

Classification of Emission Line Galaxies

Ridha Fathima Mohideen Malik

Mentors: Dragana Ilic, Isidora Jankov

Department of Astronomy, University of Belgrade, Serbia

(Dated: June 4, 2023)

I. INTRODUCTION

In the era of Big data astronomy, we need more and more efficient automation to look, analyse and classify data. This task is complicated by the fact that we do not have complete and continuous information on a source in all wavelengths. So we have to use different schemes and diagnostic tools, each with their own assumptions and uncertainties. In this work, we focus on one part of this puzzle - classification of Emission Line Galaxies (ELGs).

Spectra obtained from a source can be seen as a sum of two components: the continuum and the emission or absorption features. When a galaxy has strong absorption features, we infer that they are dominated by older stellar population as seen in elliptical or early-type galaxies. While strong emission lines indicate the presence of very hot gas, leading to ionisation of these elements. There could be different sources that heat this gas - from active star formation to massive young stars to an accreting black hole. Such galaxies with strong emission lines are classified as Emission Line Galaxies (ELGs). The task of classifying ELGs then boils down to finding schemes that capture different modes and strengths of excitation mechanisms.

This report presents the results of some such classification schemes, and compares it with the Sloan Digital Sky Survey's classification pipelines.

II. DATA AND SAMPLE

The data used in this work were taken from the Sloan Digital Sky Survey (SDSS) Data Release 18 (DR18) (Almeida et al. 2023). The survey uses the 2.5m telescope facility in the Apache Point Observatory, USA. DR18 is the first data release of its fifth phase, which aims to perform optical and infra-red spectroscopy of the entire sky.

Our sample consists of 30,000 galaxies with narrow lines ($FWHM < 500\text{km/s}$, $S/N > 5$) and redshift $z < 0.3$ with an average $z \approx 0.08(1)$. The data was obtained using an SQL query (Section A) through Astroquery Python package.

SDSS's spectroscopic pipelines fit emission lines and determine line flux and redshift, after which the objects are classified. The spectra is fit with models of different astronomical objects using a simple least-squares fit, starting from the model of the targeted class. The model is generated from templates of previous data-releases with known redshifts and classes. The templates are not an average of spectra but a set of eigen basis obtained through Principal Component Analysis (PCA) (Bolton et al. 2012).

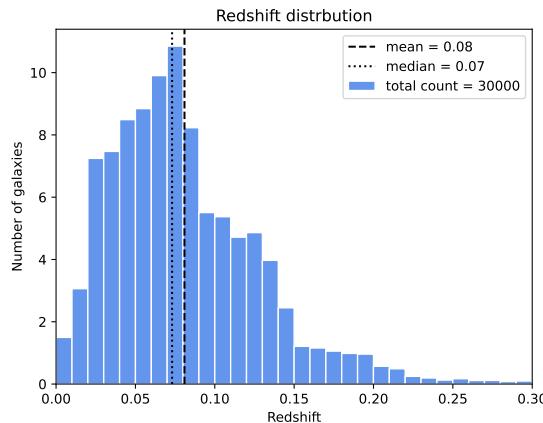


FIG. 1: Distribution of redshift in our sample

In our sample, 29719 (99.1%) objects are classified as star-forming galaxies (SFGs) and 281 (0.9%) as quasi-stellar objects (QSOs) according to SDSS's classification. Further sub-classification information in Table I. It is noticed that even though the query was made for narrow-line regions we have broadline regions. This is because often these classifications multiple features in addition to the FWHM criterion, and sometimes are ambiguous and have multiple labels associated with it. Our query result has only the primary label.

Subclass	Number
Starforming	13945
Starforming broadline	37
Starburst	12632
Starburst broadline	102
Broadline	107
AGN	2040
AGN broadline	232
No subclass	905

TABLE I: Automatic sub-classifications by SDSS pipelines

In Section V, a subsample with photometric data in SDSS g ($\lambda_{eff} \approx 4717\text{\AA}$) and z ($\lambda_{eff} \approx 8923\text{\AA}$) bands were employed. In Section VI, SDSS results were crossmatched with photometric data in W1 ($\lambda_{eff} \approx 12\mu m$), W2 ($\lambda_{eff} \approx 4.6\mu m$) and W3 ($\lambda_{eff} \approx 3.4\mu m$) bands from the Wide-field Infrared Survey Explorer(WISE) space telescope.

III. BPT DIAGRAM

The Baldwin - Phillips - Terlevich (BPT) diagram is a diagnostic tool used to distinguish HII regions from AGNs comparing line ratios of prominent lines in these regions as shown in Figure 2. In the case of BPT using the [SII] $\lambda\lambda 6717, 6731$ doublet, the flux of the 6717\AA line was picked with the assumption that their intensities do not significantly differ.

