**Independent University** **Bangladesh (IUB)**

**IUB Academic Repository**

Graduate Thesis 2025

2025-12

# RGC: A Radio AGN Classifier Based on Deep Learning

Shahal, Md Shahadat Hossain

IUB

http://ar.iub.edu.bd/handle/11348/1033

*Downloaded from IUB Academic Repository*




# RGC: A Radio AGN Classifier Based on Deep Learning

By

**Md Shahadat Hossain Shahal**

Student ID: **2222914**

**Autumn, 2025**

Supervisor:

**A K M Mahbubur Rahman, PhD**

**Associate Professor**

Department of Computer Science & Engineering

Independent University, Bangladesh

Co-Supervisor:

**Khan Muhammad Bin Asad, PhD**

**Assistant Professor**

Department of Physical Sciences

Independent University, Bangladesh

**December 2025**

**Dissertation submitted in partial fulfillment for the degree of Bachelors of Science in Computer Science & Engineering**

**Department of Computer Science & Engineering**

**Independent University, Bangladesh**

# Attestation

We are well aware that plagiarism is against the regulations of our university and is completely forbidden. We affirm that the work is entirely ours and promise its authenticity. In our project, we have used certain protected content and other people's models, which have been appropriately cited in accordance with international norms. Less than 10% of the manuscript's text was improved for readability using ChatGPT (OpenAI, 2025). We took full responsibility for the final draft of the manuscript after reviewing and editing all the content that AI had refined.

Signature

Date

Name

# Acknowledgement

I would like to thank my esteemed supervisor, A K M Mahbubur Rahman, PhD, Independent University, Bangladesh, and my co-supervisor, Khan Muhammad Bin Asad, PhD, Independent University, Bangladesh, for their guidance, wise recommendations, extraordinary tolerance, and the kind donation of their precious time. Their knowledge and guidance improved the caliber and breadth of our work, and I sincerely value their steadfast assistance during our academic career.

In addition, I would like to thank Mir Sazzat Hossain, Adrita Khan, and Fatema Akter, the other co-authors of the study. Additionally, I would like to express my gratitude to Drs. Payaswini Saikia, Amin Ahsan Ali, M Ashraful Amin, and M Arshad Momen for their insightful comments.

We also extend our sincere gratitude to the Center for Astronomy, Space Science, and Astrophysics (CASSA) for their outstanding assistance and unwavering support. The creation and implementation of our idea were greatly aided by the cooperative and inventive atmosphere that CASSA offered.

I would want to express my gratitude to everyone who helped me with this project, both directly and indirectly.

# Evaluation Committee

Supervision Panel

| | |
|---|---|
| ..... | ..... |
| Supervisor | Co-supervisor |

Examiners

| | |
|---|---|
| ..... | ..... |
| Internal Examiner 1 | Internal Examiner 2 |
| ..... | |
| External Examiner | |

Office Use

| | |
|---|---|
| ..... | ..... |
| Center Director | Head of the Department |
| ..... | |
| Director Graduate Studies, Research and Industry Relations | |
| ..... | |
| Library | |

# Abstract

No prior machine-learning classifier for bent RAGNs has used both unlabeled data and purely visually verified labels, despite the crucial role that wide-angle tail (WAT) and narrow-angle tail (NAT) radio active galactic nuclei (RAGNs) play as tracers of dense environments in galaxy groups and clusters. We provide the `RGC` Python module, which builds a semi-supervised model that combines two recently curated labeled datasets containing 639 WATs and NATs from a publicly accessible catalog of visually examined sources with 20,000 unlabeled RAGNs.

The labeled datasets in `RGC` were preprocessed using `PyBDSF`, which keeps them for comparison, and `Photutils`, which eliminates spurious sources. To create a reliable semi-supervised binary model, the underlying classifier combines a supervised E2CNN (E(2)-equivariant Convolutional Neural Network) with the self-supervised learning framework BYOL (Bootstrap YOur Latent).

The `RGC` model reaches peak performance with an accuracy of 88.88% and F1-scores of 0.90 for WATs and 0.85 for NATs when trained and assessed on a dataset free of spurious sources. The model's attention patterns point to a potential step toward physics-informed foundation models that can recognize a variety of AGN physical properties, especially when there is class imbalance.

***Keywords***— Galaxies: active – Radio continuum: galaxies – Techniques: image processing – Methods: data analysis – Methods: statistical

# Contents

**Attestation** i

**Acknowledgement** ii

**Evaluation Committee** iv

**Abstract** iv

**1 Introduction** 1

1.1 Overview . . . 1
1.2 Research Gap . . . 4
1.3 Research Objectives and Research Questions . . . 4
1.3.1 Research Objectives . . . 4
1.3.2 Research Questions . . . 5
1.4 Research Outcome . . . 5
1.5 Contribution of the thesis . . . 5
1.6 Organization of the Thesis . . . 6

**2 Literature Review** 8

2.1 Literature Selection Procedure . . . 8
2.2 Keywords Searched . . . 9
2.3 Databases for Search . . . 9
2.4 Inclusion/Exclusion Criteria . . . 10
2.5 Filtering Results . . . 11
2.6 Impact of the project . . . 12
2.7 Findings . . . 13
2.7.1 Background on Radio Active Galactic Nuclei (RAGNs) . . . 13
2.7.2 Radio Continuum Surveys and Data Sources . . . 14
2.7.3 Machine Learning in Radio Galaxy Classification . . . 15
2.7.4 Semi-Supervised Learning and Data Preprocessing . . . 16
2.7.5 Deep Learning Model Design and Training . . . 16

**3 Research Methodology** 18

3.1 Data Collection and Sources . . . 18

3.2 Preprocessing Data . . . 18
3.3 Model Architecture and Design . . . 19
3.4 Training Procedures . . . 19
3.5 Interpretability and Evaluation Metrics . . . 19

**4 Work Breakdown Structure (WBS) 21**

**5 Data and preprocessing 23**
5.1 Unlabeled data . . . 23
5.2 Labeled data . . . 24
5.3 Downloading . . . 24
5.4 Generating masks using `PyBDSF` . . . 25
5.5 Refining Masks Using `Photutils` . . . 26
5.6 Sasmal-ML: a bent RAGN catalog and dataset for ML . . . 31

**6 Architecture of the model 32**
6.1 Self-supervised pre-training . . . 32
6.2 Supervised fine-tuning . . . 34
6.3 Training details . . . 35
6.4 Class activation mapping . . . 35

**7 Results & Analysis 37**
7.1 Confusion matrix . . . 37
7.2 Classification metrics . . . 39
7.3 Discriminative ability . . . 42
7.4 Model calibration . . . 44
7.5 Findings and Challenges . . . 46
7.6 Astrophysical implications . . . 48

**8 Sustainability 50**

**9 Conclusion 51**
9.1 Project Summary . . . 51
9.2 Future Work: Multimodal Self-Supervised Radio Galaxy Classification from multi-spectral images . . . 52

**10 Appendix 55**
10.1 Sasmal-ML catalog . . . 56
10.2 Grad-CAM localization maps . . . 58

**Bibliography 72**

# List of Figures


3.1 Flowchart of the research methodology. Raw radio images are subjected to source extraction and masking only for labeled samples to prepare a clean labeled dataset of 639 bent radio AGNs. The large unlabeled set of 20,000 bent radio sources remains unprocessed for pretraining. Semi-supervised pretraining via BYOL with an E2-equivariant CNN encoder is performed on the unlabeled data, followed by supervised fine-tuning using the preprocessed labeled dataset. The trained model undergoes evaluation and interpretability analysis leading to the final bent radio AGN classification catalog. . . . . 20

5.1 An example of the process of masking an image as described in Section 5.4 and 5.5. $\mathbf{R}_n$ is the normalized original image, $\mathbf{M}_{L1}$ its mask created using `PyBDSF` and $\mathbf{M}_L$ its mask produced through `Photutils`. . . . . 30

6.1 *Bottom*:The architecture of BYOL used in our self-supervised pre-training, adapted from [1]. The encoder within BYOL in our case is E2CNN which has been used for our supervised downstream fine-tuning. *Top*: The architecture of E2CNN as used by [2], adapted from the LeNet-5 architecture of [3]. The operations GC, MP, GP and FC stand for group convolution, max-pooling, group-pooling and fully-connected, respectively. . . . . 33

