Rijksuniversiteit Groningen

Bachelor's Thesis

Quantifying AGN power fraction in JWST NIRCam imaging data

Author:
N. Pandolfo

Supervisors:
Dr. L. Wang
Dr. B. Margalef-Bentabol

CONTENTS

ABSTRACT

Accurately disentangling host-galaxy starlight from active galactic nucleus emission is essential for tracing black-hole growth in the James Webb Space Telescope COSMOS-Web survey. Whereas the pipeline of Margalef-Bentabol et al. (2025) modelled active galactic nuclei power fractions ($f_{\rm AGN}$) in the single Near-Infrared Cam band F150W, this thesis extends the approach to its remaining channels used in the COSMOS-Web mission (F115W, F277W and F444W). Blank-sky cutouts were combined with simulated galaxies from the Illustris-TNG project, injected with random point-source fluxes ($f_{\rm AGN} = [0 - 0.95]$), and convolved with the telescope's point-spread functions. A *ConvNeXt-Base* network in `Zoobot` was trained separately in F277W and F444W (F115W training proved unstable and is reserved for future work). On held-out simulations the model attains RMSE= 0.028 and RAE= 0.755 in F277W; performance is poorer in F444W, consistent with its broader point-spread function and higher thermal background. Both filters' performance proved to be slightly worse than the F150W filter in Margalef-Bentabol et al. (2025). Applied to real cutouts, the network recovers 3045 active galactic nuclei candidates above a 5σ threshold in F277W and 2110 in F444W. Merging these fractions with Spitzer/IRAC mid-infrared colour cuts and a three-colour logistic classifier yields 515 median-selected ($median(f_{AGN}) > 0.2$), 63 MIR-selected, and 103 machine-learning-selected active galactic nuclei, with only limited overlap. The divergence reflects the complementary nature of geometry-dominated photometric and ML tracers versus dust-temperature-dominated MIR criteria. Overall, a compact convolutional network recovers active galactic nuclei contributions up to ≈ 0.03 absolute accuracy. Incorporating a stable F115W model and mitigating the strong class imbalance between non-active and active galactic nuclei, are the next steps toward a fully four-band framework for forthcoming deep James Webb surveys.

1 INTRODUCTION

Supermassive black holes (SMBHs), depicted in Figure 1, have been a point of significant interest for astronomers of the current and last century. They are present in galaxies with visible bulges, where their size is directly correlated to the size of the black hole (Kormendy and Bender, 2011). In particular, they can most often be found at the centre of a galaxy, as the dense molecular gas assembles better there. The gravitational interactions between the particles of the cloud and the surrounding medium direct a substantial amount of material towards the centre (Chen, 2025). When a SMBH ($\gtrsim 10^6 M_\odot$) accretes enough gas (primarily made of hydrogen and helium) and dust, the infalling matter deepens its gravitational well. It ignites the system, transforming it into an active galactic nucleus (AGN) (Kelly et al., 2024).

Figure 1: Artist's rendition of a SMBH, massive objects located most often at the centre of galaxies. Image credit to the NASA/JPL-Caltech, 2012.

1.1 History and structure of AGNs

At the start of the 20th century, scientists like Edward A. Fath, V. M. Slipher and Edwin Hubble had observed distinct emission lines (Hβ, [O II]λ3727, [NeIII]λ3869, [OIII]λ4363, λ4959, λ5007) which were characteristic to gaseous nebulae (Osterbrock and Ferland, 2006). Carl K. Seyfert made the turning point with his publication "*Nuclear emission in spiral nebulae*" (1943). Now defined as Seyfert galaxies, the astronomer discovered the most common type of AGNs. In particular, Type I Seyfert galaxies present broad emission lines, with a FWHM $\gtrsim$ 1000 km/s, and their emissions are concentrated in the nuclear region. On the contrary, Type II Seyfert galaxies spectra present forbidden lines as well, and a FWHM$\approx$ 500 km/s (Binney and Merrifield, 1998). Since the discovery of these galaxies, other distinctions have been made. For example, with the advent of radio telescopes, more AGNs were discovered, therefore creating the necessity to divide them into radio-loud or radio-quiet AGNs. Some of the optically observed radio sources appeared as unresolved point sources instead of elliptical galaxies. These were identified as quasi-stellar radio sources, also known as quasars, which exhibit highly redshifted spectra (Binney and Merrifield, 1998). Because all these classes trace the same accretion phenomenon, no morphological or spectroscopic subdivision (e.g. Seyfert I vs II, radio-loud vs radio-quiet) is pursued in the

present work. The only additional category introduced later is the mid-infrared (MIR) selected AGN label, assigned to sources that satisfy a MIR colour criterion and intended purely as a selection flag rather than a distinct physical subtype (see Section 3.5).

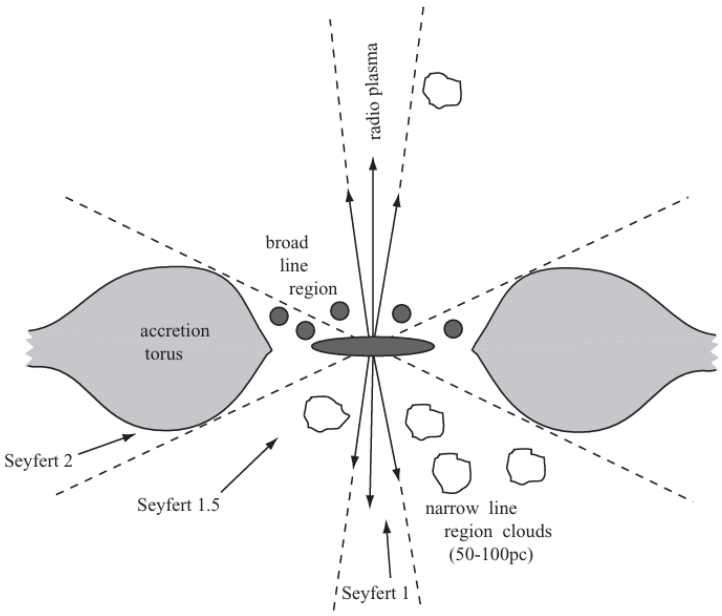

Figure 2: Schematic "unified" AGN: a black-hole/accretion-disk system (centre) is surrounded by a dusty torus. Jets emerge perpendicular to the disk and illuminate narrow-line clouds on 50-100 pc scales. Image taken from Sparke and Gallagher (2010).

Figure 2 represents the structure of an AGN according to a unified model. The black hole is surrounded by an accretion disk that contributes to the energy input in the surrounding gas through photoionisation (Osterbrock, 1993). The resulting dense clouds cause the broad-line emission that can be observed in the spectra of Type I galaxies. This region is surrounded by a dusty torus that blocks the broad-line region from certain fields of view (Urry and Padovani, 1995). Jets emerge in seemingly straight lines from the nucleus, at approximately light speed, although initially relativistic. This is also the reason why, when observing directly down through the jet, we would see a Seyfert Type 1 nucleus, and when looking through the dusty disk, a Seyfert Type 2 galaxy (Binney and Merrifield, 1998; Sparke and Gallagher, 2010).

1.2 AGNs observations

AGNs are very luminous throughout the whole electromagnetic spectrum, as different parts emit at various frequencies. The jets emit at gamma and radio frequencies, the SMBH and accretion disk in the X-Ray and UV, respectively, and the dust and obscuring mass are best visible in the infrared (Kelly et al., 2024). They are therefore easily observable across various types of telescopes, but the presence of dust and extreme surface-brightness contrast complicate matters. The bright point source can dominate the galaxy's light profile, so any attempt to measure bulge mass, star-formation rate or the galaxy's morphology must take into account the effect of the AGN. This requires both high spatial resolution to model the point-spread function (PSF) and long-wavelength sensitivity to see through the torus that blocks the ultraviolet and optical bands. Correct modelling of the PSF is essential to properly measure properties of galaxies, as

it measures how a point source spreads out across the detector of a telescope (Zhuang, Li, and Shen, 2024).

The Chandra X-Ray Observatory, with its good capabilities in mapping hot accretion, has been used to study the co-evolution between SMBHs and host galaxies (Marchesi et al., 2016). It cannot, however, measure rest-frame optical emission lines or starlight, so it provides little leverage on host-galaxy masses or morphologies. The Hubble Space Telescope has an optimal $0.05''$ resolution, measures in the UV and optical range, and is therefore able to identify AGN variability. Still, its optical coverage at $z \approx 2$ in the rest frame is not able to properly analyse structural properties of host galaxies, as the main emission lines fall out of the observed range (Zhuang, Li, and Shen, 2024). Lastly, for the Spitzer Space Telescope, although working in the MIR, the data collected has low spatial resolution and a broad PSF, so the nucleus is more blended with the bulge (González-Martín et al., 2015; Spitzer Science Center, 2008).

The James Webb Space Telescope (JWST, shown in Figure 3), launched in 2021, presents itself as one of the current best options to observe AGNs. With its Near-Infrared Camera (NIRCam) and Mid-Infrared Instrument (MIRI), it allows AGNs to be studied through heavy extinction because near- and mid-infrared photons experience far less attenuation and capture the dust-re-emitted light. More so, JWST unifies the strong characteristics of the previously mentioned telescopes, without presenting the same drawbacks. It has a resolution of around $0.07''$ at 2 μm, only slightly broader than Hubble's $0.05''$ in the optical and $0.15''$across the near-infrared, all thanks to its 6.5-meter primary mirror. The full wavelength range of the telescope goes from 0.6 to 28.5 μm, surpassing Spitzer's, delivering sharper, deeper views of rest-frame optical light from low-redshift ($z \leq 1$) galaxies, while simultaneously catching the red-shifted Balmer lines and dust emission from galaxies and quasars well beyond $z \geq 6$ (Rigby et al., 2023).

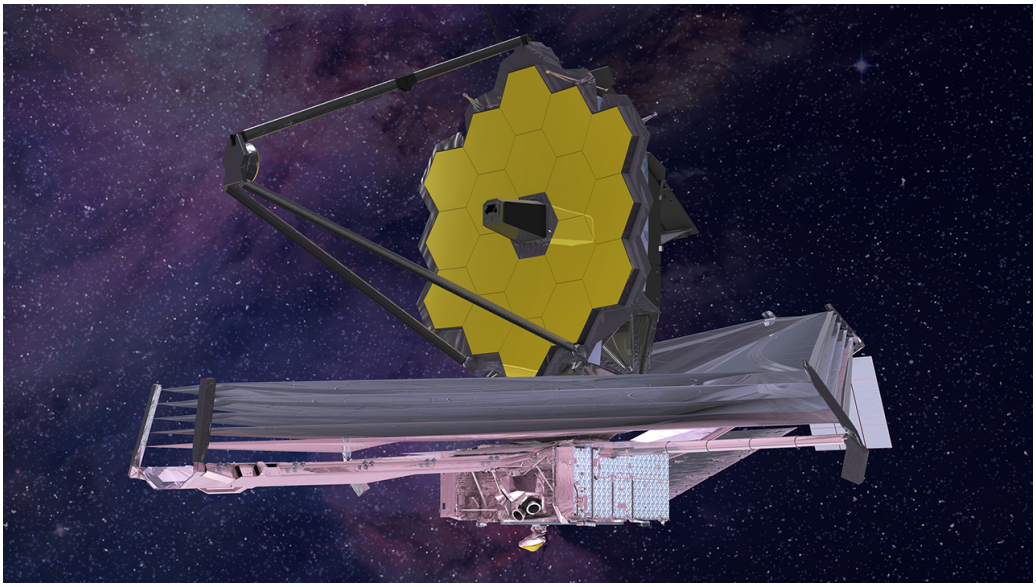

Figure 3: The James Webb Space Telescope viewed from its port (left) side. JWST displays its five-layer oblong-hexagonal sunshield with the square spacecraft bus at the center, a rear momentum flap, and a lateral solar array. Above the shield rises the gold-coated 18-segment primary mirror with its central aft-optics cone and three gray struts holding the secondary mirror out in front. All instruments, including the NIRCam, are stored behind the primary mirror. Image credit: NASA/ESA/CSA/STScI, (2022).

Recent large programmes, such as COSMOS-Web (Casey et al., 2023), confirm that JWST's high sensitivity has led to the discovery of fainter objects compared to predecessors and is tracing clear correlations between MIR and X-Ray luminosities that strengthen the case for MIR selection (Chien et al., 2024; Zhuang, Li, and Shen, 2024).

Achieving these objectives requires accurate host-galaxy photometry even when a luminous AGN dominates the central pixels. COSMOS-Web's depth and resolution provide the ideal dataset to test different AGN-host decomposition methods.

1.3 Colour-colour AGN selection

Photometric colour-colour fitting offers a profile-free complement to structural fitting. This has been done in the past across multiple wavelength ranges, the most common being X-Ray and MIR. X-Ray AGN selection has the downside of not being able to select the heavily obscured AGNs, but these can be recovered through MIR selection, which is why they are considered complementary (Chang et al., 2017; Kelly et al., 2024).

Early work with Spitzer/IRAC defined wedges in the [3.6]-[4.5] versus [5.8]-[8.0] plane, where the brackets denote the filter wavelengths in μm. Stern et al. (2005) and Lacy et al. (2004, 2007) produced broad, highly complete wedges that suffer substantial contamination from high-redshift galaxies, a problem mitigated by the stricter polygon introduced by Donley et al. (2012). The four WISE bands at 3.4, 4.6, 12 and 22 μm have likewise generated several criteria (e.g. Assef et al., 2013; Stern et al., 2012). All of these formulas are tabulated in Table 5 of Padovani et al. (2017).

This thesis uses the formulas introduced by Chang et al. (2017), which were adopted from Lacy et al. (2004, 2007). The numerical limits, adjusted to tighten the selection at faint fluxes while preserving high completeness, are listed in Section 3.5.

The absence of a near-infrared AGN criterion underscores the need to develop JWST-specific colour formulas. Because the current NIRCam filter set lacks channels beyond 5 μm, colour cuts alone cannot exhaustively find every AGN candidate. The shortcomings motivate the image-based methods discussed next.

1.4 Host-AGN decomposition techniques

Most techniques previously used are based on parametric image fitting. This method tries to fit the light distribution of a galaxy with parametric functions. This can usually be carried out either as a one-dimensional fitting of the surface-brightness profiles, or by fitting galaxies' cutouts in two dimensions (Peng et al., 2002). Within this framework, one of the most used two-dimensional algorithms is the package `GALFIT`, which minimises the pixel-by-pixel χ^2 with a Levenberg-Marquardt gradient-descent optimiser (Peng et al., 2002, 2010). As described in the user's manual, this works by fitting two-dimensional functions, usually in ellipsoid shapes, to galaxy and point sources of the cutouts. This method can be used with a different range of functions, such as Sérsic and de Vaucouleurs and the PSF function provided by the user (Peng, 2010). A similar predecessor is `GIM2D`, designed to work with batches of galaxies. It uses a Metropolis algorithm, which is based on Monte-Carlo sampling of the likelihood function (Simard, 1998).

Both of these algorithms have some downsides: `GALFIT` is not equipped to analyse a batch of data, and `GIM2D` can only fit two functions at a time and struggles in separating the light of different sources. There are also weaknesses belonging to parametric methods generally. Firstly,

the difficulty in generalising, as one model may not fit on all galaxies and a fit that worked on one cutout may not work on others. Secondly, parameter degeneracy, where two or more different sets of model parameters produce statistically indistinguishable light profiles, so the fitting routine cannot tell which combination is correct. Lastly, `GALFIT` (and similarly `GIM2D`) struggles with JWST data, as its performance downgrades as the PSF rises in complexity. A slight PSF mismatch leaves residual nuclear light that `GALFIT` "absorbs" by inflating a high-Sérsic-index bulge, so the host flux and concentration are systematically over-estimated, and this bias is even more present when the AGN dominates in the nucleus (Margalef-Bentabol et al., 2025; Zhuang and Shen, 2023).

One response to these limitations is to abandon analytic profiles altogether and measure a galaxy's light distribution without making strong assumptions about its shape. Non-parametric techniques estimate the total brightness and structure of a galaxy using methods that do not depend on a well-distinguished structure. For instance, a curve of growth simply sums the flux in concentric apertures until the cumulative light converges (Kron, 1980). On the other hand, the M_{20} statistic orders all pixels by brightness and computes the second-order moment of just the brightest 20 %, capturing how centrally concentrated a system is (Lotz et al., 2004). Once again, both metrics are not perfect: curves of growth ignore the PSF, so the unresolved nuclear light is blended into the host flux, making this model signal-to-noise-ratio (S/N) dependent. The M_{20} method, conversely, is highly sensitive to noise spikes and degrades quickly when an AGN core dominates the brightest pixels.

These caveats point to the need for a PSF-aware approach that is both flexible across diverse morphologies and fully replicable. This can be executed with machine learning (ML) techniques.

1.5 Deep learning in AGN research

Deep learning is a branch of ML that represents data hierarchically: it is built on layers that learn a description that's more abstract than what was learned in the one below (Bengio, 2009). This is typically done through non-linear transformations, whose parameters are optimised by training the model with a dataset (Dieleman et al., 2015).

In the last five years, deep learning has entered the realm of astronomy through the study of galaxy morphology (e.g. Barchi et al., 2020; Cheng et al., 2021; Dieleman et al., 2015; Domínguez Sánchez et al., 2018; Huertas-Company et al., 2015). More recently, this has expanded in the AGN research field: Tian et al. (2025) created a compositive ML framework to investigate AGN host morphologies and remove the AGN point source, and Margalef-Bentabol et al. (2025) created a convolutional neural network to identify the fraction of AGN in a galaxy.

Convolutional Neural Networks (ConvNets) are one of the first and more successful deep-learning techniques developed (Bengio, 2009). Its conceptual ancestor is Fukushima's Neocognitron (Fukushima, 1980), an unsupervised, layered network inspired by the hierarchical receptive fields that Hubel and Wiesel (1962) observed in the visual cortex. Modern ConvNets (shown in Figure 4) apply small, shared filters across the image, followed by pooling operations; stacking many such layers yields a hierarchy in which higher layers encode increasingly global features.