The theoretical division lines are taken from Kewley et al. 2001:

$$\log\left(\frac{[OIII]\lambda 5007}{H_\beta}\right) = \frac{0.61}{\log([NII]\lambda 6584/H_\alpha) - 0.47} + 1.19 \quad (1)$$

$$\log\left(\frac{[OIII]\lambda 5007}{H_\beta}\right) = \frac{0.72}{\log([SII]\lambda\lambda 6717, 6731/H_\alpha) - 0.32} + 1.30 \quad (2)$$

$$\log\left(\frac{[OIII]\lambda 5007}{H_\beta}\right) = \frac{0.73}{\log([OI]\lambda 6300/H_\alpha) + 0.59} + 1.33 \quad (3)$$

Kauffmann et al. 2003 provides empirical division line only for BPT with $[OIII]\lambda 5007/H_\beta$ vs $[NII]\lambda 6584/H_\alpha$:

$$\log\left(\frac{[OIII]\lambda 5007}{H_\beta}\right) = \frac{0.61}{\log([NII]\lambda 6584/H_\alpha) - 0.05} + 1.3 \quad (4)$$

The results are presented in Table II. It is observed that increasing number of objects are classified as AGNs with decreasing wavelength (2b,2a,2c). Instead of comparing with the ‘class’ provided by SDSS, if we group subclasses we have a higher percentage of QSO $\approx 7.6\%$. This value is closest to the values we obtain from BPT.

