

1

Learning to Identify Extragalactic 2 Radio Sources

3

4

Matthew J. Alger

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



Australian
National
University

8

9

September 2021

¹⁰
¹¹

© Matthew J. Alger 2021
All rights reserved

¹²

REVISION DIFF

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

- 17 • *Radio Galaxy Zoo: Machine learning for radio source host galaxy cross-identification*,
18 by **M. J. Alger**, J. K. Banfield, C. S. Ong, L. Rudnick, O. I. Wong, C. Wolf, H.
19 Andernach, R. P. Norris, and S. S. Shabala. Published in 2018 in the *Monthly*
20 *Notices of the Royal Astronomical Society*. Chapter 4 in this thesis.
- 21 • *Radio Galaxy Zoo: Radio luminosity functions of extended sources*, by **M. J. Alger**,
22 O. I. Wong, C. S. Ong, N. M. McClure-Griffiths, H. Andernach, L. Rudnick, S. S.
23 Shabala, A. F. Garon, J. K. Banfield, A. D. Kapińska, R. P. Norris, and A. J. M.
24 Thomson. Not yet submitted. Chapter 5 in this thesis.
- 25 • *Interpretable Faraday Complexity Classification*, by **M. J. Alger**, J. D. Livingston, N.
26 M. McClure-Griffiths, J. L. Nabaglo., O. I. Wong, and C. S. Ong. Accepted; to be
27 published in *Publications of the Astronomical Society of Australia*. Chapter 6 in this
28 thesis.

29 For all three papers I wrote the entirety of the content, produced all figures ex-
30 cept where noted, and conducted all experiments. Other authors reviewed the papers,
31 made suggestions, discussed ideas, and vitally contributed to the Radio Galaxy Zoo
32 project.

33
34

Matthew J. Alger
September 20, 2021

REVISION DIFF

REVISION DIFF

To Shirley and Bob.

REVISION DIFF

³⁶ Acknowledgments

³⁷ I could not have completed this thesis without my incredible supervisors: Naomi
³⁸ McClure-Griffiths, who stepped into my panel when circumstances required it and
³⁹ made me feel just as welcome in her group as any student who had been there from
⁴⁰ the beginning; Ivy Wong, who was always available to help answer my many inane
⁴¹ questions on the depths of radio astronomy; Julie Banfield, who helped me learn the
⁴² ropes of astronomy and pointed out all of the important websites to me; and Cheng
⁴³ Soon Ong, who advised when I needed advising, pushed when I needed pushing, and
⁴⁴ was always there when I needed support.

⁴⁵ Much of this work was done in my office at Mount Stromlo while I bothered my
⁴⁶ office-mates: Eloise Birchall, Adam Rains, Lara Cullinane, and for a time Rajika Ku-
⁴⁷ ruwita and Alex Wallace. Without their welcome distractions I would not have been
⁴⁸ able to concentrate. Of course, Mount Stromlo was more than just my office, and I
⁴⁹ also must thank Henry Zovaro, Phil Taylor, Karlie Noon, Adam Thomas, Jack Liv-
⁵⁰ ington, Fiona Panther, and Alec Thomson for their contributions to the atmosphere
⁵¹ of the observatory while I was there, as well as Howard Coyle and Michelle Cicolini for
⁵² keeping the observatory running smoothly. I would also like to thank my colleagues
⁵³ at Data61: Cody Christopher, Steve Edwards, and Angus Gruen, who provided an
⁵⁴ excellent sounding board to my many astronomical problems.

⁵⁵ Chapters 4 and 5 rely heavily on results from the Radio Galaxy Zoo citizen science
⁵⁶ project. I would like to thank all 10 000+ volunteers who contributed to the project,
⁵⁷ as well as the members of the Radio Galaxy Zoo collaboration, notably Stas Shabala,
⁵⁸ Larry Rudnick, Heinz Andernach, Avery Garon, Anna Kapinska, and Ray Norris, all of
⁵⁹ whom contributed to discussion of my work for these chapters. I must also thank those
⁶⁰ who provided me funding over this PhD: the Australian Government, the Research
⁶¹ School of Astronomy and Astrophysics, Data61/CSIRO, and the Astronomical Society
⁶² of Australia.

⁶³ A special thank you to those who proof-read this thesis: Rachel Alger, Jakub Nabaglo,
⁶⁴ Henry Zovaro, Jack Livingston, Sarah Lawson, Eloise Birchall, and Adam Rains.

⁶⁵ Finally, I would like to thank my family, Rachel, Geoff, and Kelly Alger; as well as
⁶⁶ my partner Sarah Lawson; and my friends Nicholas Hogan, Mitchell Busby, Bronnagh
⁶⁷ Norris, and Jakub Nabaglo; for their emotional support through my PhD.

68 The work and writing behind this thesis was completed in Australia on the land
69 of the Ngunnawal and Ngambri people. I would like to acknowledge and celebrate
70 the traditional custodians of this land, and pay my respects to their Elders past and
71 present.

REVISION DRAFT

72 Abstract

73 Radio observations of actively accreting supermassive black holes outside of the galaxy
74 can provide insight into the history of galaxies and their evolution. With the construc-
75 tion of fast new radio telescopes and the undertaking of large new radio surveys in the
76 lead-up to the Square Kilometre Array (SKA), radio astronomy faces a ‘data deluge’
77 where traditional methods of data analysis cannot keep up with the scale of the data.
78 Astronomers are increasingly looking to machine learning to provide ways of handling
79 large-scale data like these. This thesis introduces machine learning methods for use
80 in wide-area radio surveys and demonstrate demonstrates their application to radio
81 astronomy data. To help understand the issues facing large-scale wide-area radio sur-
82 veys, and contribute toward their solutions, we consider the problems of automated
83 radio-infrared cross-identification and Faraday complexity classification.

84 We developed an automated machine learning method for cross-identifying radio
85 objects with their infrared counterparts, training the algorithm with data from the cit-
86 izen science project Radio Galaxy Zoo. The trained result performed comparably to
87 an algorithm trained on expert cross-identifications, demonstrating the benefit of non-
88 expert labelling in radio astronomy. By examining the theoretical maximum accuracy
89 of this algorithm we showed that existing pilot studies for future surveys were not
90 sufficiently large enough to train machine learning methods. We showed the utility of
91 our cross-identification algorithm by applying it instead to a large survey, Faint Images
92 of the Radio Sky at Twenty Centimeters (FIRST), producing the largest catalogue of
93 cross-identified extended sources available at the time of writing. From this catalogue,
94 we calculated a mid-infrared-divided fractional radio luminosity function as well as
95 an estimate of energy injected into the intergalactic medium by active galactic nuclei
96 jets—one of the first applications of machine learning to radio astronomy to obtain a
97 physics result. A key result from this work was that the limitation in our sample size
98 was not due to the number of radio objects cross-identified but rather by the number
99 of available redshift measurements. Finally, we developed interpretable features for
100 spectropolarimetric measurements of radio sources and used these features to design
101 a machine learning algorithm that can identify Faraday complexity, while the features
102 themselves may be used for other tasks. The methods in this thesis will be applicable
103 to future radio surveys such as the Evolutionary Map of the Universe (EMU) contin-
104 uum survey and the Polarised Sky Survey of the Universe’s Magnetism (POSSUM),
105 as well as surveys produced with the SKA, allowing the development of higher reso-
106 lution radio luminosity functions, better estimates of the impact of radio galaxies on
107 their environments, faster analysis of polarised surveys, and better quality rotation
108 measure grids.

REVISION DIFF

Contents

109	Acknowledgments	vii
110	Abstract	ix
111	List of Figures	xv
112	List of Tables	xvii
113	List of Constants	xix
114	List of Abbreviations	xxi
115		
116	1 Introduction	1
117	1.1 Problems in extragalactic radio astronomy	2
118	1.2 Big data in astronomy	3
119	1.3 Machine learning in astronomy	4
120	1.4 How this thesis fits in	5
121	1.5 Thesis outline	6
122	1.6 Contributions	6
123	1.7 Works produced during this PhD	8
124	2 Radio Sources	11
125	2.1 The Extragalactic Radio Sky	11
126	2.2 Radio emission	13
127	2.2.1 Synchrotron radiation	14
128	2.2.2 Polarisation	14
129	2.3 Radio galaxies and active galactic nuclei	18
130	2.3.1 What we see when we look at AGN	18
131	2.3.2 Extended structure	20
132	2.3.3 The unified model	21
133	2.3.4 Polarised structure	22
134	2.3.5 AGN luminosity	23
135	2.3.6 The role of AGN	23
136	2.4 Classifying AGN	24
137	2.4.1 Statistical and manual classification of AGN	24
138	2.4.2 Machine learning classification of AGN	25
139	2.5 Cross-identification	27
140	2.5.1 Why do we need to cross-identify?	27

141	2.5.2	Methods for cross-identification	27
142	2.5.2.1	Positional matching	27
143	2.5.2.2	Other automated methods	28
144	2.6	Aggregating Radio Components	28
145	2.6.1	Missing emission in radio observations	29
146	2.6.2	Methods of aggregation	31
147	2.7	Summary: radio sources	31
148	3	Machine Learning for Astroinformatics	33
149	3.1	Prediction	33
150	3.1.1	Predictors	33
151	3.1.2	Classification	34
152	3.1.3	Regression	35
153	3.2	Data and representation	35
154	3.3	Loss functions	36
155	3.3.1	Loss function for regression	37
156	3.3.2	Loss function for binary classification	38
157	3.3.3	Gradient descent	39
158	3.4	Models	39
159	3.4.1	Logistic regression	40
160	3.4.2	Decision tree ensembles	40
161	3.4.3	Convolutional neural networks	41
162	3.5	Labels	42
163	3.5.1	Where do labels come from?	42
164	3.5.2	Label noise	43
165	3.6	Summary: machine learning	43
166	4	Radio Cross-identification	45
167	4.1	Introduction to cross-identification	46
168	4.2	Data	48
169	4.2.1	ATLAS	48
170	4.2.2	SWIRE	49
171	4.2.3	Radio Galaxy Zoo	49
172	4.3	Method	49
173	4.3.1	Cross-identification as binary classification	51
174	4.3.2	Limitations of our approach	53
175	4.3.3	Feature vector representation of infrared sources	57
176	4.3.4	Binary Classifiers <u>classifiers</u>	58
177	4.3.5	Labels	58
178	4.3.6	Experimental Setup <u>setup</u>	60
179	4.4	Results	61
180	4.4.1	Application to ATLAS-CDFS	61
181	4.4.2	Application to ATLAS-ELAIS-S1	66
182	4.5	Discussion	69

183	4.6	Summary	72
184	4.7	Acknowledgements	72
185	A	<u>Classification models</u>	73
186	A.1	<u>Logistic regression</u>	73
187	A.2	<u>Convolutional neural networks</u>	73
188	A.3	<u>Random forests</u>	74
189	B	<u>Accuracy tables</u>	74
190	C	<u>SWIRE object scores</u>	74
191	D	<u>ATLAS component cross-identifications</u>	78
192	E	<u>Cross-identification figures</u>	81
193	5	Radio Luminosity Functions	85
194	1	Introduction	86
195	2	Data	87
196	2.1	RGZ	88
197	2.2	FIRST	88
198	2.3	AllWISE	88
199	2.4	SDSS	89
200	3	Method	89
201	3.1	Visual verification	90
202	4	Radio luminosity functions	91
203	5	Discussion	96
204	5.1	Biases and uncertainties	96
205	5.2	Extended radio galaxies in the low- z Universe	97
206	5.3	Future work	99
207	6	Summary	99
208	7	Acknowledgements	100
209	F	<u>Sankey diagrams</u>	100
210	G	<u>Visual verification results</u>	100
211	H	<u>Radio luminosity function</u>	104
212	I	<u>Redshift completeness estimate</u>	104
213	J	<u>Giant radio galaxies</u>	105
214	6	Faraday Complexity	107
215	1	Introduction	108
216	2	Faraday <u>Complexity</u>	109
217	2.1	Prior work	109
218	2.2	Assumptions on Faraday dispersion functions	111
219	3	Classification approach	112
220	3.1	Features	112
221	3.2	Interpreting distances	114
222	3.3	Classifiers	115
223	4	Experimental method and results	115
224	4.1	Data	115

225	4.1.1	Simulated training and validation data	115
226	4.1.2	Observational data	116
227	4.2	Results on ‘ASKAP’ dataset	117
228	4.3	Results on ‘ATCA’ dataset	117
229	4.4	Results on observed FDFs	119
230	5	Discussion	121
231	5.1	Complexity and seeming ‘not simple’	121
232	5.2	Limitations	122
233	6	Conclusion	122
234	7	Acknowledgements	123
235	K	Simulating observed FDFs	123
236	L	2-Wasserstein begets Faraday moments	124
237	M	Euclidean distance in the no-RMSF case	125
238	N	Hyperparameters for LR and XGB	126
239	O	Predictions on real data	128
240	7	Conclusion	131
241	1	Future Work	133
242	2	Implications for radio citizen science	135
243	3	Implications for wide-area radio surveys	135
244	4	Final remarks	136
245	Appendix		136
246	5	Classification models	136
247	4.1	Logistic Regression	136
248	4.1	Convolutional neural networks	136
249	4.1	Random Forests	137
250	5	Accuracy tables	137
251	5	SWIRE object scores	138
252	5	ATLAS component cross-identifications	140
253	5	Cross-identification figures	145
254	5	Sankey diagrams	145
255	5	Radio luminosity function	145
256	5	Redshift completeness estimate	146
257	5	Giant radio galaxies	146
258	5	Visual verification results	148
259	5	2-Wasserstein begets Faraday moments	150
260	5	Euclidean distance in the no-RMSF case	151
261	5	Hyperparameters for LR and XGB	152
262	5	Predictions on real data	153
263	5	Simulating observed FDFs	153
264	Bibliography		155

List of Figures

266	2.1	False-colour image of the radio sky from the GLEAM survey.	12
267	2.2	A Faraday screen.	16
268	2.3	Examples of a FRI and a FRII radio galaxy.	19
269	2.4	Three radio galaxies with interesting structure.	20
270	2.5	The unified model of AGN.	21
271	2.6	RLF for star-forming galaxies and AGN, from Mauch and Sadler (2007).	23
272	2.7	A single dish telescope and a radio array.	29
273	2.8	An example of a ‘resolved out’ radio galaxy.	30
274	4.1	Examples showing key definitions of radio emission regions used throughout this chapter.	46
275	4.2	An example of finding the host galaxy of a radio source using our sliding-window method.	50
276	4.3	Our cross-identification method once a binary classifier has been trained.	51
277	4.4	A radio source breaking our assumption that there are no other radio sources with 1 arcmin of the source.	54
278	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.	54
279	4.6	A radio source breaking our assumption that the whole radio source is visible in the chosen radius.	55
280	4.7	Cumulative number of radio components in the expert and Radio Galaxy Zoo training sets with different signal-to-noise ratios.	58
281	4.8	CDFS field training and testing quadrants.	59
282	4.9	Balanced accuracy on the candidate classification task plotted against accuracy on the cross-identification task.	62
283	4.10	An example of predicted host galaxies in the candidate classification task.	63
284	4.11	Performance of our method with different binary classifiers on the binary classification task.	64
285	4.12	Performance of our approach using different binary classifiers on the cross-identification task.	65
286	4.13	Performance of different classification models on the binary classification task, tested on ELAIS-S1.	67
287	4.14	Performance of different classification models on the cross-identification task, tested on ELAIS-S1.	67
288	4.15	Balanced accuracies of classifiers with different SNR cutoffs.	68

301	4.16	Architecture of our CNN.	75
302	4.17	Examples of resolved sources with high disagreement between cross- identifiers.	84
304	5.1	RGZ-Ex radio luminosity function compared with the RLFs of Mauch and Sadler (2007).	91
305	5.2	<i>WISE</i> colour-colour distributions.	93
306	5.3	RLFs split by <i>WISE</i> colour regions.	94
308	5.4	Bivariate radio luminosity function showing radio luminosity against projected physical extent.	95
310	5.5	<u>Number of components removed from FIRST by each filter.</u>	101
311	5.6	<u>Number of sources removed by each filter.</u>	101
312	5.7	<u>Estimated completeness as a function of mid-infrared colour and magnitude.</u>	105
313	6.1	A simple FDF and its corresponding polarised spectra.	110
314	6.2	A complex FDF and its corresponding polarised spectra.	110
315	6.3	An illustration of our FDF features.	114
316	6.4	Mean prediction as a function of component depth separation and min- imum component amplitude for XGB and LR.	118
317	6.5	Principal component analysis for simulated data with observations over- laid.	119
319	6.6	Estimated rates of Faraday complexity for the Livingston and O'Sullivan datasets as functions of threshold.	120
320	6.7	The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex.	128
322	6.8	The 142 observed FDFs ordered by XGB-estimated probability of being Faraday complex.	129
324			
325			

List of Tables

327 4.1	Catalogues of ATLAS/SWIRE cross-identifications for the CDFS and 328 ELAIS-S1 fields.	48
329 4.2	Number of compact and resolved radio objects in each CDFS quadrant. .	59
330 4.3	Balanced accuracies for different binary classification models on CDFS. .	76
331 4.4	Balanced accuracies for different binary classification models on ELAIS- 332 S1.	76
333 4.5	Cross-identification accuracies for different classification models on CDFS. .	77
334 4.6	Cross-identification accuracies for different classification models on ELAIS- 335 S1.	77
336 4.7	Scores output by our trained classifiers for SWIRE CDFS candidate host 337 galaxies.	79
338 4.8	Scores output by our trained classifiers for SWIRE ELAIS-S1 candidate 339 host galaxies.	80
340 4.9	Cross-identifications for ATLAS CDFS components.	82
341 4.10	Cross-identifications for ATLAS ELAIS-S1 components. Columns are 342 defined in Appendix D. Full table electronic.	83
343 5.1	Medians and standard deviations used to normalise input features for 344 our classifiers.	89
345 5.2	Validation objects.	102
346 5.3	Giant radio galaxies found in RGZ-Ex.	106
347 6.1	Confusion matrix entries for LR and XGB on 'ASKAP' and 'ATCA' sim- 348 ulated datasets, and the CNN confusion matrix entries adapted from 349 Brown et al. (2018).	117
350 6.2	XGB hyperparameters for the 'ATCA' dataset.	127
351 6.3	LR hyperparameters for the 'ATCA' dataset.	127
352 6.4	XGB hyperparameters for the 'ASKAP' dataset.	127
353 6.5	LR hyperparameters for the 'ASKAP' dataset.	127

REVISION DIFF

³⁵⁴

List of Constants

³⁵⁵ The values of the following constants, except where otherwise noted, are drawn from
³⁵⁶ the NIST Reference on Constants, Units, and Uncertainty (Mohr et al., 2019) which
³⁵⁷ itself draws from the 2018 CODATA recommended values.

Symbol	Unit	Name	Value
ϵ_0	F m^{-1}	Vacuum permittivity	$8.8541878128(13) \times 10^{-12}$
G	$\text{m}^3 \text{ kg}^{-1} \text{ s}^{-2}$	Gravitational constant	$6.67430(15) \times 10^{-11}$
m_p	kg	Proton mass	$1.67262192369(51) \times 10^{-27}$
m_e	kg	Electron mass	$9.1093837015(28) \times 10^{-31}$
c	m s^{-1}	Speed of light	2.99792458×10^8
σ_T	m^2	Thomson cross section	$6.6524587321(60) \times 10^{-29}$

REVISION DIFF

358 List of Abbreviations

359 The following list summarises abbreviations that are commonly used in this thesis.

- 360 • AGN: active galactic nuclei, energetic objects at the centre of galaxies
- 361 • ASKAP: Australian Square Kilometre Array Pathfinder, a next-generation radio
362 telescope in Murchison
- 363 • ATCA: Australia Telescope Compact Array, a radio telescope in Narrabri
- 364 • CNN: convolutional neural network, a classifier which works on images and
365 spectra
- 366 • EMU: Evolutionary Map of the Universe, an upcoming large radio survey
- 367 • FDF: Faraday dispersion function, a representation of a polarisation spectrum
- 368 • FIRST: Faint Images of the Radio Sky at Twenty Centimeters, a large radio survey
- 369 • FRI: Fanaroff-Riley type I, an edge-darkened radio galaxy
- 370 • FRII: Fanaroff-Riley type II, an **edge-brightened** **edge-brightened** radio galaxy
- 371 • ISM: interstellar medium, the stuff between stars
- 372 • LR: logistic regression, a classification model
- 373 • MWA: Murchison Widefield Array, a next-generation radio telescope in Murchi-
374 son
- 375 • NVSS: NRAO VLA Sky Survey, a large radio survey
- 376 • POSSUM: Polarisation Sky Survey of the Universe's Magnetism, an upcoming
377 large radio polarisation survey
- 378 • RACS: Rapid ASKAP Continuum Survey, a new large radio survey
- 379 • RF: random forests, a classification model
- 380 • RGZ: Radio Galaxy Zoo, a citizen science project to cross-identify and aggregate
381 radio sources
- 382 • RLF: radio luminosity function, a description of how common radio galaxies of
383 different energies are

- 384 • RM: rotation measure, the amount of Faraday rotation between a polarised source
385 and an observer
- 386 • RMSF: rotation measure spread function, the kernel convolving a FDF
- 387 • SDSS: Sloan Digital Sky Survey, a large optical spectroscopic and photometric
388 survey
- 389 • SFR: star formation rate, the recent rate of star formation
- 390 • SKA: Square Kilometre Array, a next-generation radio telescope yet to be built
- 391 • SNR: signal-to-noise ratio, the ratio of total intensity to noise level
- 392 • SWIRE: *Spitzer* Wide-area Infrared Extragalactic Survey, a deep infrared survey
- 393 • VLA: Very Large Array, a radio telescope in New Mexico
- 394 • WISE: *Wide-field Infrared Survey Explorer*, a space-based infrared telescope
- 395 • XGB: extreme gradient boosted trees, a classification model

397

Introduction

398 Many great results come from study at the intersection of two fields, and the combina-
 399 tion of astronomy and informatics is no exception. The resulting interdisciplinary field
 400 is called *astroinformatics*, and concerns the application of statistical and machine learn-
 401 ing techniques to problems in astronomy and astrophysics. Machine learning, a col-
 402 lection of methods for formalising and solving data-driven problems at scale, is a nat-
 403 ural fit for radio astronomy: Radio astronomy faces a ‘data deluge’ thanks to new and
 404 upgraded telescopes and associated wide-area surveys to be undertaken with them.
 405 The goal in the near future is to be able to process data—and conduct science on that
 406 data—on the fly as the volume of data grows too large to store. This is a requirement
 407 for the Square Kilometre Array (SKA), a grand international undertaking to build a
 408 distributed radio array across Western Australia and South Africa with an intended
 409 collecting area of one square kilometre. There exists no comparable telescope today.
 410 The SKA will be able to perform new tests of general relativity, help pin down the large-
 411 scale structure of the Universe, investigate the mysteries of the epoch of reionisation,
 412 probe the history of galaxy evolution to new extents, and perhaps make unexpected
 413 new discoveries (Diamond, 2017).

414 The SKA will have technical challenges unlike any so far in radio astronomy. Raw
 415 data will stream from the telescope antennae at 2 petabytes per second, and up to 300
 416 petabytes per year of science data is expected to be generated (Diamond, 2017). This
 417 is a phenomenal amount of data, much of which won’t be stored, and the community
 418 expectation is that machine learning will provide avenues to conduct science with such
 419 a large dataset. Precursor projects to the SKA have begun to investigate these pathways
 420 (e.g. Bonaldi et al., 2020; Kapinska, 2020; Mostert et al., 2021).

421 Even without the SKA, the data deluge has already begun. Three precursor tele-
 422 scopes have been constructed: the Australian Square Kilometre Array Pathfinder
 423 (ASKAP) and Murchison Widefield Array (MWA) in Western Australia, and MeerKAT
 424 in South Africa. While MeerKAT will eventually form part of the SKA itself, all three
 425 are already online and generating science data at astonishing rates. ASKAP will soon
 426 conduct six surveys of the entire southern radio sky: the Evolutionary Map of the Uni-
 427 verse (EMU; Kapinska, 2020; Norris et al., 2011), the Widefield ASKAP L-Band Legacy
 428 All-Sky Blind Survey (Koribalski et al., 2020, WALLABY), the First Large Absorption
 429 Survey in HI (FLASH), an ASKAP Survey for Variables and Slow Transients (VAST;
 430 Murphy et al., 2013), the Galactic ASKAP Spectral Line Survey (GASKAP), and the

431 Polarisation Sky Survey of the Universe’s Magnetism (POSSUM); as well as the deeper
 432 but smaller Deep Investigation of Neutral Gas Origins (DINGO) and the Commensal
 433 Real-Time ASKAP Fast-Transients survey (CRAFT; Macquart et al., 2010). Atop these
 434 future surveys, the recent Rapid ASKAP Continuum Survey (RACS; McConnell et al.,
 435 2020) has redefined our knowledge of the southern radio sky with shallow observa-
 436 tions at 15 arcsecond resolution—compare to the previous largest radio survey, the
 437 NRAO VLA Sky Survey (NVSS; Condon et al., 1998), with 45 arcsecond resolution
 438 over the northern sky.

439 Machine learning methods for radio astronomy will be developed for and tested
 440 upon surveys like RACS and EMU. The path between an astronomical problem and a
 441 machine learning problem, however, is not a straightforward one. The goal when cast-
 442 ing an astronomy question as something mathematical or computational is to convert
 443 the question into one with a known method of solution, such as classification or regres-
 444 sion. Along the way, astronomical concepts and assumptions need to be turned into
 445 something a computer can deal with. Despite the wide availability of machine learn-
 446 ing software and tools, there is no automatic or easy way to make this transformation.
 447 This necessitates research in astroinformatics.

448 This thesis concerns applications of machine learning to radio astronomy for the
 449 identification of extended extragalactic radio sources in wide-area surveys. We will
 450 present new methods of cross-identifying radio objects with their corresponding in-
 451 frared and optical observations, demonstrate the applicability of these methods to
 452 existing wide-area radio surveys (shedding light on radio source population astron-
 453 omy along the way), and develop a new way to identify complexity in polarised radio
 454 sources.

455 1.1 Problems in extragalactic radio astronomy

456 Galaxies produce radio emission through a variety of methods. The main emission
 457 mechanisms are star formation and active galactic nuclei (AGN, Section 2.3) and only
 458 the latter show extended structure well beyond the galaxy itself. AGN are the cen-
 459 tral focus of this thesis. They are intensely energetic objects at the centre of galaxies,
 460 which actively accrete matter and eject huge jets of plasma that develop into extended
 461 lobes over huge distances. Radio astronomy has many uses for AGN: ~~their~~Their en-
 462 ergy scales provide a test-bed for high energy physics, and the extremely bright lobes
 463 and jets can be seen throughout the Universe, making AGN an accessible probe of the
 464 distant and old Universe.

465 AGN are thought to be critical to galaxy evolution and perhaps the early reionisa-
 466 tion of the Universe (Bosch-Ramon, 2018), but their exact role in their host galaxies
 467 is an open question. The radiative and mechanical energy released by AGN impacts
 468 the interstellar medium (ISM) and is a key component of contemporary galaxy sim-
 469 ulations and models (Morganti, 2017). The quenching of star formation due to AGN
 470 activity is called *AGN feedback*, the idea being that the energy expelled from an AGN is
 471 returned to the ISM of the galaxy, heating the gas so it cannot condense into stars. The

472 different impacts of radiative and mechanical energy, the location and scales within
473 the galaxy for which star formation is quenched, and whether star formation material
474 is ejected from the galaxy by the AGN are all open questions (Husemann & Harrison,
475 2018). One of our key results in Chapter 5 is an estimation of the mechanical energy
476 contributed to the intergalactic medium by AGN. Solving these questions requires an
477 understanding of AGN at both small and large scales, as well as connecting AGN to
478 their host galaxies at other wavelengths so that redshift, emission lines, star formation
479 rates, etc. can be determined. Chapter 4 describes our new method for performing
480 such cross-identifications.

481 The large scales of AGN also provide insight into the larger-scale structure of the
482 Universe. Giant radio galaxies for example (Section 2.3.1) are difficult to identify due
483 to their size and disconnected appearance (Section 2.6), but are so large that they can
484 be used to probe galaxy clusters (Banfield et al., 2016) and even the large-scale struc-
485 ture of the Universe (Reiprich et al., 2020). Other large-scale effects seem to exist, such
486 as the apparent alignment of radio galaxies ([Contigiani et al., 2017; Panwar et al., 2020](#))
487 ([Contigiani et al., 2017; Panwar et al., 2020; Taylor & Jagannathan, 2016](#)), though inves-
488 tigation continues as to whether this effect is real or due to some unknown systematic
489 bias. New radio surveys will reveal more radio sources than ever before, and if they can
490 be identified, radio structures in these surveys will allow us to investigate the structure
491 of the Universe.

492 The magnetic structure of AGN and their extended lobes may be probed by ra-
493 dio polarimetry observations (Anderson et al., 2015; Grant, 2011). Through polarisation,
494 though, extragalactic AGN can provide insight into our own [Galaxy: the galaxy:](#)
495 [The](#) Faraday depth and complexity (Section 2.2.2) of extragalactic radio sources can
496 be used to quantify local magnetic fields. With more polarised radio sources to be
497 revealed through upcoming wide-area polarisation surveys, the magnetic field of the
498 Milky Way and its surrounding intergalactic medium can be better resolved. Polarime-
499 try also allows us to determine some aspects of the structure of unresolved extragalac-
500 tic sources, even though we cannot spatially see that structure: [Spatially](#) [Spatially](#) ex-
501 tended, polarised radio sources may have different polarisation spectra to those which
502 are spatially compact when projected onto the sky.

503 Other problems in radio astronomy relate to the new level of data that we are about
504 to obtain from large telescopes like the SKA, which we discuss in Section 1.2.

505 1.2 Big data in astronomy

506 The scale of radio data underpins many of the methodology problems facing radio as-
507 tronomy. There are two main scientific benefits that come from large-scale data: better
508 statistics and more unusual objects. However, methods for dealing with radio data at
509 scale are still very much in their infancy, and need to be developed before instruments
510 like ASKAP and MeerKAT can be used to their full potential.

511 Many results in astronomy are statistical, from measuring the expansion of the
512 Universe to understanding the distribution of galaxy properties. With more observa-

513 tions, we can not ~~just only~~ narrow the uncertainty of these results, but ~~also diversify~~
 514 ~~them~~diversify them as well. When the number of objects under study is large, we can
 515 subdivide the population into subpopulations based on their physical properties and
 516 determine a statistic on each subpopulation. This can help understand the physical ba-
 517 sis behind the statistic, or remove unwanted subpopulations from analysis. Even with
 518 less data it is still possible to subdivide or filter populations, but this will dramatically
 519 raise the uncertainty in the results due to the low sample size of each bin. An example
 520 of such a statistic is the radio luminosity function (RLF), which describes the density of
 521 radio sources throughout the Universe. It can be divided into a fractional RLF to exam-
 522 ine the distribution of subpopulations, or to remove the effect of star-forming galaxies.
 523 We use large datasets to improve the uncertainty of RLFs and subdivide RLFs by the
 524 infrared properties of the population in Chapter 5.

525 With large datasets, highly unusual or rare objects are more likely to be included.
 526 Much of astronomy has been pushed forward by serendipitous discoveries, and (pro-
 527 vided we have some way of combing through the dataset) large datasets should pro-
 528 vide a wealth of such discoveries to be found. These may be found either through
 529 identifying objects where statistical methods seem to fail, or perhaps through direct
 530 searches (Norris, 2017a). By applying machine learning techniques, we ~~found~~find a
 531 number of new, rare giant radio galaxies in Chapter 5.

532 But with the benefits of big data come new challenges. At these new scales, the abil-
 533 ity to store all of our science data is no longer a given. Many methods that previously
 534 had the luxury to run over a whole dataset at their own pace will now need to process
 535 data on-the-fly. This is in contrast to how most science observations are currently per-
 536 formed, with new discoveries coming from legacy surveys many years after they were
 537 conducted. Many discoveries are unplanned, and losing the ability to make serendip-
 538 itous discoveries would be a major blow to astronomy (Norris, 2017a): ~~how~~How can
 539 we deal with so much data but still retain the ability to discover the unknown? Even
 540 with storage, the scale is tremendous: FIRST, for example, contains around 900 000
 541 sources, of which very few were manually labelled. Over 10 000 volunteers labelled in-
 542 teresting objects in FIRST over four years of the Radio Galaxy Zoo project, with 75 000
 543 aggregated labels passing quality assurance testing. While a phenomenal and, in ra-
 544 dio, unparalleled labelling effort, it pales in comparison to the estimated 70 000 000
 545 sources that EMU will find (Banfield et al., 2015).

546 1.3 Machine learning in astronomy

547 Machine learning is the process and practice of designing algorithms that automati-
 548 cally find and exploit useful patterns in data (Deisenroth et al., 2020). It is best suited
 549 for scenarios where these patterns are hard to encode by hand. Typically hard-to-
 550 describe problems might include data exploration, data visualisation, or interpolation.
 551 We face a deluge of data, and our existing methods for data analysis and astronomical
 552 discovery are hard to automate due to the often complex, ~~visual-imagery-based~~ nature
 553 of much of astronomy. This is where machine learning comes in: ~~if we can~~If we could

554 find a way to encode the processes underpinning astronomy as algorithms, we could
555 apply them at scale. Machine learning has found much success in other fields with sim-
556 ilar difficult-to-describe problems like image classification and outlier identification.

557 From the other direction, machine learning finds interesting and unique applica-
558 tions in radio astronomy. Radio images are usually quite different from images in the
559 typical image analysis problems faced in computer science. Standard assumptions are
560 frequently broken: ~~noise~~ Noise is correlated across the image, objects can be different
561 shapes when viewed at different wavelengths, and the number of colours in an image
562 may be in the hundreds. Data points are spatially located and can be close together or
563 far apart in ways that are not solely based on how they look or behave. We often do not
564 have the option of obtaining more data, sometimes because it is prohibitively expen-
565 sive (e.g. to undertake a new survey) and sometimes because it simply does not exist
566 (as we only have one Universe to look at or because the event is a transient one-off).
567 The burgeoning field of astroinformatics promises to prove interesting.

568 1.4 How this thesis fits in

569 Current methods of identifying radio sources in wide-area radio surveys work best for
570 compact, isolated objects. As resolution and depth of radio surveys ~~increases~~ increase
571 in the lead-up to the SKA, these limitations are becoming more apparent: ~~resolution~~
572 Resolution increases apparent complexity and depth increases the number of radio
573 sources visible in any given patch of sky.

574 Source identification is critical for analysis of radio sources in large-scale surveys
575 whether one is interested in individual sources or their bulk properties. Without solv-
576 ing issues like cross-identification and Faraday complexity we will not be able to make
577 full use of the SKA when it arrives. As we will discuss in Section 2.5, cross-identification
578 of radio sources with their counterparts in infrared or optical is key to understanding
579 these sources. Many radio galaxy properties can only be derived from non-radio ob-
580 servations of the galaxy, the stellar component of which can be traced by emission in
581 the infrared and optical. This can yield properties including distance, star formation
582 rate, and even mass of the central black hole powering the radio source itself. We de-
583 velop an automated machine learning method for cross-identification in Chapter 4 and
584 apply it to obtain physical results in Chapter 5.

585 Bulk analysis of polarised sources greatly benefits from being able to assess their
586 Faraday complexity. The most important application is perhaps in the development of
587 so-called *rotation measure grids*, which help characterise the magnetic field of the Milky
588 Way and surrounding intergalactic medium. Without knowing which sources are com-
589 plex and which are not, we cannot estimate the grid reliably. A fast way to estimate
590 Faraday complexity also allows us to quickly determine which sources may need fol-
591 lowing up, either with more expensive algorithms or further observations. Chapter 6
592 details our development of an automated Faraday complexity classifier.

593 **1.5 Thesis outline**

594 We begin by introducing key concepts from radio astronomy in Chapter 2, including
 595 radio active galactic nuclei (AGN), as well as motivations for and difficulties in cross-
 596 identifying observed radio emission from AGN. In Chapter 3 ~~introduces we introduce~~
 597 machine learning and ~~describes~~describe the machine learning background required
 598 for the remainder of the thesis. These chapters together comprise the background
 599 knowledge.

600 Chapter 4 is my paper *Radio Galaxy Zoo: Machine learning for radio source host galaxy*
 601 *cross-identification* (Alger et al., 2018). This chapter describes a new ~~,~~ machine learning
 602 method for cross-identifying extended radio emission with host galaxies in the
 603 infrared. We apply this approach to cross-identifying all extended radio sources in
 604 the Faint Images of the Radio Sky at Twenty Centimeters radio survey (FIRST; Becker
 605 et al., 1995) with their infrared counterparts in AllWISE (Cutri et al., 2013) and use the
 606 resulting catalogue of cross-identifications to create a fractional radio luminosity func-
 607 tion in Chapter 5, which itself is my paper *Radio Galaxy Zoo: Radio luminosity functions*
 608 *of extended sources* (Alger et al., in prep.). Chapter 6 is my paper *Interpretable Faraday*
 609 *Complexity Classification* (Alger et al., 2021), which introduces an interpretable method
 610 for classifying radio emission as Faraday complex or simple, which could be used to
 611 identify whether it is an extended source that is below the resolution limit. In Chap-
 612 ter 7 ~~provides we provide~~ a discussion of the insights gained from the findings of this
 613 thesis and ~~suggests~~suggest future directions for research building on our results.

614 **1.6 Contributions**

615 My main contributions to radio astroinformatics in this thesis are:

- 616 • I introduce a new method for cross-identifying radio emission which can learn
 617 from existing catalogues, the first application of machine learning to radio cross-
 618 identification (Chapter 4);
- 619 • I demonstrate an application of this new method to the creation of fractional ra-
 620 dio luminosity functions, which require considerably more cross-identifications
 621 than non-fractional radio luminosity functions, in the process creating the largest
 622 available catalogue of extended, cross-identified radio sources (Chapter 5);
- 623 • I produce a fractional radio luminosity function with divisions based on mid-
 624 infrared colours associated with the host galaxies of the radio emission, helping
 625 to understand how radio galaxies evolve throughout the Universe (Chapter 5);
- 626 • I highlight the requirement of considerably more redshifts in understanding fu-
 627 ture wide-area radio surveys (Chapter 5);
- 628 • I introduce a new method to identify Faraday complexity using an interpretable
 629 classifier, as well as features for Faraday dispersion functions which can be used
 630 for other machine learning tasks (Chapter 6); and

- ⁶³¹ • I apply machine learning to real radio spectropolarimetric data for the first time
⁶³² (Chapter 6).

REVISION DIFF

633 1.7 Works produced during this PhD

634 During my PhD, I had the good fortune to collaborate with many talented authors
 635 and contribute to a number of related works. The following is a list of papers which I
 636 co-authored in this time:

- 637 • *Radio Galaxy Zoo: Machine learning for radio source host galaxy cross-identification*
 638 (2018). **M. J. Alger**, J. K. Banfield, C. S. Ong, L. Rudnick, O. I. Wong, C. Wolf,
 639 H. Andernach, R. P. Norris, and S. S. Shabala. *Monthly Notices of the Royal Astro-*
 640 *nomical Society* 478, pp. 5547–5563. [This is the content of Chapter 4 with minor](#)
 641 [modifications.](#)
- 642 • *Radio Galaxy Zoo: Radio luminosity functions of extended sources* (in prep.). **M. J.**
 643 **Alger**, O. I. Wong, C. S. Ong, N. M. McClure-Griffiths, H. Andernach, L. Rudnick,
 644 S. S. Shabala, A. F. Garon, J. K. Banfield, A. D. Kapińska, R. P. Norris, and A. J. M.
 645 Thomson. [This is the content of Chapter 5 with minor modifications.](#)
- 646 • *Interpretable Faraday Complexity Classification* ([in press](#) 2021). **M. J. Alger**, J. D. Liv-
 647 ington, N. M. McClure-Griffiths, J. L. Nabaglo., O. I. Wong, and C. S. Ong. *Pub-*
 648 *lications of the Astronomical Society of Australia* –38, E022. [This is the content of](#)
 649 [Chapter 6 with minor modifications.](#)
- 650 • *Radio Galaxy Zoo Data Release 1: visual identification of 75,641 radio morphologies from*
 651 *the FIRST and ATLAS surveys* (in prep.). O. I. Wong, A. F. Garon, **M. J. Alger**, K.
 652 W. Willett, L. Rudnick, J. K. Banfield, J. Swan, S. S. Shabala, H. Andernach, R. P.
 653 Norris, B. D. Simmons, A. D. Kapińska, N. Seymour, et al.
- 654 • *Radio Galaxy Zoo: CLARAN - a deep learning classifier for radio morphologies* (2019).
 655 C. Wu, O. I. Wong, L. Rudnick, S. S. Shabala, **M. J. Alger**, J. K. Banfield, C. S. Ong,
 656 S. V. White, A. F. Garon, R. P. Norris, H. Andernach, J. Tate, V. Lukic, H. Tang,
 657 K. Schawinski, and F. I. Diakogiannis. *Monthly Notices of the Royal Astronomical*
 658 *Society* 482, pp. 1211–1230.
- 659 • *Heightened Faraday Complexity in the inner 1 kpc of the Galactic Centre* (2021), by J.
 660 D. Livingston, N. M. McClure-Griffiths, B. M. Gaensler, A. Seta, and **M. J. Alger**.
 661 *Monthly Notices of the Royal Astronomical Society* 502, pp. 3814–3828.
- 662 • *Radio Galaxy Zoo: Unsupervised Clustering of Convolutionally Auto-encoded Radio-*
 663 *astronomical Images* (2019). N. O. Ralph, R. P. Norris, G. Fang, L. A. F. Park, T. J.
 664 Galvin, **M. J. Alger**, H. Andernach, C. Lintott, L. Rudnick, S. S. Shabala, and O. I.
 665 Wong. *Publications of the Astronomical Society of the Pacific* 131, 108011.
- 666 • *Radio Galaxy Zoo: Knowledge Transfer Using Rotationally Invariant Self-organizing*
 667 *Maps* (2019). T. J. Galvin, M. Huynh, R. P. Norris, X. R. Wang, E. Hopkins, O. I.
 668 Wong, S. S. Shabala, L. Rudnick, **M. J. Alger**, and K. L. Polsterer. *Publications of*
 669 *the Astronomical Society of the Pacific* 131, 108009.

670 I also gave a number of conference talks:

- 671 • *Machine Learning Methods for Radio Host Cross-Identification with Crowdsourced La-*
672 *bels*, presented in 2017 at the seventh SKA Pathfinder Radio Continuum Survey
673 meeting, Perth, Australia.
- 674 • *Learning to Cross-identify Wide-area Radio Surveys with Radio Galaxy Zoo: Data chal-*
675 *lenges in citizen science*, presented in 2018 at the Collaborative Conference on Com-
676 putational and Data Intensive Science, Melbourne, Australia.
- 677 • *Radio luminosity functions with Radio Galaxy Zoo and machine learning*, presented
678 in 2019 at the ninth SKA Pathfinder Radio Continuum Survey meeting, Lisbon,
679 Portugal.
- 680 • *Radio luminosity functions with machine learning and Radio Galaxy Zoo*, presented in
681 2019 at the Annual Scientific Meeting of the Astronomical Society of Australia.
- 682 • *Extracting Meaningful Features from Early-Science Radio Data*, presented in 2019 at
683 the Collaborative Conference on Computational and Data Intensive Science, Can-
684 berra, Australia.
- 685 • *Extracting Meaningful Features from Early-Science Radio Data*, presented in 2019 at
686 Artificial Intelligence in Astronomy, Garching, Germany.

REVISION DIFF

688 Radio Sources

689 As its title suggests, this thesis ~~is focused~~focuses on the identification of extended ra-
 690 dio sources. This chapter introduces extended radio sources, describing what we see
 691 when we look at the sky with radio eyes and radio telescopes. We will discuss the dif-
 692 ferent kinds of radio sources that we can observe, how they are distributed throughout
 693 the Universe, and key issues surrounding their identification. We will start by looking
 694 at the extragalactic radio sky, introducing what it is that we see through radio tele-
 695 scopes (Section 2.1). Then, we will describe the physics behind radio emission, includ-
 696 ing the emission mechanisms that allow us to observe active galactic nuclei in radio,
 697 and how radio polarisation can help us determine aspects of distant magnetic fields
 698 (Section 2.2). From there we will turn to active galactic nuclei as they ~~will be are~~are the
 699 objects of interest in this thesis (Section 2.3): ~~what~~What do they look like, how are
 700 they structured, and what do they do? Finally, we will cover three different tasks we
 701 may face when presented with a collection of observed AGN, including classification
 702 (Section 2.4), cross-identification (Section 2.5), and source aggregation (Section 2.6).

703 2.1 The Extragalactic Radio Sky

704 The extragalactic sky appears quite different at different wavelengths. While an op-
 705 tical observer may look at a distant galaxy and see spirals and halos, an infrared ob-
 706 server will see discs and dust. What does the radio astronomer see? Figure 2.1 shows
 707 a false-colour image of the radio sky from 72–231 MHz. The plane of the Milky Way
 708 is clearly visible through the centre, but nearly every other object in this image is a
 709 galaxy. These galaxies fall into two main categories: those that emit radio due to star
 710 formation (called *star-forming galaxies*), and those that emit radio due to *active galactic*
 711 *nuclei* (AGN; called *radio galaxies* in this thesis). AGN can be observed at many fre-
 712 quencies: ~~for~~For the remainder of this thesis, except where otherwise noted, ‘AGN’
 713 refers to *radio AGN*.

714 Non-AGN emission from distant galaxies traces the recent star-formation rate (SFR).
 715 Besides low-power thermal emission, stellar radio emission from galaxies mainly comes
 716 from massive ($\gtrsim 8 M_{\odot}$) stars ~~through~~ two emission mechanisms. The first is through
 717 H II regions, which are ionised by such stars. The ionised electrons emit bremsstrahlung
 718 radiation at radio wavelengths. The second emission mechanism is supernovae. Mas-

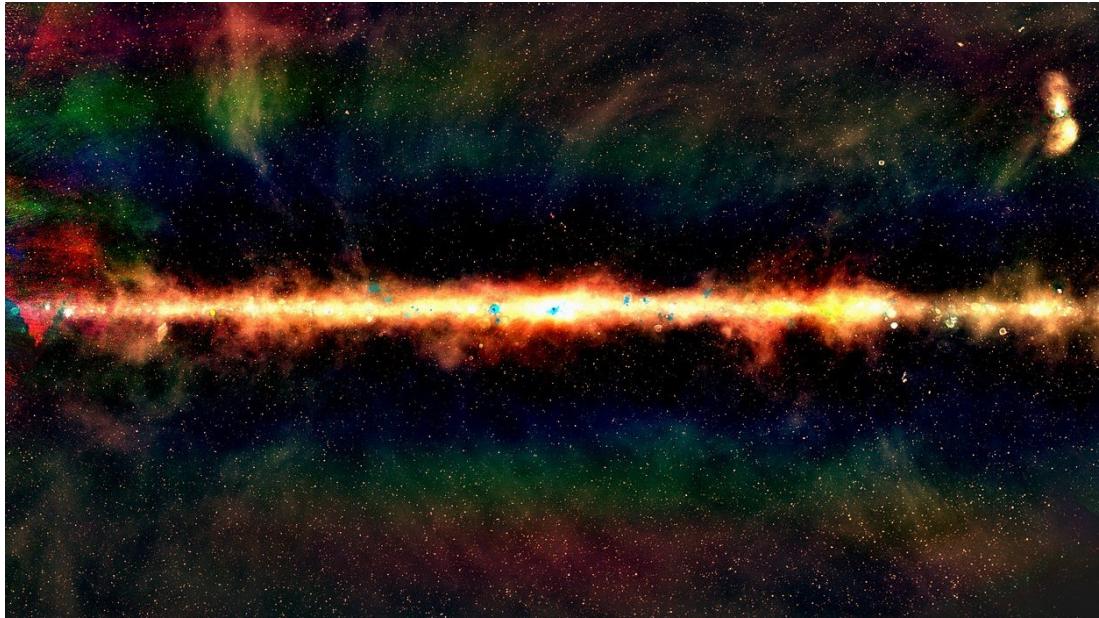


Figure 2.1: False-colour image of the radio sky from the GLEAM survey. (Image: Natasha Hurley-Walker, Curtin University/ICRAR; Hurley-Walker et al., 2017)

719 massive stars may end their lives in Type II and Type Ib supernovae, which can result in su-
 720 pernova remnants. These remnants emit Interaction between the supernova remnant
 721 and the interstellar medium (ISM) causes the emission of synchrotron radiation. Mas-
 722 sive stars like these are short-lived (a few 10^6 yr), and the corresponding emitting
 723 electrons have similarly short lifetimes ($\lesssim 10^8$ yr). The radio effects of these stars are
 724 therefore also short-lived, which is why radio emission traces the recent SFR (Condon,
 725 1992). Star formation-associated emission is mainly found in the dise discs of spiral
 726 galaxies, as this is where the star formation rate is highest. In particular, there is no
 727 star-forming radio emission extending outside of the galaxy proper. The radio power
 728 emitted by these galaxies at 1.4 GHz is on the order of 10^{18} – 10^{23} W Hz $^{-1}$ (Condon,
 729 1992). For a radio survey like the NRAO VLA Sky Survey (NVSS; Condon et al., 1998),
 730 with a detection limit of 2.3 mJy, this luminosity range corresponds to a maximum red-
 731 shift range of 0.0004–0.1272 (corresponding to 6×10^6 – 1.646×10^9 yr lookback time¹).
 732 Upcoming surveys such as the Evolutionary Map of the Universe (EMU; Norris et al.,
 733 2011), with 5σ detection thresholds of 50 μ Jy (Norris et al., 2011), will push this red-
 734 shift range to 0.0030–0.6684 (corresponding to 4.2×10^7 – 6.261×10^9 yr lookback time).

735 AGN are energetic objects at the centre of galaxies, powered by accretion into su-
 736 permassive black holes. The extended, strongly-magnetised plasma they eject emits
 737 synchrotron radiation from accelerating relativistic electrons, which is what we see
 738 when we observe a radio galaxy. The radio luminosity of a radio galaxy can range
 739 from 10^{20} – 10^{28} W Hz $^{-1}$ (Pracy et al., 2016) at 1.4 GHz, making them some of the most

¹Assuming cosmological parameters from Chapter 5. Calculated using “A Cosmology Calculator for the World Wide Web” (Wright, 2006).

740 luminous objects in the Universe. They are therefore visible throughout the Universe,
741 with the most distant AGN detected at a redshift of 7.5 (Bañados et al., 2018). De-
742 pending on the orientation and type of AGN, as well as its interaction with its host
743 galaxy, the radio emission may extend far beyond the galaxy itself—up to megaparsec
744 scales—and this emission may have complex structure. Perhaps the most impressive
745 local example is Centaurus A (Cen A), the prominent double-lobed cloud in the upper-
746 right of Figure 2.1 extending over 8 degrees across the sky. Section 2.3 discusses AGN
747 in more detail.

748 Most ~~radio galaxies are~~ AGN emission is compact and unresolved in any given
749 radio survey due to the distance at which ~~they~~ it can be detected and ~~their~~ the orien-
750 tation or type of the galaxy. This means that ~~their~~ observed structure does not always
751 help to distinguish AGN radio emission from star-forming radio emission. How can
752 we tell these apart? Synchrotron emission has a considerably steeper spectral index
753 than bremsstrahlung, but synchrotron emission dominates the bremsstrahlung in star-
754 forming galaxies at 1.4 GHz (Condon, 1992). Truly star-forming galaxies can be distin-
755 guished from AGN host galaxies by using optical spectroscopy (e.g. Groves & Kewley,
756 2007; Mauch & Sadler, 2007), but radio emission is detectable at much greater dis-
757 tances than those at which good quality optical spectra can be obtained~~at~~, making this
758 solution impractical for many galaxies. Separating star-forming galaxies from AGN
759 host galaxies at radio wavelengths remains a difficult problem in radio astronomy.

760 Polarised radio surveys can provide extra information. While radio emission due
761 to star formation tends to not have detectable polarisation, AGN may be very strongly
762 polarised. This makes polarisation an excellent indicator of whether a source is an
763 AGN, though very incomplete: ~~many AGN will also~~ Many AGN do not have detectable
764 polarisation, and the polarised intensity is usually less than ~~ten~~ 10 per cent of the total
765 radio intensity, meaning we detect far fewer polarised radio sources than we do radio
766 sources in general.

767 From the size scales described above, it should be clear that a survey of extended ra-
768 dio sources ~~will be is~~ dominated by AGN. Nevertheless, star-forming galaxies present
769 a significant part of the radio population, and the fraction of the radio sky they com-
770 prise varies significantly with survey parameters.

771 2.2 Radio emission

772 Electromagnetic radiation in radio frequencies—about 10 MHz–1 THz (Condon & Ran-
773 som, 2016)—is called *radio emission*. This is a very broad range of frequencies and so
774 radio astronomy covers a very broad range of astrophysical phenomena, from cosmo-
775 logical background radiation to neutron stars. The focus of this thesis is the excit-
776 ing, dynamic, and so-called ‘violent universe’ of radio galaxies. These galaxies are ob-
777 served through their emission of synchrotron radiation and are studied through their
778 observed physical structure, the intensity and spectroscopic properties of their radia-
779 tion, and the polarisation and spectropolarimetric properties that are uniquely visible
780 in radio. This section introduces synchrotron radiation and radio polarisation.

781 **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 spirals~~ 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~~-~~It~~ can be thought of as gyro radiation which has been Lorentz transformed to energies much greater than mc^2 . The spectrum of ~~optically thin~~ synchrotron radiation follows a power law (Condon & Ransom, 2016):

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

where ν is the frequency of radiation and α is called the *spectral index*². It is related to the energy distribution of the emitting electrons: ~~assuming~~. ~~Assuming~~ that the electron energy distribution follows a power law (~~which it generally does, Rybicki & Lightman, 2008~~) (~~which it generally does; Rybicki & Lightman, 2008~~), where the number density of electrons at a given energy E is given by

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

then

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

- 782 The spectral index for synchrotron radiation tends to range from -2 to 0 (Condon &
783 with spectral indices greater than 0 called ‘inverted’ spectra.

784 **2.2.2 Polarisation**

785 Electromagnetic radiation consists of waves of self-propagating, orthogonal electric
786 and magnetic fields. The orthogonality of these two waves allows us to characterise
787 the radiation just by the electric field. As a transverse wave, the electric field travels at
788 an angle in the plane perpendicular to the line-of-sight. This angle and its behaviour
789 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

²Note that the sign of α varies by convention, and both $S \propto \nu^\alpha$ and $S \propto \nu^{-\alpha}$ exist in the literature.

also possible. For this reason, we decompose the polarisation into linearly polarised components and a circularly polarised component, called *Stokes parameters* (Condon & Ransom, 2016; Stokes, 1851). These are:

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

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

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

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

\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~~ are 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 contribution of circular polarisation is tremendously small, especially compared to that of linear polarisation, and can be assumed to be zero (Rayner et al., 2000; Saikia & Salter, 1988). Incoherent radiation may be composed of radiation with many different polarisations, and these polarisations may fully or partially cancel out: ~~this~~. This is called *unpolarised* or ~~partially-polarised~~ 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 (O'Sullivan et al., 2015)~~ (on the order of 10 per cent; O'Sullivan et al., 2015). Additionally, the most common non-AGN cause for radio emission is star formation, which does not generally have detectable polarisation in extragalactic surveys. Polarisation is therefore an excellent way to confirm that a radio source is an AGN.

