



# Star Formation and Morphological Properties of Galaxies in the Pan-STARRS $3\pi$ Survey. I. A Machine-learning Approach to Galaxy and Supernova Classification

A. Baldeschi<sup>1</sup>, A. Miller<sup>1,2</sup> , M. Stroh<sup>1</sup> , R. Margutti<sup>1,3,4</sup> , and D. L. Coppejans<sup>1</sup>

<sup>1</sup> Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA; [adriano.baldeschi@northwestern.edu](mailto:adriano.baldeschi@northwestern.edu)

<sup>2</sup> The Adler Planetarium, Chicago, IL 60605, USA

<sup>3</sup> CIFAR Azrieli Global Scholars program, CIFAR, Toronto, Canada

Received 2020 April 30; revised 2020 July 16; accepted 2020 July 17; published 2020 October 12

## Abstract

We present a classification of galaxies in the Pan-STARRS1 (PS1)  $3\pi$  survey based on their recent star formation history and morphology. Specifically, we train and test two Random Forest (RF) classifiers using photometric features (colors and moments) from the PS1 data release 2. The labels for the morphological classification are taken from Huertas-Company et al., while labels for the star formation fraction (SFF) are from the Blanton et al. catalog. We find that colors provide more predictive accuracy than photometric moments. We morphologically classify galaxies as either early- or late-type, and our RF model achieves a 78% classification accuracy. Our second model classifies galaxies as having either a low-to-moderate or high SFF. This model achieves an 89% classification accuracy. We apply both RF classifiers to the entire PS1  $3\pi$  dataset, which allows us to assign two scores to each PS1 source:  $P_{\text{HSFF}}$ , which quantifies the probability of having a high SFF; and  $P_{\text{spiral}}$ , which quantifies the probability of having a late-type morphology. Finally, as a proof of concept, we apply our classification framework to supernova (SN) host galaxies from the Zwicky Transient Factory and the Lick Observatory Supernova Search samples. We show that by selecting  $P_{\text{HSFF}}$  or  $P_{\text{spiral}}$ , it is possible to significantly enhance or suppress the fraction of core-collapse SNe (or thermonuclear SNe) in the sample with respect to random guessing. This result demonstrates how contextual information can aid transient classifications at the time of first detection. In the current era of spectroscopically starved time-domain astronomy, prompt automated classification is paramount. Our table is available at [10.5281/zenodo.3990545](https://doi.org/10.5281/zenodo.3990545).

*Unified Astronomy Thesaurus concepts:* Classification (1907); Supernovae (1668); Core-collapse supernovae (304); Type Ia supernovae (1728); Random Forests (1935)

## 1. Introduction

The improved sensitivity, cadence, and field of view of recent and current astronomical transient surveys have led to the discovery of new types of rare transients that sample the extremes of the luminosity and timescale parameter space of cosmic explosions, and have dramatically enhanced our understanding of the classes of transients that were already known to exist. The classes of exotic transients that were recently discovered include superluminous supernovae (SLSNe, e.g., Quimby et al. 2011; Gal-Yam 2019), and fast and blue optical transients (FBOTs, e.g., Drout et al. 2014; Arcavi et al. 2016; Shivvers et al. 2016; Tanaka et al. 2016; Pursiainen et al. 2018; Tampo et al. 2020). Furthermore, astronomical surveys are now capable of routinely discovering supernovae (SNe) within one day (or less) of the explosion (e.g., Gal-Yam et al. 2011). Future surveys such as the Legacy Survey of Space and Time (LSST, Ivezić et al. 2019), which is carried out on the Vera C. Rubin Observatory, will drastically increase the discovery rate of new transients. However, this will make prompt spectroscopic classification of a sizeable fraction of transients effectively unfeasible. Indeed, even intrinsically rare events such as SLSNe are expected to be detected at a rate of  $\sim 10^4 \text{ yr}^{-1}$  (Villar et al. 2018). It is thus imperative to develop new pathways for prompt transient classification.

The most common classification methods consist of leveraging the transient photometry by using state-of-the-art

machine-learning algorithms. A first generation of photometric transient classifiers (mostly developed in response to the Supernova Photometric Classification Challenge, SPCC, Kessler et al. 2010a, 2010b) can be broadly divided into empirical template-fitting methods (e.g., Sullivan et al. 2006; Sako et al. 2008, 2011), and algorithms that rely on the derivation of (computationally expensive) features from extended photometry (e.g., Newling et al. 2011; Karpenka et al. 2013; Lochner et al. 2016; Möller et al. 2016; Narayan et al. 2018; Sooknunan et al. 2018; Ishida et al. 2019; Villar et al. 2019). A new generation of transient classifiers that do not require complete light-curve phase coverage and/or feature extraction have emerged in the last few years, and have been applied to transient images (Carrasco-Davis et al. 2019) and SN photometric time series (Charnock & Moss 2017; Moss 2018; Muthukrishna et al. 2019; Pasquet et al. 2019; Möller & de Boissière 2020). The advanced non-feature based neural network architectures that are used in these works include recurrent neural networks (RNN), and deep neural networks (DNN). More recently, Muthukrishna et al. (2019) developed RAPID (Real-time Automated Photometric IDentification) and tested this RNN-based tool on a PLAsTiCC-based dataset. In contrast to the other classification schemes, RAPID employs a deep learning model that is able to promptly classify different types of transients with very limited light-curve information and without requiring complete phase coverage.<sup>5</sup>

<sup>4</sup> Alfred P. Sloan Fellow.

<sup>5</sup> See Sullivan et al. (2006) for earlier attempts to prompt transient classification with minimal light-curve information.

Complementary methods for prompt transient classification employ contextual information; i.e., the properties of the galactic environments where transients are discovered (e.g., Foley & Mandel 2013 and the *sherlock* package<sup>6</sup>). These methods rely on very well-known correlations between transient types and their host-galaxy environments (e.g., core-collapse SNe tend to trace star formation, while SNe Ia occur both in early-type and late-type host galaxies).

Building on the early results from Foley & Mandel (2013), which showed that (among other contextual properties) the host-galaxy morphology has the largest predictive power for SN typing, we present the classification of all of the galaxies detected by the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)  $3\pi$  survey. We adopt a random forest (RF) machine-learning approach that leverages galaxy features to classify the galaxies based on their morphology (i.e., elliptical/S0 versus spiral) and their star formation properties. We train and test two supervised machine-learning RF algorithms, starting from the morphological classifications of SDSS galaxies by Huertas-Company et al. (2011) and the star formation properties from Blanton et al. (2005) and Blanton & Roweis (2007). We present a catalog where we provide two scores for each Pan-STARRS source that quantify the probability of having spiral morphology ( $P_{\text{spiral}}$ ) and the probability of having a high recent star formation rate ( $P_{\text{HSFF}}$ ). Finally, as a proof of concept, we use both scores to classify host galaxies of SNe from the Zwicky Transient Facility (ZTF) Bright Transient Survey (BTS, Fremling et al. 2019) and from the Lick Observatory Supernova Search (LOSS, Leaman et al. 2011). In addition, we show how  $P_{\text{spiral}}$  and  $P_{\text{HSFF}}$  correlate with the fraction of core-collapse SNe (or thermonuclear SNe) in the sample. Our catalog of Pan-STARRS galaxies classifications will be made publicly available and can be ingested by transient brokers (e.g., ALERCE,<sup>7</sup> ANTARES,<sup>8</sup> LASAIR,<sup>9</sup>) that sort, cross-reference, and value-add streams of alerts from astronomical surveys. The inferred host-galaxy properties can be used by recommender engines, such as REFIT (Sravan et al. 2020), together with the photometric information to promptly inform decisions on transient follow-up at the time of their first detection.

This paper is organized as follows. In Section 2, we describe the datasets used, while in Section 3 we train and test the first RF model by cross-matching the Pan-STARRS DR2 catalog with the Huertas-Company et al. (2011) dataset. In Section 4, we train and test the second RF model by cross-matching the Pan-STARRS DR2 catalog with the New York University Value-Added Galaxy Catalog (NYU-VAGC of Blanton et al. 2005; Blanton & Roweis 2007). In Section 5, we apply the second RF to the entire Pan-STARRS dataset. Finally, in Section 6 we classify the host galaxies of ZTF and LOSS SNe by leveraging the results of Section 5. Conclusions are drawn in Section 7.

## 2. Description of the Datasets

For our analysis, we use three data catalogs. Specifically, we utilize the second Pan-STARRS1 (PS1, Chambers et al. 2016)

data release of the  $3\pi$  survey (PS1-DR2 hereafter), the Huertas-Company dataset (HC hereafter, Huertas-Company et al. 2011), and the New York University Value-Added Galaxy Catalog (NYU-VAGC hereafter, Blanton et al. 2005; Blanton & Roweis 2007). We used the HC, NYU-VAGC and a subsample of PS1-DR2 to train and test the two RF algorithms that we will later apply to the entire PS1-DR2 catalog.

PS1 is a system for wide-field astronomical imaging that was developed at the University of Hawaii and is located on the island of Maui. PS1 data have been acquired with a 1.8 meter telescope and a 1.4 Gigapixel camera to capture images of the sky through five optical filters ( $g_{P1}$  [4866 Å],  $r_{P1}$  [6215 Å],  $i_{P1}$  [7545 Å],  $z_{P1}$  [8679 Å],  $y_{P1}$  [9633 Å]). Two surveys have been carried out with the PS1 telescope: the medium deep survey and the  $3\pi$  survey ( $3\pi$ S). In this paper, we use data from the  $3\pi$ S, which covers a larger fraction of the sky. The  $3\pi$ S covers the sky north of decl.  $\delta = -30^\circ$  with the five filters listed previously, and includes data acquired between 2009 June 2 and 2014 March 31. The maximum depth of the  $3\pi$ S for the stack images is  $\sim 23.5$  mag for the  $g_{P1}$ ,  $r_{P1}$ , and  $i_{P1}$  filters, while it is  $\sim 22.5$  and  $\sim 21.5$  mag for the  $z_{P1}$  and  $y_{P1}$  filters, respectively. There have been two data releases for the  $3\pi$ S (Chambers et al. 2016). Data Releases 1 and 2 (DR1 and DR2) include stacked images and a database with the photometry of all the sources detected in  $3\pi$ S (both extended and point-like). DR2 also contains forced photometry for each epoch. In this paper, we use the data from the StackObjectAttributes table within DR2.<sup>10</sup> This table contains the photometric information (e.g., PSF-flux, Kron-flux) of the stacked data, which is calculated as described in Magnier et al. (2013). To be included in this table, a source must be detected with a signal-to-noise  $S/N > 20$  in an individual exposure. Multiple detections are often made of the same source from subsequent exposures, which means that there may be multiple photometric entries for a single source. In Section 3.1 we describe the columns of the StackObjectAttributes table that we used as features to train our RF classifiers.

The HC dataset (Huertas-Company et al. 2011) consists of 699,684 galaxies from the SDSS-DR7 spectroscopic sample (Abazajian et al. 2009) with redshift  $z \leq 0.25$  and observed  $r$ -band magnitude  $m_r \leq 18$  mag. Huertas-Company et al. (2011) provide a morphological classification of the SDSS-DR7 galaxies through a supported-vector-machine (SVM) classifier (e.g., Hastie et al. 2009), which was trained with a subsample of 2253 SDSS visually classified galaxies from Fukugita et al. (2007). The SVM classifier provides a score/probability for a galaxy to be either elliptical-lenticular (E/S0) or spiral (S). In the HC catalog, 44% of the S galaxies are red, while only 7% of the E/S0 are blue.<sup>11</sup> We note that the GZ project (Lintott et al. 2008) has released a catalog of morphological classifications of a large sample of galaxies (e.g., Schawinski et al. 2009; Skibba et al. 2009; Masters et al. 2010; Willett et al. 2013). In Appendix B, we discuss the reasoning behind our choice of using the HC dataset.

The second galaxy catalog is the NYU-VAGC (Blanton et al. 2005). The NYU-VAGC provides the Star Formation Fraction (SFF) of 2506754 SDSS-DR2 galaxies (see Equation (1) for the SFF definition), which is estimated over the last 300 Myr. We use the SFF value to divide the galaxies

<sup>6</sup> <https://github.com/thespacedoctor/sherlock>

<sup>7</sup> Smith et al. (2019), <http://alerce.science>.

<sup>8</sup> Saha et al. (2014, 2016), <https://antares.noao.edu/>.

<sup>9</sup> <https://lasair.roe.ac.uk>

<sup>10</sup> StackObjectAttributes table link

<sup>11</sup> We followed Equation (1) of Masters et al. (2010) to discriminate between red and blue galaxies.

into two classes: galaxies with low-to-moderate SFF, and galaxies with high SFF. The quantitative details on the two classes are provided in Section 4. In the following two sections, we train and test two RF classifiers with these data catalogs (PS1-DR2, HC and NYU-VAGC).

### 3. RF Model for Sources in PS1-DR2 Based on the HC Catalog

In this section, we analyze the PS1-DR2 and HC catalogs. We start by cross-matching the two catalogs. For each common source, we retain features that are relevant to our subsequent analysis (Section 3.1). We then train and test an RF algorithm that classifies the galaxies as E/S0 or S (Sections 3.2, 3.3).

#### 3.1. Feature Selection and Preprocessing

In the first step, we identify sources that are common to HC and PS1-DR2 by utilizing a  $0''.8$  search radius. This search radius is optimized to account for a slight difference in astrometry between the two catalogs, without introducing a significant number of spurious associations. Tachibana & Miller (2018), for example, use a similar matching radius for the PS1-DR1 catalog. Furthermore, we discard PS1-DR2 sources with missing data. After cross-matching, the dataset consists of 659,460 galaxies. For each common source, we identify which properties of those listed in PS1-DR2 are the most relevant to our analysis, and associate the labels (E/S0 or S) from the HC catalog. Specifically, for each PS1 photometry filter, we identify the meaningful features as the PSF-flux, the Kron-flux (Kron 1980), and the second moment of the radiation intensity, defined as  $\langle XY \rangle = \int_{S_{xy}} uvI(u, v) du dv$  or  $\langle X^2 \rangle = \int_{S_x} u^2 I(u) du$ , where  $I$  is the radiation intensity. The moments of the radiation intensity are directly related to the distribution of light within a source and hence provide information on the source shape. These properties (which constitute features for our machine-learning tools of Section 3.2) are listed in the “StackObjectAttributes” table of the PS1-DR2 catalog as PSFFlux, KronFlux, momentYY, momentXY and momentXX.

