Wavelet Scattering Networks for Identifying Radio Galaxy Morphologies

Emma Tolley

Abstract – Classifying the morphologies of radio galaxies is important to understand their physical properties and evolutionary histories. A galaxy's morphology is often determined by visual inspection, but as survey size increases, robust automated techniques are needed. Deep neural networks are an attractive method for automated classification but have many free parameters, require extensive training data, and are subject to overfitting and generalization issues. We explore hybrid classification methods using the scattering transform, the recursive wavelet decomposition of an input image. We analyze the performance of the scattering transform for the Fanaroff–Riley classification of radio galaxies for convolutional neural networks (CNNs) and other machine learning algorithms. We test the robustness of the different classification methods with training data truncation and noise injection and find that the scattering transform can offer competitive performance with the most accurate CNNs.

1. Introduction

New facilities for radio interferometry such as the Low-Frequency Array [1], the Murchison Widefield Array [2], the MeerKAT telescope [3], and the Australian Square Kilometre Array Pathfinder telescope [4] produce petabytes of data per year. This data flow rate will only increase; the next-generation Square Kilometre Array Observatory is expected to produce 600 PB of calibrated data products per year [5]. In this era of data-intensive astronomy, advances in statistics, computer science, and machine learning are beginning to be used in investigations of astronomical data [6, 7].

Deep neural networks (DNNs) have been extensively used for a wide range of classification tasks in astronomy, including classifying radio galaxy morphologies [8–15]. However, DNNs face problems in *generalizability* (the ability of the network to generalize to new data sets or domains) and *interpretability* (the ability for users to understand the decisions of the network) [16–18]. DNNs must be trained on large data sets that represent the entire domain of each class of objects. However, radio astronomy data sets are often small, with only a few examples of morphologically diverse objects. Regularization techniques such as data augmentation can help with this issue, but out-of-domain generalizability [19] remains an unsolved problem.

It may also be possible to improve generalizability and interpretability by hard coding feature extraction techniques into the network structure, minimizing the amount of trainable network parameters. This has been explored for radio galaxy data with group equivariant convolutional neural networks (CNNs) [11], attention gating [12], Combination Of Shifted FIlter REsponses (COSFIRE) filters [13], and rotational standardization [14]. Of particular interest is the scattering transform [20] that embeds wavelet transforms into a recursive structure similar to a CNN. These wavelet scattering networks are computationally efficient [21] and demonstrated to be competitive with traditional neural networks for measuring cosmological parameters [22] but have not yet been evaluated for radio galaxy classification.

1.1 The Scattering Transform

The scattering transform is the convolution of an input image $I(x)$ by a series of rotated and dilated wavelet filters. For a complete discussion of the scattering transform and its properties, we refer the reader to [20, 23–26].

In this study, we construct the scattering transform using the Morlet wavelet $\psi(x) \in \mathbb{R}^2$, which is the product of a Gaussian envelope with a sinusoidal wave. We define filters $\psi_{j,l}(x)$ by rotating $\psi(x)$ with L rotations $r_l = 2\pi l/L$ in $\mathbb{R}^2$ and by dilating it with J scales $2^j > 1$:

$$\psi_{j,l}(x) = 2^{-2j}\psi(2^{-j}r_l^{-1}x) \tag{1}$$

Each wavelet filter measures variations of scale 2^j in the direction given by r_l. The wavelet transform is stable and invertible if $\psi_{j,l}$ satisfies a Littlewood–Paley condition; this requires an additional convolution with a low-pass filter $\phi(x) \in \mathbb{R}^2$, for which we use a 2D Gaussian function. The scattering transform coefficients are then constructed by convolving the wavelet filters with the input image $I(x)$. To second order, the coefficients are

$$S_0(x) = (I \star \phi)(x) \tag{2}$$

$$S_1(j_1, l_1)(x) = (|\, I \star \psi_{j_1,l_1} \,| \star \phi)(x) \tag{3}$$

$$S_2(j_1, l_1, j_2, l_2)(x) = (|\, |\, I \star \psi_{j_1,l_1} \,| \star \psi_{j_2,l_2} \,| \star \phi)(x) \tag{4}$$

These scattering transform coefficients provide a hierarchical multiscale approach to examining $I(x)$ by

Manuscript received 23 December 2023.

