Machine Learning on the Radio Galaxy Zoo

Matthew Alger
Supervisor: Cheng Soon Ong

June 3, 2016

Abstract

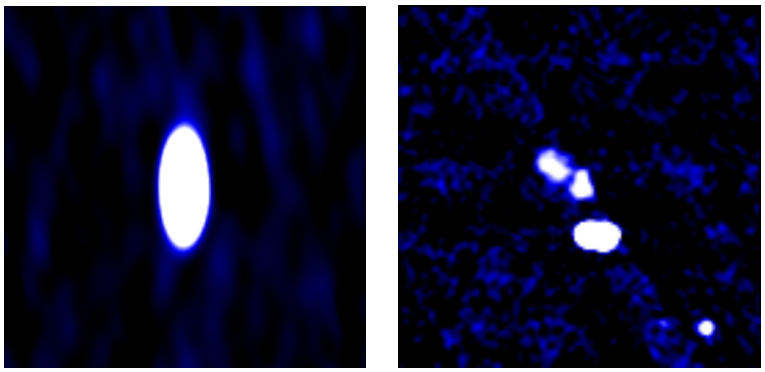
The problem of cross-identifying radio sources and their host galaxies is becoming increasingly difficult, with newer, wider, and more detailed surveys detecting more radio sources than current methods can handle. This report describes a machine learning approach to the cross-identification problem, using crowdsourced cross-identifications from the Radio Galaxy Zoo as supervised training data. While the approach is not currently cabable of distinguishing between multiple nearby radio sources, it does achieve 80.75% accuracy on locating the host for a given radio source in the ATLAS survey and does not use any astrophysical models.

1 Introduction

1.1 Cross-identification of Radio Sources and Host Galaxies

Radio surveys such as Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) [14, 2] and the Australian Telescope Large Area Survey (ATLAS) [4] have found many sources of radio emissions. These radio sources are dominated by *active galactic nuclei* (AGNs) [1], galactic centres with supermassive black holes that emit radio waves[9]. Galaxies containing a radio source are referred to as *host galaxies*. These galaxies are found in infrared surveys such as the Wide-field Infrared Survey Explorer (WISE) [15] and the SIRTF Wide-area Infrared Extragalactic survey (SWIRE) [12, 6].

Astrophysicists are interested in the properties of both AGNs and their host galaxies, but to investigate either, the radio sources need to be matched to their host galaxies. This is called *cross-identification*. Many radio sources are *compact radio sources*, where the radio emissions directly and simply overlap the host galaxy (Figure 1a). These radio sources are easy to cross-identify[1]. However, many radio sources are instead *complex radio sources*, where radio emissions can be large, sprawling, and not relate to the host galaxy in any simple way (Figure 1b).



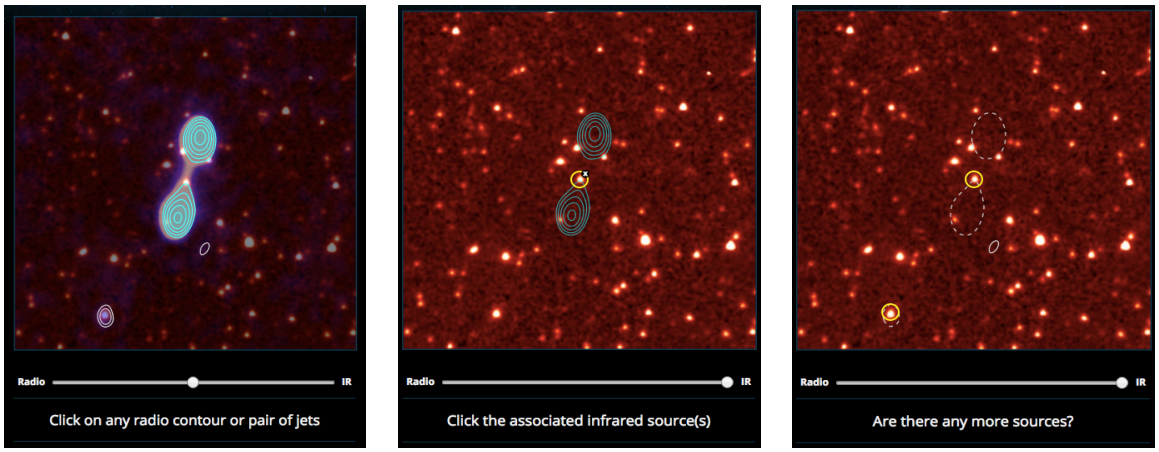
(a) A compact radio source. (b) A complex radio source.

Figure 1: Example radio emissions.

Radio Galaxy Zoo[1] is an online citizen science project that aims to crowdsource the cross-identification problem[1]. Volunteers are presented with a radio image of a small part of the sky (from FIRST or

[1] Radio Galaxy Zoo

ATLAS) and the corresponding infrared image (from WISE or SWIRE). Each part of the sky presented in this way is called a *subject*, and contains at least one radio emitter. Volunteers are asked to select which radio emissions are part of the same radio source, and which galaxy in the infrared image emitted the radio source. The workflow is shown in Figure 2.



(a) Select radio emissions. (b) Locate the host galaxy. (c) Repeat for all emissions.

Figure 2: Radio Galaxy Zoo volunteer workflow.

To increase cross-identification accuracy, each compact radio source is presented to 5 volunteers, and each complex radio source is presented to 20 volunteers[1].

Over 100 000 radio sources have been cross-identified by volunteers so far[2] out of the Radio Galaxy Zoo database of around 177 000 radio sources, compared to a few thousand classifications by experts[1]. However, new surveys such as the Evolutionary Map of the Universe (EMU) [8] and Westerbork Observations of the Deep APERTIF Northern-Sky (WODAN) [11] are expected to detect over 100 million radio sources[1], making crowdsourcing an intractable solution to the cross-identification problem.

I have developed an automated, machine learning approach to solving the cross-identification problem, framing the problem as supervised binary classification and using classifications made by Radio Galaxy Zoo volunteers as training data.

1.2 Related Work

Most existing algorithms for the cross-identification task are model-based, assuming that radio sources will have some specific structure (usually double-lobed sources, meaning that there is some central object ejecting a jet of material on each side). Van Velzen et al. [13] developed a method of identifying double-lobed radio sources in FIRST, which fits weighted lines to the two lobes and assumes that the host galaxy is located at the centre of the line. This method is clearly best suited to radio sources that form a straight line, while many other source morphologies (e.g. bent double-lobes) exist. Fan et al. [3] also developed a model-based method. Their method is probabilistic and aims to cross-identify radio sources and host galaxies based on modelling the radio source as a central radio object and two surrounding lobes in an approximately straight line. This morphology is similar to that of van Velzen et al. [13], but explicitly models the angle between lobes, and so can also cross-identify bent double-lobes. They achieved 85.8% cross-identification accuracy on the ATLAS data set against the manual cross-identifications made by Norris et al. [7]. While their method is limited by its dependence on the model, it is also easily extended to different models, and so may provide a good basis for future astronomical model-based work. Their method is also able to determine whether nearby radio objects are part of the same source. This is one of the problems Radio Galaxy Zoo is attempting to solve, as collating radio objects is a difficult task for computers. It is estimated that around 10% of radio sources detected by EMU will not be able to be cross-identified with existing algorithms like these[1, 8].