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

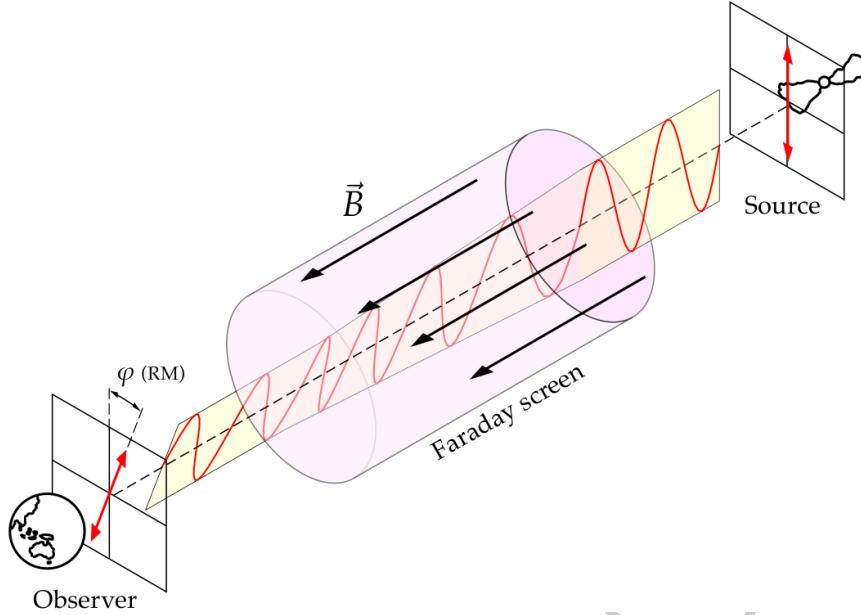


Figure 2.2: The Faraday effect for a single Faraday screen with magnetic field \vec{B} inducing a Faraday rotation of ϕ . Adapted from a figure by Bob Mellish.

field strength $\vec{B} \cdot \hat{z}$ of the intervening medium:

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

Here $d\vec{r}$ is the infinitesimal path length in pc (Brentjens & de 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 :

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

⁷⁹⁶ One useful way of thinking about this equation is that F is the decomposition of $P(\lambda^2)$ into complex sinusoids of the form $e^{2i\lambda^2\phi}$.

⁷⁹⁷ 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:

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

and the FDF is a delta distribution:

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

798 ϕ is then called the *rotation measure* (RM). If the source is not Faraday simple, then it
799 is called *Faraday complex*, and the question of whether a source is Faraday simple or
800 Faraday complex is called *Faraday complexity*. A diagram of a Faraday screen is shown
801 in Figure 2.2. Until very recently, the frequency resolution of polarised surveys was
802 insufficient to meaningfully separate most complex arrangements of Faraday depths,
803 and so most sources were assumed to be simple and characterised entirely in terms of
804 their rotation measure (e.g. Taylor et al., 2009). Advancing telescope technology and
805 emphasis on polarisation science has opened new frontiers in spectropolarimetry and
806 upcoming and ongoing surveys (e.g. RACS and POSSUM) will likely report Faraday
807 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 & de 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* (or *windowing function*, e.g. Heald, 2008) $W(\lambda^2)$. $W(\lambda^2)$ 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:

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

808 $P(\lambda_j^2)$ is the observed polarisation at the j th value of wavelength, $W(\lambda_j^2)$ is the corre-
809 sponding j th weight, and J is the total number of wavelengths for which measurements
810 were taken. The weighting function W is analogous to the weighting function in radio
811 synthesis imaging. The most common choices of W are 1) uniform weighting³ with
812 $W(\lambda_j^2) = 1$ for all nonzero values, and 2) weighting by the inverse variance at each
813 wavelength.

Of course, no physical source has a precise Faraday depth, as there is always in-

³The analogous weighting scheme in radio synthesis imaging would be natural weighting, rather than uniform—an unfortunate overlap in terminology.

trinsic scatter. Along the ~~line of sight~~^{line-of-sight}, if we assume ~~Gaussian noise in an otherwise constant that~~^{n_e is observed with random Gaussian noise} i.e. $n_e(z) \sim \mathcal{N}(\bar{n}_e, \sigma_{n_e}^2)$, and ~~a constant that~~^{B is constant} 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:

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

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

817 2.3 Radio galaxies and active galactic nuclei

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

825 2.3.1 What we see when we look at AGN

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

831 As powerful sources of radio emission, radio AGN and their associated extended
 832 structure can be seen throughout the Universe. Sufficiently close or large radio galax-
 833 ies can be resolved by telescopes and their structure examined, while more distant
 834 or smaller radio galaxies may be unresolved and point-like. A well-resolved radio
 835 galaxy can be a striking thing: ~~from~~^{From} the central AGN extend two opposing,
 836 tightly-collimated jets, which widen into huge lobes of radio-bright plasma. These
 837 lobes may have further structure, particularly bright regions called *hot-spots*, and the
 838 jets and lobes may be bent and distorted as they travel away from their host galaxy.

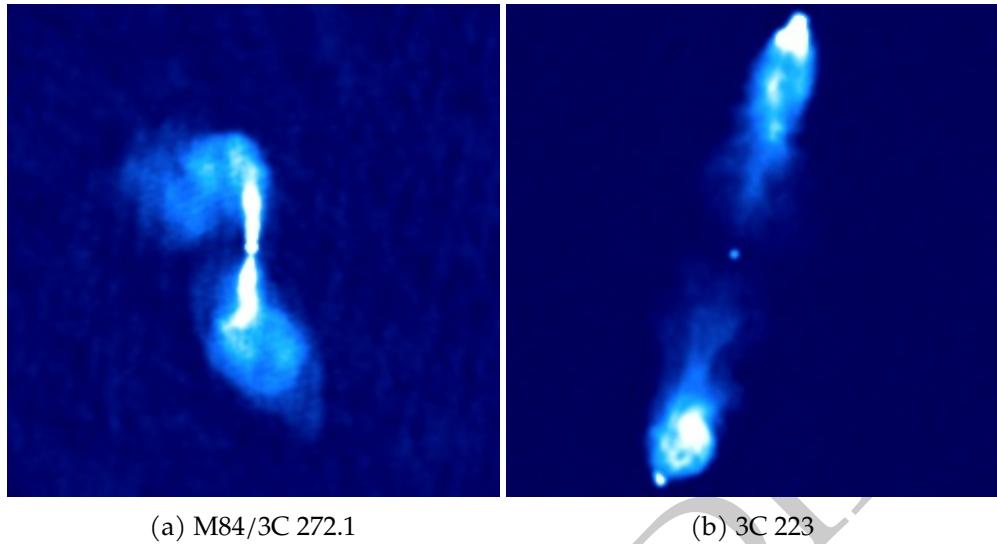


Figure 2.3: Examples of (a) a FRI (Laing & Bridle, 1987) and (b) a FRII radio galaxy (Leahy & Perley, 1991). Both are shown with an arcsinh stretch and were observed with the VLA.

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 2.3a). FRII, on the other hand, have long, tightly-collimated jets and sharp-edged lobes with bright hot-spots (Urry & Padovani, 1995) at the very end of the lobes, and are brighter further away from the host galaxy (e.g. Figure 2.3b). FRII are also generally higher-luminosity (Fanaroff & Riley, 1974) than FRI, and therefore make up the majority of observed extended radio sources throughout the Universe. However, 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 & Croston, 2020). 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 structure in Section 2.3.2. Appendix J lists some giants discovered during the work of this thesis.

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 (Bicknell, 1995; Ledlow & Owen, 1996) though more recent work suggests this may

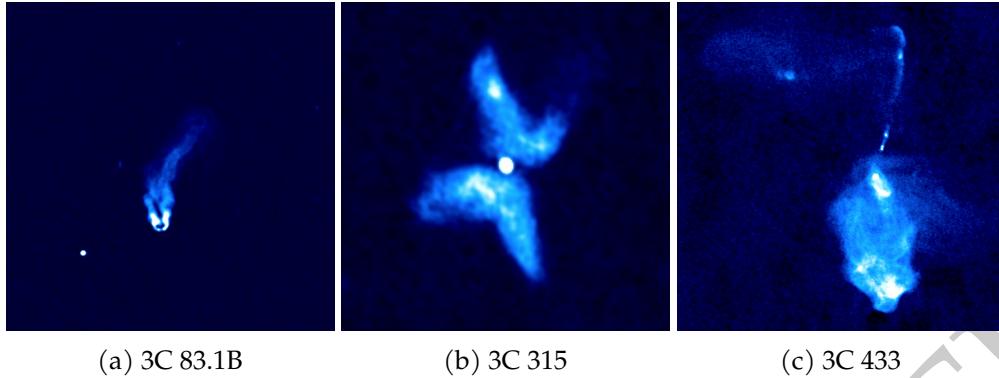


Figure 2.4: Radio galaxies, displayed with an arcsinh colour scale. All images were taken with the VLA. (a) is a narrow-angled tail radio galaxy (Leahy et al., n.d.), (b) is an X-shaped radio galaxy (Leahy et al., 1986), and (c) is a very unusually-shaped radio galaxy (Black et al., 1992).

not be a strong effect if it exists at all (Hardcastle & Croston, 2020). Chapter 5 investigates the distribution of radio luminosities conditioned on the mid-infrared colour of the host galaxy.

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 a much larger size than the radii of the host galaxies, and so the jets and lobes of AGN are uniquely poised 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 \(ISM; Mukherjee et al., 2016\)](#) [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., 2019; Rodman et al., 2019) and this structure may even be used as a probe for cluster environments (Banfield et al., 2016; Sakelliou et al., 2008).

The strong interaction of AGN with their [environment environments](#) 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 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 2.4. Large-scale automated identification of these galaxies can be tricky owing to their variety, extent, and often disconnected structure.

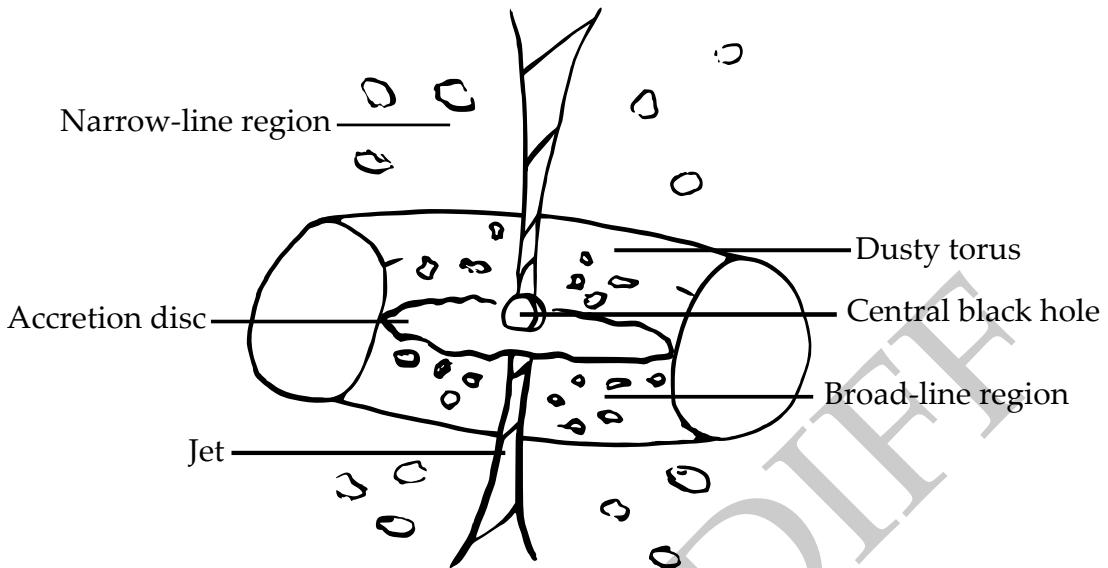


Figure 2.5: The unified model of AGN.

892 AGN cores tend to have flat or inverted spectral indices around -0.5–1 (Condon
 893 & Ransom, 2016; Randall et al., 2012). Moving out from the host galaxy, the spectral
 894 index steepens as the electrons are older and less energetic, with the spectral index
 895 of the lobes usually at about -0.7 (Condon & Ransom, 2016). The hot-spots of FRII
 896 galaxies have spectral indices between -0.5 – -0.7, becoming shallower as the electrons
 897 reaccelerate. These are only general trends: ~~the~~^{The} spectral structure within lobes
 898 can be very complex (Treichel et al., 2001). The jets do not strongly emit and are only
 899 detectable for particularly deep observations or nearby radio galaxies.

900 2.3.3 The unified model

901 At their core, AGN are an accreting *supermassive black hole*: a body so dense that even
 902 light cannot escape its gravitational pull, with mass on the order of 10^7 – $10^9 M_\odot$ (Mar-
 903 coni & Hunt, 2003). Such black holes seem to exist at the centres of galaxies and these
 904 galaxies are called *host galaxies*. The current understanding of the structure of an AGN
 905 is as follows (Urry & Padovani, 1995). The black hole is surrounded by an accretion
 906 disc emitting in ultraviolet and X-ray. Beyond this is the broad-line region, named for
 907 the Doppler-broadened emission lines emitted by the energetic clouds of material sur-
 908 rounding the accretion disc. The broad-line region and accretion disc are themselves
 909 surrounded by a dusty torus (or some other disc-like structure) which prevents light
 910 from the centre of the AGN being observed from the sides. Further still from the accre-
 911 tion disc is the narrow-line region, where lower-energy gas produces narrow emission
 912 lines. From either side of the disc, an AGN produces two collimated outflows of rel-
 913 ativistic plasma called jets, and these jets may interact with gas in the host galaxy to
 914 produce bright radio emission. The jets are not always visible. As the jets disperse fur-
 915 ther out from the centre of the AGN they widen into plumes of plasma known as *lobes*.

916 This model of AGN unifies different observed classes of AGN by their orientation and
 917 luminosity, and is hence known as the *unified model* (Antonucci, 1993). Recent work
 918 suggests that the unified model of AGN is not the full story (e.g. Zhuang & Ho, 2020).

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 & Best, 2014). Radiative-mode AGN produce radiative energy in amounts higher than 1 per cent 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 & Lightman, 2008):

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

919 where M is the mass of the compact object.

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

927 2.3.4 Polarised structure

928 The magnetic field of AGN is thought to be critical to their structure (Sikora & Begelman, 2013). A strong magnetic field is required to eject and collimate the jets (Lovelace,
 929 1976) and the magnetic environment influences the structure of the jets (O'Sullivan et
 930 al., 2015). Polarisation provides a probe for measuring this magnetic field. Radiative-
 931 and jet-mode AGN have different fractional polarisations, with jet-mode AGN having
 932 a much wider range of fractional polarisations ($p \sim [0, 30]$ per cent) compared to
 933 radiative-mode AGN (limited to $p \lesssim 15$ per cent), with this difference attributable to
 934 the magnetic environment (O'Sullivan et al., 2015). Steep-spectrum ($\alpha > 0.5$) and flat-
 935 spectrum ($\alpha < 0.5$) AGN have differing fractional polarisations, with steep-spectrum
 936 sources having much higher fractional polarisation for frequencies > 5 GHz and flat-
 937 spectrum sources having higher fractional polarisation for frequencies < 1 GHz due
 938 to frequency-dependent depolarisation of the steep-spectrum sources (Saikia & Salter,
 939 1988). **Hotspots** of FRII radio galaxies have low polarisation (< 10 per cent)
 940 while the more diffuse sections may have much greater polarisation (> 20 per cent)
 941 (Saikia & Salter, 1988). The direction of the magnetic field is correlated with the direc-
 942 tion of **patterns** in the total intensity of the source (Saikia & Salter, 1988).

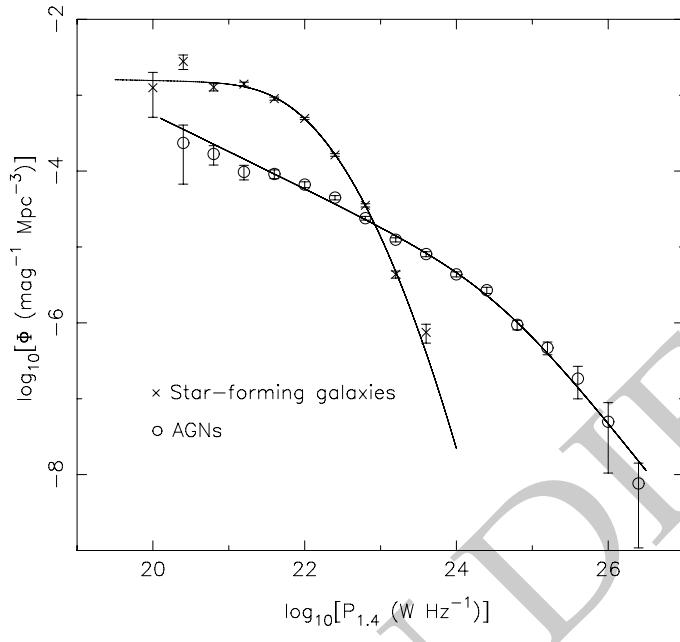


Figure 2.6: RLF for star-forming galaxies and AGN, from Mauch and Sadler (2007).

2.3.5 AGN luminosity

The distribution of AGN luminosities throughout the Universe can be characterised by the *radio luminosity function* (RLF). This describes the density of AGN within luminosity bins at a given frequency. The RLF at 1.4 GHz from Mauch and Sadler (2007) is shown in Figure 2.6. 1.4 GHz is arguably the most common large-scale survey frequency as it detects the easily observed 21 cm hydrogen line, but this RLF can be scaled to different frequencies by assuming a constant spectral index ($\alpha = -0.7$ for a typical AGN). In Chapter 5, we will use a machine learning-derived radio source catalogue to estimate a RLF for extended radio objects.

2.3.6 The role of AGN

AGN are an important part of galaxy evolution, though the exact mechanisms of this are not currently known. Understanding galaxy evolution requires understanding the links between it and AGN, and thus requires the study of AGN.

There are established relationships between AGN parameters and galaxy parameters: ~~the~~ The mass of the central black hole is correlated with the stellar mass and velocity dispersion near the core in elliptical galaxies (Cattaneo et al., 2009) as well as the luminosity of the bulge (Kormendy & Richstone, 1995). Massive galaxies seem to be more likely to contain an AGN rather than an inactive galactic nucleus (Hardcastle & Croston, 2020). The large-scale behaviour of AGN (e.g. the Fanaroff-Riley divide) may be caused by interactions within the host galaxy (Hardcastle & Croston, 2020) and this interaction would be a significant energy contribution to the ~~interstellar~~

965 medium ISM.

966 Perhaps the biggest topic in this field is that of AGN *feedback*, the role of AGN in the
 967 enhancement or quenching of star formation within their host galaxies. Star formation
 968 requires cold gas; AGN can push gas out of the galaxy (Zubovas & King, 2012) or heat
 969 it so that it can no longer form stars. But the rapid expansion of AGN jets can produce
 970 shocks in the local environment, compressing gas, triggering collapse of gas clouds,
 971 and potentially increasing the star formation rate (Zubovas et al., 2014).

972 For a review of the role of AGN within galaxy evolution, see Kormendy and Ho
 973 (2013). For a detailed review of the current understanding of AGN, see Hardcastle
 974 and Croston (2020).

975 2.4 Classifying AGN

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

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

995 2.4.1 Statistical and manual classification of AGN

996 Manual and statistical approaches to classifying AGN have dominated the radio as-
 997 tronomy literature until very recently, due to the comparative lack of computational
 998 power as well as a lack of good automated methods. Manual methods amount to ex-
 999 amining the structure of a resolved source and determining its class: ~~this~~. This is how
 1000 we usually identify bent radio galaxies, head-tail radio galaxies, X-shaped radio galax-
 1001 ies, and those radio galaxies with more unusual morphologies. Statistical approaches
 1002 identify properties of the source that can be combined and thresholded to separate
 1003 the sources into categories en masse. Modern machine learning techniques for classi-
 1004 fication of radio sources can be thought of as an extension of these statistical methods,

1005 where the properties and their combinations are identified automatically, but we will
1006 discuss these separately in Section 2.4.2.

1007 Arguably the most well-known radio classification scheme, FRI and FRII, was orig-
1008 inally defined on well-resolved radio galaxies by computing the ratio of the distance
1009 between the regions of highest brightness on opposite lobes and the total extent of the
1010 radio emission (Fanaroff & Riley, 1974). Sources with a ratio under 0.5 were called FRI
1011 and those with a ratio greater than 0.5 were called FRII. This classification has over time
1012 evolved into a less precise divide, with classification generally now morphological and
1013 based on the structure (diffuse, wavy plumes versus hot-spots and lobes for FRI and
1014 FRII respectively). The FRI and FRII divide has been further complicated by other re-
1015 lated categorisations such as the so-called “Fanaroff-Riley type 0” sources which seem
1016 to be the lower end of a continuum of radio sources with diffuse plumes (Capetti et al.,
1017 2020; Garofalo & Singh, 2019) as well as hybrid morphology radio sources (HyMoRS)
1018 which appear to be FRI on one side and FRII on the other (Gopal-Krishna & Wiita, 2000;
1019 Kapińska et al., 2017). Many classes are defined by explicitly statistical means; for ex-
1020 ample, steep- and flat-spectrum sources are divided by spectral index at $\alpha = 0.5$ (Urry
1021 & Padovani, 1995). For convenient analysis, radio sources are often also grouped into
1022 “observational” classes that don’t have a physical analogue based on their apparent
1023 structure, e.g. the GLEAM survey classifies radio sources into the number of apparent
1024 components, which is highly dependent on the observational parameters (White et al.,
1025 2020).

1026 More unusual or more loosely defined classes, such as X-shaped radio galaxies
1027 and giants, have often been identified by manual searches through large datasets, e.g.
1028 Cheung (2007), Dabhade et al. (2020) and notably the recent ROGUE I catalogue of
1029 32 616 morphologically classified radio galaxies (Żywucka et al., 2020). These searches
1030 are often aided by computer algorithms (e.g. Dabhade et al., 2020; Proctor, 2011).

1031 Radio sources are also more generally classified, such as into AGN or non-AGN
1032 emission (Koziel-Wierzbowska et al., 2020), often using optical emission lines or opti-
1033 cal/infrared magnitude.

1034 2.4.2 Machine learning classification of AGN

1035 Machine learning based approaches for radio source classification are rapidly evolv-
1036 ing as the amount of radio data available through big surveys increases. Advances
1037 in tooling, such as the wide availability of hardware-accelerated automatic differentia-
1038 tion software, have also contributed to an explosion in machine learning applications
1039 in astronomy by making machine learning techniques more available to astronomy
1040 researchers.

1041 Morphological classification of galaxies with machine learning began in optical as-
1042 tronomy, probably due to the large sample sizes of well-resolved galaxies previously
1043 available. The earliest such paper is likely the application of neural networks to the
1044 task by Storrie-Lombardi et al. in 1992. From here, the field applied other classifica-
1045 tion algorithms such as decision trees (e.g. Owens et al., 1996). The Sloan Digital Sky
1046 Survey (SDSS) brought an explosion of new data in 2003, and new experiments in clas-

1047 sification soon followed (e.g. Ball et al., 2006; Ball et al., 2004). The Galaxy Zoo project
 1048 leveraged hundreds of thousands of volunteers to produce an astonishingly large set
 1049 of labelled optical galaxies from SDSS and subsequent papers used this as a training
 1050 set for machine learning methods (Banerji et al., 2010; Dieleman et al., 2015; Zhu et al.,
 1051 2019).

1052 While machine learning has been used in radio astronomy for some time ([Condon et al., 1998](#), e.g. the NVSS
 1053 ([e.g. the NVSS used neural networks to detect sidelobes](#); [Condon et al., 1998](#)) its first ap-
 1054 plication to radio source classification was most likely to identifying quasar candi-
 1055 dates (Carballo et al., 2004) in Faint Images of the Radio Sky at Twenty Centimeters
 1056 (FIRST; Becker et al., 1995). Proctor (2006) applied decision tree ensembles to identify
 1057 bent double morphologies in FIRST, manually selecting features to characterise radio
 1058 sources, while Bastien et al. (2017) used shapelet analysis to obtain features to feed into
 1059 their decision tree ensembles. 2011–12 marked a revolution in computer vision with
 1060 the discovery that deep convolutional neural networks (known as early as 1989, see
 1061 LeCun et al., 1989), boosted dramatically by widely available training data generated
 1062 by the internet and a huge increase in computational power from GPUs, could achieve
 1063 greater-than-human performance on image classification tasks. Deep neural networks
 1064 have since found use for morphological classification of radio sources, such as FRI vs.
 1065 FRII (Aniyan & Thorat, 2017; Bowles et al., 2020; Lukic et al., 2019; Ma et al., 2019b;
 1066 Samudre et al., 2020; Tang et al., 2019; see also Ma et al., 2018), compact vs. extended
 1067 sources (Alhassan et al., 2018; Lukic et al., 2018; Lukic et al., 2019), and observational
 1068 classes (Galvin et al., 2019; Ralph et al., 2019).

1069 There are also many works on classification of radio sources besides morphology.
 1070 Machine learning has been applied to AGN classification tasks including blazar classi-
 1071 fication (Arsioli & Dedin, 2020) and radio loudness (Beaklini et al., 2020). Deep learn-
 1072 ing is also prevalent on this topic, with deep learning finding applications in Faraday
 1073 complexity classification (Brown et al., 2018) and notably in transient detection (Agar-
 1074 wal et al., 2020; Balakrishnan et al., 2020; Connor & van Leeuwen, 2018; Guo et al.,
 1075 2019; Lin et al., 2020; Wang et al., 2019; Zhang et al., 2020).

1076 It is worth contrasting these machine learning approaches with non-machine learn-
 1077 ing automated approaches, as the two are often conflated in the literature. Mingo et
 1078 al. (2019), for example, use an automated version of detecting the brightness gradi-
 1079 ent of extended radio sources to determine whether they are FRI or FRII en masse
 1080 and apply this approach to the LoTSS survey. Segal et al. (2019) apply an information
 1081 theoretic approach to estimating morphological complexity of a source. The key dif-
 1082 ference between a machine learning automated approach and a non-machine learning
 1083 automated approach is that the former has the capacity to change its behaviour based
 1084 on available data, while the latter does not—though note that this is not necessarily a
 1085 bad thing.

1086 2.5 Cross-identification

1087 *Cross-identification* is the problem of matching an observed astronomical object at one
1088 wavelength with its counterpart in a survey at some other wavelength. Sometimes
1089 the counterpart in question can be obvious, such as when the object being matched is
1090 compact and well-aligned in both wavelengths. This is not the case for radio, though:
1091 **radio**~~Radio~~ galaxies can be very extended and, due to environmental interactions, this
1092 extended structure may be complicated and have no obvious relationship to the galaxy
1093 at other wavelengths. The galaxy itself in radio cross-identification is called the *host*
1094 *galaxy* as it hosts the central AGN.

1095 2.5.1 Why do we need to cross-identify?

1096 Radio spectra do not generally contain much information besides the spectral index: as
1097 Jim Condon purportedly said, "There's nothing as useless as a radio source". Most in-
1098 formation about extragalactic sources is gleaned from other wavelengths, with optical
1099 and infrared showing physically meaningful differences due to emission and absorp-
1100 tion at these wavelengths. For this reason radio sources are typically cross-identified
1101 with their counterparts at other wavelengths to get information. Key features of a host
1102 galaxy that may be obtained from optical or infrared observations include its redshift,
1103 mass, star formation rate, and composition. From these we may also determine prop-
1104 erties of the radio plasma, most notably its physical extent and radio luminosity, both
1105 of which require knowing the redshift.

1106 2.5.2 Methods for cross-identification

1107 Most small radio surveys are cross-identified by astronomers, visually comparing the
1108 radio image to the optical or infrared image (e.g. Middelberg et al., 2008; Norris et
1109 al., 2006). The largest such catalogue is the ROGUE I catalogue with over 32 000
1110 cross-identifications (Żywucka et al., 2020). The Radio Galaxy Zoo project (Banfield
1111 et al., 2015) extended manual cross-identification to larger scales by crowdsourcing,
1112 asking volunteers online to cross-identify FIRST and ATLAS resulting in 75 000 cross-
1113 identified extended sources. This is the largest manually cross-identified source cat-
1114 alogue ever, by citizen scientists rather than professional astronomers—though this
1115 distinction does not seem to matter as the accuracy of cross-identifications in Radio
1116 Galaxy Zoo is still quite high, and there is an upper limit to how well even astronomers
1117 can cross-identify large surveys (Banfield et al., 2015; Wong et al., in prep.).

1118 2.5.2.1 Positional matching

1119 The simplest automated method for cross-identification is simply to look for the closest
1120 optical or infrared galaxy to some radio emission, a strategy called *positional matching* or
1121 *nearest neighbours*⁴. The distance to the nearest potential host galaxy is usually limited

⁴Not to be confused with the machine learning algorithm also called 'nearest neighbours', which does not make an appearance in this thesis.

to some small value, so that distant spurious relations are excluded. This distance is usually about $5''$ for infrared and $1''$ for optical but may be much further. Some authors modify this approach ~~to~~by selecting the nearest *bright* galaxy or adding some other constraints (e.g. Kimball & Ivezić, 2008). This approach generally works quite well for compact objects but extended radio structures may overlap with unrelated galaxies on the sky, leading to spurious cross-identifications.

The rate of spurious identification with this approach can be estimated by a simple model, assuming that the true host galaxy is not visible (either too faint or not nearby). Assume that a small circular region of the sky with radius a contains K potential host galaxies. The probability that a randomly selected potential host galaxy is within r rad of a given point is r^2/a^2 , with $r < a$. The probability of no potential host galaxy falling within r rad of a given point is therefore $(1 - r^2/a^2)^K$ and the probability of coincidental cross-identification is $1 - (1 - r^2/a^2)^K$. The AllWISE survey contains 747 634 026 infrared sources, over 4π sr, so it has an average source density of approximately $6 \times 10^7 \text{ sr}^{-1}$. With $r = 5''$, and $a = 1^\circ$ (corresponding to an area of $9.57 \times 10^{-4} \text{ sr}$ and thus 57 420 sources), we have a coincidental cross-identification rate of 10.5 per cent when the host galaxy is not visible.

2.5.2.2 Other automated methods

With very few non-trivial cross-identification methods in existence, the field is still very much in its infancy. Other automated methods for cross-identification can be grouped into machine learning and non-machine learning approaches. The two non-machine learning approaches are a Bayesian geometric model fitting approach (Fan et al., 2020; Fan et al., 2015) and a likelihood ratio approach (Weston et al., 2018; Weston, 2020). In Chapter 4 we introduce the very first machine learning approach for radio-infrared cross-identification (Alger et al., 2018). Galvin et al. (2020) later developed a hybrid machine learning and manual labelling approach to cross-identification.

2.6 Aggregating Radio Components

Unlike emission from galaxies observed at other wavelengths, the extended emission from radio galaxies can be disconnected when observed: ~~a~~A single radio galaxy may appear in observations as multiple discrete components. This is partly due to inhomogeneous emission over the radio galaxy structure—e.g. FRII hot-spots can be much brighter than the rest of the galaxy, so they ~~might~~may 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 our paper (Alger et al., 2018) based on the Radio Galaxy Zoo terminology: ~~a~~A *radio source*



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

(b) ATCA near Narrabri

Figure 2.7: (a) A single-dish telescope and (b) an array. Images: CSIRO.

1162 or *radio galaxy* refers to all radio emission observed that comes from a single physical
 1163 object, and a *radio component* refers to a single, contiguous region of radio emission
 1164 above the noise level of an observation. ~~We will avoid using *radio object*.~~ Note that in
 1165 the literature a *radio component* may be either a Gaussian fit to an observation or it may
 1166 be a region of connected emission; in works where the former is the case (particularly
 1167 in source-finding literature) the latter is often called a *radio island*.

1168 Since radio galaxies can appear disconnected, aggregating observed radio com-
 1169 ponents into physical sources is integral to understanding radio galaxies. Aggregat-
 1170 ing disparate components into a single source is important for two reasons. Firstly,
 1171 without all components, the total luminosity of a source is dramatically underesti-
 1172 mated. Secondly, we need all components to accurately examine the morphology of
 1173 the source.

1174 **2.6.1 Missing emission in radio observations**

1175 Radio observations ~~may be made either with~~ are made with either single-dish telescopes,
 1176 like the famous Parkes Radio Telescope (Murriyang), or radio arrays, like the Australia
 1177 Telescope Compact Array (ATCA), both shown in Figure 2.7. Both have their advan-
 1178 tages. Single-dish telescopes are able to measure absolute brightnesses (while arrays
 1179 can only measure relative brightnesses, and must therefore be calibrated to a source of
 1180 known brightness). Interferometric arrays can achieve incredibly high resolution, as
 1181 the resolution is inversely proportional to the distance between the most distant array
 1182 elements (while the resolution of single dish telescopes is inversely proportional to the
 1183 diameter of the dish).

1184 Radio telescopes can be thought of as sampling the *u-v plane*, the Fourier transform
 1185 of the sky. The *u-v plane* is perpendicular to the line-of-sight. Each pair of antennae in
 1186 an array samples two points on this plane, each corresponding to the vector between
 1187 the antennae projected onto the *u-v plane*, called a *baseline*. Longer baselines there-
 1188 fore correspond to higher (spatial) frequencies, which is why long baselines provide

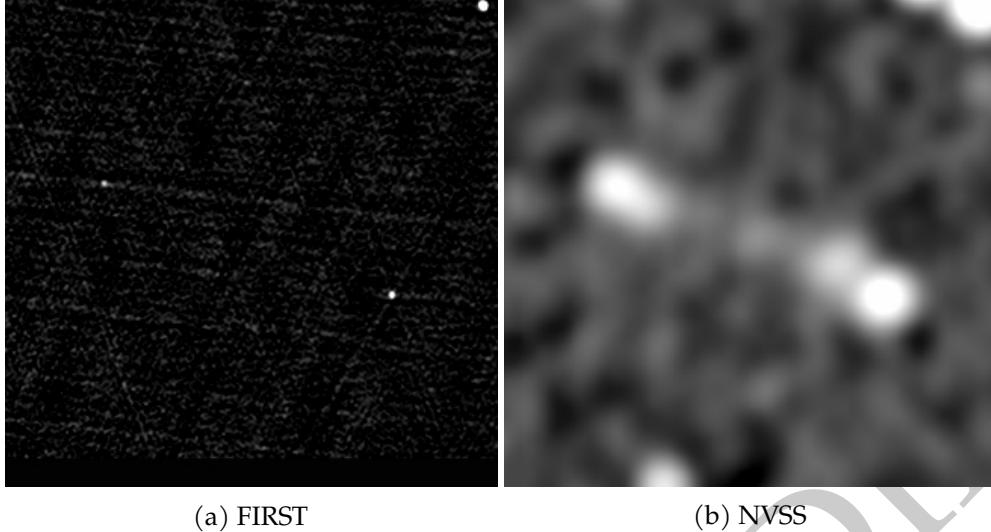


Figure 2.8: A fairly diffuse FRII, J0016+0420, observed with the VLA in the (a) FIRST (Becker et al., 1995) and (b) NVSS (Condon et al., 1998) surveys. (GRG1 from Dabhade et al., 2017)

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*. This means that large angular scale, 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 (*small σ*) would be quite broad, taking up many frequencies in the u - v plane, while a very *diffuse* Gaussian (*large σ*) 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 origin⁵. This loss of larger scale diffuse emission is often called *resolving out*. An example of this is shown in Figure 2.8, 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 *the* 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.

⁵This is, incidentally, why single-dish telescopes can measure the absolute brightness while arrays cannot: ~~there~~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.

1195 **2.6.2 Methods of aggregation**

1196 Most large radio catalogues are of *components* rather than sources, most likely because
1197 manual methods remain the best way to aggregate components into sources. Many
1198 component catalogues have source catalogues of some interesting subset later derived
1199 from them by manual inspection (e.g. Dabhadé et al., 2017). Such manual searches
1200 are often assisted by automated methods (e.g. Proctor, 2011). The Radio Galaxy Zoo
1201 citizen science project has also, with the help of over 10 000 volunteers, aggregated
1202 over 75 000 sources from FIRST (Wong et al. in prep.) and ATLAS (Banfield et al.,
1203 2015), which is one of the largest (if not the largest) extant catalogues of manually
1204 aggregated sources.

1205 Automated methods have been developing steadily as data volumes ~~have grown~~ grow.
1206 A Bayesian approach, fitting models of sources to component locations, was intro-
1207 duced by Fan et al. (2015) and later expanded upon (Fan et al., 2020). Machine learn-
1208 ing methods have also become increasingly popular. Proctor (2016) aggregated radio
1209 components using a decision tree classifier and used the results to search for giant
1210 radio galaxies (see also Proctor, 2006, 2011). Wu et al. (2019) applied a deep learn-
1211 ing model to identify radio sources from an image of radio components, training this
1212 model on the Radio Galaxy Zoo FIRST dataset. Most recently Galvin et al. (2020) used
1213 a nested self-organising map, a kind of neural network, to cluster similar sources and
1214 then aggregated them into sources using manual labelling of the clusters and some
1215 heuristics.

1216 Any method for cross-identification of components can also be used to aggregate
1217 components. Cross-identify all components with their host galaxies, then components
1218 that share a host are almost always part of the same radio source. It is this approach
1219 ~~that~~ we will later take to aggregating sources in Chapter 5. There is no obvious in-
1220 verse algorithm (to go from aggregate sources to cross-identifications) although some
1221 algorithms attempt to solve both problems simultaneously (Fan et al., 2020; Fan et al.,
1222 2015).

1223 **2.7 Summary: radio sources**

1224 The radio sky is vibrant and exciting, and current and new wide-area radio surveys are
1225 opening up new avenues of exploring this space. In this chapter we introduced radio
1226 sources that you might see in the radio sky, including the often-extended AGN. We dis-
1227 cussed their radio emission and emission mechanisms, and described some aspects of
1228 their morphology and structure. We also looked at ways that radio sources can be clas-
1229 sified, aggregated, and cross-identified with their counterparts at other wavelengths.
1230 In the next chapter, we will introduce machine learning concepts that we will use to
1231 develop methods for exploiting the vast amount of data ~~that~~ radio telescopes provide
1232 on radio sources.

REVISION DIFF

1234 **Machine Learning for
1235 Astroinformatics**

1236 Machine learning was once described to me by an anonymous supervisor as “the statistics
1237 kept at the back of the textbook”. But even accepting its grounding in statistics,
1238 is this really an accurate description of the field? I think of machine learning as a
1239 combination of three pieces: a data-driven way of formalising predictive problems and
1240 a way of converting between different kinds of statistical problems, as well as an
1241 accompanying and a set of methods and practices for handling data and uncertainty.
1242 The eventual goal is to design some method or algorithm that automatically discov-
1243 ers useful patterns in (potentially very large) data-sets datasets. There are three core
1244 components of machine learning: the data, the model, and learning (Deisenroth et al.,
1245 2020). Before discussing these, we will look at the kinds of problems that machine
1246 learning solves.

1247 **3.1 Prediction**

1248 Machine learning aims to solve *prediction tasks*: problems where we have some data
1249 and we seek some kind of output based on that data. Central to prediction tasks are
1250 predictors, the objects we train based on data.

1251 **3.1.1 Predictors**

1252 A *predictor* is an object that makes predictions based on an input. A predictor can be a
1253 function or a probabilistic model, depending on the machine learning approach being
1254 undertaken.

As a function, a predictor maps from some input domain \mathcal{X} into some output do-
 main \mathcal{Y} , and is usually written as

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

1255 \mathcal{X} and \mathcal{Y} are commonly (but certainly not always) a real vector space \mathbb{R}^n . Because the
1256 goal of machine learning involves finding a suitable function f for the task at hand, the
1257 set of functions is usually constrained. For example, if $\mathcal{X} = \mathbb{R}^n$, we might require that

1258 f is a linear function $\mathbb{R}^n \rightarrow \mathbb{R}$, easily parametrised by $n + 1$ constants. This constraint is
1259 called a *model*. Models are often (but not always) parametrised and training the model
1260 refers to finding a good set of parameters. The parameters are sometimes called *weights*.
1261 To avoid ambiguity, parameters of the model that are not changed during training are
1262 called *hyperparameters* and may include regularisation terms, constraints, or variations
1263 of the model.

1264 Some predictors can be described as a probabilistic model. In this case a predictor is
1265 a joint probability distribution between observations and hidden parameters (Deisen-
1266 roth et al., 2020). Using a probabilistic predictor allows us to formally describe and
1267 work with uncertainty both in the input space and output space. Such a predictor is
1268 usually parametrised by a finite set of parameters, which already includes most com-
1269 mon probability distributions.

1270 We ~~will~~ generally assume that our data are generated from some unobserved, true
1271 function called the *groundtruth*. This might be a physical process, or a complicated
1272 sampling function from some unknown vector space. The assumptions we make on
1273 this generative function can greatly change the way we approach machine learning
1274 problems.

1275 In some sense, the goal of machine learning is to identify a good predictor from
1276 within the space of all possible predictors. Of course, this begs the question: ~~what~~
1277 What is a ‘good’ predictor? We will return to this when we discuss learning, but for
1278 now, a good predictor is one that approximates the groundtruth well.

1279 3.1.2 Classification

1280 *Classification* is the machine learning task of predicting discrete, unstructured values
1281 (Deisenroth et al., 2020). These values are called *classes*. Classification is arguably the
1282 most important prediction task, as many other problems can be formalised as classi-
1283 fication. Astronomy has its fair share of classification tasks, from classical astronomy
1284 tasks like galaxy morphology classification (appearing in machine learning literature
1285 as e.g. Dieleman et al., 2015) to transient detection (e.g. Scalzo et al., 2017); see Sec-
1286 tion 2.4 for more examples.

1287 A classification problem seeks a predictor where \mathcal{Y} represents a finite, discrete set
1288 of classes. Classification tasks are usually delineated by the number of classes: ~~there~~
1289 There are either two classes or more than two classes. The former are called *binary*
1290 *classification* tasks and the latter are *multiclass classification* tasks. The reason for this
1291 split is that binary classes are dramatically easier to reason about and analyse, and
1292 many special cases exist for binary where they do not for multiclass.

1293 \mathcal{Y} for a binary task is usually represented as $\mathcal{Y} = \{0, 1\}$. 1 is called the *positive class*;
1294 0 is called the *negative class*.

1295 An easy way to see why many tasks can be formalised as classification can be found
1296 by taking any prediction problem $\mathcal{X} \rightarrow \mathcal{Y}$ and reinterpreting it as the binary classifica-
1297 tion problem $\mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$, i.e. instead of taking an input and predicting an output,
1298 take an input and a potential output and determine if they should be related. Of course
1299 this is not always the most efficient way to solve a prediction problem but the many

1300 known properties of classification make it an appealing framework to cast problems
1301 into. In Chapter 4, we will cast the radio astronomy problem of cross-matching galaxies
1302 seen in different wavelengths into a binary classification problem, and in Chapter 6
1303 we will classify radio observations as Faraday complex or Faraday simple.

1304 There are three different formalisms for the output of a binary classifier. The output
1305 may be a *score*, a *class probability estimate*, or a *predicted class*. A predicted class is what
1306 first comes to mind as a representation of binary classes: ~~the~~^{The} classifier outputs ei-
1307 ther 0 or 1, with no way of representing ambiguity. A class probability estimate is a
1308 number from 0 to 1 which represents the probability that an example is in the positive
1309 class, i.e. if f is a classifier then $f(x) = p(y = 1 | x)$. Class probability estimates may
1310 be output by probabilistic classifiers, or may be estimated from non-probabilistic classi-
1311 fiers. Finally, a score is a value that tends to be larger for elements of the positive class,
1312 with less ambiguous examples being given a higher score. Scores are often converted
1313 into class probability estimates by applying a sigmoid (see Section 3.4.1). There are
1314 multiple ways to generalise these concepts to multiclass classification. Having a score
1315 or a class probability estimate can be more useful than only having a predicted class.
1316 For example, you could sort your examples by how likely they are to be a member
1317 of the positive class. This underpins our approach to casting cross-identification as
1318 binary classification in Chapter 4.

1319 3.1.3 Regression

1320 The other main kind of supervised prediction task is *regression*, which is the machine
1321 learning task of predicting ordered (and usually continuous) values. In a regression
1322 problem, we seek a predictor where \mathcal{Y} is a set of ordered values, usually a subset of
1323 \mathbb{R}^k for some positive natural k . Regression is ubiquitous in astronomy, from simple
1324 linear relationships like the famous Tully-Fisher relation (Tully & Fisher, 1977) to es-
1325 timation of redshifts from photometric observations (called *photometric redshifts*; first
1326 introduced by Baum, 1962). ~~We will~~^{This thesis does} not directly address any regres-
1327 sion problems~~in this thesis, but we will~~^{but does} make use of their results.

1328 3.2 Data and representation

1329 Machine learning is centred on data and the extraction of useful information from that
1330 data. Data can include anything from numeric information, documents, or images,
1331 to spectra or galaxies. A collection of data is called a *dataset* and an element of this
1332 dataset is (interchangably) called an *example* or *instance*. Generally, data are not easy to
1333 work with in their original form and must be converted into a numerical representation
1334 before use. We usually convert our data into real vectors in \mathbb{R}^n as it is relatively easy
1335 to work with both numerically and analytically. Each axis of this vector space is called
1336 a *feature* and the space as a whole is called the *feature space*. Features are non-trivial
1337 to choose, and finding good features often requires the expertise of a human who is
1338 well-versed in the original dataset (a *domain expert*). The process of finding features is
1339 called *feature selection*, *feature design*, or *feature engineering*.

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, easy to extract, behave nicely, and work well with our model. 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 finding good methods to automatically develop 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 et al., 2015) can be viewed as developing their own task-specific features as part of their training (see Section 3.4.3). These deep learning features can be useful (e.g. when it is hard or slow to define good features manually), but are usually hard to interpret. A good feature space ~~will have has~~ 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.

Labels comprise another very important piece of the machine learning puzzle. 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 has~~ only two possible outputs. Common representations for ~~this would be these are~~ $\{0, 1\}$ 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-In~~ essence, features are cheap and labels are expensive. We will discuss labels in more detail in Section 3.5.

3.3 Loss functions

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.

Put simply, a loss function is a function of a predictor and a dataset which is chosen to be a proxy for how good the predictor is at predicting that dataset. We try to choose loss functions that are high-valued for a predictor that poorly describes the dataset, and are low-valued for a predictor that well-describes the dataset. Sometimes (and

in both cases listed in this section) the loss is minimised at zero, when the predictor perfectly captures the dataset (though whether this is possible, or whether this is even 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. The loss function we derive depends on the model we assume for the noise; for example, assuming Gaussian noise gives us the common least-squares loss.

3.3.1 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.

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

for the true mean μ and standard deviation σ where \mathcal{N} is the normal distribution:

$$\mathcal{N}(a | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}. \quad (3.3)$$

Under this assumption the probability that y is equal to a target t given an example x is

$$p(y = t | x) = \mathcal{N}(t | f(x), \sigma^2). \quad (3.4)$$

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

$$p(Y = T | X) = \prod_{i=1}^n p(y_i = t_i | x_i) \quad (3.5)$$

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

$$p(Y = T | X) = \prod_{i=1}^n \mathcal{N}(t_i | f(x_i), \sigma^2). \quad (3.6)$$

$p(Y | X)$ is called the *likelihood*. We would like to maximise this likelihood 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 likelihood is equivalent to minimising its negative logarithm, so we write:

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

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:

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

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:

$$\mathcal{L}(f; T, X) = \frac{1}{2} \sum_{i=1}^n (t_i - f(x_i))^2. \quad (3.9)$$

The factor of half helps keep the derivative tidy:

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

1403 3.3.2 Loss function for binary classification

Now we will calculate a loss function for binary classification. As for regression, we first assume a form for the noise. Assume that our labels are a random variable $y \in \{0, 1\}$ and that the prediction y is drawn from a Bernoulli distribution based on a predictor $f(x)$:

$$p(y = t | x) = \mathcal{B}(t; f(x)). \quad (3.11)$$

The Bernoulli distribution is parametrised by one parameter, usually called $p \in (0, 1)$, and in this case set to $f(x)$. It is:

$$\mathcal{B}(a; p) = p^a (1-p)^{1-a}. \quad (3.12)$$

It can be thought of as a biased coin toss with a probability p of tossing heads. To gain some intuition into how this expression works, imagine setting a to 0 and then to 1. Continuing to derive the loss function, we once again determine the likelihood

making the IID assumption:

$$p(Y = T \mid X) = \prod_{i=1}^n p(y_i = t_i \mid x_i) = \prod_{i=1}^n f(x_i)^{t_i} (1 - f(x_i))^{1-t_i}. \quad (3.13)$$

Then we find the negative log-likelihood and hence what is known as the *binary cross-entropy loss* for binary classification:

$$\mathcal{L}(f; T, X) = - \sum_{i=1}^n \log \left(f(x_i)^{t_i} (1 - f(x_i))^{1-t_i} \right) \quad (3.14)$$

$$= - \sum_{i=1}^n t_i \log f(x_i) + (1 - t_i) \log(1 - f(x_i)). \quad (3.15)$$

1404 This is the standard loss used for logistic regression (Section 3.4.1).

1405 **3.3.3 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

$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \lambda \nabla_{\vec{w}} \mathcal{L}(\vec{w}^{(k)}; T, X) \quad (3.16)$$

1406 where $\lambda > 0$ is a small scalar called the *learning rate*. With an appropriately small choice
1407 of λ \vec{w} will converge to a local minimum of \mathcal{L} . Many variations on this concept exist
1408 which attempt to avoid local minima, such as introducing a ‘momentum’ term that
1409 accumulates as multiple iterations move \vec{w} in the same direction. If the loss function
1410 is convex, then any minimum is the global minimum (there are no local minima).

1411 **3.4 Models**

1412 This section describes some common models for classification. There are a plethora of
1413 different classification models and variations on these models, but I will present here
1414 only those relevant to this thesis: logistic regression, decision tree ensembles, and neu-
1415 ral networks. These are, not coincidentally, also the most common models in astroinfor-
1416 matics. Logistic regression provides reliable and interpretable results. Decision tree
1417 ensembles are a fantastic off-the-shelf choice which work on a large variety of datasets.
1418 Neural networks have ~~proved~~-proven extremely effective for a wide variety of tasks,
1419 especially in computer vision.

1420 **3.4.1 Logistic regression**

1421 *Logistic regression* is a linear, binary, probabilistic classifier. Linear classifiers can only
 1422 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. Probabilistic classifiers, as discussed in Section 3.1.2, have outputs interpretable as class probabilities.

1425 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:

$$f(x; w) = \sigma(w^T x) \quad (3.17)$$

where σ is the *logistic function* or *sigmoid*, which is a monotonic and bijective function:

$$\sigma(a) = \frac{1}{1 + e^{-a}}. \quad (3.18)$$

1426 The output of logistic regression applied to an example x is the probability that x is
 1427 in the positive class. σ , and thus logistic regression, has a domain of $(-\infty, \infty)$ and a
 1428 range of $(0, 1)$. This enforces the output to be like a probability. $w^T x = 0$ defines a
 1429 d -dimensional hyperplane, called the *separating hyperplane* or *decision surface*. Logistic
 1430 regression is differentiable, which allows us to optimise its parameters w using
 1431 gradient descent. Interpreting the classifier is possible through examining the weights
 1432 vector, with a larger absolute value of a weight corresponding to a ‘more important’
 1433 feature. The weights from logistic regression can be used to produce scores or class
 1434 probability estimates (Section 3.1.2): $w^T x$ is a score, $\sigma(x^T x)$ is a class probability esti-
 1435 mate, and thresholding either the score or the class probability estimate at 0 gives a
 1436 discrete class.

1437 A limitation of logistic regression is its sensitivity to scale. Features need to be of
 1438 approximately the same order of magnitude and should have a standard deviation of
 1439 approximately 1. An implicit assumption is that each ~~features~~ feature has a mean of 0
 1440 across the dataset. This can be enforced by normalising and scaling: subtract the mean
 1441 of the dataset and divide by the new standard deviation.

1442 We apply logistic regression to binary classification problems in Chapter 4 and
 1443 Chapter 6.

1444 **3.4.2 Decision tree ensembles**

1445 A *decision tree* is a non-linear classifier. It repeatedly splits a dataset based on binary
 1446 comparisons until every subset contains only one class (or mostly one class, with the
 1447 amount of purity left as a hyperparameter). Each split only uses one feature for the
 1448 comparison, making decision trees relatively easy to visualise and interpret. However,
 1449 because of this, each split is axis-parallel, which can be a limitation for some datasets.
 1450 They are not sensitive to scale and do not require a zero mean, making them easy to
 1451 apply without preprocessing a dataset.

1452 Key limitations of a decision tree are:

- 1453 • They can only output a prediction, not a confidence of this prediction or a score
1454 of how likely an instance is to be found within each class.
- 1455 • Small changes to the dataset or training method can result in large changes to
1456 the tree.
- 1457 • They tend toward overfitting the training set.
- 1458 • With many low-information features, decision trees have quite poor performance
1459 (Breiman, 2001).

1460 A *decision tree ensemble* aims to reduce some of these limitations by training mul-
1461 tiple, slightly different, independently-trained decision trees. Depending on the im-
1462 plementation each constituent decision tree may only have access to some of the fea-
1463 tures or some of the data. To predict, each tree produces a prediction and ‘votes’ for
1464 this prediction; the votes ~~can be~~ are combined to produce the overall prediction (e.g.
1465 with majority voting). A simple example of such an ensemble is decision tree bagging
1466 (Breiman, 1996), which trains each tree with a random subset of the training data and
1467 takes a plurality vote. Decision tree ensembles decrease variance, increase the usabil-
1468 ity of low-information features, and increase stability of the trained model (Breiman,
1469 2001). Averaging over the classifications produced by each tree gives a number that
1470 can be interpreted as a class probability estimate.

1471 The most well-known description of decision tree ensembles is the *random forest*
1472 (Breiman, 2001), which has found common use in astronomy partly to its readily avail-
1473 able Python implementation in `scikit-learn` (Pedregosa et al., 2011). Splits are de-
1474 cided from a subset of features and training samples are randomly drawn with replace-
1475 ment from the total training set. One downside of random forests is the large number
1476 of hyperparameters that need to be set, and these vary a lot depending on the problem
1477 being addressed.

