





# Spectroscopic Confirmation of Obscured AGN Populations from Unsupervised Machine Learning

RAPHAEL E. HVIDING <sup>1,2</sup> KEVIN N. HAINLINE <sup>1</sup> ANDY D. GOULDING <sup>3</sup> AND JENNY E. GREENE <sup>3</sup>

<sup>1</sup>*Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*

<sup>2</sup>*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*

<sup>3</sup>*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*

(Received November 20, 2023; Revised January 12, 2024; Accepted February 6, 2024)

Submitted to AJ

## ABSTRACT

We present the result of a spectroscopic campaign targeting Active Galactic Nucleus (AGN) candidates selected using a novel unsupervised machine-learning (ML) algorithm trained on optical and mid-infrared (mid-IR) photometry. AGN candidates are chosen without incorporating prior AGN selection criteria and are fainter, redder, and more numerous,  $\sim 340$  AGN deg<sup>-2</sup>, than comparable photometric and spectroscopic samples. In this work we obtain 178 rest-optical spectra from two candidate ML-identified AGN classes with the Hectospec spectrograph on the MMT Observatory. We find that our first ML-identified group, is dominated by Type I AGNs (85%) with a  $< 3\%$  contamination rate from non-AGNs. Our second ML-identified group is comprised mostly of Type II AGNs (65%) with a moderate contamination rate of 15% primarily from star-forming galaxies. Our spectroscopic analyses suggest that the classes recover more obscured AGNs, confirming that ML techniques are effective at recovering large populations of AGNs at high levels of extinction. We demonstrate the efficacy of pairing existing WISE data with large-area and deep optical/near-infrared photometric surveys to select large populations of AGNs and recover obscured SMBH growth. This approach is well suited to upcoming photometric surveys, such as Euclid, Rubin, and Roman.

**Keywords:** Active galactic nuclei (16), Dimensionality reduction (1943), Galaxy spectroscopy (2171)

## 1. INTRODUCTION

Active galactic nuclei (AGN) are the most luminous sustained phenomena in the universe. Powered by accretion onto a central supermassive black hole (SMBH), AGNs are thought to have a significant impact on their host galaxy (for reviews, see Fabian 2012; Alexander & Hickox 2012). Despite their role in galaxy evolution, a uniform selection methodology for AGNs remains relatively challenging as selected samples of accreting SMBHs using differing techniques return distinct objects, sometimes with little-to-no overlap between criteria (see discussions in Hickox et al. 2009; Padovani et al. 2017; Hickox & Alexander 2018).

One of the most common forms of AGN identification comes from ultraviolet (UV) and optical photometric selection, which has been used over large areas to search for compact blue emission, a signature of power-law emission from the accretion disk (Schmidt & Green 1983). However, UV and optical photometric selection is sensitive to the obscuration of the central engine. Nuclear- or galactic-scale attenuating material can mask the signature of AGN activity, especially for low-luminosity or low-accretion-rate systems. Optical spectroscopy, on the other hand, is well suited for determining the contribution of an AGN even in the presence of obscuration through the measurement of enhanced nebular or coronal narrow emission lines that reveal the ionization signature of SMBH activity beyond the extincted nucleus, i.e. Type II AGNs. Unobscured AGNs can be detected through the identification of broad emission lines, referred to as Type I AGNs.

Large spectroscopic surveys have enabled the detailed characterization of considerable numbers of AGN. In particular, the Sloan Digital Sky Survey (SDSS; York et al. 2000) has facilitated the study of SMBH accretion, especially in Type I systems, through the analysis of over 750,000 AGNs (Lyke et al. 2020) detected as a part of the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) and the extended-BOSS (eBOSS; Zhao et al. 2016). However, these spectroscopic campaigns that target photometrically selected AGNs are biased against low-luminosity and/or obscured systems which may be below the brightness limit of the instrumentation and/or lack the characteristic compact blue emission that the surveys target. Mass- or magnitude-complete spectroscopic surveys, such as the SDSS Legacy Survey (Strauss et al. 2002), are limited to bright populations and to low redshifts ( $z \lesssim 0.3$ ).

Mid-IR photometric selection of AGNs, conversely, targets re-processed emission from circumnuclear dust heated by SMBH accretion and has traditionally been used as an effective technique for assembling large samples of AGNs at higher levels of obscuration (e.g. Lacy et al. 2004; Stern et al. 2005; Hickox et al. 2007). The recovery of obscured systems is of particular importance for exploring their role in galaxy evolution as obscured AGNs may represent a distinct phase of galaxy-SMBH coevolution. In these systems, material driven into the nucleus contributes both to rapidly feeding the central engine and extinguishing the optical, UV, and even X-ray signatures (Hickox & Alexander 2018).

The Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) satellite, which performed an all-sky survey at 3.4, 4.6, 12, and 22  $\mu\text{m}$  ( $W1$ ,  $W2$ ,  $W3$ , and  $W4$  respectively), enabled the selection of tens of millions of AGN candidates (e.g. Jarrett et al. 2011; Stern et al. 2012; Mateos et al. 2012; Assef et al. 2013, 2018). However, despite its ability to select obscured sources, mid-IR broadband photometry suffers severely from contamination from star-forming galaxies for all but the brightest AGNs (Donley et al. 2012; Mendez et al. 2013; LaMassa et al. 2019). Hviding et al. (2022) demonstrated that  $>80\%$  of spectroscopically-identified AGNs have mid-IR colors that are indistinguishable from those of star-forming galaxies down to a limiting optical brightness. In fact, the AGNs that were not selected were dominated by Type II, i.e. obscured, and lower-luminosity systems, suggesting there continues to exist a population of AGNs that cannot be selected through mid-IR- or optical-photometric selection alone.

This bias against obscured sources becomes even more urgent as ongoing and upcoming optical or near-infrared

(near-IR) imaging surveys from Hyper Suprime Cam (HSC; Miyazaki et al. 2018), Euclid (Laureijs et al. 2011), Rubin (Ivezić et al. 2019), SPHEREx (Doré et al. 2016), and Roman (Spergel et al. 2015) will result in large samples of galaxies without accompanying high-resolution ( $\mathcal{R} \sim 1000$ ) spectroscopy. In addition, without any new large-area mid-IR missions with improved spectroscopic or photometric capabilities, WISE-derived datasets remain the best source of mid-IR photometry over large-areas. It is therefore imperative to explore if existing mid-IR data may be leveraged with the next generation of survey data to develop new AGN selection techniques to recover populations of accreting SMBHs that are systematically missed in traditional selection.

Combining optical imaging and mid-IR photometry hopes to be an effective methodology for selecting obscured and low-luminosity AGNs over large-areas that cannot be selected from a single wavelength regime alone. In this work we test an unsupervised machine-learning (ML) approach for selecting AGN candidates from deep optical imaging from HSC paired with mid-IR photometry from WISE (Goulding et al., in prep.). Our approach enables us to both make use of the mid-IR data to detect obscured AGNs and the optical color and morphological data to disentangle the AGNs from star-forming galaxy contaminants. The candidate AGN classes exhibit higher number densities down to fainter magnitudes than comparable methods and it therefore becomes essential to validate our ML approach using deeper optical spectroscopy than previously available.

We obtain spectroscopic follow-up of ML-identified AGN candidates using the Hectospec spectrograph on the MMT Observatory. These observations can confirm the AGN nature of the selected sources, retrieve accurate redshifts, and measure key diagnostic emission-line ratios. By comparing the properties of the ML-selected AGNs with those of AGNs selected using traditional optical or mid-IR methods, we assess the effectiveness of the unsupervised ML techniques in identifying AGN populations that have been missed by previous surveys. This spectroscopic follow-up will provide an important test of the ML selection and advance our understanding of the diversity of AGN populations and their role in galaxy evolution.

In Section 2 we present a brief overview of the ML analysis and the datasets used to inform it that are fully described in Goulding et al. (in prep.). We select our follow-up targets and describe our follow-up spectroscopy in Section 3. Section 4 details our spectroscopic analysis, including spectral stacking and emission-line fitting. Finally, in Sections 5 and 6 we present our discussion and conclusions respectively. Throughout

Optical = Good for Type I AGN but not Obscured  
 MIR = Good for Obscured, but high SFing contamination, Also misses a lot of Type II

this work we use AB magnitudes for optical photometry, Vega magnitudes for mid-IR photometry, and air wavelengths for optical spectroscopy. We assume a flat Lambda Cold Dark Matter ( $\Lambda$ CDM) cosmology with  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_m = 0.3$ .

## 2. SAMPLE SELECTION

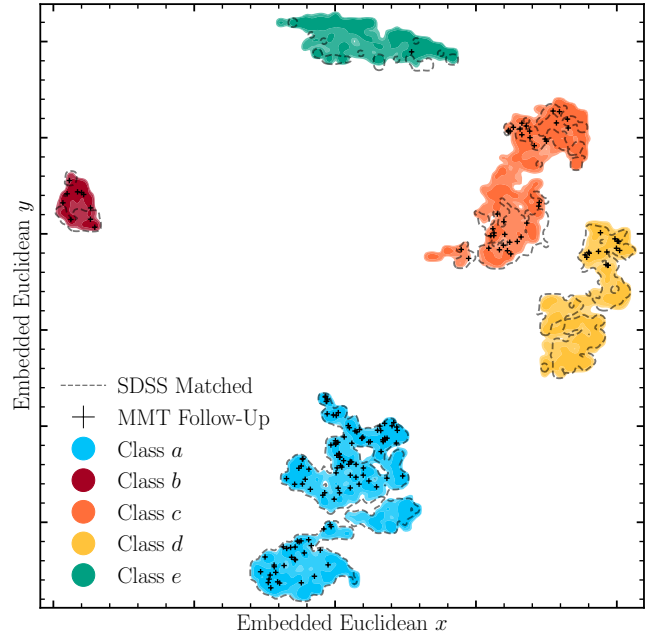
Our deep optical imaging from HSC enables us to make greater use of WISE mid-IR photometry. Typical mid-IR selection criteria require stringent SNR cuts on WISE photometry, such as  $\text{SNR} > 4-5$  in one or more WISE bands (Jarrett et al. 2011; Stern et al. 2012; Assef et al. 2018) or even  $\text{SNR} > 10$  in Mateos et al. (2012). Through the use of ML techniques to synthesize patterns in our optical and mid-IR datasets, we can incorporate mid-IR data that is at a lower significance than would be used for traditional mid-IR color criteria. By leveraging the rich photometric information available from multi-wavelength data, unsupervised ML techniques can help to uncover new and previously missed populations of AGNs over wide areas below the significance of existing WISE data. While the sample selection and ML pipeline is described in full detail in Goulding et al. (in prep.), we outline the parent sample in Section 2.1, summarize the ML procedure to generate our candidate AGN classes in Section 2.2, and present the properties of our classes in Section 2.3.

### 2.1. Parent Sample

Goulding et al. (in prep.) begin by selecting objects from the Hyper Suprime Cam Subaru Strategic Program (HSC SSP; Aihara et al. 2018), a large, deep optical imaging survey covering over  $1,000 \text{ deg}^2$ . The HSC SSP provides imaging in five broadband filters ( $g$ ,  $r$ ,  $i$ ,  $z$ , and  $y$ ; Kawanomoto et al. 2018) with a median five-sigma depth of 25.1 mag and a median seeing of  $\sim 0.6$  arcseconds. The data are taken from the HSC SSP Third Data Release (Aihara et al. 2022). Unless otherwise noted, in this work we use the HSC CModel magnitudes, calculated by fitting a linear combination of point-spread-function (PSF) convolved de Vaucouleurs and exponential profiles to the galaxy, to recover the total flux of extended sources.

The superb spatial resolution and limiting surface brightness of the HSC SSP enable morphological characterizations of the target systems. To generate a non-parametric morphology estimator we also make use of PSF magnitudes that measure the flux of an object assuming it is a point source. By taking the difference between the PSF and CModel magnitudes, we measure the deviation of the object from a point source, where a difference larger than zero, i.e. where the PSF magnitude is brighter than the CModel magnitude, indicates

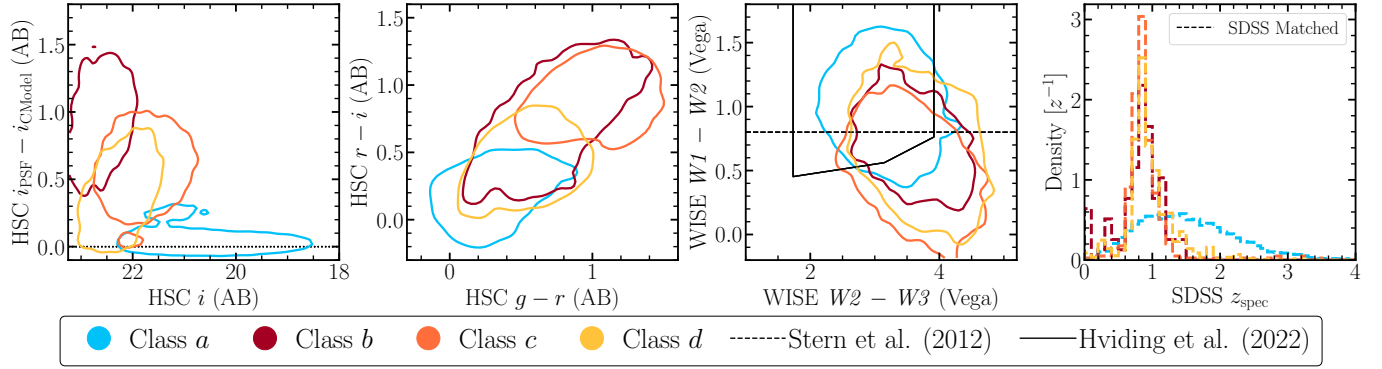
that the galaxy is more diffuse than a point source<sup>1</sup>. For a given galaxy, the larger the difference between the PSF and CModel magnitudes, the more flux lies beyond the PSF.



**Figure 1.** The embedded Euclidean space generated by UMAP from our AGN candidates classes, 3 and 10, after removing stars. We color shaded regions encompassing 68%, 95%, and 99.7% of the five subclasses,  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $e$ , identified using DBSCAN. The dashed contours encompass 99.7% of the matched SDSS spectroscopy in each subclass. Finally, we mark our MMT follow-up targets as plus symbols. Throughout this work we refer to subclass  $a$  as class I, combine subclasses  $b$ ,  $c$ , and  $d$  into class II, and identify class  $e$  as contaminants.