Since the ConvNets have a restricted connectivity pattern, their layers have fewer parameters that compute the transformation of the input, which leads to an improvement in generalisation and allows the model to be easily scaled to unseen data (Dieleman et al., 2015). Crucially, that same property lets them absorb PSF variation, noise patterns, and complex host morphologies, precisely the regimes where parametric and non-parametric fits struggle.

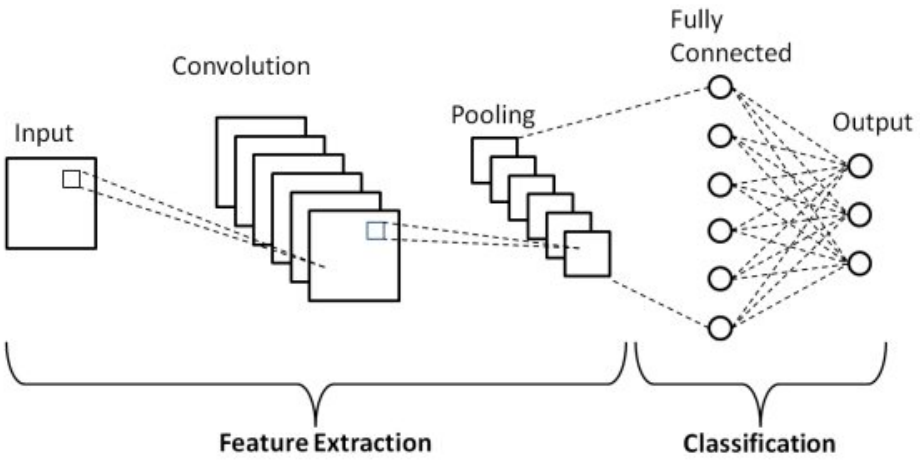

Figure 4: Schematic of a convolutional neural-network workflow: an input image passes through stacked convolution and pooling layers that extract hierarchical features, which are then flattened and fed to fully connected layers for final classification. Image credit: Hiep and Joo, 2018.

1.6 Previous work and thesis aim

The present work is based on Margalef-Bentabol et al. (2025), and is intended as an expansion of their paper. The team took into account the issues with PSF mismatch and irregular host morphologies and decided to investigate host-AGN decomposition using JWST data in the F150W NIRCam filter and a deep learning model. Their method uses simulated data from hydrodynamical simulations, injected with a random PSF and AGN fraction, to train a ConvNet algorithm integrated in the package `Zoobot` (Walmsley, Allen, et al., 2023). `Zoobot` has already powered several large-scale morphology projects, including the 0.3-M galaxy Galaxy Zoo DECaLS catalogue (Walmsley, Willett, et al., 2022) and the 8.7-M galaxy Galaxy Zoo DESI release (Walmsley, Masters, et al., 2023), demonstrating its reliability on heterogeneous survey data.

It was demonstrated that Margalef-Bentabol et al. (2025) model recovers the injected AGN fraction with negligible bias (-0.002) and an RMSE of 0.013, outperforming `GALFIT` by nearly an order of magnitude in both scatter and outlier rate. The performance remained stable across redshift, S/N, and galaxy size, whereas `GALFIT` is slower and less reliable for disturbed morphologies. Their model was then applied to the real JWST data, and their selection of galaxies containing AGNs was compared to established literature selections in the X-Ray and MIR.

This thesis' objective is to expand the proof of concept of their paper to the other NIRCam filters used in the COSMOS-Web survey (F115W, F277W, F444W) and to exploit the resulting colour distributions to derive JWST-specific AGN-selection formulas, following the idea of Chien et al. (2024). The workflow remains that of Margalef-Bentabol et al. (2025), with a few exceptions: a broader redshift range for the simulated data ($0.50 \leq z \leq 6.01$ vs. $0.50 \leq z \leq 3.01$), a linear (rather than sigmoid) activation in the regression head, and feature normalisation by the dataset's own mean and standard deviation.

This thesis is organised as follows: Sections 2 and 3 present the data and full methodology, Section 4 shows the results, which are discussed in Section 5. Finally, Section 6 offers concluding remarks.

2 DATA

2.1 Real data

Data comes from COSMOS-Web, a 255-hour JWST mission, executed in JWST's first two observing cycles to contribute to the Cosmic Evolution Survey (COSMOS) field. The NIRCam covered 0.54 deg^2 in the filters F115W, F150W, F277W and F444W, while the parallel MIRI F770W channel mapped 0.18 deg^2. The mission originated from three main scientific goals: expand the galaxies database in the range $6 \lesssim z \lesssim 11$ to map the spatial pattern and drivers of cosmic reionization, uncover massive galaxies ($M > 10^{10} M_{\odot}$) at $z > 4$, use weak gravitational lensing to trace the stellar-to-halo mass relation and its dependence on star-formation history and morphology to $z \approx 2.5$ (Casey et al., 2023). Around 800,000 galaxies were measured, and the telescope was able to cover 98% of cosmic history (COSMOS Collaboration, 2025).

NIRCam itself consists of two $2.2' \times 2.2'$ modules, each hosting a short-wavelength (0.6-2.3 μm) channel and a long-wavelength (2.4-5.0 μm) channel. The eight broad filters are F070W, F090W, F115W, F150W, F200W (SW) and F277W, F356W, F444W (LW) (Rieke et al., 2023). For the four COSMOS-Web bands used here, Zhuang, Li, and Shen (2024, Table 1) measure point-spread-function FWHM values of 56.2, 61.1, 116, and 158 mas in F115W, F150W, F277W, and F444W, respectively. Temporal drift remains below 3% in every band, while spatial non-uniformity falls from 7.1% in F115W to 1.5% in F444W. These figures set the effective angular resolution adopted later for PSF convolution of the mock galaxies.

The dataset used in this project has been pre-processed by Zhuang et al. (2024), as well as the construction of models for the point-spread functions (PSF). The data come from 80 observations with the NIRCam of the JWST telescope during the months of January, May, and June 2023, which overall covered 0.28 $\deg^2$. An overview of the position of the processed tiles is presented in Fig. 5. The images were reduced using the `jwst pipeline` (Binney and Merrifield, 1998), with custom steps that included corrections for features arising from the hardware used (such as filamentary features and noise), and background subtraction.

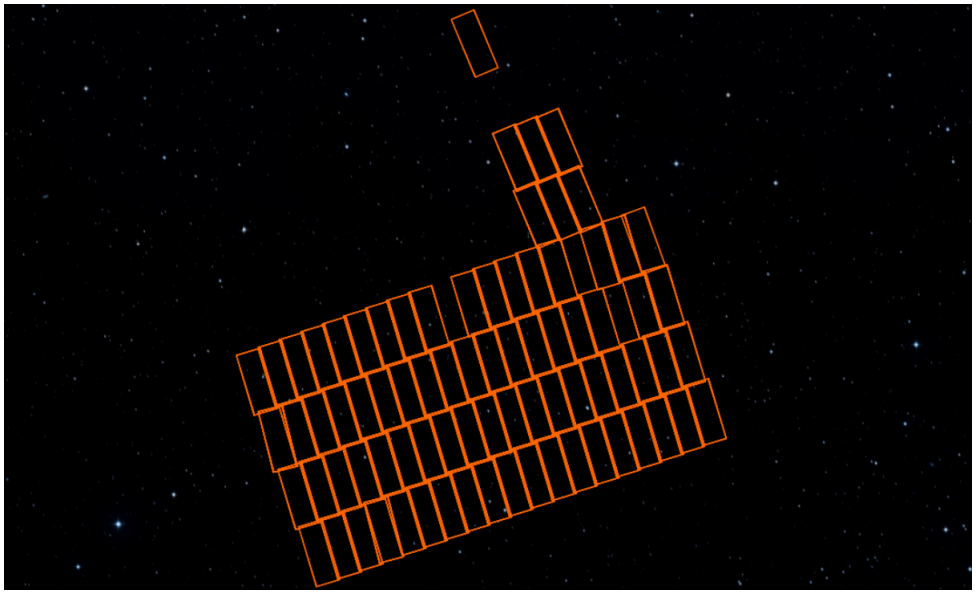

Figure 5: Map of the observation tiles processed by Zhuang et al. (2024). The observations span from a right ascension (RA) of 149.6958 to 150.4083 degrees and a declination (DEC) going from 1.7800 to 2.5880 degrees. The background comes from the Digitized Sky Survey. Image taken with the Astroview feature of the Mikulski Archive for Space Telescopes (2024).

Regarding the PSF model construction, the package `PSFEx` (Bertin, 2011) was selected as it was deemed by past research to be the best performing (Zhuang, Li, and Shen, 2024; Zhuang and Shen, 2023). For each NIRCam mosaic, Zhuang et al. (2024) created three different PSF models, meant to accompany each of the observations that they processed. The *global* model uses all the point-like sources, the *local* one accounts for the PSF spatial variation in a mosaic, and lastly, the *broad/narrow* model takes either the broader or narrower half of the point sources and is divided by the median FWHM. The authors recommend using the global model for scientific research, so this was the preferred choice for Margelef-Bentabol et al. (2025) and this thesis.

A catalogue, containing the selection of galaxies to use, was made by the team in Margalef-Bentabol et al. (2025), containing photometric information from the Farmer catalogue (Cosmos Team, 2022, January) and spectroscopy or photometric redshift value. The galaxy sample included galaxies with redshifts between $0.5 < z < 3$. Stellar masses were calculated by La Marca et al. (2025, in prep.), and were selected those bigger than $109 M_{\odot}$ or those reaching the K_s-based completeness limit described in Section 6.2 of Weaver et al. (2022), depending on which was higher.

2.2 Simulated data

The images used in the creation of the mock frames come from the Illustris-TNG project, a large set of cosmological simulations that model galaxy formation (Nelson, Springel, et al., 2019). They used the AREPO code (Springel, 2010), a moving-mesh hydrodynamics scheme. In particular, the simulation used for this work is the TNG100-1, with dimensions of 100 Mpc^3, a range of redshift from $z = 0$ to $z = 127$ that uses the Planck XIII cosmological parameters (Ade et al., 2016), and contains 18203 dark matter particles with a mass resolution of $m_{DM} = 7.5 \times 10^6 M_{\odot}$ (Nelson, Springel, et al., 2019). The dataset is made up of 99 snapshots, each with decreasing z in the range mentioned. The snapshots used in this work and their corresponding redshift are

presented in Table 1. For more information on the TNG100 simulation, the reader is redirected to the following papers: Marinacci et al., 2018; J. P. Naiman et al., 2018; Nelson, Pillepich, et al., 2018; Pillepich et al., 2018; Springel et al., 2018.

Table 1: Selected snapshots and corresponding redshifts

Snapshot	**Redshift (z)**	**Snapshot**	**Redshift (z)**
13	6.01	46	1.15
17	5.00	49	1.04
21	4.01	52	0.92
25	3.01	55	0.82
29	2.44	58	0.73
33	2.00	61	0.64
37	1.67	64	0.58
40	1.50	67	0.50
43	1.30		

2.3 Infrastructures used

Training to recognise different AGN power fractions was performed with the `Zoobot` package (Walmsley, Allen, et al., 2023). For the data processing, Python packages `pandas` and `astropy` were used (Astropy Collaboration et al., 2022; pandas development team, 2020), while all the visualisations for this work were made with `matplotlib` and `plotly` (Hunter, 2007; Inc., 2015).

Lastly, high-performance computing tasks required for this work were carried out using the Hábrók cluster, provided and supported by the Center for Information Technology (CIT) at the University of Groningen. Several of the catalogues and Python scripts used in this work were provided by the team led by Dr. Wang. Further details and acknowledgements are presented in Section 7.

3 METHODS

This section describes the full methodology used for this work. Blank sky and real galaxy cutouts are extracted, and simulated galaxies are convolved with randomly selected PSFs and injected with random AGN fractions. The *ConvNeXt-Base* model is then trained and validated using the `Zoobot` framework. Colour-colour analysis was performed with the aid of an ML model. A Random Forest ranked colour indices and a logistic regression on the top three colours produced final AGN probabilities above a chosen threshold.

3.1 Blank and real cutouts

The first step of this project was to select specific areas of the JWST dataset in which there were no galaxies or other particularly luminous sources, to serve as realistic background fields for the injection of simulated galaxies. Dr. Wang's team made available a pre-selected list of coordinates as well as an initial script to execute this. Firstly, the script checks that in the selected regions of each observation tile there are not more than a selected number of sources. For each coordinate, the code finds the 4 nearest observation tiles and attempts to extract both science and error cutouts, with shape 180×180 pixels, from the appropriate FITS maps. To improve computational efficiency, the code was modified to support parallel execution, allowing

it to run across 16 CPU cores. This significantly reduced the processing time for evaluating large sets of coordinates.

Similarly, cutouts of real galaxies were created by modifying the script described above. In this case, a different catalogue, containing galaxy positions across the JWST observations, was used, and the images were cut to a size of 128×128 pixels. Instead of searching for the nearest tiles, the code was adapted to extract cutouts centred on the exact coordinates provided. Additional filtering was applied to retain only cutouts containing a single source, while those with multiple detections were excluded. This resulted in 22353 images for the F115W filter, 16014 for F277W, and 17252 for F444W.

3.2 Mock images

Once the cutouts were created, the mock galaxies had to be united with the cutouts and convolved with a randomly selected PSF. As previously done in Margalef-Bentabol et al. (2025), the sample was reduced to galaxies with $M > 10^9 M_\odot$, and the number of images in a snapshots was reduced to randomly downsize it, with the only difference that this work also additionally included snapshots number 13, 17 and 21, which correspond to higher redshifts. After the reduction, snapshots 13, 17, and 21 yielded 684, 1892 and 4770 images respectively, while the rest provided 10000 images each. Exposure times, a necessary component for the code, were estimated by averaging the exposure times stored in the FITS headers of the dataset processed by Zhuang et al. (2024). The following values were obtained: 8245.824s for filter F115W and 2061.456s for filters F277W and F444W.

The creation of the mock images was executed with the aid of 3 Python scripts, used in Margalef-Bentabol et al. (2025). One served as a configuration file for specifying parameters related to the telescope and filters, such as applying the same pixel resolution of the telescope ($0.03''/pixel$), another built all the infrastructure needed, and the last one executed the injection process. Specifically, different functions accounted for injecting the galaxy into real-sky cutouts, the PSF convolution, converting the simulated images in the standard units of the JWST telescope (MegaJansky/steradian) and resizing the images to a final shape of 128×128 pixels.

Challenges arose due to file number quotas on the Hábrók cluster. To avoid reaching it, the cutouts were packaged into tars based on the filter. The team's code was then modified to be able to open tar packages in a temporary directory of the CPUs executing the job, as well as being optimised to run in parallel. This method was carried out for the rest of the work.

3.3 AGN injection

For each previously generated image and each filter, four different AGN power fractions were randomly selected from the interval [0,0.95] in steps of 0.05. For snapshots containing fewer images, eight fractions were chosen instead to increase sampling. Catalogues were then created following an 80/10/10 split for training, validation, and test sets. A random PSF is first resized to the image's pixel scale ($0.03''$/pixel), and normalised using aperture photometry within a $2''$ radius using the `photutils` package (Bradley et al., 2024). Using these catalogues as placeholders, the defined f_{AGN} is injected by scaling the PSF such that its flux corresponds to the desired AGN power fraction relative to the host galaxy flux. This can be represented with the formula

$$F_{AGN} = F_{gal} \cdot \frac{f_{AGN}}{1 - f_{AGN}} \tag{1}$$

where F_{AGN} is the flux of the AGN, F_{gal} the flux of the host galaxy, and f_{AGN} the AGN fraction. The composite image is constructed by summing the galaxy image and the scaled PSF. An example of the resulting AGN injected image is presented in Figure 6.

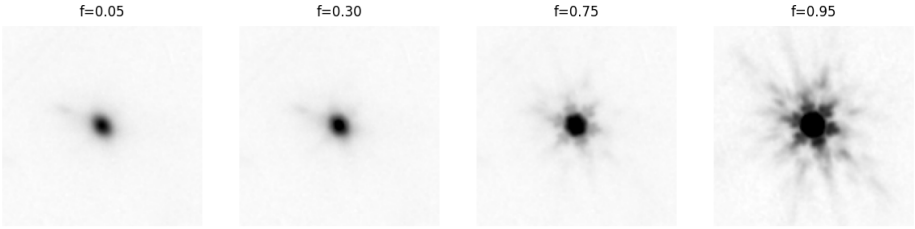

Figure 6: Example of a galaxy image with four different injected AGN fractions in the F277W filter. From left to right, the panels show increasing AGN contributions, demonstrating how the central brightness of the image changes as a function of the AGN fraction. The images are displayed in an inverse arcsin scale.

3.4 Deep learning with Zoobot

3.4.1 Training and validation

As previously mentioned, training was carried out using the `Zoobot` package (Walmsley, Allen, et al., 2023). `Zoobot` derives its weights from the Galaxy Zoo citizen-science classifications, so every backbone is already familiar with generic galaxy morphologies. The model chosen for both this thesis and Margalef-Bentabol et al. (2025) is the *ConvNeXt-Base*, a deep-learning model developed through a collaboration between Facebook AI Research and the University of California, Berkeley, as part of the larger ConvNeXt convolutional neural network family (Liu et al., 2022). ConvNeXt is a family of pure ConvNets, based on a residual neural network (ResNet) block with the added benefits of transformers. A ResNet is an architecture where the layers are learning residual functions instead of unreferenced ones (He et al., 2016). Transformers, in particular vision transformers, mix information across an image with self-attention layers (Dosovitskiy et al., 2021). ConvNeXt borrows the key ideas but keeps the faster, memory-efficient convolutional backbone, giving transformer-level accuracy while remaining a straightforward ConvNet that scales well to JWST cutouts. *ConvNeXt-Base* (Liu et al., 2022) is a modernised ResNet that replaces 3×3 kernels by 7×7 depth-wise convolutions, and inherits transformer-style GELU activations and training tricks while remaining a pure ConvNet. The *Base* model was chosen from the family of ConvNeXt architectures as it provides the best balance is performance, hardware cost and training time.

