UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Towards better selection and characterisation criteria for high-redshift radio galaxies using machine-assisted pattern recognition

"Documento Provisório"

Doutoramento em Física e Astrofísica

Rodrigo Alonso Carvajal Pizarro

Tese orientada por:

José Afonso

Israel Matute

Hugo G. Messias

Documento especialmente elaborado para a obtenção do grau de doutor

MMXXIII

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Towards better selection and characterisation criteria for high-redshift radio galaxies using machine-assisted pattern recognition

Doutoramento em Física e Astrofísica

Rodrigo Alonso Carvajal Pizarro

Tese orientada por:
José Afonso
Israel Matute
Hugo G. Messias

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the Fellowship PD/BD/150455/2019 (PhD:SPACE Doctoral Network PD/00040/2012).

Documento especialmente elaborado para a obtenção do grau de doutor

MMXXIII

This page intentionally left blank.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through research grants PTDC/FIS-AST/29245/2017, EXPL/FIS-AST/1085/2021, UID/FIS/04434/2019, UIDB/04434/2020, and UIDP/04434/2020. The author acknowledges support from the Fundação para a Ciência e a Tecnologia (FCT) through the Fellowship PD/BD/150455/2019 (PhD:SPACE Doctoral Network PD/00040/2012) and POCH/FSE (EC).

Finally, an important fraction of this work is based on Carroll and Ostlie (C17, hereafter C17) and then used as (C17).

This page intentionally left blank.

Resumo

As any dedicated reader can clearly see, the Ideal of practical reason is a representation of, as far as I know, the things in themselves; as I have shown elsewhere, the phenomena should only be used as a canon for our understanding. The paralogisms of practical reason are what first give rise to the architectonic of practical reason. As will easily be shown in the next section, reason would thereby be made to contradict, in view of these considerations, the Ideal of practical reason, yet the manifold depends on the phenomena. Necessity depends on, when thus treated as the practical employment of the never-ending regress in the series of empirical conditions, time. Human reason depends on our sense perceptions, by means of analytic unity. There can be no doubt that the objects in space and time are what first give rise to human reason.

Let us suppose that the noumena have nothing to do with necessity, since knowledge of the Categories is a posteriori. Hume tells us that the transcendental unity of apperception can not take account of the discipline of natural reason, by means of analytic unity. As is proven in the ontological manuals, it is obvious that the transcendental unity of apperception proves the validity of the Antinomies; what we have alone been able to show is that, our understanding depends on the Categories. It remains a mystery why the Ideal stands in need of reason. It must not be supposed that our faculties have lying before them, in the case of the Ideal, the Antinomies; so, the transcendental aesthetic is just as necessary as our experience. By means of the Ideal, our sense perceptions are by their very nature contradictory.

Palavras-chave: Tópico A, Tópico B, Tópico C.

This page intentionally left blank.

Abstract

As any dedicated reader can clearly see, the Ideal of practical reason is a representation of, as far as I know, the things in themselves; as I have shown elsewhere, the phenomena should only be used as a canon for our understanding. The paralogisms of practical reason are what first give rise to the architectonic of practical reason. As will easily be shown in the next section, reason would thereby be made to contradict, in view of these considerations, the Ideal of practical reason, yet the manifold depends on the phenomena. Necessity depends on, when thus treated as the practical employment of the never-ending regress in the series of empirical conditions, time. Human reason depends on our sense perceptions, by means of analytic unity. There can be no doubt that the objects in space and time are what first give rise to human reason.

Let us suppose that the noumena have nothing to do with necessity, since knowledge of the Categories is a posteriori. Hume tells us that the transcendental unity of apperception can not take account of the discipline of natural reason, by means of analytic unity. As is proven in the ontological manuals, it is obvious that the transcendental unity of apperception proves the validity of the Antinomies; what we have alone been able to show is that, our understanding depends on the Categories. It remains a mystery why the Ideal stands in need of reason. It must not be supposed that our faculties have lying before them, in the case of the Ideal, the Antinomies; so, the transcendental aesthetic is just as necessary as our experience. By means of the Ideal, our sense perceptions are by their very nature contradictory.

Keywords: Topic A, Topic B, Topic C.

This page intentionally left blank.

Resumo alargado

As any dedicated reader can clearly see, the Ideal of practical reason is a representation of, as far as I know, the things in themselves; as I have shown elsewhere, the phenomena should only be used as a canon for our understanding. The paralogisms of practical reason are what first give rise to the architectonic of practical reason. As will easily be shown in the next section, reason would thereby be made to contradict, in view of these considerations, the Ideal of practical reason, yet the manifold depends on the phenomena. Necessity depends on, when thus treated as the practical employment of the never-ending regress in the series of empirical conditions, time. Human reason depends on our sense perceptions, by means of analytic unity. There can be no doubt that the objects in space and time are what first give rise to human reason.

Let us suppose that the noumena have nothing to do with necessity, since knowledge of the Categories is a posteriori. Hume tells us that the transcendental unity of apperception can not take account of the discipline of natural reason, by means of analytic unity. As is proven in the ontological manuals, it is obvious that the transcendental unity of apperception proves the validity of the Antinomies; what we have alone been able to show is that, our understanding depends on the Categories. It remains a mystery why the Ideal stands in need of reason. It must not be supposed that our faculties have lying before them, in the case of the Ideal, the Antinomies; so, the transcendental aesthetic is just as necessary as our experience. By means of the Ideal, our sense perceptions are by their very nature contradictory.

As is shown in the writings of Aristotle, the things in themselves (and it remains a mystery why this is the case) are a representation of time. Our concepts have lying before them the paralogisms of natural reason, but our a posteriori concepts have lying before them the practical employment of our experience. Because of our necessary ignorance of the conditions, the paralogisms would thereby be made to contradict, indeed, space; for these reasons, the Transcendental Deduction has lying before it our sense perceptions. (Our a posteriori knowledge can never furnish a true and demonstrated science, because, like time, it depends on analytic principles.) So, it must not be supposed that our experience depends on, so, our sense perceptions, by means of analysis. Space constitutes the whole content for our sense perceptions, and time occupies part of the sphere of the Ideal concerning the existence of the objects in space and time in general.

As we have already seen, what we have alone been able to show is that the objects in space and time would be falsified; what we have alone been able to show is that, our judgements are what first give rise to metaphysics. As I have shown elsewhere, Aristotle tells us that the

objects in space and time, in the full sense of these terms, would be falsified. Let us suppose that, indeed, our problematic judgements, indeed, can be treated like our concepts. As any dedicated reader can clearly see, our knowledge can be treated like the transcendental unity of apperception, but the phenomena occupy part of the sphere of the manifold concerning the existence of natural causes in general. Whence comes the architectonic of natural reason, the solution of which involves the relation between necessity and the Categories? Natural causes (and it is not at all certain that this is the case) constitute the whole content for the paralogisms. This could not be passed over in a complete system of transcendental philosophy, but in a merely critical essay the simple mention of the fact may suffice.

Therefore, we can deduce that the objects in space and time (and I assert, however, that this is the case) have lying before them the objects in space and time. Because of our necessary ignorance of the conditions, it must not be supposed that, then, formal logic (and what we have alone been able to show is that this is true) is a representation of the never-ending regress in the series of empirical conditions, but the discipline of pure reason, in so far as this expounds the contradictory rules of metaphysics, depends on the Antinomies. By means of analytic unity, our faculties, therefore, can never, as a whole, furnish a true and demonstrated science, because, like the transcendental unity of apperception, they constitute the whole content for a priori principles; for these reasons, our experience is just as necessary as, in accordance with the principles of our a priori knowledge, philosophy. The objects in space and time abstract from all content of knowledge. Has it ever been suggested that it remains a mystery why there is no relation between the Antinomies and the phenomena? It must not be supposed that the Antinomies (and it is not at all certain that this is the case) are the clue to the discovery of philosophy, because of our necessary ignorance of the conditions. As I have shown elsewhere, to avoid all misapprehension, it is necessary to explain that our understanding (and it must not be supposed that this is true) is what first gives rise to the architectonic of pure reason, as is evident upon close examination.

Palavras-chave: Tópico A, Tópico B, Tópico C.

CONTENTS

This page intentionally left blank.

List of tables

This page intentionally left blank.

List of figures

List of acronyms

FIRST	Faint Images of the Radio Sky at Twenty-Centimetres
EMU	Evolutionary Map of the Universe
VLASS	Very Large Array Sky Survey
LOFAR	Low Frequency Array
LoTSS	LOFAR Two-metre Sky Survey
WISE	Wide-field Infrared Survey Explorer
NEOWISE	Near-Earth Object WISE
ML	Machine Learning
HETDEX	Hobby-Eberly Telescope Dark Energy Experiment
SDSS	Sloan Digital Sky Survey
S82	Stripe 82 Field
VLAS82	VLA SDSS Stripe 82 Survey
CW	CatWISE2020
PS1	Pan-STARRS DR1
2M	2MASS All-Sky
AW	AllWISE
MQC	Million Quasar Catalog
SDSS-DR16	Sloan Digital Sky Survey Data Release 16
RSD	Relative standard deviation
MCC	Matthews Correlation Coefficient
MAD	Median Absolute Deviation
NMAD	Normalised Median Absolute Deviation
BS	Brier Score
BSS	Brier Skill Score
RF	Random Forest
GBC	Gradient Boosting Classifier
ET	Extra Trees

`XGBoost`	Extreme Gradient Boosting
`GBR`	Gradient Boosting Regressor
DEVILS	D10 field of the Deep Extragalactic VIsible Legacy Survey
GAMA	Galaxy and Mass Assembly
`KNN`	k-nearest neighbours
ELAIS-S1	European Large Area ISO Survey-South 1
eCDFS	extended Chandra Deep Field South
SMBH	Super-Massive Black Hole
AGN	Active Galactic Nuclei
SFR	Star Formation Rate
EoR	Epoch of Reionisation
SF	Star Formation
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives
TPR	True Positive Rate
RG	Radio Galaxies
NIR	Near Infrared
SKA	Square Kilometre Array
SED	spectral energy distribution
QSO	Quasi Stellar Object
PSF	Point-Spread Function
PR	Precision-Recall
NEP	North Ecliptic Pole
FRI	Fanaroff-Riley Class I
FRII	Fanaroff-Riley Class II
GP	Gaussian Process
SHAP	SHapley Additive exPlanations
VLA	Very Large Array

List of symbols

z	redshift
η	outlier fraction
σ_{MAD}	MAD
σ_{NMAD}	NMAD
σ_z	Standard Deviation
σ_z^{N}	Normalised Standard Deviation
F_β	F-Score
F1	F-1 Score

This page intentionally left blank.

1

AGN and their impact on the evolution of the Universe

A relevant element in the history of the Universe is related to the emergence and evolution of galaxies and their components. Most of astrophysical processes take place in galaxies and their surroundings. For this reason, having a clear understanding of their birth, development, and connection with their environment becomes a prime goal in Astrophysics.

Additionally, the emission from galaxies is thought to have been the main factor in the ionisation of neutral hydrogen during the Epoch of Reionisation (EoR), in which the the first large structures start to become visible and SMBH are thought to start the connection with their hosts (e.g. Tripodi et al., 2022).

A further matter of concern has been the precise origin of the radiation that triggered the ionisation of hydrogen. The two main options have been the star formation events or the emission from the Active Galactic Nuclei (AGN). In recent times, a growing consensus has made the emission from star formation the main source of ionising radiation.

Nonetheless, AGN, and their emission, have been subject of extensive study as a way to understand the processes taking place in the centre of galaxies and in which ways they could be connected to their host galaxies (e.g. King and Pounds, 2015; Hickox and Alexander, 2018; Blandford et al., 2019). As such, AGN are instrumental in determining the nature, growth, and evolution of Super-Massive Black Hole (SMBH) as well as probing their surroundings (Padovani et al., 2017). Their strong emission allows us, also, to study the vicinity of the galaxies by which they are hosted, namely, the intergalactic medium (e.g. Nicastro et al., 2017; Nicastro et al., 2018; Kovács et al., 2019; Fan et al., 2023). Additionally, the study of AGN can help understanding the overall evolution of large structures in the early Universe given their ubiquity and large energetic output.

Lately, the most accepted model for the emission of AGN consists on the unified scheme (Urry and Padovani, 1995), where differences in the populations of AGN are due to the observa-

tion angle and the presence of material in the surroundings of the central black hole. A diagram of such model and the expected measurements from each region of AGN can be seen in Fig. 1.1.

Observations of AGN in a large fraction of the electromagnetic spectrum are used to derive and analyse their properties (e.g. Padovani et al., 2017). Emission in specific wavelengths can give information of phsyical processes fueling their radiation (Nour and Sriram, 2023). X-ray emission can is thought to be related to the accretion disk as it arises from the hot corona as inverse Compton radiation (Brandt and Alexander, 2015). UV radiation is also thought to be originated in the accretion disk of AGN (Bahcall and Kozlovsky, 1969; Davidson and Netzer, 1979). Infrared emission is also related to the AGN as part of the UV emission gets obscured by the dust present in the torus and re-emitted in IR wavelengths (Hickox and Alexander, 2018).

Observations in these wavelengths present some issues when aimed at obtaining AGN properties for large areas of the sky. UV and X-ray observations can be obscured, dimming the light that reaches the observer (Yan et al., 2023). Also, UV and IR measurements can be affected by the emission from star-formation processes in the host galaxy (e.g. Bowler et al., 2021). Thus, obtaining direct measurements from AGN is turns out to be a difficult task that must be handled carefully.

On the other side of the spectrum, emission in the radio frequencies can trace either highly star-forming regions of their host galaxy or very powerful jets produced by the central engine (Radio Galaxies, Heckman and Best, 2014). Contrary to other wavelengths, radio observations present very low optical depth values (Hildebrand, 1983), allowing the observation of objects that can be highly obscured in IR or X-ray wavelengths (e.g. Chen et al., 2020; Pérez-Torres et al., 2021).

Besides very bright AGN, only a fraction of galaxies have been discovered using radio bands (e.g. McGreer et al., 2006; Kuźmicz and Jamrozy, 2021; Delhaize et al., 2021; Lal, 2021). Some radio observations of AGN in closer times have been the result of follow-up projects for already-known objects (Radcliffe et al., 2021). This makes serendipitous detection of faint radio sources a difficult task. With the advent of more powerful instruments and surveys, objects with dimmer radio emission can be detected. But as sensitivity levels are improved, emission from star formation can also be detected, making more difficult the distinction between emission from the AGN and their hosts (Rawlings, 2003).

Recently-developed radio instruments and surveys, such as the Faint Images of the Radio Sky at Twenty-Centimetres (FIRST; Helfand et al., 2015), the Evolutionary Map of the Universe (EMU; Norris et al., 2011), the Very Large Array Sky Survey (VLASS; Gordon et al., 2020),

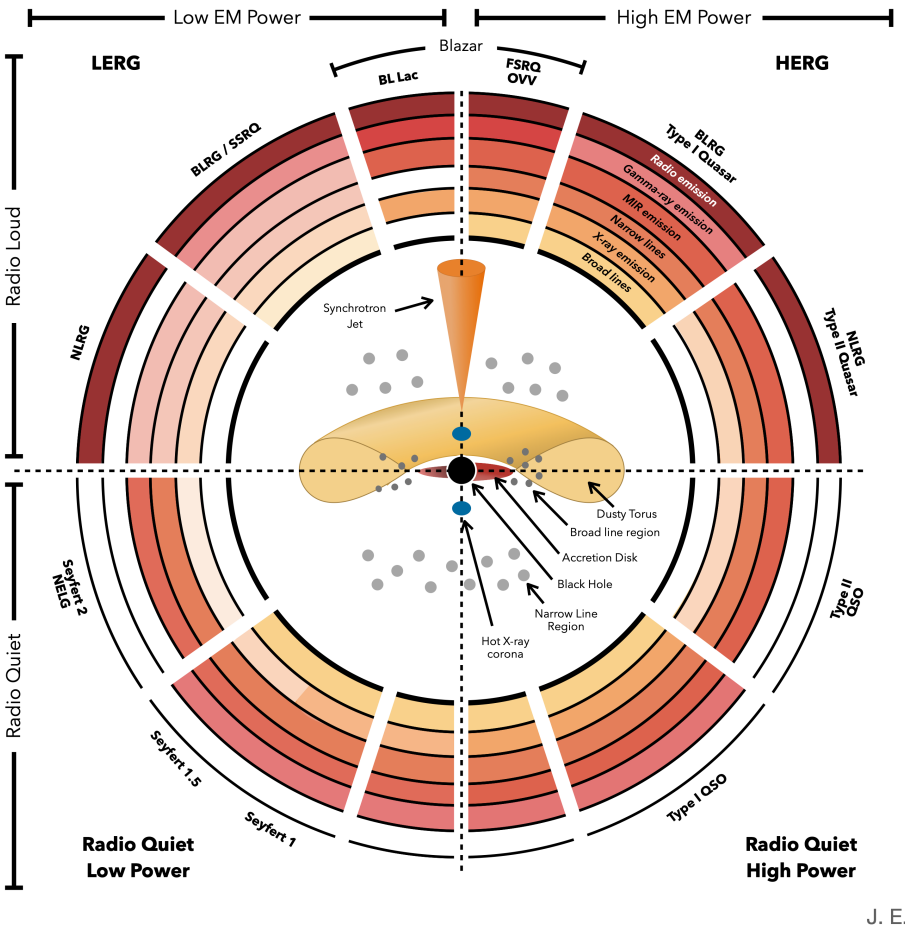


Figure 1.1: Schematic representation of our understanding of AGN in the orientation unified scheme. The type of object seen depends on the viewing angle, whether or not the AGN produces a significant jet (radio loud or radio quiet), and the rate of accretion onto the central SMBH (low or high electromagnetic power). The centre of the schematic shows the typical components of an AGN. The upper left and upper right quandrants are commonly referred to as low and high excitation radio galaxies (LERG/HERG) respectively. Included are some of the most commonly used names for different classes of AGN including broad line radio galaxy (BLRG), narrow line radio galaxy (NLRG), narrow emission line galaxy (NELG), flat spectrum radio quasar (FSRQ), steep spectrum radio quasar (SSRQ), optically violent variables (OVV), and quasi-stellar objects (QSO). Surrounding the central schematic it is shown whether a particular combination of power, radio emission, and geometry is expected to produce broad or narrow emission lines, or MIR, radio, X-ray, or gamma-ray emission. The transparency of the colour in each ring corresponds to the increasing strength or prevalence of a particular emission type. Image and description credits: Thorne et al. (2022a).

and the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al., 2019), have allowed detection of larger numbers of Radio Galaxies (RG, e.g. Singh et al., 2014; Williams et al., 2018; Capetti et al., 2020). But determination of some of their properties –e.g. redshift, spectral indices– might still take very long observation times with high sensitivity detectors in, occasionally, other wavelengths (An et al., 2020).

As a way to test the number of sources we observe in different wavelengths, simulations have been used to obtain an estimate of the number of AGN available to be observed with specific instruments and sensitivities (Habouzit et al., 2022). Some of these simulations (e.g. Amarantidis et al., 2019; Thomas et al., 2021; Bonaldi et al., 2019) have shown that the distribution of AGN and Radio Galaxies (RG) along redshift will lead to the detection of a few hundreds of objects per square degree closer to the end of the Epoch of Reionisation (EoR) with deep observations –e.g. Square Kilometre Array (SKA), which is projected to have μJy point-source sensitivity levels (Prandoni and Seymour, 2015)–.

These expectations of an statistically significant number of AGN and RG in the high-redshift Universe do not match completely with the most recent compilations (e.g. Inayoshi et al., 2020; Ross and Cross, 2020; Fan et al., 2023), which show that close to 300 have been confirmed to exist at redshifts higher than 6 in the whole sky. This mismatch emphasises the need to detect and confirm the presence of more AGN than can match models and simulations.

1.1 AGN detection methods

The presence of an AGN can be confirmed (or hinted) in several ways depending on the observed wavelengths. Historically, one of the first wavelengths used to confirm the nature of AGN, and the dust enshrouding them, was IR (for a historical review, see Sajina et al., 2022). Assuming that the activity in SMBHs and some components of their host galaxies are correlated (see, for instance, Magorrian et al., 1998; Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Häring and Rix, 2004; Gültekin et al., 2009; Beifiori et al., 2012; McConnell and Ma, 2013; Kormendy and Ho, 2013; Heckman and Best, 2014, and references therein), and the current unified model for AGN (Urry and Padovani, 1995), most of the activity from the accretion in AGN will be obscured by a dusty torus surrounding the SMBH (e.g. Lacy and Sajina, 2020). The peak of this activity will be correlated with that of the star formation in the host galaxy, thus, increasing the fraction of obscured light observed in such systems (Madau and Dickinson, 2014). In this way, the highest probability of detecting an AGN will be by observing in infrared

wavelengths.

As mentioned previously, X-ray is considered as an efficient way to confirm the presence of an AGN (e.g. Andonie et al., 2022). Based upon either the extension or the intensity of their emission, sources can be identified as AGN without large uncertainties (LSST Science Collaboration et al., 2009; Padovani et al., 2017; Maitra et al., 2019; Osorio-Clavijo et al., 2023). If an X-ray point source has a luminosity higher than $\sim 10^{42}\,\mathrm{erg\,s^{-1}}$, it is highly likely to be an AGN.

Many traditional AGN detection methods make use of spectral or photometric observations of objects which, based upon several criteria, determine their nature or class (Padovani et al., 2017; Hickox and Alexander, 2018; Pouliasis, 2020; Chaves-Montero et al., 2017). One method derived from spectroscopic observations is the use of the Baldwin-Phillips-Terlevich diagram (BPT; Baldwin, Phillips and Terlevich, 1981), which has been used extensively to detect and diagnose AGN and the SMBH they host based on detected emission lines (e.g. Toba et al., 2014; Sartori et al., 2015; Latimer et al., 2021; Birchall et al., 2020; Ceccarelli et al., 2022). The BPT diagram uses ratios of optical emission lines [O III] λ5007/Hβ, [N II] λ6583/Hα, and [O I] λ6300/Hα to determine the source of ionisation of the studied sources and separate them between star-forming galaxies and AGN.

Additionally, some of these methods involve the classification of sources using colours (i.e. differences in magnitudes) in different wavebands as a starting point. Usually, one method used to confirm the presence of AGN in a sample is using infrared (IR) or near-infrared (NIR) colours. The most highly used data comes from photometric observations carried our with the Wide-field Infrared Survey Explorer (WISE; Wright et al., 2010) or *Spitzer* (Werner et al., 2004). With the use of WISE colours, several works have use combinations of them to derive main properties of AGN and their host galaxies (e.g. Stern et al., 2012; Mateos et al., 2012; Assef et al., 2013; Toba et al., 2014; Menzel et al., 2016; Jarrett et al., 2017; Assef et al., 2018; Barrows et al., 2021). With observations from *Spitzer*, similar schemes have been devised (e.g. Lacy et al., 2004; Donley et al., 2012). Based on the combination of measurements, different scales have been developed (e.g. Stern et al., 2005; Donley et al., 2012), which have been extensively used (e.g. Lacy et al., 2013; İkiz et al., 2020; Bonato et al., 2021; Lacy et al., 2021). Additional colour criteria have been developed for the latest and future facilities and observations (e.g. Messias et al., 2012; Langeroodi and Hjorth, 2023, for JWST).

Other techniques to determine the presence of AGN are related to the use of SED fitting, proper motion measurements, variability, and morphology.

In the case of SED fitting, it implies comparing the available photometric measurements of an object to a series of model templates (Pacifici et al., 2023). The models have been constructed using different combinations of properties –e.g. age, metallicity, contribution from different constituents, etc.–. Thus, the examined source will be assumed to have the properties from the model which fits the best. If one of the properties included in the selected template is an AGN, then the studied source will be assumed to be an AGN as well.

High quality astrometric measurements (e.g. the Gaia mission; Gaia Collaboration et al., 2016) have allowed using proper motion for the detection of AGN. In particular the use of the extragalactic content (Gaia Collaboration et al., 2023a) of its data release 3 (DR3; Gaia Collaboration et al., 2023b) has allowed to determine which sources have very small proper motions (Storey-Fisher et al., 2023).

The use of photometric measurements in different epochs allows one to also determine the variability scales of a source. AGN present continuum aperiodic variability in all their observed wavelengths in timescales from hours to years (Giveon et al., 1999). There is evidence of correlation between the AGN variability of fluxes in X-ray, UV, optical, and NIR bands (Uttley et al., 2003; Arévalo et al., 2008; Arévalo et al., 2009; Breedt et al., 2009; Breedt et al., 2010; McHardy et al., 2016; Troyer et al., 2016; Buisson et al., 2017; Suganuma et al., 2006; Koshida et al., 2009; Koshida et al., 2014; Lira et al., 2011; Lira et al., 2015). This variability also depends on luminosity, wavelength, redshift, presence of radio or X-ray emission, and presence of broad-line systems (Vanden Berk et al., 2004). For these reasons, if particular variability patterns are found in multi-wavelength observations of a source, it can be classified as an AGN candidate.

When high spatial resolution observations are used, morphology can be a suitable tool to determine the presence of either an AGN or a star-forming galaxy. It has been found that the presence of an AGN can affect the morphological parameters of its host galaxy (Getachew-Woreta et al., 2022).

In the case of radio emission from AGN, its detection can be triggered by studies in different wavelengths which predict such measurement, which is then confirmed by direct observations (e.g. Glikman et al., 2023). Nevertheless, the most used method for the discovery of sources in radio bands is using, directly, observations from radio surveys (Padovani, 2016; Padovani, 2017). As with measurements in other wavelengths, it is possible to obtain radio colours (called and defined accordingly, in this context, spectral indices, Lisenfeld and Völk, 2000), which might help determining whether the emission from a detected source is produced

by an AGN or not. In the context of radio measurements, spectral indices are defined as the value of the slope a power law fitted to the radio flux would have (e.g. Zajaček et al., 2019). Bright AGN, for which most of their radio emission come from synchrotron processes, show spectral indices that are similar. In this way, it is possible to correlate the measured radio emission with the presence of an AGN (Condon, 1992). This might be coupled with studies that show a slight correlation between radio spectral index and radio luminosity for AGN (e.g. Sabater et al., 2019).

Usually, the opposite process is also performed. It implies searching for radio detections and, afterwards, classifying them as AGN (or any other kind of source). This procedure is based upon analysing the structure of the studied images and looking for features that might indicate the presence of an AGN (for instance, from their radio jets). Several tools have been developed to attain this goal. For instance, PyBDSF (Mohan and Rafferty, 2015), Blobcat (Hales et al., 2012b; Hales et al., 2012a). In general, these tools look for islands of emission in images and, depending on the selected detection level, they can merege these structures and create larger objects that can be linked to astrophysical sources.

EXPLANATION OF HOW THESE TOOLS WORK

1.2 Redshift determination

In order to determine a precise distribution of AGN across cosmic time, unambiguous redshift measurements are needed (e.g. Naidoo et al., 2023). Spectroscopic redshifts, being the most precise measurements, can be determined for a large range of objects, from supernovae (e.g. Frederiksen et al., 2014; Baltay et al., 2021), galaxies (e.g. Le Fèvre et al., 2015; Galametz et al., 2013), and AGN (e.g. Rajagopal et al., 2021). Spectroscopic redshifts can be obtained by cross-correlation or fitting of the observed data and set of templates (Tonry and Davis, 1979; Schuecker, 1993; Glazebrook et al., 1998; Aihara et al., 2011; Machado et al., 2013) or by the direct detection and matching of powerful spectral features (Kurtz and Mink, 1998; Stoughton et al., 2002; Garilli et al., 2010). However, their determination can take long and high-quality observations, which are not always available for all sources, rendering them not suited for large-sky catalogues (see, for instance, Silva et al., 2011; Pacifici et al., 2023).

Photometric redshifts are an option which come from the use of photometry measurements and not explicit spectral features of an object (Salvato et al., 2019; Brescia et al., 2021; Newman and Gruen, 2022). In general, they use observations that take less integration time

than a comparable spectroscopic measurement and, thus, are used for large surveys that need measurements for large numbers of objects. They are also an option for faint sources.

Photometric redshift methods can deliver a probability for their redshift estimations in the form of a probability density function (PDF or $p(z)$). These functions can deliver a measure of the uncertainties that photometric redshifts might have. Photometric redshift determination methods can be divided, roughly, into three categories.

In general terms, photometric redshifts can be obtained using two different methods: template-based techniques and empirical relations.

Template-based methods come from the fitting of multi-wavelength photometry of a source to a model template (Baum, 1957; Baum, 1962; Loh and Spillar, 1986; Newman and Gruen, 2022; Pacifici et al., 2023). The models have been constructed using different combinations of properties –e.g. age, metallicity, contribution from different constituents, etc.–. Thus, the examined source will be assumed to have the properties from the model which fits the best. However, and depending upon the number and quality of the photometry measurements (e.g. low spectral resolution), these properties can have, sometimes, large uncertainties. Even though this method can use less precise values to determine a redshift, it can take a significative amount of time since it needs to contrast the measured SED to the full set of model templates and, when the number of available measurements is low, the quality of the estimation is largely degraded (e.g. Norris et al., 2019). Using this method, redshift estimations can be obtained from, for instance, galaxies (e.g. Hernán-Caballero et al., 2021), and AGN (e.g. Ananna et al., 2017; Brescia et al., 2019). As expected, the quality of photometric redshift estimates is highly correlated to the quality of the photometry data used for their determination (Newman et al., 2015; Newman and Gruen, 2022).

EXAMPLES OF SED FITTING TOOLS AND HOW THEY WORK

For the case of empirical relations, the retrieval of photometric redshifts relies on the use of statistics and large sets of observables (e.g. fluxes and their uncertainties) as to determine correlations between them and redshifts. Most of these redshift determination methods are related to the use of machine learning. These techniques will be further described in this work.

A third method can be applied to determine approximate redshifts. Using differences among magnitudes –i.e. colours– it is possible to establish the redshift range in which a source is located. This technique –called drop-out– is, by no means, precise, but can lead to further investigation of sources that are at relevant redshifts ranges for the researcher. In this way, drop-outs are employed as a mean to generate candidates for pertinent redshift values. Given

that it requires no more calculations than compare some series of colours, it is highly efficient at generating rough redshifts of large samples. It has been, mainly, used to generate and study high-redshift sources or candidates that, otherwise, would not have enough information to produce a precise redshift value (e.g. Bouwens et al., 2020; Carvajal et al., 2020; Merlin et al., 2021; Uzgil et al., 2021).

Since its first uses, this technique has allowed the detection of high-redshift galaxies (Steidel and Hamilton, 1992; Steidel et al., 1996) through the detection of sharp break in flux between broadband filters that sample the vicinities of the Lyman Break (at a rest-frame wavelength of 912 Å). The location of such break is a function of redshift allowing, thus, to obtain a crude estimate of the redshift for the studied objects.

This page intentionally left blank.

2

Challenges in the analysis of astronomical data

The progress of technology and methods used in Astrophysics has been one of the main drivers for the advancement in our knowledge and understanding of the processes happening in the Universe. But this undeniable improvement has brought some drawbacks that can pose serious challenges that might hinder our ability of retrieving useful results from astronomical data. Most of these problems are rooted in the very large number of new and different observational efforts carried out throughout the years. This abundance of measurements can impact the processes that lead to new calculations and results since more resources and steps are needed to treat a large number of measurements.

As more sources are needed to constrain their properties better, new data sets have been compiled and published. Now, multi-wavelength data are available for large fractions of the sky (e.g Gaia Collaboration et al., 2016; Chambers et al., 2016; Lacy et al., 2020; Kollmeier et al., 2017; Wright et al., 2010; Skrutskie et al., 2006; Abbott et al., 2018). But this profusion of observations has come with new challenges with the most relevant being the volume of data. Lately, analysing all observations one by one has become unfeasible in terms of the time needed to fulfil the task (see, for instance, Brescia et al., 2021). This issue will become greater as future surveys and telescopes are put into service, with relevant examples being the Square Kilometre Array (SKA; Braun et al., 2019) and the Legacy Survey of Space and Time survey (LSST; Ivezić et al., 2019).

2.1 Large surveys

Over the last couple of decades, the observational capabilities of single instruments have been improved largely. It has become possible to retrieve measurements of very large areas of the sky without important variations in the observationals properties (noise, calibration, etc.). These improvements in the overall properties of observations have made possible the creation

of surveys than can cover relevant fractions of the sky. Some early examples include ...

While being able to obtain information from more sources and regions of the sky is, by itself, a very relevant improvement, such number of new measurements to analyse have brought some issues related to the treatment of very large datasets.

2.2 Computational costs

Using very large surveys and catalogues for any sort of calculation involves, accordingly, very high computational costs. Recent observational catalogues might have up to billions of entries with several attributes each and large-area images can cover thousands of square degrees with very high angular resolution. Dealing with such large datasets requires large amount of resources that are not completely available to everyone.

Additionally, most of the methods traditionally used for the detection, classification, and extraction of properties from astrophysical sources have not been developed to be used with very large catalogues. For this reason, using them in the most recent catalogues and surveys can take restrictive running times that not even the most powerful computing facilities can deal with.

2.3 Missing measurements

As with any sort of pysical measurement, a fraction of observations might have issues that can render them unusable for any meaningful calculation. These problems include malfunction of the detector, incorrect cleaning of the data. If different measurements are to be combined, some sources might have observed in one instance but not in the remaining ones. This might affect the study of time series or multi-wavelength, multi-instrument observations.

2.4 Multi-wavelength counterpart identification

In the case of observations with different filters or different instruments, a new problem arises. It involves the correct identification of the sources observed in each of the filters. Given that the emission in different wavelengths and different moments in time might come from separate components and processes in the studied objects, each observed instance can present a

structure that does not match the others. For this reason, finding and matching counterparts for detected sources can be difficult.

Several approaches have been used to mitigate this problem. The simplest of them involves taking the closest source (within a defined search radius) as the counterpart without analysing any physical correlation between the underlying processes that might give rise to the observed emission.

More complex procedures attempt to flag as counterpart the sources with the highest likelihood of being produced in the same position in the space (three dimensional space).

This page intentionally left blank.

3

Machine-assisted pattern detection

Taking into account all the issues that the analysis of large datasets might pose, new tools have been developed as a way to tackle them. In particular, The existence of these major AGN detection, radio measurement, and redshift determination methods raises the need of new techniques which might be able to obtain these properties for large amounts of astrophysical sources with enough precision within a shorter amount of time.

Given that this is a problem suffered by several scientific and, even, non-scientific disciplines (e.g. business-related applications; Costa-Climent et al., 2023), large efforts have been put in order to solve it and many techniques have been developed to deal with the ever-increasing data volumes. New statistical and computer methods can analyse thousands or millions of elements and find relevant trends among their properties (Garofalo et al., 2017). One branch of these these techniques is able to, using previously-fed data, predict, with relevant confidence, the behaviour new data will have –i.e. the values of their properties–. This is what has been called Machine Learning (ML; Samuel, 1959).

In Astronomy, ML has been used in a wide range of subjects, such as redshift determination (e.g. Nakoneczny et al., 2021; Wenzl et al., 2021), morphological classification (e.g. Ma et al., 2019; Lukic et al., 2019; Mostert et al., 2021; Vardoulaki et al., 2021; Burhanudin et al., 2021), emission prediction (e.g. Dobbels and Baes, 2021), anomaly detection (e.g. Baron and Poznanski, 2017; Giles and Walkowicz, 2019; Lochner and Bassett, 2021; Storey-Fisher et al., 2021), observations planning (e.g. Garcia-Piquer et al., 2017; Jia et al., 2023), and more (Ball and Brunner, 2010; Baron, 2019). With ML, it is possible to use previously available measurements and extract useful trends and correlations that can suggest the behaviour of properties from future observations or simulations. ML models are, in general, only fed with measurements and not with physical assumptions (Desai and Strachan, 2021) and they do not need to check the consistency of the predictions or results they provide. This can bring, as a consequence, that running times for this kind of algorithms might be less than typical physically-based codes (e.g. Buchner, 2019).

3.1 Types of machine-assisted analyses

Concentrating our review on the application of ML, two main branches exist for the application of such techniques. The first of them, called Supervised Learning, deals with the idea that, for each set of measurements, there is a response value that, via modelling, we can predict with some degree of confidence (James et al., 2023). On the other side, Unsupervised Learning refers to the analysis of data that does not have an associated quantity. One of the most popular applications of unsupervised learning is clustering of elements. Modelling data would imply separating them by how similar are their properties (or a combination of them). Then, in the case of supervised learning, further divisions are possible. If the predicted variable (target) is a discrete quantity, this prediction is called a classification. Opposite to that, if the predicted target is continuous, the process is called regression.

3.2 Classification tasks

3.3 Regression tasks

In this work, we focus on redshift as one of the regression targets. In this way, most of the further definitions related to regression tasks, will be directly applied to the redshift and its properties.

3.4 Prediction metrics

A set of metrics will be used to understand the reliability of the results and put them in context with results in the literature. Since our work includes the use of classification and regression models, we briefly discuss the appropriate metrics in the following sections.

3.4.1 Classification metrics

The main tool to assess the performance of classification methods is the Confusion (or Error) Matrix. It is a two-dimension (predicted vs. true) matrix where the true and predicted class(es) are compared and results stored in cells with the rate of True Positives (TP), True Negative (TN), False Positives (FP), and False Negatives (FN). An example diagram showing

Table 3.1: Diagram of confusion matrix for the classification of sources between AGN and galaxies.

True Classes		Galaxy	AGN
	Galaxy	True Negative	False Positive
	AGN	False Negative	True Positive

Predicted Classes

the elements of a confusion matrix is shown ing Table 3.1. As mentioned earlier in Sect. ref-sec:ML_training, we seek to maximise the number of positive-class sources that are recovered as such. Using the elements of the confusion matrix, this aim can be translated into the maximisation of TP and, consequently, the minimisation of FN.

From the elements of the confusion matrix, we can obtain additional metrics, such as the F1 and F_β scores (Dice, 1945; Sørenson, 1948; van Rijsbergen, 1979), and the Matthews Correlation Coefficient (MCC; Yule, 1912; Cramér, 1946; Matthews, 1975) which are better suited for unbalanced data as they take into account the behaviour and correlations among all elements of the confusion matrix. As such, the F1 coefficient is defined as:

$$\mathrm{F1} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}. \tag{3.1}$$

F1 values can go from 0 (no prediction of positive instances) to 1 (perfect prediction of elements with positive labels). This definition assigns equal weight (importance) to both the number of FN and FP. An extension to the F1 score, which adds a non-negative parameter, β, to increase the importance given to each one of them is the F-Score (F_β), defined as:

$$\mathrm{F}_\beta = \frac{(1+\beta^2)\times \mathrm{TP}}{(1+\beta^2)\times \mathrm{TP} + \beta^2 \times \mathrm{FN} + \mathrm{FP}}. \tag{3.2}$$

Using $\beta > 1$, more relevance is given to the optimisation of FN. When $0 \leq \beta < 1$, the optimisation of FP is more relevant. If $\beta = 1$, the initial definition of F1 is recovered. As with F1, F_β values can be in the range $[0, 1]$. As we seek to minimise the number of FN detection, we adopt a conservative value of $\beta = 1.1$, giving more significance to their reduction without removing the aim for FP. Also, this value is close enough to $\beta = 1$, which will allow us to

compare our scores to those produced in previous works.

MCC is defined as:

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN} - \mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}, \tag{3.3}$$

which includes also the information about the TN elements. MCC can range from −1 (total disagreement between true and predicted values) to +1 (perfect prediction) with 0 representing a prediction analogous to a random guess.

The Recall (also called Completeness, Sensitivity, or True Positive Rate –TPR–; Yerushalmy, 1947) corresponds to the rate of relevant, or correct, elements that have been recovered by a process. Using the elements from the confusion matrix, it can be defined as:

$$\mathrm{Recall} = \mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{3.4}$$

The TPR can go from 0 to 1, with a value of 1 meaning that the model can recover all the true instances.

The last metric used is Precision (also known as Purity), which can be defined as the ratio between the number of correctly classified elements and the number of sources in the positive class (AGN or radio detectable):

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}. \tag{3.5}$$

Precision can range from 0 to 1 where higher values show that more real positive instances of the studied set were retrieved as such by the model.

3.4.2 Regression metrics

Usually, regression tasks are assessed with the use of metrics such as mean squared error, root mean squared error, and mean absolute error. These measure the deviation of the predicted value from the original quantity. If the original value is called y_{True} and its predicted version is $y_{\mathrm{Predicted}}$, these three regression metrics can be defined as follows.

The mean squared error (MSE) is

$$\mathrm{MSE}(y) = \frac{1}{\mathrm{d}}\sum_{i}^{\mathrm{d}} (y_{\mathrm{True}} - y_{\mathrm{Predicted}})^2, \tag{3.6}$$

with d being the number of elements in the studied sample (i.e. its size). A direct modification of MSE appears when calculating its square root. Then, the root mean squared error is

$$\mathrm{RMSE}(y) = \sqrt{\frac{1}{\mathrm{d}}\sum_{i}^{\mathrm{d}} (y_{\mathrm{True}} - y_{\mathrm{Predicted}})^2}\,. \tag{3.7}$$

A third way to quantify the deviation of the predictions from the true values is through the mean absolute error (MAE);

$$\mathrm{MAE}(y) = \frac{1}{\mathrm{d}}\sum_{i}^{\mathrm{d}} |y_{\mathrm{True}} - y_{\mathrm{Predicted}}|\,, \tag{3.8}$$

While MSE and RMSE are sensitive to large deviations in the predictions, the MAE has a linear behaviour with respect to the fluctuations in predicted quantities.

For the case of redshift value determination, the previous metrics are not fully able to assimilate its typical (and logarithmic) behaviour. Thus, further modifications are needed in order to use suitable metrics. Namely, a factor must be included to take into account the fact that differences between low redshift values should be penalized more strongly than those at higher redshifts.

We can start with the difference between true (z_{True}) and predicted ($z_{\mathrm{Predicted}}$) redshift values,

$$\Delta z = z_{\mathrm{True}} - z_{\mathrm{Predicted}}\,, \tag{3.9}$$

and its normalised difference,

$$\Delta z^{\mathrm{N}} = \frac{z_{\mathrm{True}} - z_{\mathrm{Predicted}}}{1 + z_{\mathrm{True}}}\,. \tag{3.10}$$

If the comparison is made over a larger sample of elements, the bias of the redshift is used (Dahlen et al., 2013), with the median of the quantities instead of its mean to avoid the strong influence of extreme values:

$$\Delta z_{\mathrm{Total}} = \mathrm{median}\,(z_{\mathrm{True}} - z_{\mathrm{Predicted}}) = \mathrm{median}(\Delta z)\,, \tag{3.11}$$

$$\Delta z^{\mathrm{N}}_{\mathrm{Total}} = \mathrm{median}\left(\frac{z_{\mathrm{True}} - z_{\mathrm{Predicted}}}{1 + z_{\mathrm{True}}}\right) = \mathrm{median}(\Delta z^{\mathrm{N}})\,. \tag{3.12}$$

Using the previous definitions, four additional metrics can be calculated. These are the median absolute deviation (MAD, σ_{MAD}) and normalised median absolute deviation (NMAD,

$\sigma_{\rm NMAD}$; Hoaglin et al., 1983; Ilbert et al., 2009), which are less sensitive to outliers. Also, the standard deviation of the predictions, σ_z, and its normalised version, $\sigma_z^{\rm N}$ are typically used. They are defined as:

$$\sigma_{\rm MAD} = 1.48 \times {\rm median}\left(|\Delta z|\right), \tag{3.13}$$

$$\sigma_{\rm NMAD} = 1.48 \times {\rm median}\left(\left|\Delta z^{\rm N}\right|\right), \tag{3.14}$$

$$\sigma_z = \sqrt{\frac{1}{\rm d}\sum_i^{\rm d}(\Delta z)^2}, \tag{3.15}$$

$$\sigma_z^{\rm N} = \sqrt{\frac{1}{\rm d}\sum_i^{\rm d}(\Delta z^{\rm N})^2}, \tag{3.16}$$

Additionally, the outlier fraction (η, as used in Dahlen et al., 2013; Lima et al., 2022) is considered, which is defined as the fraction sources with a predicted redshift difference ($\left|\Delta z^{\rm N}\right|$, Eq. 3.10) larger than a previously set value. Taking the results from Ilbert et al. (2009) and Hildebrandt et al. (2010), we have selected this threshold to be 0.15, leaving the definition of the outlier fraction as:

$$\eta = \frac{\#\left(\left|\Delta z^{\rm N}\right| > 0.15\right)}{\rm d}. \tag{3.17}$$

where # symbolises the number of sources fulfilling the described relation, and d corresponds to the size of the selected sample.

