Identifying Active Galactic Nuclei at High Redshift in Deep JWST Fields Using Self-Organizing Maps

James Zisk[a,b], Alexander de la Vega[c]

[a] *University of Hawai'i at Hilo, 200 W Kawili St, Hilo, HI 96720*
[b] *Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822*
[c] *University of California Riverside, 900 University Ave, Riverside, CA 92521*

Abstract

The James Webb Space Telescope (JWST) has discovered a new class of galaxies called Little Red Dots (LRDs) in the early universe, with spectra suggesting they hold at their center faint active galactic nuclei (AGN). These AGN may offer further insight into the early stages of galaxy and supermassive black hole formation, making the identification of these objects imperative. The spectral energy distribution of brown dwarfs are similar to LRDs and can contaminate LRD samples when identifying using color cuts. In this study, we test the accuracy and effectiveness of using self-organizing maps (SOM), an unsupervised machine learning technique, to identify potential LRDs in the JWST Advanced Deep Extragalactic Survey (JADES). We found that the SOMs, trained on colors, are capable of separating LRDs from brown dwarf models, negating one of the pitfalls of previous LRD identification methods. Using the classification from the SOM, we obtain a sample of 34 LRDs in the Great Observatories Origin Deep Survey (GOODS) North field.

1. Introduction

The James Webb Space Telescope (JWST) has given astronomers a deeper look into the earlier epochs of the universe. From these first surveys came the discovery of faint, broad-line AGNs at redshifts of $z > 5$ (Kocevski et al., 2023; Matthee et al., 2024). These objects, fainter than quasars at similar redshifts observed in previous ground-based surveys, hold supermassive black holes several times smaller than others observed in the early universe. The broad-line spectra observed from these objects suggest they are reddened active galactic nuclei (AGN) in a stage of black hole growth that has not been previously observed due to dust obscuration (Kocevski et al., 2024). Understanding how these faint AGNs compare to the growth of their host galaxy can give the needed insight into the nature of black hole seeing and how their co-evolution with their host galaxy, as sources of reionization and sources of gravitational waves.

LRDs are hard to detect, with previous studies using color cuts and rest-wavelength cuts to identify potential LRDs in recent surveys (eg., Barro et al. (2024); Kokorev et al. (2024); Kocevski et al. (2024). The spectral energy distributions of LRDs are very similar to brown dwarfs. Cuts are made to avoid brown dwarf models, but cool brown dwarfs are still within the cuts, contaminating the sample.

By using self-organizing maps (SOMs), an unsupervised machine learning technique, we intend to develop a strategy of training SOMs to create classifications to identify LRDs while seperating the brown dwarfs entirely.

In the following section, we discuss the survey and the two fields we will be conducting our study in, the instruments of JWST, and the restrictions placed on our

Redshift	β_{UV}	β_{opt}
$2 < z < 3.25$	F606W, F814W, F115W	F150W, F200W, F277W
$3.25 < z < 4.75$	F814W, F115W, F150W	F200W, F277W, F356W
$4.75 < z < 8$	F115W, F150W, F200W	F277W, F356W, F444W
$z > 8$	F150W, F200W, F277W	F356W, F444W

Table 1: Bands Used to Derive Continuum Slope versus Redshift

data. In the methodology section we will go over how we created our sample of LRDs, trained the SOMs, applied data from the new field and produced a list of new potential LRDs. The results section displays the effectiveness of the SOMs and showcases the objects we discovered.

2. Data and Sample Selection

Our research uses the JWST Advanced Deep Extragalactic Survey (JADES) data in the Great Observatories Origins Deep Survey (GOODS) fields, a total area of 124 arcminutes squared. The GOODS-South field, released in June 2023, is the field we are using to create our sample of LRDs and to train machine learning program. The GOODS-North field, released in May 2024, will be what field we will apply to our machine learning technique to identify new potential LRDs.

