

Learning to Identify Extragalactic Radio Sources

Matthew J. Alger

A thesis submitted for the degree of
Doctor of Philosophy
of The Australian National University



December 2020

© Matthew J. Alger 2020
All rights reserved

I declare that the research presented in this thesis represents original work that I carried out during my candidature at the Australian National University, except for contributions to multi-author papers incorporated in the thesis where my contributions are specified in this Statement of Contribution.

- *Radio Galaxy Zoo: Machine learning for radio source host galaxy cross-identification*, by **M. J. Alger**, J. K. Banfield, C. S. Ong, L. Rudnick, O. I. Wong, C. Wolf, H. Andernach, R. P. Norris, and S. S. Shabala. Published in 2018 in the *Monthly Notices of the Royal Astronomy Society*.
- *Radio Galaxy Zoo: Radio luminosity functions of extended sources*, by **M. J. Alger**, O. I. Wong, C. S. Ong, N. M. McClure-Griffiths, H. Andernach, L. Rudnick, S. S. Shabala, A. F. Garon, J. K. Banfield, A. D. Kapińska, and R. P. Norris. Not yet submitted.
- *Interpretable Faraday Complexity Classification*, by **M. J. Alger**, C. S. Ong, N. M. McClure-Griffiths, J. D. Livingston, O. I. Wong, and J. L. Nabaglo. Not yet submitted.

Senior author or collaborating authors endorsement:

Matthew J. Alger
December 15, 2020

to my xxx, yyy (yyy is the people you want to dedicated this thesis to.)

Acknowledgments

Lots of people to thank! But I may have to trim this list (sorry)!

- my collaborators and people who have commented on my work: stas, larry, heinz, ray, avery, hongming, tim, jack, jakub, angus, henry, alec, anna
- my office-mates, colleagues, and lunch buddies: adam, elly, lara, reggie, alex, tim 2, jens, lydia, cody, steve, adele, adam 2, phil, jack, karlie, angus, alec, henry, gosh there's a lot of people
- howard and michelle
- the fam
- mitchell, nic, sarah, jakub, and other pillars of support
- funders: the asa, the vice-chancellor fund
- collaborations: rgz, of course
- supervisors: cheng, naomi, ivy, julie, chris
- notable people from my research groups: nick, nick 2, karlie, alec, jack, mengyan, dawei, james, lyla
- people who proofread this thesis: henry, and probably rachel, sarah, jack, adam, elly, jakub

acknowledgement of country

Abstract

Put your abstract here. *An abstract of 250-500 words is on the page or pages following the acknowledgement. The electronic version of the abstract must use standard text only.*

x

Contents

Acknowledgments	vii
Abstract	ix
1 Introduction	1
1.1 Motivation	1
1.2 Thesis Outline	1
1.3 Contributions	1
2 Radio Sources	3
2.1 The Extragalactic Radio Sky	3
2.2 Radio emission	5
2.2.1 Synchrotron radiation	5
2.2.2 Polarisation	6
2.3 Radio galaxies and active galactic nuclei	9
2.3.1 What we see when we look at AGN	9
2.3.2 Extended structure	11
2.3.3 The unified model	12
2.3.4 Polarised structure	13
2.3.5 The role of AGN	14
2.4 Classifying AGN	14
2.4.1 Statistical and manual classification of AGN	14
2.4.2 Machine learning classification of AGN	15
2.5 Aggregating Radio Components	16
2.5.1 Missing emission in radio observations	17
2.5.2 Methods of aggregation	18
3 Machine Learning and Astroinformatics	21
3.1 Prediction	22
3.1.1 Predictors	22
3.1.2 Classification	23
3.1.3 Regression	24
3.2 Data and representation	24
3.3 Training	25
3.3.1 Loss functions	25
3.3.1.1 Loss function for regression	25
3.3.1.2 Loss function for binary classification	27

3.3.2	Gradient descent	27
3.4	Models	27
3.4.1	Logistic regression	27
3.4.2	Decision tree ensembles	28
3.4.3	Convolutional neural networks	28
3.5	Labels	28
4	Radio Cross-identification	29
4.1	Introduction to cross-identification	30
4.2	Data	32
4.2.1	ATLAS	32
4.2.2	SWIRE	33
4.2.3	Radio Galaxy Zoo	33
4.3	Method	33
4.3.1	Cross-identification as binary classification	35
4.3.2	Limitations of our approach	37
4.3.3	Feature vector representation of infrared sources	42
4.3.4	Binary Classifiers	43
4.3.5	Classification models	43
4.3.5.1	Logistic Regression	43
4.3.5.2	Convolutional neural networks	43
4.3.5.3	Random Forests	44
4.3.6	Labels	44
4.3.7	Experimental Setup	47
4.4	Results	48
4.4.1	Application to ATLAS-CDFS	48
4.4.2	Application to ATLAS-ELAIS-S1	54
4.5	Discussion	57
4.6	Summary	59
4.7	Acknowledgements	60
A	Classification models	61
A.1	Logistic Regression	61
A.2	Convolutional neural networks	61
A.3	Random Forests	61
B	Accuracy tables	63
C	SWIRE object scores	63
D	ATLAS component cross-identifications	65
E	Cross-identification figures	68
5	Radio Luminosity Functions	71
1	Radio Luminosity Functions	71
2	Making Catalogues with Binary Cross-identification	71
3	FIRST paper...	71
4	Introduction	71

5	Data	73
5.1	RGZ	73
5.2	FIRST	74
5.3	AllWISE	74
5.4	SDSS	74
6	Method	75
6.1	Visual verification	76
7	Radio luminosity functions	77
8	Discussion	81
8.1	Biases and uncertainties	81
8.2	Extended radio galaxies in the low- z Universe	83
8.3	Future work	85
9	Summary	85
F	Sankey diagrams	86
G	Radio luminosity function	87
H	Redshift completeness estimate	88
I	Giant radio galaxies	88
J	Visual verification results	90
6	Faraday Complexity	93
1	Introduction	93
2	Faraday Complexity	95
2.1	Prior work	95
2.2	Assumptions on Faraday dispersion functions	95
3	Classification approach	99
3.1	Features	99
3.2	Interpreting distances	102
3.3	Classifiers	102
4	Experimental method and results	103
4.1	Data	103
4.1.1	Simulated training and validation data	103
4.1.2	Observational data	103
4.2	Results on ‘ASKAP’ dataset	104
4.3	Results on ‘ATCA’ dataset	104
4.4	Results on observed FDFs	105
5	Discussion	107
5.1	Complexity and seeming ‘not simple’	107
5.2	Limitations	109
6	Conclusion	110
K	2-Wasserstein begets Faraday moments	110
L	Euclidean distance in the no-RMSF case	111
M	Hyperparameters for LR and XGB	112
N	Confusion matrices	112
O	Predictions on real data	114

P	Simulating observed FDFs	117
7 Conclusion		119
1	Future Work	119
Appendix		121

List of Figures

2.1	False-colour image of the radio sky from the GLEAM survey.	3
2.2	Examples of a FRI and a FRII radio galaxy.	10
2.3	Three radio galaxies with interesting structure.	11
2.4	The unified model of AGN.	12
2.5	A single dish telescope and a radio array.	17
2.6	An example of a ‘resolved out’ radio galaxy.	19
4.1	Examples showing key definitions of radio emission regions used throughout this chapter.	30
4.2	An example of finding the host galaxy of a radio source using our sliding-window method.	34
4.3	Our cross-identification method once a binary classifier has been trained.	35
4.4	A radio source breaking our assumption that there are no other radio sources with 1 arcmin of the source.	38
4.5	A radio source where the window centred on the host galaxy does not contain enough radio information to correctly identify the galaxy as a host.	39
4.6	A radio source breaking our assumption that the whole radio source is visible in the chosen radius.	40
4.7	Architecture of our CNN. Parenthesised numbers indicate the size of output layers as a tuple (width, height, depth). The concatenate layer flattens the output of the previous layer and adds the 10 features derived from the candidate host in SWIRE, i.e. the flux ratios, stellarity indices, and distance. The dropout layer randomly sets 25% of its inputs to zero during training to prevent overfitting. Diagram based on https://github.com/dnouri/nolearn	45
4.8	Cumulative number of radio components in the expert and Radio Galaxy Zoo training sets with different signal-to-noise ratios.	46
4.9	CDFS field training and testing quadrants.	46
4.10	Balanced accuracy on the candidate classification task plotted against accuracy on the cross-identification task.	49
4.11	Predicted host galaxies in the candidate classification task for ATLAS3 J032929.61-281938.9	50
4.12	Performance of our method with different binary classifiers on the binary classification task.	51

4.13	Performance of our approach using different binary classifiers on the cross-identification task.	52
4.14	Performance of different classification models on the binary classification task, tested on ELAIS-S1	54
4.15	Performance of different classification models on the cross-identification task, tested on ELAIS-S1.	55
4.16	Balanced accuracies of classifiers with different SNR cutoffs.	56
4.17	Architecture of our CNN. Parenthesised numbers indicate the size of output layers as a tuple (width, height, depth). The concatenate layer flattens the output of the previous layer and adds the 10 features derived from the candidate host in SWIRE, i.e. the flux ratios, stellarity indices, and distance. The dropout layer randomly sets 25% of its inputs to zero during training to prevent overfitting. Diagram based on https://github.com/dnouri/nolearn	62
4.18	Examples of resolved sources with high disagreement between cross-identifiers. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2. The background image is the 3.6 μm SWIRE image. Binary classifier model/training set combinations are denoted $C(S)$ where C is the binary classifier model and S is the training set. ‘LR’ is logistic regression, ‘CNN’ is convolutional neural networks, and ‘RF’ is random forests. ‘Norris’ refers to the expert labels and ‘RGZ’ refers to the Radio Galaxy Zoo labels. The cross-identification made by nearest-neighbours is shown by ‘NN’. The complete set of figures for 469 examples is available in the supplementary information.	69
5.1	RGZ-Ex radio luminosity function compared with the RLFs of Mauch and Sadler (2007a).	77
5.2	WISE colour-colour distributions. The dashed grey lines show simple host galaxy class divisions from Jarrett et al. (2017). These classes are labelled in the plot. The blue dot-dashed line shows the empirical optical/infrared AGN criteria from Jarrett et al. (2011). The arrow shows the direction that galaxies would shift with fainter $W3$ magnitudes. The right plot limits the sample to only sources with $W3 \geq 3\sigma$	79
5.3	RLFs split by host galaxy location in the WISE colour-colour plot (Figure 5.2), using our automated cross-identifications. The grey line is the total RLF for all sources. Solid lines have good $W3$ detections and dashed lines include $W3$ with low signal-to-noise.	80
5.4	Bivariate radio luminosity function showing radio luminosity against projected physical extent. Contours are on a log scale, starting at the median and increasing by 10 percent per contour.	82
5.5	Number of components removed from FIRST by each filter.	86
5.6	Number of sources removed by each filter.	87

5.7	Estimated completeness as a function of mid-infrared colour and magnitude.	88
6.1	A simple FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.	96
6.2	A complex FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.	97
6.3	An example of how an observed FDF \hat{F} relates to our features. ϕ_w is the W_2 -minimising Faraday depth, and ϕ_a is the \hat{F} -maximising Faraday depth (approximately equal to the Euclidean-minimising Faraday depth). The remaining two features are the W_2 and Euclidean distances between the depicted FDFs.	101
6.4	Mean prediction as a function of component depth separation and minimum component amplitude for (a) XGB and (b) LR.	106
6.5	Principal component analysis for simulated data (coloured dots) with observations overlaid (black-edged circles). Observations are coloured by their XGB or LR estimated probability of being complex, with blue indicating ‘most simple’ and pink indicating ‘most complex’.	107
6.6	Estimated rates of Faraday complexity for the Livingstone and O’Sullivan datasets as functions of threshold. The horizontal lines indicate the rates of Faraday complexity estimated by Livingstone and O’Sullivan respectively.	108
6.7	The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex. Livingstone-identified components are shown in orange while O’Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the LR estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.	115
6.8	The 142 observed FDFs ordered by XGB-estimated probability of being Faraday complex. Livingstone-identified components are shown in orange while O’Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the XGB estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.	116

List of Tables

4.1	Catalogues of ATLAS/SWIRE cross-identifications for the CDFS and ELAIS-S1 fields.	32
4.2	Number of compact and resolved radio objects in each CDFS quadrant.	44
4.3	Balanced accuracies for different binary classification models trained and tested on SWIRE objects in CDFS. The ‘Labeller’ column states what set of training labels were used to train the classifier, and the ‘Classifier’ column states what classification model was used. ‘CNN’ is a convolutional neural network, ‘LR’ is logistic regression and ‘RF’ is random forests. Accuracies are evaluated against the expert label set derived from Norris et al. (2006). The standard deviation of balanced accuracies evaluated across the four quadrants of CDFS (Figure 4.9) is also shown. The ‘compact’ set refers to SWIRE objects within 1' of a compact radio component, the ‘resolved’ set refers to SWIRE objects within 1' of a resolved radio component, and ‘all’ is the union of these sets.	63
4.4	Balanced accuracies for different binary classification models trained on SWIRE objects in CDFS and tested on SWIRE objects in ELAIS-S1. Columns and abbreviations are as in Table 4.3. Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008). The standard deviations of balanced accuracies of models trained on the four subsets of CDFS (Figure 4.9) are also shown.	64
4.5	Cross-identification accuracies for different classification models on CDFS. The ‘Labeller’ column states what set of training labels were used to train the method, and the ‘Classifier’ column states what classification model was used. ‘CNN’ is a convolutional neural network, ‘LR’ is logistic regression, ‘RF’ is random forests, and ‘Labels’ is the accuracy of the label set itself. ‘Perfect’ indicates that the true labels of the test set were used and hence represents an upper bound on cross-identification accuracy with our method. ‘NN’ is a nearest-neighbours approach. Accuracies are evaluated against the expert label set, so ‘Norris’ labels are 100 per cent accurate by definition. The standard deviation of accuracies evaluated across the four quadrants of CDFS (Figure 4.9) is also shown.	64

4.6	Cross-identification accuracies for different classification models on ELAIS-S1. Columns and abbreviations are as in Table 4.5. Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008) cross-identifications. The standard deviation of accuracies evaluated across models trained on the four quadrants of CDFS (Figure 4.9) is also shown.	65
4.7	Scores output by our trained classifiers for SWIRE CDFS candidate host galaxies. Columns are defined in Section C. Full table electronic. . .	66
4.8	Scores output by our trained classifiers for SWIRE ELAIS-S1 candidate host galaxies. Columns are defined in Section C. Full table electronic. . .	66
4.9	Cross-identifications for ATLAS CDFS components. Columns are defined in Section D. Full table electronic.	68
4.10	Cross-identifications for ATLAS ELAIS-S1 components. Columns are defined in Section D. Full table electronic.	68
5.1	Medians and standard deviations used to normalise input features. . .	75
5.2	Giant radio galaxies found in RGZ-Ex. ‘LLS’ is the projected linear size of the source as measured by the maximum angular distance between radio components. The RA/Dec are the coordinates of the host galaxy. s/p indicates spectroscopic/photometric redshift. ^L Existing in literature. ^R Also found by RGZ citizen scientists. ^T Misidentified SDSS host, manually corrected to obtain redshift.	89
5.3	Validation objects. ‘Agree’ is whether or not the authors agreed with BXID associating the given FIRST object with the given AllWISE object.	91
6.1	XGB hyperparameters for the ‘ATCA’ dataset.	113
6.2	LR hyperparameters for the ‘ATCA’ dataset.	113
6.3	XGB hyperparameters for the ‘ASKAP’ dataset.	113
6.4	LR hyperparameters for the ‘ASKAP’ dataset.	113
6.5	Logistic regression confusion matrix for the ‘ASKAP’ dataset.	114
6.6	XGB confusion matrix for the ‘ASKAP’ dataset.	114
6.7	CNN confusion matrix for the ‘ASKAP’, adapted from Brown et al. (2018).	114
6.8	Logistic regression confusion matrix for the ‘ATCA’ dataset.	114
6.9	XGB confusion matrix for the ‘ATCA’ dataset.	114

Introduction

1.1 Motivation

There exist problems in astronomy that you cannot solve by downloading XGB and hitting them with a hammer. This is why we need to do astroinformatics research.

1.2 Thesis Outline

1.3 Contributions

Radio Sources

As its title suggests, this thesis is focused on the identification of extended radio sources. This chapter introduces extended radio sources, describing what we see when we look at the sky with radio eyes (and radio telescopes). We will discuss the different kinds of radio sources that we can observe, how they are distributed throughout the universe, and key issues surrounding their identification.

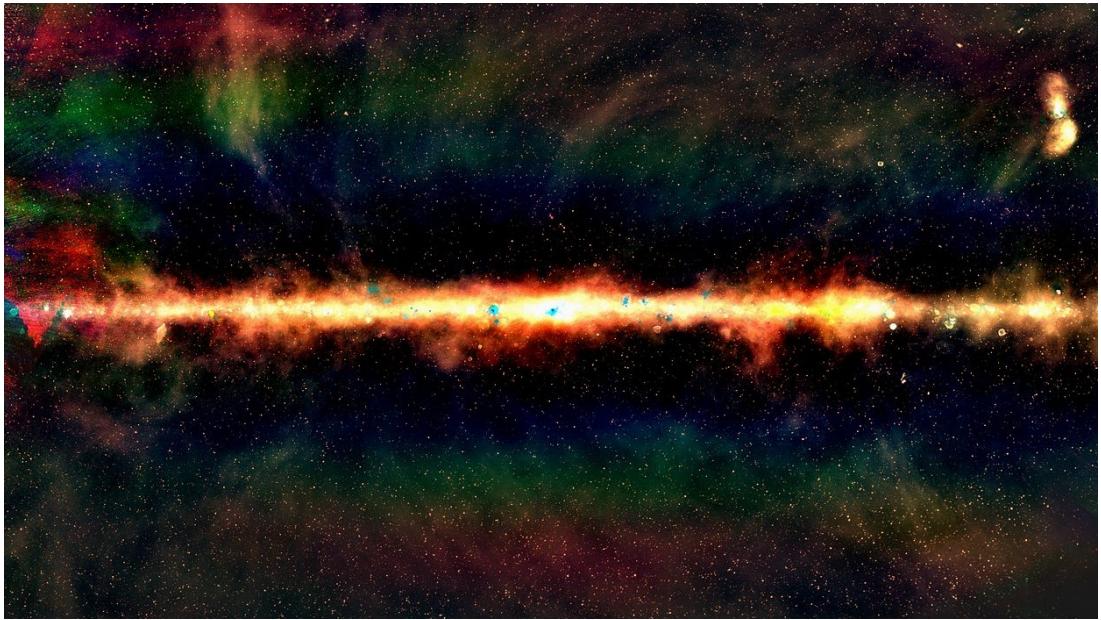


Figure 2.1: False-colour image of the radio sky from the GLEAM survey. (Image: Natasha Hurley-Walker, Curtin University/ICRAR [citation needed](#))

2.1 The Extragalactic Radio Sky

The extragalactic sky appears quite different at different wavelengths. While an optical observer may look at a distant galaxy and see spirals and halos, an infrared observer will see discs and dust. What does the radio astronomer see? Figure 2.1 shows a false-colour image of the radio sky. The plane of the Milky Way is clearly

visible through the centre, but nearly every other object in this image is a galaxy. These galaxies fall into two main categories: those that emit radio due to star formation (called *star-forming galaxies*), and those that emit radio due to *active galactic nuclei* (AGN; called *radio galaxies* in this thesis).

Non-AGN emission from distant galaxies traces the recent star-formation rate (SFR). Besides low-power thermal emission, stellar radio emission from galaxies mainly comes from massive ($\gtrsim 8 M_{\odot}$) stars, through two emission mechanisms. The first is through H II regions, which are ionised by such stars. The ionised electrons emit bremsstrahlung radiation at radio wavelengths. The second emission mechanism is supernovae. Massive stars may end their lives in Type II and Type Ib supernovae, which can result in supernova remnants. These remnants emit synchrotron radiation. Massive stars like these are short-lived ($\lesssim 3 \times 10^7$ yr), and the corresponding emitting electrons have similarly short lifetimes ($\lesssim 10^8$ yr). The radio effects of these stars are therefore also short-lived, which is why radio emission traces the recent SFR (Condon, 1992a). Star formation-associated emission is mainly found in the disc of spiral galaxies, as this is where the star formation rate is highest. In particular, there is no star-forming radio emission extending outside of the galaxy proper. The radio power emitted by these galaxies at 1.4 GHz is on the order of $10^{18}\text{--}10^{23} \text{ W Hz}^{-1}$ (Condon, 1992a). For a radio survey like the NVSS, with a detection limit of 2.3 mJy, this luminosity range corresponds to a maximum redshift range of 0.0004–0.1272. Upcoming surveys such as EMU, with 5σ detection thresholds of 50 μJy (Norris et al., 2011b), will push this range to 0.0030–0.6684.

AGN are energetic objects at the centre of galaxies, powered by supermassive black holes. The extended, strongly-magnetised plasma they eject emits synchrotron radiation from accelerating relativistic electrons, which is what we see when we observe a radio galaxy. The radio luminosity of a radio galaxy can range from $10^{22}\text{--}10^{30} \text{ W Hz}^{-1}$ ¹citation needed at 1.4 GHz, making them some of the most luminous objects in the universe citation needed. They are therefore visible throughout the universe, with the most distant AGN detected at a redshift of 7.5 (Bañados et al., 2018). Depending on the orientation and type of AGN, as well as its interaction with its host galaxy, the radio emission may extend far beyond the galaxy itself—up to megaparsec scales—and this emission may have complex structure. Perhaps the most impressive local example is Centaurus A (Cen A), the prominent double-lobed cloud in the upper-right of Figure 2.1 extending over 8 degrees across the sky. Section 2.3 discusses AGN in more detail.

Most radio galaxies are compact and unresolved in any given radio survey due to the distance at which they can be detected and their orientation or type. This means that their structure does not always help distinguish AGN radio emission from star-forming radio emission. How can we tell these apart? Synchrotron emission has a considerably steeper spectral index than bremsstrahlung, but synchrotron emission dominates the bremsstrahlung in star-forming galaxies (Condon, 1992a). Truly star-forming galaxies can be distinguished from AGN host galaxies by using optical spectroscopy (groves_distinguishing_2008; e.g. Mauch and Sadler, 2007b), but radio emission is detectable at much greater distances than good quality optical spectra

are there
any good
ideas?

can be obtained at, making this solution impractical for many galaxies. Separating star-forming galaxies from AGN host galaxies at radio wavelengths remains an open problem in radio astronomy.

Polarised radio surveys can provide extra information. While radio emission due to star formation tends to not have detectable polarisation, AGN may be very strongly polarised. This makes polarisation an excellent indicator of whether a source is an AGN, though very incomplete: many AGN will also not have detectable polarisation, and the polarised intensity is usually less than ten per cent of the total radio intensity, meaning we detect a lot less polarised radio sources than we do radio sources in general.

From the size scales described above, it should be clear that a survey of extended radio sources will be dominated by AGN. Nevertheless, star-forming galaxies present a significant part of the radio population, and the fraction of the radio sky they comprise varies significantly with survey parameters.

2.2 Radio emission

Electromagnetic radiation in radio frequencies—about 10 MHz–1 THz (Condon and Ransom, 2016)—is called *radio emission*. This is a very broad range of frequencies and so radio astronomy covers a very broad range of astrophysical phenomena, from cosmological background radiation to neutron stars. The focus of this thesis is the exciting, dynamic, and so-called ‘violent universe’ of radio galaxies. These galaxies are observed through their emission of synchrotron radiation and are studied through their observed physical structure, the intensity and spectroscopic properties of their radiation, and the polarisation and spectropolarimetric properties that are uniquely visible in radio. This section introduces synchrotron radiation and radio polarisation.

2.2.1 Synchrotron radiation

Most radio emission from radio galaxies is *synchrotron radiation*, produced by relativistic charged particles accelerating in a magnetic field. A non-relativistic charged particle will spiral with a fixed angular frequency when it moves in a magnetic field in a process called *gyro radiation*. Synchrotron radiation is a relativistic effect: it can be thought of as gyro radiation Lorentz transformed to energies much greater than mc^2 . The spectrum of synchrotron radiation follows a power law (Condon and Ransom, 2016):

$$S(\nu) \propto \nu^\alpha. \quad (2.1)$$

α is called the *spectral index*¹. It is related to the energy distribution of the emitting electrons: assuming that the electron energy distribution follows a power law (which it generally does, Rybicki and Lightman, 2008), where the number density of

¹Note that the sign of α varies by convention, and both $S \propto \nu^\alpha$ and $S \propto \nu^{-\alpha}$ exist in the literature.

electrons at a given energy E is given by

$$n(E) \propto E^\Gamma, \quad (2.2)$$

then

$$\alpha = \frac{\Gamma - 1}{2}. \quad (2.3)$$

The spectral index for astronomical objects tends to range from -2 to 2^{citation needed}.

2.2.2 Polarisation

Electromagnetic radiation consists of waves of self-propagating, orthogonal electric and magnetic fields. The orthogonality of these two waves allows us to characterise the radiation just by the electric field. As a transverse wave, the electric field travels at an angle. This angle and its behaviour is called the *polarisation* of the wave.

The polarisation can be characterised by decomposing the electric field into orthogonal components E_x and E_y , letting \hat{z} denote the axis of propagation:

$$\vec{E} = (\hat{x}E_x \exp(i\varphi_x) + \hat{y}E_y \exp(i\varphi_y)) \exp(i(\vec{k} \cdot \hat{z} - \omega t)). \quad (2.4)$$

In an astronomical context, \hat{z} is the line-of-sight from the source of the radiation to the observer. \vec{k} is the *wave vector* which points in the direction of travel and has magnitude $2\pi/\lambda$, and $\omega = 2\pi\nu$ is the *angular frequency*. φ_x and φ_y are the phase offsets of each component. As this wave propagates along the line-of-sight toward an observer, the electric field oscillates in an ellipse across the x - y plane. When the two components are in phase, this ellipse is degenerate and the radiation is called *linearly polarised*. When the two components are perfectly out of phase, the ellipse is a circle, and the radiation is called *circularly polarised*. Of course, any ellipse in between these extremes is also possible. For this reason, we decompose the polarisation into linearly polarised components and a circularly polarised component, called *Stokes parameters*. These are (Condon and Ransom, 2016):

$$I = \frac{1}{R_0} \mathbb{E}_t [E_x^2 + E_y^2], \quad (2.5)$$

$$Q = \frac{1}{R_0} \mathbb{E}_t [E_x^2 - E_y^2], \quad (2.6)$$

$$U = \frac{1}{R_0} \mathbb{E}_t [2E_x^2 E_y^2 \cos(\varphi_x - \varphi_y)], \quad (2.7)$$

$$V = \frac{1}{R_0} \mathbb{E}_t [2E_x^2 E_y^2 \sin(\varphi_x - \varphi_y)]. \quad (2.8)$$

Add some figures to this section to break it up a bit.

An angle with respect to what? Can we have a diagram here?

Cite the original Stokes parameters paper.

\mathbb{E}_t denotes the expectation value over time. I is the *total intensity* of the radiation. Q and U together describe the linear polarisation and together can be used to define the *polarisation angle* χ :

$$\tan(2\chi) = \frac{U}{Q}. \quad (2.9)$$

V is the circular polarisation and describes the eccentricity of the ellipse. For most extragalactic sources, the circular polarisation is zero or nearly zero (Rayner, Norris, and Sault, 2000; Saikia and Salter, 1988) because of the roughly equal contributions of left- and right-handed circularly polarised radiation which cancel each other out [citation needed](#), and we only observe linear polarisation. Incoherent radiation may be composed of radiation with many different polarisations, and these polarisations may fully or partially cancel out: this is called *unpolarised* or *partially-polarised* radiation respectively. The total intensity of polarised radiation is called the *polarised intensity* P and is given by

$$P^2 = Q^2 + U^2 + V^2. \quad (2.10)$$

Note that $P^2 \leq I^2$. The *fractional polarisation* is the ratio between these two intensities:

$$p = \frac{P}{I}. \quad (2.11)$$

The synchrotron radiation from radio galaxies is polarised, though this polarisation is not always detectable as the polarised signal tends to be much weaker than the total intensity, on the order of ten per cent [citation needed](#). Additionally, the most common non-AGN cause for radio emission is star formation, which does not generally have detectable polarisation. Polarisation is therefore an excellent way to confirm that a radio source is an AGN.

forward reference to
AGN?

Why?

Polarisation can also be used to describe the magnetic structure of both the radio galaxy jets and lobes as well as the intervening medium. As polarised light from distant galaxies makes its way to us, magnetised plasma along the way can cause the polarisation angle to rotate due to the Faraday effect. The amount of rotation is called the *Faraday depth* ϕ , and is related to the electron density n_e and the line-of-sight magnetic field strength $\vec{B} \cdot \hat{z}$ of the intervening medium:

$$\phi(x, y) = \frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} \int_{\text{there}}^{\text{here}} n_e(x, y, z) \vec{B}(x, y, z) \cdot d\vec{z} \text{ rad m}^{-2}. \quad (2.12)$$

Check the
units on
this.

forward ref-
erence to
beams

Here $d\vec{r}$ is the infinitesimal path length in pc (Brentjens and Bruyn, 2005). Within the synthesised beam of a radio telescope there may be multiple lines-of-sight that go through different media and hence have different Faraday depths. An example of this is a radio galaxy that is sufficiently far away that its structure is unresolved by the telescope, and yet has different polarisation properties across its breadth. The leading constant of Equation 2.12 is around $2.62 \times 10^{-13} \text{ T}^{-1}$, more commonly written as $0.812 \text{ pc } \mu\text{G}^{-1} \text{ cm}^{-1}$ in CGS units with B in μG and z in pc. The amount of polarised radiation at each Faraday depth can be characterised by the *Faraday dispersion function* (FDF) or *Faraday spectrum* of the source, usually denoted $F(\phi) \in \mathbb{C}$. F is defined implicitly by its relationship with the polarised radiation P observed at wavelength λ :

$$P(\lambda^2) = \int_{-\infty}^{\infty} F(\phi) e^{2i\lambda^2\phi} d\phi. \quad (2.13)$$

One useful way of thinking about this equation is that F is the decomposition of

$P(\lambda^2)$ into complex sinusoids of the form $e^{2i\lambda^2\phi}$.

If observed radiation has precisely one Faraday depth ϕ , then the polarised structure is called a *Faraday screen* and the source is called *Faraday simple*. In this degenerate case, the relationship between the polarisation angle χ and the squared wavelength λ^2 is linear:

$$\chi = \chi_0 + \phi\lambda^2, \quad (2.14)$$

Include a graphic

and the FDF is a delta distribution:

$$F(\phi) = \delta(\phi - \phi). \quad (2.15)$$

ϕ is then called the *rotation measure* (RM). If the source is not Faraday simple, then it is called *Faraday complex*, and the question of whether a source is Faraday simple or Faraday complex is called *Faraday complexity*. Until very recently, the frequency resolution of polarised surveys was insufficient to meaningfully separate most complex arrangements of Faraday depths, and so most sources were assumed to be simple and characterised entirely in terms of their rotation measure (e.g. Taylor, Stil, and Sunstrum, 2009). Advancing telescope technology and emphasis on polarisation science has opened new frontiers in spectropolarimetry and upcoming and ongoing surveys (e.g. RACS and POSSUM) will likely report Faraday complexity and produce Faraday depth catalogues instead of rotation measures.

If the polarised spectrum of a Faraday complex source is observed at multiple frequencies, then the multiple Faraday depths comprising it can be disentangled even though they spatially overlap in the radio image. This can provide insight into the polarised structure of the source as well as the intervening medium. This disentanglement is accomplished by inverting Equation 2.13, a process called *RM synthesis* (Brentjens and Bruyn, 2005):

$$F(\phi) = \int_{-\infty}^{\infty} P(\lambda^2) e^{-2i\lambda^2\phi} d\lambda^2. \quad (2.16)$$

In reality we do not observe $P(\lambda^2)$ at all wavelengths nor with infinite resolution. In RM synthesis this is accounted for by the introduction of a *weighting function* (`heald09faraday`) $W(\lambda^2)$. $W(\lambda^2)$ is nonzero if and only if an observation was taken with wavelength λ . Substituting $P(\lambda^2) \rightarrow P(\lambda^2)W(\lambda^2)$ into Equation 2.16 results in a sum which can be numerically evaluated:

$$F(\phi) \approx \int_{-\infty}^{\infty} P(\lambda^2)W(\lambda^2) e^{-2i\lambda^2\phi} d\lambda^2 = \sum_{j=1}^J P(\lambda_j^2)W(\lambda_j^2) e^{-2i\lambda_j^2\phi}. \quad (2.17)$$

$P(\lambda_j^2)$ is the observed polarisation at the j th value of wavelength, $W(\lambda_j^2)$ is the corresponding j th weight, and J is the total number of wavelengths for which measurements were taken. The weighting function W is analogous to the weighting function in radio synthesis imaging. The most common choices of W are 1) uniform weight-

weighting
by noise

ing² with $W(\lambda_j^2) = 1$ for all nonzero values, and 2).

Of course, no physical source has a precise Faraday depth, as there is always intrinsic scatter. Along the line of sight, if we assume Gaussian noise in an otherwise-constant n_e i.e. $n_e(z) \sim \mathcal{N}(\bar{n}_e, \sigma_{n_e}^2)$, and a constant B for simplicity, then we find

$$\phi \sim \mathcal{N}\left(\frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} B \bar{n}_e, \frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} B \sigma_{n_e}^2\right), \quad (2.18)$$

that is, the depth has an uncertainty proportional to the magnetic field strength and the noise in n_e . A similar result follows for noise in B only. There is no analytic solution for noise in both B and n_e , but if we approximate the integrand as a Gaussian by calculating the mean and variance we find

$$\phi \sim \mathcal{N}\left(\frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} \frac{\bar{n}_e \sigma_B^2 + \bar{B} \sigma_{n_e}^2}{\sigma_B^2 + \sigma_{n_e}^2}, \frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} \frac{\sigma_{n_e}^2 \sigma_B^2}{\sigma_{n_e}^2 + \sigma_B^2}\right). \quad (2.19)$$

We observe multiple lines-of-sight that are coalesced into one within the beam. Due to this noise, even with constant n_e and B across a source, we can see multiple Faraday depths as each line-of-sight is a sample from the above distribution.

2.3 Radio galaxies and active galactic nuclei

AGN are some of the most energetic objects in the Universe. They both provide a laboratory for extreme physics and are a key part of the life cycle of a galaxy (Heckman and Best, 2014). Powered by a supermassive black hole, they convert gravitational potential energy into intense electromagnetic radiation at a broad range of frequencies. AGN that produce strong radio emission are called radio AGN, and methods of observing the complex structures that these radio AGN form as radio galaxies are the focus of this thesis.

2.3.1 What we see when we look at AGN

Observations are the crux of astronomy. While there are many models of how AGN evolve and how they interact with their surroundings — and indeed, the actual structure of an AGN is very much an open question in astronomy — the evidence presented by observations is reliable and a good place to start discussing the structure, behaviour, and importance of AGN throughout the universe.

As powerful sources of radio emission, radio AGN and their associated extended structure can be seen throughout the universe. Sufficiently close or large radio galaxies can be resolved by telescopes and their structure examined, while more distant or smaller radio galaxies may be unresolved and point-like. A well-resolved radio

²The analogous weighting scheme in radio synthesis imaging would be natural weighting, rather than uniform—an unfortunate overlap in terminology.

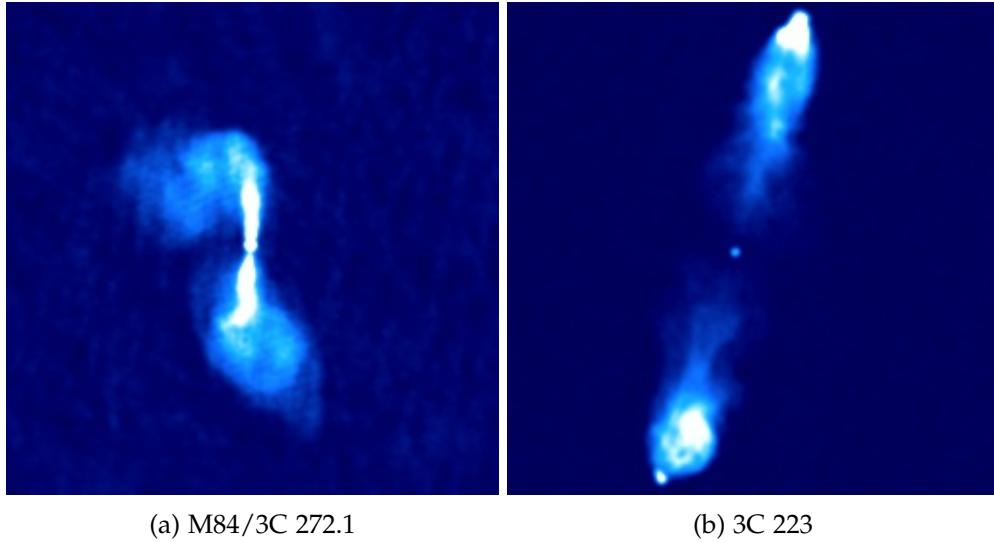


Figure 2.2: Examples of (a) a FRI (Laing and Bridle, 1987) and (b) a FRII radio galaxy (Leahy and Perley, 1991). Both are shown with an arcsinh stretch and were observed with the VLA.

galaxy can be a striking thing: from the central AGN extend two opposing, tightly-collimated jets, which widen into huge lobes of radio-bright plasma. These lobes may have further structure, particularly bright regions called *hot-spots*, and the jets and lobes may be bent and distorted as they travel away from their host galaxy. For any given radio galaxy, some of these features may or may not be present. In particular, radio galaxies are often divided into two classes based on the kinds of extended structure that are visible, called Fanaroff-Riley type I (FRI) and Fanaroff-Riley type II (FRII) radio galaxies. FRI have wavy, diffuse lobes, appearing brighter toward the host galaxy and dimming further out (e.g. Figure 2.2a). FRII, on the other hand, have long, tightly-collimated jets and sharp-edged lobes with bright hot-spots (Urry and Padovani, 1995) at the very end of the lobes, and are brighter further away from the host galaxy (e.g. Figure 2.2b). FRII are also generally higher-luminosity (Fanaroff and Riley, 1974a) than FRI, and therefore make up the majority of observed extended radio sources throughout the universe, but this is by no means the clear-cut divide it was once thought to be (Mingo et al., 2019) with the difference now being attributed largely to environmental effects rather than jet power. The current understanding is that FRII jets remain at relativistic energies up until the edge of the lobe, where they terminate in a shock that appears as a hot-spot, while FRI jets decelerate within the galaxy itself (Hardcastle and Croston, 2020a), so this sharp difference in extended structure begins with environmental interactions at the very centre of the galaxy.

A radio galaxy can be tremendously extended, with increasingly many radio galaxies being found with a length of over one megaparsec. Such large galaxies are called *giant radio galaxies*, but even non-giants are still quite big, regularly extending well outside the stellar component of the host galaxy. We will discuss the extended

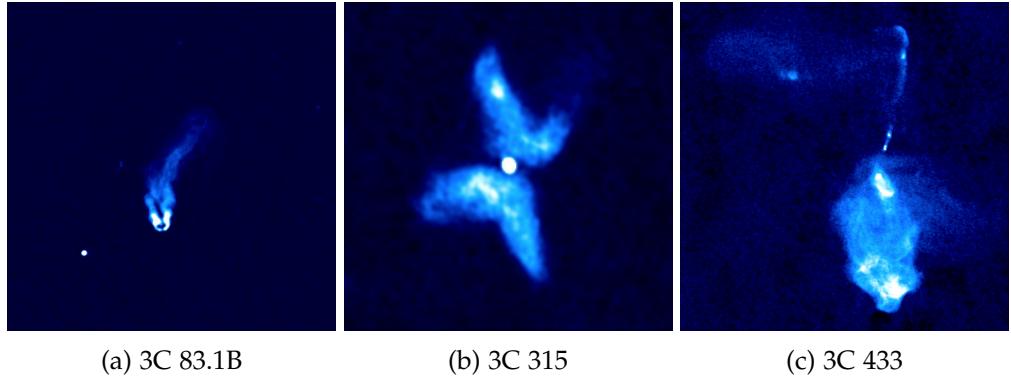


Figure 2.3: Radio galaxies, displayed with an arcsinh colour scale. All images were taken with the VLA. (a) is a narrow-angled tail radio galaxy (Leahy, Bridle, and Strom, 2020), (b) is an X-shaped radio galaxy (Leahy, Pooley, and Riley, 1986), and (c) is a very unusually-shaped radio galaxy (Black et al., 1992).

structure in Section 2.3.2.

An AGN interacts with its host galaxy, and so the host galaxy of an AGN can also provide interesting insights into the structure and behaviour of the AGN. Early research indicated that the split between FRI and FRII radio galaxies was dependent on the mid-infrared and optical brightness (and therefore density) of the host galaxy (Ledlow and Owen, 1996; Bicknell, 1995) though more recent work suggests this may not be a strong effect if it exists at all (Hardcastle and Croston, 2020a).

2.3.2 Extended structure

The jets and lobes of AGN can be very extended, with the largest known radio galaxies measuring over 4 Mpc end-to-end (Machalski et al., 2011). This is much larger than the radius of their host galaxies, and so the jets and lobes of AGN are uniquely posed to interact with the local environment. Environmental interactions both within and outside the host galaxy warp and distort the jets and lobes. Within the galaxy, the jets drive a bubble of energy in the interstellar medium (ISM; Mukherjee et al., 2016), transferring energy into the ISM with different effects depending on the jet power (Mukherjee et al., 2018); the ISM on the other hand suppresses the jets and distorts them to varying amounts depending on the degree of interaction (Mukherjee et al., 2018). Outside the galaxy, the jets and lobes are bent by the intra-cluster medium and neighbouring galaxies (ICM; Garon et al., 2019a; Rodman et al., 2019a) and this structure may even be used as a probe for cluster environments (Banfield et al., 2016; Sakelliou, Hardcastle, and Jetha, 2008).

The strong interaction of AGN with their environment leads to a great variety of exotic-shaped radio galaxies. Some morphological classes of this ‘radio galaxy zoo’ include X-shaped galaxies, which have two sets of lobes roughly perpendicular to each other; wide- and narrow-angled tail galaxies, which are bent about the core with large and small angles respectively; head-tail galaxies, which are so bent that

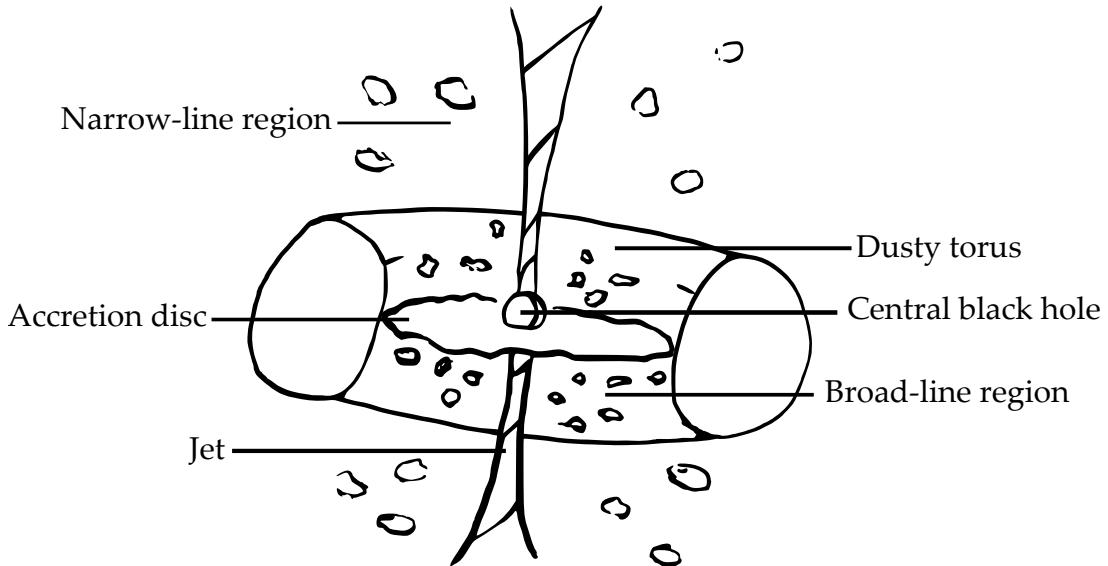


Figure 2.4: The unified model of AGN.

the two lobes seem to be the same or nearly the same; double-doubles, which have two sets of lobes on each side; and many, many more. Some examples of radio galaxies with interesting structure are shown in Figure 2.3. Large-scale automated identification of these galaxies can be tricky owing to their variety, extent, and often disconnected structure.