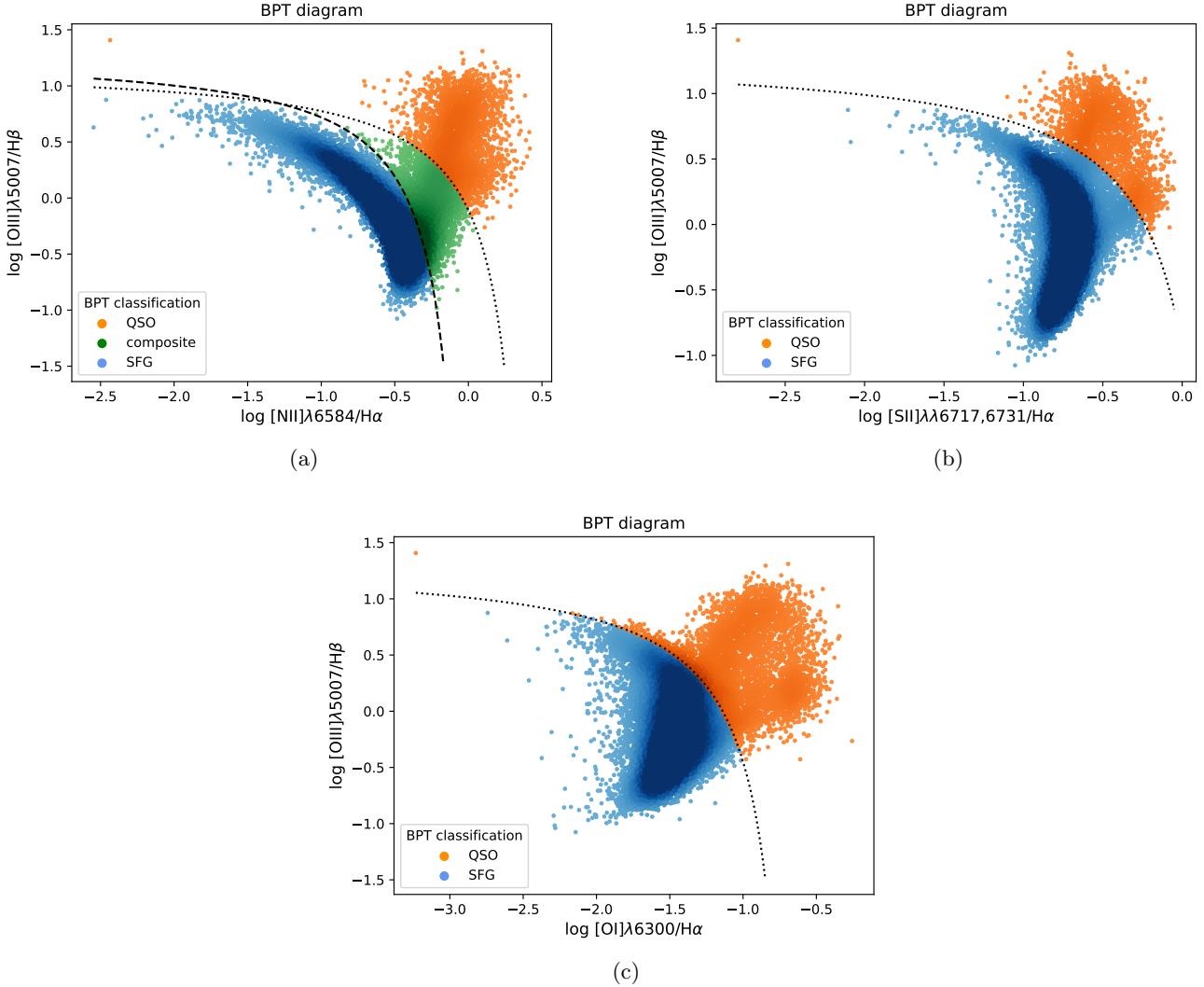


FIG. 2: BPT diagrams

Class	BPT (%)	Kewley [NII] (%)	Kewley [SI] (%)	Kewley [OI] (%)
SFG	78.3	90.3	93.3	85.4
composite	12	-	-	-
QSO	9.7	9.7	6.7	14.6

TABLE II: BPT classification results

IV. WHAN DIAGRAM

It is not always possible to have precise measurements of all the diagnostic lines required for a BPT diagram. Weak line galaxies(WLGs) are galaxies that do not have either or both of H_β and [O III] with $S/N \geq 3$ (Fernandes et al. 2010). Schemes like WHAN can cope with such galaxies (Cid Fernandes et al. 2011).

The vertical line at $\log[NII]\lambda6584/H\alpha = -0.4$ is an optimal transpose of BPT division. It is observed that while the optimal transpose is at -0.4 , the SFGs classified according to BPT extend up to -0.2 .

Besides distinguishing AGNs as strong (Seyferts) and weak (Low Ionisation Nuclear Emission Regions, or LINERs). WHAN diagram also provides subclasses according to the scheme shown in Figure 4. Applying this on our sample we