Several restrictions were applied to the GOODS-South data when creating our sample of LRDs. We set the magnitude limit to 29, just brighter the survey's limiting magnitude of 29.5 (Eisenstein et al., 2023). We masked out all objects at $z < 2$ as LRDs are high-redshift objects and to avoid any brown dwarfs that might be found at low-redshift that may be initially mistaken. Finally, we removed all objects with 'nan' or infinite values that would disrupt magnitude and slope calculations. With these restrictions, we had 22551 and 21783 in GOODS-South and -North respectively.

In this research, we will be using photometry data from the Near-Infrared Camera (NIRCam), EAZY for photometric redshift, and Near-Infrared Spectrograph (NIRSpec) low resolution PRISM and medium resolution spectroscopy to observe the Hα and Hβ lines of our objects (Hainline et al., 2024; Bunker et al., 2023; D'Eugenio et al., 2024).

3. Methodology

This study will use the criteria set in Kocevski et al. (2024) for classifying LRDs in the GOODS-North and South fields. To be classified as a potential LRD, each object must have an optical slope, β_{opt}, greater than 1.0 and a ultraviolet slope, β_{UV}, greater than -2.8. The magnitude cuts are equivalent to the color cuts of F277W-F444W > 1.0 and F150W-F200W < 0.5. An additional cut of β_{UV} less than -0.37, equivalent to F115W-F200W $<$ -0.5, accounts for the cool brown dwarfs that may resemble LRDs and reddened AGNs. The magnitudes and errors of each rest wavelength are calculated using a Monte Carlo method as chi-squared does not take into account covariance between parameters when calculating errors, creating inaccuracies in the errors. When making the cuts, we include the error into our classification, making the accuracy of the errors crucial to accurate classifications. A cut of signal to noise ratio greater than 0.12 avoids detections with too much noise, and a size cut of less than 0.12 arcseconds is used to mask out any extended objects. Two final color cuts of F277W-F356W > 0.53 and F277W-F410W > 0.84 avoid slopes potentially boosted by strong line emission.

To apply the criteria to each object we plot the spectral energy distribution

(SEDs) of that object, plotting the rest wavelengths of five or six filters based on the object's redshift bin, as listed in Table 1. Using LMFit, a tool for non-linear least-squares minimization and curve fitting, to fit a line on the ultraviolet rest wavelengths and a line on the optical rest wavelengths. LRDs typically have an increasing red continuum caused by dust obscuring the AGN, and a steady blue continuum creating a 'v-shape' with the ultraviolet and optical slopes. After plotting the SEDs of all GOODS-South objects, testing them against the criteria of our classification, we obtain a sample of 44 LRDs in GOODS-South.

Next, we trained the SOMs on colors alone. We used five filters to create colors that matched the color cuts in our LRD classification, in order: F277W, F444W, F115W, F200W, F277W, F356W. Applying the JADES GOODS-South data to the SOM produces a two-dimensional map of photometric redshift, each cell representing the mean photometric redshift of the objects mapped to that individual cell. Each cell of the SOM acts as a unique classification, placing objects of similar properties in the same cell. In this case, objects of similar redshift are found in the same and adjacent cells.

The sample of GOODS-South LRDs are applied to the SOM, showing which cells those objects were applied to. Consistently, the SOMs placed these objects in the same or adjacent cells, successfully grouping them together based solely on color. Next, we apply models of brown dwarfs created with our filters for temperatures ranging 500-1600 Kelvin. By doing so we can see where the SOM places the cool brown dwarf models in relation to the LRDs. When placed in similar cells to the LRDs, it means objects in the same bins as the LRDs share properties of brown dwarfs and could have brown dwarf contamination in the sample. When the brown dwarfs are separated, an arbitrary distance of 10 pixels is tested for, there is no brown dwarf