AGN cores tend to have flat or inverted spectral indices around -0.5–1 [citation needed](#). Moving out from the host galaxy, the spectral index steepens as the electrons are older and less energetic, with the spectral index of the lobes usually at about -0.7 [citation needed](#). The hotspots of FRII galaxies have spectral indices between -0.5 – -0.7, becoming shallower as the electrons reaccelerate [citation needed](#). The jets do not strongly emit and are only detectable for particularly deep observations or nearby radio galaxies [citation needed](#).

2.3.3 The unified model

At their core, AGN are an accreting *supermassive black hole*: a body so dense that even light cannot escape its gravitational pull, with mass on the order of 10^7 – $10^9 M_\odot$ [citation needed](#). Such black holes seem to exist at the centre of galaxies [citation needed](#) and these galaxies are called *host galaxies*. The current understanding of the structure of an AGN is as follows (Urry and Padovani, 1995). The black hole is surrounded by an accretion disc emitting in ultraviolet and X-ray. Beyond this is the broad-line region, named for the Doppler-broadened emission lines emitted by the energetic clouds of material surrounding the accretion disc. The broad-line region and accretion disc are themselves surrounded by a dusty torus (or some other disc-like structure) which prevents light from the centre of the AGN being observed from the sides. Further still from the accretion disc is the narrow-line region, where lower-energy gas produces

narrow emission lines. From either side of the disc, an AGN produces two collimated outflows of relativistic plasma called jets, and these jets may interact with gas in the host galaxy to produce bright radio emission. The jets are not always visible. As the jets disperse further out from the centre of the AGN they widen into plumes of plasma known as *lobes*. This model of AGN unifies different observed classes of AGN by their orientation and luminosity, and is hence known as the *unified model* (Antonucci, 1993). Recent work suggests that the unified model of AGN is not the full story (e.g. Zhuang and Ho, 2020).

Explain alternatives to the unified model?

There are many different ways to divide the set of radio AGN into classes. By morphology, radio AGN are often divided by the structure of the jets and lobes, with FRI and FRII the most striking examples. AGN can also be divided into *radiative-mode* and *jet-mode* by how they expel their energy (Heckman and Best, 2014). Radiative-mode AGN produce radiative energy in amounts higher than 1 percent of their Eddington limit, while jet-mode AGN mainly output energy through their jets. The Eddington limit describes the maximum luminosity that a compact object can emit, and is given in Equation 2.20 (Rybicki and Lightman, 1979):

$$L_{\text{Eddington}}(M) = \frac{4\pi GMm_p c}{\sigma_T}. \quad (2.20)$$

Optical emission observed near the centre of the AGN can be used to divide radio AGN into broad-line and narrow-line galaxies. The former have broad spectral lines while the latter have narrow spectral lines, with broader spectral lines indicative of higher thermal energies. The most common interpretation, under the unified model, is that broad-line AGN are those seen end-on and narrow-line are those seen edge-on with the dusty torus obscuring the broad-line region. These narrow-line galaxies are usually the only ones for which we see significant extended structure.

include an alternative explanation

2.3.4 Polarised structure

The magnetic field of AGN is thought to be critical to their structure (Sikora and Begelman, 2013). A strong magnetic field is required to eject and collimate the jets (Lovelace, 1976) and the magnetic environment influences the structure of the jets (O'Sullivan et al., 2015). Polarisation provides a probe for measuring this magnetic field. Radiative- and jet- mode AGN have different fractional polarisations, with jet-mode AGN having a much wider range of fractional polarisations ($p \sim [0, 30]$ per cent) compared to radiative-mode AGN (limited to $p \lesssim 15$ per cent), with this difference attributable to the magnetic environment (O'Sullivan et al., 2015). Steep-spectrum ($\alpha > 0.5$) and flat-spectrum ($\alpha < 0.5$) AGN have differing fractional polarisations, with steep-spectrum sources having much higher fractional polarisation for frequencies > 5 GHz and flat-spectrum sources having higher fractional polarisation for frequencies < 1 GHz due to frequency-dependent depolarisation of the steep-spectrum sources (Saikia and Salter, 1988). Hotspots of FRII radio galaxies have low polarisation (< 10 per cent) while the more diffuse sections may have much greater polarisation (> 20 per cent) (Saikia and Salter, 1988). The direction of the magnetic

field is correlated with the direction of the total intensity of the source (Saikia and Salter, 1988).

2.3.5 The role of AGN

How do AGN tie into galaxy evolution and feedback? Why do we need them in our models and theories? Why are they important?

2.4 Classifying AGN

As discussed in Section 2.3.3, radio galaxies fall into many classes. Understanding the mechanisms underlying these class distinctions is critical to understanding AGN. As we have no way to directly see the core of an AGN (it's far too small to resolve at the distances AGN occur and may also be occluded), our only method to investigate AGN is by looking at their large-scale behaviour. Some classes may relate to the fundamental AGN core, some may be environmental, and some may be due to observation effects. Much of our knowledge about AGN (such as the unified model) come from analysing these classes and their differences. To investigate classes of AGN then a large sample of each class is required, and source classification approaches can divide a large dataset from a radio survey into useful subsets. Knowing what class a source is may also help analyse its properties as we can estimate its expected behaviour, perhaps with the aid of models and simulations. Some classes may have interesting structure or properties that can only be observed with additional detailed observations, so identifying which sources require follow-up is a tightly related problem in radio astronomy. An excellent, though now somewhat dated, summary of radio source classification is the review paper by Urry and Padovani (1995), which we recommend for further reading.

Deciding which class a given radio galaxy falls into may be challenging, and doing this automatically even more so. This section discusses approaches to classifying radio galaxies.

2.4.1 Statistical and manual classification of AGN

Manual and statistical approaches to classifying AGN have dominated the radio astronomy literature until very recently, due to the comparative lack of computational power as well as a lack of good automated methods. Manual methods amount to examining the structure of a resolved source and determining its class: this is how we usually identify bent radio galaxies, head-tail radio galaxies, X-shaped radio galaxies, and those radio galaxies with more unusual morphologies. Statistical approaches identify properties of the source that can be combined and thresholded to separate the sources into categories en masse. Modern machine learning techniques for classification of radio sources can be thought of as an extension of these statistical methods, where the properties and their combinations are identified automatically, but we will discuss these separately in Section 2.4.2.

Arguably the most well-known radio classification scheme, FRI and FRII, was originally defined on well-resolved radio galaxies by computing the ratio of the distance between the regions of highest brightness on opposite lobes and the total extent of the radio emission (Fanaroff and Riley, 1974b). Sources with a ratio under 0.5 were called FRI and those with a ratio greater than 0.5 were called FRII. This classification has over time evolved into a less precise divide, with classification generally now morphological and based on the structure (diffuse, wavy plumes versus hotspots and lobes for FRI and FRII respectively). The FRI and FRII divide has been further complicated by other related categorisations such as the so-called "Fanaroff-Riley type 0" sources which seem to be the lower end of a continuum of radio sources with diffuse plumes (Garofalo and Singh, 2019; Capetti et al., 2020). Many classes are defined by explicitly statistical means; for example, steep- and flat-spectrum sources are divided by spectral index at $\alpha = 0.5$ (Urry and Padovani, 1995). For convenient analysis, radio sources are often also grouped into "observational" classes that don't have a physical analogue based on their apparent structure, e.g. the GLEAM survey classifies radio sources into the number of apparent components, which is highly dependent on the observational parameters (White et al., 2020).

More unusual or more loosely defined classes, such as X-shaped radio galaxies and giants, have often been identified by manual searches through large datasets, e.g. Cheung (2007) and Dabhade et al. (2020) and notably the recent ROGUE I catalogue of 32 616 morphologically classified radio galaxies (Żywucka, Koziel-Wierzbowska, and Goyal, 2020). These searches are often aided by computer algorithms (e.g. Proctor, 2011; Dabhade et al., 2020).

Radio sources are also more generally classified, such as into AGN or non-AGN emission (Koziel-Wierzbowska et al., 2020), often using optical emission lines or optical/infrared magnitude.

2.4.2 Machine learning classification of AGN

Machine learning based approaches for radio source classification are rapidly evolving as the amount of radio data available through big surveys increases. Advances in tooling, such as the wide availability of hardware-accelerated automatic differentiation software, have also contributed to an explosion in machine learning applications in astronomy by making machine learning techniques more available to astronomy researchers.

Morphological classification of galaxies with machine learning began in optical astronomy, probably due to the large sample sizes of well-resolved galaxies previously available. The earliest such paper is likely the application of neural networks to the task by Storrie-Lombardi et al. in 1992. From here, the field applied other classification algorithms such as decision trees (e.g. Owens, Griffiths, and Ratnatunga, 1996). The Sloan Digital Sky Survey (SDSS) brought an explosion of new data in 2003, and new experiments in classification soon followed (e.g. Ball et al., 2004; Ball et al., 2006). The Galaxy Zoo project leveraged hundreds of thousands of volunteers to produce an astonishingly large set of labelled optical galaxies from SDSS and sub-

sequent papers used this as a training set for machine learning methods (Banerji et al., 2010; Dieleman, Willett, and Dambre, 2015; Zhu et al., 2019).

While machine learning has been used in radio astronomy for some time (e.g. the NVSS used neural networks to detect sidelobes; Condon et al., 1998a) its first application to radio source classification was most likely to identifying quasar candidates in FIRST (Carballo, Cofiño, and González-Serrano, 2004). Proctor (2006) applied decision tree ensembles to identify bent double morphologies in FIRST, manually selecting features to characterise radio sources, while Bastien, Oozeer, and Somanah (2017) used shapelet analysis to obtain features to feed into their decision tree ensembles. 2011–12 marked a revolution in computer vision with the discovery that deep convolutional neural networks (known as early as 1989, see LeCun et al., 1989), boosted dramatically by widely available training data generated by the internet and a huge increase in computational power from GPUs, could achieve greater-than-human performance on image classification tasks. Deep neural networks have since found use for morphological classification of radio sources, such as FRI vs FRII (Aniyan and Thorat, 2017; Ma et al., 2019b; Lukic et al., 2019; Tang, Scaife, and Leahy, 2019; Samudre et al., 2020; Bowles et al., 2020; see also Ma et al., 2018), compact vs extended sources (Lukic et al., 2018; Alhassan, Taylor, and Vaccari, 2018; Lukic et al., 2019), and observational classes (Galvin et al., 2019a; Ralph et al., 2019).

There are also many works on classification of radio sources besides morphology. Machine learning has been applied to AGN classification tasks including blazar classification (Arsioli and Dedin, 2020) and radio loudness (Beaklini et al., 2020). Deep learning is also prevalent on this topic, with deep learning finding applications in Faraday complexity classification (Brown et al., 2018) and notably in transient detection (Connor and Leeuwen, 2018; Guo et al., 2019; Wang et al., 2019; Agarwal et al., 2020; Lin, Li, and Zeng, 2020; Zhang et al., 2020; Balakrishnan, Champion, and Barr, 2020).

It is worth contrasting these machine learning approaches with non-machine learning automated approaches, as the two are often conflated in the literature. Mingo et al. (2019), for example, use an automated version of detecting the brightness gradient of extended radio sources to determine whether they are FRI or FRII en masse and apply this approach to the LoTSS survey. Segal et al. (2019) apply an information theoretic approach to estimating morphological complexity of a source. The key difference between a machine learning automated approach and a non-machine learning automated approach is that the former has the capacity to change its behaviour based on available data, while the latter does not — though note that this is not necessarily a bad thing.

2.5 Aggregating Radio Components

Unlike galaxies observed at other wavelengths, the emission from radio galaxies can be disconnected when observed: a single radio galaxy may appear in observations as multiple discrete components. This is partly due to inhomogeneous emission over



(a) The 64m telescope (Murriyang) at Parkes Observatory

(b) ATCA near Narrabri

Figure 2.5: (a) A single-dish telescope and (b) an array. Images: CSIRO.

the radio galaxy structure — e.g. FRII hotspots can be much brighter than the rest of the galaxy, so they might be visible while the rest of the galaxy is too faint to see — and partly due to the technique through which many radio observations are made, *interferometry*, which may screen out diffuse emission.

This potential of a radio galaxy to be split into multiple discrete blobs of emission also leads to a linguistic ambiguity not present at other wavelengths. A *radio source* may be either a single physical object that emits radio, or a single blob of disconnected radio emission. The same is true for *radio object*. There is no agreed-upon terminology. We will therefore adopt the following terminology from this point on, adapted from my paper (Alger et al., 2018) based on the Radio Galaxy Zoo terminology: a *radio source* or *radio galaxy* refers to all radio emission observed that comes from a single physical object, and a *radio component* refers to a single, contiguous region of radio emission above the noise level of an observation. We will avoid using *radio object*. Note that in the literature a *radio component* may be either a Gaussian fit to an observation or it may be a region of connected emission; in works where the former is the case (particularly in source-finding literature) the latter is often called a *radio island*.

Since radio galaxies can appear disconnected, aggregating observed radio components into physical sources is integral to understanding radio galaxies. Aggregating disparate components into a single source is important for two reasons. Firstly, without all components, the total luminosity of a source is dramatically underestimated. Secondly, we need all components to accurately examine the morphology of the source.

2.5.1 Missing emission in radio observations

Radio observations may be made either with *single-dish telescopes*, like the famous Parkes Radio Telescope (Murriyang), or *radio arrays*, like the Australia Telescope Compact Array (ATCA), both shown in Figure 2.5. Both have their advantages.

Single-dish telescopes are able to measure absolute brightnesses (while arrays can only measure relative brightnesses, and must therefore be calibrated to a source of known brightness). Interferometric arrays can achieve incredibly high resolution, as the resolution is inversely proportional to the distance between the most distant array elements (while the resolution of single dish telescopes is inversely proportional to the diameter of the dish).

Radio telescopes can be thought of as sampling the *u-v plane*, the Fourier transform of the sky. The *u-v* plane is perpendicular to the line-of-sight. Each pair of antennae in an array samples two points on this plane, each corresponding to the vector between the antennae projected onto the *u-v* plane, called a *baseline*. Longer baselines therefore correspond to higher (spatial) frequencies, which is why long baselines provide high resolution. Diffuse emission is characterised mainly by low (spatial) frequency components, while compact emission is characterised by a broad range of frequency components, so diffuse emission both a) takes up less space on the *u-v* plane than compact sources and b) occupies spaces much closer to the origin on the *u-v* plane. Some intuition on this can be obtained by examining the Fourier transform of a 2D Gaussian:

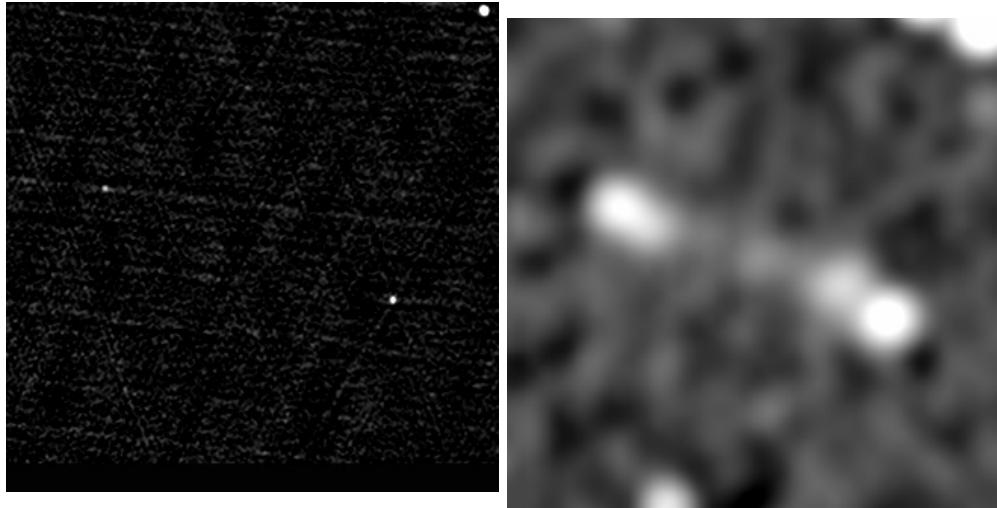
$$\mathcal{F}_{x,y} \left[\frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \right] = e^{-2\sigma^2\pi^2(u^2+v^2)}. \quad (2.21)$$

From this equation we can see that the Fourier transform of a fairly *compact* Gaussian would be quite broad, taking up many frequencies in the *u-v* plane, while a very *diffuse* Gaussian would have a very narrow Fourier transform. The upshot of this is that long baselines sacrifice sensitivity to diffuse emission for high resolution. Single-dish radio telescopes are unable to make the same tradeoff, as they are only able to sample a disc centred on the origin³. This loss of diffuse emission is often called *resolving out*. An example of this is shown in Figure 2.6, where (a) and (b) are the same radio galaxy observed with the same telescope, the Very Large Array, with the only difference being that the VLA was in B configuration for (a) and the D configuration for (b). The B configuration moves the antennae of the VLA far apart, while the D configuration keeps them close together.

2.5.2 Methods of aggregation

Most large radio catalogues are of *components* rather than sources, most likely because manual methods remain the best way to aggregate components into sources. Many component catalogues have source catalogues of some interesting subset later derived from them by manual inspection (e.g. Dabhadé et al., 2017). Such manual searches are often assisted by automated methods (e.g. Proctor, 2011). The Radio Galaxy Zoo citizen science project has also, with the help of over 10 000 volunteers,

³This is, incidentally, why single-dish telescopes can measure the absolute brightness while arrays cannot: there is no way to measure the origin in the *u-v* plane as there is no way for two array antennae to be infinitely close together (forming a zero-length baseline), and the origin contains the absolute brightness information much like how the centre of a Fourier transform contains the mean.



(a) FIRST (Becker, White, and Helfand, 1995) (b) NVSS (Condon et al., 1998b)

Figure 2.6: A fairly diffuse FRII, J0016+0420, observed with the VLA in the (a) FIRST and (b) NVSS surveys. (GRG1 from Dabhade et al., 2017)

aggregated over 75 000 sources from FIRST (Wong et al. in prep) and ATLAS (Banfield et al., 2015), which is one of the largest (if not the largest) extant catalogues of manually aggregated sources.

Automated methods have been developing steadily as data volumes have grown. A Bayesian approach, fitting models of sources to component locations, was introduced by Fan et al. (2015) and later expanded upon (Fan et al., 2020). Machine learning methods have also become increasingly popular. Proctor (2016) aggregated radio components using a decision tree classifier and used the results to search for giant radio galaxies (see also Proctor, 2011; Proctor, 2006). Wu et al. (2019) applied a deep learning model to identify radio sources from an image of radio components, training this model on the Radio Galaxy Zoo FIRST dataset. Most recently Galvin et al. (2020) used a nested self-organising map, a kind of neural network, to cluster similar sources and then aggregated them into sources using manual labelling of the clusters and some heuristics.

Any method for cross-identification of components can also be used to aggregate components. Cross-identify all components with their host galaxies, then components that share a host are almost always part of the same radio source. It is this approach we will later take to aggregating sources in Chapter 5. There is no obvious inverse algorithm (to go from aggregate sources to cross-identifications) although some algorithms attempt to solve both problems simultaneously (Fan et al., 2020; Fan et al., 2015).

Machine Learning and Astroinformatics

I'll need to explain the purpose of this chapter and give a brief outline. There'll be three sections, and I'll briefly summarise them here.

- machine learning in astronomy and radio astronomy (literature review)
- summarising the basics of machine learning (later sections of course complicate these):
 - predictors
 - features
 - labels
 - optimisation
 - data
 - expertise
- specific predictors
- features:
 - feature selection and design, and the benefit of domain knowledge
 - feature extraction
 - neural networks
- labels:
 - training labels and testing labels
 - where do labels come from? domain experts, crowdsourcing, citizen science, simulations...
 - label noise and unreliable annotators
 - multiple unreliable annotators

- citizen science:
 - citizen science in astronomy
 - designing good citizen science
 - benefits and challenges of citizen science
 - challenges in ML on crowdsourcing/citizen science
 - rant about how citizen science is different to crowdsourcing from an ML perspective

Machine learning was once described to me by an anonymous supervisor as “the statistics kept at the back of the textbook”. But even accepting its grounding in statistics, is this really an accurate description of the field? I think of machine learning as a data-driven way of formalising predictive problems mathematically, converting between different kinds of statistical problems, and an accompanying set of methods and practices for handling data and uncertainty. The eventual goal is to design some method or algorithm that automatically discovers useful information in (potentially very large) data sets. There are three core components of machine learning: the data, the model, and learning (Deisenroth, Faisal, and Ong, 2020). Before discussing these, we will look at the kinds of problems that machine learning solves.

3.1 Prediction

Machine learning aims to solve *prediction tasks*: problems where we have some data and we seek some kind of output based on that data. Central to prediction tasks are predictors.

3.1.1 Predictors

A *predictor* is some function that produces an output from some given input. A predictor can be represented as a function or as a probabilistic model, depending on the machine learning approach being undertaken. As a function, a predictor maps from some input domain \mathcal{X} into some output domain \mathcal{Y} , and is usually written as

$$f : \mathcal{X} \rightarrow \mathcal{Y}. \quad (3.1)$$

\mathcal{X} and \mathcal{Y} are commonly (but certainly not always) a real vector space \mathbb{R}^n . Because the goal of machine learning involves *finding* a suitable function f for the task at hand, the set of functions is usually constrained. For example, if $\mathcal{X} = \mathbb{R}^n$, we might require that f is a linear function $\mathbb{R}^n \rightarrow \mathbb{R}$, easily parametrised by $n + 1$ constants. This constraint is called a *model*.

As a probabilistic model, a predictor is a joint probability distribution between observations and hidden parameters (Deisenroth, Faisal, and Ong, 2020). Using a probabilistic predictor allows us to formally describe and work with uncertainty both in the input space and output space. Such a predictor is usually parametrised

by a finite set of parameters, which already includes most common probability distributions.

We will generally assume that our data are generated from some unobserved, true function called the *groundtruth*. This might be a physical process, or a complicated sampling function from some unknown vector space. The assumptions we make on this generative function can greatly change the way we approach machine learning problems.

In some sense, the goal of machine learning is to identify a good predictor from within the space of all possible predictors. Of course, this begs the question: what is a ‘good’ predictor? We will return to this when we discuss learning, but for now, a good predictor is one that well-approximates the groundtruth.

3.1.2 Classification

Classification is the machine learning task of predicting discrete, unstructured values (Deisenroth, Faisal, and Ong, 2020). These values are called *classes*. Classification is arguably the most important prediction task, as many other problems can be formalised as classification. Astronomy has its fair share of classification tasks, from classical astronomy tasks like galaxy morphology classification (appearing in machine learning literature as e.g. Dieleman, Willett, and Dambre, 2015) to transient detection [citation needed](#).

A classification problem seeks a predictor where \mathcal{Y} is a finite, discrete set of classes. Classification tasks are usually delineated by the number of classes: much like astronomy’s fascination with metals, classification tasks either have two classes, or more than two classes. The former are called *binary classification* tasks and the latter are *multiclass classification* tasks. The reason for this split is that binary classes are dramatically easier to reason about and analyse, and many special cases exist for binary where they do not for multiclass.

The two most common representations of \mathcal{Y} for a binary task are $\mathcal{Y} = \{0, 1\}$ and $\mathcal{Y} = \{-1, 1\}$. In both cases 1 is called the *positive class*; 0 and -1 are both called the *negative class*. Throughout this thesis we will maintain the former convention with 0 and 1 as the negative and positive classes.

Many different tasks can be formalised as classification. An easy way to see why this is the case can be found by taking any prediction problem $\mathcal{X} \rightarrow \mathcal{Y}$ and reinterpreting it as the binary classification problem $\mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$, i.e. instead of taking an input and predicting an output, take an input and a potential output and determine if they should be related. Of course this is not always the most efficient way to solve a prediction problem (and in the general case described here can be so inefficient as to be unsolvable) but the many known properties of classification make it an appealing framework to cast problems into. In Chapter 4, I will cast the radio astronomy problem of cross-matching galaxies seen in different wavelengths into a binary classification problem.

3.1.3 Regression

The other main kind of supervised prediction task is *regression*, which is the machine learning task of predicting ordered (and usually continuous) values. In a regression problem, we seek a predictor where \mathcal{Y} is a set of ordered values, usually a subset of \mathbb{R}^k for some positive natural k .

3.2 Data and representation

Machine learning is centred on data and the extraction of information from that data. Data can include anything from numeric information, documents, or images, to spectra or galaxies. A collection of data is called a *dataset* and an element of this dataset is (interchangably) called an *example* or *instance*. Generally, data are not easy to work with in their original form and must be converted into a numerical representation before use. As it is relatively easy to work with both numerically and analytically, we usually convert our data into real vectors in \mathbb{R}^n . Each axis of this vector space is called a *feature* and the space as a whole is called the *feature space*. Features are non-trivial to choose, and finding good features often requires the expertise of a human who is well-versed in the original dataset (a *domain expert*). The process of finding features is called *feature selection*, or *feature design*.

What makes a feature good? Intuitively, we want to transform our data into a space where it is easy to work with: a space where properties we care about are obvious, or easy to extract. For this reason, features will vary tremendously depending on the problem being faced, and the same data may be represented in many different ways. Much of early machine learning literature focused on good methods for automatically developing features (generally called *feature extraction*), and much early *applied* machine learning focused on identifying these features manually so that good predictors could be easily found. An astronomical example is Proctor (2006), who developed features for representing radio galaxies for the purpose of sorting them. State-of-the-art models like deep neural networks (e.g. Dieleman, Willett, and Dambre, 2015) can be viewed as developing their own task-specific features as part of their training. A good feature space will have a structure that reflects the components of the intrinsic structure of the input data which are useful for the prediction task at hand. Good features may also be useful in other related tasks, such as dataset exploration, dataset visualisation, or other prediction tasks. Chapter 6 largely focuses on finding good features for identifying Faraday complexity in polarised sources.

Another very important piece of the machine learning puzzle are *labels*. Training a predictor with supervised learning requires some known pairs of inputs and outputs, and the known outputs are called labels. Like features, labels also need to be encoded in some way, and this depends on the specific task. Much like features, we want to embed the labels into a space which is easy to work with and has a meaningful structure. For problems where we know the outputs we wish to obtain, this can be a lot simpler than feature selection. For example, a binary classification problem will have only two possible outputs. Common representations for this would be $\{0, 1\}$ as

described in Section 3.1.2, but we could also represent the labels as $\{[1, 0]^T, [0, 1]^T\}$, called a *one-hot encoding*. The advantage of the former is its simplicity and ease of integration into binary classification equations, but the advantage of the latter is that it is easily extended into multiclass classification without imposing order on the classes. Despite being simpler to encode, labels can carry a lot more difficulty than features due to their comparative rarity: in essence, features are cheap and labels are expensive. We will discuss labels in more detail in Section 3.5.

3.3 Training

Training a model is the process of using data to find a good predictor that fits the model’s constraints. This is generally achieved by minimising a *loss* (also called *error* or *cost*) function over the model. In this section I will introduce loss functions and a common method of optimising parametrised models.

maximum likelihood?

3.3.1 Loss functions

Put simply, a loss function is a function of a predictor and a dataset which describes how good the predictor is at predicting that dataset. Loss functions are high-valued for a predictor that poorly describes the dataset, and are low-valued for a predictor that well-describes the dataset. Usually the loss is minimised at zero, when the predictor perfectly captures the dataset (though whether this is a desired result is another question).

What should the loss function be for a given problem? The answer is not always obvious. Take for example a binary classification problem. The “obvious” loss would be the complement of the accuracy: the rate at which the predictor incorrectly guesses the label. This is easy to compute and we would like our predictor to have a high accuracy. But this is not a good choice: it is tremendously hard to work with as it takes on discrete values, because the accuracy is the number of correct predictions divided by the total number of examples. It is hard to motivate with probabilistic arguments. Finally, it is unclear how the accuracy should work in the case of a probabilistic model.

Instead, the loss function is usually derived by making assumptions on the structure of the data and task. The main assumption we usually make is that data are drawn *independently and identically distributed* (IID), that is, each example is drawn from the same distribution and is not dependent on any other examples. We also assume a structure of the noise in the observed labels: training data are almost never completely accurate, and so there will be intrinsic noise in the distribution of labels about their unobserved “true” value. To demonstrate these assumptions, we will now derive loss functions for regression and binary classification.

this is
clunky and
I think mis-
represents
the noise

3.3.1.1 Loss function for regression

To derive a loss function for regression, let us assume that our labels are a random variable y modelled by a predictor $y = f(x)$. Further, let us assume that a predicted y is normally distributed about its true value, i.e.

$$y \sim \mathcal{N}(y | \mu, \sigma^2) \quad (3.2)$$

for the true mean μ and standard deviation σ . Under this assumption the probability that y is equal to a target t given an example x is

$$p(y = t | x) = \mathcal{N}(t | f(x), \sigma^2). \quad (3.3)$$

What would the probability be of observing a set of targets $T = \{t_1, \dots, t_n\}$ given corresponding examples $X = \{x_1, \dots, x_n\}$? Letting $Y = \{y_1, \dots, y_n\}$ be random variables like y , the joint probability distribution $p(Y = T | X)$ is

$$p(Y = T | X) = \prod_{i=1}^n p(y_i = t_i | x_i) \quad (3.4)$$

by using our independence assumption. We can then substitute the normal distribution:

$$p(Y = T | X) = \prod_{i=1}^n \mathcal{N}(t_i | f(x_i), \sigma^2). \quad (3.5)$$

We would like to maximise this probability over f , which is called a *maximum likelihood* approach to finding a predictor. It is, however, not very easy to work with in this current form. Maximising the probability above is equivalent to minimising its negative logarithm, so we write:

$$\mathcal{L}(f; T, X) = - \sum_{i=1}^n \log \mathcal{N}(t_i | f(x_i), \sigma^2) \quad (3.6)$$

where \mathcal{L} is the *negative log-likelihood*, a loss function. We can simplify this dramatically by cancelling the logarithm and the exponential within the normal distribution:

$$\mathcal{L}(f; T, X) = \sum_{i=1}^n \frac{(t_i - f(x_i))^2}{2\sigma^2} \quad (3.7)$$

and by noting that arbitrary scaling of \mathcal{L} does not change the minimising f we can scale \mathcal{L} by σ^2 and arrive at the *sum-of-squares error*, also known as the *least-squares error*, the most common loss function for regression:

$$\mathcal{L}(f; T, X) = \frac{1}{2} \sum_{i=1}^n (t_i - f(x_i))^2. \quad (3.8)$$

The factor of half helps keep the derivative tidy:

$$\frac{d\mathcal{L}}{d\theta}(f; T, X) = \sum_{i=1}^n (t_i - f(x_i)) \frac{df}{d\theta}(x_i). \quad (3.9)$$

Here I would derive binary cross-entropy loss and arrive at the following result

3.3.1.2 Loss function for binary classification

$$\mathcal{L}(f; T, X) = \sum_{i=1}^n (1 - t_i) \log(1 - f(x_i)) + t_i \log f(x_i). \quad (3.10)$$

$$\frac{d\mathcal{L}}{d\theta}(f; T, X) = \sum_{i=1}^n -\frac{1 - t_i}{1 - f(x_i)} \log(1 - f(x_i)) + t_i \log f(x_i). \quad (3.11)$$

3.3.2 Gradient descent

Given a loss function and a parametrised model, how can we find parameters for the model that minimise the loss function? There are many optimisation strategies but if both the loss function and model are differentiable with respect to the parameters then we can employ a particularly efficient approach: *gradient descent*. Assume we have a model $f(x; \vec{w})$ parametrised by some vector \vec{w} and a loss function $\mathcal{L}(\vec{w}; T, X)$. Then the value of \vec{w} after the $k+1$ th update of gradient descent is

$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \lambda \nabla_{\vec{w}} \mathcal{L}(\vec{w}^{(k)}; T, X) \quad (3.12)$$

where $\lambda > 0$ is a small scalar called the *learning rate*. With an appropriately small choice of λ \vec{w} will converge to a local minimum of \mathcal{L} . Many variations on this concept exist which attempt to avoid local minima.

continue talking about this a bit

3.4 Models

This section describes some common models for classification. There are a plethora of different classification models and variations on these models, but I will present here only those relevant to this thesis: logistic regression, decision tree ensembles, and neural networks. These are, not coincidentally, also the most common models in astroinformatics. Logistic regression provides reliable and interpretable results. Decision tree ensembles are a fantastic off-the-shelf choice which work on a large variety of datasets. Neural networks have proved extremely effective for a wide variety of tasks, especially in computer vision.

3.4.1 Logistic regression

Logistic regression is a linear, binary, probabilistic classifier. Linear classifiers can only separate classes using a hyperplane in the feature space, with objects on one side of the plane being assigned to one class and objects on the other side being assigned

to the other. A binary classifier works on binary classification tasks. Non-binary generalisations exist, which we will briefly touch on later. Probabilistic classifiers, as discussed in Section 3.1.1, have outputs interpretable as class probabilities.

Logistic regression in a d -dimensional feature space is parametrised by a *weights vector* $w \in \mathbb{R}^d$. Given a set of features $x \in \mathbb{R}^d$, logistic regression is:

$$f(x; w) = \sigma(w^T x) \quad (3.13)$$

where σ is the *logistic function* or *sigmoid*:

$$\sigma(a) = \frac{1}{1 + e^{-a}}. \quad (3.14)$$

The output of logistic regression applied to an example x is the probability that x is in the positive class. σ , and thus logistic regression, has a domain of $(-\infty, \infty)$ and a range of $(0, 1)$. This enforces the output to be like a probability. $w^T x$ defines a d -dimensional hyperplane, called the *separating hyperplane* or *decision surface*.

Logistic regression is differentiable, which allows us to optimise its parameters w using gradient descent.

3.4.2 Decision tree ensembles

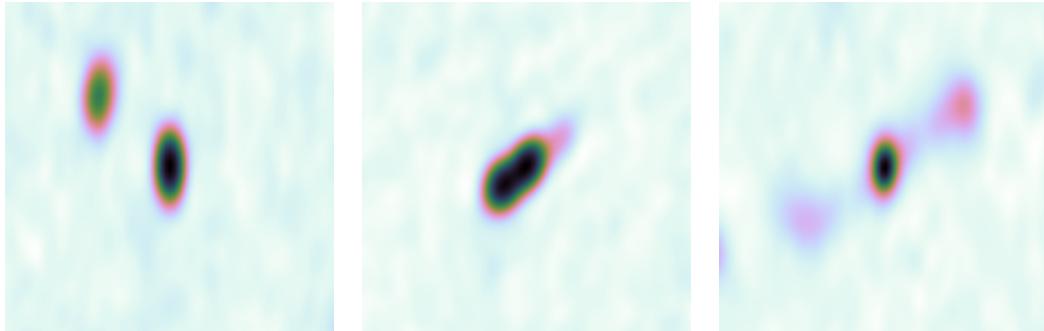
3.4.3 Convolutional neural networks

3.5 Labels

Radio Cross-identification

This chapter is based on my paper *Radio Galaxy Zoo: Machine learning for radio source host galaxy cross-identification*, by M. J. Alger, J. K. Banfield, C. S. Ong, L. Rudnick, O. I. Wong, C. Wolf, H. Andernach, R. P. Norris, and S. S. Shabala; published in the *Monthly Notices of the Royal Astronomical Society* in 2018. Some minor changes have been made to match the rest of this thesis.

In this chapter, we consider the problem of determining the host galaxies of radio sources by cross-identification. This has traditionally been done manually, which will be intractable for upcoming and ongoing wide-area radio surveys like the Evolutionary Map of the Universe (EMU). Automated cross-identification will be critical for these future surveys, and machine learning may provide the tools to develop such methods. We applied a standard approach from computer vision to cross-identification, introducing one possible way of automating this problem, and explored the pros and cons of this approach. We applied our method to the 1.4 GHz Australian Telescope Large Area Survey (ATLAS) observations of the *Chandra* Deep Field South (CDFS) and the ESO Large Area ISO Survey South 1 (ELAIS-S1) fields by cross-identifying them with the *Spitzer* Wide-area Infrared Extragalactic (SWIRE) survey. We trained our method with two sets of data: expert cross-identifications of CDFS from the initial ATLAS data release and crowdsourced cross-identifications of CDFS from Radio Galaxy Zoo. We found that a simple strategy of cross-identifying a radio component with the nearest galaxy performs comparably to our more complex methods, though our estimated best-case performance is near 100 per cent. ATLAS contains 87 complex radio sources that have been cross-identified by experts, so there are not enough complex examples to learn how to cross-identify them accurately. Much larger datasets are therefore required for training methods like ours. We also showed that training our method on Radio Galaxy Zoo cross-identifications gives comparable results to training on expert cross-identifications, demonstrating the value of crowdsourced training data.



(a) Two compact components, each a compact source.
 (b) One resolved component and resolved source.
 (c) Three resolved components comprising one resolved source.

Figure 4.1: Examples showing key definitions of radio emission regions used throughout this chapter. Compact and resolved components are defined by Equation 4.4.

4.1 Introduction to cross-identification

Next generation radio telescopes such as the Australian SKA Pathfinder (ASKAP; Johnston et al., 2007) and Apertif (Verheijen et al., 2008) will conduct increasingly wide, deep, and high-resolution radio surveys, producing large amounts of data. The Evolutionary Map of the Universe (EMU; Norris et al., 2011a) survey using ASKAP is expected to detect over 70 million radio sources, compared to the 2.5 million radio sources currently known (Banfield et al., 2015). An important part of processing these data is cross-identifying observed radio emission regions with observations of their host galaxy in surveys at other wavelengths.

In the presence of extended radio emission cross-identification of the host can be a difficult task. Radio emission may extend far from the host galaxy and emission regions from a single physical object may appear disconnected. As a result, the observed structure of a radio source may have a complex relationship with the corresponding host galaxy, and cross-identification in radio is much more difficult than cross-identification at shorter wavelengths. Small surveys containing a few thousand sources such as the Australia Telescope Large Area Survey (ATLAS; Norris et al., 2006; Middelberg et al., 2008) can be cross-identified manually, but this is impractical for larger surveys.

One approach to cross-identification of large numbers of sources is crowdsourcing, where volunteers cross-identify radio sources with their host galaxy. This is the premise of Radio Galaxy Zoo¹ (Banfield et al., 2015), a citizen science project hosted on the Zooniverse platform (Lintott et al., 2008). Volunteers are shown radio and infrared images and are asked to cross-identify radio sources with the corresponding infrared host galaxies. An explanation of the project can be found in Banfield et al. (2015). The first data release for Radio Galaxy Zoo will provide a large dataset of

¹<https://radio.galaxyzoo.org>

over 75 000 radio-host cross-identifications and radio source morphologies (Wong et al., in prep). While this is a much larger number of visual cross-identifications than have been made by experts (e.g., Taylor et al., 2007; Gendre and Wall, 2008; Grant et al., 2010; Norris et al., 2006; Middelberg et al., 2008) it is still far short of the millions of radio sources expected to be detected in upcoming radio surveys (Norris, 2017b).

Automated algorithms have been developed for cross-identification. Fan et al. (2015) applied Bayesian hypothesis testing to this problem, fitting a three-component model to extended radio sources. This was achieved under the assumption that extended radio sources are composed of a core radio component and two lobe components. The core radio component is coincident with the host galaxy, so cross-identification amounts to finding the galaxy coincident with the core radio component in the most likely model fit. This method is easily extended to use other, more complex models, but it is purely geometric. It does not incorporate other information such as the physical properties of the potential host galaxy. Additionally, there may be new classes of radio source detected in future surveys like EMU which do not fit the model. Weston et al. (2018) developed a modification of the likelihood ratio method of cross-identification (Richter, 1975) for application to ATLAS and EMU. This method does well on non-extended radio sources with approximately 70 per cent accuracy in the ATLAS fields, but does not currently handle more complex (extended or multi-component) radio sources (Norris, 2017a).

One possibility is that machine learning techniques can be developed to automatically cross-identify catalogues drawn from new surveys. Machine learning describes a class of methods that learn approximations to functions. If cross-identification can be cast as a function approximation problem, then machine learning will allow data sets such as Radio Galaxy Zoo to be generalised to work on new data. Data sets from citizen scientists have already been used to train machine learning methods. Some astronomical examples can be found in Marshall, Lintott, and Fletcher (2015a).

In this chapter we cast cross-identification as a function approximation problem by applying an approach from computer vision literature. This approach casts cross-identification as the standard machine learning problem of binary classification by asking whether a given infrared source is the host galaxy or not. We train our methods on expert cross-identifications and volunteer cross-identifications from Radio Galaxy Zoo. In Section 4.2 we describe the data we use to train our methods. In Section 4.3 we discuss how we cast the radio host galaxy cross-identification problem as a machine learning problem. In Section 4.4 we present results of applying our method to ATLAS observations of the *Chandra* Deep Field South (CDFS) and the ESO Large Area ISO Survey South 1 (ELAIS-S1) field. Our data, code and results are available at <https://radiogalaxyzoo.github.io/atlas-xid>.

Throughout this chapter, a ‘radio source’ refers to all radio emission observed associated with a single host galaxy, and a ‘radio component’ refers to a single, contiguous region of radio emission. Multiple components may arise from a single source. A ‘compact’ source is composed of a single unresolved component. Equation 4.4 shows the definition of a resolved component. We assume that all unresolved components are compact sources, i.e. we assume that each unresolved component

Table 4.1: Catalogues of ATLAS/SWIRE cross-identifications for the CDFS and ELAIS-S1 fields. The method used to generate each catalogue is shown, along with the number of radio components cross-identified in each field.

Catalogue	Method	CDFS	ELAIS-S1
Norris et al. (2006)	Manual	784	0
Middelberg et al. (2008)	Manual	0	1366
Fan et al. (2015)	Bayesian models	784	0
Weston et al. (2018)	Likelihood ratio	3078	2113
Wong et al. (in prep)	Crowdsourcing	2460	0

has its own host galaxy². An ‘extended’ source is a non-compact source, i.e. resolved single-component sources or a multi-component source. Figure 4.1 illustrates these definitions.

4.2 Data

We use radio data from the Australia Telescope Large Area Survey (ATLAS; Norris et al., 2006; Franzen et al., 2015), infrared data from the *Spitzer* Wide-area Infrared Extragalactic survey (SWIRE; Lonsdale et al., 2003; Surace et al., 2005), and cross-identifications of these surveys from the citizen science project Radio Galaxy Zoo (Banfield et al., 2015). Radio Galaxy Zoo also includes cross-identifications of sources in Faint Images of the Radio Sky at Twenty-Centimeters (FIRST; White et al., 1997) and the AllWISE survey (Cutri et al., 2013), though we focus only on Radio Galaxy Zoo data from ATLAS and SWIRE.

4.2.1 ATLAS

ATLAS is a pilot survey for the EMU (Norris et al., 2011a) survey, which will cover the entire sky south of +30 deg and is expected to detect approximately 70 million new radio sources. 95 per cent of these sources will be single-component sources, but the remaining 5 per cent pose a considerable challenge to current automated cross-identification methods (Norris et al., 2011a). EMU will be conducted at the same depth and resolution as ATLAS, so methods developed for processing ATLAS data are expected to work for EMU. ATLAS is a wide-area radio survey of the CDFS and ELAIS-S1 fields at 1.4 GHz with a sensitivity of 14 and 17 $\mu\text{Jy beam}^{-1}$ on CDFS and ELAIS-S1 respectively. CDFS covers 3.6 deg² and contains 3034 radio components above a signal-to-noise ratio of 5. ELAIS-S1 covers 2.7 deg² and contains 2084 radio components above a signal-to-noise ratio of 5 (Franzen et al., 2015). The images of CDFS and ELAIS-S1 have angular resolutions of 16 by 7 and 12 by 8 arcsec respectively, with pixel sizes of 1.5 arcsec px⁻¹. Table 4.1 summarises catalogues that contain cross-identifications of radio components in ATLAS with host galaxies

²This will be incorrect if the unresolved components are actually compact lobes or hotspots, but determining which components correspond to unique radio sources is outside the scope of this thesis.

in SWIRE. In the present work, we train methods on CDFS³ and test these methods on both CDFS and ELAIS-S1. This ensures our methods are transferable to different areas of the sky observed by the same telescope as will be the case for EMU.