The features of our training set (i.e., fluxes and moments of the radiation intensity) are distance dependent and, hence, are not properly suitable for a machine-learning algorithm. We thus engineer the fluxes and moments into a series of features that are not dependent on distance. Specifically, we consider the ratio between the fluxes and the ratio of the moments of the radiation intensity for different filters to be meaningful features. This procedure leads to 126 features associated to each of the 659,460 sources in our sample. Finally, we standardize the features according to the formula  $X_{st} = (X - \mu)/\sigma$ , where  $X$  is the input feature, while  $\mu$  and  $\sigma$  are the mean and the standard deviation of the sample, respectively.

The final step of preprocessing is the reduction of dimensionality. We use here the Principal Component Analysis (PCA, e.g., Hastie et al. 2009), which performs a linear transformation of a set of correlated features into linearly uncorrelated variables. This algorithm returns a set of eigenvectors with associated eigenvalues, where larger eigenvalues are associated to eigenvectors that describe the most of the variance of the dataset. The dimensionality reduction is achieved by discarding the components with smaller eigenvalues. We retain the 55 (of the original 126) features responsible for 99.8% of the sample variance.

Before applying machine-learning algorithms, we address the potential problem of class imbalance, which occurs when the classes of objects identified by the labels (in our case E/S0 versus S) contain markedly different numbers of elements. In the HC dataset  $\approx 39\%$  of the sources belong to the E/S0 class, while  $\approx 61\%$  belong to the S class. We apply standard undersampling because the number of objects is large (659,460) and the dataset is not heavily imbalanced, which leads to a final balanced dataset of 514,288 sources.

#### 3.2. Machine Learning with RF

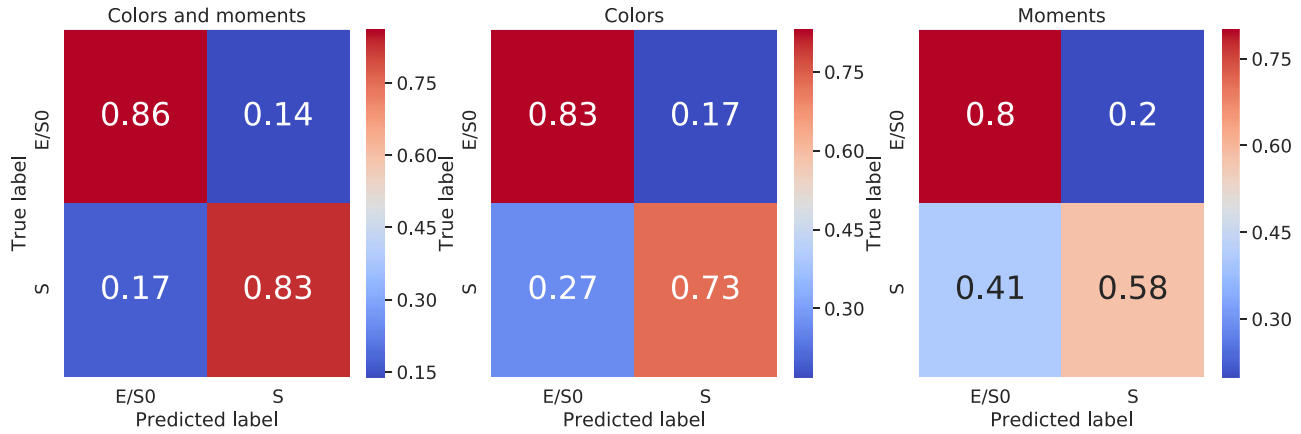
Supervised machine learning classifies objects by learning a mapping function from the training set and then applying the mapping function to previously unseen data. A large variety of machine-learning algorithms are known, and some have also been used in the astronomical literature (e.g., Dieleman et al. 2015; Lochner et al. 2016; Baldeschi et al. 2017a, 2017b; Carrasco-Davis et al. 2019; Muthukrishna et al. 2019; Ntampaka et al. 2019; Schanche et al. 2019; Villar et al. 2019; Walmsley et al. 2019; Margalef-Bentabol et al. 2020; Sravan et al. 2020; Steinhardt et al. 2020). We apply three algorithms: RF, SVM, and boosting (e.g., Hastie et al. 2009). Of these three, RF is the algorithm that leads to the highest classification accuracy, which we employ below.

The architecture of the RF construction depends on several hyperparameters. Here, we adopt the following hyperparameters: (i) the number of estimators (i.e., the number of trees) is fixed at 100; (ii) the maximum depth of a tree can be 10, 12, 14, 16, 18, 20, or 22; (iii) the minimum number of samples required to split an internal node is 2, 3, or 4; and (iv) the minimum number of samples required to be a leaf node is 2, 3, or 4. Finally, we adopt the Gini<sup>12</sup> information gain to measure the quality of a split. An RF model is produced for each combination of parameters and is then tested using a standard  $k$ -fold cross-validation procedure (e.g., Hastie et al. 2009) with  $k = 4$ . The  $k$ -fold cross-validation procedure is based on a single hyperparameter, represented by  $k$ , which indicates the number of groups (i.e., folds) into which a given sample is to be split. Cross-validation is used to estimate the performance of a machine-learning model over previously unseen data, and it generally provides a less-biased estimate of the model than other methods (e.g., a simple train/test split). The general procedure for the  $k$ -fold cross-validation includes: (i) random shuffling of the dataset; (ii) splitting the dataset into  $k$  folds; (iii) taking one fold as a test set and the remaining folds as a training set; (iv) applying a machine-learning model on the training set and evaluating the model on the test set; (v) repeating the procedure for each  $k$  fold; and (vi) averaging the results on the test set for a better evaluation score. In our specific case, each fold contains a random sample of 128,572 galaxies with  $z \leq 0.25$  and  $m_r \leq 18$  mag. Our results are described in Section 3.3.

#### 3.3. Main Results: Classification Reliability, Importance of Features and Redshift Dependence

The RF in the cross-validation set results in a classification accuracy of 85%, which means that we can reliably infer the morphological properties of galaxies as E/S0 or S by utilizing only the colors (i.e., the flux ratios) and the moments ratio of

<sup>12</sup> See Hastie et al. (2009) for a description of the Gini information gain.

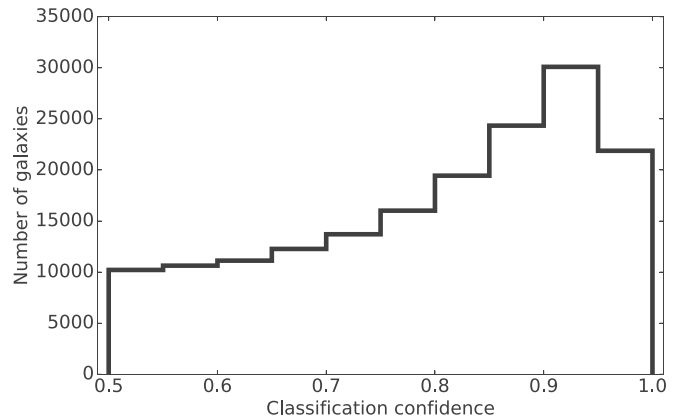


**Figure 1.** Normalized confusion matrix for the two classes of sources: Elliptical-lenticular (E/S0) and spiral (S). Here, we use a random forest (RF) algorithm to classify the objects in the cross validation set of the PS1-DR2-HC dataset. We train the RF classifier by retaining as predictive features colors and the ratio of moments of the radiation intensity (left-hand panel), only colors (middle panel), only the ratio of moments (right-hand panel).

the radiation intensity. Figure 1 (left-hand panel) displays the confusion matrix for the two classes, from which we conclude that the RF model performs equally well for both classes. The dataset is balanced and the fraction of sources that are correctly classified is 0.86 and 0.83 for the E/S0 and S types, respectively.

The RF classifier gives the probability that a source belongs to either the E/S0 or the S class. The predicted classification probabilities of an input sample are computed as the mean predicted class probabilities of the trees in the forest. The class probability of a single tree is the fraction of samples of the same class in a leaf. Figure 2 shows the distribution of the classification confidence for the sources in the test set to belong to one of the two classes (E/S0 or S). The distribution is bound between 0.5 and 1, and has a mean value of 0.8 and a median value of 0.83. We define the classification confidence as equal to the classification probability if this is  $>0.5$ , otherwise we define it as  $1-(\text{classification probability})$ . We note that 74% of the sources in the HC sample have a classification confidence (of belonging to the E/S0 or the S class) that is larger than 70%. This suggests that in most cases the classifier provides reliable classifications.

The importance of the features in an RF model is typically measured by estimating the increase in the model’s predictive accuracy after permuting the feature (see e.g., Hastie et al. 2009; Fisher et al. 2018). A feature is considered to be important if shuffling its values decreases the model’s accuracy, which implies that the model relied on that specific feature for the prediction. We estimate the importance of the colors and the ratio of moments of the radiation intensity in the morphological classification of galaxies as follows. We build two models, where one model considers only the ratio of moments to be features, and the other considers only the colors to be features. We find a classification accuracy of 78% for the RF model that retained only the colors as features, while a classification accuracy of 69% is achieved by retaining only the ratio of the moments. Figure 1 displays the confusion matrix for both the colors (middle panel) and moment classifiers (right-hand panel), respectively. We conclude that while colors are the most significant features, the combination of colors and ratio of moments leads to an improved final classification accuracy.

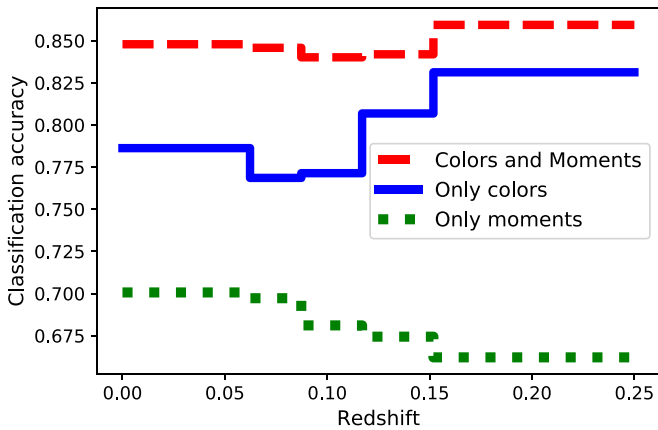


**Figure 2.** Classification confidence distribution for the HC galaxies of being elliptical-lenticular (E/S0) or spiral (S). We define the classification confidence as the classification probability if the classification probability is  $>0.5$  and as  $1-(\text{classification probability})$  if the classification probability is  $<0.5$ . The distribution has mean and median values of 0.80 and 0.83, respectively.

This result is not surprising, considering that E/S0 and S galaxies are well known to display markedly different intensity profiles (e.g., Caon et al. 1993; Young & Currie 1994). The moments of the radiation intensity are shape parameters that quantify this intrinsic difference, and can be considered as alternatives to the standard  $M_{20}$  estimator (Lotz et al. 2004).

Finally, in Figure 3 we explore the dependence of the classification accuracy on the redshift of the sources.

When the colors and moments of the radiation intensity are retained as features, (red-dashed line in Figure 3), the classification accuracy is approximately constant with redshift, suggesting that our methodology provides results that are not heavily distance dependent at  $z < 0.25$  and are mostly unbiased with distance. When colors are the only features, (blue line in Figure. 3), the classification accuracy is slightly larger at higher redshift, while when we use the ratio of moments of the radiation intensity as only features of the model (green-dotted line in Figure 3), the classification accuracy decreases at higher redshift. From Figure 3 it is also clear that the classification accuracy of the RF classifier that retains both colors and ratios of moments of the radiation intensity is significantly higher at all redshifts. The classification accuracy of the RF model that retains only colors as features shows a mild increase with  $z$ , while the classification accuracy for the



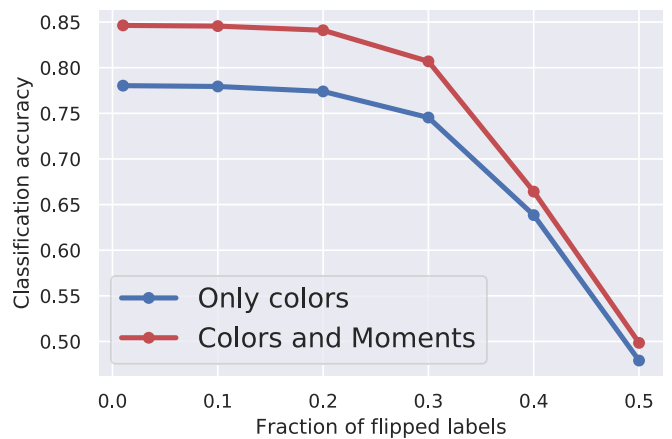
**Figure 3.** Classification accuracy as a function of the galaxy redshift for three RF models for the PS1-DR2-HC dataset. Red-dashed line: both colors and ratios of moments of the radiation intensity are used as input features for the RF classifier. Blue line: only colors are used as input features for the RF classifier. Green-dotted line: only the ratios of moments of the radiation intensity are used as input features for the RF classifier. The figure reveals that the classification accuracy is approximately constant for  $z < 0.25$  when the RF is trained using colors and ratios of the radiation intensity as features. Data at  $z > 0.15$  have been collected into a single bin due to limited statistics.

RF model that retains only the moments of the radiation intensity slightly decreases with  $z$ . We speculate that the first behavior results from the larger fraction of E/S0 galaxies at larger  $z$ ,<sup>13</sup> while the latter effect results from the degradation of the discerning power of the moments of the radiation intensity at larger  $z$ , where details of the light-distribution of galaxies becomes increasingly harder to discern. We note that this RF model discriminates better the E/S0 galaxies than the S galaxies (Figure 1 middle panel).