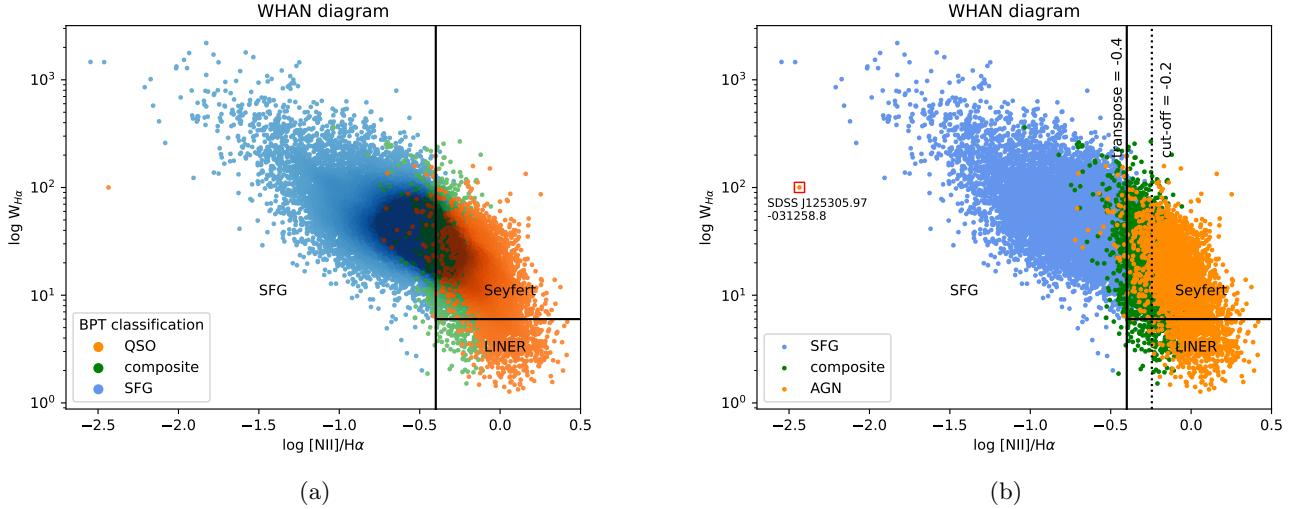


FIG. 3: WHAN diagram

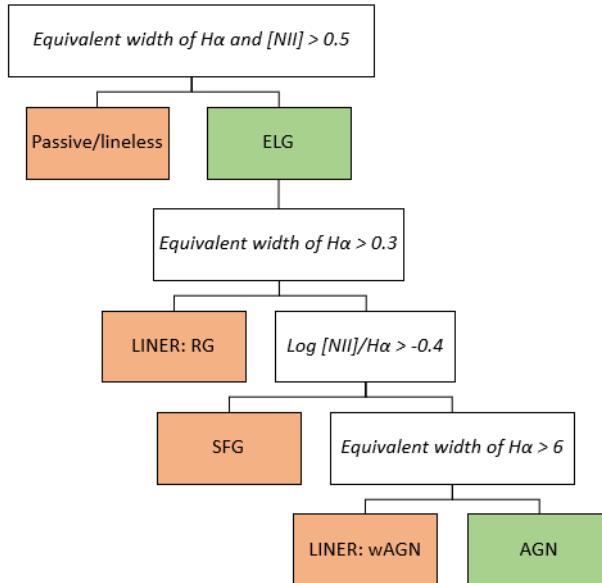


FIG. 4: WHAN classification scheme. Texts in italics are the conditions; Regular texts are classification; Orange colour box means condition(s) unsatisfied; Green colour box means condition(s) satisfied.

get Table III.

7.4 % of our galaxies are classified as AGN in both BPT and WHAN; 2.2 % are AGN in BPT but not in WHAN but 20.4 % are AGN in WHAN but not BPT.

There are 11 lineless galaxies because they did not meet the criterion on equivalent width of [NII] line and could be considered WLGs by the definition of WLGs based on line width (Fernandes et al. 2010).

There is a peculiar outlier on the right end (Figure 3b). Identified as SDSS J125305.97-031258.8, this is a case of a black hole present in a green pea galaxy (Malhotra 2018). Green peas are nearby galaxies with extreme lines, high ionisation and low metallicity, and hence the low line ratio.

Subclass	%
SFG	69.36
Seyfert	27.82
LINER: weak AGN (wAGN)	2.16
LINER: Retired Galaxy (RG)	6.33
Lineless	0.03

TABLE III: Sub-classifications by WHAN

V. TBT DIAGRAM

The limitation of using a purely spectroscopic approach as in Sections III & IV is that it could not be applied for high-z galaxies as their spectra shift to the infrared region. IR spectroscopy is challenging as this region has heavy interference from Earth's atmosphere, demanding usage of space telescopes. We do not yet have as many IR spectroscopic surveys to match our optical surveys like SDSS. This leads to the requirement of alternative diagnostic tools like the Trouille - Barger - Tremonti (TBT) diagram (Trouille, Barger, and Tremonti 2011).

It plots restframe colour ($g - z$) against $\log[\text{NeIII}]\lambda 3869]/[\text{OII}]\lambda\lambda 3726, 3729$ line ratio. One of the [OII] doublets were chosen with the same reasoning as in Section III. To obtain the restframe colour, the queried apparent magnitudes were corrected for extinction using the colour excess information present in SDSS. The extinction coefficient required to do dust correction in our given bands, R_λ , are taken from Schlafly and Finkbeiner 2011 for SDSS and Zhang and Yuan 2023 for WISE. Then, to account for the redshift, K-correction is done with the program provided by Chilingarian, Melchior, and Zolotukhin 2010. Thus the restframe colour is given as,

$$(g - z)_0 = (g - z)_{app} - R_\lambda * E(B - V) - K_{corr}(g), \quad (5)$$

where $(g - z)_{app}$ is the apparent colour, $E(B - V)$ is the colour excess, $K_{corr}(g)$ is the K correction in g-band, and $(g - z)_0$ is the restframe colour. Distance modulus corrections cancel out in the subtraction of magnitudes. Correcting to restframe allows TBT to reach out to $z < 1.4$ compared to $z < 0.5$ of BPT.

The division line is given by,

$$(g - z)_0 = -1.2 * \log \frac{[\text{NeIII}]\lambda 3869}{[\text{OII}]\lambda\lambda 3726, 3729} - 0.4 \quad (6)$$

26.5 % of our galaxies are classified as AGN in both BPT and TBT; 0.1 % are AGN in BPT but not in TBT but 7.9 % are AGN in TBT but not BPT.

VI. MIR CLASSIFICATION WITH WISE

Another way to get around the IR spectroscopy limitation is using IR photometry. The mid-infrared colors can be used to identify the presence of AGN-heated dust. Using WISE bands we get the colour-colour diagram in Figure 6a, plotted over the classification provided by Wright et al. 2010. Figure 6b & 6c explore classification criteria provided by Assef et al. 2013 and Mateos et al. 2012. Magnitude colours have been used instead of flux so that extinction correction could be performed consistently. The extinction coefficients for WISE colours were taken from Zhang and Yuan 2023. The results presented in Figure 7 show that MIR classifications have a stricter criteria for a 'pure' AGN. It is noted that MIR classification is close to the 'class' total of 0.9%. The percentage of strong AGNs in MIR classification could be a little higher if K-correction is done for the sample in IR bands.

