RGC-BENT: A NOVEL DATASET FOR BENT RADIO GALAXY CLASSIFICATION

Mir Sazzat Hossain[1,⋆] *K. M. B. Asad*[2,†] *Payaswini Saikia*[3,†] *Adrita Khan*[1]
Md Akil Raihan Iftee[1] *Rakibul Hasan Rajib*[1] *Arshad Momen*[1] *Jewel Kumar Ghosh*[1]
Md Ashraful Amin[1] *Amin Ahsan Ali*[1] *AKM Mahbubur Rahman*[1]

[1] Center for Computational & Data Sciences, Independent University, Bangladesh
[2] Center for Astronomy, Space Science and Astrophysics, Independent University, Bangladesh
[3] Center for Astrophysics and Space Science, New York University Abu Dhabi

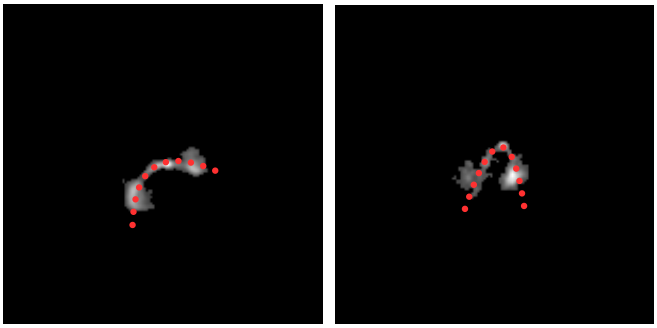
ABSTRACT

We introduce a novel machine learning dataset tailored for the classification of bent radio active galactic nuclei (AGN) in astronomical observations. Bent radio AGN, distinguished by their curved jet structures, provide critical insights into galaxy cluster dynamics, interactions within the intracluster medium, and the broader physics of AGN. Despite their astrophysical significance, the classification of bent AGN remains a challenge due to the scarcity of specialized datasets and benchmarks. To address this, we present a dataset derived from a well recognized radio astronomy survey, designed to support the classification of NAT (Narrow-Angle Tail) and WAT (Wide-Angle Tail) categories, along with detailed data processing steps. We further evaluate the performance of state-of-the-art deep learning models on the dataset, including Convolutional Neural Networks (CNNs) and transformer-based architectures. Our results demonstrate the effectiveness of advanced machine learning models in classifying bent radio AGN, with ConvNeXT achieving the highest F1-scores for both NAT and WAT sources. By sharing this dataset and benchmarks, we aim to facilitate the advancement of research in bent AGN classification, AGN and cluster environments, galaxy evolution, and more.

***Index Terms*—** Bent Active Galactic Nuclei (AGN), Astronomical Image Classification, Astrophysical Data Analysis, Radio Galaxy Morphology, Radio Galaxy Classification

1. INTRODUCTION

Active Galactic Nuclei (AGNs) are some of the most luminous objects in the universe, powered by the accretion of matter onto supermassive black holes. The classification of AGN is important for understanding the evolution of galaxies in clusters and the growth of supermassive black holes [1, 2]. Bent radio AGN (sometimes also called bent radio galaxies, head-tail galaxies, or bent-tailed galaxies) are subsets of radio-loud AGN, characterized by distinctly curved jet structures resulting from interactions with the intergalactic medium [3]. These structures are typically observed in galaxy clusters, where the interaction between the galaxy's jets and the intracluster medium (ICM) plays a significant role in shaping their appearance. The jet's bending is typically linked to the relative motion between the host galaxy and the surrounding ICM, which results in ram pressure distorting the jets' trajectories [4].



(a) WAT (b) NAT

Fig. 1: Examples of Wide-Angle Tailed (WAT) and Narrow-Angle Tailed (NAT) radio galaxies. (a) WAT galaxies exhibit a wider 'C' shaped structure with jets extending over 90 degrees, while (b) NAT galaxies show a narrower 'V' shape with jets within a 90-degree angle.