These results assume that all of the labels are exact and accurate. In fact, since the labels are the result of a probabilistic classification by Huertas-Company et al. (2011) an unknown fraction of galaxies is mislabeled. Even though a fraction of galaxies may be mislabeled, our RF algorithm is robust to the presence of partially mislabeled data. A path to evaluate the strength of the model is to randomly flip a fraction of labels in the training set and then estimate the classification accuracy. In Figure 4, we show the classification accuracy in the test set versus the fraction of flipped labels in the training set for the RF that leverages both color and moments and the RF that leverages only the colors as input features. This figure reveals that even if 30% of the labels are flipped, the classification accuracy drops by only  $\approx 5\%$ . For larger fractions of flipped labels, the classification accuracy is  $\approx 50\%$  (equivalent to random guessing), as expected.

### 3.4. Comparison to Simpler ( $g-r$ ) Color Cuts

The correlation between a galaxy’s morphology and colors is well known and studied (e.g., Blanton & Berlind 2007; Ball et al. 2008; van der Wel 2008; Bamford et al. 2009; Schawinski et al. 2009; Skibba et al. 2009; Masters et al. 2010). Specifically, Skibba et al. (2009) found a correlation between the  $g-r$  color and the likelihood for a galaxy to be S or E/S0 in the Galaxy-zoo (GZ) dataset: red galaxies with  $(g-r) > 0.95$  mag are more likely to be elliptical ( $P_{\text{ell}} \approx 70\%$ ), while blue galaxies with  $(g-r) < 0.7$  mag are more likely to have spiral



**Figure 4.** Classification accuracy in the test set (S and E/S0) vs. fraction of flipped labels in the training set in the HC dataset. The figure shows that the RF model is robust because the classification accuracy drops by only 5% when 30% of the labels are flipped. Red-solid line: RF model with both colors and moments of the radiation intensity. Blue-solid line: RF model with only colors as features.

morphology ( $P_{\text{spir}} \approx 90\%$ ). In this work, we consider a more complex classifier that leverages multiple colors and moments of the radiation intensity. However, it is important to compare our RF classifiers to a simpler  $g-r$  color-splitting classifier. We build a color-splitting classifier as follows: (i) we split the HC dataset into a training and test set; (ii) we select a threshold value  $(g-r)_{\text{thres}}$ ; (iii) sources with  $(g-r) > (g-r)_{\text{thres}}$  are classified as E/S0, and sources with  $(g-r) < (g-r)_{\text{thres}}$  are classified as S; and (iv) we select the appropriate  $(g-r)_{\text{thres}}$  as the value that maximizes the classification accuracy in the training set (which we find to be  $(g-r)_{\text{thres}} = 0.74$  mag). We obtain a classification accuracy of 75% in the test set. The RF classifier that we trained on multiple colors as input features yields a 78% classification accuracy, while the RF trained on both colors on moments of the radiation intensity leads to a 85% classification accuracy. The RF trained on multiple colors is slightly more accurate than a color-splitting classifier, but an RF trained on both colors and moments of the radiation intensity is significantly more accurate.

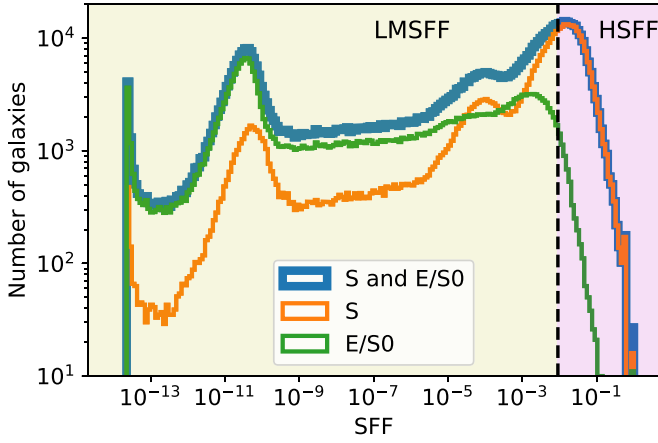
## 4. RF Model for Sources in PS1-DR2 Based on the NYU-VAGC

In this section, we train and test an RF classifier by combining the features in PS1-DR2 with the galaxies star formation properties derived from the NYU-VAGC. The goal is to build an RF model that can classify sources in the PS1-DR2 into two distinct classes consisting of galaxies with a low-to-moderate star formation fraction (SFF, Section 4.1), and galaxies with high SFF.

### 4.1. Selection of Sources and Definition of Labels for the NYU-VAGC

The NYU-VAGC consists of 250,6754 sources. Among these, we select sources with redshifts in the range  $0.002 \leq z \leq 0.5$  to limit the contamination by stars in the Galaxy and lower-quality observations. We discard sources with bad photometry (e.g., missing data and high uncertainty in the fluxes). These cuts result in a catalog of 662,804 sources that are used in the following analysis.

<sup>13</sup> 30% of the galaxies with  $z < 0.12$  are E/S0 while 53% of the galaxies with  $z > 0.12$  are E/S0.



**Figure 5.** SFF distribution of NYU-VAGC sources. Green: elliptical-lenticular (E/S0) sources. Orange: spiral (S) sources. Blue: complete sample of E/S0 and S sources. We perform object classification (E/S0 vs. S) using the RF classifier developed in Section 3.2. We labeled the sources on the left of the vertical black-dashed line as LMSFF and sources on the right of the vertical black line as HSFF. Most of the HSFF sources are S, while the LMSFF sources are both S and E/S0 (mainly E/S0).

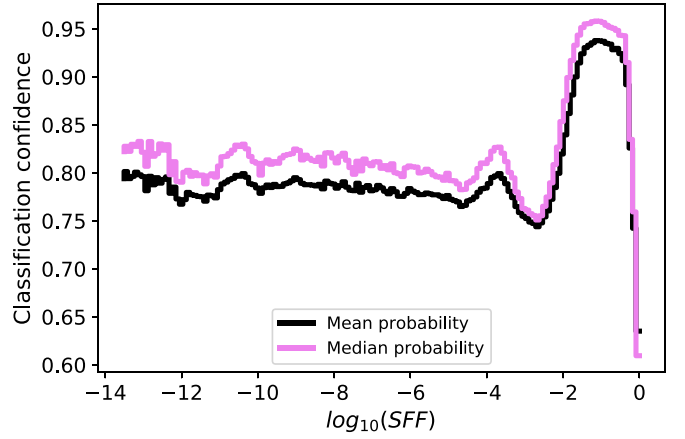
The NYU-VAGC catalog provides the star formation fraction (SFF<sup>14</sup>) for each source, defined as (Blanton & Roweis 2007):

$$\text{SFF} \equiv \frac{\int_{t_0-0.3 \text{ Gyr}}^{t_0} \text{SFR}(t) dt}{\int_0^{t_0} \text{SFR}(t) dt}, \quad (1)$$

where  $t_0$  is the present epoch and  $\text{SFR}(t)$  is the star formation rate as a function of time. SFF is thus a continuous unitless variable that represents the fraction of star formation that has occurred in the past 0.3 Gyr. We use the RF classifier developed in Section 3, which retains both the colors and ratios of moments of the radiation intensity as input features, to divide the NYU-VAGC sources into two SFF classes based on their early-type (i.e., E/S0) or late-type (i.e., S) classification (i.e., to convert the continuous SFF variable above into a discrete feature that can be used as a label).

Figure 5 displays the distribution of SFF of galaxies in the NYU-VAGC classified as E/S0 (green line) and S (orange line). Figure 5 reveals that most of the E/S0 galaxies are associated with a low-to-moderate SFF, while most of the S-classified galaxies are associated with larger SFFs, as expected. In particular, for  $\text{SFF} \gtrsim 10^{-2}$  most sources are late-type galaxies, while for  $\text{SFF} \lesssim 10^{-2}$  the galaxy types consist of a mixture of both early-type and late-type (where the early-type galaxies are predominant). Therefore, we split the population of galaxies into two distinct classes: for  $\text{SFF} > 9 \times 10^{-3}$ , the sources are labeled as galaxies with high star formation fraction (HSFF); whereas for smaller values ( $\text{SFF} < 9 \times 10^{-3}$ ), the sources are labeled as galaxies with low-to-moderate star formation fraction (LMSFF).

We can also estimate the mean and median classification confidence of the NYU-VAGC galaxies to be either E/S0 or S for each bin of the SFF distribution, and we show the results in Figure 6. The classification confidence is estimated by employing the RF classifier developed in Section 3.2. Figure 6 reveals that the median classification confidence lies between 0.8 and



**Figure 6.** Mean (black) and median (pink) classification confidence for each of the SFF distribution bins in the NYU-VAGC catalog. Galaxies with very large  $\text{SFF} > 10^{-2}$  are classified with higher confidence.

0.85 for  $\text{SFF} < 10^{-4}$ , decreases to 0.75 at  $\text{SFF} \approx 10^{-3}$  and increases above 0.9 at  $\text{SFF} \approx 10^{-1}$ . HSFF galaxies are thus classified with higher confidence.

#### 4.2. Preprocessing, Training, Testing and Results

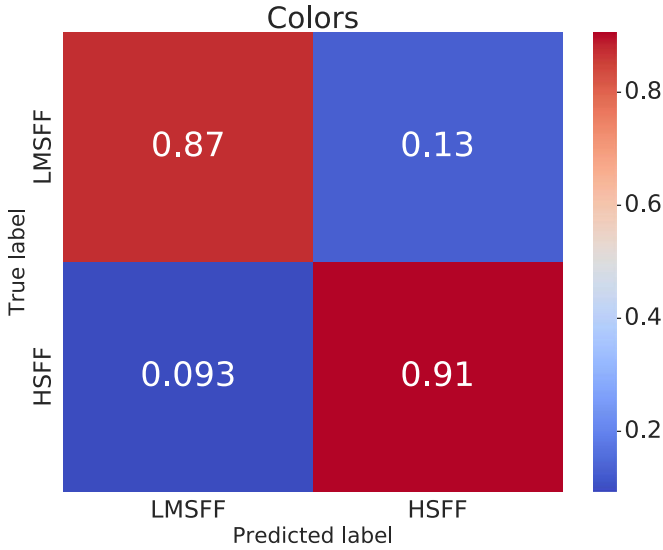
In Section 4.1, we defined the labels for the classification process (HSFF versus LMSFF). Here, we follow the methodology developed in Section 3 (i.e., preprocessing, training, and testing by cross-matching sources in the PS1-DR2 and HC dataset) for the NYU-VAGC catalog. First, we select common sources between NYU-VAGC and PS1-DR2 by cross-matching the two catalogs using a  $0''.8$  radius. We only consider colors as predictive features for the training/testing set because considering colors and ratios of moments of the radiation intensity would lead to a lower classification accuracy (see the discussion below). The features are then standardized and a PCA is performed to reduce the dimensionality of the dataset, as described in Section 3. We start with 25 features (the ratio of fluxes in different photometric bands). The PCA reduces the number of meaningful features to 16, which are responsible for 99.7% of the sample variance.

This procedure results in a dataset where  $\approx 25\%$  of sources belong to the HSFF class, while  $\approx 75\%$  belong to the LMSFF class. We adopt undersampling to balance the dataset, resulting in a final dataset of 323220 sources. After data standardization and balancing, we train and test an RF using 4-fold cross-validation. The RF classifier is optimized employing the grid-search method discussed in Section 3.2. The RF classifier reaches a classification accuracy of 89% in the 4-fold cross-validation.

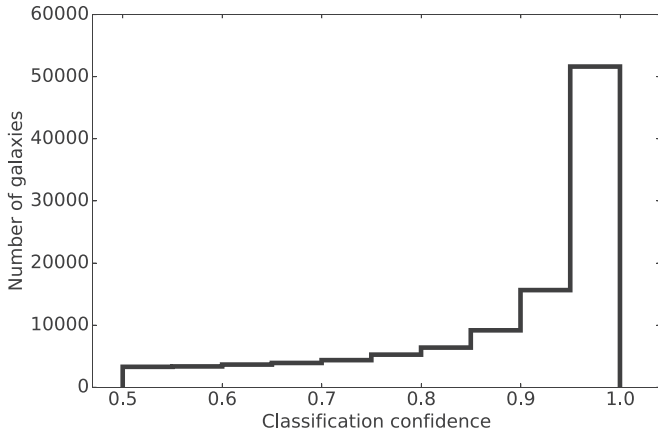
Figure 7 shows the confusion matrix for the HSFF versus LMSFF classification. We find that the fraction of sources that are correctly classified is 0.91 and 0.87 for the HSFF and LMSFF types, respectively. In Figure 8, we display the distribution of the classification confidence for each source in the test set. The distribution is bound between 0.5 and 1, and the mean and median values of the distribution are 0.88 and 0.94, respectively. These results suggest that the RF model can predict the label of the galaxy (HSFF or LMSFF) with a reasonably high probability by using the colors as the only meaningful features.

Although the high classification accuracy within the cross validation set is a reliable measure for estimating the validity of

<sup>14</sup> The SFF parameter was named as B300 in Blanton & Roweis (2007).



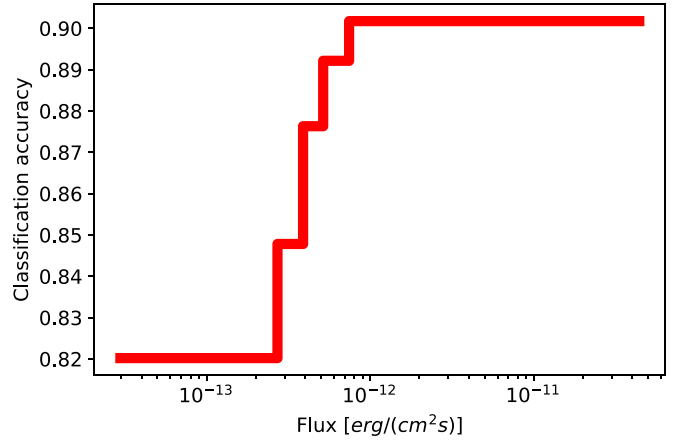
**Figure 7.** Normalized confusion matrix for the two classes HSFF and LMSFF of the NYU-VAGC catalog. Only colors were used as input features for the RF classifier described in Section 4.2.