For our mid-infrared dataset we make use of both the allWISE (Cutri et al. 2021) and unWISE (Schlafly et al. 2019) catalogs. While both catalogs leverage data from the WISE cryogenic and NEOWISE (Mainzer et al. 2011) post-cryogenic surveys, the unWISE catalog provides higher resolution photometry down to fainter fluxes in  $W1$  and  $W2$  due to the coadding procedure and the preservation of the native WISE resolution. We make use of unWISE measurements for  $W1$  and  $W2$ , and allWISE measurements for  $W3$ . Due to the low sensitivity in  $W4$  and the brightness range in our follow-up sample, our analysis does not make use of any  $W4$  measurements.

<sup>1</sup> A value of less than zero may occur when site conditions at Subaru are better than the fitted PSF shape.



**Figure 2.** Optical magnitude-morphology (left), optical color-color (center left), mid-IR color-color (center right), and matched SDSS spectroscopic redshift (right) distributions for subclasses *a* (blue), *b* (red), *c* (orange), and *d* (yellow). We plot a contour encompassing 75% of each class. Mid-IR color-color contours are restricted to objects with  $\text{SNR} > 3$  in *W1*, *W2*, and *W3*. We plot the Hviding et al. (2022, solid) and Stern et al. (2012, dashed) mid-IR selections to highlight regions where typical mid-IR selected AGNs inhabit. In the leftmost panel we plot a dashed line corresponding to objects with equal PSF magnitudes, corresponding to point sources.

Goulding et al. (in prep.) matches to all galaxies in SDSS Data Release 16 (DR16; Ahumada et al. 2020) spectroscopy using a maximum likelihood approach that takes into account both position and flux. Through this work we will compare to the properties of the SDSS-matched spectroscopic samples as a comparison to our candidate AGN selections. SDSS is representative of optical selection of AGN and therefore highlights the limitations of traditional photometric selection paired with follow-up spectroscopy, which we discuss further in Section 2.3.

## 2.2. Machine Learning

Goulding et al. (in prep.) begins with the subsample of objects that have detections in all three mid-infrared bands. With our eight-band photometry (*g*, *r*, *i*, *z*, *y*, *W1*, *W2*, *W3*), Goulding et al. (in prep.) produces the following features:

- The 28 possible optical to optical, optical to mid-IR, or mid-IR to mid-IR colors.
- Five non-parametric morphology estimators for each of the optical bands, calculated by taking the difference between PSF and CModel magnitudes. A value larger than zero indicates the galaxy is more diffuse than a point source.
- The *r* band photometry as an absolute flux measurement.

Goulding et al. (in prep.) uses a dimensionality reduction algorithm, Uniform Manifold Approximation and Projection (UMAP; McInnes et al. 2018), along with a clustering algorithm, Density-Based Spatial Clustering of Applications with Noise (DBSCAN; Ester et al.

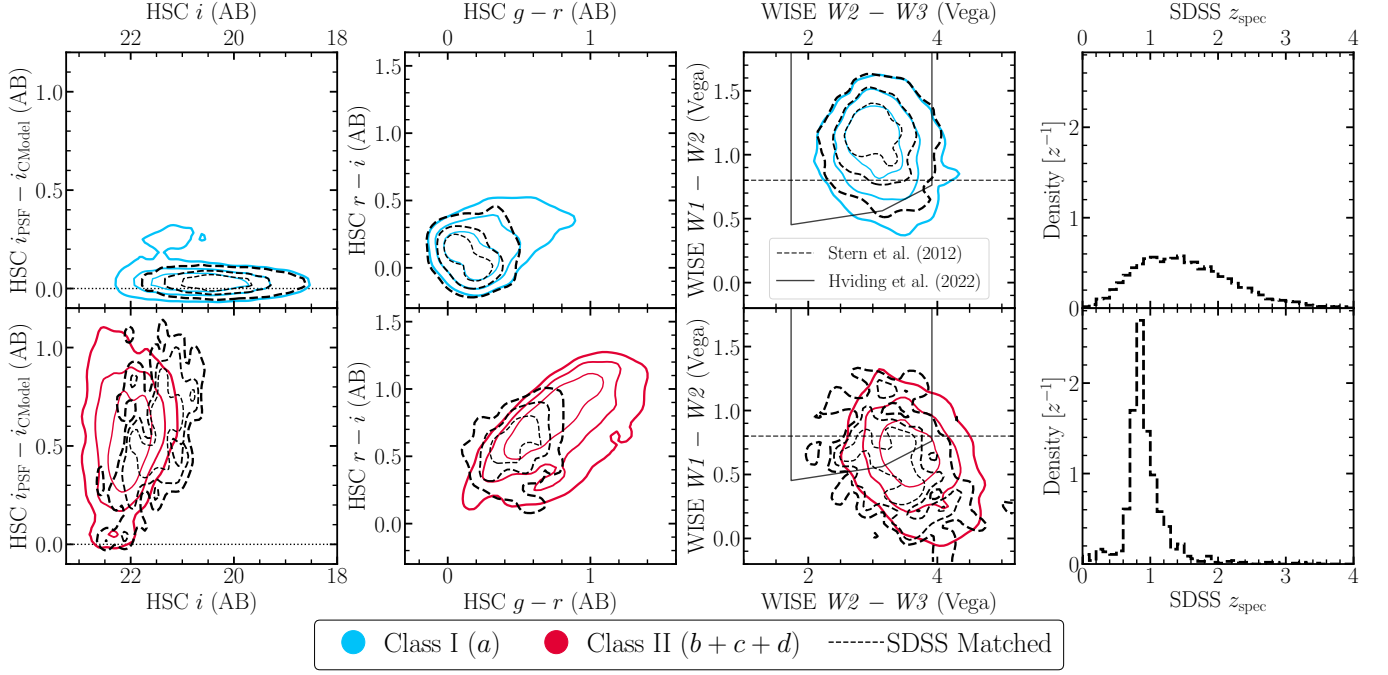
1996), to recover 16 classes, two of which are identified as potential AGN classes based on matched SDSS spectroscopy and X-ray datasets and together represent  $\sim 13\%$  of the parent sample.

Following the removal of stellar contaminants using a K-nearest neighbours approach (KNN; Fix & Hodges 1951; Cover & Hart 1967), Goulding et al. (in prep.) re-runs UMAP and DBSCAN on the candidate AGN classes to retrieve five total AGN candidate subclasses, *a*, *b*, *c*, *d*, and *e*, which we present in the embedded Euclidean space in Figure 1. By using existing SDSS spectroscopy, Goulding et al. (in prep.) concludes that class *e*, which makes up  $\sim 16\%$  of the AGN classes, is dominated by non-AGN sources, and label the subclass as contaminants. We therefore do not consider subclass *e* for the remainder of this work.

In Figure 2 we present the optical magnitude-morphology, optical color-color, mid-IR color-color distributions of subclasses *a*, *b*, *c*, and *d*. While our MMT spectroscopy, described in Section 3, is designed to select targets from subclasses *a*, *b*, *c*, and *d*, the majority ( $\sim 64\%$ ) of our follow-up targeted subclass *a* due to its relative brightness. In this work we aim to quantify the spectroscopic properties of the ML-selected AGN candidates, including fitting, stacking, and placing constraints on the AGN fractions in our ML-identified classes. It is therefore necessary to combine some of the subclasses to have sufficient spectra in each, albeit at the cost of characterizing each subclass individually.

We group subclasses *b*, *c*, and *d* together as these subclasses have less follow-up. Between subclasses *b*, *c*, and *d* we observe 63 spectra compared to 114 spectra observed in subclass *a*. While this risks combining subclasses with different intrinsic properties, we justify the combination due to the similarity in the *i*-band mag-





**Figure 3.** Optical magnitude-morphology (left), optical color-color (center left), mid-IR color-color (center right), and matched SDSS spectroscopic redshift (right) distributions for classes I (top; blue) and II (bottom; red). We plot contours encompassing 25%, 50%, and 75% of the classes (solid) and matched SDSS spectroscopy (dashed). Mid-IR color-color contours are restricted to objects with  $\text{SNR} > 3$  in  $W1$ ,  $W2$ , and  $W3$ . We plot the [Hviding et al. \(2022\)](#), (solid) and [Stern et al. \(2012\)](#), (dashed) mid-IR selections to highlight regions where typical mid-IR selected AGNs inhabit. Due to SDSS spectroscopic selection, matched objects do not span the entirety of the parameter space of class I and II. In particular SDSS spectroscopy is biased towards the bluest objects in optical color-color space.

nitude, WISE color, and SDSS-matched spectroscopic redshifts, as shown in Figure 2. For the remainder of this work, we will refer to subclasses  $b$ ,  $c$ , and  $d$  into class II.

### 2.3. Class Properties

Classes I and II represent potential AGN classes identified through photometry alone informed by unsupervised ML. If validated, the classes would contain extensive numbers of AGNs over the HSC footprint. Class I is characterized by a median  $i$ -band magnitude of  $\sim 21$  and an on-sky density of  $\sim 148 \text{ sources deg}^{-2}$ . On the other hand, class II is, on average, fainter than class I with a median  $i$ -band magnitude of  $\sim 22$  but has nearly double the on-sky density with  $\sim 235 \text{ sources deg}^{-2}$ . The SDSS Quasar Catalog, for comparison, has an average on-sky density of  $\sim 80 \text{ AGN deg}^{-2}$  over the SDSS fields with a maximum density of  $125 \text{ AGN deg}^{-2}$  ([Pâris et al. 2018](#); [Lyke et al. 2020](#)). In addition, the 90% Reliability (R90) WISE AGN Catalog, one of the largest and most accurate AGN catalogs in the literature, averages  $151 \text{ AGN deg}^{-2}$  ([Assef et al. 2018](#)). If their AGN-selection accuracy is high, classes I and II would represent a substantial increase in the number density of AGNs selected via WISE combined with deep optical imaging.

To demonstrate the difference between classes I and II and typical AGN selected samples, we present the optical magnitude-morphology, optical color-color, mid-IR color-color distributions of classes I and II compared to those with existing SDSS spectra along with the corresponding SDSS redshift distributions in Figure 3. Existing spectroscopic surveys from SDSS and mid-IR selection criteria do not span the full parameter space of our candidate AGN classes. To highlight the differences in the spectroscopic properties of our ML selected classes and matched SDSS spectroscopy, we compare spectral stacks of these populations in Section 4.2.2.

In Figure 3 we present optical magnitude-morphology, optical color-color, and mid-IR color-color distributions for classes I and II along with the matches for their SDSS matched counterparts demonstrating that SDSS is limited to the brightest, most centrally concentrated, and bluest galaxies for both classes. For example,  $\sim 30\%$  of class I objects are sufficiently red to have  $g - i > 0.75$ , whereas this is only true for  $\sim 15\%$  of the SDSS-matched subsample. For class I this is due to the fact that nearly all SDSS-matched spectroscopy in this class is targeted as part of quasar (90%) or stellar mapping (10%) sub-

**Table 1.** Follow-Up Hectospec Fields

Field #	R.A.	Dec	Obs. Date(s)	Proposal ID	# Spectra	# Class I	# Class II
0	14:43:25	−00:38:25	2020-06-20	2020b-UAO-S139	193	34 (62.96%)	20 (37.04%)
1	14:38:15	−00:51:41	2020-06-21+22	2020b-UAO-S139	212	36 (63.16%)	21 (36.84%)
2	14:39:23	+00:17:39	2020-06-22+23	2020b-UAO-S139	198	38 (70.37%)	15 (27.78%)
3	22:47:50	+00:49:35	2020-10-13	2020c-UAO-S125	127	6 (46.15%)	7 (53.85%)

surveys within SDSS, BOSS, and eBOSS which target compact blue emission.

For class II the limitations of SDSS-matched spectroscopy is especially noticeable, where  $\sim 80\%$  of matches are limited to  $i < 22$ , while half of class II is fainter than this limit. While  $\sim 50\%$  of class II galaxies have  $g - i > 1.5$ , only  $\sim 15\%$  of SDSS-matched objects are sufficiently red. However, despite having nearly twice the on-sky density as class I, class II has only 11% the number of SDSS matches, driven, in part, by the fact that class I is brighter than class II on average by  $\sim 1.43$  mag in the  $i$  band ( $\sim 3.73$  times brighter). While matched SDSS spectroscopy for class II are also targeted by quasar subsurveys (20%), they are primarily observed in eBOSS emission-line galaxy subsurveys (80%) that target galaxies at  $z \simeq 0.9$  with strong [O II] emission (Comparat et al. 2016). Therefore class II has over 90% of matched SDSS spectra from eBOSS, with 10% from BOSS, and less than a percent from other SDSS surveys.

Figure 3 emphasizes that existing SDSS spectroscopy has focused on the bluest subset of targets, i.e. SDSS spectroscopy is therefore not able to inform the AGN-selection accuracy of our candidate classes as it primarily targets bright, blue, and compact galaxies when compared to our selection. Additional spectroscopic follow-up is therefore necessary to fully span the parameter space of our ML-selected AGN classes.

### 3. MMT HECTOSPEC FOLLOW UP

To validate the efficacy of the ML pipeline at selecting AGNs, we were awarded two nights in the spring of 2020 to observe five fields<sup>2</sup> with Hectospec, a multi-object fiber-fed spectrograph installed on the 6.5m MMT Observatory (Fabricant et al. 2005). Hectospec is the ideal instrument to follow-up large numbers of photometrically-selected AGN candidates owing to its multiplexing capability to obtain moderate-resolution spectra of up to 300 objects simultaneously over a  $1^\circ$

diameter field using  $1.5''$ -diameter fibers. We utilized the 270 lines  $\text{mm}^{-1}$  grating, affording us a  $\mathcal{R} \sim 1000$  over a 3650–9200Å wavelength range. Targets are selected from across all UMAP identified classes, with a higher priority designated for MMT follow-up targets from classes I and II, which are the focus of this work. We select our MMT Hectospec targets in Section 3.1, describe our data reduction in Section 3.2, and determine spectroscopic redshifts in Section 3.3.

#### 3.1. Target Selection

Our target selection was motivated to find fields with relatively high densities of AGN candidates, e.g. objects from classes I and II. In addition, fields were selected to ensure an appropriate density of potential guide stars and standard stars. Guide stars are selected from Gaia Data Release 2 (DR2; Collaboration et al. 2018) to have G-band magnitudes in the range 13 – 15. Standard stars were selected from SDSS DR16 following the “Target Selection of F Star Photometric Standards”<sup>3</sup>. Fiber placements are generated for candidate fields using the `xfitfibs`<sup>4</sup> utility which allows for the ranking of the target galaxies by their class and assigns fibers for the determination of the sky-background.