Two notable subtypes are identified within this category based on the angle between their jets: Wide-Angle Tailed (WAT) Galaxies and Narrow-Angle Tailed (NAT) Galaxies, as shown in Figure 1. WAT galaxies feature two-sided jets that expand outward, creating a wide 'C' shape with an opening angle exceeding 90 degrees, usually associated with massive elliptical galaxies [5]. As the galaxy moves through the dense intracluster medium, interactions with it are believed to cause the bending of their jets. In contrast, NAT galaxies, also called head-tail galaxies, have closely aligned jets forming a

⋆ Corresponding author: mirsazzathossain@gmail.com
† Expert Annotator.

narrow 'V' shape with an opening angle under 90 degrees and are usually located on the outskirts of galaxy clusters, moving at high velocities through the intracluster medium, where ram pressure bends their jets [5, 6]. The study of bent radio AGN, including WATs and NATs, is crucial for understanding the dynamics of galaxy clusters, the behavior of the intracluster medium, the interactions between galaxies and their environments, and the dynamics of AGN jets and their surrounding conditions [7].

In this work, we make the following contributions:

- We present a dataset of bent radio AGN images, specifically compiled to facilitate machine learning-based classification.
- We evaluate the performance of five machine learning models on this dataset, establishing a baseline for automated AGN classification.
- We open source all code for key data preprocessing tasks, including background estimation, source identification, mask generation, and background removal via masking.

2. EXISTING DATASETS AND LIMITATIONS

Machine learning has been widely applied in astronomy, but the lack of specialized datasets remains a challenge. Existing datasets like Galaxy Zoo [8] focus on general galaxy classification and do not include specific radio morphologies such as Fanaroff-Riley (FR) types [9] or bent AGNs. The MiraBest dataset [10] provides labeled FR-I and FR-II sources but lacks bent AGNs, which have different morphological features.

While no dedicated machine learning dataset exist for bent radio AGNs, several catalogs provide relevant information. The Proctor et al. [11] catalog was among the first, but it contained automated annotations which cannot be relied upon for high-precision tasks without manual verification. In contrast, Sasmal et al. [12] introduced a manually curated catalog, offering more accurate source identification. This catalog serves as the primary data source for our dataset creation. Converting it into a machine learning-ready dataset requires careful data acquisition, preprocessing, and expert validation. In this paper, we document this process in detail to facilitate future research in bent radio AGN classification.

3. DATASET CREATION

3.1. Data Acquisition

The raw data is obtained from the Faint Images of the Radio Sky at Twenty-cm (FIRST) survey using source coordinates from the catalog proposed by Sasmal et.al. [12]. The FIRST survey, conducted at 1.4 GHz with the Karl G. Jansky Very Large Array (VLA) in New Mexico, USA, covers the sky north of $\delta = -40^\circ$ with a 5-arcsecond resolution and a 0.15 mJy sensitivity.

The data is stored in Flexible Image Transport System (FITS) format and includes flux density measurements at different frequencies. We retrieve the images from NASA SkyView[1], a virtual observatory providing access to multi-survey astronomical data. Using the provided source coordinates, we query SkyView to download the corresponding FIRST survey radio images, which are then stored in FITS format for subsequent processing.

3.2. Data Processing

The first few steps (up to and including Section 3.2.3) of data processing were performed using the Python package `PyBDSF` [13].

3.2.1. Background Estimation

The first step in processing is estimating the background noise level for each image. We compute the background level and its standard deviation by segmenting the image into smaller subregions and calculating the local mean (μ_{local}) and standard deviation (σ_{local}) of the pixel intensities within each subregion. The background estimation is modeled as:

$$B(x, y) = \mu_{\text{local}} + k \cdot \sigma_{\text{local}}, \tag{1}$$

where $B(x, y)$ denotes the local background intensity, and k is a scaling factor that accounts for typical noise fluctuations.

3.2.2. Source Identification

After background estimation, sources are identified by detecting contiguous regions (or "islands") where the intensity significantly exceeds the local background. Two thresholds are applied during this process:

- **Island Threshold (T_{isl}):** This parameter, typically set to 3, defines the minimum signal level for considering a group of contiguous pixels as part of an island. A pixel is included in an island if:

  $$I(x, y) > B(x, y) + T_{\text{isl}} \cdot \sigma_{\text{local}},$$

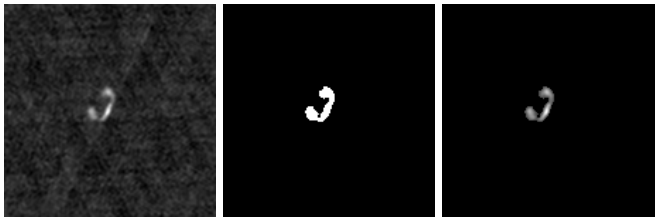
  helping to delineate regions where emission is likely associated with a source.