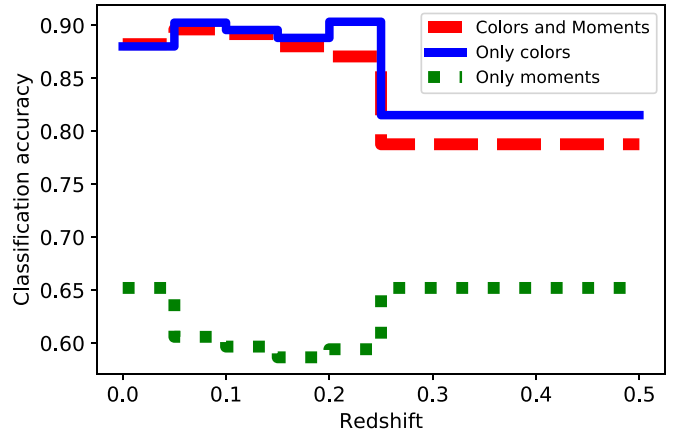
**Figure 8.** Distribution of the classification confidence of the NYU-VAGC galaxies labeled as HSFF or LMSFF. The probability that an individual galaxy is HSFF or LMSFF is calculated by employing the RF classifier that was developed in Section 4. The distribution has mean and median values of 0.88 and 0.94, respectively.

the RF model, it is also important to explore the dependence of the classification accuracy on the brightness of the galaxies. In Figure 9, we display the classification accuracy as a function of the integrated flux in the five PS1 filters. The integrated flux is estimated by computing the area under the spectral energy distribution with the trapezoidal rule. The figure shows that our classification accuracy is a monotonically increasing function of the source flux, as expected (i.e., brighter sources are easier to classify).

In Section 3.3 and Figure 3, we demonstrated that the E/S0 versus S galaxy classification is most accurate when using both colors and ratios of moments of the radiation intensity as features, and we explored the dependency of the classification accuracy on the sources redshift. Here, we reproduce the exercise of Section 3 for this second RF classifier. Figure 10 illustrates the dependence of the classification accuracy on redshift for each of the three choices of training features. The



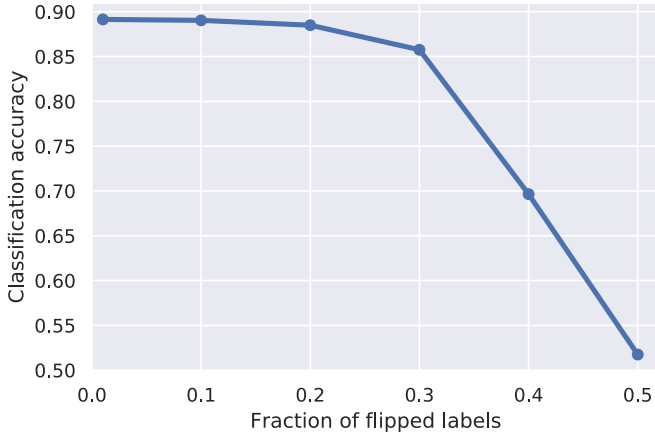
**Figure 9.** Classification accuracy (HSFF and LMSFF) vs. integrated flux of the NYU-VAGC galaxies. Each bin in this histogram contains the same number of sources. As expected, brighter sources are easier to classify.



**Figure 10.** Classification accuracy (HSFF vs. LMSFF) as a function of  $z$  for NYU-VAGC galaxies. Dashed-red line: both colors and ratios of moments of the radiation intensity have been used as input features for the RF classifier. Blue line: only colors as input features. Green-dotted line: only the ratios of moments as input features. Light-blue line: for this model we employed  $K$ -corrected colors as input features from the NYU-VAGC dataset. Data at  $z \geq 0.25$  have been collected into a single bin due to limited statistics.

figure clearly demonstrates that the highest classification accuracy is reached by considering only colors as input features. The classification accuracy is approximately constant up to  $z \approx 0.25$  and it then decreases at larger values.

Sources at different redshifts are sampled at different rest-frame wavelengths, which means that the observed colors map to intrinsically different colors in the source rest-frame. In general, it is not possible to apply a  $K$ -correction to the observed PS1-DR2 colors because the source redshift is unknown. However, we show here that training and testing on intrinsic colors leads to an increased classification accuracy, as expected. To this aim, we employ the  $K$ -corrected absolute magnitudes provided by the NYU-VAGC catalog, and we then train and test an RF classifier using  $K$ -corrected colors. In Figure 10, we show the classification accuracy in the test set as a function of redshift for this RF classifier. Unsurprisingly, we achieve higher classification accuracy at all redshifts with  $K$ -corrected colors, demonstrating the larger discerning power

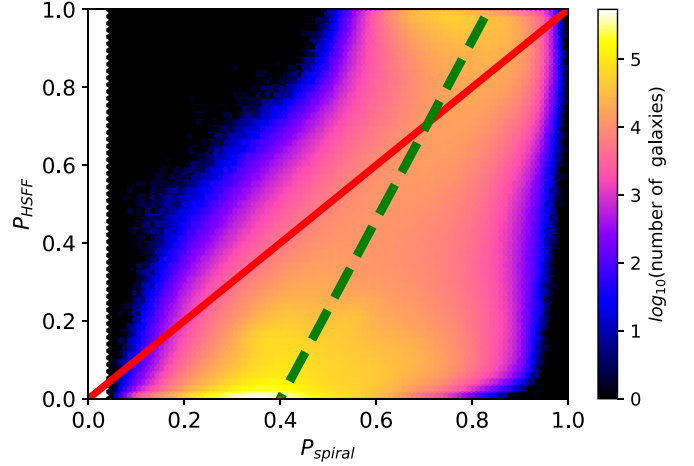


**Figure 11.** Classification accuracy in the test set (HSFF and LMSFF) vs. fraction of flipped labels in the training set in the NYU-VAGC dataset. The figure shows that the RF model is robust because the classification accuracy is  $>85\%$ , even when 30% of the labels are flipped.

of intrinsic versus observed properties. It is unfortunately not possible to apply an accurate  $K$ -correction to the majority of the PS1-DR2 sources because the redshift of most galaxies is currently unknown or is not well constrained.

These results implicitly assume that the training/testing labels are, indeed, accurate. In fact, the labels are likely to be inaccurate: a galaxy that is labeled as having HSFF may actually have LMSFF. This happens because the SFF is a difficult parameter to estimate and is subject to some degree of uncertainty. Therefore, an unknown fraction of galaxies may be mislabeled. Even though a fraction of galaxies may be mislabeled, our RF algorithm is robust to the presence of partially mislabeled data. The robustness of the model can be assessed by randomly flipping a fraction of labels in the training set and then estimating the classification accuracy. In Figure 11, we show the classification accuracy in the test set versus the fraction of flipped labels in the training set. The figure reveals that even if 30% of the labels are flipped, the classification accuracy is still very high ( $\approx 85\%$ ). For larger fractions of flipped labels, the classification accuracy approaches 50% (equivalent to random guessing), as expected.

We end with a consideration on the performance of “lighter” models that are based on a significantly smaller number of features but reach interesting levels of accuracy. In this section, we trained a large model with 16 meaningful features, and reached a classification accuracy of 0.89 in the cross validation set. It is also possible to train and test a much simpler model that uses only two features and still obtain reliable results. If we train an RF model with two features (i.e., the Kron-flux ratio between the  $g$ - and  $r$ -band filter  $F_{K,g}/F_{K,r}$ , and the  $i$  and  $z$  bands,  $F_{K,i}/F_{K,z}$ ), then we obtain a classification accuracy of 0.86 in the cross-validation set. Using PSF fluxes, the RF model achieves a classification accuracy of 0.81. Therefore, even with a highly simplified model, we obtain classification accuracies that are comparable to the main RF model of this section. In the next section, we apply the more complex and sophisticated 16-feature RF model developed in this section to the entire PS1-DR2 catalog because of its larger classification accuracy (0.89) and its intrinsic flexibility that allows the model to be applied when either (some of) the Kron or the PSF fluxes are missing for a given PS1-DR2 source.



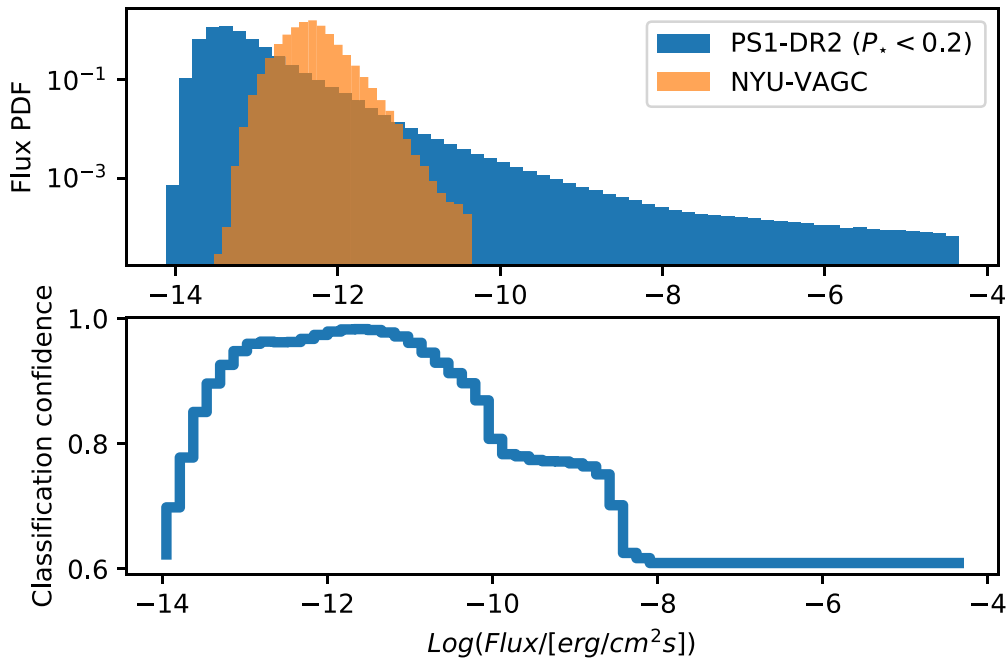
**Figure 12.**  $P_{\text{HSFF}}$  vs.  $P_{\text{spiral}}$  for PS1-DR2 galaxies (sources with  $P_* < 0.2$ ). The red line identifies the locus of the plane for which  $P_{\text{HSFF}} = P_{\text{spiral}}$ . The two probability scores show some degree of correlation, as expected (i.e., spiral galaxies tend to have HSFF, while elliptical and lenticular galaxies cluster at LMSFF). The green-dashed line ( $P_{\text{HSFF}} = 2.3 P_{\text{spiral}} - 0.9$ ) represents the approximate correlation between  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$ . The  $P_{\text{spiral}}$  score never approaches 0 because of an intrinsic difference between the cross-validation set of the HC catalog and the PS1-DR2 dataset.

## 5. Classification of Pan-STARRS Sources Based on their Star Formation Properties and Morphology

In Section 3, we developed RF models to classify galaxies as E/S0 versus S, while in Section 4 we trained an RF classifier that discriminates between HSFF and LMSFF galaxies. In this section, we apply both classification frameworks to the entire PS1-DR2 dataset and we will also build a catalog with the classification of Pan-STARRS sources based on their morphology and SFF. Specifically, we preprocess the PS1-DR2 features (in this case colors<sup>15</sup>) through the standard procedure (scaling and PCA) outlined in Sections 3 and 4 using the PCA models estimated on the HC (Section 3) and NYU-VAGC (Section 4) catalogs, respectively. We then apply the RF models developed in Sections 3 and 4 to the PS1-DR2, and we present the results of our classification in a catalog that is described in detail in Appendix A.

The average classification confidence that we obtain for the E/S0 versus S classification is lower than the classification confidence from the HSFF versus LMSFF classifier. The two classifications do, however, show some degree of correlation, as expected (Figure 12), because S-galaxies tend to be associated with HSFF sources and E/S0 galaxies to LMSFF sources. Figure 12 shows that the relationship between the two indicators is not  $P_{\text{HSFF}} = P_{\text{spiral}}$  (with scatter) but closer to  $P_{\text{HSFF}} = 2.3 P_{\text{spiral}} - 0.9$  (with large scatter), which suggests that sources with  $P_{\text{spiral}} \leq 0.5$  can effectively be mapped into sources with low  $P_{\text{HSFF}} \leq 0.25$  (and hence limited star formation). While we do not necessarily recommend the use of this RF classifier in quantitative studies of galaxy morphology, in Section 6 we find that  $P_{\text{spiral}}$  can be empirically

<sup>15</sup> In Section 3 we trained three RF models with a different set of features: (i) colors and ratio of moments of the radiation intensity; (ii) colors only; and (iii) ratio of moments of the radiation intensity only. The largest classification accuracy was achieved with the first RF model (0.85), while we achieve a classification accuracy of 0.78 for the second model and 0.69 for the third model. Despite the larger classification accuracy of the first model, here we employ the second RF model based on colors only because most sources in PS1-DR2 lack measurements of the moments of the radiation intensity.



**Figure 13.** Upper panel: empirical probability density function (PDF) of the integrated flux over the PS1 bandpass of galaxies (i.e., sources with  $P_* < 0.2$ ) in the PS1-DR2 catalog (blue) and the NYU-VAGC (orange). Log scale is used for the y-axis. Lower panel: median classification confidence that PS1-DR2 galaxies are correctly labeled as HSFF or LMSFF, as a function of the integrated flux. The classification confidence decreases when the flux of the PS1-DR2 galaxies significantly diverges from the flux of sources in the cross-validation set.

used in transient surveys to control the purity of core-collapse versus SN Ia samples.

In the rest of this section, we discuss the effects of star-galaxy misclassification, brightness, missing data and Galactic extinction on our results, and we will also provide the reader with guidelines on how to interpret and use the results from our RF models (Sections 5.1–5.2).