We observed our four selected fields in queue mode over two semesters in 2020, which are listed in Table 1. Each field was observed for a total of three hours comprised of six 30-minute integrations. Our targets were drawn from classes I and II while any remaining fibers were assigned to galaxies from other ML-identified UMAP classes. We observe a total of 178 galaxies across our ML-selected classes, with 114 ( $\sim 64\%$ ) from class I and 63 ( $\sim 36\%$ ) from class II: 10, 38, and 15 from sub-classes *b*, *c*, and *d* respectively. Due to the limit of MMT Hectospec, our spectroscopy is restricted to sources with  $i \lesssim 22.5$ . Consequently our analysis of class II will be limited the brightest 75% of objects.

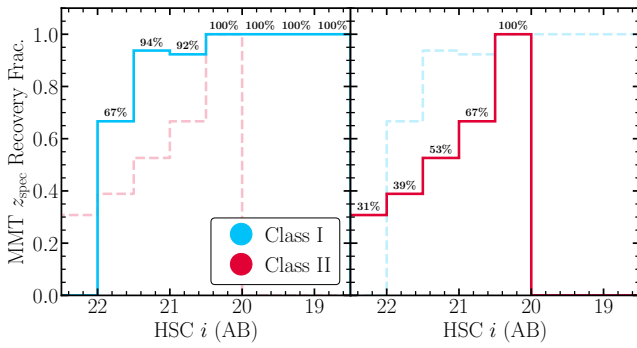
#### 3.2. Data Reduction

<sup>2</sup> Due to the outbreak of COVID-19 in early 2020, our observing time was split over two semesters. In addition, due to instrument repairs, we were only able to observe four out of our five fields.

<sup>3</sup> [https://www.sdss.org/dr16/algorithms/boss\\_std\\_tst/](https://www.sdss.org/dr16/algorithms/boss_std_tst/)

<sup>4</sup> <https://lweb.cfa.harvard.edu/mmti/hectospec/xfitfibs/>

The Hectospec spectra were reduced using the **HSRED**<sup>5</sup> reduction pipeline. The pipeline bias corrects, dark subtracts, and flat fields the data using the provided calibration images. Each spectrum is extracted, associated with the input catalog objects, and then wavelength calibrated using arc lamp spectra. Sky-background and flux corrections are then applied using the data from the fibers reserved for sky background and F-Star Photometric Standards. Objects which are observed over multiple nights are coadded after running the **HSRED** pipeline on each night individually. Finally, the spectra are corrected for Milky Way extinction based on the [Schlafly & Finkbeiner \(2011\)](#) maps assuming a [Fitzpatrick \(1999\)](#) dust law ( $R(V) = 3.1$ )



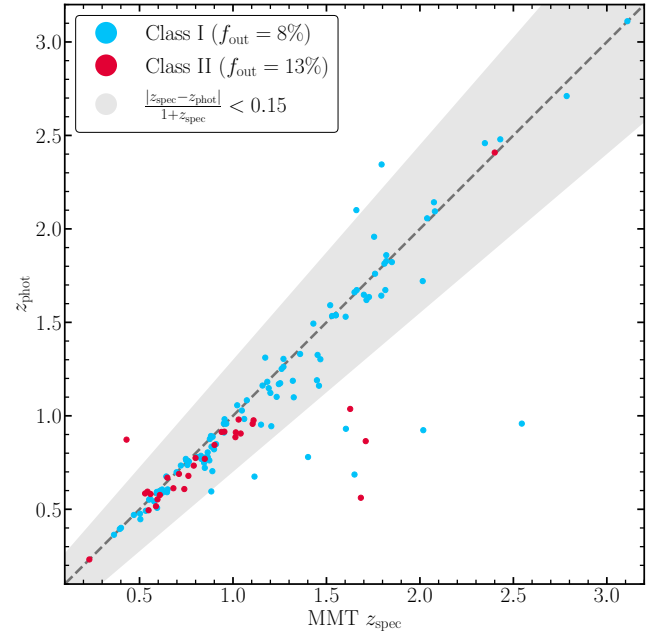
**Figure 4.** MMT redshift recovery fraction against optical magnitude distributions for classes I (blue; left) and II (red; right). We note that our follow-up spectroscopy are limited to  $i < 22.5$ , which particularly effects class II.

In addition, 17 of our objects have existing spectroscopy from SDSS, though none of which are in our candidate AGN classes. As some of our fields did not have sufficient F-stars to construct a flux calibration, we generate and apply an additional calibration to our spectra based on the reobserved targets for the relevant fields. This is primarily used to retrieve physically motivated continuum fits as described in Section 4.1. We note this has a negligible effect on our subsequent analysis as we are concerned with the detection of broad emission lines or measurement of flux ratios of nearby lines, which are minimally affected by color corrections.

### 3.3. Redshift Determination

MMT spectroscopy enables us to accurately determine the redshift of our sample by searching for prominent emission lines, which we describe at the start of Section 4, or absorption features such as the Ca H&K lines or the 4000Å break. To determine the redshift of our sam-

ple, all reduced spectra are visually inspected to generate an initial estimate for where two or more spectral features can be clearly identified. We make use of the **redshifting**<sup>6</sup> code which performs a grid-search using eigenspectrum templates from [Bolton et al. \(2012\)](#) to determine redshift from optical spectroscopy ([Johnson et al. 2018; Helton et al. 2021; Johnson et al. 2022](#)). The spectra are fit over a range of redshifts using the quasar ( $0 < z < 3$ ), galaxy ( $0 < z < 1.5$ ), and star ( $-0.005 < z < 0.005$ ) eigenspectra and we compare the retrieved  $\chi^2$  and  $\chi^2_\nu$  distributions to refine and improve our visual inspection redshifts.



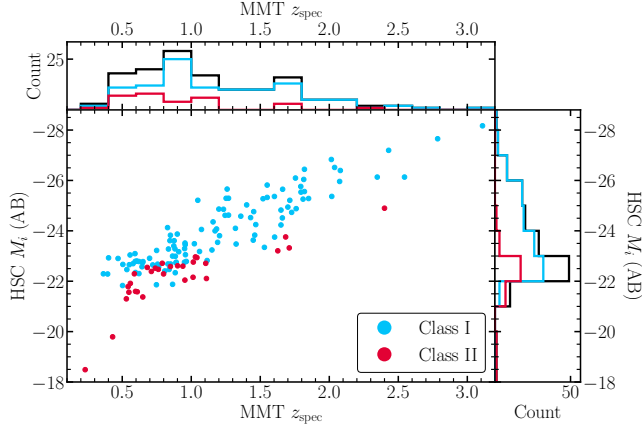
**Figure 5.** Photometric redshift against spectroscopic redshift for classes I (blue) and II (red).

We retrieve redshifts for 139 ( $\sim 78\%$ ) of our class I and II follow-up targets. In Figure 4 we show the redshift recovery fraction as a function of optical magnitude. As expected, we are less successful at identifying redshifts for fainter sources and are most successful at  $i < 21.5$ , with redshift recovery fractions of 97% and 60% for classes I and II respectively. In addition, we are less able to identify redshifts for class II, especially for the faintest objects, which we attribute to a relative lack of strong emission lines as compared to class I, as discussed in Section 4.2. Combined with our spectroscopic limit of  $i < 22.5$ , our analysis of class II is limited to the brightest 75% of objects in the class. In Figure 5 we compare the photometric redshift estimates from

<sup>5</sup> <http://www.mmt.org/hsred-reduction-pipeline/>

<sup>6</sup> <https://github.com/sdjohnson-astro/redshifting>

Goulding et al. (in prep.) designed for AGN and quasar candidates with our recovered spectroscopic redshifts for both classes I and II. Briefly, the combined HSC and WISE photometry for AGN with archival spectroscopy are used to train a set of augmented Random Forest algorithms to construct and assess the accuracy and precision of photometric redshifts for both the unobscured and obscured AGN populations. They show that these ML-based photo- $z$ s perform equally well on Type I and Type II AGN alike out to  $z \sim 3$  with an average precision of  $\delta z / (1 + z) \simeq 0.02$  and  $0.03$ , respectively, and that they significantly out-perform the standard HSC Mizuki photometric redshifts that are provided by Subaru-HSC (Nishizawa et al. 2020). Using our sample of AGN with new redshifts from MMT/Hectospec, we find that both AGN classes have good agreement with moderate outlier fractions  $\sim 10\%$ , reinforcing the efficacy of the photometric redshifts for AGN presented in Goulding et al. (in prep.).



**Figure 6.** Absolute observed-frame  $i$ -band magnitude versus spectroscopic redshift for class I (blue) and class II (red) objects with recovered MMT redshifts. The top and right panels present the redshift and magnitude histograms respectively for classes I and II along with an overall histograms in black.

In Figure 6 we plot the absolute magnitude versus spectroscopic redshift for our targets. Overall class I sources are more luminous and at higher redshifts than those in class II. Class I is distributed across redshifts from  $z = 0.5 - 3$ . The redshift distribution for class II is clustered around  $z = 0.5 - 1$ , with potentially a tail out to  $z \sim 2$ , though we are limited by the number of targets with redshift in this class. For both classes the distributions are similar to those shown in Figure 3 for matched SDSS samples.

#### 4. SPECTROSCOPIC ANALYSIS

In this Section we use our MMT follow-up to determine (a) the AGN fraction per class and (b) the properties of these AGNs in our ML-selected targets. Out to  $z \lesssim 1$  optical spectroscopy enables the characterization of both the [O III] $\lambda\lambda 5008, 4960\text{\AA}$  doublet, which is sensitive to the ionization state of nebular gas and to potential AGN-driven outflows, as well as the detection of broad Balmer or Mg II $\lambda 2799\text{\AA}$  line emission. At intermediate redshifts ( $0.5 \lesssim z \lesssim 1.5$ ) [Ne III] $\lambda 3870$  to [O II] $\lambda\lambda 3730, 3727\text{\AA}$  nebular emission-line ratios and the detection of the [Ne V] $\lambda\lambda 3426.85, 3346.79\text{\AA}$  coronal emission doublet are diagnostics for AGN activity. Finally, at the highest redshifts probed by our analysis ( $1.5 \lesssim z \lesssim 3$ ), the strengths and widths of rest-ultraviolet (rest-UV) emission lines such as Ly $\alpha$ , Si V $\lambda 1398\text{\AA}$ , C IV $\lambda 1549\text{\AA}$ , He II $\lambda 1640\text{\AA}$ , and C III] $\lambda 1909\text{\AA}$  trace SMBH accretion.

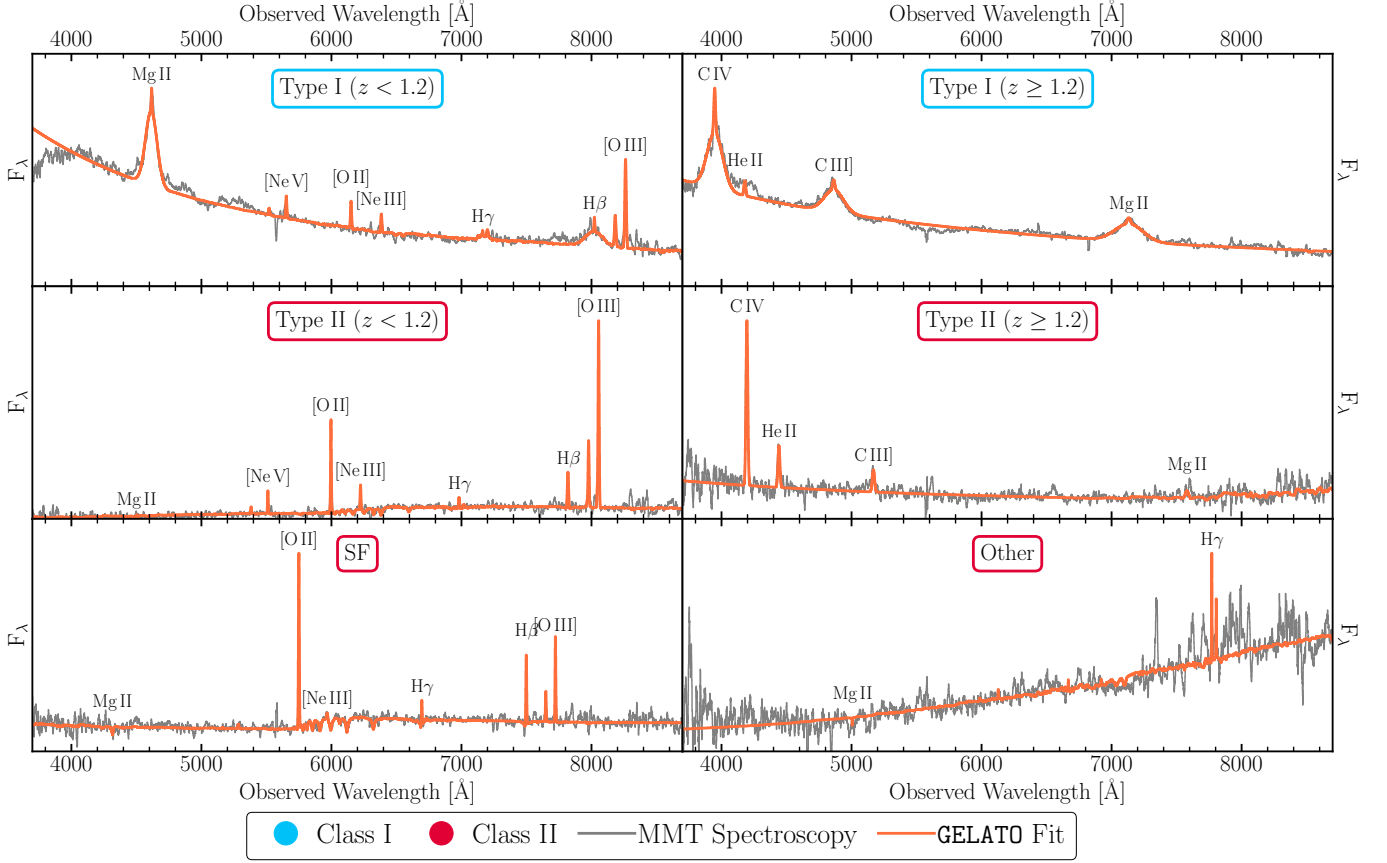
To determine the AGN properties and prevalence in our ML-selected candidate AGN classes, in this section we measure the spectroscopic properties of the 139 galaxies with identified redshifts. Emission-line fitting is performed in Section 4.1 to measure the ionization properties and kinematics in our galaxies to determine the presence of AGN activity and compute the the AGN fraction in our ML selected classes. We perform spectral stacking to determine average properties for classes I and II (Section 4.2) and present Balmer decrements to measure galaxy-scale attenuation in Section 4.3.