- **Peak Threshold (T_{pix}):** While T_{isl} defines the overall extent of an island, T_{pix}, typically set to 5, establishes the detection threshold for source peaks. Only islands with peaks exceeding:

  $$I(x, y) > B(x, y) + T_{\text{pix}} \cdot \sigma_{\text{local}},$$

  are considered valid detections.

[1] `https://skyview.gsfc.nasa.gov`



(a) Original Image (b) Generated Mask (c) Masked Image

Fig. 2: Masking process overview. (a) Original image, (b) generated mask, and (c) resulting (final) masked image.

After thresholding, contiguous pixels satisfying these criteria are grouped into islands. For each island, a two-dimensional Gaussian model is fitted to extract key source parameters, including the centroid (x_0, y_0), peak intensity I_0, and morphological descriptors (major axis σ_{maj}, minor axis σ_{min}, and position angle θ). The Gaussian model is given by:

$$I(x,y) = I_0 \exp\left(-\frac{(x'-x_0)^2}{2\sigma_{\text{maj}}^2} - \frac{(y'-y_0)^2}{2\sigma_{\text{min}}^2}\right), \quad (2)$$

where (x', y') are the coordinates rotated by the angle θ. This model effectively characterizes the morphology of each source and helps distinguish genuine sources from noise artifacts.

Table 1: Distribution of NAT and WAT sources across training and test batches. The first nine batches (576 images) form the training set, while the last batch (63 images) serves as the test set.

Batch No.	Source Count		Total Sources	Cumulative Total (Train/Test)
	WAT	NAT		
Training Set				
0	39	25	64	
1	39	25	64	
2	39	25	64	
3	39	25	64	
4	39	25	64	576
5	38	26	64	
6	38	26	64	
7	38	26	64	
8	38	26	64	
Test Set				
9	38	25	63	63
Total	385	254	639	639

3.2.3. Mask Generation

Once sources are detected and characterized, a binary mask is generated to delineate the spatial extent of each source. The mask $M(x, y)$ is produced by thresholding the fitted Gaussian model:

$$M(x,y) = \begin{cases} 1, & \text{if } I(x,y) > B(x,y) + T_{\text{pix}} \cdot \sigma_{\text{local}}, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

This criterion ensures that only regions with intensities significantly above the background are included in the mask. For sources exhibiting extended or diffuse emission, a dilation operation is applied to the mask. Dilation expands the mask boundary by some pixels d, ensuring complete coverage of the source.

For most sources, the default parameters in `PyBDSF` are $T_{\text{isl}} = 3$, $T_{\text{pix}} = 5$, and a dilation parameter $d = 0$ pixels. As a supplementary material, we give the list of sources for which these parameters were modified, along with the corresponding values of T_{isl}, T_{pix}, and d. The final binary masks are stored in FITS format for subsequent analysis. The next steps were not implemented using PyBDSF.

3.2.4. FITS to PNG Conversion

To facilitate visualization and further processing, both the FITS images and the corresponding binary masks are converted to Portable Network Graphics (PNG) format. The pixel values in the FITS images are linearly scaled to the range [0, 255] using the following transformation:

$$\text{img}_{\text{PNG}} = \frac{\text{img} - \min(\text{img})}{\max(\text{img}) - \min(\text{img})} \times 255, \quad (4)$$

where img denotes the original FITS image, and $\min(\text{img})$ and $\max(\text{img})$ are the minimum and maximum pixel values in the image, respectively.

3.2.5. Background Removal via Masking

Background removal from the radio images is achieved by applying the binary masks generated in the previous step. By multiplying the original image with its corresponding binary mask, the background is suppressed, retaining only the source emissions. This operation is expressed as:

$$\text{img}_{\text{masked}} = \text{img} \times M, \quad (5)$$

where $\text{img}_{\text{masked}}$ represents the background-subtracted image, img is the original radio image, and M denotes the binary mask.

This process is applied to all images in the dataset, resulting in a series of background-subtracted images that highlight the radio sources. Figure 2 illustrates the background removal process for a sample radio image.

Table 2: Quantitative evaluation (performance comparison) of different deep learning models on the bent AGN dataset. The table presents the accuracy of the models, the class-wise precision, recall, and F1-score for NAT and WAT sources. ConvNeXT achieves the highest F1-scores for both WAT and NAT sources, indicating its effectiveness in identifying both morphologies.