Some work has been done on feature selection for radio source morphologies, such as by Proctor [10], who investigated feature selection for sorting bent double-lobed radio sources in FIRST. This form of feature selection has not been used for this project.

[2]Based on the data supplied to me.

2 Data Sources

2.1 ATLAS

ATLAS is a radio-wavelength survey of the Chandra Deep Field South (CDFS) and the European Large Area ISO Survey – South 1 (ELAIS-S1) fields, which were chosen as they are both areas of the sky covered by the earlier SWIRE survey. This means that the ATLAS observations have corresponding observations in infrared wavelengths[4]. Infrared observations are necessary for cross-identification, since the distant galaxies we want to cross-identify the radio objects with emit infrared radiation.

While the Radio Galaxy Zoo data include classifications of objects in both the ATLAS and FIRST surveys, here I have only focused on the ATLAS observations of CDFS. There are several reasons for this. Firstly, ATLAS contains only 2400 objects. This is a good size — large enough to train supervised machine learning models, but small enough to quickly iterate on these models as is required for such exploratory research. Secondly, ATLAS is mostly well-behaved, compact objects. While the eventual goal of this project is to be able to automatically cross-identify even very complicated complex sources, being able to accurately cross-identify simple compact sources is a necessary step. Thirdly, ATLAS is considered a "test run" for the much larger EMU survey, as EMU is similar in resolution to ATLAS[4]. EMU is where many new radio observations will be made, and the scale of the survey is a key motivator behind this project. Finally, ATLAS subjects have been cross-identified by experts already[7], meaning that I do not need to rely entirely on volunteer consensuses from the crowdsourced Radio Galaxy Zoo, and I can thus better validate and evaluate my results.

ATLAS observations of CDFS consist of a 3.6 deg^2 mosaic of radio images between $3^{\mathrm{h}}26^{\mathrm{m}}27^{\circ}00'$ and $3^{\mathrm{h}}36^{\mathrm{m}}29^{\circ}00'$. The full ATLAS image of the CDFS field is shown in Figure 3.

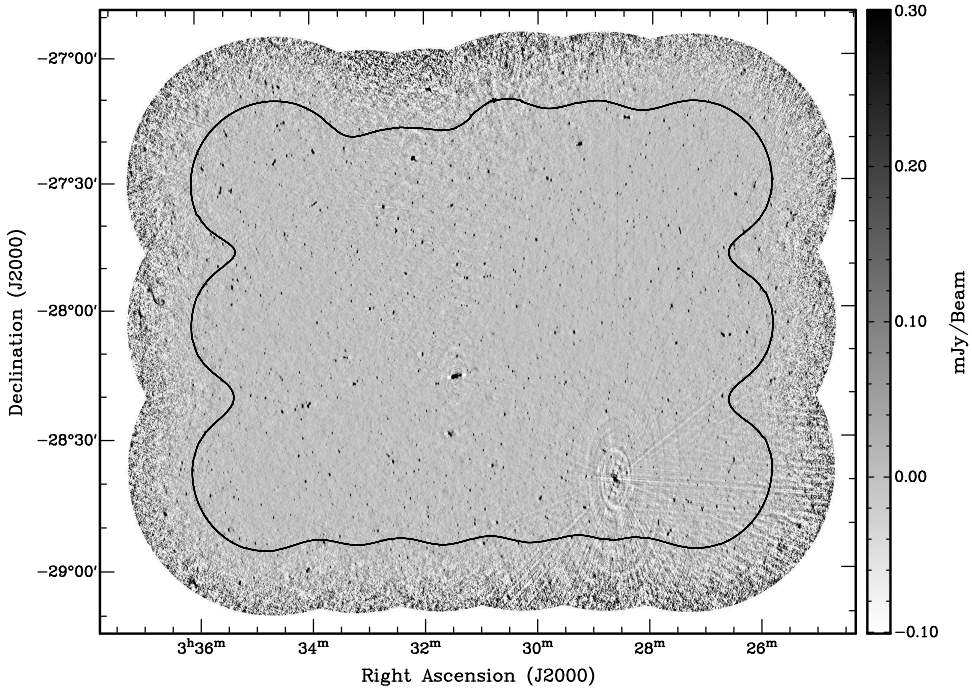


Figure 3: ATLAS observations of CDFS. Reproduced from Franzen et al. [4].

.

Each ATLAS object forms a subject. Each subject consists of a $2' \times 2'$ image patch from Figure 3 and a corresponding image patch from SWIRE centred on the associated ATLAS object.

2.2 SWIRE

SWIRE is an infrared-wavelength survey of seven regions of the sky in seven infrared bands. Of these regions, CDFS and ELAIS-S1 overlap with ATLAS and are hence used in Radio Galaxy Zoo.

SWIRE observations are infrared images in the various fields, and are provided in Radio Galaxy Zoo as $2' \times 2'$ image patches centred on ATLAS subjects. In addition, I make use of the SWIRE CDFS Region

Fall '05 Spitzer Catalog[12], which describes all objects detected in the CDFS field in the SWIRE survey.

For each object in CDFS, the catalogue provides the name, location, infrared fluxes, and stellarity index associated with that object. The location is specified in right ascension and declination. The fluxes are given in five bands (3.6 μJy, 4.5 μJy, 5.8 μJy, 8.0 μJy, and 24 μJy) and describe how bright each object is in the corresponding flux band. Finally, the stellarity index is an indicator of how star-like each object is according to the SExtractor software package, where 0 denotes an object that is totally non-star-like, and 1 denotes an object that is totally star-like[12].

2.3 Radio Galaxy Zoo

Each ATLAS subject in the Radio Galaxy Zoo data is associated with a set of *crowd classifications*, cross-identifications performed by volunteers. Each classification describes which radio objects near the subject the volunteer believes are part of the same radio source (called a *radio combination*), as well as the location that the volunteer believes the corresponding host galaxy is located at.

