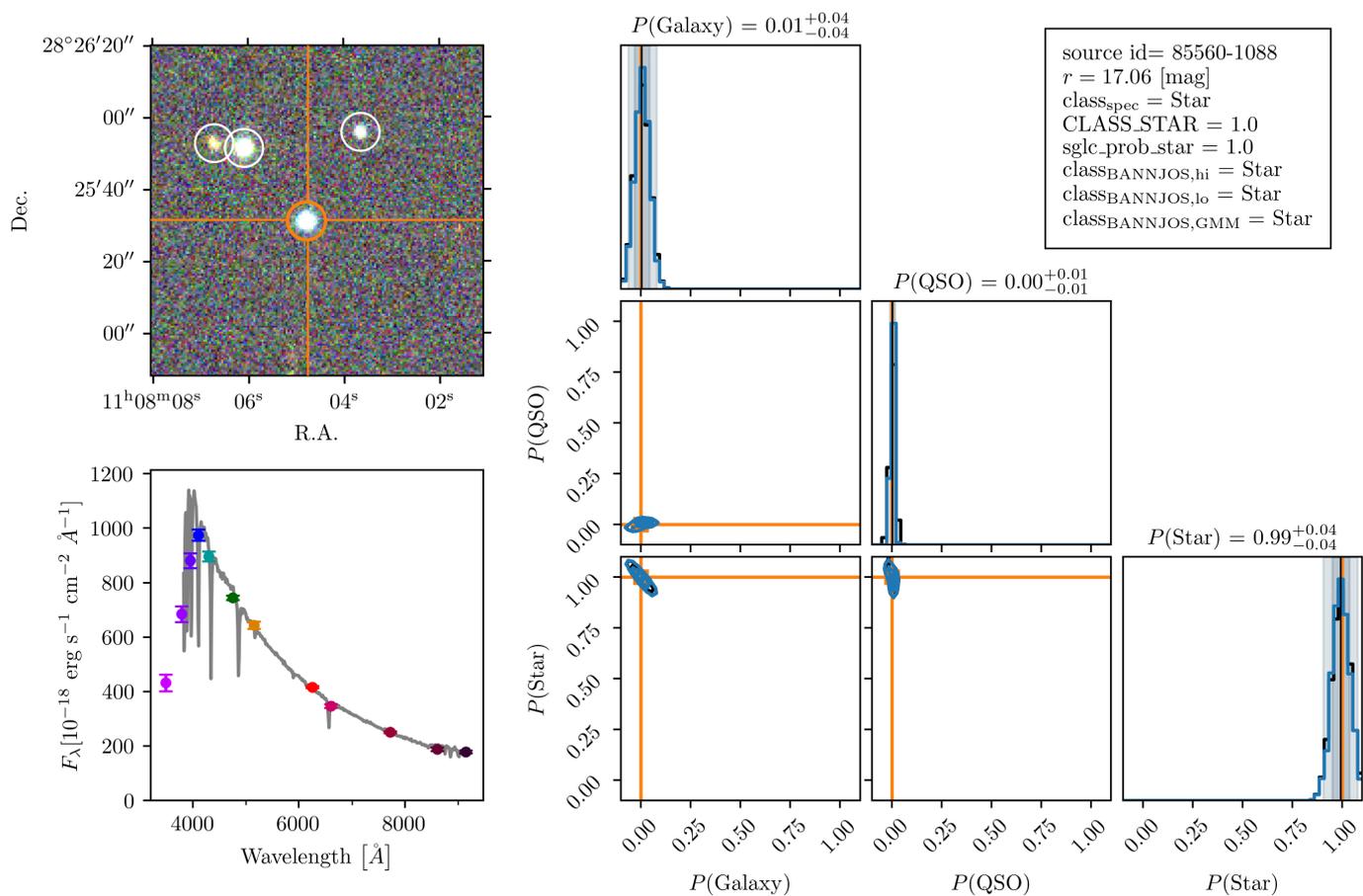


**Fig. 5.** Example of results obtained with BANNJOS for the source Tile Id = 101797, Number = 25111. *Upper left:* J-PLUS color image composed using the  $r$ ,  $g$ , and  $i$  bands, normalized between the 1st and 99th percentile of the total flux. The classified source is at the center of the reticle, marked by an orange open circle, while other sources detected in the J-PLUS catalog are marked with white open circles. *Lower left:* SDSS spectrum undersampled by a factor of 10 for improved visibility (in gray), alongside J-PLUS photometry across its 12 bands. *Right:* Corner plot illustrating the three-dimensional posterior probability distributions for  $P(\text{class} = \text{Galaxy})$ ,  $P(\text{class} = \text{QSO})$ , and  $P(\text{class} = \text{Star})$ . Orange lines denote the spectroscopic class,  $y_{\text{true}}$ . Black contours and histograms represent the posterior probability distribution sampled 5000 times ( $\text{class}_{\text{BANNJOS,hi}}$ ), with black lines indicating the median probability for each class and the gray-shaded area covering the 2nd to 98th (lighter gray) and the 16th to 84th (darker gray) percentile ranges. Blue contours, histograms, and shaded areas depict the reconstructed posterior probability distribution from the GMM model fitted to  $N = 300$  points. A text box in the upper right corner lists the complete source ID, its  $r$  magnitude, the ‘true’ classification from spectroscopy ( $\text{class}_{\text{spec}}$ ), its CLASS\_STAR and sglc\_prob\_star scores, and the BANNJOS classifications from the high-quality posterior sampling ( $N = 5000$ ,  $\text{class}_{\text{BANNJOS,hi}}$ ), the regular-quality posterior sampling ( $N = 300$ ,  $\text{class}_{\text{BANNJOS,lo}}$ ), and the classification from the reconstructed PDF following the GMM compression method,  $\text{class}_{\text{BANNJOS,GMM}}$ . The classification is determined as the one with the highest median probability value. Despite its complexity, the PDFs obtained from sampling 5000 times and the one after reconstruction from the GMM are nearly indistinguishable, with gray and blue contours and shaded areas covering the same areas. The corresponding classifications,  $\text{class}_{\text{BANNJOS,hi}}$  (black) and  $\text{class}_{\text{BANNJOS,GMM}}$  (blue), also match the spectroscopic classification. The GMM compression procedure is described in Appendix C.

and  $P(\text{class} = \text{Star})$ . Sampling the model multiple times,  $N$ , in a Monte-Carlo fashion, BANNJOS outputs three probabilities for each sample, corresponding to the likelihood of the object belonging to each class. Due to the model’s stochastic nature, different samples yield unique outcomes, resulting in  $N \times 3$  data points. Figure C.1 shows a sampling example with  $N = 5000$ .

While maintaining and analyzing all points permits comprehensive PDF reconstruction and facilitates informed class determination, extensive posterior sampling (large  $N$ ) is computationally intensive and data-heavy, becoming impractical for the entire J-PLUS catalog. To alleviate computation time and storage, we employed a reduced sampling count ( $N = 300$ ,

denoted as  $\text{class}_{\text{BANNJOS,lo}}$ ) and projected the 3-dimensional probabilities onto the plane defined by  $P(\text{class} = \text{Galaxy}) + P(\text{class} = \text{QSO}) + P(\text{class} = \text{Star}) = 1$ , thereby reducing the dimensionality to two. We then applied a 2-dimensional Gaussian Mixture Model (GMM) with three components to model the projected probabilities. The GMM parameters—covariance matrices, means, and weights—are compactly stored, effectively capturing the posterior’s essence with significantly fewer parameters. BANNJOS additionally computes mean, median absolute deviation (MAD), and specified percentiles for each class’s cumulative distribution, along with Pearson’s correlation coefficients between classes. The compression method and its efficacy are explained in more detail in Appendix C.



**Fig. 6.** Example of results obtained with BANNJOS for the source Tile Id = 85560, Number = 1088. At  $r = 17.06$  mag, BANNJOS classifies this source as a star with high confidence. The markers correspond with those used in Figure 5. Most sources classified with BANNJOS will show PDFs similar to this one, with very little dispersion around the predicted class.

In Figure 5, we showcase a faint source ( $r = 21.12$  mag) classified by BANNJOS. Due to its high photometric and astrometric uncertainties, BANNJOS assigns a low-confidence QSO classification to the source with  $P(\text{class} = \text{QSO}) = 0.65^{+0.27}_{-0.51}$ . The derived PDF exhibits multiple peaks and a pronounced anticorrelation between  $P(\text{class} = \text{QSO})$  and  $P(\text{class} = \text{Galaxy})$ , indicating a considerable probability that the object could also be a galaxy with  $P(\text{class} = \text{Galaxy}) = 0.36^{+0.45}_{-0.28}$ . This ambiguity underscores the intricate nature of distinguishing between QSOs and active galaxies, particularly when faced with unresolved targets. Indeed, when checking the SDSS’s SUBCLASS field for this and other similar objects, we found all to be unresolved active galaxies or QSOs. Interestingly, in this particular case the `sglc_prob_star` classifier wrongly assigned a very high probability of being a star to the object, showcasing how the combination of photometric, astrometric and morphological information used by BANNJOS can help to distinguish sources based on their nature, rather than just their apparent aspect, even in low S/N conditions.

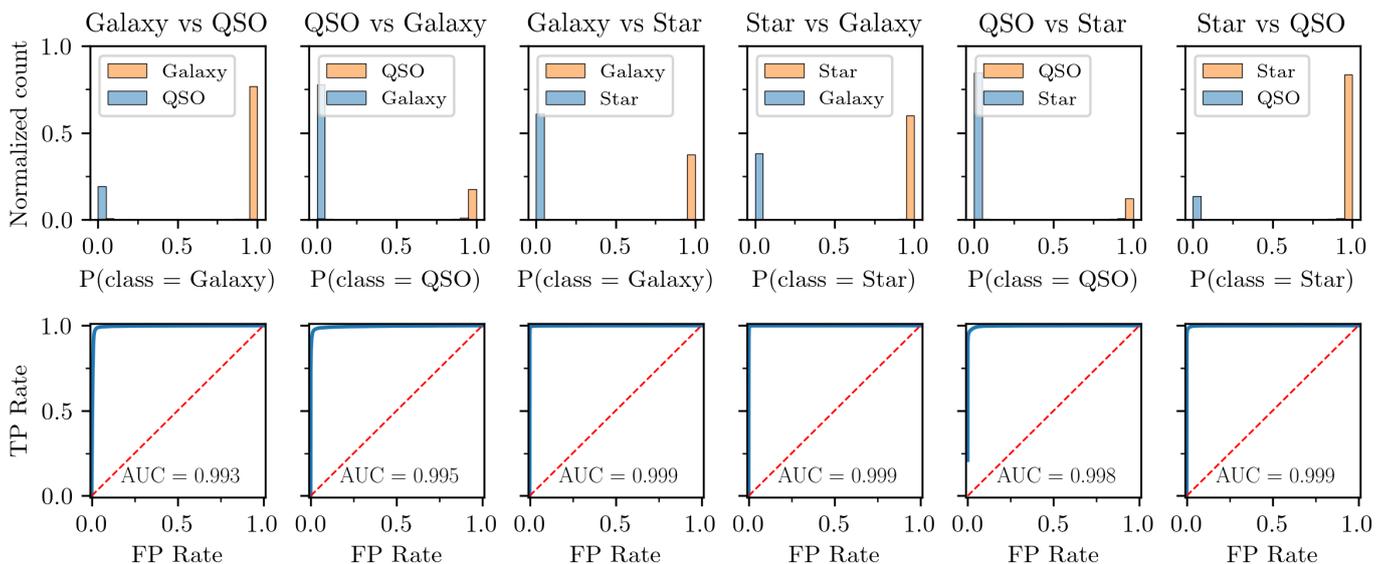
The implemented GMM compression method was tested against an additional validation sample with  $N = 5000$  BANNJOS posterior samples (`classBANNJOS,hi`). The GMM faithfully reproduces the original results. The high-quality sampling (black histograms in Figure 5), and the ones using  $N = 300$  compressed with the GMM (`classBANNJOS,GMM`, blue histograms in

Figure 5) are consistent across the entire test sample, barring rare exceptions. Only in 4 objects with very low S/N (0.003% of the cases, see Appendix C), we obtained a different classification using the highest median probability of the three classes, indicating that the implemented compression method provides a reliable classification.

In contrast with the low S/N example, in Figure 6 we show an example of a source classified by BANNJOS with high-confidence. The relatively bright source ( $r = 17.06$  mag), is unequivocally identified as a star by BANNJOS:  $P(\text{class} = \text{Star}) = 0.99^{+0.04}_{-0.04}$ ,  $P(\text{class} = \text{Galaxy}) = 0.01^{+0.04}_{-0.04}$ ,  $P(\text{class} = \text{QSO}) = 0.00^{+0.01}_{-0.01}$ . Most sources brighter than  $r \sim 20$  mag will show posteriors that tightly distribute around the predicted class with very little dispersion or correlation to other classes probabilities.

#### 4. Validation of the Model

In this section, we validate the classification outcomes of BANNJOS using the identical test sample and criteria established in Section 3.2. This involves assigning the predicted class of the object,  $y_{\text{pred}}$ , and comparing these with the spectroscopic classifications,  $y_{\text{true}}$ . Since BANNJOS provides the full PDF of the classification, there are multiple criteria that could be used in order to assign a class to each source. However,



**Fig. 7.** Model performance evaluated up to  $r = 21.9$  mag, the median completeness for compact sources in J-PLUS, assessed on the test sample consisting of 136,570 objects. The top panels illustrate the distribution of probabilities for objects being classified according to their spectroscopic category. The difference in bar heights reflex the different amount of sources present in the test sample. The bottom panels depict the Receiver Operating Characteristic (ROC) curves in blue for each class combination. These curves approach the maximum True Positive Rate (1) almost immediately, demonstrating excellent model performance. The Area Under the Curve (AUC) exceeds 0.99 for all six class combinations.

we will focus now on the most simple approach that is assigning classes based on the highest median probability value across the three PDFs. In Section 5, we explore how the purity of the selected samples can be improved using more sophisticated selection criteria. It is worth mentioning that the test sample was subtracted from the training list before any preprocessing of the data, and thus it is unbalanced with proportions very similar to those of the original training set (see Table 2).

#### 4.1. Average Validation

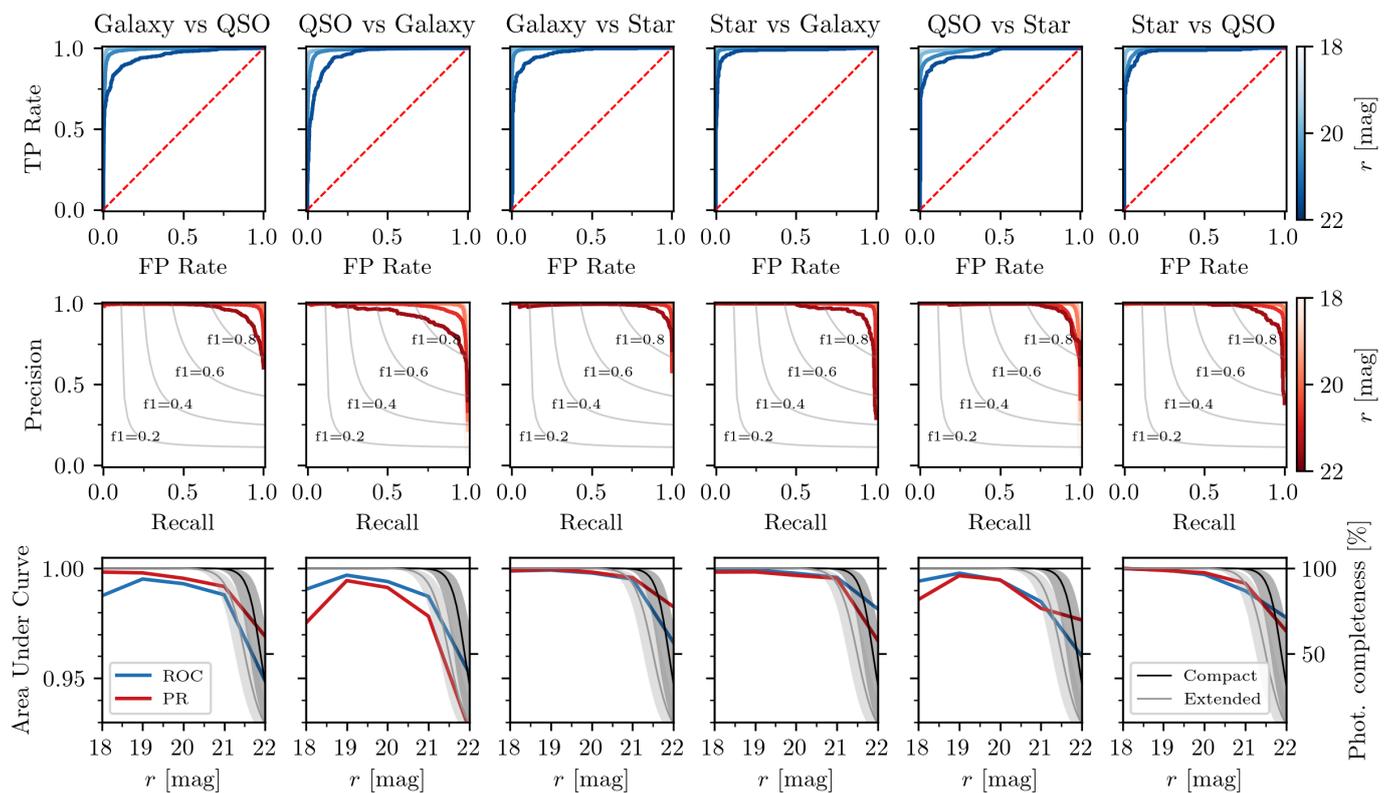
To gauge the model’s average efficacy, we compared the predicted and true classifications for objects with  $r \leq 21.9$  mag, corresponding to the median 50% completeness threshold of J-PLUS for compact sources across 1,642 tiles. In Figure 7, we show the normalized counts for each predicted class in pairs and their respective Receiver Operating Characteristic (ROC) curves. Each pair shows the statistics of True and False positives (TPs and FPs, respectively), therefore showcasing the ability of the model to avoid confusion between species. For example, the Galaxy vs. QSO pair shows how many real galaxies were classified as such (TPs) and how many as QSOs (FPs), while the QSO vs. Galaxy shows the same statistics, but for actual QSOs that are classified as such or as a galaxy.