Model	Accuracy [%]	WAT			NAT		
		Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
VGG-16 [14]	74.60	0.73	0.92	0.81	0.80	0.48	0.60
ResNet-50 [15]	77.77	0.79	0.87	0.82	0.76	0.64	0.70
ViT-B-16 [16]	76.19	0.76	0.89	0.82	0.77	0.73	0.74
SWIN-B [17]	80.95	0.80	0.92	0.85	0.84	0.64	0.73
ConvNeXT [18]	84.12	0.85	0.89	0.87	0.83	0.76	0.79

3.3. Final Dataset Curation

3.3.1. Expert Validation

After processing the data, two expert astronomers from our group manually reviewed each image in the dataset to verify the presence of bent AGNs, the validity of the binary masks, the quality of the background removal, and the overall image quality. They annotated each image with a binary label indicating whether the source was suitable for inclusion in the dataset. In cases of poor-quality samples, the experts provided feedback explaining the reasons for exclusion. Only images unanimously approved by both experts were included in the final dataset. Consequently, 64 sources were rejected, leaving a total of 639 bent radio AGNs in the dataset.

3.3.2. Dataset Splitting

The dataset was split into 10 batches, each containing 64 images, except for the last batch, which contained 63 images. The dataset was stratified based on the source type (NAT or WAT) to ensure an even distribution of source types across the batches. Each batch was saved as a separate pickle file for easy loading during training and testing. The first nine batches were used for training, and the last batch was reserved for testing. Table 1 summarizes the distribution of sources across the batches.

4. BENCHMARKING EXPERIMENTS

4.1. Experimental Setup

4.1.1. Baseline Models

We evaluate the performance of several state-of-the-art deep learning models for image classification. These include two widely used convolutional neural networks (CNNs), VGG-16 [14] and ResNet-50 [15], along with two transformer-based architectures, Vision Transformer (ViT) [16] and SWIN Transformer [17]. Additionally, we consider ConvNeXT [18], a recent CNN-based model that has demonstrated competitive results on various image classification tasks. All models are pre-trained on ImageNet and fine-tuned on our dataset.

4.1.2. Implementation Details

The models are implemented using the PyTorch deep learning framework and trained on the curated dataset. The dataset consists of grayscale images of size 150×150 pixels, which are resized to 224×224 pixels to match the input requirements of the models. Since the selected architectures expect three-channel inputs, the images are converted to RGB format by duplicating the grayscale channel. We employ the Adam optimizer for training the CNN-based models (VGG-16, ResNet-50, and ConvNeXT) with a learning rate of 0.001. The learning rate is scheduled using the StepLR scheduler with a step size of 5 epochs and a decay factor of 0.1. For transformer-based models (ViT and SWIN, specifically ViT-B-16 and SWIN-B), we use the AdamW optimizer with the same learning rate but apply the CosineAnnealingWarmRestarts scheduler with $T_{\text{max}} = 5$ epochs and $\eta_{\text{min}} = 0$. All models are initialized with pre-trained weights on ImageNet and fine-tuned on our dataset. Training is conducted with a batch size of 32 for 20 epochs. Early stopping is applied based on validation loss to prevent overfitting. All experiments were conducted on a single NVIDIA T4 GPU with 16GB of VRAM.

4.1.3. Evaluation Metrics

The performance of the models is assessed using standard classification metrics: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly classified samples. Precision quantifies the ratio of correctly predicted positive samples to total predicted positives, while recall (sensitivity) calculates the ratio of correctly predicted positive samples to total actual positives. The F1-score, the harmonic mean of precision and recall, provides a balanced measure of model performance. While accuracy is computed for the entire dataset, precision, recall, and F1-score are calculated separately for NAT and WAT sources to evaluate class-wise performance.