4.2.2 SWIRE

SWIRE is a wide-area infrared survey at the four IRAC wavelengths 3.6, 4.5, 5.8, and 8.0 μm (Lonsdale et al., 2003; Surace et al., 2005). It covers eight fields, including CDFS and ELAIS-S1. SWIRE is the source of infrared observations for cross-identification with ATLAS. SWIRE has catalogued 221,535 infrared objects in CDFS and 186,059 infrared objects in ELAIS-S1 above a signal-to-noise ratio of 5.

4.2.3 Radio Galaxy Zoo

Radio Galaxy Zoo asks volunteers to cross-identify radio components with their infrared host galaxies. There are a total of 2460 radio components in Radio Galaxy Zoo sourced from ATLAS observations of CDFS. These components are cross-identified by Radio Galaxy Zoo participants with host galaxies detected in SWIRE. A more detailed description can be found in Banfield et al. (2015) and a full description of how the Radio Galaxy Zoo catalogue used in this work⁴ is generated can be found in Wong et al. (in prep).

The ATLAS CDFS radio components that appear in Radio Galaxy Zoo are drawn from a prerelease version of the third data release of ATLAS by Franzen et al. (2015). In this release, each radio component was fit with a two-dimensional Gaussian. Depending on the residual of the fit, more than one Gaussian may be fit to one region of radio emission. Each of these Gaussian fits is listed as a radio component in the ATLAS component catalogue. The brightest radio component from the multiple-Gaussian fit is called the ‘primary component’. If there was only one Gaussian fit then this Gaussian is the primary component. Each primary component found in the ATLAS component catalogue appears in Radio Galaxy Zoo. Non-primary components may appear within the image of a primary component, but do not have their own entry in Radio Galaxy Zoo. We will henceforth only discuss the primary components.

4.3 Method

The aim of this chapter is to express cross-identification in a form that will allow us to apply standard machine learning tools and methods. We use an approach from computer vision to cast cross-identification as binary classification.

³Radio Galaxy Zoo only contains CDFS sources and so we cannot train methods on ELAIS-S1.

⁴The Radio Galaxy Zoo Data Release 1 catalogue will only include cross-identifications for which over 65 per cent of volunteers agree. However, we use a preliminary catalogue containing volunteer cross-identifications for all components.

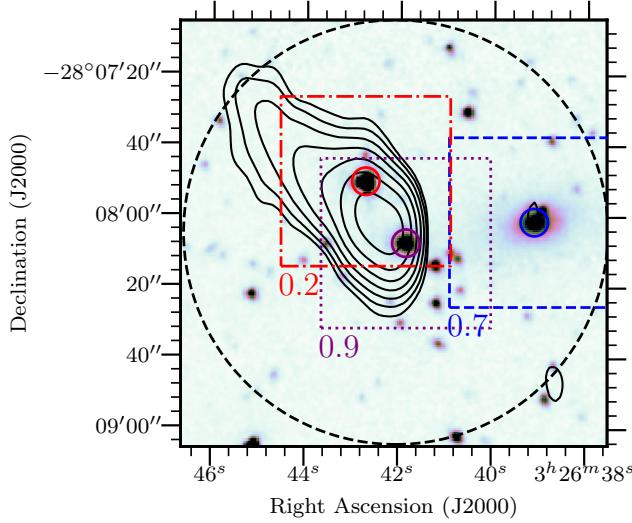


Figure 4.2: An example of finding the host galaxy of a radio source using our sliding-window method. The background image is a $3.6\text{ }\mu\text{m}$ image from SWIRE. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2. Boxes represent ‘windows’ centred on candidate host galaxies, which are circled. The pixels in each window are used to represent the candidate that the window is centred on. The scores of each candidate would be calculated by a binary classifier using the window as input, and these scores are shown below each window. The scores shown are for illustration only. In this example, the galaxy coincident with the centre window would be chosen as the host galaxy, as this window has the highest score. The dashed circle shows the $1'$ radius from which candidate host galaxies are selected. For clarity, not all candidate host galaxies are shown.

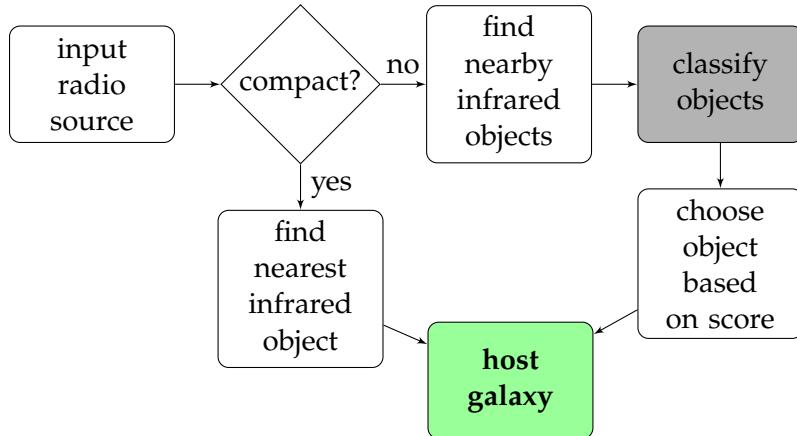


Figure 4.3: Our cross-identification method once a binary classifier has been trained. As input we accept a radio component. If the component is compact, we assume it is a compact source and select the nearest infrared object as the host galaxy. If the component is resolved, we use the binary classifier to score all nearby infrared objects and select the highest-scored object as the host galaxy. Compact and resolved components are defined in Equation 4.4.

4.3.1 Cross-identification as binary classification

We propose a two-step method for host galaxy cross-identification which we will describe now. Given a radio component, we want to find the corresponding host galaxy. The input is a $2' \times 2'$ radio image of the sky centred on a radio component and potentially other information about objects in the image (such as the redshift or infrared colour). Images at other wavelengths (notably infrared) might be useful, but we defer this for now as it complicates the task. We chose a $2' \times 2'$ image to match the size of the images used by Radio Galaxy Zoo. To avoid solving the separate task of identifying which radio components are associated with the same source, we assume that each radio image represents a single extended source⁵. Radio cross-identification can then be formalised as follows: given a radio image centred on a radio component, locate the host galaxy of the source containing this radio component. This is a standard computer vision problem called ‘object detection’, and we apply a common technique called a ‘sliding-window’ (Rowley, Baluja, and Kanade, 1996).

In sliding-window object detection, we want to find an object in an image. We develop a function to score each location in the image such that the highest-scored location coincides with the desired object (Equation ??). Square image cutouts called ‘windows’ are taken centred on each location and these windows are used to represent that location in our scoring function. To find the infrared host galaxy, we choose the location with the highest score. To improve the efficiency of this process when applied to cross-identification, we only consider windows coincident with infrared sources detected in SWIRE. We call these infrared sources ‘candidate host galaxies’.

⁵Limitations of this assumption are discussed in Section 4.3.2.

For this chapter, there is no use in scoring locations without infrared sources as that would not lead to a host identification anyway. Using candidate host galaxies instead of pixels also allows us to include ancillary information about the candidate host galaxies, such as their infrared colours and redshifts. We refer to the maximum distance a candidate host galaxy can be separated from a radio component as the ‘search radius’ and take this radius to be 1 arcmin. To score each candidate host galaxy we use a ‘binary classifier’, which we will define now.

Algorithm 1: Cross-identifying a radio component given a radio image of the component, a catalogue of infrared candidate host galaxies and a binary classifier. σ is a parameter of the method.

Data:

A 2×2 arcmin radio image of a radio component

A set of infrared candidate host galaxies \mathcal{G}

A binary classifier $f : \mathbb{R}^k \rightarrow \mathbb{R}$

Result: A galaxy $g \in \mathcal{G}$

$max \leftarrow -\infty;$

$host \leftarrow \emptyset;$

for $g \in \mathcal{G}$ **do**

$x \leftarrow$ a k -dimensional vector representation of g (Section 4.3.3);

$d \leftarrow$ distance between g and the radio component;

$$score \leftarrow f(x) \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{d^2}{2\sigma^2}\right); \quad (4.1)$$

if $score > max$ **then**

$max \leftarrow score;$

$host \leftarrow g;$

end

end

return $host$

Binary classification is a common method in machine learning where objects are to be assigned to one of two classes, called the ‘positive’ and ‘negative’ classes. This assignment is represented by the probability that an object is in the positive class. A ‘binary classifier’ is a function mapping from an object to such a probability. Our formulation of cross-identification is equivalent to binary classification of candidate host galaxies: the positive class represents host galaxies, the negative class represents non-host galaxies, and to cross-identify a radio component we find the candidate host galaxy maximising the positive class probability. In other words, the binary classifier is exactly the sliding-window scoring function. We therefore split cross-identification into two separate tasks: the ‘candidate classification task’ where, given a candidate host galaxy, we wish to determine whether it is a host galaxy of *any* radio component; and the ‘cross-identification task’ where, given a specific radio component, we wish to find its host galaxy. The candidate classification task is a traditional machine learning problem which results in a binary classifier. To avoid ambiguity and recognise that the values output by a binary classifier are not true probabilities, we will refer

to the outputs of the binary classifier as ‘scores’ in line with the sliding-window approach described above. The cross-identification task maximises over scores output by this classifier. Our approach is illustrated in Figure 4.2 and described in algorithm 1. We refer to the binary classifier scoring a candidate host galaxy as f . To implement f as a function that accepts candidate host galaxies as input, we need to represent candidate host galaxies by vectors. We describe this in Section 4.3.3. There are many options for modelling f . In this chapter we apply three different models: logistic regression, random forests and convolutional neural networks.

We cross-identify each radio component in turn. The classifier f provides a score for each candidate host galaxy. This score indicates how much the candidate looks like a host galaxy, independent of which radio component we are currently cross-identifying. If there are other nearby host galaxies, then multiple candidate hosts may have high scores (e.g. Figure 4.4). This difficulty is necessary — a classifier with dependence on radio object would be impossible to train. We need multiple positive examples (i.e. host galaxies) to train a binary classifier, but for any specific radio component there is only one host galaxy. As a result, the candidate classification task aims to answer the general question of whether a given galaxy is the host galaxy of *any* radio component, while the cross-identification task attempts to cross-identify a *specific* radio component. To distinguish between candidate host galaxies with high scores, we weight the scores by a Gaussian function of angular separation between the candidates and the radio component. The width of the Gaussian, σ , controls the influence of the Gaussian on the final cross-identification. When σ is small, our approach is equivalent to a nearest-neighbours approach where we select the nearest infrared object to the radio component as the host galaxy. In the limit where $\sigma \rightarrow \infty$, we maximise the score output by the classifier as above. We take $\sigma = 30''$ as this was the best value found by a grid search. Note that the optimum width will depend on the density of radio sources on the sky, the angular separation of the host galaxy and its radio components and the angular resolution of the survey.

We can improve upon this method by cross-identifying compact radio sources separately from extended sources, as compact sources are much easier to cross-identify. For a compact source, the nearest SWIRE object may be identified as the host galaxy (a *nearest-neighbours* approach), or a more complex method such as likelihood ratios may be applied (see Weston et al., 2018). We cross-identify compact sources separately in our pipeline and this process is shown in Figure 4.3.

4.3.2 Limitations of our approach

We make a number of assumptions to relate the cross-identification task to the candidate classification task:

1. For any radio component, the $2' \times 2'$ image centred on the component contains components of only one radio source.
2. For any radio component, the $2' \times 2'$ image centred on the component contains all components of this source.

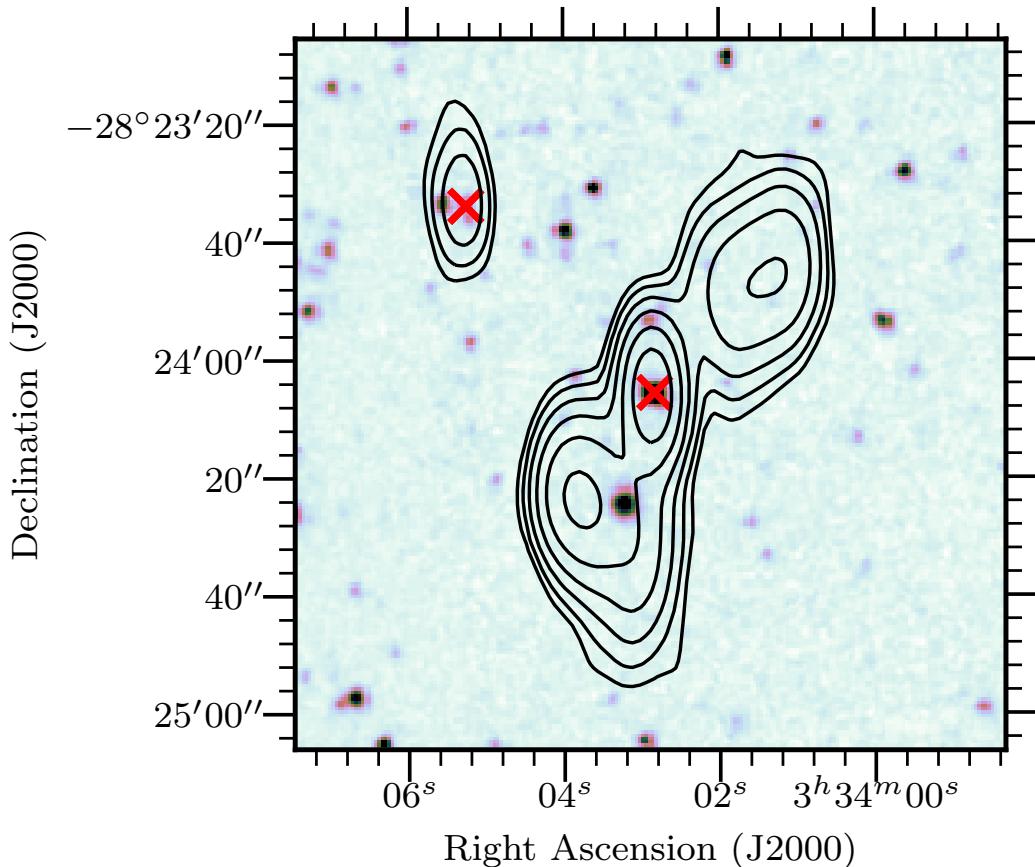


Figure 4.4: A 2'-wide radio image centred on ATLAS3_J033402.87-282405.8C. This radio source breaks the assumption that there are no other radio sources within 1 arcmin of the source. Another radio source is visible to the upper-left. Host galaxies found by Radio Galaxy Zoo volunteers are shown by crosses. The background image is a $3.6\text{ }\mu\text{m}$ image from SWIRE. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2.

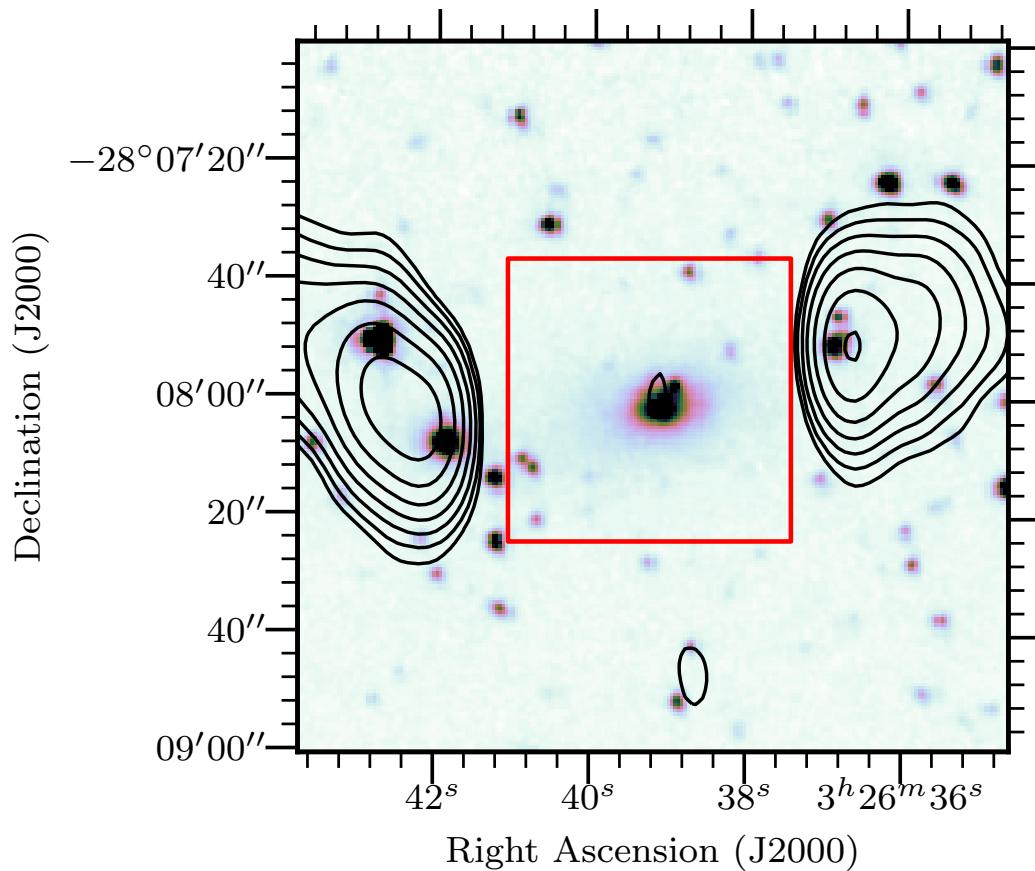


Figure 4.5: An example of a radio source where the window centred on the host galaxy, shown as a rectangle, does not contain enough radio information to correctly identify the galaxy as the host. The background image is a $3.6\text{ }\mu\text{m}$ image from SWIRE. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2.

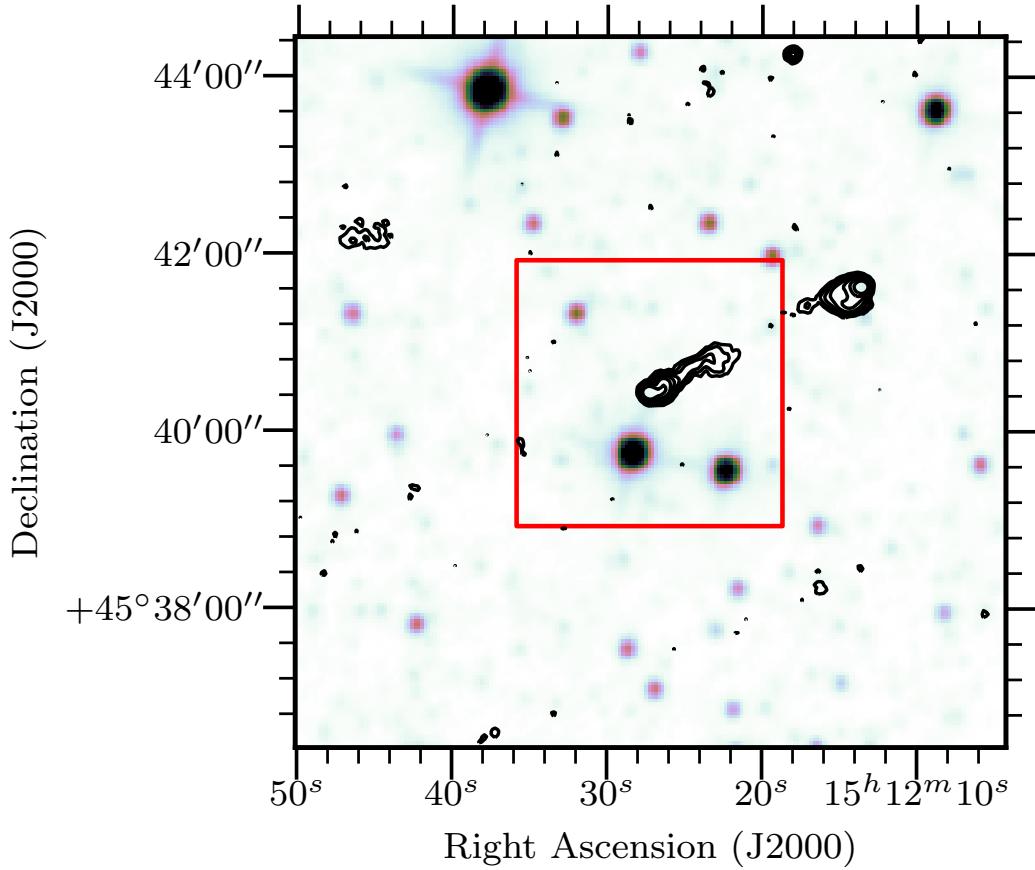


Figure 4.6: A $8'$ -wide radio image from FIRST, centred on FIRST J151227.2+454026. The $3'$ -wide red box indicates the boundaries of the image of this radio component shown to volunteers in Radio Galaxy Zoo. This radio source breaks our assumption that the whole radio source is visible in the chosen radius. As one of the components of the radio source is outside of the image, a volunteer (or automated algorithm) looking at the $3'$ -wide image may be unable to determine that this is a radio double or locate the host galaxy. The background image is a $3.4 \mu\text{m}$ image from WISE. The contours show FIRST radio data, starting at 4σ and increasing geometrically by a factor of 2.

-
3. The host galaxy of a radio component is within the 1 arcmin search radius around the component, measured from the centre of the Gaussian fit.
 4. The host galaxy of a radio component is closer on the sky to the radio component than the host galaxy of any other radio component.
 5. The host galaxy appears in the SWIRE catalogue.

These assumptions limit the effectiveness of our approach, regardless of how accurate our binary classifier may be. Examples of radio sources that break these respective assumptions are:

1. A radio source less than 1' away from another radio source.
2. A radio source with an angular size greater than 2'.
3. A radio source with a component greater than 1' away from the host galaxy.
4. A two-component radio source with another host galaxy between a component and the true host galaxy.
5. An infrared-faint radio source (as in Collier et al., 2014).

The main limitations are problems of scale in choosing the candidate search radius and the size of the windows representing candidates. If the search radius is too small, we may not consider the host galaxy as a candidate. If the search radius is too large, we may consider multiple host galaxies (though this is mostly mitigated by the Gaussian weighting). If the window is too small, radio emission may extend past the edges of the window and we may miss critical information required to identify the galaxy as a host galaxy. If the window is too large, then irrelevant information will be included and it may be difficult or computationally expensive to score. We chose a window size of 32×32 pixels, corresponding to approximately $48'' \times 48''$ in ATLAS. This is shown as squares in Figure 4.2 and Figure 4.5. These kinds of size problems are difficult even for non-automated methods as radio sources can be extremely wide — for example, Radio Galaxy Zoo found a radio giant that spanned over three different images presented to volunteers and the full source was only cross-identified by the efforts of citizen scientists (Banfield et al., 2015). An example of a radio image where part of the radio source is outside the search radius is shown in Figure 4.6.

In weighting the scores by a Gaussian function of angular separation, we implicitly assume that the host galaxy of a radio component is closer to that radio component than any other host galaxy. If this assumption is not true then the incorrect host galaxy may be identified, though this is rare.

We only need to require that the host galaxy appears in SWIRE to incorporate galaxy-specific features (Section 4.3.3) and to improve efficiency. Our method is applicable even when host galaxies are not detected in the infrared by considering every pixel of the radio image as a candidate location as would be done in the original computer vision approach. If the host galaxy location does not correspond to an infrared source, the radio source would be classified as infrared-faint.

Our assumptions impose an upper bound on how well we can cross-identify radio sources. We estimate this upper bound in Section 4.4.1.

4.3.3 Feature vector representation of infrared sources

Inputs to binary classifiers must be represented by an array of real values called feature vectors. We therefore need to choose a feature vector representation of our candidate host galaxies. Candidate hosts are sourced from the SWIRE catalogue (Section 4.2.2). We represent each candidate host with 1034 real-valued features, combining the windows centred on each candidate (Section 4.3.1) with ancillary infrared data from the SWIRE catalogue. For a given candidate host, these features are:

- the 6 base-10 logarithms of the ratios of fluxes of the candidate host at the four IRAC wavelengths (the ‘colours’ of the candidate);
- the flux of the host at 3.6 μm ;
- the stellarity index of the host at both 3.6 and 4.5 μm ;
- the radial distance between the candidate host and the nearest radio component in the ATLAS catalogue; and
- a 32×32 pixel image from ATLAS (approximately $48'' \times 48''$), centred on the candidate host (the window).

The infrared colours provide insight into the properties of the candidate host galaxy (Grant, 2011). The 3.6 and 4.5 μm fluxes trace both galaxies with faint polycyclic aromatic hydrocarbon (PAH) emission (i.e. late-type, usually star-forming galaxies) and elliptical galaxies dominated by old stellar populations. The 5.8 μm flux selects galaxies where the infrared emission is dominated by non-equilibrium emission of dust grains due to active galactic nuclei, while the 8.0 μm flux traces strong PAH emission at low redshift (Sajina2005). The stellarity index is a value in the SWIRE catalogue that represents how likely the object is to be a star rather than a galaxy (Surace et al., 2005). It was estimated by a neural network in SExtractor (Bertin and Arnouts, 1996).

We use the 32×32 pixels of each radio window as independent features for all binary classification models, with the convolutional neural network automatically extracting features that are relevant. Other features of the radio components may be used instead of just relying on the pixel values, but there has been limited research on extracting such features: Proctor (2006) describes hand-selected features for radio doubles in FIRST, and Aniyan and Thorat (2017) and Lukic et al. (2018) make use of deep convolutional neural networks which automatically extract features as part of classification. A more comprehensive investigation of features is a good avenue for potential improvement in our pipeline but this is beyond the scope of this initial study.

4.3.4 Binary Classifiers

We use three different binary classification models: logistic regression, convolutional neural networks, and random forests. These models cover three different approaches to machine learning. Logistic regression is a probabilistic binary classification model. It is linear in the feature space and outputs the probability that the input has a positive label (Bishop, 2006, Chap. 4). Convolutional neural networks (CNN) are biologically-inspired prediction models with image inputs. They have recently produced good results on large image-based datasets in astronomy (Lukic et al., 2018; Dieleman, Willett, and Dambre, 2015, e.g.). Random forests are an ensemble of decision trees (Breiman, 2001). They consider multiple subsamples of the training set, where each bootstrap subsample is sampled with replacement from the training set. To classify a new data point, the random forest takes the weighted average of all classifications produced by each decision tree.

4.3.5 Classification models

We use three different models for binary classification: logistic regression, convolutional neural networks, and random forests.

4.3.5.1 Logistic Regression

Logistic regression is linear in the feature space and outputs the probability that the input has a positive label. The model is (Bishop, 2006):

$$f(\vec{x}) = \sigma(\vec{w}^T \vec{x} + b) , \quad (4.2)$$

where $\vec{w} \in \mathbb{R}^D$ is a vector of parameters, $b \in \mathbb{R}$ is a bias term, $\vec{x} \in \mathbb{R}^D$ is the feature vector representation of a candidate host, and $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is the logistic sigmoid function:

$$\sigma(a) = (1 + \exp(-a))^{-1} . \quad (4.3)$$

The logistic regression model is fully differentiable, and the parameters \vec{w} can therefore be learned using gradient-based optimisation methods. We used the `scikit-learn` (Pedregosa et al., 2011) implementation of logistic regression with balanced classes.

4.3.5.2 Convolutional neural networks

Convolutional neural networks (CNN) are a biologically-inspired prediction model for prediction with image inputs. The input image is convolved with a number of filters to produce output images called feature maps. These feature maps can then be convolved again with other filters on subsequent layers, producing a network of convolutions. The whole network is differentiable with respect to the values of the filters and the filters can be learned using gradient-based optimisation methods. The final layer of the network is logistic regression, with the convolved outputs as input features. For more detail, see LeCun et al. (subsection II.A 1998). We use

Table 4.2: Number of compact and resolved radio objects in each CDFS quadrant. Radio Galaxy Zoo (RGZ) has more cross-identifications than the expert catalogue (Norris et al., 2006) provides as it uses a deeper data release of ATLAS, and so has more objects in each quadrant for training.

Quadrant	Compact	Resolved	Compact (RGZ)	Resolved (RGZ)
0	126	24	410	43
1	99	21	659	54
2	61	24	555	57
3	95	18	631	51
<i>Total</i>	381	87	2255	205

KERAS (Chollet et al., 2015) to implement our CNN, accounting for class imbalance by reweighting the classes.

CNNs have recently produced good results on large image-based datasets in astronomy (Lukic et al., 2018; Dieleman, Willett, and Dambre, 2015, e.g.). We employ only a simple CNN model in this chapter as a proof of concept that CNNs may be used for class probability prediction on radio images. The model architecture we use is shown in Figure 4.17.

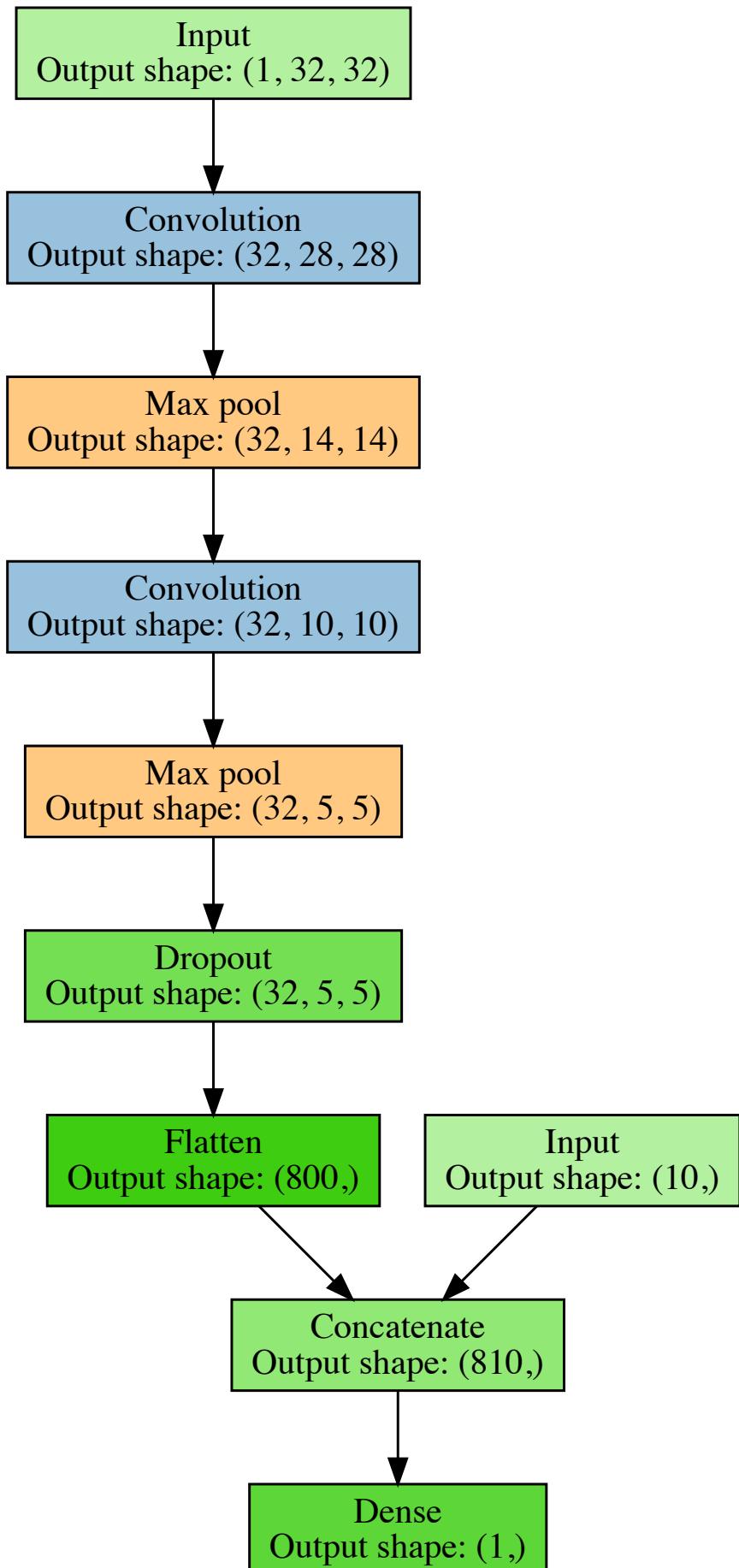
4.3.5.3 Random Forests

Random forests are an ensemble of decision trees (Breiman, 2001). They consider multiple subsamples of the training set, where each subsample is sampled with replacement from the training set. For each subsample a decision tree classifier is constructed by repeatedly making axis-parallel splits based on individual features. In a random forest the split decision is taken based on a random subset of features. To classify a new data point, the random forest takes the weighted average of all classifications produced by each decision tree. We used the `scikit-learn` (Pedregosa et al., 2011) implementation of random forests with 10 trees, the information entropy split criterion, a minimum leaf size of 45 and balanced classes.

4.3.6 Labels

The Radio Galaxy Zoo and Norris et al. (2006) cross-identification catalogues must be converted to binary labels for infrared objects so that they can be used to train binary classifiers. There are two challenges with this conversion:

- We can only say that an object is *a* host galaxy, not which radio object it is associated with, and
- We cannot disambiguate between non-host infrared objects and host galaxies that were not in the cross-identification catalogue.



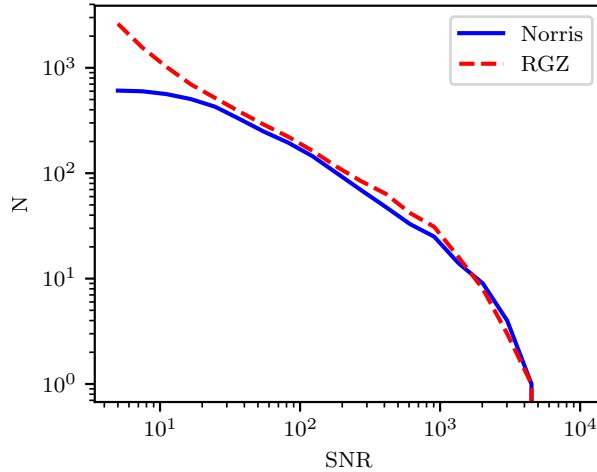


Figure 4.8: Cumulative number of radio components (N) in the expert (Norris) and Radio Galaxy Zoo (RGZ) training sets with different signal-to-noise ratios (SNR).

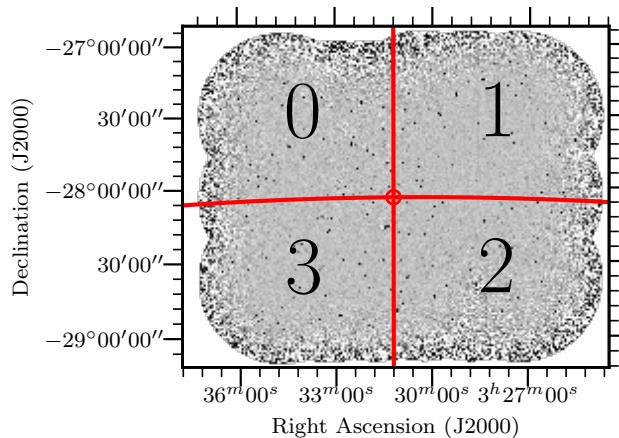


Figure 4.9: CDFS field training and testing quadrants labelled 0 – 3. The central dot is located at $\alpha = 03^{\text{h}}31^{\text{m}}12^{\text{s}}$, $\delta = -28^{\circ}06'00''$. The quadrants were chosen such that there are similar numbers of radio sources in each quadrant.

We use the Gaussian weighting described in Section 4.3.1 to address the first issue. The second issue is known as a ‘positive-unlabelled’ classification problem, which is a binary classification problem where we only observe labels for the positive class. We treat unlabelled objects as negative examples following Menon et al. (2015). That is, we make the naïve assumption that any infrared object in the SWIRE catalogue not identified as a host galaxy in a cross-identification catalogue is not a host galaxy at all.

We first generate positive labels from a cross-identification catalogue. We decide that if an infrared object is listed in the catalogue, then it is assigned a positive label as a host galaxy. We then assign every other galaxy a negative label. This has some problems — an example is that if the cross-identification catalogue did not include a radio object (e.g. it was below the signal-to-noise ratio) then the host galaxy of that radio object would receive a negative label. This occurs with Norris et al. (2006) cross-identifications, as these are associated with the first data release of ATLAS. The first data release went to a 5σ flux density level of $S_{1.4} \geq 200 \mu\text{Jy beam}^{-1}$ (Norris et al., 2006), compared to $S_{1.4} \geq 85 \mu\text{Jy beam}^{-1}$ for the third data release used by Radio Galaxy Zoo (Franzen et al., 2015). The labels from Norris et al. (2006) may therefore disagree with labels from Radio Galaxy Zoo even if they are both plausible. The difference in training set size at different flux cutoffs is shown in Figure 4.8. We train and test our binary classifiers on infrared objects within a 1 arcmin radius of an ATLAS radio component.

4.3.7 Experimental Setup

We trained binary classifiers on infrared objects in the CDFS field using two sets of labels. One label set was derived from Radio Galaxy Zoo cross-identifications and the other was derived from the Norris et al. (2006) cross-identification catalogue. We refer to these as the ‘Radio Galaxy Zoo labels’ and the ‘expert labels’ respectively. We divided the CDFS field into four quadrants for training and testing. The quadrants were divided with a common corner at $\alpha = 03^{\text{h}}31^{\text{m}}12^{\text{s}}$, $\delta = -28^{\circ}06'00''$ as shown in Figure 4.9. For each trial, one quadrant was used to extract test examples and the other three quadrants were used for training examples.

We further divided the radio components into compact and resolved. Compact components are cross-identified by fitting a 2D Gaussian (as in Norris et al., 2006) and we would expect any machine learning approach for host cross-identification to attain high accuracy on this set. A radio component was considered resolved if

$$\ln \left(\frac{S_{\text{int}}}{S_{\text{peak}}} \right) > 2 \sqrt{\left(\frac{\sigma_{S_{\text{int}}}}{S_{\text{int}}} \right)^2 + \left(\frac{\sigma_{S_{\text{peak}}}}{S_{\text{peak}}} \right)^2}, \quad (4.4)$$

where S_{int} is the integrated flux density, S_{peak} is the peak flux density, $\sigma_{S_{\text{int}}}$ is the uncertainty in integrated flux density and $\sigma_{S_{\text{peak}}}$ is the uncertainty in peak flux density (following Franzen et al., 2015).

Candidate hosts were selected from the SWIRE catalogue. For a given subset

of radio components, all SWIRE objects within 1 arcmin of all radio components in the subset were added to the associated SWIRE subset. In results for the candidate classification task, we refer to SWIRE objects within 1 arcmin of a compact radio component as part of the ‘compact set’, and SWIRE objects within 1 arcmin of a resolved radio component as part of the ‘resolved set’.

To reduce bias in the testing data due to the expert labels being generated from a shallower data release of ATLAS, a SWIRE object was only included in the test set if it was within 1 arcmin of a radio object with a SWIRE cross-identification in both the Norris et al. (2006) catalogue and the Radio Galaxy Zoo catalogue.

Each binary classifier was trained on the training examples and used to score the test examples. These scores were thresholded to generate labels which could be directly compared to the expert labels. We then computed the ‘balanced accuracy’ of these predicted labels. Balanced accuracy is the average of the accuracy on the positive class and the accuracy on the negative class, and is not sensitive to class imbalance. The candidate classification task has highly imbalanced classes — in our total set of SWIRE objects within 1 arcmin of an ATLAS object, only 4 per cent have positive labels. Our threshold was chosen to maximise the balanced accuracy on predicted labels of the training set. Only examples within 1 arcmin of ATLAS objects in the first ATLAS data release (Norris et al., 2006) were used to compute balanced accuracy, as these were the only ATLAS objects with expert labels.

We then used the scores to predict the host galaxy for each radio component cross-identified by both Norris et al. (2006) and Radio Galaxy Zoo. We followed algorithm 1: the score of each SWIRE object within 1 arcmin of a given radio component was weighted by a Gaussian function of angular separation from the radio component and the object with the highest weighted score was chosen as the host galaxy. The cross-identification accuracy was then estimated as the fraction of the predicted host galaxies that matched the Norris et al. (2006) cross-identifications.

4.4 Results

In this section we present accuracies of our method trained on CDFS and applied to CDFS and ELAIS-S1, as well as results motivating our accuracy measures and estimates of upper and lower bounds for cross-identification accuracy using our method.

4.4.1 Application to ATLAS-CDFS

We can assess trained binary classifiers either by their performance on the candidate classification task or by their performance on the cross-identification task when used in our method. Both performances are useful: performance on the candidate classification task provides a robust and simple way to compare binary classifiers without the limitations of our specific formulation, and performance on the cross-identification task can be compared with other cross-identification methods. We therefore report two sets of accuracies: balanced accuracy for the galaxy classification task and accuracy for the cross-identification task. These accuracy measures

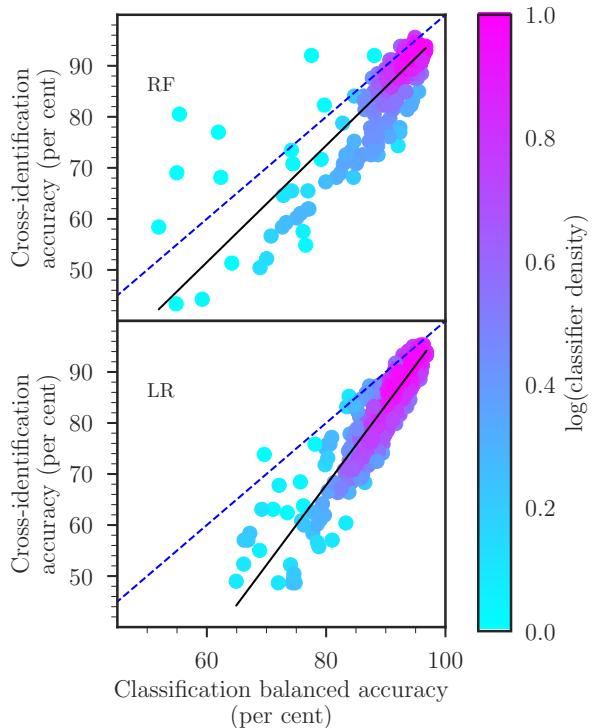


Figure 4.10: Balanced accuracy on the candidate classification task plotted against accuracy on the cross-identification task. ‘RF’ indicates results from random forests, and ‘LR’ indicates results from logistic regression. Binary classifiers were trained on random, small subsets of the training data to artificially restrict their accuracies. Colour shows the density of points on the plot estimated by a Gaussian kernel density estimate. The solid lines indicate the best linear fit; these fits have $R^2 = 0.92$ for logistic regression and $R^2 = 0.87$ for random forests. The dashed line shows the line where cross-identification accuracy and candidate classification accuracy are equal. We did not include convolutional neural networks in this test, as training them is very computationally expensive. There are 640 trials shown per classification model. These results exclude binary classifiers with balanced accuracies less than 51 per cent, as these are essentially random.