An ideal classification would yield an area under the curve (AUC) of 1, with no intermixing between classes in the count histograms, e.g. objects with  $y_{\text{true}} = \text{Galaxy}$  would exhibit a probability  $P(\text{class} = \text{Galaxy}) = 1$ , while probabilities for other classes would be zero. Our model’s results are nearly perfect, with all ROC AUC values exceeding 0.99 for all six possible inter-class combinations. The results also seem to be symmetric between switching classes, with very similar ROC AUC values, indicating that the rate of FPs and False Negatives

(FNs) is close to one. Asymmetric results are not desirable, since they reflect biases in the model’s prediction towards specific classes. For contrast, BANNJOS shows its ability to recover well the ratios between classes in all possible combinations, even though the distribution of the test sample is not balanced. The histograms in the upper panels further corroborate the accuracy of BANNJOS, displaying nearly all objects congregating at the extremes, indicating  $P(\text{class} = x) \sim 1$  for the evaluated predicted class and  $P(\text{class} = y) \sim 0$  for the other class. As anticipated, the pairs exhibiting the lowest accuracy are Galaxy versus QSO and vice versa, with ROC AUCs of 0.993 and 0.995, respectively, due to the actual dual nature of these sources.

#### 4.2. Signal-to-Noise Dependence

The predictive capability of BANNJOS is expected to be influenced by the quality of the data and its signal-to-noise ratio (S/N). We studied the potential impact of lower S/Ns by evaluating ROC curves across various magnitude bins, as shown in Figure 8. This figure also displays Precision-Recall (PR) curves, which balance the purity and completeness of the model’s classifications. Ideally, predictions should be both 100% pure and complete (full Recall). A decline in predictive power results in a compromise between these metrics. As anticipated, the S/N ratio of source measurements in  $\mathbf{x}$  markedly affects prediction quality, with dimmer objects proving more challenging to classify accurately. Deeper magnitude bins, indicated by darker shades, exhibit progressively deviating ROC curves from the ideal upper-left corner as magnitudes increase. This reduction in accuracy is mirrored in the PR curves, with fainter magnitudes resulting in a more obvious trade-off between purity and completeness. The third row in Figure 8 illustrates the AUC for both ROC and PR curves as a function of magnitude, alongside the median completeness for both



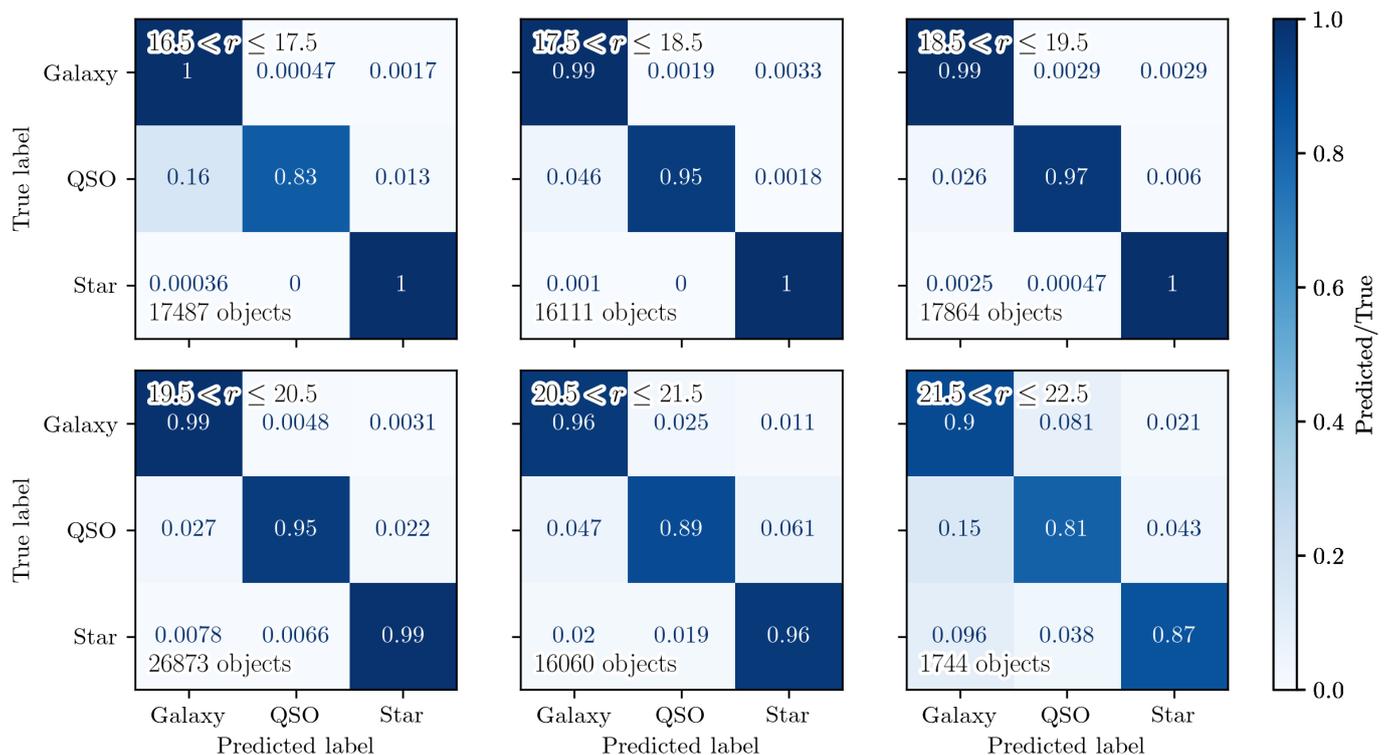
**Fig. 8.** Dependence of model performance on the signal-to-noise ratio of the sources. The top panels present the Receiver Operating Characteristic (ROC) curves for objects within different magnitude bins, shaded in varying tones of blue. Magnitude bins range from  $r = 17.5$  to  $r = 22.5$  mag in 0.5 mag increments (10 bins). The middle panels exhibit the Precision-Recall (PR) curves, illustrating the trade-off between sample purity and completeness, shaded in different red tones according to magnitude bin. Both quantities are related by  $f1$ , the harmonic mean of precision and recall, for which several values are plotted. The bottom panels summarize the results, depicting the evolution of the area under the curve (AUC) for both ROC and PR curves relative to the central magnitude of each bin. Blue and red colors denote the AUCs for ROC and PR, respectively. The median photometric completeness for compact and extended sources within J-PLUS is shown by dark and light gray curves, respectively, with surrounding gray shaded areas indicating the 16th and 84th percentiles of completeness. Classification remains near-perfect up to  $r \sim 21$  mag ( $AUC \geq 0.99$  for both curves), thereafter beginning to diminish, aligning with J-PLUS’s limiting magnitude and the ensuing loss of information in some bands. However, even under the most challenging conditions, such as distinguishing QSOs from galaxies or stars, ROC and PR AUC values maintain levels of  $\sim 0.98$  and  $\sim 0.93$  at  $r = 22$  mag, respectively, showcasing outstanding model performance. Notably, BANNJOS exhibits a propensity for mistaking QSOs for galaxies and vice versa at magnitudes brighter than  $r \sim 18$  mag, possibly due to active galaxies within the test (and training) sample being variably classified as galaxies or QSOs based on the specific spectroscopic survey.

compact and extended J-PLUS sources, depicted in light and darker gray, respectively (do not confuse with the Recall of the classification). The model’s classification remains nearly flawless up to  $r \sim 21$  mag ( $AUC \geq 0.99$  for both curves), before experiencing a decline, coinciding roughly with J-PLUS’s limiting magnitude due to diminished S/N and information loss in certain bands. Despite these challenges, even in the least favorable scenarios, such as differentiating QSOs from galaxies or stars, the model maintains high ROC and PR AUC values ( $\sim 0.98$  and  $\sim 0.93$  at  $r = 22$  mag, respectively), underscoring its exceptional performance. Also noteworthy is the increased confusion between QSOs and galaxies at magnitudes brighter than  $r \sim 18$  mag, likely stemming from the presence of active galaxies in both the test and training samples, classified differently across surveys.

Further insight into the model’s classification performance across different brightness levels is provided by Figure 9, which illustrates the confusion matrices between  $y_{\text{true}}$  and  $y_{\text{pred}}$  for varying magnitude ranges. Overall, the model achieves an average accuracy of approximately 95% for objects with

$r \sim 21.5$  mag. However, it faces challenges in accurately classifying QSOs at fainter magnitudes, a foreseeable issue since QSOs are morphologically similar to stars and their color differentiation becomes less reliable at  $r \geq 21$  mag due to the lower S/N ratios. Nevertheless, 81% of QSOs were correctly identified at  $r \geq 21.5$  mag, a good result given that J-PLUS’s nominal limiting magnitude is around  $r \sim 21.9$  mag. Additionally, since the sample is predominantly composed of stars and galaxies, the average error rate at the faint end remains below 12-13%.

Similar to observations from Figure 8, there is a noticeable increase in the misclassification of QSOs as galaxies at magnitudes brighter than  $r \sim 18$  mag. This trend is attributed to the inclusion in both the test and training samples of active galaxies, which exhibit similar characteristics but are categorized differently as either galaxies or QSOs, depending on the spectroscopic survey in question.



**Fig. 9.** Confusion matrices depicting the correlation between the true classes and the predicted classes for objects in the test sample across different magnitude bins. These bins range from  $r = 16.5$  to  $r = 22.5$  mag in 1 mag increments, totaling six bins. The proportion of objects within each bin relative to the true class is indicated by varying shades of blue, with values represented by real numbers within the interval  $[0, 1]$ . Additionally, the confusion matrices include the count of objects contributing to each calculation.

### 4.3. Position Dependence

To ensure our model remains unbiased, we deliberately excluded all positional information; i.e. positions in the sky, two-dimensional proper motions, etc; during its training phase. Nevertheless, due to the incorporation of dust attenuation data from Schlafly & Finkbeiner (2011), some positional data (albeit significantly degraded) may have influenced the model. To identify and mitigate potential biases and to study the presence of contaminants from other classes, we analyzed the prediction errors relative to position and class. This is shown in Figure 10, where the diagonal panels show the object surface density for each class in the test sample, and the off-diagonal panels the relative classification error for each possible combination between the three classes, i.e. the percentage of pollutants from each other class. Here, errors are calculated as the number of incorrectly classified objects divided by the total in the corresponding true class for each spatial bin, approximately  $3.5^\circ$  on each side.

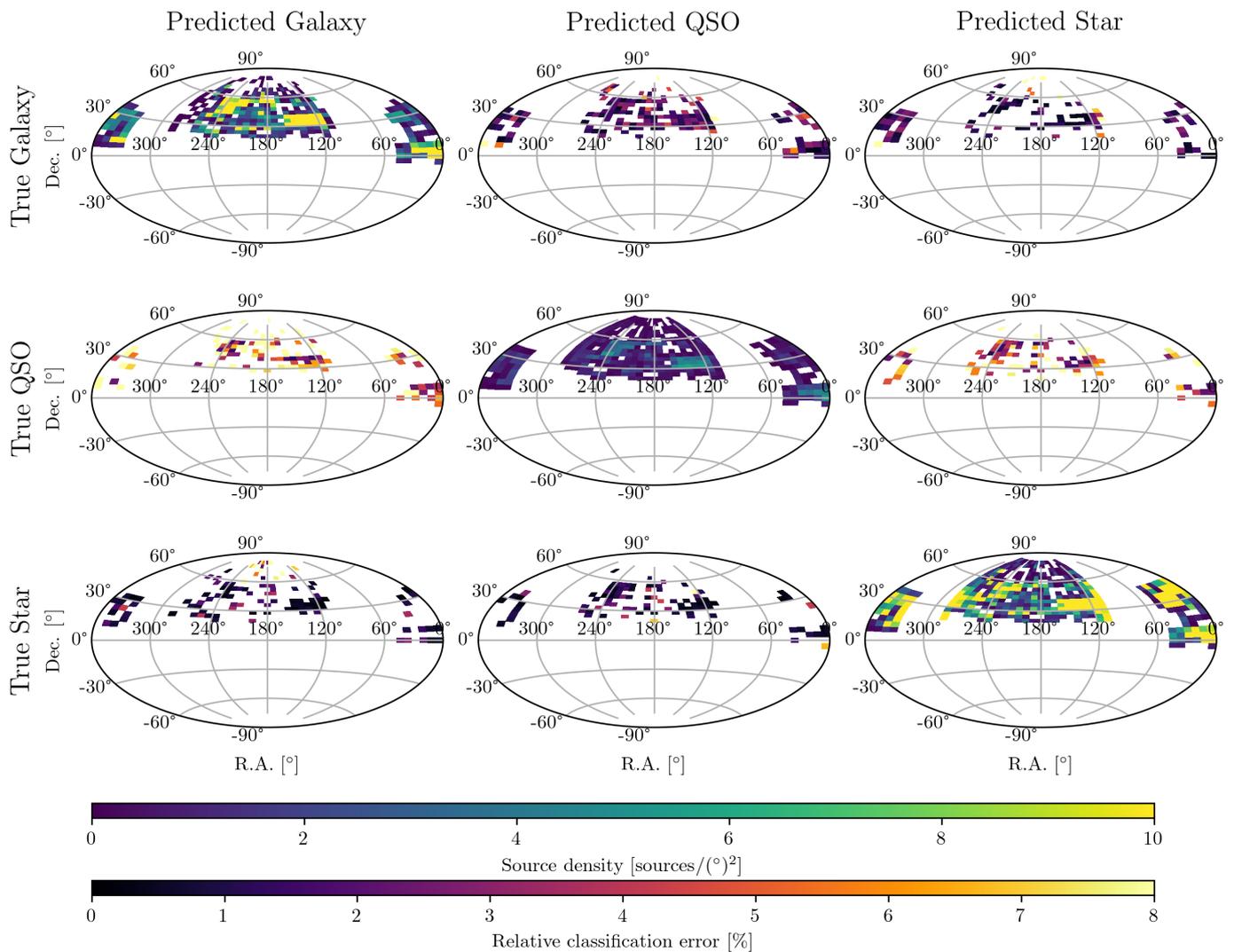
Our analysis reveals no significant correlation between source positions and classification error for any of the six class combinations, with error variations across different sky regions consistent with a normal distribution. This includes areas near the Galactic plane, where we might expect an increased prediction of stars if the model were biased by position. The existence of areas with higher surface density, clearly visible in the diagonal panels, is also noteworthy. These correspond to areas that were covered by more surveys or to deeper magnitudes (see Figure 2), hence the higher number of observed sources. This inhomogeneity in the training list is the

key reason why positional information must not be passed to the model. Otherwise it could learn the specific classification from each spectroscopic survey within their footprints, without generalizing the solution to the entire sky.