6.2 Grad-Cam implementation . . . . . 36

7.1 Confusion matrices for the classification of bent RAGNs into WATs and NATs, showing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for the RGC model trained on the datasets $\mathbf{R}_{L1}$ (left) and $\mathbf{R}_{L2}$ (right). . . . . 39

7.2 *R*OC curves and AUC scores for the RGC model trained and tested on two datasets $\mathbf{R}_{L1}$ (red) and $\mathbf{R}_{L2}$ (green), for both WATs (solid) and NATs (dashed), in comparison to the random guess (blue). . . . . 42

7.3 *P*recision-Recall (PR) curves and the corresponding AUC-PR scores for the same datasets and sources. . . . . 43

7.4 ECE curves (described in Section 7.4) for our RGC model trained and tested on two datasets for classifying WATs and NATs. The colors and linestyles are the same as Fig. 7.2. . . . . 45

# List of Tables

1.1 The surveys conducted by array radio telescopes (ART) that have found most of the bent RAGNs so far. . . . . 3

2.1 Summary of Inclusion and Exclusion Criteria for Literature Selection . . . . . 11

5.1 Distribution of WATs and NATs in our labeled dataset $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$, created following the Sasmal-ML catalog demonstrated in Table 10.1. . . . . 29

6.1 Comparison of the hyperparameters used in our self-supervised pre-training and supervised fine-tuning. . . . . 35

7.1 Performance metrics of the RGC model on two datasets, with ($\mathbf{R}_{L1}$) and without ($\mathbf{R}_{L2}$) spurious sources, along with five additional fully supervised models trained and tested on $\mathbf{R}_{L1}$ [4]. See Section 7.2 and Section 7.2 for details. . . . . 40

10.1 Information about the 63 sources from our Sasmal-ML catalog that are included in our Test Batch. . . . . 56

10.2 Notation glossary . . . . . 63

# Chapter 1

# Introduction

## 1.1 Overview

In this thesis, we present Version 1.0 of RGC, a Python package of deep-learning (DL) models and datasets for classifying radio active galactic nuclei (RAGN), [1] named after Radha Gobinda Chandra (1878–1975), an amateur astronomer from Bangladesh and India who made over 50,000 observations to the American Association of Variable Star Observers [5] and reported the observation of 1910 in Bangla [6]. RGC employs convolutional neural networks (CNN) to classify RAGNs that have straight and bent tails formed by synchrotron jets. Its performance in classifying these objects into Fanaroff–Riley (FR) types I and II was described by [7]. In this work, we present and analyze its performance in classifying bent RAGNs into wide-angle-tail (WAT) and narrow-angle-tail (NAT) sources. A background to the present work is given below.

The observable universe could contain almost two trillion galaxies [8], but only a fraction of them have been observed at different wavelengths. The X-ray space telescope eROSITA is expected to find thousands of galaxy clusters and millions of AGNs [9]. Samples of 700 thousand AGNs [10] and 12 thousand galaxy clusters and groups [11] have already been published. In visible light, SDSS-III (Sloan Digital Sky Survey) released 11th and 12th data release which contained more than 1.3 million galaxies and more than a quarter of a million quasars [12]. The mid-infrared space telescope WISE (Wide-field Infrared Survey Explorer) has identified almost 750 million radio sources which are part of its AllWISE data release [13]. At mid radio frequencies, the Faint Images of the Radio Sky at Twenty-Centimeters (FIRST), taken using the Karl G. Jansky Very Large Array (VLA), contains more than 800 thousand sources in its April 2003 release; the initial catalog of [14] contained 20 thousand sources. At a lower frequency, the Evolutionary Map of the Universe (EMU) conducted using the Australian Square Kilometre Array Pathfinder (ASKAP) contained more than 200 thousand sources in its pilot survey and it is expected to find many more [15]. Moreover, at the lowest radio frequencies, Low Frequency ARray (LOFAR) has detected almost 4.4 million sources as part of its Two-metre Sky Survey LOTSS [16].

[1] These sources are often named 'radio galaxies,' but we use the expanded term 'radio AGN' (RAGN) throughout.

Astronomers are also cross-matching the sources seen at different wavelengths, especially the AGNs with their host galaxies. For example,[17] cross-matched FIRST and SDSS with NVSS (NRAO VLA Sky Survey) to produce a catalog of more than 18,000 sources. Traditionally, cross-matching has been achieved through visual inspection by astronomers which is not feasible for the latest generation of surveys. Therefore, [18] created the citizen science project Radio Galaxy Zoo (RGZ) where users were asked to cross-identify radio images from FIRST and ATLAS (Australia Telescope Large Area Survey, a pathfinder of EMU) with their host galaxies imaged by WISE or Spitzer Space Telescope in infrared. Data Release 1 of RGZ contains classifications of 99,146 radio sources from FIRST, with an average reliability of 0.83 based on the weighted consensus levels of the citizen scientists [19].

If we focus on the history of 'radio' continuum surveys from the days of Grote Reber until the last decade, we see that the number of detected sources increased by an order of four from the 1940s to the 1970s, and since then we have seen an increase by four more orders of magnitude (Figure 2 of [20]). EMU and LOFAR are expected to find close to 100 million radio sources. The number will increase more when the Square Kilometer Array (SKA) begins to operate in two phases, SKA1 and SKA2, and at two different frequency bands, Low frequency observation from SKA (50–350 MHz) and mid-frequency observation from SKA (350 MHz – 15 GHz). At mid-frequencies, SKA1 All-Sky Survey (SASS1) will detect around 500 million galaxies and SASS2 over 3 billion galaxies spanning all redshifts [21]. The first of the 131,072 two-meter-tall Christmas-tree-shaped antennas of SKA-Low was installed in a Wajarri country in Australia on 7 March 2024 and the first of the 197 fifteen-meter-wide dishes of SKA-Mid was lifted onto a pedestal on 4 July 2024 in the Karoo region of South Africa.[2]

In this context, artificial intelligence (AI) is necessary for localizing astronomical sources in large datasets and classifying them into scientifically meaningful classes. A recent review by [22] emphasizes the new paradigm shift that has happened due to the application of machine learning (ML) for the morphological classification of RAGNs. They have reviewed 32 papers published between 2017 and 2023 that utilized both conventional ML and DL methods. The most frequently used methods were found to be based on shallow and deep convolutional neural networks (CNN). The methods were divided into model-centric and data-centric approaches. The model-centric approaches focus on applying novel architectures on existing well-curated datasets, which are scarce. Most methods use low-resolution images taken by older telescopes like the VLA. These are sometimes not suited to the new high-resolution images coming out of the latest surveys from the new generation of telescopes, for example, LOFAR and MeerKAT. The data-centric approaches leverage the availability of more numerous images and focus on transfer learning and semi-supervised learning.

[2]According to `https://www.skao.int/news`.

Table 1.1: The surveys conducted by array radio telescopes (ART) that have found most of the bent RAGNs so far.

| ART | Survey | Location | Sensitivity | Resolution | RAGNs | WATs | NATs | Reference |
|---|---|---|---|---|---|---|---|---|
| VLA | FIRST | USA | 150 | $5''$ (1400) | 717 | 430 | 287 | [23] |
| | | | | | 4876 | 4424 | 652 | [24] |
| LOFAR | LoTSS | Netherlands | 83 | $6''$ (144) | 35 | – | – | [25] |
| | | | | | 55 | 45 | 10 | [26] |
| ATCA | ATLAS | Australia | 20 | $12''$ (1400) | 45 | – | – | [27] |
| GMRT | TGSS | India | 3500 | $25''$ (150) | 264 | 203 | 61 | [28] |
| MeerKAT | MIGHTEE | South Africa | 6 | $5''$ (1250) | 359 | – | – | [29] |

Sensitivity is median RMS in $\mu$Jy beam$^{-1}$. Resolution in arcsec, with observing frequency in MHz in parentheses.

## 1.2 Research Gap

Despite significant advances in radio galaxy classification using supervised deep learning methods, several limitations remain. Existing approaches often rely on fully labeled datasets which are expensive and time-consuming to obtain. There is a scarcity of large labeled datasets of bent radio AGNs, particularly those distinguishing between Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT) morphologies. Additionally, current models struggle with class imbalance, spurious source contamination, and lack robustness to orientation variations in high-resolution images from new radio surveys. These gaps motivate the development of semi-supervised, rotation-equivariant models capable of leveraging abundant unlabeled data while addressing these challenges effectively.