There are multiple classifications for each subject (5 for compact sources and 20 for complex sources) to attempt to improve accuracy. These classifications differ from each other, so they need to be brought together in some way to identify a single radio combination/galaxy location label for each radio source in the data. The collated radio combinations for subjects are called *radio consensuses*, the collated locations for a subject are called *location consensuses*, and collated classifications on the whole are called *consensuses*. Ideally, we want to take some "majority vote" and choose the most common radio combination as the radio consensus and then the most common corresponding locations as the location consensuses. The former is simple — we just count the different combinations of radio objects and choose the most common — but finding the location consensus is considerably more difficult as many different locations could represent the same host galaxy. Banfield et al. [1] solve this problem by performing kernel density estimation on the locations associated with each radio combination, then choosing the location with the highest density. For this report, I have used a different method that uses a clustering algorithm called PG-means[5]. The different locations are assigned to clusters, and the mean of the largest cluster becomes the location consensus.

3 The Cross-identification Task

3.1 Cross-identification as Binary Classification

The first step in applying machine learning methods to the cross-identification task is to find a machine learning framework that it fits in. The cross-identification task can be modelled as binary classification, allowing the use of standard binary classification methods. This is done as follows. For an ATLAS subject, consider all SWIRE objects within $1'$ Chebyshev distance of the corresponding ATLAS object's location, i.e., all galaxies that a volunteer would be allowed to choose from in the crowdsourced cross-identification task. These objects are called *candidate hosts*. Each candidate is then either the host galaxy or not the host galaxy. "Host galaxy" and "not host galaxy" can thus be interpreted as two distinct classes, forming a binary classification problem. After training a classifier on this task, we can find the host galaxy in a subject by using the classifier to predict the probabilities that each candidate is the host galaxy, and then simply choose the candidate with the highest probability of being the host galaxy.

Note that I am ignoring the problem of having multiple host galaxies in a subject, and am assuming that any subject contains exactly one host. This is an oversimplification as there are indeed subjects that contain multiple hosts, but there are very few in the ATLAS data and they greatly increase the difficulty of the problem.

3.2 Labels

For training and evaluating the classifier, each SWIRE object is assigned a "true" label. This label comes from Radio Galaxy Zoo consensuses: For each consensus location found by Radio Galaxy Zoo volunteers, the nearest SWIRE object is labelled "host galaxy". All other SWIRE objects are labelled "not host galaxy".

I make the assumption that these labels are accurate. However, this may not be the case, since volunteers may not correctly identify host galaxies. Banfield et al. [1] found that when more than 75% of

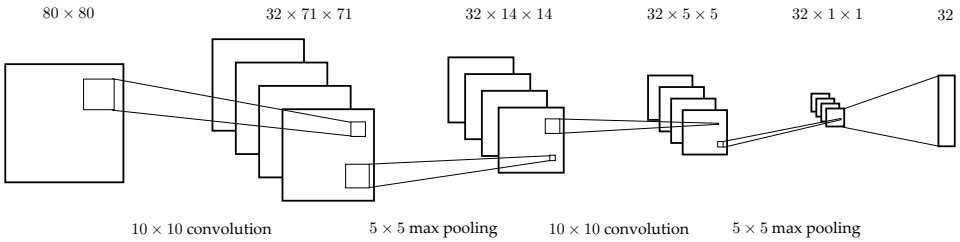


Figure 4: Convolutional neural network for radio feature extraction. An 80×80 pixel patch of radio image (corresponding to $0.8' \times 0.8'$ of radio sky) is passed through two convolutional layers to obtain 32 features.

volunteers agree, then their cross-identifications are as accurate as experts', and volunteers agree to this level on the vast majority of ATLAS subjects, so this assumption generally holds for the ATLAS Radio Galaxy Zoo classifications. It may not hold for other data sets such as the FIRST Radio Galaxy Zoo classifications.

3.3 Feature-space Representation of SWIRE Objects

To be able to classify a candidate, a feature-space representation of the candidate must be found. For a given candidate, the features used are to represent it are:

- the 3.6 μJy flux of the candidate,
- the 4.5 μJy flux of the candidate,
- the 5.8 μJy flux of the candidate,
- the 8.0 μJy flux of the candidate,
- the 24 μJy flux of the candidate,
- the Chebyshev distance from the candidate to the nearest ATLAS object, and
- features extracted from patches of radio image centred on the candidate.

The infrared images around each candidate seem to contain little predictive information, and so are not included in the features.

The radio patches are $0.8' \times 0.8'$ (80 px $\times$ 80 px) images centred on each candidate. Features are extracted from these patches by a convolutional neural network. The network architecture is shown in Figure 4. The network is pre-trained by appending a 32×64 dense layer and a 64×1 dense layer, then training the new convolutional neural network to find a map from the radio patches to the associated binary labels using backpropagation.

These features are all normalised, and the non-radio features are scaled. Scaling the radio features seems to significantly decrease classification accuracy, so I have chosen not to scale the radio features.

3.4 Classification Algorithm

As input, we consider an ATLAS subject. We wish to automatically identify where the host galaxy that emitted this radio source is located. To perform this identification, we perform the following steps.

1. Identify all candidates (SWIRE objects) within $1'$ Chebyshev distance of the ATLAS subject.
2. For each candidate, generate features (as described in 3.3).
3. Classify each candidate as being either the host galaxy or not being the host galaxy. This can be done with any classifier that can output a probability estimate.
4. Select the candidate with the highest probability of being the host galaxy.
5. Return the coordinates of the selected candidate.

The above classification algorithm is pictured in Figure 5. In this report, I have experimented with using both logistic regression and random forest classifiers.

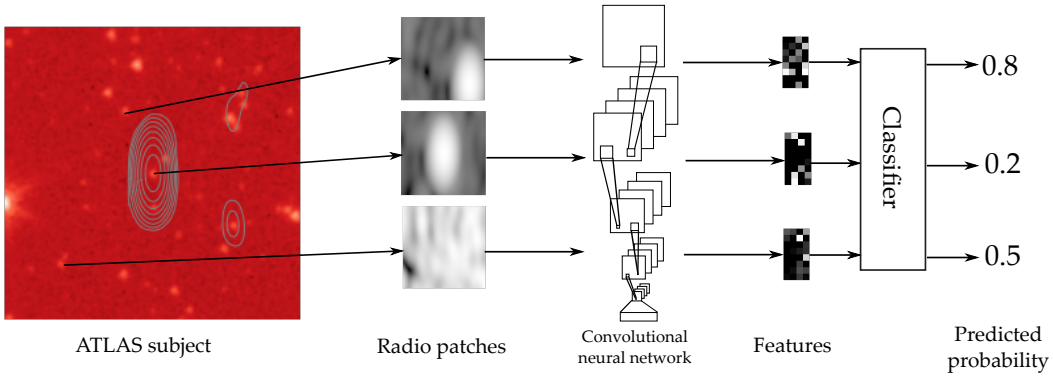


Figure 5: Classification algorithm. Non-image features are not pictured. Values are for indication only and are not the actual output of the classifier on the displayed inputs.