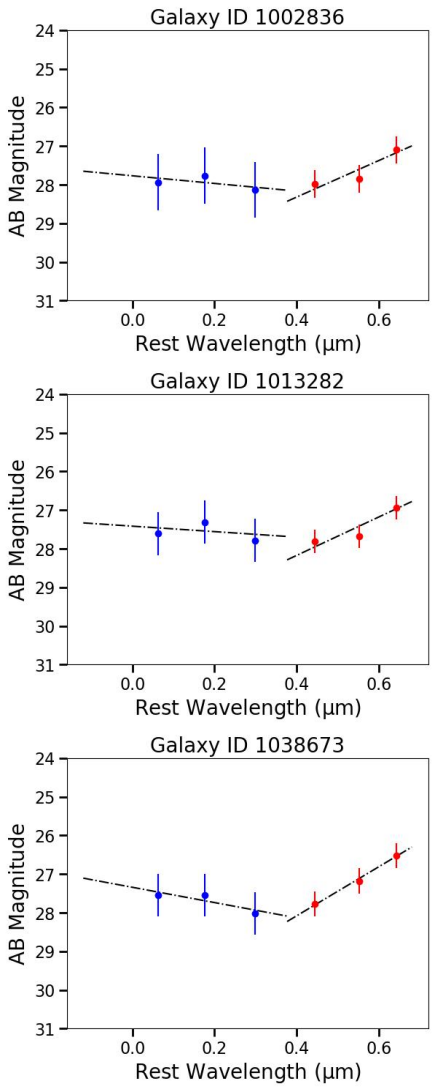

Figure 1: Three spectra energy distribution graphs of GOODS-North objects. β_{UV} line fit to rest wavelength ultraviolet, plotted in blue; β_{opt} line fit to rest wavelength optical, plotted in red.

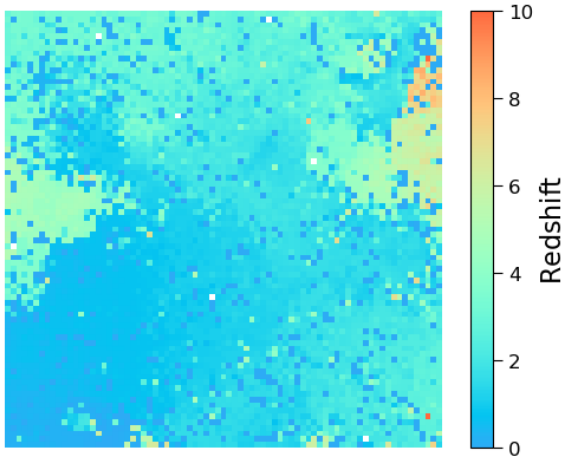

Figure 2: 2D map of mean photometric redshift produced by SOMs trained on GOODS-South data on a 77x77 grid

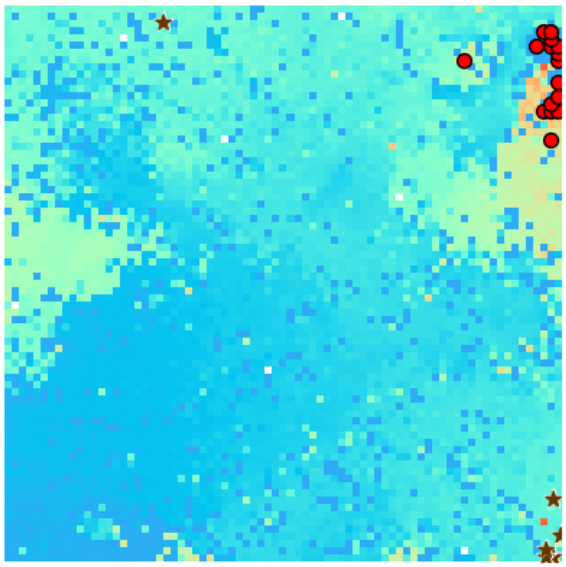

Figure 3: 2D map of mean photometric redshift from figure 2, with LRD sample applied, represented as red dots, and brown dwarf models represented by brown stars.

contamination in the LRD cells.