For consistency with the previous work, the same version of `Zoobot` (v 0.7) was used, which also meant that all supporting packages were downgraded to a compatible version. However, there were some issues with the package, and that's why the process was modified compared to Margalef-Bentabol et al. (2025).

The train and validation catalogues are unpacked, as well as the corresponding FITS files. Batch sizes are set to 32 for training, and either 16 or 32 for validation to ensure divisibility. The training images have random rotations or flips applied with a 50% percent of probability, while both training and validations are normalised by the mean and standard deviation of the dataset. The function `GalaxyDataModule` takes the images from the catalogues and applies the mentioned transformations, and returns image tensors, as well as the AGN fraction label. Then the *ConvNeXt-Base* is called on, specifying that the model is to be pre-trained and forcing the

input to be one-dimensional instead of the standard three-dimensional. The model is initialised with a *drop_path_rate*=0.4 and a learning rate of 10^{-5} in a mixed precision on a single GPU. The model is trained for 50 epochs, and only the last two ConvNeXt blocks are retrained, as well as the regression head. At the end of each epoch, `PyTorch-Lightning` switches to evaluation mode, forwards the validation batches through the unfrozen blocks, computes the mean-squared error between the predicted and injected AGN fraction, averages that error over the full set, and records the resulting loss.

Filter F115W served as the initial test bed for every stage of the pipeline. The code ran without incident on small local subsets, yet training on the full F115W catalogue consistently produced NaN values in both the training and validation losses. Because the catalogues themselves contain no missing or invalid entries, the NaNs point to a numerical instability in the tensor operations rather than to bad data. Debugging this issue would have required multiple long re-runs on the Hábrók cluster, which was not feasible within the available time and the long waiting times. In contrast, the identical code trained cleanly on the F277W and F444W datasets, converging without numerical errors. The remainder of this thesis, therefore, restricts all simulated data tests and real-data predictions to those two filters.

3.4.2 Testing

Efforts to mirror the software stack used by Margalef-Bentabol et al. (2025) succeeded for training but faltered during testing. Under `Zoobot v 0.7`, the evaluation script crashed with a channel-mismatch error because the ConvNeXt stem was still hard-wired for three-dimensional input even though the JWST cutouts are single-band. These issues disappeared after upgrading `Zoobot` to the current release (`v 2.0`), so all tests were run with that version while the trained weights themselves remained unchanged. The newer release internally patches the encoder, or repeats the single channel three times before the forward pass. Using this updated pipeline, the AGN-injected test catalogue was processed, and the mean difference and the following regression metrics were computed: root mean squared error (RMSE), relative absolute error (RAE) and the >20% and >30% outlier fractions based on the RAE. The formulas used were:

$$\Delta_{AGN} = f_{AGN,true} - f_{AGN,pred} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2} = \frac{\|y - x\|_2}{\sqrt{N}}. \tag{3}$$

$$RAE = \frac{|x_i - y_i|}{\tilde{x}_i}, \tilde{x}_i = \begin{cases} x_i, & x_i > 0, \\ 10^{-3}, & x_i = 0. \end{cases} \tag{4}$$

The results appear in Section 4.1. The same code was then applied to the real COSMOS-Web cutouts, and the predicted AGN fractions were saved for subsequent analysis. Given the absence of true values, the performance was evaluated based on different cutoffs and their Poisson 68% confidence interval.

3.5 Colour analysis

First, a master catalogue was constructed containing each predicted galaxy's coordinates and flux measurements in all four NIRCam filters. NIRCam magnitudes (in the AB system) were

computed from the standard formula for microJansky (μJy) units:

$$m = -2.5 \cdot \log_{10}(flux_{\mu Jy}) + 23.9 \tag{5}$$

Additional fluxes from Dr. Wang's team comprehensive dataset were used to yield IRAC-band magnitudes. The colour-colour criteria of Chang et al.(2017) were then applied, selecting MIR AGN candidates satisfying

$$y < 2.22x + 1.01 \tag{6}$$

$$y < 8.67x - 0.28 \tag{7}$$

$$y > -0.33x + 0.17 \tag{8}$$

$$y > 0.31x - 0.06 \tag{9}$$

where $x = m_{3.6\mu m} - m_{5.8\mu m}$ and $y = m_{4.5\mu m} - m_{8\mu m}$.

These candidates were cross-matched against the working catalogue and flagged with a Boolean `is_MIR_AGN`. Lastly, predictions from Margalef-Bentabol et al. (2025) in the F150W were also joined to both provide completeness and compensate for the lack of predictions in the filter F115W. However, only 2779 galaxies were matched out of 6829, most likely because the real cutouts catalogue was built with an updated extraction script rather than the original code base. Figures 12-14 illustrate the resulting spatial and colour distributions under various AGN-fraction thresholds. The MIR-selected galaxies were also overlapped to see if the distributions matched somewhere. These results can be observed in Figure 15.

Attempts to manually create functions to separate these AGNs were futile, prompting the implementation of an ML pipeline. For each galaxy, every plausible NIRCam colour was computed, and the source was labelled AGN only when the median of its predicted f_{AGN} in F150W, F277W and F444W exceeded 0.2. A balanced-weight Random-Forest classifier was first trained on this binary target. Its Gini importances ranked the colours by predictive power, after which the three most informative indices were passed to a logistic-regression layer that converts their linear combination, through a sigmoid, into an AGN probability. Several variants of this set-up were tested: runs that did or did not allow colours containing F115W, runs with or without the Boolean `is_MIR_AGN` flag appended as an extra feature, and probability cuts of 0.5 versus 0.7 applied to the logistic outputs. Performance for each combination, measured on the full sample, is listed in Appendix Tables 6 and 7; the configuration that maximises the F_1 score is carried forward to Section 4.3.

4 RESULTS

4.1 Results on simulated data

Here are presented the results from testing with `Zoobot` with the simulated dataset. Figure 7 showcases the distribution of the predicted values against the true values for both F277W and F444W. The results with the F277W present more outliers from the one-to-one line; however, their distribution is less sparse along the same line. Both filters suffered from overestimating at low AGN fractions and underestimating at high ones, especially filter F444W. Positively, most AGN fractions were predicted with high accuracy. For filter F277W, 57837 out of 61832 (93.5%) predictions were correctly made within ± 0.05, within ± 0.025 45287 (73.2%) were correct, and for ± 0.01 it was just 23112 out of 61832 (37.4%). For the F444W filter, 55359 predictions out of

61832 (89.5%) were within ±0.05, 37459 (60.6%) within ±0.025 and 15166 (24.5%) in the ±0.01 range.

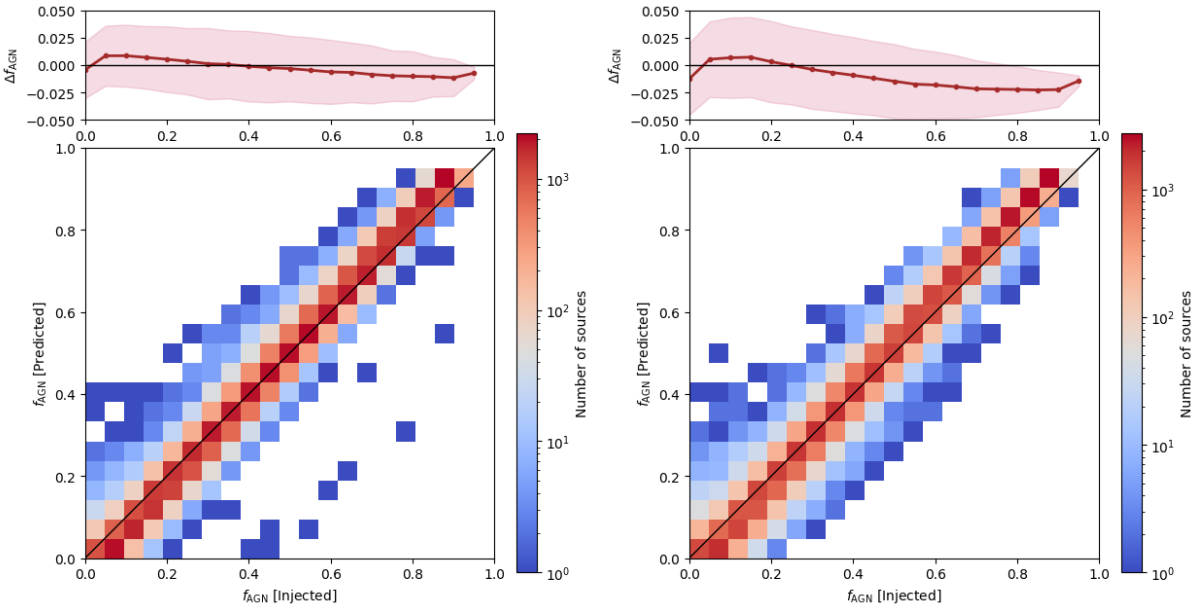

Figure 7: Joint distribution plots for F277W (left) and F444W (right). The top panels show the residual plots of the mean difference (Δ_{AGN}) between true and predicted values with shaded$\pm 1\sigma$ intervals as a function of the injected fraction. The bottom plots are density heatmaps of the predicted values against the true values, overlaid with the one-to-one line. The colourbar, scaled logarithmically, indicates the number of sources per bin.

The statistical metrics of the model are presented in Table 2, calculated with the formulae presented in Section 3.4.2. Overall, F277W performed better than F444W, but both filters performed slightly worse than F150W in Margalef-Bentabol et al. (2025).

Table 2: Regression metrics for AGN fraction prediction in F277W and F444W, as well as results in F150W from Margalef-Bentabol et al. (2025), as comparison.

Filter	**Overall RAE**	**Overall RMSE**	**% RAE > 20%**	**% RAE > 30%**
F277W	0.755 ± 0.024	0.028 ± 0.001	13.3%	9.9%
F444W	1.022 ± 0.031	0.034 ± 0.001	15.2%	11.4%
F150W	0.076 ± 0.005	0.013 ± 0.005	6.5%	6.1%

Values of RMSE and RAE for each bin of AGN fractions are plotted in Figure 8, visually comparing the metrics in both filters. The RMSE presents a sharp drop once reached $f_{AGN} = 0.9$, and F277W constantly achieves lower values of RMSE. The same can be said in the case of the RAE, although the values blow up at $f_{AGN} = 0$. Similar trends, regarding both the RAE and RMSE, were observed in Margalef-Bentabol et al. (2025). The outliers percentage of the 20% and 30% thresholds is visualised in Figure 9. Because of the RAE, both filters start with close to 100% of outliers in both thresholds. These values drop quickly and at $f_{AGN} = 0.1$ reach $\approx 20\%$ in the F277W filter, and $\approx 30\%$ in the F444W filter. Beyond $f_{AGN} > 0.5$ all curves closely approach zero.

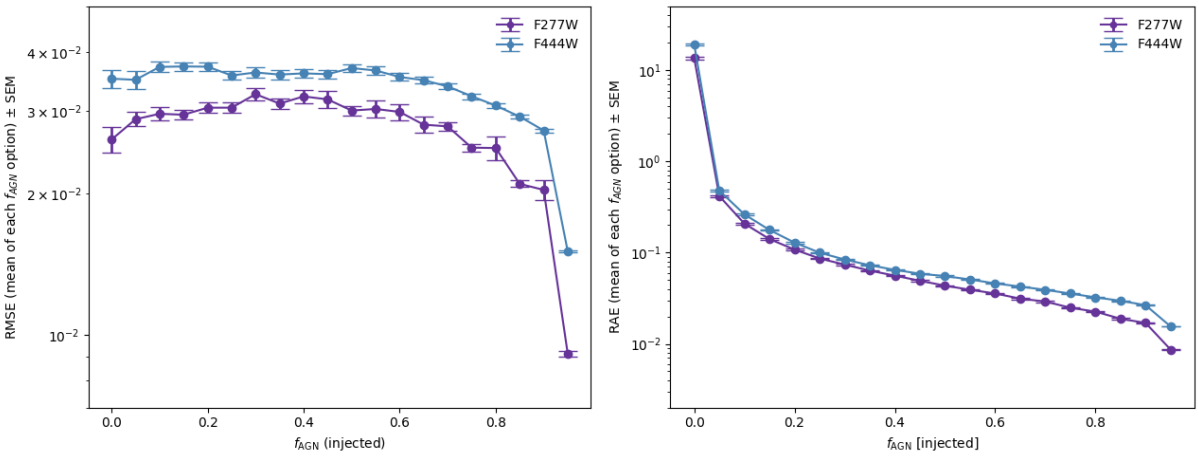

Figure 8: Plots of the RMSE (left) and RAE (right) calculated as the mean value of each bin in both filters, with error bars representing the squared error of the mean (SEM). Both y-axes are scaled logarithmically.

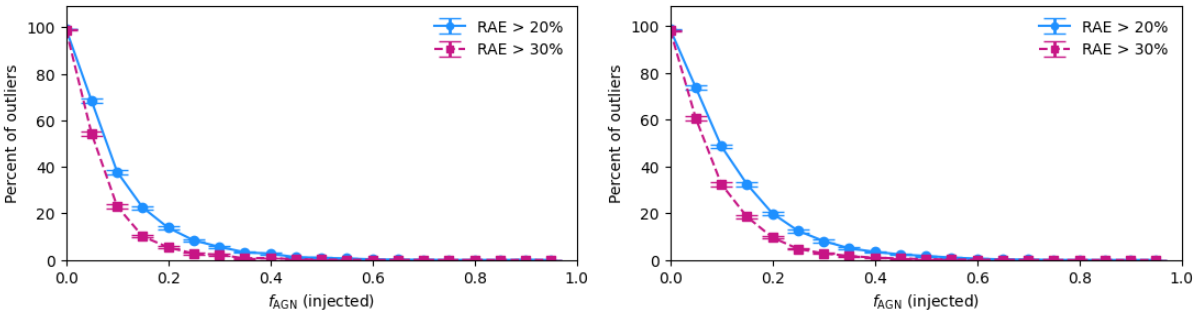

Figure 9: Percentage of sources with relative absolute error (RAE) exceeding 20% and 30% as a function of injected AGN fraction (f_{AGN}). The left panel represents F277W, and the right one F444W. Error bars denote ± SEM.

The 5σ detection thresholds were set to five times the overall RMSE in each filter, resulting in 0.1395 for F277W and 0.17 for F444W. The one-dimensional histograms of the predicted f_{AGN} are shown in Figure 10, with those lines highlighted. Both filters present peaks of distribution at the extremes of the f_{AGN} interval. This edge-peaking is an artefact of the present network design: the single linear regression neuron is optimised with mean-squared error and its outputs are hard-clipped to the physically allowed range [0,0.95]. Together these two choices encourage the optimiser to drive some predictions toward the boundaries, producing the twin peaks visible in both panels (further discussed in Section 5.1).

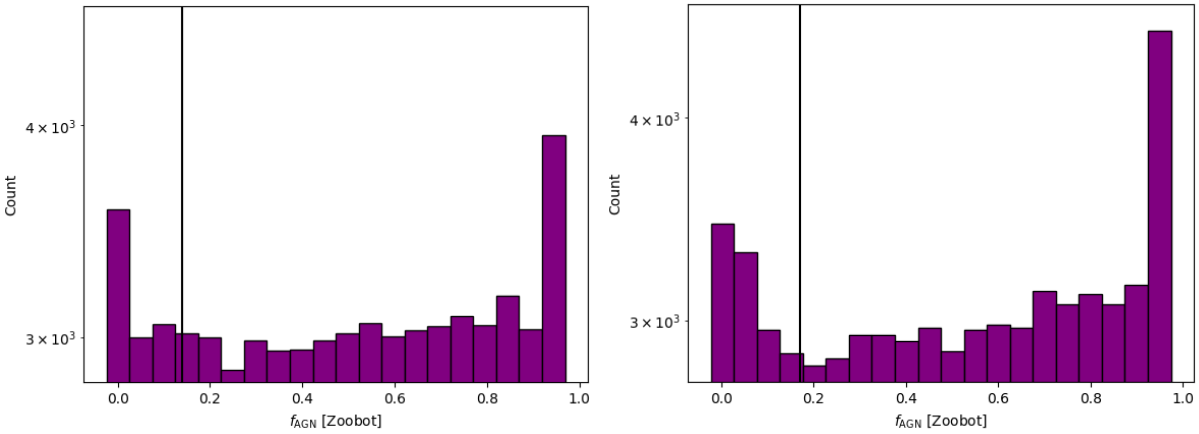

Figure 10: Log-scaled histograms of predicted f_{AGN} for the simulated test sets in F277W (left) and F444W (right). Vertical dashed lines mark the 5σ thresholds at $f_{AGN} = 0.1395$ (F277W) and $f_{AGN} = 0.17$ (F444W).

4.2 Results on real data

Results of applying the `Zoobot` model to real JWST cutouts are shown in Figure 11. In both F277W and F444W, the vast majority of galaxies are assigned $f_{AGN} < 0.1$. The F444W distribution, however, falls off more gradually with increasing f_{AGN} than does F277W, but plateaus at the mid fractions to then drop at around $f_{AGN} \approx 0.7$.

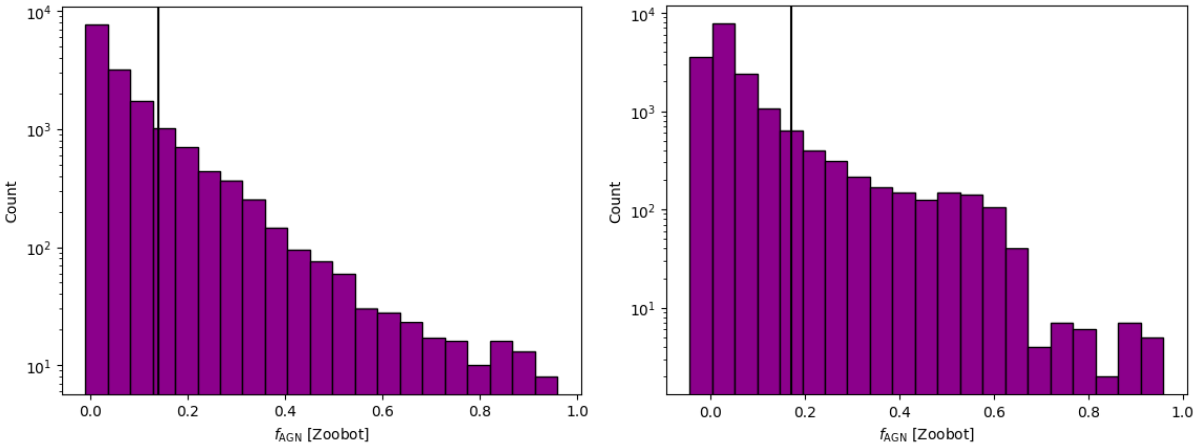

Figure 11: Log-scaled histograms of predicted f_{AGN} for the real JWST galaxies in F277W (left) and F444W (right). Vertical dashed lines mark the 5σ thresholds at $f_{AGN} = 0.1395$ (F277W) and $f_{AGN} = 0.17$ (F444W).