Emma Tolley is with Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne, Observatoire de Sauverny, Chemin Pegasi 51, 1290 Versoix, Switzerland; e-mail: emma.tolley@epfl.ch.

progressively identifying the couplings between structures at different scales. In the context of neural networks, the wavelet transform acts as a CNN, where the filters are predefined.

2. Radio Galaxy Classification Scheme

We use the Fanaroff–Riley classification of radio galaxies [27] to evaluate the performance of the scattering transform. This binary classification scheme distinguishes radio galaxies with active nuclei on the basis of the radio luminosity: *FRI* galaxies have a core-brightened structure with steep spectra, whereas *FRII* galaxies are edge brightened, with bright hot spots at the ends of the lobes. These two diverse morphologies arise from jet interactions with surrounding environments of different densities. Although there are also proposals for more semantic taxonomy of radio galaxies [28], the Fanaroff–Riley classification provides a robust benchmark for evaluating classification methods.

Two different overlapping radio galaxy data sets are used for this study. The first data set of radio galaxies is a collection and combination of several catalogs using the Very Large Array Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) survey [29]. It contains galaxies identified as FRI, FRII, compact, or bent sources, shown in Figure 1. For performance, we trim the 300 pixel $\times$ 300 pixel images to 60 pixel $\times$ 60 pixel images. The second data set we use is a collection of radio-loud active galactic nuclei created from the NRAO VLA Sky Survey and Faint Images of the Radio Sky at Twenty-cm (FIRST) surveys called MiraBest [30]. In this data set, we only use confidently classified FRI and FRII galaxies, and we keep the original 150 pixel $\times$ 150 pixel image size. For both data sets, we reserve approximately 50 instances of each class for separate validation and test data sets. The remaining data are used for training data, which we augment with 90 ° rotations and reflections of input images. The number of samples for each data set are shown in Table 1. We also evaluate performance of our classification schemes against the MNIST [31] handwritten digits data set, a common benchmark in machine learning.

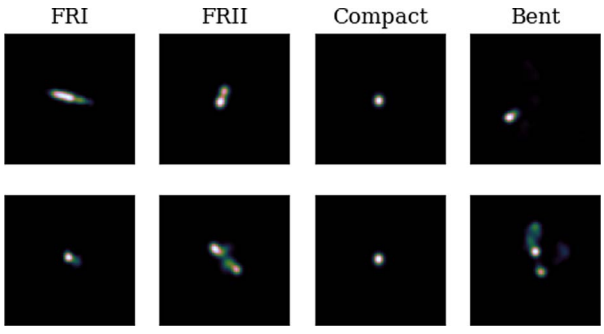

Figure 1. Examples of the four radio galaxy classes from the FIRST data set.

Table 1. Class sizes of the radio galaxy data sets

Class	Training data size	Augmented training data size	Validation data size	Test data size
FIRST data set				
FRI	395	3160	50	50
FRII	824	6592	50	50
Compact	291	2328	50	50
Bent	248	1984	50	50
MiraBest data set				
FRI	298	2384	50	49
FRII	327	2616	54	55

3. Methods

We use the Kymatio library [32] to construct the scattering transform with Morlet wavelets dilated with $J = 3$ spatial scales and $L = 8$ angular orientations. We test three different iterations of the scattering transform. First, we use the second-order scattering transform, including all the coefficients as defined in (2), (3), and (5). To evaluate the effect of the higher order coefficients, we also evaluate a scattering transform that only uses the zeroth- and first-order coefficients [(2) and (3)] and thus only outputs the modulus of coefficients of the wavelet transform. Finally, we construct a reduced scattering transform that together averages the different angular modes of the first- and second-order coefficients. For these three schemes, the scattering coefficients are input to two different classification methods.