We further examined potential Galactic plane proximity effects by assessing the relative error solely as a function of Galactic latitude,  $b$ , and class. In Figure 11, histograms for each predicted class display the number counts and the average classification error across  $10^\circ$  intervals from  $20^\circ$  to  $90^\circ$ . Similar to Figure 10, errors are defined as the quantity of objects incorrectly assigned to each class, FPs, ordered by Galactic latitude  $b$ . Mild trends emerge here; notably, the misclassification rate of QSOs as stars increases with  $b$ , suggesting the model does not entirely compensate for the declining stellar density relative to QSOs. We also notice an increase in wrongly predicted galaxies that are spectroscopically identified as stars at low Galactic latitudes, consistent with higher stellar densities. Conversely, QSO predictions remain relatively stable across latitudes, indicating consistent contamination levels in QSO-predicted samples across all Galactic latitudes. Nonetheless, it's crucial to highlight the scale of demonstrated errors, with the highest error rates being approximately 5% for predicted QSOs and only 0.8% and 0.5% for the galaxies and stars categories, respectively.

## 5. Refining Your Selection

Throughout the paper, we have conventionally assumed that the object's class corresponds to the one



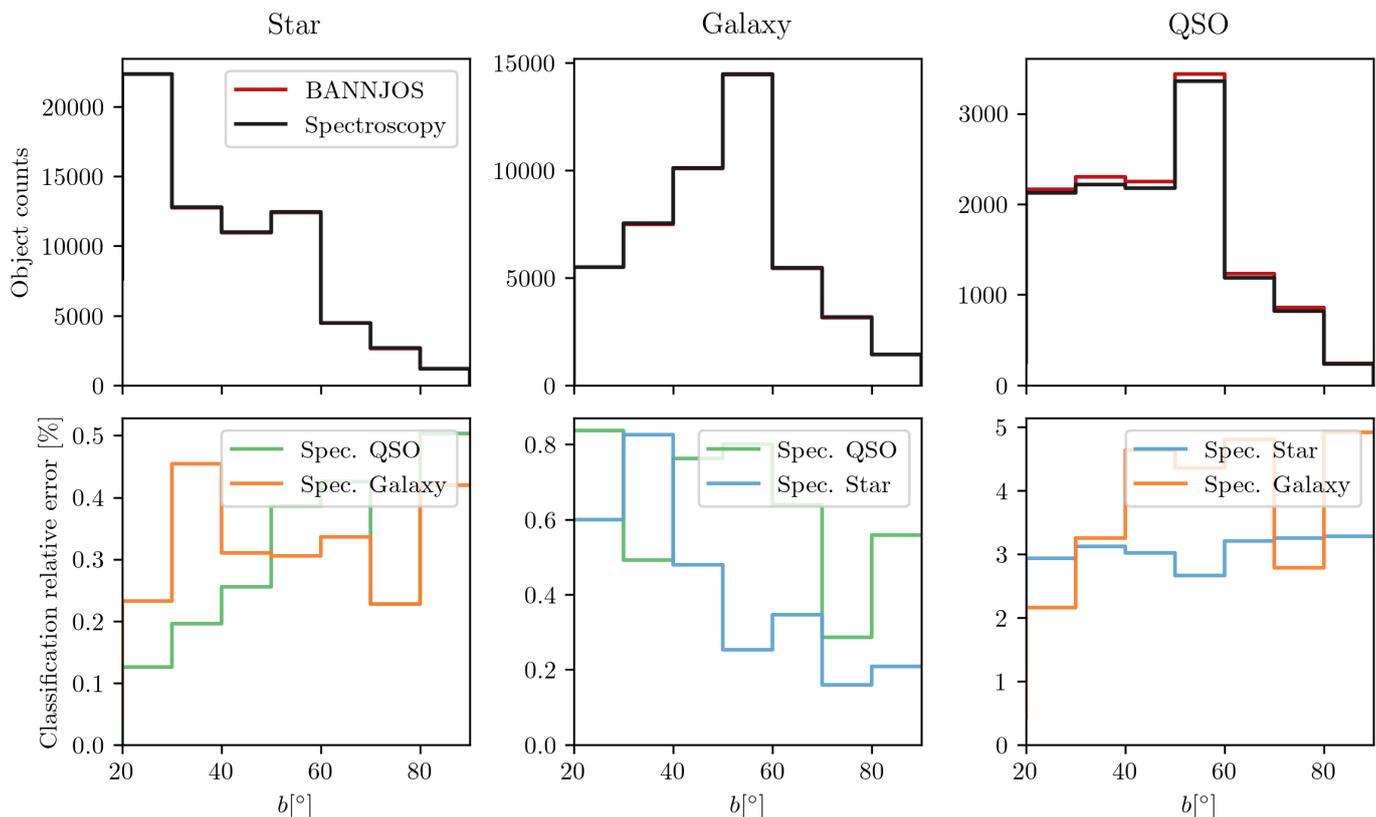
**Fig. 10.** Sky maps illustrating the surface density and expected contamination ratios for each class from the test sample. Panels are arranged as in a confusion matrix, with true classes,  $y_{\text{true}}$ , on the vertical axis and predicted classes,  $y_{\text{pred}}$ , on the horizontal axis. Diagonal panels display the sky surface density of sources classified into each category—galaxies, QSOs, and stars—uniformly scaled in color. The off-diagonal panels exhibit the relative classification error for given true and predicted classes, calculated as the number of misclassified sources divided by the total number of objects in the true class. The off-diagonal panels share a distinct color scale from the diagonal ones. No apparent trends emerge based on sky position.

with the highest median probability, i.e.,  $class = \max[PC_{\text{Galaxy}}(50), PC_{\text{QSO}}(50), PC_{\text{Star}}(50)]$ . This approach, employed in Sections 3.2 and 4, does not account for prediction uncertainties. Leveraging the complete PDF for the three classes enables more refined object selection. For instance, applying specific probability thresholds can enhance the purity of the selected sample. Moreover, apart from the median probability, BANNJOS provides the mean probability and the PDF compressed through the GMM model. These metrics can be used to select classes at any given threshold.

If the model works well, the classification success ratio (true probability) should match the probability used to select the objects (predicted probability). Specifically, the proportion of objects truly belonging to a selected class should correspond to their assigned probability of class membership. This is shown in Figure 12, where the probability predicted

by BANNJOS using three different statistics is compared to the true probability for each class across ten probability bins ranging from 0 to 1. The true probability for each class and probability bin is calculated as the ratio of the number of objects correctly classified according to their BANNJOS predicted probability and their actual spectroscopic class to the total number of objects with that particular predicted probability.

The three statistical measures yield similar results, with higher predicted probability thresholds correlating to increased true probabilities. The difference between the predicted and true probabilities remains consistently below  $\sim 0.1$  across all cases, indicating a good level of accuracy. However, there are some departures from the one-to-one relation, especially when using the median of the probability, that we should point out. For example, BANNJOS appears to be slightly underconfident when classifying stars with low probability  $0.1 \lesssim PC_{\text{Star}}(50) \lesssim 0.5$ , and slightly overconfident when clas-



**Fig. 11.** Number counts and contamination ratios as function of Galactic latitude,  $b$ . *Top panels:* Object counts for each class; Star, Galaxy QSO, from left to right; of object classified spectroscopically (black) and by BANNJOS (red). *Bottom panels:* Contamination ratios (false positives) in each class predicted by BANNJOS. The error is computed as the number of objects incorrectly assigned to each class by BANNJOS (one of the other two potential classes), divided by the number of objects of the spectroscopic class, i.e. Star, Galaxy QSO, from left to right. Bins are computed in  $10^\circ$  increments from  $20^\circ$  to  $90^\circ$  for each class. The bin  $[10^\circ, 20^\circ)$  is excluded from the analysis due to its very low number of galaxies (47).

sifying galaxies at  $0.5 \leq PC_{\text{Galaxy}}(50) \leq 0.9$ . This might result in purer and more contaminated samples of stars and galaxies, depending on the used probability threshold for the median. In contrast, the mean and the reconstructed mean from the GMM model appear to adhere more closely to a one-to-one relation. Figure 12 also highlights the probabilistic nature of our findings, with shaded areas representing the sample purity using the 16th and 84th percentiles of the PDF for source selection and the confidence regions derived from the GMM model.

We further investigate the effect of the selection threshold in Figure 13, which illustrates the impact of varying selection thresholds on the 2nd, 16th, 50th, 84th, and 98th percentiles of the cumulative PDFs,  $x$ , for each class:  $PC_{\text{Galaxy}}(x)$ ,  $PC_{\text{QSO}}(x)$ ,  $PC_{\text{Star}}(x)$ . These curves, similar to the PR curves in Figure 8, demonstrate the purity-completeness trade-off based on the selection threshold.

Increasing the probability threshold typically yields purer samples but reduces completeness. The chosen percentile for selection significantly affects the outcome. For example, selecting objects based on the 98th percentile being a QSO with  $\geq 0.5$  results in a sample with approximately 87% purity and 98% completeness. Opting for the 2nd percentile increases purity to roughly 98% at the expense of dropping completeness to about 91%. Setting higher thresholds further improves sample purity but invariably reduces completeness. Achieving over

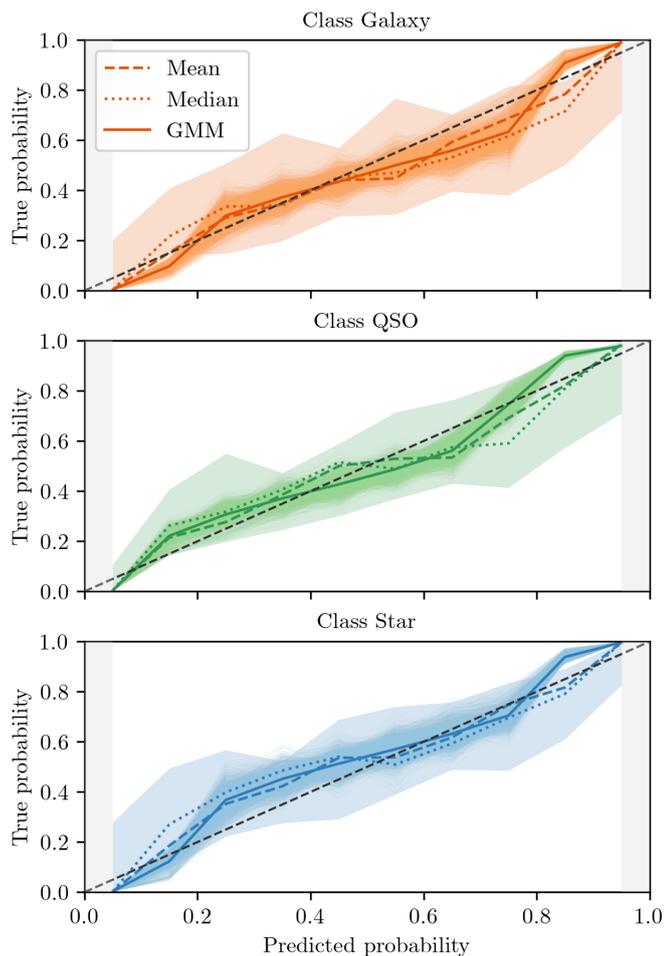
99% purity in QSO selection is possible by selecting sources whose 2nd percentile exceeds 0.8. Conditions for galaxies and stars are more lenient, with the model classifying them with greater ease. For instance, a purity above 99.5% is achievable for stars with the 2nd percentile of their PDF exceeding 33%.

Combining probabilities for the three classes enables the creation of purer samples. Typically, high purity and substantial completeness are attainable by setting the 16th and 84th percentiles above or below the random classification probability ( $1/3$  in this case). By combining these  $1\sigma$  confidence intervals, we can define the object's  $1\sigma$  class as follows:

- Star  $\iff PC_{\text{Galaxy}}(84) < 1/3 \ \& \ PC_{\text{QSO}}(84) < 1/3 \ \& \ PC_{\text{Star}}(16) > 1/3$
- Galaxy  $\iff PC_{\text{Star}}(84) < 1/3 \ \& \ PC_{\text{QSO}}(84) < 1/3 \ \& \ PC_{\text{Galaxy}}(16) > 1/3$
- QSO  $\iff PC_{\text{Galaxy}}(84) < 1/3 \ \& \ PC_{\text{Star}}(84) < 1/3 \ \& \ PC_{\text{QSO}}(16) > 1/3$

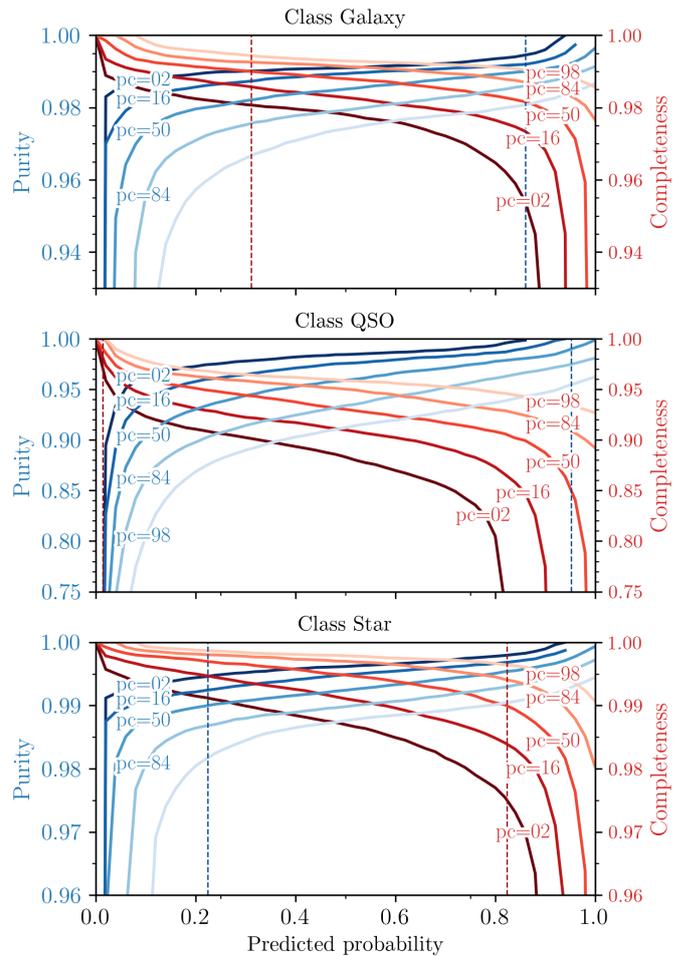
Adopting the stricter  $2\sigma$  confidence intervals, that is, the 2nd and 98th percentiles instead of the 16th and 84th, yields even purer but less complete samples.

In Figure 14, confusion matrices between  $y_{\text{true}}$  and  $y_{\text{pred}}$  are presented for sources selected using the  $2\sigma$  criteria for different magnitude ranges. Comparison with Figure 9 reveals enhanced sample purity, particularly for QSO classification, which achieves 91% accuracy in the  $21.5 < r \leq 22.5$  mag



**Fig. 12.** Predicted versus true probability for the selection based on different thresholds for each class: Galaxy (top panel), QSO (middle panel), and Star (bottom panel). For each class, the true probability is derived as the amount of objects with confirmed spectroscopic class divided by the total number of objects in each probability interval. The probability intervals are measured on the predicted probability from 0 to 1 in steps of 0.1 (10 in total). The predicted probability is the probability derived with BANNJOS using three different methods: The mean probability, the median probability, and the mean probability from the reconstructed PDF using the GMM model. The light shaded areas show the true probability of the sample if the 16th or the 84th percentile is used to select the sample. The uncertainties from the GMM model are shown by darker shaded areas and are computed by sampling the model 2000 times. The diagonal dashed line shows the one-to-one relation. Due to the binning used to create the figure, there is only data in the [0.05, 0.95] range.