## 1.3 Research Objectives and Research Questions

### 1.3.1 Research Objectives

The primary objectives guiding this research are: The main goal of this work is to use state-of-the-art semi-supervised deep learning techniques to improve automatic categorization of bending radio Active Galactic Nuclei (AGN) morphologies, especially differentiating between Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT) sources. Understanding the physical characteristics, interactions with the environment, and cosmic importance of radio galaxies depends on their morphological categorization [30, 31]. Conventional manual inspection techniques are insufficient due to the growing volume of data from next-generation radio surveys like LOFAR, MeerKAT, ASKAP, and SKA, requiring scalable, repeatable machine learning solutions [20, 32].

The following relevant scientific and technological issues are the focus of this work:

1. Creating strong semi-supervised learning frameworks that may overcome the constraints of limited annotations by utilizing large unlabeled radio astronomy datasets in addition to smaller labeled sets [33, 1].

2. Improving model generalization and physical interpretability by incorporating physical domain information with equivariant neural network topologies that follow intrinsic rotational and reflection symmetries in radio galaxy morphologies [2, 34].

3. Assessing the effects of spurious source contamination and creating preprocessing pipelines that guarantee cleaner data inputs for classification are essential for reducing misclassification brought on by imaging artifacts [35, 36].

4. Facilitating the creation of extensive, uniform, and precisely categorized databases of bent radio AGN that may be used to answer important astrophysical queries on cluster environments, AGN feedback, galaxy development, and large-scale structure tracing [37, 38].

### 1.3.2 Research Questions

1. RQ1: How effectively can a semi-supervised deep learning model classify bent radio AGNs into Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT) classes using both labeled and unlabeled VLA images?
   • Rationale: This question aims to evaluate the benefits of leveraging unlabeled data with self-supervised learning to improve classification performance over fully supervised methods.

2. RQ2: What improvements in model robustness and accuracy can be achieved by incorporating rotation and reflection equivariance into the convolutional neural network architecture?
   • Rationale: Investigating the impact of equivariant CNNs helps to understand their role in handling diverse source orientations in radio images.

3. RQ3: How do spurious sources and class imbalance in radio images affect model performance, and which preprocessing techniques best mitigate these issues?
   • Rationale: Identifying preprocessing strategies is crucial for enhancing data quality and subsequently improving classification outcomes.

4. RQ4: Can attention visualization techniques such as Grad-CAM provide interpretable insights into the features used by the model for distinguishing between WAT and NAT morphologies?
   • Rationale: This question explores the model's interpretability and its alignment with astrophysical understanding of bent radio AGNs.

## 1.4 Research Outcome

This research is expected to deliver a novel semi-supervised classification framework, the RGC model, capable of accurate and robust bent radio AGN classification with improved interpretability. The study will provide insights into the impact of equivariant neural networks and semi-supervised learning in astrophysical image analysis. Additionally, it will contribute to preprocessing pipelines for radio astronomy data and expand the catalog of reliably classified bent radio sources for subsequent scientific study. The outcome aims to facilitate handling of large unlabeled datasets anticipated from next-generation radio surveys such as SKA.

## 1.5 Contribution of the thesis

Our data-centric approach is based on semi-supervised learning, where we leverage the availability of unlabeled data for classifying bent RAGNs. The model used here is the same as that of H23, but we use it for classifying *bent* RAGNs into WATs and NATs for the first time. Automation in this classification is not merely a computational necessity in the era of big-data astronomy, but also an enabler of new physical diagnostics in extragalactic radio astronomy.

Large, homogeneous samples of WATs and NATs could make it possible to statistically trace jet–environment interactions, constrain ram-pressure conditions across redshift, and map the dynamical state of galaxy clusters and groups via bent-jet statistics.

A deep learning-based semi-supervised classification model for bent radio active galactic nuclei (RAGNs) is developed and evaluated in this thesis, with a particular focus on the binary classification of Wide-Angle Tail (WAT) and Narrow-Angle Tail (NAT) radio AGNs. Bent RAGNs are crucial astrophysical tracers of dense locations like galaxy clusters and groups because of their jets and tails, which are warped by these settings. However, the growing amount of radio survey data and the shortcomings of conventional visual examination techniques make it extremely difficult to classify large samples of bent RAGNs.

The RGC Python package, which uses a semi-supervised deep learning architecture with both labeled and unlabeled datasets from Very Large Array (VLA) FIRST survey photos, is presented in this study as a solution to this problem. The labeled datasets are meticulously preprocessed images of WAT and NAT sources that were produced by adding machine learning-friendly masks to preexisting visual catalogs in order to eliminate spurious sources and improve training data quality. For reliable and precise feature extraction and classification, the semi-supervised method combines cutting-edge self-supervised pre-training (BYOL) with equivariant convolutional neural networks (E2CNN).

The study shows that the suggested model outperforms fully supervised baseline classifiers and successfully manages class imbalance between the more frequent WATs and less frequent NATs, achieving a high classification accuracy of over 88%. Moreover, meaningful model attention patterns on pertinent source characteristics rather than false signals are shown by interpretability analyses employing Grad-CAM visualization. As a result, our study not only establishes the groundwork for future physics-informed foundation models that can thoroughly characterize AGN features across progressively larger and more varied radio astronomical datasets, but it also promotes automated radio galaxy categorization.

## 1.6 Organization of the Thesis

This thesis is structured into ten chapters, each addressing a specific aspect of the research on semi-supervised deep learning classification of bent radio active galactic nuclei (RAGNs). The organization is designed to guide the reader through the scientific background, methodology, experiments, and implications systematically.

Chapter 1 introduces the research context, motivations, objectives, and contributions. It sets the stage by discussing the astrophysical significance of bent RAGNs and challenges in their morphological classification.

Chapter 2 presents a comprehensive literature review on radio AGNs, related radio surveys, and recent developments in machine learning approaches relevant to this study.

Chapter 3 outlines the detailed research methodology. It describes data sources, preprocessing techniques including masking with PyBDSF and Photutils, and the semi-supervised learning framework integrating BYOL with E2-equivariant CNN architectures.

Chapter 4 provides the work breakdown structure.

Chapter 5 discusses the data preprocessing procedures, including source extraction, segmentation, and dataset creation for labeled and unlabeled samples.

Chapter 6 details the model design and training protocols, including hyperparameter configurations for pretraining and supervised fine-tuning.

Chapter 7 presents experimental evaluation results. It reports classification performance metrics on two datasets (with and without spurious sources), calibration analysis, and empirical robustness studies.

Chapter 8 addresses sustainability considerations, analyzing social, environmental, technical, and operational aspects in deploying artificial intelligence in radio astronomy, aligning with SDG goals.

Chapter 9 concludes the thesis with a project summary, key findings, limitations, and directions for future work toward radio foundation models and large-scale applications.

Chapter 10 includes appendices providing additional technical details, extended data catalogs, notation glossaries, and supplementary figures.

# Chapter 2

# Literature Review

A radio-active galactic nucleus (AGN) is a region at the center of a galaxy where a supermassive black hole is actively accreting matter, which creates powerful jets of plasma that emit strong radio waves. These jets can extend far beyond the host galaxy, making the galaxy an exceptionally luminous source of energy throughout the electromagnetic spectrum. Radio-loud AGN are a type of AGN that is particularly bright in radio waves.

## 2.1 Literature Selection Procedure

The collection of literature in this study covers a wide range of historical and methodological viewpoints on the classification of radio galaxies. The process starts with the basic discovery and categorization of galaxies, which dates back to the early 20th century. Edwin Hubble established the morphological classification of galaxies into spirals, ellipticals, and irregulars, which laid the groundwork for contemporary galaxy taxonomy. Understanding non-thermal radio emissions from active galactic nuclei (AGNs) was significantly advanced by the discovery of radio galaxies in the 1940s, particularly the brilliant source Cygnus A discovered by Grote Reber. By adding the FRI and FRII classes based on radio jet form and brightness distribution, the Fanaroff-Riley classification in 1974 improved radio galaxy taxonomy.

Building on this tradition, the review concentrated on current developments in deep learning and machine learning methods used for radio galaxy morphology classification, particularly the rise of semi-supervised methods utilizing sizable unlabeled datasets. In order to capture state-of-the-art techniques and challenges, especially those pertaining to the classification of Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT) sources, key academic databases such as ADS, arXiv, IEEE Xplore, and Google Scholar were systematically searched using a combination of keywords such as "radio galaxy classification," "bent radio AGNs," "deep learning in astrophysics," and "semi-supervised learning." Peer-reviewed journal articles and conference proceedings published since 2010.