##### 4.1. Emission-Line Fitting

To fit our spectra we make use of the Galaxy/AGN Emission-Line Analysis TOol (GELATO)(Hviding 2022). GELATO is designed to retrieve emission-line parameters in optical spectroscopy by fitting a series of Gaussians for emission lines, Simple Stellar Populations (SSPs) for the underlying continuum, and a power-law continuum<sup>7</sup>. Critically GELATO statistically tests for the presence of broad-lines and is therefore ideal for this work. We begin by splitting our spectra into two subsamples, a low-redshift ( $z < 1.2$ ) and a high-redshift ( $z \geq 1.2$ ) subsample. As GELATO uses the Extended MILES stellar library (E-MILES; Vazdekis et al. 2016) with SSPs that only extend to a wavelength of  $1680\text{\AA}$  we only fit our high-redshift subsample with a flexible power-law continuum and no SSP templates.

<sup>7</sup> We note that GELATO does not include Iron emission templates. While these would be necessary for in-depth study of broad-line quasars in our sample, they are not necessary for the robust detection of broad lines or the characterization of the narrow-line emission in obscured and/or low-luminosity AGN.





**Figure 7.** Example MMT spectra from our spectroscopic samples (black) along with the corresponding **GELATO** fits (orange) for select class I (blue) and class II (red) targets. The objects are chosen across the redshift range of our sample and to be representative of the spectral classifications used in this work. Spectra are smoothed with a boxcar function of width of  $10\text{\AA}$  for plotting purposes only.

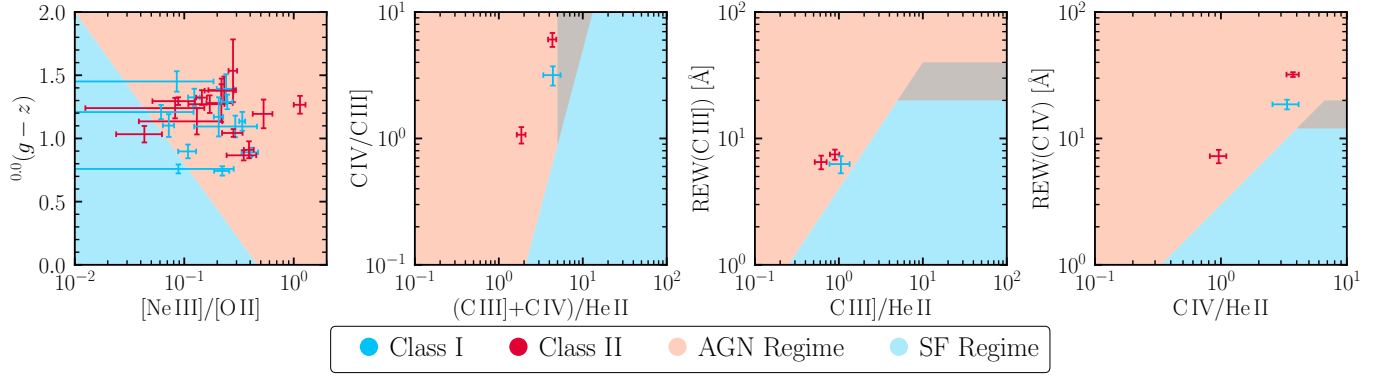
We use **GELATO** to fit all emission lines listed at the start of Section 4. To improve the accuracy of the fits and due to the signal-to-noise ratios of our spectra, we require that the narrow components of all lines share the same width and redshift. In addition, we attempt to fit the Mg II, C IV, He II, C III], and Balmer lines with a broad component. We consider a spectrum to have a broad line if the fit with a broad line statistically improves the fit at the 95% level as determined by an F-test. In this work we define a broad line to have a minimum dispersion of  $500\text{ km s}^{-1}$ , i.e. a full width at half maximum (FWHM)  $> 1200\text{ km s}^{-1}$  based on the delineation outlined in Hao et al. (2005).

Example MMT spectra from this study are presented in Figure 7 along with their corresponding **GELATO** fits. The spectra are selected to be representative examples of the spectral classifications we use throughout this work in both redshift regimes. **GELATO** proves to be an effective tool for recovering line fluxes, measuring equivalent widths, and detecting the presence of broad emission,

enabling the spectral classification of our MMT follow-up spectroscopy.

We classify a galaxy as a Type I AGN if the spectrum shows evidence for at least one broad line and find a total of 99 Type I AGNs in our sample. We target 114 class I galaxies and secure 109 redshifts. Of these, 93 are identified as Type I AGNs, 43 of which are in the low-redshift subsample and 48 in the high-redshift subsample. We target 63 class II galaxies and secure 30 redshifts. Of these, six are identified as Type I AGNs, four of which are in the low-redshift subsample and two in the high-redshift subsample. Broad emission lines for class I Type I AGNs display larger velocity widths with a median velocity FWHM of  $\sim 7800\text{ km s}^{-1}$  and a standard deviation of  $\sim 2900\text{ km s}^{-1}$  compared to a median FWHM of  $\sim 3400\text{ km s}^{-1}$  and a standard deviation of  $\sim 900\text{ km s}^{-1}$  for class II.

For the remaining 40 galaxies with a spectroscopic redshift and no detected broad line we attempt to classify the spectra based on the narrow emission lines using emission-line diagrams to distinguish ionization driven



**Figure 8.** Emission-line diagrams for MMT follow-up targets with identified redshifts that show evidence for emission lines but do not exhibit broad emission or [Ne V] emission. In the leftmost panel we present the TBT diagram for our low-redshift ( $z < 1.2$ ) AGN candidates. In the three rightmost panels we present UV emission-line diagnostic diagrams from (Nakajima et al. 2018) for our high-redshift ( $z \geq 1.2$ ) AGN candidates.

AGN activity from star formation (SF). A common tool is the Baldwin et al. (BPT; 1981) diagnostic, which uses ratios from common, strong optical emission lines to determine the source of the ionizing radiation. However, only one galaxy in our sample has spectral coverage of the necessary lines to be placed on the BPT diagram. We therefore make use of the Trouille et al. (TBT; 2011) diagram which compares the [Ne III] to [O II] ratio to the rest-frame  $g - z$  color and is more suited to the redshift range of our spectra ( $z_{\min} \simeq 0.23$ ). We find that the results are consistent for the object which can be placed on both diagrams, finding that it lies on the border between a star-forming galaxy and an AGN.

To perform the K-correction to generate the rest-frame colors necessary for the TBT diagnostic, we fit a combination of the elliptical, spiral, and irregular empirical galaxy templates from Assef et al. (2010) combined with the normal, as opposed to warm- or hot-dust deficient, AGN template from Lyu & Rieke (2017) and AGN extinction curve from Lyu (in prep.)<sup>8</sup> to the HSC and WISE photometry at the spectroscopic redshift of the galaxy using the procedure described in Section 4.1 of Hviding et al. (2022). The rest-frame colors and associated errors are computed by convolving the SDSS  $g$  and  $z$  filters (Gunn et al. 1998) with the full set of models from the MCMC chain shifted into the rest frame. The leftmost panel of Figure 8 presents the TBT diagram for the objects from our low-redshift subsample that do not have evidence for a broad line or [Ne V], 14 from class I and 16 from class II.

For the high-redshift subsample without detected broad lines, we make use of the diagnostics presented in Nakajima et al. (N18; 2018) which compare the ra-

tios and rest-equivalent widths (REWs) of the C IV, He II, and C III] emission lines. We present the three high-redshift, narrow-line galaxies on the N18 diagrams in the rightmost panels of Figure 8. For galaxies placed on either the TBT or the N18 diagrams we sample the posterior to compute the probability that each galaxy is in the AGN regime. Conversely, the complementary probability can be calculated that the emission is not consistent with AGN activity and instead due to SF.

In addition we investigate which galaxies show evidence for the coronal emission line [Ne V] that is a reliable indicator of AGN activity due to its high ionization potential of  $\sim 100$  eV (Mignoli et al. 2013; Feltre et al. 2016; Negus et al. 2023, and references therein). We find nineteen narrow-line galaxies, seven from class I and twelve from class II, with robust [Ne V] detections (flux SNR  $> 3$ ), five of which have detections at the  $10\sigma$  level, one from class I and four from class II. We note that 17 of the [Ne V]-detected galaxies are already classified as AGN using the emission line diagrams at the  $3\sigma$  confidence level, with the last two classified at the  $\sim 80\%$  and  $\sim 90\%$  level.

To place constraints on the non-AGN contamination fraction in classes I and II, we assign each object an AGN probability. Type I and Type II [Ne V]-detected AGNs are assigned a probability of 100%. For the remaining narrow-line galaxies we compute the AGN probability by sampling the posterior of each source one million times and calculating the fraction that lie in the AGN regime. As there are three N18 diagrams, we take the mean AGN probability for each of the three narrow-line high-redshift galaxies but find that the result is consistent regardless of which diagnostic is used.

Two class II galaxies in our sample show little-to-no evidence for emission lines used in AGN diagnostics. These objects cannot be robustly classified as AGNs nor

<sup>8</sup> See Section 2.2 in Lyu & Rieke (2022) for details.

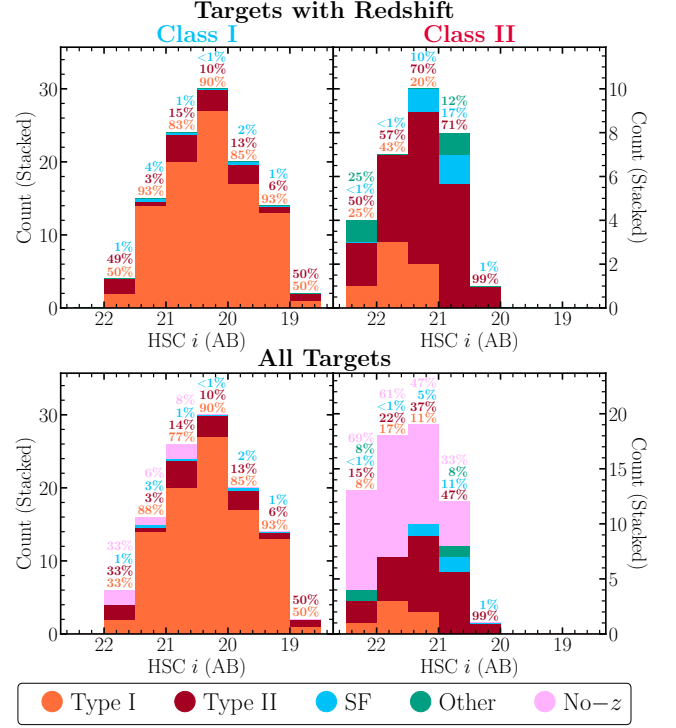
**Table 2.** AGN Fraction by Class

Class	Type I	Type II	SF	Other	No-z
Targets with Redshift					
I	86.24%	12.31%	1.45%	0.0%	—
II	20.0%	65.38%	7.95%	6.67%	—
All Targets					
I	82.46%	11.77%	1.39%	0.0%	4.39%
II	9.52%	31.13%	3.79%	3.17%	52.38%

as SF galaxies. To place lower bounds on the AGN fractions for our classes, we treat these objects as having a 0% probability of being an AGN and label them as Other in the forthcoming analysis.

By taking into account the probability that each source is an AGN, we compute the average breakdown of classes I and II by spectroscopic classification and present the results in Table 2. Figure 9 presents the spectroscopic classification as a function of optical magnitude and class with and without taking into account those objects without identified redshifts. Class I is dominated by Type I AGNs, with a <3% contamination rate, while the majority of class II is mostly comprised of Type II AGNs with a moderate contamination rate of ~15% from non-AGN sources. We note that our redshift recovery fraction is relatively low at the faintest magnitudes for class II, with over half of the class without an identified redshift. We therefore present the lower-limit AGN fractions as well, where only half of class II are identified as AGN with a confident redshift. While the spectroscopic classification of class I has little-to-no evolution with optical magnitude, class II has a higher fraction of Type I AGNs for fainter targets. This is likely driven by a bias where broad-line AGNs are easier to identify spectroscopically than Type II AGNs at the same brightness due to their strong emission, though deeper spectroscopy in class II would be needed to confirm this. Overall, our ML-selected classifications are effective at recovering AGN samples with a high level of accuracy.

MMT spectroscopy has confirmed that our ML-selected AGN classes are comprised of AGNs. Class I selects  $\sim 20 \text{ deg}^{-2}$  more Type I quasars as compared to SDSS quasar selected samples. These additional objects are primarily composed of galaxies with redder optical colors than the SDSS selection. Conversely class II predominantly selects Type II quasars with an on-sky density of  $\sim 200 \text{ AGNs deg}^{-2}$ . Even when taking

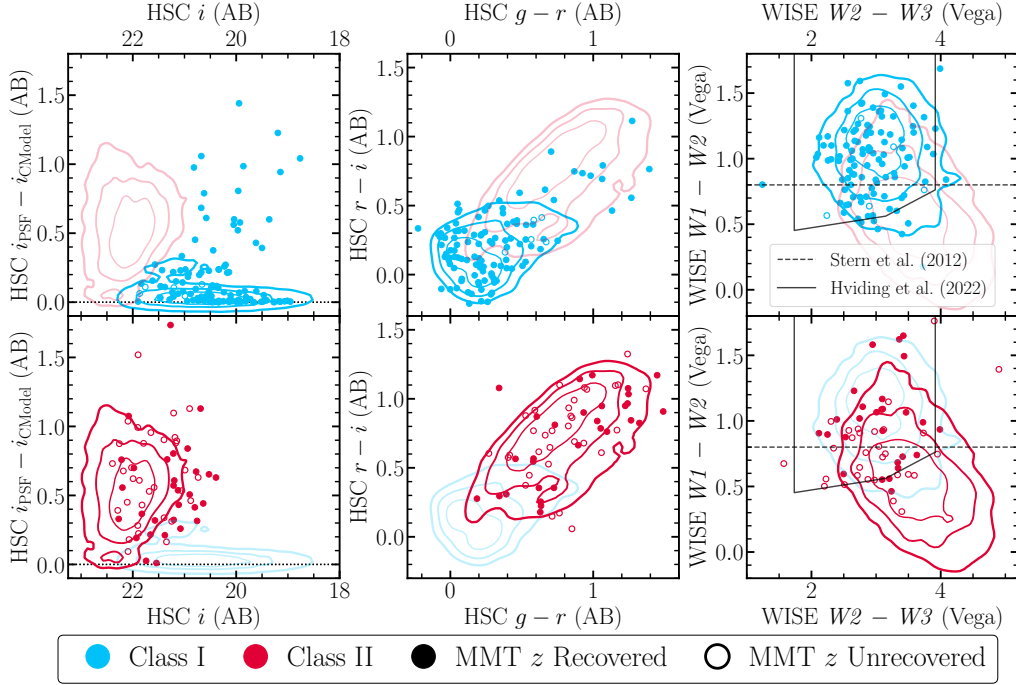


**Figure 9.** Spectroscopic classification as a function of optical magnitude for class I (left) and II (right). The top panels present the fractions for targets with a secure redshift, while the bottom panels show the percentages accounting for objects without an identified redshift. Class I is dominated by Type I AGNs with minimal contamination from non-AGN. The majority of class II are Type II AGNs with a higher fraction of non-AGN contaminants. Overall fractions for each class are presented in Table 2.

the lower limit when considering objects without a recovered redshift, class II represent an on-sky density of  $\sim 100 \text{ deg}^{-2}$  predominantly Type II AGNs. The few matched SDSS spectra that exist for class II primarily come from emission-line galaxy subsurveys and are limited to the brightest and bluest subset of the class.