3.5 Evaluating the Classifier

There are two main ways to evaluate the classifier. Firstly, the classifier can be used to label a set of training SWIRE objects with known labels, and the classification accuracy can be found. This is very easy, but does not capture the cross-identification task — labelling an individual SWIRE object may be more or less difficult than finding the host galaxy amongst a set of other candidate hosts. The second way to evaluate the classifier is to find the host coordinates for a set of training *ATLAS subjects*, and then compare these coordinates to those found by Radio Galaxy Zoo volunteers. The performance measure is then how many ATLAS subjects the classifier is able to find the correct host location for.

Both of these methods of evaluation are presented in Section 4.

4 Results

4.1 Logistic Regression

Logistic regression was used in the classification algorithm described in Section 3.4, with 1.0 L_2 regularisation. Classifications were tested against crowdsourced Radio Galaxy Zoo consensuses. This resulted in 91.25% balanced classification accuracy when classifying individual SWIRE objects, and 72.12% accuracy when finding the host galaxy for ATLAS subjects. The precision–recall and ROC curves for the individual classification task are displayed in Figure 6, and the confusion matrix for the individual classification task is displayed in Figure 7. These metrics are not well-defined for the full host galaxy task, but sample outputs are displayed in Appendix I.

4.2 Random Forests

Random forests with 10 trees were used in the classification algorithm described in Section 3.4. Classifications were tested against crowdsourced Radio Galaxy Zoo consensuses. This resulted in 97.50% balanced classification accuracy when classifying individual SWIRE objects, and 80.75% accuracy when finding the host galaxy for ATLAS subjects. The precision–recall and ROC curves for the individual classification task are displayed in Figure 8, and the confusion matrix for the individual classification task is displayed in Figure 9. These metrics are not well-defined for the full host galaxy task, but sample outputs are displayed in Appendix II.

5 Discussion

While it is not possible to directly compare against the results of Fan et al. [3], as I am making the (incorrect) assumption that there is only one host galaxy per image, their results of 85.8% accuracy are still a good baseline as they are for a more difficult task. With that in mind, my results of 80.75% accuracy with random forests are lower, but less assumptions are made on the data — my algorithm contains no astrophysical models, and all information is obtained directly from the ATLAS and SWIRE surveys. With more work, it may be possible to raise the accuracy higher.

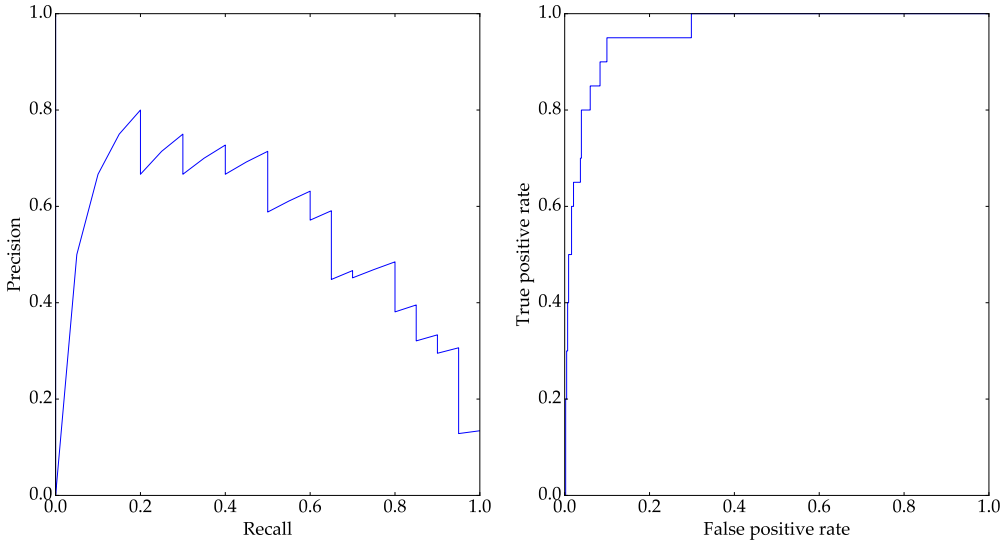


Figure 6: Precision–recall and ROC curves for logistic regression on the individual SWIRE object classification task.

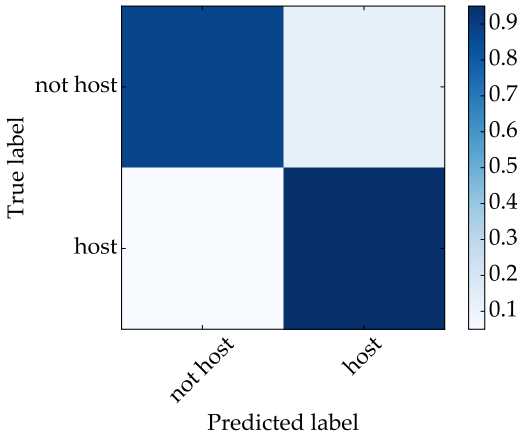


Figure 7: Confusion matrix for logistic regression on the individual SWIRE object classification task. The colour axis represents classification accuracy for each category.

It is interesting that random forests considerably outperforms logistic regression. Since logistic regression does not perform well if inputs are improperly scaled, this may be an indication that the inputs have not been properly scaled — particularly given that I did not scale the radio features at all. However, scaling the individual radio features to dropped the accuracy of logistic regression considerably. It may be the case that the radio features must be scaled in a uniform way, to preserve some scaled relationship between the different features.

Examining the sample of outputs in Appendix I, we can see that logistic regression does quite well on compact sources, confidently predicting the host galaxy to be in the middle of the source. This happens even when the source is distorted, such as in Figure 11. We can also see from Figure 12 that the logistic regression model has developed some concept of jets. The radio source pictured seems to be flowing up and to the left. Candidates to the left and below the centre of the radio source are assigned a relatively low probability, compared to the sources to the right and below of the radio source. Logistic regression does have some difficulty in handling complex sources such as Figure 13. This subject may be composed of three sources, with two compact sources in the middle and lower right, and one infrared-faint radio triple partly overlapping with the central compact source. The inability of my algorithm to handle multiple sources is clearly visible in this subject — candidates near both compact hosts are assigned high probabilities, but there is no way for the algorithm to say that these are unrelated sources. Another