Additionally, studies using rotation and reflection equivariant neural networks, data preparation approaches for spurious radio sources, and interpretability techniques like Grad-CAM to decode model decision-making were highlighted in the literature selection. To guarantee

relevance to the study questions and objectives, a thorough screening process was carried out by title, abstract, and full text. A thorough and representative list of references was produced by using citation chaining to find groundbreaking and cutting-edge research that was missed in the first searches. The construction and validation of the RGC model for automatic bent radio AGN categorization in the thesis are based on this thorough and historically grounded literature collection.

## 2.2 Keywords Searched

To ensure comprehensive coverage of relevant research, a systematic keyword search was conducted across multiple academic databases including ADS, arXiv, IEEE Xplore, and Google Scholar. The search strategy used a combination of primary and secondary keywords related to radio galaxy classification and machine learning. Primary keywords included "radio galaxy classification," "bent radio AGNs," "Wide-Angle-Tail (WAT)," and "Narrow-Angle-Tail (NAT)." Secondary keywords focused on methodological aspects such as "deep learning," "machine learning," "semi-supervised learning," "self-supervised learning," "equivariant convolutional neural networks," and "radio astronomy image preprocessing." Additional terms like "VLA survey," "LOFAR," and "radio morphology" were included to target specific datasets and observational contexts. To enhance the retrieval of relevant papers, searches were refined using boolean operators and wildcard characters. This methodical keyword search made it easier to find cutting-edge research and guaranteed that the thesis literature review included both fundamental works and more current developments in the topic.

## 2.3 Databases for Search

Several reputable academic databases were used in the literature search to guarantee a comprehensive and varied collection of pertinent research. The Astrophysics Data System (ADS), which is well-known in the astronomy and astrophysics community for its extensive collection of academic papers, conference proceedings, and preprints, particularly about galactic and extragalactic research, served as the main source. Access to important articles on radio galaxy shape and classification was made possible by ADS, which helped establish the study's foundation in both traditional and modern research.

Furthermore, arXiv was widely utilized to acquire the most recent preprints and innovative research advancements in the domains of machine learning applications to astronomy. Modern techniques like semi-supervised learning and equivariant neural networks applied to radio astronomy data were quickly acquired thanks to this library.

The technical literature on deep learning architectures, algorithmic advancements, and computational techniques pertinent to morphological categorization was covered by IEEE Xplore. Peer-reviewed conference papers and journal articles connecting the fields of computer science and astrophysics applications were available in this database.

Because of its extensive multidisciplinary indexing, which enables the acquisition of extra

research from a variety of sources, including theses, reports, and obscure journals that specialist databases would overlook, Google Scholar functioned as an additional tool.

Finally, because of their detailed dataset descriptions and cross-references that complemented bent radio AGN categorization, domain-specific sources like as the NASA/IPAC Extragalactic Database (NED) were accessed. A thorough and current selection of material that both theoretically and practically supports the study goals of this thesis was assured by combining these databases.

## 2.4 Inclusion/Exclusion Criteria

Certain inclusion and exclusion criteria were developed to guarantee the quality and applicability of the literature included in this thesis. Studies that satisfied the following criteria were included:

1. Primarily concentrated on the classification of radio galaxy morphology, particularly with regard to bent radio AGNs like Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT) sources;
2. Used machine learning, deep learning, or semi-supervised learning techniques applicable to radio astronomical image analysis;
3. Used data from major radio surveys like VLA, LOFAR, or similar facilities;
4. Were published in peer-reviewed journals or credible conference proceedings starting in 2015 to reflect contemporary developments; and
5. Offered enough methodological detail to enable critical evaluation or replication.

On the other hand, studies that:

1. Did not address radio galaxy classification or concentrated on unrelated astronomical phenomena;
2. Relied solely on fully supervised traditional methods without incorporating modern machine learning techniques;
3. Involved simulations or theoretical models without empirical validation on real survey data;
4. Were conference abstracts, poster summaries, or lacked full-text accessibility; and
5. Were published before 2015, except early critical discoveries or seminal foundational works such as the Fanaroff-Riley classification paper.

By using these criteria, the literature corpus was reduced to excellent, pertinent studies that directly influenced the research topics, methodology, and model building of the thesis while preserving a thorough understanding of the field's past and present advancements.

## 2.5 Filtering Results

To maintain a strong and pertinent corpus for the thesis, a thorough screening of the literature was directed by the criteria listed in the above table. The databases and articles that were acknowledged for their excellence and applicability to machine learning and astrophysics were the main emphasis. In order to preserve unique contributions, research articles and conference papers that presented actual data and innovative methodology were given priority over secondary publications like book chapters or reviews. The chronological range highlights contemporary advancements while recognizing important classical foundational works. Alignment with the dataset characteristics utilized in this study is ensured by including works that relate to big radio surveys.

Table 2.1: Summary of Inclusion and Exclusion Criteria for Literature Selection

| **Area Specification** | **Inclusion Criteria** | **Exclusion Criteria** |
|---|---|---|
| **Database** | ScienceDirect, Springer, IEEE Xplore/Access, Astronomical Journal, Monthly Notices of the Royal Astronomical Society, Frontiers, Astrophysical Journal, World Scientific, Astronomy and Astrophysics, and other renowned journals | Other journals including predatory journals |
| **Journal/Conference Type** | Research articles, peer-reviewed journal articles, conference proceedings | Book chapters, review articles, abstracts, poster summaries, non-peer-reviewed articles |
| **Area of Interest** | Radio galaxy classification, bent radio AGNs, machine learning in astronomy, deep learning, semi-supervised learning, equivariant CNNs | Studies unrelated to radio morphology or without relevant ML methodology |
| **Selected Time Period** | Publications from 2010 to present, including seminal papers prior to 2010 (e.g., Fanaroff-Riley 1974) | Publications before 2010, unless foundational and widely cited |
| **Data Source / App Type** | VLA, LOFAR, MeerKAT radio survey data analyses; focus on radio astronomical imaging | Studies solely involving simulations without empirical data; unrelated software applications |

## 2.6 Impact of the project

In addition to the minor in Astronomy and Astrophysics, this project has significantly contributed to the consolidation and extension of the knowledge and abilities gained throughout the undergraduate program in Computer Science and Engineering. The implementation, experimentation, and result visualization components of the work have been directly supported by the fundamental programming and software development skills acquired from courses like "Introduction to Computer Programming," "Object Oriented Programming," and "Web Application & Internet". The creation of effective data pipelines for managing massive radio survey datasets and planning experiments in a repeatable manner was made possible by foundations in data structures, algorithms, database management, operating systems, and system analysis and design

To comprehend optimization, loss functions, and performance evaluation in deep learning models, the project also makes extensive use of theoretical and mathematical background from "Discrete Mathematics," "Linear Algebra – Vectors and Matrices," "Probability & Statistics for Science & Engineering," and "Ordinary Differential Equations". The semi-supervised and equivariant deep learning architectures utilized in this study were designed, trained, and validated using the conceptual and practical foundation provided by courses like "Artificial Intelligence" and "Neural Networks". Understanding of data transfer, distributed testing, and possible deployment scenarios of machine learning models in networked contexts is supported by knowledge from "Data Communication & Computer Networks" and "Wireless Networking & Mobile Communication".

From the astronomy and astrophysics track, courses including "Introduction to Astronomy," "Introduction to Astrophysics," and "Radio Astronomy" have been crucial for understanding the physical nature of radio galaxies, emission mechanisms, and the observational characteristics of radio surveys. These lectures offered the astrophysical background necessary to evaluate model outputs, create meaningful labels, and tie morphological classifications of radio AGNs to underlying physical processes and large-scale settings. A deeper comprehension of space science and observational methods has been reinforced by further exposure to "Atmospheric and Space Physics" and "Planets and Planetary Systems," which has improved the capacity to critically assess astronomical data products.

The implementation and iterative improvement of the project reflect the practical skills in experimentation, debugging, and documentation that have been fostered by laboratory-based courses throughout the curriculum, such as programming labs, electronics labs, numerical methods labs, and AI/ML-related labs. Project management, collaboration, and communication skills—all crucial for organizing research activities, working with supervisors, and successfully presenting scientific results—were further enhanced by the "Live-in-Field Experience" course. Overall, this project serves as a capstone that combines skills in computers, mathematics, and astrophysics into a cohesive research contribution, showcasing the undergraduate coursework's practical relevance to cutting-edge issues in machine learning and radio astronomy.