1478 We apply decision tree ensembles to binary classification problems in Chapter 4
1479 and Chapter 6.

1480 3.4.3 Convolutional neural networks

A *neural network* is a directed graph of transformations, each node representing a trans-
formation that linearly combines its inputs and applies a non-linear function called the
activation function to the result. The inputs to the graph are the features. A particularly
prominent kind of neural network is the *fully-connected neural network*, where nodes
are arranged into layers, with each node in a layer taking as input every output from
the previous layer. Each layer can then be represented by a matrix multiplication of
the outputs of the previous layer by a weight matrix, composed with the activation
function. Fully-connected K -layer neural networks have the form:

$$f(x; W_K, \dots, W_1) = h_K(x; W_K, \dots, W_1) \quad (3.19)$$

$$h_i(x; W_i, \dots, W_1) = a(W_i h_{i-1}(x; W_{i-1}, \dots, W_1)) \quad (3.20)$$

$$h_1(x; W_1) = a(W_1 x) \quad (3.21)$$

1481 where a is the activation function. h_i are called *hidden layers*. In fact, neural networks
 1482 are usually described by their layer structure rather than graph structure, with the
 1483 addition of ‘concatenation layers’ to combine outputs from previous layers. Neural
 1484 networks may be used for regression or for classification; these are structured the same
 1485 but for classification the last activation function is replaced by sigmoid (for binary
 1486 classification) or its multiclass counterpart softmax.

1487 *Convolutional neural networks* (CNN; LeCun et al., 1998) are ~~a-variant~~variants of
 1488 neural networks that are particularly well-suited to inputs that have local structure,
 1489 such as images or spectra. Layers in the network may be *dense layers* of the same form
 1490 as Equation 3.20, or *convolutional layers*, where the weights are convolved with the input
 1491 rather than multiplied. These convolutional weights are called *filters* and they are small
 1492 compared to the dimensionality of the input. CNNs are translation-invariant (Waibel
 1493 et al., 1989) and derive features from local relationships thanks to the trainable filters.

1494 We apply CNNs to binary classification problems in Chapter 4 and Chapter 5, and
 1495 compare our results to a state-of-the-art CNN in Chapter 6.

1496 3.5 Labels

1497 As described in Section 3.2, labels are the known outputs of supervised prediction
 1498 tasks. They are used for two main, distinct purposes: training and validation. Labels
 1499 for training are used to evaluate loss and determine how to update the model. Labels
 1500 for validation are used to evaluate and characterise the model’s behaviour.

1501 3.5.1 Where do labels come from?

1502 We previously said that labels were ‘expensive’. This is to be interpreted as expensive
 1503 in either or both time and money: ~~labelling~~Labelling can be a slow, manual~~process~~,
 1504 ~~and labelling can be costly~~. ~~Labelling~~and process. It is usually completed by hand,
 1505 manually examining instances and determining ~~what~~which class they belong to (for
 1506 classification) or what target they ought to be associated with (for regression). In
 1507 astronomy this usually amounts to expert astronomers examining imagery at multi-
 1508 ple wavelengths and making an educated guess as to what the true label ~~should be~~is,
 1509 but labelling may also involve follow-up observations (perhaps at higher resolution,
 1510 greater sensitivity, or at a different wavelength).

1511 An increasingly popular option for labelling large amounts of data is *citizen science*:
 1512 asking volunteers who are interested in contributing to science to label our data. Citizen
 1513 science projects can be a fantastic opportunity for both science and outreach. For
 1514 example, the ABC’s ‘Stargazing Live’ television programme engaged viewers and with
 1515 their help found four exoplanets in just 48 hours (Miller, 2017) and labelled 120 000
 1516 SkyMapper images¹ in just three days (Tucker, 2017). The downside of citizen science
 1517 is that non-expert labellers may be less accurate than experts, and indeed some may
 1518 even be malicious and provide intentionally incorrect labels (Zhang et al., 2016).

¹Citizen scientists actually produced around 5 million labels—these were aggregated to 120 000 to reduce noise.

1519 Astronomers often face a large collection of unlabelled data and must choose which
1520 to label. Choosing what to label is a broad topic of research separately in machine
1521 learning (often called active learning e.g. Gilyazev & Turdakov, 2018), in astronomy
1522 ('follow-up observations'), and in citizen science project design (e.g. citizen science
1523 project Snapshot Serengeti found that showing volunteers *uninteresting* images helped
1524 retain engagement; Sieland, 2015).

1525 **3.5.2 Label noise**

1526 *Label noise* is the presence of incorrect labels in the training or validation data set. In
1527 classical machine learning there is no such thing: ~~labels~~ Labels are assumed to come
1528 from some always-correct 'oracle'. In reality, though, labels can be wrong. There is
1529 intrinsic noise in data, and even expert astronomers ~~can~~ may disagree on labels due
1530 to ambiguities (e.g. around 10 per cent of Radio Galaxy Zoo is extremely divisive
1531 amongst expert labellers; Banfield et al., 2015). All is not lost for machine learning:
1532 ~~many~~ Many optimisation targets are robust to label noise (Menon et al., 2015). One
1533 way to think about this is that the loss function for machine learning 'smooths over' or
1534 'averages out' the noise.

1535 It is important to note that label noise affects training and validation differently.
1536 While it is perfectly possible to train a good model with noisy labels, performance
1537 measures are not as robust to label noise. Noise in the validation set can change the re-
1538 ported performance in unpredictable ways and wherever possible should be avoided.

1539 **3.6 Summary: machine learning**

1540 Machine learning is a field of study concerned with ways of automatically finding use-
1541 ful patterns in large datasets, formalising predictive problems, converting between dif-
1542 ferent statistical problems, and handling data and uncertainty. With the large volumes
1543 of data radio astronomy faces in the lead-up to the SKA, machine learning provides
1544 an avenue for enabling scientific discovery at scale. In this chapter we introduced key
1545 concepts of machine learning, including predictors, features, labels, and loss functions.
1546 We also highlighted classification tasks as a key kind of prediction problem that we will
1547 use to frame astronomical questions throughout this thesis and introduced three clas-
1548 sification models: logistic regression, decision tree ensembles, and CNNs. In the next
1549 chapter, we will frame the astronomical problem of cross-identification as binary clas-
1550 sification, allowing us to train a cross-identification algorithm using an existing cata-
1551 logue of cross-identifications and hopefully allowing us to cross-identify radio sources
1552 en masse.

REVISION DIFF

1554

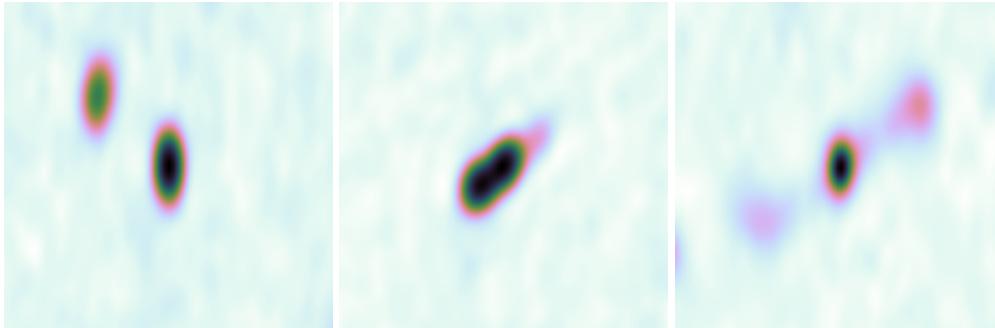
Radio Cross-identification

1555 This chapter is based on my paper *Radio Galaxy Zoo: Machine learning for radio source*
1556 *host galaxy cross-identification*, by M. J. Alger, J. K. Banfield, C. S. Ong, L. Rudnick, O. I.
1557 Wong, C. Wolf, H. Andernach, R. P. Norris, and S. S. Shabala; published in the *Monthly*
1558 *Notices of the Royal Astronomical Society* in 2018. Some minor changes have been made
1559 to match the rest of this thesis.

1560

1561 In this chapter, we consider the problem of determining the host galaxies of radio
1562 sources by cross-identification. This has traditionally been done manually, which will
1563 be intractable for upcoming and ongoing wide-area radio surveys like the Evolutionary
1564 Map of the Universe (EMU). Automated cross-identification will be critical for these
1565 future surveys, and machine learning may provide the tools to develop such methods.

1566 We applied a standard approach from computer vision to cross-identification, in-
1567 troducing one possible way of automating this problem, and explored the pros and
1568 cons of this approach. We applied our method to the 1.4 GHz Australian Telescope
1569 Large Area Survey (ATLAS) observations of the *Chandra* Deep Field South (CDFS)
1570 and the ESO Large Area ISO Survey South 1 (ELAIS-S1) fields by cross-identifying
1571 them with the *Spitzer* Wide-area Infrared Extragalactic (SWIRE) survey. We trained
1572 our method with two sets of data: expert cross-identifications of CDFS from the ini-
1573 tial ATLAS data release and crowdsourced cross-identifications of CDFS from Radio
1574 Galaxy Zoo. We found that a simple strategy of cross-identifying a radio component
1575 with the nearest galaxy performs comparably to our more complex methods, though
1576 our estimated best-case performance is near 100 per cent. ATLAS contains 87 complex
1577 radio sources that have been cross-identified by experts, so there are not enough com-
1578 plex examples to learn how to cross-identify them accurately. Much larger datasets are
1579 therefore required for training methods like ours. We also showed that training our
1580 method on Radio Galaxy Zoo cross-identifications gives comparable results to train-
1581 ing on expert cross-identifications, demonstrating the value of crowdsourced training
1582 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 4.1.

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., 2011) 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 galaxies 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; Middelberg et al., 2008; Norris et al., 2006) 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 galaxies. 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 over 75 000 radio-host cross-identifications and radio source morphologies (Wong et al., in

¹<https://radio.galaxyzoo.org>

1609 prep.). While this is a much larger number of visual cross-identifications than have
1610 been made by experts (e.g., Gendre & Wall, 2008; Grant et al., 2010; Middelberg et al.,
1611 2008; Norris et al., 2006; Taylor et al., 2007) it is still far short of the millions of radio
1612 sources expected to be detected in upcoming radio surveys (Norris, 2017b).

1613 Automated algorithms have been developed for cross-identification. Fan et al. (2015)
1614 applied Bayesian hypothesis testing to this problem, fitting a three-component model
1615 to extended radio sources. This was achieved under the assumption that extended
1616 radio sources are composed of a core radio component and two lobe components.
1617 The core radio component is coincident with the host galaxy, so cross-identification
1618 amounts to finding the galaxy coincident with the core radio component in the most
1619 likely model fit. This method ~~is could be~~ easily extended to use other, more complex
1620 models, but it is purely geometric. It does not incorporate other information such
1621 as the physical properties of the potential host galaxy. Additionally, there may be
1622 new classes of radio source detected in future surveys like EMU which do not fit the
1623 model. Weston et al. (2018) developed a modification of the likelihood ratio method of
1624 cross-identification (Richter, 1975) for application to ATLAS and EMU. This method
1625 does well on non-extended radio sources with approximately 70 per cent accuracy in
1626 the ATLAS fields, but does not currently handle more complex (extended or multi-
1627 component) radio sources (Norris, 2017a).

1628 One possibility is that machine learning techniques can be developed to automatically
1629 cross-identify catalogues drawn from new surveys. Machine learning describes
1630 a class of methods that learn approximations to functions. If cross-identification can
1631 be cast as a function approximation problem, then machine learning will allow ~~data~~
1632 ~~sets-datasets~~ such as Radio Galaxy Zoo to be generalised to work on new data. ~~Data~~
1633 ~~sets-Datasets~~ from citizen scientists have already been used to train machine learning
1634 methods. Some astronomical examples can be found in Marshall et al. (2015).

1635 In this chapter we cast cross-identification as a function approximation problem
1636 by applying an approach from computer vision literature. This approach casts cross-
1637 identification as the standard machine learning problem of binary classification by ask-
1638 ing whether a given infrared source is the host galaxy or not. We train our methods
1639 on expert cross-identifications and volunteer cross-identifications from Radio Galaxy
1640 Zoo. In Section 4.2 we describe the data we use to train our methods. In Section 4.3
1641 we discuss how we cast the radio host galaxy cross-identification problem as a ma-
1642 chine learning problem. In Section 4.4 we present results of applying our method
1643 to ATLAS observations of the *Chandra* Deep Field South (CDFS) and the ESO Large
1644 Area ISO Survey South 1 (ELAIS-S1) field. Our data, code and results are available at
1645 <https://radiogalaxyzoo.github.io/atlas-xid>.

1646 Throughout this chapter, a ‘radio source’ refers to all radio emission observed asso-
1647 ciated with a single host galaxy, and a ‘radio component’ refers to a single, contiguous
1648 region of radio emission. Multiple components may arise from a single source. A
1649 ‘compact’ source is composed of a single unresolved component. Equation 4.1 shows
1650 the definition of a resolved component. We assume that all unresolved components
1651 are compact sources, i.e. we assume that each unresolved component has its own host

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

1652 galaxy². An ‘extended’ source is a non-compact source, i.e. resolved single-component
 1653 sources or a multi-component source. Figure 4.1 illustrates these definitions.

1654 4.2 Data

1655 We use radio data from the Australia Telescope Large Area Survey (ATLAS; Franzen
 1656 et al., 2015; Norris et al., 2006), infrared data from the *Spitzer* Wide-area Infrared
 1657 Extragalactic survey (SWIRE; Lonsdale et al., 2003; Surace et al., 2005), and cross-
 1658 identifications of these surveys from the citizen science project Radio Galaxy Zoo (Ban-
 1659 field et al., 2015). Radio Galaxy Zoo also includes cross-identifications of sources in
 1660 Faint Images of the Radio Sky at Twenty Centimeters (FIRST; White et al., 1997) and
 1661 the AllWISE survey (Cutri et al., 2013), though we focus only on Radio Galaxy Zoo
 1662 data from ATLAS and SWIRE.

1663 4.2.1 ATLAS

1664 ATLAS is a pilot survey for the EMU (Norris et al., 2011) survey, which will cover
 1665 the entire sky south of +30 deg and is expected to detect approximately 70 million
 1666 new radio sources. 95 per cent of these sources will be single-component sources, but
 1667 the remaining 5 per cent pose a considerable challenge to current automated cross-
 1668 identification methods (Norris et al., 2011). EMU will be conducted at the same depth
 1669 and resolution as ATLAS, so methods developed for processing ATLAS data are ex-
 1670 pected to work for EMU. ATLAS is a wide-area radio survey of the CDFS and ELAIS-
 1671 S1 fields at 1.4 GHz with a sensitivity of 14 and 17 $\mu\text{Jy beam}^{-1}$ on CDFS and ELAIS-S1
 1672 respectively. CDFS covers 3.6 deg² and contains 3034 radio components above a signal-
 1673 to-noise ratio of 5. ELAIS-S1 covers 2.7 deg² and contains 2084 radio components above
 1674 a signal-to-noise ratio of 5 (Franzen et al., 2015). The images of CDFS and ELAIS-S1
 1675 have angular resolutions of 16 by 7 and 12 by 8 arcsec respectively, with pixel sizes
 1676 of 1.5 arcsec px⁻¹. Table 4.1 summarises catalogues that contain cross-identifications
 1677 of radio components in ATLAS with host galaxies in SWIRE. In the present work, we

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.

1678 train methods on CDFS³ and test these methods on both CDFS and ELAIS-S1. This
1679 ensures helps confirm that our methods are transferable to different areas of the sky
1680 observed by the same telescope, as will be the case for EMU.

1681 4.2.2 SWIRE

1682 SWIRE is a wide-area infrared survey at the four IRAC wavelengths 3.6, 4.5, 5.8, and
1683 8.0 μm (Lonsdale et al., 2003; Surace et al., 2005). It covers eight fields, including CDFS
1684 and ELAIS-S1. SWIRE is the source of infrared observations for cross-identification
1685 with ATLAS. SWIRE has catalogued 221,535 infrared objects in CDFS and 186,059 in-
1686 frared objects in ELAIS-S1 above a signal-to-noise ratio of 5.

1687 4.2.3 Radio Galaxy Zoo

1688 Radio Galaxy Zoo asks volunteers to cross-identify radio components with their in-
1689 frared host galaxies. There are a total of 2460 radio components in Radio Galaxy Zoo
1690 sourced from ATLAS observations of CDFS. These components are cross-identified by
1691 Radio Galaxy Zoo participants with host galaxies detected in SWIRE. A more detailed
1692 description can be found in Banfield et al. (2015) and a full description of how the Ra-
1693 dio Galaxy Zoo catalogue used in this work⁴ is generated can be found in Wong et al.
1694 (in prep.).

1695 The ATLAS CDFS radio components that appear in Radio Galaxy Zoo are drawn
1696 from a prerelease version of the third data release of ATLAS by Franzen et al. (2015).
1697 In this release, each radio component was fit with a two-dimensional Gaussian. De-
1698 pending on the residual of the fit, more than one Gaussian may be fit to one region of
1699 radio emission. Each of these Gaussian fits is listed as a radio component in the ATLAS
1700 component catalogue. The brightest radio component from the multiple-Gaussian fit
1701 is called the ‘primary component’. If there was is only one Gaussian fit then this Gaus-
1702 sian is the primary component. Each primary component found in the ATLAS compo-
1703 nent catalogue appears in Radio Galaxy Zoo. Non-primary components may appear
1704 within the image of a primary component, but do not have their own entry in Radio
1705 Galaxy Zoo. We will henceforth only discuss the primary components.

1706 4.3 Method

1707 The aim of this chapter is to express cross-identification in a form that will allow us
1708 to apply standard machine learning tools and methods. We use an approach from
1709 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.

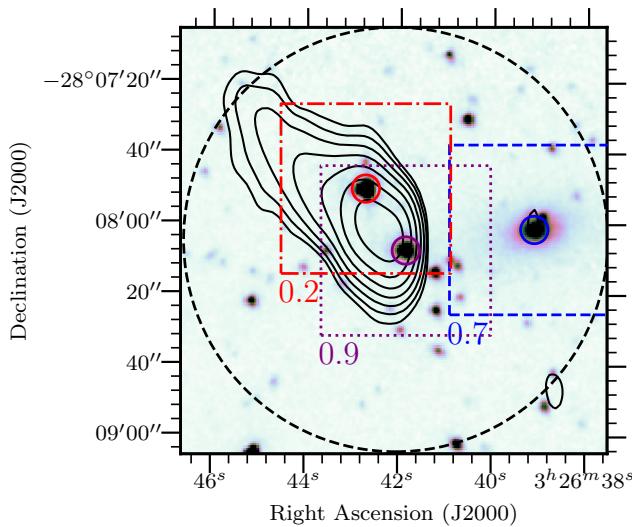


Figure 4.2: An example of finding the host galaxy of a radio source using our sliding-window method. The background image is a $3.6\text{ }\mu\text{m}$ image from SWIRE. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2. Boxes represent ‘windows’ centred on 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.

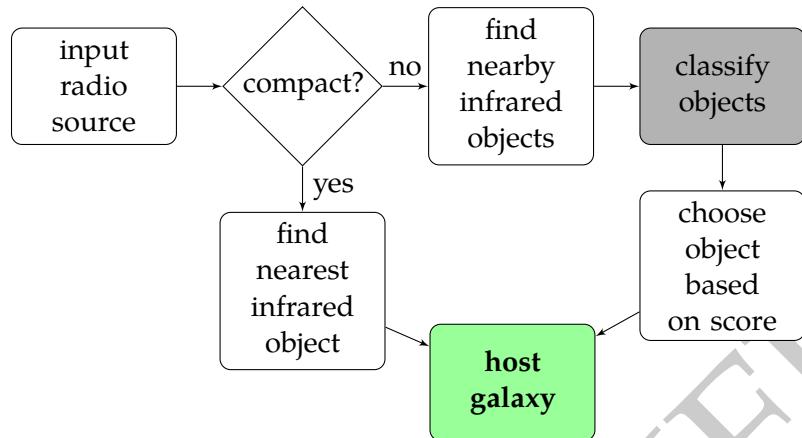


Figure 4.3: Our cross-identification method once a binary classifier has been trained. As input we accept a radio component. If the component is compact, we assume it is a compact source and select the nearest infrared object as the host galaxy. If the component is resolved, we use the binary classifier to score all nearby infrared objects and select the highest-scored object as the host galaxy. Compact and resolved components are defined in Equation 4.1.

1710 4.3.1 Cross-identification as binary classification

1711 We propose a two-step method for host galaxy cross-identification which we will de-
 1712 scribe now. Given a radio component, we want to find the corresponding host galaxy.
 1713 The input is a $2' \times 2'$ radio image of the sky centred on a radio component and po-
 1714 tentially other information about objects in the image (such as the redshift or infrared
 1715 colour). Images at other wavelengths (notably infrared) might be useful, but we defer
 1716 this for now as it complicates the task. We ~~choose~~^{choose} a $2' \times 2'$ image to match the
 1717 size of the images used by Radio Galaxy Zoo. To avoid solving the separate task of
 1718 identifying which radio components are associated with the same source, we assume
 1719 that each radio image represents a single extended source⁵. Radio cross-identification
 1720 can then be formalised as follows: ~~given~~^{Given} a radio image centred on a radio com-
 1721 ponent, locate the host galaxy of the source containing this radio component. This is a
 1722 standard computer vision problem called ‘object detection’, and we apply a common
 1723 technique called a ‘sliding-window’ (Rowley et al., 1996).

1724 In sliding-window object detection, we want to find an object in an image. We
 1725 develop a function to score each location in the image such that the highest-scored
 1726 location coincides with the desired object. Square image cutouts called ‘windows’ are
 1727 taken centred on each location and these windows are used to represent that location
 1728 in our scoring function. To find the infrared host galaxy, we choose the location with
 1729 the highest score. To improve the efficiency of this process when applied to cross-
 1730 identification, we only consider windows coincident with infrared sources detected in
 1731 SWIRE. We call these infrared sources ‘candidate host galaxies’. For this chapter, there
 1732 is no use in scoring locations without infrared sources as that would not lead to a host

⁵Limitations of this assumption are discussed in Section 4.3.2.

1733 identification anyway. Using candidate host galaxies instead of pixels also allows us to
 1734 include ancillary information about the candidate host galaxies, such as their infrared
 1735 colours and redshifts. We refer to the maximum distance a candidate host galaxy can
 1736 be separated from a radio component as the ‘search radius’ and take this radius to be 1
 1737 arcmin. To score each candidate host galaxy we use a ‘binary classifier’, which we will
 1738 define now.

Algorithm 1: Cross-identifying a radio component given a radio image of the component, a catalogue of infrared candidate host galaxies \mathcal{G} , 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;

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

if $score > max$ **then**

$max \leftarrow score;$

$host \leftarrow g;$

end

end

return $host$

1739 Binary classification is a common method in machine learning where objects are to
 1740 be assigned to one of two classes, called the ‘positive’ and ‘negative’ classes. This as-
 1741 signment is represented by the probability that an object is in the positive class. A
 1742 ‘binary classifier’ is a function mapping from an object to such a probability. Our
 1743 formulation of cross-identification is equivalent to binary classification of candidate
 1744 host galaxies: the positive class represents host galaxies, the negative class represents
 1745 non-host galaxies, and to cross-identify a radio component we find the candidate host
 1746 galaxy maximising the positive class probability. In other words, the binary classifier
 1747 is exactly the sliding-window scoring function. We therefore split cross-identification
 1748 into two separate tasks: the ‘candidate classification task’ where, given a candidate
 1749 host galaxy, we wish to determine whether it is a host galaxy of *any* radio component;
 1750 and the ‘cross-identification task’ where, given a specific radio component, we wish to
 1751 find its host galaxy. The candidate classification task is a traditional machine learning
 1752 problem which results in a binary classifier. To avoid ambiguity and recognise that
 1753 the values output by a binary classifier are not true probabilities, we will refer to the
 1754 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 the 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 maximising cross-identification accuracy. 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.

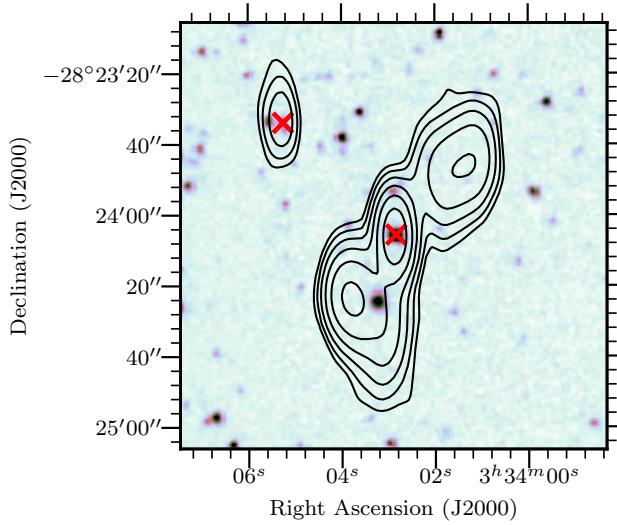


Figure 4.4: A $2'$ -wide radio image centred on ATLAS3_J033402.87-282405.8C. This radio source breaks the assumption that there are no other radio sources within 1 arcmin of the source. Another radio source is visible to the upper-left. Host galaxies found by Radio Galaxy Zoo volunteers are shown by 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.

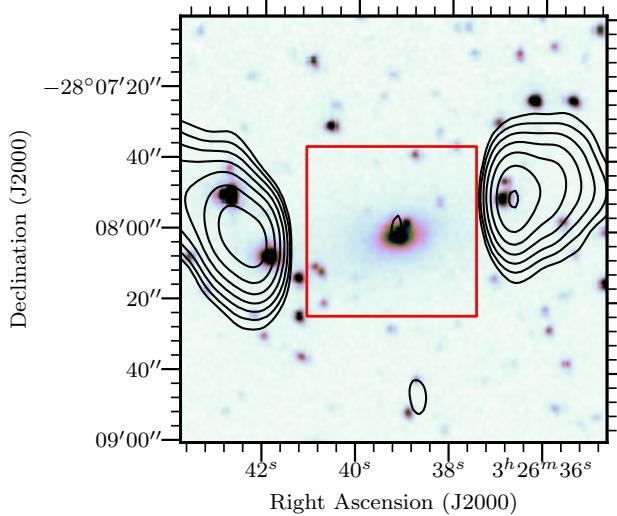


Figure 4.5: An example of a radio source where the window centred on the host galaxy, shown as a rectangle, does not contain enough radio information to correctly identify the galaxy as the host. The background image is a $3.6\text{ }\mu\text{m}$ image from SWIRE. The contours show ATLAS radio data and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

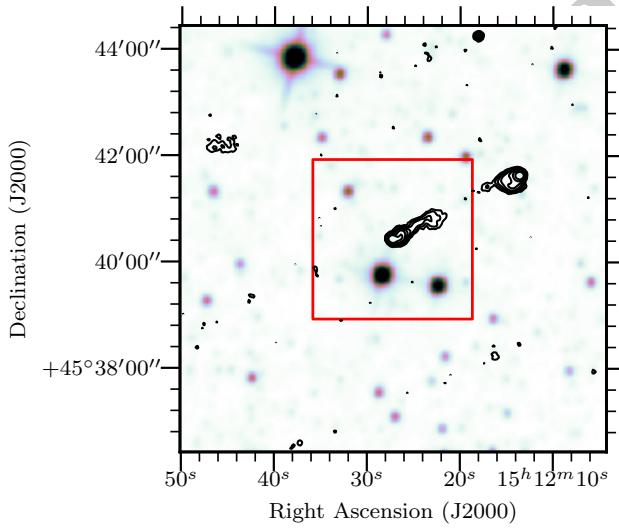


Figure 4.6: A $8'$ -wide radio image from FIRST, centred on FIRST J151227.2+454026. The $3'$ -wide red box indicates the boundaries of the image of this radio component shown to volunteers in Radio Galaxy Zoo. This radio source breaks our assumption that the whole radio source is visible in the chosen 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 $3.4\text{ }\mu\text{m}$ image from WISE. The contours show FIRST radio data, starting at 4σ and increasing geometrically by a factor of 2.

-
- 1795 3. The host galaxy of a radio component is within the 1 arcmin search radius around
 1796 the component, measured from the centre of the Gaussian fit.
- 1797 4. The host galaxy of a radio component is closer on the sky to the radio component
 1798 than the host galaxy of any other radio component.
- 1799 5. The host galaxy appears in the SWIRE catalogue.

1800 These assumptions limit the effectiveness of our approach, regardless of how accurate
 1801 our binary classifier may be. Examples of radio sources that break these respective
 1802 assumptions are:

- 1803 1. A radio source less than 1' away from another radio source.
- 1804 2. A radio source with an angular size greater than 2'.
- 1805 3. A radio source with a component greater than 1' away from the host galaxy.
- 1806 4. A two-component radio source with another host galaxy between a component
 1807 and the true host galaxy.
- 1808 5. An infrared-faint radio source (as in Collier et al., 2014).

1809 The main limitations are problems of scale in choosing the candidate search radius
 1810 and the size of the windows representing candidates. If the search radius is too small,
 1811 we may not consider the host galaxy as a candidate. If the search radius is too large, we
 1812 may consider multiple host galaxies (though this is mostly mitigated by the Gaussian
 1813 weighting). If the window is too small, radio emission may extend past the edges of
 1814 the window and we may miss critical information required to identify the galaxy as a
 1815 host galaxy. If the window is too large, then irrelevant information will be included
 1816 and it may be difficult or computationally expensive to score. We chose choose a win-
 1817 dow size of 32×32 pixels, corresponding to approximately $48'' \times 48''$ in ATLAS. This
 1818 is shown as squares in Figure 4.2 and Figure 4.5. These kinds of size problems are dif-
 1819 ficult even for non-automated methods as radio sources can be extremely wide—for
 1820 example, Radio Galaxy Zoo found a radio giant that spanned over three different im-
 1821 ages presented to volunteers and the full source was only cross-identified by the efforts
 1822 of citizen scientists (Banfield et al., 2015). An example of a radio image where part of
 1823 the radio source is outside the search radius is shown in Figure 4.6.

1824 In weighting the scores by a Gaussian function of angular separation, we implicitly
 1825 assume that the host galaxy of a radio component is closer to that radio component
 1826 than any other host galaxy. If this assumption is not true then the incorrect host galaxy
 1827 may be identified, though this is rare.

1828 We only need to require that the host galaxy appears in SWIRE to incorporate
 1829 galaxy-specific features (Section 4.3.3) and to improve efficiency. Our method is ap-
 1830 plicable even when host galaxies are not detected in the infrared by considering every
 1831 pixel of the radio image as a candidate location as would be done in the original com-
 1832 puter vision approach. If the host galaxy location does not correspond to an infrared
 1833 source, the radio source would can be classified as infrared-faint.

1834 Our assumptions impose an upper bound on how well we can cross-identify radio
1835 sources. We estimate this upper bound in Section 4.4.1.

1836 **4.3.3 Feature vector representation of infrared sources**

1837 Inputs to binary classifiers must be represented by an array of real values called feature
1838 vectors. We therefore need to choose a feature vector representation of our candidate
1839 host galaxies. Candidate hosts are sourced from the SWIRE catalogue (Section 4.2.2).
1840 We represent each candidate host with 1034 real-valued features, combining the win-
1841 dows centred on each candidate (Section 4.3.1) with ancillary infrared data from the
1842 SWIRE catalogue. For a given candidate host, these features are:

- 1843 • the 6 base-10 logarithms of the ratios of fluxes of the candidate host at the four
1844 IRAC wavelengths (the ‘colours’ of the candidate);
- 1845 • the flux of the host at 3.6 μm ;
- 1846 • the stellarity index of the host at both 3.6 and 4.5 μm ;
- 1847 • the radial distance between the candidate host and the nearest radio component
1848 in the ATLAS catalogue; and
- 1849 • a 32×32 pixel image from ATLAS (approximately $48'' \times 48''$), centred on the
1850 candidate host (the window).

1851 The infrared colours provide insight into the properties of the candidate host galaxy
1852 (Grant, 2011). The 3.6 and 4.5 μm fluxes trace both galaxies with faint polycyclic aro-
1853 matic hydrocarbon (PAH) emission (i.e. late-type, usually star-forming galaxies) and
1854 elliptical galaxies dominated by old stellar populations. The 5.8 μm flux selects galax-
1855 ies where the infrared emission is dominated by non-equilibrium emission of dust
1856 grains due to active galactic nuclei, while the 8.0 μm flux traces strong PAH emission
1857 at low redshift (Sajina et al., 2005). The stellarity index is a value in the SWIRE cata-
1858 logue that represents how likely the object is to be a star rather than a galaxy (Surace
1859 et al., 2005). It was estimated by a neural network in SExtractor (Bertin & Arnouts,
1860 1996).

1861 We use the 32×32 pixels of each radio window as independent features for all
1862 binary classification models, with the convolutional neural network automatically ex-
1863 tracting features that are relevant. Other features of the radio components may be
1864 used instead of just relying on the pixel values, but there has been limited research on
1865 extracting such features: Proctor (2006) describes hand-selected features for radio dou-
1866 bles in FIRST, and Aniyan and Thorat (2017) and Lukic et al. (2018) make use of deep
1867 convolutional neural networks which automatically extract features as part of classifi-
1868 cation. A more comprehensive investigation of features is a good avenue for potential
1869 improvement in our pipeline but this is beyond the scope of this initial study.

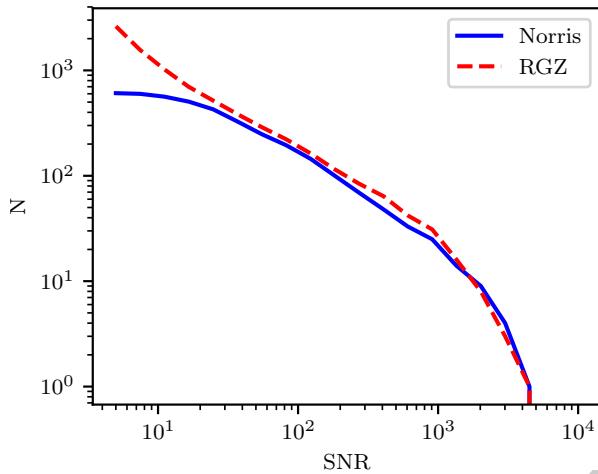


Figure 4.7: Cumulative number of radio components (N) in the expert (Norris) and Radio Galaxy Zoo (RGZ) training sets with different signal-to-noise ratios (SNR).

1870 4.3.4 Binary Classifiers

1871 We use three different binary classification models: logistic regression, convolutional
 1872 neural networks, and random forests. These models cover three different approaches
 1873 to machine learning. Logistic regression is a probabilistic binary classification model.
 1874 It is linear in the feature space and outputs the probability that the input has a positive
 1875 label (Bishop, 2006, Chap. 4). Convolutional neural networks (CNN) are biologically-
 1876 inspired prediction models with image inputs. They have recently produced good
 1877 results on large image-based datasets in astronomy (Dieleman et al., 2015; Lukic et al.,
 1878 2018, e.g.). Random forests are an ensemble of decision trees (Breiman, 2001). They
 1879 consider multiple subsamples of the training set, where each bootstrap subsample is
 1880 sampled with replacement from the training set. To classify a new data point, the ran-
 1881 dom forest takes the weighted average of all classifications produced by each decision
 1882 tree. [For a more detailed description of these models, see Appendix A.](#)

1883 4.3.5 Labels

1884 The Radio Galaxy Zoo and Norris et al. (2006) cross-identification catalogues must be
 1885 converted to binary labels for infrared objects so that they can be used to train binary
 1886 classifiers. There are two challenges with this conversion:

- 1887 • We can only say that an object is a host galaxy, not which radio object it is associ-
 1888 ated with, and
- 1889 • We cannot disambiguate between non-host infrared objects and host galaxies that
 1890 ~~were~~ are not in the cross-identification catalogue.

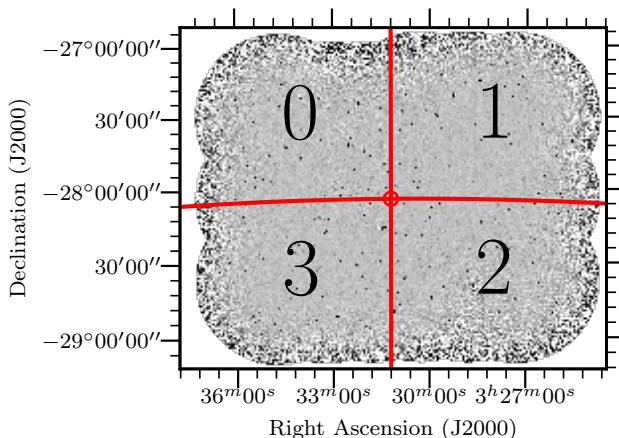


Figure 4.8: CDFS field training and testing quadrants labelled 0 – 3. The central dot is located at $\alpha = 03^{\text{h}}31^{\text{m}}12^{\text{s}}$, $\delta = -28^{\circ}06'00''$. The quadrants ~~were~~are chosen such that there are similar numbers of radio sources in each quadrant.

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

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 problems—an example is that if the cross-identification catalogue did does not include a radio object (e.g. it was-is below the signal-to-noise ratio) then the host galaxy of that radio object would receive receives 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 4.7. We train and test our binary classifiers on infrared objects within a 1 arcmin radius of an ATLAS radio component.

4.3.6 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 4.8. 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

$$\ln \left(\frac{S_{\text{int}}}{S_{\text{peak}}} \right) > 2 \sqrt{\left(\frac{\sigma_{S_{\text{int}}}}{S_{\text{int}}} \right)^2 + \left(\frac{\sigma_{S_{\text{peak}}}}{S_{\text{peak}}} \right)^2}, \quad (4.1)$$

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 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—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 are correlated and we show this correlation in Figure 4.9. Fitting a line of best fit with `scipy`

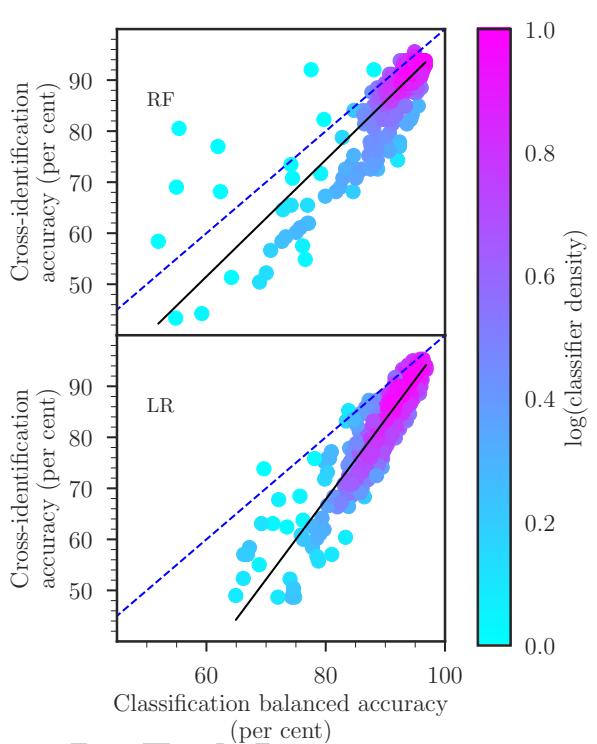


Figure 4.9: Balanced accuracy on the candidate classification task plotted against accuracy on the cross-identification task. ‘RF’ indicates results from random forests, and ‘LR’ indicates results from logistic regression. Binary classifiers were trained on random, small subsets of the training data 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.

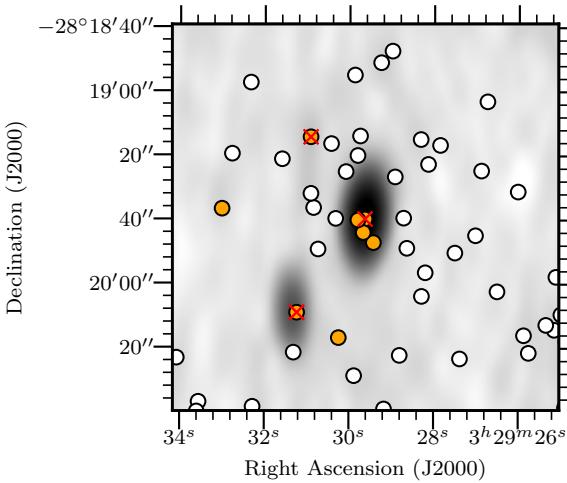


Figure 4.10: Predicted host galaxies in the candidate classification task for ATLAS3 J032929.61-281938.9. The background image is an ATLAS radio image. Radio Galaxy Zoo host galaxies are marked by crosses. SWIRE candidate host galaxies are circles coloured by the score output by a logistic regression 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.

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 4.10, where the classifier has identified 8 eight ‘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 4.11 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 classification task and

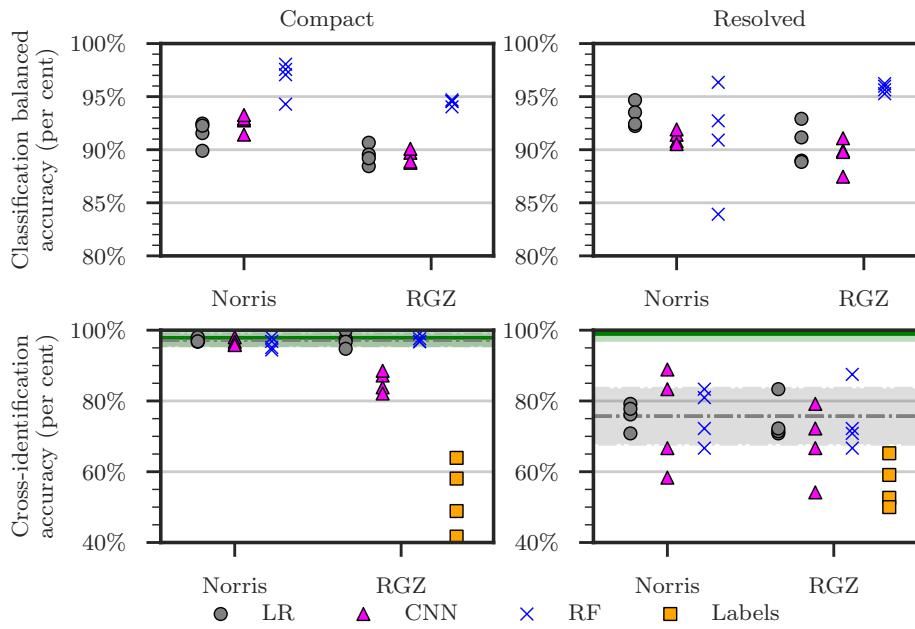


Figure 4.11: Performance of our method with logistic regression ('LR'), convolutional neural networks ('CNN') and random forest ('RF') binary classifiers. 'Norris' indicates the performance of binary classifiers trained on the expert labels and 'RGZ' indicates the performance of binary classifiers trained 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.

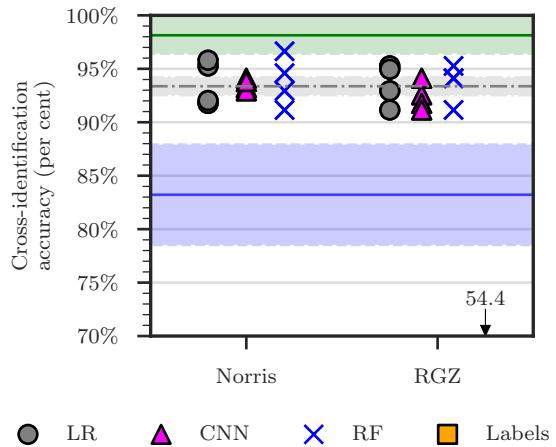


Figure 4.12: Performance of our approach using different binary classifiers on the cross-identification task. Markers and lines are as in Figure 4.11. The blue solid line indicates the performance of a random binary classifier and represents the minimum accuracy we expect to obtain. The standard deviation 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.

1985 so represents the best possible cross-identification performance achievable with our
 1986 method. We estimate the lower limit on performance by assigning random scores to
 1987 each candidate host galaxy. We expect any useful binary classifier to produce better re-
 1988 sults than this, so this represents the lowest expected cross-identification performance.
 1989 The upper estimates, lower estimates, and nearest neighbour accuracy are shown as horizontal lines in Figure 4.11.
 1990

1991 In Figure 4.12 we plot the performance of our method using different binary classi-
 1992 fication models, as well as the performance of Radio Galaxy Zoo, nearest neighbours,
 1993 and the perfect and random binary classifiers τ on the full set of ATLAS DR1 radio
 1994 components using the pipeline in Figure 4.3. The accuracy associated with each clas-
 1995 sification model and training label set averaged across all four quadrants is shown in
 1996 Appendix B.

1997 Differences between accuracies across training labels are well within one standard
 1998 deviation computed across the four quadrants, with convolutional neural networks on
 1999 compact objects as the only exception. The spread of accuracies is similar for both sets
 2000 of training labels, with the exception of random forests. The balanced accuracies of
 2001 random forests trained on expert labels have a considerably higher spread than those
 2002 trained on Radio Galaxy Zoo labels, likely because of the small size of the expert train-
 2003 ing set—there are less than half the number of objects in the expert-labelled training set
 2004 than the number of objects in the Radio Galaxy Zoo-labelled training set (Table 4.2).

2005 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-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-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 that we~~ applied and the small size of ATLAS thus limits their performance.

We noted in Section 4.3.5 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. Specifically, true host galaxies may be misidentified as non-host galaxies if the associated radio source ~~was-is~~ 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 C and the predicted cross-identification for each ATLAS object in Appendix D. Scores ~~reported-we report~~ for a given object were predicted by binary classifiers tested on the quadrant containing that object. The reported scores are not weighted.

In Figure 4.17 we show ~~5-five~~ 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 comparable performance between ELAIS-S1 and CDFS. In Figure 4.13 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 mod-

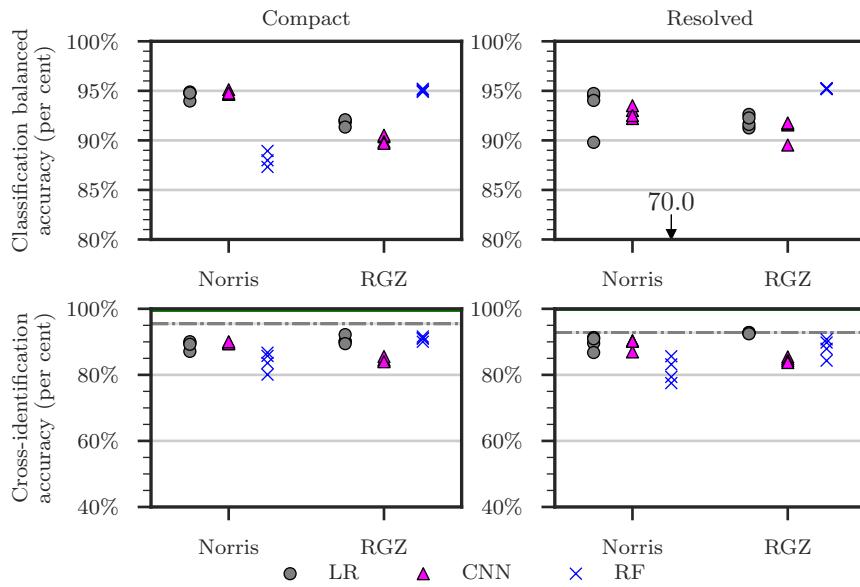


Figure 4.13: Performance of different classification models trained on CDFS and tested on resolved and compact sources in ELAIS-S1. Points represent classification models trained on different quadrants of CDFS, with markers, lines, and axes as in Figure 4.11. The balanced accuracy of expert-trained random 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.

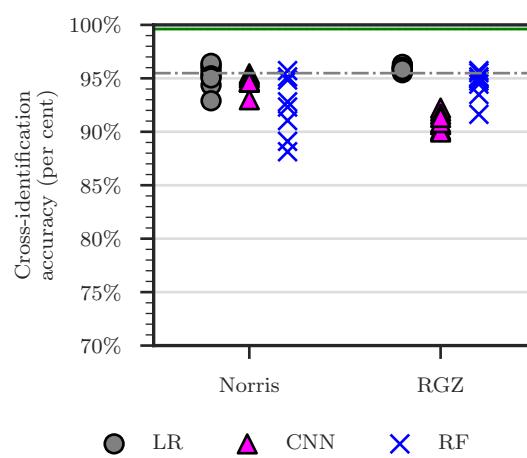


Figure 4.14: Performance of different classifiers trained on CDFS and tested on ELAIS-S1. Markers are as in Figure 4.12 and horizontal lines are as in Figure 4.13. Note that the pipeline shown in Figure 4.3 is used here, so compact objects are cross-identified in the same way regardless of binary classifier model.

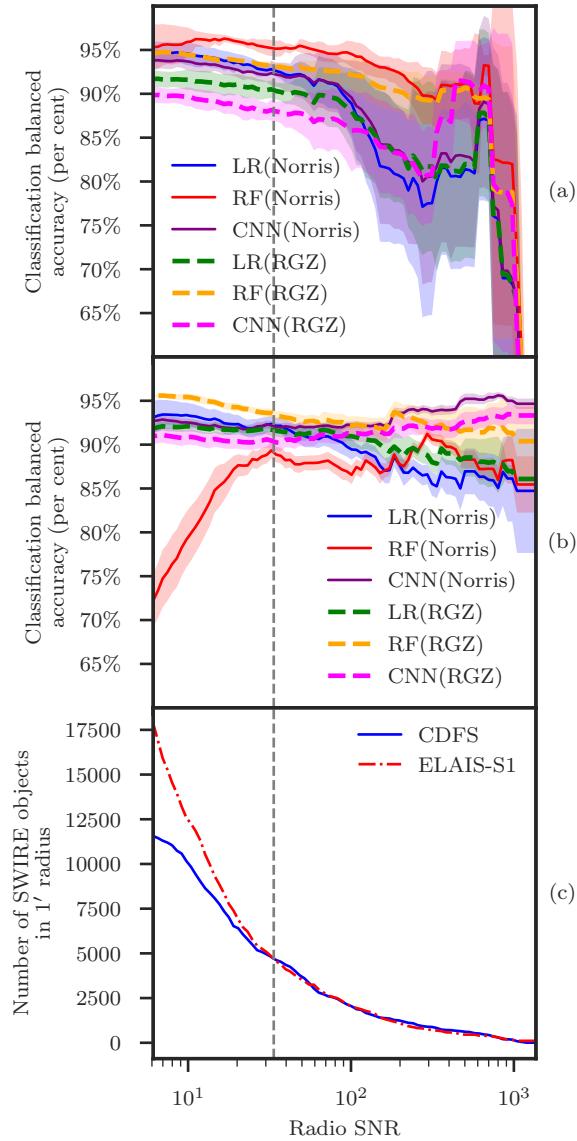


Figure 4.15: (a) Balanced accuracies of classifiers trained and tested on CDFS with different signal-to-noise ratio (SNR) cutoffs for the test set. A SWIRE object is included in the test set if it is within $1'$ of a radio component with greater SNR than the cutoff. Lines of different colour indicate 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.

2047 els. We also plot the cross-identification accuracy of a nearest neighbours approach⁶.
 2048 In Figure 4.14 we plot the performance of our method on the full set of ELAIS-S1 AT-
 2049 LAS DR1 radio components using the pipeline in Figure 4.3. We list the corresponding
 2050 accuracies in Appendix B.

2051 Cross-identification results from ELAIS-S1 are similar to those for CDFS, showing
 2052 that our method trained on CDFS performs comparably well on ELAIS-S1. However,
 2053 nearest neighbours outperforms most methods on ELAIS-S1. This is likely because
 2054 there is a much higher percentage of compact objects in ELAIS-S1 than in CDFS. The
 2055 maximum achievable accuracy we have estimated for ELAIS-S1 is very close to 100
 2056 per cent, so (as for CDFS) a very accurate binary classifier would outperform nearest
 2057 neighbours.

2058 One interesting difference between the ATLAS fields is that random forests trained
 2059 on expert labels perform well on CDFS but poorly on ELAIS-S1. This is not the case for
 2060 logistic regression or convolutional neural networks trained on expert labels, nor is it
 2061 the case for random forests trained on Radio Galaxy Zoo. We hypothesise that this is
 2062 because the ELAIS-S1 cross-identification catalogue (Middelberg et al., 2008) labelled
 2063 fainter radio components than the CDFS cross-identification catalogue (Norris et al.,
 2064 2006) due to noise from the very bright source ATCDFS_J032836.53-284156.0 in CDFS.
 2065 Classifiers trained on CDFS expert labels may thus be biased toward brighter radio
 2066 components compared to ELAIS-S1. Radio Galaxy Zoo uses a preliminary version of
 2067 the third data release of ATLAS (Franzen et al., 2015) and so classifiers trained on the
 2068 Radio Galaxy Zoo labels may be less biased toward brighter sources compared to those
 2069 trained on the expert labels. To test this hypothesis we tested each classification model
 2070 against test sets with a signal-to-noise ratio (SNR) cutoff. A SWIRE object was only
 2071 included in the test set for a given cutoff if it was located within 1' of a radio component
 2072 with a SNR above the cutoff. The balanced accuracies for each classifier at each cutoff
 2073 are shown in Figure 4.15(a) and (b) and the distribution of test set size for each cutoff
 2074 is shown in Figure 4.15(c). Figure 4.15(c) shows that ELAIS-S1 indeed has more faint
 2075 objects in its test set than the CDFS test set, with the SNR for which the two fields reach
 2076 the same test set size (approximately 34) indicated by the dashed vertical line on each
 2077 plot. For CDFS, all classifiers perform reasonably well across cutoffs, with performance
 2078 dropping as the size of the test set becomes small. For ELAIS-S1, logistic regression and
 2079 convolutional neural networks perform comparably across all SNR cutoffs, but random
 2080 forests do not. While random forests trained on Radio Galaxy Zoo labels perform
 2081 comparably to other classifiers across all SNR cutoffs, random forests trained on expert
 2082 labels show a considerable drop in performance below the dashed line.

2083 4.5 Discussion

2084 Based on the ATLAS sample, our main result is that it is possible to cast radio host
 2085 galaxy cross-identification as a machine learning task for which standard methods can

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

2086 be applied. These methods can then be trained with a variety of label sets derived
 2087 from cross-identification catalogues. While our methods have not outperformed near-
 2088 est neighbours, we have demonstrated that for a very accurate binary classifier, good
 2089 cross-identification results can be obtained using our method. Future work could com-
 2090 bine multiple catalogues or physical priors to boost performance.

2091 Nearest neighbours approaches outperform most methods we investigated, no-
 2092 tably including Radio Galaxy Zoo. This is due to the large number of compact or
 2093 partially-resolved objects in ATLAS. This result shows that for compact and partially-
 2094 resolved objects, methods that do not use machine learning such as a nearest neigh-
 2095 bours approach or likelihood ratio (Weston et al., 2018) should be preferred to ma-
 2096 chine learning methods. It also shows that ATLAS is not an ideal ~~data set~~ dataset for
 2097 developing machine learning methods like ours. Our use of ATLAS is motivated by its
 2098 status as a pilot survey for EMU, so methods developed for ATLAS should also work
 2099 for EMU. New methods developed should work well with extended radio sources, but
 2100 this goal is almost unsupported by ATLAS as it has very few examples of such sources.
 2101 This makes both training and testing difficult—there are too few extended sources to
 2102 train on and performance on such a small test set may be unreliable. Larger ~~data sets~~
 2103 datasets with many extended sources like FIRST exist, but these are considerably less
 2104 deep than and at a different resolution to EMU, so there is no reason to expect methods
 2105 trained on such ~~data sets~~ datasets to be applicable to EMU.