Table 3 presents the various thresholds tried and the number of galaxies that each yielded, together with the 68% Poisson confidence intervals. As the selection criterion tightens from the 5σ cut down to $f_{AGN} = 0.7$, the fraction of galaxies meeting it falls steadily. F277W consistently yielded a larger candidate sample than F444W at every cutoff. Regardless, even the strictest criteria retain enough sources for robust statistics, because of the small Poisson uncertainties.

Table 3: Number and fraction of galaxies above various f_{AGN} thresholds, with 68% Poisson CIs in percentages.

Filter	**Threshold**	N	**Total**	**Fraction** (%)	$+\Delta$ (%)	$-\Delta$ (%)
F277W	5σ (0.1395)	3045	16014	19.01	+0.35	-0.34
	$f > 0.1$	4280	16014	26.73	+0.41	-0.41
	$f > 0.2$	1876	16014	11.71	+0.28	-0.27
	$f > 0.7$	73	16014	0.46	+0.06	-0.05
F444W	5σ (0.17)	2110	17252	12.23	+0.27	-0.26
	$f > 0.1$	3518	17252	20.39	+0.35	-0.34
	$f > 0.2$	1777	17252	10.30	+0.25	-0.24
	$f > 0.7$	28	17252	0.16	+0.04	-0.03

4.3 Colour analysis results

This next section presents the colour mosaics constructed by taking into account the predicted AGN fractions. Although all possible colour indices, including those involving F115W, were computed and evaluated by the Random Forest feature-ranking step of the ML pipeline, no `Zoobot` predictions exist in F115W itself. Including F115W-based mosaics here would therefore break the visual consistency of mapping predicted values. For completeness, the full set of colour mosaics are provided in Appendix Section 9.1. The colour combinations were considered both as the y- and x-axis to create 6 total possible combinations and 3 colour calculated. Figures 12 through 14 show the initial thresholds considered for the ML model, separated by each filter to gather an idea of how many galaxies could be considered. The number of selected sources are presented in Table 4.

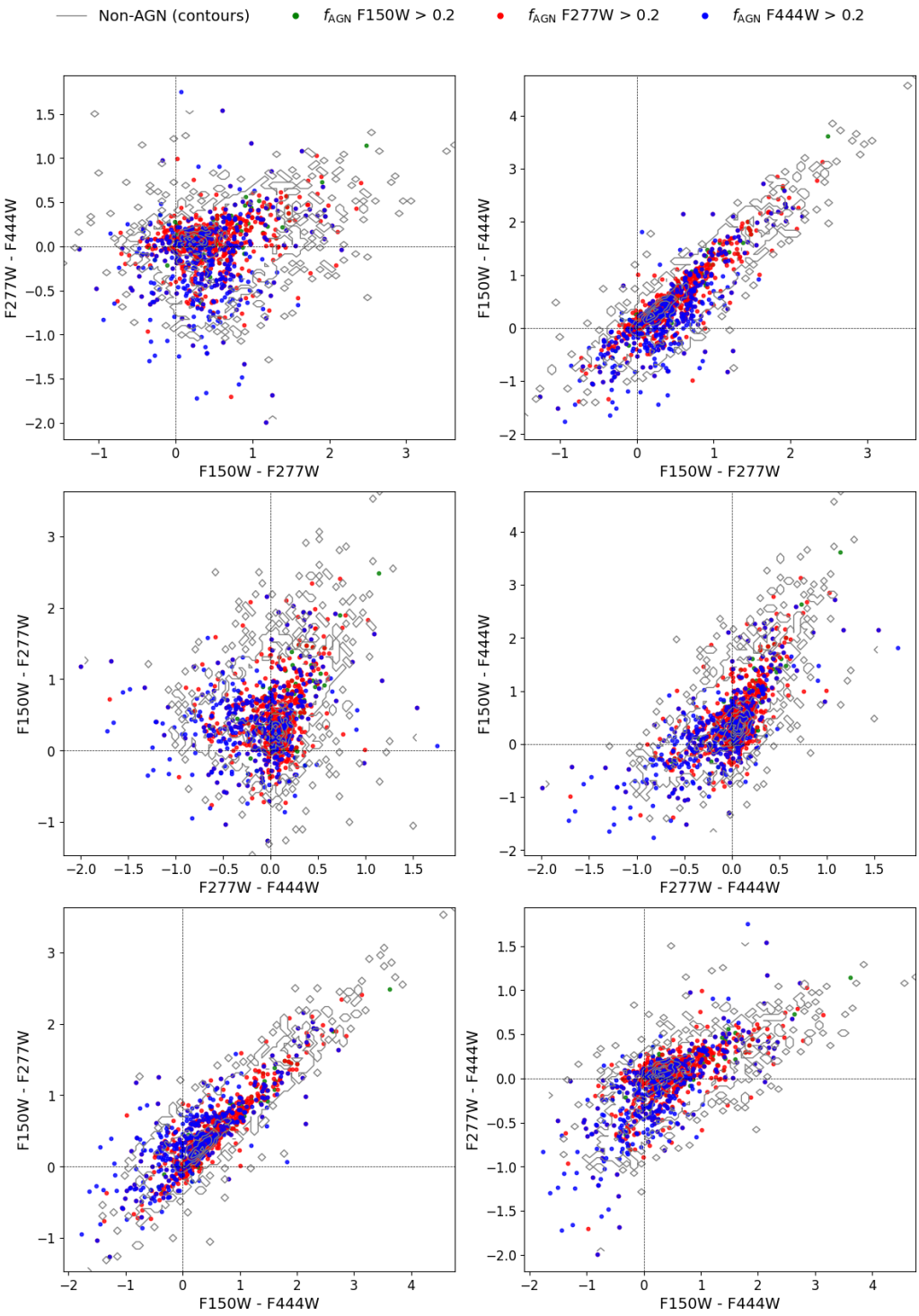

Figure 12: Six-panel mosaic of colour-colour diagrams in the NIRCam indices F150W-F277W, F150W-F444W, and F277W-F444W. Grey density contours mark where the bulk of non-AGN galaxies lie, while galaxies with $f_{\mathrm{AGN}} > 0.20$ in F150W (green), F277W (red), or F444W (blue) are plotted as filled points. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. At this low threshold the highlighted sample is very large and populates almost the entire stellar locus, so genuine overdensities are difficult to discern; every colour-colour plane appears saturated with candidates.

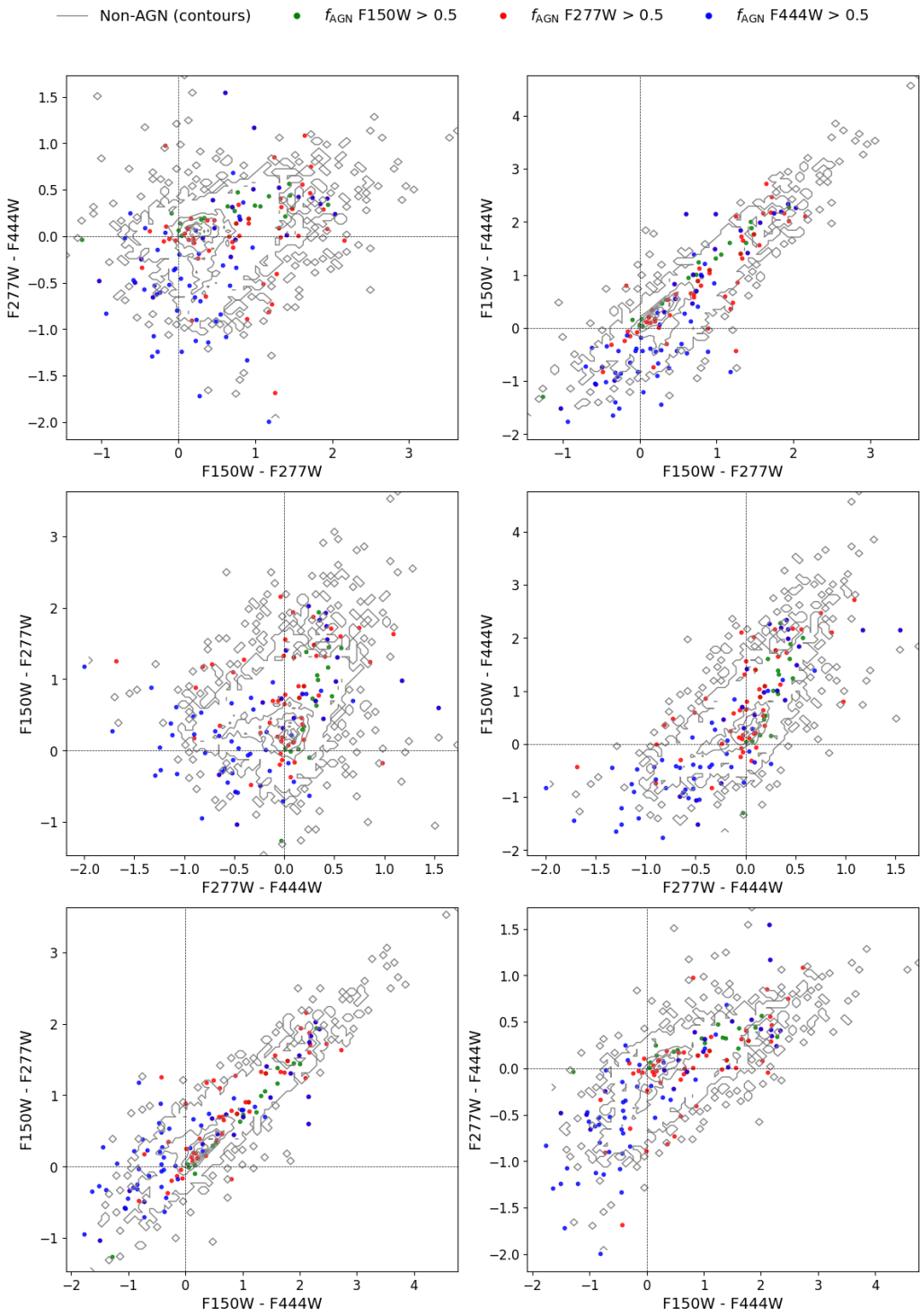

Figure 13: Six-panel mosaic of colour-colour diagrams in the NIRCam indices F150W-F277W, F150W-F444W, and F277W-F444W. Grey density contours mark where the bulk of non-AGN galaxies lie, while galaxies with $f_{\mathrm{AGN}} > 0.50$ in F150W (green), F277W (red), or F444W (blue) are plotted as filled points. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. The plots appear busy, especially in the central panels, indicating that half-fraction AGN still overlap heavily with the dense non-AGN core.

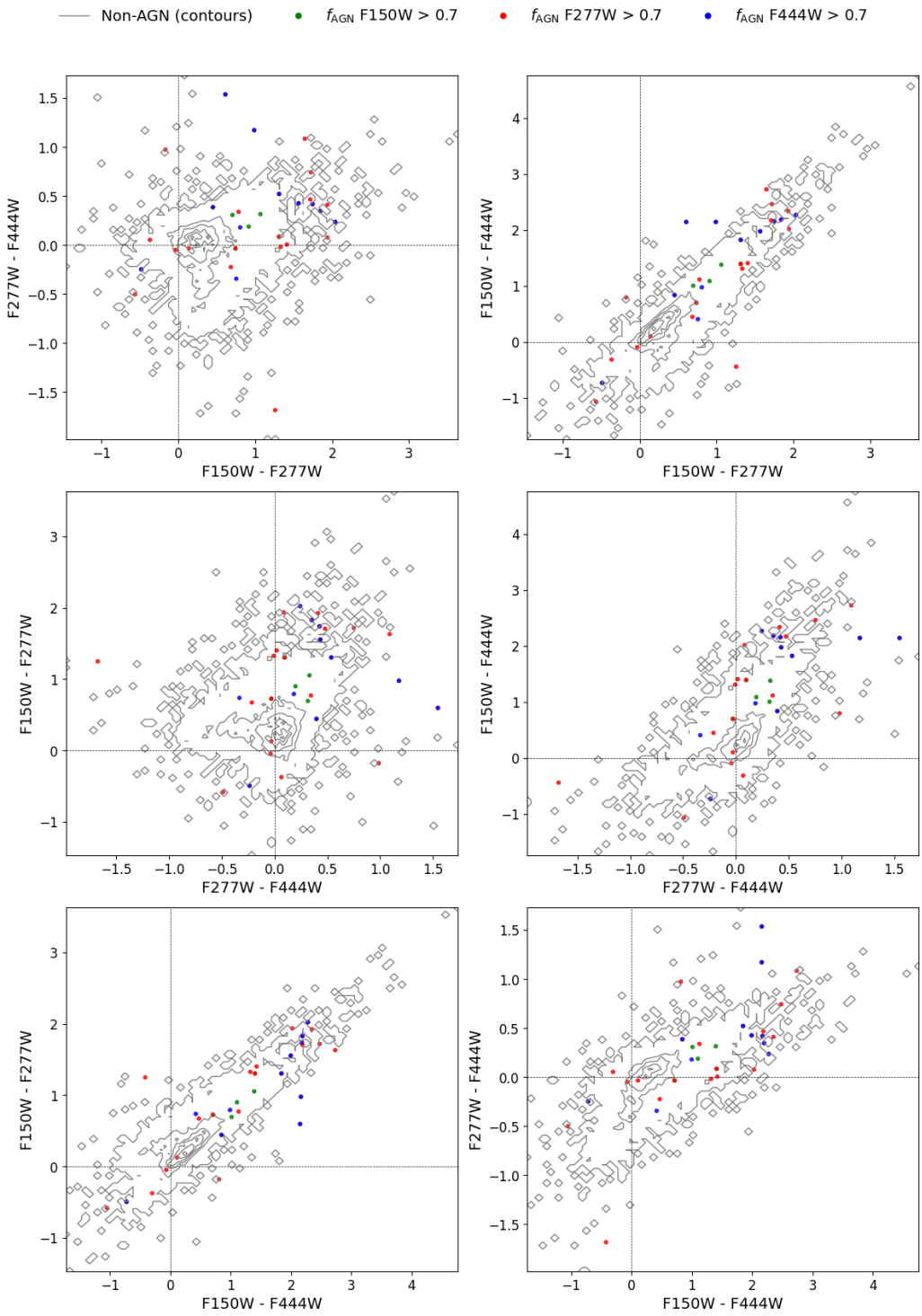

Figure 14: Six-panel mosaic of colour-colour diagrams in the NIRCam indices F150W-F277W, F150W-F444W, and F277W-F444W. Grey density contours mark where the bulk of non-AGN galaxies lie, while galaxies with $f_{\mathrm{AGN}} > 0.50$ in F150W (green), F277W (red), or F444W (blue) are plotted as filled points. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. The non-AGN population fills the colour locus while high-confidence AGNs avoid its crowded centre. The scarcity of points, however, means the diagrams become sparsely populated and statistical leverage diminishes; the 0.70 cut is therefore visually clean but too restrictive for robust sample sizes.

Table 4: Number of galaxies per filter exceeding selected AGN-fraction thresholds

Threshold	**F150W**	**F277W**	**F444W**
$f_{\mathrm{AGN}} > 0.2$	159	693	391
$f_{\mathrm{AGN}} > 0.5$	23	69	63
$f_{\mathrm{AGN}} > 0.7$	5	26	11

Figures 12 and 13 demonstrate that applying a $f_{\mathrm{AGN}} > 0.2$ or $f_{\mathrm{AGN}} > 0.5$ threshold in each filter yields an excessively large candidate set, whereas the stricter cut of $f_{\mathrm{AGN}} > 0.7$ is too restrictive. To balance sample size and purity, a simultaneous requirement of $f_{\mathrm{AGN}} > 0.2$ in all three filters was adopted for visualisations. The resulting selection gave 40 galaxies and the MIR selected AGNs are overlayed, as shown in Figure 15. In each plot, the full galaxy sample outlines a tight diagonal locus, with low colour values clustering toward the lower-left and high values toward the upper-right. The criteria-selected AGN carve out a distinct "V" within this locus: one arm descending into the lower-left (bluer-bluer combinations) and the other ascending into the upper-right (redder-redder combinations), meeting at an inflection close to the colour midpoint. By contrast, MIR-selected AGN occupy only the upper-right branch of that V, clustering among the reddest colours and leaving the lower-left arm virtually empty.