## 2.7 Findings

### 2.7.1 Background on Radio Active Galactic Nuclei (RAGNs)

Located in the cores of galaxies and driven by the accretion of matter onto supermassive black holes (SMBHs), Active Galactic Nuclei (AGNs) are among the most brilliant and energetic phenomena in the universe [39]. The tremendous gravitational potential energy generated during accretion is transformed into strong relativistic particle jets and intense electromagnetic radiation that may go far beyond the optical bounds of the host galaxy. One well-known subtype of AGNs is known as radio AGNs (RAGNs), which are distinguished mainly by intense synchrotron emission that may be seen at radio wavelengths [40].

Fanaroff and Riley (1974) defined two morphological classifications based on the brightness distribution of their radio lobes, establishing the archetypal categorization of radio galaxies. While FRII sources have edge-brightened lobes with noticeable terminal hotspots, FRI sources have jets that are brighter close to the core and fade with distance. Jet power and environmental conditions are directly related to this categorization [30]. The categorization approach has been extended beyond these conventional forms to include hybrid FRI-FRII morphologies and FR0 sources, which have minimal extended structure and compact radio emission, representing the variety of the radio AGN population [41, 42].

The bending of jets and tails, which gives birth to the Bent Radio AGN subclass, is a highly revealing morphological characteristic among radio AGNs. Narrow-Angle-Tail (NAT) and Wide-Angle-Tail (WAT) galaxies are examples of bent sources. The quantity and geometry of bending are related to the interaction between the relativistic jets and the surrounding medium. Sharply bent, "V"-shaped jets with an opening angle $\theta_{OA} < 90°$ are characteristic of NAT sources and are frequently explained as the result of high-velocity motion in the intracluster medium (ICM) [43]. In contrast, WATs have wider, "C"-shaped morphologies $\theta_{OA} > 90°$, which are often linked to slower-moving center dominating galaxies within clusters [44]. In addition to being morphological oddities, these bending geometries include important information regarding host galaxy dynamics, ambient densities, and pressures [29].

Ram pressure from the ICM acting on the jets as the host galaxy passes through the cluster gas is the main hydrodynamic physical process causing jet bending in RAGNs [45]. $P_{\text{ram}} = \rho_{\text{ICM}} v^2$, where v) is the galaxy's relative velocity and $\rho_{\text{ICM}}$ is the ICM density, yields the ram pressure. The jet bends to produce the distinctive NAT and WAT morphologies when ram pressure is greater than the internal jet pressure and collimation forces [46]. Jet morphological complexity is also influenced by complementary environmental phenomena including buoyancy, magnetic fields, and cluster weather, which includes turbulence and shocks [47]. According to early observational studies, bending morphologies were common in clusters and were useful indicators of crowded settings [48]. Since then, systematic, large-scale studies have measured the environmental dependency of bending characteristics and found relationships with dynamical state, temperature, and cluster mass [49].

Cluster expansion and galaxy evolution are closely related to the evolution of bent radio emitters. In addition to immediate environmental characteristics, they also represent the history

of accretion and dynamical events like cluster mergers, which cause shocks and turbulence that impact energy transfer and jet shape [50]. Thus, bent RAGNs may be used as natural laboratories to investigate gas dynamics and feedback processes at various sizes.

Traditionally, bending radio AGN morphological categorization relied on eye examination; however, automated approaches are now required because to the growing scope and resolution of contemporary radio surveys [18]. Leveraging bending RAGNs in environmental and evolutionary investigations requires developing reliable, repeatable categorization algorithms connected to well-understood physical interpretations [19].

### 2.7.2 Radio Continuum Surveys and Data Sources

Comprehensive radio continuum surveys that offer the required spatial resolution, sensitivity, and sky coverage are essential for studying bent radio Active Galactic Nuclei (AGNs) and their surroundings. Large samples of Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT) sources across a variety of redshifts and environments have been found and classified thanks to a number of significant radio surveys that have transformed our knowledge of radio galaxies over the past few decades [51].

The Faint Images of the Radio Sky at Twenty Centimeters (FIRST) study using the Very Large Array (VLA) [14] is one of the most important surveys in this regard. In-depth morphological examination of radio emitters spanning almost 10,000 square degrees in the northern sky was made possible by FIRST's arcsecond resolution and sensitivity down to a few mJy. Combining optical surveys like the Sloan Digital Sky Survey (SDSS) with FIRST's vast catalog of radio sources has made it easier to conduct systematic studies of bent radio galaxies, resulting in the creation of enriched catalogs with morphological classifications [52]. Because it strikes a balance between sensitivity and resolution, FIRST is still a vital tool for locating structural elements that point to environmental interaction.

In addition to FIRST, the LOw Frequency ARray (LOFAR) Two-Metre Sky Survey (LoTSS) provides previously unheard-of sensitivity to diffuse, low-surface-brightness emission at 150 MHz on greater angular scales [53]. The identification of long radio lobes and tails, which are occasionally overlooked or resolved out in higher frequency surveys, is improved by LoTSS's increased sensitivity and low-frequency observations. This makes it possible to find fainter and farther-off bent sources, providing vital information on radio AGN populations in high-redshift clusters and impoverished groupings. The complimentary functions of varying sensitivity and resolution in bent radio galaxy studies are highlighted by the synergy between FIRST and LOFAR maps.

Radio observations are crucially contextualized by multi-wavelength data. Redshift estimate, environmental characterisation, and host galaxy identification are made possible by cross-matching radio sources with optical and infrared catalogs from SDSS, the Dark Energy Survey (DES), and the Wide-field Infrared Survey Explorer (WISE) [17]. By linking radio features to local densities, star masses, and galaxy kinds, these databases provide accurate morphological categorization and environmental investigations.

Automated categorization techniques are becoming crucial due to the exponential increase

in data volume and complexity from present and future radio surveys. While machine learning algorithms are increasingly used to scale classifications with rigorous repeatability [54, 55], initiatives like Radio Galaxy Zoo use citizen science for visual morphology classification [18].

### 2.7.3 Machine Learning in Radio Galaxy Classification

Given the complexity and diversity of radio sources and the growing amount of data produced by next-generation radio telescopes, radio galaxy morphology categorization is an important but difficult challenge. When faced with massive surveys that produce millions of sources, like the LOFAR Two-Metre Sky Survey (LoTSS) and the planned SKA [20, 32], traditional approaches that rely on professional visual examination are becoming unsustainable. Machine learning (ML) and deep learning (DL) approaches, which automate categorization, use complicated features, and enhance scalability and repeatability, have become more popular as a result [56, 54]. Traditional Fanaroff-Riley (FRI/FRII) types were classified with excellent accuracy in early applications, laying the groundwork for more sophisticated morphologies like bent-tail (WAT/NAT) classifications [56]. By utilizing pretrained image models [57], transfer learning and data augmentation further alleviate limited labeled data.

Class imbalance, source label uncertainty, and dataset heterogeneity across various instruments and resolutions continue to remain obstacles despite achievements [22]. Semi-supervised learning (SSL) techniques, which integrate large amounts of unlabeled and labeled data, have been developed to solve them. With little annotation work, SSL enables the training of reliable feature extractors on unlabeled samples, enhancing downstream classification [58]. Without explicit supervision, self-supervised techniques like as contrastive learning and Bootstrap Your Latent (BYOL) generate discriminative embeddings tailored to radio galaxy characteristics [1].

Another recent development is rotation-equivariant CNN architectures, which directly encode the reflectional and rotational symmetries of radio images [59]. These models improve classification robustness and generalization by lowering the sample complexity required to learn orientation-independent features. Thus, equivariant networks augment data-driven representation learning by naturally taking use of the physics-informed symmetries seen in radio AGN topologies [2].

In radio galaxy classification, uncertainty quantification in DL models using Bayesian neural networks (BNNs) and Monte Carlo dropout is becoming more popular [60]. Identification of unclear or out-of-distribution sources is made easier by well-calibrated uncertainty estimations, which also improve trust in automated processes and direct human involvement in important circumstances [61].

In order to improve interpretability and include multi-wavelength data, recent research additionally combine classic machine learning techniques like gradient boosting and random forests with DL-extracted features [62]. Hybrid techniques adapt to increasingly rich datasets by utilizing the advantages of quick training, feature fusion, and model interpretability.