### 5.1. Effects of Star-galaxy Misclassifications and Missing Data

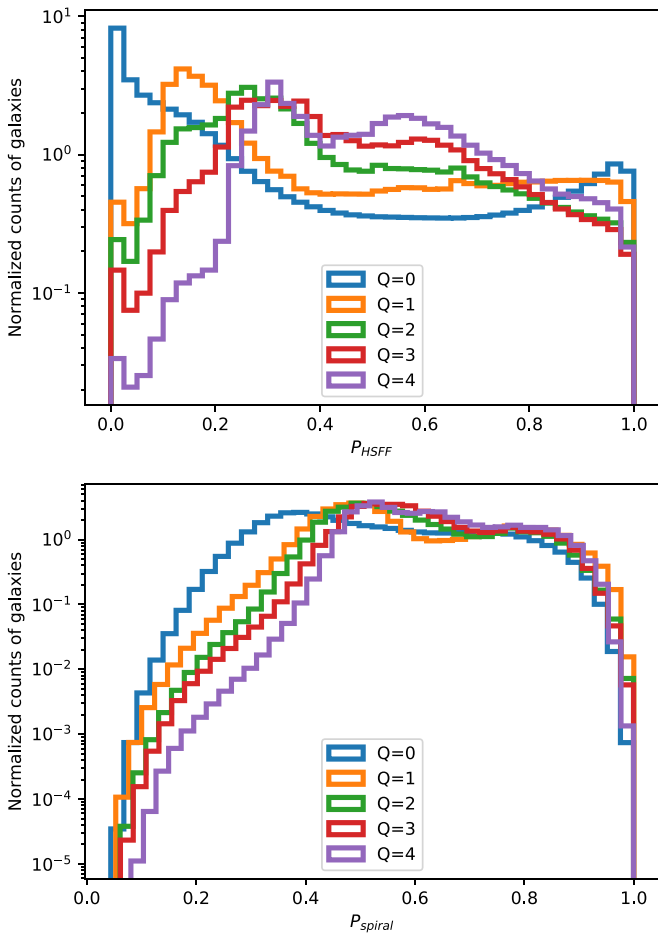
Dissimilarities between the training-testing set and PS1-DR2 include the following: (i) PS1-DR2 contains galaxies as well as non-extended objects (i.e., stars); (ii) the training/testing datasets are biased toward brighter sources, which implies that the training-testing datasets are only partially representative of sources in the PS1-DR2; and (iii) rows with missing data are present in the PS1-DR2 dataset, but in the training/testing dataset we only considered sources without missing values.

We address the star-galaxy misclassification issue using the results from Tachibana & Miller (2018). For each Pan-STARRS source, these authors provide a score  $P_*$  that quantifies the probability that the source is point-like (i.e., a star). We run the RF classifiers on the entire PS1-DR2 dataset irrespective of the  $P_*$  value, and for each source we list  $P_*$ , the probability of having a high SFF  $P_{\text{HSFF}}$  as derived by our RF classifier in Section 4, and the probability of having a spiral galaxy morphology  $P_{\text{spiral}}$  (from Section 3). Each source in our catalog thus has three separate probability scores, which allows the user to apply a custom cut on  $P_*$  as needed. For reference,  $P_* > 0.8$  indicates that the object is a star with reasonably high confidence (hence  $P_* < 0.2$  can be considered highly suggestive of a galaxy-type celestial object). In the remainder of the paper, we refer to PS1-DR2 sources with  $P_* < 0.2$  as “galaxies.”

The presence of biases between the cross validation datasets and the application set is very common in many machine-learning

implementations, and, depending on the degree of bias, cannot be easily mitigated. One way to visualize the amount of bias between the cross validation set and the whole PS1-DR2 catalog is to compare the distributions of the flux integrated over the PS1 bandpass (i.e.,  $g$ - to  $y$ -band) of sources in the two datasets, which is shown in Figure 13 (upper panel). We only consider PS1-DR2 sources with  $P_* < 0.2$  (i.e. galaxies). The median integrated flux of sources in the cross validation set (i.e., the NYU-VAGC) is  $\sim 2.6$  times larger than the median integrated flux of the entire PS1-DR2 dataset, which also shows a significantly broader distribution. Indeed, 23% of the PS1-DR2 galaxies have a lower integrated flux than the minimum value of the cross validation set, and only 0.3% of the PS1-DR2 galaxies have a larger integrated flux than the maximum value of the cross validation set. In the lower panel of Figure 13, we display the median classification confidence for the PS1-DR2 galaxies of having HSFF or LMSFF as a function of the integrated flux. This figure reveals that the classification confidence decreases when the integrated flux of the PS1-DR2 galaxies differs from the values in the cross validation set, as expected.

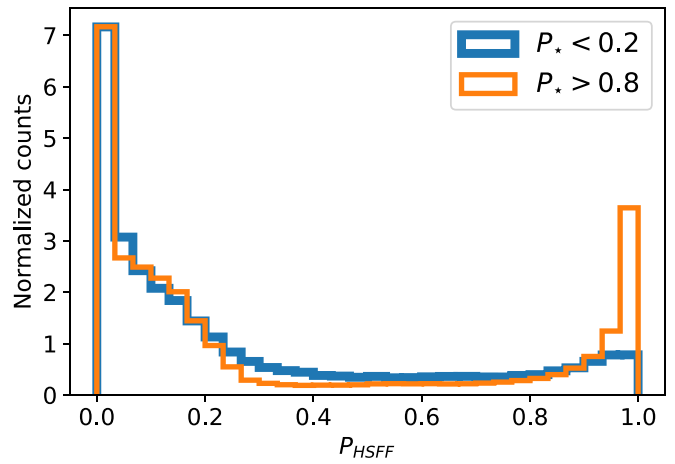
Next, we discuss the issue of missing data in PS1-DR2. For each source in the catalog of Appendix A, we add a data-quality flag  $Q$ ; where  $Q = 0$  indicates a dataset with complete information for all the five Pan-STARRS photometric filters, while  $Q > 0$  indicates that some data are missing. PS1-DR2 offers two flux measurements (i.e., PSF and Kron) for each of the five filters, for a total of 10 flux measurements per source with complete data. In the following, the value of the  $Q$  variable quantifies the number of missing flux measurements in any filter (so that, for example,  $Q = 1$  means that one flux measurement is missing, etc.). More details on the quality flag column are provided in Appendix A. For sources with  $Q > 0$ , we fill in the missing information by employing a linear interpolation of the spectral energy distribution. In total, 53% of PS1-DR2 sources have  $Q = 0$ . The classification probability



**Figure 14.** Distribution of the classification probability for PS1-DR2 galaxies (i.e., sources with  $P_* < 0.2$ ) of having HSFF ( $P_{\text{HSFF}}$ , upper panel) and of being spiral ( $P_{\text{spiral}}$ , lower panel) for different values of  $Q$ . We estimate the classification probability with the RF model developed in Sections 4 and 3 for the upper and lower panel, respectively. Large  $P_{\text{HSFF}}$  suggests a high probability that the galaxy is HSFF. Similarly, large  $P_{\text{spiral}}$  suggests a high probability that a galaxy is S. Larger font for axis labels. Poorer datasets ( $Q > 0$ ) tend to be associated with probability values closer to random guessing  $P = 0.5$ , as expected.

of the RF model for objects with  $Q > 0$  should be treated with caution. We quantify this statement below.

Figure 14 (upper panel) shows the distribution of the RF classification probability for PS1-DR2 galaxies to have HSFF. PS1-DR2 galaxies with complete information (i.e.,  $Q = 0$ ) are characterized by a bimodal classification probability distribution with one peak around 0 and a second peak around 1. This result suggests that galaxies with complete data are reliably classified as either having HSFF or LMSFF with high confidence. Instead, galaxies with  $Q > 0$  are more clustered around the region of random guessing  $P_{\text{HSFF}} = 0.5$ , as expected from their poorer data quality. The median classification confidence of PS1-DR2 galaxies with HSFF or LMSFF is as follows: galaxies with  $Q = 0$  ( $Q > 0$ ) are classified with median classification confidence of 0.9 (0.75). Specifically, galaxies with  $Q = 1, 2, 3, 4$  are classified with decreasing median classification confidence of 0.81, 0.72, 0.67, 0.65, respectively. We note that the larger median classification confidence in the cross-validation set of 0.94 of Section 4 originates from the fact that the cross-validation set is biased toward brighter sources, which are easier to classify. Another



**Figure 15.** Classification probability distribution of PS1-DR2 sources of having HSFF for different cuts on  $P_*$ . Blue and orange lines are used for  $P_* < 0.2$  and  $P_* > 0.8$ , respectively. Note that the two distributions are slightly different at  $P_{\text{HSFF}} > 0.9$ .

key difference between the two sets is that the sources in the cross-validation set are all galaxies, while the application set has some level of contamination by stars, even after we filter on  $P_*$ . A more detailed description of this effect is provided in Section 5.2.

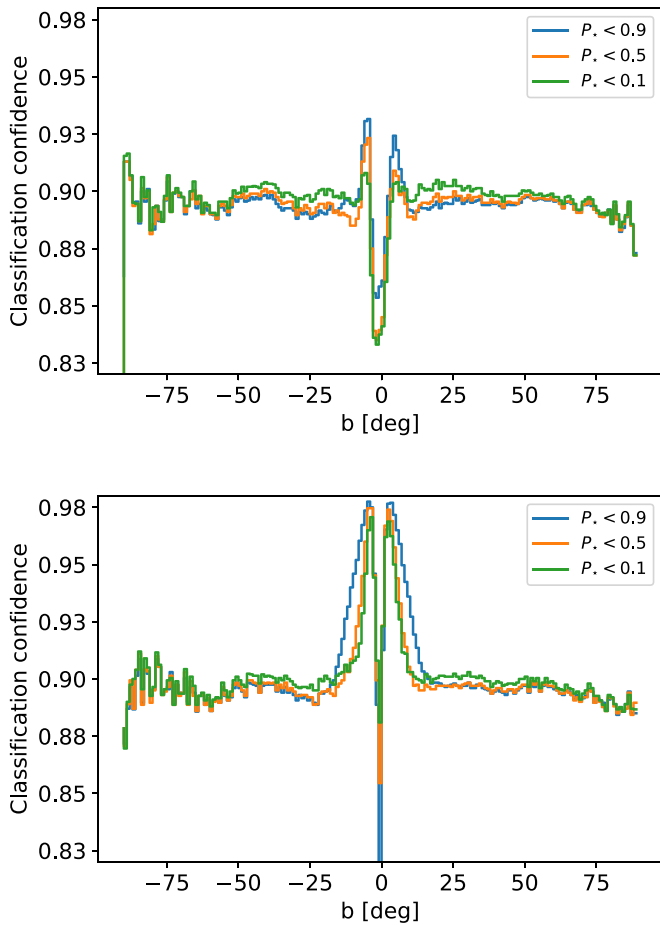
In Figure 14, lower panel, we perform a similar exercise for  $P_{\text{spiral}}$  and we compute the classification probability for different  $Q$  values. As before, datasets with missing values are associated with classification probabilities more clustered around the value of random guessing, as expected (median of 0.64 for  $Q = 0$ , and median of 0.59 for  $Q > 0$ ).  $P_{\text{spiral}}$  never approaches 0. The minimum value of  $P_{\text{spiral}}$  in the sample is 0.04. This is likely to be due to the difference between the cross-validation set and the whole PS1-DR2 dataset.

Finally, in Figure 15 we explore the effect of increasing the sample contamination with stars by applying different cuts on  $P_*$ . We consider two different scenarios: sources with  $P_* < 0.2$  (i.e., most likely galaxies) and sources with  $P_* > 0.8$  (i.e., most likely stars). The classification probability distributions of Figure 15 show some level of dependency on the cut on  $P_*$ . Not surprisingly, there are some Galactic stars that have colors which are able to perfectly mimic the colors of both highly star-forming and quiescent galaxies. For sources that are likely Galactic stars (i.e., with large  $P_*$ ) the  $P_{\text{HSFF}}$  value is also very likely to be meaningless. Taken at face value, sources with  $P_* > 0.8$  are more likely to be considered to have HSFF than sources with  $P_* < 0.2$ . Therefore, we recommend only using the  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  values of sources that have small values of  $P_* < 0.2$  (outside the plane of the Galaxy, as detailed in the following section).<sup>16</sup> We conclude by noting that despite intrinsic differences between the training/testing dataset and the entire PS1-DR2 catalog, our algorithm is able to achieve a large median classification confidence of  $\sim 0.9$  for galaxies with complete data in the PS1-DR2 release (i.e., with  $Q = 0$ ).

## 5.2. Effects of Galactic Extinction

Galactic extinction impacts the classification confidence of PS1-DR2 sources because it directly affects the observed colors of celestial objects outside the Galaxy. In this section, we

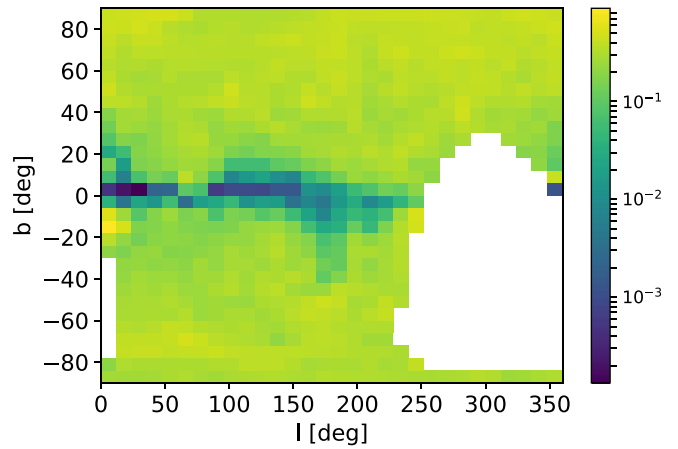
<sup>16</sup> More detailed user guidelines are provided in Appendix B.