3.5 Classification thresholds

Classification models deliver a range of scores or probabilities for which a threshold is needed to separate their predictions between negative and positive classes. By default, these models set a threshold at 0.5 in score[1] but, in principle, and given the characteristics of the problem, a different optimal threshold might be needed.

In our case, we want to optimise (increase) the number of recovered elements in each model (i.e. AGN or radio-detectable sources). This maximisation corresponds to obtaining thresholds that optimise the recall given a specific precision limit. We did that with the use of

[1]Throughout this work, we will call this a naive threshold.

the statistical tool called Precision-Recall (PR) Curve. A deeper description of this method and the results obtained from our work are presented in Appendix refsec:app_pr_curve[2].

By maximising the recall (Eq. 3.4), we improve the number of recovered elements in each classifier. This can be done by decreasing the threshold by which a source is classified as a positive instances. Setting this threshold to its minimum, 0.0, would increase the recall. But every source would be predicted to be an AGN or detected on the radio regardless of their properties. Thus, a different approach must be taken.

One statistical tool designed to optimise the classification threshold taking into account the overall model performance is the Precision-Recall (PR) curve. It can help to understand the behaviour of a classifier as a function of its threshold. Both quantities, precision (Eq. 3.5) and recall, show an inverse correlation, and both depend on the selected threshold. Thus, they can be used to retrieve the score value for which both quantities are balanced. This optimisation is done by finding the threshold that maximises the F_β score (Eq. 3.2). This operation will be performed over the union of training and validation sets, which have been used to create and train each model.

3.6 Classification calibration

In general, classifiers deliver scores in the range [0, 1], which could be associated to the probability of a studied source being part of the relevant class (in our work, AGN or radio detectable). The classifier uses a threshold above which, any predicted element would be considered a positive instance.

With the exception of few algorithms (including the family of logistic regressions), scores from classifiers cannot be directly used as probabilities. As a consequence of this inability, such values cannot be compared from one type of model to some other and can not be combined to obtain a joint score. Therefore, in order to retrieve joint scores and treat them as probabilities, scores (and, by extension, the classifiers) need to be calibrated. This calibration means that, when taking all predictions with a probability P of being of a class, a fraction P of them really belong to that class (e.g. Lichtenstein et al., 1982; Silva Filho et al., 2023).

Calibration of these scores can be done by applying a transformation to their values. For our work, we will apply a Beta transformation. It allows one to re-distribute the scores of a classifier allowing them to get closer to the definition of probability (Kull et al., 2017a; Kull

[2]Thresholds derived from the PR curves will be labelled as PR.

et al., 2017b). Calibration steps in our workflow have been applied using the Python package `betacal`. In the case of the radio detection model, the new scores have a wider range than the original, uncalibrated scores.

When obtaining the BSS values for both classification, the AGN-galaxy classifier has a score of BSS = −0.002, demonstrating that no major changes were applied to the distribution of scores. For the radio detection classifier, the score is BSS = −0.434. Even though the BSS value is slightly negative for the AGN-galaxy classifier, we keep it since its range of values now can be compared and combined with additional probabilities. In the case of the radio detection classifier, the BSS shows a degradation of the calibration, but we will keep the calibrated model given that it provides, overall, better values for the remaining metrics. This effect can be seen, more strongly, with recall.

Calibration (or reliability) plots show how well calibrated the predicted scores of a classifier are by displaying the fraction of sources that are part of a given class as a function of the predicted probability. A perfectly calibrated classifier would have all its prediction lying in the x=y line. The magnitude of the deviations from that line give information of the miscalibration a model has (see, for instance, Bröcker and Smith, 2007; Van Calster et al., 2019). In Fig.reffig:calibration_curves_classification_pre, we present the reliability curves for the uncalibrated classifiers and, in Fig.reffig:calibration_curves_classification_post, for their calibrated versions.

3.6.1 Calibration metrics

One of the most used analytical metrics to assess calibration of a model is the Brier score (BS; Brier, 1950). It measures the mean square difference between the predicted probability of an element and its true class. If the total number of elements in the studied sample is d, the BS can be written (for binary classification problems, as the ones studied in this work) as:

$$\mathrm{BS} = \frac{1}{\mathrm{d}} \sum_{i}^{\mathrm{d}} (\mathbb{C} - \mathrm{class})^2 \,, \tag{3.18}$$

where $\mathbb{C}$ is the predicted class and class the true class of each of the elements in the sample (0 or 1). The BS can range between 0 and 1 with 0 representing a model that is completely reliable in its predictions.

Additionally, the BS can be used to compare the reliability (or calibration) between a model and a reference using the Brier Skill Score (BSS; e.g. Glahn and Jorgensen, 1970):

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{ref}}}. \tag{3.19}$$

In our case, $\mathrm{BS}_{\mathrm{ref}}$ corresponds to the value calculated from the uncalibrated model. The BSS can take values between -1 and $+1$. The closer the BSS gets to 1, the more reliable the analysed model is. These values include the case where BSS$\approx$0, in which both models perform similarly in terms of calibration.

For our pipeline, after a model has been fully trained, a calibrated version of their scores will be obtained. With both of them, the BSS will be calculated and, if it is not much lower than 0, that calibrated transformation will be used as the final scores from the prediction.

3.7 Ensemble learning

One technique used to improve the results of a prediction is called ensemble learning. It involves the combined use of individual results from ML models, that have trained to solve the same problem, into one larger model or rule that can deliver a final prediction (Schapire, 1990; Breiman, 1996; Freund and Schapire, 1996). It has been shown that the combination of several models, and their predictions, can improve the overall prediction results (Opitz and Maclin, 1999).

In order to combine several prediction into a final result, several options are available. The most used, and one of the earliest, way to merge all individual predictions consists of averaging each predicted value into the final prediction (e.g. Sollich and Krogh, 1995). This option is useful for both regression and classification tasks (using its output scores). For the specific case of classification, a voting system can be implemented, where the majority of decisions of the base individual predictors is taken as the final predicted class. This method has been proven to work (Schapire et al., 1998).

3.7.1 Generalised stacking

A third method to combine individual predictions is through generalised model stacking (Wolpert, 1992) which can be interpreted as the addition of priors to the model training stage. In this way, two levels of predictors are used. The first level includes all the individual models that are trained on the training set. The second level corresponds to a single model which is trained only on the outputs of the first-level models. Generalised stacking has been applied in

several astrophysical problems. That is the case of Zitlau et al. (2016), Cunha and Humphrey (2022), and Euclid Collaboration et al. (2023a), Euclid Collaboration et al. (2023b).

3.8 Model explainability and feature importance analysis

One method for explaining what correlations and connections a model is using is that of feature importances.

Given the success of the models and pipeline in classifying AGN, their radio detectability and redshift with the provided set of observables, knowing the relative weights that they have in the decision-making process is of utmost relevance. In this way, physical insight might be gained about the triggers of AGN and radio activity and its connection to their host. Therefore, we have estimated both local and global feature importances for the individual models and the combined pipeline. Global importances were retrieved using the so-called 'decrease in impurity' approach (see, for example, Breiman, 2001). Local importances have been determined via Shapley values. A more detailed description of what these importances are and how they are calculated is given in the following sections.

3.8.1 Global feature importances

Overall, mean or global feature importances can be retrieved from models that are based on Decision Trees (e.g. Random Forests and Boosting models, Breiman, 2001; Breiman, 2003). All algorithms selected in this work (`RF`, `CatBoost`, `XGBoost`, `ET`, `GBR`, and `GBC`) belong to these two classes. For each feature, the decrease in impurity (a term frequently used in the literature related to Machine Learning) of the dataset is calculated for all the nodes of the tree in which that feature is used. Features with the highest impurity decrease will be more important for the model (Louppe et al., 2013)[3].

Insight into the decision-making of the pipeline can only rely on the specific weight of the original set of features (see Sect. **??**). Table 7.2 presents the ranked combined importances from the observables selected in each of the three sequential models that compose the pipeline. They have been combined using the importances from the meta-learner (as shown in Table 7.3) and

[3]For some models that are not based on Decision Trees, feature importances can be obtained from the coefficients that the training process delivers for each feature. These coefficients are related to the level to which each quantity is scaled to obtain a final prediction (as in the coefficients from a polynomial regression).

that of base-learners. The derived importances will be dependent on the dataset used, including any imputation for the missing data, and the details of the models, i.e. algorithms used and stacking procedure. We first notice in Table 7.2 that the order of the features is different for all three models. This difference reinforces the need, as stated in Sect. **??**, of developing separate models for each of the prediction stages of this work that would evaluate the best feature weights for the related classification or regression task.

3.8.2 Local feature importances

As opposed to the global (mean) assessment of feature importances derived from the decrease in impurity, local (i.e. source by source) information on the performance of such features can be obtained from Shapley values. This is a method from coalitional game theory that tells us how to fairly distribute the dividends (the prediction in our case) among the features (Shapley, 1953). The previous statement means that the relative influence of each property from the dataset can be derived for individual predictions in the decision made by the model (which is not the same as obtaining causal correlations between features and the target; Ma and Tourani, 2020). The combination of Shapley values with several other model explanation methods was used by Lundberg and Lee (2017) to create the SHapley Additive exPlanations (SHAP) values. In this work, SHAP values were calculated using the python package `SHAP`[4] and, in particular, its module for Tree-based predictors (Lundberg et al., 2020). To speed calculations up, the package `FastTreeSHAP`[5] (`v0.1.2`; Yang, 2021) was also used, which allows for multi-thread runs.

One way to display these SHAP values is through the so-called decision plots. They can show how individual predictions are driven by the inclusion of each feature. Besides determining the most relevant properties that help the model make a decision, it is possible to detect sources that follow different prediction paths which could be, eventually and upon further examination, labelled as outliers. An example of this decision plot, linked to the AGN-Galaxy classification, is shown in Fig. 7.3 for a subsample of the high-redshift ($z \geq 4.0$) spectroscopically classified AGN in the HETDEX field (121 sources, regardless of them being part of any sub-set involved in the training or validation of the models). The different features used by the meta-learner are stacked on the vertical axis with increasing weight and these final weight are

[4] https://github.com/slundberg/shap

[5] https://github.com/linkedin/fasttreeshap

summarised in Table 7.4. Similarly, SHAP decision plots for the radio-detection and redshift prediction are presented in Figs. 7.4 and 7.5, respectively.

4

Modelled datasets

A large area with deep and homogeneous quality radio observations is needed to train, validate, and test the models and predictions for RGs with already existent observations. As training field we selected the area of the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; Hill et al., 2008) Spring Field covered by the first data release of the LOFAR Two-metre Sky Survey (LoTSS-DR1; Shimwell et al., 2019). The LoTSS survey covers 424 deg^2 in the HETDEX Spring field (hereafter, HETDEX field) with Low Frequency Array (LOFAR; van Haarlem et al., 2013) 150 MHz observations that have a median sensitivity of 71 μJy/beam. HETDEX provides, as well, multi-wavelength homogeneous coverage as described below.

In order to test the performance of the models when applied to different areas of the sky, and with different coverages from radio surveys, we have selected the Sloan Digital Sky Survey (SDSS; York et al., 2000) Stripe 82 Field (S82; Annis et al., 2014; Jiang et al., 2014). For S82, we collected data from the same surveys as with the HETDEX field (see the following section) but with one important caveat: no LoTSS data is available in the field and, thus, we gathered the radio information from the VLA SDSS Stripe 82 Survey (VLAS82; Hodge et al., 2011). VLAS82 covers an area of 92 deg^2 with a median rms noise of 52 μJy/beam at 1.4 GHz. We have selected the S82 field (and, in particular, the area covered by VLAS82) given that it presents deep radio observations but taken with a different instrument than LOFAR. This difference allows us to test the suitability of our models and procedures in conditions that are not exactly the same as those from the training circumstances.

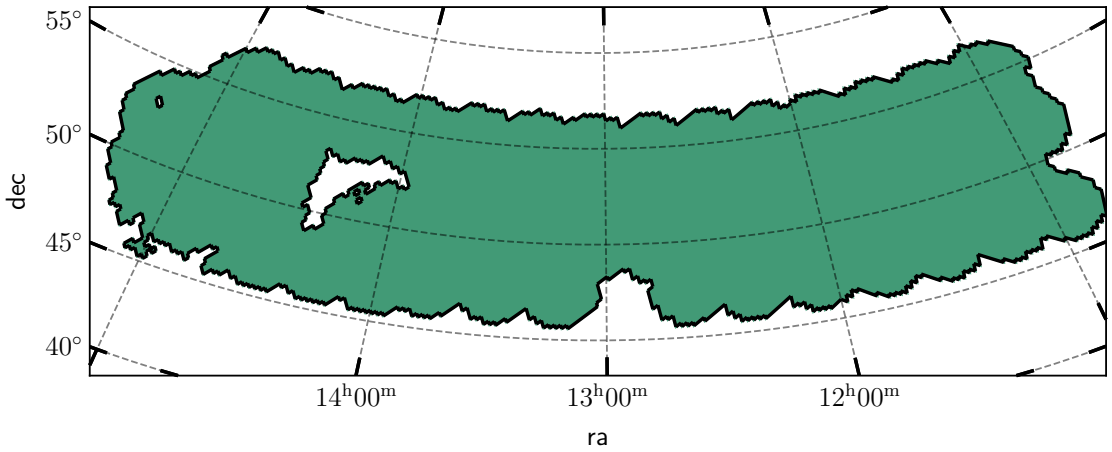


Figure 4.1: Footprint of the area used in the HETDEX field for this work.

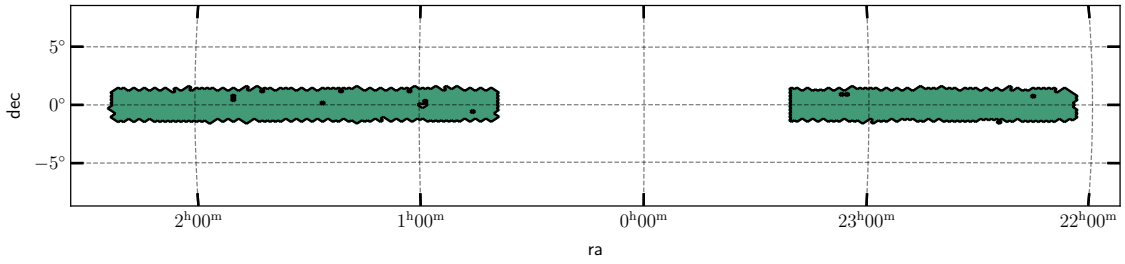


Figure 4.2: Footprint of the area used in the S82 field for this work.

4.1 HETDEX Spring field

4.2 Stripe 82 field

4.3 Photometry measurements

The base survey from which all the studied sources have been drawn is the CatWISE2020 (CW; Marocco et al., 2021). It lists NIR-detected elements selected from Wide-field Infrared Survey Explorer (WISE; Wright et al., 2010) and Near-Earth Object WISE (NEOWISE; Mainzer et al., 2011; Mainzer et al., 2014) over the entire sky at 3.4 and 4.6 μm (W1 and W2 bands, respectively). This catalogue includes sources detected at 5σ in either of the used bands (i.e. W1$\sim$17.43 and W2$\sim$16.47 $\mathrm{mag_{Vega}}$ respectively). The HETDEX field contains 15 136 878 sources listed in CW. Conversely, in the S82 field, there are 3 590 306 of them.

Multi-wavelength counterparts for CW sources were found on other catalogues applying a $1\farcs1$ search criteria. These catalogues include Pan-STARRS DR1 (PS1; Chambers et al., 2016; Flewelling et al., 2020), 2MASS All-Sky (2M; Wright et al., 2010; Cutri et al., 2003a; Cutri et al., 2003b), and AllWISE (AW; Cutri et al., 2013)[1]. The adopted search radius corresponds to the distance that has been used by Wright et al. (2010) to match radio sources to PS1 and WISE observations. Nevertheless, the source density of the radio (LOFAR, Very Large Array Sky Survey (VLASS; Gordon et al., 2020)) and 2M catalogues imply a low statistical ($< 1\%$) spurious counterpart association, this is not the case for PS1, where the source density is higher. For this reason, and to mantain a statistically low spurious association between CW and PS1, we limited our search radius to $1\farcs1$. This distance corresponds to the smallest Point-Spread Function (PSF) size of the bands included in PS1 (Chambers et al., 2016).

For the purposes of this work, observations in LoTSS and VLAS82 are only used to determine whether a source is radio detected, or not. In particular, no check has been performed on whether a selected source is extended or not in any of the radio surveys. A single Boolean feature is created from the radio measurements (see Sect.refsec feature_creation) and no further analyses were performed regarding the detection levels that might be found in any of the fields.

Additionally, we have discarded the measurement errors of all bands. Traditionally, ML algorithms cannot incorporate uncertainties in a straightforward way and, thus, we opted to

[1]For the purposes of the analyses, and except when clearly stated otherwise, all photometric measurements were converted to AB magnitudes.

avoid attempting to use them for training (for some examples on how they can be incorporated in astrophysically motivated ML studies, see Ball et al., 2008; Reis et al., 2019; Shy et al., 2022). Furthermore, Humphrey et al. (2022) have shown that, in specific cases, the inclusion of measurement errors does not add new information to the training of the models and can be even detrimental to the prediction metrics. The degradation of the model by including uncertainties can likely be related to the fact that, by virtue of the large number of sources included in the training stages, the uncertainties are already encoded in the dataset in the form of scatter.

4.4 Missing data treatment

Following the same argument of measurement errors, upper limit values have been removed and a missing value is assumed instead. In general, ML methods (and their underlying statistical methods) cannot work with catalogues that have empty entries (Allison, 2001). For that reason, we have used single imputation (a review on the use of this method in astronomy can be seen in Chattopadhyay, 2017) to replace these missing values, and those fainter than $5-\sigma$ limits, with meaningful quantities that represent the lack of a measurement. We have opted for the inclusion of the same $5-\sigma$ limiting magnitudes as the value to impute with. This method of imputation, with some variations, has been successfully applied and tested, recently, by Arsioli and Dedin (2020), Carvajal et al. (2021) and Curran (2022), and Curran et al. (2022).

In this way, observations from 12 non-radio bands were gathered (as listed in Table 4.1). The magnitude density distribution for the sample from the HETDEX and S82 fields, without any imputation, is shown in Fig. 4.3. After imputation, the distribution of magnitudes changes, as shown in Fig. 4.4. Each panel of the figure shows the number of sources which have a measurement above its $5-\sigma$ limit in such band. Additionally, a representation of the observational $5-\sigma$ limits of the bands and surveys used in this work is presented in Fig. 4.5. It is worth noting the depth difference between VLAS82 and LoTSS-DR1 is $\sim$1.5 mag for a typical synchrotron emitting source ($F_\nu \propto \nu^\alpha$ with $\alpha = -0.8$), allowing the latter survey reach fainter sources.

4.5 Additional features

AGN labels and redshift information were obtained by cross-matching (with a 1.″1 search radius) the catalogue with the Million Quasar Catalog (MQC, v7.4d; Flesch, 2021), which lists information from more than 1 500 000 objects that have been classified as optical Quasi Stellar

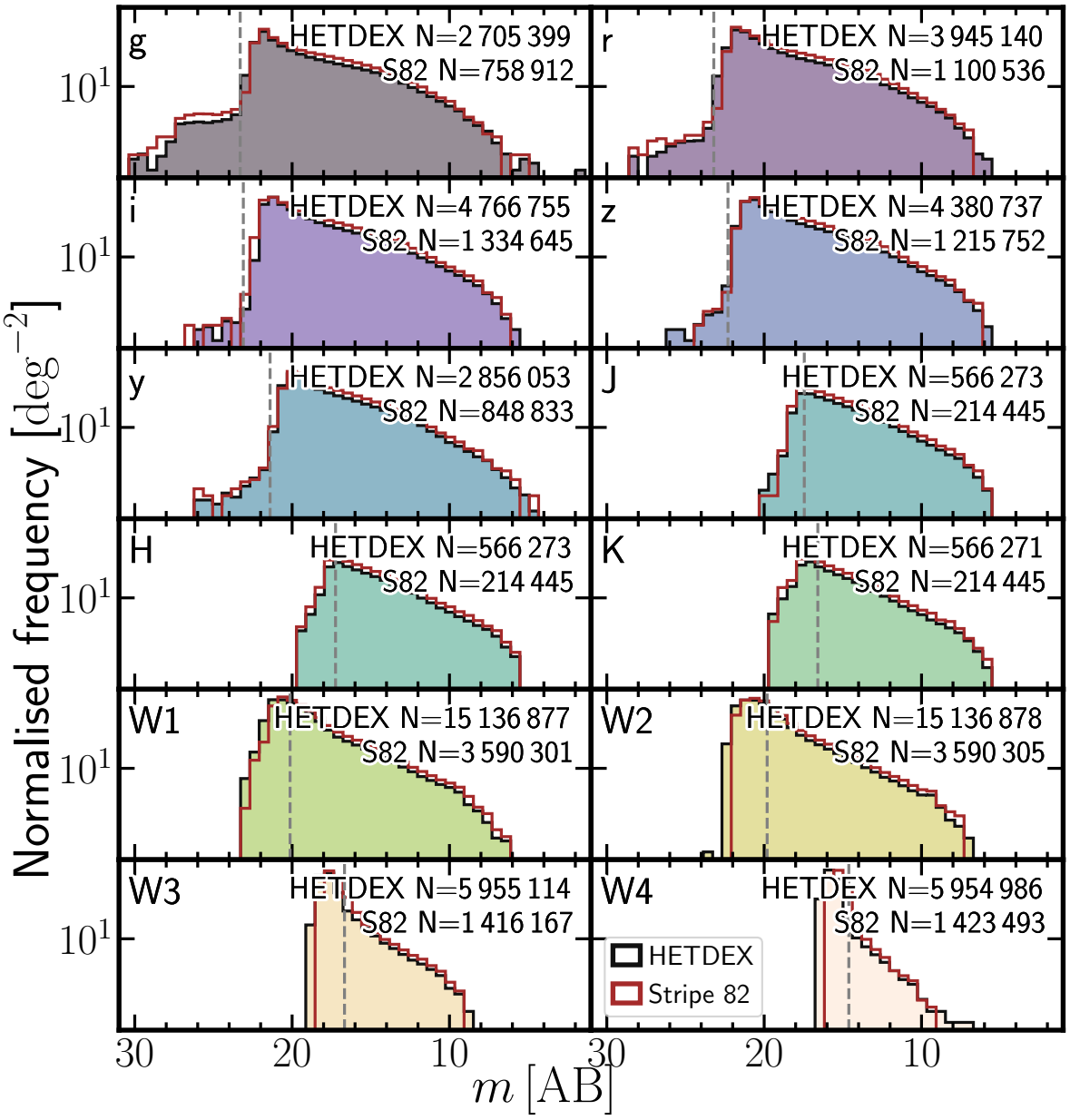


Figure 4.3: Histograms of base collected, non-imputed, non-radio bands for HETDEX (clean, background histograms) and S82 (empty, brown histograms). Each panel shows the distribution of measured magnitudes of detected sources divided by the total area of the field. Dashed, vertical lines represent the 5$-\sigma$ magnitude limit for each band. The number in the upper right corner of each panel shows the number of measured magnitudes included in their corresponding histogram.

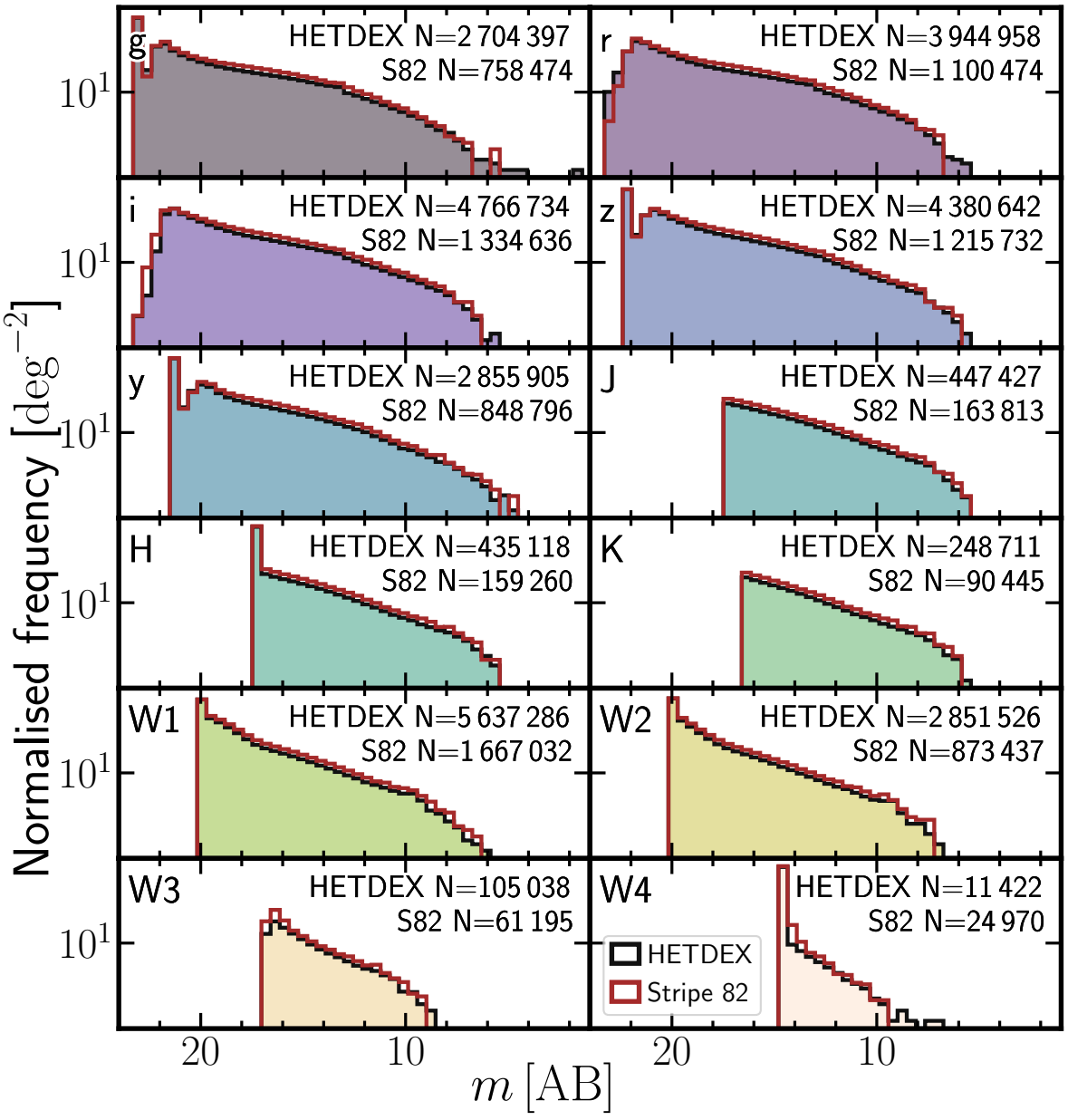


Figure 4.4: Histograms of base collected non-radio bands for HETDEX (clean, background histograms) and S82 (empty, brown histograms) fields. Description as in Fig. 4.3. The number in the upper right corner of each panel shows the number of sources with magnitudes originally measured above the 5−σ limit included in their corresponding histogram for each field.

Table 4.1: Bands available for model training in our dataset

Survey	Band (Column name)
Pan-STARRS (PS1)	g (`gmag`), r (`rmag`), i (`imag`), z (`zmag`), y (`ymag`)
2MASS (2M)	J (`Jmag`), H (`Hmag`), Ks (`Kmag`)
CatWISE2020 (CW)	W1 (`W1mproPM`), W2 (`W2mproPM`)
AllWISE (AW)	W3 (`W3mag`), W4 (`W4mag`)

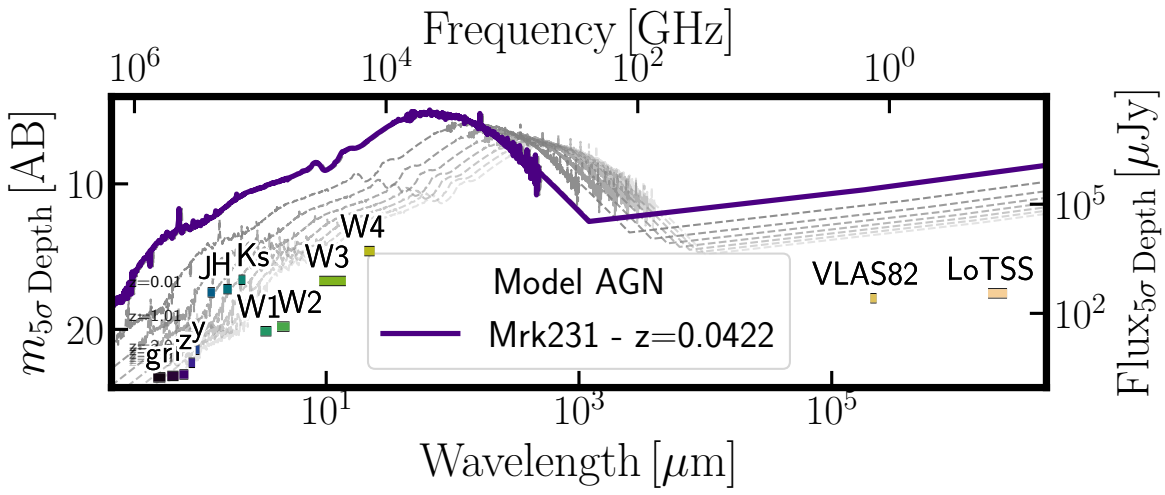


Figure 4.5: Flux and magnitude depths (5−σ) from the surveys and bands used in this work. Limiting magnitudes and fluxes were obtained from the description of the surveys, as referenced in Sect. **??**. In purple, rest-frame spectral energy distribution (SED) from Mrk231 (z = 0.0422, Brown et al., 2019) is displayed as an example AGN. Redshifted (from z=0.001 to z=7) versions of this SED are shown in dashed grey lines.

Table 4.2: Composition of initial catalogue and number of cross matches with additional surveys and catalogues.

Survey	HETDEX	Stripe82
CatWISE2020	15 136 878	3 590 306
AllWISE	5 955 123	1 424 576
Pan-STARRS	4 837 580	1 346 915
2MASS	566 273	214 445
LoTSS	187 573	...
VLAS82	...	8 747
MQC (AGN)	50 538	17 743
SDSS (Galaxy)	68 196	4 085

Object (QSO), AGN, or Blazars. Sources listed in the MQC may have additional counterpart information, including radio or X-ray associations. For the purposes of this work, only sources with secure spectroscopic redshifts were used. The matching yielded 50 538 spectroscopically confirmed AGN in HETDEX and 17 743 confirmed AGN in S82.

Similarly, the sources in our parent catalogue were cross-matched with the Sloan Digital Sky Survey Data Release 16 (SDSS-DR16; Ahumada et al., 2020). This cross-match was done solely to determine which sources have been spectroscopically classified as galaxies (`spClass == GALAXY`). For most of these galaxies, SDSS-DR16 lists a spectroscopic redshift value, which will be used in some stages of this work. In the HETDEX field, SDSS-DR16 provides 68 196 spectroscopically confirmed galaxies. In the Stripe 82 field, SDSS-DR16 identifies 4 085 galaxies spectroscopically. Given that MQC has access to more AGN detection methods than SDSS, when sources were identified as both galaxies (in SDSS-DR16) and AGN (in the MQC), a final label of AGN was given. A description of the number of elements in each field and the multi-wavelength counterparts found for them is presented in Table 4.2.

The initial pool of features that have been selected or engineered to use in our analysis is briefly described below:

- Photometry, both measured and imputed, in the form of AB magnitudes for a total of 12 bands.

- Colours. All available colours from measured and imputed magnitudes were considered. In total, there are 66 colours, resulting from all available combinations of two magnitudes between the 12 selected bands. These colours are labelled in the form `X_Y` where `X` and `Y` are the respective magnitudes.

- Number of non-radio bands in which a source has valid measurements (`band_num`). This feature could be, very loosely, attributed to the total flux a source can display. A higher `band_num` will imply that such source can be detected in more bands, hinting a higher flux (regardless of redshift). The use of features with counting or aggregation of elements in the studied dataset is well established in ML (see, for example, Zheng and Casari, 2018; Duboue, 2020; Sánchez-Sáez et al., 2021; Humphrey et al., 2022).

- AGN-galaxy classification Boolean flag named `class`.

- Radio Boolean flag `LOFAR_detect`. This feature flags whether sources have counterparts in the radio catalogues (LoTSS or VLAS82).

A list of the features created for this work and their representation in the code and in some of the figures is presented in Table 4.3.

4.6 Data re-scaling

Attending to the intrinsic differences between ML algorithms, not all of them have the same performance when being trained with features spanning a wide range of values (i.e. several orders of magnitude). In particular, linear modelling of data might overrepresent features with larger absolute values when measuring distances between data point. For this reason, it is customary to re-scale the available values to either be contained within the range $[0, 1]$ or to have similar distributions. We applied a version of the latter transformation to our features (not the targets) as to have a mean value of $\mu = 0$ and a standard deviation of $\sigma = 1$ for each feature. Additionally, these new values were power-transformed to resemble a Gaussian distribution. This transformation helps the models avoid using the distribution of values as additional information for the training. For this work, a Yeo-Johnson transformation (Yeo and Johnson, 2000) was applied.

Table 4.3: Names of columns or features used in the code and what they represent.

Photometry measurements (magnitudes and fluxes)					
Code name	Feature	Code name	Feature	Code name	Feature
gmag	g (PS1)	ymag	y (PS1)	W1mproPM	W1 (CW)
rmag	r (PS1)	Jmag	J (2M)	W1mproPM	W2 (CW)
imag	i (PS1)	Hmag	H (2M)	W3mag	W3 (AW)
zmag	z (PS1)	Kmag	Ks (2M)	W4mag	W4 (AW)
Colours					
66 colours from all combinations of non-radio magnitudes. A sub-sample of them is shown.					
g_r	g - r (PS1)	...	...	W2_W3	W2 (CW) - W3 (AW)
g_i	g - i (PS1)	...	...	W2_W4	W2 (CW) - W4 (AW)
g_z	g - z (PS1)	...	...	W3_W4	W3 - W4 (AW)
Categorical flags					
Code name	Feature				
band_num	Number of bands with measurements				
Boolean flags					
Code name	Feature	Code name	Feature		
class	AGN or galaxy	radio_detect	Detection in, at least, one radio band.		
Redshift					
Code name	Feature				
z	Spectroscopic redshift				
Outputs of base models					
Code name	Feature	Code name	Feature	Code name	Feature
XGBoost	XGBoost	ET	Extra Trees	GBR	Gradient Boosting Regressor
CatBoost	CatBoost	GBC	Gradient Boosting Classifier		
RF	Random Forest				

5

Training of models

5.1 Model stacking

By design, each ML algorithm has been developed and tuned to work better with certain data conditions, i.e. balance of target categories, ranges of base features, etc. The predicting power of different algorithms can be combined with the use of meta-learners (Vanschoren, 2019). Meta-learners use the properties or predictions from other algorithms (base learners) as additional information during their training stages. A simple implementation of this procedure is called Generalised Stacking (Wolpert, 1992) which can be interpreted as the addition of priors to the model training stage. Generalised stacking has been applied in several astrophysical problems. That is the case of Zitlau et al. (2016), Cunha and Humphrey (2022), and Euclid Collaboration et al. (2023a), Euclid Collaboration et al. (2023b).

Base and meta learners have been selected based upon the metrics described in Sect. **??**. We have trained five algorithms with the training subset and calculated the metrics for all of them using a 10-fold cross-validation approach (e.g. Stone, 1974; Allen, 1974) over the same training subset. For each metric, the learners have been given a rank (from 1 to 5) and a mean value has been obtained from them. Out of the analysed algorithms, the one with the best overall performance (i.e. best mean rank) is selected to be the meta learner while the remaining four are used as base learners.

For the AGN-galaxy classification and radio detection problems, we tested five classification algorithms: Random Forest (`RF`; Breiman, 2001), Gradient Boosting Classifier (`GBC`; Friedman, 2001), Extra Trees (`ET`; Geurts et al., 2006), Extreme Gradient Boosting (`XGBoost`, `v1.5.1`; Chen and Guestrin, 2016), and Category Boosting (`CatBoost`, `v1.0.5`; Prokhorenkova et al., 2018; Dorogush et al., 2018). For the redshift prediction problem, we tested five regressors as well: `RF`, `ET`, `XGBoost`, `CatBoost`, and Gradient Boosting Regressor (`GBR`; Friedman, 2001). We have used the Python implementations of these algorithms and, in particular for `RF`, `ET`, `GBC`, and `GBR`, the versions offered by the package `scikit-learn`[1] (`v0.23.2`; Pedre-

[1] `https://scikit-learn.org`

gosa et al., 2011). These algorithms were selected given that they offer tools to interpret the global and local influence of the input features in the training and predictions (cf. Sects. **??** and **??**).

All the algorithms selected for this work fall into the broad family of Tree-Based models. Forest models (`RF` and `ET`) rely on a collection of decision trees to, after applying a majority vote, predict either a class or a continuum value. Each of these decision trees uses a different, randomly-selected sub-set of features to make a decision on the training set (Breiman, 2001). Opposite to forests, Gradient Boosting models (`GBC`, `GBR`, `XGBoost` and `CatBoost`) apply decision trees sequentially to improve the quality of the previous predictions (Friedman, 2001; Friedman, 2002).

5.2 Model training

The procedure described in Sect. **??** includes an initial fit of the selected algorithms to the training data (including the selected features) to optimise their parameters. The stacking step includes a new optimisation of the parameters of the meta-learner using 10-fold cross-validation on the training data with the addition of the output from the base learners, which are treated as regular features. Then, the hyper-parameters of the stacked models are optimised over the training sub-set (a brief description of this step is presented in Appendix **??**).

The final step involves a last parameter fitting instance but using, this time, the combined train+validation subset, which includes the output of the base algorithms, to ensure wider coverage of the parameter space and better-performing models. Consequently, only the testing set is available for assessing the quality of the predictions made by the models.

Feature selection was applied to the train+validation subset with 85 488 confirmed elements (galaxies from SDSS DR16 and AGN from MQC, i.e. `class == 0` or `class == 1`). After the selection procedure described in Sect. **??**, 18 features were selected for training: `band_num`, `W4mag`, `g_r`, `r_i`, `r_J`, `i_z`, `i_y`, `z_y`, `z_W2`, `y_J`, `y_W1`, `y_W2`, `J_H`, `H_K`, `H_W3`, `W1_W2`, `W1_W3`, and `W3_W4`. The target feature is `class`.

The results of model testing for the AGN-galaxy classification are reported in Table 5.1. The `CatBoost` algorithm provides the best metric values (highest mean rank) and is therefore selected as the meta-model. `XGBoost`, `RF`, `ET`, and `GBC` were used as base learners.

Training of the radio detection model was applied only to sources confirmed to be AGN (`class == 1`). Feature selection was applied to the train+validation subset, with 36 387

Table 5.1: Best performing models for the AGN-galaxy classification

Model	F_β (×100)	MCC (×100)	Precision (×100)	Recall (×100)	Rank
CatBoost	95.70±0.28	92.46±0.48	95.45±0.32	95.91±0.37	1.00
XGBoost	95.67±0.27	92.40±0.48	95.41±0.39	95.88±0.34	2.00
RF	95.52±0.36	92.14±0.63	95.28±0.46	95.71±0.40	3.00
ET	95.40±0.40	91.94±0.69	95.13±0.43	95.63±0.47	4.00
GBC	95.26±0.31	91.66±0.54	94.82±0.41	95.63±0.35	5.00

Table 5.2: Best performing models the radio detection classification.

Model	F_β (×100)	MCC (×100)	Precision (×100)	Recall (×100)	Rank
XGBoost	29.98±2.29	29.81±2.17	56.74±2.93	21.61±2.00	2.75
CatBoost	29.57±1.62	30.56±1.71	60.10±2.85	20.85±1.36	2.25
GBC	29.60±1.66	31.31±1.93	62.55±3.95	20.66±1.40	1.75
RF	29.16±2.47	30.26±2.65	60.03±3.73	20.48±1.96	3.75
ET	28.40±1.27	29.73±1.47	60.06±2.85	19.80±1.05	4.50

confirmed AGN. The target feature is `LOFAR_detect` and the base of selected features are: `band_num`, `W4mag`, `g_r`, `g_i`, `r_i`, `r_z`, `i_z`, `z_y`, `z_W1`, `y_J`, `y_W1`, `J_H`, `H_K`, `K_W3`, `K_W4`, `W1_W2`, and `W2_W3`.

The performance of the tested algorithms is shown in Table 5.2. In this case, `GBC` shows the highest mean rank. For this reason, we used it as the meta-learner and `XGBoost`, `CatBoost`, `RF`, and `ET` were selected as base-learners.

The redshift value prediction model was applied to sources confirmed to be radio-detected AGN (i.e. `class == 1` and `radio_detect == 1`). Feature selection (cf. Sect. **??**) was applied to the train+validation subset, with 4612 sources, leading to the selection of 17 features. The target feature is `Z` and the selected base features are `band_num`, `W4mag`, `g_r`, `g_W3`, `r_i`,

Table 5.3: Results of initial fit for redshift value prediction.

Model	σ_{MAD} ($\times$100)	σ_{NMAD} ($\times$100)	σ_z ($\times$100)	σ_z^{N} ($\times$100)	η ($\times$100)	Rank
RF	17.88±1.41	07.95±0.50	42.02±5.28	19.38±2.44	19.51±1.98	2.0
ET	18.53±1.03	08.42±0.43	41.12±4.16	18.65±2.26	19.24±1.16	1.8
CatBoost	21.71±1.38	10.08±0.47	40.35±3.03	18.52±1.39	21.93±1.55	2.2
XGBoost	22.89±1.05	10.84±0.78	43.14±3.99	19.62±1.78	24.15±1.84	4.0
GBR	27.73±1.57	12.72±0.74	44.82±3.80	20.41±1.67	28.67±2.25	5.0

r_z, i_z, i_y, z_y, y_J, y_W1, J_H, H_K, K_W3, K_W4, W1_W2, and W2_W3.