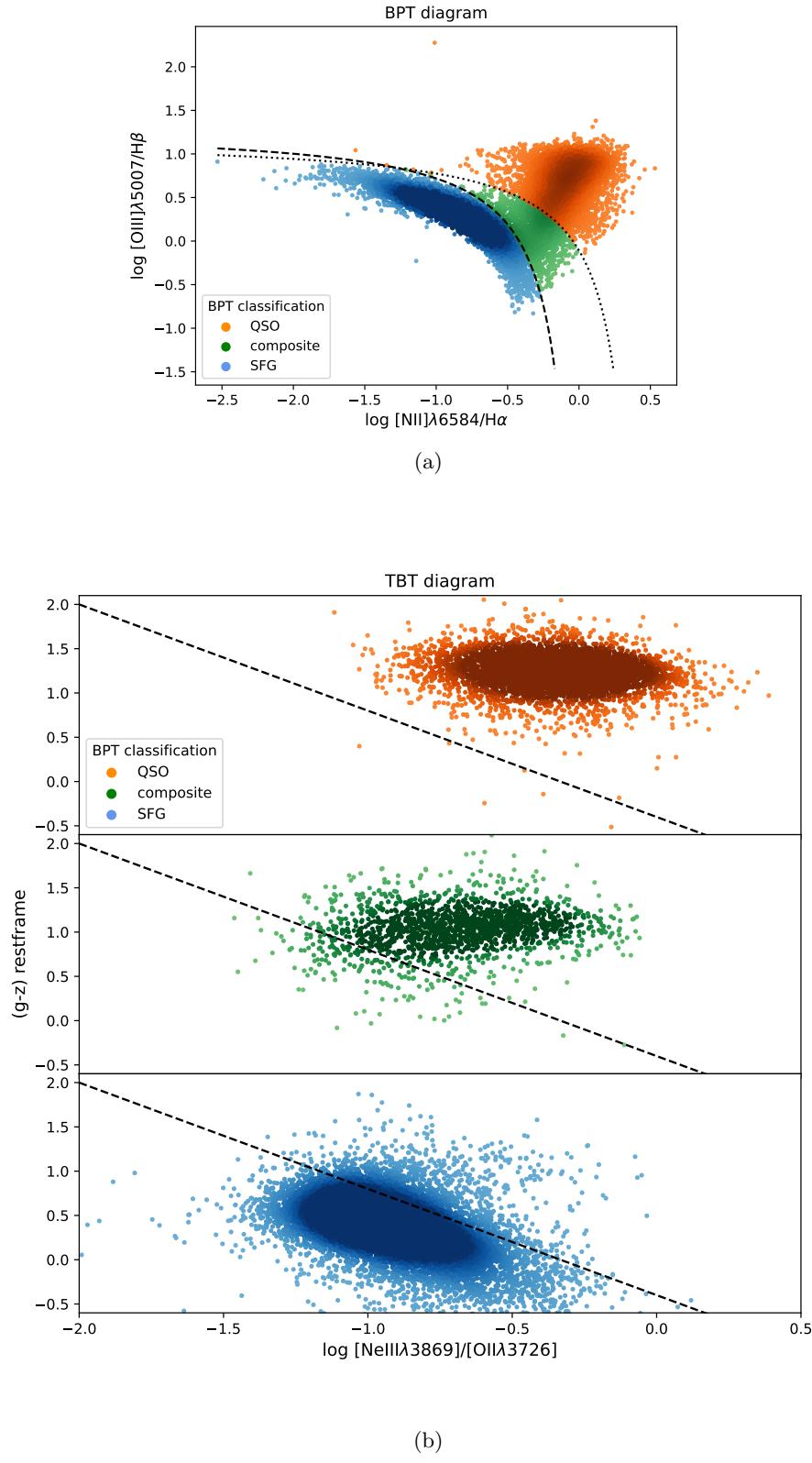
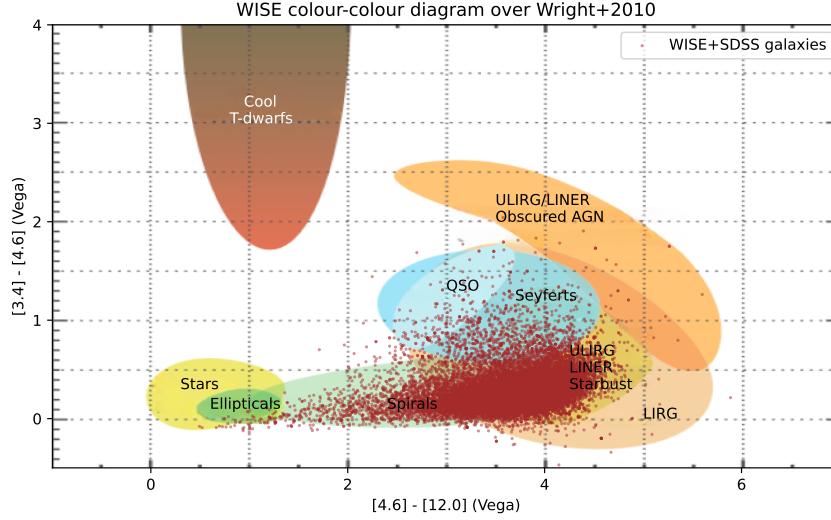