2106 The accuracies of our trained cross-identification methods generally fall far below
 2107 the estimated best possible accuracy attainable using our approach, indicated by the
 2108 green-shaded areas in Figures 4.12 and 4.14. The balanced accuracies attained by our
 2109 binary classifiers indicate that there is significant room for improvement in classifica-
 2110 tion. The classification accuracy could be improved by better model selection and more
 2111 training data, particularly for convolutional neural networks. There is a huge variety
 2112 of ways to build a convolutional neural network, and we have only investigated one ar-
 2113 chitecture. For an exploration of different convolutional neural network architectures
 2114 applied to radio astronomy, see Lukic et al. (2018). Convolutional neural networks
 2115 generally require more training data than other machine learning models and we have
 2116 only trained our networks on a few hundred sources. We would expect performance
 2117 on the classification task to greatly increase with larger training sets.

2118 Another problem is that of the window size used to select radio features. Increasing
 2119 window size would increase computational expense, but provide more information to
 2120 the models. Results are also highly sensitive to how large the window size is compared
 2121 to the size of the radio source we are trying to cross-identify, with large angular sizes
 2122 requiring large window sizes to ensure that the features contain all the information
 2123 needed to localise the host galaxy. An ideal implementation of our method would
 2124 most likely represent a galaxy using radio images taken at multiple window sizes, but
 2125 this is considerably more expensive.

2126 Larger training sets, better model selection, and larger window sizes would im-
 2127 prove performance, but only so far: ~~we~~-We would still be bounded above by the es-
 2128 timated ‘perfect’ classifier accuracy. From this point, the performance can only be
 2129 improved by addressing our broken assumptions. We detailed these assumptions in

2130 Section 4.3.2, and we will discuss here how our method could be adapted to avoid these
2131 assumptions. Our assumption that the host galaxy is contained within the search radius
2132 could be improved by dynamically choosing the search radius, perhaps based
2133 on the angular extent of the radio emission, or the redshift of candidate hosts. Radio
2134 morphology information may allow us to select relevant radio data and hence relax the
2135 assumption that a 1'-wide radio image represents just one, whole radio source. Finally,
2136 our assumption that the host galaxy is detected in infrared is technically not needed,
2137 as the sliding-window approach we have employed will still work even if there are
2138 no detected host galaxies—instead of classifying candidate hosts, simply classify each
2139 pixel in the radio image. The downside of removing candidate hosts is that we are no
2140 longer able to reliably incorporate host galaxy information such as colour and redshift,
2141 though this could be resolved by treating pixels as potentially undetected candidate
2142 hosts with noisy features.

2143 We observe that Radio Galaxy Zoo-trained methods perform comparably to meth-
2144 ods trained on expert labels. This shows that the crowdsourced labels from Radio
2145 Galaxy Zoo will provide a valuable source of training data for future machine learn-
2146 ing methods in radio astronomy.

2147 Compared to nearest neighbours, cross-identification accuracy on ELAIS-S1 is lower
2148 than on CDFS. Particularly notable is that our performance on compact objects is very
2149 low for ELAIS-S1, while it was near-optimal for CDFS. These differences may be for
2150 a number of reasons. ELAIS-S1 has beam size and noise profile different from CDFS
2151 (even though both were imaged with the same telescope), so it is possible that our
2152 methods over-adapted to the beam and noise of CDFS. Additionally, CDFS contains a
2153 very bright source which may have caused artefacts throughout the field that are not
2154 present in ELAIS-S1. Further work is required to understand the differences between
2155 the fields and their effect on performance.

2156 Figure 4.15 reveals interesting behaviour of different classifier models at different
2157 flux cutoffs. Logistic regression and convolutional neural networks seem relatively
2158 independent of flux, with these models performing well on the fainter ELAIS-S1 com-
2159 ponents even when they were trained on the generally brighter components in CDFS.
2160 Conversely, random forests were sensitive to the changes in flux distribution between
2161 datasets. This shows that not all models behave similarly on radio data, and it is
2162 therefore important to investigate multiple models when developing machine learn-
2163 ing methods for radio astronomy.

2164 Appendix E (see Figure 4.17) shows examples of incorrectly cross-identified com-
2165 ponents in CDFS. On no such component do all classifiers agree. This raises the possi-
2166 bility of using the level of disagreement of an ensemble of binary classifiers as a mea-
2167 sure of the difficulty of cross-identifying a radio component, analogous to the consen-
2168 sus level for Radio Galaxy Zoo volunteers.

2169 Our methods can be easily incorporated into other cross-identification methods or
2170 used as an extra data source for source detection. For example, the scores output by
2171 our binary classifiers could be used to disambiguate between candidate host galaxies
2172 selected by model-based algorithms, or used to weight candidate host galaxies while
2173 a source detector attempts to associate radio components. Our method can also be

2174 extended using other data sources: ~~for~~For example, information from source identifi-
2175 cation algorithms could be incorporated into the feature set of candidate host galaxies.

2176 **4.6 Summary**

2177 We presented a machine learning approach for cross-identification of radio compo-
2178 nents with their corresponding infrared host ~~galaxy~~galaxies. Using the CDFS field of
2179 ATLAS as a training set we trained our methods on expert and crowdsourced cross-
2180 identification catalogues. Applying these methods on both fields of ATLAS, we found
2181 that:

- 2182 • Our method trained on ATLAS observations of CDFS generalised to ATLAS ob-
2183 servations of ELAIS-S1, demonstrating that training on a single patch of sky is
2184 a feasible option for training machine learning methods for wide-area radio sur-
2185 veys;
- 2186 • Performance was comparable to nearest neighbours even on resolved sources,
2187 showing that nearest neighbours is useful for datasets consisting mostly of unre-
2188 solved sources such as ATLAS and EMU;
- 2189 • Radio Galaxy Zoo-trained models performed comparably to expert-trained mod-
2190 els and outperformed Radio Galaxy Zoo, showing that crowdsourced labels are
2191 useful for training machine learning methods for cross-identification even when
2192 these labels are noisy;
- 2193 • ATLAS does not contain sufficient data to train or test machine learning cross-
2194 identification methods for extended radio sources. This suggests that if machine
2195 learning methods are to be used on EMU, a larger area of sky will be required
2196 for training and testing these methods. However, existing surveys like FIRST are
2197 likely too different from EMU to expect good generalisation.

2198 While our cross-identification performance is not as high as desired, we make no
2199 assumptions on the binary classification model used in our methods and so we expect
2200 the performance to be improved by further experimentation and model selection. Our
2201 method provides a useful framework for generalising cross-identification catalogues
2202 to other areas of the sky from the same radio survey and can be incorporated into
2203 existing methods. We have shown that citizen science can provide a useful dataset for
2204 training machine learning methods in the radio domain. Chapter 5 will extend this
2205 approach and confirm that dataset size is a key limitation by successfully applying it
2206 to a considerably larger dataset: FIRST.

2207 **4.7 Acknowledgements**

2208 This chapter and corresponding publication was made possible by the participation of
2209 more than 11 000 volunteers in the Radio Galaxy Zoo project. Their contributions are

2210 individually acknowledged at <http://rgzauthors.galaxyzoo.org>. Parts of this research
 2211 were conducted by the Australian Research Council Centre of Excellence for All-sky
 2212 Astrophysics (CAASTRO), through project number CE110001020. Partial support for
 2213 LR was provided by U.S. National Science Foundation grants AST1211595 and 1714205
 2214 to the University of Minnesota. HA benefitted from grant 980/2016-2017 of Universi-
 2215 dad de Guanajuato. We thank A. Tran and the ~~reviewer~~anonymous referee for their
 2216 comments on the manuscript this chapter was based on. Radio Galaxy Zoo makes
 2217 use of data products from the Wide-field Infrared Survey Explorer and the Very Large
 2218 Array. The Wide-field Infrared Survey Explorer is a joint project of the University of
 2219 California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Tech-
 2220 nology, funded by the National Aeronautics and Space Administration. The National
 2221 Radio Astronomy Observatory is a facility of the National Science Foundation oper-
 2222 ated under cooperative agreement by Associated Universities, Inc. The figures in this
 2223 work ~~made~~make use of Astropy, a community-developed core Python package for As-
 2224 tronomy (The Astropy Collaboration et al., 2018). The Australia Telescope Compact
 2225 Array is part of the Australia Telescope, which is funded by the Commonwealth of
 2226 Australia for operation as a National Facility managed by CSIRO. We acknowledge the
 2227 Gomeroi people as the traditional custodians of the Observatory site.

2228 **A Classification models**

2229 This appendix describes the three different models we used for binary classification in
 2230 this chapter: logistic regression, convolutional neural networks, and random forests.

2231 **A.1 Logistic regression**

2232 Logistic regression is linear in the feature space and outputs the probability that the
 2233 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:

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

2234 The logistic regression model is fully differentiable, and the parameters \vec{w} can therefore
 2235 be learned using gradient-based optimisation methods. We used the scikit-learn
 2236 (Pedregosa et al., 2011) implementation of logistic regression with balanced classes.

2237 **A.2 Convolutional neural networks**

2238 Convolutional neural networks (CNN) are a biologically-inspired prediction model
 2239 for prediction with image inputs. The input image is convolved with a number of

2240 filters to produce output images called feature maps. These feature maps can then
 2241 be convolved again with other filters on subsequent layers, producing a network of
 2242 convolutions. The whole network is differentiable with respect to the values of the
 2243 filters and the filters can be learned using gradient-based optimisation methods. The
 2244 final layer of the network is logistic regression, with the convolved outputs as input
 2245 features. For more detail, see subsection II.A, LeCun et al. (1998). We used KERAS
 2246 (Chollet et al., 2015) to implement our CNN, accounting for class imbalance by reweighting
 2247 the classes.

2248 CNNs have recently produced good results on large image-based datasets in astronomy
 2249 (Dieleman et al., 2015; Lukic et al., 2018, e.g.). We employed only a simple CNN model
 2250 in Chapter 4 as a proof of concept that CNNs may be used for class probability prediction
 2251 on radio images. The model architecture we used is shown in Figure 4.16.

2252 A.3 Random forests

2253 Random forests are an ensemble of decision trees (Breiman, 2001). They consider
 2254 multiple subsamples of the training set, where each subsample is sampled with replacement
 2255 from the training set. For each subsample a decision tree classifier is constructed by
 2256 repeatedly making axis-parallel splits based on individual features. In a random forest
 2257 the split decision is taken based on a random subset of features. To classify a new data
 2258 point, the random forest takes the weighted average of all classifications produced by
 2259 each decision tree. In Chapter 4 we used the scikit-learn (Pedregosa et al., 2011)
 2260 implementation of random forests with 10 trees, the information entropy split criterion,
 2261 a minimum leaf size of 45, and balanced classes.

2262 B Accuracy tables

2263 This section contains tables of accuracy for our cross-identification method applied
 2264 to CDFS and ELAIS-S1. In Table 4.3 and Table 4.4 we list the balanced accuracies
 2265 of our Chapter 4 classifiers on the cross-identification task for CDFS and ELAIS-S1
 2266 respectively, averaged over each set of training quadrants. In Table 4.5 and Table 4.6
 2267 we list the balanced accuracies of classifiers on the cross-identification task for CDFS
 2268 and ELAIS-S1 respectively, averaged over each set of training quadrants.

2269 C SWIRE object scores

2270 This appendix contains scores predicted by our Chapter 4 binary classifiers for each
 2271 SWIRE object within 1' of a radio component in CDFS and ELAIS-S1. Scores for SWIRE CDFS
 2272 objects are shown in Table 4.7 and scores for SWIRE ELAIS-S1 are shown in Table 4.8.
 2273 For CDFS, the score for an object in a quadrant is predicted by binary classifiers trained
 2274 on all other quadrants. For ELAIS-S1, we show the scores predicted by binary classifiers
 2275 trained on each CDFS quadrant. Note that these scores have *not* been weighted by

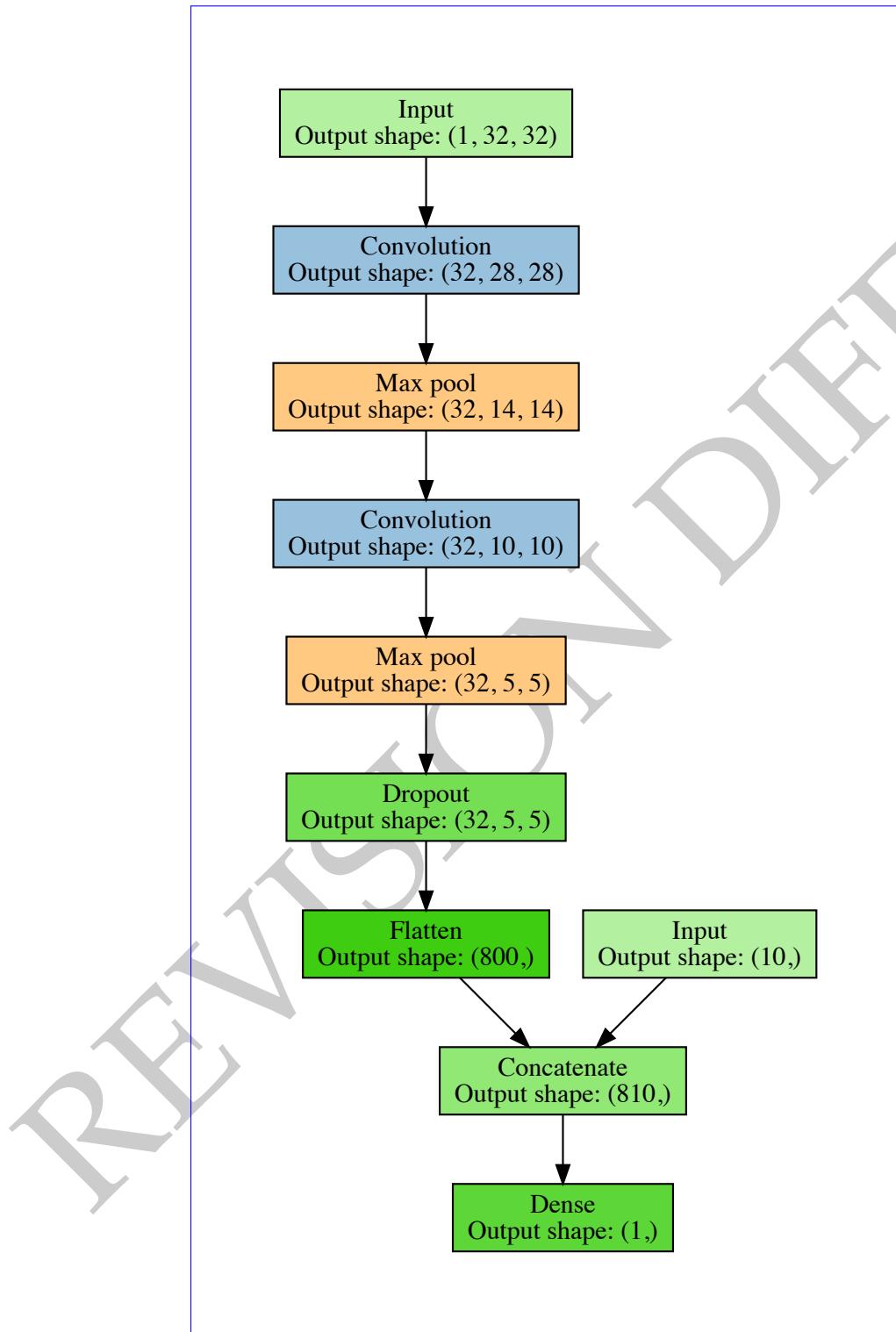


Figure 4.16: Architecture of our CNN. Parenthesised numbers indicate the size of output layers as a tuple (width, height, depth). The concatenate layer flattens the output of the previous layer and adds the 10 features derived from the candidate host in SWIRE, i.e. the flux ratios, stellarity indices, and distance. The dropout layer randomly sets 25 per cent of its inputs to zero during training to prevent overfitting. Diagram based on <https://github.com/dnouri/nolearn>.

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 4.8) 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
	CNN	92.6 ± 0.7	91.2 ± 0.5	92.0 ± 0.6
	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
	CNN	89.4 ± 0.6	89.6 ± 1.3	89.4 ± 0.5
	RF	94.5 ± 0.2	95.8 ± 0.4	94.7 ± 0.3

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 4.8) 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 4.8) 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 4.8) is also shown.

Labeller	Classifier	Mean ‘Compact’ accuracy (per cent)	Mean ‘Resolved’ accuracy (per cent)	Mean ‘All’ accuracy (per cent)
Middelberg	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
	Perfect	99.6 ± 0.0	99.8 ± 0.0	99.6 ± 0.0
	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
Norris	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
RGZ	Labels	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
	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

2276 Gaussians. These are partial tables, and the full tables are available online at the
 2277 Monthly Notices of the Royal Astronomical Society website⁷.

2278 The columns of the score tables are defined as follows:

- 2279 • *SWIRE*—SWIRE designation for candidate host galaxy.
- 2280 • *RA*—Right ascension (J2000).
- 2281 • *Dec*—Declination (J2000).
- 2282 • *Expert host*—Whether the candidate host galaxy is a host galaxy according to
 2283 Norris et al. (2006) or Middelberg et al. (2008) cross-identifications of CDFS and
 2284 ELAIS-S1 respectively.
- 2285 • *RGZ host*—Whether the candidate host galaxy is a host galaxy according to Radio
 2286 Galaxy Zoo cross-identifications (Wong et al., in prep.). This is always ‘no’ for
 2287 ELAIS-S1 objects.
- 2288 • *C/L/D*—Score assigned by binary classifier C trained on label set L of D candidate
 2289 host galaxies. C may be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression, or
 2290 random forests respectively. L may be ‘Norris’ or ‘RGZ’ for expert and Radio
 2291 Galaxy Zoo labels respectively. D may be ‘All’, ‘Compact’, or ‘Resolved’ for each
 2292 respective subset defined in Section 4.3.6.

2293 D ATLAS component cross-identifications

2294 This section contains cross-identifications predicted by our cross-identification method
 2295 for each ATLAS radio component in CDFS and ELAIS-S1. Cross-identifications for
 2296 ATLAS CDFS components are shown in Table 4.9 and cross-identifications for ATLAS ELAIS-S1
 2297 are shown in Table 4.10. For CDFS, the cross-identification for a component in a quadrant
 2298 is predicted using our method with binary classifiers trained on all other quadrants.
 2299 For ELAIS-S1, we show the cross-identifications predicted by our method using binary
 2300 classifiers trained on each CDFS quadrant. For CDFS, we also show the Radio Galaxy
 2301 Zoo consensus, which is a proxy for the difficulty of cross-identifying a component
 2302 (Wong et al., in prep.). These are partial tables, and the full tables are available online
 2303 at the Monthly Notices of the Royal Astronomical Society website⁸.

2304 The columns of the cross-identification tables are defined as follows:

- 2305 • *ATLAS*—ATLAS designation for radio component.
- 2306 • *RA*—Right ascension of radio component (J2000).
- 2307 • *Dec*—Declination of radio component (J2000).
- 2308 • *CID*—Radio Galaxy Zoo component ID.

⁷<https://doi.org/10.1093/mnras/sty1308>

⁸<https://doi.org/10.1093/mnras/sty1308>

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

SWIRE	RA	Dec	Expert host	RGZ			CNN			RGZ		
				All	Resolved	Compact	All	Resolved	Compact	All	Resolved	Compact
J032603.15-284708.5	51.5132	-28.7857	yes	no	0.5838	0.4697	0.4848	0.3754	0.3881	0.3404	0.3404	0.3404
J032603.39-284010.1	51.5142	-28.6695	no	no	0.0373	0.5814	0.4878	0.7896	0.7616	0.4668	0.4668	0.4668
J032603.44-284210.1	51.5144	-28.7028	no	no	0.0232	0.4891	0.5101	0.4319	0.4298	0.3474	0.3474	0.3474
J032603.44-284222.2	51.5143	-28.7062	no	no	0.0006	0.4164	0.5216	0.0400	0.0444	0.276	0.276	0.276
J032603.45-284748.4	51.5144	-28.7968	no	no	0.0014	0.4914	0.4865	0.1904	0.1895	0.1467	0.1467	0.1467
J032603.50-284637.0	51.5146	-28.7770	no	no	0.0074	0.4144	0.5382	0.1418	0.1515	0.1166	0.1166	0.1166
J032603.60-284627.4	51.5150	-28.7743	no	no	0.0012	0.4578	0.5165	0.0850	0.0904	0.0484	0.0484	0.0484
J032603.63-283840.5	51.5151	-28.6446	no	no	0.0021	0.4153	0.5577	0.1678	0.1746	0.323	0.323	0.323
J032603.66-283822.8	51.5153	-28.6397	no	no	0.0001	0.4752	0.5009	0.0864	0.0861	0.0613	0.0613	0.0613
J032603.75-284014.1	51.5156	-28.6706	no	no	0.0547	0.3408	0.5388	0.4889	0.5242	0.7301	0.7301	0.7301
LR				RGZ			Norris			RGZ		
All	Norris	Compact	Resolved	All	Resolved	Compact	All	Resolved	Compact	All	Resolved	Compact
0.2489	0.0009	0.1557	0.2939	0.0007	0.1174	0.8922	0.8018	0.8732	0.7167	0.6599	0.7801	0.7801
0.0183	0.1646	0.1480	0.7637	0.0605	0.6070	0.0000	0.0000	0.0000	0.1629	0.0519	0.1275	0.1275
0.0155	0.0164	0.0815	0.3714	0.5626	0.2488	0.0000	0.0734	0.0000	0.1315	0.2116	0.4150	0.4150
0.0005	0.0006	0.0175	0.0460	0.0810	0.0299	0.2656	0.1418	0.0000	0.7631	0.8166	0.5378	0.5378
0.0013	0.0037	0.0160	0.1792	0.0663	0.1821	0.0000	0.0000	0.0000	0.0255	0.0000	0.0000	0.0000
0.0047	0.0010	0.0337	0.1284	0.2198	0.0694	0.0720	0.0000	0.0000	0.6240	0.6681	0.6704	0.6704
0.0008	0.0006	0.0374	0.1053	0.1424	0.0807	0.1231	0.0876	0.0000	0.8517	0.7532	0.7019	0.7019
0.0021	0.0073	0.0386	0.1482	0.0403	0.1210	0.0000	0.0532	0.0000	0.0302	0.0000	0.0000	0.0000
0.0001	0.0004	0.0038	0.0854	0.0447	0.0514	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0542	0.2712	0.2318	0.5026	0.5631	0.5032	0.0595	0.0545	0.0000	0.4289	0.0789	0.1420	0.1420

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

SWIRE	RA			Dec			Expert host			RGZ			Norris			CNN			RGZ		
	All	Compact	Resolved	All	Compact	Resolved	All	Compact	Resolved	All	Compact	Resolved	All	Compact	Resolved	All	Compact	Resolved	All	Compact	Resolved
J002925.73-440256.2	7.3572	-44.0490	yes	no	0.9537	0.8638	0.5552	0.9195	0.9037	0.9371	0.5640	0.7740	0.7474	0.7952	0.5640	0.7740	0.7474	0.7952	0.5640	0.7740	0.7474
J002926.14-440249.0	7.3590	-44.0470	no	no	0.7361	0.8752	0.5640	0.7740	0.7474	0.7952	0.7275	0.6894	0.7197	0.7197	0.7275	0.6894	0.7197	0.7197	0.7275	0.6894	0.7197
J002926.52-440247.0	7.3605	-44.0464	no	no	0.3390	0.8338	0.5556	0.7275	0.6894	0.7197	0.3434	0.3306	0.3292	0.3292	0.3434	0.3306	0.3292	0.3292	0.3434	0.3306	0.3292
J002926.63-440301.1	7.3610	-44.0503	no	no	0.2108	0.8251	0.5623	0.7275	0.6894	0.7197	0.2228	0.2340	0.2133	0.2133	0.2228	0.2340	0.2133	0.2133	0.2228	0.2340	0.2133
J002927.13-440232.6	7.3631	-44.0424	no	no	0.0339	0.8479	0.5669	0.5853	0.5148	0.5159	0.0406	0.8345	0.5540	0.2702	0.2340	0.2133	0.2133	0.2702	0.2340	0.2133	
J002927.28-440245.3	7.3637	-44.0459	no	no	0.0406	0.8345	0.5540	0.2702	0.2340	0.2133	0.0116	0.8267	0.5746	0.2228	0.2182	0.2028	0.2028	0.2228	0.2182	0.2028	
J002927.44-440238.5	7.3644	-44.0440	no	no	0.0116	0.8266	0.5746	0.2228	0.2182	0.2028	0.0024	0.8626	0.5791	0.2297	0.2182	0.2028	0.2028	0.2297	0.2182	0.2028	
J002928.08-440230.3	7.3670	-44.0418	no	no	0.0011	0.8159	0.5514	0.2297	0.2182	0.2028	0.0003	0.8405	0.5668	0.0377	0.0384	0.0271	0.0271	0.0377	0.0384	0.0271	
J002928.80-440306.8	7.3700	-44.0519	no	no	0.0003	0.8405	0.5668	0.0236	0.0226	0.0136	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
LR			LR			RF			RF			RF			RF			RF			
All	0.9722	0.9955	0.8769	All	0.9933	0.9934	0.9658	0.8824	0.9664	0.7950	0.8078	0.9227	0.7677	All	0.9955	0.9955	0.9955	0.9955	0.9955	0.9955	
Compact	0.4669	0.0111	0.4249	Compact	0.3926	0.2220	0.5947	0.2077	0.0000	0.1613	0.1876	0.0852	0.4546	Compact	0.0111	0.4249	0.3926	0.2220	0.5947	0.2077	
Resolved	0.2264	0.0254	0.2389	Resolved	0.6275	0.3033	0.6812	0.1347	0.0857	0.0399	0.3582	0.4854	0.5347	Resolved	0.0254	0.2389	0.6275	0.3033	0.6812	0.1347	
All	0.0603	0.0007	0.0734	All	0.0688	0.0141	0.1581	0.0917	0.0000	0.0399	0.2846	0.1245	0.2833	All	0.0007	0.0734	0.0688	0.0141	0.1581	0.0917	
Compact	0.0248	0.0334	0.0301	Compact	0.5735	0.5065	0.5265	0.1977	0.1507	0.0000	0.3334	0.6593	0.3995	Resolved	0.0334	0.0248	0.0301	0.5735	0.5065	0.5265	
Resolved	0.0173	0.0016	0.0359	Resolved	0.1056	0.0492	0.1456	0.0000	0.0000	0.0000	0.0000	0.0000	0.0287	All	0.0016	0.0173	0.0359	0.1056	0.0492	0.1456	
All	0.0064	0.0049	0.0187	All	0.1981	0.1534	0.1493	0.0000	0.0000	0.0000	0.0000	0.0000	0.0287	Resolved	0.0049	0.0187	0.1981	0.1534	0.1493	0.0000	
Compact	0.0020	0.0005	0.0239	Resolved	0.1337	0.1001	0.1310	0.0000	0.0000	0.0358	0.0000	0.0000	0.0190	All	0.0005	0.0239	0.1337	0.1001	0.1310	0.0000	
Resolved	0.0008	0.0013	0.0119	Resolved	0.0280	0.0361	0.0205	0.1171	0.0000	0.0000	0.0873	0.0383	0.0000	0.0000	All	0.0013	0.0119	0.0280	0.0361	0.0205	0.1171
All	0.0004	0.0014	0.0095	All	0.0339	0.0408	0.0136	0.0000	0.0000	0.0000	0.0000	0.0000	0.1480	Resolved	0.0014	0.0095	0.0339	0.0408	0.0136	0.0000	

- 2309 • *Zooniverse ID*—Radio Galaxy Zoo Zooniverse ID.
- 2310 • *Norris/Middleberg*—Designation of SWIRE cross-identification from Norris et al. (2006)
- 2311 or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- 2312 • *Norris/Middleberg RA*—Right ascension of SWIRE cross-identification from Norris et al. (2006)
- 2313 or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- 2314 • *Norris/Middleberg Dec*—Right ascension of SWIRE cross-identification from Norris et al. (2006)
- 2315 or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- 2316 • *RGZ*—Designation of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.)
- 2317 ~
- 2318 • *RGZ RA*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo
- 2319 (Wong et al., in prep.).
- 2320 • *RGZ Dec*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo
- 2321 (Wong et al., in prep.).
- 2322 • *RGZ radio consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on
- 2323 the radio component configuration.
- 2324 • *RGZ IR consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on the
- 2325 host galaxy of this radio component.
- 2326 • *C / L / D*—Designation of SWIRE cross-identification made by our method using
- 2327 classification model C trained on label set L of D candidate host galaxies. C may
- 2328 be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or random forests respectively.
- 2329 L may be ‘Norris’ or ‘RGZ’ for expert and Radio Galaxy Zoo labels respectively.
- 2330 D may be ‘All’, ‘Compact’, or ‘Resolved’ for each respective subset defined in
- 2331 Section 4.3.6.~
- 2332 • *C / L / D RA*—Right ascension (J2000) of SWIRE cross-identification made by
- 2333 our method using classification model C trained on label set L of D candidate
- 2334 host galaxies. C, L, and D are defined as for designation.
- 2335 • *C / L / D Dec*—Declination (J2000) of SWIRE cross-identification made by our
- 2336 method using classification model C trained on label set L of D candidate host
- 2337 galaxies. C, L, and D are defined as for designation.

2338 E *Cross-identification figures*

2339 Figure 4.17 shows figures of our cross-identifications of each ATLAS radio component
 2340 in CDFS and ELAIS-S1. There are just five examples shown here, but all 469 examples
 2341 are available online at the *Monthly Notices of the Royal Astronomical Society* website⁹.

⁹<https://doi.org/10.1093/mnras/sty1308>

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

ATLAS	RA	Dec	CID	Zooniverse ID	Norris		RGZ		RGZ	
					RA	Dec	RA	Dec	radio consensus	IR consensus
J032602-32-284708.1C	51.5117	-28.7856	C10412	ARG0003rb2	J032603-15-284708.5	51.5132	-28.7857	51.5642	0.4516	0.3214
J032615-49-284629.4C	51.5646	-28.7749	C10614	ARG0003fr	J032615-41-284630.7	51.5642	-28.7752	J032615-41-284630.7	0.2941	0.8000
J032615-55-280559.8C	51.5648	-28.1000	C10820	ARG0003fb	J032615-52-280559.8	51.5647	-28.1000	J032615-52-280559.8	0.5625	0.8333
J032617-35-280710.2C	51.5723	-28.1195	C10050C1	ARG0003t2	J032617-39-280707.2	51.5746	-28.1187	0.4146	1.0000	
J032625-13-280909.8C	51.6047	-28.1527	C10409	ARG0003ra2	J032623-19-280910.1	51.6050	-28.1528	0.5158	0.6667	1.0000
J032629-10-280601.1C	51.6213	-28.1139	C10963	ARG0003rd	J032629-13-280605.7	51.6214	-28.1141	J032629-14-280606.7	0.3333	1.0000
J032629-51-284052.7C	51.6234	-28.6813	C10394	ARG0003re	J032629-54-284055.8	51.6231	-28.6822	J032629-54-284055.8	0.2676	1.0000
J032629-52-284753.5C	51.6247	-28.7982	C10229	ARG0003rw	J032629-81-284754.4	51.6242	-28.7985	J032629-81-284754.4	1.0000	0.8571
J032630-06-283657.3C	51.6278	-28.6159	C10172C1	ARG0003rz	J032630-64-283658.0	51.6277	-28.6161	J032628-56-283744.8	0.3611	0.7308
J032634-59-280228.8C	51.6441	-28.3397	C10757	ARG0003zv	J032634-58-280228.8	51.6441	-28.3397	J032631-16-281941.0	0.5781	0.5405

Compact	Norris		CNN		LR		RGZ	
	RA	Dec	RA	Dec	RA	Dec	RA	Dec
J032602-36-284711.5	51.5098	-28.7865	J032602-36-284711.5	51.5098	-28.7865	J032602-36-284711.5	51.5098	-28.7865
J032615-31-284630.7	51.5642	-28.7732	J032615-31-284630.7	51.5642	-28.7732	J032615-31-284630.7	51.5642	-28.7732
J032615-52-280559.8	51.5647	-28.1000	J032615-52-280559.8	51.5647	-28.1000	J032615-52-280559.8	51.5647	-28.1000
J032617-59-280707.2	51.5746	-28.1187	J032617-59-280707.2	51.5746	-28.1187	J032617-59-280707.2	51.5746	-28.1187
J032621-19-280901.0	51.6050	-28.1528	J032621-19-280901.0	51.6050	-28.1528	J032621-19-280901.0	51.6050	-28.1528
J032629-13-280650.7	51.6214	-28.1141	J032629-13-280650.7	51.6214	-28.1141	J032629-13-280650.7	51.6214	-28.1141
J032629-54-284651.9	51.6231	-28.6811	J032629-54-284651.9	51.6231	-28.6811	J032629-54-284651.9	51.6231	-28.6811
J032629-81-284754.4	51.6242	-28.7985	J032629-81-284754.4	51.6242	-28.7985	J032629-81-284754.4	51.6242	-28.7985
J032630-04-283658.0	51.6277	-28.6161	J032630-04-283658.0	51.6277	-28.6161	J032630-04-283658.0	51.6277	-28.6161
J032634-58-280228.8	51.6441	-28.3397	J032634-58-280228.8	51.6441	-28.3397	J032634-58-280228.8	51.6441	-28.3397

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

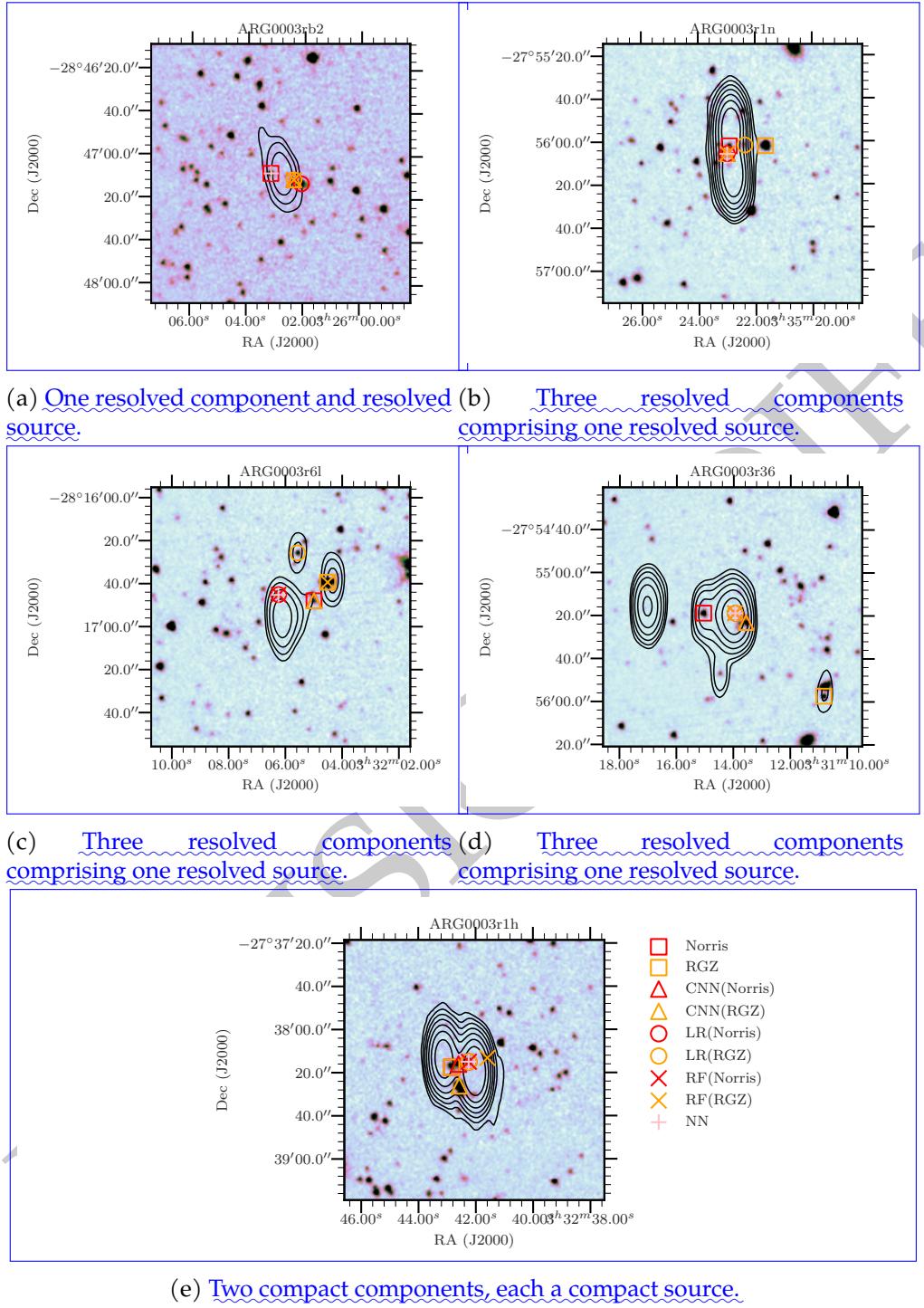


Figure 4.17: Examples of resolved sources with high disagreement between cross-identifiers. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2. The background image is the $3.6 \mu\text{m}$ SWIRE image. Binary classifier model/training set combinations are 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 and ‘RGZ’ refers to the Radio Galaxy Zoo labels. The cross-identification made by nearest neighbours is shown by ‘NN’.

2343

Radio Luminosity Functions

2344 This chapter is based on my paper *Radio Galaxy Zoo: Radio Luminosity Functions of Ex-
2345 tended Sources*, by M. J. Alger, O. I. Wong, C. S. Ong, N. M. McClure-Griffiths, H. An-
2346 dernach, L. Rudnick, S. S. Shabala, A. F. Garon, J. K. Banfield, A. D. Kapińska, R. P.
2347 Norris, and A. J. M. Thomson; to be submitted. Some minor changes have been made
2348 to match the rest of this thesis.

2349

2350 In Chapter 4, we concluded that our machine learning cross-identification approach
2351 was sound in principle, but the small size and lack of morphological complexity in AT-
2352 LAS prevented training a good classifier. To demonstrate this, we turned our attention
2353 to the Faint Images of the Radio Sky at Twenty Centimeters (FIRST) survey, which con-
2354 tained many more sources and many more morphologically complex sources due to
2355 its higher resolution and sample size. If cross-identification of this dataset was success-
2356 ful, we could use the resulting cross-identification catalogue to investigate statistical
2357 properties of radio galaxies at larger scales than previously possible. The object of in-
2358 terest would be the radio luminosity function (RLF), a statistical description of how
2359 common radio galaxies of various luminosities are throughout the universe. The RLF
2360 would be specific to only extended radio galaxies, and it is this class of galaxies that we
2361 would expect to maximally impact its local environment through direct, mechanical
2362 energy input.

2363 In this chapter, we use a machine learning model trained on the Radio Galaxy Zoo
2364 citizen science project to cross-identify 244 846 radio components of extended radio
2365 sources from FIRST with 158 337 host galaxies observed in the mid-infrared from the
2366 AllWISE source catalogue. 34 305 hosts have a spectroscopic redshift in the Sloan Digi-
2367 tal Sky Survey (SDSS). This is the largest available sample of cross-identified extended
2368 radio sources to date at the time of writing. The large sample size allows us to estimate
2369 the RLF of extended radio galaxies and to further characterise the sample with frac-
2370 tional radio luminosity functions. The extended source RLF can be used to estimate the
2371 mechanical energy injected into the local environment of active galactic nuclei (AGN).
2372 We estimate from the observed radio emission that inferred extended radio jets from
2373 AGN contribute between 1.3×10^{30} and 1.2×10^{32} W Mpc $^{-3}$ of mechanical energy to
2374 their environment throughout the low-redshift ($z < 0.6$) Universe. By further visual
2375 verification we also find 40 radio galaxies with projected size larger than 1 Mpc. Our re-

2376 sults directly demonstrate the impact that automated machine learning methods could
2377 have on future wide-area radio surveys.

2378 1 Introduction

2379 Radio active galactic nuclei (AGN) are some of the most violent and energetic objects
2380 in the Universe. Supermassive black holes at the centre of galaxies can produce large,
2381 luminous radio jets that may feed lobes extending up to megaparsec scales. As a vital
2382 part of galaxy evolution, we want to understand how radio AGN interact with their
2383 host galaxies, in what kinds of galaxies these AGN are found, and how these popula-
2384 tions change over cosmic time.

2385 Mechanical energy from AGN is needed to explain the star-formation history of
2386 galaxies throughout the Universe (e.g. Hardcastle & Croston, 2020; Raouf et al., 2017).
2387 The radio luminosity function (RLF) of extended radio sources would characterise the
2388 energy output of galaxies capable of significant energy impact into their local environ-
2389 ment. The local RLF of primarily compact sources is estimated and described in detail
2390 by Mauch and Sadler (2007), Pracy et al. (2016) and Condon et al. (2019), each with
2391 complete volume-limited samples. However, the RLF of extended radio sources has
2392 not been characterised as a separate sample, because estimating this RLF would need
2393 a large number of extended sources with known redshifts. Such redshifts are usually
2394 obtained by cross-matching the radio sources with their host galaxies in the optical
2395 or infrared. While this cross-matching is relatively straightforward for compact/unre-
2396 solved sources, and can therefore be automated (e.g. Kimball & Ivezić, 2008), extended
2397 radio emission may be complex and may not be co-located with the host galaxy. Cross-
2398 identification of complex, extended sources is therefore usually done manually, limit-
2399 ing the sample size.

2400 Wide-area radio surveys like the Very Large Array Faint Images of the Sky at Twenty
2401 Centimeters survey (FIRST; White et al., 1997) have led to catalogues of hundreds of
2402 thousands of radio objects, most of which are associated with radio AGN (Sadler et
2403 al., 2002). While individual galaxy properties vary significantly between galaxies, the
2404 large sample size of such surveys allows us to examine reliable bulk statistics that may
2405 tightly constrain models and theory of radio galaxy evolution and formation (Condon,
2406 1992). With a sufficiently large sample (of size N such that $N \gg \sqrt{N}$; Condon,
2407 1991) we can also divide the RLF into fractions based on the physical properties of
2408 each galaxy. These *fractional RLFs* show how different physical processes comprise the
2409 luminosity distribution, and can be used to investigate how these processes relate to
2410 the properties of the AGN.

In this chapter we calculate the RLF for extended radio sources in FIRST. Throughout this chapter 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 defined 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., 2011) 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 et al., 2015). 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 chapter, 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 2. Our approach is discussed in Section 3. Radio luminosity functions of extended sources are presented in Section 4 and we discuss these functions in Section 5.

Throughout this chapter 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).

2 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 shown in Figure 5.6.

2452 **2.1 RGZ**

2453 RGZ is a citizen science project that aims to cross-identify complex radio sources with
 2454 mid-infrared host galaxies with the help of volunteers. The first RGZ data release con-
 2455 tains around 75 000 cross-identifications of a random subset of extended sources in
 2456 FIRST with their host galaxies in AllWISE. For more details on RGZ see Banfield et al.
 2457 (2015). This catalogue has also been used in other machine learning contexts, includ-
 2458 ing supervised learning for source aggregation (Wu et al., 2019) and unsupervised
 2459 learning methods (Galvin et al., 2019; Ralph et al., 2019). We discard the RGZ sources
 2460 without a host galaxy detected in AllWISE and use the remaining 41 446 sources for
 2461 training.

2462 The RGZ catalogue only contains sources with at least 0.65 weighted volunteer
 2463 agreement, equivalent to approximately 80 per cent reliability. This implicitly selects
 2464 for less complex sources, since the volunteer agreement is a proxy for the difficulty of
 2465 cross-identifying a source (Wong et al. in prep).

2466 **2.2 FIRST**

2467 We use radio imagery and select radio components from the FIRST survey and associ-
 2468 ated catalogue respectively (Helfand et al., 2015; White et al., 1997). FIRST is a 1.4 GHz
 2469 radio survey covering 10 575 deg² of the sky north of Dec = −10° with an angular res-
 2470 olution of 5.4''. At a detection limit of 1 mJy, the catalogue contains 946 432 radio
 2471 components.

2472 For both training and prediction, we make use of image cutouts from FIRST cen-
 2473 tered on mid-infrared candidate hosts. We predict host galaxies for the 244 846 ex-
 2474 tended FIRST components detected at > 10 σ (about 1.5 mJy beam^{−1}, per Banfield
 2475 et al., 2015) that have complete radio imagery for all candidate hosts. [The numbers of](#)
 2476 [objects removed by our quality filters are shown in](#) Appendix F.

2477 **2.3 AllWISE**

2478 We use the sky coordinates and magnitudes from the AllWISE (Cutri et al., 2013) cata-
 2479 logue during training and prediction. AllWISE is an all-sky catalogue of mid-infrared
 2480 objects detected by the *Wide-field Infrared Survey Explorer* (WISE Wright et al., 2010) at
 2481 3.4, 4.6, 12 and 22 μm wavelengths (called W1–W4, respectively). AllWISE contains
 2482 over 747 million objects detected above 5 σ at 3.4 and 4.6 μm. WISE has an angular
 2483 resolution of 6.1'' and 6.4'' at these wavelengths, respectively.

2484 We store AllWISE objects in a MongoDB¹ database with a geospatial index on the
 2485 right ascension and declination. This allows us to perform fast spatial lookups. The
 2486 geospatial index in MongoDB assumes a perfectly spherical Earth with a fixed radius,
 2487 allowing us to use it for sky coordinate searches. We use the right ascension and decli-
 2488 nation of AllWISE sources to generate candidate hosts by searching for infrared sources
 2489 near FIRST components.

¹<https://www.mongodb.com/>

Table 5.1: Medians and standard deviations used to normalise input features [for our classifiers](#).

Feature	Median	Standard deviation
Radio image	13.2 μ Jy	3.01 mJy
W1 – W2	0.289 mag	0.378 mag
W1 – W3	4.350 mag	1.067 mag
W1 – W4	7.853 mag	1.144 mag
W2 – W3	4.016 mag	0.958 mag
W2 – W4	7.541 mag	1.046 mag
W3 – W4	3.518 mag	0.409 mag
W1	16.659 mag	1.154 mag

2490 2.4 SDSS

2491 While we do not use data from the Sloan Digital Sky Survey Data Release 15 (SDSS;
 2492 Aguado et al., 2019) for training or prediction, we do use SDSS for spectroscopic red-
 2493 shifts of our host galaxies. These redshifts are required to calculate the radio luminos-
 2494 ity of our sources. We use CDS X-Match² to match each infrared host galaxy to the
 2495 closest source imaged by SDSS to within 5''. This results in 34 305 spectroscopic red-
 2496 shifts for our 158 337 total host galaxies. To estimate the rate of false association with
 2497 SDSS sources, we add a 1' offset to all host positions and redo the matching process.
 2498 With this method we estimate a 0.4 per cent rate of false association.

2499 3 Method

2500 We apply the binary classification cross-identification method (BXID) following Alger
 2501 et al. (2018). This method casts cross-identification as a classification problem where
 2502 infrared ‘candidate host galaxies’ are classified as either being host galaxies or not. A
 2503 classifier is trained on examples of host galaxies and non-host galaxies drawn from
 2504 a cross-identification catalogue, for which we use RGZ. Other related algorithms de-
 2505 veloped to automatically cross-identify radio objects include Bayesian methods (Fan
 2506 et al., 2015), likelihood ratio (Weston et al., 2018), positional matching (e.g. Kimball &
 2507 Ivezić, 2008; Middelberg et al., 2008; Norris et al., 2006), and positional/image hybrid
 2508 approaches (van Velzen et al., 2012), but these methods do not make use of existing
 2509 cross-identification catalogues and most assume compact radio sources or that the pro-
 2510 jected 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 W3 – W4 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

²<http://cdsxmatch.u-strasbg.fr/>

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.

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

2511 The medians and standard deviations are reported in Table 5.1. These values are com-
2512 puted across the training set.

2513 For each FIRST component we generate a set of candidate host galaxies. An AllWISE
2514 object is considered a candidate host for a radio component if it is within $\sqrt{2} \times 1.5'$ of
2515 the centre of the two-dimensional Gaussian fit for that component. This search radius
2516 is the maximum angular distance that a host galaxy can be located in RGZ due to the 3'-
2517 wide square images shown to volunteers. Candidate hosts are assigned binary labels:
2518 All candidates identified as host galaxies in RGZ are assigned a positive label and all
2519 others are assigned a negative label. Following Alger et al. (2018) we train a convolutional
2520 neural network (CNN) on the labelled candidate hosts. We base our model on
2521 ResNet18 (He et al., 2016) pretrained on the ImageNet classification task, with the final
2522 layer removed and replaced by a logistic regression model. Non-image features (i.e.
2523 colours and the 3.6 μm magnitude) are concatenated with the features that are output
2524 by the final ResNet18 layer. Using Adam (Kingma & Ba, 2015) to optimise our weights,
2525 we train this model on our task until binary cross-entropy loss starts to increase on a
2526 randomly-selected 20 per cent validation set. We use PyTorch (Paszke et al., 2017) to
2527 implement this model. The scores of each FIRST component are weighted by a one-
2528 dimensional Gaussian function of angular separation, and the candidate maximising
2529 this weighted score is selected as the host galaxy. We set the standard deviation of the
2530 Gaussian to 120'' as this provides good empirical results.

2531 92 per cent of the host galaxies in RGZ are also detected as host galaxies in RGZ-Ex.
2532 The mean volunteer agreement on all RGZ sources with detected hosts was 95^{+5}_{-13} per
2533 cent, compared to 88^{+12}_{-17} per cent for sources with hosts not in RGZ-Ex. Incorrect cross-
2534 identifications can be considered a source of noise in the statistics. In future work
2535 we will design a way for BXID to output an ‘uncertainty’ so sources with uncertain
2536 cross-identifications can be removed from calculations. We quantify the reliability of
2537 RGZ-Ex in Section 3.1.

2538 We note that BXID necessarily identifies a host galaxy for all radio emission, even
2539 when it does not make sense to do so. In our current work we treat this as a source of
2540 noise. Future extensions to BXID will allow it to output ‘no detected host’.

2541 3.1 Visual verification

2542 To quantify the reliability of RGZ-Ex, ~~some of the authors we~~ (M.A., O.W., A.K., N.M.,
2543 and A.T.) visually verified a randomly selected set of 200 radio components/host galaxy
2544 pairs in RGZ-Ex. For each pair we decided whether the radio component matched its
2545 identified host or not. This allowed us to estimate the accuracy of radio component-
2546 infrared host pairs identified by BXID. Verification of component-host pairs is noisy

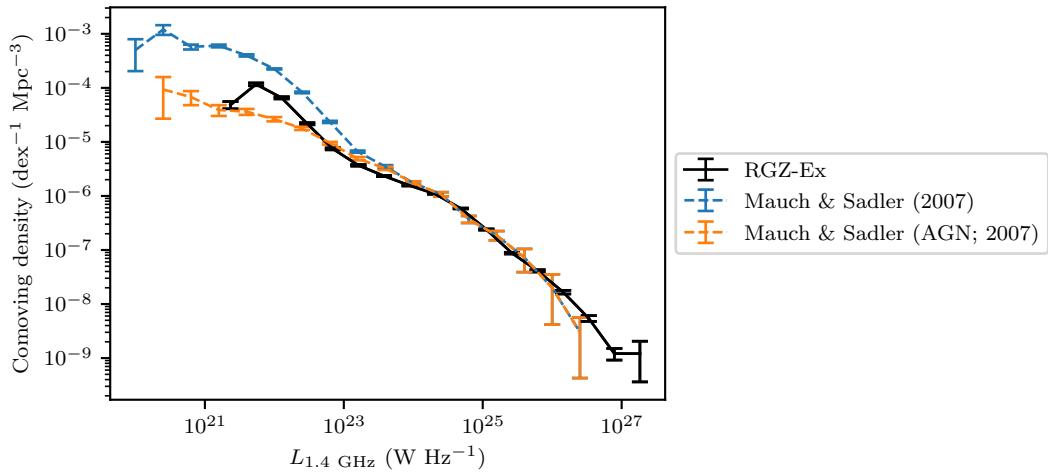


Figure 5.1: RGZ-Ex radio luminosity function compared with the RLFS of Mauch and Sadler (2007).

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 verification 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 Appendix G.

4 Radio luminosity functions

In this section we present our radio luminosity functions (RLFs) derived from the RGZ-Ex catalogue. We first eliminate sources and components with high sidelobe probability according to the FIRST catalogue and low BXID scores (see Appendix F for a diagram of this filtering process). 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 Appendix H. 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 $3.4 \mu\text{m}$, 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 et al., 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 Appendix H. Completeness estimates are shown in Appendix I. We discuss biases in our methods and results in Section 5.1.

We compare our RLFs with Mauch and Sadler (2007), who estimated RLFs from 7 824 manually cross-identified radio sources in the NRAO VLA Sky Survey (NVSS; Condon et al., 1998). 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 (2007).

In Figure 5.1 we show the RLF derived from RGZ-Ex along with the RLFs from Mauch and Sadler (2007). 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 (2007) RLF. RGZ-Ex generally finds less comoving density than Mauch and Sadler (2007), 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 (2007).

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).