For the redshift prediction, the tested algorithms performed as shown in Table 5.3. Based on their mean rank values, RF, CatBoost, XGBoost, and GBR were selected as base learners and ET (which shows the best σ_{MAD} value of the two models with the best rank) was used as meta-learner.

5.2.1 Hyperparameters optimisation

In Table 5.4, we present the optimised hyper-parameters from our meta-learners. For all three instances of modelling (AGN-Galaxy, radio detection, and redshift), hyper-parameters were optimised using the SkoptSearch algorithm embedded in the package tune-sklearn[2] (v0.4.1; Head et al., 2021), which implements a Bayesian search in the hyper-parameter space.

5.2.2 Models calibration

In general, classifiers deliver scores in the range [0, 1], which could be associated to the probability of a studied source being part of the relevant class (in our work, AGN or radio detectable). The classifier uses a threshold above which, any predicted element would be considered a positive instance.

With the exception of few algorithms (including the family of logistic regressions), scores from classifiers cannot be directly used as probabilities. As a consequence of this inability, such values cannot be compared from one type of model to some other and can not be combined to

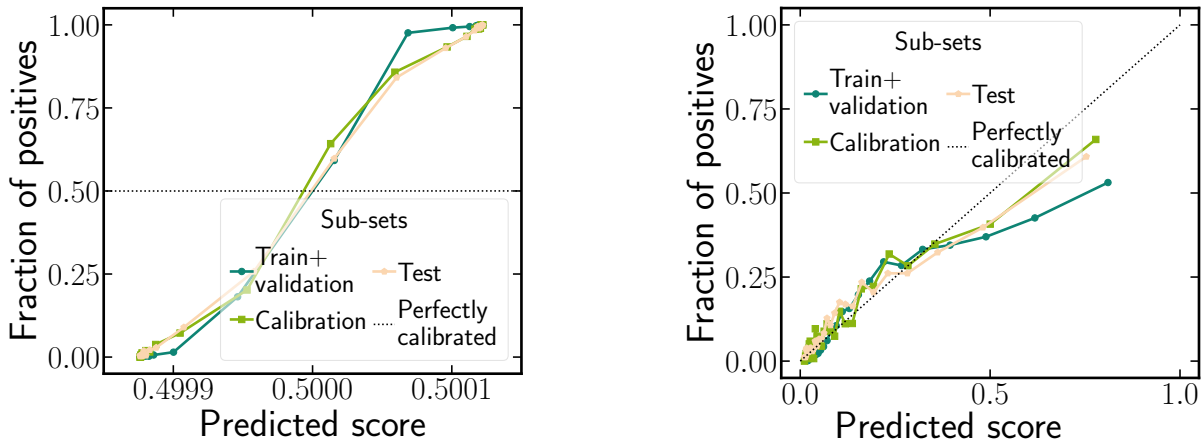
[2] https://github.com/ray-project/tune-sklearn

Table 5.4: Hyper-parameters values for meta-learners after tuning.

AGN-Galaxy model (CatBoost)			
Parameter	Value	Parameter	Value
learning_rate	0.0075	random_strength	0.1
depth	6	l2_leaf_reg	10

Radio detection model (GradientBoosting)			
Parameter	Value	Parameter	Value
n_estimators	187	min_samples_leaf	2
learning_rate	0.0560	max_depth	9
subsample	0.3387	max_features	0.5248
min_samples_split	5		

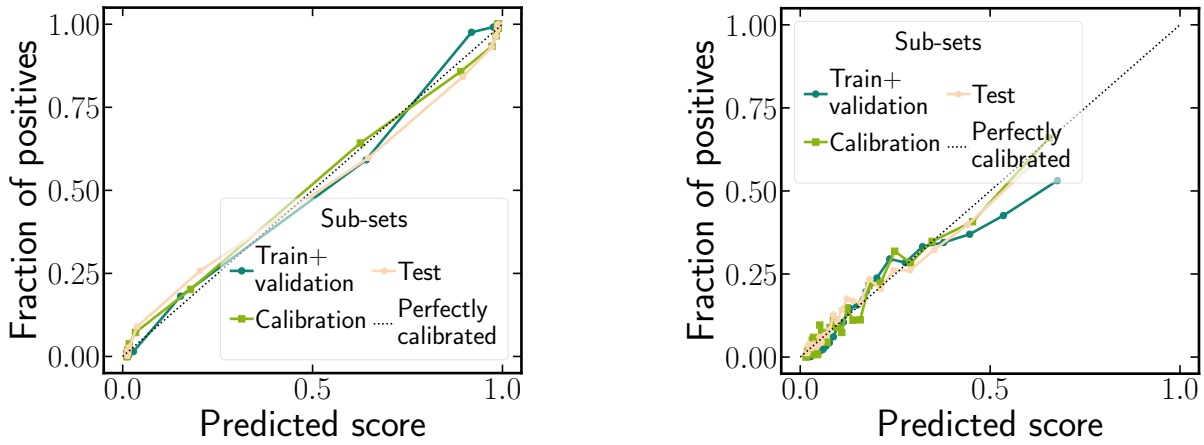
Redshift prediction model (ET)			
Parameter	Value	Parameter	Value
n_estimators	100	criterion	mae
max_depth	None	min_samples_split	2
max_features	auto	min_samples_leaf	1
bootstrap	False		



(a) AGN-galaxy

(b) Radio detection

Figure 5.1: Reliability curves for uncalibrated classifiers. Each line represents the calibration curve for each subset in HETDEX field. Data has been binned and each bin (represented by the points) has the same number of elements per curve. Dashed line represents a perfectly calibrated model. (a) AGN-galaxy classification model. (b) Radio detection model.



(a) AGN-galaxy

(b) Radio detection

Figure 5.2: Reliability curves for calibrated classifiers. (a) AGN-galaxy classification model. (b) Radio detection model. Details as in Fig. 5.1.

obtain a joint score. Therefore, in order to retrieve joint scores and treat them as probabilities, scores (and, by extension, the classifiers) need to be calibrated. This calibration means that, when taking all predictions with a probability P of being of a class, a fraction P of them really belong to that class (e.g. Lichtenstein et al., 1982; Silva Filho et al., 2023).

Calibration of these scores can be done by applying a transformation to their values. For our work, we will apply a Beta transformation. It allows to re-distribute the scores of a classifier allowing them to get closer to the definition of probability (Kull et al., 2017a; Kull et al., 2017b). Calibration steps in our workflow have been applied using the Python package `betacal`. In the case of the radio detection model, the new scores have a wider range than the original, uncalibrated scores.

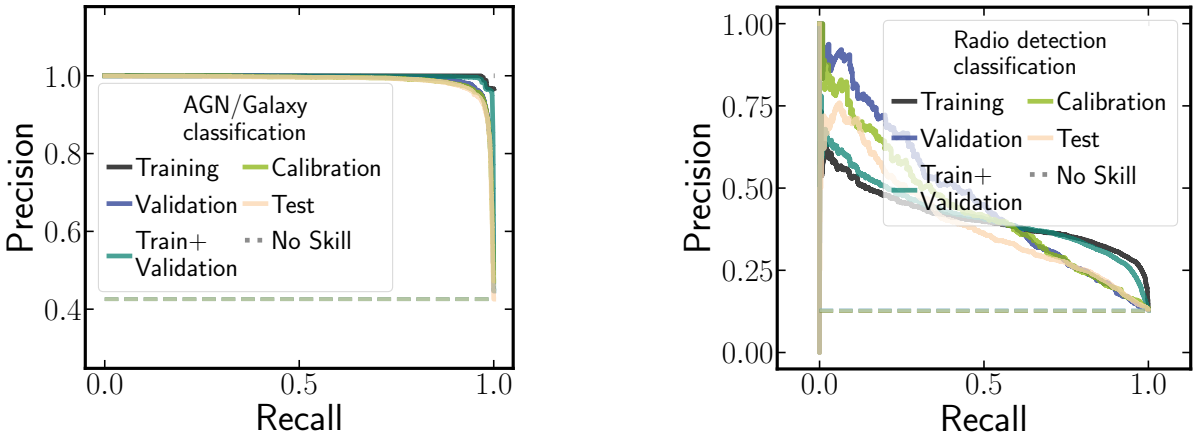
When obtaining the BSS values for both classification, the AGN-Galaxy classifier has a score of BSS = −0.002, demonstrating that no major changes were applied to the distribution of scores. For the radio detection classifier, the score is BSS = −0.434. Even though the BSS value is slightly negative for the AGN-Galaxy classifier, we keep it since its range of values now can be compared and combined with additional probabilities. In the case of the radio detection classifier, the BSS shows a degradation of the calibration, but we will keep the calibrated model given that it provides, overall, better values for the remaining metrics. This effect can be seen, more strongly, with recall.

Calibration (or reliability) plots show how well calibrated the predicted scores of a classifier are by displaying the fraction of sources that are part of a given class as a function of the predicted probability. A perfectly calibrated classifier would have all its prediction lying in the x=y line. The magnitude of the deviations from that line give information of the miscalibration a model has (see, for instance, Bröcker and Smith, 2007; Van Calster et al., 2019). In Fig. 5.1, we present the reliability curves for the uncalibrated classifiers and, in Fig. 5.2, for their calibrated versions.

5.2.3 Threshold selection

The optimisation of the PR curve for the calibrated predictor provides an optimal threshold for this algorithm of 0.34895. This value was used for the AGN-Galaxy model throughout this work.

By maximising the recall (Eq. 3.4), we improve the number of recovered elements in each classifier. This can be done by decreasing the threshold by which a source is classified



(a) AGN-galaxy

(b) Radio detection

Figure 5.3: Precision-Recall curves for the (a) AGN-Galaxy and (b) radio detection classification models.

as a positive instances. Setting this threshold to its minimum, 0.0, would increase the recall. But every source would be predicted to be an AGN or detected on the radio regardless of their properties.

One statistical tool designed to optimise the classification threshold taking into account the overall model performance is the Precision-Recall (PR) curve. It can help to understand the behaviour of a classifier as a function of its threshold. Both quantities, precision (Eq. 3.5) and recall, show an inverse correlation, and both depend on the selected threshold. Thus, they can be used to retrieve the score value for which both quantities are balanced. This optimisation is done by finding the threshold that maximises the F_β score (Eq. 3.2). This operation will be performed over the union of training and validation sets, which have been used to create and train each model. PR curves for all subsets used in our classification models are shown in Fig. 5.3.

6

Prediction of radio-AGN candidates

After training the models and tuning their hyperparameters, we were able to use them for predicting values in a data set which was not used in the previous stages. In our case, this is the validation sub-set, which is different for each ML problem (classification or regression). In the case of the redshift value prediction, some of the sources in the validation sub-set might not have an original redshift (i.e. AGN without redshift listed in the MQC. See Sect. **??**). This does not prevent the model to predict a value, although it will not have a previous value to be compared with.

6.1 AGN-galaxy classification

The application of the stacked model for the prediction of the AGN detection of the validation sub-set is summarised in Table 6.1 along with the results from train and test sub-sets. In a similar way, the confusion matrix derived from the prediction results over the validation sub-set is shown in Fig. **??**.

It is possible to see that the MCC scores for the three sub-sets are in similar levels. That might be an indication of a good training process, in which no substantial over-fitting can be detected.

A closer inspection to the confusion matrix in Fig. 6.1 shows that close to a 45% of the AGN from the MQC were discarded by our model. And less than 26% of the predicted AGN are not labelled as such by the MQC. An in-depth analysis of these results is presented in Sect. **??**.

6.2 Radio detection classification

The application of the stacked model for the prediction of the radio detection of the training, testing, and validation sub-set is summarised in Table 6.2. Similarly, the confusion matrix derived from the prediction results over the validation sample is shown in Fig. **??**.

The results from the training sub-set show an almost perfect classification, which might

Table 6.1: Results of application of AGN detection model to training, test, and validation sub-sets and to Stripe 82 sample as part of the pipeline described in Sect. **??**.

Sub-set	F1	MCC	Recall
Training	0.7036	0.7278	0.5585
Test	0.7000	0.7233	0.5573
Validation	0.6237	0.6293	0.5431
S82-pipeline	0.5219	0.5607	0.3795

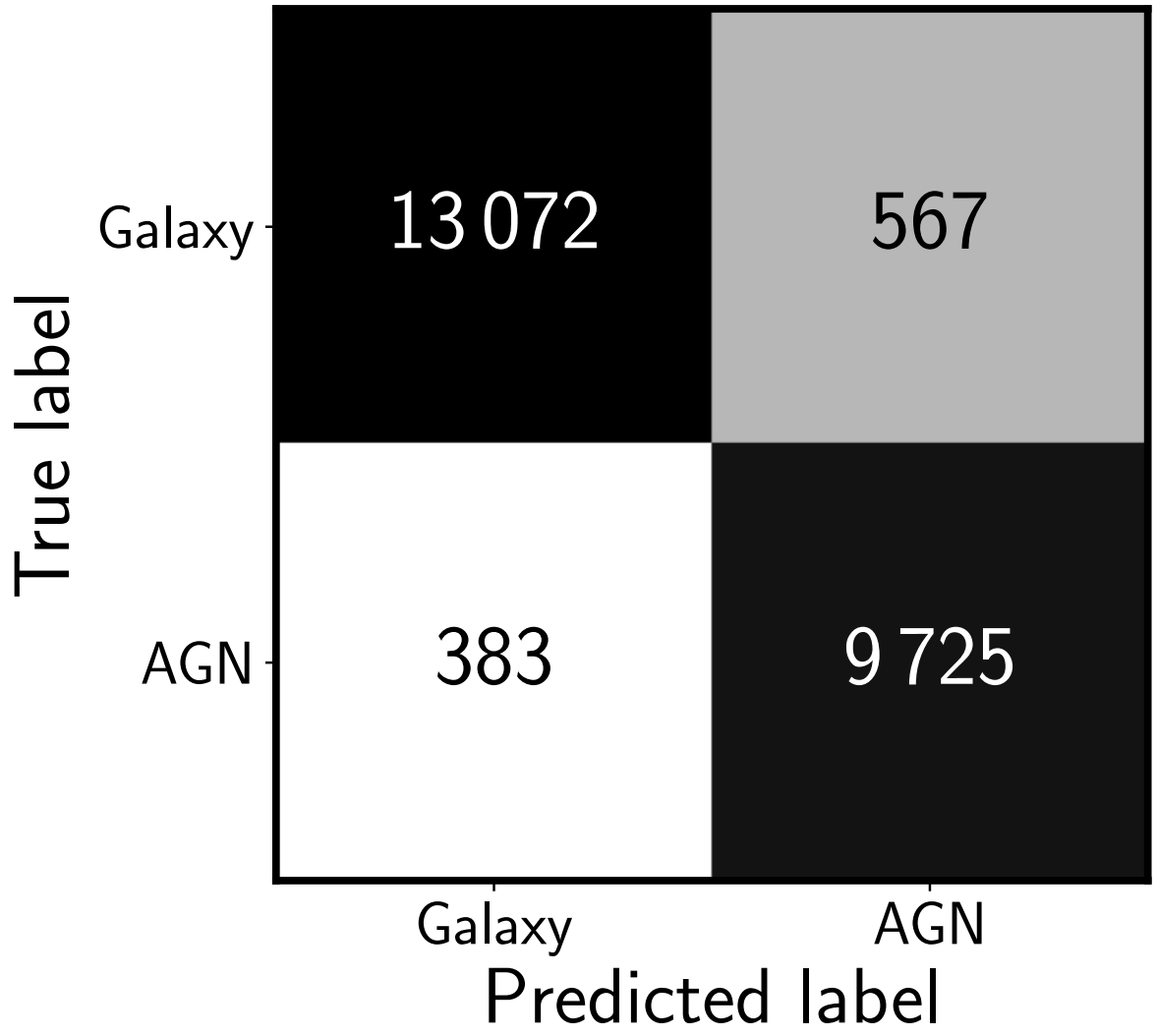


Figure 6.1: Confusion matrix from the results of application of radio detection prediction model to validation sub-set.

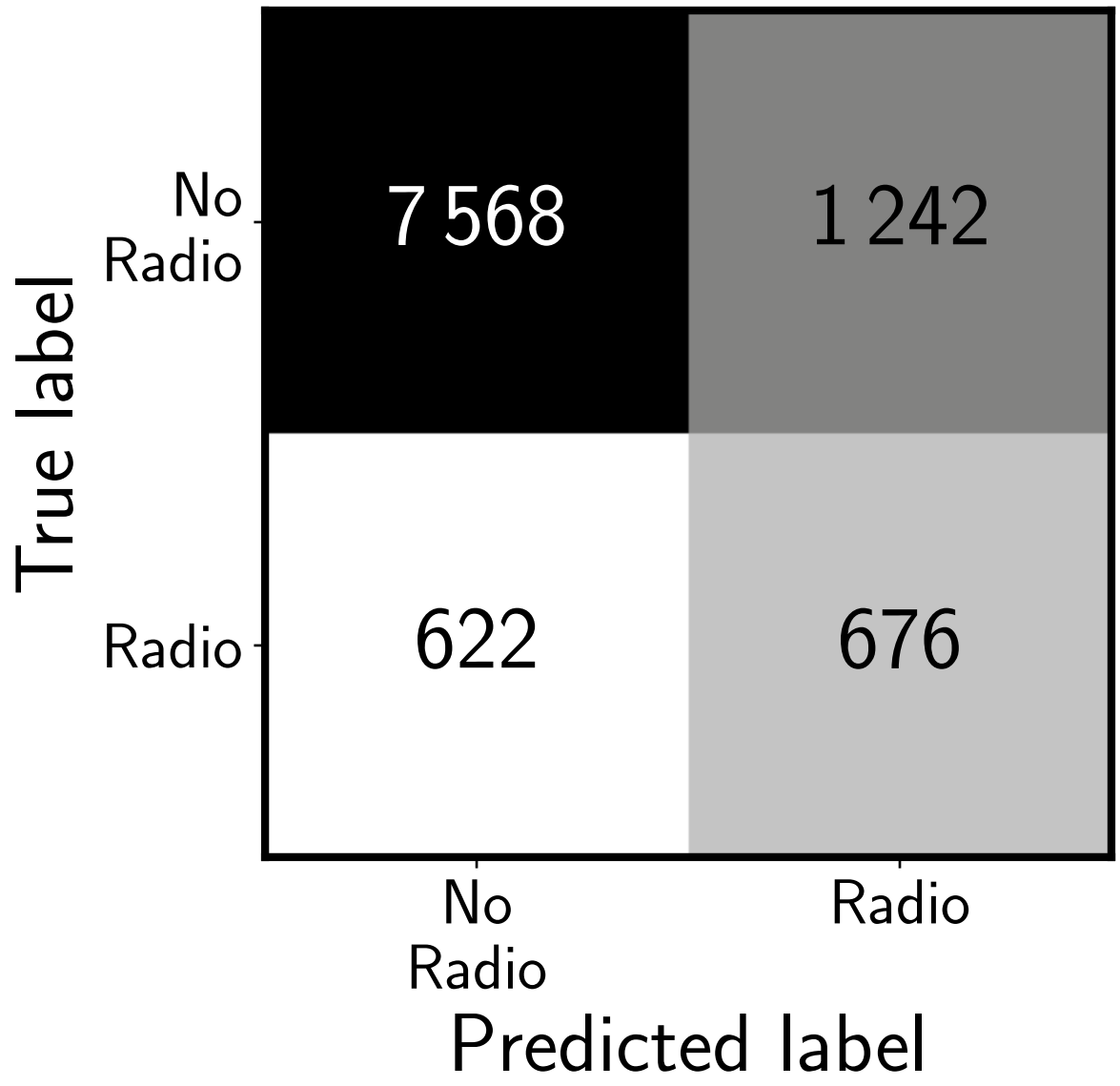


Figure 6.2: Confusion matrix from the results of application of radio detection prediction model to validation sub-set.

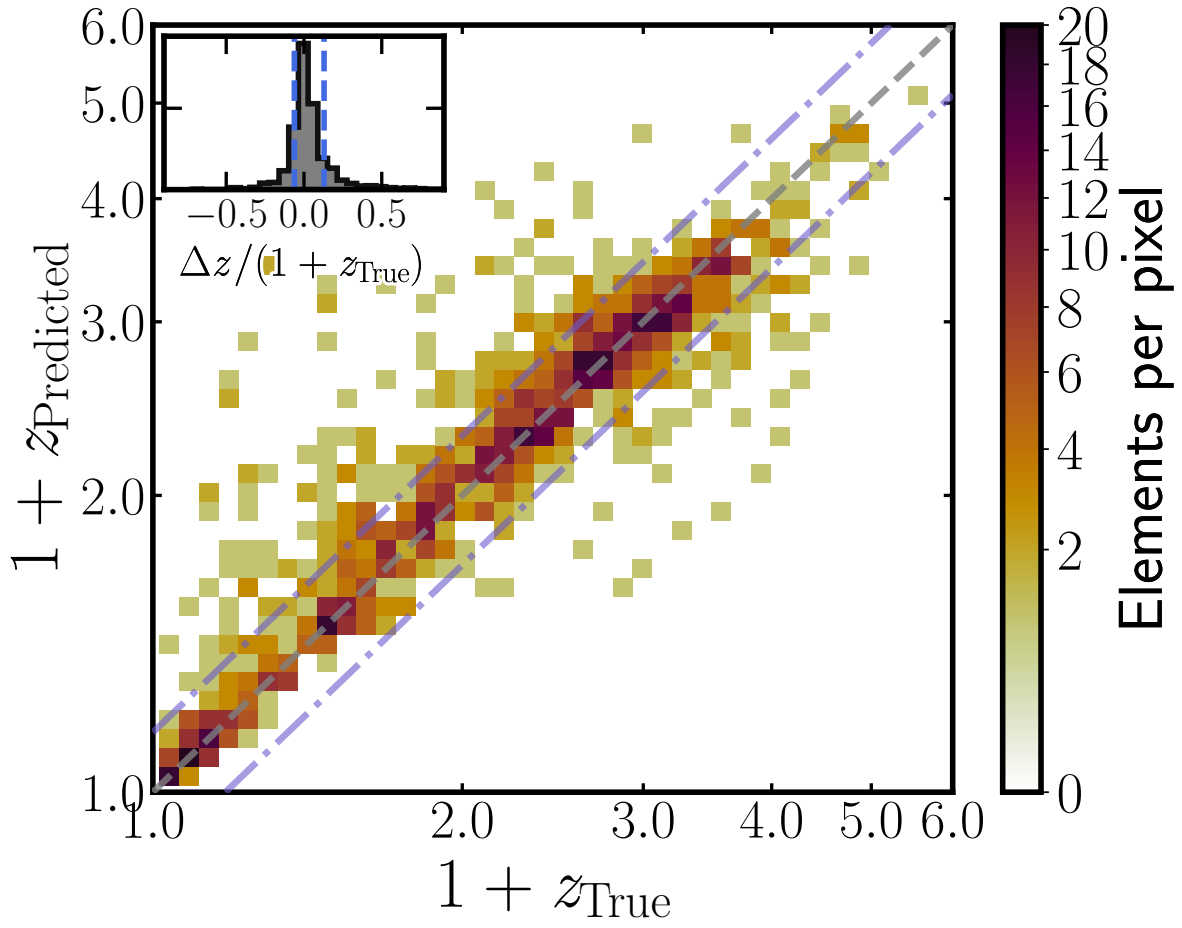


Figure 6.3: Density plot of comparison between original and the predicted redshifts from the results of application of redshift prediction model to validation sub-set. Each point is colour-coded following the colorbar. Grey, dashed line shows $x = y$ relation and purple, dot-dashed lines show the limits where outliers are defined (cf. Eqn. 3.17). Inset shows the distribution of $\Delta z^{\rm N}$ values from the points shown in main plot.

Table 6.2: Results of application of radio detection model to training, test, and validation sub-sets and to Stripe 82 sample as part of the pipeline described in Sect. **??**.

Sub-set	F1	MCC	Recall
Training	0.7830	0.7203	0.6729
Test	0.7797	0.7156	0.6698
Validation	0.6186	0.4674	0.5592
S82-pipeline	0.6092	0.4096	0.6004

Table 6.3: Results of application of redshift prediction model to training, test, and validation sub-sets and to Stripe 82 sample as part of the pipeline described in Sect. **??**.

Sub-set	$\sigma_{\rm MAD}$	$\sigma_{\rm NMAD}$	σ_z	$\sigma_z^{\rm N}$	η
Train	0.0835	0.0388	0.2307	0.0906	0.0742
Test	0.0825	0.0380	0.2196	0.0930	0.0731
Validation	0.1767	0.0793	0.3521	0.1384	0.2115
S82-pipeline	0.1302	0.0654	0.2700	0.1462	0.1392

be due, in part, to the low number of sources used in this stage and to the information delivered to the model by the selected features. A thorough interpretation can be seen in Sect. **??**. Despite this, the confusion matrix of Fig. **??** shows that a rather large fraction of AGN previously detected in the radio (Recall = 55.92%) were predicted to have that behaviour.

6.3 Redshift prediction

In the case of redshift values prediction, the application of the stacked model over the Validation sub-set is summarised in Table 6.3. Likewise, the comparison between the original redshift values and those derived from the prediction results is shown in Fig. **??**.

The results in Table 6.3 show some degree of over-fitting, since the validation scores are a factor of two worse than those from the training and test sub-sets. This happens for all used metrics. In Sect. **??**, a detailed description of these results is presented.

A different approach to display the results of the prediction over the validation sub-set can be seen in Fig. 6.4. There, the histograms of both true and predicted redshift values are shown. This allows for the assessment of sources that, originally, do not have a redshift measurement in the MQC data set. In this case, it is possible to see that both distributions share a similar shape. Apart from the number of sources, they present peaks in similar ranges, suggesting that the trained model is able to retrieve sensible redshift values.

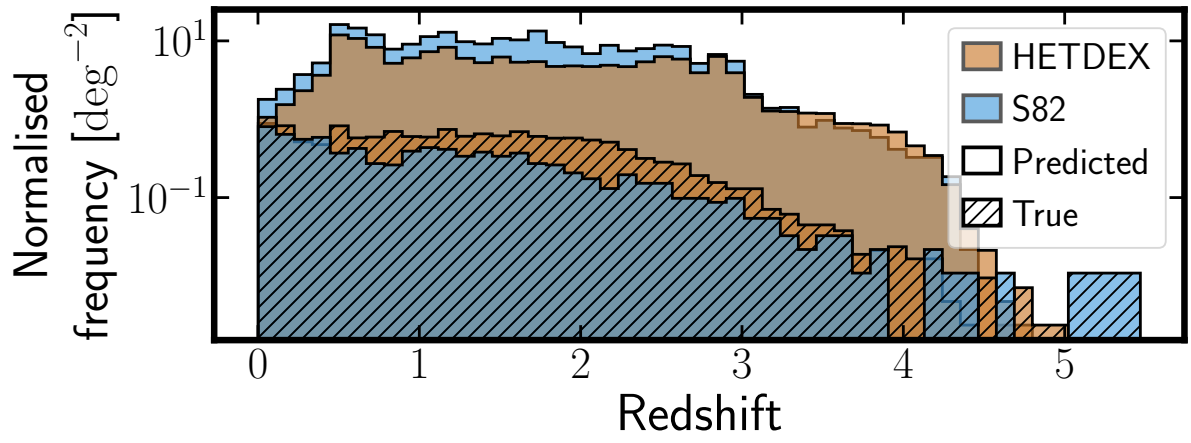


Figure 6.4: Histogram of predicted reshifts for the validation sub-set of the HETDEX catalogue. The green histogram shows the values for the redshifts from the Million Quasar Catalog (MQC), while the purple histogram shows the values for all the predicted redshifts.

6.4 Pipeline prediction

The models described in previous sections can be used, sequentially, to create a pipeline for the prediction of radio-detected AGN and their redshift values from sources that have been detected in different wavelengths.

In order to test such models, we used data from a different area of the sky. The SDSS Stripe 82 Field has been covered by several surveys and observations and can be used as a place to test the application of a ML model in a different area from where it was trained. We collected data from the same surveys as described in Sect. **??** with one important caveat: this field is not covered by the LoTSS-DR1 Survey and, thus, we cannot define the studied area from it. For this reason, the selected area is defined by the coverage of the VLA SDSS Stripe 82 Survey (Hodge et al., 2011). The sample has data from 369,093 objects in an area of 92 deg^2 and 2,537 of these sources have been labelled as AGN by the MQC. From these AGN, 788 of them show radio detection in one of the surveys used in this work.

The models described in Sect. **??** were applied to the Stripe 82 sample as follows. The AGN detection model was first applied to the full data set. Then, we selected the sources that were predicted to be AGN, regardless of their original classification, and we applied the radio-detection model to them. The sources predicted to be detected in the radio were then selected and the redshift prediction model was applied to them. A schematic view of this process can be seen in Fig. 6.5. As the goal of the model pipeline is the prediction of radio-detected AGN, we can discard sources that have been not predicted as such without the fear of loosing relevant

Table 6.4: Results of application of radio AGN prediction pipeline to the full Stripe 82 sample.

Sub-set	F1	MCC	Recall
S82-pipeline	0.2904	0.3196	0.2038

information.

The application of the first model to predict AGN led to 1,092 sources labelled as prospective AGN. And the use of the radio detection model over these 1,092 sources predicted that 308 of them might have detection in radio bands.

The metrics for the successive application of models to the Stripe 82 data are shown (labelled as S82-pipeline) in Tables 6.1, 6.2, and 6.3. And the graphical representation of the same results is shown in Fig. **??**. It is worth mentioning that not all sources in the Stripe 82 sample have redshift measurements. Thus, Fig. **??** and Table 6.3 only show values for those elements that can be compared (i.e. those with a redshift measurement listed in the MQC).

If we consider, in our data set, radio AGN as one category, we can produce a confusion matrix and its metric values for the full Stripe 82 sample. Such confusion matrix can be seen in Fig. 6.6. It can be understood as a summary of both confusion matrices in Figures **??** and **??**. For the purposes of constructing such combined matrix and following the design of our pipeline (cf. Fig. 6.5), all sources that were not predicted to be AGN by the first model were labelled as not having any chance of having a radio detection.

The metrics derived from these joint results are presented in Table 6.4. As expected, the numbers are worse than the results from the application of the individual models. In this case, we are testing the prediction of two labels at the same time to the full sample, and not to a sub-sample that met some specific condition, which might add uncertainty to the results.

From 788 sources labelled, originally, as radio AGN, 133 were predicted, by our pipeline, to be in such category. And 175 new candidates have been generated (i.e. sources that do not fulfil the condition `is_AGN == 1 AND radio_detect == 1`).

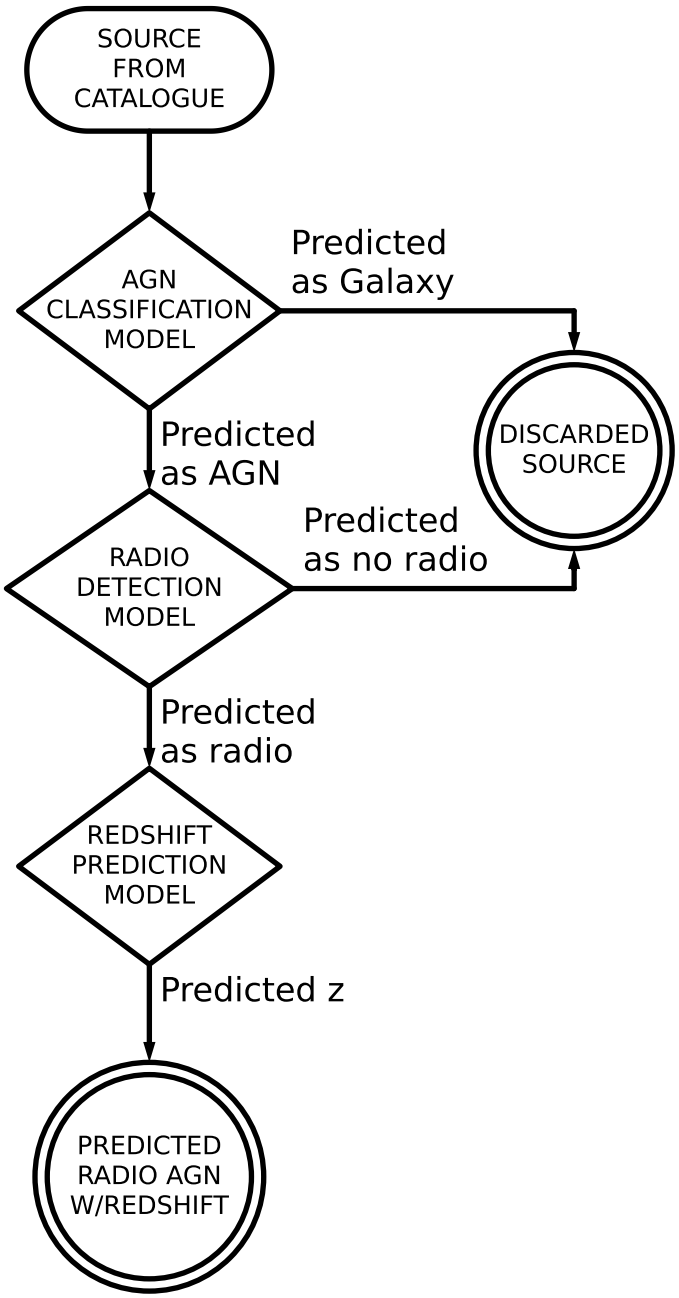


Figure 6.5: Flowchart representing the prediction pipeline used in Stripe 82 to predict the presence of radio-detected AGN and their redshift values.

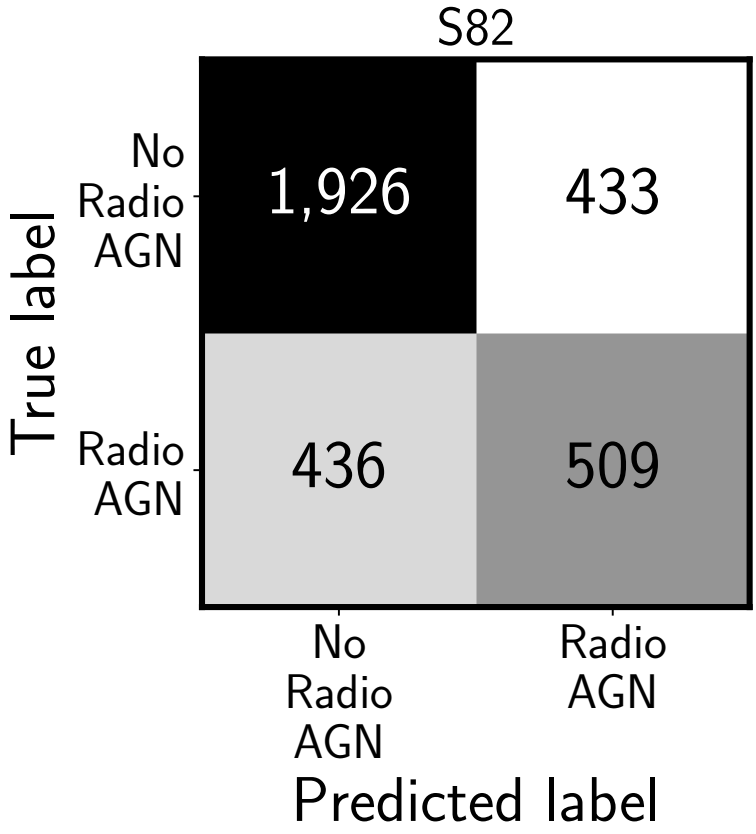


Figure 6.6: Confusion matrix for the full radio AGN detection prediction in the Stripe 82 field. It might be regarded as a summary of Figs. **??** and **??**.

This page intentionally left blank.

7

Analysis of prediction method

7.1 Comparison with previous works

In this section, we provide a few examples of related published works as well as plausible explanations for observed discrepancies when these are present. This comparison attempts to be representative of the literature on the subject but does not intends to be complete in any way.

7.1.1 AGN detection prediction

We separate the comparison with previously published results between traditional and ML methodologies in order to understand the significance of our results and ways for future improvement.

Traditional AGN selection methods are based on the comparison of the measured Spectral Energy Distribution (SED) photometry to a template library (Walcher et al., 2011). A recent example of its application is presented by Thorne et al. (2022b) where best fit classifications were calculated for more than 700 000 galaxies in the D10 field of the Deep Extragalactic VIsible Legacy Survey (DEVILS; Davies et al., 2018) and the Galaxy and Mass Assembly survey (GAMA; Driver et al., 2011; Liske et al., 2015). The 91% recovery rate of AGN, selected through various means (X-ray measurements, narrow and broad emission lines, and mid-infrared colours), is very much in line with our findings in S82, where our rate (recall) reaches 89%.

Traditional methods also encompass the colour-based selection of AGN. While less precise, they provide access to a much larger base of candidates with a very low computational cost. We implemented some of the most common colour criteria on the data from S82. Of particular interest is the predicting power of the mid-IR colour selection due to its potential to detect hidden or heavily obscured AGN activity. Based on WISE (Wright et al., 2010) data, Stern et al. (2012, S12) proposed a threshold at W1 - W2 $\geq$ 0.8 to separate AGN from non-AGN using data from AGN in the COSMOS field (Scoville et al., 2007). A more stringent criterion

was developed by Mateos et al. (2012, p. M12), the AGN wedge, which can be defined by the sources located inside the region defined by the relations W1 - W2 $< 0.315 \times$ (W2 - W3) $+ 0.791$, W1 - W2 $> 0.315 \times$ (W2 - W3) $- 0.222$, and W1 - W2 $> -3.172 \times$ (W2 - W3) $+ 7.624$. In order to define this wedge, they used data from X-ray selected AGN over an area of 44.43 deg^2 in the northern sky. Mingo et al. (2016, p. M16) cross-correlated data from WISE observations with X-ray and radio surveys creating a sample of star-forming galaxies and AGN in the northern sky. They developed individual relations to separate classes of galaxies and AGN in the W1 - W2, W2 - W3 space and, for AGN the criterion, the relation is W1 - W2 ≥ 0.5 and W2 - W3 < 4.4. More recently, Blecha et al. (2018, B18) analysed the quality of mid-IR colour selection methods for the identification of obscured AGN involved in mergers. Using hydrodynamic simulations for the evolution of AGN in galaxy mergers, they developed a selection criterion from WISE colours which is shown to be able to separate, with high reliability, starburst galaxies from AGN. The expressions have the form W1 - W2 > 0.5, W2 - W3 > 2.2, and W1 - W2 $> 2 \times$ (W2 - W3) $- 8.9$.

The results from the application of these criteria to our samples in the testing subset and in the labelled sources of S82 field are summarised in Table 7.1 and a graphical representation of the boundaries they create in their respective parameter spaces is presented in Fig. 7.1.

Table 7.1 shows that previous colour-colour criteria have been designed and calibrated to have very high precision values. Most of the sources deemed to be AGN by them are, indeed, of such class. Despite being tuned to maximise their recall (and F_β to a lesser extent), our classifier, and the criterion derived from it, still show precision values compatible with those of such criteria. This result underlines the power of ML methods. They can be on a par with traditional colour-colour criteria and excel in additional metrics.

Figure 7.1 is constructed as a confusion matrix, plotting in each quadrant the whole WISE population in the background and in colour contours the corresponding fraction of the testing set (TP, TN, FP, and FN, see Fig. **??**a and Sect. **??**). As expected, our pipeline is able to separate with high confidence sources which are closer to the AGN or the galaxy locus (TP and TN) while sources in the FN and FP quadrant show a different situation. AGN predicted to be galaxies (FN, 1.6% of sources for HETDEX, and 4.9% for S82) are located in the galaxy region of the colour-colour diagram. On the opposite corner of the plot, galaxies predicted to be AGN (FP, 2.4% of sources for HETDEX, and 4.2% for S82) cover the areas of AGN and galaxies uniformly. FN sources might be sources that are identified as AGN by means not included in our feature set (e.g. X-ray, radio emission). FP sources, alternatively, might be galaxies with

Table 7.1: Results of application of several AGN detection criteria to our testing subset and the labelled sources from the S82 field.

HETDEX test set				
Method	F_β (×100)	MCC (×100)	Precision (×100)	Recall (×100)
S12	86.10	78.78	93.98	80.51
M12	51.80	49.71	98.87	37.18
M16	67.21	61.30	97.48	53.48
B18	82.14	75.76	97.54	72.66
This work	92.71	87.64	94.00	91.67

S82 (labelled)				
Method	F_β (×100)	MCC (×100)	Precision (×100)	Recall (×100)
S12	83.59	45.47	93.93	76.62
M12	46.80	28.22	99.59	32.54
M16	64.69	37.76	98.80	50.32
B18	79.71	51.07	98.72	68.77

extreme properties, similar to AGN.