We use linear kernel support vector classification (SVC) [33, 34] with the default hyperparameters defined in scikit-learn version 0.24.1. SVC is a simple machine learning algorithm that partitions p-dimensional input data into different classes with $(p-1)$ dimensional hyperplanes. SVC can operate directly on the radio galaxy images with no feature extraction, in which case $p = N_{\text{pixel}}$ or directly on the scattering transform coefficients.

We also define a *classifier DNN* network architecture with two hidden layers with 120 and 84 nodes, rectified linear unit (ReLU) activation functions [35], and 50% dropout [36] between last hidden layer and output layer. The final output layer is changed, depending on the data set used. For the multiclass classification with the FIRST (MNIST) data set, we use four (10) output nodes with the softmax activation function [37]. For binary classification with the MiraBest data set, we use a single output node with a sigmoid activation function [38]. The network architecture is shown in Table 2 (lines five to eight). The network is trained to classify morphologies using the scattering transform coefficients as input.

Finally, we develop two complete feature extraction and classification DNN pipelines. The architecture of *CNN1* is inspired by the standard network in [11, 39]. An overview of the network architecture is shown in Table 2. We also define a second neural network called *CNN2* by replacing the global average pooling layer with a second max pooling layer. In total CNN1 has 15,116 free parameters, while CNN2 has 289,676 free parameters. All these implementations and the

Table 2. CNN1 architecture

No.	Layer[a]	Description	Activation function
1	Conv2D	16 5×5 filters	ReLU
2	MaxPooling2D	2×2 pool size	
3	Conv2D	16 5×5 filters	ReLU
4	GlobalAveragePooling2D		
5	Dense	120 nodes	ReLU
6	Dense	84 nodes	ReLU
7	Dropout	50% dropout rate	
8	Dense	1/4/10 nodes	Sigmoid/softmax

[a] Conv2D: 2D convolution layer; MaxPooling2D: max pooling operation for 2D spatial data; GlobalAveragePooling2D: global average pooling operation for 2D data; Dense: densely-connected NN layer; Dropout: applies dropout to the previous layer.

corresponding training scripts are publicly available on GitHub (https://github.com/epfl-radio-astro/scatternet).

4. Results

We fit each classification method to the training data sets and evaluate performance on the blinded test data sets. The validation data set is used to evaluate the performance of DNN components during training. We use the Adam optimizer [40] with binary cross-entropy loss when training on MiraBest data and categorical cross-entropy loss when training on the FIRST and MNIST data sets. We use sample weights in the loss function to account for class imbalance [41] in the training data sets. DNNs are fit for 50 epochs with early stopping on the basis of the average of recall obtained on each class ("balanced accuracy") for the validation data set.

We show classification performance for the FIRST data set in Figure 2 for the pure DNN pipelines CNN1 and CNN2, as well as the hybrid scattering transform + classifier methods Scattering2D+SVC and Scattering2D+NN. Scattering2D+NN and CNN2 show the best performance, with identical balanced accuracy scores. For the MiraBest data set, we find that the balanced accuracy of the Scattering2D+NN method is 92%, slightly better than the CNN2 balanced accuracy of 91%.