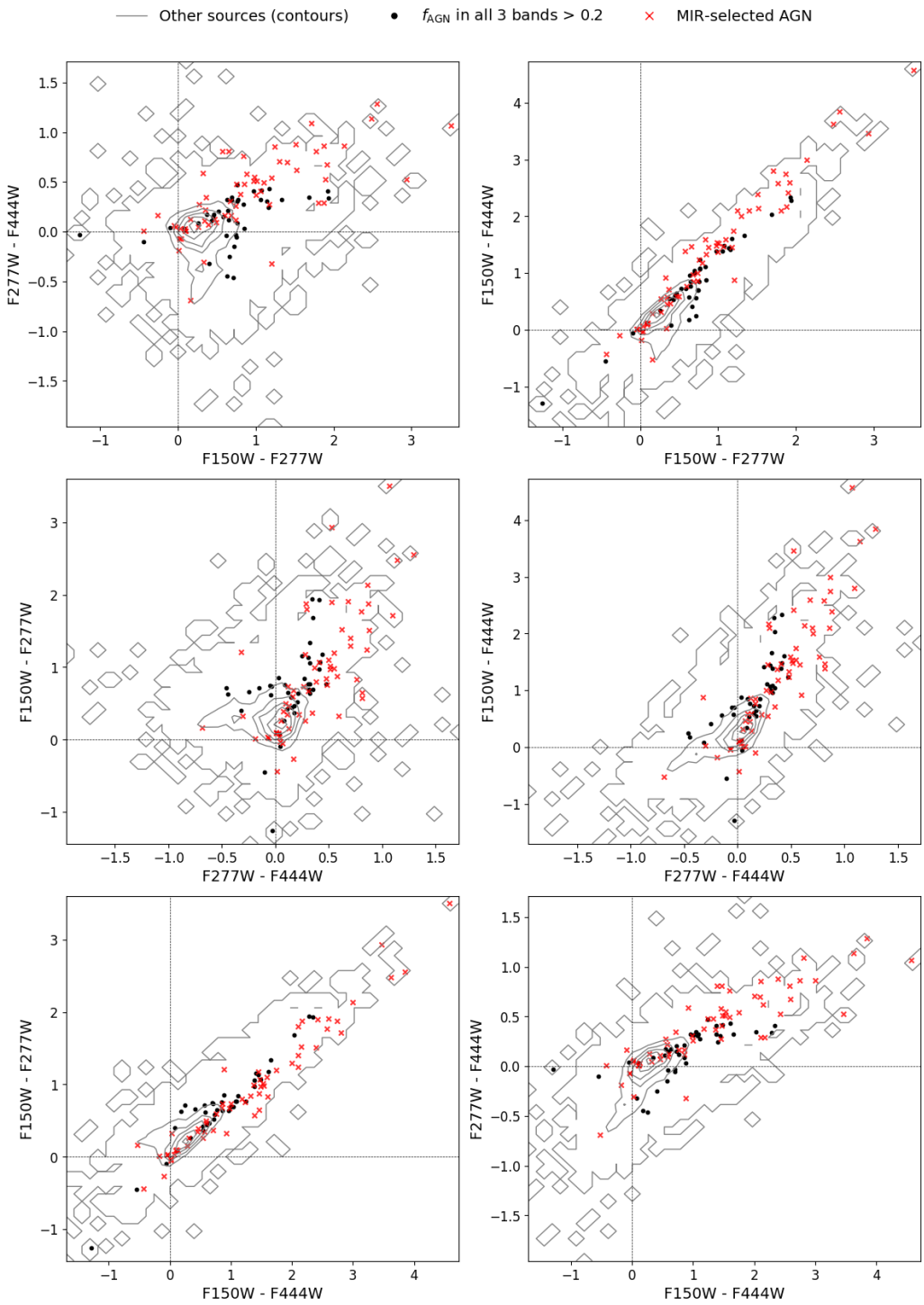

Figure 15: Six-panel mosaic of NIRCam colour-colour diagrams (F150W-F277W, F150W-F444W, F277W-F444W). Grey contours trace the full galaxy population, black dots mark galaxies with $f_{\mathrm{AGN}} > 0.2$ in all three broad bands and red crosses are MIR-selected AGN. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. Requiring $f_{\mathrm{AGN}} > 0.20$ simultaneously in all bands balances completeness and purity, retaining 40 candidates that populate both arms of the "V" locus, while the MIR cut selects only the reddest arm.

Out of the grid of trials described in the Appendix Tables 6 and 7, although all models exhibit modest precision and recall, Trial 5 achieved the strongest overall balance, with the highest F1 score (0.174), an accuracy of 92.5%, precision of 20.4% and recall of 15.2%. This configuration employs three colour indices (including F115W), applies the criterion median($f_{\rm AGN}$) > 0.2**,** and uses a probability threshold of 0.7.

The very high accuracy, coupled with low precision and recall values, shows an imbalance in the data sample (as most galaxies are non-AGN). To account for this, the model was re-evaluated on a balanced subset containing the same number of non-AGN as AGN. With the original probability threshold of 0.7, the balanced test yields a 56.9% accuracy, 91.3% precision, 15.2% recall, and an F1 score of 0.261, revealing that the model is highly conservative. Lowering the threshold to 0.5 improves recall (and F1) at the expense of precision, giving a 63.4% accuracy, 66.4% precision, 54.3% recall, and a F1 score of 0.598. However, it is important to keep in mind that the true values that these metrics are based on are the AGN-selected galaxies with the criterion median($f_{\rm AGN}$) > 0.2, which are not an absolute truth. A more direct check is to compare how the three selection methods overlap, as shown in Table 5. Out of 515 sources with median($f_{\rm AGN}$) > 0.2, 21 were selected from the ML pipeline, only 5 are shared with the MIR selection, and similarly, only 4 sources are shared between MIR and ML-selected. Zero sources were identified as AGN by all three sources.

Table 5: Overlap between median, MIR, and ML selections.

Category	**Count**	**Percentage (out of 6828 galaxies)**
Only median($f_{\rm AGN}$) > 0.2	489	7.16%
Only MIR-selected	54	0.79%
Only ML-selected	78	1.14%
median($f_{\rm AGN}$) > 0.2 & MIR-selected	5	0.07%
median($f_{\rm AGN}$) > 0.2 & ML-selected	21	0.30%
MIR-selected & ML-selected	4	0.05%
All three	0	0%

Since the median-based selection retains 515 galaxies, the resulting mosaics are visually crowded; Figure 15 therefore presents the cleaner alternative that requires $f_{\rm AGN} > 0.2$ in every filter. The full set of colour mosaics for the median-based cut is provided in Appendix Figure 22. Trial 5 ranked the colour indices by Gini importance, yielding F115W-F277W (0.1850), F150W-F277W (0.1813) and F115W-F150W (0.1674) as the three most informative, while all other combinations clustered lower, between 0.1525 and 0.1579. The logistic regression decision function is:

$$\begin{aligned} z = &-0.107 + 0.184\,(F115W - F277W) \\ &+ 0.017\,(F150W - F277W) + 0.270\,(F115W - F150W) \end{aligned} \tag{10}$$

The ML model selected 103 galaxies that respected the criteria mentioned, and the mosaic with those overlayed with the previous selections is presented in Figure 16. The ML-selected objects follow the same redward branch of the colour locus as the MIR-selected AGNs: they are concentrated in the upper-right quadrant of every panel and are almost completely absent from the lower-left (blue-dominated) arm. Because the model did not use the MIR flag as an input, this area overlap demonstrates that the three colours alone carry enough information for the classifier to reproduce the MIR-based selection and to enlarge it modestly.

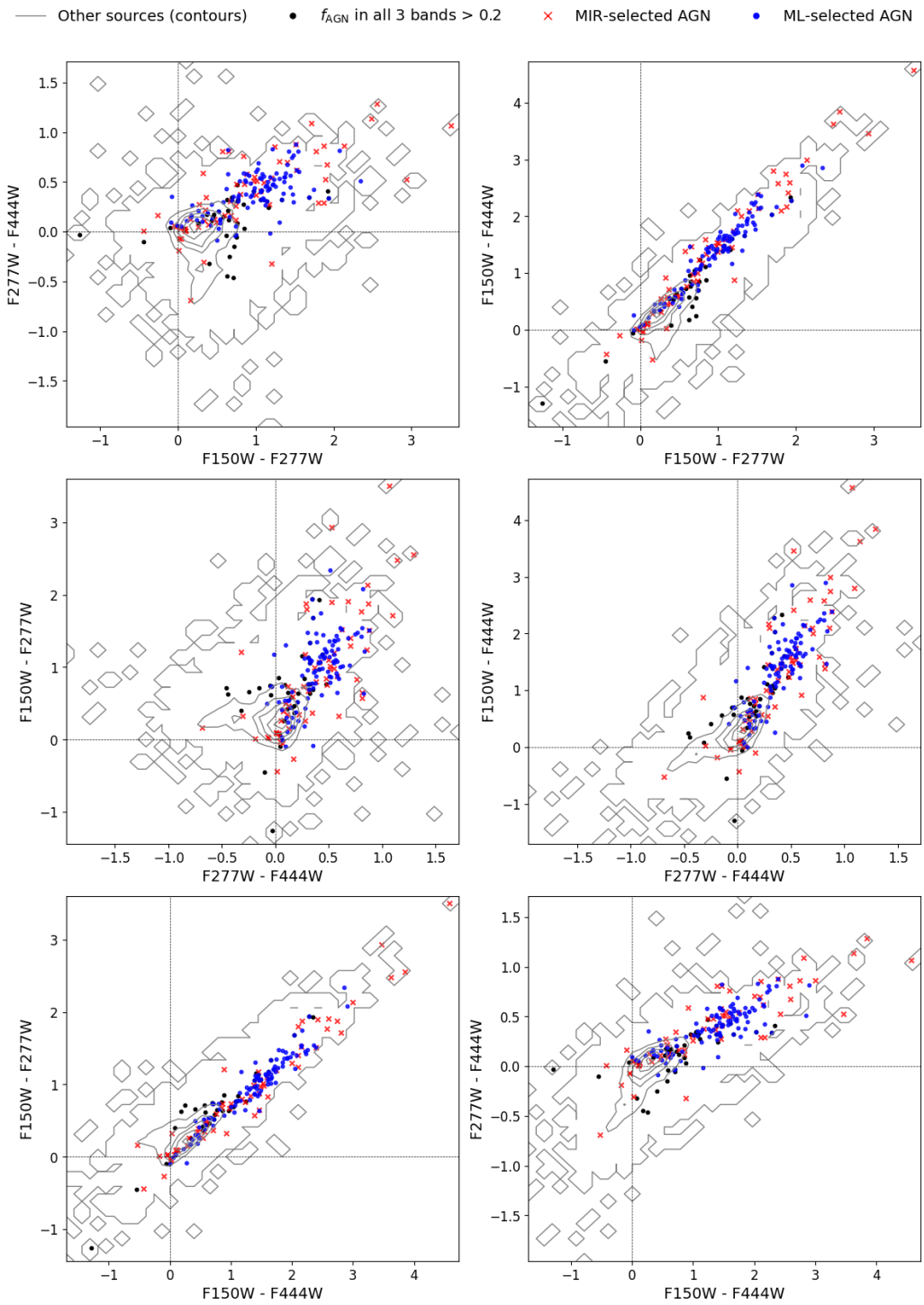

Figure 16: Six-panel mosaic of colour-colour diagrams for the NIRCam pairs F150W-F277W, F150W-F444W and F277W-F444W. Grey contours trace the full galaxy population, black dots mark galaxies with $f_{\mathrm{AGN}} > 0.2$ in all three broad bands, red crosses are MIR-selected AGN, and blue circles are AGN flagged by the 3-colour logistic classifier. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. All three AGN tracers crowd the redward arm of the locus; MIR and ML selections organise themselves almost entirely in the upper-right quadrant while avoiding the blue-dominated lower-left branch.

Figure 17 maps every two-colour combination of the four NIRCam bands while colouring each galaxy by the raw output of the logistic-regression model, corresponding to Eq. 10. The resulting gradient is remarkably coherent across all panels: points shift smoothly from deep blue (negative z values) in the lower-left of each diagram to deep red (positive z values) in the upper-right. In the Appendix, Figure 23 shows a poly-line graph visualising the same trends, as a way to represent multi-dimensionality in two dimensions.

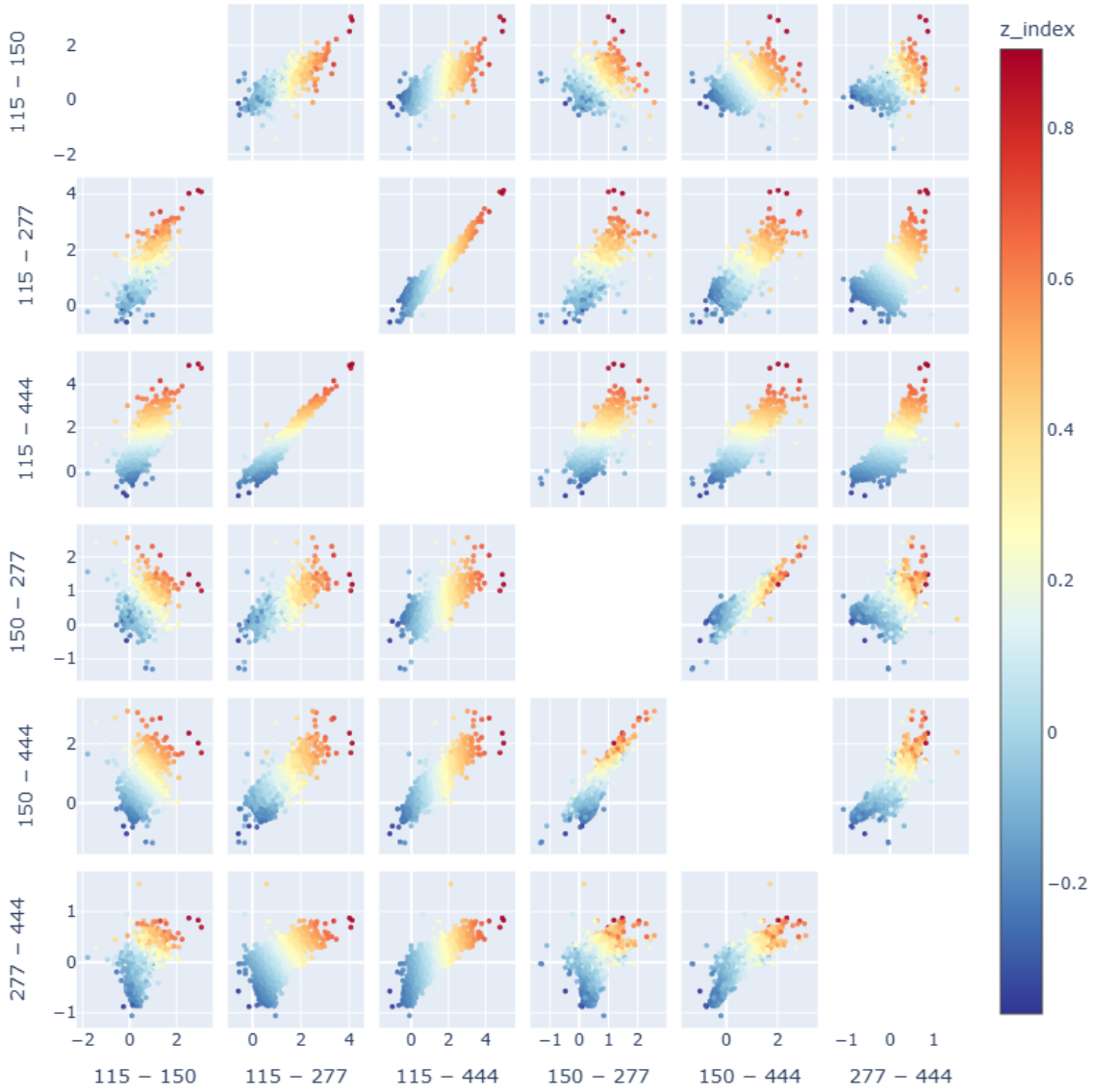

Figure 17: Pairwise colour-colour mosaic for all four NIRCam broad-band indices, with each galaxy point coloured by the logistic-regression equation expressed in Eq. 10. Bolded grid lines at zero colour guide the transition from bluer to redder systems, and the shared colour bar at right applies to every panel.

5 DISCUSSION

5.1 Data and model limitations

The results shown in Section 4.1 resemble those reported by Margalef-Bentabol et al. (2025): RMSE and RAE curves follow the same shape, and both studies show a systematic over-prediction at the lowest injected fractions, tapering to a slight under-prediction toward the top end. The slightly wider spread around the 1:1 line in this work is expected because the present model is trained and evaluated on a much larger sample (≈ 61000 versus ≈ 4800 galaxies).

One striking feature is the distribution of the predictions at the extremes of the AGN power fraction range. Two factors likely drive this behaviour. First, the regression head is a linear neuron optimised with mean-squared error, so errors at mid-fractions are penalised more heavily than errors near the ends, with the optimiser tending towards the extremes. Second, any output outside the physically allowed interval [0,0.95] is hard-clipped during post-processing, further piling up predictions at the boundaries. Adopting a sigmoid function, as done in Margalef-Bentabol et al. (2025), would mitigate this tendency.

Regarding the real COSMOS-Web cutouts, the distributions observed in Figure 11 seem to reflect the previous work. In particular, the F277W histogram declines steadily with increasing f_{AGN}, as was the case for the filter F150W in the reference paper. However, F444W exhibits a plateau from $0.35 < f_{AGN} < 0.55$ before dropping. At the F277W central wavelength (2.77 μm) the simulated PSF FWHM is $\approx 0.09''$, whereas at F444W (4.44 μm) it grows to $\approx 0.14''$ (Rigby et al., 2023; Space Telescope Science Institute, 2024b). At the same time, the glow from warm telescope components and zodiacal dust gets stronger from 4 μm (Space Telescope Science Institute, 2024a). The larger, more structured PSF dilutes the central point source and reduces the contrast that the network relies on, while the higher background lowers the effective signal-to-noise ratio of faint AGN signatures. The fact that, in all thresholds presented in Table 3, F277W yielded more AGN candidates than F444W reflects the higher sensitivity of this filter. When the requirement climbs to 0.20, the disparity narrows (11.7% vs. 10.3%), reflecting the mid-fraction "plateau" seen in the F444W histogram. At the $f_{AGN} > 0.70$ cut, F277W still contributes 73 candidates, whereas F444W collapses to 28. Because the Poisson 68% errors are $<$0.05% in every case, the divergence is real and must be driven by instrument physics rather than small-number noise. Conversely, the fact that the two filters agree to within a percentage point at the moderate 0.2 cut suggests that mid-level AGN identifications are robust against wavelength-dependent systematics, while very high-contrast cases are not.

Finally, because the true AGN fractions of real JWST galaxies are unknown, all real-data metrics remain provisional. Examining performance as a function of redshift, S/N, and host brightness, as explored in Margalef-Bentabol et al. (2025), could have added further information on the validity of this model.