To test the consistacy of the SOM in seperating LRDs and brown dwarfs, we perturbed the dimensions of the SOM, which generates a new map for each dimension combination. Adjusting each axes between units of 70 to 100, we train SOMs with 900 unique dimensions creating a new map for each. From this, 383 of the maps had a minimum of 10 pixels of seperation between the LRD cells and brown dwarf cells, enough distance to confidently determine there is no overlap between the two.

Applying the GOODS-North data to the 383 maps with brown dwarf separation, the SOM plotted 120 unique objects in the same cells as the GOODS-South LRDs. Testing the GOODS-North objects to our criteria produces a sample of 34 potential LRDs in GOODS-North.

4. Results

The initial result of our research was that the SOM was able to consistently organize similar objects when trained only on colors, grouping LRDs from our sample file into a group of adjacent cells. This remained true after perturbing the axes of the SOMs over 900 maps with unique dimensions, proving the SOM can reliably organize the catalog of objects and their respective properties. Making brown dwarf models with the filters we used to train the SOMs, the SOM mapped the cells that contained objects similar to the models. After checking each unique map through all 900 iterations we find 383 maps, over a third, with a separation between brown dwarfs and LRDs of ten pixels or more. The distance of ten pixels is arbitrary, but large enough to give confidence there is no brown dwarf contamination in the LRD cells.

From the 383 maps of unique dimensions with a brown dwarf separation of at least ten pixels, 120 objects from GOODS-North were mapped to the same cells as the sample LRDs from GOODS-South. After applying the Kocevski et al. (2024) criteria to this list, we have a list of 34 objects that satisfy the conditions as potential LRDs.

Six of our GOODS-North LRDs had spectra from the NIRSpec multi-object spectragraph of which two spectra had the Hα line in frame. The 1D spectra of both objects, JADES 1010816 and 1061888 at $z = 6.72$ and $z = 5.97$ respectively, show a broad component of the Hα line, with JADES 1010816 showing a large broad component of the Hβ line as seen in the medium resolution spectra. These broad components are significant in identifying these objects as accreting AGNs, potentially being supermassive black holes in its early stages of growth.