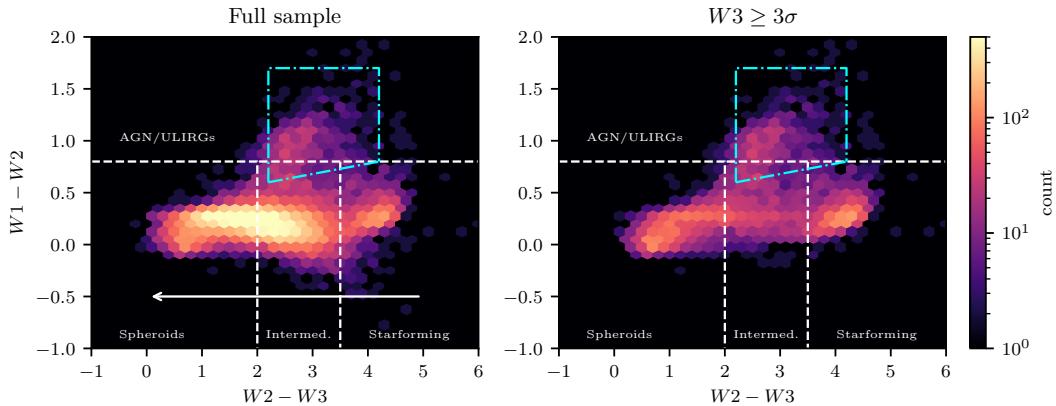


Figure 5.2: *WISE* colour-colour distributions. The dashed grey lines show simple host galaxy class divisions from Jarrett et al. (2017). These classes are labelled in the plot. The blue dot-dashed line shows the empirical optical/infrared AGN criteria from Jarrett et al. (2011). The arrow shows the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

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, 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 & Sadler, 2007). 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 (2007), 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 contribution 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 $W3$

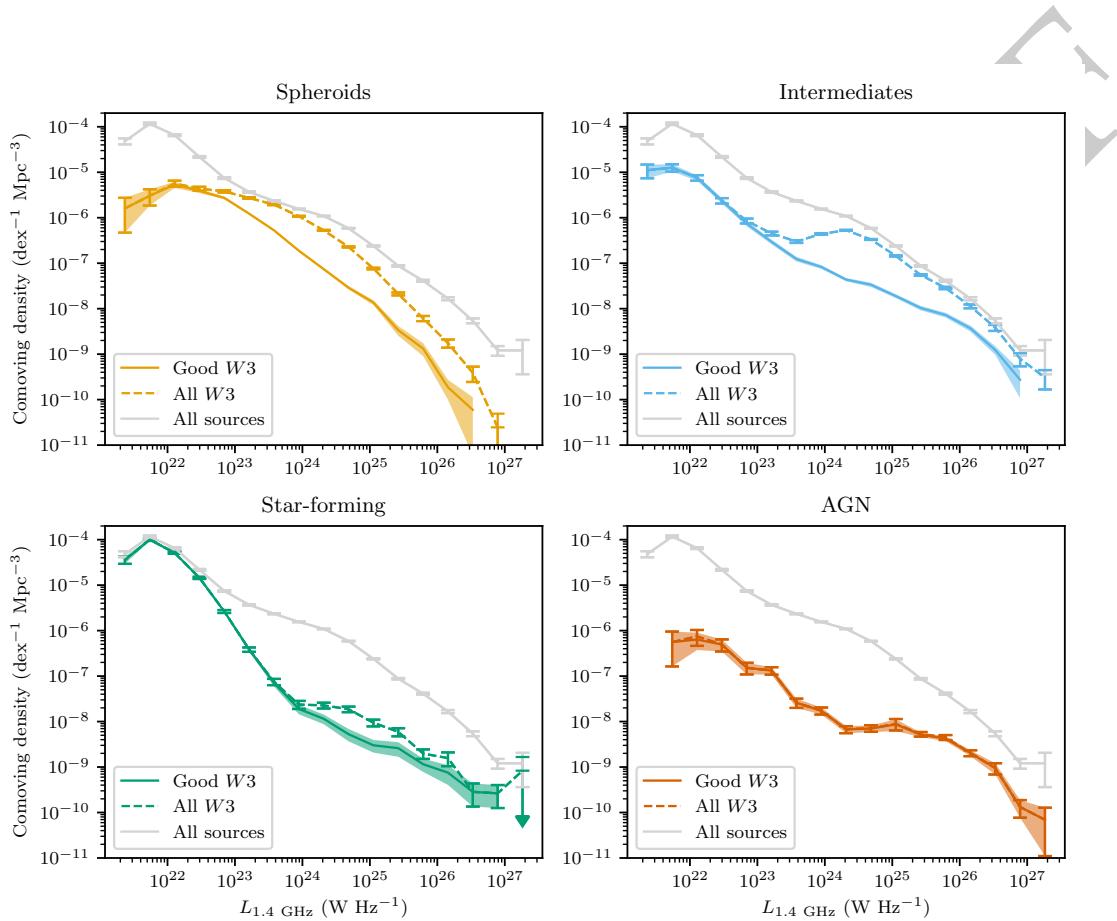


Figure 5.3: RLFs split by host galaxy location in the WISE colour-colour plot (Figure 5.2), using our automated cross-identifications. The grey line is the total RLF for all sources. Solid lines have good W3 detections and dashed lines include W3 with low signal-to-noise.

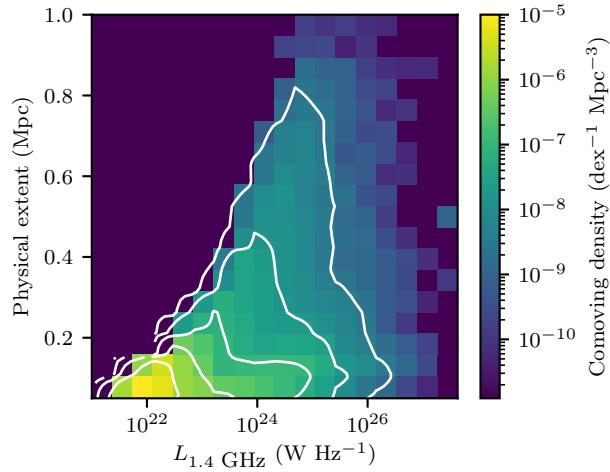


Figure 5.4: Bivariate radio luminosity function showing radio luminosity against projected physical extent. Contours are on a log scale, starting at the median and increasing by 10 per cent per contour.

2621 observations spheroids may dominate above $10^{25} \text{ W Hz}^{-1}$.

2622 Sources with hosts in the mid-infrared AGN region of the colour-colour diagram
 2623 (Figure 5.2) make up the smallest contribution to the radio luminosity function. They
 2624 have a steadily decreasing density from their lowest observed $L_{1.4 \text{ GHz}}$ of $10^{22} \text{ W Hz}^{-1}$
 2625 to their highest of $10^{27} \text{ W Hz}^{-1}$, but are present in all luminosity bins except for the
 2626 very lowest. This is a set with a very low fraction of spectroscopic SDSS matches for
 2627 the *WISE* host galaxies. 26 per cent of hosts outside the *WISE* AGN region have an
 2628 SDSS match, compared to just 12 per cent of hosts inside the *WISE* AGN region. This
 2629 is likely due to the incomplete sampling of QSOs in the SDSS spectroscopic survey or
 2630 redshift evolution effects (Strauss et al., 2002). The fraction of the RLF contributed
 2631 by galaxies classed as mid-infrared AGN increases above $10^{25} \text{ W Hz}^{-1}$, meaning that
 2632 high-luminosity radio AGN are also more likely to be infrared AGN than at lower radio
 2633 luminosities. Note that the AGN set is unaffected by missing $W3$ detections, as the
 2634 AGN set is based only on $W1 - W2$.

2635 Galaxies residing in the intermediate mid-infrared colour region can be populated
 2636 by both early- and late-type galaxies, which have a mix of processes occurring within
 2637 them. These ‘intermediate sources’ dominate in most luminosity ranges, and above
 2638 $10^{24} \text{ W Hz}^{-1}$ they comprise the vast majority of our sample. As intermediate-type
 2639 galaxies fall between star-forming galaxies and passive ellipticals on the mid-infrared
 2640 colour-colour plane, they do not have a clear morphological class and are composed
 2641 of overlapping subsets of sources. The most luminous radio AGN are almost entirely
 2642 found in this set of galaxies. In fact, as radio luminosity increases the density fraction
 2643 shifts from spheroids toward intermediate galaxies, likely due to missing $W3$ moving
 2644 objects from the spheroid set into the intermediate set.

2645 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 (Garon et al., 2019; Rodman et al., 2019).

5 Discussion

5.1 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 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 et al. (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 & Riley, 1974) galaxies tend to be toward the low end of the radio-loud luminosity distribution, about $10^{23} \text{ W Hz}^{-1}$ (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 & Riley, 1974) 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 et al., 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 undersample 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 et al., 2017).

5.2 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 (2007), except at radio luminosities below $10^{22} \text{ W Hz}^{-1}$. 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 & Shabala, 2016; Hardcastle & 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 con-

tribution 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 et al., 1997), the surrounding IGM/CGM may interact with the radio lobe expansion in non-trivial ways (e.g. Hardcastle & Krause, 2013) and the relationship between the mechanical energy and radio luminosity has high scatter on individual radio sources (Hardcastle & 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 \(2016\)](#) [Shabala and Godfrey \(2013\)](#) show that the scalings presented by Willott et al. (1999) are consistent with independent theoretical modelling for high-power radio galaxies. [Godfrey and Shabala \(2016\)](#) on the other hand 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. The relationship between luminosity and kinetic energy is not yet settled, but we can still use this popular scaling method both to explore the consequences of our RLFs and for comparison with previous work.

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 $7 \times 10^{31} \text{ W Mpc}^{-3}$).

2740 **5.3 Future work**

2741 With such a large sample size, further partitioning of the RLF into subsamples is pos-
2742 sible. Any combination of the features investigated here, plus further host galaxy and
2743 radio properties, could be used to generate fractional RLFs. Automated classifiers such
2744 as ClaRAN (Wu et al., 2019) or feature extractors such as PINK (Galvin et al., 2019; Pol-
2745 sterer et al., 2015; Ralph et al., 2019) could provide a way to divide the RLF by radio
2746 morphology. These methods provide a way of dividing galaxy classes based directly
2747 on the radio image, rather than the host galaxy like we have done here, and so should
2748 not be affected by extinction or redshift in the same way as our sample. Such sub-
2749 samples would lend insight into how radio power is connected to radio morphology
2750 and generation mechanisms. Cross-matching with other surveys such as NVSS or the
2751 150 MHz TIFR GMRT Sky Survey would provide properties such as the spectral index
2752 and observations of diffuse emission missed by FIRST (as used by Kimball & Ivezić,
2753 2008). Such properties could also be used to create interesting and insightful fractional
2754 RLFs.

2755 While we have not investigated the link between extended sources and their local
2756 environments, this will be the focus of future work. Environment will heavily factor
2757 into the source sizes, morphologies, and so on, following work such as Rodman et al.
2758 (2019) and Garon et al. (2019).

2759 Ongoing radio surveys such as EMU, VLASS (Lacy et al., 2020), and LoTSS
2760 (Shimwell et al., 2019) will greatly increase the number of extended sources. How-
2761 ever, our sample size limitations in this chapter are not from FIRST, but from SDSS:
2762 until next-generation spectroscopic surveys are available, redshifts will be the limiting
2763 factor. To significantly increase our sample size would require much greater numbers
2764 of redshifts.

2765 **6 Summary**

2766 Extended radio sources provide an opportunity to study the interaction between AGN
2767 and their large-scale environments. We trained the binary cross-identification method
2768 on the Radio Galaxy Zoo to generate the largest sample of reliably cross-identified,
2769 extended radio sources, and this large sample allows us to investigate their bulk dis-
2770 tributions in new, detailed ways. We estimated radio luminosity functions split by
2771 mid-infrared colour, physical extent and redshift. Despite our extendedness criterion,
2772 we found a significant star-forming population. We estimated that extended AGN
2773 contribute between 1.3×10^{30} and 1.2×10^{32} W Mpc $^{-3}$ of mechanical energy to their
2774 environment. Ongoing and future surveys such as EMU will provide even greater
2775 numbers of extended radio sources, and our combination of machine learning and as-
2776 tronomy methodology will allow these samples to be cross-identified and investigated
2777 efficiently and reliably.

2778 7 Acknowledgements

2779 This publication has been made possible by the participation of more than 11 000 volunteers
 2780 in the Radio Galaxy Zoo project. Their contributions are individually acknowledged at <http://rgzauthors.galaxyzoo.org>. Radio Galaxy Zoo makes use of data products
 2781 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
 2782 the National Aeronautics and Space Administration. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative
 2783 agreement by Associated Universities, Inc. The figures in this work made use of Astropy, a community-developed core Python package for Astronomy (The Astropy
 2784 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.
 2785 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
 2786 J. Nabaglo and A. Gruen for their help with efficient data preprocessing. We thank
 2787 H. Zovaro, T. J. Galvin, and H. Tang for their comments on the draft of this chapter.
 2788

2795 F Sankey diagrams

2796 This section presents Sankey diagrams showing the filtering of components and sources
 2797 from the full FIRST sample in this chapter. A Sankey diagram shows the order and
 2798 number of objects removed from a sample. Figure 5.5 shows the filtering of components
 2799 and Figure 5.6 shows the filtering of sources. The component filters are ‘Bad FIRST’ for
 2800 components on the edge of FIRST with incomplete images, ‘Sidelobe’ for components
 2801 with high sidelobe probability, ‘Low score’ for components with only low-scoring candidate
 2802 hosts, ‘Faint’ for components with less than 10 signal-to-noise according to the FIRST
 2803 catalogue, and ‘Compact’ for components that do not have extended radio emission
 2804 according to Equation 5.1. Sources were removed after each component filter if they
 2805 no longer contained any components.

2806 G Visual verification results

2807 In Section 3.1 we described our visual verification of the BXID method. We list the radio
 2808 components in the verification set in Table 5.2. Each row of the table contains the FIRST
 2809 component, its AllWISE host galaxy according to BXID, and whether the association
 2810 is correct according to our visual verification. If an author was particularly unsure
 2811 about an object, they were able to skip this object, and so are not accounted for in the
 2812 verification for that object. Verification was weighted by the Dawid and Skene (1979)
 2813 maximum likelihood model.

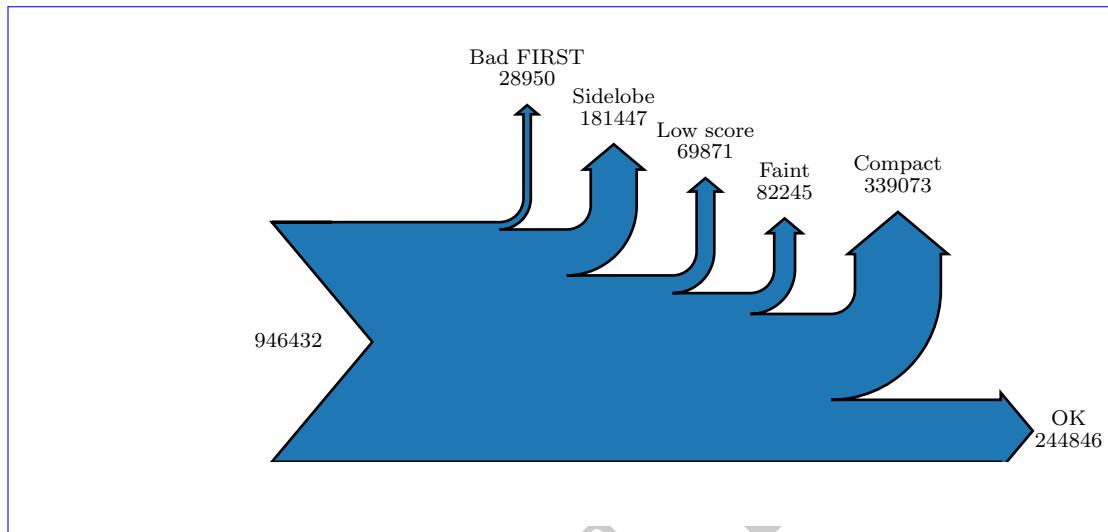


Figure 5.5: Number of components removed from FIRST by each filter.

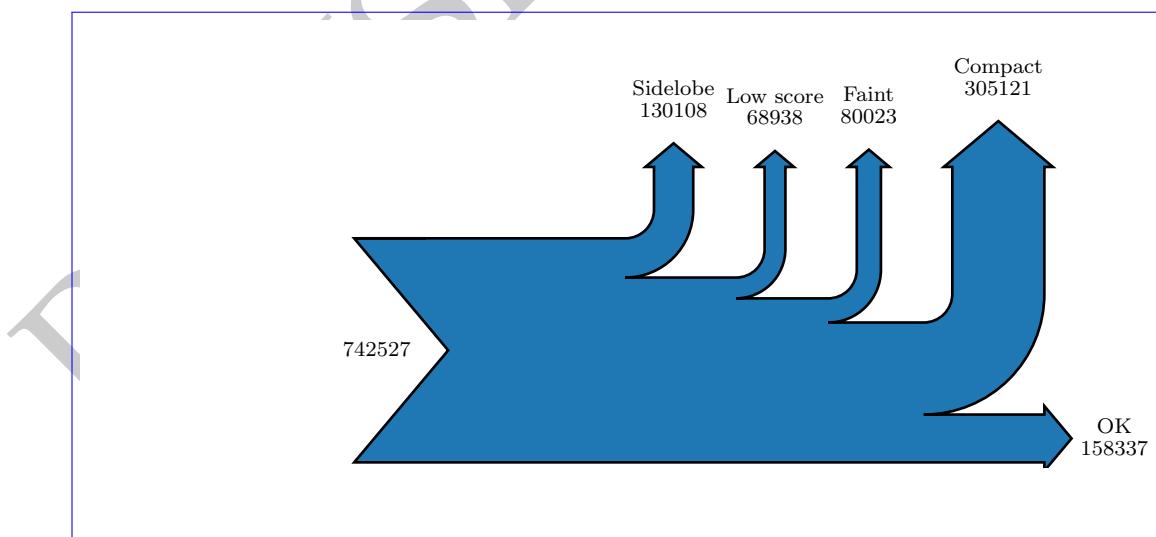


Figure 5.6: Number of sources removed by each filter.

Table 5.2: Validation objects. ‘Agree’ is whether or not the authors of Alger et al. (in prep.) agreed with BXID associating the given FIRST object with the given AllWISE object.

FIRST	AllWISE	Agree	FIRST	AllWISE	Agree
J000234.9-001421	J000242.35-001320.5	n	J094009.5+600403	J094011.55+600357.6	n
J002841.1+141654	J002840.37+141652.7	y	J094023.7+135123	J094023.73+135125.2	y
J003731.4+000156	J003731.26+000146.7	y	J094324.5+435341	J094324.61+435342.0	y
J005407.5-011158	J005407.61-011158.9	y	J094650.8+382015	J094650.44+382010.9	y
J011210.3+002203	J011210.41+002201.9	y	J095011.8+455319	J095011.82+455320.0	y
J012342.4+015849	J012342.24+015850.4	x	J095113.5+180211	J095113.82+180204.2	n
J013015.1+110653	J013015.16+110653.4	y	J095242.4+222638	J095242.45+222638.0	y
J013107.7+070343	J013102.02+070332.0	y	J095538.7+013546	J095539.20+013546.1	y
J014247.9-000039	J014247.81-000040.3	y	J095609.9+363441	J095609.30+363445.4	y
J014250.0-000032	J014247.81-000040.3	n	J095811.8+225056	J095811.90+225055.5	y
J020222.3+030138	J020223.20+030150.4	y	J100019.2+263516	J100018.84+263527.5	y
J020333.8+000853	J020336.94+000759.3	y	J101315.9+064520	J101316.51+064519.0	y
J021840.1-032311	J021840.13-032306.0	y	J101455.2-004716	J101455.30-004718.3	y
J023022.0+010834	J023022.11+010840.0	y	J102153.5+260429	J102153.52+260429.6	y
J024245.3-022535	J024245.35-022534.6	y	J102354.7+390653	J102354.88+390654.0	y
J025901.0+005350	J025901.50+005346.1	y	J102620.4+303600	J102620.46+303550.4	y
J033204.1-004757	J033204.15-004757.1	y	J102710.4+460254	J102714.81+460256.4	n
J073033.2+390413	J073033.21+390412.9	y	J102955.9+424906	J102955.96+424906.7	y
J073954.1+481810	J073954.87+481759.5	y	J103503.9+102404	J103503.92+102403.6	y
J074504.9+331247	J074504.81+331256.2	y	J103839.9+331200	J103839.94+331201.1	y
J074640.4+421709	J074640.45+421709.1	x	J104030.5+211624	J104031.09+211620.6	n
J074707.9+171719	J074708.35+171726.5	y	J104533.8+430025	J104535.22+430020.8	y
J075043.6+274838	J075043.35+274844.8	n	J104907.5+322903	J104907.91+322906.6	y
J075050.3+331937	J075051.25+331905.0	x	J105146.9+552257	J105147.40+552308.4	y
J075422.2+311253	J075422.35+311252.5	x	J105257.5+105418	J105257.53+105421.5	y
J075637.0+212006	J075636.65+212001.4	y	J105521.6+372641	J105521.24+372652.4	y
J082326.1+141438	J082326.34+141435.9	y	J105758.8+321605	J105758.84+321605.3	y
J082422.5+351121	J082422.65+351114.6	y	J110104.9+151618	J110104.90+151618.2	y
J082925.9+462618	J082926.02+462618.5	y	J110353.2+352320	J110353.37+352319.9	y
J083512.4+175441	J083512.45+175441.1	x	J110414.4+481345	J110423.08+481311.0	n
J084133.5+402035	J084133.40+402042.8	x	J111057.7+220756	J111057.18+220758.3	y
J084238.4+405305	J084238.38+405306.6	n	J111208.5+275207	J111201.79+275053.8	n
J084417.3+315845	J084417.92+315845.9	y	J111225.2+233159	J111225.30+233157.9	y
J084728.5+360700	J084728.24+360714.6	y	J111726.3+375336	J111726.35+375337.0	y
J084905.5+111448	J084905.51+111447.8	y	J111746.1+261151	J111746.18+261150.9	y
J085236.8+262006	J085236.11+262013.4	y	J111854.3+424708	J111854.45+424652.8	y
J085415.6+524930	J085415.62+524936.7	x	J112124.4+640417	J112125.02+640408.6	y
J090623.2+300746	J090622.87+300743.9	x	J112135.3+352330	J112135.44+352324.9	y
J091745.1+275049	J091745.89+275103.8	y	J112550.9+200631	J112558.75+200554.3	y
J091752.0+431614	J091752.14+431612.7	y	J112859.7+260923	J112859.86+260911.3	y
J092014.4+302907	J092013.95+302859.3	y	J113201.1+442639	J113201.23+442639.4	y
J092140.5+540118	J092140.24+540121.1	y	J113302.5+355408	J113301.80+355415.3	y
J092213.0+542157	J092213.03+542157.2	x	J113712.7+263301	J113711.86+263335.1	y
J092406.9+562703	J092406.47+562656.2	x	J113756.3+471314	J113756.31+471314.1	y
J092713.1+105841	J092713.14+105839.8	y	J113906.6+230602	J113906.68+230602.1	y
J093108.6+613447	J093108.63+613447.2	y	J114325.0+600721	J114323.90+600737.1	y
J093239.6+052308	J093237.71+052240.7	n	J114759.7+370305	J114759.22+370311.2	y
J093627.8+103610	J093627.87+103609.7	y	J114916.7+083022	J114916.33+083040.5	n
J093645.2+561435	J093645.89+561434.2	y	J115010.9+063340	J115010.93+063340.5	y
J094006.8+482651	J094006.92+482649.2	y	J115308.6+374851	J115316.96+374850.0	y

FIRST	AllWISE	Agree	FIRST	AllWISE	Agree
J115448.7+472222	J115448.67+472223.7	y	J142829.5+070836	J142829.60+070836.3	y
J115603.7+584704	J115603.48+584706.1	y	J143411.0+170036	J143411.18+170035.7	y
J115605.9+343230	J115605.64+343229.4	y	J143624.0-001057	J143623.89-001100.8	y
J115653.0+572338	J115645.38+572151.7	y	J143742.6+104412	J143742.69+104412.8	y
J120138.0+230922	J120137.97+230922.2	y	J143840.8+475355	J143841.08+475356.1	y
J120752.8+533808	J120752.85+533807.3	y	J143909.1+430847	J143909.08+430847.8	y
J120943.3-021934	J120942.89-021943.0	y	J144135.8+102246	J144135.91+102245.1	y
J121045.6+190225	J121045.68+190227.0	y	J144333.6+275229	J144333.02+275250.2	y
J121207.6+115412	J121207.72+115413.8	y	J145012.3+471739	J145012.33+471738.7	y
J121211.3+485951	J121211.86+485952.0	y	J145103.7+452459	J145102.66+452520.5	n
J121406.7+002634	J121406.73+002635.0	y	J145401.6+141009	J145401.70+141009.6	y
J122518.0+350258	J122517.85+350301.9	y	J150158.7+191413	J150158.87+191405.3	y
J122525.1+451530	J122524.71+451508.5	y	J150743.9+352720	J150743.62+352724.1	y
J122640.9+430508	J122640.82+430509.2	y	J151141.6-003209	J151142.01-003213.0	y
J123429.8+260107	J123434.79+260134.3	n	J151315.5+403107	J151315.56+403107.7	y
J123633.1+100928	J123633.12+100928.7	y	J151518.7+230256	J151518.67+230257.3	y
J124839.3+411522	J124839.42+411522.3	n	J151703.6+105947	J151703.68+105947.6	y
J125129.2+551012	J125128.76+551009.3	y	J151736.8+610856	J151736.83+610857.7	y
J130005.8+524801	J130006.14+524803.0	y	J152121.6+281635	J152120.68+281626.2	y
J130132.1+511351	J130132.32+511352.5	y	J152714.8+310425	J152714.88+310424.7	y
J131104.4+464936	J131104.45+464934.0	y	J153428.9+272134	J153429.68+272120.8	y
J131452.2+252811	J131446.81+252820.8	n	J154245.3+100919	J154245.71+100917.8	y
J132033.8+332639	J132033.59+332639.0	n	J154901.6+103159	J154901.40+103152.6	y
J132257.5+191134	J132257.53+191133.9	y	J154925.2+395316	J154926.17+395303.7	y
J132529.3+230734	J132529.35+230733.8	y	J155206.3-005348	J155206.58-005339.3	y
J132546.8+052453	J132546.86+052454.1	y	J155457.3+344637	J155458.45+344644.7	y
J132637.7+112110	J132637.92+112108.8	y	J155743.5+272752	J155743.52+272752.8	y
J132831.8+104339	J132831.88+104338.8	y	J160130.0+083848	J160130.07+083850.7	y
J132932.3+131839	J132932.32+131839.6	y	J160534.8+441220	J160535.55+441221.5	y
J133022.8+311904	J133022.83+311902.8	y	J160859.2+400135	J160901.32+400230.7	n
J133453.3+405653	J133454.13+405650.6	y	J161545.4+231617	J161545.14+231617.2	y
J133741.1+124302	J133741.13+124303.1	y	J161930.4+085533	J161930.51+085532.6	y
J133823.6+103337	J133823.67+103341.9	y	J162228.0+264743	J162228.70+264736.7	y
J134651.2+415154	J134651.06+415156.1	y	J162750.4+473624	J162750.55+473623.5	y
J134704.3+110622	J134704.35+110622.7	y	J162904.2+470852	J162904.34+470853.0	y
J134752.7+555046	J134752.71+555048.6	y	J163038.7+214740	J163037.43+214748.9	n
J134831.7+164325	J134831.57+164328.2	y	J163323.6+424051	J163323.61+424051.9	y
J134949.8+385539	J134949.93+385542.8	y	J163327.5+242426	J163327.87+242427.4	y
J135106.5+074534	J135106.50+074534.2	y	J163533.8+454557	J163534.00+454554.3	y
J135107.7+615502	J135107.75+615502.1	y	J164211.2+512029	J164211.27+512029.3	y
J135658.5+134028	J135659.15+134017.0	y	J165549.1+375923	J165549.01+375923.6	y
J135833.9+180021	J135834.03+180020.4	y	J165620.0+363402	J165619.89+363403.9	y
J140630.7+554017	J140629.32+554009.9	y	J165700.5+474820	J165659.58+474809.0	y
J140804.2+503019	J140804.10+503021.1	y	J171406.2+292712	J171404.16+292704.0	n
J141226.7+454125	J141226.54+454125.5	y	J172126.4+374446	J172126.46+374446.6	y
J141245.0+495213	J141243.84+495206.4	y	J222627.7-005010	J222627.77-005010.8	y
J141317.4+325306	J141317.50+325306.8	y	J223636.4-013827	J223636.48-013827.2	y
J141723.8+543639	J141724.33+543629.5	y	J225619.0+143257	J225621.96+143351.4	y
J141938.8+312146	J141940.16+312138.8	y	J232410.1+001315	J232410.15+001314.5	y
J142515.3+175526	J142513.89+175525.7	y	J234727.9-000919	J234727.65-000912.9	y

2814 H Radio luminosity function

2815 We computed the radio luminosity function following the $1/V_{\max}$ method (Schmidt, 1968).
 2816 We performed the following steps:

- 2817 1. Remove all radio sources that do not fit the selection criteria. This applies for
 2818 both radio and infrared properties, so we choose a minimum radio flux density
 2819 f_{\min} and a maximum infrared magnitude $m_{\max, \text{ir}}$, as well as redshift limits z_{lower}
 2820 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 similarly find the maximum redshift that the infrared host galaxy could be observed at within the selection criteria, z_{ir} , by numerically solving Equation 5.6. $d(z_{\text{ir}})$ is the luminosity distance at a redshift z_{ir} , d is the luminosity distance of the host galaxy, and m is the apparent magnitude of the host galaxy, all in the infrared.

$$2821 5 \log_{10} \left(\frac{d(z_{\text{ir}})}{d} \right) + m = m_{\max, \text{ir}} \quad (5.6)$$

2822 The maximum redshift that the source could have been observed within the selection criteria is then $z_{\max} = \min(z_{\text{ir}}, z_{\text{radio}}, z_{\text{upper}})$.

- 2823 3. For each source, compute the comoving volume V_{\max} at redshift z_{\max} .
- 2824 4. The count for each luminosity bin is the sum over $1/V_{\max}$ for each source in the
 2825 bin. We divided these counts by the estimated completeness (Appendix I) to
 2826 account for redshift incompleteness. We account for the fact FIRST does not cover
 2827 the whole sky by multiplying by the total area of the sky divided by the area of
 2828 our selection.

2829 After computing the luminosity function, we estimate the uncertainty in each bin
 2830 using Poisson statistics, \sqrt{N} for a bin count N .

2831 I Redshift completeness estimate

2832 Figure 5.7 shows the estimated completeness of our RLF sample in Chapter 5 as a
 2833 function of W_1 and $W_1 - W_2$. We followed the same method as Pracy et al. (2016)
 2834 for this estimation, averaging completeness over circles centred on each source. Each
 2835 source is associated with a circle of radius equal to the distance to its 50th nearest
 2836 neighbour in the W_1 and $W_1 - W_2$ plane. This appendix was originally part of Alger et al. (in prep.).
 2837

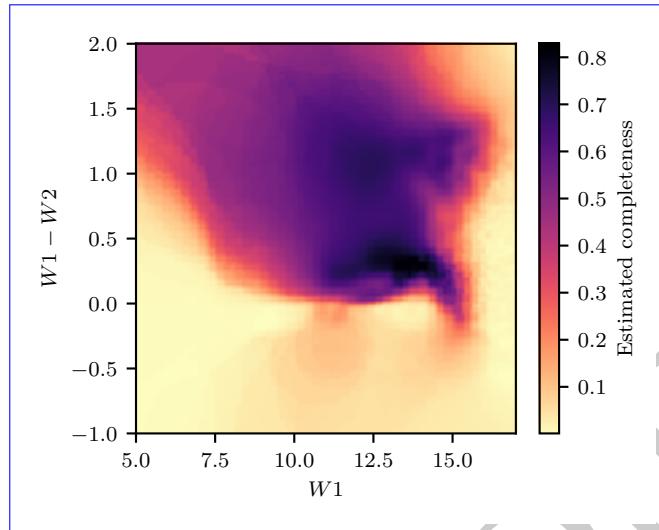


Figure 5.7: Estimated completeness as a function of mid-infrared colour and magnitude.

2838 J Giant radio galaxies

2839 This appendix describes our search for giant radio galaxies in RGZ-Ex, and the results
 2840 of this search. To identify radio sources we assumed that if any two components had
 2841 the same host galaxy then they were part of the same source. This is a reasonable
 2842 assumption if all host galaxies are correctly identified, which was not the case. This
 2843 assumption therefore introduced spurious sources due to galaxies incorrectly identified
 2844 as host galaxies: not all sources used in this chapter are real sources, and in particular
 2845 sources of large angular size are likely to be incorrect. Nevertheless RGZ-Ex provides a
 2846 useful catalogue of *candidate* radio sources, and visual follow-up can confirm whether
 2847 sources of interest are real.

2848 H.A. and M.J.A. examined all 296 candidate sources in the RGZ-Ex catalogue with
 2849 an estimated physical extent larger than 1 Mpc. Of these, 40 were real giant radio
 2850 galaxies, which we show in Table 5.3. We defined ‘giant radio galaxy’ as a radio galaxy
 2851 with emission extended to physical sizes ≥ 1.0 Mpc. Other thresholds, such as 0.7 Mpc,
 2852 also exist in literature. The physical extents of the remaining 256 candidate sources
 2853 were overestimated mostly due to sidelobes/artefacts (103), incorrect source grouping
 2854 (82), or incorrect SDSS matches (21). The citizen scientists who identified giants are:
 2855 WizardHowl, DolorousEdd, antikodon, csunjoto, sisifolibre, JeanTate, JKJD, PADV, and
 2856 firejuggler. H.A., together with his summer students, had previously identified 29 of
 2857 these giants.

2858 Note that this is a particularly challenging set: sources that are misidentified will
 2859 often have unusually large estimated extents due to the inclusion of spurious components.
 2860 The error rate in this set therefore does not reflect the rest of the catalogue.

Table 5.3: 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 ± 0.00001	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

2862

Faraday Complexity

2863 This chapter is based on my paper *Interpretable Faraday Complexity Classification*, by M.
2864 J. Alger, J. D. Livingston, N. M. McClure-Griffiths, J. L. Nabaglo, O. I. Wong, and C.
2865 S. Ong; accepted and to be published by the *Publications of the Astronomical Society of
2866 Australia*.

2867 In the last two chapters we developed and applied a way of automatically cross-
2868 identifying radio sources, which will allow us to quickly extract useful information
2869 from wide-area radio continuum surveys like EMU. These are not the only radio sur-
2870 veys that will be conducted in the lead-up to the SKA, however. Polarisation sky sur-
2871 veys like POSSUM can provide very different information about the radio sky. One
2872 piece of information that could be determined from polarised observations is the Far-
2873 day complexity of a radio source. A complex source may be separated from us as ob-
2874 servers by some magnetised structure, or it might have extended structure across the
2875 sky, even if we cannot resolve the source spatially. This tells us something about the
2876 intervening Universe or the physical structure of the source, respectively. This chapter
2877 develops a new machine learning method of classifying sources as Faraday complex
2878 or simple.

2879 Faraday complexity describes whether a spectropolarimetric observation has sim-
2880 ple or complex magnetic structure. Quickly determining the Faraday complexity of a
2881 spectropolarimetric observation is important for processing large, polarised radio sur-
2882 veys. Finding simple sources lets us build rotation measure grids, and finding complex
2883 sources lets us follow these sources up with slower analysis techniques or further ob-
2884 servations. We introduce five features that can be used to train simple, interpretable
2885 machine learning classifiers for estimating Faraday complexity. We train logistic re-
2886 gression and extreme gradient boosted tree classifiers on simulated polarised spectra
2887 using our features, analyse their behaviour, and demonstrate our features are effective
2888 for both simulated and real data. This is the first application of machine learning meth-
2889 ods to real spectropolarimetry data. With 95 per cent accuracy on simulated ASKAP
2890 data and 90 per cent accuracy on simulated ATCA data, our method performs compa-
2891 rably to state-of-the-art convolutional neural networks while being simpler and easier
2892 to interpret. Logistic regression trained with our features behaves sensibly on real data
2893 and its outputs are useful for sorting polarised sources by apparent Faraday complex-
2894 ity.

2895 1 Introduction

2896 As polarised radiation from distant galaxies makes its way to us, magnetised plasma
 2897 along the way can cause the polarisation angle to change due to the Faraday effect.
 2898 The amount of rotation depends on the squared wavelength of the radiation, and the
 2899 rotation per squared wavelength is called the Faraday depth. Multiple Faraday depths
 2900 may exist along one line-of-sight, and if a polarised source is observed at multiple
 2901 wavelengths then these multiple depths can be disentangled. This can provide insight
 2902 into the polarised structure of the source or the intervening medium.

2903 Faraday rotation measure synthesis (RM synthesis) is a technique for decomposing
 2904 a spectropolarimetric observation into flux at its Faraday depths ϕ , the resulting distri-
 2905 bution of depths being called a *Faraday dispersion function* (FDF) or a *Faraday spectrum*.
 2906 It was introduced by Brentjens and de Bruyn (2005) as a way to rapidly and reliably
 2907 analyse the polarisation structure of complex and high-Faraday depth polarised obser-
 2908 vations.

2909 A *Faraday simple* observation is one for which there is only one Faraday depth, and
 2910 in this simple case the Faraday depth is also known as a *rotation measure* (RM). All Far-
 2911 day simple observations can be modelled as a polarised source with a thermal plasma
 2912 of constant electron density and magnetic field (a ‘Faraday screen’; Anderson et al.,
 2913 2015; Brentjens & de Bruyn, 2005) between the observer and the source. A *Faraday*
 2914 *complex* observation is one which is not Faraday simple, and may differ from a Faraday
 2915 simple source due to plasma emission or composition of multiple screens (Brentjens
 2916 & de Bruyn, 2005). The complexity of a source tells us important details about the
 2917 polarised structure of the source and along the line-of-sight, such as whether the in-
 2918 tervening medium emits polarised radiation, or whether there are turbulent magnetic
 2919 fields or different electron densities in the neighbourhood. The complexity of nearby
 2920 sources taken together can tell us about the magneto-ionic structure of the galactic and
 2921 intergalactic medium between the sources and us as observers. O’Sullivan et al. (2017)
 2922 show examples of simple and complex sources, and Figure 6.1 and Figure 6.2 show an
 2923 example of a simulated simple and complex FDF respectively.

2924 Identifying when an observation is Faraday complex is an important problem in po-
 2925 larised surveys (Sun et al., 2015), and with current surveys such as the Polarised Sky
 2926 Survey of the Universe’s Magnetism (POSSUM) larger than ever before, methods that
 2927 can quickly characterise Faraday complexity en masse are increasingly useful. Being
 2928 able to identify which sources are simple lets us produce a reliable rotation measure
 2929 grid from background sources, and being able to identify which sources might be com-
 2930 plex allows us to find sources to follow-up with slower polarisation analysis methods
 2931 that may require manual oversight, such as QU fitting (as seen in e.g. Miyashita et al.,
 2932 2019; O’Sullivan et al., 2017). In this [paper](#)[chapter](#), we introduce five simple, inter-
 2933 pretable features representing polarised spectra, use these features to train machine
 2934 learning classifiers to identify Faraday complexity, and demonstrate their effective-
 2935 ness on real and simulated data. We construct our features by comparing observed
 2936 polarised sources to idealised polarised sources. The features are intuitive and can be
 2937 estimated from real FDFs.

2938 Section 2 provides a background to our work, including a summary of prior work
2939 and our assumptions on FDFs. Section 3 describes our approach to the Faraday com-
2940 plexity problem. Section 4 explains how we trained and evaluated our method. Finally,
2941 Section 5 discusses these results.

2942 2 Faraday Complexity

2943 Faraday complexity is an observational property of a source: if multiple Faraday depths
2944 are observed within the same apparent source (e.g. due to multiple lines-of-sight being
2945 combined within a beam), then the source is complex. A source composed of multiple
2946 Faraday screens may produce observations consistent with many models (Sun et al.,
2947 2015), including simple sources, so there is some overlap between simple and complex
2948 sources. Faraday thickness is also a source of Faraday complexity: when the interven-
2949 ing medium between a polarised source and the observer also emits polarised light, the
2950 FDF cannot be characterised by a simple Faraday screen. As discussed in Section 2.2
2951 we defer Faraday thick sources to future work. In this section we summarise existing
2952 methods of Faraday complexity estimation and explain our assumptions and model of
2953 simple and complex polarised FDFs.

2954 2.1 Prior work

2955 There are multiple ways to estimate Faraday complexity, including detecting non-
2956 linearity in $\chi(\lambda^2)$ (Goldstein & Reed, 1984), change in fractional polarisation as a func-
2957 tion of frequency (Farnes et al., 2014), non-sinusoidal variation in fractional polari-
2958 sation in Stokes Q and U (O’Sullivan et al., 2012), counting components in the FDF
2959 (Law et al., 2011), minimising the Bayesian information criterion (BIC) over a range of
2960 simple and complex models (called QU fitting; O’Sullivan et al., 2017), the method of
2961 Faraday moments (Anderson et al., 2015; Brown, 2011), and deep convolutional neural
2962 network classifiers (CNNs; Brown et al., 2018). See Sun et al. (2015) for a comparison
2963 of these methods.

2964 The most common approaches to estimating complexity are QU fitting
2965 (e.g. O’Sullivan et al., 2017) and Faraday moments (e.g. Anderson et al., 2015). To our
2966 knowledge there is currently no literature examining the accuracy of QU fitting when
2967 applied to complexity classification specifically, though Miyashita et al. (2019) analyse
2968 its effectiveness on identifying the structure of two-component sources. Brown (2011)
2969 suggested Faraday moments as a method to identify complexity, a method later used
2970 by Farnes et al. (2014) and Anderson et al. (2015), but again no literature examines
2971 the accuracy. CNNs are the current state-of-the-art with an accuracy of 94.9 per cent
2972 (Brown et al., 2018) on simulated ASKAP Band 1 and 3 data, and we will compare our
2973 results to this method.

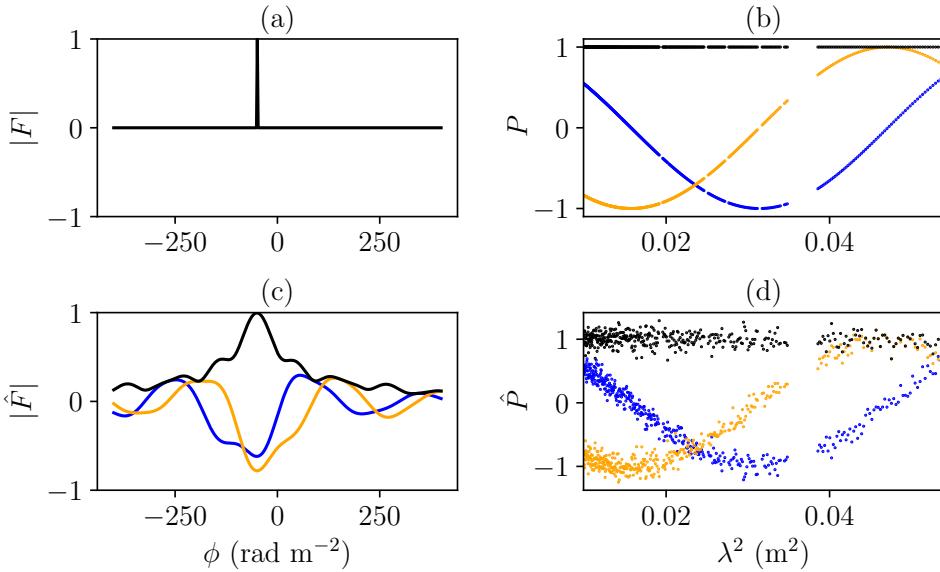


Figure 6.1: A simple FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.

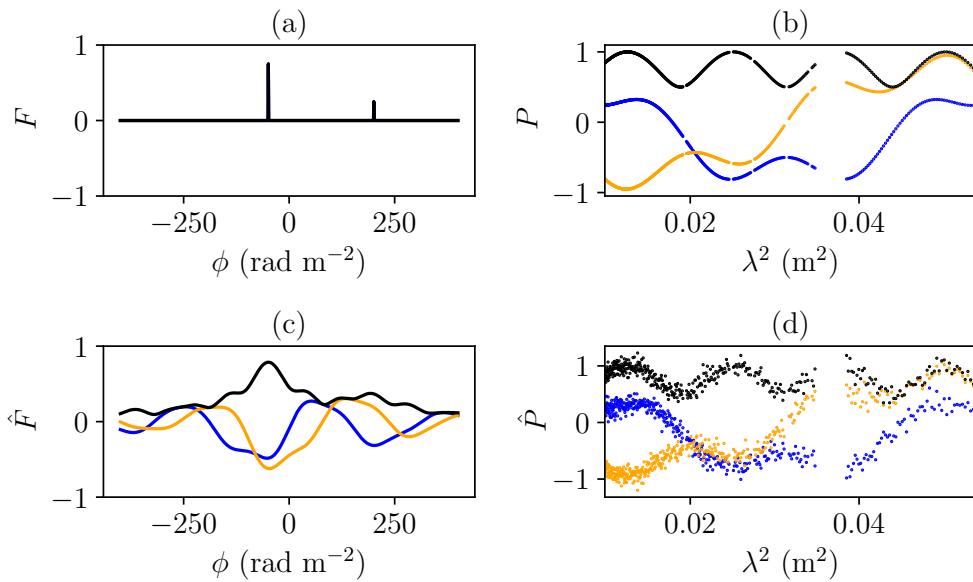


Figure 6.2: A complex FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.

²⁹⁷⁴ **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 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 & de Bruyn, 2005) on \hat{P} with uniform weighting we arrive at an observed FDF:

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

²⁹⁷⁵ Examples of F , \hat{F} , P , and \hat{P} for simple and complex observations are shown in Figure 6.1
²⁹⁷⁶ and Figure 6.2 respectively. Note that there are two reasons that the observed FDF \hat{F}
²⁹⁷⁷ does not match the groundtruth FDF F . The first is the noise in \hat{P} . The second arises
²⁹⁷⁸ from the incomplete sampling of \hat{P} .

²⁹⁷⁹ We do not consider external or internal Faraday dispersion in this work. External
²⁹⁸⁰ Faraday dispersion would broaden the delta functions of Equation 6.1 into peaks, and
²⁹⁸¹ internal Faraday dispersion would broaden them into top-hat functions. All sources
²⁹⁸² have at least a small amount of dispersion as the Faraday depth is a bulk property of
²⁹⁸³ the intervening medium and is subject to noise, but the assumption we make is that
²⁹⁸⁴ this dispersion is sufficiently small that the groundtruth FDFs are well-modelled with

2985 delta functions. Faraday thick sources would also invalidate our assumptions, and we
 2986 assume that there are none in our data as Faraday thickness can be consistent with a
 2987 two-component model depending on the wavelength sampling (e.g. Brentjens & de
 2988 Bruyn, 2005; Ma et al., 2019a). Nevertheless some external Faraday dispersion would
 2989 be covered by our model, as depending on observing parameters Faraday thick sources
 2990 may appear as two screens (Van Eck et al., 2017).

2991 To simulate observed FDFs we follow the method of Brown et al. (2018), which we
 2992 describe in Appendix K.

2993 3 Classification approach

2994 The Faraday complexity classification problem is as follows: Given an FDF \hat{F} , is it Far-
 2995 day complex or Faraday simple? In this section we describe the features that we have
 2996 developed to address this problem, which can be used in any standard machine learn-
 2997 ing classifier. We trained two classifiers on these features, which we describe here also.

2998 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:

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

2999 We propose two distances that have nice properties:

- 3000 • invariant over changes in complex phase,
- 3001 • translationally invariant in Faraday depth,
- 3002 • zero for Faraday simple sources (i.e. when $A_0 = 0, A_1 = 0$, or $\phi_0 = \phi_1$) when
 3003 there is no noise,
- 3004 • symmetric in components (i.e. swapping $A_0 \leftrightarrow A_1$ and $\phi_0 \leftrightarrow \phi_1$ should not
 3005 change the distance),

- 3006 • increasing as A_0 and A_1 become closer to each other, and
 3007 • increasing as screen separation $|\phi_0 - \phi_1|$ increases over a large range.

3008 Our features are constructed from this distance and its minimiser. In other words we
 3009 look for the simple FDF \hat{F}_{simple} that is “closest” to the observed FDF \hat{F} . The minimiser
 3010 ϕ_s is the Faraday depth of the simple FDF.

3011 While we could choose any distance that operates on functions, we used the 2-
 3012 Wasserstein (W_2) distance (Equation 6.7) and the Euclidean distance (Equation 6.9).
 3013 The W_2 distance operates on probability distributions and can be thought of as the min-
 3014 imum cost to ‘move’ one probability distribution to the other, where the cost of moving
 3015 one unit of probability mass is the squared distance it is moved. Under W_2 distance,
 3016 the minimiser ϕ_w in Equation 6.6 can be interpreted as the Faraday depth that the FDF
 3017 \hat{F} would be observed to have if its complexity was unresolved (i.e. the weighted mean
 3018 of its components). The Euclidean distance is the square root of the least-squares loss
 3019 which is often used for fitting \hat{F}_{simple} to the FDF \hat{F} . Under Euclidean distance, the
 3020 minimiser ϕ_s is equivalent to the depth of the best-fitting single component under as-
 3021 sumption of Gaussian noise in \hat{F} . We calculated the W_2 distance using Python Optimal
 3022 Transport (Flamary & Courty, 2017), and we calculated the Euclidean distance using
 3023 `scipy.spatial.distance.euclidean` (Virtanen et al., 2020). Further intuition about
 3024 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.

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

3025 Note that in practice $\phi_a \approx \phi_e$. For complex observed FDFs, the values of Faraday
 3026 depths ϕ_w and ϕ_a tend to differ (essentially by a proportion of the location of the sec-
 3027 ond screen). The difference between ϕ_w and ϕ_a therefore provides useful information
 3028 to identify complex FDFs. When the observed FDF is simple, the 2-Wasserstein fit will
 3029 overlap significantly, hence the observed magnitudes $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$ will be similar.
 3030 However, for complex FDFs ϕ_w and ϕ_a are at different depths, leading to different val-
 3031 ues of $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$. Therefore the magnitudes of the observed FDFs at the depths
 3032 ϕ_w and ϕ_a indicate how different the observed FDF is from a simple FDF.

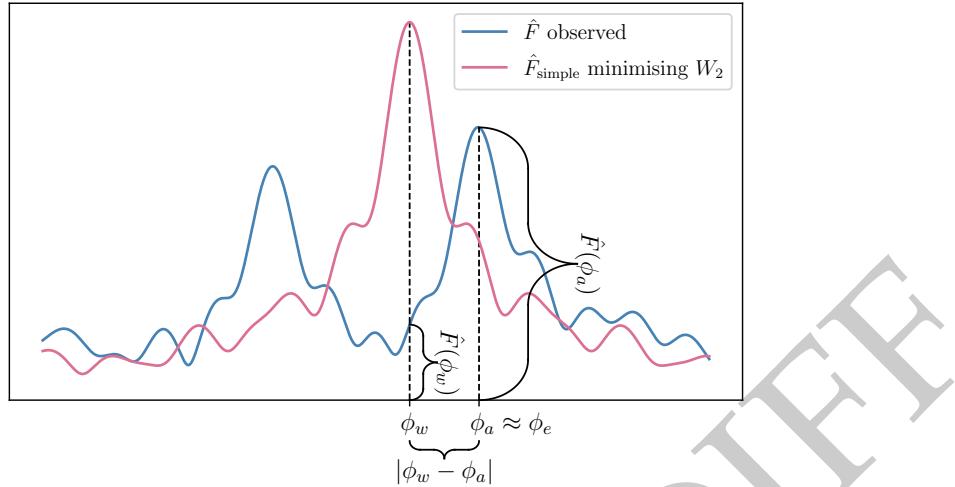


Figure 6.3: An example of how an observed FDF \hat{F} relates to our features. ϕ_w is the W_2 -minimising Faraday depth, and ϕ_a is the \hat{F} -maximising Faraday depth (approximately equal to the Euclidean-minimising Faraday depth). The remaining two features are the W_2 and Euclidean distances between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

3033 In summary, we provide the following features to the classifier:

3034 • $\log |\phi_w - \phi_a|$,

3035 • $\log \hat{F}(\phi_w)$,

3036 • $\log \hat{F}(\phi_a)$,

3037 • $\log D_{W_2}(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_w))$,

3038 • $\log D_E(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_e))$,

3039 where D_E is the Euclidean distance, D_{W_2} is the W_2 distance, ϕ_a is the Faraday depth of
3040 the FDF peak, ϕ_w is the minimiser for W_2 distance, and ϕ_e is the minimiser for Euclidean
3041 distance.

3042 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 Appendix L 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:

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

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

3043 See Appendix M for the corresponding calculation.

3044 3.3 Classifiers

3045 We trained two classifiers on simulated observations using these features: logistic re-
3046 gression (LR) and extreme gradient boosted trees (XGB). These classifiers are useful
3047 together for understanding Faraday complexity classification. LR is a linear classi-
3048 fier that is readily interpretable by examining the weights it applies to each feature,
3049 and is one of the simplest possible classifiers. XGB is a powerful off-the-shelf non-
3050 linear ensemble classifier, and is an example of a decision tree ensemble which are
3051 widely used in astronomy (e.g. Hložek et al., 2020; Machado Poletti Valle et al., 2020).
3052 We used the `scikit-learn` implementation of LR and we use the `XGBoost` library
3053 for XGB. We optimised hyperparameters for XGB using a fork of `xgboost-tuner`¹
3054 as utilised by Zhu et al. (2020). We used 1 000 iterations of randomised parameter
3055 tuning and the hyperparameters we found are tabulated in Table 6.2. We optimised
3056 hyperparameters for LR using a 5-fold cross-validation grid search implemented in
3057 `sklearn.model_selection.GridSearchCV`. The resulting hyperparameters are tabu-
3058 lated in Table 6.3 in the [Appendix](#) Appendix N.

3059 4 Experimental method and results

3060 We applied our classifiers to classify simulated (Section 4.2 and 4.3) and real (Sec-
3061 tion 4.4) FDFs. We replicated the experimental setup of Brown et al. (2018) for com-
3062 parison with the state-of-the-art CNN classification method, and we also applied our
3063 method to 142 real FDFs observed with the Australia Telescope Compact Array (ATCA)
3064 from Livingston et al. (2021) and O’Sullivan et al. (2017).

3065 4.1 Data

3066 4.1.1 Simulated training and validation data

3067 Our classifiers were trained and validated on simulated FDFs. We produced two sets
3068 of simulated FDFs, one for comparison with the state-of-the-art method in the litera-
3069 ture and one for application to our observed FDFs (described in Section 4.1.2). We
3070 refer to the former as the ‘ASKAP’ dataset as it uses frequencies from the Australian

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

3071 Square Kilometre Array Pathfinder 12-antenna early science configuration. These fre-
 3072 quencies included 900 channels from 700–1300 and 1500–1800 MHz and were used to
 3073 generate simulated training and validation data by Brown et al. (2018). We refer to
 3074 the latter as the ‘ATCA’ dataset as it uses frequencies from the 1–3 GHz configuration
 3075 of the ATCA. These frequencies included 394 channels from 1.29–3.02 GHz and match
 3076 our real data. We simulated Faraday depths from -50 to 50 rad m^{-2} for the ‘ASKAP’
 3077 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 :

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

3078 4.1.2 Observational data

3079 We used two real datasets containing a total of 142 sources: 42 polarised spectra from
 3080 Livingston et al. (2020, submitted) and 100 polarised spectra from O’Sullivan et al.
 3081 (2017). These datasets were observed in similar frequency ranges on the same tele-
 3082 scope (with different binning), but are in different parts of the sky. The Livingston
 3083 data were taken near the Galactic Centre, and the O’Sullivan data were taken away
 3084 from the plane of the Galaxy. There are more Faraday complex sources near the Galac-
 3085 tic Centre compared to more Faraday simple sources away from the plane of the Galaxy
 3086 (Livingston et al.). The similar frequency channels used in the two datasets result in
 3087 almost identical RMSFs over the Faraday depth range we considered (-500 to 500 rad
 3088 m^{-2}), so we expected that the classifiers would work equally well on both datasets
 3089 with no need to re-train. We discarded the 26 Livingston sources with modelled Far-
 3090 day depths outside of this Faraday depth range, which we do not expect to affect the
 3091 applicability of our methods to wide-area surveys because these fairly high depths are
 3092 not common.

3093 Livingston et al. (2021) used RM-CLEAN (Heald, 2008) to identify significant com-
 3094 ponents in their FDFs. Some of these components had very high Faraday depths up to
 3095 2000 rad m^{-2} , but we chose to ignore these components in this [paper](#)-[chapter](#) as they
 3096 are much larger than might be expected in a wide-area survey like POSSUM. They used
 3097 the second Faraday moment (Brown, 2011) to estimate Faraday complexity, with Far-
 3098 day depths determined using `scipy.signal.find_peaks` on the cleaned FDFs, with
 3099 a cutoff of 7 times the noise of the polarised spectrum. Using this method, they es-
 3100 timated that 89 per cent of their sources were Faraday complex i.e. had a Faraday
 3101 moment greater than 0.