In order to eliminate artifacts and false sources from radio pictures that might deceive classifiers, data preparation and masking approaches are still crucial [35]. Tools like PyBDSF are used in automated pipelines for source extraction and masking, which enhances model

robustness and subsequent classification performance [63].

Community initiatives like the Radio Galaxy Zoo (RGZ) citizen science project offer useful labeled data to support machine learning model training and validation. citebanfield2015. In order to advance morphological categorization in the face of an increase in data volume, human and machine intelligence must work together.

### 2.7.4 Semi-Supervised Learning and Data Preprocessing

The increasing amount of radio astronomy data from contemporary surveys calls for effective classification techniques that lessen reliance on properly labeled datasets, which are labor-intensive and time-consuming to create. In order to increase classification performance, semi-supervised learning (SSL) has gained popularity as a potent technique that blends a little quantity of labeled data with a large amount of unlabeled data. [33]. SSL uses the data structure and implicit labels included in unlabeled samples to bridge the gap between supervised and unsupervised learning.

Contrastive learning frameworks like SimCLR [64] and self-distillation techniques like Bootstrap Your Own Latent (BYOL) [1], which learn rich data representations without needing manual labeling, are examples of recent advancements in SSL. These methods are increasingly being used to astronomical picture classification challenges after being used in computer vision domains [65, 66]. They enable models to successfully represent intrinsic morphology by promoting consistency and invariance in feature space across various data augmentations.

By explicitly incorporating symmetric transformations like rotations and reflections, which are often seen in radio source pictures, rotation-equivariant convolutional neural networks (CNNs), first shown by [34], expand these self-supervised frameworks. By guaranteeing that learnt features are invariant or equivariant under reasonable physical transformations, such designs improve model resilience and decrease sample complexity [2]. In radio galaxy morphology classification, where orientations can differ greatly, this is especially helpful.

Due to complicated imaging artifacts and false sources produced by interferometric calibration and deconvolution processes, data preparation is crucial in radio astronomy [35]. Conventional source extraction methods, such as PyBDSF [35], fit source models and identify radio components based on local maxima, but they frequently contain artifacts or blended sources that might interfere with categorization.

To increase classifier accuracy and generalizability, studies have highlighted the significance of robust mask creation and spurious source removal [67]. The efficacy of semi-supervised feature learning on actual survey data is increased by appropriate preprocessing, which stabilizes input data distributions and lessens noise effects.

### 2.7.5 Deep Learning Model Design and Training

Because deep learning (DL) models can automatically build hierarchical feature representations from image data, they have become essential tools for morphological categorization in radio astronomy. In particular, Convolutional Neural Networks (CNNs) have revolutionized the

categorization of radio galaxies by capturing intricate spatial patterns without requiring manual feature engineering. [56]. In order to differentiate between morphological classes like Fanaroff-Riley types and bending source subclasses like Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT), the standard architecture consists of several layers of convolution, nonlinear activation, pooling, and fully linked neurons.

Input preprocessing, architectural depth and complexity, loss function selection, and integration of domain-specific invariances are all factors to be taken into account while designing a model for radio galaxy classification. Standard architectures like AlexNet or VGGNet [68], modified for radio images, were used in early CNN applications. In order to highlight pertinent pixels and capture delicate morphological traits across scales, more recent architectures use residual connections [69] and attention processes [70]. For instance, [71] achieved state-of-the-art results on radio galaxy datasets by combining deep residual networks with batch normalization and dropout.

It is essential to take into consideration rotational and reflection symmetries in radio pictures. Convolutional kernels in symmetry-aware architectures, such E(2)-equivariant CNNs (E2CNNs) [34, 59], incorporate intrinsic invariance or equivariance to group transformations. This increases resilience and physical interpretability in addition to sampling efficiency. According to recent research, using such equivariances is very useful for categorizing bent and elongated jet morphologies regardless of direction [2].

Class imbalance, noise, and label uncertainty must be carefully handled when training deep learning models using radio galaxy data since these factors can skew models toward dominant or simple-to-classify categories [72]. Generalization across uncommon bent-tail morphologies is enhanced by data augmentation techniques and loss functions designed for imbalance, such as focused loss [73].

Large unlabeled radio datasets may be used to pretrain feature extractors using semi-supervised and self-supervised techniques. Supervised fine-tuning on carefully chosen labels comes next. Recent studies, such as [58] and [1], show that self-supervised techniques like SimCLR and Bootstrapped Your Latent (BYOL) may achieve competition-class performance. For radio data, where orientation-invariant features are crucial and human annotations are costly, these techniques promote representation learning through consistency under augmentations.

Visualization techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) [74] support model explainability, which is essential for scientific acceptability. Interpretability also helps identify spurious correlations or imaging artifacts that may mislead classifiers [75]. Model evaluation involves multiple metrics to assess classification quality comprehensively, including accuracy, precision, recall, F1-score, Area Under the Receiver Operating Characteristic Curve (ROC-AUC), and calibration errors The latter is critical because well-calibrated models provide trustworthy confidence estimates, which are necessary for follow-up observations

# Chapter 3

# Research Methodology

Using curated labeled samples and large-scale radio survey data, this study uses a semi-supervised deep learning architecture to categorize bending radio Active Galactic Nuclei (AGN) into Wide-Angle-Tail (WAT) and Narrow-Angle-Tail (NAT) morphologies. In order to create a reliable and accurate classifier appropriate for massive astronomical datasets, the technique combines sophisticated data preparation, model design that prioritizes symmetry-aware structures, and stringent training and assessment procedures.

## 3.1 Data Collection and Sources

The main sources of data are labeled bent radio AGN samples selected from the literature, particularly the Sasmal et al. (2022a) catalog, and unlabeled radio galaxy pictures taken from the Radio Galaxy Zoo (RGZ) project. About 20,000 Radio Galaxy Zoo photos with different morphologies make up the unlabeled corpus, which offers a rich but sparsely labeled dataset perfect for self-supervised learning techniques [36]. Supervised fine-tuning and performance benchmarking are made possible by the labeled sets, which comprise 639 carefully examined WAT and NAT sources.

## 3.2 Preprocessing Data

To guarantee high-quality inputs, accurate source extraction and artifact abatement are essential. To create initial masks of radio emission components, the technique makes use of PyBDSF, a well-known radio source extraction tool [35]. Nevertheless, Photutils, a Python library intended for astronomical picture segmentation and background estimation [76], is used to further enhance PyBDSF, which can retain misleading sources like noise peaks and sidelobes. Two datasets are created using this dual-mask technique: RL1, which keeps spurious sources, and RL2, which eliminates them. The impact of preprocessing on model robustness is evaluated by comparing outcomes across various datasets.

To make it easier for the neural network to import and process images, they are normalized and transformed from FITS to PNG formats. By simulating observational and inherent

feature variability, data augmentation techniques like random rotations, flips, and brightness modifications improve the model's capacity for generalization.

## 3.3 Model Architecture and Design

For pretraining on unlabeled data, the research approach uses Bootstrap Your Own Latent (BYOL), a self-supervised learning paradigm [1]. Given the large amount of unlabeled data available, this pretraining makes it possible for the model to learn invariant and meaningful feature representations without the need for explicit labeling.

To make it easier for the neural network to import and process images, they are normalized and transformed from FITS to PNG formats. By simulating observational and inherent feature variability, data augmentation techniques like random rotations, flips, and brightness modifications improve the model's capacity for generalization.

## 3.4 Training Procedures

Pretraining and fine-tuning are the two stages of training. Pretraining makes use of the unlabeled RGZ dataset, where feature embeddings are learned by feeding picture pairings with random augmentations via the BYOL architecture. To train the classification head for binary WAT/NAT labels, fine-tuning makes use of the labeled bending morphology datasets.

Batch sizes, learning rates, and epoch numbers are optimized experimentally. Regularization strategies such as weight decay and dropout are employed to promote generalization. Based on performance metrics on a held-out validation set, model selection and early halting are determined by accuracy, F1-score, and expected calibration error.