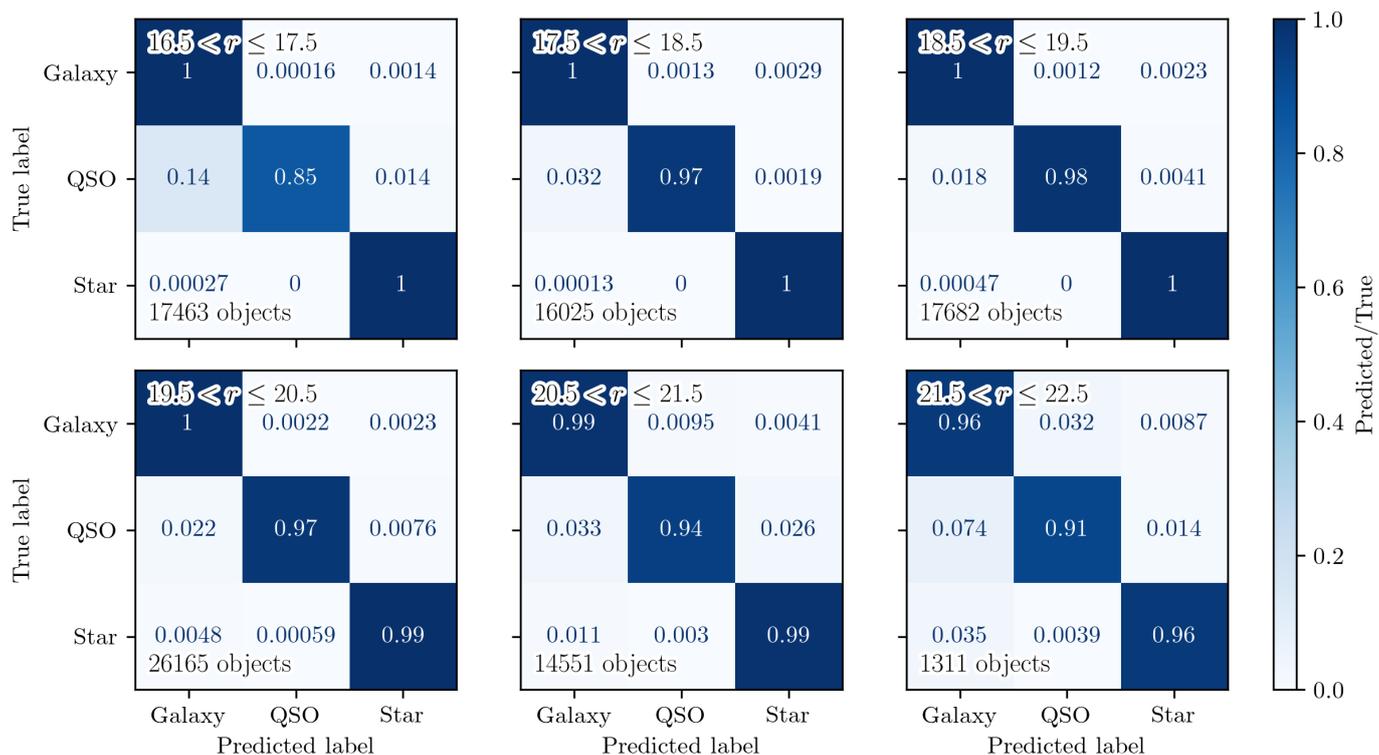
range and 96% for the other classes. However, this precision is gained at the expense of excluding low-confidence classified sources, evidenced by the reduced count of objects in the same magnitude bin (only 1311 from 1744 objects). The persistent galaxy-QSO confusion at  $r \gtrsim 18$  mag corroborates that these objects are confidently classified by BANNJOS, suggesting that variation in classification stems from the disparate spectroscopic surveys constituting the training and test sets. The drop from 15% to 7% in galaxy-QSO misclassifications at fainter magnitudes supports this notion, as most active galaxies at such depths are identified as QSOs by the employed surveys.



**Fig. 13.** Purity versus completeness for the selection based on different thresholds for each class: Galaxy (top panel), QSO (middle panel), and Star (bottom panel). The curves represent purity (blue) and completeness (red) of the sample as functions of the chosen threshold and percentile used to select the class. Vertical dashed lines in corresponding colors indicate the probability thresholds required to achieve 99% purity (blue) and 99% completeness (red), respectively.

The three selection strategies discussed serve merely as examples of the possible approaches. Higher (lower) selection thresholds can produce more pure (more complete) samples. Even more refined object selections can be achieved by exploiting the full covariance matrix and various correlation coefficients between the classes. For instance, active galaxies could be identified as objects with high probabilities of being either a galaxy or a QSO, accompanied by a significant negative correlation between these probabilities, as discussed in Section 3.3. If very pure samples are required, the correlation coefficients between probabilities can be useful too, by selecting sources with low uncertainties and little degeneration between species. Lastly, since the full PDF is recoverable from BANNJOS's output, users can also opt to use it for weighting the PDFs, finding particular objects, etc.

In Appendix D, we provide examples on how to query BANNJOS data to obtain pure samples of specific objects. In particular, we demonstrate how to select a pure sample of QSOs that could be candidates for spectroscopic follow up.



**Fig. 14.** Confusion Matrices between the true class and the predicted class for objects in the test sample for different magnitude bins selected with the  $2\sigma$  criteria. Markers and symbols coincide with those in Figure 9. The accuracy of the classification is greatly improved when using the predicted uncertainties to select high confidence sources.

## 6. Comparisons with other classifiers

At the moment of writing this paper, there are three other classifications available for J-PLUS. All of them are deterministic, meaning that they provide a single value for the probability of each object belonging to a certain class. The most important difference between them lies in the basis of their classification, morphological or nature-based, and thus in the number of classes they manage. In the following subsections, we compare BANNJOS's predictions to these previous works by using the highest median as the assigned class.

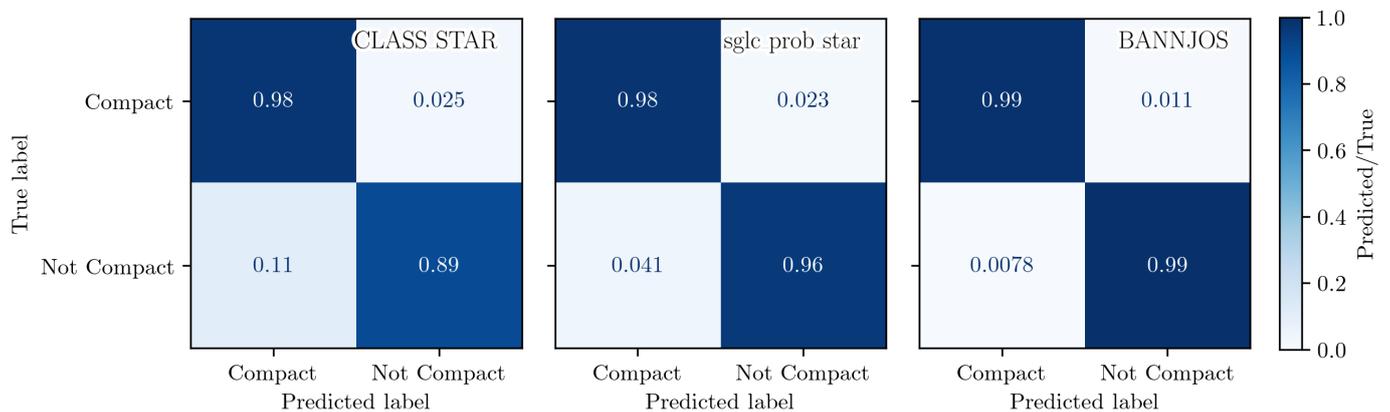
### 6.1. two-classes classifiers

Two classifications are currently available along J-PLUS tables at CEFCA's archive. The first, CLASS\_STAR, is derived from the SExtractor photometry package. The second, sglc\_prob\_star, introduced by López-Sanjuan et al. (2019), applies Bayesian priors to enhance the classification accuracy beyond that provided by SExtractor. Both classifiers categorize sources into two morphological classes: compact or extended, thereby assigning the probability of a source being star-like. However, this binary scheme complicates direct comparisons with our three-class approach. To address this, we recategorized the true and predicted classes from our test sample ( $y_{\text{true}}$  and  $y_{\text{pred}}$ ) into these two morphological categories. We classified objects as compact if their CLASS\_STAR, sglc\_prob\_star scores, or their median  $PC_{\text{Star|QSO}}(50)$  exceeded 0.5. Conversely, objects were deemed extended other-

wise. This approach enables direct comparison between our results and those of the pre-existing binary classifiers.

Figure 15 presents confusion matrices for these morphological classification derived from the three classifiers. All three classifiers perform commendably in classifying compact sources, with correct classification rates of approximately 98%. However, CLASS\_STAR exhibits notable shortcomings in identifying extended sources, correctly classifying only about 89% as galaxies. The sglc\_prob\_star classifier significantly improves upon this, correctly identifying around 96% of galaxies. When examining the misclassification of galaxies as compact (star-like) sources, sglc\_prob\_star shows a threefold improvement over CLASS\_STAR. Nevertheless, BANNJOS demonstrates superior accuracy, significantly reducing the misclassification rate of extended objects compared to sglc\_prob\_star.

The left part of Figure 16 (blue) illustrates the classification error rates for the three classifiers across various magnitude bins. Errors for each classifier are quantified as the ratio of misclassified objects to the total number of objects per bin, and are further categorized into False Positives (FPs) and False Negatives (FNs). Since only two classes are involved, the FNs in compact sources correspond to the FPs in extended sources. BANNJOS consistently outperforms the other classifiers across all bins, maintaining lower error rates. This holds true except possibly for FPs (Compact) at the faintest magnitudes, where the error rates for all three classifiers converge. A notable aspect in this comparison is the significant asymmetry displayed by the sglc\_prob\_star and CLASS\_STAR classifiers between FPs and FNs, indicating a



**Fig. 15.** Confusion matrices comparing the true class (spectroscopic) against the predicted class for objects in the test sample up to  $r = 21.9$  mag, corresponding to the median 50% completeness threshold of J-PLUS for compact sources across 1,642 tiles, for three different classifiers. These matrices are based on the assumption that QSOs and stars are compact sources, while galaxies are extended. A source is considered compact if its corresponding CLASS\_STAR, sg1c\_prob\_star scores, or its median  $PC_{\text{Star|QSO}}(50)$  exceeds 0.5. Labels coincide with those of Figure 9.

tendency to misclassify QSOs or stars as extended sources as the S/N decreases, with CLASS\_STAR being the most pronounced in this regard.

Adding FPs and FNs yields the total classification error, where BANNJOS outperforms the other two classifiers by a significant margin. For example, CLASS\_STAR incorrectly classifies approximately 30% of objects beyond  $r > 20.5$  mag, resulting in a cumulative error rate of about 11% at  $r = 22.5$  mag. In contrast, sg1c\_prob\_star reduces the error rate to approximately 12-13% at similar magnitudes, with an accumulated error of about 5% at  $r = 22.5$  mag. BANNJOS, however, achieves an error rate of merely around 3.5% in the same range, with a cumulative error below 2% for all objects. Remarkably, BANNJOS maintains near-perfect performance for objects brighter than  $r = 19.5$  mag, with error rates around 0.3% and cumulative errors below 0.1%. While the purity of the classification could potentially be enhanced by adjusting the probability thresholds for the three classifiers, our experiments indicate that the relative differences between the classifiers remained almost unchanged, or favored BANNJOS even more.

## 6.2. Three-classes classifiers

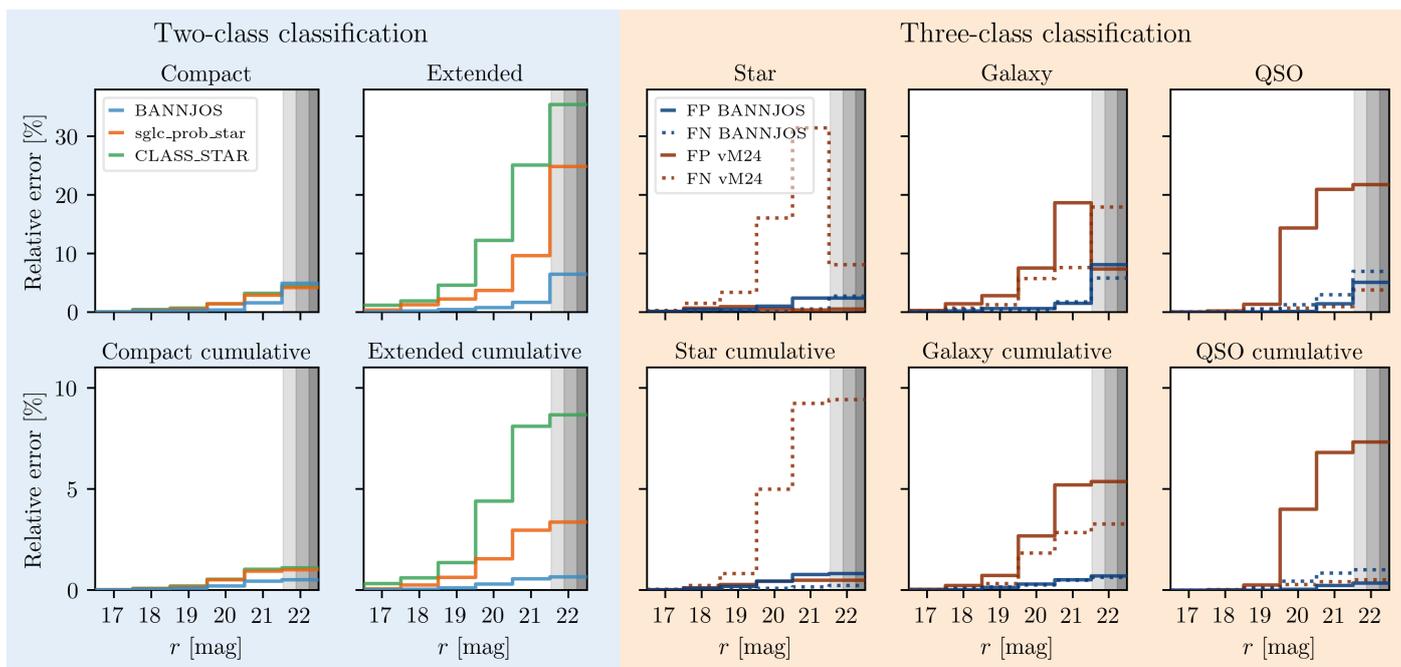
Differently from CLASS\_STAR and sg1c\_prob\_star, the classification introduced in von Martens et al. (2024), hereafter referred to as vM24, bases its results not only on the morphology of the source but also on its colors. This approach, coupled with the utilization of a more sophisticated algorithm, XGBoost, enables the authors to differentiate sources based on their inherent nature, adding the QSO class to their classification. Consequently, their classification encompasses three categories: Galaxy, QSO, and stars; making it a natural counterpart for comparison with our classification.

A new test sample is required for this comparison, one that includes only objects never seen during the training phases of any of the models, i.e., BANNJOS and XGBoost. This new test sample was composed by cross-matching our own test sample with the training sample used in vM24, retaining only sources present in our test sample but absent in their training sample. From our original test sample of 136,570 objects, we

identified 52,105 sources not included in the vM24 training sample. While significantly reduced, this number of objects is still sufficiently large to conduct a comparative analysis. This dataset, composed entirely of sources unfamiliar to both models, allows us to assess their general performance on independent data, thus facilitating a fair comparison between the two. As in previous sections, we assign the class predicted by BANNJOS as the one with the highest median probability,  $\max[PC_{\text{class}}(50)]$ . In vM24, the classification is presented similarly to that of BANNJOS, albeit with a single probability value per class. We assigned the class in vM24 as the one with the highest probability. While simple, this is the only criterion ensuring 100% completeness in the sample selection.

Following the analysis from the previous section, we proceeded to bin the new test sample according to the  $r$ -magnitude of its sources. We then computed the relative and cumulative errors as in the left sub-figure of Figure 16, i.e., the relative classification error per class for each classifier is defined as the ratio between misclassified objects and the total number of objects, determined by their spectroscopic class, within each magnitude bin. The results are shown in the right part of Figure 16 (orange), where upper panels detail the relative FP and FN errors for each classifier by class, and bottom panels present cumulative errors for FPs and FNs.

BANNJOS significantly outperforms the performance presented in vM24 across all magnitude bins, with accuracies varying from 2 to 10 times better, depending on the class and magnitude bin. The higher error rate of vM24 is clearly visible as a generally higher rate of FPs and FNs in the vM24 results compared to those of BANNJOS. For instance, the typical total classification error (FPs+FNs) for Galaxies at  $r = 21$  mag is around 27% in vM24, while BANNJOS reduces this to merely 3%. These differences diminish at the faintest magnitudes, where BANNJOS still typically yields classification accuracies twice as higher as those obtained in vM24. To provide some figures, the average total classification error for QSOs in vM24 is approximately 26% at  $r \sim 22$ , compared to around 12% for BANNJOS, while cumulative errors typically escalate to 8% at  $r \sim 22.5$  in vM24 but remain at  $\sim 1.5\%$  for BANNJOS.