3102 O’Sullivan et al. (2017) used the QU-fitting and model selection technique described
 3103 in O’Sullivan et al. (2012). The QU-fitting models contained up to three Faraday screen
 3104 components as well as a term for internal and external Faraday dispersion. We ignore
 3105 the Faraday thickness and dispersion for the purposes of this paperchapter, as most
 3106 sources were not found to have Faraday thickness and dispersion is beyond the scope
 3107 of our current work. 37 sources had just one component, 52 had two, and the remaining
 3108 11 had three.

3109 4.2 Results on ‘ASKAP’ dataset

Table 6.1: Confusion matrix entries for LR and XGB on ‘ASKAP’ and ‘ATCA’ simulated datasets, and the CNN confusion matrix entries adapted from Brown et al. (2018).

	‘ASKAP’			‘ATCA’	
	LR	XGB	CNN	LR	XGB
True negative rate	0.99	0.99	0.97	0.92	0.91
False positive rate	0.01	0.01	0.03	0.08	0.09
False negative rate	0.10	0.09	0.07	0.16	0.10
True positive rate	0.90	0.91	0.93	0.84	0.90

3110 The accuracy of the LR and XGB classifiers on the ‘ASKAP’ testing set was 94.4 and
 3111 95.1 per cent respectively. The rates of true and false identifications are summarised
 3112 in Table 6.1. These results are very close to the CNN presented by Brown et al. (2018),
 3113 with a slightly higher true negative rate and a slightly lower true positive rate (recall-
 3114 ing that positive sources are complex, and negative sources are simple). The accuracy
 3115 of the CNN was 94.9, slightly lower than our XGB classifier and slightly higher than
 3116 our LR classifier. Both of our classifiers therefore produce similar classification perfor-
 3117 mance to the CNN, with faster training time and easier interpretation.

3118 4.3 Results on ‘ATCA’ dataset

3119 The accuracy of the LR and XGB classifiers on the ‘ATCA’ dataset was 89.2 and 90.5
 3120 per cent respectively. The major differences between the ‘ATCA’ and the ‘ASKAP’ ex-
 3121 periments are the range of the simulated Faraday depths and the distribution of noise
 3122 levels. The ‘ASKAP’ dataset, to match past CNN work, only included depths from
 3123 -50 to 50 rad m^{-2} , while the ‘ATCA’ dataset includes depths from -500 to 500 rad
 3124 m^{-2} . The rates of true and false identifications are again shown in Table 6.1.

3125 As we know the true Faraday depths of the components in our simulation, we can
 3126 investigate the behaviour of these classifiers as a function of physical properties. Fig-
 3127 ure 6.4 shows the mean classifier prediction as a function of component depth separa-
 3128 tion and minimum component amplitude. This is tightly related to the mean accuracy,
 3129 as the entire plot domain contains complex spectra besides the left and bottom edge:
 3130 by thresholding the classifier prediction to a certain value, the accuracy will be one
 3131 hundred per cent on the non-edge for all sources with higher prediction values.

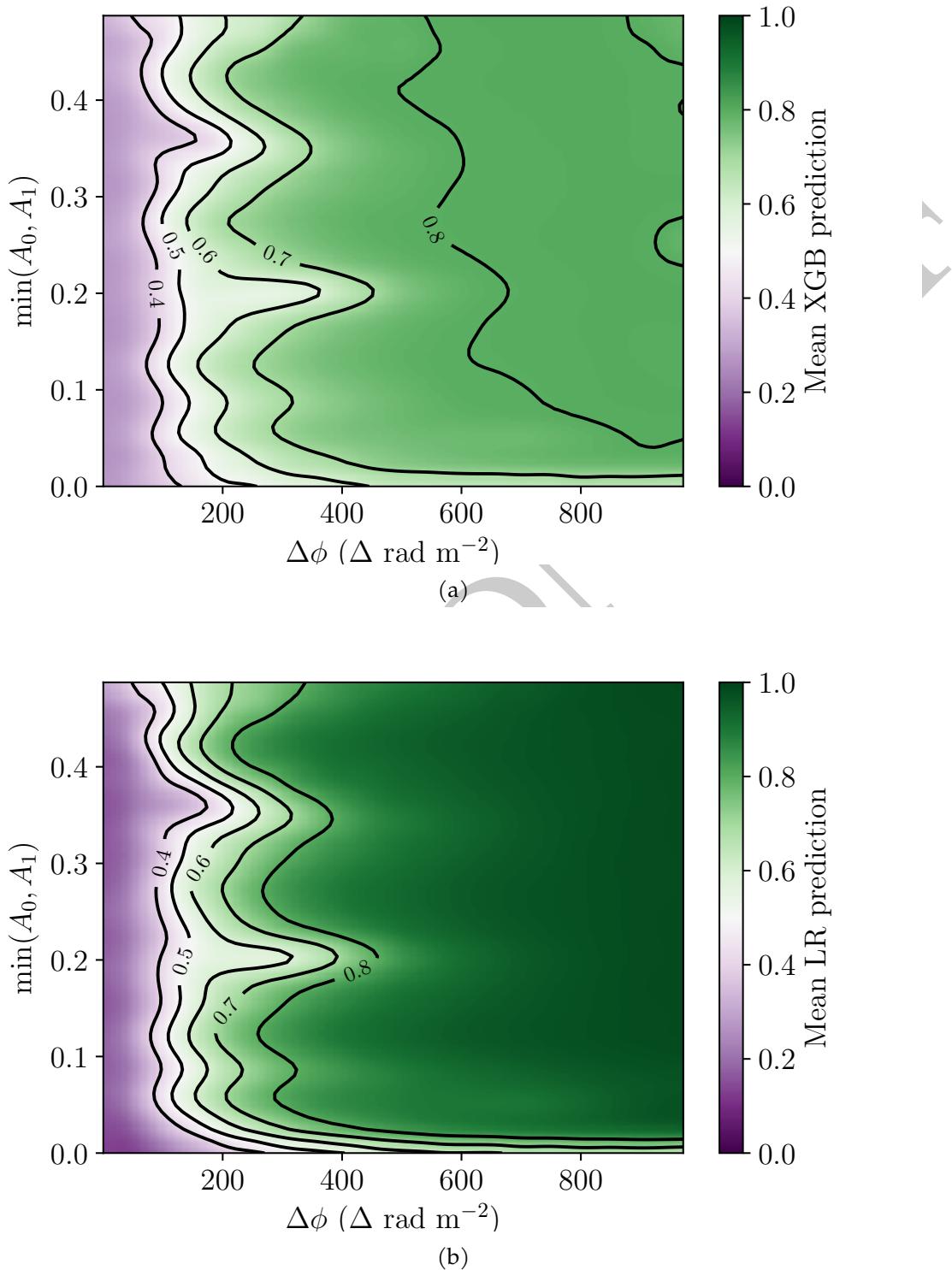


Figure 6.4: Mean prediction as a function of component depth separation and minimum component amplitude for (a) XGB and (b) LR.

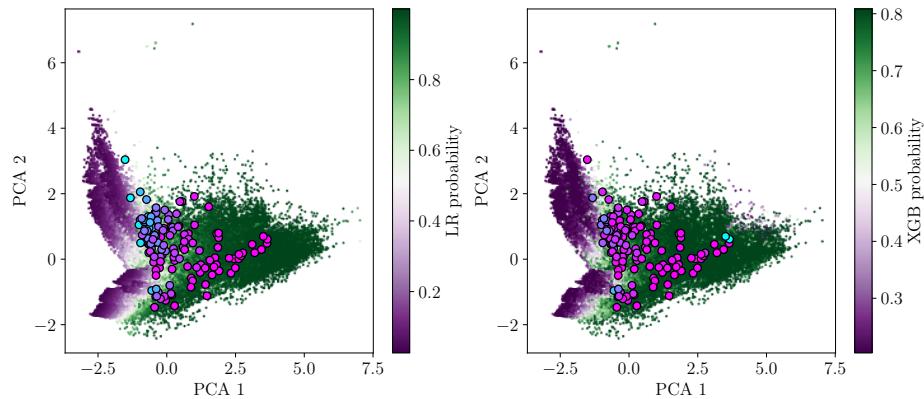


Figure 6.5: Principal component analysis for simulated data (coloured dots) with observations overlaid (black-edged circles). Observations are coloured by their XGB or LR estimated probability of being complex, with blue indicating ‘most simple’ and pink indicating ‘most complex’.

3132 4.4 Results on observed FDFs

3133 We used the LR and XGB classifiers which were trained on the ‘ATCA’ dataset to esti-
 3134 mate the probability that our 142 observed FDFs (Section 4.1.2) were Faraday complex.
 3135 As these classifiers were trained on simulated data, they face the issue of the ‘domain
 3136 gap’: the distribution of samples from a simulation differs from the distribution of real
 3137 sources, and this affects performance on real data. Solving this issue is called ‘domain
 3138 adaptation’ and how to do this is an open research question in machine learning (Pan
 3139 & Yang, 2010; Zhang, 2020). Nevertheless, the features of our observations mostly fall
 3140 in the same region of feature space as the simulations (Figure 6.5) and so we expect
 3141 reasonably good domain transfer.

3142 Two apparently complex sources in the Livingston sample are classified as simple
 3143 with high probability by XGB. These outliers are on the very edge of the training sam-
 3144 ple (Figure 6.5) and the underdensity of training data here is likely the cause of this
 3145 issue. LR does not suffer the same issue, producing plausible predictions for the entire
 3146 dataset, and these sources are instead classified as complex with high probability.

3147 With a threshold of 0.5, LR predicted that 96 and 83 per cent of the Livingston and
 3148 O’Sullivan sources were complex respectively. This is in line with expectations that the
 3149 Livingston data should have more Faraday complex sources than the O’Sullivan data
 3150 due to their location near the Galactic Centre. XGB predicted that 93 and 100 per cent
 3151 of the Livingston and O’Sullivan sources were complex respectively. Livingston et al.
 3152 (2021) found that 90 per cent of their sources were complex, and O’Sullivan et al. (2017)
 3153 found that 64 per cent of their sources were complex. This suggests that our classifiers
 3154 are overestimating complexity, though it could also be the case that the methods used
 3155 by Livingston and O’Sullivan underestimate complexity. Modifying the prediction
 3156 threshold from 0.5 changes the estimated rate of Faraday complexity, and we show

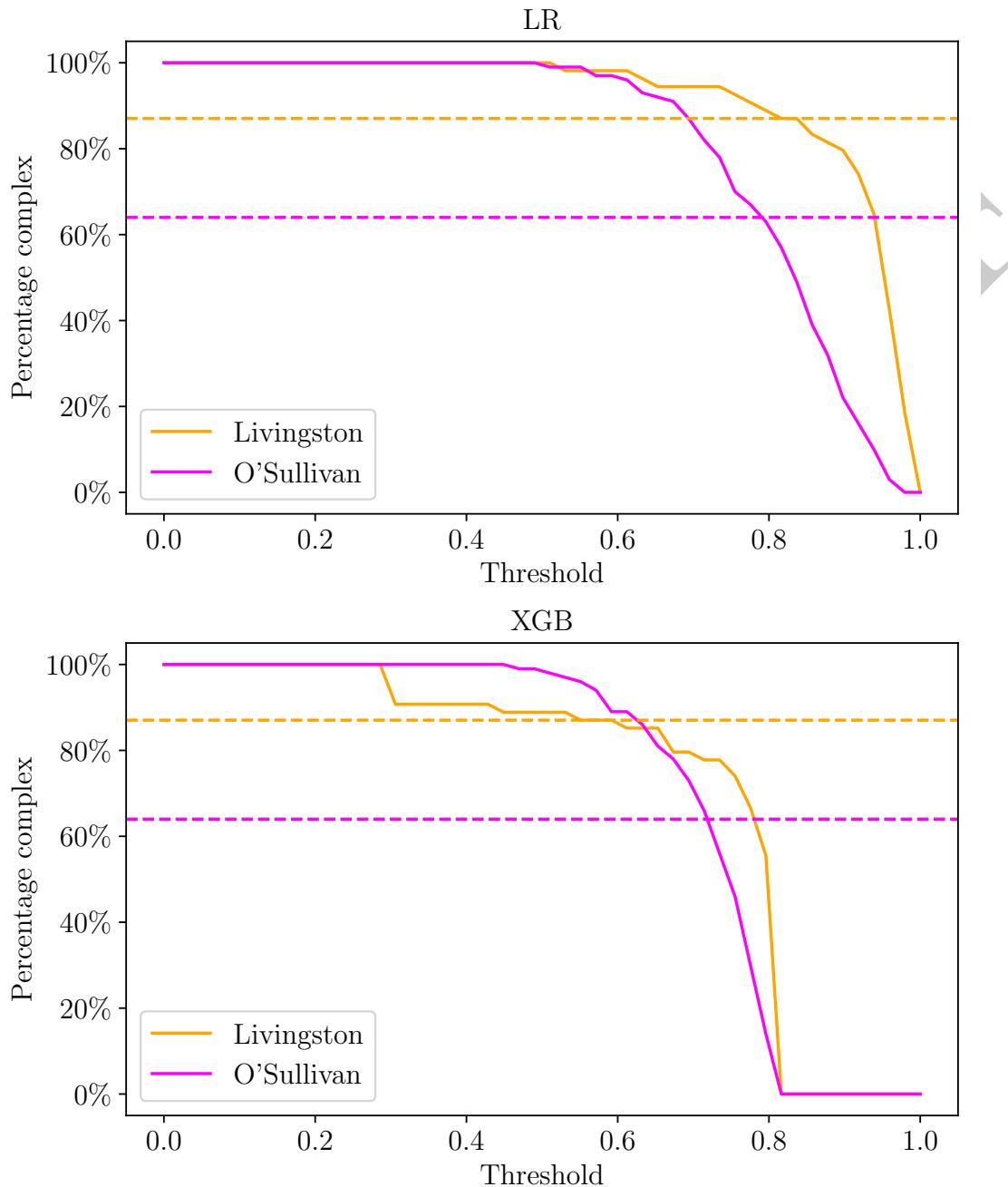


Figure 6.6: Estimated rates of Faraday complexity for the Livingston and O'Sullivan datasets as functions of threshold. The horizontal lines indicate the rates of Faraday complexity estimated by Livingston and O'Sullivan respectively.

3157 the estimated rates against threshold for both classifiers in Figure 6.6. We suggest that
3158 this result is indicative of our probabilities being uncalibrated, and a higher threshold
3159 should be chosen in practice. We chose to keep the threshold at 0.5 as this had the
3160 highest accuracy on the simulated validation data. The very high complexity rates of
3161 XGB and two outlying classifications indicate that the XGB classifier may be overfitting
3162 to the simulation and that it is unable to generalise across the domain gap.

3163 Figure 6.7 and Figure 6.8 in Appendix O show every observed FDF ordered by esti-
3164 mated Faraday complexity, alongside the models predicted by Livingston and O’Sullivan
3165 et al. (2017), for LR and XGB respectively. There is a clear visual trend of increasingly
3166 complex sources with increasing predicted probability of being complex.

3167 5 Discussion

3168 On simulated data (Section 4.3) we achieve state-of-the-art accuracy. Our results on
3169 observed FDFs show that our classifiers produce plausible results, with Figure 6.7 and
3170 Figure 6.8 showing a clear trend of apparent complexity. Some issues remain: we
3171 discuss the intrinsic overlap between simple and complex FDFs in Section 5.1 and the
3172 limitations of our method in Section 5.2.

3173 5.1 Complexity and seeming ‘not simple’

3174 Through this work we found our methods limited by the significant overlap between
3175 complex and simple FDFs. Complex FDFs can be consistent with simple FDFs due to
3176 close Faraday components or very small amplitudes on the secondary component, and
3177 vice versa due to noise.

3178 The main failure mode of our classifiers is misclassifying a complex source as sim-
3179 ple (Table 6.1). Whether sources with close components or small amplitudes should
3180 be considered complex is not clear, since for practical purposes they can be treated as
3181 simple: assuming the source is simple yields a very similar RM to the RM of the pri-
3182 mary component, and thus would not negatively impact further data products such as
3183 an RM grid. The scenarios where we would want a Faraday complexity classifier rather
3184 than a polarisation structure model – large-scale analysis and wide-area surveys – do
3185 not seem to be disadvantaged by considering such sources simple. Additional sources
3186 similar to these are likely hidden in presumably ‘simple’ FDFs by the frequency range
3187 and spacing of the observations, just as how these complex sources would be hidden
3188 in lower-resolution observations. Note also that misidentification of complex sources
3189 as simple is intrinsically a problem with complexity estimation even for models not
3190 well-represented by a simple FDF, as complex sources may conspire to appear as a
3191 wide range of viable models including simple (Sun et al., 2015).

3192 Conversely, high-noise simple FDFs may be consistent with complex FDFs. One
3193 key question is how Faraday complexity estimators should behave as the noise in-
3194 creases: should high noise result in a complex prediction or a simple prediction, given
3195 that a complex or simple FDF would both be consistent with a noisy FDF? Occam’s
3196 razor suggests that we should choose the simplest suitable model, and so increasing

3197 noise should lead to predictions of less complexity. This is not how our classifiers op-
 3198 erate, however: high-noise FDFs are different to the model simple FDFs and so are
 3199 predicted to be ‘not simple’. In some sense our classifiers are not looking for complex
 3200 sources, but are rather looking for ‘not simple’ sources.

3201 5.2 Limitations

3202 Our main limitations are our simplifying assumptions on FDFs and the domain gap be-
 3203 tween simulated and real observations. However, our proposed features (Section Sec-
 3204 tion 3.1) can be applied to future improved simulations.

3205 It is unclear what the effect of our simplifying assumptions are on the effective-
 3206 ness of our simulation. The three main simplifications that may negatively affect our
 3207 simulations are 1) limiting to two components, 2) assuming no external Faraday dis-
 3208 persions, and 3) assuming no internal Faraday dispersion (Faraday thickness). Future
 3209 work will explore removing these simplifying assumptions, but will need to account
 3210 for the increased difficulty in characterising the simulation with more components and
 3211 no longer having Faraday screens as components. Additionally, more work will be re-
 3212 quired to make sure that the rates of internal and external Faraday dispersion match
 3213 what might be expected from real sources, or risk making a simulation that has too
 3214 large a range of consistent models for a given source: for example, a two-component
 3215 source could also be explained as a sufficiently wide or resolved-out Faraday thick
 3216 source or a three-component source with a small third component. This greatly com-
 3217 plifies the classification task.

3218 Previous machine learning work (e.g. Brown et al., 2018) has not been run before
 3219 on real FDF data, so this [paper](#) [chapter](#) is the first example of the domain gap arising
 3220 in Faraday complexity classification. This is a problem that requires further research
 3221 to solve. We have no good way to ensure that our simulation matches reality, so some
 3222 amount of domain adaptation will always be necessary to train classifiers on simulated
 3223 data and then apply these classifiers to real data. But with the low source counts in po-
 3224 larisation science (high-resolution spectropolarimetric data currently numbers in the
 3225 few hundreds) any machine learning method will need to be trained on simulations.
 3226 This is not just a problem in Faraday complexity estimation, and domain adaptation
 3227 is also an issue faced in the wider astroinformatics community: large quantities of la-
 3228 belled data are hard to come by, and some sources are very rare (e.g. gravitational
 3229 wave detections or fast radio bursts; Agarwal et al., 2020; Gebhard et al., 2019; Zevin
 3230 et al., 2017). LR seems to handle the domain adaptation better than XGB, with only a
 3231 slightly lower accuracy on simulated data. Our results are plausible and the distribu-
 3232 tion of our simulation well overlaps the distribution of our real data (Figure 6.5).

3233 6 Conclusion

3234 We developed a simple, interpretable machine learning method for estimating Faraday
 3235 complexity. Our interpretable features were derived by comparing observed FDFs to

3236 idealised simple FDFs, which we could determine both for simulated and real obser-
 3237 vations. We demonstrated the effectiveness of our method on both simulated and real
 3238 data. Using simulated data, we found that our classifiers were 95 per cent accurate,
 3239 with near perfect recall (specificity) of Faraday simple sources. On simulated data
 3240 that matched existing observations, our classifiers obtained an accuracy of 90 per cent.
 3241 Evaluating our classifiers on real data gave the plausible results shown in Figure 6.7,
 3242 and marks the first application of machine learning to observed FDFs. Future work
 3243 will need to narrow the domain gap to improve transfer of classifiers trained on simu-
 3244 lations to real, observed data.

3245 **7 Acknowledgements**

3246 This research was conducted in Canberra, on land for which the Ngunnawal and Ngam-
 3247 bri people are the traditional and ongoing custodians. M.J.A. and J.D.L. were sup-
 3248 ported by the Australian Government Research Training Program. M.J.A. was sup-
 3249 ported by the Astronomical Society of Australia. The Australia Telescope Compact
 3250 Array is part of the Australia Telescope National Facility which is funded by the Austra-
 3251 lian Government for operation as a National Facility managed by CSIRO. We acknowl-
 3252 edge the Gomeroi people as the traditional ~~owners~~custodians of the Observatory site.
 3253 We thank the anonymous referee for their comments on this work.

3254 **K Simulating observed FDFs**

This appendix describes how we simulated FDFs in Chapter 6. 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 $\vec{F} \in \mathbb{R}^d$:

$$\vec{F}_j = \sum_{k=0}^1 A_k \delta(-\phi_{\max} + j\delta\phi - \phi_k) \quad (6.12)$$

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:

$$\phi_k \in [\phi_{\min}, \phi_{\min} + \delta\phi, \dots, \phi_{\max}] \quad (6.13)$$

$$A_k \sim \mathcal{U}(0, 1). \quad (6.14)$$

We then generate a vector polarisation spectrum $\vec{P} \in \mathbb{R}^m$ from \vec{F} using a Equation 6.15:

$$\vec{P}_\ell = \sum_{j=0}^j F_j e^{2i(\phi_{\min} + j\delta\phi)\lambda_\ell^2} d\phi. \quad (6.15)$$

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:

$$\hat{F}_j = m^{-1} \sum_{\ell=1}^m \vec{P}_\ell e^{-2i(\phi_{\min} + j\delta_\phi)\lambda_\ell^2}. \quad (6.16)$$

3255 L 2-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. This appendix demonstrates that fact. Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the sum-normalised simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1} \quad (6.17)$$

$$\tilde{S}(\phi; \phi_w) = \delta(\phi - \phi_w). \quad (6.18)$$

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

$$D_{W_2}(A \parallel B)^2 = \inf_{\gamma \in \Gamma(A, B)} \iint_{\phi_{\min}}^{\phi_{\max}} |x - y|^2 d\gamma(x, y) \quad (6.19)$$

$$\tilde{A}(\phi) = \frac{|A(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |A(\theta)| d\theta} \quad (6.20)$$

$$\tilde{B}(\phi) = \frac{|B(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |B(\theta)| d\theta} \quad (6.21)$$

3256 where $\Gamma(A, B)$ is the set of couplings of A and B , i.e. the set of joint probability distributions
 3257 that marginalise to A and B ; and $\inf_{\gamma \in \Gamma(A, B)}$ is the infimum over $\Gamma(A, B)$. This can be
 3258 interpreted as the minimum cost to ‘move’ one probability distribution to the other,
 3259 where the cost of moving one unit of probability mass is the squared distance it is
 3260 moved.

²<https://github.com/CIRADA-Tools/RM>

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

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\phi = \tilde{S}(\varphi; \phi_w), \quad (6.22)$$

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\varphi = \tilde{F}(\phi). \quad (6.23)$$

The coupling that minimises the integral in Equation 6.19 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:

$$\gamma(\phi, \varphi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1} \delta(\varphi - \phi_w). \quad (6.24)$$

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:

$$D_{W_2}(\tilde{F} \parallel \tilde{S})^2 = \iint_{\phi_{\min}}^{\phi_{\max}} |\phi - \varphi|^2 d\gamma(\phi, \varphi) \quad (6.25)$$

$$= \frac{A_0(\phi_0 - \phi_w)^2 + A_1(\phi_1 - \phi_w)^2}{A_0 + A_1}. \quad (6.26)$$

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

$$\phi_w = \frac{A_0 \phi_0 + A_1 \phi_1}{A_0 + A_1}. \quad (6.27)$$

Substituting this back in, we find

$$\zeta_{W_2}(F)^2 = \frac{A_0 A_1}{A_0 + A_1} (\phi_0 - \phi_1)^2 \quad (6.28)$$

³²⁶¹ which is the Faraday moment.

3262 M Euclidean distance in the no-RMSF case

In this appendix we calculate the minimised 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:

$$\tilde{F}(\phi) = \frac{A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1)}{A_0 + A_1} \quad (6.29)$$

$$\tilde{S}(\phi; \phi_e) = \delta(\phi - \phi_e). \quad (6.30)$$

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

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (6.31)$$

$$= \int_{\phi_{\min}}^{\phi_{\max}} |\tilde{F}(\phi) - \delta(\phi - \phi_e)|^2 d\phi. \quad (6.32)$$

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

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (6.33)$$

$$= \frac{1}{(A_0 + A_1)^2} \int_{\phi_{\min}}^{\phi_{\max}} A_1^2 |\delta(\phi - \phi_1) - \delta(\phi - \phi_0)|^2 d\phi \quad (6.34)$$

$$= \frac{2A_1^2}{(A_0 + A_1)^2} \quad (6.35)$$

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

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 = \frac{A_0^2 + A_1^2 + 1}{(A_0 + A_1)^2}. \quad (6.36)$$

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

$$D_E(F) = \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \quad (6.37)$$

$$= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \quad (6.38)$$

3263 If $\phi_0 = \phi_1$, then the minimised Euclidean distance is 0.

3264 N Hyperparameters for LR and XGB

3265 This section contains tables of the hyperparameters that we used for our classifiers
 3266 in Chapter 6. Table 6.2 and Table 6.3 tabulate the hyperparameters for XGB and LR
 3267 respectively for the 'ATCA' dataset. Table 6.4 and Table 6.5 tabulate the hyperparameters
 3268 for XGB and LR respectively for the 'ASKAP' dataset.

Table 6.2: 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 6.3: LR hyperparameters for the 'ATCA' dataset.

Parameter	Value
penalty	L1
C	1.668

Table 6.4: 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 6.5: LR hyperparameters for the 'ASKAP' dataset.

Parameter	Value
penalty	L2
C	0.464

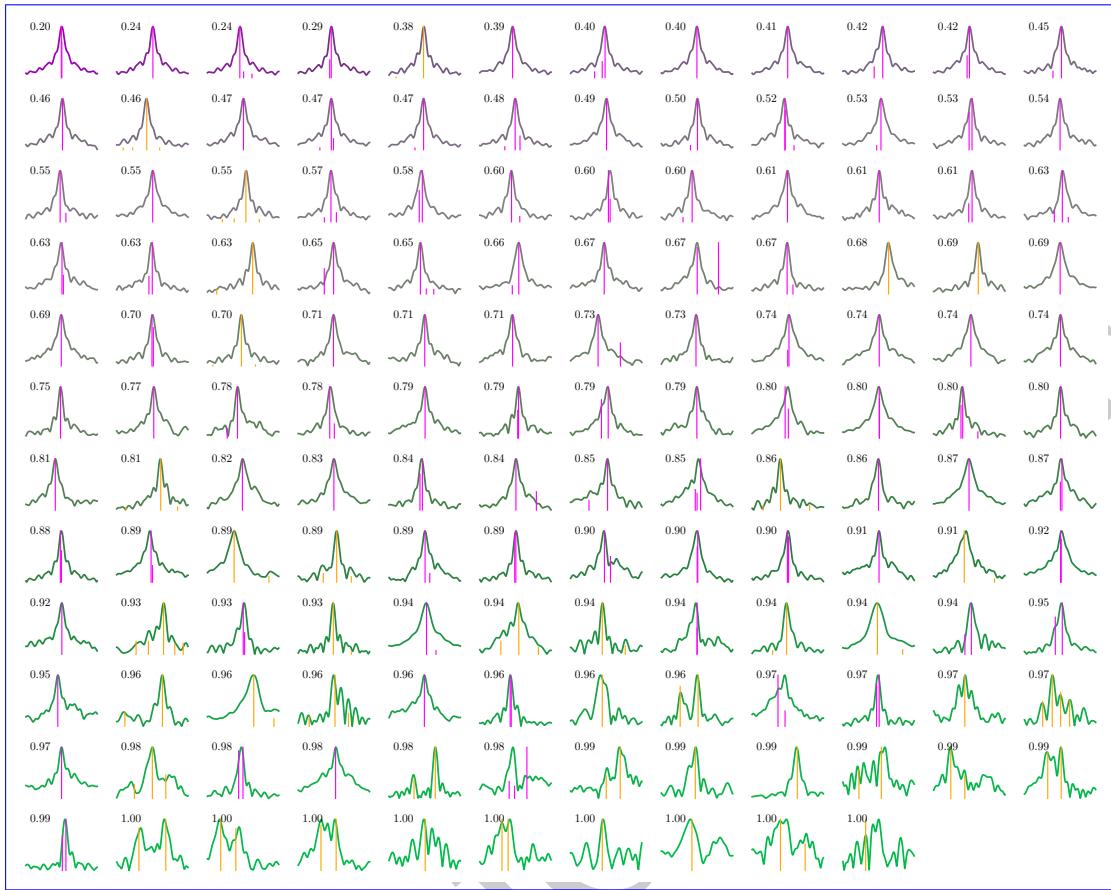


Figure 6.7: The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O'Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are 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.

O Predictions on real data

This appendix contains Figure 6.7 and Figure 6.8. These show the predicted probability of being Faraday complex for all real data used in Chapter 6, drawn from Livingston et al. (2021) and O'Sullivan et al. (2017).

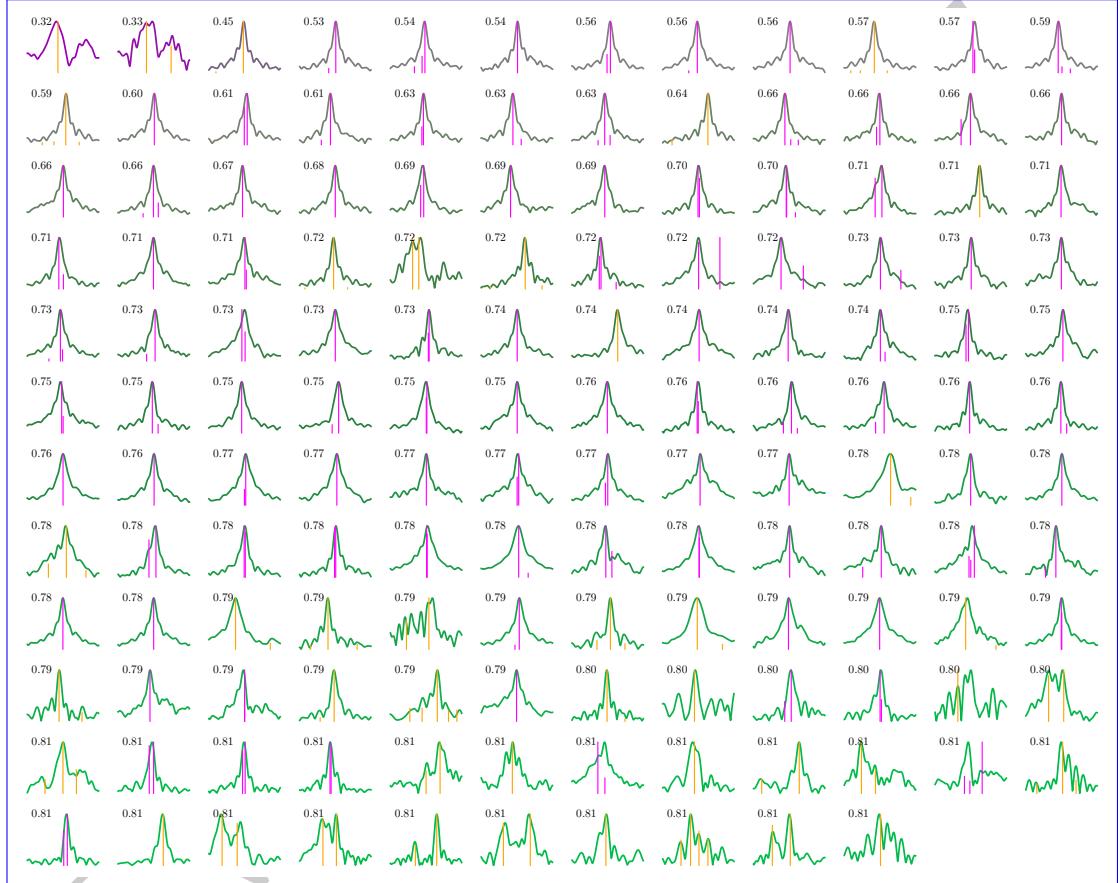


Figure 6.8: The 142 observed FDFs ordered by XGB-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O'Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are 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.

REVISION DIFF

Conclusion

3275 The future of radio astronomy is immensely exciting, with upcoming radio surveys
3276 like EMU and POSSUM sure to revolutionise our understanding of the violent Uni-
3277 verse occupied by radio active galactic nuclei. Without innovative new methods for
3278 processing astronomical data at scale, however, we will be limited in what interesting
3279 physics we can learn about with these surveys. In this thesis we have explored meth-
3280 ods for identifying extragalactic radio sources in these future wide-area radio surveys.
3281 We developed a new automated cross-identification approach and a new way of clas-
3282 sifying radio sources as Faraday complex or Faraday simple. We applied our methods
3283 to real data, and used our cross-identification method to directly probe the mechani-
3284 cal energy contribution of active galactic nuclei (AGN) to the intergalactic medium
3285 (IGM). Our contributions here are all aimed at extracting more information out of the
3286 very large radio surveys we have begun to face.

3287 Chapter 4 developed a new approach to cross-identifying radio components with
3288 their infrared host galaxies and applied this method to the CDFS and ELAIS-S1 AT-
3289 LAS survey, a pilot survey for the upcoming ASKAP key survey EMU. This was the
3290 first application of machine learning to radio cross-identification. In the process, we
3291 demonstrated that our methods worked on ELAIS-S1 with similar performance to how
3292 they worked on CDFS, despite these being different parts of the sky. We also eval-
3293 uated simple positional matching cross-identification on the ATLAS fields, showing
3294 better performance than either machine learning or Radio Galaxy Zoo volunteers on
3295 this dataset, which we suggest is because of how most objects in ATLAS are very com-
3296 pact. Our machine learning methods performed similarly whether they were trained
3297 on expert-provided or citizen science-provided labelled datasets, clearly demonstrat-
3298 ing the benefit of citizen science labels to radio astroinformatics even if they may be
3299 less accurate than labels provided by experts. We showed that the algorithm we had
3300 developed could in principle perform much better than positional matching if only its
3301 binary accuracy could be raised, perhaps with a more complex or better-trained classi-
3302 fier. Finally, we concluded that ATLAS was an insufficiently complex dataset to train
3303 machine learning algorithms for future radio surveys. ATLAS contains many compact
3304 sources, and while there will be many such sources in EMU and other future surveys
3305 there will also be a zoo of partially resolved, extended, strange, or unusual morpholo-
3306 gies. Methods designed or trained on a dataset like ATLAS, which lacks diversity of
3307 non-compact sources, will not work on the wide range of extended radio sources that

3308 will appear. More complex and larger training sets, perhaps real or perhaps simulated,
 3309 will be necessary for producing algorithms intended for use in future surveys.

3310 To further investigate our classification-based cross-identification algorithm we
 3311 needed a more complex dataset, and in Chapter 5 we turned to FIRST, a legacy sur-
 3312 vey containing around 900 000 radio components. Of these, 250 000 or so were not
 3313 compact, compared to the vast majority of ATLAS. We increased the complexity of
 3314 our machine learning model and trained it on this dataset along with 75 000 Radio
 3315 Galaxy Zoo labels, far more than we had available for ATLAS. The resulting classifier
 3316 could then be used to cross-identify every extended object in FIRST. Using the fact
 3317 that any two radio components in the same radio source should also have the same
 3318 host galaxy, we identified not only the host galaxies of our radio components but also
 3319 their association to other radio components. This allowed us to produce the RGZ-Ex
 3320 catalogue, containing over 150 000 extended radio sources—the largest existing cata-
 3321 logue of extended radio sources at the time of writing. In this catalogue we identified
 3322 40 giant radio galaxies, most of which were new to literature.

3323 Such a large catalogue enabled us to estimate a radio luminosity function (RLF)
 3324 for extended radio sources, the first time a RLF has been produced for just extended
 3325 sources. We could also subdivide this RLF into a fractional RLF, and we were able
 3326 to produce RLFs divided by the mid-infrared colour of the host galaxy, their physi-
 3327 cal extent, and their redshift. Extended radio sources ought to be the sources which
 3328 contribute the most mechanical energy to their surrounding IGM, and with an RLF
 3329 dedicated solely to extended sources we were able to estimate this energy contribu-
 3330 tion as 1.3×10^{30} and $1.2 \times 10^{32} \text{ W Mpc}^{-3}$. Perhaps most importantly, we showed our
 3331 cross-identification approach works and used it to obtain a physically meaningful re-
 3332 sult. Our methods can easily be extended to new surveys, as long as sufficient host
 3333 galaxy/radio component pairs are known so that the algorithm can be trained.

3334 Our work on cross-identification allowed us to extract more information from wide-
 3335 area continuum radio surveys. To help gain more use out of large polarisation surveys
 3336 as well, we developed an interpretable Faraday complexity classifier for Faraday dis-
 3337 persions functions (FDFs) in Chapter 6. We constructed features that were easy to un-
 3338 derstand by measuring the distance of observations from a simple model of Faraday
 3339 simple sources. Our resulting features could be calculated from both simulated and
 3340 real observations. We demonstrated the effectiveness of our method on both simulated
 3341 and real data and showed that on simulated data our simple methods matched the
 3342 state-of-the-art convolutional neural network (CNN) classifier despite having far less
 3343 parameters. This was the first application of machine learning to real spectropolarimet-
 3344 ric data. We highlighted the domain gap between simulated and observed FDFs and
 3345 suggested reframing the Faraday complexity classification task as finding simple and
 3346 non-simple sources. This method will be applicable to future surveys like POSSUM.

3347 1 Future Work

3348 There are many ways that the research in this thesis can be extended in future. We will
3349 summarise some of these here.

3350 Our methods can be extended in a number of ways. These fall into two categories:
3351 further applications and extensions to the algorithms. An obvious target for future
3352 work is to apply our binary cross-identification algorithm (BXID; Section 4.3.1 and Sec-
3353 tion 3) to new and upcoming surveys like LotSS and EMU. These promise a tremen-
3354 dous amount of data with new discoveries certain to be lying in wait within, and cross-
3355 identifying the radio emission with its corresponding infrared or optical host galaxy
3356 will be vital for uncovering those secrets. Similarly, we would like to apply our Faraday
3357 complexity classifier to future spectropolarimetric surveys like POSSUM. Our method-
3358 ology can also be improved. Better models almost certainly exist than the CNN we
3359 applied to BXID. As we demonstrated in Chapter 6, a well thought-out model and fea-
3360 tures may best a complex model like the CNN. How would our cross-identification
3361 approach worked if, say, we applied it to hand-selected features such as those chosen
3362 by Proctor (2006)? Would a search over more CNN architectures, like that described
3363 by Lukic et al. (2019), result in better classifiers and hence better cross-identifications?
3364 Perhaps we could even improve performance by pre-training on some unlabelled but
3365 larger dataset? A less obvious improvement to our BXID approach would be to change
3366 how the classification scores are aggregated. Currently this is a weighted maximisa-
3367 tion over candidate host galaxies, but other methods are possible. The weights could
3368 be something other than a Gaussian function of distance, from other functions of dis-
3369 tance to an entire separate classification model. Maybe we could aggregate the scores
3370 in bulk, using some kind of algorithm that assigns radio-host relationships based on
3371 not just the radio source itself, but also on the other radio sources around it and how
3372 they have been paired to their own host galaxies.

3373 The way that our labels were generated for BXID could be improved. Our algo-
3374 rithms in Chapter 4 and Chapter 5 were trained on labels generated by Radio Galaxy
3375 Zoo. These labels were aggregated from multiple different labellers (usually 20) by
3376 majority vote, with the most common label for any given radio object being assigned
3377 as the true label in Radio Galaxy Zoo. This is not the only possible aggregation strategy,
3378 though. We employed the Dawid-Skene method (Dawid & Skene, 1979) ourselves in
3379 Section 3.1 to help assess the performance of our cross-identification algorithm, and
3380 this model for example may also be applied to Radio Galaxy Zoo. There are in fact ag-
3381 gregation strategies that work in tandem with a machine learning model to get better
3382 labels, such as Raykar et al. (2010). These methods simultaneously take into account
3383 the labels and the labellers, and can accommodate for different levels of ability in the
3384 labellers, or different levels of difficulty in the examples being labelled.

3385 Our RLFs could be improved. The RLF calculations in Chapter 5 are severely lim-
3386 ited by the availability of redshifts. We limit our analyses to host galaxies that do have
3387 available spectroscopic redshifts in SDSS, but we could also employ the less-reliable
3388 but considerably more prolific photometric redshifts. These are derived from regres-
3389 sion models rather than direct observations of redshifted spectral lines, and so can

3390 be produced from photometric surveys without dedicated spectroscopy. However,
3391 without methods to handle the uncertainty introduced by photometric redshifts, the
3392 resulting RLFs would be unreliable. Developing ways to not only decrease the uncer-
3393 tainty in photometric redshifts but also to understand and incorporate the uncertainty
3394 into downstream calculations like those of the RLFs, for example using probabilistic
3395 programming (e.g. Bingham et al., 2019), will allow these photometric redshifts to be
3396 used and tremendously increase sample sizes. Besides existing photometric redshifts,
3397 future surveys will also produce many more redshifts, both spectroscopic and photo-
3398 metric. This will be very important for surveys like EMU, which are both deep and
3399 wide with low redshift availability.

3400 The RGZ-Ex dataset (Chapter 5) also lets us pose many other interesting science
3401 questions. We demonstrated in Appendix J that rare galaxy classes can be identified
3402 from within this dataset, including examples that have never before been identified in
3403 the literature. Our dataset may be augmented with other features and used to iden-
3404 tify unusual objects in a similar way. Besides this, our fractional RLFs could also be
3405 extended with any number of galaxy properties. One particularly interesting prop-
3406 erty could be morphology, as other algorithms in radio astroinformatics are developed
3407 which can automatically identify morphologies (e.g. Wu et al., 2019): such a classifier
3408 could be used to segment RGZ-Ex and a fractional morphological RLF could be ob-
3409 tained. These morphologies may even be classes that are not easily separated, such as
3410 those found by self-organising maps (e.g. Polsterer et al., 2015). Of course, there are
3411 other properties that are more easily extracted, such as optical lines and colours which
3412 could be taken from SDSS using our SDSS cross-identifications.

3413 While creating features for FDFs in Chapter 6 we demonstrated that W2 distance
3414 was a sensible distance measure between FDFs. This is useful for more than just feature
3415 construction, as it implies a geometry on the space of FDFs. This distance could be
3416 used to help gain insight on the behaviours of future algorithms that work on FDFs. A
3417 particularly exciting idea is to improve QU fitting by modifying the distance function
3418 to match ours. Our features could also be used to develop other methods for FDF
3419 analysis, like outlier detection or data visualisation.

3420 Further research is needed to close the domain gap for FDFs. This is an interest-
3421 ing case study as it is such a simple case, where we know essentially all the physics
3422 behind the observations and the observations are one-dimensional. Even this is not
3423 enough, and whether through unmodelled physics (e.g. more than two screens, depo-
3424 larisation) or unmodelled observational properties (e.g. radio frequency interference)
3425 simulation and observation do not fully line up. This is critical if we want to train
3426 machine learning algorithms on simulations in the future, and we very much want to
3427 do this to augment our limited observational training data. Similarly, the domain gap
3428 must be reduced for radio continuum observations. Our results in Chapter 4 show that
3429 pilot surveys like ATLAS may not contain enough complex sources to train machine
3430 learning models, and while larger surveys like FIRST exist, transferring models from a
3431 survey undertaken with one set of observing parameters (telescope, frequency, depth,
3432 resolution...) is both non-trivial and as-yet relatively unexplored.

3433 2 Implications for radio citizen science

3434 Our results in Chapter 4 demonstrated that machine learning methods trained on citizen
3435 science labels perform comparably to those trained on expert labels, even when
3436 those labels are lower quality than expert labels. We applied this insight in Chap-
3437 ter 5 to obtain scientific results, using machine learning to extrapolate the labels to a
3438 larger dataset. While Radio Galaxy Zoo alone was not enough to fully label FIRST, Ra-
3439 dio Galaxy Zoo working in tandem with machine learning was. This is a pattern that
3440 may hold true for future surveys and applications, too. Future astronomical research
3441 at-scale may leverage the idea of people working with machine learning, sometimes
3442 called *human-in-the-loop* (e.g. Holzinger, 2016) learning, to pore through data-at-scale.

3443 3 Implications for wide-area radio surveys

3444 As we move toward larger and larger datasets, an important question is how applica-
3445 ble our models will be across the sky. Our results in Chapter 4 showed that we can
3446 expect some generalisation, as our model trained on one patch of sky was applicable
3447 to another without a great loss of performance. Similarly, our classifier trained on part
3448 of FIRST seemed to work well on the rest of the dataset. This is promising as it implies
3449 that limited area surveys may help develop training sets that generalise to the whole
3450 sky, potentially making the process of generating training sets considerably cheaper.

3451 With our work in Chapter 4 and 5 we demonstrated that a large set of good quality,
3452 complex data is required for training good astroinformatics algorithms. Pilot datasets
3453 like ATLAS will not work by themselves: the sources they contain are too simple and
3454 their complex sources are too few. A sensible question to ask is, could we simulate
3455 data for training purposes? We trained our classifier in Chapter 6 on simulated data
3456 and found that it was difficult to bridge the domain gap between simulation and ob-
3457 servation, even in a well-understood, one-dimensional case—let alone the complex
3458 three-dimensional projected morphologies we observe in imagery. Getting across this
3459 domain gap will be difficult and will necessarily be a major topic of research in the
3460 astroinformatics field in the near future.

3461 Tying observations, simulations, and models together are the representation of the
3462 data: features. Our results in Chapter 6 show that judicious choice of features can out-
3463 perform even complex and powerful models. This is good for two reasons. The first
3464 reason is that these features may be more easily interpreted. The meaning of the fea-
3465 tures may be understood to be representative of some physical property, or at least the
3466 relationship between physical reality and predictions may be more easily found. The
3467 second reason is that features may be selected which can generalise well to datasets
3468 beyond just the training set. In other words, features that are less overfit to the train-
3469 ing set. This is of particular concern when developing predictive models on simulated
3470 training data, as features being less suited for real data than for simulations is one as-
3471 pect that may contribute to the domain gap. Choosing good features in astronomy may
3472 be more important than in many other fields for which machine learning is applied, as
3473 while in most fields it is possible to conduct experiments, in astronomy we only have

3474 one Universe to look at. We need to make the best use we can of the limited radio sky.

3475 4 Final remarks

3476 Radio astronomy faces a deluge of data, with current and upcoming surveys delivering
 3477 incredible amounts of data for science use. While we can get a lot out of these data—
 3478 from investigations into new physics, to finding rare and unusual objects—doing so
 3479 is dependent on the development of new methods for astronomy at scale. These fu-
 3480 ture methods will necessarily be computational, and so the challenge lies in encoding
 3481 the abstract concepts of astronomy and astrophysics into a rigorously defined set of
 3482 rules that a computer can interpret and execute at scale. This is decidedly non-trivial,
 3483 and the nuance and unique skills required to do so motivates the burgeoning field of
 3484 astroinformatics. By combining concepts from deep within the often disparate fields
 3485 of astronomy and machine learning, we hope that we will be able to make the absolute
 3486 most of the incredible new technologies and instruments that will arise in the future
 3487 of radio astronomy.

3488 5 Classification models

3489 This appendix describes the three different models we used for binary classification in
 3490 (logistic regression, convolutional neural networks, and random forests) and was part
 3491 of Alger et al. (2018).

3492 4.1 Logistic Regression

3493 Logistic regression is linear in the feature space and outputs the probability that the
 3494 input has a positive label. The model is (Bishop, 2006):

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

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:

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

3495 The logistic regression model is fully differentiable, and the parameters \vec{w} can therefore
 3496 be learned using gradient-based optimisation methods. We used the scikit-learn
 3497 (Pedregosa et al., 2011) implementation of logistic regression with balanced classes.

3498 4.1 Convolutional neural networks

3499 Convolutional neural networks (CNN) are a biologically-inspired prediction model
 3500 for prediction with image inputs. The input image is convolved with a number of
 3501 filters to produce output images called feature maps. These feature maps can then

3502 be convolved again with other filters on subsequent layers, producing a network of
3503 convolutions. The whole network is differentiable with respect to the values of the
3504 filters and the filters can be learned using gradient-based optimisation methods. The
3505 final layer of the network is logistic regression, with the convolved outputs as input
3506 features. For more detail, see subsection II.A, LeCun et al. (1998). We used KERAS
3507 (Chollet et al., 2015) to implement our CNN, accounting for class imbalance by reweighting
3508 the classes.

3509 Architecture of our CNN. Parenthesised numbers indicate the size of output layers
3510 as a tuple (width, height, depth). The concatenate layer flattens the output of the previous
3511 layer and adds the 10 features derived from the candidate host in SWIRE, i.e. the flux
3512 ratios, stellarity indices, and distance. The dropout layer randomly sets 25% of its
3513 inputs to zero during training to prevent overfitting. Diagram based on .

3514 CNNs have recently produced good results on large image-based datasets in astronomy
3515 (Dieleman et al., 2015; Lukic et al., 2018, e.g.). We employed only a simple CNN model
3516 in as a proof of concept that CNNs may be used for class probability prediction on radio
3517 images. The model architecture we used is shown in .

3518 4.1 Random Forests

3519 Random forests are an ensemble of decision trees (Breiman, 2001). They consider
3520 multiple subsamples of the training set, where each subsample is sampled with replacement
3521 from the training set. For each subsample a decision tree classifier is constructed
3522 by repeatedly making axis-parallel splits based on individual features. In a random
3523 forest the split decision is taken based on a random subset of features. To classify
3524 a new data point, the random forest takes the weighted average of all classifications
3525 produced by each decision tree. In we used the scikit-learn (Pedregosa et al., 2011)
3526 implementation of random forests with 10 trees, the information entropy split criterion,
3527 a minimum leaf size of 45, and balanced classes.

3528 5 Accuracy tables

3529 This section contains tables of accuracy for our cross-identification method applied to
3530 CDFS and ELAIS-S1, and was originally presented in Alger et al. (2018). In and we
3531 list the balanced accuracies of our classifiers on the cross-identification task for CDFS
3532 and ELAIS-S1 respectively, averaged over each set of training quadrants. In and we
3533 list the balanced accuracies of classifiers on the cross-identification task for CDFS and
3534 ELAIS-S1 respectively, averaged over each set of training quadrants.

3535 Balanced accuracies for different binary classification models trained and tested on
3536 SWIRE objects in CDFS. The ‘Labeller’ column states what set of training labels were
3537 used to train the classifier, and the ‘Classifier’ column states what classification model
3538 was used. ‘CNN’ is a convolutional neural network, ‘LR’ is logistic regression and ‘RF’
3539 is random forests. Accuracies are evaluated against the expert label set derived from
3540 Norris et al. (2006). The standard deviation of balanced accuracies evaluated across
3541 the four quadrants of CDFS () is also shown. The ‘compact’ set refers to SWIRE objects

3542 within 1' of a compact radio component, the 'resolved' set refers to SWIRE objects
 3543 within 1' of a resolved radio component, and 'all' is the union of these sets. Labeller
 3544 Classifier Mean 'Compact' Mean 'Resolved' Mean 'All' accuracy accuracy accuracy (per
 3545 cent) (per cent) (per cent) Norris LR 91.5 ± 1.0 93.2 ± 1.0 93.0 ± 1.2 CNN 92.6 ± 0.7
 3546 91.2 ± 0.5 92.0 ± 0.6 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 CNN
 3547 89.4 ± 0.6 89.6 ± 1.3 89.4 ± 0.5 RF 94.5 ± 0.2 95.8 ± 0.4 94.7 ± 0.3

3548 Balanced accuracies for different binary classification models trained on SWIRE
 3549 objects in CDFS and tested on SWIRE objects in ELAIS-S1. Columns and abbreviations
 3550 are as in . Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008)
 3551 . The standard deviations of balanced accuracies of models trained on the four subsets
 3552 of CDFS () are also shown. Labeller Classifier Mean 'Compact' Mean 'Resolved' Mean
 3553 'All' accuracy accuracy accuracy (per cent) (per cent) (per cent) Norris LR 94.6 ± 0.4
 3554 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
 3555 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
 3556 95.2 ± 0.3

3557 Cross-identification accuracies for different classification models on CDFS. The 'Labeller'
 3558 column states what set of training labels were used to train the method, and the 'Classifier'
 3559 column states what classification model was used. 'CNN' is a convolutional neural
 3560 network, 'LR' is logistic regression, 'RF' is random forests, and 'Labels' is the accuracy
 3561 of the label set itself. 'Perfect' indicates that the true labels of the test set were used and
 3562 hence represents an upper bound on cross-identification accuracy with our method.
 3563 'NN' is a nearest neighbours approach. Accuracies are evaluated against the expert
 3564 label set, so 'Norris' labels are 100 per cent accurate by definition. The standard deviation
 3565 of accuracies evaluated across the four quadrants of CDFS () is also shown. Labeller
 3566 Classifier Mean 'Compact' Mean 'Resolved' Mean 'All' accuracy accuracy accuracy (per
 3567 cent) (per cent) (per cent) NN 97.2 ± 1.7 75.7 ± 7.9 93.4 ± 0.8 Random 97.9 ± 2.2
 3568 22.3 ± 9.2 83.2 ± 4.7 Norris Labels 100.0 ± 0.0 100.0 ± 0.0 100.0 ± 0.0 Perfect 97.9 ± 2.2
 3569 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 RF
 3570 96.1 ± 1.4 75.8 ± 6.7 93.8 ± 2.0 RGZ Labels 53.1 ± 8.5 56.7 ± 5.9 54.4 ± 5.9 LR 97.3 ± 1.9
 3571 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

3572 Cross-identification accuracies for different classification models on ELAIS-S1. Columns
 3573 and abbreviations are as in . Accuracies are evaluated against the expert label set
 3574 derived from Middelberg et al. (2008) cross-identifications. The standard deviation
 3575 of accuracies evaluated across models trained on the four quadrants of CDFS () is
 3576 also shown. Labeller Classifier Mean 'Compact' Mean 'Resolved' Mean 'All' accuracy
 3577 accuracy accuracy (per cent) (per cent) (per cent) NN 95.5 ± 0.0 92.8 ± 0.0 95.5 ± 0.0 Random
 3578 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
 3579 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
 3580 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
 3581 91.3 ± 1.0 90.3 ± 2.4 94.7 ± 1.2

3582 **5 SWIRE object scores**

3583 This appendix is from Alger et al. (2018), and contains scores predicted by our binary
 3584 classifiers for each SWIRE object within 1' of a radio component in CDFS and ELAIS-S1.
 3585 Scores for SWIRE CDFS objects are shown in and scores for SWIRE ELAIS-S1 are
 3586 shown in . For CDFS, the score for an object in a quadrant is predicted by binary
 3587 classifiers trained on all other quadrants. For ELAIS-S1, we show the scores predicted
 3588 by binary classifiers trained on each CDFS quadrant. Note that these scores have *not*
 3589 been weighted by Gaussians. These are partial tables, and the full tables are available
 3590 online at the *Monthly Notices of the Royal Astronomical Society* website¹.