#### 4.2. Spectral Stacking

To determine if the AGN properties are consistent across our ML-selected classes, we investigate our spectroscopy in bins of magnitude, redshift, and optical color and compare to the spectral stacks of matched SDSS galaxies. In Figure 10 we present the optical magnitude-morphology, optical color-color and mid-IR color-color distributions of our selected targets with respect to our parent HSC sample. Our follow-up targets are representative across the parent classes' morphological and optical color-color spaces. We therefore stack the spectra of AGNs in our follow-up to accurately measure the properties of classes I and II. We consider an object an AGN if it has an AGN probability greater than 50% as



**Figure 10.** Optical magnitude-morphology (left), optical color-color (center), and mid-IR color-color (right) for classes I (blue; top) and II (red; bottom). We plot contours encompassing 25%, 50%, and 75% of the classes and plot our MMT follow-up targets with recovered and unrecovered redshifts as filled and unfilled circles respectively. Mid-IR color-color contours are restricted to objects with  $\text{SNR} > 3$  in  $W1$ ,  $W2$ , and  $W3$ . We plot the Hviding et al. (2022, solid) and Stern et al. (2012, dashed) mid-IR selections to highlight regions where typical mid-IR selected AGNs inhabit. We note that our follow-up spectroscopy span the majority of each class’s parameter space, but are limited to  $i < 22.5$ , which particularly effects class II.

determined in Section 4.1. By combining spectra of similar objects, we can enhance the SNR of specific spectral features, such as emission lines, and study their properties as a function ML-selected class.

Spectra are stacked in the low-redshift and high-redshift subsamples to determine what impact, if any, redshift has on the optical colors of the classes and to ensure all spectra in the subsample share a region of continuum free from spectral features. We normalize all spectra to a region of the continuum: 3500-3600Å for the low-redshift subsample and 2000-2100Å for the high-redshift subsample. All spectra are interpolated onto a wavelength grid of 1Å intervals using the **SpectRes** (Carnall 2017, 2021) flux-preserving spectral resampling code. The weighted arithmetic mean is taken at each wavelength to generate the stacked spectrum.

#### 4.2.1. Stacks by Optical Magnitude

In Figure 11 we present the stacks of class I AGNs in bins of optical magnitude along with the distributions of the retrieved spectroscopic redshift. We observe that class I is dominated by Type I AGNs with broad optical and UV emission lines and strong blue optical/UV continua. This trend continues across the entire optical magnitude range of the sample. In addition, we find no

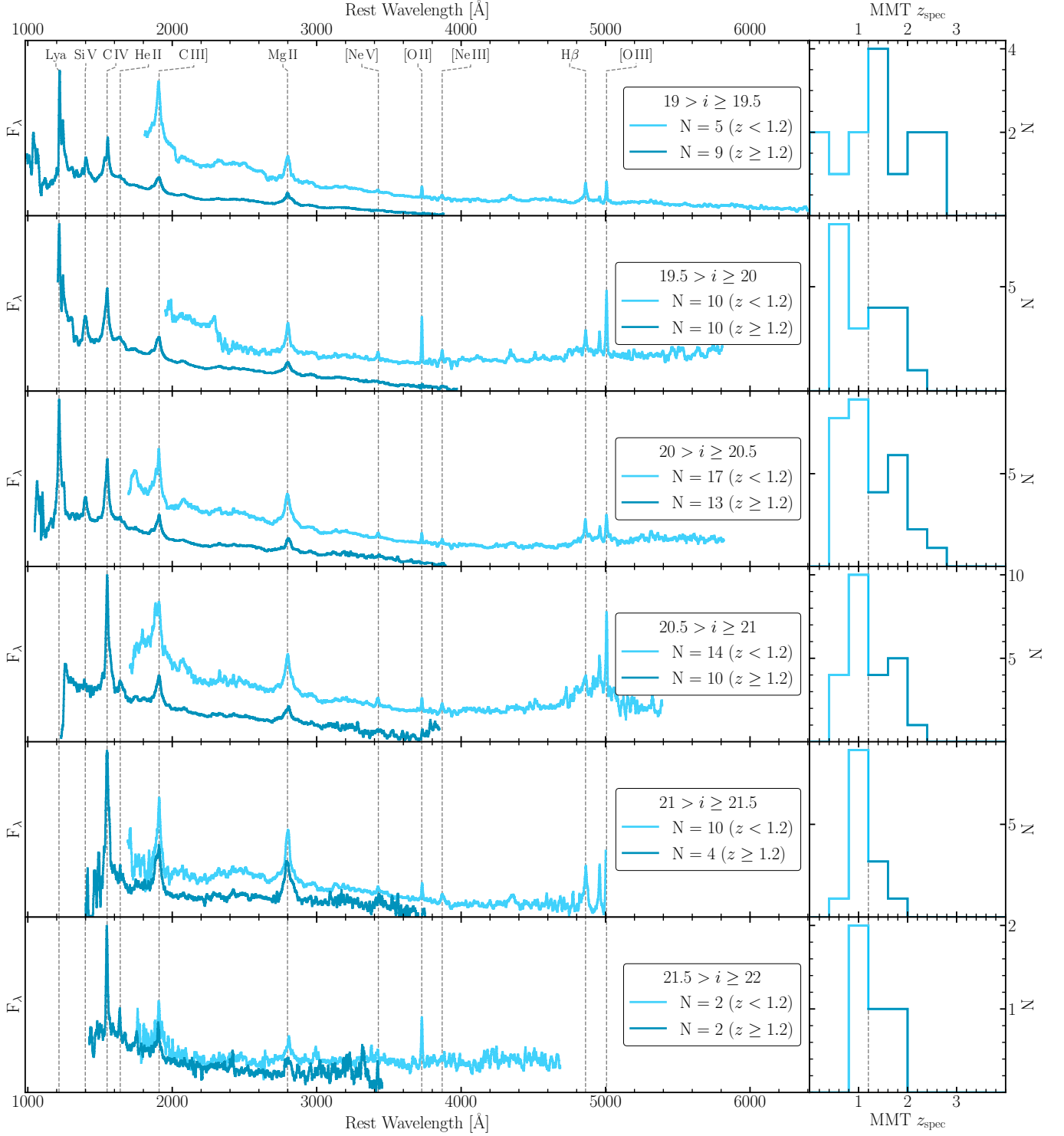
difference of the spectroscopic properties between the high-redshift and low-redshift subsamples.

We repeat our stacking procedure for those sources in class II. Figure 12 presents the stacks of class II AGNs in bins of optical magnitude along with the distributions of the retrieved spectroscopic redshift. We observe that class II is predominantly comprised of Type II AGNs with strong narrow emission lines and clear stellar continuum as evidenced by stellar absorption features. This trend continues across the entire optical magnitude range of the sample but with a variety in the emission-line strength of the stacks. In addition, while the properties of the low- and high-redshift samples of class II appear consistent, it is difficult to confirm with relatively few high-redshift class II objects.

#### 4.2.2. Comparison to SDSS Spectral Stacks

While SDSS-matched class I galaxies are nearly all ( $> 99\%$ ) identified as QSOs by the automated SDSS pipeline, this is only true for 22% of SDSS-matched class II galaxies. This does not mean that the remaining 79% do not host AGN, but rather that there are no detectable broad-lines in the spectrum or that they suffer from greater contamination from the host. We thus create spectral stacks for class II galaxies with existing SDSS

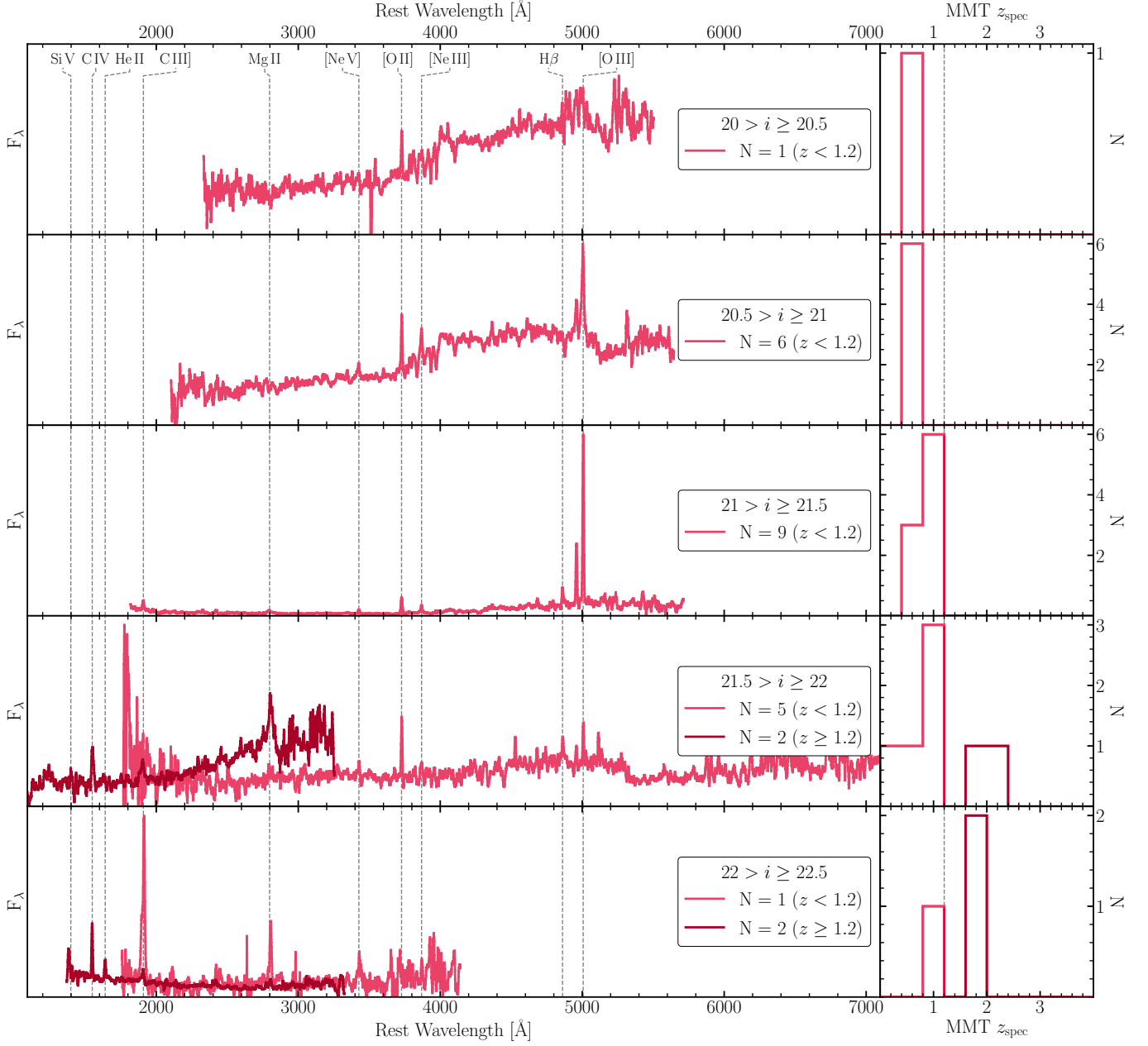




**Figure 11.** Low-redshift ( $z < 1.2$ ; light blue) and high-redshift ( $z \geq 1.2$ ; dark blue) stacked MMT spectra (left) and redshift distributions (right) for class I by  $i$ -band magnitude. The class is primarily comprised of Type I broad-line AGNs. Stacked spectra are smoothed with a boxcar function of width of  $10\text{\AA}$  for plotting purposes only.

spectroscopy. Figure 13 presents these stacks compared to stacks of our class II MMT spectra. While it is immediately apparent that the MMT follow-up presented in this work samples a redder distribution of galaxies than

the SDSS-matched sample, it is also clear that the SDSS subsample hosts a high fraction of AGN at both high redshift, as evidenced by broad emission features, and at low redshift, as evidenced by the presence of [Ne V]



**Figure 12.** Low-redshift ( $z < 1.2$ ; light red) and high-redshift ( $z \geq 1.2$ ; dark red) stacked MMT spectra (left) and redshift distributions (right) for class II by  $i$ -band magnitude. The class is primarily comprised of Type II narrow-line AGNs. Stacked spectra are smoothed with a boxcar function of width of  $10\text{\AA}$  for plotting purposes only.