**Fig. 16.** Classification error rates per magnitude bin for all evaluated classifiers. The magnitude bins coincide with those in Figure 8, and error rates are defined as the ratio of objects incorrectly classified as False Positives (FPs) or False Negatives (FNs) to the total number of objects within each bin. *Left sub-figure (Two-class classification, light blue):* The error rates for compact objects, specifically galaxies objects misclassified as stars or QSOs (left) and vice versa (right), are presented. Each classifier is denoted by a unique color: BANNJOS in blue, `sglc_prob_star` in orange, and CLASS\_STAR in green. An object is considered compact if its corresponding CLASS\_STAR or `sglc_prob_star` score exceeds 0.5, or its median  $PC_{\text{Star|QSO}}(50)$  is greater than  $1/3$ . The statistics are obtained using the test sample of 136,570 objects. *Right sub-figure (light orange):* Shows the classification error rate for the three classes available in BANNJOS (dark blue) and von Martens et al. (2024) (vM24, dark orange). FPs are depicted with solid lines, while FNs are shown with dotted lines in the respective colors. In this case, the error classification ratios are obtained with another test sample of 52,105 sources never seen during the training phases of the two models. The distribution of errors over cumulative magnitude bins is shown in the corresponding bottom panels. The median 90%, 50%, and 10% completeness levels for J-PLUS compact sources are indicated by vertical gray shaded areas, with the darkest shade representing the 10% level. BANNJOS surpasses all other classifiers by a large margin at all magnitude range, both in total classification error and in symmetry between FPs and FNs.

Asymmetries between FPs and FNs could indicate a bias in the model’s predictions towards a specific class. Notably, stars exhibit a very high rate of FNs in vM24, which seems to translate into FPs for galaxies and QSOs, indicating that at magnitudes fainter than  $r \sim 18$  mag, the XGBoost used in vM24 tends to misclassify a significant number of stars as galaxies and QSOs. Confusion also seems to occur between the galaxy and QSO classes at the faintest magnitudes. This effect is also observable in BANNJOS’s results, albeit with a lower error rate and more symmetry between FPs, FNs, and the classes themselves. However, as discussed in Section 4.1, this is expected and likely caused by the dual nature of active galaxies.

We conducted an additional test using a refined sample, specifically selecting sources classified with over 95% probability in vM24 in any of the three classes. This criterion reduced the sample size to 37,355 objects (approximately 72% of the original sample) that are classified with high confidence in vM24. Upon reevaluating both models with this high-probability sample, improvements were noticeable, although BANNJOS continued to exhibit superior performance across all magnitude bins. In vM24, typical error rates peaked at around 8% at  $r \sim 21$  mag for both stars and galaxies, with QSOs remaining lower at 2% at the same magnitude. However, the error distribution between these classes remained very asymmetric, with classification errors for stars being almost entirely

FNs and for galaxies mostly FPs. The high asymmetry shown between the galaxy and star classes at magnitudes fainter than  $r \sim 19$  mag indicates that the vM24 classification systematically misclassifies stars as galaxies at these magnitudes when applying high-probability selection criteria. Interestingly, the numbers of FPs and FNs are consistent for QSOs under this high-probability selection. In contrast, BANNJOS exhibits its maximum error at the faintest magnitudes,  $r = 22.5$  mag, at around 2% for galaxies and around 1% for the other classes, demonstrating higher consistency between FPs and FNs across the entire magnitude range. In summary, using this particular high-confidence sample from the vM24 catalog, BANNJOS still outperforms XGBoost significantly.

The largest difference between the training of the models stems from the addition of DESI data to our training list. The test sample used for this comparison also contains sources classified by DESI, which could potentially represent an advantage for BANNJOS. To test whether the presence of these sources in the test sample was beneficial to BANNJOS, we removed all the DESI-only-measured sources from the list, leaving a total of 31,124 sources, and repeated the experiment. While BANNJOS’s errors remained the same or even decreased in some cases, the classification errors in vM24’s results increased, particularly in the star and galaxy classes.

All previous tests were performed covering scenarios going from neutral to favorable to vM24. In a last test, instead of

selecting objects based on  $vM24$  scores, we repeated the experiment using the  $2\sigma$  criteria presented in Section 5, including sources classified by DESI. This criterion, which could potentially be favorable to BANNJOS, yielded 51,214 sources (approximately 98% of the original). As expected, the results using this list show a significant improvement in the classification accuracy of BANNJOS (with cumulative errors well below 1% for the three classes) while still maintaining a large portion of the sources. In contrast, the XGBoost model used in  $vM24$  yielded slightly worse results than those obtained using the entire sample, potentially due to the marginal reduction in sample size without a corresponding decrease in the number of misclassified objects.

In Appendix E, we offer a comparison between the four classifiers, i.e. CLASS\_STAR,  $sglc\_prob\_star$ ,  $vM24$ , and BANNJOS, based on ROC curves for this new test sample of 52,105 sources.

## 7. Results and Statistical Validation

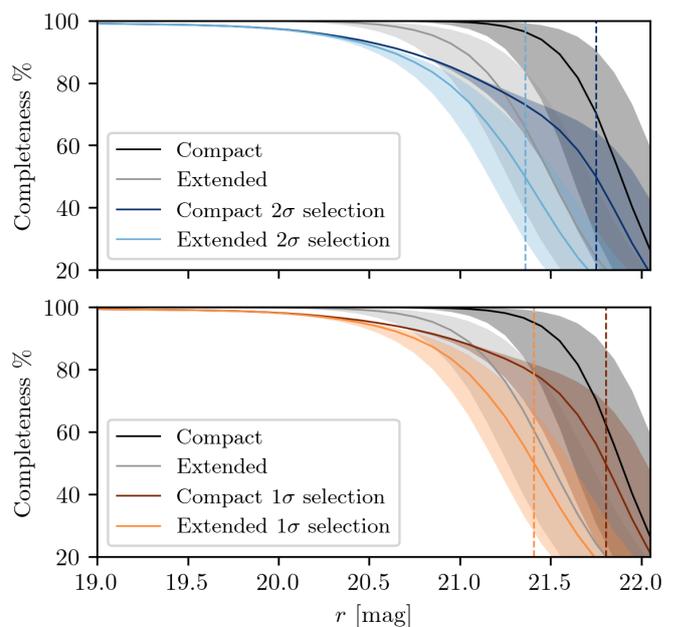
In the previous section, Section 4, we validated the classification using the test sample, composed of sources with a spectroscopic classification *never seen* by our model. The test sample represents a valuable asset for validating our classification. However, it is still a subset of the training set, hence the model might have certain advantages when classifying its objects. For instance, the observed properties of sources in both the test and training sets are similar, and the S/N is relatively high for most sources. This scenario could lead to overly optimistic validation results, as the model might merely learn to reproduce the classifications for the types of objects found in the training set.

To further validate the classification performance of BANNJOS, we compared its predictions with those from previous independent works across the entire J-PLUS DR3 catalog. We employed BANNJOS to classify all 47.4 million objects in J-PLUS DR3 into stars, QSOs, and galaxies. In the following subsections, we statistically analyze the classification results for the entire J-PLUS dataset.

### 7.1. Completeness

Initially, we examine the completeness of the classification based on the three basic selection criteria presented in Section 5. Figure 17 shows the average completeness of J-PLUS as a function of  $r$  magnitude for the three criteria: highest median probability,  $1\sigma$ , and  $2\sigma$ . The completeness is calculated as the number of sources remaining after classification divided by the total number of sources, then multiplied by the actual photometric completeness of J-PLUS at each specific magnitude. Thus, the curves represent the expected real completeness for compact and extended sources at a given magnitude  $r$  when using the BANNJOS classification with these selection criteria. We avoid detailed calculations for the three individual classes since we only have access to the classes predicted by BANNJOS, which could contain misclassified objects, especially at fainter magnitudes.

Using the highest median value of the PDFs ensures no object is rejected, hence the completeness remains equivalent to the photometric completeness of J-PLUS for both compact and extended sources (dark and grey curves, respectively). However, selecting sources based on their confidence intervals leads to the exclusion of uncertain sources. As expected,

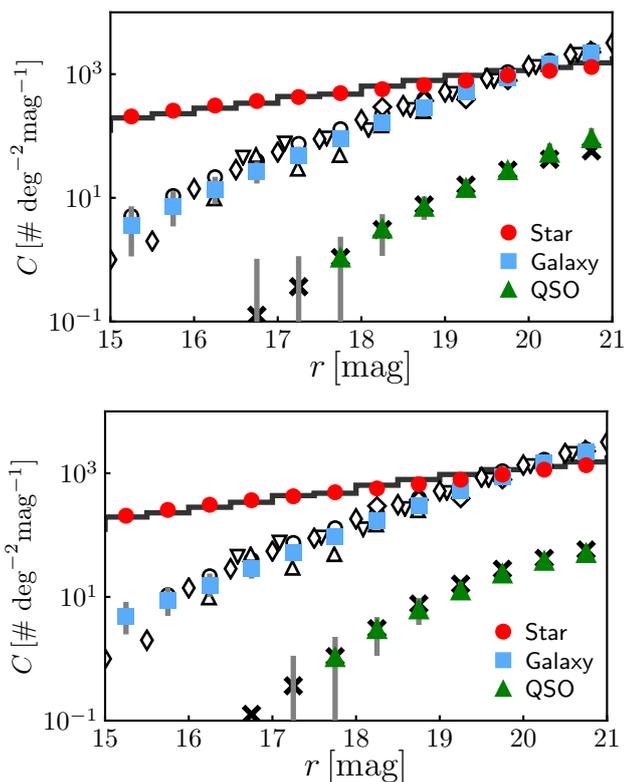


**Fig. 17.** Final completeness of the J-PLUS survey using different selection criteria in BANNJOS. The black and grey curves represent the median photometric completeness of J-PLUS for compact and extended sources, respectively, measured across its 1,642 tiles in 0.1 mag wide magnitude bins. This is equivalent to the completeness achieved when selecting sources based on BANNJOS’s maximum median probability. The shaded areas indicate the 16th and 84th percentiles of this completeness. The orange curves and their corresponding shaded areas in the bottom panel illustrate the completeness for compact and extended sources using  $1\sigma$  significance in their Probability Density Functions (PDFs). The blue curves in the upper panel represent the completeness using the  $2\sigma$  selection criteria (see Section 5). Vertical dashed lines indicate the 50% completeness level for each configuration, and are colored correspondingly.

the number of sources excluded depends on the visual morphology of the sources and the considered magnitude range. In general, the  $1\sigma$  selection is less complete than using the maximum median value. Yet, completeness values above 95% are still achievable at  $r \approx 20.3$  mag, and 90% at  $r \approx 21.2$  mag for compact sources. The selection based on  $2\sigma$  confidence intervals is more restrictive, placing the same levels of completeness at  $r \approx 20.0$  and  $r \approx 20.5$  mag, respectively.

### 7.2. Number Counts

Table 3 presents the final numbers of objects classified by BANNJOS: 47,463,878, divided into galaxies, QSOs, and stars using the three different criteria presented in Section 5. As criteria become more restrictive, fewer objects are classified with sufficient confidence. For instance, of the approximately  $2.0 \times 10^7$  galaxies observed in J-PLUS and classified with the  $\max[PC_{\text{class}}(50)]$  criterion, only about  $1.7 \times 10^7$  are classified as such with a  $2\sigma$  confidence. This effect is most pronounced for QSOs, where fewer than half of those classified with the  $\max[PC_{\text{class}}(50)]$  criterion meet the high confidence threshold of the  $2\sigma$  criteria. Stars, however, suffer the least impact from poor S/N; thus, the three classification criteria produce nearly the same number of stars.



**Fig. 18.** *Top panel:* Number counts for stars (red dots), galaxies (blue squares), and QSOs (green triangles) as a function of  $r$ -band magnitude, estimated as the median of the 1642 J-PLUS DR3 pointings, using the maximum median probability criteria to assign classes. Vertical gray lines show the variance in number counts between pointings. Black solid histograms represent stellar number counts at the median pointing position, estimated using the TRILEGAL model of the Milky Way (Girardi et al. 2005). Open symbols denote galaxy number counts from literature sources: Yasuda et al. (2001, circles); Huang et al. (2001, triangles); Kümmel & Wagner (2001, inverted triangles); and Kashikawa et al. (2004, diamonds). Black crosses show QSO number counts as predicted by Palanque-Delabrouille et al. (2016). *Bottom panel:* Similar to the top panel, but applying the more restrictive  $2\sigma$  selection criteria.

**Table 3.** Object counts for the three different criteria discussed in Section 5

Class	$\max[PC_{\text{class}}(50)]$	$1\sigma$	$2\sigma$
Galaxies	20,479,247	18,466,693	17,391,099
QSOs	997,436	508,440	404,010
Stars	25,987,195	24,712,265	24,175,950
Total	47,463,878	43,687,398	41,971,059

BANNJOS’s classification should yield number counts consistent with expectations from models and previous works. We computed the number counts for each object type as a function of  $r$  magnitude using the  $\max[PC_{\text{class}}(50)]$  criterion, which includes all sources defining the class as the one with the highest median probability. We defined the  $r$ -band number counts in each J-PLUS pointing as the Probability Density Function (PDF)-weighted histogram normalized by area and magnitude (see Eq. 17 in López-Sanjuan et al. 2019). The median and dispersion of counts from the 1642 pointings were then computed. We present the J-PLUS DR3 stellar, galaxy, and QSO number counts in Fig. 18. The stellar number counts match the TRILEGAL model (Girardi et al.

2005) predictions for the Milky Way at the median (RA, Dec) of the J-PLUS DR3 tiles, despite a large dispersion reflecting the variation in stellar density across the surveyed area. The galaxy number counts is consistent with literature results. However, QSO number counts exhibit an excess over expectations from Palanque-Delabrouille et al. (2016) at  $r > 20$ , with approximately 30% and 60% more QSOs at  $r = 20.25$  and  $20.75$  mag, respectively. Notably, QSOs are 100 times less numerous than galaxies and stars at faint magnitudes, and even a low misclassification rate can significantly overestimate the number of QSOs.

Repeating the analysis with a more stringent  $2\sigma$  selection scheme for QSOs (bottom panel), we found that while stellar and galaxy number counts remained unchanged, QSO number counts at faint magnitudes decreased. In this case, the QSO counts are at 90% of the expectations for sources within  $19 < r < 21$  mag. This value is comparable to the predicted 91% completeness for the  $2\sigma$  selection outlined in Section 5, suggesting that the selection could indeed be of high purity. This underscores the utility of classification PDFs beyond the best solution, illustrating that BANNJOS can statistically provide accurate densities of stars, galaxies, and QSOs for  $r < 21$  mag.

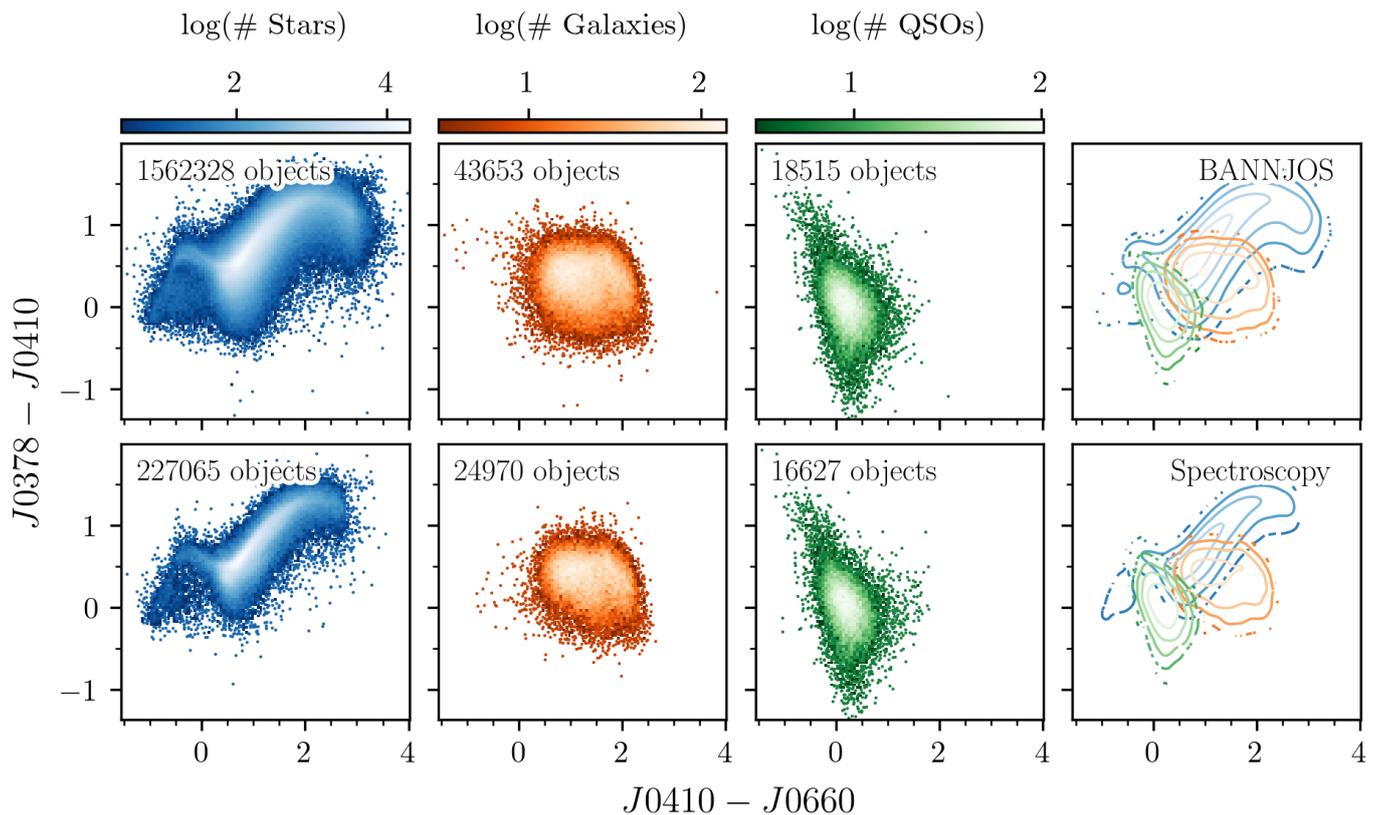
It is also important to point out that BANNJOS was trained using a list that is balanced, containing an equal number of objects for each class. The fact that BANNJOS accurately estimates the number counts for the three classes showcases its ability to distinguish between classes without propagating biases regarding the expected number of objects.