**Figure 16.** Classification confidence for the PS1-DR2 sources of having HSFF before (upper panel) and after (lower panel) Galactic extinction correction. We plot the classification confidence for different values of  $P_*$  ( $P_* < 0.1$ ,  $P_* < 0.5$ ,  $P_* < 0.9$ ) and for  $Q = 0$ . Upper panel: for low Galactic latitudes  $|b| \lesssim 8^\circ$  the classification confidence differs significantly from the mean value due to the large Galactic reddening. By constraining on the non-stellar nature of the objects of interest  $P_* < 0.1$  results in less pronounced peaks, and vice versa. Lower panel: applying the Galactic extinction correction has the effect of mitigating the large decrease of classification confidence around  $b \sim 0^\circ$ . However, it also produces more pronounced “wings,” which are likely to be associated with contamination by stellar objects.

discuss the effects of Galactic extinction on our classification capabilities and we quantitatively explore the possibility of applying an extinction correction to PS1-DR2 data to improve on the performance of our algorithm at low Galactic latitudes. Galactic extinction has no effect on the training/testing sets because the data contains only objects at high Galactic latitudes  $|b| > 15^\circ$ , which are not significantly affected by Galactic reddening. In this section, we explore the effects of Galactic reddening on our classification performance using the HSFF versus LMSFF classifier of Section 4. Analogous results hold for the morphology classifier of Section 3. Consequently, we do not include a specific discussion on the effects of the Galactic extinction correction for the RF morphology classifier (which is also outperformed by the HSFF versus LMSFF classifier, as discussed below).

We explore the performance of our classification algorithm as a function of Galactic latitude  $b$  in Figure 16, where we plot the classification confidence of PS1-DR2 sources (for  $P_* < k$  with  $k = 0.1, 0.5, 0.9$ ) with complete data. We find that before applying any extinction correction for low Galactic latitudes in



**Figure 17.** Bidimensional histogram of the ratio between the number of HSFF galaxies (here defined as  $P_{\text{HSFF}} > 0.6$ ) and LMSFF galaxies ( $P_{\text{HSFF}} < 0.4$ ) for PS1-DR2 galaxies ( $P_* < 0.2$ ) as a function of galactic coordinates. The white region is outside the PS1-DR2 footprint (corresponding to  $\delta < -30^\circ$ ). The HSFF/LMSFF ratio in the Galactic plane is smaller than in the rest of the sky. This effect is an expected result from severe Galactic reddening, which artificially reduces the number of objects with observed blue colors.

the range  $-8^\circ \lesssim b \lesssim 8^\circ$ , the classification confidence significantly differs from the mean value and lies below the mean value at  $|b| \lesssim 1^\circ$ ; reaching a minimum of  $\sim 0.84$  (Figure 16, upper panel). Furthermore, there are two peaks of high classification confidence around  $b = -5.5$  and  $b = 5.5$  that result from the combined effects of high extinction and large contamination by stars. The amplitude of these two peaks is sensitive to the assumed upper cut on  $P_*$  (and hence to the allowed level of contamination by stellar objects). By constraining on the non-stellar nature of the objects of interest,  $P_* < 0.1$  directly results in less pronounced peaks, and vice versa (Figure 16, upper panel).

One way of looking at this result is that in the absence of any extinction correction, along the Galactic plane our algorithm very confidently mistakes highly reddened stars as early-type galaxies. Therefore, we may infer that the peaks originate from the fact that there is a *certain amount* of extinction that makes reddened stars mimic the colors of early-type galaxies. This effect happens at a small range of Galactic latitudes. The reddening is more extreme at even lower Galactic latitudes, and the reddened stars no longer look like early-type galaxies. This explains the location and presence of the peaks, the presence of the deep minimum at very small Galactic latitudes, and the fact that by filtering out stars we remove the peaks. In this latter case, we are changing the underlying colors of the population, which is no longer a population of stars but galaxies. This implies that that amount of galactic reddening will no longer be able to accurately mimic the colors of an early-type galaxy.

We further visualize the impact of Galactic extinction on our classification in Figure 17 by showing the bidimensional distribution of the ratio between the number of HSFF galaxies (defined here as  $P_{\text{HSFF}} > 0.6$ ) and galaxies with LMSFF ( $P_{\text{HSFF}} < 0.4$ ) in the sky. As before, for this test we select galaxies ( $P_* < 0.2$ ) with complete information ( $Q = 0$ ). As expected, in proximity to the Galactic plane, the HSFF/LMSFF ratio is significantly lower than average. In this region of the sky, the number of sources classified as LMSFF is significantly (and artificially) larger due to the redder observed colors.

Next, we quantify the amount of Galactic extinction along the line of sight for each source in the PS1-DR2 using the

extinction map by Schlafly & Finkbeiner (2011). We then run our RF classification algorithm on the extinction corrected PS1-DR2 photometry and we compare our galaxy-classification results to the pre-extinction correction results. We are also aware that most of the sources in the Galactic plane are stars for which the extinction correction along the line of sight is only approximate. Figure 16, lower panel, shows the resulting classification confidence after the Galactic extinction correction has been applied. We find that while the depth and width of the absolute minimum of the classification confidence at  $|b| \sim 0^\circ$  seem to benefit from the extinction correction, the two peaks at  $|b| \sim 5^\circ$  are largely unaffected (if not even strengthened). The larger width of the peaks at  $|b| \sim 5^\circ$  most likely results from the fact that we are artificially creating more stars with the same colors as early-type galaxies. We conclude that the anomalous behavior of the classification confidence around the Galactic plane is mainly driven by a large contamination of stars. Since we do not obtain significantly better performance with extinction corrections, we present the classification catalog without applying any Galactic extinction correction and we advise the user to be very selective on  $P_*$ , especially in the Galactic plane. A reasonable cut may be  $P_* < 0.2$  outside the Galactic plane ( $b < -8^\circ$  or  $b > 8^\circ$ ) and  $P_* < 0.1$  in the plane (here defined as  $-8^\circ < b < 8^\circ$ ). We also suggest trusting more our inferences on sources with  $Q = 0$ .

## 6. Using $P_{\text{HSFF}}$ and $P_{\text{spiral}}$ for Prompt Supernova Classification

The immediate goal of this paper is to characterize the star formation properties of galaxies within PS1-DR2. In this section we carry out a simple exercise that serves as a proof of concept to highlight the predictive power of  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  in the context of supernova (SN) typing (core-collapse versus thermonuclear SNe) at the time of their first detection. This is part of a larger effort aimed at classifying transients by combining information on the transient’s environment and their photometric evolution. Specifically, we will show how the  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  scores can be used to statistically infer the SN type. Because these scores are available from pre-explosion PS1 images and can readily be associated to the host-galaxy of a newly identified transient, these scores may be used as a useful tool to improve our capabilities of prompt classification of transients at the time of their first detection. However, a detailed analysis of the host-galaxy properties and their connection to the transient properties is beyond the scope of this work and will be addressed in a forthcoming paper. Here, we focus our analysis on the relationship between the star formation properties and morphological properties of a galaxy, and on the probability that it will host core-collapse or thermonuclear stellar explosions. We will use SN spectroscopic classifications from both a magnitude-limited untargeted transient survey (ZTF) and a galaxy-targeted nearby supernova survey (LOSS).

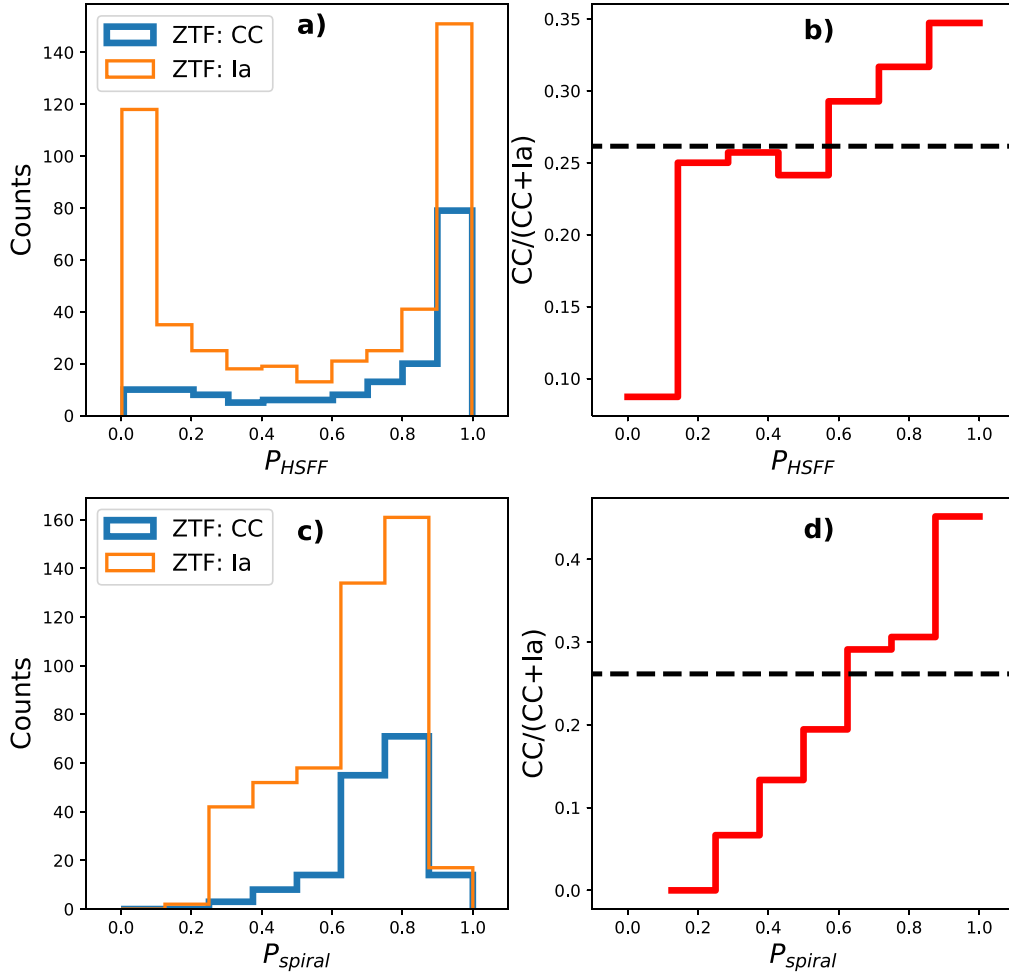
We start by considering spectroscopically classified SNe from the ZTF Bright Transient Survey (BTS, Fremling et al. 2019). The ZTF-BTS contains transients brighter than 18.5 mag at peak, at a distance corresponding to  $z \lesssim 0.15$  ( $d \lesssim 700$  Mpc). We associate each spectroscopically classified SN in the ZTF-BTS catalog with its host galaxy in PS1-DR2 (and its respective  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  scores). We carry out the host/SN association with a method developed by M. Stroh et al. (2020, in

preparation), which is based on Bloom & Kulkarni (2001). The ZTF-BTS catalog provides the PS1-DR2 host galaxies. Here, we use the Stroh et al. association method because we plan to expand this work to other transients and other surveys. This completely automatic procedure leads to an association that is consistent with the host galaxies provided by ZTF-BTS. This algorithm identifies the likely host galaxy as the galaxy with the lowest chance coincidence probability ( $P_{\text{cc}} = 1 - e^{-\pi R_e^2 \sigma(\leq m)}$ ) where  $\sigma(\leq m)$  is the galaxy number density as given by Berger (2010). Following Blanchard et al. (2016), the effective radius,  $R_e = \sqrt{R^2 + (2.5R_{\text{kron}})^2}$  where  $R$  is the angular separation between the ZTF host galaxy position and the PS1-DR2 potential host galaxy, while  $R_{\text{kron}}$  is the PS1-DR2 g Kron radius of the galaxy. We selected host galaxies with complete data in PS1-DR2 and obtain a sample of 162 core-collapse SNe (CCSNe, including types II, II-87A, IIb, IIn, Ib, Ib/c, Ibn, Ic, Ic-BL, Ic-pec, SLSN-I, SLSN-II) and 464 thermonuclear SNe “Ia” in short, including branch-normal Ia, Ia-02cx, Ia-91T, Ia-91bg, Ia-SC, Ia-CSM). The final sample contains 26% CCSNe by number. The median SN distance of the sample is  $\sim 250$  Mpc.

Figure 18 (panel (a)) shows the distribution of  $P_{\text{HSFF}}$  for the host galaxies of CCSNe and SNe Ia. Most host galaxies of SNe Ia have  $P_{\text{HSFF}} < 0.1$  or  $P_{\text{HSFF}} > 0.9$ , while CCSNe are mainly associated with actively star-forming galaxies with  $P_{\text{HSFF}} > 0.9$ . In Figure 18, panel (b), we plot the fractional number of CCSNe (i.e., the ratio between the number of CCSNe and the number of SNe for each bin of the histogram) as a function of  $P_{\text{HSFF}}$ . Because CCSNe constitute  $\sim 26\%$  of the sample, this fraction indicates the level that one would obtain by random guessing (indicated by the horizontal dashed line in Figure 18). This figure shows that if we select galaxies with  $P_{\text{HSFF}} < 0.1$ , the fraction of CCSNe drops to  $\text{CC}/(\text{CC} + \text{Ia}) \sim 7\%$ . At higher  $P_{\text{HSFF}}$ , the fraction of CCSNe increases. In particular, we find that galaxies with  $P_{\text{HSFF}} > 0.8$  have a large CCSNe fraction of  $\text{CC}/(\text{CC} + \text{Ia}) \sim 30\%–35\%$ . In Figure 18 (panels (c) and (d)), we perform a similar exercise using  $P_{\text{spiral}}$ . We find that the host galaxies of Ia SNe cover a wide range of  $P_{\text{spiral}}$  (i.e., SNe Ia are hosted in early and late-type galaxies), while CCSNe are mainly hosted in galaxies with large  $P_{\text{spiral}}$ , as expected. As before, the  $\text{CC}/(\text{CC} + \text{Ia})$  fraction is a monotonically increasing function of the  $P_{\text{spiral}}$  score. Specifically, for  $P_{\text{spiral}} > 0.9$  we find a ratio of  $\text{CC}/(\text{CC} + \text{Ia}) \approx 50\%$ , while for  $P_{\text{spiral}} < 0.4$  the ratio is  $\text{CC}/(\text{CC} + \text{Ia}) \approx 10\%$ .