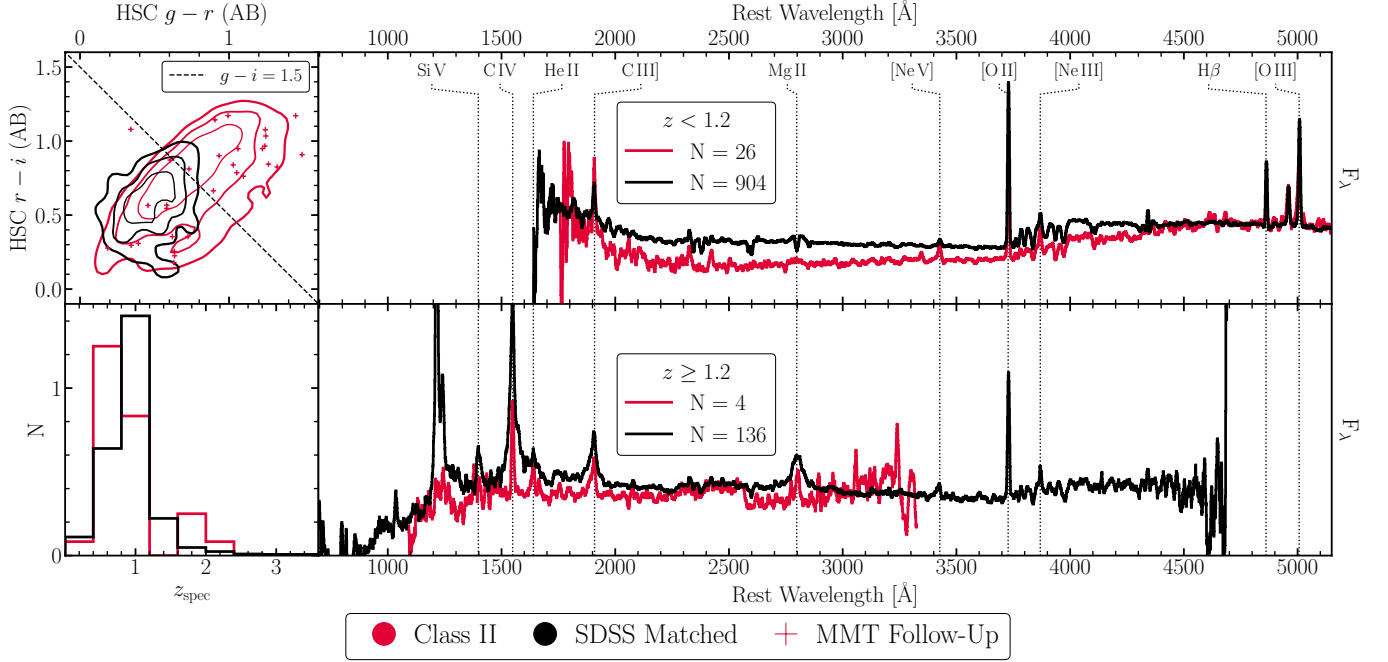
coronal emission and high-ionization emission line ratios from  $[\text{O III}]/\text{H}\beta$  and  $[\text{Ne III}]/[\text{O II}]$ . Therefore, the SDSS-matched class II galaxies are indeed comprised of Type II AGN, similar to the bulk of class II, but are bluer than the majority of the parent class.

#### 4.2.3. Stacks by Optical Color

A clear advantage of our ML-identified classes is the identification of redder AGN candidates than typical spectroscopic surveys as highlighted in Section 2.1 and 13. In particular, it is of interest to determine the AGN

properties of the candidates that are redder than the matched SDSS spectroscopy, to determine the differences to their bluer counterparts. While our stacking analysis has confirmed that classes I and II are indeed comprised of AGNs across all magnitudes, we pursue stacking in bins of optical color to determine the difference in AGN properties, if any, between the bluest and reddest AGNs in our sample.

To investigate the properties of our AGNs in class I as a function of optical color, we split the class into a



**Figure 13.** Optical color-color (top left) and spectroscopic redshift (bottom left) distributions of class II (red) and SDSS matched (black) subsamples. In addition, we present stacked spectra for low-redshift ( $z < 1.2$ ; top right) and high-redshift ( $z \geq 1.2$ ; bottom right) subsamples. Stacked spectra are smoothed with a boxcar function of width of  $10\text{\AA}$  for plotting purposes only.

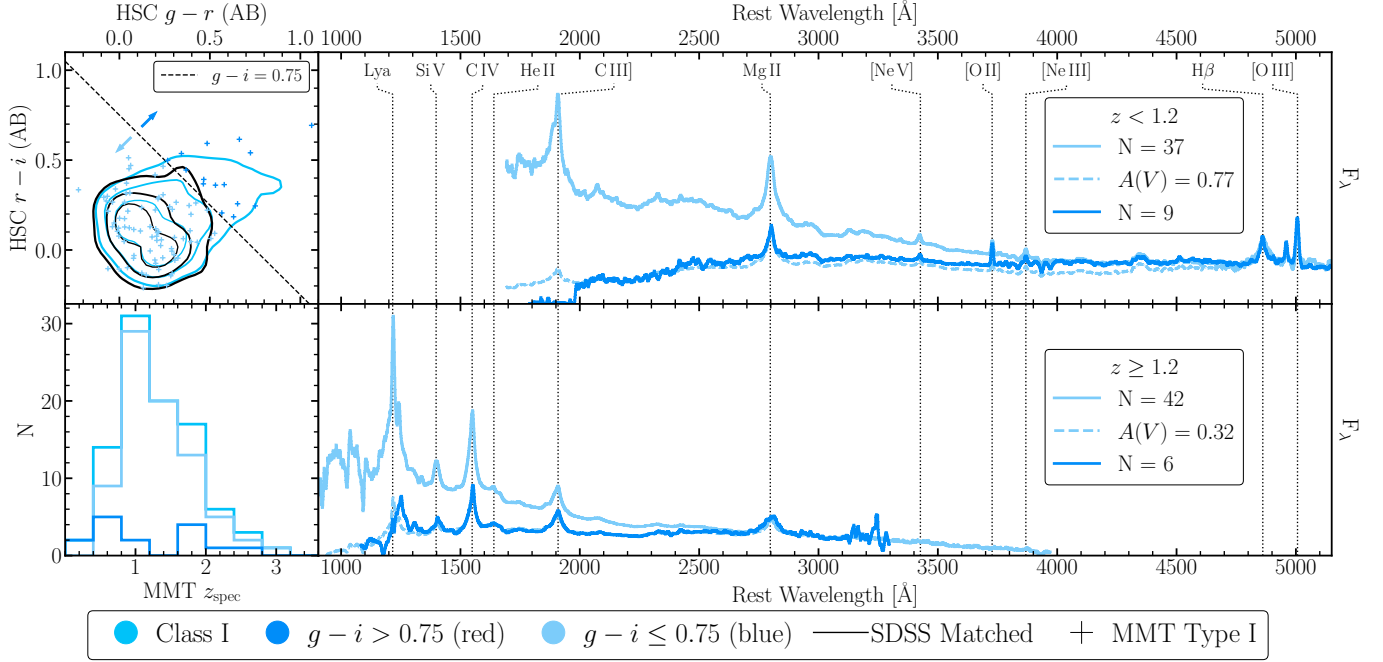
“red” subsample ( $g-i > 0.75$ ) and a “blue” subsample ( $g-i \leq 0.75$ ) based on where we observe class I deviates from SDSS-matched spectroscopy in Figure 3. In addition, we restrict our analysis to Type I AGNs, the dominant constituent (85%) of class I. Figure 14 presents the stacks as a function of redshift and optical magnitude. **Regardless of optical color or redshift, the emission-line shape and strength are consistent, suggesting Type I AGNs in class I are drawn from a similar underlying population of galaxies.**

To determine if the differences between the “red” and “blue” subsamples can be explained solely through obscuration, we fit the red stacked spectrum by taking the the blue stacked spectrum and applying an extinction curve. We make use of the Small Magellanic Cloud (SMC) Bar extinction curve (Gordon et al. 2003,  $R(V) = 2.74$ ) as it has been found to best describe the extinction in dust-reddened quasars and AGNs (Richards et al. 2003; Hopkins et al. 2004). By minimizing the  $\chi^2$  we retrieve a best fit extinction of  $A(V) = 0.77$  mag and  $0.33$  mag for the low- and high-redshift  $g-i > 0.75$  subsamples respectively. **The extinguished blue stacked spectra reproduces the red stacked spectra, suggesting the diversity in Type I class I objects is driven by extinction, potentially nuclear or galaxy-wide. Our ML selection methodology is therefore able to recover AGNs at higher obscuration levels.** Although

reddening can naturally explain the spectral shapes of the red tail of Class I objects, we cannot entirely rule out the contribution of a redder stellar continuum for these sources.

To investigate the properties of our class II follow-up spectroscopy as a function of optical color, we split the class based on the  $g-i$  color. Since class II is optically redder than class I, we define the “red” subsample to have  $g-i > 1.5$ , and the “blue” subsample as  $g-i \leq 1.5$ . In addition, we restrict our analysis to Type II AGN, the majority constituent (65%) of class II. Figure 15 presents the stacks as a function of redshift and optical magnitude. Similar to class I, the emission-line shape and strength are consistent across optical magnitude, although we are unable to draw any strong conclusions across redshift as we only have two blue high-redshift Type II class II AGNs, and no corresponding red galaxies. We again fit the red stacked spectra with the blue stacked spectra by applying the SMC bar extinction curve, however this can only be done at low-redshift as there are no ‘red’, high-redshift, spectra in our sample. We find a best-fit  $A(V) = 1.26$  mag, suggesting Type II class II AGNs are distinguished primarily by varying levels of extinction, but deeper spectroscopy would be required to constrain the differences, if any, in the underlying stellar continua.

#### 4.3. Balmer Decrement



**Figure 14.** Optical color-color (top left) and MMT redshift (bottom left) distributions of class I (blue) and “red” ( $g-i > 0.75$ ; dark blue) and “blue” ( $g-i \leq 0.75$ ; light blue) subsamples. In addition, we present stacked spectra for low-redshift ( $z < 1.2$ ; top right) and high-redshift ( $z \geq 1.2$ ; bottom right) subsamples for red and blue class I targets identified as Type I AGNs. We fit the red stacked spectra with the blue stacked spectra combined with the SMC Bar extinction curve. Stacked spectra are smoothed with a boxcar function of width of  $10\text{\AA}$  for plotting purposes only.

To further investigate the level of extinction in the ML-selected classes, we measure the Balmer decrement in our spectroscopic sample. Typically the  $H\alpha/H\beta$  line ratio is used as the Balmer decrement however the redshift of our galaxies combined with our spectral coverage requires that we measure the  $H\gamma/H\beta$  line ratio to compute the effects of attenuation. While measurements of  $A(V)$  from  $H\gamma/H\beta$  are not as robust as those measured from the traditional  $H\alpha/H\beta$  ratio, we aim to bolster our interpretation that AGN in our sample exhibit obscuration, rather than precise measurements of extinction in specific galaxies. In this work we assume case B recombination in which the value of the  $H\gamma/H\beta$  ratio is 0.47 (Osterbrock & Ferland 2006).

We plot the Balmer decrement for spectra with confidently measured  $H\beta$  and  $H\gamma$  narrow-line fluxes ( $\text{SNR} > 5$ ) against spectroscopic redshift in Figure 16. Using the SMC Bar extinction curve, we compute the requisite level of attenuation implied by the measured Balmer decrement. While we can only obtain the requisite SNR for 16 objects, 11 from class I and five from class II, the spectra exhibit attenuations ranging from  $A(V) \simeq 0-2.5$  mag, further indicating that extinction is what drives the diversity in the optical colors of AGNs within classes I and II.

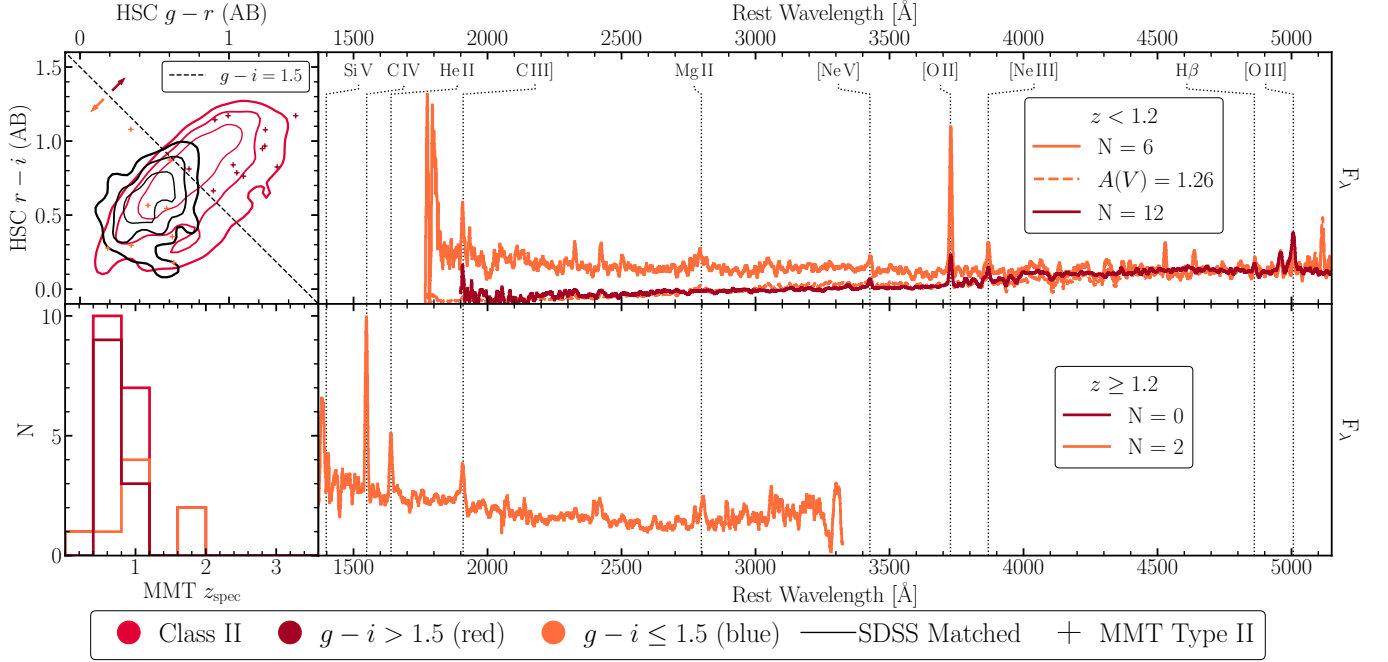
As we investigate the narrow-line Balmer decrement, our measurements are sensitive to extinction on larger

scales than the nucleus. We measure a weighted-mean attenuation of  $A(V) \simeq 1.3$  for both class I and class II, representative of significant galaxy-scale obscuration for these sources. However, due to the few galaxies for which we can perform this analysis and the larger uncertainties on the recovered lines, especially for the weaker  $H\gamma$ , further investigation would be required to draw conclusions about each class. This is especially true for class I, which is comprised nearly entirely of Type I AGN, as our analysis relies on the decomposition of broad and narrow emission, which can contribute an additional source of uncertainty.