### 7.3. Color-Color Diagrams

In this section, we present color-color diagrams for all sources within the J-PLUS dataset that meet the  $2\sigma$  selection criteria. It is important to note that colors were not directly employed in the model’s training process. Consequently, observable color distinctions between different types of sources would indicate that the model has autonomously learned the relation between the J-PLUS photometric bands for each category. Thus, these diagrams act as additional validation of the classification. Ideally, a proficiently trained model would accurately identify the specific loci associated with each class; as a result, each class should occupy distinct areas within the color-color diagrams.

The top panels of Figure 19 display the  $J0378 - J0410$  versus  $J0410 - J0660$  color-color diagram<sup>9</sup> for the three classes selected with BANNJOS using the  $2\sigma$  criterion. The distribution of objects from the training list, which are spectroscopically confirmed, is shown in the bottom panels. For clarity, we limited both samples to only sources with photometric uncertainties and reddening below 0.25 mag. We also excluded sources with non-zero photometric flags, which would indicate poor measurements. Up to 1,624,496 sources predicted by BANNJOS (top panels) and 268,662 in the spectroscopic sample (bottom panels) met these criteria. BANNJOS appears to perform well in separating the three classes: stars, galaxies, and QSOs occupy their expected regions. Although some overlap is anticipated, the highest concentration for each individual class differs (rightmost panels). No significant trends in the isodensity contours suggest major misclassification from one species to another. The consistency of the locus of the

<sup>9</sup> Colors were chosen randomly between those showing a more obvious separation between classes, for illustrative purposes.



**Fig. 19.** Color-color diagrams for the three available classes. Stars, Galaxies, and QSOs are shown in blue, orange, and green respectively, with their distributions in the left, middle-left, and middle-right columns. The color shading corresponds to object density on a logarithmic scale, computed separately for each class. The right column shows the concentration of objects for each class using correspondingly colored contours. The top row shows BANNJOS classification for the entire J-PLUS catalog based on the  $2\sigma$  selection criterion. The bottom row shows the spectroscopic classification in the training sample. For clarity, we restricted the sample to sources with photometric uncertainties and reddening below 0.25 mag, and no photometric flags: 1,624,496 sources predicted by BANNJOS (top panels) and 268,662 in the spectroscopic sample (bottom panels). The distribution of sources is very similar, indicating that BANNJOS is effectively recovering the object classes even though colors were not used during model training.

three different classes, classified by BANNJOS and by spectroscopy, indicates that our classifier can provide reasonably clean samples, comparable even to those obtained with spectroscopy.

Due to the imposed photometric quality cuts, Figure 19 only includes sources with relatively high S/N. In fact, results using the  $1\sigma$  and the  $\max[PC_{\text{class}}(50)]$  criteria are nearly identical for this subset, as the limiting factor is the quality of the photometry. Relaxing the criteria for maximum allowed photometric error and reddening results in a more blurred locus for all three cases, yet consistent between the spectroscopic and photometric samples. Even without any quality selection in the photometry, the results remain consistent between BANNJOS and the spectroscopic sample, though loci become much more blurred for BANNJOS objects, which include very low S/N sources. In this scenario, the number of sources classified by BANNJOS varies according to Table 3.

## 8. Caveats

### 8.1. Peculiar Tiles

During a thorough analysis of our results, we noticed that certain Tiles produced number counts far from expectations. We checked them in detail and we provide the list here for the shake of the user:

- Tiles 95100, 95118, 95126, 95136, 95162, 95175, 95186, 95197. The number counts for stars are lower than expected at  $r < 17$  mag. These Tiles have a factor of 10 larger exposure time in the  $r$  band than the rest of the Tiles, producing saturated stars to fainter magnitudes.
- Tile 98155. An excess of galaxies and a lack of stars are observed at  $r > 18$  mag. We noticed a relatively high photometric noise in the all the narrow bands that could be affecting the prediction.
- Tile 102263. An excess of both QSOs at all magnitudes and galaxies at  $r > 18$  mag is observed. This pointing is dominated by M33, and the detection of structures of the galaxy as individual sources produces the observed overestimation.
- Tile 103605. Idem, but for M31.

– Tiles 103874, 103884, 103908, 103919, 103930. A lack of galaxies and QSOs at  $r > 16$  mag is observed. These Tiles have the lower galactic latitude in the survey and the larger density of stars. We found that the detection from SExtractor is missing a large amount of faint sources and performed a deficient deblending in several cases.

We recommend using caution when utilizing data from these Tiles and leveraging the information from all other available classifiers.

### 8.2. Multiple Entries

We added data from other all-sky surveys into our data sets, including astrometry and photometry from the *Gaia* third data release and the CatWISE2020 catalog. The match between catalogs is offered along other scientific tables at J-PLUS data portal and is based on the sky positions of the sources, with a maximum match radius defined by the standard PSF of the surveys. Due to varying resolving powers of the surveys and atmospheric effects in J-PLUS, multiple matches are sometimes possible, especially for *Gaia*, where the standard match radius of 1.5 arcsec allows more than one *Gaia* source to be matched to a single J-PLUS source. Similar issues arise with CatWISE2020, whose wider PSF can result in multiple J-PLUS objects being matched to a single source. This leads to some objects having multiple entries in the training and output catalogs, potentially resulting in different classifications as the feature vector,  $\mathbf{x}$ , varies between entries (see Appendix A).

To be consistent with the J-PLUS archive and how cross-matches with external catalogs are provided, we decided to retain all possible classifications for any given object. This resulted in a list of 47,794,839 objects, 330,961 more than the original list. For some of these objects, the classification is consistent between entries. However, many vary depending on the matched source. We view this as beneficial: two different objects, blended by the J-PLUS PSF, can potentially be distinguished in such cases. Alternatively, users can select the most robust classification or simply exclude objects with multiple classifications.

### 8.3. Robustness of Results

We meticulously selected, configured, and trained BANNJOS to minimize potential biases and errors in its predictions. Despite our extensive cross-validation tests affirming its accuracy, machine learning models are limited by the quality of their target data,  $y_{\text{true}}$ , and the features used for training,  $\mathbf{x}$ . Observational errors in these quantities can degrade the model's predictive performance. Additionally, systematic errors likely affecting the training data can lead to propagated biases in the final predictions,  $y_{\text{pred}}$ .

Extensive testing was conducted to identify and analyze systematics in  $y_{\text{true}}$  within our training list. By construction, this list shows a large variety in terms of depth and color coverage. However, we found data to be generally consistent and of high quality. We manually reviewed the predictions from our best candidate models, which provided results qualitatively consistent between them, with our training lists, and previous studies. However, we cannot rule out the possibility of incorrect classifications by our method for given particular objects. Consequently, we advise users to exercise caution and utilize the covariance matrices and various percentiles to en-

sure sample purity, especially at low S/N, where the training and target samples differ the most.

## 9. Conclusions

In this study, we introduced BANNJOS, the Bayesian Artificial Neural Network for the Javalambre Observatory Surveys. As a versatile machine learning pipeline, BANNJOS leverages advanced deep Bayesian Neural Networks for regression tasks, proving itself as a comprehensive tool for analyzing and predicting a broad spectrum of numerical data. Specifically, we harnessed BANNJOS's capabilities to systematically classify an extensive dataset from the J-PLUS survey, categorizing an ensemble of 47.4 million astronomical objects into stars, QSOs, and galaxies. This classification is not merely categorical; BANNJOS furnished detailed probability distribution functions (PDFs) for each source across the three categories, thus facilitating a nuanced selection process based on probabilistic confidence, inter-class correlations, and more.

The training of BANNJOS was meticulously conducted using a list of approximately 1.2 million objects including approximately 430,000 galaxies (35%), 120,000 QSOs (9%), and 680,000 stars (55%), all with reliable classifications from SDSS DR18, LAMOST DR9, DESI EDR, or *Gaia* DR3. We employed data augmentation techniques to balance the training sample, and incorporated information from 445 variables to the final training list, including photometric information in the 12 filters of J-PLUS across 8 different apertures, morphology variables, masking flags, image quality variables, etc., alongside infrared photometric information from CatWISE2020 and astrometry from *Gaia*.

The results were validated using a test sample of approximately  $1.4 \times 10^5$  sources. BANNJOS demonstrated exceptional performance, with ROC AUC values greater than 0.99 for all six possible class combinations up to  $r = 21.9$  mag, which corresponds to the 50% completeness level in J-PLUS for compact sources. BANNJOS also performs well at fainter magnitudes, correctly classifying 90%, 81%, and 87% of galaxies, QSOs, and stars, respectively, at  $21.5 < r \leq 22.5$  mag. The total accumulated error falls below 2% for sources with  $r \leq 22.5$  mag, and about 0.9% at  $r \leq 20.5$  mag. BANNJOS significantly outperforms the three currently available classifiers in J-PLUS, CLASS\_STAR, `sg1c_prob_star`, and the classification presented in von Marttens et al. (2024) with relative classification errors between 8 and 4 times smaller at  $r \sim 22.5$  mag. Extensive tests revealed no significant biases or spatial trends.

We have used BANNJOS to classify all J-PLUS sources into around 20 million galaxies, 1 million QSOs, and 26 million stars. The resulting classification is consistent in number counts with results from previous works and model predictions. This consistency extends to magnitudes up to  $r \sim 20.5$  mag for the three classes using the simplest classification criterion. However, results significantly improve when applying more restrictive criteria, aligning in number counts with model predictions up to  $r \sim 21.5$  mag for the three species.

The full PDF provided by BANNJOS enables J-PLUS users to refine their object selection. We have demonstrated its potential by selecting objects using three different criteria, allowing for the creation of very pure samples, even at faint magnitudes. Utilizing the covariance matrix allows for even finer selection of sources, enabling users to identify species not considered in the original classification, such as partially resolved active galaxies.

As general-purpose regressor, BANNJOS can be utilized in a wide variety of scientific cases. For example to derive stellar chemical abundances, or photometric redshifts for galaxies or QSOs. Its potential will be further explored with the upcoming J-PAS survey, where the information from an astonishing amount of 56 colors will allow BANNJOS to investigate the nature of each source in great detail.

The value-added catalog with the classification and whose content is described in Appendix F is accessible at the the J-PLUS archive through CEFCA's catalogues portal<sup>10</sup>. This table, called ClassBANNJOS, has all the necessary information to reconstruct the full PDF, enabling astronomers to easily select sources based on their criteria and to resample BANNJOS' predictions at their convenience. BANNJOS and the code to decompress the PDF are publicly available at <https://github.com/AndresdPM/BANNJOS>.

*Acknowledgements.* Based on observations made with the JAST80 telescope and T80Cam camera for the J-PLUS project at the Observatorio Astrofísico de Javalambre (OAJ), in Teruel, owned, managed, and operated by the Centro de Estudios de Física del Cosmos de Aragón (CEFCA). We acknowledge the OAJ Data Processing and Archiving Unit (UPAD) for reducing and calibrating the OAJ data used in this work. Funding for OAJ, UPAD, and CEFCA has been provided by the Governments of Spain and Aragón through the Fondo de Inversiones de Teruel and their general budgets; the Aragonese Government through the Research Groups E96, E103, E16\_17R, E16\_20R and E16\_23R; the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033 y FEDER, Una manera de hacer Europa) with grants PID2021-124918NB-C41, PID2021-124918NB-C42, PID2021-124918NA-C43, and PID2021-124918NB-C44; the Spanish Ministry of Science, Innovation and Universities (MCIU/AEI/FEDER, UE) with grant PGC2018-097585-B-C21; the Spanish Ministry of Economy and Competitiveness (MINECO) under AYA2015-66211-C2-1-P, AYA2015-66211-C2-2, AYA2012-30789, and ICTS-2009-14; and European FEDER funding (FCDD10-4E-867, FCDD13-4E-2685). The Brazilian agencies FINEP, FAPESP, and the National Observatory of Brazil have also contributed to this project. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. F.J.E and P.C. acknowledge financial support from MCIN/AEI/10.13039/501100011033 through grant PID2020-112949GB-I00. C.H.M. also acknowledges the support of the Spanish Ministry of Science and Innovation via project grant PID2021-126616NB-I00. RvM is supported by Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) grant TO APP0039/2023. VM thanks CNPq (Brazil) and FAPES (Brazil) for partial financial support. MQ is supported by the Brazilian research agencies FAPERJ and CNPq. A. del Pino acknowledges the financial support from the European Union - NextGenerationEU and the Spanish Ministry of Science and Innovation through the Recovery and Resilience Facility project ICTS-MRR-2021-03-CEFCA and the project PID2021-124918NB-C41. A. del Pino also thanks Dr. Bertran de Lis for her support and help during the realization of this project. *Software:* *numpy* (Harris et al. 2020), *scipy* (Virtanen et al. 2020), *matplotlib* (Hunter 2007), *astropy* (Astropy Collaboration et al. 2013, 2018, 2022), *tensorflow* (Abadi et al. 2015)

## References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org  
 Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543  
 Almeida, A., Anderson, S. F., Argudo-Fernández, M., et al. 2023, ApJS, 267, 44  
 Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167  
 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123  
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, MNRAS, 404, 86  
 Ball, N. M., Brunner, R. J., Myers, A. D., & Tchenguiz, D. 2006, ApJ, 650, 497  
 Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393  
 Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, A&A, 622, A176  
 Cenarro, A. J., Moles, M., Marín-Franch, A., et al. 2014, in Proc. SPIE, Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V, 914911  
 Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, J. Artif. Intell. Res., 16, 321  
 Chen, T. & Guestrin, C. 2016, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16 (New York, NY, USA: Association for Computing Machinery), 785–794  
 Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, Research in Astronomy and Astrophysics, 12, 1197  
 DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2023, arXiv e-prints, arXiv:2306.06308  
 Dye, S., Lawrence, A., Read, M. A., et al. 2018, MNRAS, 473, 5113  
 Elston, R. J., Gonzalez, A. H., McKenzie, E., et al. 2006, ApJ, 639, 816  
 Flaugh, B. 2012, in APS April Meeting Abstracts, D7007  
 Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1  
 Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1  
 Gal, R. R., de Carvalho, R. R., Odewahn, S. C., et al. 2004, AJ, 128, 3082  
 Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E., & da Costa, L. 2005, A&A, 436, 895  
 Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357  
 Henrion, M., Mortlock, D. J., Hand, D. J., & Gandy, A. 2011, MNRAS, 412, 2286  
 Huang, J. S., Cowie, L. L., Gardner, J. P., et al. 1997, ApJ, 476, 12  
 Huang, J.-S., Thompson, D., Kümmel, M. W., et al. 2001, A&A, 368, 787  
 Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90  
 Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111  
 Kashikawa, N., Shimasaku, K., Yasuda, N., et al. 2004, PASJ, 56, 1011  
 Kron, R. G. 1980, ApJS, 43, 305  
 Kümmel, M. W. & Wagner, S. J. 2001, A&A, 370, 384  
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193  
 López-Sanjuan, C., Vázquez Ramió, H., Varela, J., et al. 2019, A&A, 622, A177  
 López-Sanjuan, C., Vázquez Ramió, H., Xiao, K., et al. 2023, arXiv e-prints, arXiv:2301.12395  
 Matek, K., Solarz, A., Pollo, A., et al. 2013, A&A, 557, A16  
 Marín-Franch, A., Taylor, K., Cenarro, J., Cristóbal-Hornillos, D., & Moles, M. 2015, in IAU General Assembly, Vol. 29, 2257381  
 Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., et al. 2021, ApJS, 253, 8  
 McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, The Messenger, 154, 35  
 Miller, A. A., Kulkarni, M. K., Cao, Y., et al. 2017, AJ, 153, 73  
 Molino, A., Benítez, N., Moles, M., et al. 2014, MNRAS, 441, 2891  
 Odewahn, S. C., de Carvalho, R. R., Gal, R. R., et al. 2004, AJ, 128, 3092  
 Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2016, A&A, 589, C2  
 Reid, I. N., Yan, L., Majewski, S., Thompson, I., & Smail, I. 1996, AJ, 112, 1472  
 Saglia, R. P., Tonry, J. L., Bender, R., et al. 2012, ApJ, 746, 128  
 Schlafly, E. F. & Finkbeiner, D. P. 2011, ApJ, 737, 103  
 Scranton, R., Johnston, D., Dodelson, S., et al. 2002, ApJ, 579, 48  
 Seaborn, W. L. 1979, AJ, 84, 1526  
 Vasconcellos, E. C., de Carvalho, R. R., Gal, R. R., et al. 2011, AJ, 141, 189  
 Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261  
 von Martens, R., Marra, V., Quartín, M., et al. 2024, MNRAS, 527, 3347  
 Wang, C., Bai, Y., López-Sanjuan, C., et al. 2022, A&A, 659, A144  
 Yasuda, N., Fukugita, M., Narayanan, V. K., et al. 2001, AJ, 122, 1104