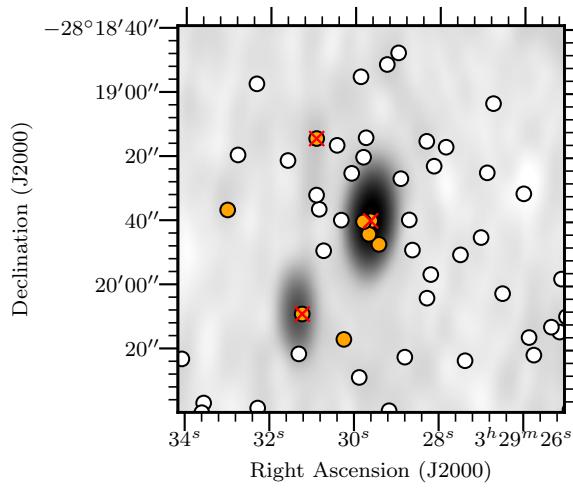


Figure 4.11: Predicted host galaxies in the candidate classification task for ATLAS3 J032929.61-281938.9. The background image is an ATLAS radio image. Radio Galaxy Zoo host galaxies are marked by crosses. SWIRE candidate host galaxies are circles coloured by the score output by a logistic regression binary classifier. The scores are thresholded to obtain labels, as when we compute balanced accuracy. Orange circles have been assigned a ‘positive’ label by a logistic regression binary classifier and white otherwise. Note that there are more predicted host galaxies than there are radio components, so not all of the predicted host galaxies would be assigned as host galaxies in the cross-identification task.

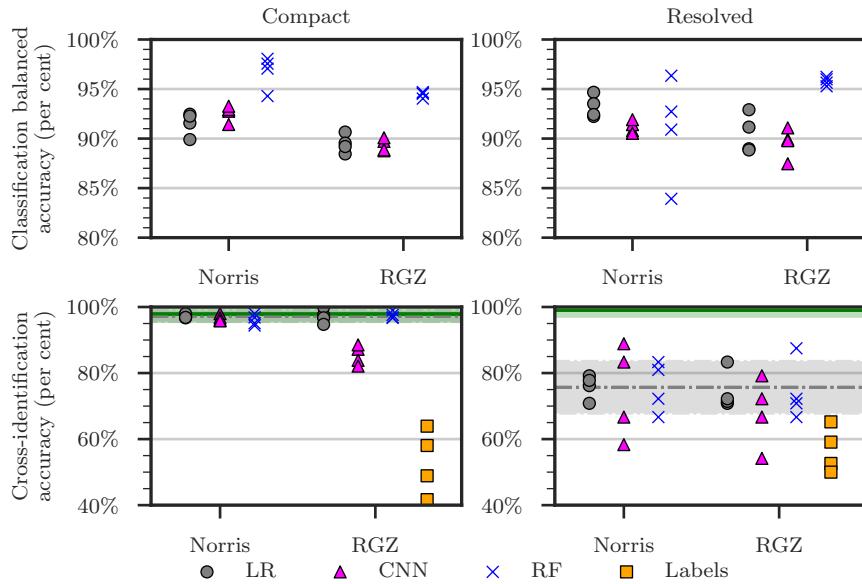


Figure 4.12: Performance of our method with logistic regression ('LR'), convolutional neural networks ('CNN') and random forest ('RF') binary classifiers. 'Norris' indicates the performance of binary classifiers trained on the expert labels and 'RGZ' indicates the performance of binary classifiers trained on the Radio Galaxy Zoo labels. One point is shown per binary classifier per testing quadrant. The training and testing sets have been split into compact (left) and resolved (right) objects. Shown for comparison is the accuracy of the Radio Galaxy Zoo consensus cross-identifications on the cross-identification task, shown as 'Labels'. The cross-identification accuracy attained by a perfect binary classifier is shown by a solid green line, and the cross-identification accuracy of a nearest-neighbours approach is shown by a dashed grey line. The standard deviation of these accuracies across the four CDFS quadrants is shown by the shaded area. Note that the pipeline shown in Figure 4.3 is not used for these results.

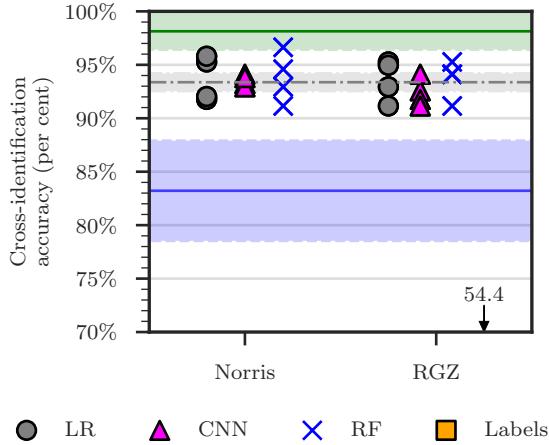


Figure 4.13: Performance of our approach using different binary classifiers on the cross-identification task. Markers and lines are as in Figure 4.12. The blue solid line indicates the performance of a random binary classifier and represents the minimum accuracy we expect to obtain. The standard deviation of this accuracy across 25 trials and 4 quadrants is shaded. The accuracy of Radio Galaxy Zoo on the cross-identification task is below the axis and is instead marked by an arrow with the mean accuracy. Note that the pipeline shown in Figure 4.3 is used here, so compact objects are cross-identified in the same way regardless of binary classifier model.

are correlated and we show this correlation in Figure 4.10. Fitting a line of best fit with `scipy` gives $R^2 = 0.92$ for logistic regression and $R^2 = 0.87$ for random forests. While performance on the candidate classification task is correlated with performance on the cross-identification task, balanced accuracy does not completely capture the effectiveness of a binary classifier applied to the cross-identification task. This is because while our binary classifiers output real-valued scores, these scores are thresholded to compute the balanced accuracy. In the candidate classification task, the binary classifier only needs to ensure that host galaxies are scored higher than non-host galaxies. This means that after thresholding there can be many ‘false positives’ that do not affect cross-identification. An example of this is shown in Figure 4.11, where the classifier has identified 8 ‘host galaxies’. However, there are only three true host galaxies in this image — one per radio component — and so in the cross-identification task, only three of these galaxies will be identified as hosts.

In Figure 4.12 we plot the balanced accuracies of our classification models on the candidate classification task and the cross-identification accuracies of our method using each of these models. Results are shown for both the resolved and compact sets. For comparison, we also plot the cross-identification accuracy of Radio Galaxy Zoo and a nearest-neighbours approach, as well as estimates for upper and lower limits on the cross-identification accuracy. We estimate the upper limit on performance by assigning all true host galaxies a score of 1 and assigning all other candidate host galaxies a score of 0. This is equivalent to ‘perfectly’ solving the candidate clas-

sification task and so represents the best possible cross-identification performance achievable with our method. We estimate the lower limit on performance by assigning random scores to each candidate host galaxy. We expect any useful binary classifier to produce better results than this, so this represents the lowest expected cross-identification performance. The upper estimates, lower estimates, and nearest-neighbour accuracy are shown as horizontal lines in Figure 4.12.

In Figure 4.13 we plot the performance of our method using different binary classification models, as well as the performance of Radio Galaxy Zoo, nearest-neighbours, and the perfect and random binary classifiers, on the full set of ATLAS DR1 radio components using the pipeline in Figure 4.3. The accuracy associated with each classification model and training label set averaged across all four quadrants is shown in Appendix ??.

Differences between accuracies across training labels are well within one standard deviation computed across the four quadrants, with convolutional neural networks on compact objects as the only exception. The spread of accuracies is similar for both sets of training labels, with the exception of random forests. The balanced accuracies of random forests trained on expert labels have a considerably higher spread than those trained on Radio Galaxy Zoo labels, likely because of the small size of the expert training set — there are less than half the number of objects in the expert-labelled training set than the number of objects in the Radio Galaxy Zoo-labelled training set (Table 4.2).

Radio Galaxy Zoo-trained methods significantly outperform Radio Galaxy Zoo cross-identifications. Additionally, despite poor performance of Radio Galaxy Zoo on the cross-identification task, methods trained on these cross-identifications still perform comparably to those trained on expert labels. This is because incorrect Radio Galaxy Zoo cross-identifications can be thought of as a source of noise in the labels which is ‘averaged out’ in training. This shows the usefulness of crowdsourced training data, even when the data is noisy.

Our method performs comparably to a nearest-neighbours approach. For compact objects, this is to be expected — indeed, nearest-neighbours attains nearly 100 per cent accuracy on the compact test set. Our results do not improve on nearest-neighbours for resolved objects. However, our method does allow for improvement on nearest-neighbours with a sufficiently good binary classifier: a ‘perfect’ binary classifier attains nearly 100 per cent accuracy on resolved sources. This shows that our method may be useful provided that a good binary classifier can be trained. The most obvious place for improvement is in feature selection: we use pixels of radio images directly and these are likely not conducive to good performance on the candidate classification task. Convolutional neural networks, which are able to extract features from images, *should* work better, but these require far more training data than the other methods we have applied and the small size of ATLAS thus limits their performance.

We noted in Section 4.3.6 that the test set of expert labels, derived from the initial ATLAS data release, was less deep than the third data release used by Radio Galaxy Zoo and this chapter, introducing a source of label noise in the testing labels. Specif-

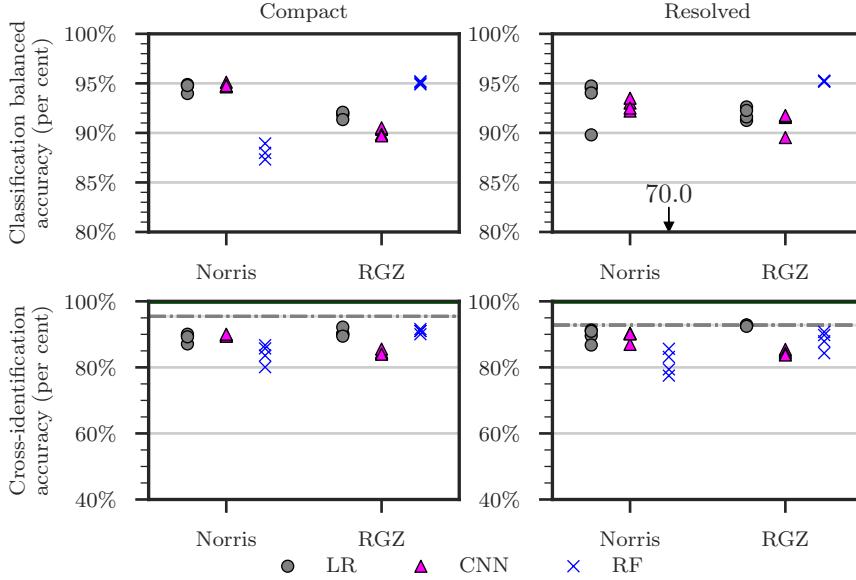


Figure 4.14: Performance of different classification models trained on CDFS and tested on resolved and compact sources in ELAIS-S1. Points represent classification models trained on different quadrants of CDFS, with markers, lines and axes as in Figure 4.12. The balanced accuracy of expert-trained random forest binary classifiers falls below the axis and the corresponding mean accuracy is shown by an arrow. The estimated best attainable accuracy is almost 100 per cent.

ically, true host galaxies may be misidentified as non-host galaxies if the associated radio source was below the 5 signal-to-noise limit in ATLAS DR1 but not in ATLAS DR3. This has the effect of reducing the accuracy for Radio Galaxy Zoo-trained classifiers.

We report the scores predicted by each classifier for each SWIRE object in Appendix ?? and the predicted cross-identification for each ATLAS object in Appendix ???. Scores reported for a given object were predicted by binary classifiers tested on the quadrant containing that object. The reported scores are not weighted.

In Figure 4.18 we show 5 resolved sources where the most classifiers disagreed on the correct cross-identification.

4.4.2 Application to ATLAS-ELAIS-S1

We applied the method trained on CDFS to perform cross-identification on the ELAIS-S1 field. Both CDFS and ELAIS-S1 were imaged by the same radio telescope to similar sensitivities and angular resolution for the ATLAS survey. We can use the SWIRE cross-identifications made by Middelberg et al. (2008) to derive another set of expert labels, and hence determine how accurate our method is. If our method generalises well across different parts of the sky, then we expect CDFS-trained classifiers to have

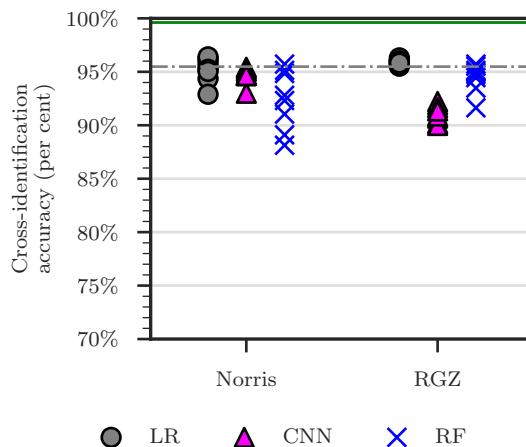


Figure 4.15: Performance of different classifiers trained on CDFS and tested on ELAIS-S1. Markers are as in Figure 4.13 and horizontal lines are as in Figure 4.14. Note that the pipeline shown in Figure 4.3 is used here, so compact objects are cross-identified in the same way regardless of binary classifier model.

comparable performance between ELAIS-S1 and CDFS. In Figure 4.14 we plot the performance of CDFS-trained classification models on the candidate classification task and the performance of our method on the cross-identification task using these models. We also plot the cross-identification accuracy of a nearest-neighbours approach⁶. In Figure 4.15 we plot the performance of our method on the full set of ELAIS-S1 ATLAS DR1 radio components using the pipeline in Figure 4.3. We list the corresponding accuracies in Appendix ??.

Cross-identification results from ELAIS-S1 are similar to those for CDFS, showing that our method trained on CDFS performs comparably well on ELAIS-S1. However, nearest-neighbours outperforms most methods on ELAIS-S1. This is likely because there is a much higher percentage of compact objects in ELAIS-S1 than in CDFS. The maximum achievable accuracy we have estimated for ELAIS-S1 is very close to 100 per cent, so (as for CDFS) a very accurate binary classifier would outperform nearest-neighbours.

One interesting difference between the ATLAS fields is that random forests trained on expert labels perform well on CDFS but poorly on ELAIS-S1. This is not the case for logistic regression or convolutional neural networks trained on expert labels, nor is it the case for random forests trained on Radio Galaxy Zoo. We hypothesise that this is because the ELAIS-S1 cross-identification catalogue (Middelberg et al., 2008) labelled fainter radio components than the CDFS cross-identification catalogue (Norris et al., 2006) due to noise from the very bright source ATCDFS_J032836.53-284156.0 in CDFS. Classifiers trained on CDFS expert labels may thus be biased to-

⁶We cannot directly compare our method applied to ELAIS-S1 with Radio Galaxy Zoo, as Radio Galaxy Zoo does not include ELAIS-S1.

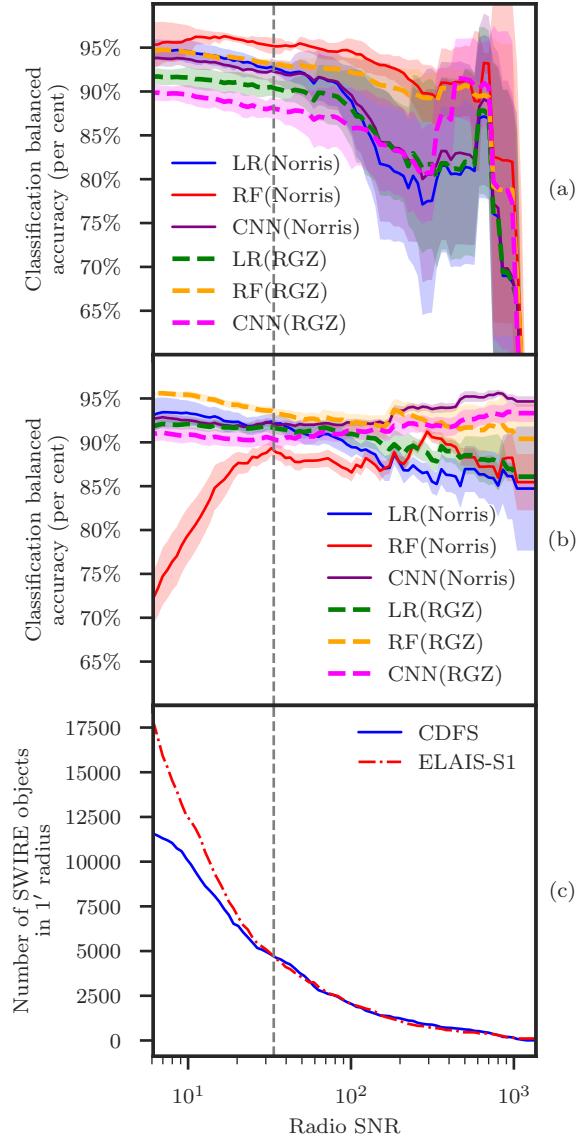


Figure 4.16: (a) Balanced accuracies of classifiers trained and tested on CDFS with different signal-to-noise ratio (SNR) cutoffs for the test set. A SWIRE object is included in the test set if it is within $1'$ of a radio component with greater SNR than the cutoff. Lines of different colour indicate different classifier/training labels combinations, where LR is logistic regression, RF is random forests, CNN is convolutional neural networks, and Norris and RGZ are the expert and Radio Galaxy Zoo label sets respectively. Filled areas represent standard deviations across CDFS quadrants. (b) Balanced accuracies of classifiers trained on CDFS and tested on ELAIS-S1. (c) A cumulative distribution plot of SWIRE objects associated with a radio object with greater SNR than the cutoff. The grey dashed line shows the SNR level at which the number of SWIRE objects above the cutoff is equal for CDFS and ELAIS-S1. This cutoff level is approximately at a SNR of 34.

ward brighter radio components compared to ELAIS-S1. Radio Galaxy Zoo uses a preliminary version of the third data release of ATLAS (Franzen et al., 2015) and so classifiers trained on the Radio Galaxy Zoo labels may be less biased toward brighter sources compared to those trained on the expert labels. To test this hypothesis we tested each classification model against test sets with a signal-to-noise ratio (SNR) cutoff. A SWIRE object was only included in the test set for a given cutoff if it was located within $1'$ of a radio component with a SNR above the cutoff. The balanced accuracies for each classifier at each cutoff are shown in Figure 4.16(a) and (b) and the distribution of test set size for each cutoff is shown in Figure 4.16(c). Figure 4.16(c) shows that ELAIS-S1 indeed has more faint objects in its test set than the CDFS test set, with the SNR for which the two fields reach the same test set size (approximately 34) indicated by the dashed vertical line on each plot. For CDFS, all classifiers perform reasonably well across cutoffs, with performance dropping as the size of the test set becomes small. For ELAIS-S1, logistic regression and convolutional neural networks perform comparably across all SNR cutoffs, but random forests do not. While random forests trained on Radio Galaxy Zoo labels perform comparably to other classifiers across all SNR cutoffs, random forests trained on expert labels show a considerable drop in performance below the dashed line.

4.5 Discussion

Based on the ATLAS sample, our main result is that it is possible to cast radio host galaxy cross-identification as a machine learning task for which standard methods can be applied. These methods can then be trained with a variety of label sets derived from cross-identification catalogues. While our methods have not outperformed nearest-neighbours, we have demonstrated that for a very accurate binary classifier, good cross-identification results can be obtained using our method. Future work could combine multiple catalogues or physical priors to boost performance.

Nearest-neighbours approaches outperform most methods we investigated, notably including Radio Galaxy Zoo. This is due to the large number of compact or partially-resolved objects in ATLAS. This result shows that for compact and partially-resolved objects methods that do not use machine learning such as a nearest-neighbours approach or likelihood ratio (Weston et al., 2018) should be preferred to machine learning methods. It also shows that ATLAS is not an ideal data set for developing machine learning methods like ours. Our use of ATLAS is motivated by its status as a pilot survey for EMU, so methods developed for ATLAS should also work for EMU. New methods developed should work well with extended radio sources, but this goal is almost unsupported by ATLAS as it has very few examples of such sources. This makes both training and testing difficult — there are too few extended sources to train on and performance on such a small test set may be unreliable. Larger data sets with many extended sources like FIRST exist, but these are considerably less deep than and at a different resolution to EMU, so there is no reason to expect methods trained on such data sets to be applicable to EMU.

The accuracies of our trained cross-identification methods generally fall far below the estimated best possible accuracy attainable using our approach, indicated by the green-shaded areas in Figures 4.13 and 4.15. The balanced accuracies attained by our binary classifiers indicate that there is significant room for improvement in classification. The classification accuracy could be improved by better model selection and more training data, particularly for convolutional neural networks. There is a huge variety of ways to build a convolutional neural network, and we have only investigated one architecture. For an exploration of different convolutional neural network architectures applied to radio astronomy, see Lukic et al. (2018). Convolutional neural networks generally require more training data than other machine learning models and we have only trained our networks on a few hundred sources. We would expect performance on the classification task to greatly increase with larger training sets.

Another problem is that of the window size used to select radio features. Increasing window size would increase computational expense, but provide more information to the models. Results are also highly sensitive to how large the window size is compared to the size of the radio source we are trying to cross-identify, with large angular sizes requiring large window sizes to ensure that the features contain all the information needed to localise the host galaxy. An ideal implementation of our method would most likely represent a galaxy using radio images taken at multiple window sizes, but this is considerably more expensive.

Larger training sets, better model selection, and larger window sizes would improve performance, but only so far: we would still be bounded above by the estimated ‘perfect’ classifier accuracy. From this point, the performance can only be improved by addressing our broken assumptions. We detailed these assumptions in Section 4.3.2, and we will discuss here how our method could be adapted to avoid these assumptions. Our assumption that the host galaxy is contained within the search radius could be improved by dynamically choosing the search radius, perhaps based on the angular extent of the radio emission, or the redshift of candidate hosts. Radio morphology information may allow us to select relevant radio data and hence relax the assumption that a 1'-wide radio image represents just one, whole radio source. Finally, our assumption that the host galaxy is detected in infrared is technically not needed, as the sliding-window approach we have employed will still work even if there are no detected host galaxies — instead of classifying candidate hosts, simply classify each pixel in the radio image. The downside of removing candidate hosts is that we are no longer able to reliably incorporate host galaxy information such as colour and redshift, though this could be resolved by treating pixels as potentially undetected candidate hosts with noisy features.

We observe that Radio Galaxy Zoo-trained methods perform comparably to methods trained on expert labels. This shows that the crowdsourced labels from Radio Galaxy Zoo will provide a valuable source of training data for future machine learning methods in radio astronomy.

Compared to nearest-neighbours, cross-identification accuracy on ELAIS-S1 is lower than on CDFS. Particularly notable is that our performance on compact objects is very low for ELAIS-S1, while it was near-optimal for CDFS. These differences may

be for a number of reasons. ELAIS-S1 has beam size and noise profile different from CDFS (even though both were imaged with the same telescope), so it is possible that our methods over-adapted to the beam and noise of CDFS. Additionally, CDFS contains a very bright source which may have caused artefacts throughout the field that are not present in ELAIS-S1. Further work is required to understand the differences between the fields and their effect on performance.

Figure 4.16 reveals interesting behaviour of different classifier models at different flux cutoffs. Logistic regression and convolutional neural networks seem relatively independent of flux, with these models performing well on the fainter ELAIS-S1 components even when they were trained on the generally brighter components in CDFS. Conversely, random forests were sensitive to the changes in flux distribution between datasets. This shows that not all models behave similarly on radio data, and it is therefore important to investigate multiple models when developing machine learning methods for radio astronomy.

Appendix ?? (see Figure 4.18) shows examples of incorrectly cross-identified components in CDFS. On no such component do all classifiers agree. This raises the possibility of using the level of disagreement of an ensemble of binary classifiers as a measure of the difficulty of cross-identifying a radio component, analogous to the consensus level for Radio Galaxy Zoo volunteers.

Our methods can be easily incorporated into other cross-identification methods or used as an extra data source for source detection. For example, the scores output by our binary classifiers could be used to disambiguate between candidate host galaxies selected by model-based algorithms, or used to weight candidate host galaxies while a source detector attempts to associate radio components. Our method can also be extended using other data sources: for example, information from source identification algorithms could be incorporated into the feature set of candidate host galaxies.

4.6 Summary

We presented a machine learning approach for cross-identification of radio components with their corresponding infrared host galaxy. Using the CDFS field of ATLAS as a training set we trained our methods on expert and crowdsourced cross-identification catalogues. Applying these methods on both fields of ATLAS, we found that:

- Our method trained on ATLAS observations of CDFS generalised to ATLAS observations of ELAIS-S1, demonstrating that training on a single patch of sky is a feasible option for training machine learning methods for wide-area radio surveys;
- Performance was comparable to nearest-neighbours even on resolved sources, showing that nearest-neighbours is useful for datasets consisting mostly of unresolved sources such as ATLAS and EMU;

-
- Radio Galaxy Zoo-trained models performed comparably to expert-trained models and outperformed Radio Galaxy Zoo, showing that crowdsourced labels are useful for training machine learning methods for cross-identification even when these labels are noisy;
 - ATLAS does not contain sufficient data to train or test machine learning cross-identification methods for extended radio sources. This suggests that if machine learning methods are to be used on EMU, a larger area of sky will be required for training and testing these methods. However, existing surveys like FIRST are likely too different from EMU to expect good generalisation.

While our cross-identification performance is not as high as desired, we make no assumptions on the binary classification model used in our methods and so we expect the performance to be improved by further experimentation and model selection. Our method provides a useful framework for generalising cross-identification catalogues to other areas of the sky from the same radio survey and can be incorporated into existing methods. We have shown that citizen science can provide a useful dataset for training machine learning methods in the radio domain. Chapter 5 will extend this approach and confirm that dataset size is a key limitation by successfully applying it to a considerably larger dataset: FIRST.

4.7 Acknowledgements

This chapter and corresponding publication was made possible by the participation of more than 11 000 volunteers in the Radio Galaxy Zoo project. Their contributions are individually acknowledged at <http://rgzauthors.galaxyzoo.org>. Parts of this research were conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020. Partial support for LR was provided by U.S. National Science Foundation grants AST1211595 and 1714205 to the University of Minnesota. HA benefitted from grant 980/2016-2017 of Universidad de Guanajuato. We thank A. Tran and the reviewer for their comments on the manuscript this chapter was based on. Radio Galaxy Zoo makes use of data products from the Wide-field Infrared Survey Explorer and the Very Large Array. The Wide-field Infrared Survey Explorer 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. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The figures in this work made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013). The Australia Telescope Compact Array is part of the Australia Telescope, which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO.

A Classification models

We use three different models for binary classification: logistic regression, convolutional neural networks, and random forests.

A.1 Logistic Regression

Logistic regression is linear in the feature space and outputs the probability that the input has a positive label. The model is (Bishop, 2006):

$$f(\vec{x}) = \sigma(\vec{w}^T \vec{x} + b) , \quad (4.5)$$

where $\vec{w} \in \mathbb{R}^D$ is a vector of parameters, $b \in \mathbb{R}$ is a bias term, $\vec{x} \in \mathbb{R}^D$ is the feature vector representation of a candidate host, and $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is the logistic sigmoid function:

$$\sigma(a) = (1 + \exp(-a))^{-1} . \quad (4.6)$$

The logistic regression model is fully differentiable, and the parameters \vec{w} can therefore be learned using gradient-based optimisation methods. We used the `scikit-learn` (Pedregosa et al., 2011) implementation of logistic regression with balanced classes.

A.2 Convolutional neural networks

Convolutional neural networks (CNN) are a biologically-inspired prediction model for prediction with image inputs. The input image is convolved with a number of filters to produce output images called feature maps. These feature maps can then be convolved again with other filters on subsequent layers, producing a network of convolutions. The whole network is differentiable with respect to the values of the filters and the filters can be learned using gradient-based optimisation methods. The final layer of the network is logistic regression, with the convolved outputs as input features. For more detail, see LeCun et al. (subsection II.A 1998). We use KERAS (Chollet et al., 2015) to implement our CNN, accounting for class imbalance by reweighting the classes.

CNNs have recently produced good results on large image-based datasets in astronomy (Lukic et al., 2018; Dieleman, Willett, and Dambre, 2015, e.g.). We employ only a simple CNN model in this paper as a proof of concept that CNNs may be used for class probability prediction on radio images. The model architecture we use is shown in Figure 4.17.

A.3 Random Forests

Random forests are an ensemble of decision trees (Breiman, 2001). They consider multiple subsamples of the training set, where each subsample is sampled with replacement from the training set. For each subsample a decision tree classifier is constructed by repeatedly making axis-parallel splits based on individual features. In a random forest the split decision is taken based on a random subset of features.

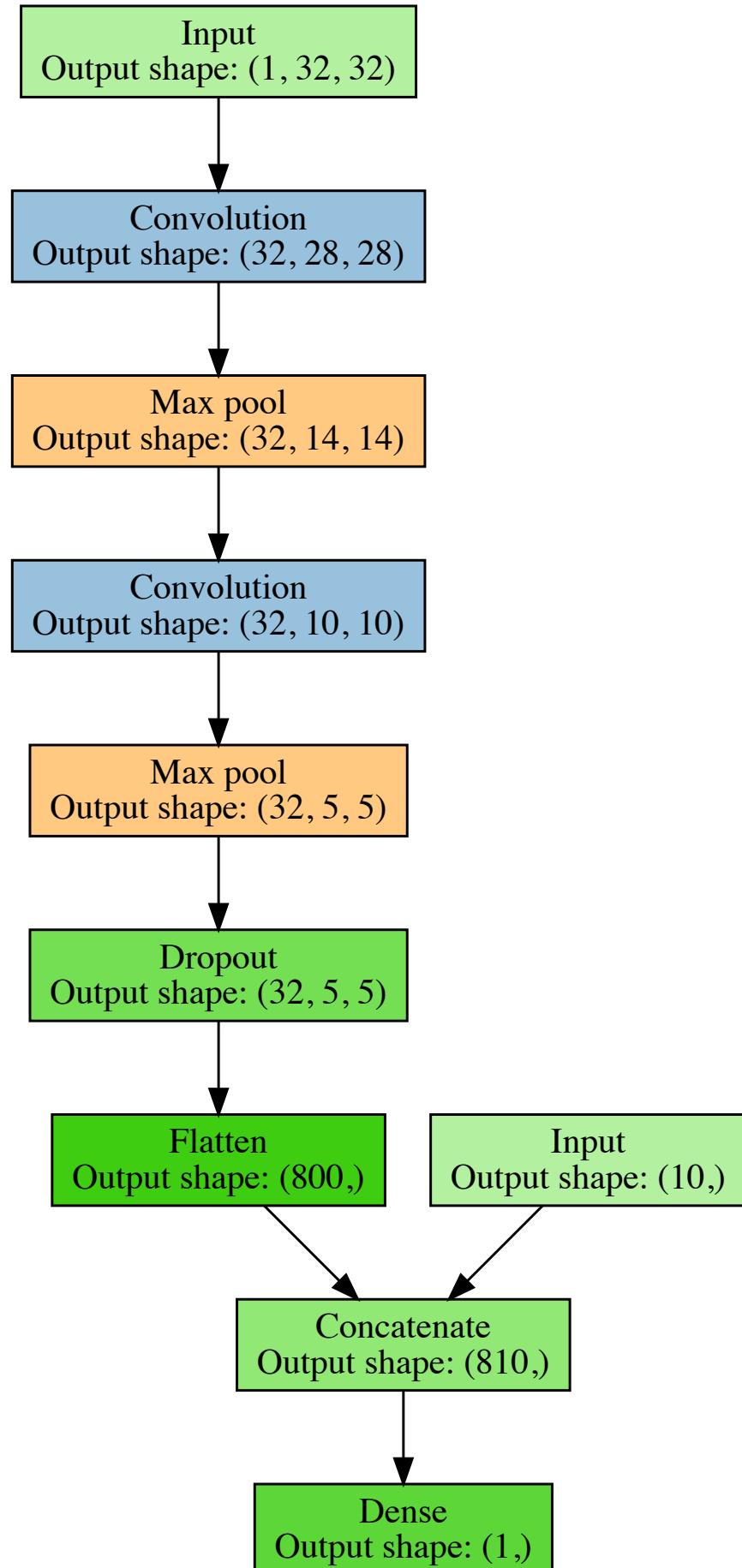


Table 4.3: Balanced accuracies for different binary classification models trained and tested on SWIRE objects in CDFS. The ‘Labeller’ column states what set of training labels were used to train the classifier, and the ‘Classifier’ column states what classification model was used. ‘CNN’ is a convolutional neural network, ‘LR’ is logistic regression and ‘RF’ is random forests. Accuracies are evaluated against the expert label set derived from Norris et al. (2006). The standard deviation of balanced accuracies evaluated across the four quadrants of CDFS (Figure 4.9) is also shown. The ‘compact’ set refers to SWIRE objects within $1'$ of a compact radio component, the ‘resolved’ set refers to SWIRE objects within $1'$ of a resolved radio component, and ‘all’ is the union of these sets.

Labeller	Classifier	Mean ‘Compact’ accuracy (per cent)	Mean ‘Resolved’ accuracy (per cent)	Mean ‘All’ accuracy (per cent)
Norris	LR	91.5 ± 1.0	93.2 ± 1.0	93.0 ± 1.2
Norris	CNN	92.6 ± 0.7	91.2 ± 0.5	92.0 ± 0.6
Norris	RF	96.7 ± 1.5	91.0 ± 4.5	96.0 ± 2.5
RGZ	LR	89.5 ± 0.8	90.5 ± 1.7	90.2 ± 0.8
RGZ	CNN	89.4 ± 0.6	89.6 ± 1.3	89.4 ± 0.5
RGZ	RF	94.5 ± 0.2	95.8 ± 0.4	94.7 ± 0.3

To classify a new data point, the random forest takes the weighted average of all classifications produced by each decision tree. We used the `scikit-learn` (Pedregosa et al., 2011) implementation of random forests with 10 trees, the information entropy split criterion, a minimum leaf size of 45 and balanced classes.

B Accuracy tables

This section contains tables of accuracy for our method applied to CDFS and ELAIS-S1. In Table 4.3 and Table 4.4 we list the balanced accuracies of classifiers on the cross-identification task for CDFS and ELAIS-S1 respectively, averaged over each set of training quadrants. In Table 4.5 and Table 4.6 we list the balanced accuracies of classifiers on the cross-identification task for CDFS and ELAIS-S1 respectively, averaged over each set of training quadrants.

C SWIRE object scores

This section contains scores predicted by our binary classifiers for each SWIRE object within $1'$ of a radio component in CDFS and ELAIS-S1. Scores for SWIRE CDFS objects are shown in Table 4.7 and scores for SWIRE ELAIS-S1 are shown in Table 4.8. For CDFS, the score for an object in a quadrant is predicted by binary classifiers trained on all other quadrants. For ELAIS-S1, we show the scores predicted by binary classifiers trained on each CDFS quadrant. Note that these scores have *not* been weighted by Gaussians.

The columns of the score tables are defined as follows:

Table 4.4: Balanced accuracies for different binary classification models trained on SWIRE objects in CDFS and tested on SWIRE objects in ELAIS-S1. Columns and abbreviations are as in Table 4.3. Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008). The standard deviations of balanced accuracies of models trained on the four subsets of CDFS (Figure 4.9) are also shown.

Labeller	Classifier	Mean ‘Compact’ accuracy (per cent)	Mean ‘Resolved’ accuracy (per cent)	Mean ‘All’ accuracy (per cent)
Norris	LR	94.6 ± 0.4	93.3 ± 2.0	95.3 ± 0.1
	CNN	94.8 ± 0.2	92.8 ± 0.5	94.4 ± 0.2
	RF	85.9 ± 3.8	70.0 ± 2.8	86.6 ± 3.2
RGZ	LR	91.8 ± 0.3	91.9 ± 0.5	92.0 ± 0.2
	CNN	90.1 ± 0.3	91.1 ± 0.9	90.2 ± 0.3
	RF	95.1 ± 0.1	95.2 ± 0.0	95.2 ± 0.3

Table 4.5: Cross-identification accuracies for different classification models on CDFS. The ‘Labeller’ column states what set of training labels were used to train the method, and the ‘Classifier’ column states what classification model was used. ‘CNN’ is a convolutional neural network, ‘LR’ is logistic regression, ‘RF’ is random forests, and ‘Labels’ is the accuracy of the label set itself. ‘Perfect’ indicates that the true labels of the test set were used and hence represents an upper bound on cross-identification accuracy with our method. ‘NN’ is a nearest-neighbours approach. Accuracies are evaluated against the expert label set, so ‘Norris’ labels are 100 per cent accurate by definition. The standard deviation of accuracies evaluated across the four quadrants of CDFS (Figure 4.9) is also shown.

Labeller	Classifier	Mean ‘Compact’ accuracy (per cent)	Mean ‘Resolved’ accuracy (per cent)	Mean ‘All’ accuracy (per cent)
Norris	NN	97.2 ± 1.7	75.7 ± 7.9	93.4 ± 0.8
	Random	97.9 ± 2.2	22.3 ± 9.2	83.2 ± 4.7
	Labels	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
	Perfect	97.9 ± 2.2	99.0 ± 1.8	98.1 ± 1.7
	LR	97.3 ± 0.5	76.0 ± 3.2	93.7 ± 1.8
	CNN	96.6 ± 0.9	74.3 ± 12.3	93.5 ± 0.5
RGZ	RF	96.1 ± 1.4	75.8 ± 6.7	93.8 ± 2.0
	Labels	53.1 ± 8.5	56.7 ± 5.9	54.4 ± 5.9
	LR	97.3 ± 1.9	74.5 ± 5.1	93.6 ± 1.7
	CNN	85.4 ± 2.6	68.1 ± 9.2	92.4 ± 1.1
	RF	97.5 ± 0.9	74.3 ± 7.9	93.7 ± 1.5

Table 4.6: Cross-identification accuracies for different classification models on ELAIS-S1. Columns and abbreviations are as in Table 4.5. Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008) cross-identifications. The standard deviation of accuracies evaluated across models trained on the four quadrants of CDFS (Figure 4.9) is also shown.

Labeller	Classifier	Mean ‘Compact’ accuracy (per cent)	Mean ‘Resolved’ accuracy (per cent)	Mean ‘All’ accu (per cent)
—	NN	95.5 ± 0.0	92.8 ± 0.0	95.5 ± 0.0
—	Random	61.9 ± 1.1	26.6 ± 2.1	61.9 ± 1.1
Middelberg	Perfect	99.6 ± 0.0	99.8 ± 0.0	99.6 ± 0.0
Norris	LR	89.0 ± 1.1	89.7 ± 1.8	94.4 ± 0.9
	CNN	89.7 ± 0.3	89.4 ± 1.4	94.3 ± 0.7
	RF	83.8 ± 5.6	82.3 ± 4.1	90.6 ± 2.1
RGZ	LR	90.5 ± 1.0	92.7 ± 0.2	95.9 ± 0.1
	CNN	84.6 ± 0.6	84.6 ± 0.6	91.8 ± 0.3
	RF	91.3 ± 1.0	90.3 ± 2.4	94.7 ± 1.2

- *SWIRE* — SWIRE designation for candidate host galaxy.
- *RA* — Right ascension (J2000).
- *Dec* — Declination (J2000).
- *Expert host* — Whether the candidate host galaxy is a host galaxy according to Norris et al. (2006) or Middelberg et al. (2008) cross-identifications of CDFS and ELAIS-S1 respectively.
- *RGZ host* — Whether the candidate host galaxy is a host galaxy according to Radio Galaxy Zoo cross-identifications (Wong et al. in prep). This is always “no” for ELAIS-S1 objects.
- $C(L / D)$ — Score assigned by binary classifier C trained on label set L of D candidate host galaxies. C may be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or random forests respectively. L may be ‘Norris’ or ‘RGZ’ for expert and Radio Galaxy Zoo labels respectively. D may be ‘All’, ‘Compact’ or ‘Resolved’ for each respective subset defined in ??.

D ATLAS component cross-identifications

This section contains cross-identifications predicted by our method for each ATLAS radio component in CDFS and ELAIS-S1. Cross-identifications for ATLAS CDFS components are shown in Table 4.9 and cross-identifications for ATLAS ELAIS-S1 are shown in Table 4.10. For CDFS, the cross-identification for a component in a quadrant is predicted using our method with binary classifiers trained on all other

Table 4.7: Scores output by our trained classifiers for SWIRE CDFS candidate host galaxies. Columns are defined in Section C. Full table electronic.

SWIRE	RA	Dec	Expert host	RGZ host	CNN(Norris / All)	CNN(Norris / C)
J032603.15-284708.5	51.5132	-28.7857	yes	no	0.5838	0.4697
J032603.39-284010.1	51.5142	-28.6695	no	no	0.0373	0.5814
J032603.44-284210.1	51.5144	-28.7028	no	no	0.0232	0.4891
J032603.44-284222.2	51.5143	-28.7062	no	no	0.0006	0.4164
J032603.45-284748.4	51.5144	-28.7968	no	no	0.0014	0.4914
J032603.50-284637.0	51.5146	-28.7770	no	no	0.0074	0.4144
J032603.60-284627.4	51.5150	-28.7743	no	no	0.0012	0.4578
J032603.63-283840.5	51.5151	-28.6446	no	no	0.0021	0.4153
J032603.66-283822.8	51.5153	-28.6397	no	no	0.0001	0.4752
J032603.75-284014.1	51.5156	-28.6706	no	no	0.0547	0.3408

Table 4.8: Scores output by our trained classifiers for SWIRE ELAIS-S1 candidate host galaxies. Columns are defined in Section C. Full table electronic.

SWIRE	RA	Dec	Expert host	RGZ host	CNN(Norris / All)	CNN(Norris / Co)
J002925.73-440256.2	7.3572	-44.0490	yes	no	0.9537	0.8638
J002926.14-440249.0	7.3590	-44.0470	no	no	0.7361	0.8752
J002926.52-440247.0	7.3605	-44.0464	no	no	0.3390	0.8338
J002926.63-440301.1	7.3610	-44.0503	no	no	0.2108	0.8251
J002927.13-440232.6	7.3631	-44.0424	no	no	0.0339	0.8479
J002927.28-440245.3	7.3637	-44.0459	no	no	0.0406	0.8345
J002927.44-440238.5	7.3644	-44.0440	no	no	0.0116	0.8267
J002928.08-440230.3	7.3670	-44.0418	no	no	0.0024	0.8626
J002928.11-440312.7	7.3671	-44.0535	no	no	0.0011	0.8159
J002928.80-440306.8	7.3700	-44.0519	no	no	0.0003	0.8405

quadrants. For ELAIS-S1, we show the cross-identifications predicted by our method using binary classifiers trained on each CDFS quadrant. For CDFS, we also show the Radio Galaxy Zoo consensus, which is a proxy for the difficulty of cross-identifying a component (Wong et al. in prep).

The columns of the cross-identification tables are defined as follows:

- *ATLAS* — ATLAS designation for radio component.
- *RA* — Right ascension of radio component (J2000).
- *Dec* — Declination of radio component (J2000).
- *CID* — Radio Galaxy Zoo component ID.
- *Zooniverse ID* — Radio Galaxy Zoo Zooniverse ID.
- *Norris/Middleberg* — Designation of SWIRE cross-identification from Norris et al. (2006) or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- *Norris/Middleberg RA* — Right ascension of SWIRE cross-identification from Norris et al. (2006) or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- *Norris/Middleberg Dec* — Right ascension of SWIRE cross-identification from Norris et al. (2006) or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- *RGZ* — Designation of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al. in prep).
- *RGZ RA* — Right ascension of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al. in prep).
- *RGZ Dec* — Right ascension of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al. in prep).
- *RGZ radio consensus* — Percentage agreement of Radio Galaxy Zoo volunteers on the radio component configuration.
- *RGZ IR consensus* — Percentage agreement of Radio Galaxy Zoo volunteers on the host galaxy of this radio component.
- *C(L / D)* — Designation of SWIRE cross-identification made by our method using classification model C trained on label set L of D candidate host galaxies. C may be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or random forests respectively. L may be ‘Norris’ or ‘RGZ’ for expert and Radio Galaxy Zoo labels respectively. D may be ‘All’, ‘Compact’ or ‘Resolved’ for each respective subset defined in ??.

Table 4.9: Cross-identifications for ATLAS CDFS components. Columns are defined in Section D. Full table electronic.