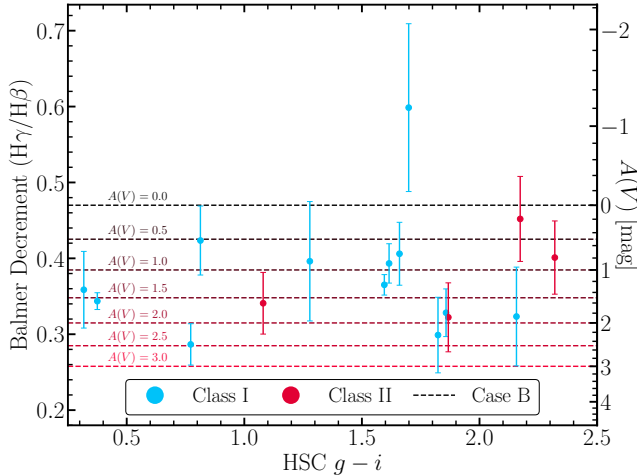
## 5. DISCUSSION

Our ML-selected classes represent substantial AGN populations over the entire  $\sim 1,000\text{ deg}^2$  HSC footprint. The spectroscopic analysis conducted in this work demonstrates that the high number densities of class I and II are not driven by contamination and are indeed comprised of accreting SMBHs. Accounting for the AGN-selection accuracy, the on-sky density of class I equates to  $\sim 145\text{ AGNs deg}^{-2}$  while class II represents over  $\sim 200\text{ AGNs deg}^{-2}$ . Even when taking into account targets for which we could not retrieve a redshift, which primarily affects class II, the classes together make up  $\gtrsim 240\text{ AGNs deg}^{-2}$ . Classes I and II therefore represent some of the densest AGN samples over large areas, es-





**Figure 15.** Optical color-color (top left) and MMT redshift (bottom left) distributions of class II (red) and “red” ( $g-i > 1.5$ ; dark red) and “blue” ( $g-i \leq 1.5$ ; orange) subsamples. In addition, we present stacked spectra for low-redshift ( $z < 1.2$ ; top right) and high-redshift ( $z \geq 1.2$ ; bottom right) subsamples for red and blue class II targets identified as Type II AGNs. We fit the ‘red’ stacked spectra with the ‘blue’ stacked spectra combined with the SMC Bar extinction curve. Stacked spectra are smoothed with a boxcar function of width of  $10\text{\AA}$  for plotting purposes only.



**Figure 16.** Balmer decrement ( $H\gamma/H\beta$  (left) and attenuation (right) versus  $g-i$  color for class I (blue) and class II (red) objects with  $H\beta$  and  $H\gamma$  flux SNRs  $> 5$ . We plot horizontal lines corresponding to the necessary  $A(V)$  required to achieve a specific Balmer decrement assuming the SMC Bar extinction curve.

pecially compared to samples selected from SDSS or WISE.

Critically, a substantial fraction of these galaxies lie below the detection limits of typical WISE criteria or inhabit in regions of mid-IR and optical color space that

typically suffer heavy contamination from star-forming galaxies. This is especially true at the reddest optical colors and faintest optical magnitude, where there is limited existing SDSS spectroscopy that recovers these AGN populations. Given that our spectroscopic analysis suggests that the red optical colors in our samples are driven primarily by dust extinction, rather than stellar properties or redshift, our ML-selected AGNs represent a window into low-luminosity and/or obscured SMBH growth for large samples.

In particular, the reddest class I Type I AGNs appear to exhibit similar spectral properties to the reddest quasars studied in Glikman et al. (2007, 2022); Banerji et al. (2015); Hamann et al. (2017). Fawcett et al. (2022) find that red quasars are distinguished primarily from ‘blue quasars’ through extinction with measured values of  $A(V)$  reaching 0.7 mag, consistent with the extinction of our low-redshift red spectral stack. A similar trend with extinction was additionally observed for Type I quasars in (Pat et al. 2022), also through the use of unsupervised ML. Our ML approach may present an effective methodology to finding red quasars along with blue quasars at higher levels of AGN-selection accuracy than traditional spectroscopic quasar searches.

Class II is predominantly comprised of Type II AGNs. Historically, optical spectroscopy has been an effective tool for assembling samples of Type II AGNs, though

their demographics are not fully understood (Reyes et al. 2008; Yuan et al. 2016). Class II appears to recover Type II AGNs similar to previous study and may additionally encompass AGNs similar to those selected from small-area X-ray surveys (Lusso et al. 2011) or to the Type II AGNs identified in Hviding et al. (2022) which cannot be selected using mid-IR colors alone. Our selection suggests these sources can be directly targeted using a combination of optical imaging and mid-IR data without resorting to X-ray observations or subsets of emission line galaxy samples.

Critically, these galaxies likely represent an opportunity to understand weaker AGN activity relative to the host galaxy and at higher redshifts. As simulations suggest a galaxy’s SMBH spends a larger fraction of its life as a low-Eddington-ratio AGNs rather than a powerful quasar, it remains a priority to fully detect and characterize the low-luminosity AGN population (Novak et al. 2011; Schawinski et al. 2015). Class II AGNs may reflect the impact of low-luminosity AGN feedback on the host galaxy as they co-evolve through cosmic time, or even for testing the accuracy of model predictions for the time-scales of low-Eddington-rate SMBH accretion.

The combination of optical imaging with mid-IR photometry presents a powerful AGN selection methodology. Mid-IR colors, sensitive to obscured and unobscured SMBH accretion, paired with optical colors, sensitive to host-galaxy properties, enables the retrieval of accurate AGN samples. This is especially true when paired with additional discerning information, such as the optical morphologies used in this work, to provide additional insight into the galaxy’s dynamical history. The inclusion of additional, non-color, information, such as morphology, proves to be an effective approach for selecting accurate samples of AGNs across the entire sky that could not previously be recovered solely by using photometric colors from one wavelength regime.

## 6. CONCLUSIONS & FUTURE WORK

In this work we perform optical spectroscopic follow up of AGN candidates selected from a parent sample of WISE-matched HSC targets using ML. The optical morphological data paired with the longer wavelengths afforded by WISE present a powerful dataset for selecting AGN candidates which can be fully explored through our use of ML algorithms. We present the results of our MMT spectroscopy which enable the characterization of our ML-selected classes, specifically with respect to AGN type and selection accuracy.

We recover redshifts in  $\sim 80\%$  of our 178 MMT spectra drawn from our target classes. Our spectroscopy confirms that classes I and II are predominately com-

prised of AGNs, reiterating that deep optical imaging joined with mid-infrared data is effective at selecting AGN candidates from photometry alone. Class I has a small contamination rate ( $< 3\%$ ) and is dominated by Type I AGNs ( $> 90\%$ ) while class II has a moderate contamination rate ( $\sim 15\%$ ) and is mostly made up of Type II AGNs ( $\sim 65\%$ ). The AGN-selection accuracy of our follow-up spectroscopy highlights the efficacy of deep optical imaging combined with mid-IR photometry especially when paired with ML techniques to analyze the multi-dimensional color-morphology parameter space.

Critically, our AGN classes occupy an area of the parameter space that has not been explored by traditional spectroscopic surveys such as SDSS. Our selected AGNs live in galaxies that are more diffuse and exhibit redder colors than the traditional spectroscopic samples. In addition, we demonstrate that the difference between the “red” and “blue” stacked spectra can be explained by applying an extinction curve, suggesting that AGNs from the ML-selected classes are drawn from a continuum in extinction. This is further supported by our investigation of the Balmer decrements in these sources that exhibit a range of attenuations from  $A(V) \simeq 0$  mag to  $A(V) \simeq 2.5$  mag. Together this suggests our selection probes more obscured sources with a more easily detectable host galaxy, i.e. low-luminosity AGN activity.

In the future, we aim to extend our spectroscopy to fainter optical magnitudes. This is especially relevant for class II, where we were restricted to the brightest 75% of objects in this work, and primarily retrieved redshifts for the brightest 20% of the class. Class II follow-up would be ideal from MMT Binospec, a multiplexed high-throughput spectrograph that can probe fainter targets than MMT Hectospec, albeit over a smaller area (Fabricant et al. 2019). Additional spectroscopy in class II will also enable the individual characterization of the parent UMAP subclasses *b*, *c*, and *d*.

The next generation of spectroscopic surveys from the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2023) or the upcoming 4-metre Multi-Object Spectrograph Telescope (4MOST; De Jong et al. 2019) will further inform the properties of ML-selected classes with spectra of fainter and redder galaxies/AGN. Furthermore, ML selection techniques may also prove useful for mitigating quasar misidentification in spectroscopic validation pipelines which can impact the AGN recovery fraction by a few percent as seen in the SDSS and DESI surveys (Farr et al. 2020; Alexander et al. 2023). In addition, studying the X-ray and radio properties of our ML-identified AGN candidates may provide additional insight into the accretion

properties of our sample and consequently their context relative to other AGN samples.

Finally, once the next generation of optical and near-IR survey telescopes come online, the ML methodology can be applied to select orders of magnitude more AGN candidates. Rubin, for example, presents a similar filter set and depth to HSC, but whose Legacy Survey of Space and Time (LSST) will cover an area of the sky that is nearly an order of magnitude larger. The recovery of these new AGN populations will inform our understanding of galaxy and SMBH coevolution and provide follow-up targets for further in-depth study. The unique composition of optical and mid-IR data paired with ML techniques proves to be a powerful tool for selecting AGN candidates from the next generation of surveys that are currently missing from current selection techniques.

We would like to thank the anonymous reviewer for their constructive comments which improved the final manuscript.

REH acknowledges support from the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1746060. KNH was supported by the National Aeronautics and Space Administration (NASA) Contract NAS50210 to the University of Arizona. ADG was supported through NASA Astrophysical Data Analysis Program (ADAP) award 80NSSC23K0485.

This work makes use of color palettes created by Martin Krzywinski designed for colorblindness. The color palettes and more information can be found at <http://mkweb.bcgsc.ca/colorblind/>.

Observations reported here were obtained at the MMT Observatory, a joint facility of the University of Arizona and the Smithsonian Institution.

We respectfully acknowledge the University of Arizona is on the land and territories of Indigenous peoples. Today, Arizona is home to 22 federally recognized tribes, with Tucson being home to the O’odham and the Yaqui. Committed to diversity and inclusion, the University strives to build sustainable relationships with sovereign Native Nations and Indigenous communities through education offerings, partnerships, and community service.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High

Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at <http://dm.lsst.org>.

The Pan-STARRS1 Surveys (PS1) have been made possible through contributions of 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, 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 under Grant No. AST-1238877, the University of Maryland, and Eotvos Lorand University (ELTE) and the Los Alamos National Laboratory.

Based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center at National Astronomical Observatory of Japan.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

This publication also makes use of data products from NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology, funded by the Planetary Science Division of the National Aeronautics and Space Administration.

This work has made use of data from the European Space Agency (ESA) mission Gaia (<https://www>.

[cosmos.esa.int/gaia](https://cosmos.esa.int/gaia)), processed by the Gaia Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions.

SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is [www.sdss4.org](http://www.sdss4.org).

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics — Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam

(AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

*Facilities:* MMT (Hectospec), Sloan, Subaru (HSC), WISE, NEOWISE

*Software:* **Astropy** (Collaboration et al. 2013), **GELATO** (Hviding 2022), **HSRED** (Fabricant et al. 2005, Section 7), **L<sup>A</sup>T<sub>E</sub>X** (Lamport 1994), **Matplotlib** (Hunter 2007), **NumPy** (Oliphant 2006; van der Walt et al. 2011; Harris et al. 2020), **redshifting** (Johnson et al. 2018; Helton et al. 2021; Johnson et al. 2022), **SciPy** (Virtanen et al. 2020), **SpectRes** (Carnall 2017, 2021)

## REFERENCES

- Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, The Astrophysical Journal Supplement Series, 249, 3, doi: [10.3847/1538-4365/ab929e](https://doi.org/10.3847/1538-4365/ab929e)
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, Publications of the Astronomical Society of Japan, 70, S4, doi: [10.1093/pasj/psx066](https://doi.org/10.1093/pasj/psx066)
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2022, Publications of the Astronomical Society of Japan, 74, 247, doi: [10.1093/pasj/psab122](https://doi.org/10.1093/pasj/psab122)
- Alexander, D. M., & Hickox, R. C. 2012, New Astronomy Reviews, 56, 93, doi: [10.1016/j.newar.2011.11.003](https://doi.org/10.1016/j.newar.2011.11.003)
- Alexander, D. M., Davis, T. M., Chaussidon, E., et al. 2023, The Astronomical Journal, 165, 124, doi: [10.3847/1538-3881/acacfc](https://doi.org/10.3847/1538-3881/acacfc)
- Assef, R. J., Stern, D., Noirot, G., et al. 2018, The Astrophysical Journal Supplement Series, 234, 23, doi: [10.3847/1538-4365/aaa00a](https://doi.org/10.3847/1538-4365/aaa00a)
- Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2010, The Astrophysical Journal, 713, 970, doi: [10.1088/0004-637X/713/2/970](https://doi.org/10.1088/0004-637X/713/2/970)
- Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, The Astrophysical Journal, 772, 26, doi: [10.1088/0004-637X/772/1/26](https://doi.org/10.1088/0004-637X/772/1/26)
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, Publications of the Astronomical Society of the Pacific, 93, 5, doi: [10.1086/130766](https://doi.org/10.1086/130766)
- Banerji, M., Alaghband-Zadeh, S., Hewett, P. C., & McMahon, R. G. 2015, Monthly Notices of the Royal Astronomical Society, 447, 3368, doi: [10.1093/mnras/stu2649](https://doi.org/10.1093/mnras/stu2649)
- Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, The Astronomical Journal, 144, 144, doi: [10.1088/0004-6256/144/5/144](https://doi.org/10.1088/0004-6256/144/5/144)
- Carnall, A. 2021, Astrophysics Source Code Library, ascl:2104.019
- Carnall, A. C. 2017, SpectRes: A Fast Spectral Resampling Tool in Python, doi: [10.48550/arXiv.1705.05165](https://doi.org/10.48550/arXiv.1705.05165)
- Collaboration, A., Robitaille, T. P., Tollerud, E. J., et al. 2013, Astronomy and Astrophysics, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- Collaboration, G., Brown, A. G. A., Vallenari, A., et al. 2018, Astronomy & Astrophysics, 616, A1, doi: [10.1051/0004-6361/201833051](https://doi.org/10.1051/0004-6361/201833051)
- Comparat, J., Delubac, T., Jovel, S., et al. 2016, Astronomy and Astrophysics, 592, A121, doi: [10.1051/0004-6361/201527377](https://doi.org/10.1051/0004-6361/201527377)