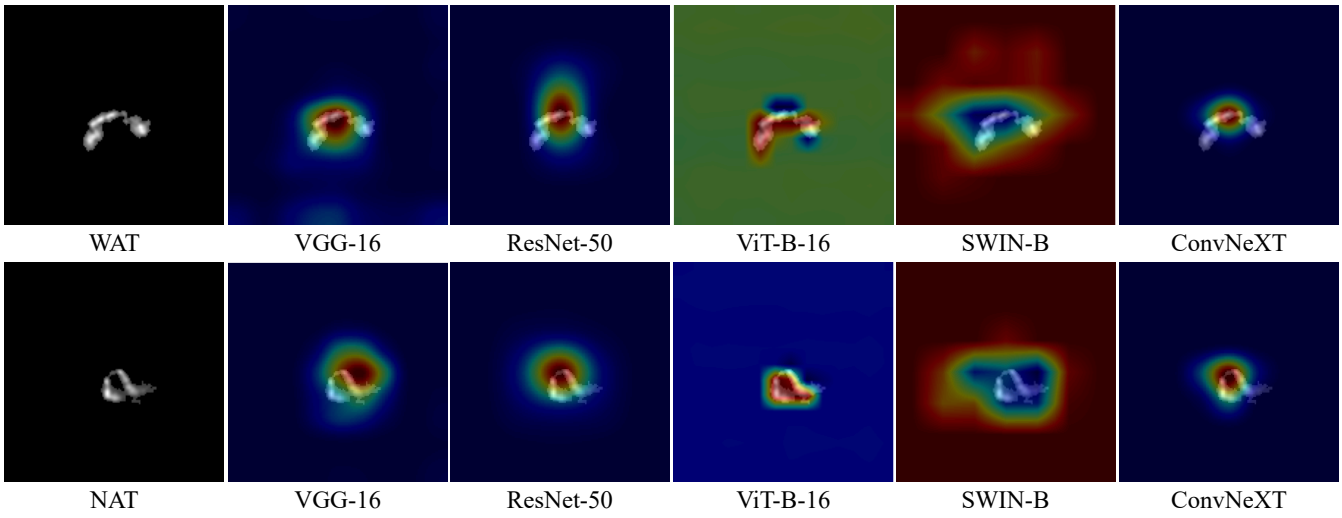


Fig. 3: Class activation maps (CAMs) generated by different models for a sample NAT and WAT source using Grad-CAM [19]. The CAMs highlight the regions in the image that contribute most to the model's classification decision.

4.2. Benchmarking Results

4.2.1. Quantitative Results

We present the quantitative results of the different models in Table 2 as benchmarking results. The table shows the overall accuracy of the models, as well as the class-wise precision, recall, and F1-score for NAT and WAT sources. ConvNeXT achieves the highest F1-scores for both WAT (0.87) and NAT (0.79), demonstrating its robustness in classifying both morphologies. SWIN-B follows closely, particularly excelling in WAT classification with an F1-score of 0.85, benefiting from its strong recall. Transformer-based models, such as ViT-B-16, also show high recall for WAT sources, suggesting their effectiveness in capturing elongated and complex structures. However, NAT classification remains challenging, with VGG-16 performing the worst (F1-score = 0.60) due to its low recall (0.48). While ResNet-50 provides a balanced performance, it falls short of ConvNeXT and SWIN-B in overall classification effectiveness.

4.2.2. Qualitative Analysis

To gain insights into the models' decision-making processes, we visualize the class activation maps (CAMs) using Grad-CAM [19]. Figure 3 presents the CAMs generated by all models for a sample NAT and WAT source. The CAMs highlight the regions of the image that contribute most to the model's classification decision. Convolutional neural networks like VGG-16, ResNet-50, and ConvNeXT tend to focus on the core regions of the source, while transformer-based models such as ViT-B-16 and SWIN-B capture the elongated structure of the source more effectively. This observation aligns with the models' performance in classifying WAT sources, where transformers exhibit higher recall compared to CNNs. The CAMs thus provide valuable insights into the models' interpretability, emphasizing the importance of capturing source morphology for accurate classification.

4.3. Discussion

The dataset exhibits a notable class imbalance, with 385 WAT and 254 NAT sources, making overall accuracy an inadequate metric for evaluation. Given the small training size (576 samples), F1-score provides a more reliable measure of model effectiveness. Transformer-based models' limitations may stem from the relatively small dataset size, which could hinder their ability to generalize effectively without larger-scale pretraining on similar astronomical images. On the other hand, conventional CNN-based models like VGG-16 and ResNet-50, though outperformed by ConvNeXt, provided baseline insights into the efficacy of simpler architectures for bent AGN detection. These results highlight that CNN-based architectures, particularly ConvNeXT, offer better consistency across both classes, whereas transformers may be advantageous in scenarios requiring high recall, especially for WAT detection.

5. LICENSING AND ETHICAL CONSIDERATIONS

The dataset is available online at `https://drive.google.com/file/d/1s3u9HKXCJmX_1DKv5Hp9fJS_b5DyfNHV` under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. This license permits unrestricted use, distribution, and reproduction in any medium, provided that the original authors and source are credited. The code used to process the dataset and train the models is included in the sup-

plementary material and will be made publicly available upon acceptance of the paper. The final version of the paper will contain the link to access the code. The data is anonymized, and no personal information is included, ensuring compliance with ethical standards.