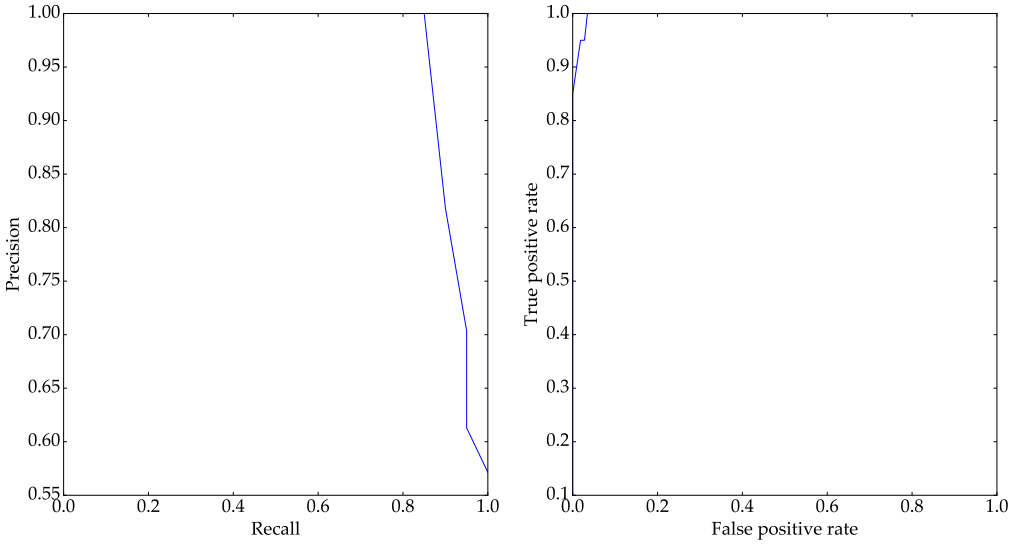


Figure 8: Precision–recall and ROC curves for random forests on the individual SWIRE object classification task.

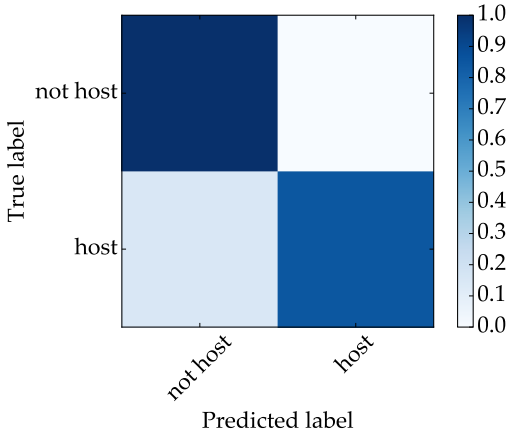


Figure 9: Confusion matrix for random forests on the individual SWIRE object classification task. The colour axis represents classification accuracy for each category.

flaw in my algorithm is also apparent from this subject in that there is no capability to accurately cross-identify infrared-faint radio sources, since they have no associated SWIRE host.

Figures 14, 16, and 17 show that random forest tends to be much more certain about the *un*likeliness of a candidate to be the host, as indicated by the very low probabilities assigned to almost all candidates. This is reflected in the confusion matrix (Figure 9), where random forests tends to under-predict whether a candidate is a host. Random forests is also remarkably decisive about Figure 17, even ignoring the large compact host in the lower right of the subject. This may indicate that the classifier places a high emphasis on feature indicating distance from the radio object, especially when compared to logistic regression. Random forests performs poorly on Figure 15, assigning equal probabilities to two candidates despite a relatively clear compact host in the centre. This most likely indicates that the classifier is "confused" by the infrared-weak subject in the lower right, and — as random forests is not a probabilistic classifier — is unable to determine that the central source should be focused on.

6 Future Work

In this report, my algorithm was only trained and tested against data from the ATLAS survey. While there were good reasons for this (Section 2.1), the algorithm should also be evaluated against the FIRST survey. FIRST is considerably higher resolution than ATLAS[7, 2], much larger than ATLAS, and contains more varied radio sources.

There are a number of potentially useful features I have not made use of. This most notably includes the SWIRE infrared images, but may also include explicit morphological radio features such as those presented in Proctor [10]. It would be interesting to see if these features are useful, and in particular if the explicit features are more useful than the features extracted by the convolutional neural network.

I have evaluated classification performance against the Radio Galaxy Zoo consensuses, mainly for simplicity as these are also the source of the training data. This means that the testing data is noisy, and hence the evaluation is also noisy. Future testing should be done against the expert data set from Norris et al. [7].

So far, little has been done to combat the possibility of inherent noise in the crowdsourced labels. While the crowdsourced labels for the ATLAS survey are not very noisy, this is not the case for the FIRST survey and will probably not be the case for future surveys like EMU. This noise could be modelled to estimate its effect on the classifier, or reduced somehow. Additionally, there is a possibility that Radio Galaxy Zoo itself has too little redundancy, and that 20 crowd classifications simply isn't enough for complex radio sources. One solution may be to predict the redundancy needed on a per-subject basis, to maximise the use of the limited number of volunteers.

One very interesting future direction is in active learning. Radio Galaxy Zoo, as a crowdsourced project, is well-positioned to benefit from active learning. Currently, radio subjects are classified by volunteers at random, but an intelligent way of finding unclassified radio subjects to show to volunteers may increase the impact of the volunteers' classifications on the training data. This is particularly important as we move toward larger surveys like EMU.

7 Conclusion

While the accuracy obtained through this method is not better than that obtained through existing methods, the algorithm presented here utilises only a supervised machine learning framework, and without the use of astrophysical models. This method will thus provide a good base for future work on the cross-identification problem on large and detailed radio surveys.

References

[1] J. Banfield, O. Wong, K. Willett, R. Norris, L. Rudnick, S. Shabala, B. Simmons, C. Snyder, A. Garon, N. Seymour, et al. Radio Galaxy Zoo: host galaxies and radio morphologies derived from visual inspection. *Monthly Notices of the Royal Astronomical Society*, 453(3):2326–2340, 2015.

[2] R. H. Becker, R. L. White, and D. J. Helfand. The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *Astrophysical Journal*, 450:559, Sept. 1995. doi: 10.1086/176166.

[3] D. Fan, T. Budavári, R. P. Norris, and A. M. Hopkins. Matching radio catalogues with realistic geometry: application to swire and atlas. *Monthly Notices of the Royal Astronomical Society*, 451(2): 1299–1305, 2015. doi: 10.1093/mnras/stv994. URL http://mnras.oxfordjournals.org/content/451/2/1299.abstract.

[4] T. Franzen, J. Banfield, C. Hales, A. Hopkins, R. Norris, N. Seymour, K. Chow, A. Herzog, M. Huynh, E. Lenc, et al. ATLAS–I. third release of 1.4 GHz mosaics and component catalogues. *Monthly Notices of the Royal Astronomical Society*, 453(4):4020–4036, 2015.

[5] Y. F. G. Hamerly. PG-means: learning the number of clusters in data. *Advances in neural information processing systems*, 19:393–400, 2007.