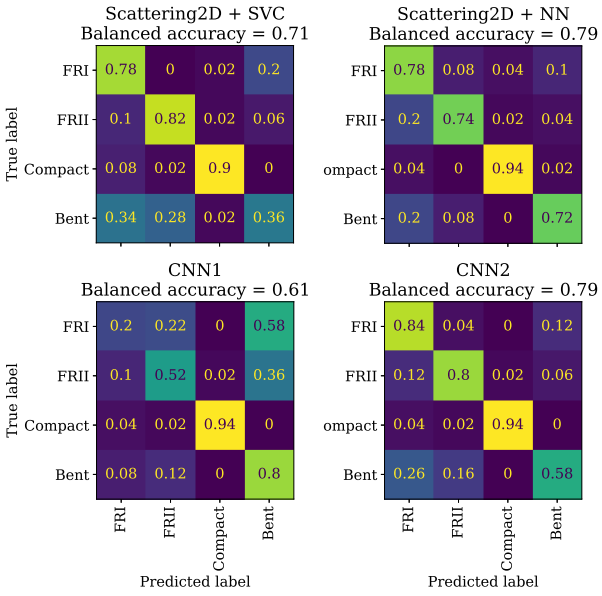

Figure 2. Confusion matrix results for the neural network and second-order scattering transform classification methods on the FIRST data set. The confusion matrix is normalized over the true conditions.

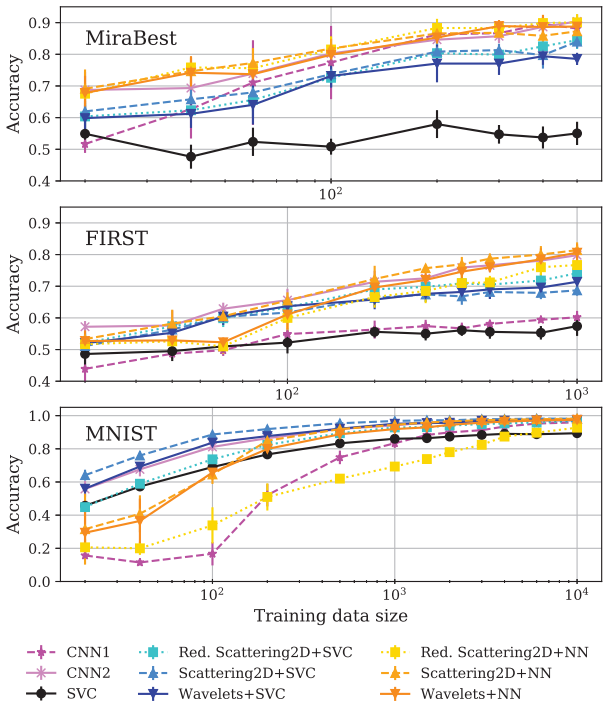

Figure 3. Training data truncation classification results. The x-axis shows the total size of the training data (cumulative for all classes) before applying data augmentation. The y-axis shows the balanced class accuracy. Error bars represent the standard deviation of five trials.

However, comparing performance on a single data set is not enough to demonstrate the robustness of a given classification technique. We test the generalizability of the different methods by truncating the full training data sets. In this test, data augmentation is only applied to the truncated radio galaxy training data after the training data have been truncated, allowing us to evaluate how a network trained on tens or hundreds of training examples might generalize to the test and validation data. We do not apply data augmentation to the MNIST data, as the classes are not invariant under rotations or reflections.

The results of the truncation tests on the radio galaxy data sets are shown in the top two panels of Figure 3. We see the solutions that combine the scattering transform with a DNN classifier, and the larger neural network CNN2 consistently performs the best for both radio galaxy data sets. In addition, we observe that the second-order scattering transform consistently outperforms the first-order scattering transform, demonstrating the added value of using higher order coefficients. For the FIRST data set, the Scattering2D+NN method consistently has the best accuracy. The smaller neural network CNN1 struggles with the multiclass classification task, even when trained on the full $\mathcal{O}(1000)$ size training data.

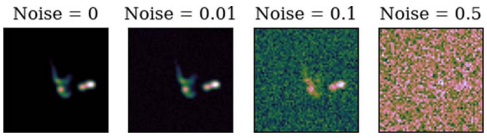

Figure 4. Example of the noise injection levels for a single FIRST radio galaxy. The noise parameter corresponds to the σ value of the Gaussian-injected noise, relative to the maximum pixel value of the image.

However, the performance of these different methods is quite different for the MNIST handwritten digits data set. Here, the second-order scattering transform combined with the SVC classifier is consistently the best method, outperforming even CNN2.