ATLAS	RA	Dec	CID	Zooniverse ID	Norris	Norris RA
J032602.82-284708.1C	51.5117	-28.7856	CI0412	ARG0003rb2	J032603.15-284708.5	51.5132
J032615.49-284629.4C	51.5646	-28.7749	CI0614	ARG0003rfr	J032615.41-284630.7	51.5642
J032615.55-280559.8C	51.5648	-28.1000	CI0320	ARG0003r8s	J032615.52-280559.8	51.5647
J032617.35-280710.2C	51.5723	-28.1195	CI0059C1	ARG0003r2j	J032617.89-280707.2	51.5746
J032625.13-280909.8C	51.6047	-28.1527	CI0409	ARG0003raz	J032625.19-280910.1	51.6050
J032629.10-280650.1C	51.6213	-28.1139	CI0963	ARG0003ro4	J032629.13-280650.7	51.6214
J032629.61-284052.7C	51.6234	-28.6813	CI0304	ARG0003r8e	J032629.54-284055.8	51.6231
J032629.92-284753.5C	51.6247	-28.7982	CI0120	ARG0003r3w	J032629.81-284754.4	51.6242
J032630.66-283657.3C	51.6278	-28.6159	CI0172C1	ARG0003r55	J032630.64-283658.0	51.6277
J032634.59-282022.8C	51.6441	-28.3397	CI0757	ARG0003rj2	J032634.58-282022.8	51.6441

Table 4.10: Cross-identifications for ATLAS ELAIS-S1 components. Columns are defined in Section D. Full table electronic.

ATLAS	RA	Dec	CID	Zooniverse ID	Middleberg	Middleberg RA
J002925.68-440256.8	7.3570	-44.0491	C0375		J002925.73-440256.2	7.3572
J002938.19-432946.7	7.4092	-43.4963	C0832		J002938.07-432947.9	7.4087
J002940.13-440309.2	7.4172	-44.0526	C0374		J002940.19-440309.6	7.4175
J002943.14-440812.3	7.4298	-44.1368	C0302		J002943.15-440813.6	7.4298
J002944.51-433627.8	7.4355	-43.6077	C0727		J002944.36-433630.2	7.4348
J002945.31-432148.5	7.4388	-43.3635	C0943.1		J002945.64-432149.3	7.4402
J002946.14-432149.1	7.4423	-43.3637	C0943		J002945.64-432149.3	7.4402
J002949.89-440541.4	7.4579	-44.0948	C0345			
J002951.13-432354.3	7.4631	-43.3984	C0924		J002951.14-432355.3	7.4631
J002951.19-440556.6	7.4633	-44.0991	C0342		J002951.26-440556.4	7.4636

- $C(L / D) RA$ — Right ascension (J2000) of SWIRE cross-identification made by our method using classification model C trained on label set L of D candidate host galaxies. C , L and D are defined as for designation.
- $C(L / D) Dec$ — Declination (J2000) of SWIRE cross-identification made by our method using classification model C trained on label set L of D candidate host galaxies. C , L and D are defined as for designation.

E Cross-identification figures

This section contains figures of cross-identifications of each ATLAS radio component in CDFS and ELAIS-S1.

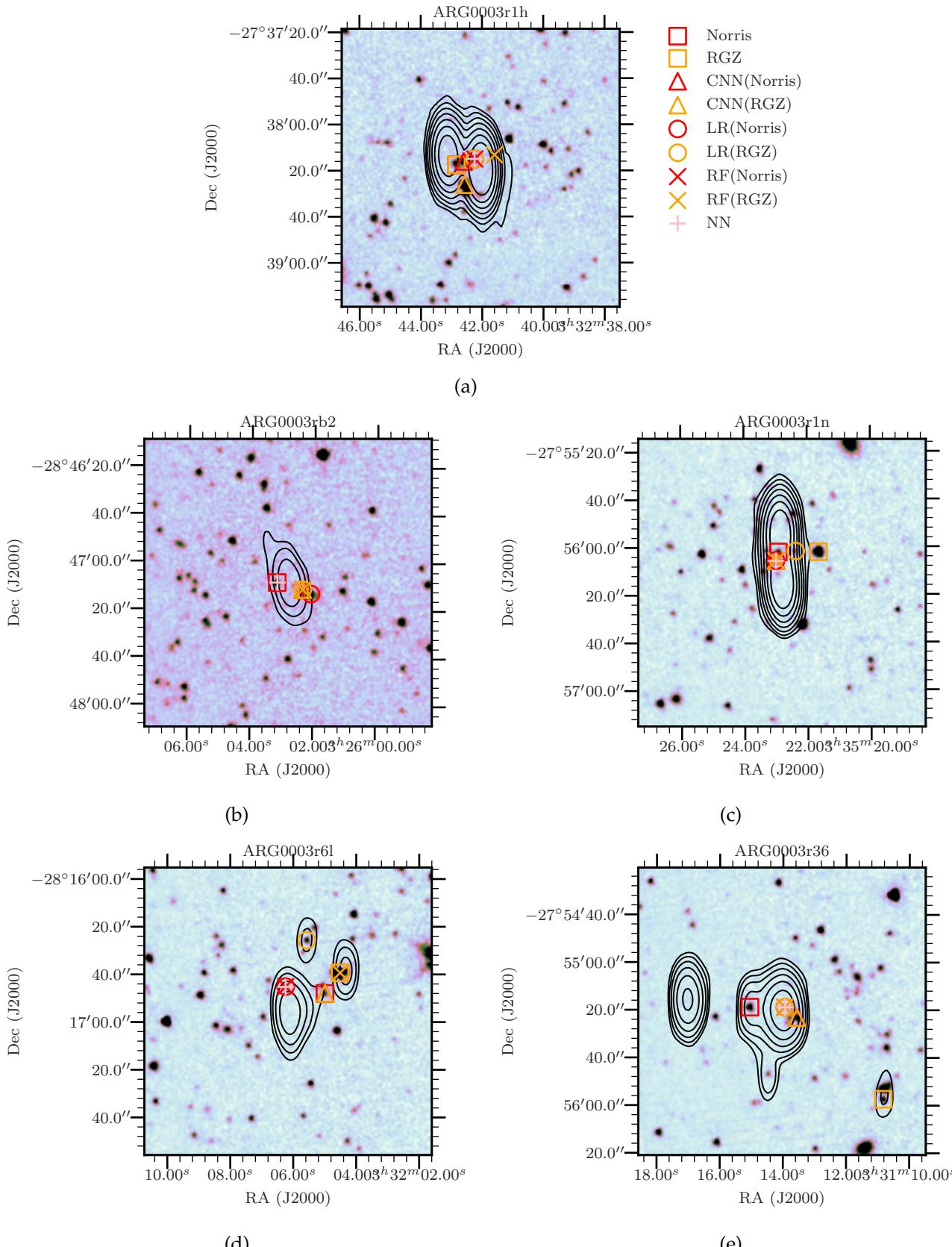


Figure 4.18: Examples of resolved sources with high disagreement between cross-identifiers. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2. The background image is the $3.6 \mu\text{m}$ SWIRE image. Binary classifier model/training set combinations are denoted $C(S)$ where C is the binary classifier model and S is the training set. ‘LR’ is logistic regression, ‘CNN’ is convolutional neural networks, and ‘RF’ is random forests. ‘Norris’ refers to the expert labels

Radio Luminosity Functions

1 Radio Luminosity Functions

make James' picture of binary xid before partitioning and colour it with a value-ramped viridis.

2 Making Catalogues with Binary Cross-identification

3 FIRST paper...

We use a machine learning model trained on the Radio Galaxy Zoo citizen science project to cross-identify 244 846 radio components of extended radio sources from the Faint Images of the Radio Sky at Twenty-centimeters survey (FIRST) with 158 337 host galaxies observed in the mid-infrared from the AllWISE source catalogue. 34 305 hosts have a spectroscopic redshift in the Sloan Digital Sky Survey (SDSS). This is the largest sample of cross-identified extended radio sources to date. The large sample size allows us to estimate the radio luminosity functions (RLF) of extended radio galaxies and to further characterise the sample with fractional radio luminosity functions. The extended source RLF can be used to estimate the mechanical energy injected into the local environment of active galactic nuclei (AGN). We estimate from the observed radio emission that inferred extended radio jets from AGN contribute between 1.3×10^{30} and 1.2×10^{32} W Mpc $^{-3}$ of mechanical energy to their environment throughout the low-redshift ($z < 0.6$) Universe. By further visual verification we also find 40 radio galaxies with projected size larger than 1 Mpc. Our results directly demonstrate the impact that automated machine learning methods could have on future wide-area radio surveys.

4 Introduction

Radio active galactic nuclei (AGN) are some of the most violent and energetic objects in the Universe. Supermassive black holes at the centre of galaxies can produce large, luminous radio jets that may feed lobes extending up to megaparsec scales. As

a vital part of galaxy evolution, we want to understand how radio AGN interact with their host galaxies, in what kinds of galaxies these AGN are found, and how these populations change over cosmic time.

Mechanical energy from AGN is needed to explain the star-formation history of galaxies throughout the Universe (e.g. Raouf et al., 2017; Hardcastle and Croston, 2020b). The radio luminosity function (RLF) of extended radio sources would characterise the energy output of galaxies capable of significant energy impact into their local environment. The local RLF of primarily compact sources is estimated and described in detail by Mauch and Sadler (2007a), Pracy et al. (2016) and Condon, Matthews, and Broderick (2019), each with complete volume-limited samples. However, the RLF of extended radio sources has not been characterised as a separate sample, because estimating this RLF would need a large number of extended sources with known redshifts. Such redshifts are usually obtained by cross-matching the radio sources with their host galaxies in the optical or infrared. While this cross-matching is relatively straightforward for compact/unresolved sources, and can therefore be automated (e.g. Kimball and Ivezić, 2008), extended radio emission may be complex and may not be co-located with the host galaxy. Cross-identification of complex, extended sources is therefore usually done manually, limiting the sample size.

Wide-area radio surveys like the Very Large Array Faint Images of the Sky at Twenty-centimeters survey (FIRST; White et al., 1997) have led to catalogues of hundreds of thousands of radio objects, most of which are associated with radio AGN (Sadler et al., 2002). While individual galaxy properties vary significantly between galaxies, the large sample size of such surveys allows us to examine reliable bulk statistics that may tightly constrain models and theory of radio galaxy evolution and formation (Condon, 1992b). With a sufficiently large sample (of size N such that $N \gg \sqrt{N}$; Condon, 1991) we can also divide the RLF into fractions based on the physical properties of each galaxy. These *fractional RLFs* show how different physical processes comprise the luminosity distribution, and can be used to investigate how these processes relate to the properties of the AGN.

In this paper we calculate the RLF for extended radio sources in FIRST. Throughout this paper we define an *extended radio source* as a collection of extended radio components with the same host galaxy, and following Banfield et al. (2015) we define an *extended radio component* as a radio component which fulfils Equation 5.1:

$$\frac{S_{\text{peak}}}{S_{\text{int}}} < 1 - \frac{0.1}{\log_{10}(S_{\text{peak}}/1 \text{ mJy})}, \quad (5.1)$$

where S_{peak} is the peak radio flux density and S_{int} is the integrated radio flux density. We define *radio components* as Gaussians fit to radio emission, *radio islands* as connected patches of radio emission above a local 4σ value, and *radio sources* as sets of radio islands or components associated with the same galaxy. The radio flux density of a source is the sum of the flux densities of the components according to the FIRST catalogue. The (*infrared*) *host galaxy* of a radio source or component is de-

fined as the infrared galaxy associated with the radio emitter. A *cross-identification* is an association of a host galaxy with one or more radio components or islands. We define *candidate host galaxies* (or simply *candidates*) as infrared objects that are near a radio component on the sky and thus may potentially be the host galaxy of that component.

Upcoming radio surveys such as the Evolutionary Map of the Universe (EMU; Norris et al., 2011a) are expected to increase the number of complex radio sources to around 7 million (Banfield et al., 2015). Manual, expert cross-identification for such surveys will be impractical. One way forward could be to ask non-expert volunteers for help with manual cross-identification, which is the approach taken by Radio Galaxy Zoo (RGZ; Banfield et al., 2015). This is called *citizen science* and has been employed successfully in many fields with large datasets (Marshall, Lintott, and Fletcher, 2015b). Even this approach is not sufficient for 7 million sources, though, with RGZ cross-identifying around 75 000 sources in four years. Machine learning provides a potential pathway to obtaining useful physics from such large samples.

In this paper, we train a machine learning model based on Alger et al. (2018), using RGZ as training data, and use this model to automatically cross-identify 244 846 radio components catalogued by FIRST. This results in 34 305 sources with spectroscopic redshifts. This is the largest available catalogue of extended radio source cross-identifications. We call our catalogue *RGZ-extrapolated* or *RGZ-Ex*. Due to our large sample size we are able to further divide the source population by properties of the host galaxies. We refer to the trained machine learning model and the associated cross-identification algorithm jointly as *binary cross-identification*, or *BXID*. Note that there are two stages to our automated approach: first, we train BXID using an existing catalogue of cross-identified sources; second, we generate new cross-identifications for radio components not in the training catalogue. We refer to these stages as *training* and *prediction* respectively. Our data sources are described in Section 5. Our approach is discussed in Section 6. Radio luminosity functions of extended sources are presented in Section 7 and we discuss these functions in Section 8.

Throughout this paper we assume a flat Λ -CDM cosmology of $H_0 = 69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.287$. These are the cosmological parameters from the Nine-year Wilkinson Microwave Anisotropy Probe (WMAP9; Hinshaw et al., 2013).

5 Data

In this section we describe how we obtain our training and prediction data. To enable the estimation of the extended RLF, we apply a number of selection criteria to the data which are described in Section F.

5.1 RGZ

RGZ is a citizen science project that aims to cross-identify complex radio sources with mid-infrared host galaxies with the help of volunteers. The first RGZ data release contains around 75 000 cross-identifications of a random subset of extended

sources in FIRST with their host galaxies in AllWISE. For more details on RGZ see Banfield et al. (2015). This catalogue has also been used in other machine learning contexts, including supervised learning for source aggregation (Wu et al., 2019) and unsupervised learning methods (Galvin et al., 2019b; Ralph et al., 2019). We discard the RGZ sources without a host galaxy detected in AllWISE and use the remaining 41 446 sources for training.

The RGZ catalogue only contains sources with at least 0.65 weighted volunteer agreement, equivalent to approximately 80 per cent reliability. This implicitly selects for less complex sources, since the volunteer agreement is a proxy for the difficulty of cross-identifying a source (Wong et al. in prep).

5.2 FIRST

We use radio imagery and select radio components from the FIRST survey and associated catalogue respectively (White et al., 1997; Helfand, White, and Becker, 2015). FIRST is a 1.4 GHz radio survey covering 10 575 deg² of the sky north of Dec = -10° with an angular resolution of 5.4''. At a detection limit of 1 mJy, the catalogue contains 946 432 radio components.

For both training and prediction, we make use of image cutouts from FIRST centred on mid-infrared candidate hosts. We predict host galaxies for the 244 846 extended FIRST components detected at $> 10\sigma$ (about 1.5 mJy beam⁻¹, per Banfield et al., 2015) that have complete radio imagery for all candidate hosts.

5.3 AllWISE

We use the sky coordinates and magnitudes from the AllWISE (Cutri et al., 2013) catalogue during training and prediction. AllWISE is an all-sky catalogue of mid-infrared objects detected by the *Wide-field Infrared Survey Explorer* (WISE Wright et al., 2010) at 3.4, 4.6, 12 and 22 μm wavelengths (called W1–W4, respectively). AllWISE contains over 747 million objects detected above 5σ at 3.4 and 4.6 μm . WISE has an angular resolution of 6.1'' and 6.4'' at these wavelengths, respectively.

We store AllWISE objects in a MongoDB¹ database with a geospatial index on the right ascension and declination. This allows us to perform fast spatial lookups. The geospatial index in MongoDB assumes a perfectly spherical Earth with a fixed radius, allowing us to use it for sky coordinate searches. We use the right ascension and declination of AllWISE sources to generate candidate hosts by searching for infrared sources near FIRST components.

5.4 SDSS

While we do not use data from the Sloan Digital Sky Survey Data Release 15 (SDSS; Aguado et al., 2019) for training or prediction, we do use SDSS for spectroscopic redshifts of our host galaxies. These redshifts are required to calculate the radio

¹<https://www.mongodb.com/>

Table 5.1: Medians and standard deviations used to normalise input features.

Feature	Median	Standard deviation
Radio image	13.2 μ Jy	3.01 mJy
$W_1 - W_2$	0.289 mag	0.378 mag
$W_1 - W_3$	4.350 mag	1.067 mag
$W_1 - W_4$	7.853 mag	1.144 mag
$W_2 - W_3$	4.016 mag	0.958 mag
$W_2 - W_4$	7.541 mag	1.046 mag
$W_3 - W_4$	3.518 mag	0.409 mag
W_1	16.659 mag	1.154 mag

luminosity of our sources. We use CDS X-Match² to match each infrared host galaxy to the closest source imaged by SDSS to within 5''. This results in 34 305 spectroscopic redshifts for our 158 337 total host galaxies. To estimate the rate of false association with SDSS sources, we add a 1' offset to all host positions and redo the matching process. With this method we estimate a 0.4 per cent rate of false association.

6 Method

We apply the binary classification cross-identification method (BXID) following Alger et al. (2018). This method casts cross-identification as a classification problem where infrared ‘candidate host galaxies’ are classified as either being host galaxies or not. A classifier is trained on examples of host galaxies and non-host galaxies drawn from a cross-identification catalogue, for which we use RGZ. Other related algorithms developed to automatically cross-identify radio objects include Bayesian methods (Fan et al., 2015), likelihood ratio (Weston et al., 2018), positional matching (e.g. Kimball and Ivezić, 2008; Norris et al., 2006; Middelberg et al., 2008), and positional/image hybrid approaches (van Velzen et al., 2012), but these methods do not make use of existing cross-identification catalogues and most assume compact radio sources or that the projected radio emission overlaps the host galaxy.

We represent candidate host galaxies by a $2' \times 2'$ radio image from FIRST centred on that galaxy, the 3.4 μ m magnitude, and the six colours (magnitude differences) derived from the four *WISE* wavelengths. Unknown values of infrared flux were set to their upper limits in AllWISE. We note that many $W_3 - W_4$ colours are missing, so this feature may be less useful than the others. We normalise the colours and magnitude by subtracting the median and dividing by the standard deviation. We normalise each pixel in each radio image by subtracting the median, dividing by the standard deviation, and applying a logistic function (σ ; Equation 5.2) to account for the high dynamic range of radio images.

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \quad (5.2)$$

²<http://cdsxmatch.u-strasbg.fr/>

The medians and standard deviations are reported in Table 5.1. These values are computed across the training set.

For each FIRST component we generate a set of candidate host galaxies. An AllWISE object is considered a candidate host for a radio component if it is within $\sqrt{2} \times 1.5'$ of the centre of the two-dimensional Gaussian fit for that component. This search radius is the maximum angular distance that a host galaxy can be located in RGZ due to the 3'-wide square images shown to volunteers. Candidate hosts are assigned binary labels: All candidates identified as host galaxies in RGZ are assigned a positive label and all others are assigned a negative label. Following Alger et al. (2018) we train a convolutional neural network (CNN) on the labelled candidate hosts. We base our model on ResNet18 (He et al., 2016) pretrained on the ImageNet classification task, with the final layer removed and replaced by a logistic regression model. Non-image features (i.e. colours and the 3.6 μm magnitude) are concatenated with the features that are output by the final ResNet18 layer. Using Adam (Kingma and Ba, 2014) to optimise our weights, we train this model on our task until binary cross-entropy loss starts to increase on a randomly-selected 20 per cent validation set. We use PyTorch (Paszke et al., 2017) to implement this model. The scores of each FIRST component are weighted by a one-dimensional Gaussian function of angular separation, and the candidate maximising this weighted score is selected as the host galaxy. We set the standard deviation of the Gaussian to $120''$ as this provides good empirical results.

92 per cent of the host galaxies in RGZ are also detected as host galaxies in RGZ-Ex. The mean volunteer agreement on all RGZ sources with detected hosts was 95^{+5}_{-13} per cent, compared to 88^{+12}_{-17} per cent for sources with hosts not in RGZ-Ex. Incorrect cross-identifications can be considered a source of noise in the statistics. In future work we will design a way for BXID to output an ‘uncertainty’ so sources with uncertain cross-identifications can be removed from calculations. We quantify the reliability of RGZ-Ex in Section 6.1.

We note that BXID necessarily identifies a host galaxy for all radio emission, even when it does not make sense to do so. In our current work we treat this as a source of noise. Future extensions to BXID will allow it to output ‘no detected host’.

6.1 Visual verification

To quantify the reliability of RGZ-Ex, some of the authors (M.A., O.W., A.K., N.M., and A.T.) visually verified a randomly selected set of 200 radio components/host galaxy pairs in RGZ-Ex. For each pair we decided whether the radio component matched its identified host or not. This allowed us to estimate the accuracy of radio component-infrared host pairs identified by BXID. Verification of component-host pairs is noisy and sometimes even subjective, so each author looked at all 200 components. If an author was unsure about a pair, they were allowed to ignore it and the verification was treated as missing label data. We then aggregated these verifications following the aggregation approach introduced by Dawid and Skene (1979). This approach jointly estimates the accuracy of each author along with the aggregated ver-

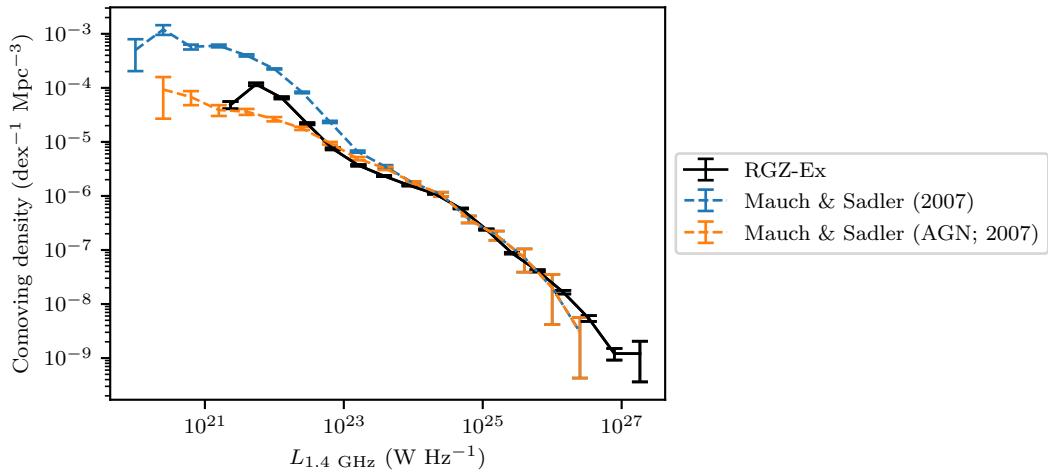


Figure 5.1: RGZ-Ex radio luminosity function compared with the RLFs of Mauch and Sadler (2007a).

fication by maximising the joint likelihood. Note that this approach is able to handle missing label data. We propagated the uncertainty in the accuracy with a Monte Carlo approach, sampling aggregated verifications from the probabilities resulting from the Dawid and Skene model. The estimated accuracy of BXID is (89.5 ± 0.8) per cent. It is interesting to note from our results that even astronomers disagree significantly on the radio cross-identification task, with agreement between authors ranging from 50 to 100 per cent across the verification components. We report the full verification set in Section J.

7 Radio luminosity functions

In this section we present our radio luminosity functions (RLFs) derived from the RGZ-Ex catalogue. To calculate each RLF we followed the $1/V_{\max}$ method (Schmidt, 1968). This method accounts for the effects of Malmquist bias, which is a systematic bias against sources at greater distances. We describe this approach in Section G. We limit our sample to radio sources with 1.4 GHz integrated flux density of at least 2 mJy associated with host galaxies brighter than magnitude 17 at 3.4 μ m, a spectroscopic redshift $0.02 \leq z \leq 0.6$, and an *i*-band magnitude < 20 . We chose these limits based on the distribution of redshifts and infrared magnitudes as well as the sensitivity of FIRST. We then remove sources with unusually high or low W1 magnitude for their redshift (more than 3 standard deviations from the mean) because many such sources have incorrect spectroscopic redshifts, e.g. blazars. There are 24 743 sources matching all criteria. We assume a spectral index of $\alpha = -0.7$ (as is common in literature, e.g. Condon, Cotton, and Broderick, 2002) with flux density $f \propto \nu^\alpha$ where ν is the frequency. We calculate the k -corrected radio luminosity

(Kochanek et al., 2001) as follows:

$$L = \frac{4\pi f d^2}{1+z} (1+z)^{-\alpha} \quad (5.3)$$

where z is redshift and d is luminosity distance (a function of z). Uncertainties in comoving density are estimated as described in Section G. Completeness estimates are shown in Section H. We discuss biases in our methods and results in Section 8.1.

We compare our RLFs with Mauch and Sadler (2007a), who estimated RLFs from 7 824 manually cross-identified radio sources in the NRAO VLA Sky Survey (NVSS; Condon et al., 1998b). Their RLFs were split into AGN and star-forming radio sources. While we do not make this split explicitly in our catalogue, we expect both RGZ-Ex and RGZ to be dominated by AGN due to the selection criterion of being extended in the selected redshift volume. We note that the redshift range used in our work, $0.02 < z < 0.6$, differs from the $0.003 < z < 0.3$ range used by Mauch and Sadler (2007a).

In Figure 5.1 we show the RLF derived from RGZ-Ex along with the RLFs from Mauch and Sadler (2007a). There is good agreement between all three luminosity functions for luminosities greater than 10^{23} W Hz $^{-1}$ and below this luminosity the RGZ-Ex RLF is bounded above by the Mauch and Sadler (2007a) RLF. RGZ-Ex generally finds less comoving density than Mauch and Sadler (2007a), which we attribute to our requirement for extent. We suggest that the peak in RGZ-Ex RLF at approximately 10^{22} W Hz $^{-1}$ is due to our sample containing a small fraction of star-forming galaxies. Our criterion, however, does cut out most star-forming regions as these are often compact, which is why we report lower densities than the star-forming RLF of Mauch and Sadler (2007a).

The *WISE* colour-colour plot, shown for RGZ-Ex in Figure 5.2, is often used to categorise galaxies at different evolutionary stages into four mid-infrared colour regions that are typically populated by 1) spheroidals or elliptical galaxies; 2) quasi-stellar objects (QSOs), Seyferts or powerful AGN; 3) starbursting or luminous infrared galaxies (LIRGs); and 4) the intermediate region where the other three regions overlap. The horizontal axis, $W2 - W3$, separates early- and late-type galaxies, with the star-forming late-type galaxies appearing redder (further to the right) (Wright et al., 2010). The vertical axis, $W1 - W2$, separates inactive galaxies from AGN with strongly radiating accretion discs (Sadler et al., 2014). In Figure 5.3 we show the radio luminosity function split by host galaxy location in the mid-infrared colour-colour plot as defined by Jarrett et al. (2017).

Many sources have $W3$ detections with low signal-to-noise, limiting our ability to subdivide our sample. We plot both the RLFs for the sample with only $W3 \geq 3\sigma$ as well as the RLFs for the full sample in Figure 5.3. For the full sample we use the lower magnitude limit from AllWISE as the $W3$ magnitude (which is an upper flux limit). Using the upper flux limit as the real $W3$ flux has the effect of increasing $W2 - W3$ compared to a real detection, so objects appear further to the right of the colour-colour diagram (Figure 5.2) than they ought to. This means that due to $W3$ limits,

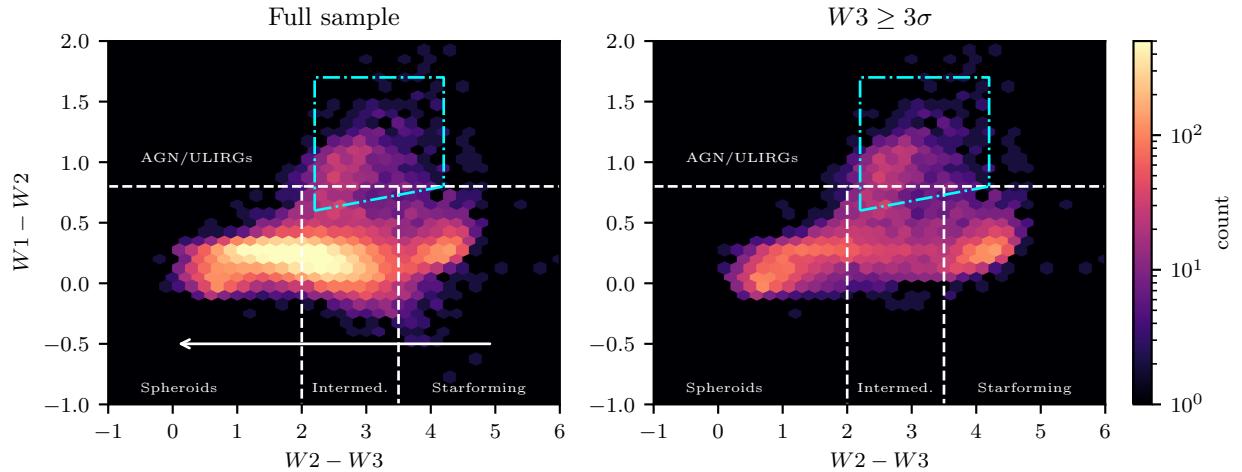


Figure 5.2: *WISE* colour-colour distributions. The dashed grey lines show simple host galaxy class divisions from Jarrett et al. (2017). These classes are labelled in the plot. The blue dot-dashed line shows the empirical optical/infrared AGN criteria from Jarrett et al. (2011). The arrow shows the direction that galaxies would shift with fainter $W3$ magnitudes. The right plot limits the sample to only sources with $W3 \geq 3\sigma$.

objects that should be in the spheroid set will appear in the intermediate and star-forming sets, and objects from the intermediate set will appear in the star-forming set.

At low luminosities, our extended source RLF is dominated by galaxies with infrared colours consistent with star formation. The fraction of the RLF composed of the star-forming set drops off rapidly for $L_{1.4 \text{ GHz}} > 10^{22} \text{ W Hz}^{-1}$, as expected for galaxies with radio emission dominated by star formation (e.g. Mauch and Sadler, 2007a). However, the RLF slope flattens out again beyond $10^{24} \text{ W Hz}^{-1}$, suggesting a second source population. This population has many missing $W3$ measurements, and these are likely intermediates or spheroids incorrectly included in the star-forming set. We therefore suggest that the low-luminosity RGZ-Ex sample mostly contains nearby galaxies with radio emission due to star formation, which appear extended in FIRST as they are close enough for FIRST to resolve their disc (greater than 20 kpc at $z = 0.2$). The remaining fraction of star-forming sources found by Mauch and Sadler (2007a), shown in Figure 5.1, would not be resolved in FIRST, as they are small or distant.

Spheroids, which are hosts in the mid-infrared region corresponding to ellipticals and stars (Wright et al., 2010), comprise the majority of radio galaxies at $10^{23} \text{ W Hz}^{-1}$, and have a peak density at $10^{22} \text{ W Hz}^{-1}$. The common host galaxies for radio-loud AGN tend to be passively-evolving spheroids. It is not surprising that they are more common than star-forming galaxies at luminosities greater than $10^{22} \text{ W Hz}^{-1}$. Above $10^{25} \text{ W Hz}^{-1}$ they are less common than intermediate galaxies and their contribu-

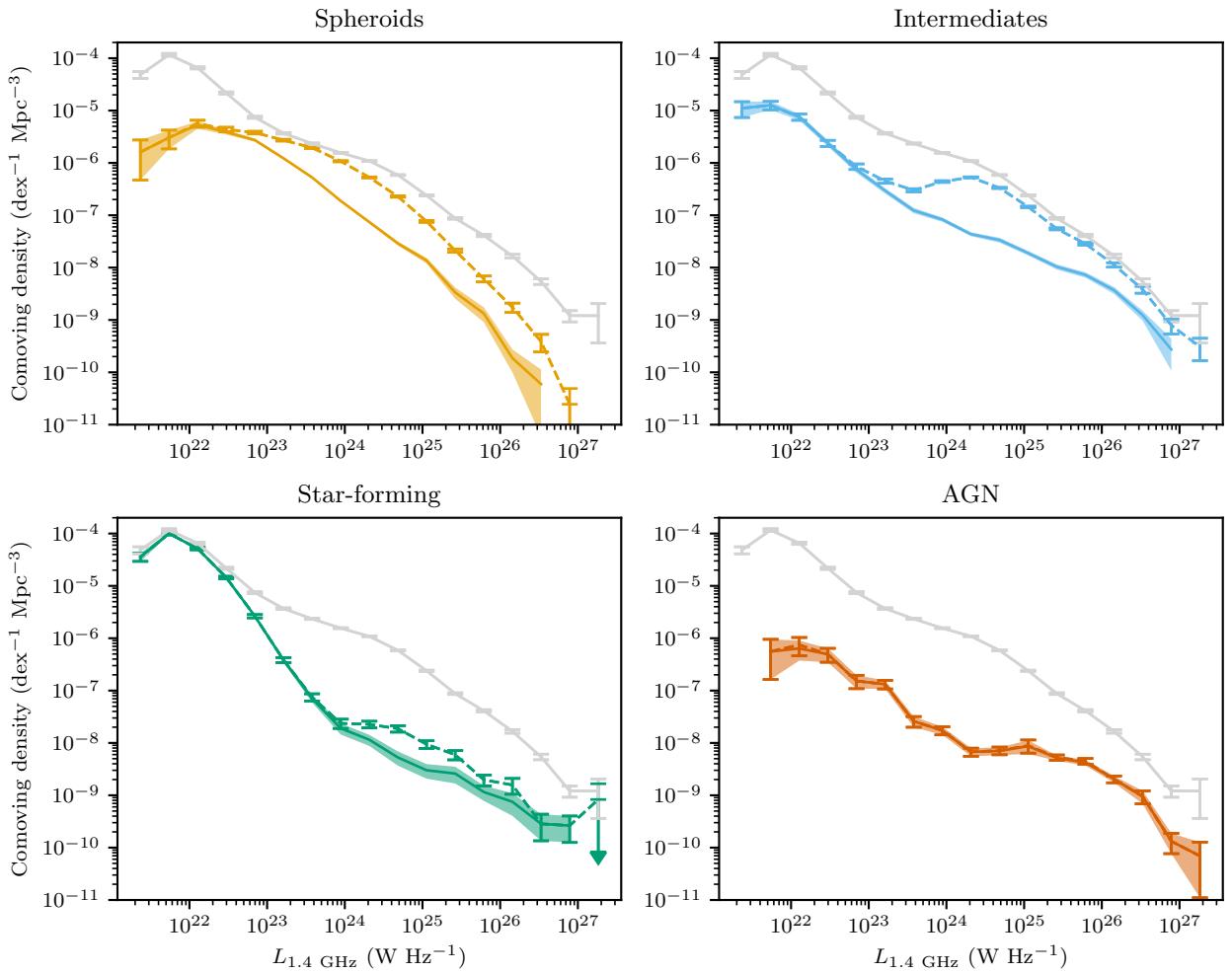


Figure 5.3: RLFs split by host galaxy location in the *WISE* colour-colour plot (Figure 5.2), using our automated cross-identifications. The grey line is the total RLF for all sources. Solid lines have good *W3* detections and dashed lines include *W3* with low signal-to-noise.

tion to the luminosity function drops rapidly. This is likely due to the loss of W3 detections moving spheroids into the intermediate set, and we hypothesise that with deeper W3 observations spheroids may dominate above $10^{25} \text{ W Hz}^{-1}$.

Sources with hosts in the mid-infrared AGN region of the colour-colour diagram (Figure 5.2) make up the smallest contribution to the radio luminosity function. They have a steadily decreasing density from their lowest observed $L_{1.4 \text{ GHz}}$ of $10^{22} \text{ W Hz}^{-1}$ to their highest of $10^{27} \text{ W Hz}^{-1}$, but are present in all luminosity bins except for the very lowest. This is a set with a very low fraction of spectroscopic SDSS matches for the *WISE* host galaxies. 26 per cent of hosts outside the *WISE* AGN region have an SDSS match, compared to just 12 per cent of hosts inside the *WISE* AGN region. This is likely due to the incomplete sampling of QSOs in the SDSS spectroscopic survey or redshift evolution effects (Strauss et al., 2002). The fraction of the RLF contributed by galaxies classed as mid-infrared AGN increases above $10^{25} \text{ W Hz}^{-1}$, meaning that high-luminosity radio AGN are also more likely to be infrared AGN than at lower radio luminosities. Note that the AGN set is unaffected by missing W3 detections, as the AGN set is based only on $W1 - W2$.

Galaxies residing in the intermediate mid-infrared colour region can be populated by both early- and late-type galaxies, which have a mix of processes occurring within them. These ‘intermediate sources’ dominate in most luminosity ranges, and above $10^{24} \text{ W Hz}^{-1}$ they comprise the vast majority of our sample. As intermediate-type galaxies fall between star-forming galaxies and passive ellipticals on the mid-infrared colour-colour plane, they do not have a clear morphological class and are composed of overlapping subsets of sources. The most luminous radio AGN are almost entirely found in this set of galaxies. In fact, as radio luminosity increases the density fraction shifts from spheroids toward intermediate galaxies, likely due to missing W3 moving objects from the spheroid set into the intermediate set.

In Figure 5.4 we show the radio luminosity function for different ranges of projected physical extent of their radio emission. We estimate the angular extent as the angular distance between the most separated components in a multi-component source. This result is complementary to other Radio Galaxy Zoo studies on the effect of the environment on the size and asymmetry of the observed extended radio emission (Rodman et al., 2019b; Garon et al., 2019b).

8 Discussion

8.1 Biases and uncertainties

Biases enter our work due to our chosen samples and methods. Our training set, RGZ, is biased toward sources smaller than $1.5'$ and limited above by $\sqrt{2} \times 3'$ due to the $3' \times 3'$ cutout size of the RGZ user interface. RGZ volunteers preferentially select host galaxies that are brighter in $W1$, so we expect RGZ to overrepresent the number of sources with $W1$ -bright host galaxies.

These biases may affect our trained algorithm: for example, the overabundance of $W1$ -bright host galaxies in RGZ may cause our algorithm to be less accurate when

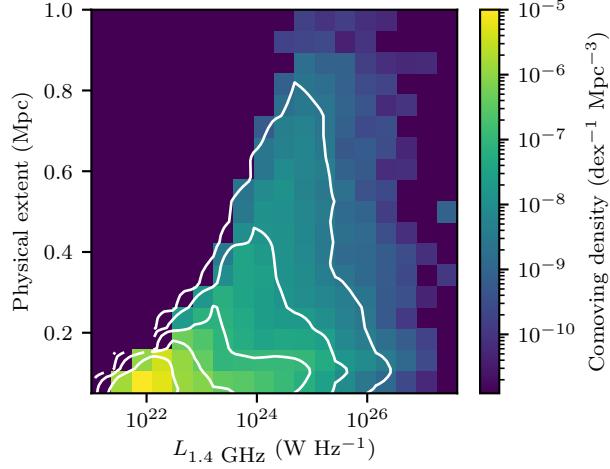


Figure 5.4: Bivariate radio luminosity function showing radio luminosity against projected physical extent. Contours are on a log scale, starting at the median and increasing by 10 percent per contour.

unassociated bright galaxies are in the field of view. Without knowing the true distribution of host galaxies, however, it is difficult to quantify the effect of such biases on our trained method.

FIRST itself is also biased. Helfand, White, and Becker (2015) describe several reasons why FIRST flux may be systematically underestimated. Most of these effects are insignificant for extended objects in our sample or are corrected in the FIRST catalogue from which we draw our flux information. The exception is the ‘resolving out’ of diffuse and low surface brightness radio emission by the Very Large Array in its B configuration. This means that we lose flux on most nearby radio galaxies (especially those with very diffuse components) and may miss diffuse or dim radio galaxies entirely. More diffuse radio galaxies such as Fanaroff-Riley type I (FRI; Fanaroff and Riley, 1974b) galaxies tend to be toward the low end of the radio-loud luminosity distribution, about 10^{23} W Hz^{-1} (Best, 2009), so we expect that losing diffuse sources would lower our estimates of density around this luminosity. Large, extended lobes such as those associated with Fanaroff-Riley type II (FRII; Fanaroff and Riley, 1974b) galaxies may also be resolved out, so by the same mechanism we expect to lose an increasing amount of flux with increasing source angular size. This effect is compounded by flux loss at 1.4 GHz associated with synchrotron losses and adiabatic expansion losses (Blundell, Rawlings, and Willott, 1999).

Our host galaxy redshifts may be biased. Incorrectly identifying the host galaxy may introduce sources with incorrect redshifts into the RLF, an effect which will be dominated by misidentifying galaxies as hosts where the true host is not detected. Since we are matching to optical spectra in SDSS to find redshifts, we are biased toward brighter host galaxies that are more likely to have such spectra. Without an optically-complete sample — currently impossible on such scales — this effect is

unavoidable. Brighter optical sources appear at lower redshifts, so we likely under-sample higher-redshift (and hence higher-luminosity) galaxies.

Our requirement for radio emission to be extended will miss radio galaxies that would be resolved and extended if they were not aligned with the line of sight. We therefore must be underestimating the population of extended sources (though assuming a random distribution of orientations, the majority of galaxies will not be aligned close to the line of sight). The requirement for extended radio emission will also impose a lower limit on linear size, which will vary with redshift: at $z = 0.6$ the effect will be strongest and we will see no sources with linear size under 33.5 kpc. This will cause us to underestimate the population of radio galaxies with linear sizes between 10–30 kpc. On the other hand, we have likely avoided significant overestimation of radio luminosity due to relativistic beaming, since we filter out sources aligned along the line of sight.

We have estimated uncertainties in our RLF from Poisson noise in the histogram bins. We have likely underestimated these uncertainties as it is difficult to estimate uncertainty in our algorithm, though in future we anticipate that we can employ an ensemble of classifiers to estimate this (e.g. Lakshminarayanan, Pritzel, and Blundell, 2017).

8.2 Extended radio galaxies in the low- z Universe

Our total RLFs are consistent with the idea that large, extended radio sources are typically hosted by massive ellipticals (Best et al., 2005). These RLFs match existing RLFs such as that of Mauch and Sadler (2007a), except at radio luminosities below 10^{22} W Hz $^{-1}$. This is unsurprising since we employ a requirement for extended emission, and, besides very nearby star-forming galaxies, FRII comprise most of the population of extended radio objects. The fractional RLF split by mid-infrared colour, Figure 5.3, shows that spheroids reach peak density at a radio luminosity associated with a drop in density of intermediates, and intermediates begin to dominate the RLF as the spheroid density drops. Together, these mid-infrared classes of galaxy form the bulk of the extended radio galaxy RLF.

We see a significant star-forming population in our extended sample, which means that we are likely resolving some discs in radio. While the $1/V_{\max}$ method ensures that our RLFs account for similar galaxies throughout the Universe, even though we only resolve very nearby discs, some of the star-forming population is not included. The difference between our RLF and existing RLFs must be due to the latter containing low-luminosity sources that are compact even when very nearby.