5.2 Colour selection

One serious limitation in the colour-colour analysis is the absence of reliable F115W predictions. Without that band, the full-panel mosaic cannot be explored in its entirety, and any visual separation of AGN and non-AGN is necessarily incomplete. To obtain at least three independent colours, the F150W predictions catalogue from Margalef-Bentabol et al. (2025) was merged with the one related to this work. However, only $\approx$ 2800 objects intersected the present sample. Using the criterion median(f_{AGN}) > 0.2 was therefore a pragmatic compromise: it retains every galaxy that has predictions in any two of the three filters (F150W, F277W, F444W) instead of

discarding around 60% of the data.

The distribution of the `Zoobot` predictions and the MIR-selected galaxies in Figure 15 show that the `Zoobot`-selected AGNs almost span the full colour range, whereas the MIR-selected sources cluster almost exclusively on the reddest branch. This divergence is expected. The MIR selection relies on the IRAC 3.6-8 μm criteria of Chang et al. (2017), which are sensitive to the hot dusty torus and thus pick out AGN with either substantial extinction or intrinsically strong mid-IR continua. The `Zoobot` fractions, in contrast, are measured directly from the spatial distribution of light in rest-frame optical and NIR images: they remain high for unobscured nuclei even when the broadband colours are still blue.

5.3 Selection comparison and limitations

The ML-selected galaxies are found in the same loci as the MIR-selected, as well as the galaxies with an $f_{\mathrm{AGN}} > 0.2$. The fact that no model is found in another area is a strong confirmation of the correct general distribution of AGNs in colour-colour diagrams, although, as mentioned, all models would benefit from the inclusion of the F115W filter.

The ML pipeline suffered from the extreme difference in sample size between AGNs and non-AGNs, affecting its performance. One remedy would be to train the network on colour-defined subsets or to use balanced smaller batches, so the overwhelming excess of blue, non-AGN galaxies no longer skews the optimisation and degrades performance in the redder, AGN-rich regime.

In Figure 17, the steepest colour change aligns with the F115W-F150W axis, reflecting that term's dominant coefficient in the linear regression decision function. Even projections that do not explicitly appear in the formula, such as F277W-F444W, inherit the same trends because those colours are correlated with the weighted ones. The three-colour linear model therefore imposes a single tilted plane through the full four-band space; moving orthogonally to that plane shifts the value of Eq. 10 smoothly, from confident non-AGN to confident AGN.

The fact that very few sources are selected by two or all methods could be explained by two different points of view. The first is that photometric-AGN selection and the ML selection are geometry-driven, whereas MIR selection is dust-temperature driven. The MIR method could therefore be considered complementary, and the modest overlap seen in Table 5 can be considered a natural outcome rather than a failure of either technique.

A second explanation is the bandpass mismatch itself. The MIR cut relies on Spitzer/IRAC photometry with longer wavelengths than JWST NIRCam's 5 μm limit, so at any given redshift it samples hotter dust and a different part of the spectral energy distribution than the JWST-only colours. The observed partial overlap could therefore reflect genuine physical differences between geometry-driven and temperature-driven AGN tracers and/or the fact that the two selections probe non-overlapping wavelength regimes.

6 CONCLUSION AND FUTURE DIRECTIONS

The purpose of this thesis was to build a JWST-only pipeline for estimating the fractional AGN contribution (f_{AGN}) in COSMOS-Web galaxies in the filters F115W, F277W and F444W of the NIRCam on the JWST, and to compare those estimates against MIR and other colour-colour selections. A *ConvNeXt-Base* network was trained with Illustris-TNG cutouts that were PSF-convolved and AGN-injected. It attained a RAE$= 0.755 \pm 0.024$ and a RMSE$= 0.028 \pm 0.001$ in the F277W filter, followed by F444W with RAE$= 1.022 \pm 0.031$ and a RMSE$= 0.034 \pm 0.001$, although not yet at the 0.013 RMSE reported for F150W by Margalef-Bentabol et al. (2025).

Both filters had cases of overestimation for low f_{AGN} and underestimation for high f_{AGN}. Around 93 % of F277W predictions lie within ± 0.05 of the truth and, for various thresholds, more sources were contained in those ranges than in the F444W filter, tracing this disparity to the poor spatial resolution in the latter. The linear regression head, together with post-processing hard-clipping, drives the excess of predictions at the extreme ends of the f_{AGN} range.

On real COSMOS-Web cutouts the network flags 3045 galaxies above a 5σ threshold in F277W and 2110 in F444W. Combining these fractions with every possible NIRCam colour index shows that MIR-selected AGN occupy the upper-right, reddest branch of each colour-colour panel, while `Zoobot` predictions trace the same area.

A Random Forest ranking identifies F115W-F277W, F150W-F277W, and F115W-F150W as the most informative indices. A logistic regression based on those three colours, thresholded at $P_{AGN} > 0.7$, returns 103 ML AGNs that extend the MIR sequence but leave the blue arm largely untouched, underscoring the complementary nature of dust-temperature and geometry-based tracers.

Several avenues could transform this prototype into a survey-grade AGN finder. On the network side, retraining F277W and F444W with a sigmoid activation would soften the edge piling, while debugging the numerical instability that blocked F115W would restore the short-wavelength channel and the colours that depend on it. A homogeneous re-run of F150W within the present pipeline, would give source-by-source consistency, and the expanded set of high-redshift TNG snapshots already assembled here invites a systematic test of accuracy versus redshift, S/N and host size, as already done in Margalef-Bentabol et al. (2025). Spectra of the $f_{AGN} > 0.7$ candidates, where available, would further ground the photometric estimates if the characteristic features of an AGN are present.

On the colour-selection side, repeating the Random Forest ranking once F115W is recovered and the F150W sample is enlarged, should refine the optimal selection of AGNs. The addition of X-Ray selected AGNs and MIRI photometry will reveal whether the pipeline misses heavily obscured nuclei or can be tuned to match the longer-wavelength IRAC wedges. A balanced batch strategy during training could also reduce the precision-recall tension seen in the current, highly imbalanced data.

Together, these steps would couple JWST's unique spatial resolution and wavelength leverage to a robust, data-driven classifier, ready for the far larger galaxy samples expected from future NIRCam programmes.

7 ACKNOWLEDGMENTS

A special thank you to my supervisors, Dr. Lingyu Wang and Dr. Berta Margalef-Bentabol, for their invaluable support and guidance throughout this thesis. It has been a pleasure working and learning with you.

Thank you to the whole team lead by Dr. Wang, for their advice and for being so welcoming. Specifically, thank you to the following people for providing essential resources used in this thesis:

- Antonio La Marca for the code and catalogue to execute the blank sky cutouts;
- Dr. Margalef-Bentabol for providing the reduced Illustris-TNG snapshots, the code to execute the AGN injection and the catalogue for the real cutouts;
- Mi Chen for the full catalogue containing flux calculations in the Spitzer channels, as well as in the different NIRCam bands.

I would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high-performance computing cluster.

I wish to acknowledge the use of the following AI-powered tools: Grammarly for support in validating the grammar correct use of the English language in this thesis and ChatGPT for syntax adjustments and code debugging.

8 REFERENCES

Ade, P. A., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Bartolo, N., & et al. (2016). Planck 2015 results. xiii. cosmological parameters. *Astronomy & Astrophysics*, *594*. https://doi.org/10.1051/0004-6361/201525830

Assef, R. J., Stern, D., Kochanek, C. S., Blain, A. W., Brodwin, M., Brown, M. J. I., Donoso, E., Eisenhardt, P. R. M., Jannuzi, B. T., Jarrett, T. H., Stanford, S. A., Tsai, C. .-., Wu, J., & Yan, L. (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. *The Astrophysical Journal*, *772*, 26. https://doi.org/10.1088/0004-637X/772/1/26

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., Earl, N., Starkman, N., Bradley, L., Shupe, D. L., Patil, A. A., Corrales, L., Brasseur, C. E., N"othe, M., Donath, A., Tollerud, E., Morris, B. M., Ginsburg, A., Vaher, E., Weaver, B. A., Tocknell, J., Jamieson, W., ... Astropy Project Contributors. (2022). The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *ApJ*, *935*(2), Article 167, 167. https://doi.org/10.3847/1538-4357/ac7c74

Barchi, P. H., de Carvalho, R. R., Rosa, R. R., Sautter, R. A., Soares-Santos, M., Marques, B. A. D., Clua, E., Gonçalves, T. S., de Sá-Freitas, C., & Moura, T. C. (2020). Machine and Deep Learning applied to galaxy morphology - A comparative study. *Astronomy and Computing*, *30*, Article 100334, 100334. https://doi.org/10.1016/j.ascom.2019.100334

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, *2*(1), 1–127. https://doi.org/10.1561/2200000006

Bertin, E. (2011). Automated morphometry with sextractor and psfex (I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, Eds.). *442*, 435.

Binney, J., & Merrifield, M. (1998). *Galactic astronomy*. Princeton University Press.

Bradley, L., Sipőcz, B., Robitaille, T., Tollerud, E., Vinícius, Z., Deil, C., Barbary, K., Wilson, T. J., Busko, I., Donath, A., Günther, H. M., Cara, M., Lim, P. L., Meßlinger, S., Burnett, Z., Conseil, S., Droettboom, M., Bostroem, A., Bray, E. M., ... Perren, G. (2024). Astropy/photutils: 1.12.0 [Accessed 2025-06-18]. https://doi.org/10.5281/zenodo.10967176

Casey, C. M., Kartaltepe, J. S., Drakos, N. E., Franco, M., Harish, S., Paquereau, L., Ilbert, O., Rose, C., Cox, I. G., Nightingale, J. W., Robertson, B. E., Silverman, J. D., Koekemoer, A. M., Massey, R., McCracken, H. J., Rhodes, J., Akins, H. B., Amvrosiadis, A., Arango-Toro, R. C., ... Zavala, J. A. (2023). Cosmos-web: An overview of the jwst cosmic origins survey. (arXiv:2211.07865). https://doi.org/10.48550/arXiv.2211.07865

Chang, Y.-Y., Le Floc'h, E., Juneau, S., da Cunha, E., Salvato, M., Civano, F., Marchesi, S., Ilbert, O., Toba, Y., Lim, C.-F., Tang, J.-J., Wang, W.-H., Ferraro, N., Urry, M. C., Griffiths, R. E., & Kartaltepe, J. S. (2017). Infrared selection of obscured active galactic nuclei in the cosmos field. *The Astrophysical Journal Supplement Series*, *233*, 19. https://doi.org/10.3847/1538-4365/aa97da

Chen, K.-J. (2025). Origins of supermassive black holes in galactic centers. (arXiv:2505.19039). https://doi.org/10.48550/arXiv.2505.19039

Cheng, T.-Y., Huertas-Company, M., Conselice, C. J., Aragón-Salamanca, A., Robertson, B. E., & Ramachandra, N. (2021). Beyond the hubble sequence - exploring galaxy morphology with unsupervised machine learning. *MNRAS*, *503*(3), 4446–4465. https://doi.org/10.1093/mnras/stab734

Chien, T. C. .-., Ling, C.-T., Goto, T., Wu, C. K. .-., Kim, S. J., Hashimoto, T., Lin, Y.-W., Kilerci, E., Ho, S. C. .-., Wang, P.-Y., & Raquel, B. J. R. (2024). Finding dusty agns

from the jwst ceers survey with mid-infrared photometry. *Monthly Notices of the Royal Astronomical Society*, *532*(1), 719–733. https://doi.org/10.1093/mnras/stae1550

COSMOS Collaboration. (2025). COSMOS–Web Unveils Largest Look Ever Into the Deep Universe. https://cosmos.astro.caltech.edu/news/87

Cosmos Team. (2022, January). *COSMOS2020 Farmer Catalog*. https://doi.org/10.26131/IRSA564

Dieleman, S., Willett, K. W., & Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *MNRAS*, *450*(2), 1441–1459. https://doi.org/10.1093/mnras/stv632

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. (2018). Improving galaxy morphologies for SDSS with Deep Learning. *MNRAS*, *476*(3), 3661–3676. https://doi.org/10.1093/mnras/sty338

Donley, J. L., Koekemoer, A. M., Brusa, M., Capak, P., Cardamone, C. N., Civano, F., Ilbert, O., Impey, C. D., Kartaltepe, J. S., Miyaji, T., et al. (2012). Identifying luminous active galactic nuclei in deep surveys: Revised irac selection criteria. *The Astrophysical Journal*, *748*(2), 142. https://doi.org/10.1088/0004-637X/748/2/142

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. https://arxiv.org/abs/2010.11929

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

González-Martín, O., Masegosa, J., Márquez, I., Rodríguez-Espinosa, J. M., Acosta-Pulido, J. A., Ramos Almeida, C., Dultzin, D., Hernández-García, L., Ruschel-Dutra, D., & Alonso-Herrero, A. (2015). Nuclear obscuration in liners. *Astronomy & Astrophysics*, *578*. https://doi.org/10.1051/0004-6361/201425254

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/cvpr.2016.90

Hiep, P., & Joo, R. (2018). A deep learning approach for classification of cloud image patches on small datasets. *Journal of Information and Communication Convergence Engineering*, *16*(3), 173–178.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., Pérez-González, P. G., Kartaltepe, J. S., Barro, G., Bernardi, M., Mei, S., Shankar, F., Dimauro, P., Bell, E. F., Kocevski, D., Koo, D. C., Faber, S. M., & Mcintosh, D. H. (2015). A Catalog of Visual-like Morphologies in the 5 CANDELS Fields Using Deep Learning. *ApJS*, *221*(1), Article 8, 8. https://doi.org/10.1088/0067-0049/221/1/8

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Inc., P. T. (2015). *Collaborative data science*. https://plot.ly

Kelly, T., Auge, C., Sanders, D., & Kirkpatrick, A. (2024). Accretion history of agn: A multiwavelength comparison of x-ray and infrared selected active galactic nuclei. *American Astronomical Society Meeting Abstracts #243*, *56*(2), Article 243.06, 243.06. https://ui.adsabs.harvard.edu/abs/2024AAS...24324306K

Kormendy, J., & Bender, R. (2011). Supermassive black holes do not correlate with dark matter haloes of galaxies. *Nature*, *469*(7330), 377–380. https://doi.org/10.1038/nature09695

Kron, R. G. (1980). Photometry of a complete sample of faint galaxies. *The Astrophysical Journal Supplement Series*, *43*, 305. https://doi.org/10.1086/190669

Lacy, M., Petric, A. O., Sajina, A., Canalizo, G., Storrie-Lombardi, L. J., Armus, L., Fadda, D., & Marleau, F. R. (2007). Optical spectroscopy and x-ray detections of a sample of quasars and active galactic nuclei selected in the mid-infrared from two *Spitzer Space Telescope* wide-area surveys. *The Astronomical Journal*, *133*(1), 186–205. https://doi.org/10.1086/509617

Lacy, M., Storrie-Lombardi, L. J., Sajina, A., Appleton, P. N., Armus, L., Chapman, S. C., Choi, P. I., Fadda, D., Fang, F., Frayer, D. T., Heinrichsen, I., Helou, G., Im, M., Marleau, F. R., Masci, F., Shupe, D. L., Soifer, B. T., Surace, J., Teplitz, H. I., ... Yan, L. (2004). Obscured and Unobscured Active Galactic Nuclei in the Spitzer Space Telescope First Look Survey. *ApJS*, *154*(1), 166–169. https://doi.org/10.1086/422816

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. (arXiv:2201.03545). https://doi.org/10.48550/arXiv.2201.03545

Lotz, J. M., Primack, J., & Madau, P. (2004). A new nonparametric approach to galaxy morphological classification. *The Astronomical Journal*, *128*(1), 163–182. https://doi.org/10.1086/421849

Marchesi, S., Civano, F., Elvis, M., Salvato, M., Brusa, M., Comastri, A., Gilli, R., Hasinger, G., Lanzuisi, G., Miyaji, T., Treister, E., Urry, C. M., Vignali, C., Zamorani, G., Allevato, V., Cappelluti, N., Cardamone, C., Finoguenov, A., Griffiths, R. E., ... Trakhtenbrot, B. (2016). The chandra cosmos legacy survey: Optical/ir identifications. *The Astrophysical Journal*, *817*, 34. https://doi.org/10.3847/0004-637X/817/1/34

Margalef-Bentabol, B., Wang, L., Marca, A. L., & Rodriguez-Gomez, V. (2025). Agn – host galaxy photometric decomposition using a fast, accurate and precise deep learning approach. (arXiv:2410.01437). https://doi.org/10.48550/arXiv.2410.01437

Marinacci, F., Vogelsberger, M., Pakmor, R., Torrey, P., Springel, V., Hernquist, L., Nelson, D., Weinberger, R., Pillepich, A., Naiman, J., et al. (2018). First results from the illustristng simulations: Radio haloes and magnetic fields. *Monthly Notices of the Royal Astronomical Society*, *480*(4), 5113–5139.

Mikulski Archive for Space Telescopes (MAST). (2024). Mast portal. https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html

Naiman, J. P., Pillepich, A., Springel, V., Ramirez-Ruiz, E., Torrey, P., Vogelsberger, M., Pakmor, R., Nelson, D., Marinacci, F., Hernquist, L., et al. (2018). First results from the illustristng simulations: A tale of two elements–chemical evolution of magnesium and europium. *Monthly Notices of the Royal Astronomical Society*, *477*(1), 1206–1224.

NASA/ESA/CSA/STScI. (2022). Webb telescope image 4177 — [insert official image title here] [Image credit: NASA, ESA, CSA, and STScI. Accessed 1 July 2025].

NASA/JPL-Caltech. (2012). Black holes - monsters in space [Image credit: NASA/JPL-Caltech. Accessed 27 June 2025]. https://d2pn8kiwq2w21t.cloudfront.net/images/jpegPIA16695.width-1600.jpg

Nelson, D., Pillepich, A., Springel, V., Weinberger, R., Hernquist, L., Pakmor, R., Genel, S., Torrey, P., Vogelsberger, M., Kauffmann, G., et al. (2018). First results from the illustristng simulations: The galaxy colour bimodality. *Monthly Notices of the Royal Astronomical Society*, *475*(1), 624–647.