3591 The columns of the score tables are defined as follows:

- 3592 • *SWIRE*—SWIRE designation for candidate host galaxy.
- 3593 • *RA*—Right ascension (J2000).
- 3594 • *Dec*—Declination (J2000).
- 3595 • *Expert host*—Whether the candidate host galaxy is a host galaxy according to
 3596 Norris et al. (2006) or Middelberg et al. (2008) cross-identifications of CDFS and
 3597 ELAIS-S1 respectively.
- 3598 • *RGZ host*—Whether the candidate host galaxy is a host galaxy according to Radio
 3599 Galaxy Zoo cross-identifications (Wong et al., in prep.). This is always ‘no’ for
 3600 ELAIS-S1 objects.
- 3601 • *C/L/D*—Score assigned by binary classifier *C* trained on label set *L* of *D* candidate
 3602 host galaxies. *C* may be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or
 3603 random forests respectively. *L* may be ‘Norris’ or ‘RGZ’ for expert and Radio
 3604 Galaxy Zoo labels respectively. *D* may be ‘All’, ‘Compact’ or ‘Resolved’ for each
 3605 respective subset defined in .

3606 Scores output by our trained classifiers for SWIRE CDFS candidate host galaxies.
 3607 Columns are defined in . Full table electronic. SWIRE RA Dec Expert RGZ host host All
 3608 Compact Resolved All Compact Resolved J032603.15-284708.5 51.5132-28.7857 yes no
 3609 0.5838-0.4697 0.4848-0.3754 0.3881-0.3404 J032603.39-284010.1 51.5142-28.6695 no no
 3610 0.0373-0.5814 0.4878-0.7896 0.7616-0.4668 J032603.44-284210.1 51.5144-28.7028 no no
 3611 0.0232-0.4891 0.5101-0.4319 0.4298-0.3474 J032603.44-284222.2 51.5143-28.7062 no no
 3612 0.0006-0.4164 0.5216-0.0400 0.0444-0.0276 J032603.45-284748.4 51.5144-28.7968 no no
 3613 0.0014-0.4914 0.4865-0.1904 0.1895-0.1467 J032603.50-284637.0 51.5146-28.7770 no no
 3614 0.0074-0.4144 0.5382-0.1418 0.1515-0.1166 J032603.60-284627.4 51.5150-28.7743 no no
 3615 0.0012-0.4578 0.5165-0.0850 0.0904-0.0484 J032603.63-283840.5 51.5151-28.6446 no no
 3616 0.0021-0.4153 0.5577-0.1678 0.1746-0.1323 J032603.66-283822.8 51.5153-28.6397 no no
 3617 0.0001-0.4752 0.5009-0.0864 0.0861-0.0613 J032603.75-284014.1 51.5156-28.6706 no no
 3618 0.0547-0.3408 0.5388-0.4889 0.5242-0.7301 All Compact Resolved All Compact Resolved
 3619 All Compact Resolved All Compact Resolved 0.2489-0.0009 0.1557-0.2939 0.0007-0.1174
 3620 0.8922-0.8018 0.8732-0.7167 0.6599-0.7801 0.0183-0.1646 0.1480-0.7637 0.7065-0.6070 0.0000

3621 0.0000 0.0000 0.1629 0.0519 0.1275 0.0155 0.0164 0.0815 0.3714 0.5626 0.2488 0.0000 0.0734
 3622 0.0000 0.1315 0.2116 0.4150 0.0005 0.0006 0.0175 0.0460 0.0810 0.0299 0.2656 0.1418 0.0000
 3623 0.7631 0.8166 0.5378 0.0013 0.0037 0.0160 0.1792 0.0663 0.1821 0.0000 0.0000 0.0000 0.0255
 3624 0.0000 0.0000 0.0047 0.0010 0.0337 0.1284 0.2198 0.0694 0.0720 0.0000 0.0000 0.6240 0.6681
 3625 0.6704 0.0008 0.0006 0.0374 0.1053 0.1424 0.0807 0.1231 0.0876 0.0000 0.8517 0.7532 0.7019
 3626 0.0021 0.0073 0.0386 0.1482 0.0403 0.1210 0.0000 0.0532 0.0000 0.0000 0.0302 0.0000 0.0001
 3627 0.0004 0.0038 0.0854 0.0447 0.0514 0.0000 0.0000 0.0000 0.0000 0.0000 0.0542 0.2712
 3628 0.2318 0.5026 0.5631 0.5032 0.0595 0.0545 0.0000 0.4289 0.0789 0.1420
 3629 Scores output by our trained classifiers for SWIRE-ELAIS-S1 candidate host galaxies.
 3630 Columns are defined in . Full table electronic. SWIRE-RA Dec Expert RGZ host host
 3631 All Compact Resolved All Compact Resolved J002925.73-440256.27.3572-44.0490 yes
 3632 no 0.9537 0.8638 0.5552 0.9195 0.9037 0.9371 J002926.14-440249.07.3590-44.0470 no no
 3633 0.7361 0.8752 0.5640 0.7740 0.7474 0.7952 J002926.52-440247.07.3605-44.0464 no no 0.3390
 3634 0.8338 0.5556 0.7275 0.6894 0.7197 J002926.63-440301.17.3610-44.0503 no no 0.2108 0.8251
 3635 0.5623 0.3434 0.3306 0.3292 J002927.13-440232.67.3631-44.0424 no no 0.0339 0.8479 0.5669
 3636 0.5853 0.5148 0.5159 J002927.28-440245.37.3637-44.0459 no no 0.0406 0.8345 0.5540 0.2702
 3637 0.2340 0.2133 J002927.44-440238.57.3644-44.0440 no no 0.0116 0.8267 0.5746 0.2228 0.2182
 3638 0.2028 J002928.08-440230.37.3670-44.0418 no no 0.0024 0.8626 0.5791 0.2297 0.1963 0.1549 J002928.11-440312.7
 3639 7.3671-44.0535 no no 0.0011 0.8159 0.5514 0.0377 0.0384 0.0271 J002928.80-440306.87.3700
 3640 -44.0519 no no 0.0003 0.8405 0.5668 0.0236 0.0226 0.0136 All Compact Resolved All Compact
 3641 Resolved All Compact Resolved All Compact Resolved 0.9722 0.9955 0.8769 0.9933
 3642 0.9934 0.9658 0.8824 0.9664 0.7950 0.8078 0.9227 0.7677 0.4669 0.0111 0.4249 0.3926 0.2220
 3643 0.5947 0.2077 0.0000 0.1613 0.1876 0.0852 0.4546 0.2264 0.0254 0.2389 0.6275 0.3033 0.6812
 3644 0.1347 0.0857 0.0399 0.3582 0.4854 0.5347 0.0603 0.0007 0.0734 0.0688 0.0141 0.1581 0.0917
 3645 0.0000 0.0399 0.2846 0.1245 0.2833 0.0248 0.0334 0.0301 0.5735 0.5065 0.5265 0.1977 0.1507
 3646 0.0000 0.3334 0.6593 0.3995 0.0173 0.0016 0.0359 0.1056 0.0492 0.1456 0.0000 0.0000 0.0000
 3647 0.0000 0.0000 0.0287 0.0064 0.0049 0.0187 0.1981 0.1534 0.1493 0.0000 0.0000 0.0000 0.1565
 3648 0.1634 0.1284 0.0020 0.0005 0.0239 0.1337 0.1001 0.1310 0.0000 0.0000 0.0358 0.0000 0.0000
 3649 0.0190 0.0008 0.0013 0.0119 0.0280 0.0361 0.0205 0.1171 0.0000 0.0000 0.0873 0.0383 0.0000
 3650 0.0004 0.0014 0.0095 0.0339 0.0408 0.0136 0.0000 0.0000 0.0000 0.1114 0.1480 0.1584

3651 5 ATLAS component cross-identifications

3652 This section contains cross-identifications predicted by our cross-identification method
 3653 for each ATLAS radio component in CDFS and ELAIS-S1. Cross-identifications for
 3654 ATLAS-CDFS components are shown in and cross-identifications for ATLAS-ELAIS-S1
 3655 are shown in . For CDFS, the cross-identification for a component in a quadrant is
 3656 predicted using our method with binary classifiers trained on all other quadrants. For
 3657 ELAIS-S1, we show the cross-identifications predicted by our method using binary
 3658 classifiers trained on each CDFS quadrant. For CDFS, we also show the Radio-Galaxy
 3659 Zoo consensus, which is a proxy for the difficulty of cross-identifying a component
 3660 (Wong et al., in prep.). These are partial tables, and the full tables are available online

3661 at the *Monthly Notices of the Royal Astronomical Society* website¹.

3662 The columns of the cross-identification tables are defined as follows:

3663 • *ATLAS*—ATLAS designation for radio component.

3664 • *RA*—Right ascension of radio component (J2000).

3665 • *Dec*—Declination of radio component (J2000).

3666 • *CID*—Radio Galaxy Zoo component ID.

3667 • *Zooniverse ID*—Radio Galaxy Zoo Zooniverse ID.

3668 • *Norris/Middleberg*—Designation of SWIRE cross-identification from Norris et al. (2006) or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.

3670 • *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.

3672 • *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.

3674 • *RGZ*—Designation of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.).

3675 .

3676 • *RGZ RA*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.).

3678 • *RGZ Dec*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.).

3680 • *RGZ radio consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on the radio component configuration.

3682 • *RGZ IR consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on the host galaxy of this radio component.

3684 • *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 .

3689 • *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.

-
- 3692 • $C / L / D$ Dec—Declination (J2000) of SWIRE cross-identification made by our
 3693 method using classification model C trained on label set L of D candidate host
 3694 galaxies. C , L and D are defined as for designation.

3695 Cross-identifications for ATLAS CDFS components. Columns are defined in . Full
 3696 table electronic. ATLAS RA Dec CID Zooniverse ID RA Dec RA Dec radio IR RA Dec
 3697 RA Dec consensus consensus J032602.82-284708.1C 51.5117-28.7856 CI0412 ARG0003rb2
 3698 J032603.15-284708.551.5132-28.7857 0.4516 0.3214 J032615.49-284629.4C 51.5646-28.7749
 3699 CI0614 ARG0003rfr J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642
 3700 -28.7752 0.2941 0.8000 J032615.55-280559.8C 51.5648-28.1000 CI0320 ARG0003r8s J032615.52-280559.8
 3701 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 0.5625 0.8333 J032617.35-280710.2C
 3702 51.5723-28.1195 CI0059C1 ARG0003r2j J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2
 3703 51.5746-28.1187 0.4146 1.0000 J032625.13-280909.8C 51.6047-28.1527 CI0409 ARG0003raz
 3704 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 0.3158 0.6667
 3705 J032629.10-280650.1C 51.6213-28.1139 CI0963 ARG0003ro4 J032629.13-280650.7 51.6214
 3706 -28.1141 J032626.74-280636.7 51.6114-28.1102 0.3333 1.0000 J032629.61-284052.7C 51.6234
 3707 -28.6813 CI0304 ARG0003r8e J032629.54-284055.8 51.6231-28.6822 J032629.54-284055.8
 3708 51.6231-28.6822 0.2676 1.0000 J032629.92-284753.5C 51.6247-28.7982 CI0120 ARG0003r3w
 3709 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 1.0000 0.8571
 3710 J032630.66-283657.3C 51.6278-28.6159 CI0172C1 ARG0003r55 J032630.64-283658.0 51.6277
 3711 -28.6161 J032628.56-283744.8 51.619-28.6291 0.3611 0.7308 J032634.59-282022.8C 51.6441
 3712 -28.3397 CI0757 ARG0003rj2 J032634.58-282022.8 51.6441-28.3397 J032631.96-281941.0
 3713 51.6332-28.3281 0.5781 0.5405 RA Dec RA Dec RA Dec RA Dec J032602.36-284711.5
 3714 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865
 3715 J032602.36-284711.5 51.5098-28.7865 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7
 3716 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.52-280559.8
 3717 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000
 3718 J032615.52-280559.8 51.5647-28.1000 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2
 3719 51.5746-28.1187 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2 51.5746-28.1187 J032625.19-280910.1
 3720 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032624.50-280905.9 51.6021-28.1517
 3721 J032625.19-280910.1 51.6050-28.1528 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7
 3722 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.54-284051.9
 3723 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811
 3724 J032629.54-284051.9 51.6231-28.6811 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4
 3725 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032630.64-283658.0
 3726 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161
 3727 J032630.64-283658.0 51.6277-28.6161 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8
 3728 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397 RA
 3729 Dec RA Dec RA Dec J032604.58-284650.9 51.5191-28.7808 J032602.08-284713.1
 3730 51.5087-28.787 J032602.36-284711.5 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865 J032615.41-284630.7
 3731 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752
 3732 J032615.41-284630.7 51.5642-28.7752 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8
 3733 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.86-280628.8
 3734 51.5661-28.1080 J032615.16-280742.2 51.5632-28.1284 J032615.86-280628.8 51.5661-28.1080

3735 J032618.84-280722.6 51.5785-28.1230 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1
 3736 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032629.13-
 3737 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141
 3738 J032629.13-280650.7 51.6214-28.1141 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9
 3739 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.81-
 3740 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985
 3741 J032629.81-284754.4 51.6242-28.7985 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0
 3742 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032634.58-
 3743 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397
 3744 J032634.58-282022.8 51.6441-28.3397 RA Dec RA Dec RA Dec J032603.15-284708.5
 3745 51.5132-28.7857 J032602.36-284711.5 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865
 3746 J032602.36-284711.5 51.5098-28.7865 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7
 3747 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752
 3748 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8
 3749 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032617.89-280707.2 51.5746-28.1187
 3750 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2
 3751 51.5746-28.1187 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528
 3752 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032629.13-280650.7
 3753 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141
 3754 J032629.13-280650.7 51.6214-28.1141 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9
 3755 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811
 3756 J032630.12-284751.2 51.6255-28.7976 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4
 3757 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032630.64-283658.0 51.6277-28.6161
 3758 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0
 3759 51.6277-28.6161 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397
 3760 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397

3761 Cross-identifications for ATLAS ELAIS-S1 components. Columns are defined in
 3762 . Full-table electronic. ATLAS RA Dec CID Zooniverse ID RA Dec RA Dec radio
 3763 IR RA Dec RA Dec consensus consensus J002925.68-440256.8 7.3570-44.0491-C0375
 3764 J002925.73-440256.2 7.3572-44.0490 J002938.19-432946.77.4092-43.4963 C0832 J002938.07-432947.9
 3765 7.4087-43.4967 J002940.13-440309.2 7.4172-44.0526 C0374 J002940.19-440309.6 7.4175
 3766 -44.0527 J002943.14-440812.3 7.4298-44.1368 C0302 J002943.15-440813.6 7.4298-44.1371
 3767 J002944.51-433627.8 7.4355-43.6077 C0727 J002944.36-433630.2 7.4348-43.6084 J002945.31-432148.5
 3768 7.4388-43.3635 C0943.1 J002945.64-432149.3 7.4402-43.3637 J002946.14-432149.1 7.4423
 3769 -43.3637 C0943 J002945.64-432149.3 7.4402-43.3637 J002949.89-440541.4 7.4579-44.0948
 3770 C0345 J002951.13-432354.3 7.4631-43.3984 C0924 J002951.14-432355.3 7.4631-43.3987
 3771 J002951.19-440556.6 7.4633-44.0991 C0342 J002951.26-440556.4 7.4636-44.0990 RA Dec
 3772 RA Dec RA Dec RA Dec J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572
 3773 -44.0490 J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490 J002938.07-432947.9
 3774 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002937.50-432945.4 7.4063-43.4959
 3775 J002937.50-432945.4 7.4063-43.4959 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6
 3776 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002943.15-4408
 3777 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371
 3778 J002943.15-440813.6 7.4298-44.1371 J002944.36-433630.2 7.4348-43.6084 J002944.36-433630.2

3779 7.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002945.64-432149.3
 3780 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637
 3781 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3782 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002951.44-440546.1
 3783 7.4644-44.0962 J002951.44-440546.1 7.4644-44.0962 J002951.44-440546.1 7.4644-44.0962
 3784 J002951.44-440546.1 7.4644-44.0962 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3
 3785 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.26-440556.4
 3786 7.4636-44.0990 J002951.44-440546.1 7.4644-44.0962 J002951.51-440617.1 7.4646-44.1048
 3787 J002951.44-440546.1 7.4644-44.0962 RA Dee RA Dee RA Dee J002925.73-440256.2
 3788 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490
 3789 J002925.73-440256.2 7.3572-44.0490 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9
 3790 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002940.19-440309.6
 3791 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527
 3792 J002940.19-440309.6 7.4175-44.0527 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6
 3793 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002944.36-433630.2
 3794 7.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002944.36-433630.27.4348-43.6084
 3795 J002944.36-433630.27.4348-43.6084 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3796 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3797 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637
 3798 J002945.64-432149.3 7.4402-43.3637 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4
 3799 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.14-432355.3
 3800 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987
 3801 J002951.14-432355.3 7.4631-43.3987 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4
 3802 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 RA
 3803 Dee RA Dee RA Dec J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490
 3804 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490
 3805 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9
 3806 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002940.19-440309.6 7.4175-44.0527
 3807 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6
 3808 7.4175-44.0527 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371
 3809 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002944.36-433630.2
 3810 7.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002944.36-433630.27.4348-43.6084
 3811 J002944.36-433630.27.4348-43.6084 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3812 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637
 3813 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3814 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002951.26-440556.4 7.4636-44.0990
 3815 J002951.26-440556.4 7.4636-44.0990 J002949.13-440536.5 7.4547-44.0935 J002949.13-440536.5
 3816 7.4547-44.0935 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987
 3817 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.26-440556.4
 3818 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990
 3819 J002951.26-440556.4 7.4636-44.0990

3820 5 Cross-identification figures

3821 This section contains figures of our cross-identifications of each ATLAS radio component
 3822 in CDFS and ELAIS-S1, and was originally an appendix to Alger et al. (2018). These
 3823 are just five examples, but all 469 examples are available online at the *Monthly Notices*
 3824 of the Royal Astronomical Society website¹.

3825 One resolved component and resolved source. Three resolved components comprising
 3826 one resolved source. Three resolved components comprising one resolved source.
 3827 Three resolved components comprising one resolved source. Two compact components,
 3828 each a compact source. Examples of resolved sources with high disagreement between
 3829 cross-identifiers. The contours show ATLAS radio data and start at 4σ , increasing
 3830 geometrically by a factor of 2. The background image is the SWIRE image. Binary
 3831 classifier model/training set combinations are denoted $C(S)$ where C is the binary
 3832 classifier model and S is the training set. ‘LR’ is logistic regression, ‘CNN’ is convolutional
 3833 neural networks, and ‘RF’ is random forests. ‘Norris’ refers to the expert labels and
 3834 ‘RGZ’ refers to the Radio Galaxy Zoo labels. The cross-identification made by nearest
 3835 neighbours is shown by ‘NN’. The complete set of figures for 469 examples is available
 3836 in the supplementary information online.

3837 5 Sankey diagrams

3838 This section presents Sankey diagrams showing the filtering of components and sources
 3839 from the full FIRST sample in, and was originally an appendix to Alger et al. (in prep.)
 3840 . A Sankey diagram shows the order and number of objects removed from a sample.
 3841 shows the filtering of components and shows the filtering of sources. The component
 3842 filters are ‘Bad FIRST’ for components on the edge of FIRST with incomplete images,
 3843 ‘Sidelobe’ for components with high sidelobe probability, ‘Low score’ for components
 3844 with only low-scoring candidate hosts, ‘Faint’ for components with less than 10 signal-to-noise
 3845 according to the FIRST catalogue, and ‘Compact’ for components that do not have
 3846 extended radio emission according to . Sources were removed after each component
 3847 filter if they no longer contained any components.

3848 Number of components removed from FIRST by each filter.

3849 Number of sources removed by each filter.

3850 5 Radio luminosity function

3851 We computed the radio luminosity function following the $1/V_{\max}$ method (Schmidt, 1968)
 3852 . This appendix explains our implementation in and was originally an appendix to
 3853 Alger et al. (in prep.). We performed the following steps:

- 3854 1. Remove all radio sources that do not fit the selection criteria. This applies for
 3855 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 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 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.

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

~~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 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 .~~

3866 5 Redshift completeness estimate

~~shows the estimated completeness of our RLF sample in as a function of W1 and W1 – W2. 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 W1 and W1 – W2 plane. This appendix was originally part of Alger et al. (in prep.).~~

~~Estimated completeness as a function of mid-infrared colour and magnitude.~~

3873 5 Giant radio galaxies

~~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~~

3881 $1.0\text{sJ}080831.68+473523.9^R$ $122.13\ 47.59\ 0.58854 \pm 0.00016$ $1.1\text{sJ}083034.78+231124.6$ 127.64
 3882 $23.19\ 0.94 \pm 0.13$ $1.1\text{pJ}090604.03+011114.2$ $136.52\ 1.19\ 0.7975 \pm 0.0004$ $1.6\text{sJ}093256.81+074212.2$
 3883 $143.24\ 7.70$ 1.0032 ± 0.0003 $1.1\text{sJ}093526.80+051729.8^R$ $143.86\ 5.29\ 0.84 \pm 0.04$ $1.2\text{pJ}094238.72+114337.9$
 3884 $145.66\ 11.73$ 0.49 ± 0.05 $1.2\text{pJ}094835.60+535946.4^R$ $147.15\ 54.00\ 0.64 \pm 0.10$ $1.2\text{pJ}095706.12+292439.2$
 3885 $149.28\ 29.41$ 0.71 ± 0.12 $1.5\text{pJ}102335.25+433208.0$ $155.90\ 43.54\ 0.75 \pm 0.09$ $1.5\text{pJ}102933.99+210345.8^R$
 3886 $157.39\ 21.06$ 0.82407 ± 0.00008 $1.1\text{sJ}103043.98+355451.2^R$ $157.68\ 35.91\ 0.64074 \pm 0.00008$
 3887 $1.2\text{sJ}104449.92+234525.6^+$ $161.20\ 23.76\ 0.57712 \pm 0.00009$ $1.6\text{sJ}110655.98+624759.8^R$
 3888 $166.73\ 62.80$ 0.84379 ± 0.00004 $1.1\text{sJ}112900.68+635543.2$ $172.25\ 63.93\ 0.71 \pm 0.06$ $1.1\text{pJ}112948.20+243922.6$
 3889 $172.45\ 24.66\ 0.79 \pm 0.07$ $1.1\text{pJ}114553.67-003304.7$ $176.47\ 0.55$
 3890 2.0522 ± 0.0006 $1.3\text{sJ}121111.26+534840.4$ $182.80\ 53.81\ 0.74 \pm 0.14$ $1.1\text{pJ}121152.04+304232.4^R$
 3891 $182.97\ 30.71\ 0.47102 \pm 0.00012$ $1.3\text{sJ}121944.73+174121.3$ $184.94\ 17.69\ 1.5129 \pm 0.0009$
 3892 $1.0\text{sJ}123735.89+544814.4^R$ $189.40\ 54.80\ 1.0271 \pm 0.0006$ $1.2\text{sJ}123819.16+113444.8$ 189.58
 3893 $11.58\ 0.80 \pm 0.08$ $1.2\text{pJ}123846.84-032857.5^+$ $189.70\ 3.48\ 0.67 \pm 0.07$ $1.5\text{pJ}131625.00+272042.8$
 3894 $199.10\ 27.35\ 0.69092 \pm 0.00004$ $1.0\text{sJ}133307.00+045048.6^R$ $203.28\ 4.85\ 1.40534 \pm 0.00016$
 3895 $1.1\text{sJ}141933.36+104706.4^R$ $214.89\ 10.79\ 0.33973 \pm 0.00003$ $1.0\text{sJ}142008.45+185422.7^R$
 3896 $215.04\ 18.91\ 0.63 \pm 0.04$ $1.4\text{pJ}145057.28+530007.7^L$ $222.74\ 53.00\ 0.91662 \pm 0.00009$ $1.3\text{sJ}150012.18+604941.3$
 3897 $225.05\ 60.83\ 1.6626 \pm 0.0007$ $1.2\text{sJ}153547.13+432245.0^R$ 233.95
 3898 $43.38\ 0.63891 \pm 0.00007$ $1.3\text{sJ}154631.18+194819.9$ $236.63\ 19.81\ 0.5917 \pm 0.0002$ $1.4\text{sJ}160852.10+561110.7$
 3899 $242.22\ 56.19\ 1.3196 \pm 0.0003$ $1.3\text{sJ}162200.48+364044.0$ $245.50\ 36.68\ 1.9994 \pm 0.0002$ $1.1\text{sJ}163004.35+103321.9^R$
 3900 $247.52\ 10.56\ 0.85 \pm 0.09$ $1.2\text{pJ}163125.75+200224.1^R$ $247.86\ 20.04$
 3901 0.62662 ± 0.00013 $1.0\text{sJ}165055.46+394446.6$ $252.73\ 39.75\ 0.58829 \pm 0.00013$ $1.1\text{sJ}232410.33+045309.6$
 3902 $351.04\ 4.89\ 0.76 \pm 0.06$ $1.4\text{pJ}234440.02-003231.6$ $356.17\ 0.54\ 0.5014 \pm 0.0001$ 1.0s

3903 This appendix describes our search for giant radio galaxies in RGZ-Ex, and the
 3904 results of this search. It was originally an appendix to Alger et al. (in prep.). To identify
 3905 radio sources we assumed that if any two components had the same host galaxy then
 3906 they were part of the same source. This is a reasonable assumption if all host galaxies
 3907 are correctly identified, which was not the case. This assumption therefore introduced
 3908 spurious sources due to galaxies incorrectly identified as host galaxies: not all sources
 3909 used in are real sources, and in particular sources of large angular size are likely to be
 3910 incorrect. Nevertheless RGZ-Ex provides a useful catalogue of *candidate* radio sources,
 3911 and visual follow-up can confirm whether sources of interest are real.

3912 H.A. and M.J.A. examined all 296 candidate sources in the RGZ-Ex catalogue with
 3913 an estimated physical extent larger than 1 Mpc. Of these, 40 were real giant radio
 3914 galaxies, which we show in. We defined ‘giant radio galaxy’ as a radio galaxy with
 3915 emission extended to physical sizes ≥ 1.0 Mpc. Other thresholds, such as 0.7 Mpc,
 3916 also exist in literature. The physical extents of the remaining 256 candidate sources
 3917 were overestimated mostly due to sidelobes/artefacts (103), incorrect source grouping
 3918 (82), or incorrect SDSS matches (21). The citizen scientists who identified giants are:
 3919 WizardHowl, DolorousEdd, antikodon, csunjoto, sisifolibre, JeanTate, JKD, PADV, and
 3920 firejuggler. H.A., together with his summer students, had previously identified 29 of
 3921 these giants.

3922 Note that this is a particularly challenging set: sources that are misidentified will
 3923 often have unusually large estimated extents due to the inclusion of spurious components.
 3924 The error rate in this set therefore does not reflect the rest of the catalogue.

3925 **5 Visual verification results**

3926 In we described our visual verification of the BXID method from . We list the radio
 3927 components in the verification set in . Each row of the table contains the FIRST component,
 3928 its AllWISE host galaxy according to BXID, and whether the association is correct
 3929 according to our visual verification. If an author was particularly unsure about an
 3930 object, they were able to skip this object, and so are not accounted for in the verification
 3931 for that object. Verification was weighted by the Dawid and Skene (1979) maximum
 3932 likelihood model. This appendix was originally part of (Alger et al., in prep.).

3933 Validation objects. ‘Agree’ is whether or not the authors of Alger et al. (in prep.)
 3934 agreed with BXID associating the given FIRST object with the given AllWISE object.
 3935 FIRST AllWISE AgreeJ000234.9-001421 J000242.35-001320.5 nJ002841.1+141654 J002840.37+141652.7
 3936 yJ003731.4+000156 J003731.26+000146.7 yJ005407.5-011158 J005407.61-011158.9 yJ011210.3+002203
 3937 J011210.41+002201.9 yJ012342.4+015849 J012342.24+015850.4 yJ013015.1+110653 J013015.16+110653.4
 3938 yJ013107.7+070343 J013102.02+070332.0 yJ014247.9-000039 J014247.81-000040.3 yJ014250.0-000032
 3939 J014247.81-000040.3 nJ020222.3+030138 J020223.20+030150.4 yJ020333.8+000853 J020336.94+000759.3
 3940 yJ021840.1-032311 J021840.13-032306.0 yJ023022.0+010834 J023022.11+010840.0 yJ024245.3-022535
 3941 J024245.35-022534.6 yJ025901.0+005350 J025901.50+005346.1 yJ033204.1-004757 J033204.15-004757.1
 3942 yJ073033.2+390413 J073033.21+390412.9 yJ073954.1+481810 J073954.87+481759.5 yJ074504.9+331247
 3943 J074504.81+331256.2 yJ074640.4+421709 J074640.45+421709.1 yJ074707.9+171719 J074708.35+171726.5
 3944 yJ075043.6+274838 J075043.35+274844.8 nJ075050.3+331937 J075051.25+331905.0 yJ075422.2+311253
 3945 J075422.35+311252.5 yJ075637.0+212006 J075636.65+212001.4 yJ082326.1+141438 J082326.34+141435.9
 3946 yJ082422.5+351121 J082422.65+351114.6 yJ082925.9+462618 J082926.02+462618.5 yJ083512.4+175441
 3947 J083512.45+175441.1 yJ084133.5+402035 J084133.40+402042.8 yJ084238.4+405305 J084238.38+405306.6
 3948 nJ084417.3+315845 J084417.92+315845.9 yJ084728.5+360700 J084728.24+360714.6 yJ084905.5+111448
 3949 J084905.51+111447.8 yJ085236.8+262006 J085236.11+262013.4 yJ085415.6+524930 J085415.62+524936.7
 3950 yJ090623.2+300746 J090622.87+300743.9 yJ091745.1+275049 J091745.89+275103.8 yJ091752.0+431614
 3951 J091752.14+431612.7 yJ092014.4+302907 J092013.95+302859.3 yJ092140.5+540118 J092140.24+540121.1
 3952 yJ092213.0+542157 J092213.03+542157.2 yJ092406.9+562703 J092406.47+562656.2 yJ092713.1+105841
 3953 J092713.14+105839.8 yJ093108.6+613447 J093108.63+613447.2 yJ093239.6+052308 J093237.71+052240.7
 3954 nJ093627.8+103610 J093627.87+103609.7 yJ093645.2+561435 J093645.89+561434.2 yJ094006.8+482651
 3955 J094006.92+482649.2 yFIRST AllWISE AgreeJ094009.5+600403 J094011.55+600357.6 nJ094023.7+135123
 3956 J094023.73+135125.2 yJ094324.5+435341 J094324.61+435342.0 yJ094650.8+382015 J094650.44+382010.9
 3957 yJ095011.8+455319 J095011.82+455320.0 yJ095113.5+180211 J095113.82+180204.2 nJ095242.4+222638
 3958 J095242.45+222638.0 yJ095538.7+013546 J095539.20+013546.1 yJ095609.9+363441 J095609.30+363445.4
 3959 yJ095811.8+225056 J095811.90+225055.5 yJ100019.2+263516 J100018.84+263527.5 yJ101315.9+064520
 3960 J101316.51+064519.0 yJ101455.2-004716 J101455.30-004718.3 yJ102153.5+260429 J102153.52+260429.6
 3961 yJ102354.7+390653 J102354.88+390654.0 yJ102620.4+303600 J102620.46+303550.4 yJ102710.4+460254
 3962 J102714.81+460256.4 nJ102955.9+424906 J102955.96+424906.7 yJ103503.9+102404 J103503.92+102403.6
 3963 yJ103839.9+331200 J103839.94+331201.1 yJ104030.5+211624 J104031.09+211620.6 nJ104533.8+430025
 3964 J104535.22+430020.8 yJ104907.5+322903 J104907.91+322906.6 yJ105146.9+552257 J105147.40+552308.4
 3965 yJ105257.5+105418 J105257.53+105421.5 yJ105521.6+372641 J105521.24+372652.4 yJ105758.8+321605
 3966 J105758.84+321605.3 yJ110104.9+151618 J110104.90+151618.2 yJ110353.2+352320 J110353.37+352319.9
 3967 yJ110414.4+481345 J110423.08+481311.0 nJ111057.7+220756 J111057.18+220758.3 yJ111208.5+275207

3968 J111201.79+275053.8 nJ111225.2+233159 J111225.30+233157.9 yJ111726.3+375336 J111726.35+375337.
 3969 yJ111746.1+261151 J111746.18+261150.9 yJ111854.3+424708 J111854.45+424652.8 yJ112124.4+640417
 3970 J112125.02+640408.6 yJ112135.3+352330 J112135.44+352324.9 yJ112550.9+200631 J112558.75+200554.
 3971 yJ112859.7+260923 J112859.86+260911.3 yJ113201.1+442639 J113201.23+442639.4 yJ113302.5+355408
 3972 J113301.80+355415.3 yJ113712.7+263301 J113711.86+263335.1 yJ113756.3+471314 J113756.31+471314.
 3973 yJ113906.6+230602 J113906.68+230602.1 yJ114325.0+600721 J114323.90+600737.1 yJ114759.7+370305
 3974 J114759.22+370311.2 yJ114916.7+083022 J114916.33+083040.5 nJ115010.9+063340 J115010.93+063340
 3975 yJ115308.6+374851 J115316.96+374850.0 yFIRST AllWISE AgreeJ115448.7+472222 J115448.67+472223
 3976 yJ115603.7+584704 J115603.48+584706.1 yJ115605.9+343230 J115605.64+343229.4 yJ115653.0+572338
 3977 J115645.38+572151.7 yJ120138.0+230922 J120137.97+230922.2 yJ120752.8+533808 J120752.85+533807.
 3978 yJ120943.3-021934 J120942.89-021943.0 yJ121045.6+190225 J121045.68+190227.0 yJ121207.6+115412
 3979 J121207.72+115413.8 yJ121211.3+485951 J121211.86+485952.0 yJ121406.7+002634 J121406.73+002635.
 3980 yJ122518.0+350258 J122517.85+350301.9 yJ122525.1+451530 J122524.71+451508.5 yJ122640.9+430508
 3981 J122640.82+430509.2 yJ123429.8+260107 J123434.79+260134.3 nJ123633.1+100928 J123633.12+100928
 3982 yJ124839.3+411522 J124839.42+411522.3 nJ125129.2+551012 J125128.76+551009.3 yJ130005.8+524801
 3983 J130006.14+524803.0 yJ130132.1+511351 J130132.32+511352.5 yJ131104.4+464936 J131104.45+464934
 3984 yJ131452.2+252811 J131446.81+252820.8 nJ132033.8+332639 J132033.59+332639.0 nJ132257.5+191134
 3985 J132257.53+191133.9 yJ132529.3+230734 J132529.35+230733.8 yJ132546.8+052453 J132546.86+052454
 3986 yJ132637.7+112110 J132637.92+112108.8 yJ132831.8+104339 J132831.88+104338.8 yJ132932.3+131839
 3987 J132932.32+131839.6 yJ133022.8+311904 J133022.83+311902.8 yJ133453.3+405653 J133454.13+405650
 3988 yJ133741.1+124302 J133741.13+124303.1 yJ133823.6+103337 J133823.67+103341.9 yJ134651.2+415154
 3989 J134651.06+415156.1 yJ134704.3+110622 J134704.35+110622.7 yJ134752.7+555046 J134752.71+555048
 3990 yJ134831.7+164325 J134831.57+164328.2 yJ134949.8+385539 J134949.93+385542.8 yJ135106.5+074534
 3991 J135106.50+074534.2 yJ135107.7+615502 J135107.75+615502.1 yJ135658.5+134028 J135659.15+134017
 3992 yJ135833.9+180021 J135834.03+180020.4 yJ140630.7+554017 J140629.32+554009.9 yJ140804.2+503019
 3993 J140804.10+503021.1 yJ141226.7+454125 J141226.54+454125.5 yJ141245.0+495213 J141243.84+495206
 3994 yJ141317.4+325306 J141317.50+325306.8 yJ141723.8+543639 J141724.33+543629.5 yJ141938.8+312146
 3995 J141940.16+312138.8 yJ142515.3+175526 J142513.89+175525.7 yFIRST AllWISE AgreeJ142829.5+070833
 3996 J142829.60+070836.3 yJ143411.0+170036 J143411.18+170035.7 yJ143624.0-001057 J143623.89-001100.8
 3997 yJ143742.6+104412 J143742.69+104412.8 yJ143840.8+475355 J143841.08+475356.1 yJ143909.1+430847
 3998 J143909.08+430847.8 yJ144135.8+102246 J144135.91+102245.1 yJ144333.6+275229 J144333.02+275250
 3999 yJ145012.3+471739 J145012.33+471738.7 yJ145103.7+452459 J145102.66+452520.5 nJ145401.6+141009
 4000 J145401.70+141009.6 yJ150158.7+191413 J150158.87+191405.3 yJ150743.9+352720 J150743.62+352724
 4001 yJ151141.6-003209 J151142.01-003213.0 yJ151315.5+403107 J151315.56+403107.7 yJ151518.7+230256
 4002 J151518.67+230257.3 yJ151703.6+105947 J151703.68+105947.6 yJ151736.8+610856 J151736.83+610857
 4003 yJ152121.6+281635 J152120.68+281626.2 yJ152714.8+310425 J152714.88+310424.7 yJ153428.9+272134
 4004 J153429.68+272120.8 yJ154245.3+100919 J154245.71+100917.8 yJ154901.6+103159 J154901.40+103152
 4005 yJ154925.2+395316 J154926.17+395303.7 yJ155206.3-005348 J155206.58-005339.3 yJ155457.3+344637
 4006 J155458.45+344644.7 yJ155743.5+272752 J155743.52+272752.8 yJ160130.0+083848 J160130.07+083850
 4007 yJ160534.8+441220 J160535.55+441221.5 yJ160859.2+400135 J160901.32+400230.7 nJ161545.4+231617
 4008 J161545.14+231617.2 yJ161930.4+085533 J161930.51+085532.6 yJ162228.0+264743 J162228.70+264736
 4009 yJ162750.4+473624 J162750.55+473623.5 yJ162904.2+470852 J162904.34+470853.0 yJ163038.7+214740
 4010 J163037.43+214748.9 nJ163323.6+424051 J163323.61+424051.9 yJ163327.5+242426 J163327.87+242427
 4011 yJ163533.8+454557 J163534.00+454554.3 yJ164211.2+512029 J164211.27+512029.3 yJ165549.1+375923

4012 J165549.01+375923.6 y J165620.0+363402 J165619.89+363403.9 y J165700.5+474820 J165659.58+474809.0
 4013 y J171406.2+292712 J171404.16+292704.0 n J172126.4+374446 J172126.46+374446.6 y J222627.7-005010
 4014 J222627.77-005010.8 y J223636.4-013827 J223636.48-013827.2 y J225619.0+143257 J225621.96+143351.4
 4015 y J232410.1+001315 J232410.15+001314.5 y J234727.9-000919 J234727.65-000912.9 y

4016 5 2-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. This appendix demonstrates that fact, and was originally part of Alger et al. (2021). Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the sum-normalised simple model FDF:

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

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

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

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

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

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

4017 where $\Gamma(A, B)$ is the set of couplings of A and B , i.e. the set of joint probability distributions
 4018 that marginalise to A and B ; and $\inf_{\gamma \in \Gamma(A, B)}$ is the infimum over $\Gamma(A, B)$. This can be
 4019 interpreted as the minimum cost to ‘move’ one probability distribution to the other,
 4020 where the cost of moving one unit of probability mass is the squared distance it is
 4021 moved.

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

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

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

The coupling that minimises the integral in 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:

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

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:

$$\begin{aligned} D_{W_2}(\tilde{F} \parallel \tilde{S})^2 &= \iint_{\phi_{\min}}^{\phi_{\max}} |\phi - \varphi|^2 d\gamma(\phi, \varphi) \\ &= \frac{A_0(\phi_0 - \phi_w)^2 + A_1(\phi_1 - \phi_w)^2}{A_0 + A_1}. \end{aligned}$$

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

$$\phi_w = \frac{A_0\phi_0 + A_1\phi_1}{A_0 + A_1}.$$

Substituting this back in, we find

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

4022 which is the Faraday moment.

4023 5 Euclidean distance in the no-RMSF case

In this appendix, originally from Alger et al. (2021), we calculate the minimised Euclidean distance evaluated on a model FDF (). Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the normalised simple model FDF:

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

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

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

$$\begin{aligned} & D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \\ &= \int_{\phi_{\min}}^{\phi_{\max}} |\tilde{F}(\phi) - \delta(\phi - \phi_e)|^2 d\phi. \end{aligned}$$

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

$$\begin{aligned} & D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \\ &= \frac{1}{(A_0 + A_1)^2} \int_{\phi_{\min}}^{\phi_{\max}} A_1^2 |\delta(\phi - \phi_1) - \delta(\phi - \phi_0)|^2 d\phi \\ &= \frac{2A_1^2}{(A_0 + A_1)^2} \end{aligned}$$

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

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

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

$$\begin{aligned} D_E(F) &= \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \\ &= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \end{aligned}$$

4024 If $\phi_0 = \phi_1$, then the minimised Euclidean distance is 0.

4025 5 Hyperparameters for LR and XGB

4026 This section contains tables of the hyperparameters that we used for our classifiers in
 4027 and was originally an appendix to Alger et al. (2021). and tabulate the hyperparameters
 4028 for XGB and LR respectively for the ‘ATCA’ dataset. and tabulate the hyperparameters
 4029 for XGB and LR respectively for the ‘ASKAP’ dataset.

4030 XGB hyperparameters for the ‘ATCA’ dataset. Parameter Value colsample_bytree
 4031 0.912 gamma 0.532 learning_rate 0.1 max_depth 7 min_child_weight 2 scale_pos_weight
 4032 1 subsample 0.557 n_estimators 135 reg_alpha 0.968 reg_lambda 1.420

4033 LR hyperparameters for the ‘ATCA’ dataset. Parameter Value penalty L1 C 1.668

4034 XGB hyperparameters for the ‘ASKAP’ dataset. Parameter Value colsample_bytree
 4035 0.865 gamma 0.256 learning_rate 0.1 max_depth 6 min_child_weight 1 scale_pos_weight
 4036 1 subsample 0.819 n_estimators 108 reg_alpha 0.049 reg_lambda 0.454

4037 LR hyperparameters for the ‘ASKAP’ dataset. Parameter Valuepenalty L2C 0.464

4038 **5 Predictions on real data**

4039 This appendix, originally part of Alger et al. (2021), contains and . These show the
 4040 predicted probability of being Faraday complex for all real data used in , drawn from
 4041 Livingston et al. (2021) and O’Sullivan et al. (2017).

4042 The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex.
 4043 Livingston-identified components are shown in orange while O’Sullivan-identified components
 4044 are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple
 4045 while more complex FDFs are shown in green, and the numbers overlaid indicate the
 4046 LR estimate. A lower number indicates a lower probability that the corresponding
 4047 source is complex, i.e. lower numbers correspond to simpler spectra.

4048 The 142 observed FDFs ordered by XGB-estimated probability of being Faraday
 4049 complex. Livingston-identified components are shown in orange while O’Sullivan-identified
 4050 components are shown in magenta. Simpler FDFs (as deemed by the classifier) are
 4051 shown in purple while more complex FDFs are shown in green, and the numbers
 4052 overlaid indicate the XGB estimate. A lower number indicates a lower probability
 4053 that the corresponding source is complex, i.e. lower numbers correspond to simpler
 4054 spectra.

4055 **5 Simulating observed FDFs**

This appendix was originally part of Alger et al. (2021) and describes how we simulated FDFs in . 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 $\vec{F} \in \mathbb{R}^d$:

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

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:

$$\underline{\phi_k \in [\phi_{\min}, \phi_{\min} + \delta\phi, \dots, \phi_{\max}]}$$

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

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

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

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 package¹, which is a Python module that implements a discrete version of RM synthesis:

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

REVISION DRAFT

¹

Bibliography

- 4058 Agarwal, D., Aggarwal, K., Burke-Spolaor, S., Lorimer, D. R., & Garver-Daniels, N.
 4059 (2020). FETCH: A deep-learning based classifier for fast transient classification.
 4060 *MNRAS*, 497, 1661–1674. <https://doi.org/10.1093/mnras/staa1856>
- 4061 Aguado, D. S., Ahumada, R., Almeida, A., Anderson, S. F., Andrews, B. H., Anguiano,
 4062 B., Aquino Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., & Aubert,
 4063 M. e. a. (2019). The fifteenth data release of the Sloan Digital Sky Surveys: First
 4064 release of MaNGA-derived quantities, data visualization tools, and stellar li-
 4065 brary. *ApJS*, 240(2). <https://doi.org/10.3847/1538-4365/aaf651>
- 4066 Alger, M. J., Banfield, J. K., Ong, C. S., Rudnick, L., Wong, O. I., Wolf, C., Andernach, H.,
 4067 Norris, R. P., & Shabala, S. S. (2018). Radio Galaxy Zoo: Machine learning for
 4068 radio source host galaxy cross-identification. *MNRAS*, 478, 5547–5563. <https://doi.org/10.1093/mnras/sty1308>
- 4069 4070 Alger, M. J., Livingston, J. D., McClure-Griffiths, N. M., Nabaglo, J. L., Wong, O. I.,
 4071 & Ong, C. S. (2021). Interpretable Faraday complexity classification. *PASA*, 38.
 4072 <https://doi.org/10.1017/pasa.2021.10>
- 4073 4074 Alger, M. J., Wong, O. I., Ong, C. S., McClure-Griffiths, N. M., Andernach, H., Rudnick,
 4075 L., Shabala, S. S., Garon, A. F., Banfield, J. K., Kapińska, A. D., Norris, R. P., &
 4076 Thomson, A. J. M. (in prep.). Radio Galaxy Zoo: Radio luminosity functions of
 extended sources.
- 4077 4078 Alhassan, W., Taylor, A. R., & Vaccari, M. (2018). The FIRST Classifier: Compact and
 4079 extended radio galaxy classification using deep convolutional neural networks.
MNRAS, 480, 2085–2093. <https://doi.org/10.1093/mnras/sty2038>
- 4080 4081 Anderson, C. S., Gaensler, B. M., Feain, I. J., & Franzen, T. M. O. (2015). Broadband
 4082 radio polarimetry and Faraday rotation of 563 extragalactic radio sources. *ApJ*,
 815(1). <https://doi.org/10.1088/0004-637X/815/1/49>
- 4083 4084 Aniyan, A. K., & Thorat, K. (2017). Classifying radio galaxies with the convolutional
 4085 neural network. *ApJS*, 230. <https://doi.org/10.3847/1538-4365/aa7333>
- 4086 Antonucci, R. (1993). Unified models for active galactic nuclei and quasars. *ARA&A*,
 31(1), 473–521. <https://doi.org/10.1146/annurev.aa.31.090193.002353>
- 4087 4088 Arsioli, B., & Dedin, P. (2020). Machine learning applied to multifrequency data in
 4089 astrophysics: Blazar classification. *MNRAS*, 498, 1750–1764. <https://doi.org/10.1093/mnras/staa2449>

- 4090 Balakrishnan, V., Champion, D., & Barr, E. (2020). Pulsar candidate identification us-
4091 ing semi-supervised generative adversarial networks. *arXiv e-prints*, arXiv:
4092 2010.07457v1 [astro-ph.IM].
- 4093 Ball, N. M., Brunner, R. J., Myers, A. D., & Tcheng, D. (2006). Robust machine learn-
4094 ing applied to astronomical data sets. I. Star-galaxy classification of the Sloan
4095 Digital Sky Survey DR3 using decision trees. *ApJ*, 650(1), 497–509. <https://doi.org/10.1086/507440>
- 4097 Ball, N. M., Loveday, J., Fukugita, M., Nakamura, O., Okamura, S., Brinkmann, J., &
4098 Brunner, R. J. (2004). Galaxy types in the Sloan Digital Sky Survey using su-
4099 pervised artificial neural networks. *MNRAS*, 348, 1038–1046. <https://doi.org/10.1111/j.1365-2966.2004.07429.x>
- 4101 Bañados, E., Venemans, B. P., Mazzucchelli, C., Farina, E. P., Walter, F., Wang, F., Decarli,
4102 R., Stern, D., Fan, X., Davies, F. B., Hennawi, J. F., Simcoe, R. A., Turner, M. L.,
4103 Rix, H.-W., Yang, J., Kelson, D. D., Rudie, G. C., & Winters, J. M. (2018). An 800-
4104 million-solar-mass black hole in a significantly neutral Universe at a redshift of
4105 7.5. *Nature*, 553(7689), 473–476. <https://doi.org/10.1038/nature25180>
- 4106 Banerji, M., Lahav, O., Lintott, C. J., Abdalla, F. B., Schawinski, K., Bamford, S. P., An-
4107 dreescu, D., Murray, P., Raddick, M. J., Slosar, A., Szalay, A., Thomas, D., &
4108 Vandenberg, J. (2010). Galaxy Zoo: Reproducing galaxy morphologies via ma-
4109 chine learning. *MNRAS*, 406, 342–353. <https://doi.org/10.1111/j.1365-2966.2010.16713.x>
- 4111 Banfield, J. K., Andernach, H., Kapińska, A. D., Rudnick, L., Hardcastle, M. J., Cotter,
4112 G., Vaughan, S., Jones, T. W., Heywood, I., Wing, J. D., Wong, O. I., Matorny,
4113 T., Terentev, I. A., López-Sánchez, Á. R., Norris, R. P., Seymour, N., Shabala,
4114 S. S., & Willett, K. W. (2016). Radio Galaxy Zoo: Discovery of a poor cluster
4115 through a giant wide-angle tail radio galaxy. *MNRAS*, 460, 2376–2384. <https://doi.org/10.1093/mnras/stw1067>
- 4117 Banfield, J. K., Wong, O. I., Willett, K. W., Norris, R. P., Rudnick, L., Shabala, S. S., Sim-
4118 mons, B. D., Snyder, C., Garon, A., Seymour, N., Middelberg, E., Andernach,
4119 H., Lintott, C. J., Jacob, K., Kapińska, A. D., Mao, M. Y., Masters, K. L., Jarvis,
4120 M. J., Schawinski, K., ... Whyte, L. (2015). Radio Galaxy Zoo: Host galaxies and
4121 radio morphologies derived from visual inspection. *MNRAS*, 453, 2326–2340.
4122 <https://doi.org/10.1093/mnras/stv1688>
- 4123 Bastien, D., Oozeer, N., & Somanah, R. (2017). Classifying bent radio galaxies from a
4124 mixture of point-like/extended images with machine learning. *IOP Conference
4125 Series: Materials Science and Engineering*, 198. [https://doi.org/10.1088/1757-899X/198/1/012013](https://doi.org/10.1088/1757-
4126 899X/198/1/012013)
- 4127 Baum, W. A. (1962). Photoelectric magnitudes and red-shifts. *Proceedings from the In-
4128 ternational Astronomical Union Symposium 15: Problems of Extra-Galactic Research*,
4129 390.
- 4130 Beaklini, P. P. B., Quadros, A. V. C., de Avellar, M. G. B., Dantas, M. L. L., & Cançado,
4131 A. L. F. (2020). AGN dichotomy beyond radio loudness: A Gaussian mixture
4132 model analysis. *MNRAS*, 497, 1463–1474. <https://doi.org/10.1093/mnras/staa2072>

- 4134 Becker, R. H., White, R. L., & Helfand, D. J. (1995). The FIRST survey: Faint images of
 4135 the radio sky at twenty centimeters. *ApJ*, *450*(2), 559. <https://doi.org/10.1086/176166>
- 4137 Bertin, E., & Arnouts, S. (1996). SExtractor: Software for source extraction. *A&AS*,
 4138 *117*(2), 393–404. <https://doi.org/10.1051/aas:1996164>
- 4139 Best, P. N. (2009). Radio source populations: Results from SDSS. *Astronomische
 4140 Nachrichten*, *330*, 184–189. <https://doi.org/10.1002/asna.200811152>
- 4141 Best, P. N., Kauffmann, G., Heckman, T. M., Brinchmann, J., Charlot, S., Ivezić, Ž., &
 4142 White, S. D. M. (2005). The host galaxies of radio-loud active galactic nuclei:
 4143 Mass dependences, gas cooling and active galactic nuclei feedback. *MNRAS*,
 4144 *362*, 25–40. <https://doi.org/10.1111/j.1365-2966.2005.09192.x>
- 4145 Bicknell, G. V., Dopita, M. A., & O'Dea, C. P. O. (1997). Unification of the radio and opti-
 4146 cal properties of gigahertz peak spectrum and compact steep-spectrum radio
 4147 sources. *ApJ*, *485*(1), 112–124. <https://doi.org/10.1086/304400>
- 4148 Bicknell, G. V. (1995). Relativistic Jets and the Fanaroff-Riley classification of radio
 4149 galaxies. *ApJS*, *101*, 29–39. <https://doi.org/10.1086/192232>
- 4150 Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T.,
 4151 Singh, R., Szerlip, P., Horsfall, P., & Goodman, N. D. (2019). Pyro: Deep Uni-
 4152 versal Probabilistic Programming. *Journal of Machine Learning Research*, *20*.
- 4153 Bîrzan, L., McNamara, B. R., Nulsen, P. E. J., Carilli, C. L., & Wise, M. W. (2008). Ra-
 4154 diative efficiency and content of extragalactic radio sources: Toward a univer-
 4155 sal scaling relation between jet power and radio power. *ApJ*, *686*(2), 859–880.
 4156 <https://doi.org/10.1086/591416>
- 4157 Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- 4158 Black, A. R. S., Baum, S. A., Leahy, J. P., Perley, R. A., Riley, J. M., & Scheuer, P. A. G.
 4159 (1992). A study of FRII radio galaxies with z less than 0.15 — I. high-resolution
 4160 maps of eight sources at 3.6 cm. *MNRAS*, *256*, 186–208. <https://doi.org/10.1093/mnras/256.2.186>
- 4162 Blundell, K. M., Rawlings, S., & Willott, C. J. (1999). The nature and evolution of clas-
 4163 sical double radio sources from complete samples. *AJ*, *117*(2), 677–706. <https://doi.org/10.1086/300721>
- 4165 Bonaldi, A., An, T., Bruggen, M., Burkutean, S., Coelho, B., Goodarzi, H., Hartley, P.,
 4166 Sandhu, P. K., Wu, C., Yu, L., Haghghi, M. H. Z., Anton, S., Bagheri, Z., Barbosa,
 4167 D., Barraca, J. P., Bartashevich, D., Bergano, M., Bonato, M., Brand, J., ... Wong,
 4168 O. I. (2020). Square Kilometre Array Science Data Challenge 1: Analysis and
 4169 results. *MNRAS*, *500*(3), 3821–3837. <https://doi.org/10.1093/mnras/staa3023>
- 4170 Bosch-Ramon, V. (2018). The role of AGN jets in the reionization epoch. *A&A*, *617*.
 4171 <https://doi.org/10.1051/0004-6361/201833952>
- 4172 Bowles, M., Scaife, A. M. M., Porter, F., Tang, H., & Bastien, D. J. (2020). Attention-
 4173 gating for improved radio galaxy classification. *MNRAS*, *501*, 4579–4595. <https://doi.org/10.1093/mnras/staa3946>
- 4175 Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. <https://doi.org/10.1023/A:1018054314350>