## 3.5 Interpretability and Evaluation Metrics

Accuracy, precision, recall, F1-score, ROC-AUC, and calibration metrics are used in model evaluation to provide a complete assessment of classification quality [77, 78]. Gradient-weighted Class Activation Mapping (Grad-CAM) improves interpretability by visualizing important picture areas that influence classification choices [74]

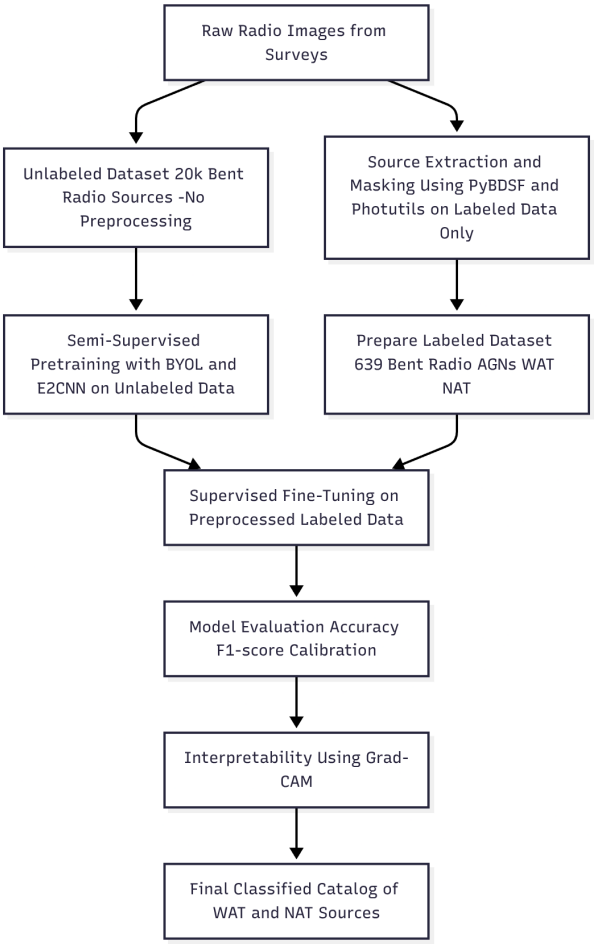



Figure 3.1: Flowchart of the research methodology. Raw radio images are subjected to source extraction and masking only for labeled samples to prepare a clean labeled dataset of 639 bent radio AGNs. The large unlabeled set of 20,000 bent radio sources remains unprocessed for pretraining. Semi-supervised pretraining via BYOL with an E2-equivariant CNN encoder is performed on the unlabeled data, followed by supervised fine-tuning using the preprocessed labeled dataset. The trained model undergoes evaluation and interpretability analysis leading to the final bent radio AGN classification catalog.

# Chapter 4

# Work Breakdown Structure (WBS)

The Work Breakdown Structure (WBS) provides a hierarchical decomposition of the research project into manageable components, facilitating effective planning, organization, and monitoring of progress. For this study on semi-supervised deep learning classification of bent radio galaxies, the WBS encapsulates all major research activities spanning data acquisition, preprocessing, model development, evaluation, and dissemination.

- **1. Data Collection and Curation**
  - 1.1 Acquisition of radio survey data
  - 1.2 Compilation of labeled bent radio AGN datasets from literature catalogs.
- **2. Data Preprocessing**
  - 2.1 Source detection and masking using PyBDSF.
  - 2.2 Removal of spurious sources using Photutils and artifact filtering.
- **3. Model Design and Development**
  - 3.1 Implementation of self-supervised learning framework (e.g., BYOL).
  - 3.2 Integration of rotation-equivariant CNN architectures.
  - 3.3 Development of supervised fine-tuning pipeline for WAT/NAT classification.
- **4. Training and Optimization**
  - 4.1 Hyperparameter tuning including learning rate, batch size, optimizer settings.
  - 4.2 Application of regularization methods such as dropout and early stopping.
  - 4.3 Evaluation using cross-validation and holdout datasets.
- **5. Model Evaluation and Interpretability**
  - 5.1 Computation of performance metrics: accuracy, precision, recall, F1-score, ROC-AUC, and calibration errors.

– 5.2 Implementation of interpretability techniques such as Grad-CAM visualization.

– 5.3 Analysis of uncertainty quantification and model confidence scores.

# Chapter 5

# Data and preprocessing

As mentioned before, both our labeled ($\mathbf{R}_L$) and unlabeled ($\mathbf{R}_U$) datasets are created using FIRST images. Although the angular resolution of FIRST was 5 arcsec at 1400 MHz (Table 1.1), the final images have a pixel resolution of 1.8 arcsec. The images we use in training and testing have a dimension of 150 pixels on either side, giving an angular size of 4.5 arcmin.

## 5.1 Unlabeled data

We have used the batched dataset of 20,000 sources created for use in ML by [79] from RGZ. As described in their Section 3.2 and in [19], RGZ data release 1 (DR1) contains approximately 100,000 sources, 99.2% of which were taken from FIRST. For each FIRST object in this catalog with the largest angular size ($\theta_{LAS}$) between 15 and 270 arcsec, [79] downloaded the corresponding image using the `Python API` of `SkyView`, a virtual observatory that provides access to a wide range of astronomical images from different surveys.[1] They cropped the inner $150^2$ pixels from the original $300^2$-pixel images, and put all pixels outside a radius of $0.6\theta_{LAS}$ to zero. After applying a $3\sigma$ amplitude thresholding, each original image $\mathbf{R}$ was normalized to

$$\mathbf{R}_n = 255 \times \frac{\mathbf{R} - min(\mathbf{R})}{max(\mathbf{R}) - min(\mathbf{R})} \tag{5.1}$$

in order to convert them to PNG (Portable Network Graphics) format, where $min(\mathbf{R})$ is the minimum value of $\mathbf{R}$, and $max(\mathbf{R})$ is the maximum. Thereafter all unresolved sources were removed, and the dataset was further reduced by discarding sources that were too dissimilar to their labeled dataset of FRI and FRII sources, the MiraBest dataset [61], which was also used in H23. The similarity between RGZ DR1 and MiraBest was analyzed using Fréchet inception distance (FID), as described in their Section 4.3. Most importantly, the labeled sources of MiraBest that were present in RGZ were also removed, so that there is no overlap between the labeled and unlabeled data. The final unlabeled dataset contained 20,000 sources organized in 10 batches of training data, and one batch of test data. We downloaded these from their GitHub

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

repository[2] and used them without any further processing. Hereafter, this batched dataset will be referred to as $\mathbf{R}_U$.

[80] uses a larger unlabeled dataset of 108,000 sources from RGZ for self-supervised learning, but pre-training on such a large dataset is beyond the scope of our present work.

## 5.2 Labeled data

Unlike $\mathbf{R}_U$, our labeled dataset $\mathbf{R}_L$ has been preprocessed for efficient use in ML in this work for the first time. As mentioned in Section 2.7.1, this dataset is created using the catalog of more than 700 WATs and NATs published by [23] (hereafter, the Sasmal catalog). However, it was created through visual inspection and, hence, there was no batched dataset ready for use in ML based on the catalog. We describe below how we refined the Sasmal catalog and prepared an ML-friendly batched dataset.

## 5.3 Downloading

We accessed the Sasmal catalog from `VizieR` using the `astropy` module `astroquery`. It contains two tables: Table 1 lists 430 WATs, while Table 2 lists 287 NATs. The different columns of the table have been described above in Section 2.7.1. The catalog was downloaded in `VOTable` format and converted to a `Pandas DataFrame` for further processing. To facilitate automated catalog retrieval, we developed the `catalog_quest` function within `RGC`, which directly queries VizieR and returns the data in a structured `DataFrame` format.

The coordinates of the sources were used to download the corresponding FIRST images through the `Python API` of `SkyView` within `astroquery`. The images were retrieved in `FITS` (Flexible Image Transport System) format. To facilitate the automated download of images, we have developed a function within `RGC` that takes the survey name (e. g., FIRST), source coordinates and output file path as input, and downloads the images of the sources to the specified file path. Additionally, we provide a function which enables batch downloading of all sources in a given catalog. This function takes the catalog `DataFrame`, survey name, and directory path as input and downloads the sources in the catalog to the specified directory.

Among the 717 bent RAGNs of the Sasmal catalog, we could successfully retrieve 703. The remaining 14 sources could not be retrieved due to either server errors during the download, or the downloaded image being blank. Among the 703 images, 281 were NATs and 422 were WATs. We preprocessed these images and created a refined annotated catalog for efficient use in ML which will be called Sasmal-ML hereafter. The preprocessing was performed using two separate software: `PyBDSF` [35] and `Photutils` [81]. The labeled dataset $\mathbf{R}_L$ created through `PyBDSF` processing will be called $\mathbf{R}_{L1}$ and the one created through `Photutils` processing will be called $\mathbf{R}_{L2}$.