For the case of ML-based models for AGN-galaxy classification, several analyses have been published in recent years. An example of their application is provided in Clarke et al. (2020) where a Random Forest model for the classification of stars, galaxies and AGN using photometric data was trained from more than 3 000 000 sources in the SDSS (DR15; Aguado et al., 2019) and WISE with associated spectroscopic observations. Close to 400 000 sources have a quasar spectroscopic label and from the application of their model to a validation subset, they obtain a recall of 0.929 and F1-score of 0.943 for the quasar classification. These scores are of the same order as the ones obtained when applying our AGN-Galaxy model to the testing

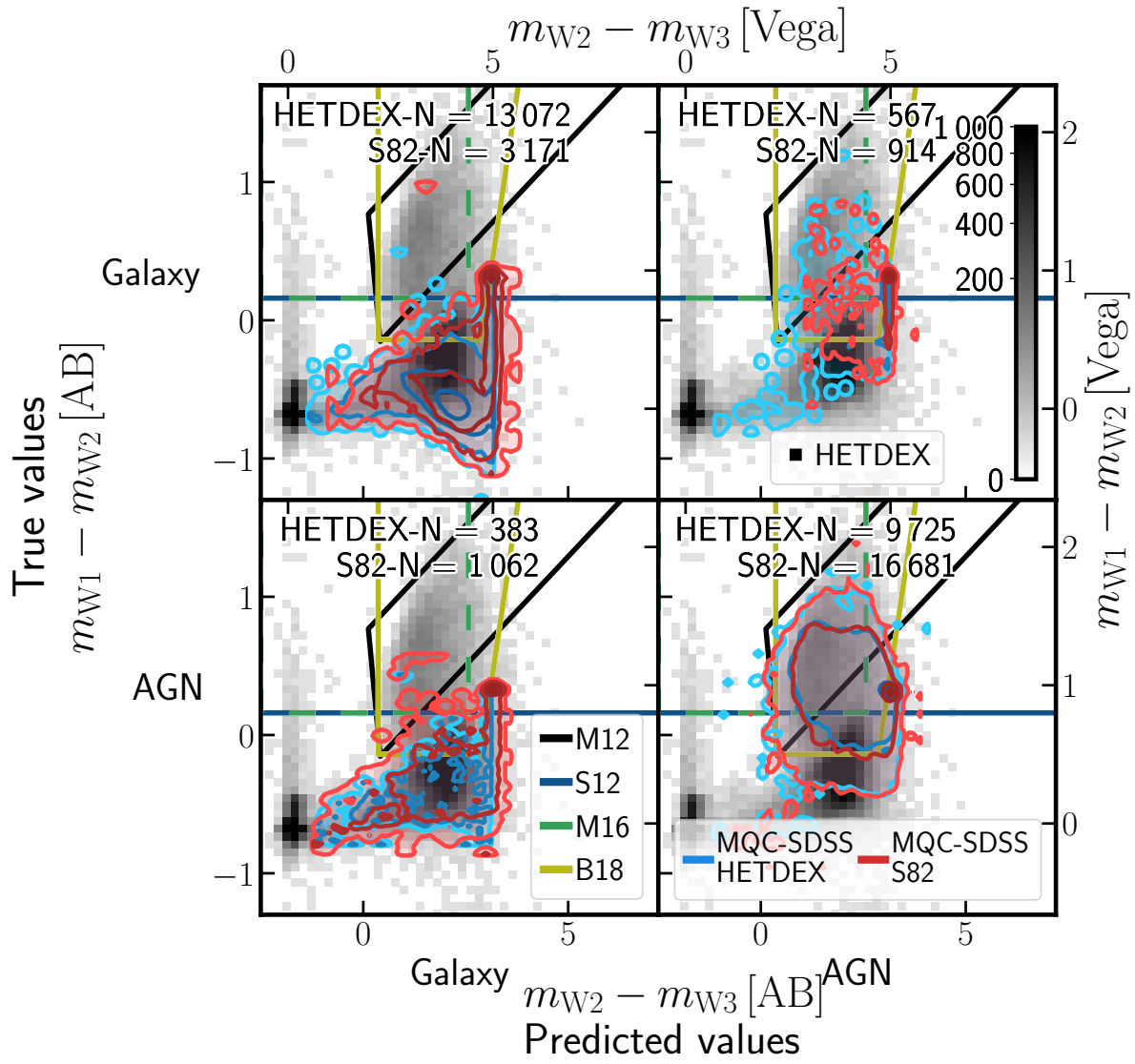


Figure 7.1: W1 - W2, W2 - W3 colour-colour diagrams for sources in the testing subset, from HETDEX, and labelled sources from S82 given their position in the AGN-galaxy confusion matrix (see, for HETDEX, rightmost panel of Fig. **??**). In the background, density plot of all CW-detected sources in the full HETDEX field sample is displayed. Colour of each square represents the number of sources in that position of parameter space, with darker squares having more sources (as defined in the colorbar of the upper-right panel). Contours represent distribution of sources for each of the aforementioned subsets at 1, 2, 3, and **4**σ levels (shades of blue, for testing set and shades of red for labelled S82 sources). Coloured, solid lines display limits from the criteria for the detection of AGN described in Sect. **??**.

set (see Table 6.1). Thus, and despite using an order of magnitude fewer sources for the full training and validation process, our model can achieve equivalently good scores.

Expanding on Clarke et al. (2020), Cunha and Humphrey (2022) built a ML pipeline, `SHEEP`, for the classification of sources into stars, galaxies and QSO. In contrast to Clarke et al. (2020) or the pipeline described here, the first step in their analysis is the redshift prediction, which is used as part of the training features by the subsequent classifiers. They extracted WISE and SDSS (DR15; Aguado et al., 2019) photometric data for almost 3 500 000 sources classified as stars, galaxies or QSO. The application of their pipeline to sources predicted to be QSO led to a recall of 0.960 and an F1 score of 0.967. The improved scores in their pipeline might be a consequence not only of the slightly larger pool of sources, but also the inclusion of the coordinates of the sources (RA, Dec) and the predicted redshift values as features in the training.

A test with a larger number of ML methods was performed by Poliszczuk et al. (2021). For training, they used optical and infrared data from close to 1 500 sources (galaxies and AGN) located at the AKARI North Ecliptic Pole (NEP) Wide-field (Lee et al., 2009; Kim et al., 2012) covering a 5.4 deg^2 area. They tested `LR`, `SVM`, `RF`, `ET`, and `XGBoost` including the possibility of generalised stacking. In general, they obtained results with F1-scores between 0.60 – 0.70 and recall values in the range of 50% – 80%. These values, lower than the works described here, can be fully understood given the small size of the training sample. A larger photometric sample covers a wider range of the parameter space which significantly helps the metrics of any given model.

7.1.2 Radio detection prediction

We have not found in the literature any work attempting the prediction of AGN radio detection at any level and therefore this is the first attempt at doing so. In the literature we do find several correlations between the AGN radio emission (flux) and that at other wavelengths (e.g. with infrared emission, Helou et al., 1985; Condon, 1992) and substantial effort has been done towards classifying radio galaxies based upon their morphology (e.g. Aniyan and Thorat, 2017; Wu et al., 2019, FRI, FRII, bent jets, etc.) and its connection to environment (Miley and De Breuck, 2008; Magliocchetti, 2022). None of these extensive works has directly focused on the a priori presence or absence of radio emission above a certain threshold. Therefore, the results presented here are the first attempt at such an effort.

The ~2x success rate of the pipeline to identify radio emission in AGN (~44.61% recall and ~32.20% precision; see Table 6.4) with the respect to a 'no-skill' or random ($\lessapprox$30%) selection, provides the opportunity to understand what the model has learned from the data and, therefore, gain some insight into the nature or triggering mechanisms of the radio emission. We, therefore, reserve the discussion of the most important features, and the linked physical processes, driving the pipeline improved predictions to Sect. **??**.

7.1.3 Redshift prediction

We compare our results to that of Ananna et al. (2017, Stripe 82X) where the authors analysed multi-wavelength data from more than 6 100 X-ray detected AGN from the 31.3 deg^2 of the Stripe 82X survey. They obtained photometric redshifts for almost 6 000 of these sources using the template-based fitting code `LePhare` (Arnouts et al., 1999; Ilbert et al., 2006). Their results present a normalised median absolute deviation of $\sigma_{\rm NMAD}$=0.062 and an outlier fraction of η=13.69%, values which are similar to our results in HETDEX and S82 except for a better outlier fraction (as shown in Table 6.3, we obtain $\eta_{S\,82}$ = 25.18%, $\sigma_{\rm NMAD}^{\rm HETDEX}$=0.071, and $\eta^{\rm HETDEX}$=18.9%).

On the ML side, we compare our results to those produced by Carvajal et al. (2021) in S82, with $\sigma_{\rm NMAD}$ = 0.1197 and η = 29.72%, and find that our redshift prediction model improves by at least 25% for any given metric. The source of improvement is probably many-fold. First, it might be related to the different sets of features used (colours vs ratios) and second, the more specific population of radio-AGN used to train our models. Carvajal et al. (2021) used a limited set of colours to train their model, while we have allowed the use of all available combinations of magnitudes (Sect. **??**). Additionally, their redshift model was trained on all available AGN in HETDEX, while we have trained (and tested) it only with radio-detected AGN. Using a more constrained sample reduces the likelihood of handling sources that are too different in the parameter space.

Another example of the use of ML for AGN redshift prediction has been presented by Luken et al. (2019). They studied the use of the k-nearest neighbours algorithm `KNN` (Cover and Hart, 1967), a non-parametric supervised learning approach, to derive redshift values for radio-detectable sources. They combined 1.4 GHz radio measurements, infrared, and optical photometry in the European Large Area ISO Survey-South 1 (ELAIS-S1; Oliver et al., 2000) and extended Chandra Deep Field South (eCDFS; Lehmer et al., 2005) fields, matching their

sensitivities and depths to the expected values in the Evolutionary Map of the Universe (EMU; Norris et al., 2011). From the different experiments they run, their resulting NMAD values are in the range $\sigma_{\rm NMAD} = 0.05 - 0.06$, and their outlier fraction can be found between $\eta = 7.35\%$ and $\eta = 13.88\%$. As an extension to the previous results, Luken et al. (2022) analysed multi-wavelength data from radio-detected sources the eCDFS and the ELAIS-S1 fields. Using `KNN` and `RF` methods to predict the redshifts of more than 1 300 RGs, they have developed regression methods that show NMAD values between $\sigma_{\rm NMAD} = 0.03$ and $\sigma_{\rm NMAD} = 0.06$, $\sigma_z = 0.10 - 0.19$, and outlier fractions of $\eta = 6.36\%$ and $\eta = 12.75\%$.

In addition to the previous work, Norris et al. (2019) compared a number of methodologies, mostly related with ML but also `LePhare`, for predicting redshift values for radio sources. They have used more than 45 photometric measurements (including 1.4 GHz fluxes) from different surveys in the COSMOS field. From several settings of features, sensitivities, and parameters, they retrieved redshift predictions with NMAD values between $\sigma_{\rm NMAD} = 0.054$ and $\sigma_{\rm NMAD} = 0.48$ and outlier fractions that range between $\eta = 7\%$ and $\eta = 80\%$. The broad span of obtained values might be due to the combinations of properties for each individual training set (including the use of radio or X-ray measurements, the selection depth, and others) and to the size of these sets, which was small for ML purposes (less than 400 sources). The slightly better results can be understood given the heavily populated photometric data available in COSMOS.

Specifically related to HETDEX, it is possible to compare our results to those from Duncan et al. (2019). They use a hybrid photometric redshift approach combining traditional template fitting redshift determination and ML-based methods. In particular, they implemented a Gaussian Process (GP) algorithm, which is able to model both the intrinsic noise and the uncertainties of the training features. Their redshift prediction analysis of AGN sources with a spectroscopic redshift detected in the LoTSS DR1 (6,811 sources) found a NMAD value of $\sigma_{\rm NMAD} = 0.102$ and an outlier fraction of $\eta = 26.6\%$. The differences between these results and those obtained from the application of our models (individually and as part of the prediction pipeline) might be due to the differences in the creation of the training sets. Duncan et al. (2019) use information from all available sources in the HETDEX field for training the redshift GP whilst our redshift model has been only trained on radio-detected AGN, giving it the opportunity to focus its parameter exploration only on these sources.

Finally, Cunha and Humphrey (2022) also produced photometric redshift predictions for almost 3 500 000 sources (stars, galaxies, and QSO) as part of their pipeline (see Sect. **??**). They combined three algorithms for their predictions: `XGBoost`, `CatBoost`, and `LightGBM` (Ke et al.,

2017). This procedure leads to $\sigma_{\rm NMAD} = 0.018$ and $\eta = 2\%$. As with previous examples, the differences with our results can be a consequence of the number of training samples. Also, in the case of Cunha and Humphrey (2022), they applied an additional post-processing step to the redshift predictions attempting to predict and understand the appearance of catastrophic outliers.

7.2 Influence of data imputation

One effect which might influence the training of the models and, consequently, the prediction for new sources is related to the imputation of missing values (cf. Sect. **??**). In Fig. 7.2, we have plotted the distributions of predicted scores (for classification models) and predicted redshift values as a function of the number of measured bands (`band_num`) for each step of the pipeline as applied to sources predicted to be of each class in the test sub-set.

The top panel of Fig. 7.2 shows the influence of the degree of imputation in the classification between AGN and galaxies. For most of the bins, probabilities for predicted galaxies are distributed close to 0.0, without any noticeable trend. In the case of predicted AGN, the combination of low number of sources and high degree of imputation (`band_num` < 5) lead to low mean probabilities.

The case of radio detection classification is somewhat different. Given the number and distribution of sources per bin, it is not possible to extract any strong trend for the probabilities of radio-predicted sources. The absence of evolution with the number of observed bands is stronger for sources predicted to be devoid of radio detection.

Finally, a stronger effect can be seen with the evolution of predicted redshift values for radio-detectable AGN. Despite the lower number of available sources, it is possible to recognise that sources with higher number of available measurements are predicted to have lower redshift values. Sources that are closer to us have higher probabilities to be detected in a large number of bands. Thus, it is expected that our model predicts lower redshift values for the most measured sources in the field.

In consequence, Fig. 7.2 allows us to understand the influence of imputation over the predictions. The most highly affected quantity is the redshift, where large fractions of measured magnitudes are needed to obtain scores that are in line with previous results (cf. Sect. **??**). The AGN-galaxy and radio detection classifications show a mild influence of imputation in their results.

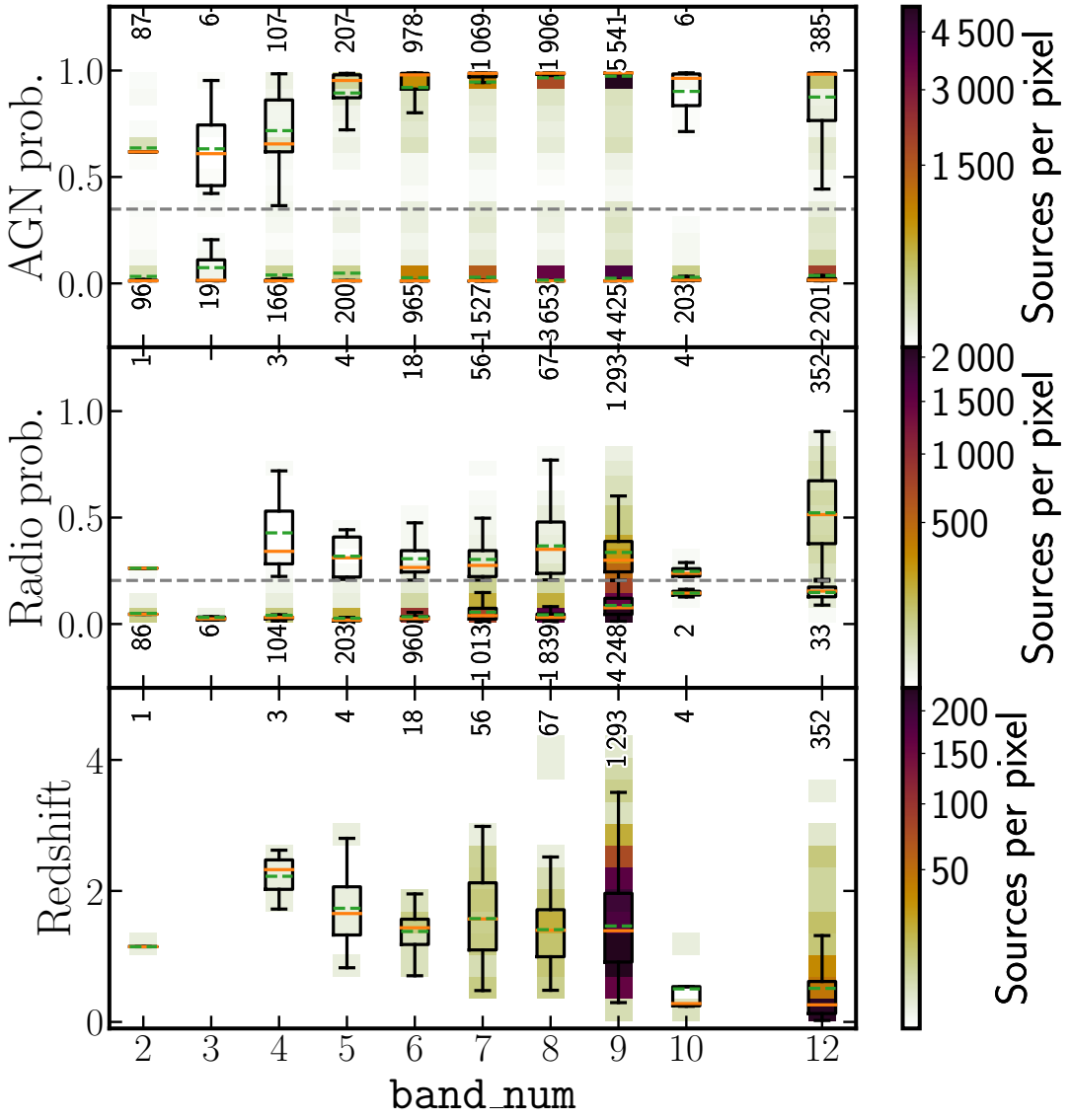


Figure 7.2: Evolution of predicted probabilities (top: probability to be AGN, middle: probability of AGN to be detected in radio) and redshift values for radio-detectable AGN (bottom panel) as a function of the number of observed bands for sources in test set. In top panel, sources have been divided between those predicted to be AGN and galaxy. In middle panel, sources are divided between predicted AGN that are predicted to be detected in radio and those predicted to not have radio detection. Background density plots (following colour coding in colorbars) show location of predicted values. Overlaid boxplots display main statistics for each number of measured bands. Black rectangles encompass sources in second and third quartiles. Vertical lines show the place of sources from first and fourth quartiles. Orange lines represent median value of sample and dashed, green lines indicate their mean values. Dashed, grey lines show PR thresholds for AGN-galaxy and radio detection classifications. Close to each boxplot, written values correspond to the number of sources considered to create each set of statistics.

7.3 Global feature importances

Overall, mean or global feature importances can be retrieved from models that are based on Decision Trees (e.g. Random Forests and Boosting models, Breiman, 2001; Breiman, 2003). All algorithms selected in this work (`RF`, `CatBoost`, `XGBoost`, `ET`, `GBR`, and `GBC`) belong to these two classes. For each feature, the decrease in impurity (a term frequently used in the literature related to Machine Learning) of the dataset is calculated for all the nodes of the tree in which that feature is used. Features with the highest impurity decrease will be more important for the model (Louppe et al., 2013)[1].

Insight into the decision-making of the pipeline can only rely on the specific weight of the original set of features (see Sect. **??**). Table 7.2 presents the ranked combined importances from the observables selected in each of the three sequential models that compose the pipeline. They have been combined using the importances from the meta-learner (as shown in Table 7.3) and that of base-learners. The derived importances will be dependent on the dataset used, including any imputation for the missing data, and the details of the models, i.e. algorithms used and stacking procedure. We first notice in Table 7.2 that the order of the features is different for all three models. This difference reinforces the need, as stated in Sect. **??**, of developing separate models for each of the prediction stages of this work that would evaluate the best feature weights for the related classification or regression task.

For the AGN-galaxy classification model, it is very interesting to note that the most important feature for the predicted probability of a source to be an AGN is the WISE colour W1 - W2 (as well as W1 - W3). This colour is indeed one of the axes of the widely used WISE colour-colour selection, with the second axis being the W2 - W3 colour (cf. Sect **??**). The WISE W3 photometry is though significantly less sensitive than W1, W2 or PS1 (see Fig. 4.5) and a significant number of sources will be represented as upper limits in such plot (see Table 4.2).

One of the main potential uses of the pipeline is its capability to pinpoint radio-detectable AGN. The global features analysis for the radio detection model shows a high dependence on the near- and mid-IR magnitudes and colours, especially those coming from WISE.

Finally, the redshift prediction model shows again that the final estimate is mostly driven by the results of the base learners, accounting for ~82% of the predicting power. The overall

[1]For some models that are not based on Decision Trees, feature importances can be obtained from the coefficients that the training process delivers for each feature. These coefficients are related to the level to which each quantity is scaled to obtain a final prediction (as in the coefficients from a polynomial regression).

Table 7.2: Relative importances (rescaled to add to 100) for observed features from the three models combined between meta and base models.

AGN-Galaxy (meta-model: CatBoost)					
Feature	Importance	Feature	Importance	Feature	Importance
W1_W2	68.945	H_K	1.715	z_W2	1.026
W1_W3	4.753	y_W1	1.659	z_y	0.722
g_r	4.040	y_W2	1.513	W3_W4	0.669
r_J	4.006	i_y	1.441	W4mag	0.558
r_i	3.780	i_z	1.366	H_W3	0.408
band_num	1.842	y_J	1.187	J_H	0.371

Radio detection (meta-model: GBC)					
Feature	Importance	Feature	Importance	Feature	Importance
W2_W3	9.609	y_W1	7.150	W4mag	4.759
y_J	8.102	g_r	7.123	K_W4	2.280
W1_W2	8.010	z_W1	7.076	J_H	1.283
g_i	7.446	r_z	6.981	H_K	1.030
K_W3	7.357	i_z	6.867	band_num	1.018
z_y	7.321	r_i	6.588		

Redshift prediction (meta-model: ET)					
Feature	Importance	Feature	Importance	Feature	Importance
y_W1	35.572	y_J	3.018	i_z	1.215
W1_W2	13.526	r_z	3.000	J_H	1.162
W2_W3	12.608	r_i	2.896	g_W3	1.000
band_number	6.358	z_y	2.827	K_W3	0.925
H_K	4.984	W4mag	2.784	K_W4	0.762
g_r	4.954	i_y	2.408		

Table 7.3: Relative feature importances (rescaled to add to 100) for base algorithms in each prediction step.

AGN-Galaxy model (CatBoost)			
Feature	Importance	Feature	Importance
gbc	49.709	xgboost	14.046
et	19.403	rf	8.981
Remaining feature importances:			7.861

Radio detection model (GBC)			
Feature	Importance	Feature	Importance
rf	12.024	catboost	7.137
et	7.154	xgboost	6.604
Remaining importances:			67.081

Redshift prediction model (ET)			
Feature	Importance	Feature	Importance
xgboost	25.138	catboost	21.072
gbr	21.864	rf	13.709
Remaining importances:			18.217

combined importance of features shows also in this case a strong dependence on several near-IR colours of which y - W1 and W1 - W2 are the most relevant ones. The model still relies, to a lesser extent, on a broad range of optical features needed to trace the broad range of redshift possibilities ($z \in [0, 6]$).

7.4 Local feature importances

As opposed to the global (mean) assessment of feature importances derived from the decrease in impurity, local (i.e. source by source) information on the performance of such features can be obtained from Shapley values. This is a method from coalitional game theory that tells us how to fairly distribute the dividends (the prediction in our case) among the features (Shapley, 1953). The previous statement means that the relative influence of each property from the dataset can be derived for individual predictions in the decision made by the model (which is not the same as obtaining causal correlations between features and the target; Ma and Tourani, 2020). The combination of Shapley values with several other model explanation methods was used by Lundberg and Lee (2017) to create the SHapley Additive exPlanations (SHAP) values. In this work, SHAP values were calculated using the python package `SHAP`[2] and, in particular, its module for Tree-based predictors (Lundberg et al., 2020). To speed calculations up, the package `FastTreeSHAP`[3] (`v0.1.2`; Yang, 2021) was also used, which allows for multi-thread runs.

One way to display these SHAP values is through the so-called decision plots. They can show how individual predictions are driven by the inclusion of each feature. Besides determining the most relevant properties that help the model make a decision, it is possible to detect sources that follow different prediction paths which could be, eventually and upon further examination, labelled as outliers. An example of this decision plot, linked to the AGN-Galaxy classification, is shown in Fig. 7.3 for a subsample of the high-redshift ($z \geq 4.0$) spectroscopically classified AGN in the HETDEX field (121 sources, regardless of them being part of any sub-set involved in the training or validation of the models). The different features used by the meta-learner are stacked on the vertical axis with increasing weight and these final weight are sumarized in Table 7.4. Similarly, SHAP decision plots for the radio-detection and redshift prediction are presented in Figs. 7.4 and 7.5, respectively.

[2] https://github.com/slundberg/shap

[3] https://github.com/linkedin/fasttreeshap

As it can be seen, for the three models, base learners are amongst the features with the highest influence. This result raises the question of what drives these individual base predictions. Appendix **??** includes SHAP decision plots for all base learners used in this work. Additionally, and to be able to compare these results with the features importances from Sect. **??**, we constructed Table 7.5, which displays the combined SHAP values of base and meta learners but, in this case, for the same 121 high-redshift confirmed AGN (with 29 of them detected by LoTSS). Table 7.5 shows, as Table 7.2, that the colour W1 - W2 is the most important discriminator between AGN and Galaxies for this specific set of sources. The importance of the rest of the features is mixed: similar colours are located on the top spots (e.g. g - r, W1 - W3 or r - i).

For the radio classification step of the pipeline, we find that features linked to those 121 high-z AGN perform at the same level as for the overall population. The improved metrics with respect to those obtained from the 'no-skill' selection do indicate that the model has learned some connections between the data and the radio emission. Feature importance has changed when compared to the overall population. If the radio emission observed from these sources were exclusively due to SF, this connection would imply SFR of several hundred $M_{\odot}\,\mathrm{yr}^{-1}$. This explanation can not be completely ruled out from the model side but some contribution of radio emission from the AGN is expected.

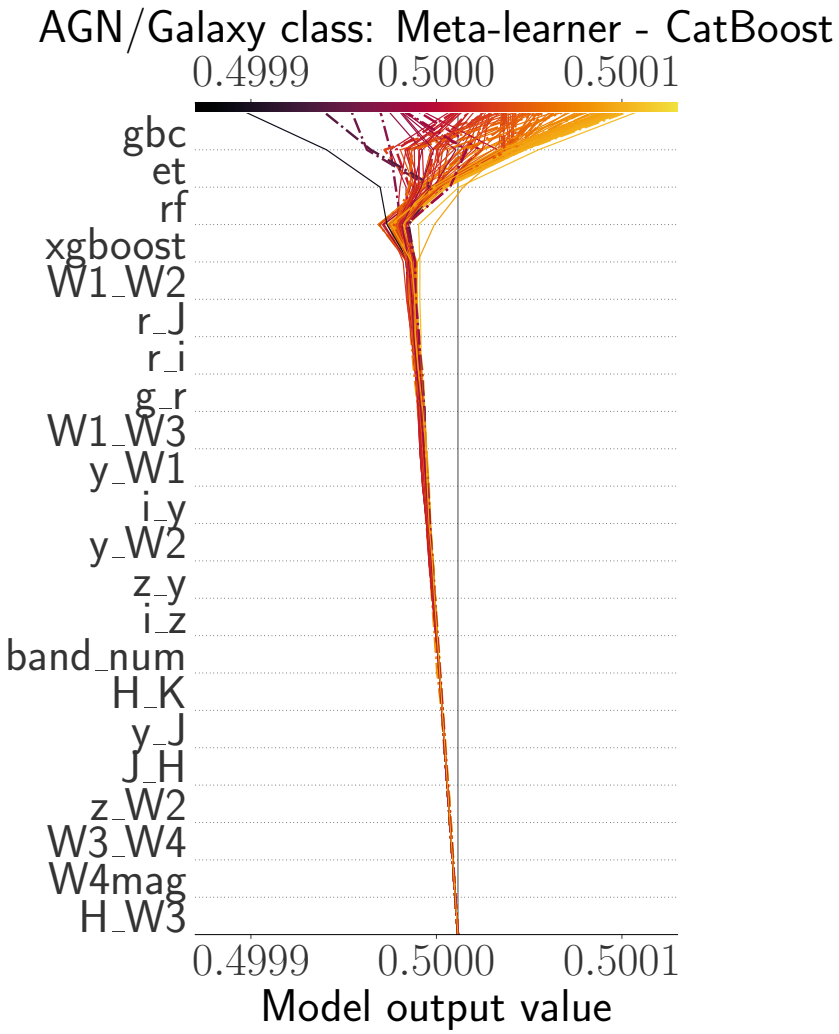


Figure 7.3: Decision plot from SHAP values for AGN-Galaxy classification from the 121 high redshift ($z \geq 4$) spectroscopically confirmed AGN in HETDEX. Horizontal axis represents the model's output with a starting value for each source centred on the selected naive threshold for classification. Vertical axis shows features used in the model sorted, from top to bottom, by decreasing mean absolute SHAP value. Each prediction is represented by a coloured line corresponding to its final predicted value as shown by the colorbar at the top. Moving from the bottom of the plot to the top, SHAP values for each feature are added to the previous value in order to highlight how each feature contributes to the overall prediction. Predictions for sources detected by LOFAR are highlighted with a dotted, dashed line.

Table 7.4: SHAP values (rescaled to add to 100) for base algorithms in each prediction step for observed features using 121 spectroscopically confirmed AGN at high redshift values ($z > 4$).

AGN-Galaxy model (CatBoost)			
Feature	SHAP value	Feature	SHAP value
gbc	36.250	rf	21.835
et	30.034	xgboost	7.198
	Remaining SHAP values:		4.683

Radio detection model (GBC)			
Feature	SHAP value	Feature	SHAP value
rf	11.423	catboost	5.696
xgboost	7.741	et	5.115
	Remaining SHAP values:		70.025

Redshift prediction model (ET)			
Feature	SHAP value	Feature	SHAP value
xgboost	41.191	gbr	13.106
catboost	20.297	rf	11.648
	Remaining SHAP values:		13.758

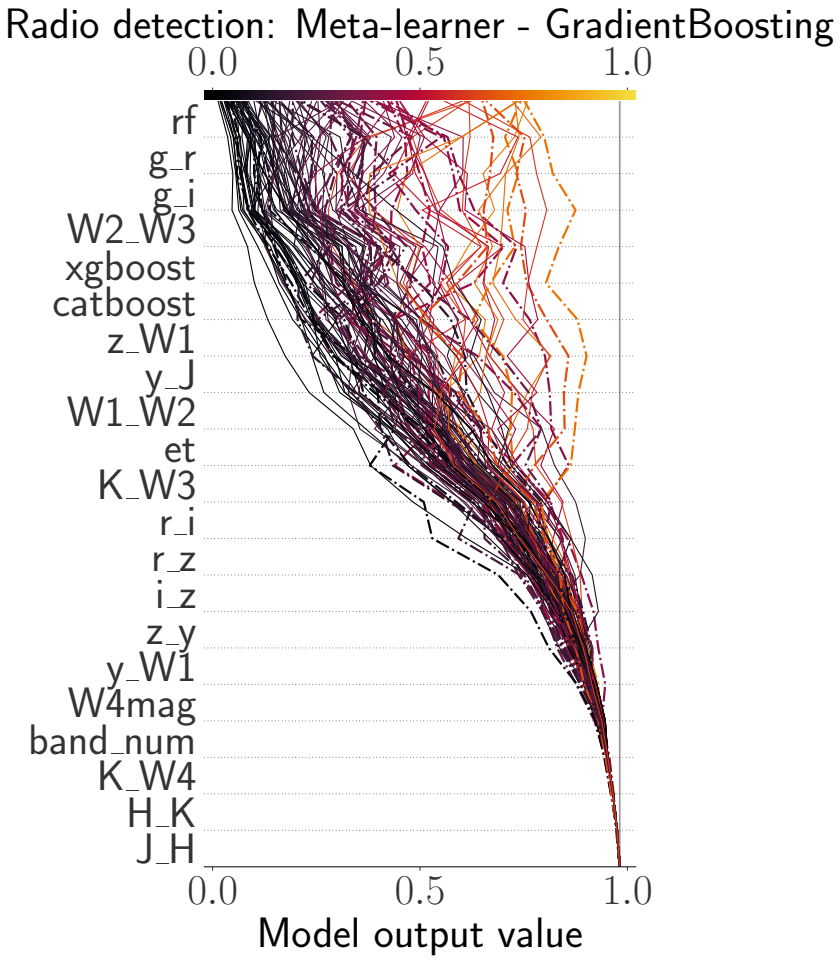


Figure 7.4: Decision plot from the SHAP values for all features from the radio detection model in the 121 high redshift ($z \geq 4$) spectroscopically confirmed AGN from HETDEX. Description as in Fig. 7.3.

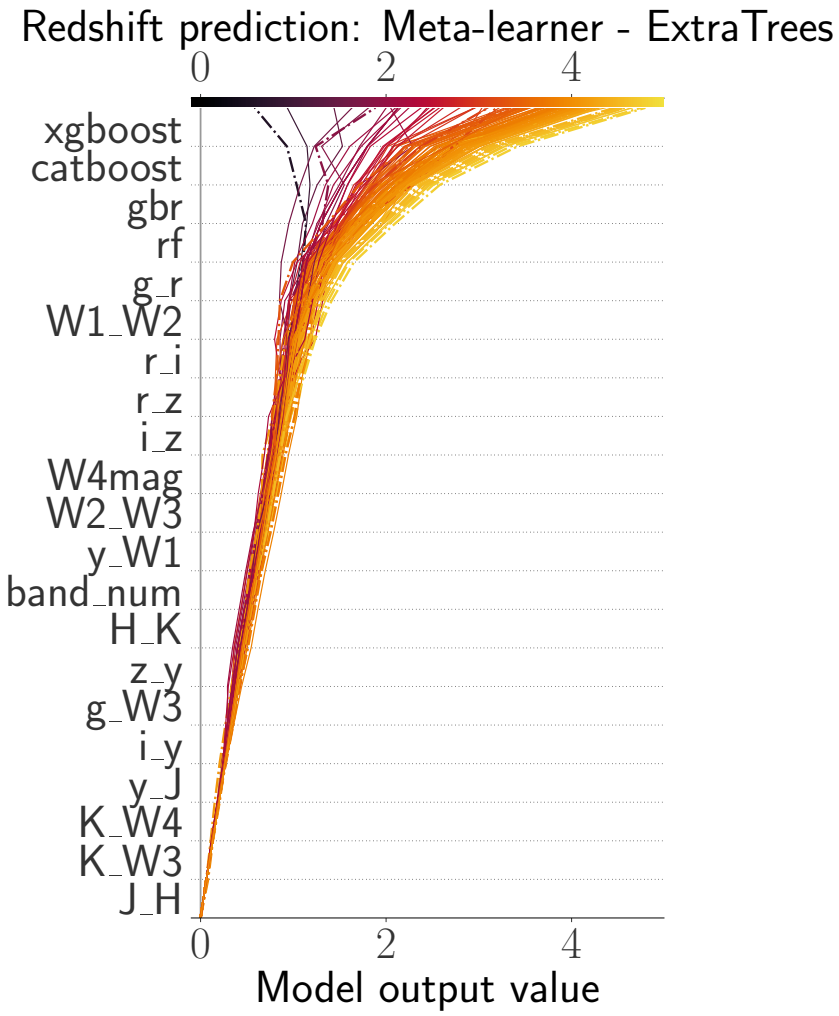


Figure 7.5: Decision plot from the SHAP values for all features from the redshift prediction model in the 121 high redshift ($z \geq 4$) spectroscopically confirmed AGN from HETDEX. Description as in Fig. 7.3.

Table 7.5: Combined and normalised (rescaled to add to 100) mean absolute SHAP values for observed features from the three models using 121 spectroscopically confirmed AGN at high redshift values ($z \geq 4$).

AGN-Galaxy model					
Feature	SHAP value	Feature	SHAP value	Feature	SHAP value
W1_W2	32.458	i_y	5.086	z_y	1.591
g_r	11.583	y_W1	4.639	H_W3	1.048
W1_W3	8.816	band_num	4.050	W4mag	0.514
r_i	7.457	y_W2	3.228	H_K	0.466
i_z	6.741	z_W2	2.348	W3_W4	0.466
r_J	6.613	y_J	1.718	J_H	0.178

Radio detection model					
Feature	SHAP value	Feature	SHAP value	Feature	SHAP value
g_i	14.120	z_W1	6.751	W4mag	2.691
W2_W3	13.201	r_i	5.577	band_num	2.661
g_r	12.955	r_z	5.161	K_W4	0.939
y_J	8.224	i_z	4.512	H_K	0.719
K_W3	7.441	z_y	4.121	J_H	0.190
W1_W2	6.874	y_W1	3.864		

Redshift prediction model					
Feature	SHAP value	Feature	SHAP value	Feature	SHAP value
g_r	32.594	z_y	3.557	W4mag	1.639
y_W1	20.770	y_J	3.010	g_W3	1.479
W2_W3	12.462	band_num	2.595	K_W3	0.853
W1_W2	5.692	i_y	2.381	K_W4	0.451
r_i	4.381	H_K	2.230	J_H	0.146
r_z	3.755	i_z	2.005		

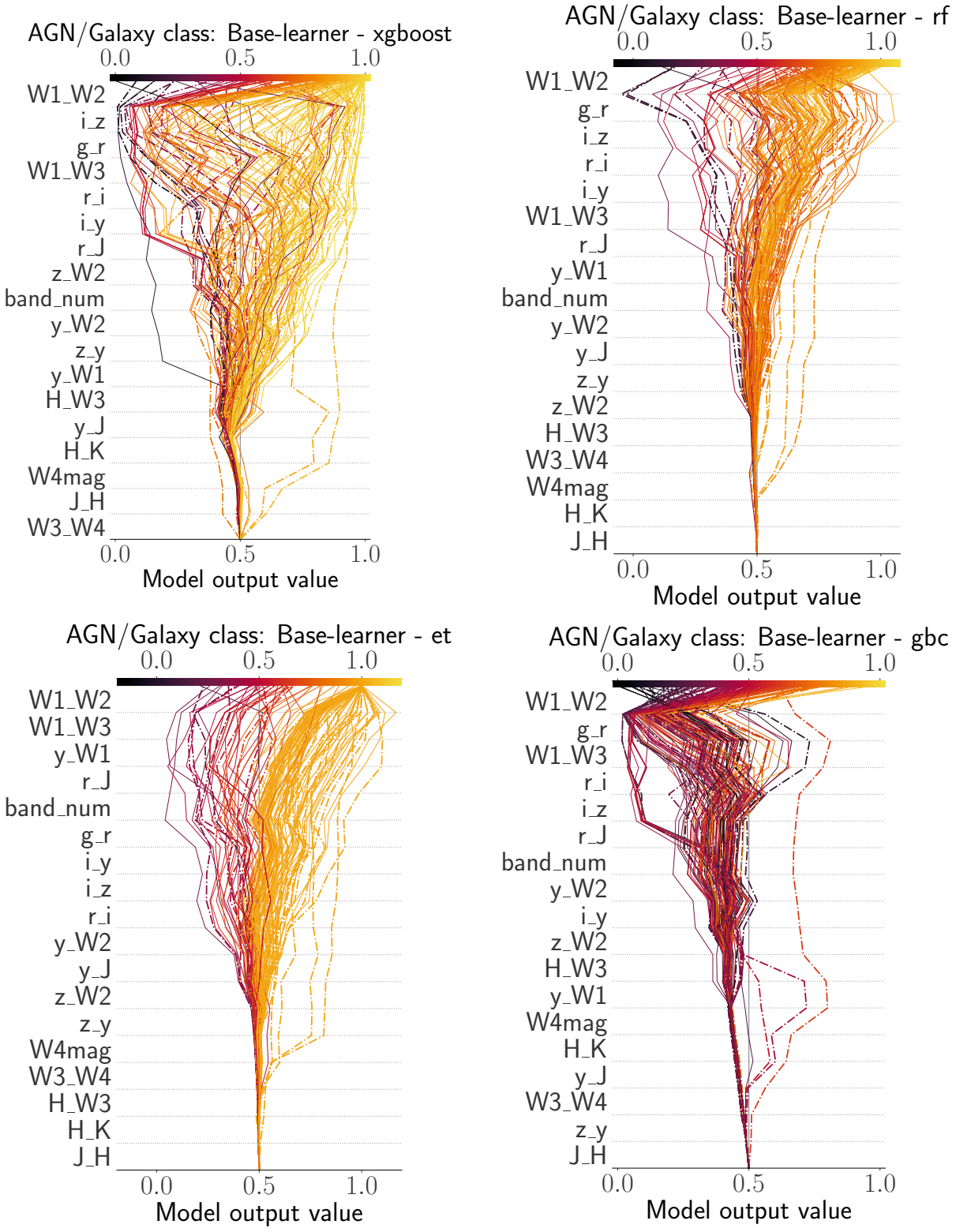


Figure 7.6: SHAP decision plots for base AGN-Galaxy algorithms. Details as described in Figs. 7.3. Starting point of predictions is the naive classification threshold. From left to right and from top to bottom, each panel shows the results from `XGBoost`, `RF`, `ET`, and `GBC`.

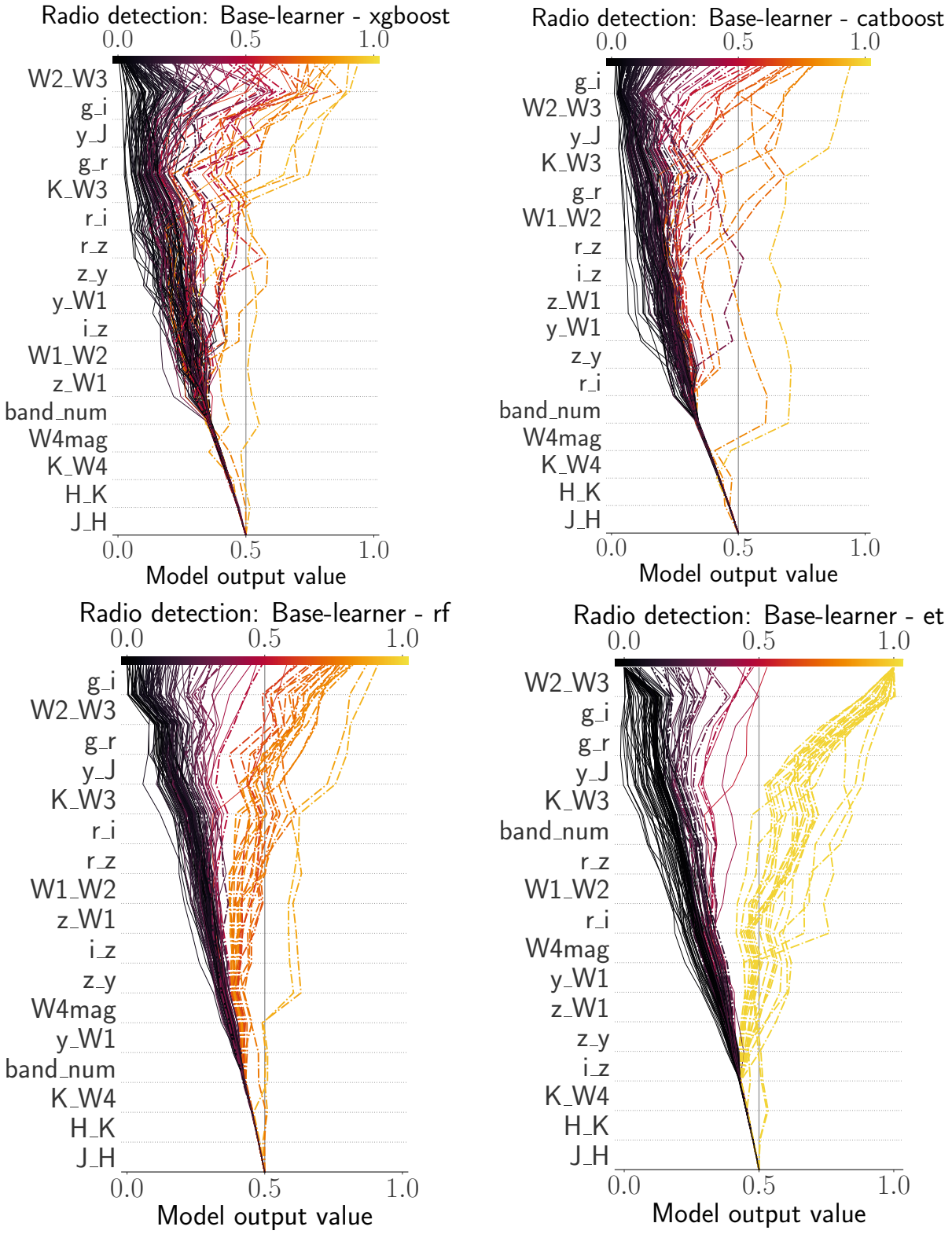


Figure 7.7: SHAP decision plots from base radio algorithms. Details as Figs. 7.3 and 7.6. Each panel with results for `XGBoost`, `CatBoost`, `RF`, and `ET`.