- Cover, T., & Hart, P. 1967, *IEEE Transactions on Information Theory*, 13, 21, doi: [10.1109/TIT.1967.1053964](https://doi.org/10.1109/TIT.1967.1053964)
- Cutri, R. M., Wright, E. L., Conrow, T., et al. 2021, *VizieR Online Data Catalog*, II/328
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *The Astronomical Journal*, 145, 10, doi: [10.1088/0004-6256/145/1/10](https://doi.org/10.1088/0004-6256/145/1/10)
- De Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, *Published in The Messenger* vol. 175, pp. 3-11, 9 pages, doi: [10.18727/0722-6691/5117](https://doi.org/10.18727/0722-6691/5117)
- DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2023, *The Early Data Release of the Dark Energy Spectroscopic Instrument*, doi: [10.48550/arXiv.2306.06308](https://doi.org/10.48550/arXiv.2306.06308)
- Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, *The Astrophysical Journal*, 748, 142, doi: [10.1088/0004-637X/748/2/142](https://doi.org/10.1088/0004-637X/748/2/142)
- Doré, O., Werner, M. W., Ashby, M., et al. 2016, *arXiv e-prints*, 1606, arXiv:1606.07039
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in *Second International Conference on Knowledge Discovery and Data Mining (Association for the Advancement of Artificial Intelligence Press)*, 226–331
- Fabian, A. C. 2012, *Annual Review of Astronomy and Astrophysics*, 50, 455, doi: [10.1146/annurev-astro-081811-125521](https://doi.org/10.1146/annurev-astro-081811-125521)
- Fabricant, D., Fata, R., Roll, J., et al. 2005, *Publications of the Astronomical Society of the Pacific*, 117, 1411, doi: [10.1086/497385](https://doi.org/10.1086/497385)
- Fabricant, D., Fata, R., Epps, H., et al. 2019, *Publications of the Astronomical Society of the Pacific*, 131, 075004, doi: [10.1088/1538-3873/ab1d78](https://doi.org/10.1088/1538-3873/ab1d78)
- Farr, J., Font-Ribera, A., & Pontzen, A. 2020, *Journal of Cosmology and Astroparticle Physics*, 2020, 015, doi: [10.1088/1475-7516/2020/11/015](https://doi.org/10.1088/1475-7516/2020/11/015)
- Fawcett, V. A., Alexander, D. M., Rosario, D. J., et al. 2022, *Monthly Notices of the Royal Astronomical Society*, 513, 1254, doi: [10.1093/mnras/stac945](https://doi.org/10.1093/mnras/stac945)
- Feltre, A., Charlot, S., & Gutkin, J. 2016, *Monthly Notices of the Royal Astronomical Society*, 456, 3354, doi: [10.1093/mnras/stv2794](https://doi.org/10.1093/mnras/stv2794)
- Fitzpatrick, E. L. 1999, *Publications of the Astronomical Society of the Pacific*, 111, 63, doi: [10.1086/316293](https://doi.org/10.1086/316293)
- Fix, E., & Hodges, J. L. 1951, *USAF School of Aviation Medicine*
- Glikman, E., Helfand, D. J., White, R. L., et al. 2007, *The Astrophysical Journal*, 667, 673, doi: [10.1086/521073](https://doi.org/10.1086/521073)
- Glikman, E., Lacy, M., LaMassa, S., et al. 2022, *The Astrophysical Journal*, 934, 119, doi: [10.3847/1538-4357/ac6bee](https://doi.org/10.3847/1538-4357/ac6bee)
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, *The Astrophysical Journal*, 594, 279, doi: [10.1086/376774](https://doi.org/10.1086/376774)
- Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, *The Astronomical Journal*, 116, 3040, doi: [10.1086/300645](https://doi.org/10.1086/300645)
- Hamann, F., Zakamska, N. L., Ross, N., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 3431, doi: [10.1093/mnras/stw2387](https://doi.org/10.1093/mnras/stw2387)
- Hao, L., Strauss, M. A., Tremonti, C. A., et al. 2005, *The Astronomical Journal*, 129, 1783, doi: [10.1086/428485](https://doi.org/10.1086/428485)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)
- Helton, J. M., Johnson, S. D., Greene, J. E., & Chen, H.-W. 2021, *Monthly Notices of the Royal Astronomical Society*, 505, 5497, doi: [10.1093/mnras/stab1647](https://doi.org/10.1093/mnras/stab1647)
- Hickox, R. C., & Alexander, D. M. 2018, *Annual Review of Astronomy and Astrophysics*, 56, 625, doi: [10.1146/annurev-astro-081817-051803](https://doi.org/10.1146/annurev-astro-081817-051803)
- Hickox, R. C., Jones, C., Forman, W. R., et al. 2007, *The Astrophysical Journal*, 671, 1365, doi: [10.1086/523082](https://doi.org/10.1086/523082)
- . 2009, *The Astrophysical Journal*, 696, 891, doi: [10.1088/0004-637X/696/1/891](https://doi.org/10.1088/0004-637X/696/1/891)
- Hopkins, P. F., Strauss, M. A., Hall, P. B., et al. 2004, *The Astronomical Journal*, 128, 1112, doi: [10.1086/423291](https://doi.org/10.1086/423291)
- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Hvinding, R. E. 2022, *TheSkyentist/GELATO: GELATO v2.5.2*, Zenodo, doi: [10.5281/zenodo.7439084](https://doi.org/10.5281/zenodo.7439084)
- Hvinding, R. E., Hainline, K. N., Rieke, M., et al. 2022, *The Astronomical Journal*, 163, 224, doi: [10.3847/1538-3881/ac5e33](https://doi.org/10.3847/1538-3881/ac5e33)
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *The Astrophysical Journal*, 873, 111, doi: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c)
- Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, *The Astrophysical Journal*, 735, 112, doi: [10.1088/0004-637X/735/2/112](https://doi.org/10.1088/0004-637X/735/2/112)
- Johnson, S. D., Chen, H.-W., Straka, L. A., et al. 2018, *The Astrophysical Journal*, 869, L1, doi: [10.3847/2041-8213/aaf1cf](https://doi.org/10.3847/2041-8213/aaf1cf)
- Johnson, S. D., Schaye, J., Walth, G. L., et al. 2022, *The Astrophysical Journal*, 940, L40, doi: [10.3847/2041-8213/aca28e](https://doi.org/10.3847/2041-8213/aca28e)
- Kawanomoto, S., Uraguchi, F., Komiyama, Y., et al. 2018, *Publications of the Astronomical Society of Japan*, 70, 66, doi: [10.1093/pasj/psy056](https://doi.org/10.1093/pasj/psy056)

- Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, The Astrophysical Journal Supplement Series, 154, 166, doi: [10.1086/422816](https://doi.org/10.1086/422816)
- LaMassa, S. M., Georgakakis, A., Vivek, M., et al. 2019, The Astrophysical Journal, 876, 50, doi: [10.3847/1538-4357/ab108b](https://doi.org/10.3847/1538-4357/ab108b)
- Lamport, L. 1994, LaTeX: A Document Preparation System, 2nd edn. (Addison-Wesley Professional)
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, 1110, arXiv:1110.3193
- Lusso, E., Comastri, A., Vignali, C., et al. 2011, Astronomy and Astrophysics, 534, A110, doi: [10.1051/0004-6361/201117175](https://doi.org/10.1051/0004-6361/201117175)
- Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, The Astrophysical Journal Supplement Series, 250, 8, doi: [10.3847/1538-4365/aba623](https://doi.org/10.3847/1538-4365/aba623)
- Lyu, J., & Rieke, G. 2022, Universe, 8, 304, doi: [10.3390/universe8060304](https://doi.org/10.3390/universe8060304)
- Lyu, J., & Rieke, G. H. 2017, The Astrophysical Journal, 841, 76, doi: [10.3847/1538-4357/aa7051](https://doi.org/10.3847/1538-4357/aa7051)
- Mainzer, A., Bauer, J., Grav, T., et al. 2011, The Astrophysical Journal, 731, 53, doi: [10.1088/0004-637X/731/1/53](https://doi.org/10.1088/0004-637X/731/1/53)
- Mateos, S., Alonso-Herrero, A., Carrera, F. J., et al. 2012, Monthly Notices of the Royal Astronomical Society, 426, 3271, doi: [10.1111/j.1365-2966.2012.21843.x](https://doi.org/10.1111/j.1365-2966.2012.21843.x)
- McInnes, L., Healy, J., & Melville, J. 2018, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, doi: [10.48550/arXiv.1802.03426](https://doi.org/10.48550/arXiv.1802.03426)
- Mendez, A. J., Coil, A. L., Aird, J., et al. 2013, The Astrophysical Journal, 770, 40, doi: [10.1088/0004-637X/770/1/40](https://doi.org/10.1088/0004-637X/770/1/40)
- Mignoli, M., Vignali, C., Gilli, R., et al. 2013, Astronomy and Astrophysics, 556, A29, doi: [10.1051/0004-6361/201220846](https://doi.org/10.1051/0004-6361/201220846)
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, Publications of the Astronomical Society of Japan, 70, S1, doi: [10.1093/pasj/psx063](https://doi.org/10.1093/pasj/psx063)
- Nakajima, K., Schaerer, D., Le Fèvre, O., et al. 2018, Astronomy and Astrophysics, 612, A94, doi: [10.1051/0004-6361/201731935](https://doi.org/10.1051/0004-6361/201731935)
- Negus, J., Comerford, J. M., Sánchez, F. M., et al. 2023, The Astrophysical Journal, 945, 127, doi: [10.3847/1538-4357/acb772](https://doi.org/10.3847/1538-4357/acb772)
- Nishizawa, A. J., Hsieh, B.-C., Tanaka, M., & Takata, T. 2020, Photometric Redshifts for the Hyper Suprime-Cam Subaru Strategic Program Data Release 2, doi: [10.48550/arXiv.2003.01511](https://doi.org/10.48550/arXiv.2003.01511)
- Novak, G. S., Ostriker, J. P., & Ciotti, L. 2011, The Astrophysical Journal, 737, 26, doi: [10.1088/0004-637X/737/1/26](https://doi.org/10.1088/0004-637X/737/1/26)
- Oliphant, T. E. 2006, A Guide to NumPy, Vol. 1 (Trelgol Publishing USA)
- Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei, 2nd. ed. by D.E. Osterbrock and G.J. Ferland. Sausalito, CA: University Science Books, 2006
- Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, Astronomy and Astrophysics Review, 25, 2, doi: [10.1007/s00159-017-0102-9](https://doi.org/10.1007/s00159-017-0102-9)
- Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, Astronomy and Astrophysics, 613, A51, doi: [10.1051/0004-6361/201732445](https://doi.org/10.1051/0004-6361/201732445)
- Pat, F., Juneau, S., Böhm, V., et al. 2022, Reconstructing and Classifying SDSS DR16 Galaxy Spectra with Machine-Learning and Dimensionality Reduction Algorithms, doi: [10.48550/arXiv.2211.11783](https://doi.org/10.48550/arXiv.2211.11783)
- Reyes, R., Zakamska, N. L., Strauss, M. A., et al. 2008, The Astronomical Journal, 136, 2373, doi: [10.1088/0004-6256/136/6/2373](https://doi.org/10.1088/0004-6256/136/6/2373)
- Richards, G. T., Hall, P. B., Vanden Berk, D. E., et al. 2003, The Astronomical Journal, 126, 1131, doi: [10.1086/377014](https://doi.org/10.1086/377014)
- Schawinski, K., Koss, M., Berney, S., & Sartori, L. F. 2015, Monthly Notices of the Royal Astronomical Society, 451, 2517, doi: [10.1093/mnras/stv1136](https://doi.org/10.1093/mnras/stv1136)
- Schlafly, E. F., & Finkbeiner, D. P. 2011, The Astrophysical Journal, 737, 103, doi: [10.1088/0004-637X/737/2/103](https://doi.org/10.1088/0004-637X/737/2/103)
- Schlafly, E. F., Meisner, A. M., & Green, G. M. 2019, The Astrophysical Journal Supplement Series, 240, 30, doi: [10.3847/1538-4365/aafbea](https://doi.org/10.3847/1538-4365/aafbea)
- Schmidt, M., & Green, R. F. 1983, The Astrophysical Journal, 269, 352, doi: [10.1086/161048](https://doi.org/10.1086/161048)
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, 1503, arXiv:1503.03757
- Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, The Astrophysical Journal, 631, 163, doi: [10.1086/432523](https://doi.org/10.1086/432523)
- Stern, D., Assef, R. J., Benford, D. J., et al. 2012, The Astrophysical Journal, 753, 30, doi: [10.1088/0004-637X/753/1/30](https://doi.org/10.1088/0004-637X/753/1/30)
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, The Astronomical Journal, 124, 1810, doi: [10.1086/342343](https://doi.org/10.1086/342343)
- Trouille, L., Barger, A. J., & Tremonti, C. 2011, The Astrophysical Journal, 742, 46, doi: [10.1088/0004-637X/742/1/46](https://doi.org/10.1088/0004-637X/742/1/46)

- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science Engineering, 13, 22, doi: [10.1109/MCSE.2011.37](https://doi.org/10.1109/MCSE.2011.37)
- Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, Monthly Notices of the Royal Astronomical Society, 463, 3409, doi: [10.1093/mnras/stw2231](https://doi.org/10.1093/mnras/stw2231)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, The Astronomical Journal, 140, 1868, doi: [10.1088/0004-6256/140/6/1868](https://doi.org/10.1088/0004-6256/140/6/1868)
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, The Astronomical Journal, 120, 1579, doi: [10.1086/301513](https://doi.org/10.1086/301513)
- Yuan, S., Strauss, M. A., & Zakamska, N. L. 2016, Monthly Notices of the Royal Astronomical Society, 462, 1603, doi: [10.1093/mnras/stw1747](https://doi.org/10.1093/mnras/stw1747)
- Zhao, G.-B., Wang, Y., Ross, A. J., et al. 2016, Monthly Notices of the Royal Astronomical Society, 457, 2377, doi: [10.1093/mnras/stw135](https://doi.org/10.1093/mnras/stw135)