We also tested the robustness of the different techniques with noise injection. We modified the test data by injecting $\sigma = 1\%$, 10%, or 50% Gaussian noise to each image. An example of the noise injection is shown in Figure 4. Each classification method is fit to the original training data and evaluated on the different noise-injected test data, with results shown in Figure 5. For the MiraBest data set, again we see that CNN1, CNN2, and the hybrid scattering transform and DNN solutions have similar performance. On the FIRST data set, none of the classifiers are particularly stable against noise injection, even with 1% injected noise. Surprisingly, the performance of the simple SVC classifier is the most robust to noise injection across both radio galaxy data sets. On the MNIST data set, methods that use DNNs are more robust to noise injection compared with SVC classifiers, an inversion of the truncation test.

5. Discussion and Conclusions

We have shown that the scattering transform can be effectively used for identifying galactic morphology and is competitive with the best-performing CNN in this work. We have also shown that including second-order coefficients improves feature extraction and classification accuracy. We find that our Scattering2D+NN architecture has an accuracy of 79% on the FIRST data set, with similar performance to the CNN (79%) and vision transformer (81%) classifiers from [15]. On the MiraBest data set, the Scattering2D+NN architecture has an accuracy of 92%, comparing favorably to the attention-gated CNN (92%) in [12] but not outperforming the standard CNN (94%) or group equivariant CNNs from [11]. Using more comprehensive data augmentation as in [11, 12] may improve our results, which seem to be limited by training data size, as shown in Figure 3.

We observe strikingly different performance results when evaluating the classification methods on our radio galaxy data sets versus the MNIST data set. The radio galaxy data propose a much more difficult classification task compared with handwritten digits: galaxy classes are more morphologically diverse, galaxy images are inherently noisier and subject to misclassification error, and contain inherent rotational and translational symmetries that must be learned by any classification method.

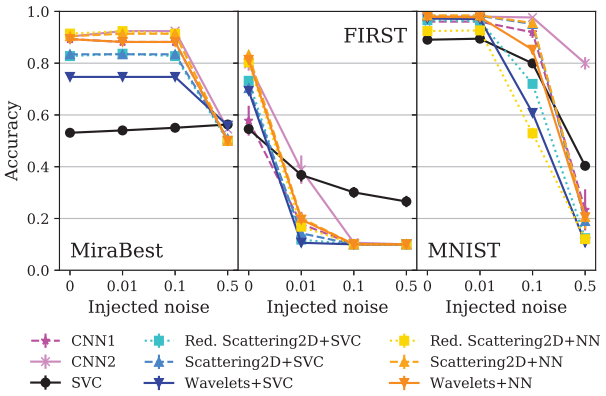

Figure 5. Noise injection test classification results. The y-axis shows the balanced class accuracy.

We do not find that the scattering transform improves classifier performance in the data truncation or noise injection generalizability tests. Nevertheless, inclusion of the scattering transform improves network interpretability and drastically reduces the number of trainable parameters, improving the computational and memory requirements. CNN2 and Scattering2D+NN have the same balanced accuracy when evaluated on the FIRST data set, but Scattering2D+NN uses $\sim 11,000$ parameters, whereas CNN2 uses $\sim 290,000$ parameters.

We expect that dedicated hyperparameter optimization on each data set may improve the performance of the classifiers. Because the scattering coefficients are highly correlated, an additional processing step using principle component analysis to reduce dimensionality may further improve performance.