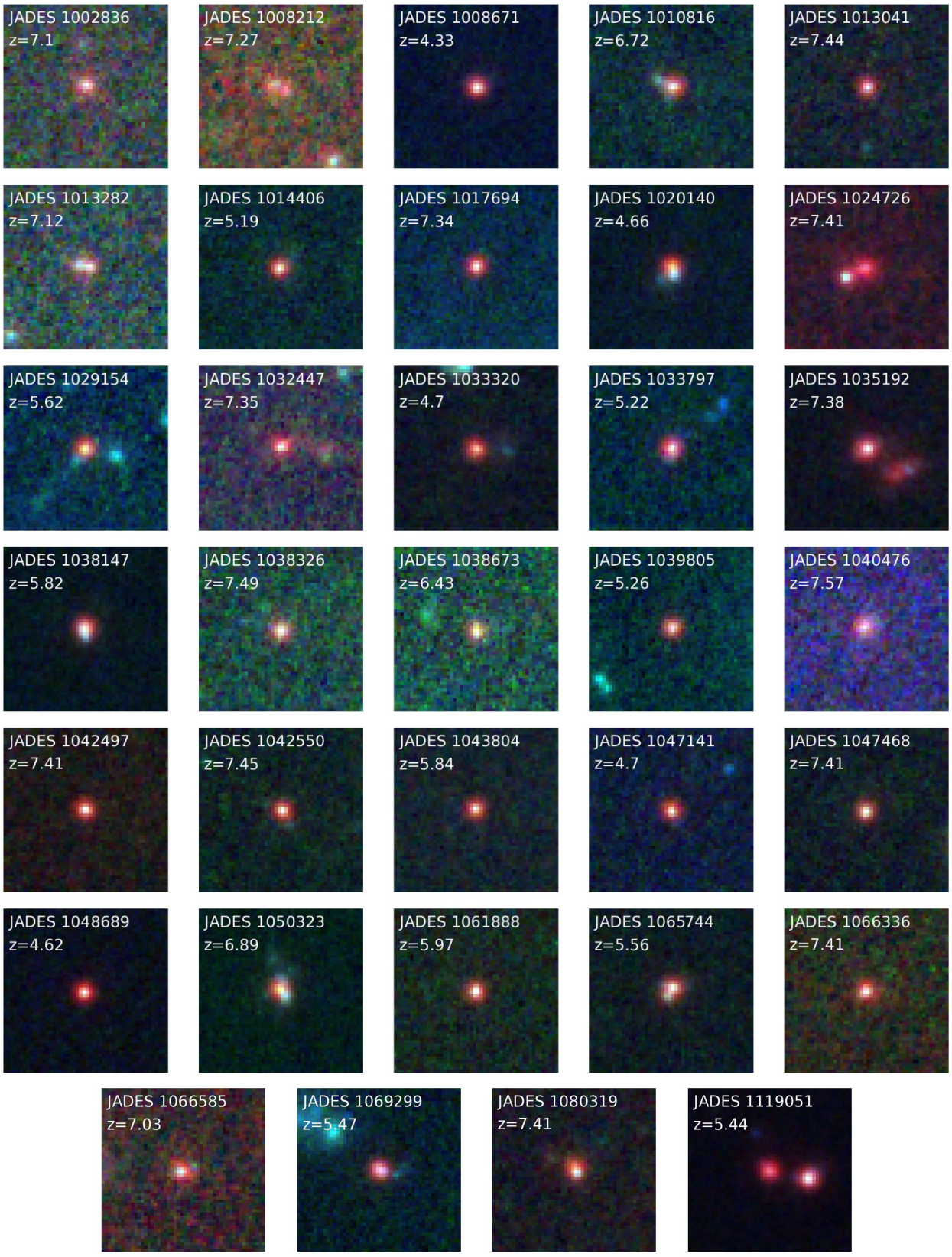

Figure 4: NIRCam images of potential LRDs in GOODS-North, F115W, F200W, F444W filters.

5. Conclusion

In this study we were able to successfully showcase the use of SOMs in the application of identifying objects in a large survey. First, the SOMs were trained only on colors and return a two-dimensional map of mean photometric redshift, mapping objects of similar redshift in the same and adjacent cells. LRDs in our GOODS-South sample file were succesfully grouped in the same or adjacent cells on each unique map. Additionally, SOMs consistantly grouped brown dwarf models in cells with a ten-pixel or greater seperation from the group of LRDs. Applying GOODS-North data to the SOMs with brown dwarf seperation gave us the classifications to create a list of 34 potential LRDs confidently without brown dwarf contamination.

From the low and medium resolution spectra available for two of our objects, we can get some insight about what kind of object it is, whether its an accreting AGN or an early galaxy experiencing large star formation. Because of the nature of the JADES survey and the targeting of the multi-object spectragraph, we have a very limited amount of spectra for these objects, of which only a low- and medium-resolution spectra are available. After identifying these objects as potential LRDs, we can now move forward with future missions to target these objects to record higher-resolution spectra to further study the characteristics of these objects and the supermassive black holes at their centers.