6. CONCLUSION

In this paper, we introduced a novel dataset for the classification of bent AGNs, focusing on WAT and NAT sources. With 639 curated radio images, the dataset provides valuable resources for advancing research in astronomical image classification, including detailed data acquisition, processing, and expert validation. We benchmarked various deep learning models, highlighting the strengths of CNN and transformer-based architectures for bent AGN classification. This work lays the foundation for future efforts, including expanding the dataset with additional sources from broader radio surveys, contributing to the development of comprehensive resources for astronomical research across multiple domains.

7. REFERENCES

[1] W. N. Brandt and D. M. Alexander, "Supermassive black-hole growth over cosmic time: Active galaxy demography, physics, and ecology from Chandra surveys," *Proceedings of the National Academy of Science*, vol. 107, pp. 7184–7189, Apr. 2010.

[2] Emmet Golden-Marx, E. Moravec, L. Shen, Z. Cai, E. L. Blanton, M. L. Gendron-Marsolais, H. J. A. Röttgering, R. J. van Weeren, V. Buiten, R. D. P. Grumitt, J. Golden-Marx, S. Pinjarkar, and H. Tang, "The High-redshift Clusters Occupied by Bent Radio AGN (COBRA) Survey: Investigating the Role of Environment on Bent Radio AGNs Using LOFAR," , vol. 956, no. 2, pp. 87, Oct. 2023.

[3] James McBride and Michael McCourt, "Bent radio jets reveal a stripped interstellar medium in NGC 1272," *Monthly Notices of the Royal Astronomical Society*, vol. 442, no. 1, pp. 838–843, July 2014.

[4] W. Domainko, M. Mair, W. Kapferer, E. van Kampen, T. Kronberger, S. Schindler, S. Kimeswenger, M. Ruffert, and O. E. Mangete, "Enrichment of the ICM of galaxy clusters due to ram-pressure stripping," *Astronomy & Astrophysics*, vol. 452, no. 3, pp. 795–802, June 2006.

[5] Christopher P. O'Dea and Stefi A. Baum, "Wide-Angle-Tail (WAT) Radio Sources," *Galaxies*, vol. 11, no. 3, pp. 67, May 2023.

[6] B Terni de Gregory, Luigina Feretti, Gabriele Giovannini, Federica Govoni, Matteo Murgia, Rick A Perley, and Valentina Vacca, "Narrow head-tail radio galaxies at very high resolution," *Astronomy & Astrophysics*, vol. 608, pp. A58, 2017.

[7] Zhaoming Gan, Hui Li, Shengtai Li, and Feng Yuan, "Three-dimensional magnetohydrodynamical simulations of the morphology of head–tail radio galaxies based on the magnetic tower jet model," *The Astrophysical Journal*, vol. 839, no. 1, pp. 14, apr 2017.

[8] Chris J. Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, Robert C. Nichol, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg, "Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey*," *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, pp. 1179–1189, 09 2008.

[9] 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*, vol. 167, pp. 31P–36P, May 1974.

[10] Fiona A M Porter and Anna M M Scaife, "Mirabest: a data set of morphologically classified radio galaxies for machine learning," *RAS Techniques and Instruments*, vol. 2, no. 1, pp. 293–306, 06 2023.

[11] DD Proctor, "Morphological annotations for groups in the first database," *The Astrophysical Journal Supplement Series*, vol. 194, no. 2, pp. 31, June 2011.

[12] Tapan K. Sasmal, Soumen Bera, Sabyasachi Pal, and Soumen Mondal, "A New Catalog of Head-Tail Radio Galaxies from the VLA FIRST Survey," *The Astrophysical Journal Supplement Series*, vol. 259, no. 2, pp. 31, Apr. 2022.

[13] Niruj Mohan and David Rafferty, "PyBDSF: Python Blob Detection and Source Finder," Astrophysics Source Code Library, record ascl:1502.007, Feb. 2015.

[14] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun, Eds., 2015.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer,

Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.

[17] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 10012–10022.

[18] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11976–11986.

[19] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.

8. SUPPLEMENTARY MATERIALS

The expert-annotated dataset and the custom codes for data processing, annotation conversion, and label refinement are available in this anonymous link (link).