6. References

1. M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, et al., "LOFAR: The LOw-Frequency ARray," *Astronomy & Astrophysics*, **556**, X, July 2013, p. A2.
2. A. P. Beardsley, M. Johnston-Hollitt, C. M. Trott, et al., "Science with the Murchison Widefield Array: Phase I results and Phase II opportunities," *Publications of the Astronomical Society of Australia*, **36**, 2019, p. e050, doi:10.1017/pasa.2019.41.
3. M. Jarvis, "The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey," *MeerKAT Science: On the Pathway to the SKA*, 2016, doi:10.22323/1.277.0006.
4. S. Johnston, R. Taylor, M. Bailes, N. Bartel, C. Baugh, et al., "Science With ASKAP. The Australian Square-Kilometre-Array Pathfinder," *Experimental Astronomy*, **22**, 3, December 2008, p. 151-273.
5. Square Kilometre Array Organization, *SKA1 Scheduling and Archive Constraints*, SKA-TEL-SKO-0000951, United Kingdom, March 2019.
6. K. D. Borne, "Astroinformatics: Data-Oriented Astronomy Research and Education," *Earth Science Informatics*, **3**, 5, May 2010, pp. 5-17.
7. E. D. Feigelson and G. J. Babu, "Statistical Methods for Astronomy," in T. D. Oswalt and H. E. Bond (eds.), *Planets, Stars and Stellar Systems. Volume 2: Astronomical*