JADES ID	R.A	Dec.	z
1002836	189.128	62.233	7.10
1008212	189.031	62.249	7.27
1008671	189.162	62.251	4.33
1010816	189.152	62.260	6.72
1013041	189.204	62.268	7.44
1013282	189.038	62.269	7.12
1014406	189.072	62.273	5.19
1017694	189.061	62.284	7.34
1020140	188.988	62.291	4.66
1024726	189.084	62.222	7.41
1029154	189.159	62.260	5.62
1032447	189.180	62.282	7.35
1033320	189.126	62.287	4.70
1033797	189.087	62.291	5.22
1035192	189.158	62.302	7.38
1038147	189.271	62.148	5.82
1038326	189.243	62.149	7.49
1038673	189.206	62.151	6.43
1039805	189.244	62.155	5.26
1040476	189.318	62.157	7.57
1042497	189.308	62.163	7.41
1042550	189.322	62.163	7.45
1043804	189.274	62.166	5.84
1047141	189.270	62.176	4.70
1047468	189.198	62.177	7.41
1048689	189.275	62.180	4.62
1050323	189.339	62.185	6.89
1061888	189.168	62.217	5.97
1065744	189.261	62.232	5.56
1066336	189.259	62.235	7.41
1066585	189.259	62.237	7.03
1069299	189.279	62.150	5.47
1080319	189.262	62.212	7.41
1119051	189.214	62.149	5.44

Table 2: GOODS-North objects that comply with Kocevski criteria for LRDs

References

G. Barro, P. G. Pérez-González, D. D. Kocevski, E. J. McGrath, J. R. Trump, R. C. Simons, R. S. Somerville, L. Y. A. Yung, P. Arrabal Haro, H. B. Akins, M. B. Bagley, N. J. Cleri, L. Costantin, K. Davis, M. Dickinson, S. L. Finkelstein, M. Giavalisco, C. Gómez-Guijarro, N. P. Hathi, M. Hirschmann, B. W. Holwerda, M. Huertas-Company, J. S. Kartaltepe, A. M. Koekemoer, R. A. Lucas, C. Papovich, N. Pirzkal, L.-M. Seillé, S. Tacchella, S. Wuyts, S. M. Wilkins, A. de la Vega, G. Yang, and J. A. Zavala. Extremely Red Galaxies at z = 5–9 with MIRI and NIRSpec: Dusty Galaxies or Obscured Active Galactic Nuclei? , 963 (2):128, Mar. 2024. doi: 10.3847/1538-4357/ad167e.

A. J. Bunker, A. J. Cameron, E. Curtis-Lake, P. Jakobsen, S. Carniani, M. Curti, J. Witstok, R. Maiolino, F. D'Eugenio, T. J. Looser, C. Willott, N. Bonaventura, K. Hainline, H. Uebler, C. N. A. Willmer, A. Saxena, R. Smit, S. Alberts, S. Arribas, W. M. Baker, S. Baum, R. Bhatawdekar, R. A. A. Bowler, K. Boyett, S. Charlot, Z. Chen, J. Chevallard, C. Circosta, C. DeCoursey, A. de Graaff, E. Egami, D. J. Eisenstein, R. Endsley, P. Ferruit, G. Giardino, R. Hausen, J. M. Helton, R. E. Hviding, Z. Ji, B. D. Johnson, G. C. Jones, N. Kumari, I. Laseter, N. Luetzgendorf, M. V. Maseda, E. Nelson, E. Parlanti, M. Perna, B. J. Rauscher, T. Rawle, H.-W. Rix, M. Rieke, B. Robertson, B. Rodriguez Del Pino, L. Sandles, J. Scholtz, K. Sharpe, M. Skarbinski, D. P. Stark, F. Sun, S. Tacchella, M. W. Topping, N. C. Villanueva, I. E. B. Wallace, C. C. Williams, and C. Woodrum. JADES NIRSpec Initial Data Release for the Hubble Ultra Deep Field: Redshifts and Line Fluxes of Distant Galaxies from the Deepest JWST Cycle 1 NIRSpec Multi-Object Spectroscopy. *arXiv e-prints*, art. arXiv:2306.02467, June 2023. doi: 10.48550/arXiv.2306.02467.