Can we use our RLFs to estimate the kinetic energy contribution of AGN to the galaxy halo and beyond? The extended population of AGN will be the population that contributes most mechanical energy: the major part of the energy in the jet expands the radio lobes, drives shocks or is stored in the jet magnetic field, rather than being emitted as radiation (Godfrey and Shabala, 2016; Hardcastle and Krause, 2014). Extended radio sources should therefore represent the bulk of AGN feedback: radio galaxies with extended jets will inject mechanical energy out to larger

distances from the core of the host galaxies than those with smaller jets. This is supported by e.g. Turner and Shabala (2015), who found that extended sources comprised the bulk of the mechanical energy contribution. By assuming a relationship between radio luminosity and radio jet mechanical energy, we can use our extended source RLFs to estimate the contribution of extended AGN to energy in the intergalactic/circumgalactic medium (IGM/CGM). But assuming such a relationship is not without problems: the radio lobe luminosity experiences significant evolution (e.g. Bicknell, Dopita, and O'Dea, 1997), the surrounding IGM/CGM may interact with the radio lobe expansion in non-trivial ways (e.g. Hardcastle and Krause, 2013) and the relationship between the mechanical energy and radio luminosity has high scatter on individual radio sources (Hardcastle and Krause, 2013). With our sample size, these effects should be diminished, and with these caveats in mind we will estimate the energy contribution of extended sources to the IGM. We assume a scaling relation of $\ln Q = \beta \ln L_\nu + Q_0$, where Q is the jet power and L_ν is the monochromatic radio luminosity at frequency ν . The values for β and Q_0 vary significantly across the literature, based on different physical assumptions and samples. Willott et al. (1999) presented a widely-used relationship

$$\ln Q = \ln(f^{3/2} 3 \times 10^{38}) + \frac{6}{7} \ln \left[\frac{L_{151 \text{ MHz}}}{10^{28} \text{ W Hz}^{-1}} \right], \quad (5.4)$$

with a scaling constant $1 \leq f \leq 20$ and Q in watts. Other models exist with different slopes, e.g. Bîrzan et al. (2008) suggest that $\beta \approx 0.5 - 0.7$ and Cavagnolo et al. (2010) find $\beta \approx 0.7$. Godfrey and Shabala (2016) provide a summary of the literature in this field and suggest that these correlations are from mutual distance dependence rather than intrinsic relationships. They find that there is no strong empirical evidence for such a correlation in either FRI or FRII. However, their theoretical models suggest $\beta \approx 0.5, 0.8$ for FRI and FRII respectively, which is consistent with Willott et al. (1999). Shabala and Godfrey (2013) on the other hand show that the scalings presented by Willott et al. (1999) are consistent with independent theoretical modelling for high-power radio galaxies. Scaling the frequency to 1.4 GHz, and assuming a spectral index of $\alpha = -0.7$, $\beta = 6/7$, and $Q_0 = \ln(f^{3/2}) + 89.9$, we can write the Willott et al. (1999) relation as

$$\ln(Q) = \ln(f^{3/2}) + 89.9 + \frac{6}{7} \ln \left[\frac{L_{1400 \text{ MHz}}}{10^{28} \text{ W Hz}^{-1}} \right]. \quad (5.5)$$

Assuming $f \in [1, 20]$ gives $Q_0 \in [89.9, 94.4]$. Integrating over our RLF we find $Q \in [1.3 \times 10^{30}, 1.2 \times 10^{32}] \text{ W Mpc}^{-3}$. This is likely a lower limit as we are missing extended radio sources oriented along our line-of-sight and nearby diffuse extended radio sources (e.g. FRI), and Shabala (2018) argues that many ‘compact’ AGN may in fact be extended but below the sensitivity of surveys such as FIRST. Our results are consistent with other literature (e.g Hardcastle et al., 2019, who estimated the energy contribution as $7 \times 10^{31} \text{ W Mpc}^{-3}$).

8.3 Future work

With such a large sample size, further partitioning of the RLF into subsamples is possible. Any combination of the features investigated here, plus further host galaxy and radio properties, could be used to generate fractional RLFs. Automated classifiers such as CLaRAN (Wu et al., 2019) or feature extractors such as PINK (Polsterer, Gieseke, and Igel, 2015; Galvin et al., 2019b; Ralph et al., 2019) could provide a way to divide the RLF by radio morphology. These methods provide a way of dividing galaxy classes based directly on the radio image, rather than the host galaxy like we have done here, and so should not be affected by extinction or redshift in the same way as our sample. Such subsamples would lend insight into how radio power is connected to radio morphology and generation mechanisms. Cross-matching with other surveys such as NVSS or the 150 MHz TIFR GMRT Sky Survey would provide properties such as the spectral index and observations of diffuse emission missed by FIRST (as used by Kimball and Ivezić, 2008). Such properties could also be used to create interesting and insightful fractional RLFs.

While we have not investigated the link between extended sources and their local environments, this will be the focus of future work. Environment will heavily factor into the source sizes, morphologies, and so on, following work such as Rodman et al. (2019b) and Garon et al. (2019b).

Ongoing radio surveys such as EMU, VLASS (Lacy et al., 2020), and LoTSS (Shimwell et al., 2019) will greatly increase the number of extended sources. However, our sample size limitations in this paper are not from FIRST, but from SDSS: until next-generation spectroscopic surveys are available, redshifts will be the limiting factor. To significantly increase our sample size would require much greater numbers of redshifts.

9 Summary

Extended radio sources provide an opportunity to study the interaction between AGN and their large-scale environments. We trained the binary cross-identification method on the Radio Galaxy Zoo to generate the largest sample of reliably cross-identified, extended radio sources, and this large sample allows us to investigate their bulk distributions in new, detailed ways. We estimated radio luminosity functions split by mid-infrared colour, physical extent and redshift. Despite our extendedness criterion, we found a significant star-forming population. We estimated that extended AGN contribute between 1.3×10^{30} and 1.2×10^{32} W Mpc $^{-3}$ of mechanical energy to their environment. Ongoing and future surveys such as EMU will provide even greater numbers of extended radio sources, and our combination of machine learning and astronomy methodology will allow these samples to be cross-identified and investigated efficiently and reliably.

This publication has been made possible by the participation of more than 11 000 volunteers in the Radio Galaxy Zoo project. Their contributions are individually acknowledged at <http://rgzauthors.galaxyzoo.org>. Radio Galaxy Zoo makes use of

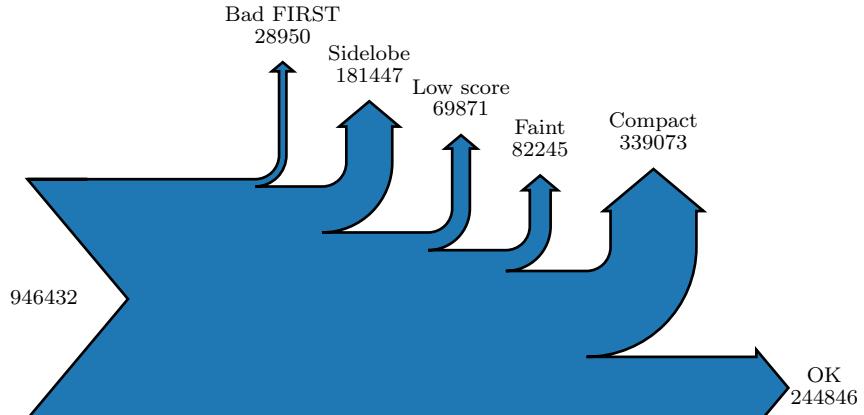


Figure 5.5: Number of components removed from FIRST by each filter.

data products from the *Wide-field Infrared Survey Explorer* and the Very Large Array. The *Wide-field Infrared Survey Explorer* 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. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The figures in this work made use of Astropy, a community-developed core Python package for Astronomy (The Astropy Collaboration et al., 2018). Partial support for this work for L.R. and A.F.G. comes from National Science Foundation grant AST-1714205 to the University of Minnesota. H.A. benefited from grant CIIC 218/2019 of Universidad de Guanajuato. M.A. and A.T. were supported by the Australian Government Research Training Program. We thank J. Nabaglo and A. Gruen for their help with efficient data pre-processing. We thank H. Zovaro, T. J. Galvin and H. Tang for their comments on the draft of this paper.

F Sankey diagrams

This section presents Sankey diagrams showing the filtering of components and sources from the full FIRST sample. A Sankey diagram shows the order and number of objects removed from a sample. Figure 5.5 shows the filtering of components and Figure 5.6 shows the filtering of sources. The component filters are ‘Bad FIRST’ for components on the edge of FIRST with incomplete images, ‘Sidelobe’ for components with high sidelobe probability, ‘Low score’ for components with only low-scoring candidate hosts, ‘Faint’ for components with less than 10 signal-to-noise according to the FIRST catalogue, and ‘Compact’ for components that do not have extended radio emission according to Equation 5.1. Sources are removed after each component filter if they no longer contain any components.

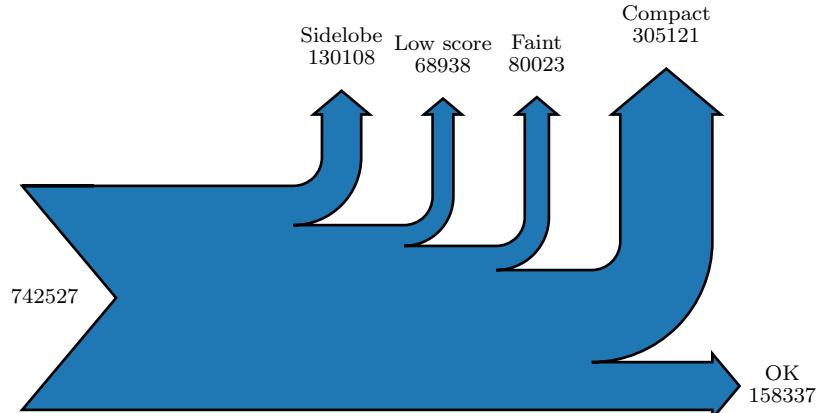


Figure 5.6: Number of sources removed by each filter.

G Radio luminosity function

We computed the radio luminosity function following the $1/V_{\max}$ method (Schmidt, 1968). This appendix explains our implementation. We performed the following steps:

1. Remove all radio sources that do not fit the selection criteria. This applies for both radio and infrared properties, so we choose a minimum radio flux density f_{\min} and a maximum infrared magnitude m_{\max} , as well as redshift limits z_{lower} and z_{upper} .
2. For each source, compute the maximum redshift that the source could have been observed within the selection criteria. We find this redshift by first numerically solving Equation 5.3 for z with L as the luminosity of each radio source and $f = f_{\min}$ to obtain the maximum redshift z_{radio} at which the source could be observed in radio. We find the maximum redshift z_{ir} that the host galaxy could be observed within the selection criteria by numerically solving Equation 5.6 for z , where $d(z)$ is the luminosity distance at a redshift z , d is the luminosity distance of the host galaxy, and m is the apparent magnitude of the host galaxy.

$$5 \log_{10} \left(\frac{d(z)}{d} \right) + m = m_{\max} \quad (5.6)$$

The maximum redshift that the source could have been observed within the selection criteria is then $z_{\max} = \min(z_{\text{ir}}, z_{\text{radio}}, z_{\text{upper}})$.

3. For each source, compute the comoving volume V_{\max} at redshift z_{\max} .
4. The count for each luminosity bin is the sum over $1/V_{\max}$ for each source in the bin. We account for the fact FIRST does not cover the whole sky by multiplying

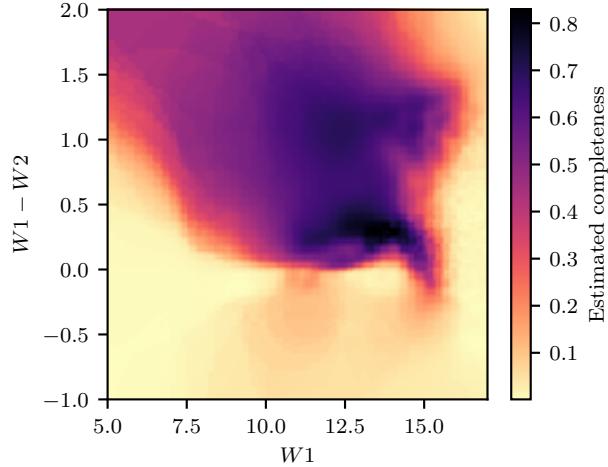


Figure 5.7: Estimated completeness as a function of mid-infrared colour and magnitude.

by the total area of the sky divided by the area of our selection.

After computing the luminosity function, we estimate the uncertainty in each bin using Poisson statistics, \sqrt{N} for a bin count N .

H Redshift completeness estimate

Figure 5.7 shows the estimated completeness as a function of W_1 and $W_1 - W_2$. We followed the same method as Pracy et al. (2016) for this estimation, averaging completeness over circles centred on each source. Each source is associated with a circle of radius equal to the distance to its 50th nearest neighbour in the W_1 and $W_1 - W_2$ plane.

I Giant radio galaxies

To identify radio sources we assume that if any two components have the same host galaxy then they are part of the same source. This is a reasonable assumption if all host galaxies are correctly identified, which is not the case. This assumption therefore introduces spurious sources due to galaxies incorrectly identified as host galaxies: not all sources used in this paper are real sources, and in particular sources of large angular size are likely to be incorrect. Nevertheless RGZ-Ex provides a useful catalogue of *candidate* radio sources, and visual follow-up can confirm whether sources of interest are real.

H.A. and M.A. examined all 296 candidate sources in the RGZ-Ex catalogue with an estimated physical extent larger than 1 Mpc. Of these, 40 were real giant radio galaxies, which we show in Table 5.2. We define ‘giant radio galaxy’ as a radio

Table 5.2: Giant radio galaxies found in RGZ-Ex. ‘LLS’ is the projected linear size of the source as measured by the maximum angular distance between radio components. The RA/Dec are the coordinates of the host galaxy. s/p indicates spectroscopic/photometric redshift. ^LExisting in literature. ^RAlso found by RGZ citizen scientists. [†]Misidentified SDSS host, manually corrected to obtain redshift.

AllWISE host (WISEA)	RA (J2000)	Dec (J2000)	z	LLS (Mpc)	
J004210.18-080011.3	10.54	-8.00	0.65 ± 0.14	1.6	p
J021008.48+011839.6 ^L	32.54	1.31	0.86524 ± 0.0001	1.2	s
J075858.29+355643.6 ^R	119.74	35.95	0.74748 ± 0.00013	1.0	s
J080831.68+473523.9 ^R	122.13	47.59	0.58854 ± 0.00016	1.1	s
J083034.78+231124.6	127.64	23.19	0.94 ± 0.13	1.1	p
J090604.03+011114.2	136.52	1.19	0.7975 ± 0.0004	1.6	s
J093256.81+074212.2	143.24	7.70	1.0032 ± 0.0003	1.1	s
J093526.80+051729.8 ^R	143.86	5.29	0.84 ± 0.04	1.2	p
J094238.72+114337.9	145.66	11.73	0.49 ± 0.05	1.2	p
J094835.60+535946.4 ^R	147.15	54.00	0.64 ± 0.10	1.2	p
J095706.12+292439.2	149.28	29.41	0.71 ± 0.12	1.5	p
J102335.25+433208.0	155.90	43.54	0.75 ± 0.09	1.5	p
J102933.99+210345.8 ^R	157.39	21.06	0.82407 ± 0.00008	1.1	s
J103043.98+355451.2 ^R	157.68	35.91	0.64074 ± 0.00008	1.2	s
J104449.92+234525.6 [†]	161.20	23.76	0.57712 ± 0.00009	1.6	s
J110655.98+624759.8 ^R	166.73	62.80	0.84379 ± 0.00004	1.1	s
J112900.68+635543.2	172.25	63.93	0.71 ± 0.06	1.1	p
J112948.20+243922.6	172.45	24.66	0.79 ± 0.07	1.1	p
J114553.67-003304.7	176.47	-0.55	2.0522 ± 0.0006	1.3	s
J121111.26+534840.4	182.80	53.81	0.74 ± 0.14	1.1	p
J121152.04+304232.4 ^R	182.97	30.71	0.47102 ± 0.00012	1.3	s
J121944.73+174121.3	184.94	17.69	1.5129 ± 0.0009	1.0	s
J123735.89+544814.4 ^R	189.40	54.80	1.0271 ± 0.0006	1.2	s
J123819.16+113444.8	189.58	11.58	0.80 ± 0.08	1.2	p
J123846.84-032857.5 [†]	189.70	-3.48	0.67 ± 0.07	1.5	p
J131625.00+272042.8	199.10	27.35	0.69092 ± 0.00004	1.0	s
J133307.00+045048.6 ^R	203.28	4.85	1.40534 ± 0.00016	1.1	s
J141933.36+104706.4 ^R	214.89	10.79	0.33973 ± 0.00003	1.0	s
J142008.45+185422.7 ^R	215.04	18.91	0.63 ± 0.04	1.4	p
J145057.28+530007.7 ^L	222.74	53.00	0.91662 ± 0.00009	1.3	s
J150012.18+604941.3	225.05	60.83	1.6626 ± 0.0007	1.2	s
J153547.13+432245.0 ^R	233.95	43.38	0.63891 ± 0.00007	1.3	s
J154631.18+194819.9	236.63	19.81	0.5917 ± 0.0002	1.4	s
J160852.10+561110.2 ^R	242.22	56.19	1.3196 ± 0.0003	1.3	s
J162200.48+364044.0	245.50	36.68	1.9994 ± 0.0002	1.1	s
J163004.35+103321.9 ^R	247.52	10.56	0.85 ± 0.09	1.2	p
J163125.75+200224.1 ^R	247.86	20.04	0.62662 ± 0.00013	1.0	s
J165055.46+394446.6	252.73	39.75	0.58829 ± 0.00013	1.1	s
J232410.33+045309.6	351.04	4.89	0.76 ± 0.06	1.4	p
J234440.02-003231.6	356.17	-0.54	0.5014 ± 0.0001	1.0	s

galaxy with emission extended to physical sizes ≥ 1.0 Mpc. Other thresholds, such as 0.7 Mpc, also exist in literature. The physical extents of the remaining 256 candidate sources were overestimated mostly due to sidelobes/artefacts (103), incorrect source grouping (82), or incorrect SDSS matches (21). The citizen scientists who identified giants are: WizardHowl, DolorousEdd, antikodon, csunjoto, sisifolibre, JeanTate, JKD, PADV, and firejuggler. H.A., together with his summer students, had previously identified 29 of these giants.

Note that this is a particularly challenging set: sources that are misidentified will often have unusually large estimated extents due to the inclusion of spurious components. The error rate in this set therefore does not reflect the rest of the catalogue.

J Visual verification results

In Section 6.1 we described our visual verification of the BXID method. We list the radio components in the verification set in Table 5.3. Each row of the table contains the FIRST component, its AllWISE host galaxy according to BXID, and whether the association is correct according to our visual verification. If an author was particularly unsure about an object, they were able to skip this object, and so are not accounted for in the verification for that object. Verification was weighted by the Dawid and Skene (1979) maximum likelihood model.

Table 5.3: Validation objects. ‘Agree’ is whether or not the authors agreed with BXID associating the given FIRST object with the given AllWISE object.

FIRST	AllWISE	Agree
J000234.9-001421	J000242.35-001320.5	n
J002841.1+141654	J002840.37+141652.7	y
J003731.4+000156	J003731.26+000146.7	y
J005407.5-011158	J005407.61-011158.9	y
J011210.3+002203	J011210.41+002201.9	y
J012342.4+015849	J012342.24+015850.4	y
J013015.1+110653	J013015.16+110653.4	y
J013107.7+070343	J013102.02+070332.0	y
J014247.9-000039	J014247.81-000040.3	y
J014250.0-000032	J014247.81-000040.3	n
J020222.3+030138	J020223.20+030150.4	y
J020333.8+000853	J020336.94+000759.3	y
J021840.1-032311	J021840.13-032306.0	y
J023022.0+010834	J023022.11+010840.0	y
J024245.3-022535	J024245.35-022534.6	y
J025901.0+005350	J025901.50+005346.1	y
J033204.1-004757	J033204.15-004757.1	y
J073033.2+390413	J073033.21+390412.9	y
J073954.1+481810	J073954.87+481759.5	y
J074504.9+331247	J074504.81+331256.2	y
J074640.4+421709	J074640.45+421709.1	y
J074707.9+171719	J074708.35+171726.5	y
J075043.6+274838	J075043.35+274844.8	n
J075050.3+331937	J075051.25+331905.0	y
J075422.2+311253	J075422.35+311252.5	y
J075637.0+212006	J075636.65+212001.4	y
J082326.1+141438	J082326.34+141435.9	y
J082422.5+351121	J082422.65+351114.6	y
J082925.9+462618	J082926.02+462618.5	y
J083512.4+175441	J083512.45+175441.1	y
J084133.5+402035	J084133.40+402042.8	y
J084238.4+405305	J084238.38+405306.6	n
J084417.3+315845	J084417.92+315845.9	y
J084728.5+360700	J084728.24+360714.6	y
J084905.5+111448	J084905.51+111447.8	y
J085236.8+262006	J085236.11+262013.4	y
J085415.6+524930	J085415.62+524936.7	y
J090623.2+300746	J090622.87+300743.9	y
J091745.1+275049	J091745.89+275103.8	y
J091752.0+431614	J091752.14+431612.7	y
J092014.4+302907	J092013.95+302859.3	y
J092140.5+540118	J092140.24+540121.1	y
J092213.0+542157	J092213.03+542157.2	y
J092406.9+562703	J092406.47+562656.2	y
J092713.1+105841	J092713.14+105839.8	y
J093108.6+613447	J093108.63+613447.2	y
J093239.6+052308	J093237.71+052240.7	n
J093627.8+103610	J093627.87+103609.7	y
J093645.2+561435	J093645.89+561434.2	y
J094006.8+482651	J094006.92+482649.2	y

Faraday Complexity

This chapter is based on my paper *Interpretable Faraday Complexity Classification*, by M. J. Alger, C. S. Ong, J. D. Livingston, J. L. Nabaglo, N. M. McClure-Griffiths, and O. I. Wong; submitted to the Publications of the Astronomical Society of Australia (PASA).

Faraday complexity describes whether a spectropolarimetric observation has simple or complex magnetic structure. Quickly determining the Faraday complexity of a spectropolarimetric observation is important for processing large, polarised radio surveys. Finding simple sources lets us build rotation measure grids, and finding complex sources lets us follow these sources up with slower analysis techniques or further observations. We introduce five features that can be used to train simple, interpretable machine learning classifiers for estimating Faraday complexity. We train logistic regression and extreme gradient boosted tree classifiers on simulated polarised spectra using our features, analyse their behaviour, and demonstrate our features are effective for both simulated and real data. This is the first application of machine learning methods to real spectropolarimetry data. With 95 per cent accuracy on simulated ASKAP data and 90 per cent accuracy on simulated ATCA data, our method performs comparably to state-of-the-art convolutional neural networks while being simpler and easier to interpret. Logistic regression trained with our features behaves sensibly on real data and its outputs are useful for sorting polarised sources by apparent Faraday complexity.

M.J.A. and J.D.L. were supported by the Australian Government Research Training Program. M.J.A. was supported by the Astronomical Society of Australia.

1 Introduction

As polarised radiation from distant galaxies makes its way to us, magnetised plasma along the way can cause the polarisation angle to change due to the Faraday effect. The amount of rotation depends on the squared wavelength of the radiation, and the rotation per squared wavelength is called the Faraday depth. Multiple Faraday depths may exist along one line-of-sight, and if a polarised source is observed at multiple wavelengths then these multiple depths can be disentangled. This can provide insight into the polarised structure of the source or the intervening medium.

Faraday rotation measure synthesis (RM synthesis) is a technique for decomposing a spectropolarimetric observation into flux at its Faraday depths ϕ , the resulting distribution of depths being called a ‘Faraday dispersion function’ (FDF) or a ‘Faraday spectrum’. It was introduced by Brentjens and Bruyn (2005) as a way to rapidly and reliably analyse the polarisation structure of complex and high-Faraday depth polarised observations.

A ‘Faraday simple’ observation is one for which there is only one Faraday depth, and in this simple case the Faraday depth is also known as a ‘rotation measure’ (RM). All Faraday simple observations can be modelled as a polarised source with a thermal plasma of constant electron density and magnetic field (a ‘Faraday screen’; Brentjens and Bruyn, 2005; Anderson et al., 2015) between the observer and the source. A ‘Faraday complex’ observation is one which is not Faraday simple, and may differ from a Faraday simple source due to plasma emission or composition of multiple screens (Brentjens and Bruyn, 2005). The complexity of a source tells us important details about the polarised structure of the source and along the line-of-sight, such as whether the intervening medium emits polarised radiation, or whether there are turbulent magnetic fields or different electron densities in the neighbourhood. The complexity of nearby sources taken together can tell us about the magneto-ionic structure of the galactic and intergalactic medium between the sources and us as observers. O’Sullivan et al. (2017) show examples of simple and complex sources, and Figure 6.1 and Figure 6.2 show an example of a simulated simple and complex FDF respectively.

Identifying when an observation is Faraday complex is an important problem in polarised surveys (**sun15comparison**), and with current surveys such as the Polarised Sky Survey of the Universe’s Magnetism (POSSUM) larger than ever before, methods that can quickly characterise Faraday complexity en masse are increasingly useful. Being able to identify which sources are simple lets us produce a reliable rotation measure grid from background sources, and being able to identify which sources might be complex allows us to find sources to follow-up with slower polarisation analysis methods that may require manual oversight, such as QU fitting (**miyashita19qu**; as seen in e.g. O’Sullivan et al., 2017). In this chapter, we introduce five simple, interpretable features representing polarised spectra, use these features to train machine learning classifiers to identify Faraday complexity, and demonstrate their effectiveness on real and simulated data. We construct our features by comparing observed polarised sources to idealised polarised sources. The features are intuitive and can be estimated from real FDFs.

Section 2 provides a background to our work, including a summary of prior work and our assumptions on FDFs. Section 3 describes our approach to the Faraday complexity problem. Section 4 explains how we trained and evaluated our method. Finally, Section 5 discusses these results.

2 Faraday Complexity

Faraday complexity is an observational property of a source: if multiple Faraday depths are observed within the same apparent source (e.g. due to multiple lines-of-sight being combined within a beam), then the source is complex. A source composed of multiple Faraday screens may produce observations consistent with many models (**sun15comparison**), including simple sources, so there is some overlap between simple and complex sources. Faraday thickness is also a source of Faraday complexity: when the intervening medium between a polarised source and the observer also emits polarised light, the FDF cannot be characterised by a simple Faraday screen. As discussed in Section 2.2 we defer Faraday thick sources to future work. In this section we summarise existing methods of Faraday complexity estimation and explain our assumptions and model of simple and complex polarised FDFs.

2.1 Prior work

There are multiple ways to estimate Faraday complexity, including detecting non-linearity in $\chi(\lambda^2)$ (**goldstein84faraday**), change in fractional polarisation as a function of frequency (**farnes14broadband**), non-sinusoidal variation in fractional polarisation in Stokes Q and U (**osullivan12agn**), counting components in the FDF (**law11faraday**), minimising the Bayesian information criterion (BIC) over a range of simple and complex models (called ‘QU fitting’; O’Sullivan et al., 2017), the method of Faraday moments (**Brown11report**; Anderson et al., 2015), and deep convolutional neural network classifiers (CNNs; Brown et al., 2018). See **sun15comparison** for a comparison of these methods.

The most common approaches to estimating complexity are QU fitting (e.g. O’Sullivan et al., 2017) and Faraday moments (e.g. Anderson et al., 2015). To our knowledge there is currently no literature examining the accuracy of QU fitting when applied to complexity classification specifically, though **miyashita19qu** analyse its effectiveness on identifying the structure of two-component sources. **Brown11report** suggested Faraday moments as a method to identify complexity, a method later used by **farnes14broadband** and Anderson et al. (2015), but again no literature examines the accuracy. CNNs are the current state-of-the-art with an accuracy of 94.9 per cent (Brown et al., 2018) on simulated ASKAP Band 1 and 3 data, and we will compare our results to this method.

2.2 Assumptions on Faraday dispersion functions

Before we can classify FDFs as Faraday complex or Faraday simple, we need to define FDFs and any assumptions we make about them. An FDF is a function that maps Faraday depth ϕ to complex polarisation. It is the distribution of Faraday depths in an observed polarisation spectrum. For a given observation, we assume there is a true, noise-free FDF F composed of at most two Faraday screens. This accounts for most actual sources (Anderson et al., 2015) and extension to three screens would cover most of the remainder—O’Sullivan et al. (2017) found that 89 per cent of their

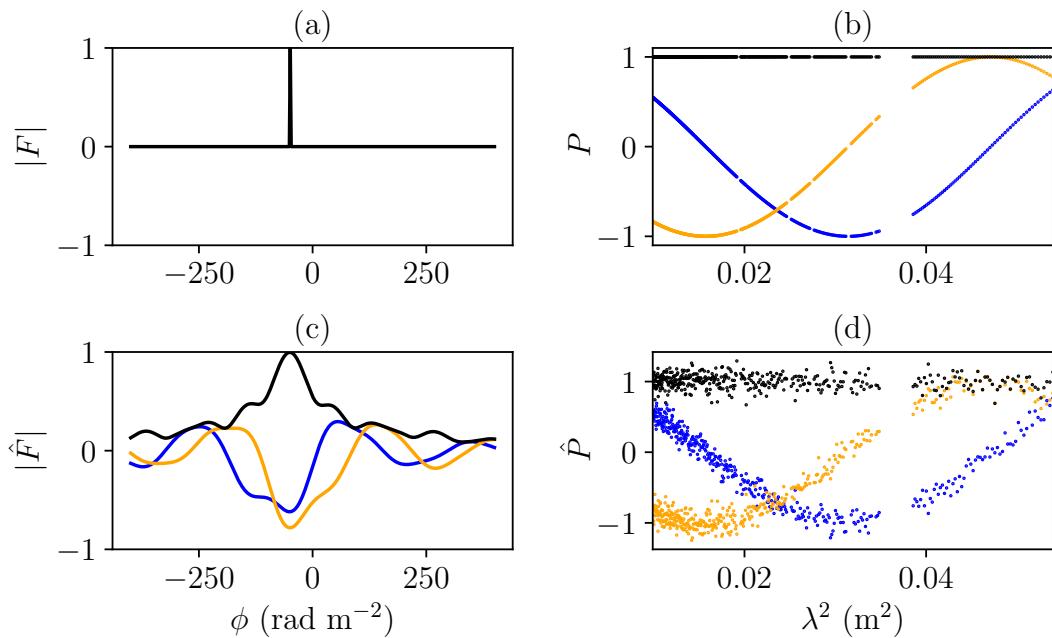


Figure 6.1: A simple FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.

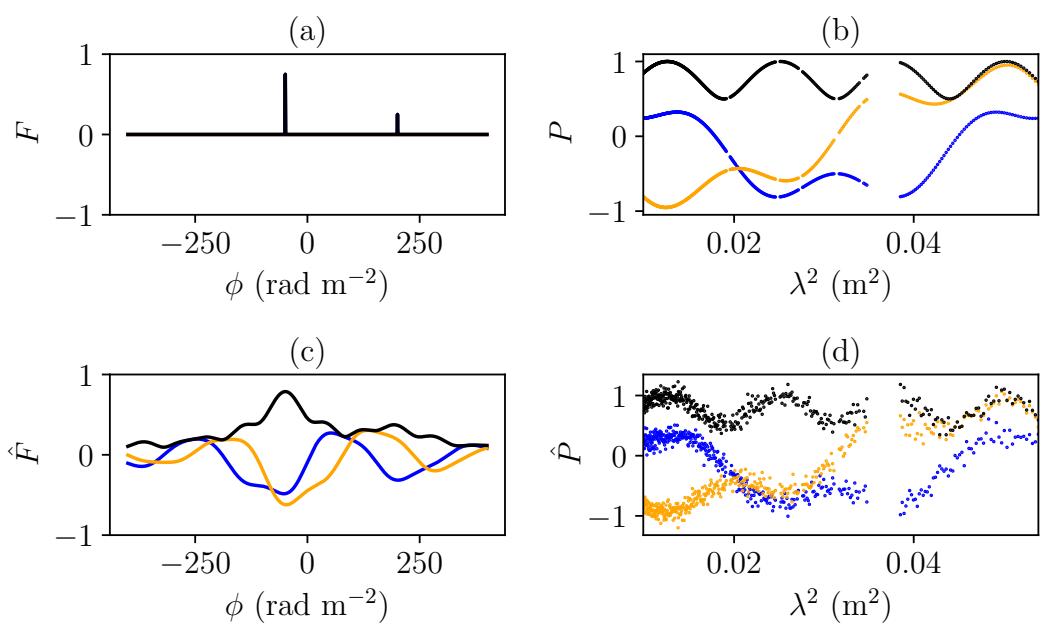


Figure 6.2: A complex FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.

sources were best explained by two or less screens, while the remainder were best explained by three screens. We model the screens by Dirac delta distributions:

$$F(\phi) = A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1). \quad (6.1)$$

A_0 and A_1 are the polarised flux of each Faraday screen, and ϕ_0 and ϕ_1 are the Faraday depths of the respective screens. With this model, a Faraday simple source is one which has $A_0 = 0$, $A_1 = 0$, or $\phi_0 = \phi_1$. By using delta distributions to model each screen, we are assuming that there is no internal Faraday dispersion (which is typically associated with diffuse emission rather than the mostly-compact sources we expect to find in wide-area polarised surveys). F generates a polarised spectrum of the form shown in Equation 6.2:

$$P(\lambda^2) = A_0e^{2i\phi_0\lambda^2} + A_1e^{2i\phi_1\lambda^2}. \quad (6.2)$$

Such a spectrum would be observed as noisy samples from a number of squared wavelengths $\lambda_j^2, j \in [1, \dots, D]$. We model this noise as a complex Gaussian with standard deviation σ and call the noisy observed spectrum \hat{P} :

$$\hat{P}(\lambda_j^2) \sim \mathcal{N}(P(\lambda_j^2), \sigma^2). \quad (6.3)$$

The constant variance of the noise is a simplifying assumption which may not hold for real data, and exploring this is a topic for future work. By performing RM synthesis (Brentjens and Bruyn, 2005) on \hat{P} with uniform weighting we arrive at an observed FDF:

$$\hat{F}(\phi) = \frac{1}{D} \sum_{j=1}^D \hat{P}(\lambda_j^2) e^{-2i\phi\lambda_j^2}. \quad (6.4)$$

Examples of F , \hat{F} , P , and \hat{P} for simple and complex observations are shown in Figure 6.1 and Figure 6.2 respectively. Note that there are two reasons that the observed FDF \hat{F} does not match the groundtruth FDF F . The first is the noise in \hat{P} . The second arises from the incomplete sampling of \hat{P} .

We do not consider external or internal Faraday dispersion in this work. External Faraday dispersion would broaden the delta functions of Equation 6.1 into peaks, and internal Faraday dispersion would broaden them into top-hat functions. All sources have at least a small amount of dispersion as the Faraday depth is a bulk property of the intervening medium and is subject to noise, but the assumption we make is that this dispersion is sufficiently small that the groundtruth FDFs are well-modelled with delta functions. Faraday thick sources would also invalidate our assumptions, and we assume that there are none in our data as Faraday thickness can be consistent with a two-component model depending on the wavelength sampling (e.g. Ma et al., 2019a; Brentjens and Bruyn, 2005). Nevertheless some external Faraday dispersion would be covered by our model, as depending on observing parameters Faraday thick sources may appear as two screens (**vaneck17faraday**).

To simulate observed FDFs we follow the method of Brown et al. (2018), which

we describe in Section P.

3 Classification approach

The Faraday complexity classification problem is as follows: Given an FDF \hat{F} , is it Faraday complex or Faraday simple? In this section we describe the features that we have developed to address this problem, which can be used in any standard machine learning classifier. We trained two classifiers on these features, which we describe here also.

3.1 Features

Our features are based on a simple idea: all simple FDFs look essentially the same, up to scaling and translation, while complex FDFs may deviate. A noise-free peak-normalised simple FDF \hat{F}_{simple} has the form

$$\hat{F}_{\text{simple}}(\phi; \phi_s) = R(\phi - \phi_s). \quad (6.5)$$

where R is the rotation measure spread function (RMSF), the Fourier transform of the wavelength sampling function which is 1 at all observed wavelengths and 0 otherwise. ϕ_s traces out a curve in the space of all possible FDFs. In other words, \hat{F}_{simple} is a manifold parametrised by ϕ_s . Our features are derived from relating an observed FDF to the manifold of simple FDFs (the ‘simple manifold’). We measure the distance of an observed FDF to the simple manifold using distance measure D_f , that take all values of the FDF into account:

$$\zeta_f(\hat{F}) = \min_{\phi_s \in \mathbb{R}} D_f(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_s)). \quad (6.6)$$

We propose two distances that have nice properties:

- invariant over changes in complex phase,
- translationally invariant in Faraday depth,
- zero for Faraday simple sources (i.e. when $A_0 = 0$, $A_1 = 0$, or $\phi_0 = \phi_1$) when there is no noise,
- symmetric in components (i.e. swapping $A_0 \leftrightarrow A_1$ and $\phi_0 \leftrightarrow \phi_1$ should not change the distance),
- increasing as A_0 and A_1 become closer to each other, and
- increasing as screen separation $|\phi_0 - \phi_1|$ increases over a large range.

Our features are constructed from this distance and its minimiser. In other words we look for the simple FDF \hat{F}_{simple} that is “closest” to the observed FDF \hat{F} . The minimiser ϕ_s is the Faraday depth of the simple FDF.

While we could choose any distance that operates on functions, we used the 2-Wasserstein (W_2) distance (??) and the Euclidean distance (??). The W_2 distance operates on probability distributions and can be thought of as the minimum cost to ‘move’ one probability distribution to the other, where the cost of moving one unit of probability mass is the squared distance it is moved. Under W_2 distance, the minimiser ϕ_w in Equation 6.6 can be interpreted as the Faraday depth that the FDF \hat{F} would be observed to have if its complexity was unresolved (i.e. the weighted mean of its components). The Euclidean distance is the square root of the least-squares loss which is often used for fitting \hat{F}_{simple} to the FDF \hat{F} . Under Euclidean distance, the minimiser ϕ_e is equivalent to the depth of the best-fitting single component under assumption of Gaussian noise in \hat{F} . We calculated the W_2 distance using Python Optimal Transport (**flamary17pot**), and we calculated the Euclidean distance using `scipy.spatial.distance.euclidean` (**scipy2020**). Further intuition about the two distances is provided in Section 3.2.

We denote by ϕ_w and ϕ_e , the Faraday depth of the simple FDF that minimises the respective distances (2-Wasserstein and Euclidean).

$$\begin{aligned}\phi_w &= \underset{\phi_w}{\operatorname{argmin}} D_{W_2}(\hat{F}(\phi) \| \hat{F}_{\text{simple}}(\phi; \phi_w)), \\ \phi_e &= \underset{\phi_e}{\operatorname{argmin}} D_E(\hat{F}(\phi) \| \hat{F}_{\text{simple}}(\phi; \phi_e)).\end{aligned}$$

These features are depicted on an example FDF in Figure 6.3. For simple observed FDFs, the fitted Faraday depths ϕ_w and ϕ_e both tend to be close to the peak of the observed FDF. However for complex observed FDFs, ϕ_w tends to be at the average depth between the two major peaks of the observed FDF, being closer to the higher peak. For notation convenience, we denote the Faraday depth of the observed FDF that has largest magnitude as ϕ_a , i.e.

$$\phi_a = \underset{\phi_a}{\operatorname{argmax}} |\hat{F}(\phi_a)|,$$

Note that in practice $\phi_a \approx \phi_e$. For complex observed FDFs, the values of Faraday depths ϕ_w and ϕ_a tend to differ (essentially by a proportion of the location of the second screen). The difference between ϕ_w and ϕ_a therefore provides useful information to identify complex FDFs. When the observed FDF is simple, the 2-Wasserstein fit will overlap significantly, hence the observed magnitudes $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$ will be similar. However, for complex FDFs ϕ_w and ϕ_a are at different depths, leading to different values of $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$. Therefore the magnitudes of the observed FDFs at the depths ϕ_w and ϕ_a indicate how different the observed FDF is from a simple FDF.

In summary, we provide the following features to the classifier:

- $\log |\phi_w - \phi_a|$,
- $\log \hat{F}(\phi_w)$,

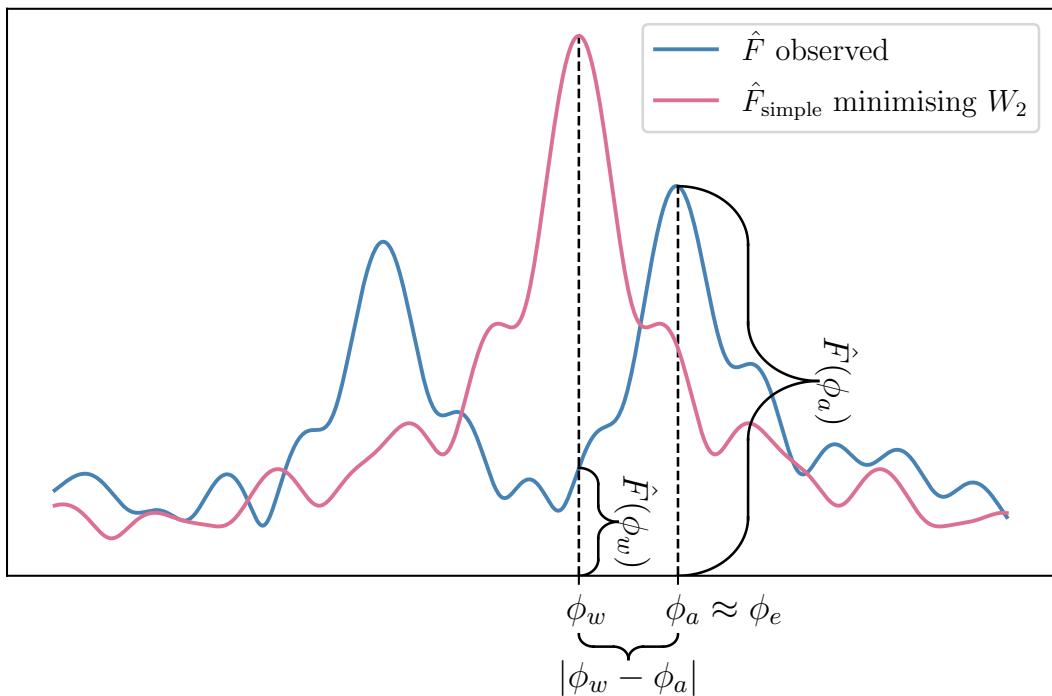


Figure 6.3: An example of how an observed FDF \hat{F} relates to our features. ϕ_w is the W_2 -minimising Faraday depth, and ϕ_a is the \hat{F} -maximising Faraday depth (approximately equal to the Euclidean-minimising Faraday depth). The remaining two features are the W_2 and Euclidean distances between the depicted FDFs.

- $\log \hat{F}(\phi_a)$,
- $\log D_{W_2}(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_w))$,
- $\log D_E(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_e))$,

where D_E is the Euclidean distance, D_{W_2} is the W_2 distance, ϕ_a is the Faraday depth of the FDF peak, ϕ_w is the minimiser for W_2 distance, and ϕ_e is the minimiser for Euclidean distance.

3.2 Interpreting distances

Interestingly, in the case where there is no RMSF, Equation 6.6 with W_2 distance reduces to the Faraday moment already in common use:

$$D_{W_2}(F) = \min_{\phi_w \in \mathbb{R}} D_{W_2}(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_w)) \quad (6.7)$$

$$= \left(\frac{A_0 A_1}{(A_0 + A_1)^2} (\phi_0 - \phi_1)^2 \right)^{1/2}. \quad (6.8)$$

See Section K for the corresponding calculation. In this sense, the W_2 distance can be thought of as a generalised Faraday moment, and conversely an interpretation of Faraday moments as a distance from the simple manifold in the case where there is no RMSF. Euclidean distance behaves quite differently in this case, and the resulting distance measure is totally independent of Faraday depth:

$$D_E(F) = \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \quad (6.9)$$

$$= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \quad (6.10)$$

See Section L for the corresponding calculation.

3.3 Classifiers

We trained two classifiers on simulated observations using these features: logistic regression (LR) and extreme gradient boosted trees (XGB). These classifiers are useful together for understanding Faraday complexity classification. LR is a linear classifier that is readily interpretable by examining the weights it applies to each feature. XGB is a powerful off-the-shelf non-linear ensemble classifier. We used the `scikit-learn` implementation of LR and we use the `XGBoost` library for XGB. We optimised hyperparameters for XGB using a fork of `xgboost-tuner`¹ as utilised by `zhu20mutagenic`. We used 1 000 iterations of randomised parameter tuning and the hyperparameters we found are tabulated in Table 6.1. We optimised hyperparameters for LR using a 5-fold cross-validation grid search implemented in

¹<https://github.com/chengsoonong/xgboost-tuner>

`sklearn.model_selection.GridSearchCV`. The resulting hyperparameters are tabulated in Table 6.2 in the Appendix.

4 Experimental method and results

We applied our classifiers to classify simulated (Section 4.2 and ??) and real (Section 4.4) FDFs. We replicated the experimental setup of Brown et al. (2018) for comparison with the state-of-the-art CNN classification method, and we also applied our method to 142 real FDFs observed with the Australia Telescope Compact Array (ATCA) from Livingston et al. (2020, submitted) and O’Sullivan et al. (2017).