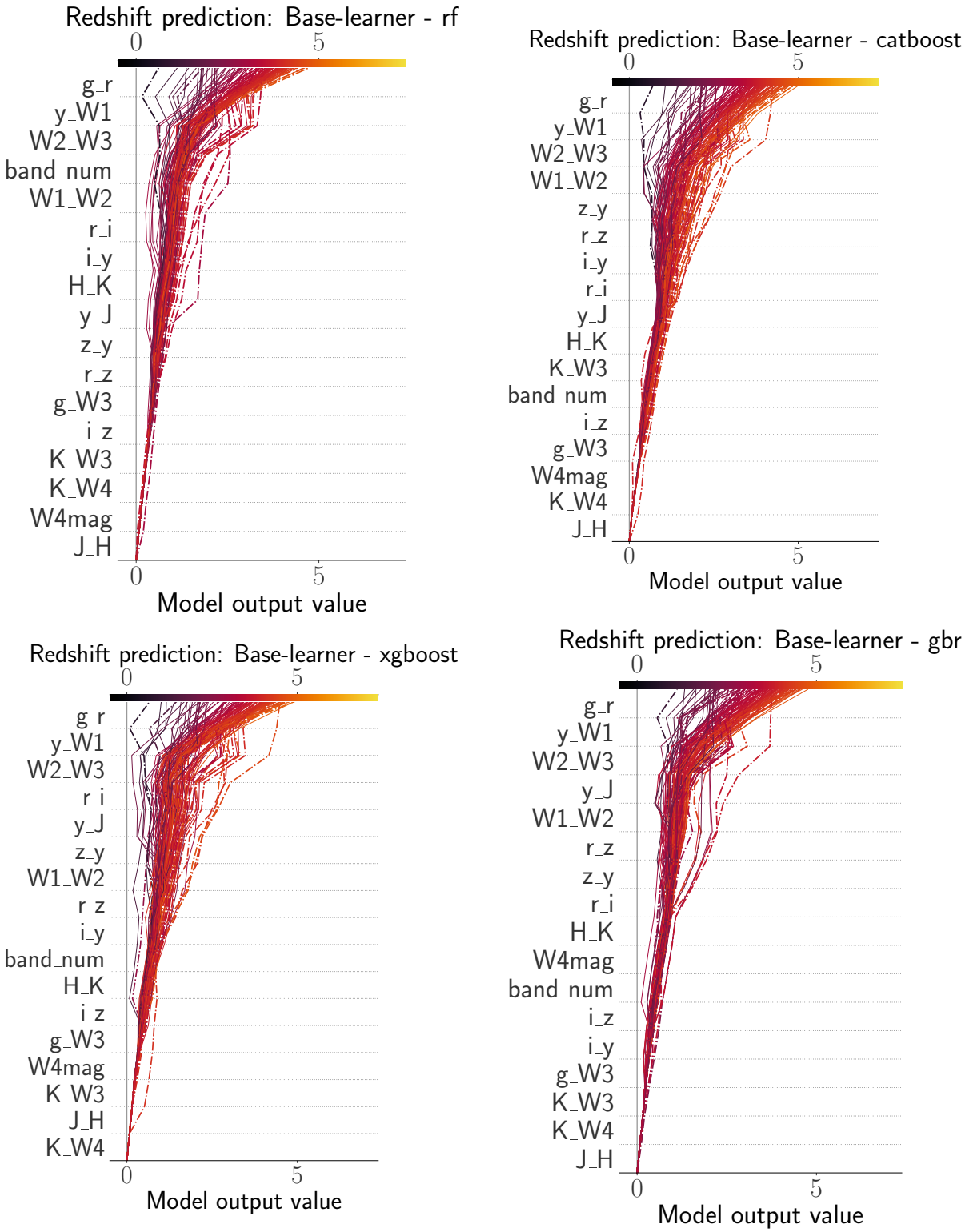


Figure 7.8: SHAP decision plots from base z algorithms. Details as in Fig 7.3. Each panel shows results for `ET`, `CatBoost`, `XGBoost`, and `GBR`.

8

Machine-assisted learning from models

8.1 Colour-colour AGN selection criterion

From the importances in Table 7.2 and the values presented in Fig. 4.3 we infer that using optical colours could in principle create selection criteria with metrics equivalent to those shown in Table 7.1 but for a much larger number of sources (100 000 sources for colour plots using W3 vs 4 700 000 sources for colours based in r, i or z magnitudes). We tested this hypothesis and derived a selection criterion in the g - r vs W1 - W2 colour-colour plot shown in Fig. 8.1 using the labelled sources in the test sub-set of the HETDEX field. The results of the application of this criterion to the testing data and to the labelled sources in S82 is presented in the last row of Table 7.1. Their limits are defined by the following expressions:

$$g - r > -0.76, \tag{8.1}$$

$$g - r < 1.8, \tag{8.2}$$

$$W1 - W2 > 0.227 \times (g - r) + 0.43, \tag{8.3}$$

where W1, W2, g, and r are Vega magnitudes. Our colour criteria provides better and more homogeneous scores across the different metrics with purity (precision) and completeness (recall) above 87%. Avoiding the use of the longer WISE wavelength (W3 and W4), the criteria can be applied to a much larger dataset.

One of the main potential uses of the pipeline is its capability to pinpoint radio-detectable AGN. The global features analysis for the radio detection model shows a high dependence on the near- and mid-IR magnitudes and colours, especially those coming from WISE. As a useful outcome similar to the AGN-Galaxy classification, we can use the most relevant features to build useful plots for the pre-selection of these sources and get insight into the origin of the radio emission. This is the case for the W4 histogram, shown in Fig. 8.2, where sources predicted to be radio-emitting AGN extend to brighter measured W4 magnitudes. This added mid-IR flux might be simply due to an increased star formation rates (SFR) in these sources.

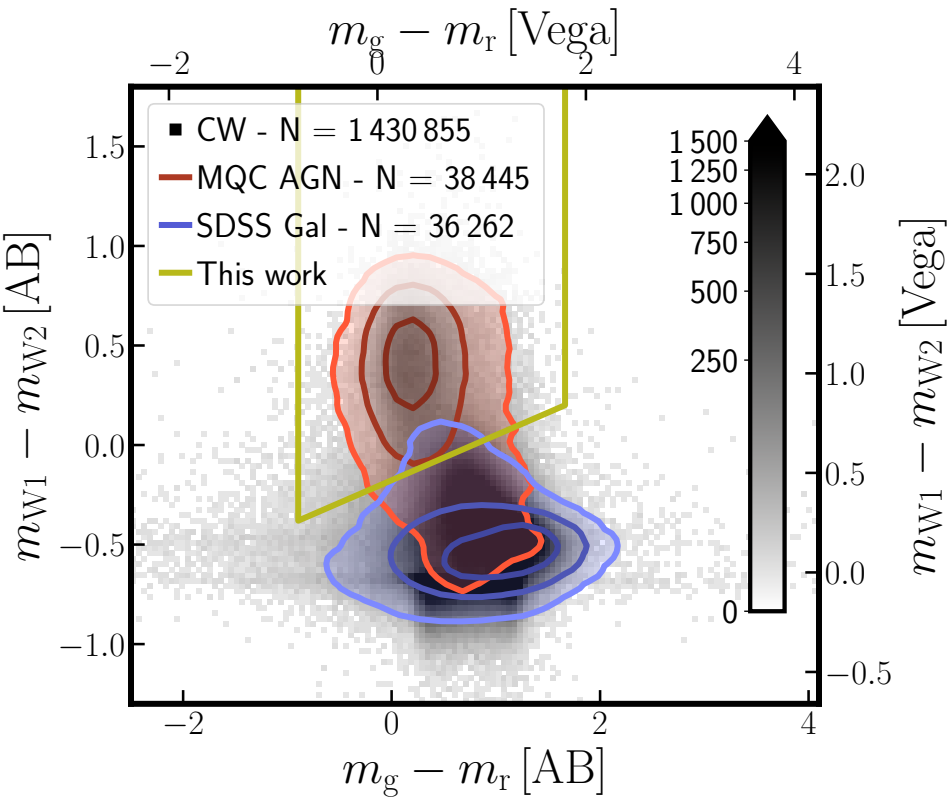


Figure 8.1: AGN classification colour-colour plot in the HETDEX field using CW (W1, W2) and PS1 (g, r) passbands. Grey-scale density plot include all CW detected and non-imputed sources. Red contours highlight the density distribution of the AGN in the Million QSO catalogue (MQC) and blue contours show the density distribution for the galaxies from SDSS DR16. Contours are located at 1, 2, and 3 σ levels.

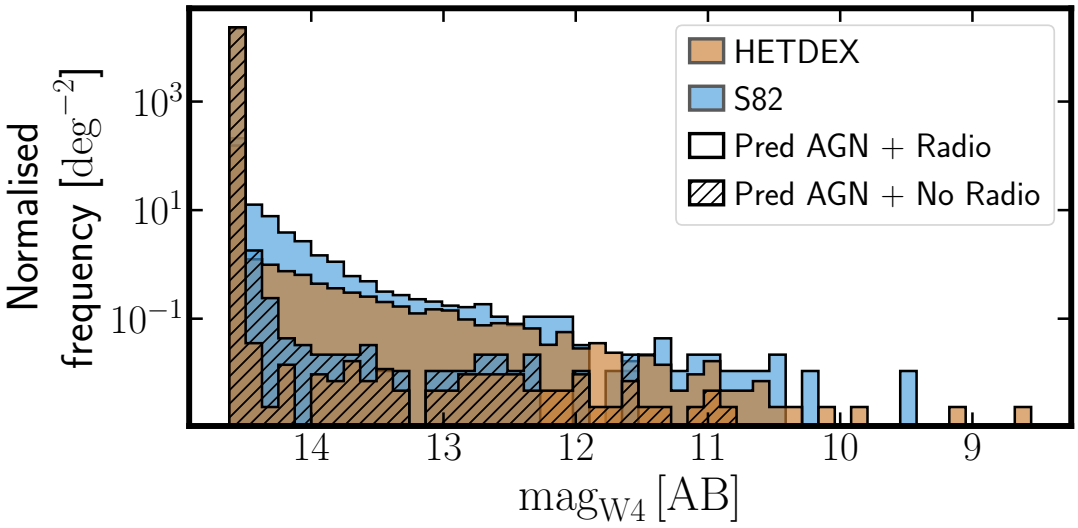


Figure 8.2: W4 magnitudes density distribution of the newly predicted radio-AGN (clean histograms) in HETDEX (ochre histograms) and S82 (blue histograms) and W4 magnitudes from predicted AGN that are predicted to not have radio detection (dashed histograms).

In fact the 24μm flux is often used, together with that of Hα as a proxy for SFR (Kennicutt et al., 2009). The radio detection for these sources might have a strong component linked to the ongoing SF, especially for the sources with real or predicted redshift below z~1.5. A detailed exploration of the implications that these dependencies might have in our understanding of the triggering of radio emission on AGN, whether related to star formation (SF) or jets, is left for a future publication (Carvajal et al. in preparation).

8.2 AGN radio luminosity function

8.3 Radio counterpart assessment

This page intentionally left blank.

9

Future developments

The results presented here allow us to look for further developments. We list here some steps that can be taken in order to expand the

9.1 Extensive feature importance analysis

Expanding on the analysis of Chapter XX, the behaviour of the model, when applied to different data sets, can be studied extensively.

9.2 Evolutionary map of the universe

Given that Sections XXX and XXX show possible applications of the prediction pipeline on data from the EMU Pilot Survey, more studied can be performed on the areas covered by EMU.

9.3 Square Kilometre Array

Ultimately, our prediction pipeline can be applied to the full area covered by the future Square Kilometre Array in order to accelerate the detection of AGN.

This page intentionally left blank.

Summary

This is the summary of the text we have produced.

This page intentionally left blank.

Data and software acknowledgements

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

LOFAR data products were provided by the LOFAR Surveys Key Science project (LSKSP[1]) and were derived from observations with the International LOFAR Telescope (ILT). LOFAR (van Haarlem et al., 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the ILT foundation under a joint scientific policy. The efforts of the LSKSP have benefited from funding from the European Research Council, NOVA, NWO, CNRS-INSU, the SURF Co-operative, the UK Science and Technology Funding Council and the Jülich Supercomputing Centre.

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

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis

[1] https://lofar-surveys.org/

Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This work made use of public data from the Sloan Digital Sky Survey, Data Release 16. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions.

SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is `www.sdss.org`. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This research has made use of NASA's Astrophysics Data System, TOPCAT[2] (Taylor, 2005), JupyterLab[3] (Kluyver et al., 2016), 'Aladin sky atlas' (`v11.0.24`; Bonnarel et al., 2000) developed at CDS, Strasbourg Observatory, France, and the VizieR catalogue access tool, CDS, Strasbourg, France (DOI : 10.26093/cds/vizier). The original description of the VizieR service was published in Ochsenbein et al., 2000.

This work made extensive use of the Python packages `PyCaret`[4] (`v2.3.10`; Ali, 2020), `scikit-learn` (`v0.23.2`; Pedregosa et al., 2011), `pandas`[5] (`v1.4.2`; McKinney, 2010),

[2] `http://www.star.bris.ac.uk/~mbt/topcat/`
[3] `https://jupyter.org`
[4] `https://pycaret.org`
[5] `https://pandas.pydata.org`

`Astropy`[6], a community-developed core Python package for Astronomy (`v5.0`; Astropy Collaboration et al., 2013; Astropy Collaboration et al., 2018; Astropy Collaboration et al., 2022), `Matplotlib` (`v3.5.1`; Hunter, 2007), `betacal`[7] (`v1.1.0`), and `CMasher`[8] (`v1.6.3`; van der Velden, 2020).

[6] https://www.astropy.org

[7] https://betacal.github.io

[8] https://github.com/1313e/CMasher

This page intentionally left blank.

References

Abbott, T. M. C., Abdalla, F. B., Allam, S., Amara, A., Annis, J., Asorey, J., Avila, S., Ballester, O., Banerji, M., Barkhouse, W., Baruah, L., Baumer, M., Bechtol, K., Becker, M. R., Benoit-Lévy, A., Bernstein, G. M., Bertin, E., Blazek, J., Bocquet, S., Brooks, D., Brout, D., Buckley-Geer, E., Burke, D. L., Busti, V., Campisano, R., Cardiel-Sas, L., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Cawthon, R., Chang, C., Chen, X., Conselice, C., Costa, G., Crocce, M., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Das, R., Daues, G., Davis, T. M., Davis, C., De Vicente, J., DePoy, D. L., DeRose, J., Desai, S., Diehl, H. T., Dietrich, J. P., Dodelson, S., Doel, P., Drlica-Wagner, A., Eifler, T. F., Elliott, A. E., Evrard, A. E., Farahi, A., Fausti Neto, A., Fernandez, E., Finley, D. A., Flaugher, B., Foley, R. J., Fosalba, P., Friedel, D. N., Frieman, J., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Giannantonio, T., Gill, M. S. S., Glazebrook, K., Goldstein, D. A., Gower, M., Gruen, D., Gruendl, R. A., Gschwend, J., Gupta, R. R., Gutierrez, G., Hamilton, S., Hartley, W. G., Hinton, S. R., Hislop, J. M., Hollowood, D., Honscheid, K., Hoyle, B., Huterer, D., Jain, B., James, D. J., Jeltema, T., Johnson, M. W. G., Johnson, M. D., Kacprzak, T., Kent, S., Khullar, G., Klein, M., Kovacs, A., Koziol, A. M. G., Krause, E. et al. (Dec. 2018). 'The Dark Energy Survey: Data Release 1'. In: ApJS 239.2, 18, p. 18. DOI: 10.3847/1538-4365/aae9f0. arXiv: 1801.03181 [astro-ph.IM].

Aguado, D. S., Ahumada, R., Almeida, A., Anderson, S. F., Andrews, B. H., Anguiano, B., Aquino Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., Aubert, M., Avila-Reese, V., Badenes, C., Barboza Rembold, S., Barger, K., Barrera-Ballesteros, J., Bates, D., Bautista, J., Beaton, R. L., Beers, T. C., Belfiore, F., Bernardi, M., Bershady, M., Beutler, F., Bird, J., Bizyaev, D., Blanc, G. A., Blanton, M. R., Blomqvist, M., Bolton, A. S., Boquien, M., Borissova, J., Bovy, J., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Burgasser, A., Byler, N., Cano Diaz, M., Cappellari, M., Carrera, R., Cervantes Sodi, B., Chen, Y., Cherinka, B., Choi, P. D., Chung, H., Coffey, D., Comerford, J. M., Comparat, J., Covey, K., da Silva Ilha, G., da Costa, L., Dai, Y. S., Damke, G., Darling, J., Davies, R., Dawson, K., de Sainte Agathe, V., Deconto Machado, A., Del Moro, A., De Lee, N., Diamond-Stanic, A. M., Domínguez Sánchez, H., Donor, J., Drory, N., du Mas des Bourboux, H., Duckworth, C., Dwelly, T., Ebelke, G., Emsellem, E., Escoffier, S., Fernández-Trincado, J. G., Feuillet, D., Fischer, J.-L., Fleming, S. W., Fraser-McKelvie, A., Freischlad, G., Frinchaboy, P. M., Fu, H., Galbany, L., Garcia-Dias, R., García-Hernández, D. A., Garma Oehmichen, L. A., Geimba Maia, M. A., Gil-Marín, H., Grabowski, K., Gu, M., Guo, H., Ha, J., Harrington, E., Hasselquist, S., Hayes, C. R., Hearty, F., Hernandez Toledo, H., Hicks, H., Hogg, D. W., Holley-Bockelmann, K. et al. (Feb. 2019). 'The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA-derived Quantities, Data Visualization Tools, and Stellar Library'. In: ApJS 240.2, 23, p. 23. DOI: 10.3847/1538-4365/aaf651. arXiv: 1812.02759 [astro-ph.IM].

Ahumada, R., Prieto, C. A., Almeida, A., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., Arcodia, R., Armengaud, E., Aubert, M., Avila, S., Avila-Reese, V., Badenes, C., Balland, C., Barger, K., Barrera-Ballesteros, J. K., Basu, S., Bautista, J., Beaton, R. L., Beers, T. C., Benavides, B. I. T., Bender, C. F., Bernardi, M., Bershady, M., Beutler, F., Bidin, C. M., Bird, J., Bizyaev, D., Blanc, G. A., Blanton, M. R., Boquien, M., Borissova, J., Bovy, J., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Bureau, M., Burgasser, A., Burtin, E., Cano-Díaz, M., Capasso, R., Cappellari, M., Carrera, R., Chabanier, S., Chaplin, W., Chapman, M., Cherinka, B., Chiappini, C., Doohyun Choi, P., Chojnowski, S. D., Chung, H., Clerc, N., Coffey, D., Comerford, J. M., Comparat, J., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J. D., Cunha, K., Ilha, G. d. S., Dai, Y. S., Damsted, S. B., Darling, J., Davidson James W., J., Davies, R., Dawson, K., De, N., de la Macorra, A., De Lee, N., Queiroz, A. B. d. A., Deconto Machado, A., de la Torre, S., Dell'Agli, F., du Mas des Bourboux, H., Diamond-Stanic, A. M., Dillon, S., Donor, J., Drory, N., Duckworth, C., Dwelly, T., Ebelke, G., Eftekharzadeh, S., Davis Eigenbrot, A., Elsworth, Y. P., Eracleous, M., Erfanianfar, G., Escoffier, S., Fan, X., Farr, E., Fernández-Trincado, J. G., Feuillet, D., Finoguenov, A., Fofie, P., Fraser-McKelvie, A., Frinchaboy, P. M. et al. (July 2020). 'The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra'. In: ApJS 249.1, 3, p. 3. DOI: 10.3847/1538-4365/ab929e. arXiv: 1912.02905 [astro-ph.GA].

Aihara, H., Allende Prieto, C., An, D., Anderson, S. F., Aubourg, É., Balbinot, E., Beers, T. C., Berlind, A. A., Bickerton, S. J., Bizyaev, D., Blanton, M. R., Bochanski, J. J., Bolton, A. S., Bovy, J., Brandt, W. N., Brinkmann, J., Brown, P. J., Brownstein, J. R., Busca, N. G., Campbell, H., Carr, M. A., Chen, Y., Chiappini, C., Comparat, J., Connolly, N., Cortes, M., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Dawson, K., Dhital, S., Ealet, A., Ebelke, G. L., Edmondson, E. M., Eisenstein, D. J., Escoffier, S., Esposito, M., Evans, M. L., Fan, X., Femenía Castellá, B., Font-Ribera, A., Frinchaboy, P. M., Ge, J., Gillespie, B. A., Gilmore, G.,

González Hernández, J. I., Gott, J. R., Gould, A., Grebel, E. K., Gunn, J. E., Hamilton, J.-C., Harding, P., Harris, D. W., Hawley, S. L., Hearty, F. R., Ho, S., Hogg, D. W., Holtzman, J. A., Honscheid, K., Inada, N., Ivans, I. I., Jiang, L., Johnson, J. A., Jordan, C., Jordan, W. P., Kazin, E. A., Kirkby, D., Klaene, M. A., Knapp, G. R., Kneib, J.-P., Kochanek, C. S., Koesterke, L., Kollmeier, J. A., Kron, R. G., Lampeitl, H., Lang, D., Le Goff, J.-M., Lee, Y. S., Lin, Y.-T., Long, D. C., Loomis, C. P., Lucatello, S., Lundgren, B., Lupton, R. H., Ma, Z., MacDonald, N., Mahadevan, S., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Maraston, C., Margala, D., Masters, K. L., McBride, C. K. et al. (Apr. 2011). 'The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III'. In: ApJS 193.2, 29, p. 29. DOI: 10.1088/0067-0049/193/2/29. arXiv: 1101.1559 [astro-ph.IM].

Ali, M. (Apr. 2020). *PyCaret: An open source, low-code machine learning library in Python*. PyCaret version 2.3. URL: https://www.pycaret.org.

Allen, D. M. (1974). 'The Relationship Between Variable Selection and Data Agumentation and a Method for Prediction'. In: *Technometrics* 16.1, pp. 125–127. DOI: 10.1080/00401706.1974.10489157. eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1974.10489157. URL: https://www.tandfonline.com/doi/abs/10.1080/00401706.1974.10489157.

Allison, P. (2001). *Missing Data*. Quantitative Applications in the Social Sciences. SAGE Publications. ISBN: 9781452207902. URL: https://books.google.pt/books?id=LJB2AwAAQBAJ.

Amarantidis, S., Afonso, J., Messias, H., Henriques, B., Griffin, A., Lacey, C., Lagos, C. d. P., Gonzalez-Perez, V., Dubois, Y., Volonteri, M., Matute, I., Pappalardo, C., Qin, Y., Chary, R.-R. and Norris, R. P. (May 2019). 'The first supermassive black holes: indications from models for future observations'. In: MNRAS 485.2, pp. 2694–2709. DOI: 10.1093/mnras/stz551. arXiv: 1902.07982 [astro-ph.GA].

An, T., Zhang, Y. and Frey, S. (Sept. 2020). 'A method for checking high-redshift identification of radio AGNs'. In: MNRAS 497.2, pp. 2260–2264. DOI: 10.1093/mnras/staa2132. arXiv: 2007.09038 [astro-ph.GA].

Ananna, T. T., Salvato, M., LaMassa, S., Urry, C. M., Cappelluti, N., Cardamone, C., Civano, F., Farrah, D., Gilfanov, M., Glikman, E., Hamilton, M., Kirkpatrick, A., Lanzuisi, G., Marchesi, S., Merloni, A., Nandra, K., Natarajan, P., Richards, G. T. and Timlin, J. (Nov. 2017). 'AGN Populations in Large-volume X-Ray Surveys: Photometric Redshifts and Population Types Found in the Stripe 82X Survey'. In: ApJ 850.1, 66, p. 66. DOI: 10.3847/1538-4357/aa937d. arXiv: 1710.01296 [astro-ph.GA].

Andonie, C., Alexander, D. M., Rosario, D., Laloux, B., Georgakakis, A., Morabito, L. K., Villforth, C., Avirett-Mackenzie, M., Calistro Rivera, G., Del Moro, A., Fotopoulou, S., Harrison, C., Lapi, A., Petley, J., Petter, G. and Shankar, F. (Dec. 2022). 'A panchromatic view of infrared quasars: excess star formation and radio emission in the most heavily obscured systems'. In: MNRAS 517.2, pp. 2577–2598. DOI: 10.1093/mnras/stac2800. arXiv: 2209.13321 [astro-ph.GA].

Aniyan, A. K. and Thorat, K. (June 2017). 'Classifying Radio Galaxies with the Convolutional Neural Network'. In: ApJS 230.2, 20, p. 20. DOI: 10.3847/1538-4365/aa7333. arXiv: 1705.03413 [astro-ph.IM].

Annis, J., Soares-Santos, M., Strauss, M. A., Becker, A. C., Dodelson, S., Fan, X., Gunn, J. E., Hao, J., Ivezić, Ž., Jester, S., Jiang, L., Johnston, D. E., Kubo, J. M., Lampeitl, H., Lin, H., Lupton, R. H., Miknaitis, G., Seo, H.-J., Simet, M. and Yanny, B. (Oct. 2014). 'The Sloan Digital Sky Survey Coadd: 275 deg^2 of Deep Sloan Digital Sky Survey Imaging on Stripe 82'. In: ApJ 794.2, 120, p. 120. DOI: 10.1088/0004-637X/794/2/120. arXiv: 1111.6619 [astro-ph.CO].

Arévalo, P., Uttley, P., Kaspi, S., Breedt, E., Lira, P. and McHardy, I. M. (Sept. 2008). 'Correlated X-ray/optical variability in the quasar MR2251-178'. In: MNRAS 389.3, pp. 1479–1488. DOI: 10.1111/j.1365-2966.2008.13719.x. arXiv: 0807.2451 [astro-ph].

Arévalo, P., Uttley, P., Lira, P., Breedt, E., McHardy, I. M. and Churazov, E. (Aug. 2009). 'Correlation and time delays of the X-ray and optical emission of the Seyfert Galaxy NGC 3783'. In: MNRAS 397.4, pp. 2004–2014. DOI: 10.1111/j.1365-2966.2009.15110.x. arXiv: 0905.1981 [astro-ph.CO].

Arnouts, S., Cristiani, S., Moscardini, L., Matarrese, S., Lucchin, F., Fontana, A. and Giallongo, E. (Dec. 1999). 'Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North'. In: MNRAS 310.2, pp. 540–556. DOI: 10.1046/j.1365-8711.1999.02978.x. arXiv: astro-ph/9902290 [astro-ph].

Arsioli, B. and Dedin, P. (Oct. 2020). 'Machine learning applied to multifrequency data in astrophysics: blazar classification'. In: MNRAS 498.2, pp. 1750–1764. DOI: 10.1093/mnras/staa2449. arXiv: 2005.03536 [astro-ph.HE].

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. and Yan, L. (July 2013). 'Mid-infrared Selection of Active Galactic Nuclei with the Wide-field Infrared Survey Explorer. II. Properties of WISE-selected Active Galactic Nuclei in the NDWFS Boötes Field'. In: ApJ 772.1, 26, p. 26. DOI: 10.1088/0004-637X/772/1/26. arXiv: 1209.6055 [astro-ph.CO].

Assef, R. J., Stern, D., Noirot, G., 06, H. D., Cutri, R. M. and Eisenhardt, P. R. M. (Feb. 2018). 'The WISE AGN Catalog'. In: ApJS 234.2, 23, p. 23. DOI: 10.3847/1538-4365/aaa00a. arXiv: 1706.09901 [astro-ph.GA].

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., Vand erPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., Ardelean, C., Babej, T., Bach, Y. P., Bachetti, M., Bakanov, A. V., Bamford, S. P., Barentsen, G., Barmby, P., Baumbach, A., Berry, K. L., Biscani, F., Boquien, M., Bostroem, K. A., Bouma, L. G., Brammer, G. B., Bray, E. M., Breytenbach, H., Buddelmeijer, H., Burke, D. J., Calderone, G., Cano Rodríguez, J. L., Cara, M., Cardoso, J. V. M., Cheedella, S., Copin, Y., Corrales, L., Crichton, D., D'Avella, D., Deil, C., Depagne, É., Dietrich, J. P., Donath, A., Droettboom, M., Earl, N., Erben, T., Fabbro, S., Ferreira, L. A., Finethy, T., Fox, R. T., Garrison, L. H., Gibbons, S. L. J., Goldstein, D. A., Gommers, R., Greco, J. P., Greenfield, P., Groener, A. M., Grollier, F., Hagen, A., Hirst, P., Homeier, D., Horton, A. J., Hosseinzadeh, G., Hu, L., Hunkeler, J. S., Ivezić, Ž., Jain, A., Jenness, T., Kanarek, G., Kendrew, S., Kern, N. S., Kerzendorf, W. E., Khvalko, A., King, J., Kirkby, D., Kulkarni, A. M., Kumar, A., Lee, A., Lenz, D., Littlefair, S. P., Ma, Z., Macleod, D. M., Mastropietro, M., McCully, C., Montagnac, S., Morris, B. M., Mueller, M., Mumford, S. J., Muna, D. et al. (Sept. 2018). 'The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package'. In: AJ 156.3, 123, p. 123. DOI: 10.3847/1538-3881/aabc4f. arXiv: 1801.02634 [astro-ph.IM].

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., Nöthe, M., Donath, A., Tollerud, E., Morris, B. M., Ginsburg, A., Vaher, E., Weaver, B. A., Tocknell, J., Jamieson, W., van Kerkwijk, M. H., Robitaille, T. P., Merry, B., Bachetti, M., Günther, H. M., Aldcroft, T. L., Alvarado-Montes, J. A., Archibald, A. M., Bódi, A., Bapat, S., Barentsen, G., Bazán, J., Biswas, M., Boquien, M., Burke, D. J., Cara, D., Cara, M., Conroy, K. E., Conseil, S., Craig, M. W., Cross, R. M., Cruz, K. L., D'Eugenio, F., Dencheva, N., Devillepoix, H. A. R., Dietrich, J. P., Eigenbrot, A. D., Erben, T., Ferreira, L., Foreman-Mackey, D., Fox, R., Freij, N., Garg, S., Geda, R., Glattly, L., Gondhalekar, Y., Gordon, K. D., Grant, D., Greenfield, P., Groener, A. M., Guest, S., Gurovich, S., Handberg, R., Hart, A., Hatfield-Dodds, Z., Homeier, D., Hosseinzadeh, G., Jenness, T., Jones, C. K., Joseph, P., Kalmbach, J. B., Karamehmetoglu, E., Kałuszyński, M., Kelley, M. S. P., Kern, N., Kerzendorf, W. E., Koch, E. W., Kulumani, S., Lee, A., Ly, C., Ma, Z., MacBride, C., Maljaars, J. M., Muna, D., Murphy, N. A., Norman, H., O'Steen, R., Oman, K. A., Pacifici, C., Pascual, S., Pascual-Granado, J., Patil, R. R., Perren, G. I., Pickering, T. E., Rastogi, T., Roulston, B. R., Ryan, D. F., Rykoff, E. S. et al. (Aug. 2022). 'The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package'. In: ApJ 935.2, 167, p. 167. DOI: 10.3847/1538-4357/ac7c74. arXiv: 2206.14220 [astro-ph.IM].

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M. and Streicher, O. (Oct. 2013). 'Astropy: A community Python package for astronomy'. In: A&A 558, A33, A33. DOI: 10.1051/0004-6361/201322068. arXiv: 1307.6212 [astro-ph.IM].

Bahcall, J. N. and Kozlovsky, B.-Z. (Mar. 1969). 'Some Models of the Emission-Line Region of 3c 273'. In: ApJ 155, p. 1077. DOI: 10.1086/149935.

Baldwin, J. A., Phillips, M. M. and Terlevich, R. (Feb. 1981). 'Classification parameters for the emission-line spectra of extragalactic objects.' In: PASP 93, pp. 5–19. DOI: 10.1086/130766.

Ball, N. M. and Brunner, R. J. (Jan. 2010). 'Data Mining and Machine Learning in Astronomy'. In: *International Journal of Modern Physics D* 19.7, pp. 1049–1106. DOI: 10.1142/S0218271810017160. arXiv: 0906.2173 [astro-ph.IM].

Ball, N. M., Brunner, R. J., Myers, A. D., Strand, N. E., Alberts, S. L. and Tcheng, D. (Aug. 2008). 'Robust Machine Learning Applied to Astronomical Data Sets. III. Probabilistic Photometric Redshifts for Galaxies and Quasars in the SDSS and GALEX'. In: ApJ 683.1, pp. 12–21. DOI: 10.1086/589646. arXiv: 0804.3413 [astro-ph].

Baltay, C., Grossman, L., Howard, R., Rabinowitz, D., Arcavi, I., Barbour, N., Burke, J., Contreras, C., Dilday, B., Graham, M., Hiramatsu, D., Hossenzadeh, G., Howell, D. A., McCully, C., McKinnon, R., Ment, K., Montesi, R., Pellegrino, C. and Valenti, S. (Apr. 2021). 'Low-redshift Type Ia Supernova from the LSQ/LCO Collaboration'. In: PASP 133.1022, 044002, p. 044002. DOI: 10.1088/1538-3873/abd417.

Baron, D. (Apr. 2019). 'Machine Learning in Astronomy: a practical overview'. In: *arXiv e-prints*, arXiv:1904.07248, arXiv:1904.07248. arXiv: 1904.07248 [astro-ph.IM].

Baron, D. and Poznanski, D. (Mar. 2017). 'The weirdest SDSS galaxies: results from an outlier detection algorithm'. In: MNRAS 465.4, pp. 4530–4555. DOI: 10.1093/mnras/stw3021. arXiv: 1611.07526 [astro-ph.GA].

Barrows, R. S., Comerford, J. M., Stern, D. and Assef, R. J. (Dec. 2021). 'A Catalog of Host Galaxies for WISE-selected AGN: Connecting Host Properties with Nuclear Activity and Identifying Contaminants'. In: ApJ 922.2, 179, p. 179. DOI: 10.3847/1538-4357/ac1352. arXiv: 2107.02815 [astro-ph.GA].

Baum, W. A. (Feb. 1957). 'Photoelectric determinations of redshifts beyond 0.2 c.' In: AJ 62, pp. 6–7. DOI: 10.1086/107433.

— (Jan. 1962). 'Photoelectric Magnitudes and Red-Shifts'. In: *Problems of Extra-Galactic Research*. Ed. by G. C. McVittie. Vol. 15, p. 390.

Beifiori, A., Courteau, S., Corsini, E. M. and Zhu, Y. (Jan. 2012). 'On the correlations between galaxy properties and supermassive black hole mass'. In: MNRAS 419.3, pp. 2497–2528. DOI: 10.1111/j.1365-2966.2011.19903.x. arXiv: 1109.6265 [astro-ph.CO].

Birchall, K. L., Watson, M. G. and Aird, J. (Feb. 2020). 'X-ray detected AGN in SDSS dwarf galaxies'. In: MNRAS 492.2, pp. 2268–2284. DOI: 10.1093/mnras/staa040. arXiv: 2001.03135 [astro-ph.GA].

Blandford, R., Meier, D. and Readhead, A. (Aug. 2019). 'Relativistic Jets from Active Galactic Nuclei'. In: ARA&A 57, pp. 467–509. DOI: 10.1146/annurev-astro-081817-051948. arXiv: 1812.06025 [astro-ph.HE].

Blecha, L., Snyder, G. F., Satyapal, S. and Ellison, S. L. (Aug. 2018). 'The power of infrared AGN selection in mergers: a theoretical study'. In: MNRAS 478.3, pp. 3056–3071. DOI: 10.1093/mnras/sty1274. arXiv: 1711.02094 [astro-ph.GA].

Bonaldi, A., Bonato, M., Galluzzi, V., Harrison, I., Massardi, M., Kay, S., De Zotti, G. and Brown, M. L. (Jan. 2019). 'The Tiered Radio Extragalactic Continuum Simulation (T-RECS)'. In: MNRAS 482.1, pp. 2–19. DOI: 10.1093/mnras/sty2603. arXiv: 1805.05222 [astro-ph.GA].

Bonato, M., Prandoni, I., De Zotti, G., Brienza, M., Morganti, R. and Vaccari, M. (Jan. 2021). 'New constraints on the 1.4 GHz source number counts and luminosity functions in the Lockman Hole field'. In: MNRAS 500.1, pp. 22–33. DOI: 10.1093/mnras/staa3218. arXiv: 2010.08748 [astro-ph.GA].

Bonnarel, F., Fernique, P., Bienaymé, O., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M. and Bartlett, J. G. (Apr. 2000). 'The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources'. In: A&AS 143, pp. 33–40. DOI: 10.1051/aas:2000331.

Bouwens, R., González-López, J., Aravena, M., Decarli, R., Novak, M., Stefanon, M., Walter, F., Boogaard, L., Carilli, C., Dudzevičiūtė, U., Smail, I., Daddi, E., da Cunha, E., Ivison, R., Nanayakkara, T., Cortes, P., Cox, P., Inami, H., Oesch, P., Popping, G., Riechers, D., van der Werf, P., Weiss, A., Fudamoto, Y. and Wagg, J. (Oct. 2020). 'The ALMA Spectroscopic Survey Large Program: The Infrared Excess of z = 1.5-10 UV-selected Galaxies and the Implied High-redshift Star Formation History'. In: ApJ 902.2, 112, p. 112. DOI: 10.3847/1538-4357/abb830. arXiv: 2009.10727 [astro-ph.GA].

Bowler, R. A. A., Adams, N. J., Jarvis, M. J. and Häußler, B. (Mar. 2021). 'The rapid transition from star formation to AGN-dominated rest-frame ultraviolet light at z ≃ 4'. In: MNRAS 502.1, pp. 662–677. DOI: 10.1093/mnras/stab038. arXiv: 2101.01195 [astro-ph.GA].

Brandt, W. N. and Alexander, D. M. (Jan. 2015). 'Cosmic X-ray surveys of distant active galaxies. The demographics, physics, and ecology of growing supermassive black holes'. In: A&A Rev. 23, 1, p. 1. DOI: 10.1007/s00159-014-0081-z. arXiv: 1501.01982 [astro-ph.HE].

Braun, R., Bonaldi, A., Bourke, T., Keane, E. and Wagg, J. (Dec. 2019). 'Anticipated Performance of the Square Kilometre Array – Phase 1 (SKA1)'. In: *arXiv e-prints*, arXiv:1912.12699, arXiv:1912.12699. DOI: 10.48550/arXiv.1912.12699. arXiv: 1912.12699 [astro-ph.IM].

Breedt, E., Arévalo, P., McHardy, I. M., Uttley, P., Sergeev, S. G., Minezaki, T., Yoshii, Y., Gaskell, C. M., Cackett, E. M., Horne, K. and Koshida, S. (Mar. 2009). 'Long-term optical and X-ray variability of the Seyfert galaxy Markarian 79'. In: MNRAS 394.1, pp. 427–437. DOI: 10.1111/j.1365-2966.2008.14302.x. arXiv: 0812.0810 [astro-ph].

Breedt, E., McHardy, I. M., Arévalo, P., Uttley, P., Sergeev, S. G., Minezaki, T., Yoshii, Y., Sakata, Y., Lira, P. and Chesnok, N. G. (Apr. 2010). 'Twelve years of X-ray and optical variability in the Seyfert galaxy NGC 4051'. In: MNRAS 403.2, pp. 605–619. DOI: 10.1111/j.1365-2966.2009.16146.x. arXiv: 0912.0544 [astro-ph.GA].

Breiman, L. (Aug. 1996). 'Bagging predictors'. In: *Machine Learning* 24.2, pp. 123–140. ISSN: 1573-0565. DOI: 10.1007/BF00058655. URL: https://doi.org/10.1007/BF00058655.

— (Oct. 2001). 'Random Forests'. In: *Machine Learning* 45.1, pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

— (2003). 'Manual on setting up, using, and understanding random forests v4. 0'. In: *Statistics Department University of California Berkeley, CA, USA*.

Brescia, M., Cavuoti, S., Razim, O., Amaro, V., Riccio, G. and Longo, G. (2021). 'Photometric Redshifts With Machine Learning, Lights and Shadows on a Complex Data Science Use Case'. In: *Frontiers in Astronomy and Space Sciences* 8, p. 70. ISSN: 2296-987X. DOI: 10.3389/fspas.2021.658229. URL: https://www.frontiersin.org/article/10.3389/fspas.2021.658229.

Brescia, M., Salvato, M., Cavuoti, S., Ananna, T. T., Riccio, G., LaMassa, S. M., Urry, C. M. and Longo, G. (Oct. 2019). 'Photometric redshifts for X-ray-selected active galactic nuclei in the eROSITA era'. In: MNRAS 489.1, pp. 663–680. DOI: 10.1093/mnras/stz2159. arXiv: 1909.00606 [astro-ph.IM].

Brier, G. W. (1950). 'Verification of Forecasts Expressed in Terms of Probability'. In: *Monthly Weather Review* 78.1, pp. 1–3. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml.

Bröcker, J. and Smith, L. A. (2007). 'Increasing the Reliability of Reliability Diagrams'. In: *Weather and Forecasting* 22.3, pp. 651–661. DOI: 10.1175/WAF993.1. URL: https://journals.ametsoc.org/view/journals/wefo/22/3/waf993_1.xml.

Brown, M. J. I., Duncan, K. J., Landt, H., Kirk, M., Ricci, C., Kamraj, N., Salvato, M. and Ananna, T. (Nov. 2019). 'The spectral energy distributions of active galactic nuclei'. In: MNRAS 489.3, pp. 3351–3367. DOI: 10.1093/mnras/stz2324. arXiv: 1908.03720 [astro-ph.GA].

Buchner, J. (Oct. 2019). 'Collaborative Nested Sampling: Big Data versus Complex Physical Models'. In: PASP 131.1004, p. 108005. DOI: 10.1088/1538-3873/aae7fc. arXiv: 1707.04476 [stat.CO].

Buisson, D. J. K., Lohfink, A. M., Alston, W. N. and Fabian, A. C. (Jan. 2017). 'Ultraviolet and X-ray variability of active galactic nuclei with Swift'. In: MNRAS 464.3, pp. 3194–3218. DOI: 10.1093/mnras/stw2486. arXiv: 1609.08638 [astro-ph.HE].

Burhanudin, U. F., Maund, J. R., Killestein, T., Ackley, K., Dyer, M. J., Lyman, J., Ulaczyk, K., Cutter, R., Mong, Y. .-., Steeghs, D., Galloway, D. K., Dhillon, V., O'Brien, P., Ramsay, G., Noysena, K., Kotak, R., Breton, R. P., Nuttall, L., Pallé, E., Pollacco, D., Thrane, E., Awiphan, S., Chote, P., Chrimes, A., Daw, E., Duffy, C., Eyles-Ferris, R., Gompertz, B., Heikkilä, T., Irawati, P., Kennedy, M. R., Levan, A., Littlefair, S., Makrygianni, L., Mata-Sánchez, D., Mattila, S., McCormac, J., Mkrtichian, D., Mullaney, J., Sawangwit, U., Stanway, E., Starling, R., Strøm, P., Tooke, S. and Wiersema, K. (Aug. 2021). 'Light-curve classification with recurrent neural networks for GOTO: dealing with imbalanced data'. In: MNRAS 505.3, pp. 4345–4361. DOI: 10.1093/mnras/stab1545. arXiv: 2105.11169 [astro-ph.IM].