F. D'Eugenio, A. J. Cameron, J. Scholtz, S. Carniani, C. J. Willott, E. Curtis-Lake, A. J. Bunker, E. Parlanti, R. Maiolino, C. N. A. Willmer, P. Jakobsen, B. E. Robertson, B. D. Johnson, S. Tacchella, P. A. Cargile, T. Rawle, S. Arribas, J. Chevallard, M. Curti, E. Egami, D. J. Eisenstein, N. Kumari, T. J. Looser, M. J. Rieke, B. Rodríguez Del Pino, A. Saxena, H. Übler, G. Venturi, J. Witstok, W. M. Baker, R. Bhatawdekar, N. Bonaventura, K. Boyett, S. Charlot, A. L. Danhaive, K. N. Hainline, R. Hausen, J. M. Helton, X. Ji, Z. Ji, G. C. Jones, I. Joudžbalis, M. V. Maseda, P. G. Pérez-González, M. Perna, D. Puskás, I. Shivaei, M. S. Silcock, C. Simmonds, R. Smit, F. Sun, N. C. Villanueva, C. C. Williams, and Y. Zhu. JADES Data Release 3 – NIRSpec/MSA spectroscopy for 4,000 galaxies in the GOODS fields. *arXiv e-prints*, art. arXiv:2404.06531, Apr. 2024. doi: 10.48550/arXiv.2404.06531.

D. J. Eisenstein, C. Willott, S. Alberts, S. Arribas, N. Bonaventura, A. J. Bunker, A. J. Cameron, S. Carniani, S. Charlot, E. Curtis-Lake, F. D'Eugenio, R. Endsley, P. Ferruit, G. Giardino, K. Hainline, R. Hausen, P. Jakobsen, B. D. Johnson, R. Maiolino, M. Rieke, G. Rieke, H.-W. Rix, B. Robertson, D. P. Stark, S. Tacchella, C. C. Williams, C. N. A. Willmer, W. M. Baker, S. Baum, R. Bhatawdekar, K. Boyett, Z. Chen, J. Chevallard, C. Circosta, M. Curti, A. L. Danhaive, C. DeCoursey, A. de Graaff, A. Dressler, E. Egami, J. M. Helton, R. E. Hviding, Z. Ji, G. C. Jones, N. Kumari, N. Lützgendorf, I. Laseter, T. J. Looser, J. Lyu, M. V. Maseda, E. Nelson, E. Parlanti, M. Perna, D. Puskás, T. Rawle, B. Rodríguez Del Pino, L. Sandles, A. Saxena, J. Scholtz, K. Sharpe, I. Shivaei, M. S. Silcock, C. Simmonds, M. Skarbinski, R. Smit, M. Stone, K. A. Suess, F. Sun, M. Tang, M. W. Topping, H. Übler, N. C. Villanueva, I. E. B. Wallace, L. Whitler, J. Witstok, and C. Woodrum. Overview of the JWST Advanced Deep Extragalactic Survey (JADES). *arXiv e-prints*, art. arXiv:2306.02465, June 2023. doi: 10.48550/arXiv.2306.02465.

K. N. Hainline, B. D. Johnson, B. Robertson, S. Tacchella, J. M. Helton, F. Sun, D. J. Eisenstein, C. Simmonds, M. W. Topping, L. Whitler, C. N. A. Willmer, M. Rieke, K. A. Suess, R. E. Hviding, A. J. Cameron, S. Alberts, W. M. Baker, S. Baum, R. Bhatawdekar, N. Bonaventura, K. Boyett, A. J. Bunker, S. Carniani, S. Charlot, J. Chevallard, Z. Chen, M. Curti, E. Curtis-Lake, F. D'Eugenio, E. Egami, R. Endsley, R. Hausen, Z. Ji, T. J. Looser, J. Lyu, R. Maiolino, E. Nelson, D. Puskás, T. Rawle, L. Sandles, A. Saxena, R. Smit, D. P. Stark, C. C. Williams, C. Willott, and J. Witstok. The Cosmos in Its Infancy: JADES Galaxy Candidates at z ¿ 8 in GOODS-S and GOODS-N. , 964(1):71, Mar. 2024. doi: 10.3847/1538-4357/ad1ee4.