FIG. 5: TBT diagram



(a)

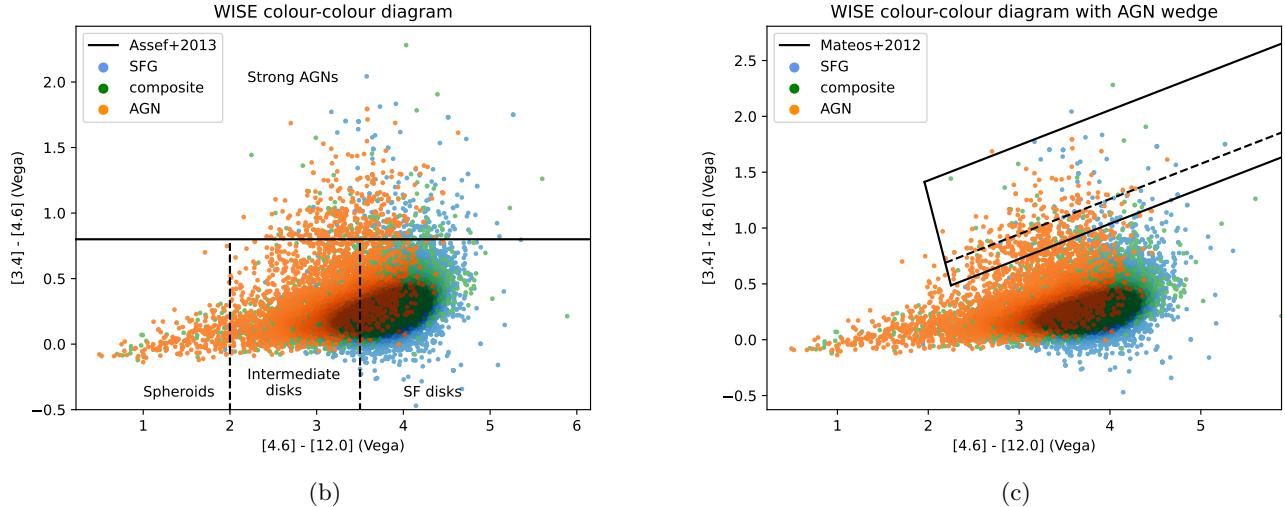


FIG. 6: WISE colour-colour diagram

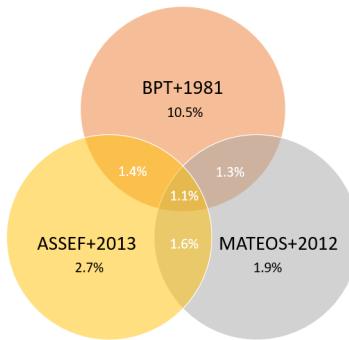


FIG. 7: Comparison of BPT with MIR classifications. The black texts give the total percentages; The white texts give the intersections.

VII. CONCLUSION

In this report, we looked at some of the classification schemes and diagnostic tools for Emission Line Galaxies. Each classification tries to overcome the limitations of the other. As these classifications capture different physical processes, ‘misclassifications’ between them could lead to interesting observations as in the case of SDSS J125305.97-031258.8. The ultimate goal of such explorations is faster and reliable classification pipelines that could help understand the physics of the respective population classes.