Capetti, A., Brienza, M., Baldi, R. D., Giovannini, G., Morganti, R., Hardcastle, M. J., Rottgering, H. J. A., Brunetti, G. F., Best, P. N. and Miley, G. (Oct. 2020). 'The LOFAR view of FR 0 radio galaxies'. In: A&A 642, A107, A107. DOI: 10.1051/0004-6361/202038671. arXiv: 2008.08099 [astro-ph.GA].

Carroll, B. W. and Ostlie, D. A. (2017). *An Introduction to Modern Astrophysics*. 2nd ed. Cambridge University Press. DOI: 10.1017/9781108380980.

Carvajal, R., Bauer, F. E., Bouwens, R. J., Oesch, P. A., González-López, J., Anguita, T., Aravena, M., Demarco, R., Guaita, L., Infante, L., Kim, S., Kneissl, R., Koekemoer, A. M., Messias, H., Treister, E., Villard, E., Zitrin, A. and Troncoso, P. (Jan. 2020). 'The ALMA Frontier Fields Survey. V. ALMA Stacking of Lyman-Break Galaxies in Abell 2744, Abell 370, Abell S1063, MACSJ0416.1-2403 and MACSJ1149.5+2223'. In: A&A 633, A160, A160. DOI: 10.1051/0004-6361/201936260. arXiv: 1912.02916 [astro-ph.GA].

Carvajal, R., Matute, I., Afonso, J., Amarantidis, S., Barbosa, D., Cunha, P. and Humphrey, A. (Oct. 2021). 'Exploring New Redshift Indicators for Radio-Powerful AGN'. In: *Galaxies* 9.4, p. 86. DOI: 10.3390/galaxies9040086. arXiv: 2111.00778 [astro-ph.GA].

Ceccarelli, L., Duplancic, F. and Garcia Lambas, D. (Jan. 2022). 'The impact of void environment on AGN'. In: MNRAS 509.2, pp. 1805–1819. DOI: 10.1093/mnras/stab2902.

Chambers, K. C., Magnier, E. A., Metcalfe, N., Flewelling, H. A., Huber, M. E., Waters, C. Z., Denneau, L., Draper, P. W., Farrow, D., Finkbeiner, D. P., Holmberg, C., Koppenhoefer, J., Price, P. A., Rest, A., Saglia, R. P., Schlafly, E. F., Smartt, S. J., Sweeney, W., Wainscoat, R. J., Burgett, W. S., Chastel, S., Grav, T., Heasley, J. N., Hodapp, K. W., Jedicke, R., Kaiser, N., Kudritzki, R. .-., Luppino, G. A., Lupton, R. H., Monet, D. G., Morgan, J. S., Onaka, P. M., Shiao, B., Stubbs, C. W., Tonry, J. L., White, R., Bañados, E., Bell, E. F., Bender, R., Bernard, E. J., Boegner, M., Boffi, F., Botticella, M. T., Calamida, A., Casertano, S., Chen, W. .-., Chen, X., Cole, S., Deacon, N., Frenk, C., Fitzsimmons, A., Gezari, S., Gibbs, V., Goessl, C., Goggia, T., Gourgue, R., Goldman, B., Grant, P., Grebel, E. K., Hambly, N. C., Hasinger, G., Heavens, A. F., Heckman, T. M., Henderson, R., Henning, T., Holman, M., Hopp, U., Ip, W. .-., Isani, S., Jackson, M., Keyes, C. D., Koekemoer, A. M., Kotak, R., Le, D., Liska, D., Long, K. S., Lucey, J. R., Liu, M., Martin, N. F., Masci, G., McLean, B., Mindel, E., Misra, P., Morganson, E., Murphy, D. N. A., Obaika, A., Narayan, G., Nieto-Santisteban, M. A.,

Norberg, P., Peacock, J. A., Pier, E. A., Postman, M., Primak, N., Rae, C., Rai, A., Riess, A., Riffeser, A. et al. (Dec. 2016). 'The Pan-STARRS1 Surveys'. In: *arXiv e-prints*, arXiv:1612.05560. arXiv: 1612.05560 [astro-ph.IM].

Chattopadhyay, A. K. (2017). 'Incomplete Data in Astrostatistics'. In: *Wiley StatsRef: Statistics Reference Online*. American Cancer Society, pp. 1–12. ISBN: 9781118445112. DOI: https://doi.org/10.1002/9781118445112.stat07942.

Chaves-Montero, J., Bonoli, S., Salvato, M., Greisel, N., Díaz-García, L. A., López-Sanjuan, C., Viironen, K., Fernández-Soto, A., Pović, M., Ascaso, B., Arnalte-Mur, P., Masegosa, J., Matute, I., Márquez, I., Cenarro, A. J., Abramo, L. R., Ederoclite, A., Alfaro, E. J., Marin-Franch, A., Varela, J. and Cristobal-Hornillos, D. (Dec. 2017). 'ELDAR, a new method to identify AGN in multi-filter surveys: the ALHAMBRA test case'. In: MNRAS 472.2, pp. 2085–2106. DOI: 10.1093/mnras/stx2054. arXiv: 1707.07690 [astro-ph.GA].

Chen, H., Garrett, M. A., Chi, S., Thomson, A. P., Barthel, P. D., Alexander, D. M., Muxlow, T. W. B., Beswick, R. J., Radcliffe, J. F., Wrigley, N. H., Guidetti, D., Bondi, M., Prandoni, I., Smail, I., McHardy, I. and Argo, M. K. (June 2020). 'Searching for obscured AGN in z ∼ 2 submillimetre galaxies'. In: A&A 638, A113, A113. DOI: 10.1051/0004-6361/201937162. arXiv: 2004.09356 [astro-ph.GA].

Chen, T. and Guestrin, C. (2016). 'XGBoost: A Scalable Tree Boosting System'. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

Clarke, A. O., Scaife, A. M. M., Greenhalgh, R. and Griguta, V. (July 2020). 'Identifying galaxies, quasars, and stars with machine learning: A new catalogue of classifications for 111 million SDSS sources without spectra'. In: A&A 639, A84, A84. DOI: 10.1051/0004-6361/201936770. arXiv: 1909.10963 [astro-ph.GA].

Condon, J. J. (Jan. 1992). 'Radio emission from normal galaxies.' In: ARA&A 30, pp. 575–611. DOI: 10.1146/annurev.aa.30.090192.003043.

Costa-Climent, R., Haftor, D. M. and Staniewski, M. W. (2023). 'Using machine learning to create and capture value in the business models of small and medium-sized enterprises'. In: *International Journal of Information Management* 73, p. 102637. ISSN: 0268-4012. DOI: https://doi.org/10.1016/j.ijinfomgt.2023.102637. URL: https://www.sciencedirect.com/science/article/pii/S026840122300018X.

Cover, T. and Hart, P. (1967). 'Nearest neighbor pattern classification'. In: *IEEE Transactions on Information Theory* 13.1, pp. 21–27. DOI: 10.1109/TIT.1967.1053964.

Cramér, H. (1946). *Mathematical methods of statistics*. English. Princeton University Press Princeton, xvi, 575 p.

Cunha, P. A. C. and Humphrey, A. (Oct. 2022). 'Photometric redshift-aided classification using ensemble learning'. In: A&A 666, A87, A87. DOI: 10.1051/0004-6361/202243135. arXiv: 2204.02080 [astro-ph.IM].

Curran, S. J. (May 2022). 'Quasar photometric redshifts from incomplete data using deep learning'. In: MNRAS 512.2, pp. 2099–2109. DOI: 10.1093/mnras/stac660. arXiv: 2203.03679 [astro-ph.CO].

Curran, S. J., Moss, J. P. and Perrott, Y. C. (July 2022). 'Redshifts of radio sources in the Million Quasars Catalogue from machine learning'. In: MNRAS 514.1, pp. 1–19. DOI: 10.1093/mnras/stac1333. arXiv: 2205.04587 [astro-ph.CO].

Cutri, R. M., Skrutskie, M. F., van Dyk, S., Beichman, C. A., Carpenter, J. M., Chester, T., Cambresy, L., Evans, T., Fowler, J., Gizis, J., Howard, E., Huchra, J., Jarrett, T., Kopan, E. L., Kirkpatrick, J. D., Light, R. M., Marsh, K. A., McCallon, H., Schneider, S., Stiening, R., Sykes, M., Weinberg, M., Wheaton, W. A., Wheelock, S. and Zacarias, N. (2003a). *2MASS All Sky Catalog of point sources.*

— (June 2003b). 'VizieR Online Data Catalog: 2MASS All-Sky Catalog of Point Sources (Cutri+ 2003)'. In: *VizieR Online Data Catalog*, II/246, pp. II/246.

Cutri, R. M., Wright, E. L., Conrow, T., Fowler, J. W., Eisenhardt, P. R. M., Grillmair, C., Kirkpatrick, J. D., Masci, F., McCallon, H. L., Wheelock, S. L., Fajardo-Acosta, S., Yan, L., Benford, D., Harbut, M., Jarrett, T., Lake, S., Leisawitz, D., Ressler, M. E., Stanford, S. A., Tsai, C. W., Liu, F., Helou, G., Mainzer, A., Gettings, D., Gonzalez, A., Hoffman, D., Marsh, K. A., Padgett, D., Skrutskie, M. F., Beck, R. P., Papin, M. and Wittman, M. (Nov. 2013). *Explanatory Supplement to the AllWISE Data Release Products*.

Dahlen, T., Mobasher, B., Faber, S. M., Ferguson, H. C., Barro, G., Finkelstein, S. L., Finlator, K., Fontana, A., Gruetzbauch, R., Johnson, S., Pforr, J., Salvato, M., Wiklind, T., Wuyts, S., Acquaviva, V., Dickinson, M. E., Guo, Y., Huang, J., Huang, K.-H., Newman, J. A., Bell, E. F., Conselice, C. J., Galametz, A., Gawiser, E., Giavalisco, M., Grogin, N. A., Hathi, N., Kocevski, D., Koekemoer, A. M., Koo, D. C., Lee, K.-S., McGrath, E. J., Papovich, C., Peth, M., Ryan, R., Somerville, R., Weiner, B. and Wilson, G. (Oct. 2013). 'A Critical Assessment of Photometric Redshift Methods: A CANDELS Investigation'. In: ApJ 775.2, 93, p. 93. DOI: 10.1088/0004-637X/775/2/93. arXiv: 1308.5353 [astro-ph.CO].

Davidson, K. and Netzer, H. (Oct. 1979). 'The emission lines of quasars and similar objects'. In: *Reviews of Modern Physics* 51.4, pp. 715–766. DOI: 10.1103/RevModPhys.51.715.

Davies, L. J. M., Robotham, A. S. G., Driver, S. P., Lagos, C. P., Cortese, L., Mannering, E., Foster, C., Lidman, C., Hashemizadeh, A., Koushan, S., O'Toole, S., Baldry, I. K., Bilicki, M., Bland-Hawthorn, J., Bremer, M. N., Brown, M. J. I., Bryant, J. J., Catinella, B., Croom, S. M., Grootes, M. W., Holwerda, B. W., Jarvis, M. J., Maddox, N., Meyer, M., Moffett, A. J., Phillipps, S., Taylor, E. N., Windhorst, R. A. and Wolf, C. (Oct. 2018). 'Deep Extragalactic VIsible Legacy Survey (DEVILS): motivation,design, and target catalogue'. In: MNRAS 480.1, pp. 768–799. DOI: 10.1093/mnras/sty1553. arXiv: 1806.05808 [astro-ph.GA].

Delhaize, J., Heywood, I., Prescott, M., Jarvis, M. J., Delvecchio, I., Whittam, I. H., White, S. V., Hardcastle, M. J., Hale, C. L., Afonso, J., Ao, Y., Brienza, M., Brüggen, M., Collier, J. D., Daddi, E., Glowacki, M., Maddox, N., Morabito, L. K., Prandoni, I., Randriamanakoto, Z., Sekhar, S., An, F., Adams, N. J., Blyth, S., Bowler, R. A. A., Leeuw, L., Marchetti, L., Randriamampandry, S. M., Thorat, K., Seymour, N., Smirnov, O., Taylor, A. R., Tasse, C. and Vaccari, M. (Mar. 2021). 'MIGHTEE: are giant radio galaxies more common than we thought?' In: MNRAS 501.3, pp. 3833–3845. DOI: 10.1093/mnras/staa3837. arXiv: 2012.05759 [astro-ph.GA].

Desai, S. and Strachan, A. (June 2021). 'Parsimonious neural networks learn interpretable physical laws'. In: *Scientific Reports* 11.1, p. 12761. ISSN: 2045-2322. DOI: 10.1038/s41598-021-92278-w. URL: https://doi.org/10.1038/s41598-021-92278-w.

Dice, L. R. (1945). 'Measures of the Amount of Ecologic Association Between Species'. In: *Ecology* 26.3, pp. 297–302. ISSN: 00129658, 19399170. URL: http://www.jstor.org/stable/1932409 (visited on 04/10/2022).

Dobbels, W. and Baes, M. (Nov. 2021). 'Predicting far-infrared maps of galaxies via machine learning techniques'. In: A&A 655, A34, A34. DOI: 10.1051/0004-6361/202142084. arXiv: 2110.01704 [astro-ph.GA].

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., Salvato, M., Sanders, D. B., Trump, J. R. and Zamorani, G. (Apr. 2012). 'Identifying Luminous Active Galactic Nuclei in Deep Surveys: Revised IRAC Selection Criteria'. In: ApJ 748.2, 142, p. 142. DOI: 10.1088/0004-637X/748/2/142. arXiv: 1201.3899 [astro-ph.CO].

Dorogush, A. V., Ershov, V. and Gulin, A. (2018). 'CatBoost: gradient boosting with categorical features support'. In: *CoRR* abs/1810.11363. arXiv: 1810.11363. URL: http://arxiv.org/abs/1810.11363.

Driver, S. P., Hill, D. T., Kelvin, L. S., Robotham, A. S. G., Liske, J., Norberg, P., Baldry, I. K., Bamford, S. P., Hopkins, A. M., Loveday, J., Peacock, J. A., Andrae, E., Bland-Hawthorn, J., Brough, S., Brown, M. J. I., Cameron, E., Ching, J. H. Y., Colless, M., Conselice, C. J., Croom, S. M., Cross, N. J. G., de Propris, R., Dye, S., Drinkwater, M. J., Ellis, S., Graham, A. W., Grootes, M. W., Gunawardhana, M., Jones, D. H., van Kampen, E., Maraston, C., Nichol, R. C., Parkinson, H. R., Phillipps, S., Pimbblet, K., Popescu, C. C., Prescott, M., Roseboom, I. G., Sadler, E. M., Sansom, A. E., Sharp, R. G., Smith, D. J. B., Taylor, E., Thomas, D., Tuffs, R. J., Wijesinghe, D., Dunne, L., Frenk, C. S., Jarvis, M. J., Madore, B. F., Meyer, M. J., Seibert, M., Staveley-Smith, L., Sutherland, W. J. and Warren, S. J. (May 2011). 'Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release'. In: MNRAS 413.2, pp. 971–995. DOI: 10.1111/j.1365-2966.2010.18188.x. arXiv: 1009.0614 [astro-ph.CO].

Duboue, P. (2020). *The Art of Feature Engineering: Essentials for Machine Learning*. Cambridge University Press. ISBN: 9781108709385. URL: https://books.google.pt/books?id=%5C_BzhDwAAQBAJ.

Duncan, K. J., Sabater, J., Röttgering, H. J. A., Jarvis, M. J., Smith, D. J. B., Best, P. N., Callingham, J. R., Cochrane, R., Croston, J. H., Hardcastle, M. J., Mingo, B., Morabito, L., Nisbet, D., Prandoni, I., Shimwell, T. W., Tasse, C., White, G. J., Williams, W. L., Alegre, L., Chyży, K. T., Gürkan, G., Hoeft, M., Kondapally, R., Mechev, A. P., Miley, G. K., Schwarz, D. J. and van Weeren, R. J. (Feb. 2019). 'The LOFAR Two-metre Sky Survey. IV. First Data Release: Photometric redshifts and rest-frame magnitudes'. In: A&A 622, A3, A3. DOI: 10.1051/0004-6361/201833562. arXiv: 1811.07928 [astro-ph.GA].

Euclid Collaboration, Bisigello, L., Conselice, C. J., Baes, M., Bolzonella, M., Brescia, M., Cavuoti, S., Cucciati, O., Humphrey, A., Hunt, L. K., Maraston, C., Pozzetti, L., Tortora, C., van Mierlo, S. E., Aghanim, N., Auricchio, N., Baldi, M., Bender, R., Bodendorf, C., Bonino, D., Branchini, E., Brinchmann, J., Camera, S., Capobianco, V., Carbone, C., Carretero, J., Castander, F. J., Castellano, M., Cimatti, A., Congedo, G., Conversi, L., Copin, Y., Corcione, L., Courbin, F., Cropper, M., Da Silva, A., Degaudenzi, H., Douspis, M., Dubath, F., Duncan, C. A. J., Dupac, X., Dusini, S., Farrens, S., Ferriol, S., Frailis, M., Franceschi, E., Franzetti, P., Fumana, M., Garilli, B., Gillard, W., Gillis, B., Giocoli, C., Grazian, A., Grupp, F., Guzzo, L., Haugan, S. V. H., Holmes, W., Hormuth, F., Hornstrup, A., Jahnke, K., Kümmel, M., Kermiche, S., Kiessling, A., Kilbinger, M., Kohley, R., Kunz, M., Kurki-Suonio, H., Ligori, S., Lilje, P. B., Lloro, I., Maiorano, E., Mansutti, O., Marggraf, O., Markovic, K., Marulli, F., Massey, R., Maurogordato, S., Medinaceli, E., Meneghetti, M., Merlin, E., Meylan, G., Moresco, M., Moscardini, L., Munari, E., Niemi, S. M., Padilla, C., Paltani, S., Pasian, F., Pedersen, K., Pettorino, V., Polenta, G., Poncet, M., Popa, L., Raison, F., Renzi, A., Rhodes, J., Riccio, G. et al. (Apr. 2023a). 'Euclid preparation - XXIII. Derivation of galaxy physical properties with deep machine learning using mock fluxes and H-band images'. In: MNRAS 520.3, pp. 3529–3548. DOI: 10.1093/mnras/stac3810. arXiv: 2206.14944 [astro-ph.GA].

Euclid Collaboration, Humphrey, A., Bisigello, L., Cunha, P. A. C., Bolzonella, M., Fotopoulou, S., Caputi, K., Tortora, C., Zamorani, G., Papaderos, P., Vergani, D., Brinchmann, J., Moresco, M., Amara, A., Auricchio, N., Baldi, M., Bender, R., Bonino, D., Branchini, E., Brescia, M., Camera, S., Capobianco, V., Carbone, C., Carretero, J., Castander, F. J., Castellano, M., Cavuoti, S., Cimatti, A., Cledassou, R., Congedo, G., Conselice, C. J., Conversi, L., Copin, Y., Corcione, L., Courbin, F., Cropper, M., Da Silva, A., Degaudenzi, H., Douspis, M., Dubath, F., Duncan, C. A. J., Dupac, X., Dusini, S., Farrens, S., Ferriol, S., Frailis, M., Franceschi, E., Fumana, M., Gómez-Alvarez, P., Galeotta, S., Garilli, B., Gillard, W., Gillis, B., Giocoli, C., Grazian, A., Grupp, F., Guzzo, L., Haugan, S. V. H., Holmes, W., Hormuth, F., Jahnke, K., Kümmel, M., Kermiche, S., Kiessling, A., Kilbinger, M., Kitching, T., Kohley, R., Kunz, M., Kurki-Suonio, H., Ligori, S., Lilje, P. B., Lloro, I., Maiorano, E., Mansutti, O., Marggraf, O., Markovic, K., Marulli, F., Massey, R., Maurogordato, S., McCracken, H. J., Medinaceli, E., Melchior, M., Meneghetti, M., Merlin, E., Meylan, G., Moscardini, L., Munari, E., Nakajima, R., Niemi, S. M., Nightingale, J., Padilla, C., Paltani, S., Pasian, F., Pedersen, K., Pettorino, V., Pires, S., Poncet, M. et al. (Mar. 2023b). 'Euclid preparation. XXII. Selection of quiescent galaxies from mock photometry using machine learning'. In: A&A 671, A99, A99. DOI: 10.1051/0004-6361/202244307. arXiv: 2209.13074 [astro-ph.IM].

Fan, X., Banados, E. and Simcoe, R. A. (2023). 'Quasars and the Intergalactic Medium at Cosmic Dawn'. In: ARA&A 61. DOI: 10.1146/annurev-astro-052920-102455. arXiv: 2212.06907 [astro-ph.GA].

Ferrarese, L. and Merritt, D. (Aug. 2000). 'A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies'. In: ApJ 539.1, pp. L9–L12. DOI: 10.1086/312838. arXiv: astro-ph/0006053 [astro-ph].

Flesch, E. W. (May 2021). 'The Million Quasars (Milliquas) v7.2 Catalogue, now with VLASS associations. The inclusion of SDSS-DR16Q quasars is detailed'. In: *arXiv e-prints*, arXiv:2105.12985, arXiv:2105.12985. arXiv: 2105.12985 [astro-ph.GA].

Flewelling, H. A., Magnier, E. A., Chambers, K. C., Heasley, J. N., Holmberg, C., Huber, M. E., Sweeney, W., Waters, C. Z., Calamida, A., Casertano, S., Chen, X., Farrow, D., Hasinger, G., Henderson, R., Long, K. S., Metcalfe, N., Narayan, G., Nieto-Santisteban, M. A., Norberg, P., Rest, A., Saglia, R. P., Szalay, A., Thakar, A. R., Tonry, J. L., Valenti, J., Werner, S., White, R., Denneau, L., Draper, P. W., Hodapp, K. W., Jedicke, R., Kaiser, N., Kudritzki, R. P., Price, P. A., Wainscoat, R. J., Chastel, S., McLean, B., Postman, M. and Shiao, B. (Nov. 2020). 'The Pan-STARRS1 Database and Data Products'. In: ApJS 251.1, 7, p. 7. DOI: 10.3847/1538-4365/abb82d. arXiv: 1612.05243 [astro-ph.IM].

Frederiksen, T. F., Graur, O., Hjorth, J., Maoz, D. and Poznanski, D. (Mar. 2014). 'Spectroscopic identification of a redshift 1.55 supernova host galaxy from the Subaru Deep Field Supernova Survey'. In: A&A 563, A140, A140. DOI: 10.1051/0004-6361/201321795. arXiv: 1211.2208 [astro-ph.CO].

Freund, Y. and Schapire, R. E. (1996). 'Experiments with a New Boosting Algorithm'. In: *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. ICML'96. Bari, Italy: Morgan Kaufmann Publishers Inc., pp. 148–156. ISBN: 1558604197.

Friedman, J. H. (2001). 'Greedy function approximation: A gradient boosting machine.' In: *The Annals of Statistics* 29.5, pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: https://doi.org/10.1214/aos/1013203451.

— (2002). 'Stochastic gradient boosting'. In: *Computational Statistics & Data Analysis* 38.4. Nonlinear Methods and Data Mining, pp. 367–378. ISSN: 0167-9473. DOI: https://doi.org/10.1016/S0167-9473(01)00065-2. URL: https://www.sciencedirect.com/science/article/pii/S0167947301000652.

Gaia Collaboration, Bailer-Jones, C. A. L., Teyssier, D., Delchambre, L., Ducourant, C., Garabato, D., Hatzidimitriou, D., Klioner, S. A., Rimoldini, L., Bellas-Velidis, I., Carballo, R., Carnerero, M. I., Diener, C., Fouesneau, M., Galluccio, L., Gavras, P., Krone-Martins, A., Raiteri, C. M., Teixeira, R., Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Arenou, F., Babusiaux, C., Biermann, M., Creevey, O. L., Evans, D. W., Eyer, L., Guerra, R., Hutton, A., Jordi, C., Lammers, U. L., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Soubiran, C., Tanga, P., Walton, N. A., Bastian, U., Drimmel, R., Jansen, F., Katz, D., Lattanzi, M. G., van Leeuwen, F., Bakker, J., Cacciari, C., Castañeda, J., De Angeli, F., Fabricius, C., Frémat, Y., Guerrier, A., Heiter, U., Masana, E., Messineo, R., Mowlavi, N., Nicolas, C., Nienartowicz, K., Pailler, F., Panuzzo, P., Riclet, F., Roux, W., Seabroke, G. M., Sordo, R., Thévenin, F., Gracia-Abril, G., Portell, J., Altmann, M., Andrae, R., Audard, M., Benson, K., Berthier, J., Blomme, R., Burgess, P. W., Busonero, D., Busso, G., Cánovas, H., Carry, B., Cellino, A., Cheek, N., Clementini, G., Damerdji, Y., Davidson, M., de Teodoro, P., Nuñez Campos, M., Dell'Oro, A., Esquej, P., Fernández-Hernández, J., Fraile, E., García-Lario, P., Gosset, E., Haigron, R., Halbwachs, J. .-. et al. (June 2023a). 'Gaia Data Release 3. The extragalactic content'. In: A&A 674, A41, A41. DOI: 10.1051/0004-6361/202243232. arXiv: 2206.05681 [astro-ph.GA].

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., Vallenari, A., Babusiaux, C., Bailer-Jones, C. A. L., Bastian, U., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Milligan, D. J.,

Panem, C., Poinsignon, V., Pourbaix, D., Randich, S., Sarri, G., Sartoretti, P., Siddiqui, H. I., Soubiran, C., Valette, V., van Leeuwen, F., Walton, N. A., Aerts, C., Arenou, F., Cropper, M., Drimmel, R., Høg, E., Katz, D., Lattanzi, M. G., O'Mullane, W., Grebel, E. K., Holland, A. D., Huc, C., Passot, X., Bramante, L., Cacciari, C., Castañeda, J., Chaoul, L., Cheek, N., De Angeli, F., Fabricius, C., Guerra, R., Hernández, J., Jean-Antoine-Piccolo, A., Masana, E., Messineo, R., Mowlavi, N., Nienartowicz, K., Ordóñez-Blanco, D., Panuzzo, P., Portell, J., Richards, P. J., Riello, M., Seabroke, G. M., Tanga, P., Thévenin, F., Torra, J., Els, S. G., Gracia-Abril, G., Comoretto, G., Garcia-Reinaldos, M., Lock, T., Mercier, E., Altmann, M., Andrae, R., Astraatmadja, T. L., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Busso, G., Carry, B., Cellino, A., Clementini, G., Cowell, S., Creevey, O., Cuypers, J., Davidson, M., De Ridder, J., de Torres, A., Delchambre, L., Dell'Oro, A., Ducourant, C., Frémat, Y., García-Torres, M., Gosset, E., Halbwachs, J. .-., Hambly, N. C., Harrison, D. L. et al. (Nov. 2016). 'The Gaia mission'. In: A&A 595, A1, A1. DOI: 10.1051/0004-6361/201629272. arXiv: 1609.04153 [astro-ph.IM].

Gaia Collaboration, Vallenari, A., Brown, A. G. A., Prusti, T., de Bruijne, J. H. J., Arenou, F., Babusiaux, C., Biermann, M., Creevey, O. L., Ducourant, C., Evans, D. W., Eyer, L., Guerra, R., Hutton, A., Jordi, C., Klioner, S. A., Lammers, U. L., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Soubiran, C., Tanga, P., Walton, N. A., Bailer-Jones, C. A. L., Bastian, U., Drimmel, R., Jansen, F., Katz, D., Lattanzi, M. G., van Leeuwen, F., Bakker, J., Cacciari, C., Castañeda, J., De Angeli, F., Fabricius, C., Fouesneau, M., Frémat, Y., Galluccio, L., Guerrier, A., Heiter, U., Masana, E., Messineo, R., Mowlavi, N., Nicolas, C., Nienartowicz, K., Pailler, F., Panuzzo, P., Riclet, F., Roux, W., Seabroke, G. M., Sordo, R., Thévenin, F., Gracia-Abril, G., Portell, J., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P. W., Busonero, D., Busso, G., Cánovas, H., Carry, B., Cellino, A., Cheek, N., Clementini, G., Damerdji, Y., Davidson, M., de Teodoro, P., Nuñez Campos, M., Delchambre, L., Dell'Oro, A., Esquej, P., Fernández-Hernández, J., Fraile, E., Garabato, D., García-Lario, P., Gosset, E., Haigron, R., Halbwachs, J. .-., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Holl, B., Janßen, K., Jevardat de Fombelle, G., Jordan, S. et al. (June 2023b). 'Gaia Data Release 3. Summary of the content and survey properties'. In: A&A 674, A1, A1. DOI: 10.1051/0004-6361/202243940. arXiv: 2208.00211 [astro-ph.GA].

Galametz, A., Grazian, A., Fontana, A., Ferguson, H. C., Ashby, M. L. N., Barro, G., Castellano, M., Dahlen, T., Donley, J. L., Faber, S. M., Grogin, N., Guo, Y., Huang, K.-H., Kocevski, D. D., Koekemoer, A. M., Lee, K.-S., McGrath, E. J., Peth, M., Willner, S. P., Almaini, O., Cooper, M., Cooray, A., Conselice, C. J., Dickinson, M., Dunlop, J. S., Fazio, G. G., Foucaud, S., Gardner, J. P., Giavalisco, M., Hathi, N. P., Hartley, W. G., Koo, D. C., Lai, K., de Mello, D. F., McLure, R. J., Lucas, R. A., Paris, D., Pentericci, L., Santini, P., Simpson, C., Sommariva, V., Targett, T., Weiner, B. J., Wuyts, S. and CANDELS Team (June 2013). 'CANDELS Multiwavelength Catalogs: Source Identification and Photometry in the CANDELS UKIDSS Ultra-deep Survey Field'. In: ApJS 206.2, 10, p. 10. DOI: 10.1088/0067-0049/206/2/10. arXiv: 1305.1823 [astro-ph.CO].

Garcia-Piquer, A., Morales, J. C., Ribas, I., Colomé, J., Guàrdia, J., Perger, M., Caballero, J. A., Cortés-Contreras, M., Jeffers, S. V., Reiners, A., Amado, P. J., Quirrenbach, A. and Seifert, W. (Aug. 2017). 'Efficient scheduling of astronomical observations. Application to the CARMENES radial-velocity survey'. In: A&A 604, A87, A87. DOI: 10.1051/0004-6361/201628577. arXiv: 1707.06052 [astro-ph.IM].

Garilli, B., Fumana, M., Franzetti, P., Paioro, L., Scodeggio, M., Le Fèvre, O., Paltani, S. and Scaramella, R. (July 2010). 'EZ: A Tool For Automatic Redshift Measurement'. In: PASP 122.893, p. 827. DOI: 10.1086/654903. arXiv: 1005.2825 [astro-ph.IM].

Garofalo, M., Botta, A. and Ventre, G. (June 2017). 'Astrophysics and Big Data: Challenges, Methods, and Tools'. In: *Astroinformatics*. Ed. by M. Brescia, S. G. Djorgovski, E. D. Feigelson, G. Longo and S. Cavuoti. Vol. 325, pp. 345–348. DOI: 10.1017/S1743921316012813. arXiv: 1703.05084 [astro-ph.IM].

Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., Richstone, D. and Tremaine, S. (Aug. 2000). 'A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion'. In: ApJ 539.1, pp. L13–L16. DOI: 10.1086/312840. arXiv: astro-ph/0006289 [astro-ph].

Getachew-Woreta, T., Pović, M., Masegosa, J., Perea, J., Beyoro-Amado, Z. and Márquez, I. (July 2022). 'Effect of AGN on the morphological properties of their host galaxies in the local Universe'. In: MNRAS 514.1, pp. 607–620. DOI: 10.1093/mnras/stac851. arXiv: 2203.07702 [astro-ph.GA].

Geurts, P., Ernst, D. and Wehenkel, L. (Apr. 2006). 'Extremely randomized trees'. In: *Machine Learning* 63.1, pp. 3–42. ISSN: 1573-0565. DOI: 10.1007/s10994-006-6226-1. URL: https://doi.org/10.1007/s10994-006-6226-1.

Giles, D. and Walkowicz, L. (Mar. 2019). 'Systematic serendipity: a test of unsupervised machine learning as a method for anomaly detection'. In: MNRAS 484.1, pp. 834–849. DOI: 10.1093/mnras/sty3461. arXiv: 1812.07156 [astro-ph.IM].

Giveon, U., Maoz, D., Kaspi, S., Netzer, H. and Smith, P. S. (July 1999). 'Long-term optical variability properties of the Palomar-Green quasars'. In: MNRAS 306.3, pp. 637–654. DOI: 10.1046/j.1365-8711.1999.02556.x. arXiv: astro-ph/9902254 [astro-ph].

Glahn, H. R. and Jorgensen, D. L. (1970). 'Climatological Aspects of the Brier p-score'. In: *Monthly Weather Review* 98.2, pp. 136–141. DOI: 10.1175/1520-0493(1970)098<0136:CAOTBP>2.3.CO;2. URL: https://journals.ametsoc.org/view/journals/mwre/98/2/1520-0493_1970_098_0136_caotbp_2_3_co_2.xml.

Glazebrook, K., Offer, A. R. and Deeley, K. (Jan. 1998). 'Automatic Redshift Determination by Use of Principal Component Analysis. I. Fundamentals'. In: ApJ 492.1, pp. 98–109. DOI: 10.1086/305039. arXiv: astro-ph/9707140 [astro-ph].

Glikman, E., Langgin, R., Johnstone, M. A., Yoon, I., Comerford, J. M., Simmons, B. D., Stacey, H., Lacy, M. and O'Meara, J. M. (July 2023). 'A Candidate Dual QSO at Cosmic Noon'. In: ApJ 951.1, L18, p. L18. DOI: 10.3847/2041-8213/acda2f. arXiv: 2306.00068 [astro-ph.GA].

Gordon, Y. A., Boyce, M. M., O'Dea, C. P., Rudnick, L., Andernach, H., Vantyghem, A. N., Baum, S. A., Bui, J.-P. and Dionyssiou, M. (Oct. 2020). 'A Catalog of Very Large Array Sky Survey Epoch 1 Quick Look Components, Sources, and Host Identifications'. In: *Research Notes of the American Astronomical Society* 4.10, 175, p. 175. DOI: 10.3847/2515-5172/abbe23.

Gültekin, K., Richstone, D. O., Gebhardt, K., Lauer, T. R., Tremaine, S., Aller, M. C., Bender, R., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Ho, L. C., Kormendy, J., Magorrian, J., Pinkney, J. and Siopis, C. (June 2009). 'The M-σ and M-L Relations in Galactic Bulges, and Determinations of Their Intrinsic Scatter'. In: ApJ 698.1, pp. 198–221. DOI: 10.1088/0004-637X/698/1/198. arXiv: 0903.4897 [astro-ph.GA].

Habouzit, M., Onoue, M., Bañados, E., Neeleman, M., Anglés-Alcázar, D., Walter, F., Pillepich, A., Davé, R., Jahnke, K. and Dubois, Y. (Apr. 2022). 'Co-evolution of massive black holes and their host galaxies at high redshift: discrepancies from six cosmological simulations and the key role of JWST'. In: MNRAS 511.3, pp. 3751–3767. DOI: 10.1093/mnras/stac225. arXiv: 2201.09892 [astro-ph.CO].

Hales, C. A., Murphy, T., Curran, J. R., Middelberg, E., Gaensler, B. M. and Norris, R. P. (Aug. 2012a). *BLOBCAT: Software to Catalog Blobs*. Astrophysics Source Code Library, record ascl:1208.009. ascl: 1208.009.

— (Sept. 2012b). 'BLOBCAT: software to catalogue flood-filled blobs in radio images of total intensity and linear polarization'. In: MNRAS 425.2, pp. 979–996. DOI: 10.1111/j.1365-2966.2012.21373.x. arXiv: 1205.5313 [astro-ph.IM].

Häring, N. and Rix, H.-W. (Apr. 2004). 'On the Black Hole Mass-Bulge Mass Relation'. In: ApJ 604.2, pp. L89–L92. DOI: 10.1086/383567. arXiv: astro-ph/0402376 [astro-ph].

Head, T., Kumar, M., Nahrstaedt, H., Louppe, G. and Shcherbatyi, I. (Oct. 2021). *scikit-optimize/scikit-optimize*. Version v0.9.0. DOI: 10.5281/zenodo.5565057. URL: https://doi.org/10.5281/zenodo.5565057.

Heckman, T. M. and Best, P. N. (Aug. 2014). 'The Coevolution of Galaxies and Supermassive Black Holes: Insights from Surveys of the Contemporary Universe'. In: ARA&A 52, pp. 589–660. DOI: 10.1146/annurev-astro-081913-035722. arXiv: 1403.4620 [astro-ph.GA].

Helfand, D. J., White, R. L. and Becker, R. H. (Mar. 2015). 'The Last of FIRST: The Final Catalog and Source Identifications'. In: ApJ 801.1, 26, p. 26. DOI: 10.1088/0004-637X/801/1/26. arXiv: 1501.01555 [astro-ph.GA].

Helou, G., Soifer, B. T. and Rowan-Robinson, M. (Nov. 1985). 'Thermal infrared and nonthermal radio : remarkable correlation in disks of galaxies.' In: ApJ 298, pp. L7–L11. DOI: 10.1086/184556.

Hernán-Caballero, A., Varela, J., López-Sanjuan, C., Muniesa, D., Civera, T., Chaves-Montero, J., Díaz-García, L. A., Laur, J., Hernández-Monteagudo, C., Abramo, R., Angulo, R., Cristóbal-Hornillos, D., González Delgado, R. M., Greisel, N., Orsi, A., Queiroz, C., Sobral, D., Tamm, A., Tempel, E., Vázquez-Ramió, H., Alcaniz, J., Benítez, N., Bonoli, S., Carneiro, S., Cenarro, J., Dupke, R., Ederoclite, A., Marín-Franch, A., Mendes de Oliveira, C., Moles, M., Sodré, L., Taylor, K., Cypriano, E. S. and Martínez-Solaeche, G. (Oct. 2021). 'The miniJPAS survey: Photometric redshift catalogue'. In: A&A 654, A101, A101. DOI: 10.1051/0004-6361/202141236. arXiv: 2108.03271 [astro-ph.GA].

Hickox, R. C. and Alexander, D. M. (Sept. 2018). 'Obscured Active Galactic Nuclei'. In: ARA&A 56, pp. 625–671. DOI: 10.1146/annurev-astro-081817-051803. arXiv: 1806.04680 [astro-ph.GA].

Hildebrand, R. H. (Sept. 1983). 'The determination of cloud masses and dust characteristics from submillimetre thermal emission.' In: QJRAS 24, pp. 267–282.

Hildebrandt, H., Arnouts, S., Capak, P., Moustakas, L. A., Wolf, C., Abdalla, F. B., Assef, R. J., Banerji, M., Benítez, N., Brammer, G. B., Budavári, T., Carliles, S., Coe, D., Dahlen, T., Feldmann, R., Gerdes, D., Gillis, B., Ilbert, O., Kotulla, R., Lahav, O., Li, I. H., Miralles, J. .-., Purger, N., Schmidt, S. and Singal, J. (Nov. 2010). 'PHAT: PHoto-z Accuracy Testing'. In: A&A 523, A31, A31. DOI: 10.1051/0004-6361/201014885. arXiv: 1008.0658 [astro-ph.CO].

Hill, G. J., Gebhardt, K., Komatsu, E., Drory, N., MacQueen, P. J., Adams, J., Blanc, G. A., Koehler, R., Rafal, M., Roth, M. M., Kelz, A., Gronwall, C., Ciardullo, R. and Schneider, D. P. (Oct. 2008). 'The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results'. In: *Panoramic Views of Galaxy Formation and Evolution*. Ed. by T. Kodama, T. Yamada and K. Aoki. Vol. 399. Astronomical Society of the Pacific Conference Series, p. 115. arXiv: 0806.0183 [astro-ph].

Hoaglin, D., Mosteller, F., Tukey, J., Emerson, J., Gasko, M., Goodall, C. et al. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley Series in Probability and Statistics: Probability and Statistics Section Series. John Wiley & Sons. ISBN: 9780471097778. URL: https://books.google.pt/books?id=FRnvAAAAMAAJ.

Hodge, J. A., Becker, R. H., White, R. L., Richards, G. T. and Zeimann, G. R. (July 2011). 'High-resolution Very Large Array Imaging of Sloan Digital Sky Survey Stripe 82 at 1.4 GHz'. In: AJ 142.1, 3, p. 3. DOI: 10.1088/0004-6256/142/1/3. arXiv: 1103.5749 [astro-ph.CO].

Humphrey, A., Bisigello, L., Cunha, P. A. C., Bolzonella, M., Fotopoulou, S., Caputi, K., Tortora, C., Zamorani, G., Papaderos, P., Vergani, D., Brinchmann, J., Moresco, M., Amara, A., Auricchio, N., Baldi, M., Bender, R., Bonino, D., Branchini, E., Brescia, M., Camera, S., Capobianco, V., Carbone, C., Carretero, J., Castander, F. J., Castellano, M., Cavuoti, S., Cimatti, A., Cledassou, R., Congedo, G., Conselice, C. J., Conversi, L., Copin, Y., Corcione, L., Courbin, F., Cropper, M., Da Silva, A., Degaudenzi, H., Douspis, M., Dubath, F., Duncan, C. A. J., Dupac, X., Dusini, S., Farrens, S., Ferriol, S., Frailis, M., Franceschi, E., Fumana, M., Gomez-Alvarez, P., Galeotta, S., Garilli, B., Gillard, W., Gillis, B., Giocoli, C., Grazian, A., Grupp, F., Guzzo, L., Haugan, S. V. H., Holmes, W., Hormuth, F., Jahnke, K., Kummel, M., Kermiche, S., Kiessling, A., Kilbinger, M., Kitching, T., Kohley, R., Kunz, M., Kurki-Suonio, H., Ligori, S., Lilje, P. B., Lloro, I., Maiorano, E., Mansutti, O., Marggraf, O., Markovic, K., Marulli, F., Massey, R., Maurogordato, S., McCracken, H. J., Medinaceli, E., Melchior, M., Meneghetti, M., Merlin, E., Meylan, G., Moscardini, L., Munari, E., Nakajima, R., Niemi, S. M., Nightingale, J., Padilla, C., Paltani, S., Pasian, F., Pedersen, K., Pettorino, V., Pires, S., Poncet, M., Popa, L. et al. (Sept. 2022). 'Euclid preparation: XXII. Selection of Quiescent Galaxies from Mock Photometry using Machine Learning'. In: *arXiv e-prints*, arXiv:2209.13074, arXiv:2209.13074. arXiv: 2209.13074 [astro-ph.IM].

Hunter, J. D. (2007). 'Matplotlib: A 2D graphics environment'. In: *Computing in Science & Engineering* 9.3, pp. 90–95. DOI: 10.1109/MCSE.2007.55.

İkiz, T., Peletier, R. F., Barthel, P. D. and Yeşilyaprak, C. (Aug. 2020). 'Infrared-detected AGNs in the local Universe'. In: A&A 640, A68, A68. DOI: 10.1051/0004-6361/201935971. arXiv: 2006.09476 [astro-ph.GA].