[6] C. J. Lonsdale, H. E. Smith, M. Rowan-Robinson, J. Surace, D. Shupe, C. Xu, S. Oliver, D. Padgett, F. Fang, T. Conrow, et al. SWIRE: The SIRTF wide-area infrared extragalactic survey. *Publications of the Astronomical Society of the Pacific*, 115(810):897, 2003.

[7] R. P. Norris, J. Afonso, P. N. Appleton, B. J. Boyle, P. Ciliegi, S. M. Croom, M. T. Huynh, C. A. Jackson, A. M. Koekemoer, C. J. Lonsdale, et al. Deep atlas radio observations of the chandra deep field-south/spitzer wide-area infrared extragalactic field. *The Astronomical Journal*, 132(6):2409, 2006.

[8] R. P. Norris, A. M. Hopkins, J. Afonso, S. Brown, J. J. Condon, L. Dunne, I. Feain, R. Hollow, M. Jarvis, M. Johnston-Hollitt, E. Lenc, E. Middelberg, P. Padovani, I. Prandoni, L. Rudnick, N. Seymour, G. Umana, H. Andernach, D. M. Alexander, P. N. Appleton, D. Bacon, J. Banfield, W. Becker, M. J. I. Brown, P. Ciliegi, C. Jackson, S. Eales, A. C. Edge, B. M. Gaensler, G. Giovannini, C. A. Hales, P. Hancock, M. T. Huynh, E. Ibar, R. J. Ivison, R. Kennicutt, A. E. Kimball, A. M. Koekemoer, B. S. Koribalski, Á. R. López-Sánchez, M. Y. Mao, T. Murphy, H. Messias, K. A. Pimbblet, A. Raccanelli, K. E. Randall, T. H. Reiprich, I. G. Roseboom, H. Röttgering, D. J. Saikia, R. G. Sharp, O. B. Slee, I. Smail, M. A. Thompson, J. S. Urquhart, J. V. Wall, and G.-B. Zhao. EMU: Evolutionary Map of the Universe. *PASA*, 28:215–248, Aug. 2011. doi: 10.1071/AS11021.

[9] B. M. Peterson. *An introduction to active galactic nuclei*. Cambridge University Press, 1997.

[10] D. Proctor. Comparing pattern recognition feature sets for sorting triples in the FIRST database. *The Astrophysical Journal Supplement Series*, 165(1):95, 2006.

[11] H. Röttgering, J. Afonso, P. Barthel, F. Batejat, P. Best, A. Bonafede, M. Brüggen, G. Brunetti, K. Chyży, J. Conway, F. D. Gasperin, C. Ferrari, M. Haverkorn, G. Heald, M. Hoeft, N. Jackson, M. Jarvis, L. Ker, M. Lehnert, G. Macario, J. McKean, G. Miley, R. Morganti, T. Oosterloo, E. Orrù, R. Pizzo, D. Rafferty, A. Shulevski, C. Tasse, I. v. Bemmel, B. Tol, R. Weeren, M. Verheijen, G. White, and M. Wise. Lofar and apertif surveys of the radio sky: Probing shocks and magnetic fields in galaxy clusters. *Journal of Astrophysics and Astronomy*, 32(4):557–566, 2012. ISSN 0973-7758. doi: 10.1007/s12036-011-9129-x. URL http://dx.doi.org/10.1007/s12036-011-9129-x.

[12] J. Surace, D. Shupe, F. Fang, C. Lonsdale, E. Gonzalez-Solares, E. Hatziminaoglou11, B. Siana, T. Babbedge, M. Polletta, G. Rodighiero, et al. The SWIRE data release 2: Image atlases and source catalogs for ELAIS-N1, ELAIS-N2, XMM-LSS, and the Lockman hole. *Spitzer Science Centre, California Institute of Technology, Pasadena, CA*, 2005.

[13] S. van Velzen, H. Falcke, and E. Körding. Nature and evolution of powerful radio galaxies at z ~1 and their link with the quasar luminosity function. *Monthly Notices of the Royal Astronomical Society*, 446:2985–3001, Jan. 2015. doi: 10.1093/mnras/stu2213.

[14] R. L. White, R. H. Becker, D. J. Helfand, and M. D. Gregg. A catalog of 1.4 GHz radio sources from the FIRST survey. *The Astrophysical Journal*, 475(2):479, 1997.

[15] E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, S. A. Stanford, M. Cohen, R. G. Walker, J. C. Mather, D. Leisawitz, T. N. Gautier, III, I. McLean, D. Benford, C. J. Lonsdale, A. Blain, B. Mendez, W. R. Irace, V. Duval, F. Liu, D. Royer, I. Heinrichsen, J. Howard, M. Shannon, M. Kendall, A. L. Walsh, M. Larsen, J. G. Cardon, S. Schick, M. Schwalm, M. Abid, B. Fabinsky, L. Naes, and C.-W. Tsai. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *The Astronomical Journal*, 140:1868-1881, Dec. 2010. doi: 10.1088/0004-6256/140/6/1868.

Appendix I

This appendix contains sample outputs from the logistic regression classifier. On the left of each figure is an ATLAS subject, with the infrared image from SWIRE in the background and intensity contours of the ATLAS radio image in the foreground. Candidate hosts are plotted on top of these images, coloured based on the predicted probability that they are the true host (where blue is least likely, and pink is most likely). On the right of each figure is a plot of each candidate's predicted probability with an arbitrary x axis. The candidates have been sorted by increasing probability. This helps to visualise the spread of the probabilities.

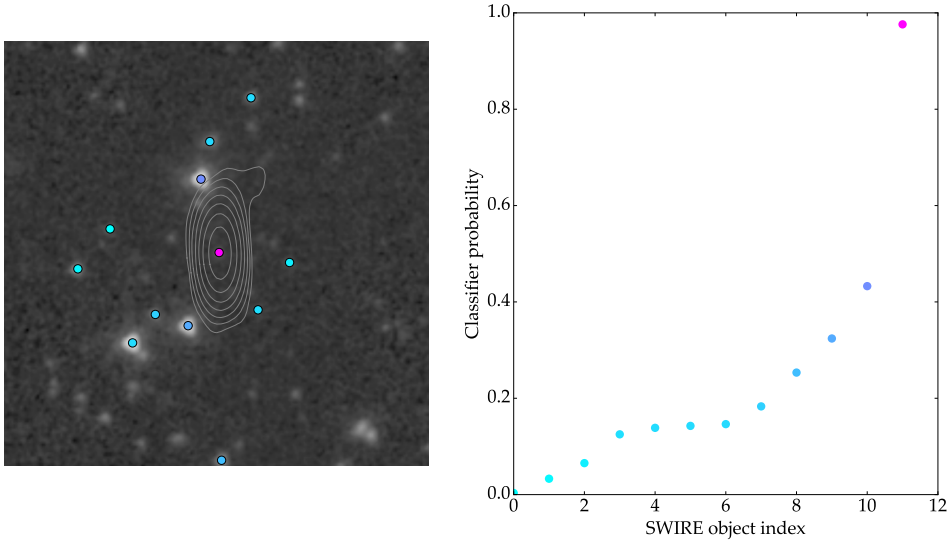


Figure 10: Logistic regression output for ARG0003r2o, an easy-to-classify compact source.