REFERENCES

- Almeida, Andrés et al. (Jan. 2023). *The Eighteenth Data Release of the Sloan Digital Sky Surveys: Targeting and First Spectra from SDSS-V*. Tech. rep. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230107688A Type: article. DOI: 10.48550/arXiv.2301.07688. URL: <https://ui.adsabs.harvard.edu/abs/2023arXiv230107688A> (visited on 05/27/2023).
- Assef, R. J. et al. (July 2013). “Mid-infrared Selection of Active Galactic Nuclei with the Wide-field Infrared Survey Explorer. II. Properties of WISE-selected Active Galactic Nuclei in the NDWFS Boötes Field”. In: *The Astrophysical Journal* 772. ADS Bibcode: 2013ApJ...772...26A, p. 26. ISSN: 0004-637X. DOI: 10.1088/0004-637X/772/1/26. URL: <https://ui.adsabs.harvard.edu/abs/2013ApJ...772...26A> (visited on 06/04/2023).
- Bolton, Adam S. et al. (Oct. 2012). “SPECTRAL CLASSIFICATION AND REDSHIFT MEASUREMENT FOR THE SDSS-III BARYON OSCILLATION SPECTROSCOPIC SURVEY”. en. In: *The Astronomical Journal* 144.5. Publisher: The American Astronomical Society, p. 144. ISSN: 1538-3881. DOI: 10.1088/0004-6256/144/5/144. URL: <https://dx.doi.org/10.1088/0004-6256/144/5/144> (visited on 05/27/2023).
- Chilingarian, Igor V., Anne-Laure Melchior, and Ivan Yu. Zolotukhin (July 2010). “Analytical approximations of K-corrections in optical and near-infrared bands”. In: *Monthly Notices of the Royal Astronomical Society* 405.3, pp. 1409–1420. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2010.16506.x. URL: <https://doi.org/10.1111/j.1365-2966.2010.16506.x> (visited on 06/04/2023).
- Cid Fernandes, R. et al. (May 2011). “A comprehensive classification of galaxies in the Sloan Digital Sky Survey: how to tell true from fake AGN?” In: *Monthly Notices of the Royal Astronomical Society* 413. ADS Bibcode: 2011MNRAS.413.1687C, pp. 1687–1699. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.18244.x. URL: <https://ui.adsabs.harvard.edu/abs/2011MNRAS.413.1687C> (visited on 06/01/2023).
- Fernandes, R. Cid et al. (Mar. 2010). “Alternative diagnostic diagrams and the ‘forgotten’ population of weak line galaxies in the SDSS”. In: *Monthly Notices of the Royal Astronomical Society* 403.2. eprint: <https://academic.oup.com/mnras/article-pdf/403/2/1036/4012938/mnras0403-1036.pdf>, pp. 1036–1053. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2009.16185.x. URL: <https://doi.org/10.1111/j.1365-2966.2009.16185.x>.
- Kauffmann, Guinevere et al. (Dec. 2003). “The host galaxies of active galactic nuclei”. In: *Monthly Notices of the Royal Astronomical Society* 346. ADS Bibcode: 2003MNRAS.346.1055K, pp. 1055–1077. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2003.07154.x. URL: <https://ui.adsabs.harvard.edu/abs/2003MNRAS.346.1055K> (visited on 06/01/2023).
- Kewley, L. J. et al. (July 2001). “Theoretical Modeling of Starburst Galaxies”. In: *The Astrophysical Journal* 556. ADS Bibcode: 2001ApJ...556..121K, pp. 121–140. ISSN: 0004-637X. DOI: 10.1086/321545. URL: <https://ui.adsabs.harvard.edu/abs/2001ApJ...556..121K> (visited on 06/01/2023).
- Malhotra, Sangeeta (Sept. 2018). “Black Holes in Green Peas?” In: *Chandra Proposal*. ADS Bibcode: 2018cxo..prop.5433M, p. 5433. URL: <https://ui.adsabs.harvard.edu/abs/2018cxo..prop.5433M> (visited on 06/01/2023).
- Mateos, S. et al. (Nov. 2012). “Using the Bright Ultrahard XMM-Newton survey to define an IR selection of luminous AGN based on WISE colours”. In: *Monthly Notices of the Royal Astronomical Society* 426. ADS Bibcode: 2012MNRAS.426.3271M, pp. 3271–3281. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.21843.x. URL: <https://ui.adsabs.harvard.edu/abs/2012MNRAS.426.3271M> (visited on 06/01/2023).
- Schlafly, Edward F. and Douglas P. Finkbeiner (Aug. 2011). “MEASURING REDDENING WITH SLOAN DIGITAL SKY SURVEY SLAR SPECTRA AND RECALIBRATING SFD”. en. In: *The Astrophysical Journal* 737.2. Publisher: The American Astronomical Society, p. 103. ISSN: 0004-637X. DOI: 10.1088/0004-637X/737/2/103. URL: <https://dx.doi.org/10.1088/0004-637X/737/2/103> (visited on 05/27/2023).
- Trouille, L., A. J. Barger, and C. Tremonti (Nov. 2011). “The OPTX Project. V. Identifying Distant Active Galactic Nuclei”. In: *The Astrophysical Journal* 742. ADS Bibcode: 2011ApJ...742...46T, p. 46. ISSN: 0004-637X. DOI: 10.1088/0004-637X/742/1/46. URL: <https://ui.adsabs.harvard.edu/abs/2011ApJ...742...46T> (visited on 06/01/2023).
- Wright, Edward L. et al. (Dec. 2010). “The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance”. In: *The Astronomical Journal* 140. ADS Bibcode: 2010AJ....140.1868W, pp. 1868–1881. ISSN: 0004-6256. DOI: 10.1088/0004-6256/140/6/1868. URL: <https://ui.adsabs.harvard.edu/abs/2010AJ....140.1868W> (visited on 06/01/2023).
- Zhang, Ruoyi and Haibo Yuan (Jan. 2023). “Empirical Temperature- and Extinction-dependent Extinction Coefficients for the GALEX, Pan-STARRS 1, Gaia, SDSS, 2MASS, and WISE Passbands”. In: *The Astrophysical Journal Supplement Series* 264. ADS Bibcode: 2023ApJS..264...14Z, p. 14. ISSN: 0067-0049. DOI: 10.3847/1538-4365/ac9dfa. URL: <https://ui.adsabs.harvard.edu/abs/2023ApJS..264...14Z> (visited on 05/27/2023).