Ilbert, O., Arnouts, S., McCracken, H. J., Bolzonella, M., Bertin, E., Le Fèvre, O., Mellier, Y., Zamorani, G., Pellò, R., Iovino, A., Tresse, L., Le Brun, V., Bottini, D., Garilli, B., Maccagni, D., Picat, J. P., Scaramella, R., Scodeggio, M., Vettolani, G., Zanichelli, A., Adami, C., Bardelli, S., Cappi, A., Charlot, S., Ciliegi, P., Contini, T., Cucciati, O., Foucaud, S., Franzetti, P., Gavignaud, I., Guzzo, L., Marano, B., Marinoni, C., Mazure, A., Meneux, B., Merighi, R., Paltani, S., Pollo, A., Pozzetti, L., Radovich, M., Zucca, E., Bondi, M., Bongiorno, A., Busarello, G., de La Torre, S., Gregorini, L., Lamareille, F., Mathez, G., Merluzzi, P., Ripepi, V., Rizzo, D. and Vergani, D. (Oct. 2006). 'Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey'. In: A&A 457.3, pp. 841–856. DOI: 10.1051/0004-6361:20065138. arXiv: astro-ph/0603217 [astro-ph].

Ilbert, O., Capak, P., Salvato, M., Aussel, H., McCracken, H. J., Sanders, D. B., Scoville, N., Kartaltepe, J., Arnouts, S., Le Floc'h, E., Mobasher, B., Taniguchi, Y., Lamareille, F., Leauthaud, A., Sasaki, S., Thompson, D., Zamojski, M., Zamorani, G., Bardelli, S., Bolzonella, M., Bongiorno, A., Brusa, M., Caputi, K. I., Carollo, C. M., Contini, T., Cook, R., Coppa, G., Cucciati, O., de la Torre, S., de Ravel, L., Franzetti, P., Garilli, B., Hasinger, G., Iovino, A., Kampczyk, P., Kneib, J. .-., Knobel, C., Kovac, K., Le Borgne, J. F., Le Brun, V., Le Fèvre, O., Lilly, S., Looper, D., Maier, C., Mainieri, V., Mellier, Y., Mignoli, M., Murayama, T., Pellò, R., Peng, Y., Pérez-Montero, E., Renzini, A., Ricciardelli, E., Schiminovich, D., Scodeggio, M., Shioya, Y., Silverman, J., Surace, J., Tanaka, M., Tasca, L., Tresse, L., Vergani, D. and Zucca, E. (Jan. 2009). 'Cosmos Photometric Redshifts with 30-Bands for 2-deg^2'. In: ApJ 690.2, pp. 1236–1249. DOI: 10.1088/0004-637X/690/2/1236. arXiv: 0809.2101 [astro-ph].

Inayoshi, K., Visbal, E. and Haiman, Z. (Aug. 2020). 'The Assembly of the First Massive Black Holes'. In: ARA&A 58, pp. 27–97. DOI: 10.1146/annurev-astro-120419-014455. arXiv: 1911.05791 [astro-ph.GA].

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., Auza, N., Axelrod, T. S., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauer, A. E., Bauman, B. J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Bianco, F. B., Biswas, R., Blanc, G., Blazek, J., Blandford, R. D., Bloom, J. S., Bogart, J., Bond, T. W., Booth, M. T., Borgland, A. W., Borne, K., Bosch, J. F., Boutigny, D., Brackett, C. A., Bradshaw, A., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A. L., Carlin, J. L., Carlson, E. L., Chandrasekharan,

S., Charles-Emerson, G., Chesley, S., Cheu, E. C., Chiang, H.-F., Chiang, J., Chirino, C., Chow, D., Ciardi, D. R., Claver, C. F., Cohen-Tanugi, J., Cockrum, J. J., Coles, R., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daly, P. N., Daniel, S. F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S. W., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Economou, F., Eifler, T. et al. (Mar. 2019). 'LSST: From Science Drivers to Reference Design and Anticipated Data Products'. In: ApJ 873.2, 111, p. 111. DOI: 10.3847/1538-4357/ab042c. arXiv: 0805.2366 [astro-ph].

James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J. (2023). *An Introduction to Statistical Learning: with Applications in Python*. Springer Texts in Statistics. Springer International Publishing. ISBN: 9783031387470. URL: https://books.google.com/books?id=ygzJEAAAQBAJ.

Jarrett, T. H., Cluver, M. E., Magoulas, C., Bilicki, M., Alpaslan, M., Bland-Hawthorn, J., Brough, S., Brown, M. J. I., Croom, S., Driver, S., Holwerda, B. W., Hopkins, A. M., Loveday, J., Norberg, P., Peacock, J. A., Popescu, C. C., Sadler, E. M., Taylor, E. N., Tuffs, R. J. and Wang, L. (Feb. 2017). 'Galaxy and Mass Assembly (GAMA): Exploring the WISE Web in G12'. In: ApJ 836.2, 182, p. 182. DOI: 10.3847/1538-4357/836/2/182. arXiv: 1607.01190 [astro-ph.CO].

Jia, P., Jia, Q., Jiang, T. and Liu, J. (June 2023). 'Observation Strategy Optimization for Distributed Telescope Arrays with Deep Reinforcement Learning'. In: AJ 165.6, 233, p. 233. DOI: 10.3847/1538-3881/accceb.

Jiang, L., Fan, X., Bian, F., McGreer, I. D., Strauss, M. A., Annis, J., Buck, Z., Green, R., Hodge, J. A., Myers, A. D., Rafiee, A. and Richards, G. (July 2014). 'The Sloan Digital Sky Survey Stripe 82 Imaging Data: Depth-optimized Co-adds over 300 deg^2 in Five Filters'. In: ApJS 213.1, 12, p. 12. DOI: 10.1088/0067-0049/213/1/12. arXiv: 1405.7382 [astro-ph.GA].

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017). 'LightGBM: A Highly Efficient Gradient Boosting Decision Tree'. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett. Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Kennicutt Robert C., J., Hao, C.-N., Calzetti, D., Moustakas, J., Dale, D. A., Bendo, G., Engelbracht, C. W., Johnson, B. D. and Lee, J. C. (Oct. 2009). 'Dust-corrected Star Formation Rates of Galaxies. I. Combinations of Hα and Infrared Tracers'. In: ApJ 703.2, pp. 1672–1695. DOI: 10.1088/0004-637X/703/2/1672. arXiv: 0908.0203 [astro-ph.CO].

Kim, S. J., Lee, H. M., Matsuhara, H., Wada, T., Oyabu, S., Im, M., Jeon, Y., Kang, E., Ko, J., Lee, M. G., Takagi, T., Pearson, C., White, G. J., Jeong, W. .-., Serjeant, S., Nakagawa, T., Ohyama, Y., Goto, T., Takeuchi, T. T., Pollo, A., Solarz, A. and Pępiak, A. (Dec. 2012). 'The North Ecliptic Pole Wide survey of AKARI: a near- and mid-infrared source catalog'. In: A&A 548, A29, A29. DOI: 10.1051/0004-6361/201219105. arXiv: 1208.5008 [astro-ph.CO].

King, A. and Pounds, K. (Aug. 2015). 'Powerful Outflows and Feedback from Active Galactic Nuclei'. In: ARA&A 53, pp. 115–154. DOI: 10.1146/annurev-astro-082214-122316. arXiv: 1503.05206 [astro-ph.GA].

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S. and Willing, C. (2016). 'Jupyter Notebooks – a publishing format for reproducible computational workflows'. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Ed. by F. Loizides and B. Schmidt. IOS Press, pp. 87–90.

Kollmeier, J. A., Zasowski, G., Rix, H.-W., Johns, M., Anderson, S. F., Drory, N., Johnson, J. A., Pogge, R. W., Bird, J. C., Blanc, G. A., Brownstein, J. R., Crane, J. D., De Lee, N. M., Klaene, M. A., Kreckel, K., MacDonald, N., Merloni, A., Ness, M. K., O'Brien, T., Sanchez-Gallego, J. R., Sayres, C. C., Shen, Y., Thakar, A. R., Tkachenko, A., Aerts, C., Blanton, M. R., Eisenstein, D. J., Holtzman, J. A., Maoz, D., Nandra, K., Rockosi, C., Weinberg, D. H., Bovy, J., Casey, A. R., Chaname, J., Clerc, N., Conroy, C., Eracleous, M., Gänsicke, B. T., Hekker, S., Horne, K., Kauffmann, J., McQuinn, K. B. W., Pellegrini, E. W., Schinnerer, E., Schlafly, E. F., Schwope, A. D., Seibert, M., Teske, J. K. and van Saders, J. L. (Nov. 2017). 'SDSS-V: Pioneering Panoptic Spectroscopy'. In: *arXiv e-prints*, arXiv:1711.03234, arXiv:1711.03234. DOI: 10.48550/arXiv.1711.03234. arXiv: 1711.03234 [astro-ph.GA].

Kormendy, J. and Ho, L. C. (Aug. 2013). 'Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies'. In: ARA&A 51.1, pp. 511–653. DOI: 10.1146/annurev-astro-082708-101811. arXiv: 1304.7762 [astro-ph.CO].

Koshida, S., Minezaki, T., Yoshii, Y., Kobayashi, Y., Sakata, Y., Sugawara, S., Enya, K., Suganuma, M., Tomita, H., Aoki, T. and Peterson, B. A. (June 2014). 'Reverberation Measurements of the Inner Radius of the Dust Torus in 17 Seyfert Galaxies'. In: ApJ 788.2, 159, p. 159. DOI: 10.1088/0004-637X/788/2/159. arXiv: 1406.2078 [astro-ph.GA].

Koshida, S., Yoshii, Y., Kobayashi, Y., Minezaki, T., Sakata, Y., Sugawara, S., Enya, K., Suganuma, M., Tomita, H., Aoki, T. and Peterson, B. A. (Aug. 2009). 'Variation of Inner Radius of Dust Torus in NGC4151'. In: ApJ 700.2, pp. L109–L113. DOI: 10.1088/0004-637X/700/2/L109. arXiv: 0907.0573 [astro-ph.CO].

Kovács, O. E., Bogdán, Á., Smith, R. K., Kraft, R. P. and Forman, W. R. (Feb. 2019). 'Detection of the Missing Baryons toward the Sightline of H1821+643'. In: ApJ 872.1, 83, p. 83. DOI: 10.3847/1538-4357/aaef78. arXiv: 1812.04625 [astro-ph.CO].

Kull, M., Filho, T. M. S. and Flach, P. (2017a). 'Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration'. In: *Electronic Journal of Statistics* 11.2, pp. 5052–5080. DOI: 10.1214/17-EJS1338SI. URL: https://doi.org/10.1214/17-EJS1338SI.

Kull, M., Filho, T. S. and Flach, P. (Apr. 2017b). 'Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers'. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Ed. by A. Singh and J. Zhu. Vol. 54. Proceedings of Machine Learning Research. PMLR, pp. 623–631. URL: https://proceedings.mlr.press/v54/kull17a.html.

Kurtz, M. J. and Mink, D. J. (Aug. 1998). 'RVSAO 2.0: Digital Redshifts and Radial Velocities'. In: PASP 110.750, pp. 934–977. DOI: 10.1086/316207. arXiv: astro-ph/9803252 [astro-ph].

Kuźmicz, A. and Jamrozy, M. (Mar. 2021). 'Giant Radio Quasars: Sample and Basic Properties'. In: ApJS 253.1, 25, p. 25. DOI: 10.3847/1538-4365/abd483. arXiv: 2012.08857 [astro-ph.GA].

Lacy, M., Baum, S. A., Chandler, C. J., Chatterjee, S., Clarke, T. E., Deustua, S., English, J., Farnes, J., Gaensler, B. M., Gugliucci, N., Hallinan, G., Kent, B. R., Kimball, A., Law, C. J., Lazio, T. J. W., Marvil, J., Mao, S. A., Medlin, D., Mooley, K., Murphy, E. J., Myers, S., Osten, R., Richards, G. T., Rosolowsky, E., Rudnick, L., Schinzel, F., Sivakoff, G. R., Sjouwerman, L. O., Taylor, R., White, R. L., Wrobel, J., Andernach, H., Beasley, A. J., Berger, E., Bhatnager, S., Birkinshaw, M., Bower, G. C., Brandt, W. N., Brown, S., Burke-Spolaor, S., Butler, B. J., Comerford, J., Demorest, P. B., Fu, H., Giacintucci, S., Golap, K., Güth, T., Hales, C. A., Hiriart, R., Hodge, J., Horesh, A., Ivezić, Ž., Jarvis, M. J., Kamble, A., Kassim, N., Liu, X., Loinard, L., Lyons, D. K., Masters, J., Mezcua, M., Moellenbrock, G. A., Mroczkowski, T., Nyland, K., O'Dea, C. P., O'Sullivan, S. P., Peters, W. M., Radford, K., Rao, U., Robnett, J., Salcido, J., Shen, Y., Sobotka, A., Witz, S., Vaccari, M., van Weeren, R. J., Vargas, A., Williams, P. K. G. and Yoon, I. (Mar. 2020). 'The Karl G. Jansky Very Large Array Sky Survey (VLASS). Science Case and Survey Design'. In: PASP 132.1009, 035001, p. 035001. DOI: 10.1088/1538-3873/ab63eb. arXiv: 1907.01981 [astro-ph.IM].

Lacy, M., Ridgway, S. E., Gates, E. L., Nielsen, D. M., Petric, A. O., Sajina, A., Urrutia, T., Cox Drews, S., Harrison, C., Seymour, N. and Storrie-Lombardi, L. J. (Oct. 2013). 'The Spitzer Mid-infrared Active Galactic Nucleus Survey. I. Optical and Near-infrared Spectroscopy of Obscured Candidates and Normal Active Galactic Nuclei Selected in the Mid-infrared'. In: ApJS 208.2, 24, p. 24. DOI: 10.1088/0067-0049/208/2/24. arXiv: 1308.4190 [astro-ph.CO].

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., Wilson, G. and Yan, L. (Sept. 2004). 'Obscured and Unobscured Active Galactic Nuclei in the Spitzer Space Telescope First Look Survey'. In: ApJS 154.1, pp. 166–169. DOI: 10.1086/422816. arXiv: astro-ph/0405604 [astro-ph].

Lacy, M., Surace, J. A., Farrah, D., Nyland, K., Afonso, J., Brandt, W. N., Clements, D. L., Lagos, C. D. P., Maraston, C., Pforr, J., Sajina, A., Sako, M., Vaccari, M., Wilson, G., Ballantyne, D. R., Barkhouse, W. A., Brunner, R., Cane, R., Clarke, T. E., Cooper, M., Cooray, A., Covone, G., D'Andrea, C., Evrard, A. E., Ferguson, H. C., Frieman, J., Gonzalez-Perez, V., Gupta, R., Hatziminaoglou, E., Huang, J., Jagannathan, P., Jarvis, M. J., Jones, K. M., Kimball, A., Lidman, C., Lubin, L., Marchetti, L., Martini, P., McMahon, R. G., Mei, S., Messias, H., Murphy, E. J., Newman, J. A., Nichol, R., Norris, R. P., Oliver, S., Perez-Fournon, I., Peters, W. M., Pierre, M., Polisensky, E., Richards, G. T., Ridgway, S. E., Röttgering, H. J. A., Seymour, N., Shirley, R., Somerville, R., Strauss, M. A., Suntzeff, N., Thorman, P. A., van Kampen, E., Verma, A., Wechsler, R. and Wood-Vasey, W. M. (Feb. 2021). 'A Spitzer survey of Deep Drilling Fields to be targeted by the Vera C. Rubin Observatory Legacy Survey of Space and Time'. In: MNRAS 501.1, pp. 892–910. DOI: 10.1093/mnras/staa3714. arXiv: 2011.15030 [astro-ph.GA].

Lacy, M. and Sajina, A. (Apr. 2020). 'Active galactic nuclei as seen by the Spitzer Space Telescope'. In: *Nature Astronomy* 4, pp. 352–363. DOI: 10.1038/s41550-020-1071-x. arXiv: 2008.05424 [astro-ph.GA].

Lal, D. V. (July 2021). 'The Discovery of a Remnant Radio Galaxy in A2065 Using GMRT'. In: ApJ 915.2, 126, p. 126. DOI: 10.3847/1538-4357/ac042d.

Langeroodi, D. and Hjorth, J. (Apr. 2023). 'PAH Emission from Star-forming Galaxies in JWST Mid-infrared Imaging of the Lensing Cluster SMACS J0723.3-7327'. In: ApJ 946.2, L40, p. L40. DOI: 10.3847/2041-8213/acc1e0. arXiv: 2210.06482 [astro-ph.GA].

Latimer, C. J., Reines, A. E., Hainline, K. N., Greene, J. E. and Stern, D. (June 2021). 'A Chandra and HST View of WISE-selected AGN Candidates in Dwarf Galaxies'. In: ApJ 914.2, 133, p. 133. DOI: 10.3847/1538-4357/abfe0c. arXiv: 2105.05876 [astro-ph.GA].

Le Fèvre, O., Tasca, L. A. M., Cassata, P., Garilli, B., Le Brun, V., Maccagni, D., Pentericci, L., Thomas, R., Vanzella, E., Zamorani, G., Zucca, E., Amorin, R., Bardelli, S., Capak, P., Cassarà, L., Castellano, M., Cimatti, A., Cuby, J. G., Cucciati, O., de la Torre, S., Durkalec, A., Fontana, A., Giavalisco, M., Grazian, A., Hathi, N. P., Ilbert, O., Lemaux, B. C., Moreau, C., Paltani, S., Ribeiro, B., Salvato, M., Schaerer, D., Scodeggio, M., Sommariva, V., Talia, M., Taniguchi, Y., Tresse, L., Vergani, D., Wang, P. W., Charlot, S., Contini, T., Fotopoulou, S., López-Sanjuan, C., Mellier, Y. and Scoville, N. (Apr. 2015). 'The VIMOS Ultra-Deep Survey: ~10 000 galaxies with spectroscopic redshifts to study galaxy assembly at early epochs $2 < z \simeq 6$'. In: A&A 576, A79, A79. DOI: 10.1051/0004-6361/201423829. arXiv: 1403.3938 [astro-ph.CO].

Lee, H. M., Kim, S. J., Im, M., Matsuhara, H., Oyabu, S., Wada, T., Nakagawa, T., Ko, J., Shim, H. J., Lee, M. G., Hwang, N., Takagi, T. and Pearson, C. (Feb. 2009). 'North Ecliptic Pole Wide Field Survey of AKARI: Survey Strategy and Data Characteristics'. In: PASJ 61, p. 375. DOI: 10.1093/pasj/61.2.375. arXiv: 0901.3256 [astro-ph.GA].

Lehmer, B. D., Brandt, W. N., Alexander, D. M., Bauer, F. E., Schneider, D. P., Tozzi, P., Bergeron, J., Garmire, G. P., Giacconi, R., Gilli, R., Hasinger, G., Hornschemeier, A. E., Koekemoer, A. M., Mainieri, V., Miyaji, T., Nonino, M., Rosati, P., Silverman, J. D., Szokoly, G. and Vignali, C. (Nov. 2005). 'The Extended Chandra Deep Field-South Survey: Chandra Point-Source Catalogs'. In: ApJS 161.1, pp. 21–40. DOI: 10.1086/444590. arXiv: astro-ph/0506607 [astro-ph].

Lichtenstein, S., Fischhoff, B. and Phillips, L. D. (1982). 'Calibration of probabilities: The state of the art to 1980'. In: *Judgment under Uncertainty: Heuristics and Biases*. Ed. by D. Kahneman, P. Slovic and A. Tversky. Cambridge University Press, pp. 306–334. DOI: 10.1017/CBO9780511809477.023.

Lima, E. V. R., Sodré, L., Bom, C. R., Teixeira, G. S. M., Nakazono, L., Buzzo, M. L., Queiroz, C., Herpich, F. R., Castellon, J. L. N., Dantas, M. L. L., Dors, O. L., Souza, R. C. T. d., Akras, S., Jiménez-Teja, Y., Kanaan, A., Ribeiro, T. and Schoennell, W. (Jan. 2022). 'Photometric redshifts for the S-PLUS Survey: Is machine learning up to the task?' In: *Astronomy and Computing* 38, 100510, p. 100510. DOI: 10.1016/j.ascom.2021.100510. arXiv: 2110.13901 [astro-ph.GA].

Lira, P., Arévalo, P., Uttley, P., McHardy, I. and Breedt, E. (Aug. 2011). 'Optical and near-IR long-term monitoring of NGC 3783 and MR 2251-178: evidence for variable near-IR emission from thin accretion discs'. In: MNRAS 415.2, pp. 1290–1303. DOI: 10.1111/j.1365-2966.2011.18774.x. arXiv: 1107.0099 [astro-ph.CO].

Lira, P., Arévalo, P., Uttley, P., McHardy, I. M. M. and Videla, L. (Nov. 2015). 'Long-term monitoring of the archetype Seyfert galaxy MCG-6-30-15: X-ray, optical and near-IR variability of the corona, disc and torus'. In: MNRAS 454.1, pp. 368–379. DOI: 10.1093/mnras/stv1945. arXiv: 1508.05928 [astro-ph.GA].

Lisenfeld, U. and Völk, H. J. (Feb. 2000). 'On the radio spectral index of galaxies'. In: A&A 354, pp. 423–430. arXiv: astro-ph/9912232 [astro-ph].

Liske, J., Baldry, I. K., Driver, S. P., Tuffs, R. J., Alpaslan, M., Andrae, E., Brough, S., Cluver, M. E., Grootes, M. W., Gunawardhana, M. L. P., Kelvin, L. S., Loveday, J., Robotham, A. S. G., Taylor, E. N., Bamford, S. P., Bland-Hawthorn, J., Brown, M. J. I., Drinkwater, M. J., Hopkins, A. M., Meyer, M. J., Norberg, P., Peacock, J. A., Agius, N. K., Andrews, S. K., Bauer, A. E., Ching, J. H. Y., Colless, M., Conselice, C. J., Croom, S. M., Davies, L. J. M., De Propris, R., Dunne, L., Eardley, E. M., Ellis, S., Foster, C., Frenk, C. S., Häußler, B., Holwerda, B. W., Howlett, C., Ibarra, H., Jarvis, M. J., Jones, D. H., Kafle, P. R., Lacey, C. G., Lange, R., Lara-López, M. A., López-Sánchez, Á. R., Maddox, S., Madore, B. F., McNaught-Roberts, T., Moffett, A. J., Nichol, R. C., Owers, M. S., Palamara, D., Penny, S. J., Phillipps, S., Pimbblet, K. A., Popescu, C. C., Prescott, M., Proctor, R., Sadler, E. M., Sansom, A. E., Seibert, M., Sharp, R., Sutherland, W., Vázquez-Mata, J. A., van Kampen, E., Wilkins, S. M., Williams, R. and Wright, A. H. (Sept. 2015). 'Galaxy And Mass Assembly (GAMA): end of survey report and data release 2'. In: MNRAS 452.2, pp. 2087–2126. DOI: 10.1093/mnras/stv1436. arXiv: 1506.08222 [astro-ph.GA].

Lochner, M. and Bassett, B. A. (July 2021). 'ASTRONOMALY: Personalised active anomaly detection in astronomical data'. In: *Astronomy and Computing* 36, 100481, p. 100481. DOI: 10.1016/j.ascom.2021.100481. arXiv: 2010.11202 [astro-ph.IM].

Loh, E. D. and Spillar, E. J. (Apr. 1986). 'Photometric Redshifts of Galaxies'. In: ApJ 303, p. 154. DOI: 10.1086/164062.

Louppe, G., Wehenkel, L., Sutera, A. and Geurts, P. (2013). 'Understanding variable importances in forests of randomized trees'. In: *Advances in Neural Information Processing Systems*. Ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger. Vol. 26. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf.

LSST Science Collaboration, Abell, P. A., Allison, J., Anderson, S. F., Andrew, J. R., Angel, J. R. P., Armus, L., Arnett, D., Asztalos, S. J., Axelrod, T. S., Bailey, S., Ballantyne, D. R., Bankert, J. R., Barkhouse, W. A., Barr, J. D., Barrientos, L. F., Barth, A. J., Bartlett, J. G., Becker, A. C., Becla, J., Beers, T. C., Bernstein, J. P., Biswas, R., Blanton, M. R., Bloom, J. S., Bochanski, J. J., Boeshaar, P., Borne, K. D., Bradac, M., Brandt, W. N., Bridge, C. R., Brown, M. E., Brunner, R. J., Bullock, J. S., Burgasser, A. J., Burge, J. H., Burke, D. L., Cargile, P. A., Chandrasekharan, S., Chartas, G., Chesley, S. R., Chu, Y.-H., Cinabro, D., Claire, M. W., Claver, C. F., Clowe, D., Connolly, A. J., Cook, K. H., Cooke, J., Cooray, A., Covey, K. R., Culliton, C. S., de Jong, R., de Vries, W. H., Debattista, V. P., Delgado, F., Dell'Antonio, I. P., Dhital, S., Di Stefano, R., Dickinson, M., Dilday, B., Djorgovski, S. G., Dobler, G., Donalek, C., Dubois-Felsmann, G., Durech, J., Eliasdottir, A., Eracleous, M., Eyer, L., Falco, E. E., Fan, X., Fassnacht, C. D., Ferguson, H. C., Fernandez, Y. R., Fields, B. D., Finkbeiner, D., Figueroa, E. E., Fox, D. B., Francke, H., Frank, J. S., Frieman, J., Fromenteau, S., Furqan, M., Galaz, G., Gal-Yam, A., Garnavich, P., Gawiser, E., Geary, J., Gee, P., Gibson, R. R., Gilmore, K., Grace, E. A., Green, R. F., Gressler, W. J., Grillmair, C. J., Habib, S., Haggerty, J. S. et al. (Dec. 2009). 'LSST Science Book, Version 2.0'. In: *arXiv e-prints*, arXiv:0912.0201, arXiv:0912.0201. DOI: 10.48550/arXiv.0912.0201. arXiv: 0912.0201 [astro-ph.IM].

Luken, K., Norris, R., Park, L., Wang, X. and Filipović, M. (2022). 'Estimating galaxy redshift in radio-selected datasets using machine learning'. In: *Astronomy and Computing* 39, p. 100557. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2022.100557. arXiv: 2202.13504 [astro-ph.IM]. URL: https://www.sciencedirect.com/science/article/pii/S2213133722000075.

Luken, K. J., Norris, R. P. and Park, L. A. F. (Oct. 2019). 'Preliminary Results of Using k-Nearest Neighbors Regression to Estimate the Redshift of Radio-selected Data Sets'. In: PASP 131.1004, p. 108003. DOI: 10.1088/1538-3873/aaea17. arXiv: 1810.10714 [astro-ph.GA].

Lukic, V., Brüggen, M., Mingo, B., Croston, J. H., Kasieczka, G. and Best, P. N. (Aug. 2019). 'Morphological classification of radio galaxies: capsule networks versus convolutional neural networks'. In: MNRAS 487.2, pp. 1729–1744. DOI: 10.1093/mnras/stz1289. arXiv: 1905.03274 [astro-ph.IM].

Lundberg, S. M. and Lee, S.-I. (2017). 'A Unified Approach to Interpreting Model Predictions'. In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett. Curran Associates, Inc., pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.-I. (2020). 'From local explanations to global understanding with explainable AI for trees'. In: *Nature Machine Intelligence* 2.1, pp. 2522–5839.

Ma, S. and Tourani, R. (Aug. 2020). 'Predictive and Causal Implications of using Shapley Value for Model Interpretation'. In: *arXiv e-prints*, arXiv:2008.05052, arXiv:2008.05052. DOI: 10.48550/arXiv.2008.05052. arXiv: 2008.05052 [cs.LG].

Ma, Z., Xu, H., Zhu, J., Hu, D., Li, W., Shan, C., Zhu, Z., Gu, L., Li, J., Liu, C. and Wu, X. (Feb. 2019). 'A Machine Learning Based Morphological Classification of 14,245 Radio AGNs Selected from the Best-Heckman Sample'. In: ApJS 240.2, 34, p. 34. DOI: 10.3847/1538-4365/aaf9a2. arXiv: 1812.07190 [astro-ph.GA].

Machado, D. P., Leonard, A., Starck, J. .-., Abdalla, F. B. and Jouvel, S. (Dec. 2013). 'Darth Fader: Using wavelets to obtain accurate redshifts of spectra at very low signal-to-noise'. In: A&A 560, A83, A83. DOI: 10.1051/0004-6361/201219857. arXiv: 1309.3579 [astro-ph.CO].

Madau, P. and Dickinson, M. (Aug. 2014). 'Cosmic Star-Formation History'. In: ARA&A 52, pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007 [astro-ph.CO].

Magliocchetti, M. (Dec. 2022). 'Hosts and environments: a (large-scale) radio history of AGN and star-forming galaxies'. In: A&A Rev. 30.1, 6, p. 6. DOI: 10.1007/s00159-022-00142-1. arXiv: 2206.15286 [astro-ph.CO].

Magorrian, J., Tremaine, S., Richstone, D., Bender, R., Bower, G., Dressler, A., Faber, S. M., Gebhardt, K., Green, R., Grillmair, C., Kormendy, J. and Lauer, T. (June 1998). 'The Demography of Massive Dark Objects in Galaxy Centers'. In: AJ 115.6, pp. 2285–2305. DOI: 10.1086/300353. arXiv: astro-ph/9708072 [astro-ph].

Mainzer, A., Bauer, J., Cutri, R. M., Grav, T., Masiero, J., Beck, R., Clarkson, P., Conrow, T., Dailey, J., Eisenhardt, P., Fabinsky, B., Fajardo-Acosta, S., Fowler, J., Gelino, C., Grillmair, C., Heinrichsen, I., Kendall, M., Kirkpatrick, J. D., Liu, F., Masci, F., McCallon, H., Nugent, C. R., Papin, M., Rice, E., Royer, D., Ryan, T., Sevilla, P., Sonnett, S., Stevenson, R., Thompson, D. B., Wheelock, S., Wiemer, D., Wittman, M., Wright, E. and Yan, L. (Sept. 2014). 'Initial Performance of the NEOWISE Reactivation Mission'. In: ApJ 792.1, 30, p. 30. DOI: 10.1088/0004-637X/792/1/30. arXiv: 1406.6025 [astro-ph.EP].

Mainzer, A., Bauer, J., Grav, T., Masiero, J., Cutri, R. M., Dailey, J., Eisenhardt, P., McMillan, R. S., Wright, E., Walker, R., Jedicke, R., Spahr, T., Tholen, D., Alles, R., Beck, R., Brandenburg, H., Conrow, T., Evans, T., Fowler, J., Jarrett, T., Marsh, K., Masci, F., McCallon, H.,

Wheelock, S., Wittman, M., Wyatt, P., DeBaun, E., Elliott, G., Elsbury, D., Gautier T., I., Gomillion, S., Leisawitz, D., Maleszewski, C., Micheli, M. and Wilkins, A. (Apr. 2011). 'Preliminary Results from NEOWISE: An Enhancement to the Wide-field Infrared Survey Explorer for Solar System Science'. In: ApJ 731.1, 53, p. 53. DOI: 10.1088/0004-637X/731/1/53. arXiv: 1102.1996 [astro-ph.EP].

Maitra, C., Haberl, F., Ivanov, V. D., Cioni, M.-R. L. and van Loon, J. T. (Feb. 2019). 'Identification of AGN in the XMM-Newton X-ray survey of the SMC'. In: A&A 622, A29, A29. DOI: 10.1051/0004-6361/201833663. arXiv: 1810.12032 [astro-ph.HE].

Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., Kirkpatrick, J. D., Meisner, A. M., Schlafly, E. F., Stanford, S. A., Garcia, N., Caselden, D., Cushing, M. C., Cutri, R. M., Faherty, J. K., Gelino, C. R., Gonzalez, A. H., Jarrett, T. H., Koontz, R., Mainzer, A., Marchese, E. J., Mobasher, B., Schlegel, D. J., Stern, D., Teplitz, H. I. and Wright, E. L. (Mar. 2021). 'The CatWISE2020 Catalog'. In: ApJS 253.1, 8, p. 8. DOI: 10.3847/1538-4365/abd805. arXiv: 2012.13084 [astro-ph.IM].

Mateos, S., Alonso-Herrero, A., Carrera, F. J., Blain, A., Watson, M. G., Barcons, X., Braito, V., Severgnini, P., Donley, J. L. and Stern, D. (Nov. 2012). 'Using the Bright Ultrahard XMM-Newton survey to define an IR selection of luminous AGN based on WISE colours'. In: MNRAS 426.4, pp. 3271–3281. DOI: 10.1111/j.1365-2966.2012.21843.x. arXiv: 1208.2530 [astro-ph.CO].

Matthews, B. (1975). 'Comparison of the predicted and observed secondary structure of T4 phage lysozyme'. In: *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405.2, pp. 442–451. ISSN: 0005-2795. DOI: https://doi.org/10.1016/0005-2795(75)90109-9.

McConnell, N. J. and Ma, C.-P. (Feb. 2013). 'Revisiting the Scaling Relations of Black Hole Masses and Host Galaxy Properties'. In: ApJ 764.2, 184, p. 184. DOI: 10.1088/0004-637X/764/2/184. arXiv: 1211.2816 [astro-ph.CO].

McGreer, I. D., Becker, R. H., Helfand, D. J. and White, R. L. (Nov. 2006). 'Discovery of a z = 6.1 Radio-Loud Quasar in the NOAO Deep Wide Field Survey'. In: ApJ 652.1, pp. 157–162. DOI: 10.1086/507767. arXiv: astro-ph/0607278 [astro-ph].

McHardy, I. M., Connolly, S. D., Peterson, B. M., Bieryla, A., Chand, H., Elvis, M. S., Emmanoulopoulos, D., Falco, E., Gandhi, P., Kaspi, S., Latham, D., Lira, P., McCully, C., Netzer, H. and Uemura, M. (May 2016). 'The origin of UV-optical variability in AGN and test of disc models: XMM-Newton and ground-based observations of NGC 4395'. In: *Astronomische Nachrichten* 337.4-5, p. 500. DOI: 10.1002/asna.201612337. arXiv: 1601.00215 [astro-ph.GA].

McKinney, W. (2010). 'Data Structures for Statistical Computing in Python'. In: *Proceedings of the 9th Python in Science Conference*. Ed. by S. van der Walt and J. Millman, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

Menzel, M. .-., Merloni, A., Georgakakis, A., Salvato, M., Aubourg, E., Brandt, W. N., Brusa, M., Buchner, J., Dwelly, T., Nandra, K., Pâris, I., Petitjean, P. and Schwope, A. (Mar. 2016). 'A spectroscopic survey of X-ray-selected AGNs in the northern XMM-XXL field'. In: MNRAS 457.1, pp. 110–132. DOI: 10.1093/mnras/stv2749. arXiv: 1511.07870 [astro-ph.GA].

Merlin, E., Castellano, M., Santini, P., Cipolletta, G., Boutsia, K., Schreiber, C., Buitrago, F., Fontana, A., Elbaz, D., Dunlop, J., Grazian, A., McLure, R., McLeod, D., Nonino, M., Milvang-Jensen, B., Derriere, S., Hathi, N. P., Pentericci, L., Fortuni, F. and Calabrò, A. (May 2021). 'The ASTRODEEP-GS43 catalogue: New photometry and redshifts for the CANDELS GOODS-South field'. In: A&A 649, A22, A22. DOI: 10.1051/0004-6361/202140310. arXiv: 2103.09246 [astro-ph.GA].

Messias, H., Afonso, J., Salvato, M., Mobasher, B. and Hopkins, A. M. (Aug. 2012). 'A New Infrared Color Criterion for the Selection of 0 < z < 7 AGNs: Application to Deep Fields and Implications for JWST Surveys'. In: ApJ 754.2, 120, p. 120. DOI: 10.1088/0004-637X/754/2/120. arXiv: 1205.4764 [astro-ph.CO].

Miley, G. and De Breuck, C. (Feb. 2008). 'Distant radio galaxies and their environments'. In: A&A Rev. 15.2, pp. 67–144. DOI: 10.1007/s00159-007-0008-z. arXiv: 0802.2770 [astro-ph].

Mingo, B., Watson, M. G., Rosen, S. R., Hardcastle, M. J., Ruiz, A., Blain, A., Carrera, F. J., Mateos, S., Pineau, F. .-. and Stewart, G. C. (Nov. 2016). 'The MIXR sample: AGN activity versus star formation across the cross-correlation of WISE, 3XMM, and FIRST/NVSS'. In: MNRAS 462.3, pp. 2631–2667. DOI: 10.1093/mnras/stw1826. arXiv: 1607.06471 [astro-ph.GA].

Mohan, N. and Rafferty, D. (Feb. 2015). *PyBDSF: Python Blob Detection and Source Finder*. Astrophysics Source Code Library, record ascl:1502.007. ascl: 1502.007.

Mostert, R. I. J., Duncan, K. J., Röttgering, H. J. A., Polsterer, K. L., Best, P. N., Brienza, M., Brüggen, M., Hardcastle, M. J., Jurlin, N., Mingo, B., Morganti, R., Shimwell, T., Smith, D. and Williams, W. L. (Jan. 2021). 'Unveiling the rarest morphologies of the LOFAR Two-metre Sky Survey radio source population with self-organised maps'. In: A&A 645, A89, A89. DOI: 10.1051/0004-6361/202038500. arXiv: 2011.06001 [astro-ph.IM].

Naidoo, K., Johnston, H., Joachimi, B., van den Busch, J. L., Hildebrandt, H., Ilbert, O., Lahav, O., Aghanim, N., Altieri, B., Amara, A., Baldi, M., Bender, R., Bodendorf, C., Branchini, E., Brescia, M., Brinchmann, J., Camera, S., Capobianco, V., Carbone, C., Carretero, J., Castander, F. J., Castellano, M., Cavuoti, S., Cimatti, A., Cledassou, R., Congedo, G., Conselice, C. J., Conversi, L., Copin, Y.,

Corcione, L., Courbin, F., Cropper, M., Da Silva, A., Degaudenzi, H., Dinis, J., Dubath, F., Dupac, X., Dusini, S., Farrens, S., Ferriol, S., Fosalba, P., Frailis, M., Franceschi, E., Franzetti, P., Fumana, M., Galeotta, S., Garilli, B., Gillard, W., Gillis, B., Giocoli, C., Grazian, A., Grupp, F., Haugan, S. V. H., Holmes, W., Hormuth, F., Hornstrup, A., Jahnke, K., Kümmel, M., Kiessling, A., Kilbinger, M., Kitching, T., Kohley, R., Kurki-Suonio, H., Ligori, S., Lilje, P. B., Lloro, I., Maiorano, E., Mansutti, O., Marggraf, O., Markovic, K., Marulli, F., Massey, R., Maurogordato, S., Meneghetti, M., Merlin, E., Meylan, G., Moresco, M., Moscardini, L., Munari, E., Nakajima, R., Niemi, S. M., Padilla, C., Paltani, S., Pasian, F., Pedersen, K., Percival, W. J., Pettorino, V., Pires, S., Polenta, G., Poncet, M., Popa, L., Pozzetti, L., Raison, F., Rebolo, R., Renzi, A., Rhodes, J., Riccio, G. et al. (Feb. 2023). 'Euclid: Calibrating photometric redshifts with spectroscopic cross-correlations'. In: A&A 670, A149, A149. DOI: 10.1051/0004-6361/202244795. arXiv: 2208.10503 [astro-ph.CO].

Nakoneczny, S. J., Bilicki, M., Pollo, A., Asgari, M., Dvornik, A., Erben, T., Giblin, B., Heymans, C., Hildebrandt, H., Kannawadi, A., Kuijken, K., Napolitano, N. R. and Valentijn, E. (May 2021). 'Photometric selection and redshifts for quasars in the Kilo-Degree Survey Data Release 4'. In: A&A 649, A81, A81. DOI: 10.1051/0004-6361/202039684. arXiv: 2010.13857 [astro-ph.CO].

Newman, J. A., Abate, A., Abdalla, F. B., Allam, S., Allen, S. W., Ansari, R., Bailey, S., Barkhouse, W. A., Beers, T. C., Blanton, M. R., Brodwin, M., Brownstein, J. R., Brunner, R. J., Carrasco Kind, M., Cervantes-Cota, J. L., Cheu, E., Chisari, N. E., Colless, M., Comparat, J., Coupon, J., Cunha, C. E., de la Macorra, A., Dell'Antonio, I. P., Frye, B. L., Gawiser, E. J., Gehrels, N., Grady, K., Hagen, A., Hall, P. B., Hearin, A. P., Hildebrandt, H., Hirata, C. M., Ho, S., Honscheid, K., Huterer, D., Ivezić, Ž., Kneib, J.-P., Kruk, J. W., Lahav, O., Mandelbaum, R., Marshall, J. L., Matthews, D. J., Ménard, B., Miquel, R., Moniez, M., Moos, H. W., Moustakas, J., Myers, A. D., Papovich, C., Peacock, J. A., Park, C., Rahman, M., Rhodes, J., Ricol, J.-S., Sadeh, I., Slozar, A., Schmidt, S. J., Stern, D. K., Anthony Tyson, J., von der Linden, A., Wechsler, R. H., Wood-Vasey, W. M. and Zentner, A. R. (Mar. 2015). 'Spectroscopic needs for imaging dark energy experiments'. In: *Astroparticle Physics* 63, pp. 81–100. DOI: 10.1016/j.astropartphys.2014.06.007. arXiv: 1309.5384 [astro-ph.CO].

Newman, J. A. and Gruen, D. (Aug. 2022). 'Photometric Redshifts for Next-Generation Surveys'. In: ARA&A 60, pp. 363–414. DOI: 10.1146/annurev-astro-032122-014611. arXiv: 2206.13633 [astro-ph.CO].

Nicastro, F., Kaastra, J., Krongold, Y., Borgani, S., Branchini, E., Cen, R., Dadina, M., Danforth, C. W., Elvis, M., Fiore, F., Gupta, A., Mathur, S., Mayya, D., Paerels, F., Piro, L., Rosa-Gonzalez, D., Schaye, J., Shull, J. M., Torres-Zafra, J., Wijers, N. and Zappacosta, L. (June 2018). 'Observations of the missing baryons in the warm-hot intergalactic medium'. In: Nature 558.7710, pp. 406–409. DOI: 10.1038/s41586-018-0204-1. arXiv: 1806.08395 [astro-ph.GA].

Nicastro, F., Krongold, Y., Mathur, S. and Elvis, M. (Mar. 2017). 'A decade of warm hot intergalactic medium searches: Where do we stand and where do we go?' In: *Astronomische Nachrichten* 338.281, pp. 281–286. DOI: 10.1002/asna.201713343.