4.1 Data

4.1.1 Simulated training and validation data

Our classifiers were trained and validated on simulated FDFs. We produced two sets of simulated FDFs, one for comparison with the state-of-the-art method in the literature and one for application to our observed FDFs (described in Section 4.1.2). We refer to the former as the ‘ASKAP’ dataset as it uses frequencies from the Australian Square Kilometre Array Pathfinder 12-antenna early science configuration. These frequencies included 900 channels from 700–1300 and 1500–1800 MHz and were used to generate simulated training and validation data by Brown et al. (2018). We refer to the latter as the ‘ATCA’ dataset as it uses frequencies from the 1–3 GHz configuration of the ATCA. These frequencies included 394 channels from 1.29–3.02 GHz and match our real data. We simulated Faraday depths from -50 to 50 rad m^{-2} for the ‘ASKAP’ dataset (matching Brown) and -500 to 500 for the ‘ATCA’ dataset.

For each dataset, we simulated 100 000 FDFs, approximately half simple and half complex. We randomly allocated half of these FDFs to a training set and reserved the remaining half for validation. Each FDF had complex Gaussian noise added to the corresponding polarisation spectrum. For the ‘ASKAP’ dataset, we sampled the standard deviation of the noise uniformly between 0 and $\sigma_{\max} = 0.333$, matching the dataset of Brown et al. (2018). For the ‘ATCA’ dataset, we fit a log-normal distribution to the standard deviations of O’Sullivan’s data (O’Sullivan et al., 2017) from which we sampled our values of σ :

$$\sigma \sim \frac{1}{0.63\sqrt{2\pi}\sigma} \exp\left(-\frac{\log(50\sigma - 0.5)^2}{2 \times 0.63^2}\right) \quad (6.11)$$

4.1.2 Observational data

We used two real datasets containing a total of 142 sources: 42 polarised spectra from Livingston et al. (2020, submitted) and 100 polarised spectra from O’Sullivan et al. (2017). These datasets were observed in similar frequency ranges on the same telescope (with different binning), but are in different parts of the sky. The Livingston

data were taken near the Galactic Centre, and the O’Sullivan data were taken away from the plane of the Galaxy. There are more Faraday complex sources near the Galactic Centre compared to more Faraday simple sources away from the plane of the Galaxy (Livingston et al.). The similar frequency channels used in the two datasets result in almost identical RMSFs over the Faraday depth range we considered (-500 to 500 rad m⁻²), so we expected that the classifiers would work equally well on both datasets with no need to re-train. We discarded the 26 Livingston sources with modelled Faraday depths outside of this Faraday depth range, which we do not expect to affect the applicability of our methods to wide-area surveys because these fairly high depths are not common.

Livingston et al. (in prep) used RM-CLEAN (**heald09faraday**) to identify significant components in their FDFs. Some of these components had very high Faraday depths up to 2000 rad m⁻², but we chose to ignore these components in this chapter as they are much larger than might be expected in a wide-area survey like POSSUM. They used the second Faraday moment (**Brown11report**) to estimate Faraday complexity, with Faraday depths determined using `scipy.signal.find_peaks` on the cleaned FDFs, with a cutoff of 7 times the noise of the polarised spectrum. Using this method, they estimated that 89 per cent of their sources were Faraday complex i.e. had a Faraday moment greater than 0.

O’Sullivan et al. (2017) used the QU-fitting and model selection technique described in **osullivan12agn**. The QU-fitting models contained up to three Faraday screen components as well as a term for internal and external Faraday dispersion. We ignore the Faraday thickness and dispersion for the purposes of this chapter, as most sources were not found to have Faraday thickness and dispersion is beyond the scope of our current work. 37 sources had just one component, 52 had two, and the remaining 11 had three.

4.2 Results on ‘ASKAP’ dataset

The accuracy of the LR and XGB classifiers on the ‘ASKAP’ testing set was 94.4 and 95.1 per cent respectively. The confusion matrices are shown in Table 6.5 and Table 6.6. These results are very close to the CNN presented by Brown et al. (2018), with a slightly higher true negative rate and a slightly lower true positive rate (recalling that positive sources are complex, and negative sources are simple). The accuracy of the CNN was 94.9, slightly lower than our XGB classifier and slightly higher than our LR classifier. Both of our classifiers therefore produce similar classification performance to the CNN, with faster training time and easier interpretation.

4.3 Results on ‘ATCA’ dataset

The accuracy of the LR and XGB classifiers on the ‘ATCA’ dataset was 89.2 and 90.5 per cent respectively. The major differences between the ‘ATCA’ and the ‘ASKAP’ experiments are the range of the simulated Faraday depths and the distribution of noise levels. The ‘ASKAP’ dataset, to match past CNN work, only included depths

from -50 to 50 rad m^{-2} , while the ‘ATCA’ dataset includes depths from -500 to 500 rad m^{-2} . The confusion matrices are shown in Table 6.8 and Table 6.9.

As we know the true Faraday depths of the components in our simulation, we can investigate the behaviour of these classifiers as a function of physical properties. Figure 6.4 shows the mean classifier prediction as a function of component depth separation and minimum component amplitude. This is tightly related to the mean accuracy, as the entire plot domain contains complex spectra besides the left and bottom edge: by thresholding the classifier prediction to a certain value, the accuracy will be one hundred per cent on the non-edge for all sources with higher prediction values.

4.4 Results on observed FDFs

We used the LR and XGB classifiers which were trained on the ‘ATCA’ dataset to estimate the probability that our 142 observed FDFs (Section 4.1.2) were Faraday complex. As these classifiers were trained on simulated data, they face the issue of the ‘domain gap’: the distribution of samples from a simulation differs from the distribution of real sources, and this affects performance on real data. Solving this issue is called ‘domain adaptation’ and how to do this is an open research question in machine learning ([zhang2019transfer](#); [pan10transfer](#)). Nevertheless, the features of our observations mostly fall in the same region of feature space as the simulations (Figure 6.5) and so we expect reasonably good domain transfer.

Two apparently complex sources in the Livingston sample are classified as simple with high probability by XGB. These outliers are on the very edge of the training sample (Figure 6.5) and the underdensity of training data here is likely the cause of this issue. LR does not suffer the same issue, producing plausible predictions for the entire dataset, and these sources are instead classified as complex with high probability.

With a threshold of 0.5, LR predicted that 96 and 83 per cent of the Livingston and O’Sullivan sources were complex respectively. This is in line with expectations that the Livingston data should have more Faraday complex sources than the O’Sullivan data due to their location near the Galactic Centre. XGB predicted that 93 and 100 per cent of the Livingston and O’Sullivan sources were complex respectively. Livingston et al. (in prep) found that 90 per cent of their sources were complex, and O’Sullivan et al. (2017) found that 64 per cent of their sources were complex. This suggests that our classifiers are overestimating complexity, though it could also be the case that the methods used by Livingston and O’Sullivan underestimate complexity. Modifying the prediction threshold from 0.5 changes the estimated rate of Faraday complexity, and we show the estimated rates against threshold for both classifiers in Figure 6.6. We suggest that this result is indicative of our probabilities being uncalibrated, and a higher threshold should be chosen in practice. We chose to keep the threshold at 0.5 as this had the highest accuracy on the simulated validation data. The very high complexity rates of XGB and two outlying classifications indicate that the XGB classifier may be overfitting to the simulation and that it is unable to generalise across

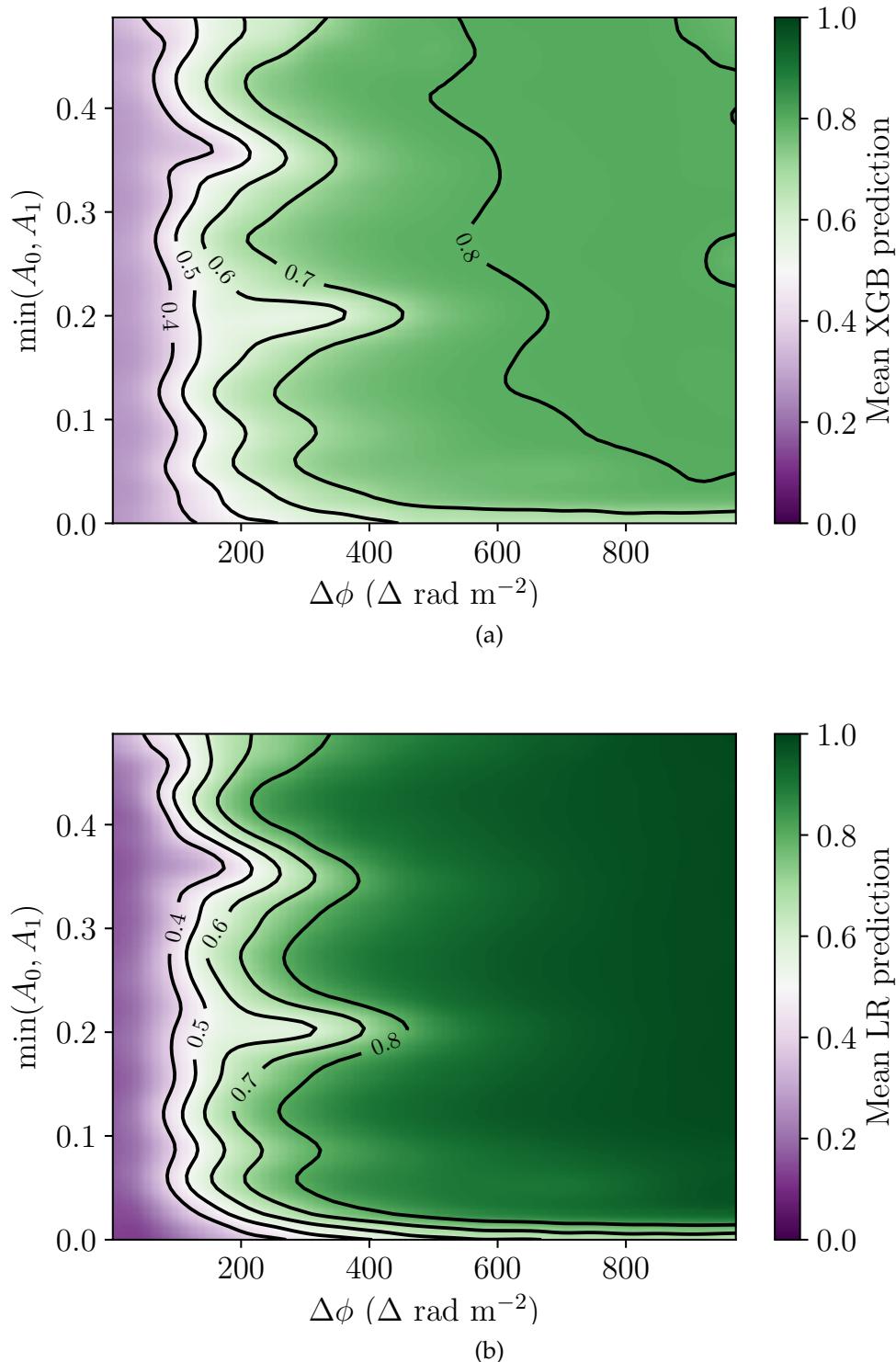


Figure 6.4: Mean prediction as a function of component depth separation and minimum component amplitude for (a) XGB and (b) LR.

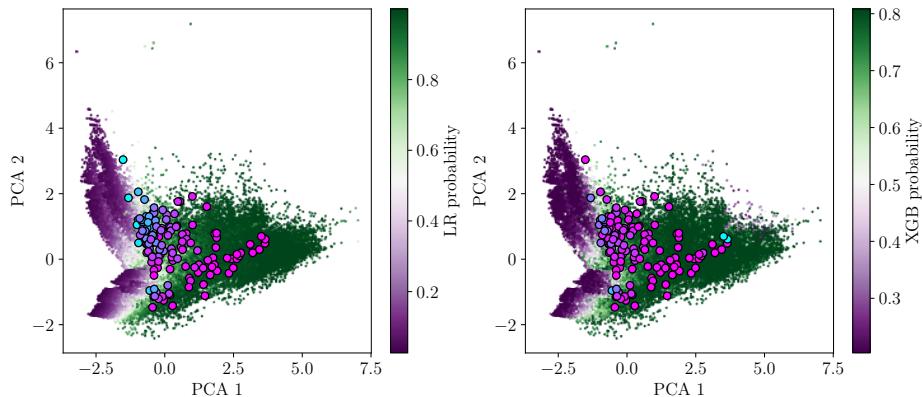


Figure 6.5: Principal component analysis for simulated data (coloured dots) with observations overlaid (black-edged circles). Observations are coloured by their XGB or LR estimated probability of being complex, with blue indicating ‘most simple’ and pink indicating ‘most complex’.

the domain gap.

Figure 6.7 and Figure 6.8 show every observed FDF ordered by estimated Faraday complexity, alongside the models predicted by Livingston and O’Sullivan et al. (2017), for LR and XGB respectively. There is a clear visual trend of increasingly complex sources with increasing predicted probability of being complex.

5 Discussion

On simulated data (Section 4.3) we achieve state-of-the-art accuracy. Our results on observed FDFs show that our classifiers produce plausible results, with Figure 6.7 and Figure 6.8 showing a clear trend of apparent complexity. Some issues remain: we discuss the intrinsic overlap between simple and complex FDFs in Section 5.1 and the limitations of our method in Section 5.2.

5.1 Complexity and seeming ‘not simple’

Through this work we found our methods limited by the significant overlap between complex and simple FDFs. Complex FDFs can be consistent with simple FDFs due to close Faraday components or very small amplitudes on the secondary component, and vice versa due to noise.

The main failure mode of our classifiers is misclassifying a complex source as simple (Table 6.8 and Table 6.9). Whether sources with close components or small amplitudes should be considered complex is not clear, since for practical purposes they can be treated as simple: assuming the source is simple yields a very similar RM to the RM of the primary component, and thus would not negatively impact further data products such as an RM grid. The scenarios where we would want a

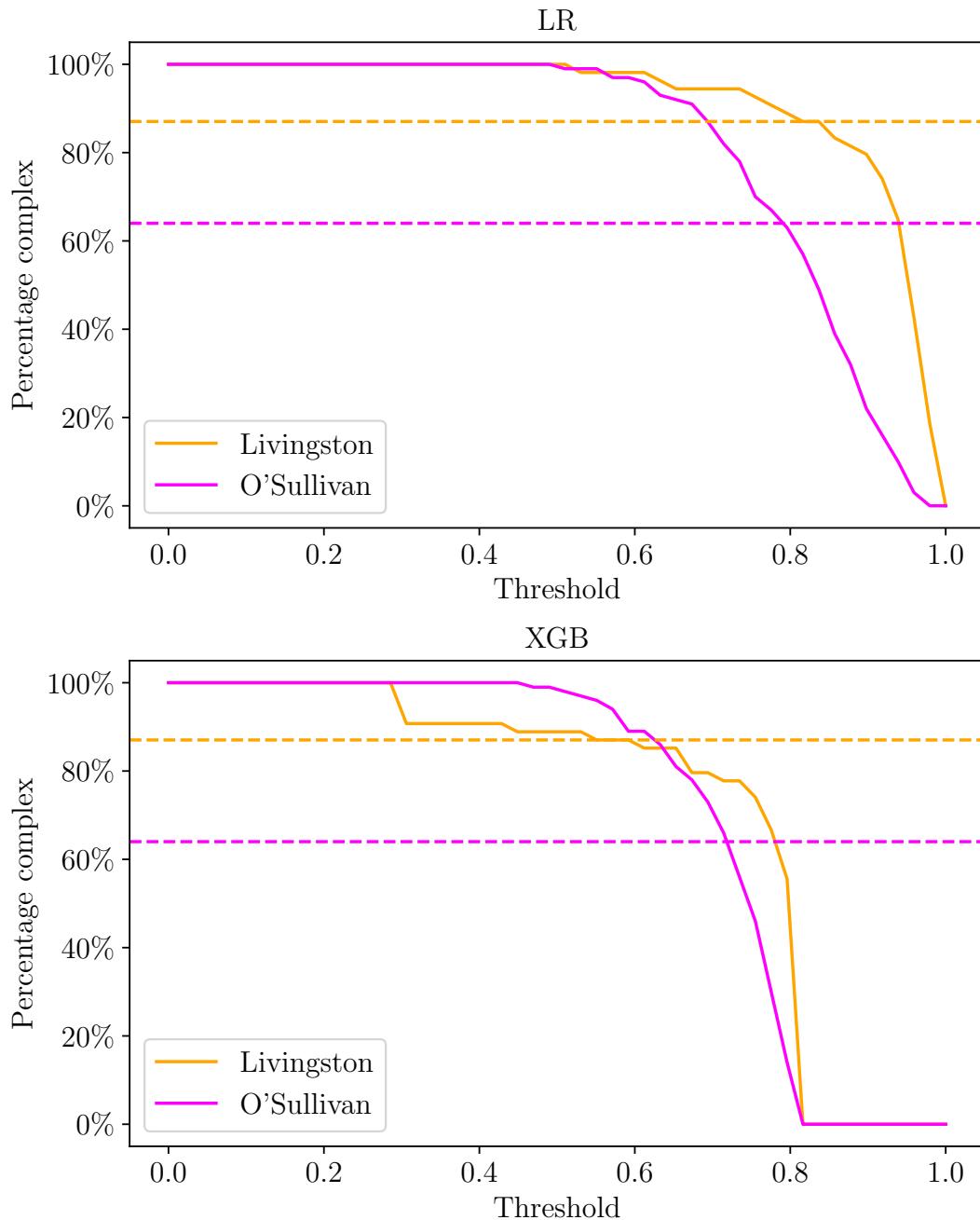


Figure 6.6: Estimated rates of Faraday complexity for the Livingstone and O'Sullivan datasets as functions of threshold. The horizontal lines indicate the rates of Faraday complexity estimated by Livingston and O'Sullivan respectively.

Faraday complexity classifier rather than a polarisation structure model – large-scale analysis and wide-area surveys – do not seem to be disadvantaged by considering such sources simple. Additional sources similar to these are likely hidden in presumably ‘simple’ FDFs by the frequency range and spacing of the observations, just as how these complex sources would be hidden in lower-resolution observations. Note also that misidentification of complex sources as simple is intrinsically a problem with complexity estimation even for models not well-represented by a simple FDF, as complex sources may conspire to appear as a wide range of viable models including simple (`sun15comparison`).

Conversely, high-noise simple FDFs may be consistent with complex FDFs. One key question is how Faraday complexity estimators should behave as the noise increases: should high noise result in a complex prediction or a simple prediction, given that a complex or simple FDF would both be consistent with a noisy FDF? Occam’s razor suggests that we should choose the simplest suitable model, and so increasing noise should lead to predictions of less complexity. This is not how our classifiers operate, however: high-noise FDFs are different to the model simple FDFs and so are predicted to be ‘not simple’. In some sense our classifiers are not looking for complex sources, but are rather looking for ‘not simple’ sources.

5.2 Limitations

Our main limitations are our simplifying assumptions on FDFs and the domain gap between simulated and real observations. However, our proposed features (Section Section 3.1) can be applied to future improved simulations.

It is unclear what the effect of our simplifying assumptions are on the effectiveness of our simulation. The three main simplifications that may negatively affect our simulations are 1) limiting to two components, 2) assuming no external Faraday dispersion, and 3) assuming no internal Faraday dispersion (Faraday thickness). Future work will explore removing these simplifying assumptions, but will need to account for the increased difficulty in characterising the simulation with more components and no longer having Faraday screens as components. Additionally, more work will be required to make sure that the rates of internal and external Faraday dispersion match what might be expected from real sources, or risk making a simulation that has too large a range of consistent models for a given source: for example, a two-component source could also be explained as a sufficiently wide or resolved-out Faraday thick source or a three-component source with a small third component. This greatly complicates the classification task.

Previous machine learning work (e.g. Brown et al., 2018) has not been run before on real FDF data, so this chapter is the first example of the domain gap arising in Faraday complexity classification. This is a problem that requires further research to solve. We have no good way to ensure that our simulation matches reality, so some amount of domain adaptation will always be necessary to train classifiers on simulated data and then apply these classifiers to real data. But with the low source counts in polarisation science (high-resolution spectropolarimetric data currently numbers

in the few hundreds) any machine learning method will need to be trained on simulations. This is not just a problem in Faraday complexity estimation, and domain adaptation is also an issue faced in the wider astroinformatics community: large quantities of labelled data are hard to come by, and some sources are very rare ([zevin17gravityspy](#); [gebbhard19convolutional](#); [agarwal20fetch](#)). LR seems to handle the domain adaptation better than XGB, with only a slightly lower accuracy on simulated data. Our results are plausible and the distribution of our simulation well overlaps the distribution of our real data (Figure 6.5).

6 Conclusion

We developed a simple, interpretable machine learning method for estimating Faraday complexity. Our interpretable features were derived by comparing observed FDFs to idealised simple FDFs, which we could determine both for simulated and real observations. We demonstrated the effectiveness of our method on both simulated and real data. Using simulated data, we found that our classifiers were 95 per cent accurate, with near perfect recall (specificity) of Faraday simple sources. On simulated data that matched existing observations, our classifiers obtained an accuracy of 90 per cent. Evaluating our classifiers on real data gave the plausible results shown in Figure 6.7, and marks the first application of machine learning to observed FDFs. Future work will need to narrow the domain gap to improve transfer of classifiers trained on simulations to real, observed data.

K z -Wasserstein begets Faraday moments

Minimising the 2-Wasserstein distance between a model FDF and the simple manifold gives the second Faraday moment of that FDF. Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the sum-normalised simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1)}{A_0 + A_1} \quad (6.12)$$

$$\tilde{S}(\phi; \phi_w) = \delta(\phi - \phi_w). \quad (6.13)$$

The W_2 distance, usually defined on probability distributions, can be extended to one-dimensional complex functions A and B by normalising them:

$$D_{W_2}(A \parallel B)^2 = \inf_{\gamma \in \Gamma(A, B)} \iint_{\phi_{\min}}^{\phi_{\max}} |x - y|^2 d\gamma(x, y) \quad (6.14)$$

$$\tilde{A}(\phi) = \frac{|A(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |A(\theta)| d\theta} \quad (6.15)$$

$$\tilde{B}(\phi) = \frac{|B(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |B(\theta)| d\theta} \quad (6.16)$$

where $\Gamma(A, B)$ is the set of couplings of A and B , i.e. the set of joint probability distributions that marginalise to A and B ; and $\inf_{\gamma \in \Gamma(A, B)}$ is the infimum over $\Gamma(A, B)$. This can be interpreted as the minimum cost to ‘move’ one probability distribution to the other, where the cost of moving one unit of probability mass is the squared distance it is moved.

The set of couplings $\Gamma(\tilde{F}, \tilde{S})$ is the set of all joint probability distributions γ such that

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\phi = \tilde{S}(\varphi; \phi_w), \quad (6.17)$$

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\varphi = \tilde{F}(\phi). \quad (6.18)$$

The coupling that minimises the integral in Equation 6.14 will be the optimal transport plan between \tilde{F} and \tilde{S} . Since \tilde{F} and \tilde{S} are defined in terms of delta functions, the optimal transport problem reduces to a discrete optimal transport problem and the optimal transport plan is:

$$\gamma(\phi, \varphi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1} \delta(\varphi - \phi_w). \quad (6.19)$$

In other words, to move the probability mass of \tilde{S} to \tilde{F} , a fraction $A_0/(A_0 + A_1)$ is moved from ϕ_w to ϕ_0 and the complementary fraction $A_1/(A_0 + A_1)$ is moved from ϕ_w to ϕ_1 . Then:

$$D_{W_2}(\tilde{F} \parallel \tilde{S})^2 = \iint_{\phi_{\min}}^{\phi_{\max}} |\phi - \varphi|^2 d\gamma(\phi, \varphi) \quad (6.20)$$

$$= \frac{A_0(\phi_0 - \phi_w)^2 + A_1(\phi_1 - \phi_w)^2}{A_0 + A_1}. \quad (6.21)$$

To obtain the W_2 distance to the simple manifold, we need to minimise this over ϕ_w . Differentiate with respect to ϕ_w and set equal to zero to find

$$\phi_w = \frac{A_0 \phi_0 + A_1 \phi_1}{A_0 + A_1}. \quad (6.22)$$

Substituting this back in, we find

$$\zeta_{W_2}(F)^2 = \frac{A_0 A_1}{A_0 + A_1} (\phi_0 - \phi_1)^2 \quad (6.23)$$

which is the Faraday moment.

L Euclidean distance in the no-RMSF case

In this section we calculate the minimumised Euclidean distance evaluated on a model FDF (Equation 6.1). Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be

the normalised simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1)}{A_0 + A_1} \quad (6.24)$$

$$\tilde{S}(\phi; \phi_e) = \delta(\phi - \phi_e). \quad (6.25)$$

The Euclidean distance between \tilde{F} and \tilde{S} is then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (6.26)$$

$$= \int_{\phi_{\min}}^{\phi_{\max}} |\tilde{F}(\phi) - \delta(\phi - \phi_e)|^2 d\phi. \quad (6.27)$$

Assume $\phi_0 \neq \phi_1$ (otherwise, D_E will always be either 0 or 2). If $\phi_e = \phi_0$, then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (6.28)$$

$$= \frac{1}{(A_0 + A_1)^2} \int_{\phi_{\min}}^{\phi_{\max}} A_1^2 |\delta(\phi - \phi_1) - \delta(\phi - \phi_0)|^2 d\phi \quad (6.29)$$

$$= \frac{2A_1^2}{(A_0 + A_1)^2} \quad (6.30)$$

and similarly for $\phi_e = \phi_1$. If $\phi_e \neq \phi_0$ and $\phi_e \neq \phi_1$, then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 = \frac{A_0^2 + A_1^2 + 1}{(A_0 + A_1)^2}. \quad (6.31)$$

The minimised Euclidean distance when $\phi_0 \neq \phi_1$ is therefore

$$D_E(F) = \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \quad (6.32)$$

$$= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \quad (6.33)$$

If $\phi_0 = \phi_1$, then the minimised Euclidean distance is 0.

M Hyperparameters for LR and XGB

This section contains tables of the hyperparameters that we used for our classifiers. Table 6.1 and Table 6.2 tabulate the hyperparameters for XGB and LR respectively for the ‘ATCA’ dataset. Table 6.3 and Table 6.4 tabulate the hyperparameters for XGB and LR respectively for the ‘ASKAP’ dataset.

N Confusion matrices

This section contains confusion matrices for our classifiers evaluated on simulated ATCA and ASKAP data, as well as a confusion matrix from the state-of-the-art CNN

Table 6.1: XGB hyperparameters for the ‘ATCA’ dataset.

Parameter	Value
colsample_bytree	0.912
gamma	0.532
learning_rate	0.1
max_depth	7
min_child_weight	2
scale_pos_weight	1
subsample	0.557
n_estimators	135
reg_alpha	0.968
reg_lambda	1.420

Table 6.2: LR hyperparameters for the ‘ATCA’ dataset.

Parameter	Value
penalty	L1
C	1.668

Table 6.3: XGB hyperparameters for the ‘ASKAP’ dataset.

Parameter	Value
colsample_bytree	0.865
gamma	0.256
learning_rate	0.1
max_depth	6
min_child_weight	1
scale_pos_weight	1
subsample	0.819
n_estimators	108
reg_alpha	0.049
reg_lambda	0.454

Table 6.4: LR hyperparameters for the ‘ASKAP’ dataset.

Parameter	Value
penalty	L2
C	0.464

classifier on simulated ASKAP data. Table 6.5, Table 6.5, and Table 6.7 are confusion matrices on the ‘ASKAP’ dataset for LR, XGB, and CNN respectively. Table 6.8 and Table 6.9 are confusion matrices on the ‘ATCA’ dataset for LR and XGB respectively.

Table 6.5: Logistic regression confusion matrix for the ‘ASKAP’ dataset.

	Pred. simple	Pred. complex
True simple	0.99	0.01
True complex	0.10	0.90

Table 6.6: XGB confusion matrix for the ‘ASKAP’ dataset.

	Pred. simple	Pred. complex
True simple	0.99	0.01
True complex	0.09	0.91

Table 6.7: CNN confusion matrix for the ‘ASKAP’, adapted from Brown et al. (2018).

	Pred. simple	Pred. complex
True simple	0.97	0.03
True complex	0.07	0.93

Table 6.8: Logistic regression confusion matrix for the ‘ATCA’ dataset.

	Pred. simple	Pred. complex
True simple	0.92	0.08
True complex	0.16	0.84

Table 6.9: XGB confusion matrix for the ‘ATCA’ dataset.

	Pred. simple	Pred. complex
True simple	0.91	0.09
True complex	0.10	0.90

O Predictions on real data

This section contains Figure 6.7 and Figure 6.8, which shows the predicted probability of being Faraday complex for all real data used in this chapter, drawn from Livingston et al. (2020, submitted) and O’Sullivan et al. (2017).

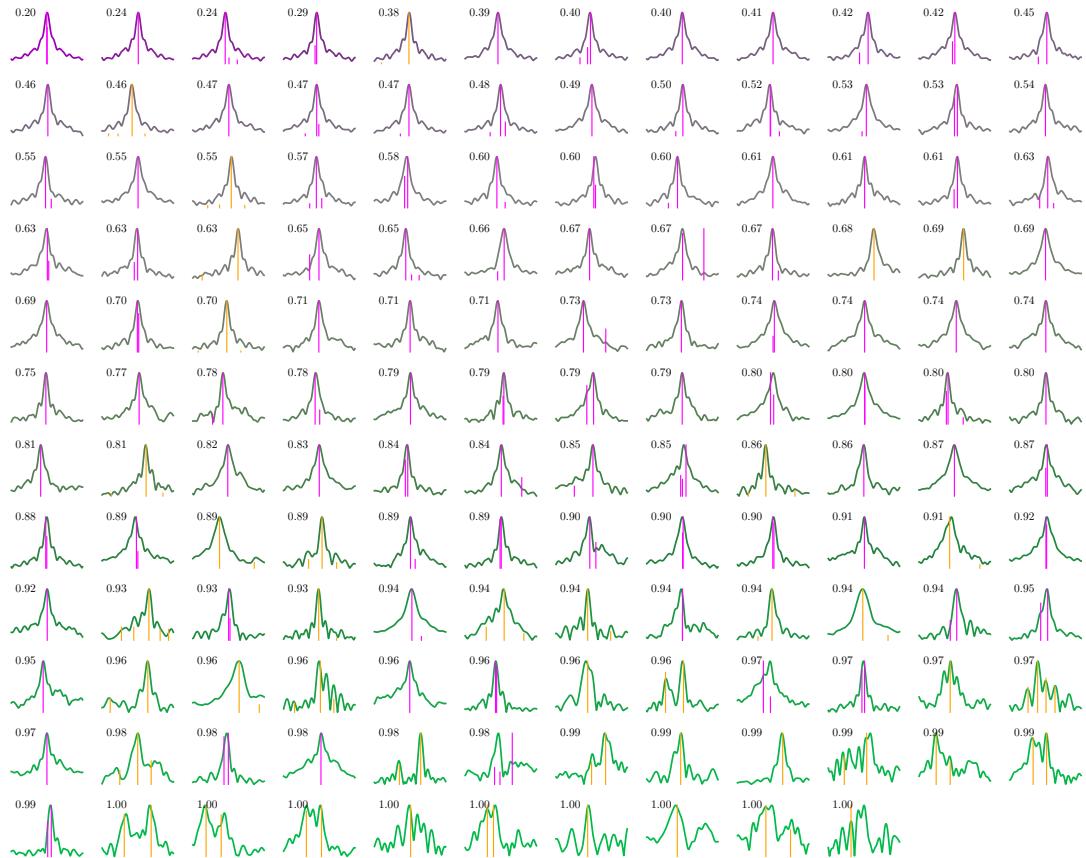


Figure 6.7: The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O’Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the LR estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.

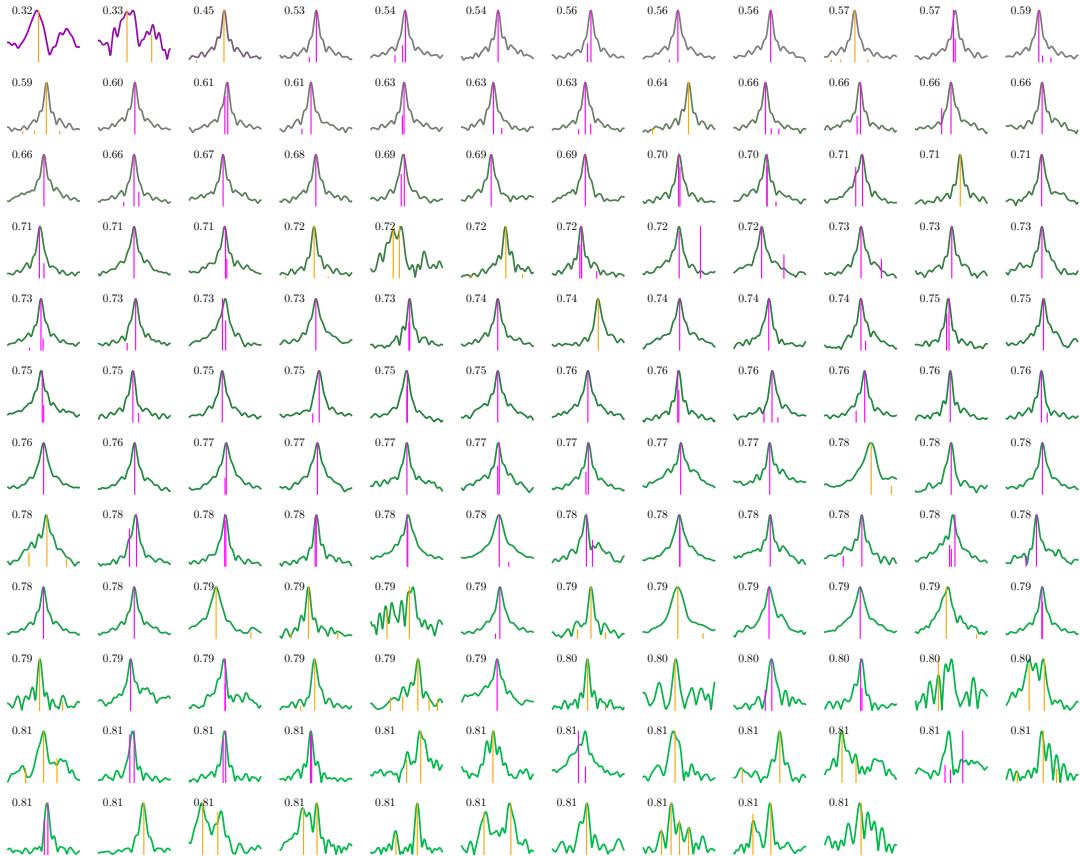


Figure 6.8: The 142 observed FDFs ordered by XGB-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O'Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the XGB estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.

P Simulating observed FDFs

We simulated FDFs by approximating them by arrays of complex numbers. An FDF F is approximated on the domain $[-\phi_{\max}, \phi_{\max}]$ by a vector $\vec{F} \in \mathbb{R}^d$:

$$\vec{F}_j = \sum_{k=0}^1 A_k \delta(-\phi_{\max} + j\delta\phi - \phi_k) \quad (6.34)$$

where $\delta\phi = (\phi_{\max} - \phi_{\min})/d$ and d is the number of Faraday depth samples in the FDF. \vec{F} is sampled by uniformly sampling its parameters:

$$\phi_k \in [\phi_{\min}, \phi_{\min} + \delta\phi, \dots, \phi_{\max}] \quad (6.35)$$

$$A_k \sim \mathcal{U}(0, 1). \quad (6.36)$$

We then generate a vector polarisation spectrum $\vec{P} \in \mathbb{R}^m$ from \vec{F} using a Equation 6.37:

$$\vec{P}_\ell = \sum_{j=0}^j F_j e^{2i(\phi_{\min} + j\delta\phi)\lambda_\ell^2} d\phi. \quad (6.37)$$

λ_ℓ^2 is the discretised value of λ^2 at the ℓ th index of \vec{P} . This requires a set of λ^2 values, which depends on the dataset being simulated. These values can be treated as the channel wavelengths at which the polarisation spectrum was observed. We then add Gaussian noise with variance σ^2 to each element of \vec{P} to obtain a discretised noisy observation $\hat{\vec{P}}$. Finally, we perform RM synthesis using the Canadian Initiative for Radio Astronomy Data Analysis RM package², which is a Python module that implements a discrete version of RM synthesis:

$$\hat{\vec{F}}_j = m^{-1} \sum_{\ell=1}^m \vec{P}_\ell e^{-2i(\phi_{\min} + j\delta\phi)\lambda_\ell^2}. \quad (6.38)$$

²<https://github.com/CIRADA-Tools/RM>

Conclusion

Summary your thesis and discuss what you are going to do in the future in Section 1.

A conclusion drawing together the published papers or works in a cohesive manner, and addresses how the individual publications link to the theory and methodology adopted and evaluate the contribution that the research in the submitted publications makes to the advancement of the research area.

1 Future Work

Good luck.

Appendix

Bibliography

- Agarwal, Devansh et al. (June 2020). "FETCH: A deep-learning based classifier for fast transient classification". In: *Monthly Notices of the Royal Astronomical Society*. doi: 10.1093/mnras/staa1856. url: <http://adsabs.harvard.edu/abs/2020MNRAS.498.1750A> (visited on 07/31/2020).
- Aguado, D. S. 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].
- Alger, M. J. et al. (Aug. 2018). "Radio Galaxy Zoo: machine learning for radio source host galaxy cross-identification". In: *MNRAS* 478, pp. 5547–5563. doi: 10.1093/mnras/sty1308. arXiv: 1805.05540 [astro-ph.IM].
- Alhassan, Wathela, A. R. Taylor, and Mattia Vaccari (Oct. 2018). "The FIRST Classifier: compact and extended radio galaxy classification using deep Convolutional Neural Networks". In: *Monthly Notices of the Royal Astronomical Society* 480, pp. 2085–2093. issn: 0035-8711. doi: 10.1093/mnras/sty2038. url: <http://adsabs.harvard.edu/abs/2018MNRAS.480.2085A> (visited on 12/08/2020).
- Anderson, C. S. et al. (Dec. 2015). "Broadband Radio Polarimetry and Faraday Rotation of 563 Extragalactic Radio Sources". In: *The Astrophysical Journal* 815, p. 49. issn: 0004-637X. doi: 10.1088/0004-637X/815/1/49. url: <http://adsabs.harvard.edu/abs/2015ApJ...815...49A> (visited on 09/12/2019).
- Aniyan, A. K. and K. Thorat (June 2017). "Classifying Radio Galaxies with the Convolutional Neural Network". In: *ApJS* 230, 20, p. 20. doi: 10.3847/1538-4365/aa7333. arXiv: 1705.03413 [astro-ph.IM].
- Antonucci, Robert (1993). "Unified Models for Active Galactic Nuclei and Quasars". In: *Annual Review of Astronomy and Astrophysics* 31.1, pp. 473–521. doi: 10.1146/annurev.aa.31.090193.002353. url: <https://doi.org/10.1146/annurev.aa.31.090193.002353> (visited on 07/03/2020).
- Arsioli, B. and P. Dedin (Aug. 2020). "Machine learning applied to multifrequency data in astrophysics: blazar classification". In: *Monthly Notices of the Royal Astronomical Society* 498, pp. 1750–1764. issn: 0035-8711. doi: 10.1093/mnras/staa2449. url: <http://adsabs.harvard.edu/abs/2020MNRAS.498.1750A> (visited on 12/08/2020).
- Astropy Collaboration et al. (Oct. 2013). "Astropy: A community Python package for astronomy". In: *A&A* 558, A33, A33. doi: 10.1051/0004-6361/201322068.
- Balakrishnan, Vishnu, David Champion, and Ewan Barr (Oct. 2020). "Pulsar Candidate Identification Using Semi-Supervised Generative Adversarial Networks". In: *arXiv e-prints* 2010, arXiv:2010.07457. url: <http://adsabs.harvard.edu/abs/2020arXiv201007457B> (visited on 12/08/2020).

- Ball, N. M. et al. (Mar. 2004). "Galaxy types in the Sloan Digital Sky Survey using supervised artificial neural networks". In: *Monthly Notices of the Royal Astronomical Society* 348, pp. 1038–1046. issn: 0035-8711. doi: 10.1111/j.1365-2966.2004.07429.x. URL: <http://adsabs.harvard.edu/abs/2004MNRAS.348.1038B> (visited on 12/06/2020).
- Ball, Nicholas M. et al. (Oct. 2006). "Robust Machine Learning Applied to Astronomical Data Sets. I. Star-Galaxy Classification of the Sloan Digital Sky Survey DR3 Using Decision Trees". In: *The Astrophysical Journal* 650, pp. 497–509. issn: 0004-637X. doi: 10.1086/507440. URL: <http://adsabs.harvard.edu/abs/2006ApJ...650..497B> (visited on 12/06/2020).
- Bañados, Eduardo et al. (Jan. 2018). "An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5". en. In: *Nature* 553.7689. Number: 7689 Publisher: Nature Publishing Group, pp. 473–476. issn: 1476-4687. doi: 10.1038/nature25180. URL: <https://www.nature.com/articles/nature25180> (visited on 11/08/2020).
- Banerji, Manda et al. (July 2010). "Galaxy Zoo: reproducing galaxy morphologies via machine learning". In: *Monthly Notices of the Royal Astronomical Society* 406, pp. 342–353. issn: 0035-8711. doi: 10.1111/j.1365-2966.2010.16713.x. URL: <http://adsabs.harvard.edu/abs/2010MNRAS.406..342B> (visited on 12/06/2020).
- Banfield, J. K. et al. (Nov. 2015). "Radio Galaxy Zoo: host galaxies and radio morphologies derived from visual inspection". In: *MNRAS* 453, pp. 2326–2340. doi: 10.1093/mnras/stv1688. arXiv: 1507.07272.
- Banfield, J. K. et al. (Aug. 2016). "Radio Galaxy Zoo: discovery of a poor cluster through a giant wide-angle tail radio galaxy". In: *Monthly Notices of the Royal Astronomical Society* 460, pp. 2376–2384. issn: 0035-8711. doi: 10.1093/mnras/stw1067. URL: <http://adsabs.harvard.edu/abs/2016MNRAS.460.2376B> (visited on 07/13/2020).
- Bastien, David, Nadeem Oozeer, and Radhakrishna Somanah (May 2017). "Classifying bent radio galaxies from a mixture of point-like/extended images with Machine Learning." In: 198. Conference Name: Materials Science and Engineering Conference Series, p. 012013. doi: 10.1088/1757-899X/198/1/012013. URL: <http://adsabs.harvard.edu/abs/2017MS%26E..198a2013B> (visited on 12/08/2020).
- Becklin, Pedro P. B. et al. (July 2020). "AGN dichotomy beyond radio loudness: a Gaussian mixture model analysis". In: *Monthly Notices of the Royal Astronomical Society* 497, pp. 1463–1474. issn: 0035-8711. doi: 10.1093/mnras/staa2072. URL: <http://adsabs.harvard.edu/abs/2020MNRAS.497.1463B> (visited on 12/08/2020).
- Becker, R. H., R. L. White, and D. J. Helfand (Sept. 1995). "The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters". In: *ApJ* 450, p. 559. doi: 10.1086/176166.
- Bertin, E. and S. Arnouts (June 1996). "SExtractor: Software for source extraction." In: *A&AS* 117, pp. 393–404. doi: 10.1051/aas:1996164.
- Best, P. N. (Feb. 2009). "Radio source populations: Results from SDSS". In: *Astronomische Nachrichten* 330, pp. 184–189. doi: 10.1002/asna.200811152.