Appendix A: SQL query

This is the combined SQL query. In this work, the SQL query was carried out in parts to obtain maximum data possible given the conditions for the diagnostic scheme that was used.

```

SELECT TOP 30000
s.plate, s.mjd, s.fiberid, s.class, g.subclass, s.z, g.e_bv_sfd,
l.h_alpha_flux, l.h_beta_flux, l.oiii_5007_flux, l.nii_6584_flux,
l.oi_6300_flux, l.sii_6717_flux, l.sii_6731_flux,
l.neiii_3869_flux, l.oii_3726_flux, l.oii_3729_flux,
l.h_alpha_eqw, l.nii_6584_eqw,
p.psfMag_g, p.psfMag_z,
w.w1mpo, w.w2mpo, w.w3mpo, w.w1flux, w.w2flux, w.w3flux

FROM SpecObjAll AS s
JOIN GalSpecInfo AS g ON s.specobjid = g.specobjid
JOIN GalSpecLine AS l ON s.specobjid = l.specobjid
JOIN PhotoObjAll as p on s.bestobjid = p.objid
JOIN wise_xmatch AS x ON s.bestobjid = x.sdss_objid
JOIN wise_allsky AS w ON x.wise_cntr = w.cntr

WHERE (s.class = "QSO" OR s.class = "GALAXY")
AND s.z < 0.3
AND (l.sigma_balmer * 2.355) < 500
AND l.h_alpha_flux > 0
AND l.h_beta_flux > 0
AND l.oiii_5007_flux > 0
AND l.nii_6584_flux > 0
AND l.oi_6300_flux > 0
AND l.sii_6717_flux > 0
AND l.sii_6731_flux > 0
AND l.neiii_3869_flux > 0
AND l.oii_3726_flux > 0
AND l.oii_3729_flux > 0
AND l.h_alpha_flux/l.h_alpha_flux_err > 5
AND l.h_beta_flux/l.h_beta_flux_err > 5
AND l.oiii_5007_flux/l.oiii_5007_flux_err > 5
AND l.nii_6584_flux/l.nii_6584_flux_err > 5
AND l.oi_6300_flux/l.oi_6300_flux_err > 5
AND l.sii_6717_flux/l.sii_6717_flux_err > 5
AND l.sii_6731_flux/l.sii_6731_flux_err > 5
AND l.neiii_3869_flux/l.neiii_3869_flux_err > 5
AND l.oii_3726_flux/l.oii_3726_flux_err > 5
AND l.oii_3729_flux/l.oii_3729_flux_err > 5
AND p.psfMag_g/p.psfMagErr_g > 5
AND p.psfMag_z/p.psfMagErr_z > 5
AND w.w1snr > 5
AND w.w2snr > 5
AND w.w3snr > 5
AND w.w1flux / w.w1sigflux > 5
AND w.w2flux / w.w2sigflux > 5
AND w.w3flux / w.w3sigflux > 5

```

The code to generate the classifications and plots along with the dataset can be found in this GitHub link.