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., & et al. (2019). The illustris tng sim-

ulations: Public data release. *Computational Astrophysics and Cosmology*, *6*(1). https://doi.org/10.1186/s40668-019-0028-x

Osterbrock, D. E. (1993). The nature and structure of active galactic nuclei. *The Astrophysical Journal*, *404*, 551. https://doi.org/10.1086/172307

Osterbrock, D. E., & Ferland, G. J. (2006). *Astrophysics of gaseous nebulae and active galactic nuclei* (2nd ed.). University Science Books.

Padovani, P., Alexander, D. M., Assef, R. J., De Marco, B., Giommi, P., Hickox, R. C., Richards, G. T., Smolčić, V., Hatziminaoglou, E., Mainieri, V., & Salvato, M. (2017). Active galactic nuclei: What's in a name? *The Astronomy and Astrophysics Review*, *25*(1). https://doi.org/10.1007/s00159-017-0102-9

pandas development team, T. (2020). *Pandas-dev/pandas: Pandas* (Version latest). Zenodo. https://doi.org/10.5281/zenodo.3509134

Peng, C. Y. (2010). *Galfit: User's manual and reference guide (version 3.0.4 anbd higher)*. Carnegie Observatories. https://users.obs.carnegiescience.edu/peng/work/galfit/README.pdf

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. (2002). Detailed structural decomposition of galaxy images. *The Astronomical Journal*, *124*(1), 266–293. https://doi.org/10.1086/340952

Peng, C. Y., Lilly, S. J., Kovač, K., Bolzonella, M., Pozzetti, L., Renzini, A., Zamorani, G., Ilbert, O., Knobel, C., Iovino, A., Maier, C., Cucciati, O., Tasca, L., Carollo, C. M., Silverman, J., Kampczyk, P., de Ravel, L., Sanders, D., Scoville, N., ... Scaramella, R. (2010). Mass and environment as drivers of galaxy evolution in sdss and zcosmos and the origin of the schechter function. *The Astrophysical Journal*, *721*(1), 193–221. https://doi.org/10.1088/0004-637x/721/1/193

Pillepich, A., Nelson, D., Hernquist, L., Springel, V., Pakmor, R., Torrey, P., Weinberger, R., Genel, S., Naiman, J. P., Marinacci, F., et al. (2018). First results from the illustristng simulations: The stellar mass content of groups and clusters of galaxies. *Monthly Notices of the Royal Astronomical Society*, *475*(1), 648–675.

Rieke, M. J., Kelly, D. M., Misselt, K., Stansberry, J., Boyer, M., Beatty, T., Egami, E., Florian, M., Greene, T. P., Hainline, K., Leisenring, J., Roellig, T., Schlawin, E., Sun, F., Tinnin, L., Williams, C. C., Willmer, C. N. A., Wilson, D., Clark, C. R., ... Young, E. T. (2023). Performance of NIRCam on JWST in Flight., *135*(1044), Article 028001, 028001. https://doi.org/10.1088/1538-3873/acac53

Rigby, J., Perrin, M., McElwain, M., Kimble, R., Friedman, S., Lallo, M., Doyon, R., Feinberg, L., Ferruit, P., Glasse, A., Rieke, M., Rieke, G., Wright, G., Willott, C., Colon, K., Milam, S., Neff, S., Stark, C., Valenti, J., ... Zonak, S. (2023). The science performance of jwst as characterized in commissioning. *Publications of the Astronomical Society of the Pacific*, *135*(1046), 048001. https://doi.org/10.1088/1538-3873/acb293

Seyfert, C. K. (1943). Nuclear emission in spiral nebulae. *The Astrophysical Journal*, *97*, 28. https://doi.org/10.1086/144488

Simard, L. (1998). GIM2D: An IRAF Package for the Quantitative Morphology Analysis of Distant Galaxies. In R. Albrecht, R. N. Hook, & H. A. Bushouse (Eds.), *Astronomical data analysis software and systems vii* (p. 108, Vol. 145). Astronomical Society of the Pacific. http://adsabs.harvard.edu/abs/1998ASPC..145..108S

Space Telescope Science Institute. (2024a). *Jwst background model* [JWST User Documentation]. Retrieved June 27, 2025, from https://jwst-docs.stsci.edu/jwst-general-support/jwst-background-model

Space Telescope Science Institute. (2024b). *Nircam point-spread functions* [JWST User Documentation]. Retrieved June 27, 2025, from https://jwst-docs.stsci.edu/jwst-near-infrared-camera/nircam-performance/nircam-point-spread-functions

Sparke, L. S., & Gallagher, J. S. (2010). *Galaxies in the universe: An introduction*. Cambridge Univ. Press.

Spitzer Science Center. (2008). IRAC Point Response Function (PRF) and Point Spread Function (PSF) Calibration Files. https://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/calibrationfiles/psfprf/

Springel, V. (2010). E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Monthly Notices of the Royal Astronomical Society*, *401*(2), 791–851. https://doi.org/10.1111/j.1365-2966.2009.15715.x

Springel, V., Pakmor, R., Pillepich, A., Weinberger, R., Nelson, D., Hernquist, L., Vogelsberger, M., Genel, S., Torrey, P., Marinacci, F., et al. (2018). First results from the illustristng simulations: Matter and galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, *475*(1), 676–698.

Stern, D., Assef, R. J., Benford, D. J., Blain, A., Cutri, R., Dey, A., Eisenhardt, P., Griffith, R. L., Jarrett, T. H., Lake, S., Masci, F., Petty, S., Stanford, S. A., Tsai, C.-W., Wright, E. L., Yan, L., Harrison, F., & Madsen, K. (2012). Mid-infrared selection of active galactic nuclei with the wide-field infrared survey explorer. i. characterizing wise-selected active galactic nuclei in cosmos. *The Astrophysical Journal*, *753*(1), 30. https://doi.org/10.1088/0004-637x/753/1/30

Stern, D., Eisenhardt, P., Gorjian, V., Kochanek, C. S., Caldwell, N., Eisenstein, D., Brodwin, M., Brown, M. J. I., Cool, R., Dey, A., Green, P., Jannuzi, B. T., Murray, S. S., Pahre, M. A., & Willner, S. P. (2005). Mid-infrared selection of active galaxies. *The Astrophysical Journal*, *631*(1), 163–168. https://doi.org/10.1086/432523

Tian, C., Urry, C. M., Ghosh, A., Nagai, D., Ananna, T. T., Powell, M. C., Auge, C., Mishra, A., Sanders, D. B., Cappelluti, N., & Schawinski, K. (2025). Automatic machine learning framework to study morphological parameters of agn host galaxies within $z < 1.4$ in the hyper supreme-cam wide survey. *The Astrophysical Journal*, *981*(1), 5. https://doi.org/10.3847/1538-4357/adaec0

Urry, C. M., & Padovani, P. (1995). Unified schemes for radio-loud active galactic nuclei. *Publications of the Astronomical Society of the Pacific*, *107*, 803. https://doi.org/10.1086/133630

Walmsley, M., Allen, C., Aussel, B., Bowles, M., Gregorowicz, K., Slijepcevic, I. V., Lintott, C. J., m. Scaife, A. M., Jabłońska, M., Karchev, K., Lanzieri, D., Mohan, D., O'Ryan, D., Saiguhan, B., Suárez, C., Guerra-Varas, N., & Velu, R. (2023). Zoobot: Adaptable deep learning models for galaxy morphology. *Journal of Open Source Software*, *8*(85), 5312. https://doi.org/10.21105/joss.05312

Walmsley, M., Masters, K. L., Willett, K. W., Smethurst, R. J., Lingard, T. E., et al. (2023). Galaxy Zoo DESI: Detailed morphology measurements for 8.7M galaxies in the DESI Legacy Imaging Surveys. *Monthly Notices of the Royal Astronomical Society*, *525*(1), 894–912. https://doi.org/10.1093/mnras/stad2596

Walmsley, M., Willett, K. W., Masters, K. L., Smethurst, R. J., Galloway, C. E., et al. (2022). Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314,000 galaxies. *Monthly Notices of the Royal Astronomical Society*, *514*(2), 2594–2613. https://doi.org/10.1093/mnras/stac1539

Weaver, J. R., Kauffmann, O. B., Ilbert, O., McCracken, H. J., Moneti, A., Toft, S., Brammer, G., Shuntov, M., Davidzon, I., Hsieh, B. C., Laigle, C., Anastasiou, A., Jespersen, C. K., Vinther, J., Capak, P., Casey, C. M., McPartland, C. J. R., Milvang-Jensen, B.,

Mobasher, B., ... Zamorani, G. (2022). COSMOS2020: A Panchromatic View of the Universe to z∼10 from Two Complementary Catalogs. *ApJS*, *258*(1), Article 11, 11. https://doi.org/10.3847/1538-4365/ac3078

Zhuang, M.-Y., Li, J., & Shen, Y. (2024). Active galactic nuclei and host galaxies in cosmos-web. i. nircam images, point-spread-function models and initial results on x-ray-selected broad-line agns at $0.35 < z < 3.5$. *The Astrophysical Journal*, *962*(1), 93. https://doi.org/10.3847/1538-4357/ad1517

Zhuang, M.-Y., & Shen, Y. (2023). Characterization of jwst nircam psfs and implications for agn+host image decomposition. (arXiv:2304.13776). https://doi.org/10.48550/arXiv.2304.13776

9 APPENDIX

9.1 Complete colour-colour mosaics

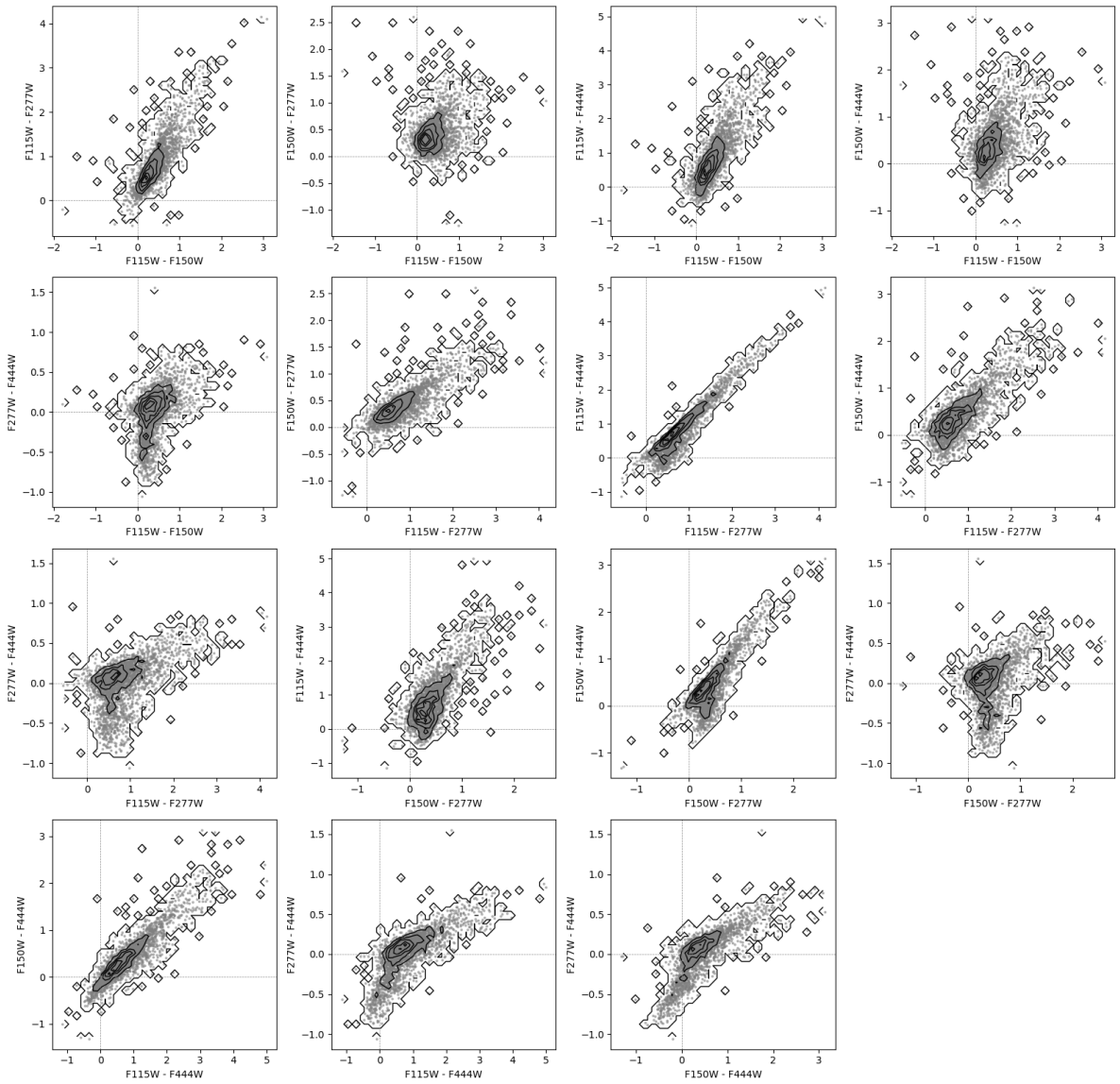

Figure 18: Fifteen-panel mosaic of pair-wise colour-colour diagrams in the four broad NIRCam bands F115W, F150W, F277W and F444W. Each panel plots one colour index against another (axes labelled), with individual sources shown by grey dots and a smoothed kernel-density contour set highlighting the regions of highest source density. The sequence reveals a tight, elongated locus running from the lower-left (bluest colours) toward the upper-right (reddest colours) in most projections, reflecting the strong mutual correlations among the four NIRCam bands. Dashed lines at zero colour mark the transition between blue- and red-dominated regimes.

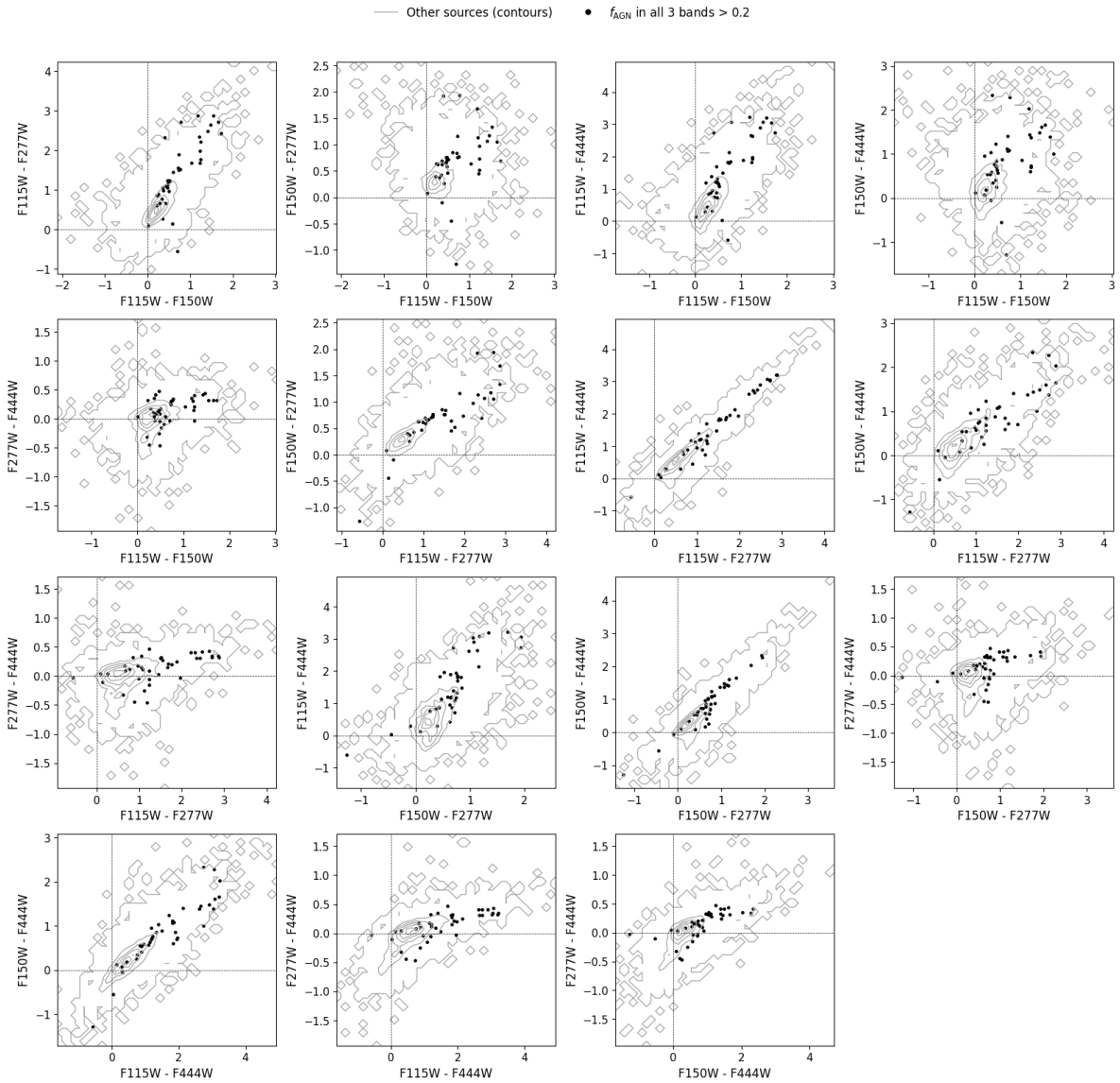

Figure 19: Fifteen colour-colour panels built from all unique pairs of the four broad NIRCam bands. Grey contours mark the locus of the full galaxy sample; black dots highlight the galaxies satisfying fAGN > 0.2 simultaneously in F115W, F150W, F277W and F444W. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye.