We perform a similar analysis using the nearby SN sample from LOSS, which is a galaxy-targeted search for SNe in the local universe at  $d \leq 200$  Mpc (Leaman et al. 2011). This sample includes SNe that are significantly closer than those in the ZTF sample. As before, we associate each SN with its host galaxy in PS1-DR2, and the relative  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  scores. The original LOSS sample of spectroscopically classified SNe consists of 929 sources. Of these, we selected 517 associations with good-quality PS1-DR2 photometry ( $Q = 0$ ). Our final sample consists of 249 SNe Ia and 268 CCSNe (i.e., the CCSN fraction by number is  $\approx 48\%$ ).

Figure 19 (panels (a) and (c)) shows the  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  distributions of the host galaxies in our sample, while panels (b) and (d) show the fractional number of CCSNe ( $\text{CC}/(\text{CC} + \text{Ia})$ ) as a function of both scores. As for the ZTF sample,  $\text{CC}/(\text{CC} + \text{Ia})$  is a (mostly) monotonic function of  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$ . We interpret that the roughly “flat” distribution of  $P_{\text{HSFF}}$  of CCSN host-galaxies is due to the fact that the galaxies in the

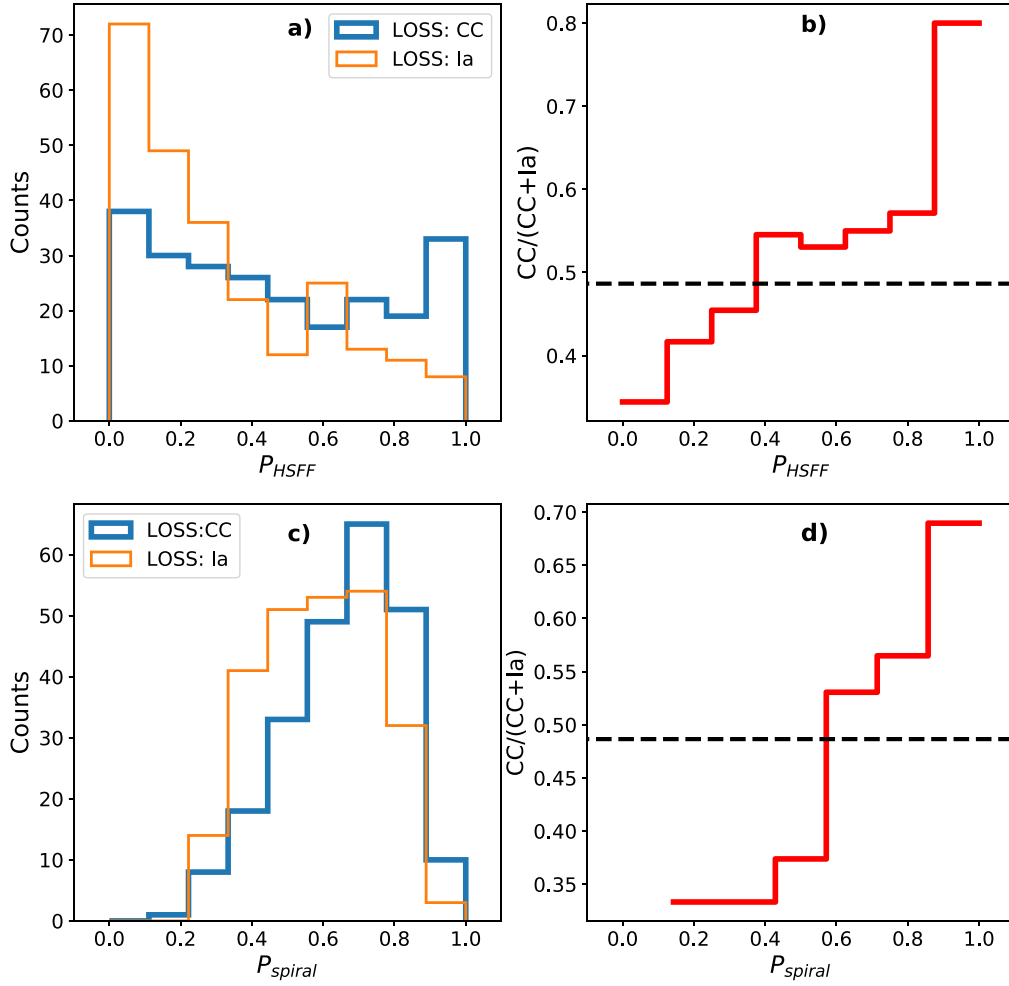


**Figure 18.** Panel (a):  $P_{\text{HSFF}}$  distribution of the host galaxies of SNe from the ZTF-BTS. Orange (Blue) thick lines show galaxies associated with SNe Ia (CC-SNe). Panel (b) red line: fraction of CC-SNe ( $\text{CC}/(\text{CC} + \text{Ia})$ ) as a function of  $P_{\text{HSFF}}$ . Horizontal black-dashed line: fraction of CC-SNe in the ZTF sample that we consider here, which corresponds to the random guessing level. Panel (c):  $P_{\text{spiral}}$  distribution of SN host galaxies from the ZTF magnitude-limited catalog. Panel (d), red line: fraction of CC supernovae ( $\text{CC}/(\text{CC} + \text{Ia})$ ) as a function of  $P_{\text{spiral}}$ . In panels (b) and (d), any significant departure of the  $\text{CC}/(\text{CC} + \text{Ia})$  fraction from the horizontal black-dashed line can be considered an improvement over random guessing. As discussed in Section 5, the  $P_{\text{spiral}}$  score never reaches 0 because of the intrinsic difference between the cross-validation and the entire PS1-DR2 dataset.

training sample have significantly larger distances than the nearby galaxies targeted by the LOSS. Indeed, the most distant SN in the LOSS sample (at  $d \sim 200$  Mpc) is closer than  $\sim 90\%$  of the galaxies in the training set. Consequently, in the training sample the relationship between the Kron and PSF photometry is different to that of the large, well-resolved, nearby galaxies in LOSS. Any significant difference between the training set and the actual sample leads the algorithm to “confusion,” the manifestation of which is this flat distribution of scores. With this caveat in mind, it is still interesting to note that  $\text{CC}/(\text{CC} + \text{Ia}) \sim 80\%$  for  $P_{\text{HSFF}} > 0.9$  (which is significantly above the 52% value expected for random guessing), and that the fraction of CCSNe is suppressed to  $\text{CC}/(\text{CC} + \text{Ia}) \sim 40\%$  for host galaxies with  $P_{\text{HSFF}} < 0.2$ . As expected, the  $P_{\text{spiral}}$  distribution is skewed toward large values for CCSNe. For  $P_{\text{spiral}} < 0.4$  we find  $\text{CC}/(\text{CC} + \text{Ia}) \sim 33\%$  below the 52% value of random guessing), while at large  $P_{\text{spiral}} > 0.85$  the fraction of CCSNe is highly enhanced to  $\text{CC}/(\text{CC} + \text{Ia}) \sim 70\%$ .

The predictive power of the host galaxy morphology in SN typing (CC versus Ia SNe) was first quantified by Foley & Mandel (2013) on the LOSS sample. Foley & Mandel (2013) (their Figure 1, upper panel) showed that in their “full” LOSS

sample which contains 41% of Ia SNe by number, the fraction of SNe Ia in E/S0 galaxies is in the range  $\sim 65\%–100\%$  (corresponding to  $\text{CC}/(\text{CC} + \text{Ia}) \sim 0\%–35\%$ ), decreasing to  $\lesssim 20\%$  (or  $\text{CC}/(\text{CC} + \text{Ia}) \gtrsim 80\%$ ) in Sbc/Sb/Scd/Irr galaxies. While it is not possible to directly compare our results to Figure 1 of Foley & Mandel (2013), it is interesting to note that our RF classifiers that are uniquely based on host-galaxy colors reach comparable purity levels at the extremes of the  $P_{\text{spiral}}$  or  $P_{\text{HSFF}}$  distributions. Foley & Mandel (2013) further employed a Naive Bayes classifier that leverages the transient’s contextual information (e.g., the host galaxy’s morphology, absolute magnitude ( $M_k$ ), colors ( $B_0 - K$ ), offset from host-galaxy’s nucleus and pixel rank) for SN Ia identification. Their Naive Bayes classifier returns the probability ( $p_{\text{Ia}}$ ) for each LOSS SN of being a Ia SN. These authors found that 30% of SNe in their sample have  $p_{\text{Ia}} > 0.5$ ; of these, 71% are SNe Ia. This result compares favorably to the random guessing level of  $P(\text{Ia}) = 41\%$ . For the same sample, 21% of SNe have  $p_{\text{Ia}} < 0.1$ , 84% of which are CC SNe. These findings are the result of the combination of inferences obtained from the different sources of contextual information listed earlier (including detailed host-galaxy morphology classification, the transient’s distance, and absolute magnitudes  $M_k$ ). These



**Figure 19.** Panel (a):  $P_{\text{HSFF}}$  distribution of host galaxies of SNe from the galaxy-targeted LOSS. Orange (Blue) thick lines: galaxies associated with SNe Ia (CC-SNe). Panel (b), red line: fraction of CC-SNe ( $\text{CC}/(\text{CC} + \text{Ia})$ ) as a function of  $P_{\text{HSFF}}$ . Horizontal black-dashed line: fraction of CC-SNe in the LOSS sample that we study here, which corresponds to the random guessing level. Panel (c):  $P_{\text{spiral}}$  distribution of host galaxies of SNe from the LOSS catalog. Panel (d), red line: fraction of CC supernovae ( $\text{CC}/(\text{CC} + \text{Ia})$ ) as a function of  $P_{\text{spiral}}$ . In panels (b) and (d), any significant departure of the  $\text{CC}/(\text{CC} + \text{Ia})$  fraction from the horizontal black-dashed line may be considered an improvement over random guessing. We interpret that the “flat” distribution of  $P_{\text{HSFF}}$  of CCSN host-galaxies is due to the fact that the galaxies in the training sample have significantly larger distances than the nearby galaxies targeted by the LOSS.

features are not available in main wide-field transient surveys. This implies that the Foley & Mandel (2013) methodology in its current form cannot easily be extended to very large datasets, such as those of the LSST. Simplified approaches that rely on minimal contextual information (e.g., colors) have the advantage that they are directly applicable to most transients surveys.

The important conclusion from these two exercises on the LOSS and ZTF-BTS samples is that by selecting on the  $P_{\text{HSFF}}$  or  $P_{\text{spiral}}$  scores of SN host galaxies, it is possible to artificially and significantly enhance or suppress the fraction of CCSNe (or thermonuclear SNe) with respect to random guessing. This result demonstrates that it is possible to improve on the SN classification at the time of their first detection by using the available information on their large-scale environments processed with machine-learning algorithms (i.e., no-human in the loop).

## 7. Summary and Conclusions

Machine learning is becoming a fundamental tool in a variety of fields in astrophysics, from exoplanet discovery to galaxy and transient classification. In this paper, we have

developed two machine-learning algorithms and presented the classification of galaxies in the Pan-STARRS  $3\pi$  survey based on their morphology and recent star formation history. Specifically, we have trained and tested two RF models on a subsample of the PS1-DR2 galaxies using PS1-DR2 colors as input features for the RF classifiers, and using labels from the Huertas-Company dataset (for galaxy morphologies) and from the New York University Value-Added Galaxy Catalog (NYU-VAGC, for the fraction of star formation occurred in the last 300 Myr). We have obtained a classification accuracy of 78% when discriminating between elliptical and spiral galaxies in the cross validation set. The classification accuracy is 89% when discriminating between galaxies with high and low-to-moderate star formation fraction (HSFF versus LMSFF) in the cross-validation set. We then applied both RF models to the entire PS1-DR2 catalog to determine the probability that each galaxy is spiral ( $P_{\text{spiral}}$ ) and whether or not it has a HSFF ( $P_{\text{HSFF}}$ ). We present our classifications in a catalog with a structure as outlined in Appendix A. User guidelines are also described in Appendix A.

We have applied the two RF classifiers to host galaxies of two SN samples from the ZTF-BTS and LOSS. We have also

demonstrated that the colors of the transient’s host galaxies can be used to statistically infer their star formation and morphological properties in a way that can be used to aid transient classification at the time of the first detection (in line with the initial study by Foley & Mandel 2013). The ZTF-BTS and the LOSS samples contain core-collapse SNe (CCSNe) and stellar explosion of thermonuclear origin. For both the ZTF-BTS (Fremling et al. 2019) and LOSS (Leaman et al. 2011) samples, we find that  $P_{\text{spiral}}$  and  $P_{\text{HSFF}}$  are highly correlated with the fraction of CCSNe.

In particular, for the brightness-limited SN sample from ZTF-BTS: by selecting host galaxies with  $P_{\text{HSFF}} > 0.8$ , we obtain a  $\approx 10\%$  larger fraction of CCSNe with respect to random guessing; while for  $P_{\text{HSFF}} < 0.1$ , we obtain a  $\approx 20\%$  lower CCSN fraction with respect to random guessing. Furthermore, by selecting host galaxies with  $P_{\text{spiral}} > 0.9$ , we obtain a  $\approx 50\%$  fraction of CCSNe (which constitutes a  $\sim 24\%$  improvement with respect to random guessing). We obtain similar results for the galaxy-targeted SN sample from LOSS. In this case,  $\sim 70\%–80\%$  of SNe associated with likely spiral host galaxies ( $P_{\text{spiral}} > 0.9$ ) or galaxies with high star formation fraction ( $P_{\text{HSFF}} > 0.9$ ) are of core-collapse origin, compared with the 48% fraction of CCSNe in the sample.