- 4177 Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- 4179 Brentjens, M. A., & de Bruyn, A. G. (2005). Faraday rotation measure synthesis. *A&A*, 441(3), 1217–1228. <https://doi.org/10.1051/0004-6361:20052990>
- 4181 Brown, S. (2011). *POSSUM report 9: Assess the complexity of an RM synthesis spectrum*.
- 4182 Brown, S., Bergerud, B., Costa, A., Gaensler, B. M., Isbell, J., LaRocca, D., Norris, R.,
4183 Purcell, C., Rudnick, L., & Sun, X. (2018). Classifying complex Faraday spectra
4184 with convolutional neural networks. *MNRAS*, 483, 964–970. <https://doi.org/10.1093/mnras/sty2908>
- 4186 Capetti, A., Brienza, M., Baldi, R. D., Giovannini, G., Morganti, R., Hardcastle, M. J.,
4187 Rottgering, H. J. A., Brunetti, G. F., Best, P. N., & Miley, G. (2020). The LOFAR
4188 view of FR 0 radio galaxies. *A&A*, 642. <https://doi.org/10.1051/0004-6361/202038671>
- 4190 Carballo, R., Cofiño, A. S., & González-Serrano, J. I. (2004). Selection of quasar can-
4191 didates from combined radio and optical surveys using neural networks. *MNRAS*, 353, 211–220. <https://doi.org/10.1111/j.1365-2966.2004.08056.x>
- 4193 Cattaneo, A., Faber, S. M., Binney, J., Dekel, A., Kormendy, J., Mushotzky, R., Babul,
4194 A., Best, P. N., Brüggen, M., Fabian, A. C., Frenk, C. S., Khalatyan, A., Netzer,
4195 H., Mahdavi, A., Silk, J., Steinmetz, M., & Wisotzki, L. (2009). The role of black
4196 holes in galaxy formation and evolution. *Nature*, 460(7252), 213–219. <https://doi.org/10.1038/nature08135>
- 4198 Cavagnolo, K. W., McNamara, B. R., Nulsen, P. E. J., Carilli, C. L., Jones, C., & Bîrzan, L.
4199 (2010). A relationship between AGN jet power and radio power. *ApJ*, 720(2),
4200 1066–1072. <https://doi.org/10.1088/0004-637X/720/2/1066>
- 4201 Cheung, C. C. (2007). FIRST “winged” and X-shaped radio source candidates. *AJ*,
4202 133(5), 2097–2121. <https://doi.org/10.1086/513095>
- 4203 Chollet, F. et al. (2015). Keras. GitHub. <https://github.com/fchollet/keras>
- 4204 Collier, J. D., Banfield, J. K., Norris, R. P., Schnitzeler, D. H. F. M., Kimball, A. E., Fil-
4205 ipović, M. D., Jarrett, T. H., Lonsdale, C. J., & Tothill, N. F. H. (2014). Infrared-
4206 faint radio sources: A new population of high-redshift radio galaxies. *MNRAS*,
4207 439, 545–565. <https://doi.org/10.1093/mnras/stt2485>
- 4208 Condon, J. J. (1991). Radio luminosity functions (N. Duric & P. C. Crane, Eds.). In N.
4209 Duric & P. C. Crane (Eds.), *The interpretation of modern synthesis observations of*
4210 *spiral galaxies*.
- 4211 Condon, J. J. (1992). Radio emission from normal galaxies. *ARA&A*, 30, 575–611. <https://doi.org/10.1146/annurev.aa.30.090192.003043>
- 4213 Condon, J. J., Cotton, W. D., & Broderick, J. J. (2002). Radio sources and star formation
4214 in the local universe. *AJ*, 124(2), 675–689. <https://doi.org/10.1086/341650>
- 4215 Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., &
4216 Broderick, J. J. (1998). The NRAO VLA Sky Survey. *AJ*, 115(5), 1693–1716. <https://doi.org/10.1086/300337>
- 4218 Condon, J. J., Matthews, A. M., & Broderick, J. J. (2019). Radio sources in the nearby
4219 universe. *ApJ*, 872(2). <https://doi.org/10.3847/1538-4357/ab0301>

- 4220 Condon, J. J., & Ransom, S. M. (2016). *Essential radio astronomy*. Princeton University
4221 Press.
- 4222 Connor, L., & van Leeuwen, J. (2018). Applying deep learning to fast radio burst clas-
4223 sification. AJ, 156(6). <https://doi.org/10.3847/1538-3881/aae649>
- 4224 Contigiani, O., de Gasperin, F., Miley, G. K., Rudnick, L., Andernach, H., Banfield, J. K.,
4225 Kapińska, A. D., Shabala, S. S., & Wong, O. I. (2017). Radio Galaxy Zoo: Cos-
4226 mological alignment of radio sources. MNRAS, 472, 636–646. <https://doi.org/10.1093/mnras/stx1977>
- 4227 Cutri, R., Wright, E., Conrow, T., Fowler, J., Eisenhardt, P., Grillmair, C., Kirkpatrick, J.,
4228 Masci, F., McCallon, H., Wheelock, S., et al. (2013). Explanatory supplement to
4229 the allwise data release products. *Explanatory Supplement to the AllWISE Data*
4230 *Release Products*.
- 4231 Dabhade, P., Mahato, M., Bagchi, J., Saikia, D. J., Combes, F., Sankhyayan, S., Röttgering,
4232 H. J. A., Ho, L. C., Gaikwad, M., Raychaudhury, S., Vaidya, B., & Guiderdoni,
4233 B. (2020). Search and analysis of giant radio galaxies with associated nuclei
4234 (SAGAN). I. new sample and multi-wavelength studies. A&A, 642. <https://doi.org/10.1051/0004-6361/202038344>
- 4235 Dabhade, P., Gaikwad, M., Bagchi, J., Pandey-Pommier, M., Sankhyayan, S., & Ray-
4236 chaudhury, S. (2017). Discovery of giant radio galaxies from NVSS: Radio and
4237 infrared properties. MNRAS, 469, 2886–2906. <https://doi.org/10.1093/mnras/stx860>
- 4238 Dawid, A. P., & Skene, A. M. (1979). Maximum likelihood estimation of observer error-
4239 rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C (Ap-
4240 plied Statistics)*, 28(1), 20–28.
- 4241 Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*
4242 (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108679930>
- 4243 Diamond, P. (2017). SKA community briefing. Retrieved February 1, 2021, from <https://www.skatelescope.org/ska-community-briefing-18jan2017/>
- 4244 Dieleman, S., Willett, K. W., & Dambre, J. (2015). Rotation-invariant convolutional neu-
4245 ral networks for galaxy morphology prediction. MNRAS, 450, 1441–1459. <https://doi.org/10.1093/mnras/stv632>
- 4246 Fan, D., Budavári, T., Norris, R. P., & Basu, A. (2020). Optimal probabilistic catalogue
4247 matching for radio sources. MNRAS, 498, 565–573. <https://doi.org/10.1093/mnras/staa2447>
- 4248 Fan, D., Budavári, T., Norris, R. P., & Hopkins, A. M. (2015). Matching radio catalogues
4249 with realistic geometry: Application to SWIRE and ATLAS. MNRAS, 451(2),
4250 1299–1305. <https://doi.org/10.1093/mnras/stv994>
- 4251 Fanaroff, B. L., & Riley, J. M. (1974). The morphology of extragalactic radio sources
4252 of high and low luminosity. MNRAS, 167, 31P–36P. <https://doi.org/10.1093/mnras/167.1.31P>
- 4253 Farnes, J. S., Gaensler, B. M., & Carretti, E. (2014). A broadband polarization catalog
4254 of extragalactic radio sources. ApJS, 212(1). <https://doi.org/10.1088/0067-0049/212/1/15>

- 4263 Flamary, R., & Courty, N. (2017). POT: Python optimal transport library. <https://github.com/rflamary/POT>
- 4264
- 4265 Franzen, T. M. O., Banfield, J. K., Hales, C. A., Hopkins, A., Norris, R. P., Seymour, N.,
4266 Chow, K. E., Herzog, A., Huynh, M. T., Lenc, E., Mao, M. Y., & Middelberg, E.
4267 (2015). ATLAS - I. third release of 1.4 GHz mosaics and component catalogues.
4268 *MNRAS*, 453, 4020–4036. <https://doi.org/10.1093/mnras/stv1866>
- 4269 Galvin, T. J., Huynh, M., Norris, R. P., Wang, X. R., Hopkins, E., Wong, O. I., Sha-
4270 bala, S., Rudnick, L., Alger, M. J., & Polsterer, K. L. (2019). Radio Galaxy Zoo:
4271 Knowledge transfer using rotationally invariant self-organizing maps. *PASP*,
4272 131(1004). <https://doi.org/10.1088/1538-3873/ab150b>
- 4273 Galvin, T. J., Huynh, M. T., Norris, R. P., Wang, X. R., Hopkins, E., Polsterer, K., Ralph,
4274 N. O., O'Brien, A. N., & Heald, G. H. (2020). Cataloguing the radio-sky with
4275 unsupervised machine learning: A new approach for the SKA era. *MNRAS*,
4276 497, 2730–2758. <https://doi.org/10.1093/mnras/staa1890>
- 4277 Garofalo, D., & Singh, C. B. (2019). FR0 radio galaxies and their place in the radio
4278 morphology classification. *ApJ*, 871(2). <https://doi.org/10.3847/1538-4357/aaf056>
- 4279
- 4280 Garon, A. F., Rudnick, L., Wong, O. I., Jones, T. W., Kim, J.-A., Andernach, H., Sha-
4281 bala, S. S., Kapińska, A. D., Norris, R. P., de Gasperin, F., Tate, J., & Tang, H.
4282 (2019). Radio Galaxy Zoo: The distortion of radio galaxies by galaxy clusters.
4283 *AJ*, 157(3). <https://doi.org/10.3847/1538-3881/aaff62>
- 4284 Gebhard, T. D., Kilbertus, N., Harry, I., & Schölkopf, B. (2019). Convolutional neural
4285 networks: A magic bullet for gravitational-wave detection? *Phys. Rev. D*, 100.
4286 <https://doi.org/10.1103/PhysRevD.100.063015>
- 4287 Gendre, M. A., & Wall, J. V. (2008). The Combined NVSS-FIRST Galaxies (CoNFIG)
4288 sample — I. sample definition, classification and evolution. *MNRAS*, 390, 819–
4289 828. <https://doi.org/10.1111/j.1365-2966.2008.13792.x>
- 4290 Gilyazev, R. A., & Turdakov, D. Y. (2018). Active learning and crowdsourcing: A survey
4291 of optimization methods for data labeling. *Programming and Computer Software*,
4292 44(6), 476–491. <https://doi.org/10.1134/S0361768818060142>
- 4293 Godfrey, L. E. H., & Shabala, S. S. (2016). Mutual distance dependence drives the ob-
4294 served jet-power-radio-luminosity scaling relations in radio galaxies. *MNRAS*,
4295 456, 1172–1184. <https://doi.org/10.1093/mnras/stv2712>
- 4296 Goldstein, J., S. J., & Reed, J. A. (1984). Double faraday rotation toward 3C 27. *ApJ*, 283,
4297 540–545. <https://doi.org/10.1086/162337>
- 4298 Gopal-Krishna, & Wiita, P. J. (2000). Extragalactic radio sources with hybrid morphol-
4299 ogy: Implications for the Fanaroff-Riley dichotomy. *A&A*, 363, 507–516.
- 4300 Grant, J. K., Taylor, A. R., Stil, J. M., Landecker, T. L., Kothes, R., Ransom, R. R., &
4301 Scott, D. (2010). The DRAO planck deep fields: The polarization properties of
4302 radio galaxies at 1.4 GHz. *ApJ*, 714, 1689–1701. [https://doi.org/10.1088/0004-637X/714/2/1689](https://doi.org/10.1088/0004-
4303 637X/714/2/1689)
- 4304 Grant, J. K. (2011). *Polarised radio sources: A study of luminosity, redshift and flux density*
4305 (Doctoral dissertation). University of Calgary.

- 4306 Groves, B., & Kewley, L. (2007). Distinguishing active galactic nuclei and star forma-
 4307 tion. *Pathways Through an Eclectic Universe*.
- 4308 Guo, P., Duan, F., Wang, P., Yao, Y., Yin, Q., Xin, X., Li, D., Qian, L., Wang, S., Pan, Z.,
 4309 & Zhang, L. (2019). Pulsar candidate classification using generative adversary
 4310 networks. MNRAS, 490(4), 5424–5439. <https://doi.org/10.1093/mnras/stz2975>
- 4311 Hardcastle, M. J., & Croston, J. H. (2020). Radio galaxies and feedback from agn jets.
 4312 *New Astronomy Reviews*, 88. <https://doi.org/10.1016/j.newar.2020.101539>
- 4313 Hardcastle, M. J., & Krause, M. G. H. (2013). Numerical modelling of the lobes of radio
 4314 galaxies in cluster environments. MNRAS, 430, 174–196. <https://doi.org/10.1093/mnras/sts564>
- 4315 Hardcastle, M. J., & Krause, M. G. H. (2014). Numerical modelling of the lobes of radio
 4316 galaxies in cluster environments — II. magnetic field configuration and observ-
 4317 ability. MNRAS, 443, 1482–1499. <https://doi.org/10.1093/mnras/stu1229>
- 4318 Hardcastle, M. J., Williams, W. L., Best, P. N., Croston, J. H., Duncan, K. J., Röttgering,
 4319 H. J. A., Sabater, J., Shimwell, T. W., Tasse, C., Callingham, J. R., & et al. (2019).
 4320 Radio-loud AGN in the first LoTSS data release: The lifetimes and environmen-
 4321 tal impact of jet-driven sources. A&A, 622. <https://doi.org/10.1051/0004-6361/201833893>
- 4322 He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recogni-
 4323 tion, In *Proceedings of the ieee conference on computer vision and pattern recognition*.
- 4324 Heald, G. (2008). The Faraday rotation measure synthesis technique. *Proceedings of the*
 4325 *International Astronomical Union*, 4(S259), 591–602. <https://doi.org/10.1017/S1743921309031421>
- 4326 Heckman, T. M., & Best, P. N. (2014). The coevolution of galaxies and supermassive
 4327 black holes: Insights from surveys of the contemporary universe. ARA&A, 52(1),
 4328 589–660. <https://doi.org/10.1146/annurev-astro-081913-035722>
- 4329 Helfand, D. J., White, R. L., & Becker, R. H. (2015). The last of FIRST: The final catalog
 4330 and source identifications. ApJ, 801(1). <https://doi.org/10.1088/0004-637x/801/1/26>
- 4331 Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., Nolta,
 4332 M. R., Halpern, M., Hill, R. S., Odegard, N., Page, L., Smith, K. M., Weiland, J. L.,
 4333 Gold, B., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., ... Wright,
 4334 E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) ob-
 4335 servations: Cosmological parameter results. ApJS, 208. <https://doi.org/10.1088/0067-0049/208/2/19>
- 4336 Hložek, R., Ponder, K. A., Malz, A. I., Dai, M., Narayan, G., Ishida, E. E. O., Allam, J., T.,
 4337 Bahmanyar, A., Biswas, R., Galbany, L., & et al. (2020). Results of the photometric
 4338 LSST astronomical time-series classification challenge (PLAsTiCC). *arXiv e-prints*, arXiv:2012.12392v1 [astro-ph.IM].
- 4339 Holzinger, A. (2016). Interactive machine learning for health informatics: When do we
 4340 need the human-in-the-loop? *Brain Informatics*, 3(2), 119–131.
- 4341 Hurley-Walker, N., Callingham, J. R., Hancock, P. J., Franzen, T. M. O., Hindson, L.,
 4342 Kapińska, A. D., Morgan, J., Offringa, A. R., Wayth, R. B., Wu, C., Zheng, Q.,
 4343 Murphy, T., Bell, M. E., Dwarakanath, K. S., For, B., Gaensler, B. M., Johnston-
 4344

- 4350 Hollitt, M., Lenc, E., Procopio, P., ... Williams, C. L. (2017). GaLactic and Ex-
 4351 tragalactic All-sky Murchison Widefield Array (GLEAM) survey - I. a low-
 4352 frequency extragalactic catalogue. *MNRAS*, *464*, 1146–1167. <https://doi.org/10.1093/mnras/stw2337>
- 4354 Husemann, B., & Harrison, C. M. (2018). Reality and myths of AGN feedback. *Nature Astronomy*, *2*(3), 196–197. <https://doi.org/10.1038/s41550-018-0407-2>
- 4355 Jarrett, T. H., Cluver, M. E., Magoulas, C., Bilicki, M., Alpaslan, M., Bland-Hawthorn,
 4356 J., Brough, S., Brown, M. J. I., Croom, S., Driver, S., Holwerda, B. W., Hopkins,
 4357 A. M., Loveday, J., Norberg, P., Peacock, J. A., Popescu, C. C., Sadler, E. M., Tay-
 4358 lor, E. N., Tuffs, R. J., & Wang, L. (2017). Galaxy and mass assembly (GAMA):
 4359 Exploring the WISE web in G12. *ApJ*, *836*(2).
- 4360 Jarrett, T. H., Cohen, M., Masci, F., Wright, E., Stern, D., Benford, D., Blain, A., Carey, S.,
 4361 Cutri, R. M., Eisenhardt, P., Lonsdale, C., Mainzer, A., Marsh, K., Padgett, D.,
 4362 Petty, S., Ressler, M., Skrutskie, M., Stanford, S., Surace, J., ... Yan, D. L. (2011).
 4363 The Spitzer-WISE survey of the ecliptic poles. *ApJ*, *735*(2).
- 4364 Johnston, S., Bailes, M., Bartel, N., Baugh, C., Bietenholz, M., Blake, C., Braun, R., Brown,
 4365 J., Chatterjee, S., Darling, J., Deller, A., Dodson, R., Edwards, P. G., Ekers, R.,
 4366 Ellingsen, S., Feain, I., Gaensler, B. M., Havercorn, M., Hobbs, G., ... Wolleben,
 4367 M. (2007). Science with the Australian Square Kilometre Array Pathfinder. *PASA*,
 4368 *24*, 174–188. <https://doi.org/10.1071/AS07033>
- 4370 Kapinska, A. D. (2020). EMU: The Evolutionary Map of the Universe radio sky survey.
 4371 *American Astronomical Society Meeting Abstracts*, *236*.
- 4372 Kapińska, A. D., Terentev, I., Wong, O. I., Shabala, S. S., Andernach, H., Rudnick, L.,
 4373 Storer, L., Banfield, J. K., Willett, K. W., Gasperin, F. d., Lintott, C. J., López-
 4374 Sánchez, Á. R., Middelberg, E., Norris, R. P., Schawinski, K., Seymour, N., &
 4375 Simmons, B. (2017). Radio Galaxy Zoo: A search for hybrid morphology radio
 4376 galaxies. *AJ*, *154*(6), 253. <https://doi.org/10.3847/1538-3881/aa90b7>
- 4377 Kimball, A. E., & Ivezić, Ž. (2008). A unified catalog of radio objects detected by NVSS,
 4378 FIRST, WENSS, GB6, and SDSS. *AJ*, *136*, 684–712. [https://doi.org/10.1088/0004-6256/136/2/684](https://doi.org/10.1088/0004-

 4379 6256/136/2/684)
- 4380 Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *3rd Interna-
 4381 tional Conference for Learning Representations*, arXiv:1412.6980v9 [cs.LG].
- 4382 Kochanek, C. S., Pahre, M. A., Falco, E. E., Huchra, J. P., Mader, J., Jarrett, T. H., Chester,
 4383 T., Cutri, R., & Schneider, S. E. (2001). The K-band galaxy luminosity function.
 4384 *ApJ*, *560*, 566–579. <https://doi.org/10.1086/322488>
- 4385 Koribalski, B. S., Staveley-Smith, L., Westmeier, T., Serra, P., Spekkens, K., Wong, O. I.,
 4386 Lee-Waddell, K., Lagos, C. D. P., Obreschkow, D., Ryan-Weber, E. V., Zwaan,
 4387 M., Kilborn, V., Bekiaris, G., Bekki, K., Bigiel, F., Boselli, A., Bosma, A., Catinella,
 4388 B., Chauhan, G., ... Wolf, C. (2020). WALLABY - an SKA pathfinder HI survey.
 4389 *Ap&SS*, *365*. <https://doi.org/10.1007/s10509-020-03831-4>
- 4390 Kormendy, J., & Ho, L. C. (2013). Coevolution (or not) of supermassive black holes
 4391 and host galaxies. *ARA&A*, *51*(1), 511–653.

- 4392 Kormendy, J., & Richstone, D. (1995). Inward bound—the search for supermassive
 4393 black holes in galactic nuclei. *ARA&A*, 33. <https://doi.org/10.1146/annurev.aa.33.090195.003053>
- 4395 Kozieł-Wierzbowska, D., Vale Asari, N., Stasińska, G., Herpich, F. R., Sikora, M., Ży-
 4396 wucka, N., & Goyal, A. (2020). Identifying radio active galactic nuclei among
 4397 radio-emitting galaxies. *arXiv e-prints*, arXiv:2012.00367v1 [astro-ph.GA].
- 4398 Lacy, M., Baum, S. A., Chandler, C. J., Chatterjee, S., Clarke, T. E., Deustua, S., English,
 4399 J., Farnes, J., Gaensler, B. M., Gugliucci, N., Hallinan, G., Kent, B. R., Kimball, A.,
 4400 Law, C. J., Lazio, T. J. W., Marvil, J., Mao, S. A., Medlin, D., Mooley, K., ... Yoon,
 4401 I. (2020). The Karl G. Jansky Very Large Array Sky Survey (VLASS). science
 4402 case and survey design. *PASP*, 132(1009). <https://doi.org/10.1088/1538-3873/ab63eb>
- 4404 Laing, R. A., & Bridle, A. H. (1987). Rotation measure variation across M84. *MNRAS*,
 4405 228, 557–571. <https://doi.org/10.1093/mnras/228.3.557>
- 4406 Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predic-
 4407 tive uncertainty estimation using deep ensembles, In *Advances in neural infor-
 4408 mation processing systems*.
- 4409 Law, C. J., Gaensler, B. M., Bower, G. C., Backer, D. C., Bauermeister, A., Croft, S.,
 4410 Forster, R., Gutierrez-Kraybill, C., Harvey-Smith, L., Heiles, C., Hull, C., Keat-
 4411 ing, G., MacMahon, D., Whysong, D., Williams, P. K. G., & Wright, M. (2011).
 4412 Spectropolarimetry with the Allen Telescope Array: Faraday rotation toward
 4413 bright polarized radio galaxies. *ApJ*, 728(1). <https://doi.org/10.1088/0004-637X/728/1/57>
- 4415 Leahy, J. P., Bridle, A. H., & Strom, R. G. (n.d.). An Atlas of DRAGNs. Retrieved June
 4416 13, 2020, from <http://www.jb.man.ac.uk/atlas/index.html>
- 4417 Leahy, J. P., & Perley, R. A. (1991). VLA images of 23 extragalactic radio sources. *AJ*,
 4418 102, 537–561. <https://doi.org/10.1086/115892>
- 4419 Leahy, J. P., Pooley, G. G., & Riley, J. M. (1986). The polarization of classical double-
 4420 radio sources. *MNRAS*, 222, 753–785. <https://doi.org/10.1093/mnras/222.4.753>
- 4421 LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel,
 4422 L. D. (1989). Backpropagation applied to handwritten zip code recognition.
Neural Computation, 1(4), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
- 4424 LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied
 4425 to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- 4426 Ledlow, M. J., & Owen, F. N. (1996). 20 cm VLA survey of Abell clusters of galaxies.
 4427 VI. radio/optical luminosity functions. *AJ*, 112. <https://doi.org/10.1086/117985>
- 4428 Lin, H., Li, X., & Zeng, Q. (2020). Pulsar candidate sifting using multi-input convolu-
 4429 tion neural networks. *ApJ*, 899. <https://doi.org/10.3847/1538-4357/aba838>
- 4430 Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick,
 4431 M. J., Nichol, R. C., Szalay, A., Andreeescu, D., Murray, P., & Vandenberg, J.
 4432 (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies
 4433 from the Sloan Digital Sky Survey. *MNRAS*, 389, 1179–1189. <https://doi.org/10.1111/j.1365-2966.2008.13689.x>

- 4435 Livingston, J. D., McClure-Griffiths, N. M., Gaensler, B. M., Seta, A., & Alger, M. J.
4436 (2021). Heightened Faraday complexity in the inner 1 kpc of the Galactic cen-
4437 tre. MNRAS. <https://doi.org/10.1093/mnras/stab253>
- 4438 Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., Surace, J., Shupe, D., Xu, C., Oliver,
4439 S., Padgett, D., Fang, F., Conrow, T., Franceschini, A., Gautier, N., Griffin, M.,
4440 Hacking, P., Masci, F., Morrison, G., O'Linger, J., Owen, F., Pérez-Fournon, I., ...
4441 Serjeant, S. (2003). SWIRE: The SIRTF Wide-Area Infrared Extragalactic Survey.
4442 PASP, 115, 897–927. <https://doi.org/10.1086/376850>
- 4443 Lovelace, R. V. E. (1976). Dynamo model of double radio sources. Nature, 262(5570),
4444 649–652. <https://doi.org/10.1038/262649a0>
- 4445 Lukic, V., Brüggen, M., Banfield, J. K., Wong, O. I., Rudnick, L., Norris, R. P., & Simmons,
4446 B. (2018). Radio Galaxy Zoo: Compact and extended radio source classification
4447 with deep learning. MNRAS, 476, 246–260. <https://doi.org/10.1093/mnras/sty163>
- 4448 Lukic, V., Brüggen, M., Mingo, B., Croston, J. H., Kasieczka, G., & Best, P. N. (2019).
4449 Morphological classification of radio galaxies: Capsule networks versus convo-
4450 lutional neural networks. MNRAS, 487, 1729–1744. <https://doi.org/10.1093/mnras/stz1289>
- 4451 Ma, Y. K., Mao, S. A., Stil, J., Basu, A., West, J., Heiles, C., Hill, A. S., & Betti, S. K.
4452 (2019a). A broad-band spectro-polarimetric view of the NVSS rotation mea-
4453 sure catalogue - I. breaking the $n\pi$ -ambiguity. MNRAS, 487(3), 3432–3453. <https://doi.org/10.1093/mnras/stz1325>
- 4454 Ma, Z., Xu, H., Zhu, J., Hu, D., Li, W., Shan, C., Zhu, Z., Gu, L., Li, J., Liu, C., & Wu,
4455 X. (2019b). A machine learning based morphological classification of 14,245
4456 radio AGNs selected from the Best-Heckman sample. ApJS, 240(2), 34. <https://doi.org/10.3847/1538-4365/aaf9a2>
- 4457 Ma, Z., Zhu, J., Li, W., & Xu, H. (2018). Radio galaxy morphology generation using
4458 DNN autoencoder and Gaussian mixture models. arXiv e-prints, 1806, arXiv:
4459 1806.00398v1 [cs.CV].
- 4460 Machado Poletti Valle, L. F., Avestruz, C., Barnes, D. J., Farahi, A., Lau, E. T., & Nagai,
4461 D. (2020). SHAPing the gas: Understanding gas shapes in dark matter haloes
4462 with interpretable machine learning. arXiv e-prints, 2011, arXiv:2011.12987v1
4463 [astro-ph.CO].
- 4464 Machalski, J., Jamrozy, M., Stawarz, L., & Koziel-Wierzbowska, D. (2011). Under-
4465 standing giant radio galaxy J1420-0545: Large-scale morphology, environment, and
4466 energetics. ApJ, 740(2). <https://doi.org/10.1088/0004-637X/740/2/58>
- 4467 Macquart, J.-P., Bailes, M., Bhat, N. D. R., Bower, G. C., Bunton, J. D., Chatterjee, S.,
4468 Colegate, T., Cordes, J. M., D'Addario, L., Deller, A., & et al. (2010). The Com-
4469 mensal Real-Time ASKAP Fast-Transients (CRAFT) survey. PASA, 27(3), 272–
4470 282. <https://doi.org/10.1071/AS09082>
- 4471 Marconi, A., & Hunt, L. K. (2003). The relation between black hole mass, bulge mass,
4472 and near-infrared luminosity. ApJ, 589(1), L21–L24. <https://doi.org/10.1086/375804>

- 4478 Marshall, P. J., Lintott, C. J., & Fletcher, L. N. (2015). Ideas for citizen science in astronomy. *ARA&A*, 53, 247–278. <https://doi.org/10.1146/annurev-astro-081913-035959>
- 4481 Mauch, T., & Sadler, E. M. (2007). Radio sources in the 6dFGS: Local luminosity functions at 1.4GHz for star-forming galaxies and radio-loud AGN. *MNRAS*, 375, 931–950. <https://doi.org/10.1111/j.1365-2966.2006.11353.x>
- 4484 McConnell, D., Hale, C. L., Lenc, E., Banfield, J. K., Heald, G., Hotan, A. W., Leung, J. K., Moss, V. A., Murphy, T., O'Brien, A., Pritchard, J., Raja, W., Sadler, E. M., Stewart, A., Thomson, A. J. M., Whiting, M., Allison, J. R., Amy, S. W., Anderson, C., ... Westmeier, T. (2020). The Rapid ASKAP Continuum Survey I: Design and first results. *PASA*, 37. <https://doi.org/10.1017/pasa.2020.41>
- 4489 Menon, A. K., Van Rooyen, B., Ong, C. S., & Williamson, R. C. (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*.
- 4493 Middelberg, E., Norris, R. P., Cornwell, T. J., Voronkov, M. A., Siana, B. D., Boyle, B. J., Ciliegi, P., Jackson, C. A., Huynh, M. T., Berta, S., Ruble, S., Lonsdale, C. J., Ivison, R. J., & Smail, I. (2008). Deep Australia Telescope large area survey radio observations of the European large area ISO survey S1/Spitzer wide-area infrared extragalactic field. *AJ*, 135, 1276–1290. <https://doi.org/10.1088/0004-6256/135/4/1276>
- 4499 Miller, D. (2017). Stargazing Live viewers find four-planet solar system. Retrieved January 16, 2021, from <https://www.abc.net.au/news/2017-04-06/stargazing-live-four-planets-discovered-in-new-solar-system/8423142>
- 4502 Mingo, B., Croston, J. H., Hardcastle, M. J., Best, P. N., Duncan, K. J., Morganti, R., Röttgering, H. J. A., Sabater, J., Shimwell, T. W., Williams, W. L., Brienza, M., Gurkan, G., Mahatma, V. H., Morabito, L. K., Prandoni, I., Bondi, M., Ineson, J., & Mooney, S. (2019). Revisiting the Fanaroff–Riley dichotomy and radio-galaxy morphology with the LOFAR Two-Metre Sky Survey (LoTSS). *MNRAS*, 488(2), 2701–2721. <https://doi.org/10.1093/mnras/stz1901>
- 4508 Miyashita, Y., Ideguchi, S., Nakagawa, S., Akahori, T., & Takahashi, K. (2019). Performance test of QU-fitting in cosmic magnetism study. *MNRAS*, 482(2), 2739–2749. <https://doi.org/10.1093/mnras/sty2862>
- 4511 Mohr, P. J., Newell, D. B., Barry N. Taylor, & Tiesinga, E. (2019). The NIST reference on constants, units, and uncertainty. NIST Physical Measurement Laboratory. <https://physics.nist.gov/cuu/Reference/about.html>
- 4514 Morganti, R. (2017). The many routes to AGN feedback. *Frontiers in Astronomy and Space Sciences*, 4. <https://doi.org/10.3389/fspas.2017.00042>
- 4516 Mostert, R. I. J., Duncan, K. J., Röttgering, H. J. A., Polsterer, K. L., Best, P. N., Brienza, M., Brüggen, M., Hardcastle, M. J., Jurlin, N., Mingo, B., Morganti, R., Shimwell, T., Smith, D., & Williams, W. L. (2021). Unveiling the rarest morphologies of the LOFAR Two-metre Sky Survey radio source population with self-organised maps. *A&A*, 645. <https://doi.org/10.1051/0004-6361/202038500>

- 4521 Mukherjee, D., Bicknell, G. V., Sutherland, R., & Wagner, A. (2016). Relativistic jet
4522 feedback in high-redshift galaxies - I. dynamics. MNRAS, 461, 967–983. <https://doi.org/10.1093/mnras/stw1368>
- 4524 Mukherjee, D., Bicknell, G. V., Wagner, A. Y., Sutherland, R. S., & Silk, J. (2018). Rela-
4525 tivistic jet feedback - III. feedback on gas discs. MNRAS, 479, 5544–5566. <https://doi.org/10.1093/mnras/sty1776>
- 4527 Murphy, T., Chatterjee, S., Kaplan, D. L., Banyer, J., Bell, M. E., Bignall, H. E., Bower,
4528 G. C., Cameron, R. A., Coward, D. M., Cordes, J. M., Croft, S., Curran, J. R.,
4529 Djorgovski, S. G., Farrell, S. A., Frail, D. A., Gaensler, B. M., Galloway, D. K.,
4530 Gendre, B., Green, A. J., ... Williams, P. K. G. (2013). VAST: An ASKAP survey
4531 for variables and slow transients. PASA, 30. <https://doi.org/10.1017/pasa.2012.006>
- 4533 Norris, R. P. (2017a). Discovering the unexpected in astronomical survey data. PASA, 34. <https://doi.org/10.1017/pasa.2016.63>
- 4535 Norris, R. P. (2017b). Extragalactic radio continuum surveys and the transformation of
4536 radio astronomy. Nature Astronomy, 1, 671–678. <https://doi.org/10.1038/s41550-017-0233-y>
- 4538 Norris, R. P., Afonso, J., Appleton, P. N., Boyle, B. J., Ciliegi, P., Croom, S. M., Huynh,
4539 M. T., Jackson, C. A., Koekemoer, A. M., Lonsdale, C. J., Middelberg, E., Mobasher,
4540 B., Oliver, S. J., Polletta, M., Siana, B. D., Smail, I., & Voronkov, M. A. (2006).
4541 Deep ATLAS radio observations of the Chandra Deep Field-South/Spitzer wide-
4542 area infrared extragalactic field. AJ, 132(6), 2409–2423. <https://doi.org/10.1086/508275>
- 4544 Norris, R. P., Hopkins, A. M., Afonso, J., Brown, S., Condon, J. J., Dunne, L., Feain, I.,
4545 Hollow, R., Jarvis, M., Johnston-Hollitt, M., Lenc, E., Middelberg, E., Padovani,
4546 P., Prandoni, I., Rudnick, L., Seymour, N., Umana, G., Andernach, H., Alexan-
4547 der, D. M., ... Zhao, G.-B. (2011). EMU: Evolutionary Map of the Universe.
4548 PASA, 28, 215–248. <https://doi.org/10.1071/AS11021>
- 4549 O'Sullivan, S. P., Brown, S., Robishaw, T., Schnitzeler, D. H. F. M., McClure-Griffiths,
4550 N. M., Feain, I. J., Taylor, A. R., Gaensler, B. M., Land ecker, T. L., Harvey-Smith,
4551 L., & Carretti, E. (2012). Complex Faraday depth structure of active galactic nu-
4552 clei as revealed by broad-band radio polarimetry. MNRAS, 421(4), 3300–3315.
4553 <https://doi.org/10.1111/j.1365-2966.2012.20554.x>
- 4554 O'Sullivan, S. P., Gaensler, B. M., Lara-López, M. A., Velzen, S. v., Banfield, J. K., &
4555 Farnes, J. S. (2015). The magnetic field and polarization properties of radio
4556 galaxies in different accretion states. ApJ, 806(1). <https://doi.org/10.1088/0004-637X/806/1/83>
- 4558 O'Sullivan, S. P., Purcell, C. R., Anderson, C. S., Farnes, J. S., Sun, X. H., & Gaensler,
4559 B. M. (2017). Broad-band, radio spectro-polarimetric study of 100 radiative-
4560 mode and jet-mode AGN. MNRAS, 469. <https://doi.org/10.1093/mnras/stx1133>
- 4561 Owens, E. A., Griffiths, R. E., & Ratnatunga, K. U. (1996). Using oblique decision trees
4562 for the morphological classification of galaxies. MNRAS, 281. <https://doi.org/10.1093/mnras/281.1.153>

- 4564 Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge
4565 and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- 4566
- 4567 Panwar, M., Prabhakar, Sandhu, P. K., Wadadekar, Y., & Jain, P. (2020). Alignment of
4568 radio galaxy axes using FIRST catalogue. *MNRAS*. <https://doi.org/10.1093/mnras/staa2975>
- 4569
- 4570 Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison,
4571 A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch, In *NIPS 2017 Workshop Autodiff*.
- 4572
- 4573 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel,
4574 M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine
4575 learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- 4576
- 4577 Polsterer, K. L., Gieseke, F., & Igel, C. (2015). Automatic galaxy classification via ma-
4578 chine learning techniques: Parallelized rotation/flipping invariant kohonen maps
4579 (PINK), In *Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV)*.
- 4580
- 4581 Pracy, M. B., Ching, J. H. Y., Sadler, E. M., Croom, S. M., Baldry, I. K., Bland-Hawthorn,
4582 J., Brough, S., Brown, M. J. I., Couch, W. J., Davis, T. M., Drinkwater, M. J.,
4583 Hopkins, A. M., Jarvis, M. J., Jelliffe, B., Jurek, R. J., Loveday, J., Pimbblet, K. A.,
4584 Prescott, M., Wisnioski, E., & Woods, D. (2016). GAMA/WiggleZ: The 1.4 GHz
4585 radio luminosity functions of high- and low-excitation radio galaxies and their
4586 redshift evolution to $z = 0.75$. *MNRAS*, 460, 2–17. <https://doi.org/10.1093/mnras/stw910>
- 4587
- 4588 Proctor, D. D. (2006). Comparing pattern recognition feature sets for sorting triples in
the FIRST database. *ApJS*, 165, 95–107. <https://doi.org/10.1086/504801>
- 4589
- 4590 Proctor, D. D. (2011). Morphological annotations for groups in the FIRST database.
ApJS, 194. <https://doi.org/10.1088/0067-0049/194/2/31>
- 4591
- 4592 Proctor, D. D. (2016). A selection of giant radio sources from NVSS. *ApJS*, 224. <https://doi.org/10.3847/0067-0049/224/2/18>
- 4593
- 4594 Ralph, N. O., Norris, R. P., Fang, G., Park, L. A. F., Galvin, T. J., Alger, M. J., Andernach,
H., Lintott, C., Rudnick, L., Shabala, S., & Wong, O. I. (2019). Radio Galaxy Zoo:
4595 Unsupervised clustering of convolutionally auto-encoded radio-astronomical
4596 images. *PASP*, 131(1004). <https://doi.org/10.1088/1538-3873/ab213d>
- 4597
- 4598 Randall, K., Hopkins, A., Norris, R., Zinn, P.-C., Middelberg, E., Mao, M., & Sharp, R.
4599 (2012). Spectral index properties of milliJansky radio sources. *MNRAS*, 421(2),
1644–1660. <https://doi.org/10.1111/j.1365-2966.2012.20422.x>
- 4600
- 4601 Raouf, M., Shabala, S. S., Croton, D. J., Khosroshahi, H. G., & Bernyk, M. (2017). The
4602 many lives of active galactic nuclei — II: The formation and evolution of radio
4603 jets and their impact on galaxy evolution. *MNRAS*, 471, 658–670. <https://doi.org/10.1093/mnras/stx1598>
- 4604
- 4605 Raykar, V. C., Yu, S., Zhao, L. H., Valadez, G. H., Florin, C., Bogoni, L., & Moy, L. (2010).
Learning from crowds. *Journal of Machine Learning Research*, 11(43), 1297–1322.

- 4606 Rayner, D. P., Norris, R. P., & Sault, R. J. (2000). Radio circular polarization of active
4607 galaxies. *MNRAS*, *319*, 484–496. <https://doi.org/10.1111/j.1365-8711.2000.03854.x>
- 4608
- 4609 Reiprich, T. H., Veronica, A., Pacaud, F., Ramos-Ceja, M. E., Ota, N., Sanders, J., Kara,
4610 M., & Erben, T. (2020). The Abell 3391/95 galaxy cluster system. A 15 Mpc
4611 intergalactic medium emission filament, a warm gas bridge, infalling matter
4612 clumps, and (re-) accelerated plasma discovered by combining SRG/eROSITA
4613 data with ASKAP/EMU and DECam data. *A&A*. <https://doi.org/10.1051/0004-6361/202039590>
- 4614
- 4615 Richter, G. A. (1975). Search for optical identifications in the 5C3-radio survey. II —
4616 statistical treatment and results. *Astronomische Nachrichten*, *296*, 65–81. <https://doi.org/10.1002/asna.19752960203>
- 4616
- 4617
- 4618 Rodman, P. E., Turner, R. J., Shabala, S. S., Banfield, J. K., Wong, O. I., Andernach,
4619 H., Garon, A. F., Kapińska, A. D., Norris, R. P., & Rudnick, L. (2019). Radio
4620 Galaxy Zoo: Observational evidence for environment as the cause of radio
4621 source asymmetry. *MNRAS*, *482*, 5625–5641. <https://doi.org/10.1093/mnras/sty3070>
- 4622
- 4623 Rowley, H. A., Baluja, S., & Kanade, T. (1996). Human face detection in visual scenes,
4624 In *Advances in neural information processing systems*.
- 4624
- 4625 Rybicki, G. B., & Lightman, A. P. (2008). *Radiative processes in astrophysics*. John Wiley
4626 & Sons.
- 4626
- 4627 Sadler, E. M., Ekers, R. D., Mahony, E. K., Mauch, T., & Murphy, T. (2014). The local
4628 radio-galaxy population at 20 GHz. *MNRAS*, *438*, 796–824. <https://doi.org/10.1093/mnras/stt2239>
- 4629
- 4630 Sadler, E. M., Jackson, C. A., Cannon, R. D., McIntyre, V. J., Murphy, T., Bland-Hawthorn,
4631 J., Bridges, T., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., de Propris,
4632 R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Lahav,
4633 O., ... Taylor, K. (2002). Radio sources in the 2dF galaxy redshift survey — II.
4634 local radio luminosity functions for AGN and star-forming galaxies at 1.4 GHz.
4635 *MNRAS*, *329*(1), 227–245. <https://doi.org/10.1046/j.1365-8711.2002.04998.x>
- 4636
- 4637 Saikia, D. J., & Salter, C. J. (1988). Polarization properties of extragalactic radio sources.
4638 *ARA&A*, *26*, 93–144. <https://doi.org/10.1146/annurev.aa.26.090188.000521>
- 4638
- 4639 Sajina, A., Lacy, M., & Scott, D. (2005). Simulating the Spitzer mid-infrared color-color
4640 diagrams. *ApJ*, *621*, 256–268. <https://doi.org/10.1086/426536>
- 4640
- 4641 Sakelliou, I., Hardcastle, M. J., & Jetha, N. N. (2008). 3C40 in Abell194: Can tail radio
4642 galaxies exist in a quiescent cluster? *MNRAS*, *384*, 87–93. <https://doi.org/10.1111/j.1365-2966.2007.12465.x>
- 4642
- 4643 Samudre, A., George, L., Bansal, M., & Wadadekar, Y. (2020). Data-efficient classification
4644 of radio galaxies. *arXiv e-prints*, *2021*, arXiv:2011.13311v1 [astro-ph.IM].
- 4644
- 4645 Scalzo, R. A., Yuan, F., Childress, M. J., Möller, A., Schmidt, B. P., Tucker, B. E., Zhang,
4646 B. R., Onken, C. A., Wolf, C., Astier, P., Betoule, M., & Regnault, N. (2017). The
4647 SkyMapper Transient Survey. *PASA*, *34*. <https://doi.org/10.1017/pasa.2017.24>
- 4648
- 4649 Schmidt, M. (1968). Space distribution and luminosity functions of quasi-stellar radio
4650 sources. *ApJ*, *151*. <https://doi.org/10.1086/149446>

- 4650 Segal, G., Parkinson, D., Norris, R. P., & Swan, J. (2019). Identifying complex sources in
4651 large astronomical data sets using a coarse-grained complexity measure. *PASP*,
4652 131(1004). <https://doi.org/10.1088/1538-3873/ab0068>
- 4653 Shabala, S. S. (2018). The role of environment in the observed fundamental plane of
4654 radio active galactic nuclei. *MNRAS*, 478, 5074–5080. <https://doi.org/10.1093/mnras/sty1328>
- 4655 Shabala, S. S., & Godfrey, L. E. H. (2013). Size dependence of the radio-luminosity-
4656 mechanical-power correlation in radio galaxies. *ApJ*, 769(2). <https://doi.org/10.1088/0004-637X/769/2/129>
- 4657 Shimwell, T. W., Tasse, C., Hardcastle, M. J., Mechev, A. P., Williams, W. L., Best, P. N.,
4658 Röttgering, H. J. A., Callingham, J. R., Dijkema, T. J., de Gasperin, F., Hoang,
4659 D. N., Hugo, B., Mirmont, M., Oonk, J. B. R., Prandoni, I., Rafferty, D., Sabater,
4660 J., Smirnov, O., van Weeren, R. J., ... Wilber, A. (2019). The LOFAR Two-metre
4661 Sky Survey. II. First data release. *A&A*, 622. <https://doi.org/10.1051/0004-6361/201833559>
- 4662 Sieland, M. (2015). Snapshot Serengeti: An Unexpected Discovery. Retrieved January
4663 16, 2021, from <https://www.mico-project.eu/snapshot-serengeti-an-unexpected-discovery/>
- 4664 Sikora, M., & Begelman, M. C. (2013). Magnetic flux paradigm for radio loudness of
4665 active galactic nuclei. *ApJ*, 764. <https://doi.org/10.1088/2041-8205/764/2/L24>
- 4666 Stokes, G. G. (1851). On the composition and resolution of streams of polarized light
4667 from different sources. *Transactions of the Cambridge Philosophical Society*, 9, 399.
- 4668 Storrie-Lombardi, M. C., Lahav, O., Sodre, L., & Storrie-Lombardi, L. J. (1992). Mor-
4669 phological classification of galaxies by artificial neural networks. *MNRAS*, 259.
4670 <https://doi.org/10.1093/mnras/259.1.8P>
- 4671 Strauss, M. A., Weinberg, D. H., Lupton, R. H., Narayanan, V. K., Annis, J., Bernardi,
4672 M., Blanton, M., Burles, S., Connolly, A. J., Dalcanton, J., Doi, M., Eisenstein,
4673 D., Frieman, J. A., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Kim, R. S. J.,
4674 Knapp, G. R., ... Zehavi, I. (2002). Spectroscopic target selection in the Sloan
4675 Digital Sky Survey: The main galaxy sample. *AJ*, 124, 1810–1824. <https://doi.org/10.1086/342343>
- 4676 Sun, X. H., Rudnick, L., Akahori, T., Anderson, C. S., Bell, M. R., Bray, J. D., Farnes,
4677 J. S., Ideguchi, S., Kumazaki, K., O'Brien, T., & et al. (2015). Comparison of al-
4678 gorithms for determination of rotation measure and Faraday structure. I. 1100–
4679 1400 MHz. *AJ*, 149(2), 60. <https://doi.org/10.1088/0004-6256/149/2/60>
- 4680 Surace, J., Shupe, D., Fang, F., Lonsdale, C., Gonzalez-Solares, E., Hatziminaoglou, E.,
4681 Siana, B., Babbedge, T., Polletta, M., Rodighiero, G., et al. (2005). The SWIRE
4682 data release 2: Image atlases and source catalogs for ELAIS-N1, ELAIS-N2,
4683 XMM-LSS, and the Lockman Hole. *Spitzer Science Centre, California Institute of
4684 Technology, Pasadena, CA*.
- 4685 Tang, H., Scaife, A. M. M., & Leahy, J. P. (2019). Transfer learning for radio galaxy
4686 classification. *MNRAS*, 488, 3358–3375. <https://doi.org/10.1093/mnras/stz1883>
- 4687
- 4688
- 4689
- 4690
- 4691

- 4692 Taylor, A. R., & Jagannathan, P. (2016). Alignments of radio galaxies in deep radio
 4693 imaging of ELAIS N1. *MNRAS*, 459(1), L36–L40. <https://doi.org/10.1093/mnrasl/slw038>
- 4695 Taylor, A. R., Stil, J. M., Grant, J. K., Landecker, T. L., Kothes, R., Reid, R. I., Gray, A. D.,
 4696 Scott, D., Martin, P. G., Boothroyd, A. I., Joncas, G., Lockman, F. J., English,
 4697 J., Sajina, A., & Bond, J. R. (2007). Radio polarimetry of the ELAIS N1 field:
 4698 Polarized compact sources. *ApJ*, 666, 201–211. <https://doi.org/10.1086/519786>
- 4699 Taylor, A. R., Stil, J. M., & Sunstrum, C. (2009). A rotation measure image of the sky.
 4700 *ApJ*, 702(2), 1230–1236. <https://doi.org/10.1088/0004-637X/702/2/1230>
- 4701 The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim,
 4702 P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N.,
 4703 Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro,
 4704 M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., ... Astropy Con-
 4705 tributors. (2018). The Astropy Project: Building an open-science project and
 4706 status of the v2.0 core package. *AJ*, 156(3). <https://doi.org/10.3847/1538-3881/aabc4f>
- 4708 Treichel, K., Rudnick, L., Hardcastle, M. J., & Leahy, J. P. (2001). Spectral structure in FR
 4709 II radio galaxies and jets. *ApJ*, 561(2), 691–702. <https://doi.org/10.1086/323254>
- 4710 Tucker, B. (2017). Stargazing citizens join search for Planet Nine. Retrieved January 16,
 4711 2021, from <https://www.abc.net.au/news/2017-04-04/stargazing-live-citizens-join-search-for-planet-nine/8414832>
- 4713 Tully, R. B., & Fisher, J. R. (1977). A new method of determining distances to galaxies.
 4714 *A&A*, 54, 661–673.
- 4715 Turner, R. J., & Shabala, S. S. (2015). Energetics and lifetimes of local radio active galac-
 4716 tic nuclei. *ApJ*, 806. <https://doi.org/10.1088/0004-637X/806/1/59>
- 4717 Urry, C. M., & Padovani, P. (1995). Unified schemes for radio-loud active galactic nuclei.
 4718 *PASP*, 107(715), 803.
- 4719 Van Eck, C. L., Havercorn, M., Alves, M. I. R., Beck, R., de Bruyn, A. G., Enßlin, T.,
 4720 Farnes, J. S., Ferrière, K., Heald, G., Horellou, C., Horneffer, A., Iacobelli, M.,
 4721 Jelić, V., Martí-Vidal, I., Mulcahy, D. D., Reich, W., Röttgering, H. J. A., Scaife,
 4722 A. M. M., Schnitzeler, D. H. F. M., ... Sridhar, S. S. (2017). Faraday tomography
 4723 of the local interstellar medium with LOFAR: Galactic foregrounds towards IC
 4724 342. *A&A*, 597. <https://doi.org/10.1051/0004-6361/201629707>
- 4725 van Velzen, S., Falcke, H., Schellart, P., Nierstenhöfer, N., & Kampert, K.-H. (2012).
 4726 Radio galaxies of the local universe: All-sky catalog, luminosity functions, and
 4727 clustering. *A&A*, 544. <https://doi.org/10.1051/0004-6361/201219389>
- 4728 Verheijen, M. A. W., Oosterloo, T. A., van Cappellen, W. A., Bakker, L., Ivashina, M. V.,
 4729 & van der Hulst, J. M. (2008). Apertif, a focal plane array for the WSRT, In
 4730 *AIP Conference Series 1035, The Evolution of Galaxies Through the Neutral Hydrogen*
 4731 *Window*.
- 4732 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D.,
 4733 Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett,
 4734 M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern,
 4735 R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algo-

- 4779 - a deep learning classifier for radio morphologies. *MNRAS*, 482, 1211–1230.
4780 <https://doi.org/10.1093/mnras/sty2646>
- 4781 Zevin, M., Coughlin, S., Bahaadini, S., Besler, E., Rohani, N., Allen, S., Cabero, M.,
4782 Crowston, K., Katsaggelos, A. K., Larson, S. L., & et al. (2017). Gravity Spy:
4783 Integrating advanced LIGO detector characterization, machine learning, and
4784 citizen science. *Classical and Quantum Gravity*, 34. <https://doi.org/10.1088/1361-6382/aa5cea>
- 4785 Zhang, C., Wang, C., Hobbs, G., Russell, C. J., Li, D., Zhang, S.-B., Dai, S., Wu, J.-W.,
4786 Pan, Z.-C., Zhu, W.-W., Toomey, L., & Ren, Z.-Y. (2020). Applying saliency-map
4787 analysis in searches for pulsars and fast radio bursts. *A&A*, 642. <https://doi.org/10.1051/0004-6361/201937234>
- 4788 Zhang, J., Wu, X., & Sheng, V. S. (2016). Learning from crowdsourced labeled data: A
4789 survey. *Artificial Intelligence Review*, 46(4), 543–576. <https://doi.org/10.1007/s10462-016-9491-9>
- 4790 Zhang, L. (2020). Transfer adaptation learning: A decade survey. *arXiv e-prints*, arXiv:
4791 1903.04687v2 [cs.CV].
- 4792 Zhu, X.-P., Dai, J.-M., Bian, C.-J., Chen, Y., Chen, S., & Hu, C. (2019). Galaxy morphology
4793 classification with deep convolutional neural networks. *Astrophysics and Space Science*, 364, 55. <https://doi.org/10.1007/s10509-019-3540-1>
- 4794 Zhu, Y., Ong, C. S., & Huttley, G. A. (2020). Machine learning techniques for classifying
4795 the mutagenic origins of point mutations. *Genetics*, 215(1), 25–40. <https://doi.org/10.1534/genetics.120.303093>
- 4796 Zhuang, M.-Y., & Ho, L. C. (2020). The interplay between star formation and black
4797 hole accretion in nearby active galaxies. *ApJ*, 896. <https://doi.org/10.3847/1538-4357/ab8f2e>
- 4798 Zubovas, K., & King, A. (2012). Clearing out a galaxy. *ApJ*, 745. <https://doi.org/10.1088/2041-8205/745/2/L34>
- 4799 Zubovas, K., Sabulis, K., & Naujalis, R. (2014). Collapse and fragmentation of molecular
4800 clouds under pressure. *MNRAS*, 442(3), 2837–2854. <https://doi.org/10.1093/mnras/stu1048>
- 4801 Żywucka, N., Kozięł-Wierzbowska, D., & Goyal, A. (2020). Catalogue with visual
4802 morphological classification of 32,616 radio galaxies with optical hosts, In *Proceedings of the international astronomical union symposium 356: Nuclear activity in galaxies across cosmic time*.
- 4803
- 4804
- 4805
- 4806
- 4807
- 4808
- 4809
- 4810
- 4811
- 4812