D. D. Kocevski, M. Onoue, K. Inayoshi, J. R. Trump, P. Arrabal Haro, A. Grazian, M. Dickinson, S. L. Finkelstein, J. S. Kartaltepe, M. Hirschmann, J. Aird, B. W. Holwerda, S. Fujimoto, S. Juneau, R. O. Amorín, B. E. Backhaus, M. B. Bagley, G. Barro, E. F. Bell, L. Bisigello, A. Calabrò, N. J. Cleri, M. C. Cooper, X. Ding, N. A. Grogin, L. C. Ho, T. A. Hutchison, A. K. Inoue, L. Jiang, B. Jones, A. M. Koekemoer, W. Li, Z. Li, E. J. McGrath, J. Molina, C. Papovich, P. G. Pérez-González, N. Pirzkal, S. M. Wilkins, G. Yang, and L. Y. A. Yung. Hidden Little Monsters: Spectroscopic Identification of Low-mass, Broad-line AGNs at z ¿ 5 with CEERS. , 954(1):L4, Sept. 2023. doi: 10.3847/2041-8213/ace5a0.

D. D. Kocevski, S. L. Finkelstein, G. Barro, A. J. Taylor, A. Calabrò, B. Laloux, J. Buchner, J. R. Trump, G. C. K. Leung, G. Yang, M. Dickinson, P. G. Pérez-González, F. Pacucci, K. Inayoshi, R. S. Somerville, E. J. McGrath, H. B. Akins, M. B. Bagley, L. Bisigello, R. A. A. Bowler, A. Carnall, C. M. Casey, Y. Cheng, N. J. Cleri, L. Costantin, F. Cullen, K. Davis, C. T. Donnan, J. S. Dunlop, R. S. Ellis, H. C.

Ferguson, S. Fujimoto, A. Fontana, M. Giavalisco, A. Grazian, N. A. Grogin, N. P. Hathi, M. Hirschmann, M. Huertas-Company, B. W. Holwerda, G. Illingworth, S. Juneau, J. S. Kartaltepe, A. M. Koekemoer, W. Li, R. A. Lucas, D. Magee, C. Mason, D. J. McLeod, R. J. McLure, L. Napolitano, C. Papovich, N. Pirzkal, G. Rodighiero, P. Santini, S. M. Wilkins, and L. Y. A. Yung. The rise of faint, red agn at $z > 4$: A sample of little red dots in the jwst extragalactic legacy fields. *arXiv e-prints*, art. arXiv:2404.03576, Apr. 2024. doi: 10.48550/arXiv.2404.03576.

V. Kokorev, K. I. Caputi, J. E. Greene, P. Dayal, M. Trebitsch, S. E. Cutler, S. Fujimoto, I. Labbé, T. B. Miller, E. Iani, R. Navarro-Carrera, and P. Rinaldi. A Census of Photometrically Selected Little Red Dots at 4 ¡ z ¡ 9 in JWST Blank Fields. , 968(1):38, June 2024. doi: 10.3847/1538-4357/ad4265.

J. Matthee, R. P. Naidu, G. Brammer, J. Chisholm, A.-C. Eilers, A. Goulding, J. Greene, D. Kashino, I. Labbe, S. J. Lilly, R. Mackenzie, P. A. Oesch, A. Weibel, S. Wuyts, M. Xiao, R. Bordoloi, R. Bouwens, P. van Dokkum, G. Illingworth, I. Kramarenko, M. V. Maseda, C. Mason, R. A. Meyer, E. J. Nelson, N. A. Reddy, I. Shivaei, R. A. Simcoe, and M. Yue. Little Red Dots: An Abundant Population of Faint Active Galactic Nuclei at z $\sim$ 5 Revealed by the EIGER and FRESCO JWST Surveys. , 963(2):129, Mar. 2024. doi: 10.3847/1538-4357/ad2345.