Our work demonstrates that it is possible to achieve significant improvements in prompt SN classification by using available contextual information automatically processed with machine-learning algorithms. The host galaxy information from our catalog can thus be directly used to complement and improve the classification accuracy of existing algorithms that solely rely on the transient’s photometric properties. A key advantage of classifiers that will include inference from contextual information is related to the fact that (some of) the host galaxies properties are known at the time of the very first detection of a new transient, when the photometric information is exceedingly limited. In the current era of spectroscopically starved time-domain astronomy, the ability to promptly infer the nature of a large number of transients without spectroscopic follow up (or visual inspection of each individual host galaxy) is of paramount importance. Indeed, in the near future, surveys such as the LSST carried out on the Vera C. Rubin Observatory will dramatically increase the discovery rate of transients by producing  $\approx 10^6$  alerts per night, making a systematic transient spectroscopic-classification unviable. In a future paper, we will extend the use of contextual information for prompt transient classification to include other properties of the large-scale environments of a variety of astronomical transients.

We thank the referee for their constructive criticism and comments that improved the quality of this paper. The authors thank Vicky Kalogera, Wen-fai Fong and Cristiano Guidorzi for discussions on this project. This work is supported by the Heising-Simons Foundation under grant #2018-0911 (PI: Margutti). R.M. is grateful to KITP for hospitality during the completion of this paper. This research was supported in part by the National Science Foundation under grant No. NSF PHY-1748958. R.M. acknowledges support by the National Science Foundation under Award No. AST-1909796. Raffaella Margutti is a CIFAR Azrieli Global Scholar in the Gravity & the Extreme Universe Program, 2019. A.A.M. is funded by the Large Synoptic Survey Telescope Corporation, the Brinson Foundation, and the Moore Foundation in support of the LSSTC Data Science Fellowship Program; he also receives

support as a CIERA Fellow by the CIERA Postdoctoral Fellowship Program (Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University). The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen’s University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

## Appendix A Catalog of Classifications of PS1-DR2 Sources

This catalog is organized as follows. The first column is the ID of the Pan-STARRS object. The second and third columns are the R.A. and decl. coordinates measured in degrees. The fourth column represents the probability for an object to be a point-source ( $P_*$ ), as derived by Tachibana & Miller (2018). The fifth column represents the probability for a source to be HSFF ( $P_{\text{HSFF}}$ ). The sixth column represents the probability for a source to be spiral ( $P_{\text{spiral}}$ ). Note that when  $P_* > 0.5$ , the values of  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  are meaningless. The seventh column is a completeness flag for the data. PS1-DR2 offers ten flux density measurements for each source: five PSF fluxes (for the  $g$ ,  $r$ ,  $i$ ,  $z$  and  $y$ , respectively) and five Kron fluxes (one for each photometric band). The completeness flag is expressed as a 10 digit binary number, where each digit tells if the data in a specific filter is present (0) or missing (1). The first five digits are related to the PSF fluxes for photometric bands in this order:  $g$ ,  $r$ ,  $i$ ,  $z$ ,  $y$ . The second five digits are associated with Kron fluxes for the same order of photometric bands. As reference, the binary number 0000000000 states that there are no missing data (parameter  $Q = 0$  in the paper), 0000010000 states that the  $g$ -band Kron flux is missing ( $Q = 1$ ), and 0100000010 states that the  $r$ -band PSF and the  $i$ -band Kron are missing ( $Q = 2$ ). Here  $Q$  represents the number of missing filters for each sources.

We recommend that the user mostly trusts  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  classifications for sources with  $Q = 0$ ,  $P_* < 0.2$  and Galactic latitude outside the range  $-8^\circ < b < 8^\circ$ . Classifications of sources close to the Galactic plane and classifications of sources with  $P_* > 0.2$  should be treated with caution. Users interested in supernova classification (CC versus thermonuclear) with  $P_{\text{HSFF}}$  and  $P_{\text{spiral}}$  may use the results in Figures 18 and 19 as a guideline. In Table 1, we report the first few rows of the catalog for display. Finally, we recommend using the  $P_{\text{spiral}}$  value for quick SN typing and not for quantitative studies of galaxy morphology (see Section 5). Our catalog of Pan-STARRS galaxies classifications is publicly available at [10.5281/zenodo.3990545](https://zenodo.org/record/3990545). The catalog is subdivided into 13 .csv files

**Table 1**  
Sample Table of the Released Catalog

ID	R.A.	Decl.	$P_*$	$P_{\text{HSFF}}$	$P_{\text{spiral}}$	Completeness Flag
115573500156866067	350.0157709	6.31298624	0.0167	0.3218	0.4726	0000010000
115591651428609170	165.14281046	6.33225118	0.9958	0.0018	0.3720	0000000000
115592747546664681	274.75467722	6.32846762	0.7686	0.6024	0.6924	0000011001
115591851329443064	185.13275190	6.32728953	0.64258	0.8345	0.924	0000000011
115582187787633432	218.77857803	6.31919698	0.01426	0.3189	0.587	0000000000

**Note.** Description of the Table in Appendix A.

and contains the IDs, positions, and classification probabilities. Each file is  $\approx 3.8$  GB in size.

## Appendix B





### RF Model for Sources in PS1-DR2 Based on the GZ CATALOG

The GZ project (e.g., Lintott et al. 2008) is a citizen science project with visual morphological classifications of more than 300,000 galaxies drawn from SDSS. In the first data release (GZ1, Lintott et al. 2011), citizens were asked to visually classify SDSS galaxies as either elliptical or spiral. In the second data release (GZ2, Willett et al. 2013), citizens were asked to answer several questions. The main questions were: (i) Is the galaxy simply smooth and rounded, with no sign of a disk? (ii) How rounded is it? (iii) Could this be a disk viewed edge-on? The full list of questions can be found in Figure 1 of Willett et al. (2013).

We train an RF model with GZ2 labels and PS1-DR2 features. We start by cross-matching GZ2 with PS1-DR2. For each common source, we retain features that are relevant to our subsequent analysis (colors and moments of the radiation intensity). GZ2 labels are not directly usable for our machine-learning classification task. GZ (and implicitly the Willett et al. 2013 catalog) provides likelihoods for galaxies being “Smooth” or “Features or Disk,” which can be used to separate E/S0 from S. Were E/S0 galaxies selected with a likelihood of “Smooth”  $> 50\%$ , and S galaxies with a likelihood of “Smooth”  $< 50\%$ . We follow the procedure described in Section 3 for preprocessing, training, and testing. We train and test an RF algorithm that classifies galaxies according to their morphology (E/S0 or S), and obtain an 80% classification accuracy in the cross-validation set.

Finally, we apply the RF classifier trained with both colors and moments on the PS1-DR2-HC catalog to the PS1-DR2-GZ2 dataset, and obtain a classification accuracy of 81%. We may explain the 4% larger classification accuracy in the HC dataset considering that the GZ2 labels that we defined might not directly map into morphological classes (E/S0 and S). Our model performs comparably well on the GZ sample (where labels are exact, even if they might not directly map to morphological classes) and on the HS sample (where labels are directly relevant to the morphological classification of galaxies, but are derived from a ML effort and are therefore subject to a level of uncertainty that we quantified in Section 3).

## ORCID iDs

A. Miller  <https://orcid.org/0000-0001-9515-478X>  
M. Stroh  <https://orcid.org/0000-0002-3019-4577>  
R. Margutti  <https://orcid.org/0000-0003-4768-7586>  
D. L. Coppejans  <https://orcid.org/0000-0001-5126-6237>

## References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543
- Arcavi, I., Wolf, W. M., Howell, D. A., et al. 2016, *ApJ*, **819**, 35
- Baldeschi, A., Elia, D., Molinari, S., et al. 2017a, *MNRAS*, **466**, 3682
- Baldeschi, A., Molinari, S., Elia, D., Pezzuto, S., & Schisano, E. 2017b, *MNRAS*, **472**, 1778
- Ball, N. M., Loveday, J., & Brunner, R. J. 2008, *MNRAS*, **383**, 907
- Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, *MNRAS*, **393**, 1324
- Berger, E. 2010, *ApJ*, **722**, 1946
- Blanchard, P. K., Berger, E., & Fong, W.-f. 2016, *ApJ*, **817**, 144
- Blanton, M. R., & Berlind, A. A. 2007, *ApJ*, **664**, 791
- Blanton, M. R., & Roweis, S. 2007, *AJ*, **133**, 734
- Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, *AJ*, **129**, 2562
- Bloom, J. S., & Kulkarni, S. R. 2001, in *Gamma-ray Bursts in the Afterglow Era*, ed. E. Costa, F. Frontera, & J. Hjorth (Berlin: Springer), 209
- Caon, N., Capaccioli, M., & D’Onofrio, M. 1993, *MNRAS*, **265**, 1013
- Carrasco-Davis, R., Cabrera-Vives, G., Förster, F., et al. 2019, *PASP*, **131**, 108006
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
- Charnock, T., & Moss, A. 2017, *ApJL*, **837**, L28
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, **450**, 1441
- Drout, M. R., Chornock, R., Soderberg, A. M., et al. 2014, *ApJ*, **794**, 23
- Fisher, A., Rudin, C., & Dominici, F. 2018, arXiv:1801.01489
- Foley, R. J., & Mandel, K. 2013, *ApJ*, **778**, 167
- Fremming, U. C., Miller, A. A., Sharma, Y., et al. 2019, arXiv:1910.12973
- Fukugita, M., Nakamura, O., Okamura, S., et al. 2007, *AJ*, **134**, 579
- Gal-Yam, A. 2019, *ARA&A*, **57**, 305
- Gal-Yam, A., Kasliwal, M. M., Arcavi, I., et al. 2011, *ApJ*, **736**, 159
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.; Berlin: Springer), <http://www-stat.stanford.edu/tibs/ElemStatLearn/>
- Huertas-Company, M., Aguerri, J. A. L., Bernardi, M., Mei, S., & Sánchez Almeida, J. 2011, *A&A*, **525**, A157
- Ishida, E. E. O., Beck, R., González-Gaitán, S., et al. 2019, *MNRAS*, **483**, 2
- Ivezić, Z., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Karpenka, N. V., Feroz, F., & Hobson, M. P. 2013, *MNRAS*, **429**, 1278
- Kessler, R., Bassett, B., Belov, P., et al. 2010a, *PASP*, **122**, 1415
- Kessler, R., Conley, A., Jha, S., & Kuhlmann, S. 2010b, arXiv:1001.5210
- Kron, R. G. 1980, *ApJS*, **43**, 305
- Leaman, J., Li, W., Chornock, R., & Filippenko, A. V. 2011, *MNRAS*, **412**, 1419
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, *MNRAS*, **410**, 166
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, **389**, 1179
- Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, *ApJS*, **225**, 31
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, **128**, 163
- Magnier, E. A., Schlafly, E., Finkbeiner, D., et al. 2013, *ApJS*, **205**, 20
- Margalef-Bentabol, B., Huertas-Company, M., Charnock, T., et al. 2020, *MNRAS*, **496**, 2346
- Masters, K. L., Mosleh, M., Romer, A. K., et al. 2010, *MNRAS*, **405**, 783
- Möller, A., & de Boissière, T. 2020, *MNRAS*, **491**, 4277
- Möller, A., Ruhlmann-Kleider, V., Leloup, C., et al. 2016, *JCAP*, **2016**, 008
- Moss, A. 2018, arXiv:1810.06441
- Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, *PASP*, **131**, 118002
- Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, *ApJS*, **236**, 9
- Newling, J., Varughese, M., Bassett, B., et al. 2011, *MNRAS*, **414**, 1987
- Ntampaka, M., Avestruz, C., Boada, S., et al. 2019, *BAAS*, **51**, 14
- Pasquet, J., Pasquet, J., Chaumont, M., & Fouchez, D. 2019, *A&A*, **627**, A21
- Pursiainen, M., Childress, M., Smith, M., et al. 2018, *MNRAS*, **481**, 894

- Quimby, R. M., Kulkarni, S. R., Kasliwal, M. M., et al. 2011, *Natur*, **474**, 487
- Saha, A., Matheson, T., Snodgrass, R., et al. 2014, *Proc. SPIE*, **9149**, 914908
- Saha, A., Wang, Z., Matheson, T., et al. 2016, *Proc. SPIE*, **9910**, 99100F
- Sako, M., Bassett, B., Becker, A., et al. 2008, *AJ*, **135**, 348
- Sako, M., Bassett, B., Connolly, B., et al. 2011, *ApJ*, **738**, 162
- Schanche, N., Collier Cameron, A., Hébrard, G., et al. 2019, *MNRAS*, **483**, 5534
- Schawinski, K., Lintott, C., Thomas, D., et al. 2009, *MNRAS*, **396**, 818
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, **737**, 103
- Shivvers, I., Zheng, W. K., Mauerhan, J., et al. 2016, *MNRAS*, **461**, 3057
- Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, *MNRAS*, **399**, 966
- Smith, K. W., Williams, R. D., Young, D. R., et al. 2019, *RNAAS*, **3**, 26
- Sooknunan, K., Lochner, M., Bassett, B. A., et al. 2018, arXiv:1811.08446
- Pravan, N., Milisavljevic, D., Reynolds, J. M., Lentner, G., & Linvill, M. 2020, *ApJ*, **893**, 127
- Steinhardt, C. L., Weaver, J. R., Maxfield, J., et al. 2020, *ApJ*, **891**, 136
- Sullivan, M., Howell, D. A., Perrett, K., et al. 2006, *AJ*, **131**, 960
- Tachibana, Y., & Miller, A. A. 2018, *PASP*, **130**, 128001
- Tampo, Y., Tanaka, M., Maeda, K., et al. 2020, *ApJ*, **894**, 27
- Tanaka, M., Tominaga, N., Morokuma, T., et al. 2016, *ApJ*, **819**, 5
- van der Wel, A. 2008, *ApJL*, **675**, L13
- Villar, V. A., Berger, E., Miller, G., et al. 2019, arXiv:1905.07422
- Villar, V. A., Nicholl, M., & Berger, E. 2018, *ApJ*, **869**, 166
- Walmsley, M., Smith, L., Lintott, C., et al. 2019, arXiv:1905.07424
- Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, *MNRAS*, **435**, 2835
- Young, C. K., & Currie, M. J. 1994, *MNRAS*, **268**, L11