<sup>10</sup> <https://archive.cefca.es/catalogues/>

## Appendix A: Training Data Query and External Catalogs

We have enriched our training dataset with information from other all-sky surveys, specifically incorporating astrometry and photometry from the *Gaia* third data release and the CatWISE2020 catalog. For *Gaia*, we utilized the standard match radius of 1".5 available in the J-PLUS archive, which typically provides a single match per J-PLUS object. However, for CatWISE2020, employing the standard matching radius of 5" (nearly equivalent to two ALLWISE pixels) resulted in approximately 20% of J-PLUS sources being matched with multiple CatWISE2020 counterparts. After a detailed examination, we determined that the majority, if not all, of these cases were spurious detections associated with local intensity maxima around extended sources, primarily galaxies. The cumulative distribution of angular distances between J-PLUS and CatWISE2020 sources indicates that 99.9% of matches occur within 1.5 arcseconds. Consequently, we opted to include only those CatWISE2020 sources that are located within an angular distance of 1.5 arcseconds from a J-PLUS source, while retaining all matches that have smaller separations.

It is important to note that, by adhering to this criterion, some objects may appear more than once in the catalog. This situation could, in theory, lead to multiple different outcomes from BANNJOS. However, for the sake of completeness and consistency with the matching criteria used for other external catalogs in the J-PLUS archive, we decided to retain all such matches.

For each Tile, we performed exactly the same query. For example, our query for the Tile 103930 was:

```
SELECT FLambda.*, MagAB.MU_MAX,
MagAB.APER3_WORSTPSF, FNu.FLUX_AUTO,
FNu.FLUX_APER_0_8, FNu.FLUX_APER_1_0,
FNu.FLUX_APER_1_2, FNu.FLUX_APER_1_5
FNu.FLUX_APER_2_0, FNu.FLUX_APER_3_0,
FNu.FLUX_APER_4_0, FNu.FLUX_APER_6_0,
MagAB.MAG_AUTO, MagAB.MAG_APER_0_8,
MagAB.MAG_APER_1_0, MagAB.MAG_APER_1_2,
MagAB.MAG_APER_1_5, MagAB.MAG_APER_2_0,
MagAB.MAG_APER_3_0, MagAB.MAG_APER_4_0,
MagAB.MAG_APER_6_0, FNu.FLUX_RELERR_AUTO,
FNu.FLUX_RELERR_APER_0_8,
FNu.FLUX_RELERR_APER_1_0,
FNu.FLUX_RELERR_APER_1_2,
FNu.FLUX_RELERR_APER_1_5,
FNu.FLUX_RELERR_APER_2_0,
FNu.FLUX_RELERR_APER_3_0,
FNu.FLUX_RELERR_APER_4_0,
FNu.FLUX_RELERR_APER_6_0, MagAB.MAG_ERR_AUTO,
MagAB.MAG_ERR_APER_0_8, MagAB.MAG_ERR_APER_1_0,
MagAB.MAG_ERR_APER_1_2, MagAB.MAG_ERR_APER_1_5,
MagAB.MAG_ERR_APER_2_0, MagAB.MAG_ERR_APER_3_0,
MagAB.MAG_ERR_APER_4_0, MagAB.MAG_ERR_APER_6_0,
MWEx.ax, MWEx.ax_err, MWEx.ebv, MWEx.ebv_err,
gaia.angDist, gaia.pmra as pmra_g, gaia.pmde,
gaia.plx, gaia.ruwe, gaia.fg, gaia.fbp,
gaia.frp, gaia.e_pmra, gaia.e_pmde,
gaia.e_plx, gaia.e_fg, gaia.e_fbp, gaia.e_frp,
allwise.angDist, allwise.Jmag, allwise.Hmag,
allwise.Kmag, allwise.e_Jmag, allwise.e_Hmag,
allwise.e_Kmag, catwise.angDist, catwise.pmRA,
catwise.e_pmRA, catwise.pmDE, catwise.e_pmDE,
catwise.W1mproPM, catwise.e_W1mproPM,
```

```
catwise.W2mproPM, catwise.e_W2mproPM FROM
jplus.FLambdaDualObj AS FLambda LEFT JOIN
jplus.MagABDualObj AS MagAB ON ((FLambda.NUMBER
= MagAB.NUMBER) AND (FLambda.TILE_ID =
MagAB.TILE_ID)) LEFT JOIN jplus.FNuDualObj
AS FNu ON ((FLambda.NUMBER = FNu.NUMBER) AND
(FLambda.TILE_ID = FNu.TILE_ID)) LEFT JOIN
jplus.MWExtinction AS MWEx ON ((FLambda.NUMBER
= MWEx.NUMBER) AND (FLambda.TILE_ID =
MWEx.TILE_ID)) LEFT JOIN jplus.xmatch_gaia_dr3
AS gaia ON ((FLambda.NUMBER = gaia.NUMBER)
AND (FLambda.TILE_ID = gaia.TILE_ID)) LEFT
JOIN jplus.xmatch_allwise AS allwise ON
((FLambda.NUMBER = allwise.NUMBER) AND
(FLambda.TILE_ID = allwise.TILE_ID)) LEFT
JOIN jplus.xmatch_catwise2020 AS catwise
ON ((FLambda.NUMBER = catwise.NUMBER) AND
(FLambda.TILE_ID = catwise.TILE_ID)) WHERE
((catwise.angDist <= 1.5) AND (FLambda.TILE_ID =
103930))
```

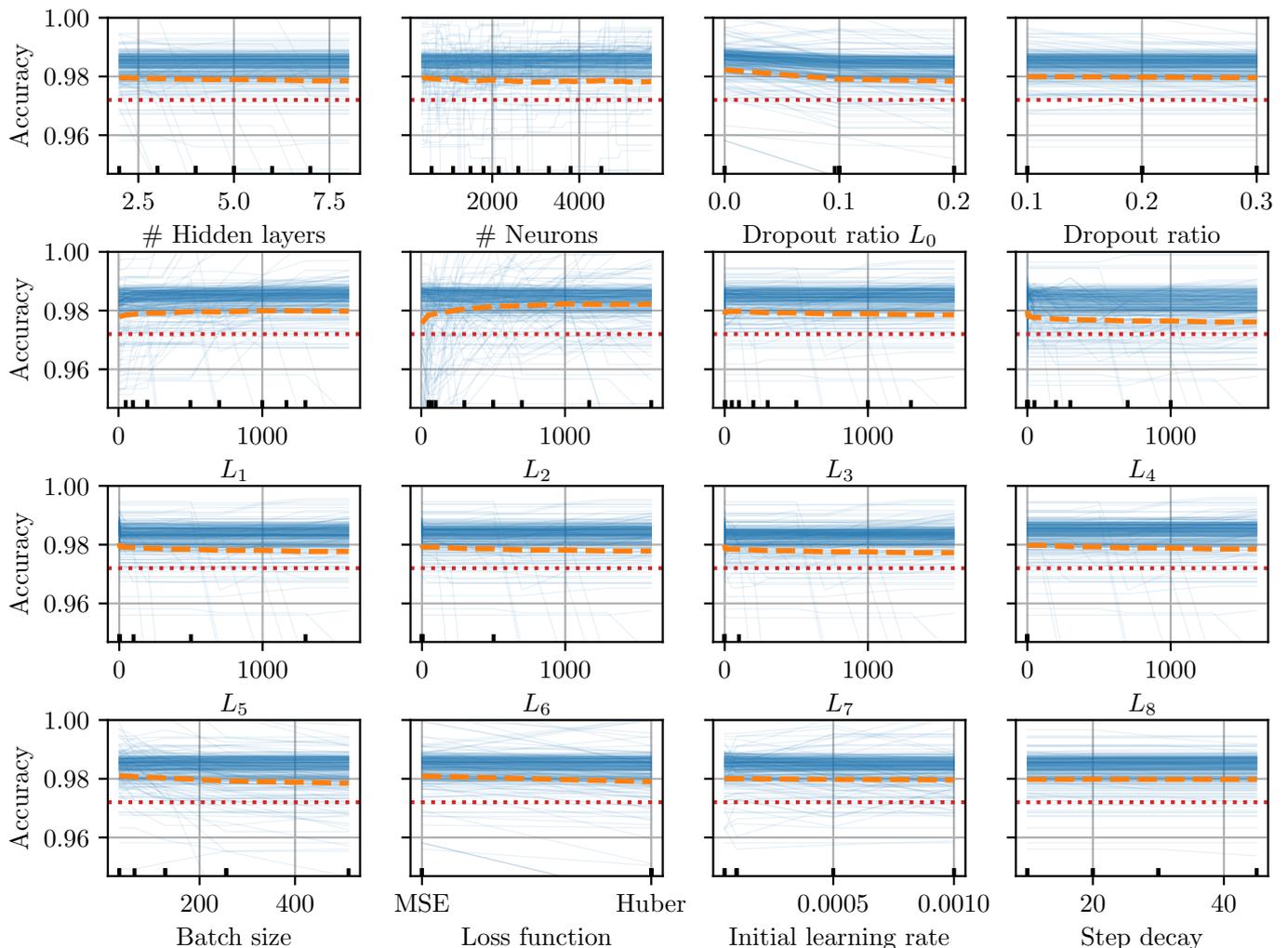
## Appendix B: Model Selection and Hyper-Parameter Tuning

Figure B.1 displays Individual Conditional Expectation (ICE) curves illustrating the model's accuracy as influenced by each hyperparameter. We note variations in performance depending on the hyperparameters employed, observing generally better outcomes with deeper architectures and notably the number of neurons in the first four hidden layers ( $L_{1-4}$ ). Optimal accuracy was achieved with models containing a large number of neurons in  $L_1$  and  $L_2$  (at least 200) followed by additional hidden layers with fewer neurons. Our tests revealed correlations between certain hyperparameters, such as simultaneous increases in  $L_1$  and  $L_2$  or  $L_2$  and  $L_5$ , leading to improved results. Conversely, some degree of anticorrelation between  $L_3$  and  $L_{4-5}$ , and between  $L_4$  and  $L_5$ , suggests a balance between the number of layers and free parameters for maintaining accuracy. The neuron count in layers  $L_5$  to  $L_8$  had minimal impact on accuracy, though a slight decline was observed with higher neuron counts ( $\geq 500$ ). Similarly, increasing the dropout ratio at  $L_0$  marginally reduced model accuracy. Other parameters had less significant effects on performance, with the dropout BANN maintaining consistent average performance for models featuring three or more hidden layers. The best hyperparameters were identified by a Histogram-based Gradient Boosting Regression Tree, and are detailed in Section 3.2.

## Appendix C: Data Compression

Figure C.1 illustrates the results for a source with ambiguous classification by the model (Tile Id = 94217, Number = 10482) with  $N = 5000$ . As anticipated, the points reside on the plane  $P(\text{class} = \text{Galaxy}) + P(\text{class} = \text{QSO}) + P(\text{class} = \text{Star}) = 1$  with minimal dispersion, indicating the model's correct learning of the correlations between the three probabilities without explicit training.

Although retaining all points permits full PDF reconstruction for the object, extensive sampling with large  $N$  entails high computational and storage costs. To mitigate this, we experimented with smaller  $N$  values, identifying a minimal consistent value,  $N = 300$ . However, this still results in sub-



**Fig. B.1.** Individual Conditional Expectation curves for the model accuracy. Each blue line represents the accuracy of the model depending on the chosen value for each hyperparameter. The orange dashed line indicates the mean accuracy value, approximately 0.98, which is below the median accuracy value, closer to 0.99. The red dashed line represents the performance of a Random Forest Regressor with 500 trees. Most hyperparameters have minimal impact on model performance, except for the Dropout at  $L_0$ , and the number of neurons at levels  $L_{1-4}$ .

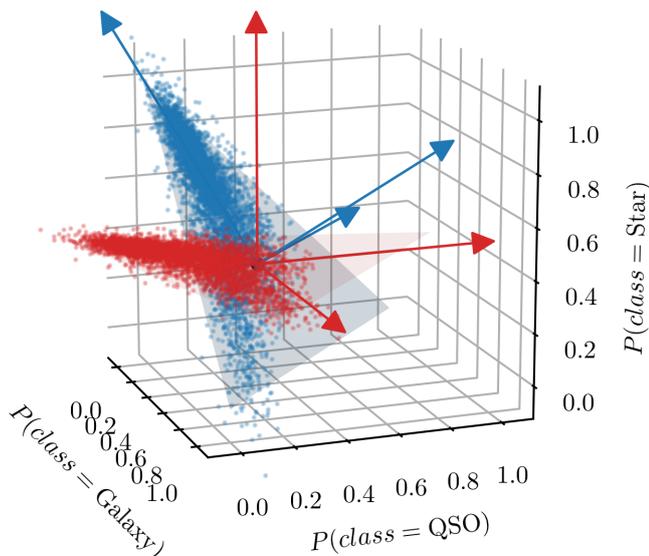
stantial amount of data. To further reduce data volume, we exploit the points' planar distribution by applying a rotation:

$$R = \begin{pmatrix} (\sqrt{3} + 3)/6 & -\sqrt{(2 - \sqrt{3})}/6 & -1/\sqrt{3} \\ -\sqrt{(2 - \sqrt{3})}/6 & (\sqrt{3} + 3)/6 & -1/\sqrt{3} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix} \quad (\text{C.1})$$

which projects the  $N \times 3$  points onto the plane defined by  $P(x = \text{Galaxy}) + P(x = \text{QSO}) + P(x = \text{Star}) = 1$ . In this projection, the vertical coordinate, perpendicular to the plane, is negligible and can be discarded. This transformation leverages the interrelated PDFs to shift from three to two dimensions, effectively reducing storage needs to  $N \times 2$ . Further compression is achieved by fitting a Gaussian Mixture Model, GMM, with three components to the  $N \times 2$  points, parameterizing the entire PDF with the means, covariances, and weights of the three Gaussian distributions—totalling 18 parameters after reducing the covariance matrix elements.

This compression approach was tested with sources exhibiting one, two, and three components in their predicted PDFs by creating an additional validation sample through  $N = 5000$  BANN posterior samples ( $\text{PDF}_{\text{hi}}$ ). We then fitted GMMs to both high and low-resolution datasets, reconstructing 3D PDFs by sampling each fitted GMM 5000 times. For each test sample object, we thus compared four different PDFs:  $\text{PDF}_{\text{hi}}$ ,  $\text{PDF}_{\text{lo}}$ ,  $\text{GMM}_{\text{hi}}$ , and  $\text{GMM}_{\text{lo}}$ , adopting the class with the highest median value from each PDF as the predicted classification.