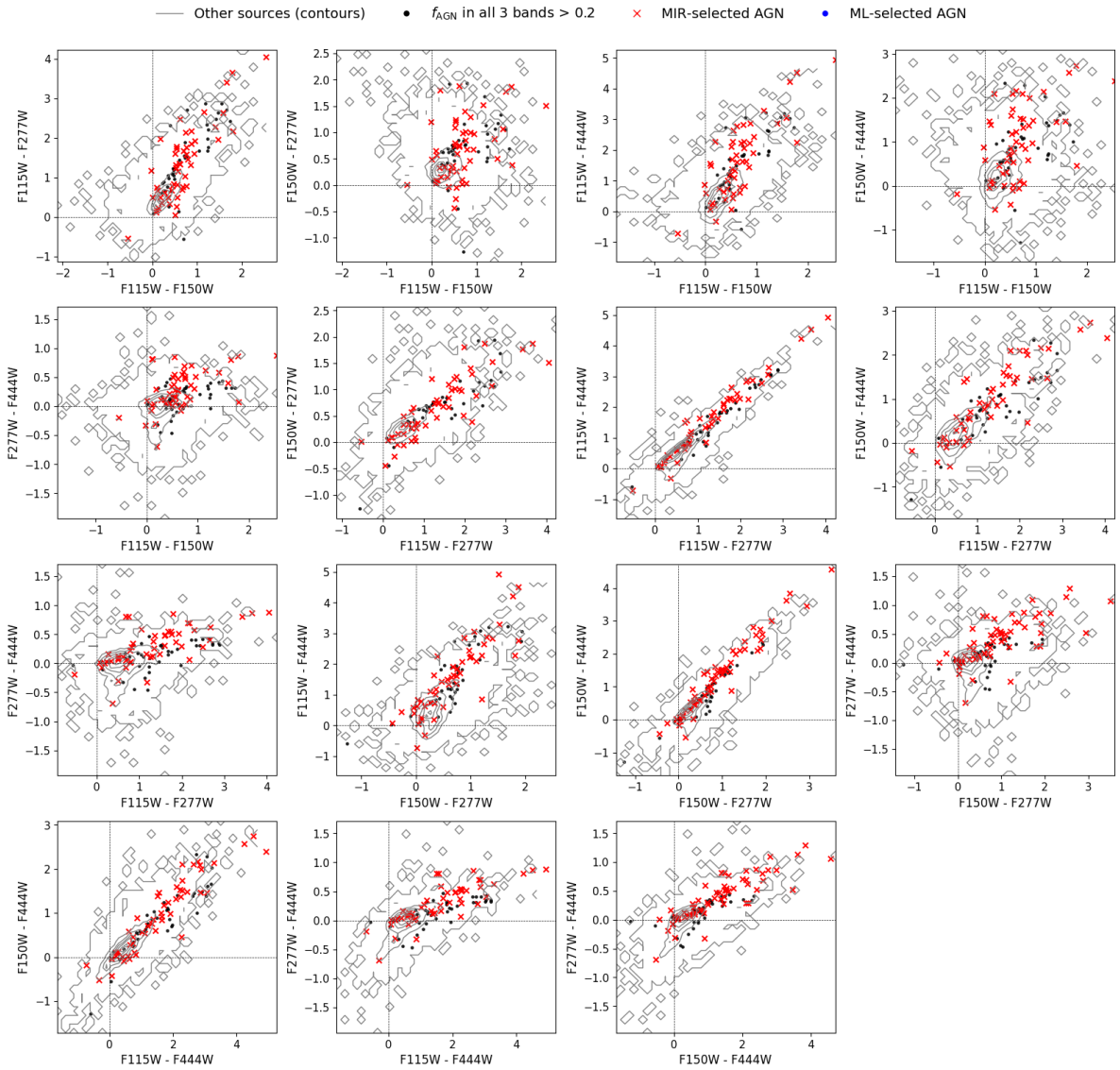

Figure 20: Fifteen colour-colour panels built from all unique pairs of the four broad NIRCam bands. Grey contours mark the locus of the full galaxy sample; black dots highlight the galaxies satisfying $f_{AGN} > 0.2$ simultaneously in F115W, F150W, F277W and F444W, while red crosses trace the MIR-selected AGNs. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. The two selections overlap almost exclusively on the redward (upper-right) branch of each panel, leaving the bluer arm practically empty.

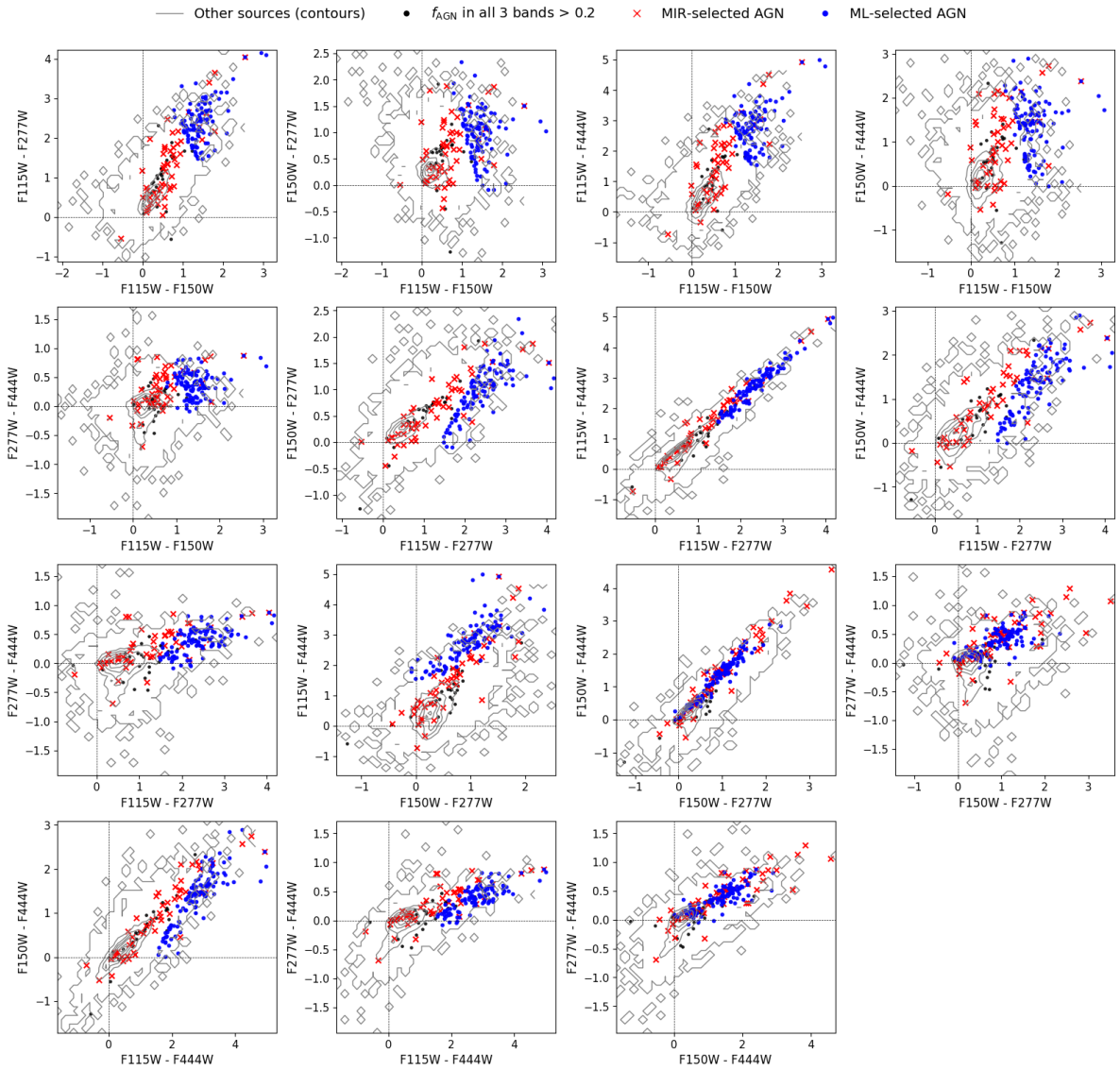

Figure 21: Fifteen colour-colour panels built from all unique pairs of the four broad NIRCam bands. Grey contours mark the locus of the full galaxy sample; black dots highlight the galaxies satisfying $f_{AGN} > 0.2$ simultaneously in F115W, F150W, F277W and F444W, while red crosses trace the MIR-selected AGNs and blue dots the ML candidates with $P_{\rm AGN} > 0.7$. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. The ML selection follows the same redward branch occupied by MIR AGNs but extends it to slightly bluer colours, recovering additional candidates while avoiding most of the blue. This overlap shows that the three optimised NIR colours alone encode sufficient information to reproduce and modestly enlarge the MIR selection.

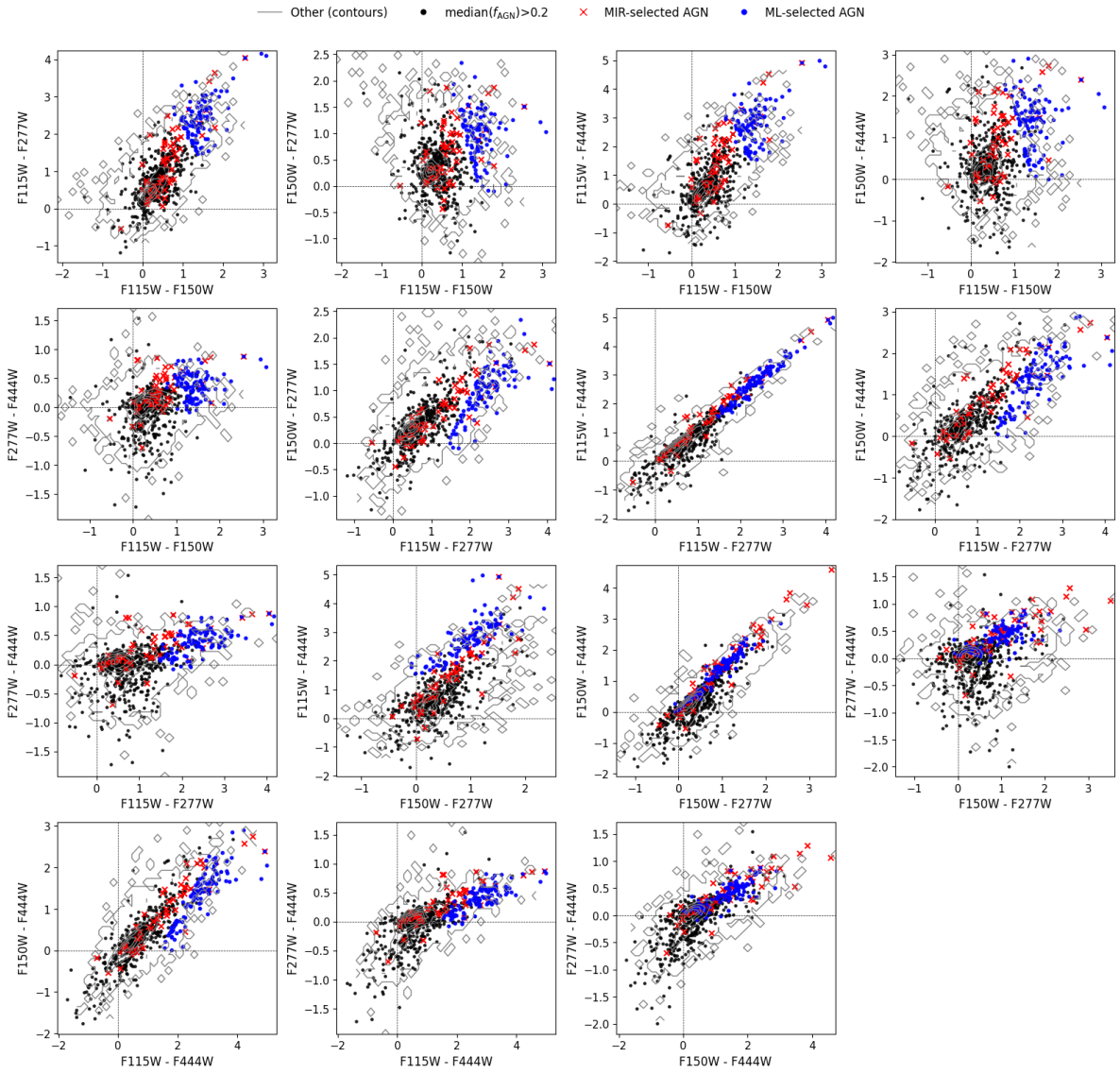

Figure 22: Fifteen colour-colour panels built from all unique pairs of the four broad NIRCam bands. Grey contours mark the locus of the full galaxy sample; black dots highlight the galaxies satisfying $median(f_{AGN}) > 0.2$ simultaneously in F115W, F150W, F277W and F444W, while red crosses trace the MIR-selected AGNs and blue dots the ML candidates with $P_{\rm AGN} > 0.7$. The single legend above the mosaic applies to every panel; dashed lines at zero colour guide the eye. The ML selection follows the same redward branch occupied by MIR AGNs but extends it to slightly bluer colours, recovering additional candidates while avoiding most of the blue. This overlap shows that the three optimised NIR colours alone encode sufficient information to reproduce and modestly enlarge the MIR selection. The median selection makes the graph look visually cluttered, hence the preference of using $f_{AGN} > 0.2$ in visualisations.

9.2 Supplementary colour-selection tables and figures

Table 6: Configuration and full metric set for the six ML trials (part 1/2).

	Trial 1	**Trial 2**	**Trial 3**
Description	4 colours(+F115W)	3 colours (+F115W)	3 colours (-F115W, MIR feature)
# colours	4	3	3
F115W included	Yes	Yes	No
MIR flag used	No	No	Yes
f_{AGN} criterion	All > 0.2	All > 0.2	Median > 0.2
Threshold (τ)	0.50	0.50	0.50
Base (no extra MIR columns)			
Accuracy	0.766	0.750	0.676
Precision	0.037	0.037	0.080
Recall	0.575	0.650	0.514
F1 score	0.070	0.071	0.138
With MIR feature			
Accuracy	-	-	0.674
Precision	-	-	0.079
Recall	-	-	0.514
F1 score	-	-	0.137
ML / MIR			
Accuracy	-	-	0.672
Precision	-	-	0.080
Recall	-	-	0.522
F1 score	-	-	0.138
ML & MIR			
Accuracy	-	-	0.941
Precision	-	-	0.037
Recall	-	-	0.007
F1 score	-	-	0.012

Table 7: Configuration and full metric set for the six ML trials (part 2/2).

	Trial 4	**Trial 5**	**Trial 6**
Description	3 colours (+F115W, MIR)	+F115W, MIR (τ=0.70)	+F115W, MIR (τ=0.80)
# colours	3	3	3
F115W included	Yes	Yes	Yes
MIR flag used	Yes	Yes	Yes
f_{AGN} criterion	Median > 0.2	Median > 0.2	Median > 0.2
Threshold (τ)	0.50	0.70	0.80
Base			
Accuracy	0.700	0.925	0.943
Precision	0.092	0.204	0.276
Recall	0.543	0.152	0.058
F1 score	0.158	0.174	0.096
With MIR feature			
Accuracy	0.699	0.925	0.943
Precision	0.092	0.194	0.276
Recall	0.543	0.145	0.058
F1 score	0.158	0.166	0.096
ML / MIR			
Accuracy	0.695	0.915	0.932
Precision	0.092	0.169	0.156
Recall	0.551	0.168	0.072
F1 score	0.157	0.168	0.099
ML & MIR			
Accuracy	0.941	-	-
Precision	0.045	-	-
Recall	0.007	-	-
F1 score	0.013	-	-

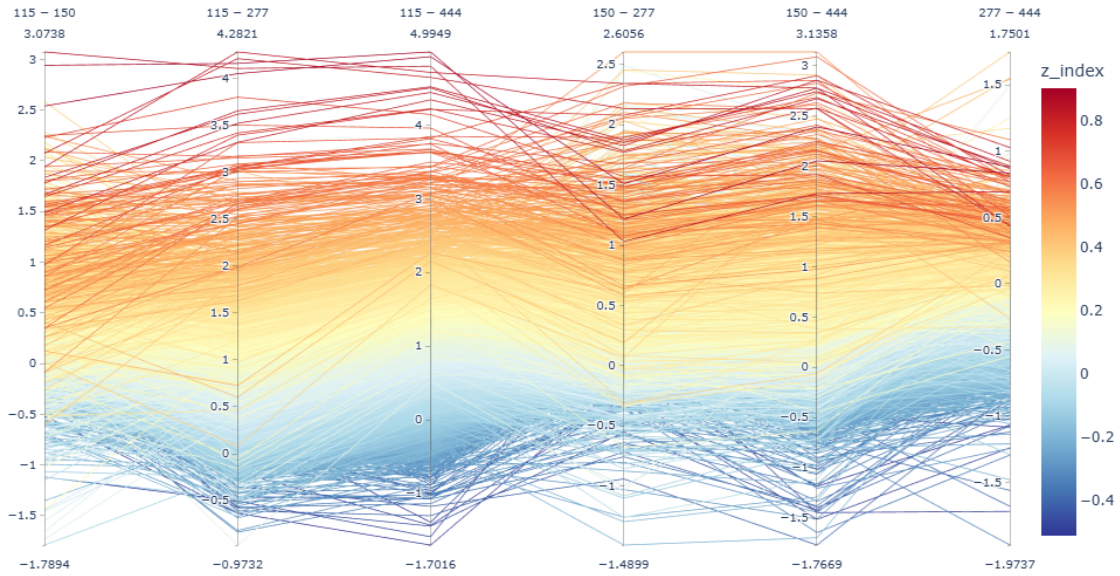

Figure 23: Parallel-coordinate diagram for all four NIRCam broad-band indices. Each poly-line represents one galaxy and is coloured by the logistic-regression equation expressed in Eq. 10. The biggest spread is in all the combinations involving F115W, while the smallest is in the F150W-F277W combination, reflecting the coefficients of the equation used.

9.3 Code produced

The original codes produced for this work (colour analysis and visualisations of test results and colour-colour diagrams), the first catalogue with the merged fluxes and the end catalogue used for the ML pipeline, as well as the predictions in the F277W and F444W filters can be found by going to this GitHub repository: https://github.com/nadiapan4/thesis-codes.