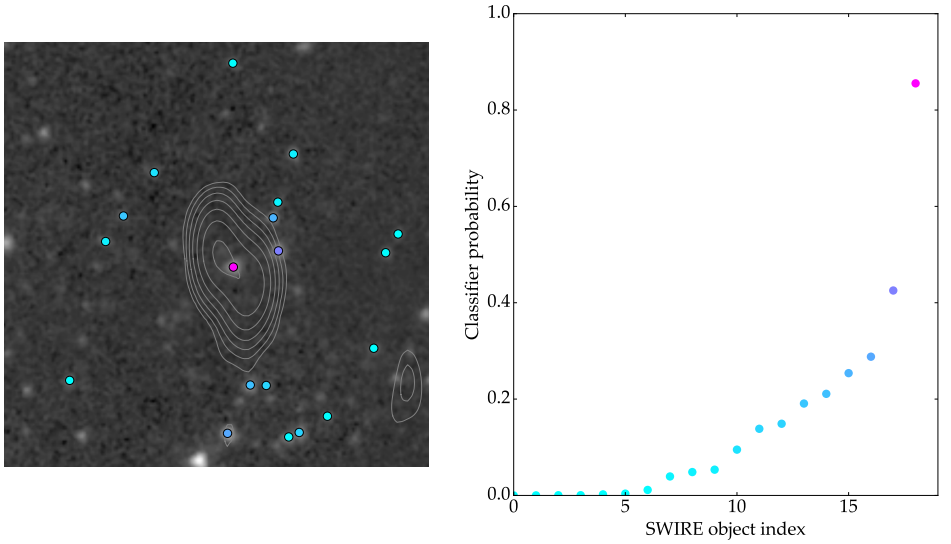


Figure 11: Logistic regression output for ARG0003r2w, an easy-to-classify compact source.

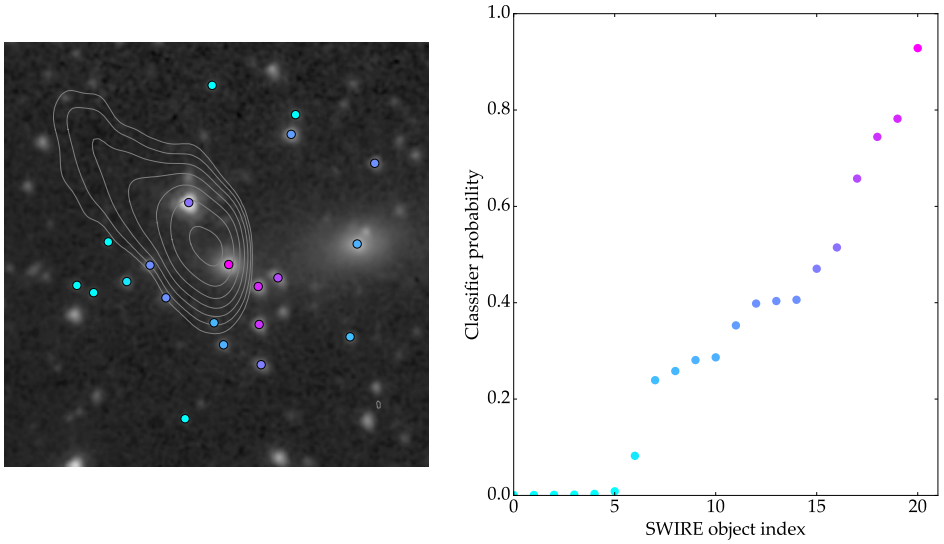


Figure 12: Logistic regression output for ARG0003r25, a complex source.

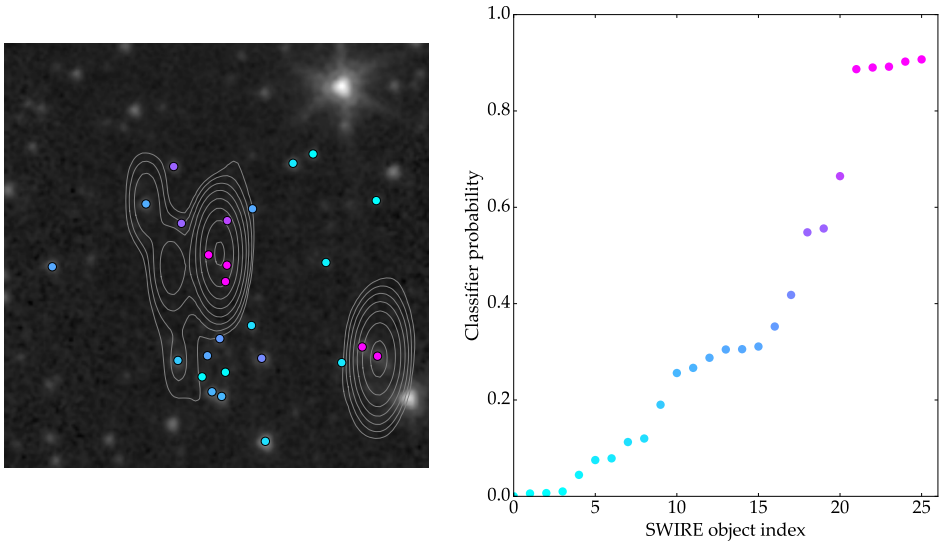


Figure 13: Logistic regression output for ARG0003r1r, consisting of two (possibly three) radio sources. This is hard to classify even for humans, and the difficulty is clear from the plot.

Appendix II

This appendix contains sample outputs from the random forest classifier. On the left of each figure is an ATLAS subject, with the infrared image from SWIRE in the background and intensity contours of the ATLAS radio image in the foreground. Candidate hosts are plotted on top of these images, coloured based on the predicted probability that they are the true host (where blue is least likely, and pink is most likely). On the right of each figure is a plot of each candidate's predicted probability with an arbitrary x axis. The candidates have been sorted by increasing probability.