The vast majority of sources showed consistent classifications across  $\text{PDF}_{\text{hi}}$ ,  $\text{PDF}_{\text{lo}}$ , and their corresponding GMMs. Furthermore,  $\text{GMM}_{\text{lo}}$  accurately replicated the original  $\text{PDF}_{\text{hi}}$  in all visually inspected cases. However, a small fraction of cases exhibited classification shifts due to reduced sample size and the GMM compression, affecting less than 0.003% of cases (4 out of 136570), typically those with low S/N and large classification uncertainties. These discrepancies were mitigated by applying quality cuts. For the remainder, BANNJOS yielded consistent results between high-quality



**Fig. C.1.** Sampling from the posterior of the model for Tile Id = 94217 and Number = 10482, a source with low signal-to-noise ratio. Each blue point represents a sample from the BANN ( $N = 5000$ ). The most probable class is ‘star’, though BANNJOS also assigns some probability to the ‘galaxy’ class. The points distribute on the plane  $P(\text{class} = \text{Galaxy}) + P(\text{class} = \text{QSO}) + P(\text{class} = \text{Star}) = 1$ , as expected. Blue arrows represent the eigenvectors of the 3-dimensional space defined by the probabilities of belonging to each of the three classes. A rotation (Equation C) applied to the blue points allows for dimension reduction, transforming them into the red points on the horizontal plane. The red arrows denote the eigenvectors of this plane.

and standard samples and between original and reconstructed PDFs. Considering the significant benefits and minimal disadvantages, we proceeded with  $N = 300$  and GMM compression.

## Appendix D: Selecting pure samples

In this Appendix, we provide some examples of Astronomical Data Query Language (ADQL) queries that yield purer object selections. Using the percentiles of the PDF, users can specify precisely the type of objects they wish to select. For example, the following query selects stars at the  $1\sigma$  confidence level and provides their sky coordinates and  $r$  magnitudes:

```
SELECT MagAB.ALPHA_J2000, MagAB.DELTA_J2000,
MagAB.MAG_PSF COR[1], BANNJOS.* FROM
jplus.MagABDualObj AS MagAB LEFT JOIN
jplus.ClassBANNJOS AS BANNJOS ON
((BANNJOS.NUMBER = MagAB.NUMBER) AND
(BANNJOS.TILE_ID = MagAB.TILE_ID)) WHERE
(BANNJOS.CLASS_STAR_prob_pc16 >= 0.333)
```

Here, we require the 16th percentile to be above  $\frac{1}{3}$  of the probability, which is equivalent to a random probability in a classification with three possible classes. Since the probability at which a certain percentile is fulfilled does not need to be complementary between classes, this query also selects objects whose probability of being a galaxy or a QSO is compatible within one sigma with  $\frac{1}{3}$ . To be even more restrictive

when selecting stars, we can select only sources with the 84th percentiles below  $\frac{1}{3}$ :

```
SELECT MagAB.ALPHA_J2000, MagAB.DELTA_J2000,
MagAB.MAG_PSF COR[1], BANNJOS.* FROM
jplus.MagABDualObj AS MagAB LEFT JOIN
jplus.ClassBANNJOS AS BANNJOS ON
((BANNJOS.NUMBER = MagAB.NUMBER) AND
(BANNJOS.TILE_ID = MagAB.TILE_ID)) WHERE
((BANNJOS.CLASS_STAR_prob_pc16 >= 0.333) AND
(BANNJOS.CLASS_QSO_prob_pc84 < 0.333) AND
(BANNJOS.CLASS_GALAXY_prob_pc84 < 0.333))
```

The same criteria can be used to select very probable QSOs at the  $2\sigma$  confidence level. Furthermore, we can use the additional information from the correlation between the probabilities to select objects with a negative correlation between being a star and a QSO. Requiring negative correlations between the classes’ probabilities can be interpreted as requiring that there is no confusion between the classes, i.e., it is either one or the other.

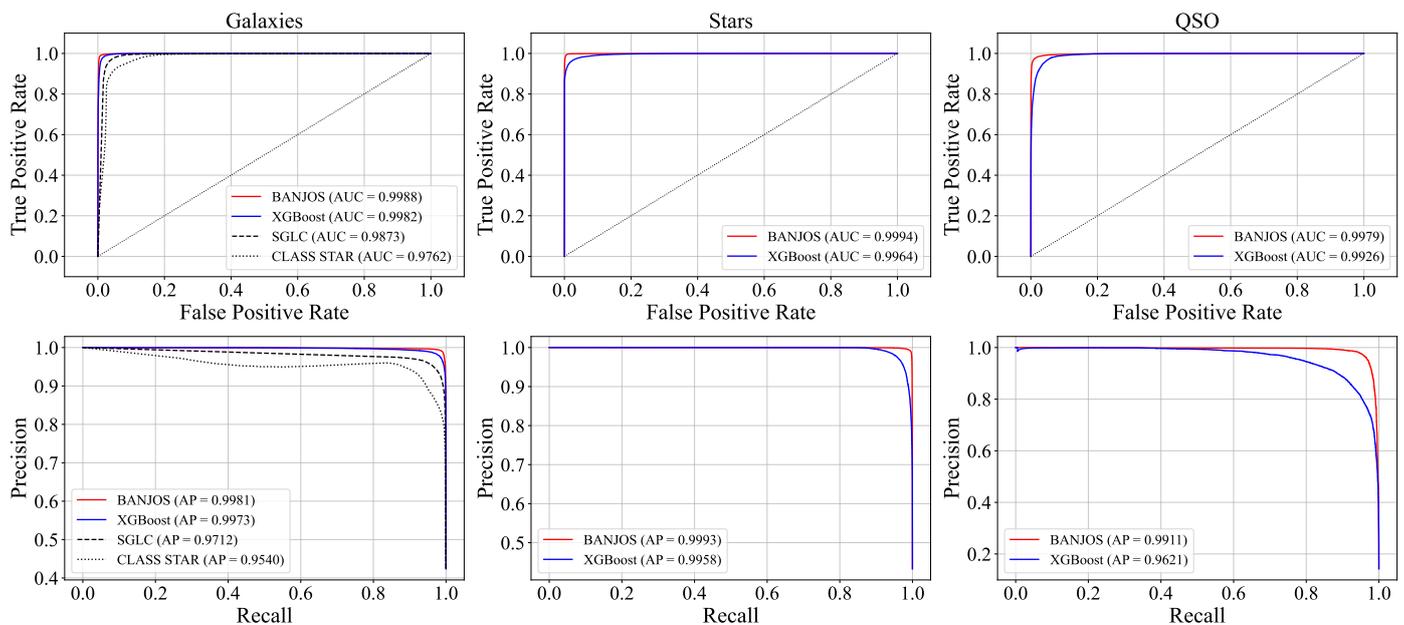
```
SELECT MagAB.ALPHA_J2000, MagAB.DELTA_J2000,
MagAB.MAG_PSF COR[1], BANNJOS.* FROM
jplus.MagABDualObj AS MagAB LEFT JOIN
jplus.ClassBANNJOS AS BANNJOS ON
((BANNJOS.NUMBER = MagAB.NUMBER) AND
(BANNJOS.TILE_ID = MagAB.TILE_ID)) WHERE
((BANNJOS.CLASS_QSO_prob_pc02 >= 0.333) AND
(BANNJOS.CLASS_STAR_prob_pc98 < 0.333) AND
(BANNJOS.CLASS_GALAXY_prob_pc98 < 0.333) AND
(BANNJOS.CLASS_QSO_CLASS_STAR_prob_corr <= 0))
```

This query will return a sample of 404,070 high-confidence QSOs in J-PLUS. We tested this sample by cross-matching its objects with those from our test sample, obtaining a total of 11,230 common sources never seen by BANNJOS during training. From these sources, 134 are spectroscopically classified as galaxies, and 18 as stars. This represents approximately 1.19% and 0.16% contamination from these other species, respectively. It is important to point out, however, that we did not explicitly reject sources with positive correlation between the galaxy and QSO classes, since we could also be interested in partially resolved active galaxies. Requiring  $\text{BANNJOS.CLASS\_GALAXY\_CLASS\_QSO\_prob\_corr} \leq 0$  would produce an even cleaner sample by reducing the contamination from these types of sources.

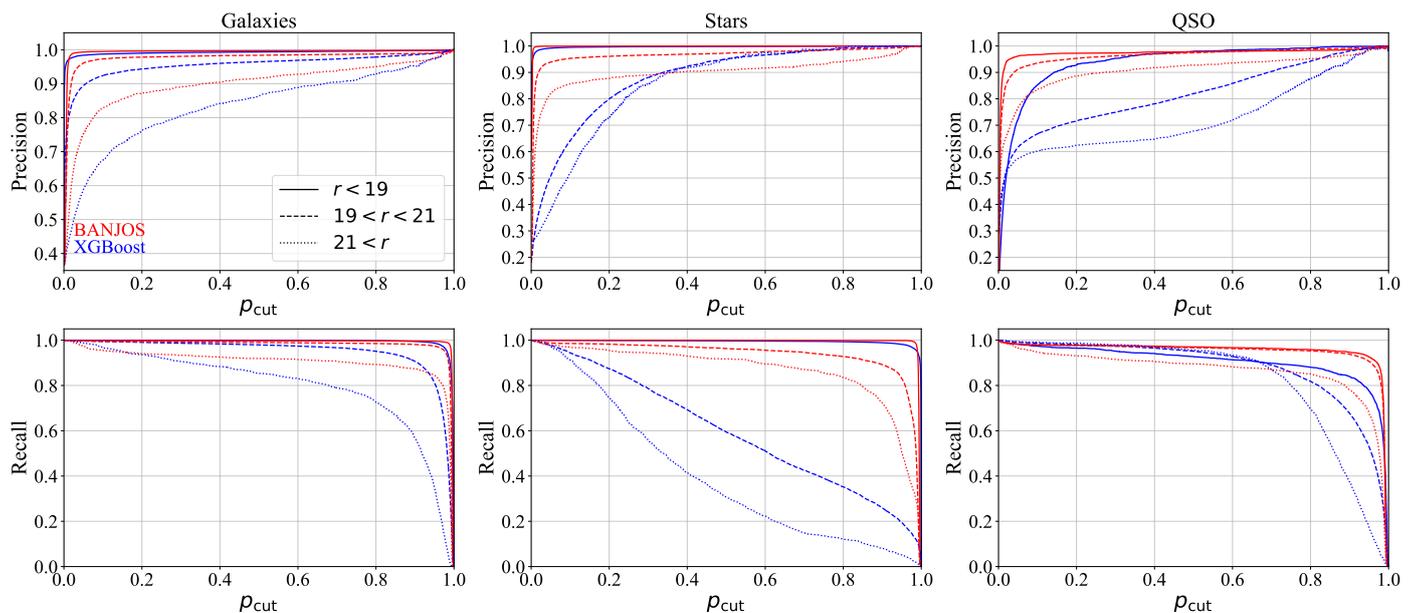
We cross-matched this list of high-confidence QSOs from our training list, obtaining up to 290,794 new QSO candidates not previously cataloged either by SDSS, LAMOST, DESI, or *Gaia*. If we relax the conditions to  $1\sigma$ , which still produces very pure samples, this number increases to 388,976 relatively high-confidence QSOs without spectroscopic confirmation that could be used for spectroscopic follow-up.

## Appendix E: Performance comparison between classifiers

In this Appendix, we present a more general analysis of the comparison between the classifiers presented in Section 6 using the test sample of sources not common to the training of BANNJOS or the XGBoost algorithm used in vM24. Figure E.1 shows the ROC and the PR curves for the four analyzed classifiers. We assign the class predicted by BANNJOS as the one with the highest median probability,  $\max[PC_{\text{class}}(50)]$ . For the rest of the classifiers, the class is selected as the one with the higher probability. Point-like sources have been



**Fig. E.1.** *Top:* ROC curves for the three classes using the four available classifiers. *sglc\_prob\_star* and *CLASS\_STAR* are only shown for extended sources. *Bottom:* Corresponding PR curves.



**Fig. E.2.** Precision and Recall. *Top:* Precision curves for *vM24* and *BANNJOS* for the three classes in three different magnitude bins. The precision is computed as function of the probability used to select each class ( $p_{cut}$ ). In the case of *BANNJOS* we use the maximum of the median probability. *Bottom:* Corresponding Recall curves.

separated into stars and QSOs only for *vM24* and *BANNJOS*. Figure E.2 shows the Precision and the Recall curves separately for the *vM24* classification and *BANNJOS* at different magnitude bins. As discussed in detail in Section 6, *BANNJOS* is superior to all other classifiers for all the considered classes and across the entire magnitude range. It is also noticeable that the *XGBoost* algorithm used in *vM24* is superior to *sglc\_prob\_star* and *CLASS\_STAR* for classifying extended

sources, with *CLASS\_STAR* being the least reliable of all four classifiers.

## Appendix F: The ClassBANNJOS table

Table F.1 details the structure of the *ClassBANNJOS* table generated by *BANNJOS*. For additional information on the significance of each field, refer to Section 3.3 and Appendix C.

The table is accessible through J-PLUS data portal<sup>11</sup>, named as ClassBANNJOS.

---

<sup>11</sup> <https://archive.cefca.es/catalogues/jplus-dr3>

**Table F.1.** Structure of the ClassBANNJOS table.

Name	Type	Description
TILE_ID	int	Identifier of the Tile image where the object was detected
NUMBER	int	Number identifier assigned by SExtractor for the object in the image
CLASS_GALAXY_prob_mean	float	Mean probability of being a galaxy
CLASS_QSO_prob_mean	float	Mean probability of being a QSO
CLASS_STAR_prob_mean	float	Mean probability of being a star
CLASS_GALAXY_prob_std	float	MAD standard deviation of the probability of being a galaxy
CLASS_QSO_prob_std	float	MAD standard deviation of the probability of being a QSO
CLASS_STAR_prob_std	float	MAD standard deviation of the probability of being a star
CLASS_GALAXY_CLASS_QSO_prob_corr	float	Correlation between the probabilities of being a galaxy or a QSO
CLASS_GALAXY_CLASS_STAR_prob_corr	float	Correlation between the probabilities of being a galaxy or a star
CLASS_QSO_CLASS_STAR_prob_corr	float	Correlation between the probabilities of being a QSO or a star
CLASS_GALAXY_prob_pc02	float	2.275th percentile of the probability of being a galaxy
CLASS_QSO_prob_pc02	float	2.275th percentile of the probability of being a QSO
CLASS_STAR_prob_pc02	float	2.275th percentile of the probability of being a star
CLASS_GALAXY_prob_pc16	float	15.865th percentile of the probability of being a galaxy
CLASS_QSO_prob_pc16	float	15.865th percentile of the probability of being a QSO
CLASS_STAR_prob_pc16	float	15.865th percentile of the probability of being a star
CLASS_GALAXY_prob_pc50	float	Median of the probability of being a galaxy
CLASS_QSO_prob_pc50	float	Median of the probability of being a QSO
CLASS_STAR_prob_pc50	float	Median of the probability of being a star
CLASS_GALAXY_prob_pc84	float	84.135th percentile of the probability of being a galaxy
CLASS_QSO_prob_pc84	float	84.135th percentile of the probability of being a QSO
CLASS_STAR_prob_pc84	float	84.135th percentile of the probability of being a star
CLASS_GALAXY_prob_pc98	float	97.725th percentile of the probability of being a galaxy
CLASS_QSO_prob_pc98	float	97.725th percentile of the probability of being a QSO
CLASS_STAR_prob_pc98	float	97.725th percentile of the probability of being a star
comp1_cov_11	float	2-dimensional covariance term 11 for Gaussian component 1
comp1_cov_12	float	2-dimensional covariance term 12 for Gaussian component 1
comp1_cov_22	float	2-dimensional covariance term 22 for Gaussian component 1
comp2_cov_11	float	2-dimensional covariance term 11 for Gaussian component 2
comp2_cov_12	float	2-dimensional covariance term 12 for Gaussian component 2
comp2_cov_22	float	2-dimensional covariance term 22 for Gaussian component 2
comp3_cov_11	float	2-dimensional covariance term 11 for Gaussian component 3
comp3_cov_12	float	2-dimensional covariance term 12 for Gaussian component 3
comp3_cov_22	float	2-dimensional covariance term 22 for Gaussian component 3
comp1_mean_1	float	2-dimensional mean term 1 for Gaussian component 1
comp1_mean_2	float	2-dimensional mean term 2 for Gaussian component 1
comp2_mean_1	float	2-dimensional mean term 1 for Gaussian component 2
comp2_mean_2	float	2-dimensional mean term 2 for Gaussian component 2
comp3_mean_1	float	2-dimensional mean term 1 for Gaussian component 3
comp3_mean_2	float	2-dimensional mean term 2 for Gaussian component 3
comp1_weight	float	2-dimensional weight for Gaussian component 1
comp2_weight	float	2-dimensional weight for Gaussian component 2
comp3_weight	float	2-dimensional weight for Gaussian component 3