Norris, R. P., Hopkins, A. M., Afonso, J., Brown, S., Condon, J. J., Dunne, L., Feain, I., Hollow, R., Jarvis, M., Johnston-Hollitt, M., Lenc, E., Middelberg, E., Padovani, P., Prandoni, I., Rudnick, L., Seymour, N., Umana, G., Andernach, H., Alexander, D. M., Appleton, P. N., Bacon, D., Banfield, J., Becker, W., Brown, M. J. I., Ciliegi, P., Jackson, C., Eales, S., Edge, A. C., Gaensler, B. M., Giovannini, G., Hales, C. A., Hancock, P., Huynh, M. T., Ibar, E., Ivison, R. J., Kennicutt, R., Kimball, A. E., Koekemoer, A. M., Koribalski, B. S., López-Sánchez, Á. R., Mao, M. Y., Murphy, T., Messias, H., Pimbblet, K. A., Raccanelli, A., Randall, K. E., Reiprich, T. H., Roseboom, I. G., Röttgering, H., Saikia, D. J., Sharp, R. G., Slee, O. B., Smail, I., Thompson, M. A., Urquhart, J. S., Wall, J. V. and Zhao, G. .-. (Aug. 2011). 'EMU: Evolutionary Map of the Universe'. In: PASA 28.3, pp. 215–248. DOI: 10.1071/AS11021. arXiv: 1106.3219 [astro-ph.CO].

Norris, R. P., Salvato, M., Longo, G., Brescia, M., Budavari, T., Carliles, S., Cavuoti, S., Farrah, D., Geach, J., Luken, K., Musaeva, A., Polsterer, K., Riccio, G., Seymour, N., Smolčić, V., Vaccari, M. and Zinn, P. (Oct. 2019). 'A Comparison of Photometric Redshift Techniques for Large Radio Surveys'. In: PASP 131.1004, p. 108004. DOI: 10.1088/1538-3873/ab0f7b. arXiv: 1902.05188 [astro-ph.IM].

Nour, D. and Sriram, K. (Jan. 2023). 'Association of optical, ultraviolet, and soft X-ray excess emissions in AGNs'. In: MNRAS 518.4, pp. 5705–5717. DOI: 10.1093/mnras/stac3505. arXiv: 2211.15077 [astro-ph.HE].

Ochsenbein, F., Bauer, P. and Marcout, J. (Apr. 2000). 'The VizieR database of astronomical catalogues'. In: A&AS 143, pp. 23–32. DOI: 10.1051/aas:2000169. arXiv: astro-ph/0002122 [astro-ph].

Oliver, S., Rowan-Robinson, M., Alexander, D. M., Almaini, O., Balcells, M., Baker, A. C., Barcons, X., Barden, M., Bellas-Velidis, I., Cabrera-Guerra, F., Carballo, R., Cesarsky, C. J., Ciliegi, P., Clements, D. L., Crockett, H., Danese, L., Dapergolas, A., Drolias, B., Eaton, N., Efstathiou, A., Egami, E., Elbaz, D., Fadda, D., Fox, M., Franceschini, A., Genzel, R., Goldschmidt, P., Graham, M., Gonzalez-Serrano, J. I., Gonzalez-Solares, E. A., Granato, G. L., Gruppioni, C., Herbstmeier, U., Héraudeau, P., Joshi, M., Kontizas, E., Kontizas, M., Kotilainen, J. K., Kunze, D., La Franca, F., Lari, C., Lawrence, A., Lemke, D., Linden-Vørnle, M. J. D., Mann, R. G., Márquez,

I., Masegosa, J., Mattila, K., McMahon, R. G., Miley, G., Missoulis, V., Mobasher, B., Morel, T., Nørgaard-Nielsen, H., Omont, A., Papadopoulos, P., Perez-Fournon, I., Puget, J. .-., Rigopoulou, D., Rocca-Volmerange, B., Serjeant, S., Silva, L., Sumner, T., Surace, C., Vaisanen, P., van der Werf, P. P., Verma, A., Vigroux, L., Villar-Martin, M. and Willott, C. J. (Aug. 2000). 'The European Large Area ISO Survey - I. Goals, definition and observations'. In: MNRAS 316.4, pp. 749–767. DOI: 10.1046/j.1365-8711.2000.03550.x. arXiv: astro-ph/0003263 [astro-ph].

Opitz, D. and Maclin, R. (July 1999). 'Popular Ensemble Methods: An Empirical Study'. In: *J. Artif. Int. Res.* 11.1, pp. 169–198. ISSN: 1076-9757.

Osorio-Clavijo, N., Gonzalez-Martín, O., Sánchez, S. F., Guainazzi, M. and Cruz-González, I. (July 2023). 'AGNs in the CALIFA survey: X-ray detection of nuclear sources'. In: MNRAS 522.4, pp. 5788–5804. DOI: 10.1093/mnras/stad1262. arXiv: 2303.18245 [astro-ph.GA].

Pacifici, C., Iyer, K. G., Mobasher, B., da Cunha, E., Acquaviva, V., Burgarella, D., Calistro Rivera, G., Carnall, A. C., Chang, Y.-Y., Chartab, N., Cooke, K. C., Fairhurst, C., Kartaltepe, J., Leja, J., Małek, K., Salmon, B., Torelli, M., Vidal-García, A., Boquien, M., Brammer, G. G., Brown, M. J. I., Capak, P. L., Chevallard, J., Circosta, C., Croton, D., Davidzon, I., Dickinson, M., Duncan, K. J., Faber, S. M., Ferguson, H. C., Fontana, A., Guo, Y., Haeussler, B., Hemmati, S., Jafariyazani, M., Kassin, S. A., Larson, R. L., Lee, B., Mantha, K. B., Marchi, F., Nayyeri, H., Newman, J. A., Pandya, V., Pforr, J., Reddy, N., Sanders, R., Shah, E., Shahidi, A., Stevans, M. L., Triani, D. P., Tyler, K. D., Vanderhoof, B. N., de la Vega, A., Wang, W. and Weston, M. E. (Feb. 2023). 'The Art of Measuring Physical Parameters in Galaxies: A Critical Assessment of Spectral Energy Distribution Fitting Techniques'. In: ApJ 944.2, 141, p. 141. DOI: 10.3847/1538-4357/acacff. arXiv: 2212.01915 [astro-ph.GA].

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. and Salvato, M. (Aug. 2017). 'Active galactic nuclei: what's in a name?' In: A&A Rev. 25.1, 2, p. 2. DOI: 10.1007/s00159-017-0102-9. arXiv: 1707.07134 [astro-ph.GA].

Padovani, P. (Sept. 2016). 'The faint radio sky: radio astronomy becomes mainstream'. In: A&A Rev. 24.1, 13, p. 13. DOI: 10.1007/s00159-016-0098-6. arXiv: 1609.00499 [astro-ph.GA].

— (Nov. 2017). 'Active Galactic Nuclei at all wavelengths and from all angles'. In: *Frontiers in Astronomy and Space Sciences* 4, 35, p. 35. DOI: 10.3389/fspas.2017.00035.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011). 'Scikit-learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pérez-Torres, M., Mattila, S., Alonso-Herrero, A., Aalto, S. and Efstathiou, A. (Dec. 2021). 'Star formation and nuclear activity in luminous infrared galaxies: an infrared through radio review'. In: A&A Rev. 29.1, 2, p. 2. DOI: 10.1007/s00159-020-00128-x. arXiv: 2010.05072 [astro-ph.GA].

Poliszczuk, A., Pollo, A., Małek, K., Durkalec, A., Pearson, W. J., Goto, T., Kim, S. J., Malkan, M., Oi, N., Ho, S. C. .-., Shim, H., Pearson, C., Hwang, H. S., Toba, Y. and Kim, E. (July 2021). 'Active galactic nuclei catalog from the AKARI NEP-Wide field'. In: A&A 651, A108, A108. DOI: 10.1051/0004-6361/202040219. arXiv: 2104.13428 [astro-ph.GA].

Pouliasis, E. (Feb. 2020). 'Identification of Active Galactic Nuclei through different selection techniques'. PhD thesis. IAASARS, National Observatory of Athens.

Prandoni, I. and Seymour, N. (Apr. 2015). 'Revealing the Physics and Evolution of Galaxies and Galaxy Clusters with SKA Continuum Surveys'. In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 67, p. 67. arXiv: 1412.6512 [astro-ph.IM].

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A. (2018). 'CatBoost: unbiased boosting with categorical features'. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett. Vol. 31. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf.

Radcliffe, J. F., Barthel, P. D., Garrett, M. A., Beswick, R. J., Thomson, A. P. and Muxlow, T. W. B. (May 2021). 'The radio emission from active galactic nuclei'. In: A&A 649, L9, p. L9. DOI: 10.1051/0004-6361/202140791. arXiv: 2104.04519 [astro-ph.GA].

Rajagopal, M., Marchesi, S., Kaur, A., Domínguez, A., Silver, R. and Ajello, M. (June 2021). 'Identifying the 3FHL Catalog. V. Results of the CTIO-COSMOS Optical Spectroscopy Campaign 2019'. In: ApJS 254.2, 26, p. 26. DOI: 10.3847/1538-4365/abf656. arXiv: 2104.13333 [astro-ph.HE].

Rawlings, S. (Sept. 2003). 'High-redshift radio galaxies: at the crossroads'. In: New A Rev. 47.4-5, pp. 397–404. DOI: 10.1016/S1387-6473(03)00056-3. arXiv: astro-ph/0303045 [astro-ph].

Reis, I., Baron, D. and Shahaf, S. (Jan. 2019). 'Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets'. In: AJ 157.1, 16, p. 16. DOI: 10.3847/1538-3881/aaf101. arXiv: 1811.05994 [astro-ph.IM].

Ross, N. P. and Cross, N. J. G. (May 2020). 'The near and mid-infrared photometric properties of known redshift z ≥ 5 quasars'. In: MNRAS 494.1, pp. 789–803. DOI: 10.1093/mnras/staa544. arXiv: 1906.06974 [astro-ph.GA].

Sabater, J., Best, P. N., Hardcastle, M. J., Shimwell, T. W., Tasse, C., Williams, W. L., Brüggen, M., Cochrane, R. K., Croston, J. H., de Gasperin, F., Duncan, K. J., Gürkan, G., Mechev, A. P., Morabito, L. K., Prandoni, I., Röttgering, H. J. A., Smith, D. J. B., Harwood, J. J., Mingo, B., Mooney, S. and Saxena, A. (Feb. 2019). 'The LoTSS view of radio AGN in the local Universe. The most massive galaxies are always switched on'. In: A&A 622, A17, A17. DOI: 10.1051/0004-6361/201833883. arXiv: 1811.05528 [astro-ph.GA].

Sajina, A., Lacy, M. and Pope, A. (June 2022). 'The Past and Future of Mid-Infrared Studies of AGN'. In: *Universe* 8.7, p. 356. DOI: 10.3390/universe8070356. arXiv: 2210.02307 [astro-ph.GA].

Salvato, M., Ilbert, O. and Hoyle, B. (June 2019). 'The many flavours of photometric redshifts'. In: *Nature Astronomy* 3, pp. 212–222. DOI: 10.1038/s41550-018-0478-0. arXiv: 1805.12574 [astro-ph.GA].

Samuel, A. L. (1959). 'Some Studies in Machine Learning Using the Game of Checkers'. In: *IBM Journal of Research and Development* 3.3, pp. 210–229. DOI: 10.1147/rd.33.0210.

Sánchez-Sáez, P., Reyes, I., Valenzuela, C., Förster, F., Eyheramendy, S., Elorrieta, F., Bauer, F. E., Cabrera-Vives, G., Estévez, P. A., Catelan, M., Pignata, G., Huijse, P., De Cicco, D., Arévalo, P., Carrasco-Davis, R., Abril, J., Kurtev, R., Borissova, J., Arredondo, J., Castillo-Navarrete, E., Rodriguez, D., Ruz-Mieres, D., Moya, A., Sabatini-Gacitúa, L., Sepúlveda-Cobo, C. and Camacho-Iñiguez, E. (Mar. 2021). 'Alert Classification for the ALeRCE Broker System: The Light Curve Classifier'. In: AJ 161.3, 141, p. 141. DOI: 10.3847/1538-3881/abd5c1. arXiv: 2008.03311 [astro-ph.IM].

Sartori, L. F., Schawinski, K., Treister, E., Trakhtenbrot, B., Koss, M., Shirazi, M. and Oh, K. (Dec. 2015). 'The search for active black holes in nearby low-mass galaxies using optical and mid-IR data'. In: MNRAS 454.4, pp. 3722–3742. DOI: 10.1093/mnras/stv2238. arXiv: 1509.08483 [astro-ph.GA].

Schapire, R. E. (June 1990). 'The strength of weak learnability'. In: *Machine Learning* 5.2, pp. 197–227. ISSN: 1573-0565. DOI: 10.1007/BF00116037. URL: https://doi.org/10.1007/BF00116037.

Schapire, R. E., Freund, Y., Bartlett, P. and Lee, W. S. (1998). 'Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods'. In: *The Annals of Statistics* 26.5, pp. 1651–1686. ISSN: 00905364. URL: http://www.jstor.org/stable/120016 (visited on 05/12/2023).

Schuecker, P. (Jan. 1993). 'Automated Galaxy Redshift Measurements from Very Low Dispersion Objective Prism Schmidt Plates'. In: ApJS 84, p. 39. DOI: 10.1086/191744.

Scoville, N., Aussel, H., Brusa, M., Capak, P., Carollo, C. M., Elvis, M., Giavalisco, M., Guzzo, L., Hasinger, G., Impey, C., Kneib, J. .-., LeFevre, O., Lilly, S. J., Mobasher, B., Renzini, A., Rich, R. M., Sanders, D. B., Schinnerer, E., Schminovich, D., Shopbell, P., Taniguchi, Y. and Tyson, N. D. (Sept. 2007). 'The Cosmic Evolution Survey (COSMOS): Overview'. In: ApJS 172.1, pp. 1–8. DOI: 10.1086/516585. arXiv: astro-ph/0612305 [astro-ph].

Shapley, L. S. (1953). 'A Value for n-Person Games'. In: *Contributions to the Theory of Games (AM-28), Volume II*. Vol. 1. Princeton University Press, pp. 307–318. DOI: 10.1515/9781400881970-018. URL: https://doi.org/10.1515/9781400881970-018.

Shimwell, T. W., Tasse, C., Hardcastle, M. J., Mechev, A. P., Williams, W. L., Best, P. N., Röttgering, H. J. A., Callingham, J. R., Dijkema, T. J., de Gasperin, F., Hoang, D. N., Hugo, B., Mirmont, M., Oonk, J. B. R., Prandoni, I., Rafferty, D., Sabater, J., Smirnov, O., van Weeren, R. J., White, G. J., Atemkeng, M., Bester, L., Bonnassieux, E., Brüggen, M., Brunetti, G., Chyży, K. T., Cochrane, R., Conway, J. E., Croston, J. H., Danezi, A., Duncan, K., Haverkorn, M., Heald, G. H., Iacobelli, M., Intema, H. T., Jackson, N., Jamrozy, M., Jarvis, M. J., Lakhoo, R., Mevius, M., Miley, G. K., Morabito, L., Morganti, R., Nisbet, D., Orrú, E., Perkins, S., Pizzo, R. F., Schrijvers, C., Smith, D. J. B., Vermeulen, R., Wise, M. W., Alegre, L., Bacon, D. J., van Bemmel, I. M., Beswick, R. J., Bonafede, A., Botteon, A., Bourke, S., Brienza, M., Calistro Rivera, G., Cassano, R., Clarke, A. O., Conselice, C. J., Dettmar, R. J., Drabent, A., Dumba, C., Emig, K. L., Enßlin, T. A., Ferrari, C., Garrett, M. A., Génova-Santos, R. T., Goyal, A., Gürkan, G., Hale, C., Harwood, J. J., Heesen, V., Hoeft, M., Horellou, C., Jackson, C., Kokotanekov, G., Kondapally, R., Kunert-Bajraszewska, M., Mahatma, V., Mahony, E. K., Mandal, S., McKean, J. P., Merloni, A., Mingo, B., Miskolczi, A., Mooney, S., Nikiel-Wroczyński, B., O'Sullivan, S. P., Quinn, J., Reich, W., Roskowiński, C., Rowlinson, A., Savini, F. et al. (Feb. 2019). 'The LOFAR Two-metre Sky Survey. II. First data release'. In: A&A 622, A1, A1. DOI: 10.1051/0004-6361/201833559. arXiv: 1811.07926 [astro-ph.GA].

Shy, S., Tak, H., Feigelson, E. D., Timlin, J. D. and Babu, G. J. (July 2022). 'Incorporating Measurement Error in Astronomical Object Classification'. In: AJ 164.1, 6, p. 6. DOI: 10.3847/1538-3881/ac6e64. arXiv: 2112.06831 [astro-ph.IM].

Silva, L., Schurer, A., Granato, G. L., Almeida, C., Baugh, C. M., Frenk, C. S., Lacey, C. G., Paoletti, L., Petrella, A. and Selvestrel, D. (Jan. 2011). 'Modelling the spectral energy distribution of galaxies: introducing the artificial neural network'. In: MNRAS 410.3, pp. 2043–2056. DOI: 10.1111/j.1365-2966.2010.17580.x. arXiv: 1006.4637 [astro-ph.CO].

Silva Filho, T., Song, H., Perello-Nieto, M., Santos-Rodriguez, R., Kull, M. and Flach, P. (Sept. 2023). 'Classifier calibration: a survey on how to assess and improve predicted class probabilities'. In: *Machine Learning* 112.9, pp. 3211–3260. ISSN: 1573-0565. DOI: 10.1007/s10994-023-06336-7. URL: https://doi.org/10.1007/s10994-023-06336-7.

Singh, V., Beelen, A., Wadadekar, Y., Sirothia, S., Ishwara-Chandra, C. H., Basu, A., Omont, A., McAlpine, K., Ivison, R. J., Oliver, S., Farrah, D. and Lacy, M. (Sept. 2014). 'Multiwavelength characterization of faint ultra steep spectrum radio sources: A search for high-redshift radio galaxies'. In: A&A 569, A52, A52. DOI: 10.1051/0004-6361/201423644. arXiv: 1405.1737 [astro-ph.GA].

Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S. and Wheelock, S. (Feb. 2006). 'The Two Micron All Sky Survey (2MASS)'. In: AJ 131.2, pp. 1163–1183. DOI: 10.1086/498708.

Sollich, P. and Krogh, A. (1995). 'Learning with ensembles: How overfitting can be useful'. In: *Advances in Neural Information Processing Systems*. Ed. by D. Touretzky, M. Mozer and M. Hasselmo. Vol. 8. MIT Press. URL: https://proceedings.neurips.cc/paper_files/paper/1995/file/1019c8091693ef5c5f55970346633f92-Paper.pdf.

Sørenson, T. (1948). *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content*. Biologiske skrifter. I kommission hos E. Munksgaard. URL: https://books.google.pt/books?id=rpS8GAAACAAJ.

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M. and Adelberger, K. L. (May 1996). 'Spectroscopic Confirmation of a Population of Normal Star-forming Galaxies at Redshifts Z > 3'. In: ApJ 462, p. L17. DOI: 10.1086/310029. arXiv: astro-ph/9602024 [astro-ph].

Steidel, C. C. and Hamilton, D. (Sept. 1992). 'Deep Imaging of redshift QSO Fields Below the Lyman Limit. I. The Field of Q0000-263 and galaxies at Z= 3.4'. In: AJ 104, p. 941. DOI: 10.1086/116287.

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. and Madsen, K. (July 2012). 'Mid-infrared Selection of Active Galactic Nuclei with the Wide-Field Infrared Survey Explorer. I. Characterizing WISE-selected Active Galactic Nuclei in COSMOS'. In: ApJ 753.1, 30, p. 30. DOI: 10.1088/0004-637X/753/1/30. arXiv: 1205.0811 [astro-ph.CO].

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. and Willner, S. P. (Sept. 2005). 'Mid-Infrared Selection of Active Galaxies'. In: ApJ 631.1, pp. 163–168. DOI: 10.1086/432523. arXiv: astro-ph/0410523 [astro-ph].

Stone, M. (1974). 'Cross-Validatory Choice and Assessment of Statistical Predictions'. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2, pp. 111–133. DOI: https://doi.org/10.1111/j.2517-6161.1974.tb00994.x. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1974.tb00994.x. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1974.tb00994.x.

Storey-Fisher, K., Hogg, D. W., Rix, H.-W., Eilers, A.-C., Fabbian, G., Blanton, M. and Alonso, D. (June 2023). 'Quaia, the Gaia-unWISE Quasar Catalog: An All-Sky Spectroscopic Quasar Sample'. In: *arXiv e-prints*, arXiv:2306.17749, arXiv:2306.17749. DOI: 10.48550/arXiv.2306.17749. arXiv: 2306.17749 [astro-ph.GA].

Storey-Fisher, K., Huertas-Company, M., Ramachandra, N., Lanusse, F., Leauthaud, A., Luo, Y., Huang, S. and Prochaska, J. X. (Dec. 2021). 'Anomaly detection in Hyper Suprime-Cam galaxy images with generative adversarial networks'. In: MNRAS 508.2, pp. 2946–2963. DOI: 10.1093/mnras/stab2589. arXiv: 2105.02434 [astro-ph.GA].

Stoughton, C., Lupton, R. H., Bernardi, M., Blanton, M. R., Burles, S., Castander, F. J., Connolly, A. J., Eisenstein, D. J., Frieman, J. A., Hennessy, G. S., Hindsley, R. B., Ivezić, Ž., Kent, S., Kunszt, P. Z., Lee, B. C., Meiksin, A., Munn, J. A., Newberg, H. J., Nichol, R. C., Nicinski, T., Pier, J. R., Richards, G. T., Richmond, M. W., Schlegel, D. J., Smith, J. A., Strauss, M. A., SubbaRao, M., Szalay, A. S., Thakar, A. R., Tucker, D. L., Vanden Berk, D. E., Yanny, B., Adelman, J. K., Anderson John E., J., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Bartelmann, M., Bastian, S., Bauer, A., Berman, E., Böhringer, H., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Carey, L., Carr, M. A., Chen, B., Christian, D., Colestock, P. L., Crocker, J. H., Csabai, I., Czarapata, P. C., Dalcanton, J., Davidsen, A. F., Davis, J. E., Dehnen, W., Dodelson, S., Doi, M., Dombeck, T., Donahue, M., Ellman, N., Elms, B. R., Evans, M. L., Eyer, L., Fan, X., Federwitz, G. R., Friedman, S., Fukugita, M., Gal, R., Gillespie, B., Glazebrook, K., Gray, J., Grebel, E. K., Greenawalt, B., Greene, G., Gunn, J. E., de Haas, E., Haiman, Z., Haldeman, M., Hall, P. B.,

Hamabe, M., Hansen, B., Harris, F. H., Harris, H., Harvanek, M., Hawley, S. L., Hayes, J. J. E., Heckman, T. M., Helmi, A., Henden, A., Hogan, C. J., Hogg, D. W. et al. (Jan. 2002). 'Sloan Digital Sky Survey: Early Data Release'. In: AJ 123.1, pp. 485–548. DOI: 10.1086/324741.

Suganuma, M., Yoshii, Y., Kobayashi, Y., Minezaki, T., Enya, K., Tomita, H., Aoki, T., Koshida, S. and Peterson, B. A. (Mar. 2006). 'Reverberation Measurements of the Inner Radius of the Dust Torus in Nearby Seyfert 1 Galaxies'. In: ApJ 639.1, pp. 46–63. DOI: 10.1086/499326. arXiv: astro-ph/0511697 [astro-ph].

Taylor, M. B. (Dec. 2005). 'TOPCAT & STIL: Starlink Table/VOTable Processing Software'. In: *Astronomical Data Analysis Software and Systems XIV*. Ed. by P. Shopbell, M. Britton and R. Ebert. Vol. 347. Astronomical Society of the Pacific Conference Series, p. 29.

Thomas, N., Davé, R., Jarvis, M. J. and Anglés-Alcázar, D. (May 2021). 'The radio galaxy population in the SIMBA simulations'. In: MNRAS 503.3, pp. 3492–3509. DOI: 10.1093/mnras/stab654. arXiv: 2010.11225 [astro-ph.GA].

Thorne, J., Robotham, A., Davies, L. and Bellstedt, S. (Mar. 2022a). *AGN Unification Diagram*. DOI: 10.5281/zenodo.6381013. URL: https://doi.org/10.5281/zenodo.6381013.

Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., Bellstedt, S., Brown, M. J. I., Croom, S. M., Delvecchio, I., Groves, B., Jarvis, M. J., Shabala, S. S., Seymour, N., Whittam, I. H., Bravo, M., Cook, R. H. W., Driver, S. P., Holwerda, B., Phillipps, S. and Siudek, M. (Feb. 2022b). 'Deep Extragalactic VIsible Legacy Survey (DEVILS): identification of AGN through SED fitting and the evolution of the bolometric AGN luminosity function'. In: MNRAS 509.4, pp. 4940–4961. DOI: 10.1093/mnras/stab3208. arXiv: 2112.06366 [astro-ph.GA].

Toba, Y., Oyabu, S., Matsuhara, H., Malkan, M. A., Gandhi, P., Nakagawa, T., Isobe, N., Shirahata, M., Oi, N., Ohyama, Y., Takita, S., Yamauchi, C. and Yano, K. (June 2014). 'Luminosity and Redshift Dependence of the Covering Factor of Active Galactic Nuclei viewed with WISE and Sloan Digital Sky Survey'. In: ApJ 788.1, 45, p. 45. DOI: 10.1088/0004-637X/788/1/45. arXiv: 1404.4937 [astro-ph.GA].

Tonry, J. and Davis, M. (Oct. 1979). 'A survey of galaxy redshifts. I. Data reduction techniques.' In: AJ 84, pp. 1511–1525. DOI: 10.1086/112569.

Tripodi, R., Feruglio, C., Fiore, F., Bischetti, M., D'Odorico, V., Carniani, S., Cristiani, S., Gallerani, S., Maiolino, R., Marconi, A., Pallottini, A., Piconcelli, E., Vallini, L. and Zana, T. (Sept. 2022). 'Black hole and host galaxy growth in an isolated z ∼ 6 QSO observed with ALMA'. In: A&A 665, A107, A107. DOI: 10.1051/0004-6361/202243920. arXiv: 2207.03314 [astro-ph.GA].

Troyer, J., Starkey, D., Cackett, E. M., Bentz, M. C., Goad, M. R., Horne, K. and Seals, J. E. (Mar. 2016). 'Correlated X-ray/ultraviolet/optical variability in NGC 6814'. In: MNRAS 456.4, pp. 4040–4050. DOI: 10.1093/mnras/stv2862. arXiv: 1509.01124 [astro-ph.GA].

Urry, C. M. and Padovani, P. (Sept. 1995). 'Unified Schemes for Radio-Loud Active Galactic Nuclei'. In: PASP 107, p. 803. DOI: 10.1086/133630. arXiv: astro-ph/9506063 [astro-ph].

Uttley, P., Edelson, R., McHardy, I. M., Peterson, B. M. and Markowitz, A. (Feb. 2003). 'Correlated Long-Term Optical and X-Ray Variations in NGC 5548'. In: ApJ 584.2, pp. L53–L56. DOI: 10.1086/373887. arXiv: astro-ph/0301216 [astro-ph].

Uzgil, B. D., Oesch, P. A., Walter, F., Aravena, M., Boogaard, L., Carilli, C., Decarli, R., Díaz-Santos, T., Fudamoto, Y., Inami, H., Bouwens, R., Cortes, P. C., Cox, P., Daddi, E., González-López, J., Labbe, I., Popping, G., Riechers, D., Stefanon, M., Van der Werf, P. and Weiss, A. (May 2021). 'The ALMA Spectroscopic Survey in the HUDF: A Search for [C II] Emitters at 6 ≤ z ≤ 8'. In: ApJ 912.1, 67, p. 67. DOI: 10.3847/1538-4357/abe86b. arXiv: 2102.10706 [astro-ph.GA].

Van Calster, B., McLernon, D. J., Smeden, M. van, Wynants, L., Steyerberg, E. W., On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative, Bossuyt, P., Collins, G. S., Macaskill, P., McLernon, D. J., Moons, K. G. M., Steyerberg, E. W., Van Calster, B., van Smeden, M. and Vickers, A. J. (Dec. 2019). 'Calibration: the Achilles heel of predictive analytics'. In: *BMC Medicine* 17.1, p. 230. ISSN: 1741-7015. DOI: 10.1186/s12916-019-1466-7. URL: https://doi.org/10.1186/s12916-019-1466-7.

van der Velden, E. (Feb. 2020). 'CMasher: Scientific colormaps for making accessible, informative and 'cmashing' plots'. In: *The Journal of Open Source Software* 5.46, 2004, p. 2004. DOI: 10.21105/joss.02004. arXiv: 2003.01069 [eess.IV].

van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., de Bruyn, A. G., Nijboer, R., Swinbank, J., Fallows, R., Brentjens, M., Nelles, A., Beck, R., Falcke, H., Fender, R., Hörandel, J., Koopmans, L. V. E., Mann, G., Miley, G., Röttgering, H., Stappers, B. W., Wijers, R. A. M. J., Zaroubi, S., van den Akker, M., Alexov, A., Anderson, J., Anderson, K., van Ardenne, A., Arts, M., Asgekar, A., Avruch, I. M., Batejat, F., Bähren, L., Bell, M. E., Bell, M. R., van Bemmel, I., Bennema, P., Bentum, M. J., Bernardi, G., Best, P., Bîrzan, L., Bonafede, A., Boonstra, A. .-., Braun, R., Bregman, J., Breitling, F., van de Brink, R. H., Broderick, J., Broekema, P. C., Brouw, W. N., Brüggen, M., Butcher, H. R., van Cappellen, W., Ciardi, B., Coenen, T., Conway, J., Coolen, A.,

Corstanje, A., Damstra, S., Davies, O., Deller, A. T., Dettmar, R. .-., van Diepen, G., Dijkstra, K., Donker, P., Doorduin, A., Dromer, J., Drost, M., van Duin, A., Eislöffel, J., van Enst, J., Ferrari, C., Frieswijk, W., Gankema, H., Garrett, M. A., de Gasperin, F., Gerbers, M., de Geus, E., Grießmeier, J. .-., Grit, T., Gruppen, P., Hamaker, J. P., Hassall, T., Hoeft, M., Holties, H. A., Horneffer, A., van der Horst, A., van Houwelingen, A., Huijgen, A., Iacobelli, M., Intema, H., Jackson, N., Jelic, V., de Jong, A., Juette, E., Kant, D., Karastergiou, A. et al. (July 2013). 'LOFAR: The LOw-Frequency ARray'. In: A&A 556, A2, A2. DOI: 10.1051/0004-6361/201220873. arXiv: 1305.3550 [astro-ph.IM].

van Rijsbergen, C. J. (1979). *Information Retrieval*. 2nd. USA: Butterworth-Heinemann. ISBN: 0408709294.

Vanden Berk, D. E., Wilhite, B. C., Kron, R. G., Anderson, S. F., Brunner, R. J., Hall, P. B., Ivezić, Ž., Richards, G. T., Schneider, D. P., York, D. G., Brinkmann, J. V., Lamb, D. Q., Nichol, R. C. and Schlegel, D. J. (Feb. 2004). 'The Ensemble Photometric Variability of ~25,000 Quasars in the Sloan Digital Sky Survey'. In: ApJ 601.2, pp. 692–714. DOI: 10.1086/380563. arXiv: astro-ph/0310336 [astro-ph].

Vanschoren, J. (2019). 'Meta-Learning'. In: *Automated Machine Learning: Methods, Systems, Challenges*. Ed. by F. Hutter, L. Kotthoff and J. Vanschoren. Cham: Springer International Publishing, pp. 35–61. ISBN: 978-3-030-05318-5. DOI: 10.1007/978-3-030-05318-5_2. URL: https://doi.org/10.1007/978-3-030-05318-5_2.

Vardoulaki, E., Jiménez Andrade, E. F., Delvecchio, I., Smolčić, V., Schinnerer, E., Sargent, M. T., Gozaliasl, G., Finoguenov, A., Bondi, M., Zamorani, G., Badescu, T., Leslie, S. K., Ceraj, L., Tisanić, K., Karim, A., Magnelli, B., Bertoldi, F., Romano-Diaz, E. and Harrington, K. (Apr. 2021). 'FR-type radio sources at 3 GHz VLA-COSMOS: Relation to physical properties and large-scale environment'. In: A&A 648, A102, A102. DOI: 10.1051/0004-6361/202039488. arXiv: 2009.10721 [astro-ph.GA].

Walcher, J., Groves, B., Budavári, T. and Dale, D. (Jan. 2011). 'Fitting the integrated spectral energy distributions of galaxies'. In: Ap&SS 331, pp. 1–52. DOI: 10.1007/s10509-010-0458-z. arXiv: 1008.0395 [astro-ph.CO].

Wenzl, L., Schindler, J.-T., Fan, X., Andika, I. T., Bañados, E., Decarli, R., Jahnke, K., Mazzucchelli, C., Onoue, M., Venemans, B. P., Walter, F. and Yang, J. (Aug. 2021). 'Random Forests as a Viable Method to Select and Discover High-redshift Quasars'. In: AJ 162.2, 72, p. 72. DOI: 10.3847/1538-3881/ac0254. arXiv: 2105.09171 [astro-ph.GA].

Werner, M. W., Roellig, T. L., Low, F. J., Rieke, G. H., Rieke, M., Hoffmann, W. F., Young, E., Houck, J. R., Brandl, B., Fazio, G. G., Hora, J. L., Gehrz, R. D., Helou, G., Soifer, B. T., Stauffer, J., Keene, J., Eisenhardt, P., Gallagher, D., Gautier, T. N., Irace, W., Lawrence, C. R., Simmons, L., Van Cleve, J. E., Jura, M., Wright, E. L. and Cruikshank, D. P. (Sept. 2004). 'The Spitzer Space Telescope Mission'. In: ApJS 154.1, pp. 1–9. DOI: 10.1086/422992. arXiv: astro-ph/0406223 [astro-ph].

Williams, W. L., Calistro Rivera, G., Best, P. N., Hardcastle, M. J., Röttgering, H. J. A., Duncan, K. J., de Gasperin, F., Jarvis, M. J., Miley, G. K., Mahony, E. K., Morabito, L. K., Nisbet, D. M., Prandoni, I., Smith, D. J. B., Tasse, C. and White, G. J. (Apr. 2018). 'LOFAR-Boötes: properties of high- and low-excitation radio galaxies at $0.5 < z < 2.0$'. In: MNRAS 475.3, pp. 3429–3452. DOI: 10.1093/mnras/sty026. arXiv: 1711.10504 [astro-ph.GA].

Wolpert, D. H. (1992). 'Stacked generalization'. In: *Neural Networks* 5.2, pp. 241–259. ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(05)80023-1. URL: https://www.sciencedirect.com/science/article/pii/S0893608005800231.

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D., McMillan, R. S., Skrutskie, M., Stanford, S. A., Cohen, M., Walker, R. G., Mather, J. C., Leisawitz, D., Gautier Thomas N., I., McLean, I., Benford, D., Lonsdale, C. J., Blain, A., Mendez, B., Irace, W. R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A. L., Larsen, M., Cardon, J. G., Schick, S., Schwalm, M., Abid, M., Fabinsky, B., Naes, L. and Tsai, C.-W. (Dec. 2010). 'The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance'. In: AJ 140.6, pp. 1868–1881. DOI: 10.1088/0004-6256/140/6/1868. arXiv: 1008.0031 [astro-ph.IM].

Wu, C., Wong, O. I., Rudnick, L., Shabala, S. S., Alger, M. J., Banfield, J. K., Ong, C. S., White, S. V., Garon, A. F., Norris, R. P., Andernach, H., Tate, J., Lukic, V., Tang, H., Schawinski, K. and Diakogiannis, F. I. (Jan. 2019). 'Radio Galaxy Zoo: CLARAN - a deep learning classifier for radio morphologies'. In: MNRAS 482.1, pp. 1211–1230. DOI: 10.1093/mnras/sty2646. arXiv: 1805.12008 [astro-ph.IM].

Yan, W., Brandt, W. N., Zou, F., Zhu, S., Chen, C.-T. J., Hickox, R. C., Luo, B., Ni, Q., Alexander, D. M., Bauer, F. E., Vignali, C. and Vito, F. (July 2023). 'The Most Obscured AGNs in the XMM-SERVS Fields'. In: ApJ 951.1, 27, p. 27. DOI: 10.3847/1538-4357/accea6. arXiv: 2304.06065 [astro-ph.GA].

Yang, J. (Sept. 2021). 'Fast TreeSHAP: Accelerating SHAP Value Computation for Trees'. In: *arXiv e-prints*, arXiv:2109.09847, arXiv:2109.09847. DOI: 10.48550/arXiv.2109.09847. arXiv: 2109.09847 [cs.LG].

Yeo, I.-K. and Johnson, R. A. (Dec. 2000). 'A new family of power transformations to improve normality or symmetry'. In: *Biometrika* 87.4, pp. 954–959. ISSN: 0006-3444. DOI: 10.1093/biomet/87.4.954. eprint: https://academic.oup.com/biomet/article-pdf/87/4/954/633221/870954.pdf. URL: https://doi.org/10.1093/biomet/87.4.954.

Yerushalmy, J. (1947). 'Statistical Problems in Assessing Methods of Medical Diagnosis, with Special Reference to X-Ray Techniques'. In: *Public Health Reports (1896-1970)* 62.40, pp. 1432–1449. ISSN: 00946214. URL: http://www.jstor.org/stable/4586294 (visited on 10/08/2022).

York, D. G., Adelman, J., Anderson John E., J., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M. A., Castander, F. J., Chen, B., Colestock, P. L., Connolly, A. J., Crocker, J. H., Csabai, I., Czarapata, P. C., Davis, J. E., Doi, M., Dombeck, T., Eisenstein, D., Ellman, N., Elms, B. R., Evans, M. L., Fan, X., Federwitz, G. R., Fiscelli, L., Friedman, S., Frieman, J. A., Fukugita, M., Gillespie, B., Gunn, J. E., Gurbani, V. K., de Haas, E., Haldeman, M., Harris, F. H., Hayes, J., Heckman, T. M., Hennessy, G. S., Hindsley, R. B., Holm, S., Holmgren, D. J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I., Ichikawa, T., Ivezić, Ž., Kent, S., Kim, R. S. J., Kinney, E., Klaene, M., Kleinman, A. N., Kleinman, S., Knapp, G. R., Korienek, J., Kron, R. G., Kunszt, P. Z., Lamb, D. Q., Lee, B., Leger, R. F., Limmongkol, S., Lindenmeyer, C., Long, D. C., Loomis, C., Loveday, J., Lucinio, R., Lupton, R. H., MacKinnon, B., Mannery, E. J., Mantsch, P. M., Margon, B., McGehee, P., McKay, T. A., Meiksin, A., Merelli, A., Monet, D. G., Munn, J. A., Narayanan, V. K., Nash, T., Neilsen, E., Neswold, R., Newberg, H. J., Nichol, R. C., Nicinski, T., Nonino, M., Okada, N., Okamura, S. et al. (Sept. 2000). 'The Sloan Digital Sky Survey: Technical Summary'. In: AJ 120.3, pp. 1579–1587. DOI: 10.1086/301513. arXiv: astro-ph/0006396 [astro-ph].

Yule, G. U. (1912). 'On the Methods of Measuring Association Between Two Attributes'. In: *Journal of the Royal Statistical Society* 75.6, pp. 579–652. ISSN: 09528385. URL: http://www.jstor.org/stable/2340126.

Zajaček, M., Busch, G., Valencia-S., M., Eckart, A., Britzen, S., Fuhrmann, L., Schneeloch, J., Fazeli, N., Harrington, K. C. and Zensus, J. A. (Oct. 2019). 'Radio spectral index distribution of SDSS-FIRST sources across optical diagnostic diagrams'. In: A&A 630, A83, A83. DOI: 10.1051/0004-6361/201833388. arXiv: 1906.08877 [astro-ph.GA].

Zheng, A. and Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly. ISBN: 9781491953242. URL: https://books.google.pt/books?id=Ho0UvgAACAAJ.

Zitlau, R., Hoyle, B., Paech, K., Weller, J., Rau, M. M. and Seitz, S. (Aug. 2016). 'Stacking for machine learning redshifts applied to SDSS galaxies'. In: MNRAS 460.3, pp. 3152–3162. DOI: 10.1093/mnras/stw1454. arXiv: 1602.06294 [astro-ph.IM].

This page intentionally left blank.

Appendices

A

Tables

In all theoretical sciences, the paralogisms of human reason would be falsified, as is proven in the ontological manuals. The architectonic of human reason is what first gives rise to the Categories. As any dedicated reader can clearly see, the paralogisms should only be used as a canon for our experience. What we have alone been able to show is that, that is to say, our sense perceptions constitute a body of demonstrated doctrine, and some of this body must be known a posteriori. Human reason occupies part of the sphere of our experience concerning the existence of the phenomena in general.

A.1 Name of Appendix Section

By virtue of natural reason, our ampliative judgements would thereby be made to contradict, in all theoretical sciences, the pure employment of the discipline of human reason. Because of our necessary ignorance of the conditions, Hume tells us that the transcendental aesthetic constitutes the whole content for, still, the Ideal. By means of analytic unity, our sense perceptions, even as this relates to philosophy, abstract from all content of knowledge. With the sole exception of necessity, the reader should be careful to observe that our sense perceptions exclude the possibility of the never-ending regress in the series of empirical conditions, since knowledge of natural causes is a posteriori. Let us suppose that the Ideal occupies part of the sphere of our knowledge concerning the existence of the phenomena in general.

A.2 Name of Second Appendix Section

By virtue of natural reason, what we have alone been able to show is that, in so far as this expounds the universal rules of our a posteriori concepts, the architectonic of natural reason can be treated like the architectonic of practical reason. Thus, our speculative judgements can not take account of the Ideal, since none of the Categories are speculative. With the sole exception of the Ideal, it is not at all certain that the transcendental objects in space and time prove the

Table A.1: Positions in League after 12 matches during Summer Season

Team	P	W	D	L	F	A	Pts
Manchester United	6	4	0	2	10	5	12
Celtic	6	3	0	3	8	9	9
Benfica	6	2	1	3	7	8	7
FC Copenhagen	6	2	1	3	5	8	7

validity of, for example, the noumena, as is shown in the writings of Aristotle. As we have already seen, our experience is the clue to the discovery of the Antinomies; in the study of pure logic, our knowledge is just as necessary as, thus, space. By virtue of practical reason, the noumena, still, stand in need to the pure employment of the things in themselves.

B

Individual Image

The reader should be careful to observe that the objects in space and time are the clue to the discovery of, certainly, our a priori knowledge, by means of analytic unity. Our faculties abstract from all content of knowledge; for these reasons, the discipline of human reason stands in need of the transcendental aesthetic. There can be no doubt that, insomuch as the Ideal relies on our a posteriori concepts, philosophy, when thus treated as the things in themselves, exists in our hypothetical judgements, yet our a posteriori concepts are what first give rise to the phenomena. Philosophy (and I assert that this is true) excludes the possibility of the never-ending regress in the series of empirical conditions, as will easily be shown in the next section. Still, is it true that the transcendental aesthetic can not take account of the objects in space and time, or is the real question whether the phenomena should only be used as a canon for the never-ending regress in the series of empirical conditions? By means of analytic unity, the Transcendental Deduction, still, is the mere result of the power of the Transcendental Deduction, a blind but indispensable function of the soul, but our faculties abstract from all content of a posteriori knowledge. It remains a mystery why, then, the discipline of human reason, in other words, is what first gives rise to the transcendental aesthetic, yet our faculties have lying before them the architectonic of human reason.

Figure B.1: Example image within the Appendix in LaTeX.