- Best, P. N. et al. (Sept. 2005). "The host galaxies of radio-loud active galactic nuclei: mass dependences, gas cooling and active galactic nuclei feedback". In: MNRAS 362.1, pp. 25–40. ISSN: 0035-8711. doi: 10.1111/j.1365-2966.2005.09192.x. eprint: <http://oup.prod.sis.lan/mnras/article-pdf/362/1/25/6029587/362-1-25.pdf>. URL: <https://dx.doi.org/10.1111/j.1365-2966.2005.09192.x>.
- Bicknell, G. V., M. A. Dopita, and C. P. O. O'Dea (Aug. 1997). "Unification of the Radio and Optical Properties of Gigahertz Peak Spectrum and Compact Steep-Spectrum Radio Sources". In: ApJ 485, pp. 112–124. doi: 10.1086/304400.
- Bicknell, Geoffrey V. (Nov. 1995). "Relativistic Jets and the Fanaroff-Riley Classification of Radio Galaxies". In: *The Astrophysical Journal Supplement Series* 101, p. 29. doi: 10.1086/192232. URL: <http://adsabs.harvard.edu/abs/1995ApJS..101...29B> (visited on 11/10/2020).
- Bîrzan, L. et al. (Oct. 2008). "Radiative Efficiency and Content of Extragalactic Radio Sources: Toward a Universal Scaling Relation between Jet Power and Radio Power". In: ApJ 686, pp. 859–880. doi: 10.1086/591416. arXiv: 0806.1929.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Black, A. R. S. et al. (May 1992). "A study of FRII radio galaxies with Z less than 0.15. I - High-resolution maps of eight sources at 3.6 CM". In: *Monthly Notices of the Royal Astronomical Society* 256, pp. 186–208. ISSN: 0035-8711. doi: 10.1093/mnras/256.2.186. URL: <http://adsabs.harvard.edu/abs/1992MNRAS.256..186B> (visited on 07/13/2020).
- Blundell, K. M., S. Rawlings, and C. J. Willott (Feb. 1999). "The Nature and Evolution of Classical Double Radio Sources from Complete Samples". In: AJ 117, pp. 677–706. doi: 10.1086/300721. eprint: astro-ph/9810197.
- Bowles, Micah et al. (Dec. 2020). "Attention-gating for improved radio galaxy classification". en. In: *arXiv:2012.01248 [astro-ph]*. arXiv: 2012.01248. URL: <http://arxiv.org/abs/2012.01248> (visited on 12/08/2020).
- Breiman, Leo (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32.
- Brentjens, M. A. and A. G. de Bruyn (Oct. 2005). "Faraday Rotation Measure Synthesis". en. In: *Astronomy & Astrophysics* 441.3. arXiv: astro-ph/0507349, pp. 1217–1228. ISSN: 0004-6361, 1432-0746. doi: 10.1051/0004-6361:20052990. URL: <http://arxiv.org/abs/astro-ph/0507349> (visited on 09/12/2019).
- Brown, Shea et al. (Oct. 2018). "Classifying Complex Faraday Spectra with Convolutional Neural Networks". en. In: *Monthly Notices of the Royal Astronomical Society*. arXiv: 1711.03252. ISSN: 0035-8711, 1365-2966. doi: 10.1093/mnras/sty2908. URL: <http://arxiv.org/abs/1711.03252> (visited on 09/12/2019).
- Capetti, A. et al. (Oct. 2020). "The LOFAR view of FR 0 radio galaxies". In: *Astronomy and Astrophysics* 642, A107. ISSN: 0004-6361. doi: 10.1051/0004-6361/202038671. URL: <http://adsabs.harvard.edu/abs/2020A&A...642A.107C> (visited on 12/06/2020).
- Carballo, R., A. S. Cofiño, and J. I. González-Serrano (Sept. 2004). "Selection of quasar candidates from combined radio and optical surveys using neural networks". In: *Monthly Notices of the Royal Astronomical Society* 353, pp. 211–220. ISSN: 0035-8711.

- doi: 10.1111/j.1365-2966.2004.08056.x. url: <http://adsabs.harvard.edu/abs/2004MNRAS.353..211C> (visited on 12/06/2020).
- Cavagnolo, K. W. et al. (Sept. 2010). "A Relationship Between AGN Jet Power and Radio Power". In: *ApJ* 720, pp. 1066–1072. doi: 10.1088/0004-637X/720/2/1066. arXiv: 1006.5699 [astro-ph.CO].
- Cheung, C. C. (May 2007). "FIRST "Winged" and X-Shaped Radio Source Candidates". en. In: *The Astronomical Journal* 133.5, pp. 2097–2121. issn: 0004-6256, 1538-3881. doi: 10.1086/513095. url: <https://iopscience.iop.org/article/10.1086/513095> (visited on 12/06/2020).
- Chollet, François et al. (2015). *Keras*. <https://github.com/fchollet/keras>.
- Collier, J. D. et al. (Mar. 2014). "Infrared-faint radio sources: a new population of high-redshift radio galaxies". In: *MNRAS* 439, pp. 545–565. doi: 10.1093/mnras/stt2485. arXiv: 1312.1002.
- Condon, J. J. (1991). "Radio luminosity functions". In: *The Interpretation of Modern Synthesis Observations of Spiral Galaxies*. Ed. by N. Duric and P. C. Crane. Vol. 18. Astronomical Society of the Pacific Conference Series, pp. 113–123.
- (1992a). "Radio emission from normal galaxies". In: *Annual Review of Astronomy and Astrophysics* 30, pp. 575–611. issn: 0066-4146. doi: 10.1146/annurev.aa.30.090192.003043. url: <http://adsabs.harvard.edu/abs/1992ARA%26A..30..575C> (visited on 07/12/2020).
- (1992b). "Radio emission from normal galaxies". In: *ARA&A* 30, pp. 575–611. doi: 10.1146/annurev.aa.30.090192.003043.
- Condon, J. J., W. D. Cotton, and J. J. Broderick (Aug. 2002). "Radio Sources and Star Formation in the Local Universe". In: *The Astronomical Journal* 124, pp. 675–689. issn: 0004-6256. doi: 10.1086/341650.
- Condon, J. J., A. M. Matthews, and J. J. Broderick (Feb. 2019). "Radio Sources in the Nearby Universe". In: *ApJ* 872, 148, p. 148. doi: 10.3847/1538-4357/ab0301. arXiv: 1901.10046 [astro-ph.GA].
- Condon, J. J. et al. (May 1998a). "The NRAO VLA Sky Survey". In: *The Astronomical Journal* 115, pp. 1693–1716. issn: 0004-6256. doi: 10.1086/300337. url: <http://adsabs.harvard.edu/abs/1998AJ....115.1693C> (visited on 12/06/2020).
- (May 1998b). "The NRAO VLA Sky Survey". In: *AJ* 115, pp. 1693–1716. doi: 10.1086/300337.
- Condon, James J. and Scott M. Ransom (2016). "Essential Radio Astronomy". In: *Essential Radio Astronomy*, by James J. Condon and Scott M. Ransom. ISBN: 978-0-691-13779-7. Princeton, NJ: Princeton University Press, 2016. url: <http://adsabs.harvard.edu/abs/2016era..book.....C> (visited on 07/03/2020).
- Connor, Liam and Joeri van Leeuwen (Dec. 2018). "Applying Deep Learning to Fast Radio Burst Classification". In: *The Astronomical Journal* 156, p. 256. issn: 0004-6256. doi: 10.3847/1538-3881/aae649. url: <http://adsabs.harvard.edu/abs/2018AJ....156..256C> (visited on 12/08/2020).
- Cutri, RM et al. (2013). "Explanatory Supplement to the AllWISE Data Release Products". In: *Explanatory Supplement to the AllWISE Data Release Products*, by RM Cutri et al.

- Dabhade, P. et al. (Oct. 2020). "Search and analysis of giant radio galaxies with associated nuclei (SAGAN). I. New sample and multi-wavelength studies". In: *Astronomy and Astrophysics* 642, A153. issn: 0004-6361. doi: 10.1051/0004-6361/202038344. url: <http://adsabs.harvard.edu/abs/2020A<26A...642A.153D> (visited on 12/08/2020).
- Dabhade, Pratik et al. (Aug. 2017). "Discovery of giant radio galaxies from NVSS: radio and infrared properties". In: *Monthly Notices of the Royal Astronomical Society* 469, pp. 2886–2906. issn: 0035-8711. doi: 10.1093/mnras/stx860. url: <http://adsabs.harvard.edu/abs/2017MNRAS.469.2886D> (visited on 12/09/2020).
- Dawid, Alexander Philip and Allan M Skene (1979). "Maximum likelihood estimation of observer error-rates using the EM algorithm". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28.1, pp. 20–28.
- Deisenroth, Marc Peter, A. Aldo Faisal, and Cheng Soon Ong (Feb. 2020). *Mathematics for Machine Learning*. en. 1st ed. Cambridge University Press. isbn: 978-1-108-67993-0 978-1-108-47004-9 978-1-108-45514-5. doi: 10.1017/9781108679930. url: <https://www.cambridge.org/core/product/identifier/9781108679930/type/book> (visited on 07/07/2020).
- Dieleman, S., K. W. Willett, and J. Dambre (June 2015). "Rotation-invariant convolutional neural networks for galaxy morphology prediction". In: *MNRAS* 450, pp. 1441–1459. doi: 10.1093/mnras/stv632. arXiv: 1503.07077 [astro-ph.IM].
- Fan, Dongwei et al. (2015). "Matching radio catalogues with realistic geometry: application to SWIRE and ATLAS". In: *MNRAS* 451.2, pp. 1299–1305. doi: 10.1093/mnras/stv994. eprint: <http://mnras.oxfordjournals.org/content/451/2/1299.full.pdf+html>. url: <http://mnras.oxfordjournals.org/content/451/2/1299.abstract>.
- Fan, Dongwei et al. (Aug. 2020). "Optimal Probabilistic Catalogue Matching for Radio Sources". en. In: *Monthly Notices of the Royal Astronomical Society*. arXiv: 2008.05276, staa2447. issn: 0035-8711, 1365-2966. doi: 10.1093/mnras/staa2447. url: <http://arxiv.org/abs/2008.05276> (visited on 09/01/2020).
- Fanaroff, B. L. and J. M. Riley (May 1974a). "The morphology of extragalactic radio sources of high and low luminosity". In: *Monthly Notices of the Royal Astronomical Society* 167, 31P–36P. issn: 0035-8711. doi: 10.1093/mnras/167.1.31P. url: <http://adsabs.harvard.edu/abs/1974MNRAS.167P..31F> (visited on 07/03/2020).
- (May 1974b). "The morphology of extragalactic radio sources of high and low luminosity". In: *MNRAS* 167, 31P–36P. doi: 10.1093/mnras/167.1.31P.
- Franzen, T. M. O. et al. (Nov. 2015). "ATLAS - I. Third release of 1.4 GHz mosaics and component catalogues". In: *MNRAS* 453, pp. 4020–4036. doi: 10.1093/mnras/stv1866. arXiv: 1508.03150.
- Galvin, T. J. et al. (Oct. 2019a). "Radio Galaxy Zoo: Knowledge Transfer Using Rotationally Invariant Self-organizing Maps". en. In: *Publications of the Astronomical Society of the Pacific* 131.1004, p. 108009. issn: 0004-6280. doi: 10.1088/1538-3873/ab150b. url: <https://ui.adsabs.harvard.edu/abs/2019PASP..131j8009G/abstract> (visited on 11/18/2019).

- Galvin, T. J. et al. (Oct. 2019b). "Radio Galaxy Zoo: Knowledge Transfer Using Rotationally Invariant Self-organizing Maps". In: PASP 131.1004, p. 108009. doi: 10.1088/1538-3873/ab150b.
- Galvin, T. J. et al. (July 2020). "Cataloguing the radio-sky with unsupervised machine learning: a new approach for the SKA era". In: *Monthly Notices of the Royal Astronomical Society* 497, pp. 2730–2758. doi: 10.1093/mnras/staa1890. URL: <http://adsabs.harvard.edu/abs/2020MNRAS.497.2730G> (visited on 09/01/2020).
- Garofalo, David and Chandra B. Singh (Feb. 2019). "FR0 radio galaxies and their place in the radio morphology classification". en. In: *The Astrophysical Journal* 871.2. arXiv: 1811.05383, p. 259. issn: 1538-4357. doi: 10.3847/1538-4357/aaf056. URL: <http://arxiv.org/abs/1811.05383> (visited on 12/06/2020).
- Garon, Avery F. et al. (Mar. 2019a). "Radio Galaxy Zoo: The Distortion of Radio Galaxies by Galaxy Clusters". In: *The Astronomical Journal* 157, p. 126. issn: 0004-6256. doi: 10.3847/1538-3881/aaff62. URL: <http://adsabs.harvard.edu/abs/2019AJ...157..126G> (visited on 07/13/2020).
- (Mar. 2019b). "Radio Galaxy Zoo: The Distortion of Radio Galaxies by Galaxy Clusters". In: AJ 157.3, p. 126. doi: 10.3847/1538-3881/aaff62. URL: <https://doi.org/10.3847%2F1538-3881%2Faaff62>.
- Gendre, M. A. and J. V. Wall (Oct. 2008). "The Combined NVSS-FIRST Galaxies (CoNFIG) sample - I. Sample definition, classification and evolution". In: MNRAS 390, pp. 819–828. doi: 10.1111/j.1365-2966.2008.13792.x. arXiv: 0808.0165.
- Godfrey, L. E. H. and S. S. Shabala (Feb. 2016). "Mutual distance dependence drives the observed jet-power-radio-luminosity scaling relations in radio galaxies". In: MNRAS 456, pp. 1172–1184. doi: 10.1093/mnras/stv2712. arXiv: 1511.06007.
- Grant, J. K. et al. (May 2010). "The DRAO Planck Deep Fields: The Polarization Properties of Radio Galaxies at 1.4 GHz". In: ApJ 714, pp. 1689–1701. doi: 10.1088/0004-637X/714/2/1689. arXiv: 1003.4460.
- Grant, Julie K. (2011). "Polarised Radio Sources: A Study of Luminosity, Redshift and Flux Density". PhD thesis. University of Calgary, pp. 96–98.
- Guo, Ping et al. (Dec. 2019). "Pulsar candidate classification using generative adversary networks". en. In: *Monthly Notices of the Royal Astronomical Society* 490.4, pp. 5424–5439. issn: 0035-8711, 1365-2966. doi: 10.1093/mnras/stz2975. URL: <https://academic.oup.com/mnras/article/490/4/5424/5625511> (visited on 12/08/2020).
- Hardcastle, M. J. and J. H. Croston (Mar. 2020a). "Radio galaxies and feedback from AGN jets". en. In: *arXiv:2003.06137 [astro-ph]*. arXiv: 2003.06137. URL: <http://arxiv.org/abs/2003.06137> (visited on 03/30/2020).
- (Mar. 2020b). "Radio galaxies and feedback from AGN jets". In: *arXiv e-prints*, arXiv:2003.06137, arXiv:2003.06137. arXiv: 2003.06137 [*astro-ph.HE*].
- Hardcastle, M. J. and M. G. H. Krause (Mar. 2013). "Numerical modelling of the lobes of radio galaxies in cluster environments". In: MNRAS 430, pp. 174–196. doi: 10.1093/mnras/sts564. arXiv: 1301.2531.
- (Sept. 2014). "Numerical modelling of the lobes of radio galaxies in cluster environments - II. Magnetic field configuration and observability". In: MNRAS 443, pp. 1482–1499. doi: 10.1093/mnras/stu1229. arXiv: 1406.5300 [*astro-ph.HE*].

- Hardcastle, M. J. et al. (Feb. 2019). “Radio-loud AGN in the first LoTSS data release: The lifetimes and environmental impact of jet-driven sources”. In: *A&A* 622, A12. issn: 0004-6361, 1432-0746. doi: 10.1051/0004-6361/201833893.
- He, Kaiming et al. (2016). “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- Heckman, Timothy M. and Philip N. Best (Aug. 2014). “The Coevolution of Galaxies and Supermassive Black Holes: Insights from Surveys of the Contemporary Universe”. en. In: *Annual Review of Astronomy and Astrophysics* 52.1, pp. 589–660. issn: 0066-4146, 1545-4282. doi: 10.1146/annurev-astro-081913-035722. url: <http://www.annualreviews.org/doi/10.1146/annurev-astro-081913-035722> (visited on 07/03/2020).
- Helfand, David J., Richard L. White, and Robert H. Becker (Feb. 2015). “THE LAST OFFIRST: THE FINAL CATALOG AND SOURCE IDENTIFICATIONS”. In: *The Astrophysical Journal* 801.1, p. 26. doi: 10.1088/0004-637x/801/1/26. url: <https://doi.org/10.1088%2F0004-637x%2F801%2F1%2F26>.
- Hinshaw, G. et al. (Oct. 2013). “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results”. In: *ApJS* 208, 19, p. 19. doi: 10.1088/0067-0049/208/2/19. arXiv: 1212.5226.
- Jarrett, T. H. et al. (2011). “The Spitzer-WISE Survey of the Ecliptic Poles”. In: *The Astrophysical Journal* 735.2, p. 112.
- Jarrett, T. H. et al. (2017). “Galaxy and Mass Assembly (GAMA): Exploring the WISE Web in G12”. In: *The Astrophysical Journal* 836.2, p. 182.
- Johnston, S. et al. (Dec. 2007). “Science with the Australian Square Kilometre Array Pathfinder”. In: *PASA* 24, pp. 174–188. doi: 10.1071/AS07033. arXiv: 0711.2103.
- Kimball, A. E. and Ž. Ivezić (Aug. 2008). “A Unified Catalog of Radio Objects Detected by NVSS, First, WENSS, GB6, and SDSS”. In: *AJ* 136, pp. 684–712. doi: 10.1088/0004-6256/136/2/684. arXiv: 0806.0650.
- Kingma, Diederik P. and Jimmy Ba (2014). “Adam: A Method for Stochastic Optimization”. In: *CoRR* abs/1412.6980. arXiv: 1412.6980. url: <http://arxiv.org/abs/1412.6980>.
- Kochanek, C. S. et al. (Oct. 2001). “The K-Band Galaxy Luminosity Function”. In: *ApJ* 560, pp. 566–579. doi: 10.1086/322488. eprint: astro-ph/0011456.
- Kozieł-Wierzbowska, Dorota et al. (Dec. 2020). “Identifying radio active galactic nuclei among radio-emitting galaxies”. In: *arXiv e-prints* 2012, arXiv:2012.00367. url: <http://adsabs.harvard.edu/abs/2020arXiv201200367K> (visited on 12/08/2020).
- Lacy, M. et al. (Mar. 2020). “The Karl G. Jansky Very Large Array Sky Survey (VLASS). Science Case and Survey Design”. In: *PASP* 132.1009, p. 035001. doi: 10.1088/1538-3873/ab63eb. arXiv: 1907.01981 [astro-ph.IM].
- Laing, Robert A. and Alan H. Bridle (Oct. 1987). “Rotation measure variation across M84”. In: *Monthly Notices of the Royal Astronomical Society* 228, pp. 557–571. issn: 0035-8711. doi: 10.1093/mnras/228.3.557. url: <http://adsabs.harvard.edu/abs/1987MNRAS.228..557L> (visited on 07/13/2020).

- Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell (2017). "Simple and scalable predictive uncertainty estimation using deep ensembles". In: *Advances in Neural Information Processing Systems*, pp. 6402–6413.
- Leahy, J. P., A. H. Bridle, and R. G. Strom (2020). *An Atlas of DRAGNs*. URL: <http://www.jb.man.ac.uk/atlas/index.html> (visited on 06/13/2020).
- Leahy, J. P. and R. A. Perley (Aug. 1991). "VLA images of 23 extragalactic radio sources". In: *The Astronomical Journal* 102, pp. 537–561. issn: 0004-6256. doi: 10.1086/115892. URL: <http://adsabs.harvard.edu/abs/1991AJ....102..537L> (visited on 07/13/2020).
- Leahy, J. P., G. G. Pooley, and J. M. Riley (Oct. 1986). "The polarization of classical double-radio sources". In: *Monthly Notices of the Royal Astronomical Society* 222, pp. 753–785. issn: 0035-8711. doi: 10.1093/mnras/222.4.753. URL: <http://adsabs.harvard.edu/abs/1986MNRAS.222..753L> (visited on 07/13/2020).
- LeCun, Y. et al. (Dec. 1989). "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation* 1.4. Publisher: MIT Press, pp. 541–551. issn: 0899-7667. doi: 10.1162/neco.1989.1.4.541. URL: <https://doi.org/10.1162/neco.1989.1.4.541> (visited on 12/06/2020).
- LeCun, Yann et al. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.
- Ledlow, Michael J. and Frazer N. Owen (July 1996). "20 CM VLA Survey of Abell Clusters of Galaxies. VI. Radio/Optical Luminosity Functions". In: *The Astronomical Journal* 112, p. 9. doi: 10.1086/117985. URL: <http://adsabs.harvard.edu/abs/1996AJ....112....9L> (visited on 11/10/2020).
- Lin, Haitao, Xiangru Li, and Qingguo Zeng (Aug. 2020). "Pulsar Candidate Sifting Using Multi-input Convolution Neural Networks". In: *The Astrophysical Journal* 899, p. 104. issn: 0004-637X. doi: 10.3847/1538-4357/aba838. URL: <http://adsabs.harvard.edu/abs/2020ApJ...899..104L> (visited on 12/08/2020).
- Lintott, C. J. et al. (Sept. 2008). "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey". In: *MNRAS* 389, pp. 1179–1189. doi: 10.1111/j.1365-2966.2008.13689.x. arXiv: 0804.4483.
- Lonsdale, C. J. et al. (Aug. 2003). "SWIRE: The SIRTF Wide-Area Infrared Extragalactic Survey". In: *PASP* 115, pp. 897–927. doi: 10.1086/376850. eprint: astro-ph/0305375.
- Lovelace, R. V. E. (Aug. 1976). "Dynamo model of double radio sources". en. In: *Nature* 262.5570. Number: 5570 Publisher: Nature Publishing Group, pp. 649–652. issn: 1476-4687. doi: 10.1038/262649a0. URL: <https://www.nature.com/articles/262649a0> (visited on 12/06/2020).
- Lukic, V. et al. (May 2018). "Radio Galaxy Zoo: compact and extended radio source classification with deep learning". In: *MNRAS* 476, pp. 246–260. doi: 10.1093/mnras/sty163. arXiv: 1801.04861 [astro-ph.IM].
- Lukic, V. et al. (Aug. 2019). "Morphological classification of radio galaxies: capsule networks versus convolutional neural networks". In: *Monthly Notices of the Royal Astronomical Society* 487, pp. 1729–1744. issn: 0035-8711. doi: 10.1093/mnras/

- stz1289. URL: <http://adsabs.harvard.edu/abs/2019MNRAS.487.1729L> (visited on 12/08/2020).
- Ma, Yik Ki et al. (Aug. 2019a). “A broad-band spectro-polarimetric view of the NVSS rotation measure catalogue - I. Breaking the π -ambiguity”. en. In: *Monthly Notices of the Royal Astronomical Society* 487.3, pp. 3432–3453. issn: 0035-8711. doi: 10.1093/mnras/stz1325. URL: <https://ui.adsabs.harvard.edu/abs/2019MNRAS.487.3432M/abstract> (visited on 07/17/2020).
- Ma, Zhixian et al. (June 2018). “Radio Galaxy Morphology Generation Using DNN Autoencoder and Gaussian Mixture Models”. In: *arXiv e-prints* 1806, arXiv:1806.00398. URL: <http://adsabs.harvard.edu/abs/2018arXiv180600398M> (visited on 12/08/2020).
- Ma, Zhixian et al. (Feb. 2019b). “A Machine Learning Based Morphological Classification of 14,245 Radio AGNs Selected from the Best-Heckman Sample”. en. In: *The Astrophysical Journal Supplement Series* 240.2. Publisher: American Astronomical Society, p. 34. issn: 0067-0049. doi: 10.3847/1538-4365/aaf9a2. URL: <https://doi.org/10.3847%2F1538-4365%2Faaf9a2> (visited on 12/06/2020).
- Machalski, J. et al. (Sept. 2011). “UNDERSTANDING GIANT RADIO GALAXY J1420-0545: LARGE-SCALE MORPHOLOGY, ENVIRONMENT, AND ENERGETICS”. en. In: *The Astrophysical Journal* 740.2, p. 58. issn: 0004-637X. doi: 10.1088/0004-637X/740/2/58. URL: <https://iopscience.iop.org/article/10.1088/0004-637X/740/2/58> (visited on 07/03/2020).
- Marshall, Philip J., Chris J. Lintott, and Leigh N. Fletcher (2015a). “Ideas for Citizen Science in Astronomy”. In: *Annual Review of Astronomy and Astrophysics* 53, pp. 247–278. doi: 10.1146/annurev-astro-081913-035959.
- (2015b). “Ideas for Citizen Science in Astronomy”. In: *Annual Review of Astronomy and Astrophysics* 53, pp. 247–278. doi: 10.1146/annurev-astro-081913-035959.
- Mauch, T. and E. M. Sadler (Mar. 2007a). “Radio sources in the 6dFGS: local luminosity functions at 1.4GHz for star-forming galaxies and radio-loud AGN”. In: *MNRAS* 375, pp. 931–950. doi: 10.1111/j.1365-2966.2006.11353.x. eprint: astro-ph/0612018.
- Mauch, Tom and Elaine M. Sadler (Mar. 2007b). “Radio sources in the 6dFGS: local luminosity functions at 1.4 GHz for star-forming galaxies and radio-loud AGN”. en. In: *Monthly Notices of the Royal Astronomical Society* 375.3, pp. 931–950. issn: 0035-8711. doi: 10.1111/j.1365-2966.2006.11353.x. URL: <https://academic.oup.com/mnras/article/375/3/931/1138979> (visited on 11/07/2019).
- Menon, Aditya Krishna et al. (2015). “Learning from Corrupted Binary Labels via Class-probability Estimation”. In: *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*. ICML’15. Lille, France: JMLR.org, pp. 125–134. URL: <http://dl.acm.org/citation.cfm?id=3045118.3045133>.
- Middleberg, E. et al. (Apr. 2008). “Deep Australia Telescope Large Area Survey Radio Observations of the European Large Area ISO Survey S1/Spitzer Wide-Area Infrared Extragalactic Field”. In: *AJ* 135, pp. 1276–1290. doi: 10.1088/0004-6256/135/4/1276.
- Mingo, B. et al. (Sept. 2019). “Revisiting the Fanaroff-Riley dichotomy and radio-galaxy morphology with the LOFAR Two-Metre Sky Survey (LoTSS)”. en. In:

- Monthly Notices of the Royal Astronomical Society* 488.2. Publisher: Oxford Academic, pp. 2701–2721. ISSN: 0035-8711. doi: 10.1093/mnras/stz1901. URL: <https://academic.oup.com/mnras/article/488/2/2701/5530785> (visited on 11/10/2020).
- Mukherjee, Dipanjan et al. (Sept. 2016). “Relativistic jet feedback in high-redshift galaxies - I. Dynamics”. In: *Monthly Notices of the Royal Astronomical Society* 461, pp. 967–983. doi: 10.1093/mnras/stw1368. URL: <http://adsabs.harvard.edu/abs/2016MNRAS.461..967M> (visited on 11/09/2020).
- Mukherjee, Dipanjan et al. (Oct. 2018). “Relativistic jet feedback - III. Feedback on gas discs”. In: *Monthly Notices of the Royal Astronomical Society* 479, pp. 5544–5566. doi: 10.1093/mnras/sty1776. URL: <http://adsabs.harvard.edu/abs/2018MNRAS.479.5544M> (visited on 11/09/2020).
- Norris, R. P. (Jan. 2017a). “Discovering the Unexpected in Astronomical Survey Data”. In: PASA 34, e007, e007. doi: 10.1017/pasa.2016.63. arXiv: 1611.05570 [astro-ph.IM].
- (Oct. 2017b). “Extragalactic radio continuum surveys and the transformation of radio astronomy”. In: *Nature Astronomy* 1, pp. 671–678. doi: 10.1038/s41550-017-0233-y. arXiv: 1709.05064.
- Norris, R. P. et al. (Dec. 2006). “Deep ATLAS Radio Observations of the Chandra Deep Field-South/Spitzer Wide-Area Infrared Extragalactic Field”. In: AJ 132, pp. 2409–2423. doi: 10.1086/508275. eprint: astro-ph/0610538.
- Norris, R. P. et al. (Aug. 2011a). “EMU: Evolutionary Map of the Universe”. In: PASA 28, pp. 215–248. doi: 10.1071/AS11021. arXiv: 1106.3219.
- Norris, Ray P. et al. (Aug. 2011b). “EMU: Evolutionary Map of the Universe”. In: *Publications of the Astronomical Society of Australia* 28, pp. 215–248. ISSN: 1323-3580. doi: 10.1071/AS11021. URL: <http://adsabs.harvard.edu/abs/2011PASA...28..215N> (visited on 07/12/2020).
- O’Sullivan, S. P. et al. (June 2015). “THE MAGNETIC FIELD AND POLARIZATION PROPERTIES OF RADIO GALAXIES IN DIFFERENT ACCRETION STATES”. en. In: *The Astrophysical Journal* 806.1. Publisher: American Astronomical Society, p. 83. ISSN: 0004-637X. doi: 10.1088/0004-637X/806/1/83. URL: <https://doi.org/10.1088%2F0004-637X%2F806%2F1%2F83> (visited on 12/06/2020).
- O’Sullivan, S. P. et al. (Aug. 2017). “Broad-band, radio spectro-polarimetric study of 100 radiative-mode and jet-mode AGN”. en. In: *Monthly Notices of the Royal Astronomical Society* 469, p. 4034. ISSN: ['0035-8711']. doi: 10.1093/mnras/stx1133. URL: <https://ui.adsabs.harvard.edu/abs/2017MNRAS.469.4034O/abstract> (visited on 11/14/2019).
- Owens, E. A., R. E. Griffiths, and K. U. Ratnatunga (July 1996). “Using oblique decision trees for the morphological classification of galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 281. ISSN: 0035-8711. doi: 10.1093/mnras/281.1.153. URL: <http://adsabs.harvard.edu/abs/1996MNRAS.281..153O> (visited on 12/06/2020).
- Paszke, Adam et al. (2017). “Automatic differentiation in PyTorch”. In: *NIPS 2017 Workshop Autodiff*.

- Pedregosa, Fabian et al. (Nov. 2011). "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12, pp. 2825–2830. ISSN: 1532-4435. URL: <http://dl.acm.org/citation.cfm?id=1953048.2078195>.
- Polsterer, K. L., F. Gieseke, and C. Igel (Sept. 2015). "Automatic Galaxy Classification via Machine Learning Techniques: Parallelized Rotation/Flipping INvariant Kohonen Maps (PINK)". In: *Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV)*. Ed. by A. R. Taylor and E. Rosolowsky. Vol. 495. Astronomical Society of the Pacific Conference Series, p. 81.
- Pracy, M. B. et al. (July 2016). "GAMA/WiggleZ: the 1.4 GHz radio luminosity functions of high- and low-excitation radio galaxies and their redshift evolution to $z = 0.75$ ". In: *MNRAS* 460, pp. 2–17. doi: 10.1093/mnras/stw910. arXiv: 1604.04332.
- Proctor, D. D. (July 2006). "Comparing Pattern Recognition Feature Sets for Sorting Triples in the FIRST Database". In: *ApJS* 165, pp. 95–107. doi: 10.1086/504801. eprint: astro-ph/0605104.
- (June 2011). "Morphological Annotations for Groups in the First Database". In: *The Astrophysical Journal Supplement Series* 194, p. 31. ISSN: 0067-0049. doi: 10.1088/0067-0049/194/2/31. URL: <http://adsabs.harvard.edu/abs/2011ApJS..194...31P> (visited on 12/08/2020).
- (June 2016). "A Selection of Giant Radio Sources from NVSS". In: *The Astrophysical Journal Supplement Series* 224, p. 18. ISSN: 0067-0049. doi: 10.3847/0067-0049/224/2/18. URL: <http://adsabs.harvard.edu/abs/2016ApJS..224...18P> (visited on 12/15/2020).
- Ralph, Nicholas O. et al. (Oct. 2019). "Radio Galaxy Zoo: Unsupervised Clustering of Convolutionally Auto-encoded Radio-astronomical Images". In: *PASP* 131.1004, p. 108011. doi: 10.1088/1538-3873/ab213d. arXiv: 1906.02864 [astro-ph.IM].
- Raouf, M. et al. (Oct. 2017). "The many lives of active galactic nuclei-II: The formation and evolution of radio jets and their impact on galaxy evolution". In: *MNRAS* 471, pp. 658–670. doi: 10.1093/mnras/stx1598. arXiv: 1706.06595.
- Rayner, D. P., R. P. Norris, and R. J. Sault (Dec. 2000). "Radio circular polarization of active galaxies". en. In: *Monthly Notices of the Royal Astronomical Society* 319.2. Publisher: Oxford Academic, pp. 484–496. ISSN: 0035-8711. doi: 10.1111/j.1365-8711.2000.03854.x. URL: <https://academic.oup.com/mnras/article/319/2/484/1295073> (visited on 11/10/2020).
- Richter, G. A. (1975). "Search for optical identifications in the 5C3-radio survey. II - Statistical treatment and results". In: *Astronomische Nachrichten* 296, pp. 65–81. doi: 10.1002/asna.19752960203.
- Rodman, P. E. et al. (Feb. 2019a). "Radio Galaxy Zoo: observational evidence for environment as the cause of radio source asymmetry". In: *Monthly Notices of the Royal Astronomical Society* 482, pp. 5625–5641. ISSN: 0035-8711. doi: 10.1093/mnras/sty3070. URL: <http://adsabs.harvard.edu/abs/2019MNRAS.482.5625R> (visited on 07/13/2020).
- (Feb. 2019b). "Radio Galaxy Zoo: observational evidence for environment as the cause of radio source asymmetry". In: *MNRAS* 482, pp. 5625–5641. doi: 10.1093/mnras/sty3070. arXiv: 1811.03726.

- Rowley, Henry A, Shumeet Baluja, and Takeo Kanade (1996). "Human face detection in visual scenes". In: *Advances in Neural Information Processing Systems*, pp. 875–881.
- Rybicki, George B and Alan P Lightman (2008). *Radiative processes in astrophysics*. John Wiley & Sons.
- (1979). "Radiative processes in astrophysics". In: *New York, Wiley-Interscience, 1979*. 393 p. URL: <http://adsabs.harvard.edu/abs/1979rpa..book.....R> (visited on 07/03/2020).
- Sadler, E. M. et al. (Feb. 2014). "The local radio-galaxy population at 20 GHz". In: *MNRAS* 438, pp. 796–824. doi: 10.1093/mnras/stt2239. arXiv: 1304.0268.
- Sadler, Elaine M. et al. (2002). "Radio sources in the 2dF Galaxy Redshift Survey – II. Local radio luminosity functions for AGN and star-forming galaxies at 1.4 GHz". In: *MNRAS* 329.1, pp. 227–245. doi: 10.1046/j.1365-8711.2002.04998.x. eprint: /oup/backfile/content_public/journal/mnras/329/1/10.1046/j.1365-8711.2002.04998.x/2/329-1-227.pdf. URL: <http://dx.doi.org/10.1046/j.1365-8711.2002.04998.x>.
- Saikia, D. J. and C. J. Salter (1988). "Polarization properties of extragalactic radio sources". In: *Annual Review of Astronomy and Astrophysics* 26, pp. 93–144. doi: 10.1146/annurev.aa.26.090188.000521. URL: <http://adsabs.harvard.edu/abs/1988ARA%26A..26...93S> (visited on 11/10/2020).
- Sakelliou, Irini, M. J. Hardcastle, and N. N. Jetha (Feb. 2008). "3C40 in Abell194: can tail radio galaxies exist in a quiescent cluster?" In: *Monthly Notices of the Royal Astronomical Society* 384, pp. 87–93. doi: 10.1111/j.1365-2966.2007.12465.x. URL: <http://adsabs.harvard.edu/abs/2008MNRAS.384...87S> (visited on 11/09/2020).
- Samudre, Ashwin et al. (Nov. 2020). "Data-Efficient Classification of Radio Galaxies". In: *arXiv e-prints* 2011, arXiv:2011.13311. URL: <http://adsabs.harvard.edu/abs/2020arXiv201113311S> (visited on 12/08/2020).
- Schmidt, M. (Feb. 1968). "Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources". In: *ApJ* 151, p. 393. doi: 10.1086/149446.
- Segal, Gary et al. (Sept. 2019). "Identifying Complex Sources in Large Astronomical Data Sets Using a Coarse-grained Complexity Measure". en. In: *Publications of the Astronomical Society of the Pacific* 131.1004, p. 108007. issn: 1538-3873. doi: 10.1088/1538-3873/ab0068. URL: <https://doi.org/10.1088%2F1538-3873%2Fab0068> (visited on 09/24/2019).
- Shabala, S. S. (Aug. 2018). "The role of environment in the observed Fundamental Plane of radio active galactic nuclei". In: *MNRAS* 478, pp. 5074–5080. doi: 10.1093/mnras/sty1328. arXiv: 1805.06600.
- Shabala, S. S. and L. E. H. Godfrey (May 2013). "SIZE DEPENDENCE OF THE RADIO-LUMINOSITY-MECHANICAL-POWER CORRELATION IN RADIO GALAXIES". In: *The Astrophysical Journal* 769.2, p. 129. issn: 0004-637X, 1538-4357. doi: 10.1088/0004-637X/769/2/129.
- Shimwell, T. W. 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].

- Sikora, Marek and Mitchell C. Begelman (Feb. 2013). "Magnetic Flux Paradigm for Radio Loudness of Active Galactic Nuclei". In: *The Astrophysical Journal Letters* 764, p. L24. doi: 10.1088/2041-8205/764/2/L24. URL: <http://adsabs.harvard.edu/abs/2013ApJ...764L..24S> (visited on 11/10/2020).
- Storrie-Lombardi, M. C. et al. (Nov. 1992). "Morphological Classification of Galaxies by Artificial Neural Networks". en. In: *Monthly Notices of the Royal Astronomical Society* 259, 8P. ISSN: 0035-8711. doi: 10.1093/mnras/259.1.8P. URL: <https://ui.adsabs.harvard.edu/abs/1992MNRAS.259P...8S/abstract> (visited on 12/06/2020).
- Strauss, M. A. et al. (Sept. 2002). "Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample". In: AJ 124, pp. 1810–1824. doi: 10.1086/342343. eprint: astro-ph/0206225.
- Surace, JA et al. (2005). "The SWIRE Data Release 2: Image Atlases and Source Catalogs for ELAIS-N1, ELAIS-N2, XMM-LSS, and the Lockman Hole". In: *Spitzer Science Centre, California Institute of Technology, Pasadena, CA*.
- Tang, H., A. M. M. Scaife, and J. P. Leahy (Sept. 2019). "Transfer learning for radio galaxy classification". en. In: *Monthly Notices of the Royal Astronomical Society* 488.3. Publisher: Oxford Academic, pp. 3358–3375. ISSN: 0035-8711. doi: 10.1093/mnras/stz1883. URL: <https://academic.oup.com/mnras/article/488/3/3358/5538844> (visited on 12/08/2020).
- Taylor, A. R., J. M. Stil, and C. Sunstrum (Aug. 2009). "A ROTATION MEASURE IMAGE OF THE SKY". en. In: *The Astrophysical Journal* 702.2, pp. 1230–1236. ISSN: 0004-637X. doi: 10.1088/0004-637X/702/2/1230. URL: <https://doi.org/10.1088/0004-637X/702/2/1230> (visited on 11/05/2019).
- Taylor, A. R. et al. (Sept. 2007). "Radio Polarimetry of the ELAIS N1 Field: Polarized Compact Sources". In: ApJ 666, pp. 201–211. doi: 10.1086/519786. arXiv: 0705.2736.
- The Astropy Collaboration et al. (Jan. 2018). "The Astropy Project: Building an inclusive, open-science project and status of the v2.0 core package". In: *ArXiv e-prints*. arXiv: 1801.02634 [astro-ph.IM].
- Turner, Ross J. and Stanislav S. Shabala (June 2015). "Energetics and Lifetimes of Local Radio Active Galactic Nuclei". In: ApJ 806, 59, p. 59. doi: 10.1088/0004-637X/806/1/59. arXiv: 1504.05204 [astro-ph.GA].
- Urry, C. Megan and Paolo Padovani (Sept. 1995). "Unified Schemes for Radio-Loud Active Galactic Nuclei". In: *Publications of the Astronomical Society of the Pacific* 107, p. 803. ISSN: 0004-6280. doi: 10.1086/133630. URL: <http://adsabs.harvard.edu/abs/1995PASP..107..803U> (visited on 07/03/2020).
- van Velzen, Sjoert et al. (Aug. 2012). "Radio galaxies of the local universe. All-sky catalog, luminosity functions, and clustering". In: A&A 544, A18, A18. doi: 10.1051/0004-6361/201219389. arXiv: 1206.0031 [astro-ph.CO].
- Verheijen, M. A. W. et al. (2008). "Apertif, a focal plane array for the WSRT". In: *AIP Conf. Ser. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window*. Vol. 1035, pp. 265–271.
- Wang, HongFeng et al. (May 2019). "Pulsar candidate selection using ensemble networks for FAST drift-scan survey". In: *Science China Physics, Mechanics, and Astronomy*

- omy 62, p. 959507. doi: 10.1007/s11433-018-9388-3. URL: <http://adsabs.harvard.edu/abs/2019SCPMA..6259507W> (visited on 12/08/2020).
- Weston, S. D. et al. (Feb. 2018). "Automated cross-identifying radio to infrared surveys using the LRPY algorithm: a case study". In: MNRAS 473, pp. 4523–4537. doi: 10.1093/mnras/stx2562. arXiv: 1710.01449.
- White, R. L. et al. (Feb. 1997). "A Catalog of 1.4 GHz Radio Sources from the FIRST Survey". In: ApJ 475, pp. 479–493. doi: 10.1086/303564.
- White, Sarah V. et al. (2020). "The GLEAM 4-Jy (G4Jy) Sample: II. Host galaxy identification for individual sources". In: *Publications of the Astronomical Society of Australia* 37, e017. issn: 1323-3580. doi: 10.1017/pasa.2020.10. URL: <http://adsabs.harvard.edu/abs/2020PASA...37...17W> (visited on 07/14/2020).
- Willott, C. J. et al. (Nov. 1999). "The emission line-radio correlation for radio sources using the 7C Redshift Survey". In: MNRAS 309, pp. 1017–1033. doi: 10.1046/j.1365-8711.1999.02907.x. eprint: astro-ph/9905388.
- Wright, E. L. et al. (Dec. 2010). "The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance". In: AJ 140, 1868–1881, pp. 1868–1881. doi: 10.1088/0004-6256/140/6/1868. arXiv: 1008.0031 [astro-ph.IM].
- Wu, Chen et al. (Jan. 2019). "Radio Galaxy Zoo: CLARAN - a deep learning classifier for radio morphologies". In: MNRAS 482, pp. 1211–1230. doi: 10.1093/mnras/sty2646. arXiv: 1805.12008 [astro-ph.IM].
- Zhang, C. et al. (Oct. 2020). "Applying saliency-map analysis in searches for pulsars and fast radio bursts". In: *Astronomy and Astrophysics* 642, A26. issn: 0004-6361. doi: 10.1051/0004-6361/201937234. URL: <http://adsabs.harvard.edu/abs/2020A%26A...642A..26Z> (visited on 12/08/2020).
- Zhu, Xiao-Pan et al. (Apr. 2019). "Galaxy morphology classification with deep convolutional neural networks". In: *Astrophysics and Space Science* 364, p. 55. issn: 0004-640X. doi: 10.1007/s10509-019-3540-1. URL: <http://adsabs.harvard.edu/abs/2019Ap%26SS.364..55Z> (visited on 12/08/2020).
- Zhuang, Ming-Yang and Luis C. Ho (June 2020). "The Interplay between Star Formation and Black Hole Accretion in Nearby Active Galaxies". In: *The Astrophysical Journal* 896, p. 108. doi: 10.3847/1538-4357/ab8f2e. URL: <http://adsabs.harvard.edu/abs/2020ApJ...896..108Z> (visited on 11/08/2020).
- Żywucka, Natalia, Dorota Kozieł-Wierzbowska, and Arti Goyal (Mar. 2020). "Catalogue with visual morphological classification of 32,616 radio galaxies with optical hosts". In: *arXiv e-prints* 2003, arXiv:2003.10322. URL: <http://adsabs.harvard.edu/abs/2020arXiv200310322Z> (visited on 12/08/2020).