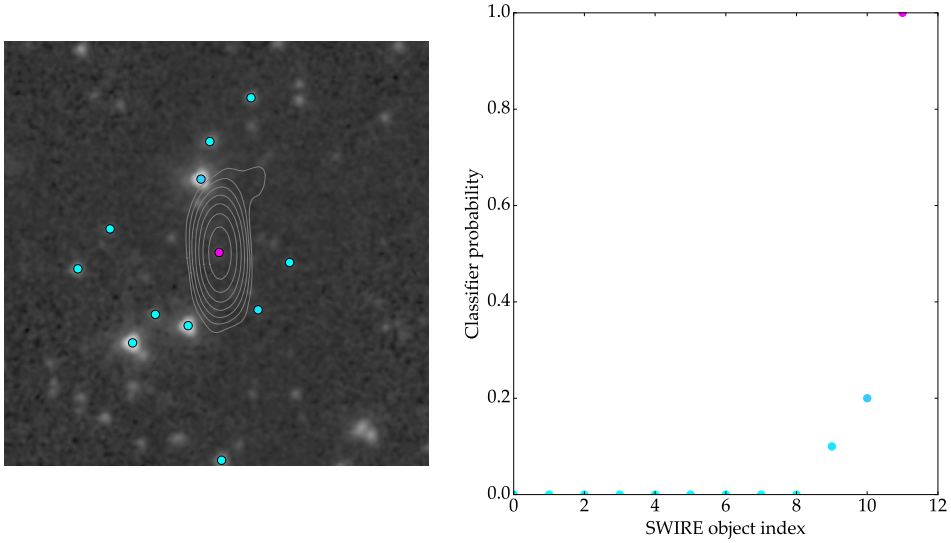


Figure 14: Random forests output for ARG0003r2o, an easy-to-classify compact source.

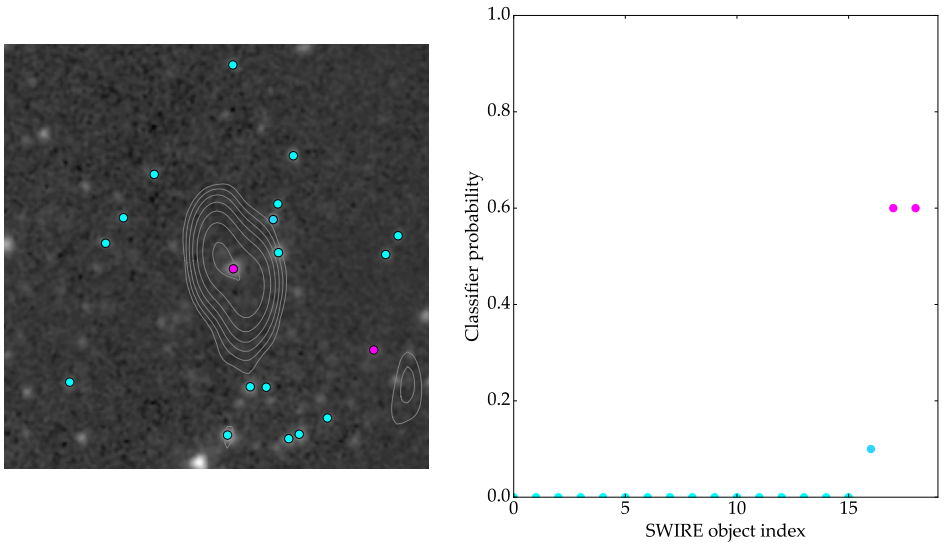


Figure 15: Random forests output for ARG0003r2w, a compact source. Logistic regression classifies this easily, but random forests finds two candidates with equal probability of being the true host.

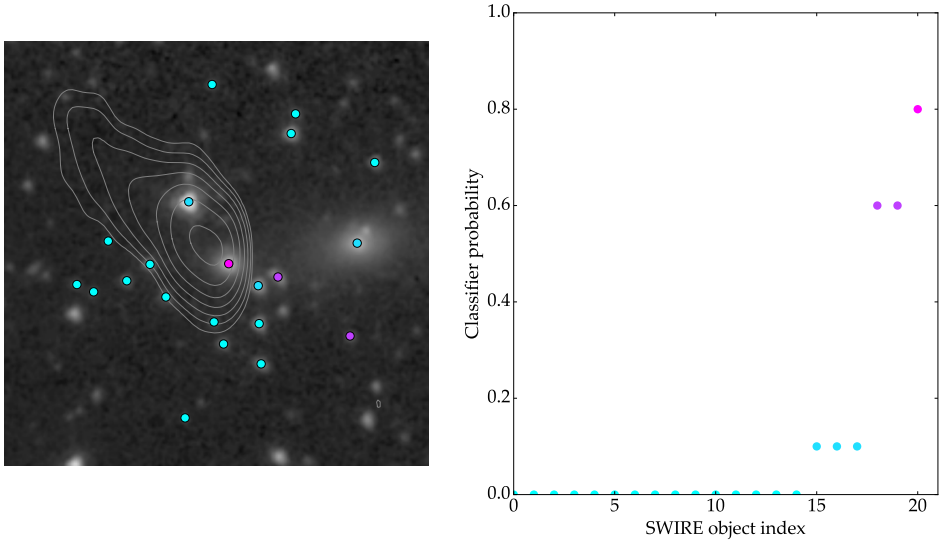


Figure 16: Random forests output for ARG0003r25, a complex source.

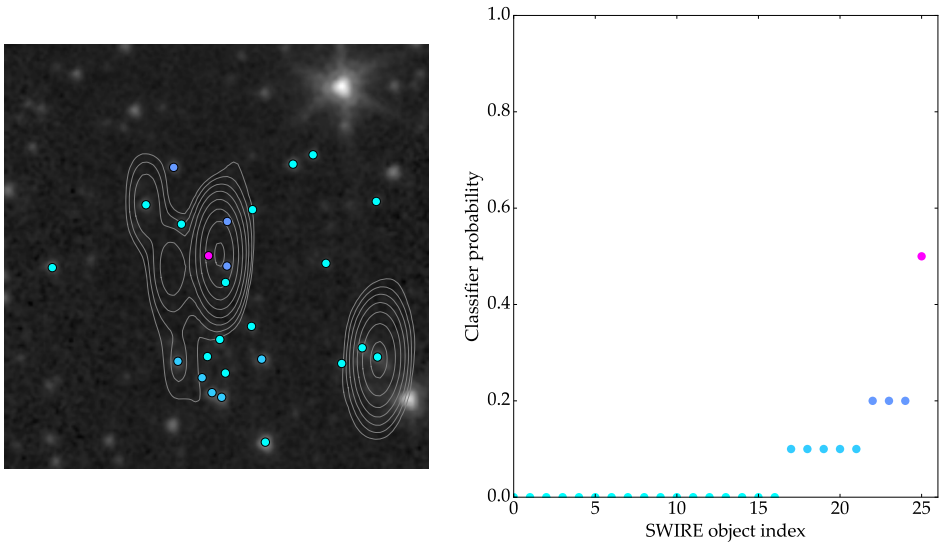


Figure 17: Random forests output for ARG0003r1r, consisting of two (possibly three) radio sources. Random forests does well on this subject despite it being quite complicated.