Techniques, Software and Data, Springer, Dordrecht, 2013, doi:10.1007/978-94-007-5618-2_10.
8. A. K. Aniyan and K. Thorat, "Classifying Radio Galaxies With the Convolutional Neural Network," *The Astrophysical Journal Supplement Series*, **230**, 2, June 2017, p. 20.
9. E. Tolley, D. Korber, A. Galan, A. Peel, M. T. Sargent, et al., "Lightweight HI source finding for next generation radio surveys," *Astronomy and Computing*, **41**, 2022, p. 10063.
10. X. Rustige, et al., *RASTI* (2023).
11. A. M. M. Scaife and F. Porter, "Fanaroff-Riley Classification of Radio Galaxies Using Group-Equivariant Convolutional Neural Networks," *Monthly Notices of the Royal Astronomical Society*, **503**, 2, May 2021, pp. 2369-2379.
12. M. Bowles, A. M. M. Scaife, F. Porter, H. Tang, and D. J. Bastien, "Attention-Gating for Improved Radio Galaxy Classification," *Monthly Notices of the Royal Astronomical Society,* **501**, 3 March 2021, pp. 4579-4595.
13. S. Ndung'u, T. Grobler, S. J. Wijnholds, D. Karastoyanova, and G. Azzopardi, "Classification of radio galaxies with trainable COSFIRE filters," *Monthly Notices of the Royal Astronomical Society*, **530**, 1, May 2024, pp. 783-794, https://doi.org/10.1093/mnras/stae821.
14. K. Brand, T. L. Grobler, W. Kleynhans, M. Vaccari, M. Prescott, et al., "Feature Guided Training and Rotational Standardization for the Morphological Classification of Radio Galaxies," *Monthly Notices of the Royal Astronomical Society*, **522**, 1, June 2023, pp 292-311.
15. L. Rustige, J. Kummer, F. Griese, K. Borras, M. Bruggen, et al., "Morphological Classification of Radio Galaxies With Wasserstein Generative Adversarial Network-Supported Augmentation," *RAS Techniques and Instruments*, **2**, 1, January 2023, pp. 264-277.
16. I. J. Goodfellow, J. Shlens, and C. Szegedy, "*Explaining and Harnessing Adversarial Examples*," International Conference on Learning Representations, May 7-9, 2015, San Diego.
17. A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 427-436, doi:10.1109/CVPR.2015.7298640.
18. T. Pang, C. Du, Y. Dong, and J. Zhu, "Towards Robust Detection of Adversarial Examples," in S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, New York, Curran Associates, Inc., 2018, Volume 31, https://proceedings.neurips.cc/paper_files/paper/2018/file/e0f7a4d0ef9b84b83b693bbf3feb8e6e-Paper.pdf.
19. Y. Balaji, S. Sankaranarayanan, and R. Chellappa, "Towards domain generalization using meta-regularization," Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, 2018, pp. 1006-1016.
20. S. Mallat, "Group Invariant Scattering," *Communications on Pure and Applied Mathematics*, **65**, 10, October 2012, pp. 1331-1398.
21. Z. Baharlouei, H. Rabbani, and G. Plonka, "Wavelet Scattering Transform Application in Classification of Retinal Abnormalities Using OCT Images," *Scientific Reports*, **13**, 1, November 2023, p. 19013.
22. S. Cheng, Y. -S. Ting, B. Ménard, and J. Bruna, "A new approach to observational cosmology using the scattering transform," *Monthly Notices of the Royal Astronomical Society,* **499**, 4, December 2020, pp. 5902-5914, doi:10.1093/mnras/staa3165.
23. J. Bruna and S. Mallet, "Invariant Scattering Convolution Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 8, August 2013, pp. 1872-1886.
24. J. Andén and S. Mallet, "Deep Scattering Spectrum," *IEEE Transactions on Signal Processing*, **62**, 16, May 2014, pp. 4114-4128.
25. S. Cheng and B. Ménard, How to quantify fields or textures? A guide to the scattering transform, arXiv:2112.01288, November, 2021.
26. S. Cheng, R. Morel, E. Allys, B. Ménard, and S. Mallat, "Scattering spectra models for physics," *PNAS Nexus*, **3**, 4, April 2024, p. 103, doi:10.1093/pnasnexus/pgae103.
27. B. L. Fanaroff and J. M. Riley, "The Morphology of Extragalactic Radio Sources of High and Low Luminosity," *Monthly Notices of the Royal Astronomical Society*, **167**, May 1974, pp. 31-36.
28. M. Bowles, H. Tang, E. Vardoulaki, E. L. Alexander, Y. Luo, et al., "Radio Galaxy Zoo EMU: Towards a Semantic Radio Galaxy Morphology Taxonomy," *Monthly Notices of the Royal Astronomical Society*, **522**, 2, June 2023, pp. 2584-2600.
29. F. Greise, J. Kummer, P. L. S. Connor, M. Brüggen, and L. Rustige, "First Radio Galaxy Data Set Containing Curated Labels of Classes FRI, FRII, Compact and Bent" *Data in Brief* **47**, X, April 2023, p. 108974.
30. F. A. M. Porter and A. M. M. Scaife, "MiraBest: A Data Set of Morphologically Classified Radio Galaxies for Machine Learning," *RAS Techniques and Instruments*, **2**, 1, January 2023, pp. 293-306.
31. L. Deng, "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]," *IEEE Signal Processing Magazine*, **29**, 6, October 2012, pp. 141-142.
32. M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, and G. Rochette, et al., "Kymatio: Scattering Transforms in Python," *Journal of Machine Learning Research*, **21**, 60, 2020, pp. 1-6.
33. B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers" Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, Pennsylvania, USA, 1992, pp. 144-152.
34. R. -E. Fan, K. -W. Chang, C. -J. Hsieh, X. -R. Wang, and C. -J. Lin, et al., "LIBLINEAR: A Library for Large Linear Classification," *Journal of Machine Learning Research*, **9**, X, August 2008, pp. 1871-1874.
35. A. F. Agarap, Deep Learning using Rectified Linear Units (ReLU), arXiv:1803.08375, 2019.
36. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks From Overfitting, *Journal of Machine Learning Research*, **15**, 56, June 2014, pp. 1929-1958.
37. I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA, USA, MIT Press, 2016.
38. S. Narayan, "The Generalized Sigmoid Activation Function: Competitive Supervised Learning," *Information Sciences*, **99**, 1–2, June 1997, pp. 69-82.
39. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, **521**, 7553, May 2015, pp. 436-444.
40. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 3rd International Conference for Learning Representations, San Diego, 2015.
41. Y. Cui, M. Jia, T. -Y. Lin, Y. Song, and S. Belongie, "Class-Balanced Loss Based on Effective Number of Samples," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 9260-9269, doi:10.1109/CVPR.2019.00949.