Because of the way the images were downloaded, the target is always positioned at the center of a $150^2$-pixel image. During the first step of preprocessing, we created a mask that can

[2] `https://github.com/inigoval/fixmatch`

be applied to the image to remove everything except the central target. This was performed in two steps: first, a mask was generated using PyBDSF for $\mathbf{R}_{L1}$ and, then, the mask was refined using Photutils for $\mathbf{R}_{L2}$. The two steps are described in the subsequent two subsections.

## 5.4 Generating masks using PyBDSF

Estimating each image's background noise level is the first step in creating a mask. By dividing the image into smaller subregions and calculating the local mean ($\mu_{\text{local}}$) and standard deviation ($\sigma_{\text{local}}$) of the pixel intensities inside each subregion, PyBDSF accomplishes this. According to the PyBDSF documentation, the background intensity of an original image $\mathbf{R}$ at a pixel $(x, y)$

$$B(x, y) = \mu_{\text{local}} + k \cdot \sigma_{\text{local}} \tag{5.2}$$

where $k$ is a scaling factor that takes common noise fluctuations into account. Following background estimate, sources are found by locating adjacent areas known as "islands" where the intensity is higher than the local background by a predetermined threshold. There are two different kinds of thresholds used. The lowest signal level required to classify a collection of adjacent pixels as an island is known as the "island threshold" $T_{\text{isl}}$. If a pixel's intensity

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

where, based on eye assessment, we employed $T_{\text{isl}} = 3$. An island's highest intensity is known as the 'peak threshold' $T_{\text{pix}}$, and only islands with peaks above

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

are regarded as legitimate findings. Based on visual inspection, we chose $T_{\text{pix}} = 5$. Contiguous pixels that satisfy the requirements are grouped into islands when the thresholds are applied. The centroid $(x_0, y_0)$, peak intensity $_0$, and morphological characteristics, such as major axis $\sigma_{\text{maj}}$, minor axis $\sigma_{\text{min}}$, and position angle $\theta$, are extracted from each island using a two-dimensional Gaussian model. The model of Gaussian

$$R_G(x, y) = R_0 \exp\left(-\frac{(x' - x_0)^2}{2\sigma_{\text{maj}}^2} - \frac{(y' - y_0)^2}{2\sigma_{\text{min}}^2}\right) \tag{5.3}$$

where the coordinates rotated by the angle $\theta$ are represented as $(x', y')$. Following source identification, a binary mask is created to show each source's spatial extent. The fitted Gaussian model is thresholded to produce the mask $\mathbf{M}_{L1}$ as

$$M_{L1}(x, y) = \begin{cases} 1, & \text{if } R_G(x, y) > B(x, y) + T_{\text{pix}} \cdot \sigma_{\text{local}} \\ 0, & \text{otherwise} \end{cases} \tag{5.4}$$

as shown in the Fig. 5.1. Our sources were expanded, thus in order to guarantee full coverage of the source, the masks were dilated by d pixels. The final binary masks are saved in FITS format.

The subsequent stages were carried out using the `process_image` function of `PyBDSF`, which accepts as inputs the beamshape, frequency, and the previously specified threshold parameters and outputs a `PyBDSF` object with the sources that were found. We export the sources' mask using the `export_fits` method, using a dilation of $d = 0$ for the majority of the sources and other values for a few instances where modifications were required. Sasmal-ML, the catalog included with this paper, identifies the photos for which various dilations have to be applied.

We automate this procedure with the function `generate_mask` in `RGC` for convenience. It creates the matching mask as an output after receiving the picture file name, beamshape, frequency, threshold values, dilation, and output file name as inputs. We also provide `generate_mask_bulk` to process an entire image catalog.

## 5.5 Refining Masks Using `Photutils`

We implemented an extra, independent masking stage based on the `Photutils` package [82] to enhance the isolation of the target radio galaxy and decrease irrelevant artifacts. This refinement is done in two stages: first, we use `Photutils` to create a source segmentation–based mask $\mathbf{M}_L$ from each FIRST FITS image; second, we take the Hadamard (element-wise) product of this mask with the original `PyBDSF` mask to obtain a conservative final mask that keeps only regions jointly supported by both methods.

Starting with the calibrated FIRST cutoff $R(x, y)$, we create the `Photutils` mask $\mathbf{M}_L$ by substituting zeros for any incorrect pixel values (NaN or infinite). In order to provide a well-defined numerical image for later statistics and segmentation, flagged or missing data are set to the background level in several radio-imaging pipelines. Operationally, this is implemented via `numpy`'s `nan_to_num` function, which ensures that

$$R'(x, y) = \begin{cases} R(x, y), & \text{if } R(x, y) \text{ is finite,} \\ 0, & \text{if } R(x, y) \text{ is NaN or } \pm\infty \end{cases} \tag{5.5}$$

and all subsequent processing is carried out on $R'$.

Next, each image's global background properties are estimated using sigma-clipped statistics, which are implemented in `Astropy`'s `sigma_clipped_stats` function [83]. For radio interferometric pictures, where bright sources and sidelobes cause hefty non-Gaussian tails, this approach iteratively excludes pixels that significantly vary from a robust central value. By removing any pixel that deviates more than $3\sigma$ from the mean in each iteration, we are able to determine the image's median $m$, mean $\mu$, and standard deviation $\sigma$:

$$(\mu, m, \sigma) = \text{SigmaClippedStats}\big(R', \sigma_{\text{clip}} = 3.0\big) \tag{5.6}$$

By reducing the impact of bright sources and artifacts, sigma clipping offers a reliable approximation of the usual background and noise level

A global detection threshold is then defined as

$$T = m + k'\sigma \tag{5.7}$$

where $k'$ is an adjustable parameter that regulates the trade-off between purity (avoiding noise fluctuations) and completeness (recording faint extended emission). We chose $k' = 2.5$ or 3.0 for the majority of photos based on visual examination of a representative subset of WAT and NAT sources and testing with other values. Since the objective is to track the long tails and lobes of bent radio galaxies rather than to create a catalogue of isolated compact components, this is purposefully lower than standard peak-detection thresholds used in catalog production (e.g., $5\sigma$).

The `detect_sources` function of `Photutils` is used for source detection. It finds groupings of linked pixels over the threshold $T$ and gives them segment labels [82]. First, conceptually, we establish a binary threshold image

$$S(x, y) = \begin{cases} 1, & R'(x, y) > T \\ 0, & \text{otherwise} \end{cases} \tag{5.8}$$

where pixels that are noticeably above the predicted backdrop are indicated by $S(x, y) = 1$. Then, using a selected connection criteria (in this case, the default 8-connectivity), pixels with $S(x, y) = 1$ are grouped into contiguous sections ("segments'). We need a segment to have at least $N_{\min}$ linked pixels in order to reduce solitary noise spikes and extremely minor artifacts. Using the `npixels` parameter in the code, we implemented $N_{\min} = 10$–$20$ pixels, which is similar to the angular size of the smallest reliably identified FIRST sources at the corresponding resolution. Segments with fewer pixels than $N_{\min}$ are eliminated.

All pixels that do not belong to any segment are set to zero, and each remaining segment is given an integer label $l_i$. The labeled segmentation picture that results, denoted by $\mathbb{L}(x, y)$, may be expressed as

$$\mathbb{L}(x, y) = \begin{cases} l_i, & \text{if } (x, y) \text{ lies in segment } i \\ 0, & \text{otherwise} \end{cases} \tag{5.9}$$

Every source found in the field by `Photutils` is included in this segmentation.

We take use of the fact that each FITS cutout has been recentered such that the host galaxy sits close to the picture center $(x_c, y_c)$ in order to identify the target object associated with the WAT/NAT in the Sasmal et al. catalogue. First, we calculate the image's center as

$$(x_c, y_c) = \left( \frac{N_x - 1}{2}, \frac{N_y - 1}{2} \right) \tag{5.10}$$

where the picture dimensions in pixels are denoted by $N_x$ and $N_y$. Next, we look at $\mathbb{L}(x_c, y_c)$, the label at this center pixel. We consider a segment to be the core component associated with the target source if it is present in the precise center (i.e., $\mathbb{L}(x_c, y_c) = l_i \neq 0$).

Sometimes, though, no segment is found in the precise center (for instance, if the source

is broken up in a way that leaves the central pixel below threshold, or if the core is dim in comparison to the lobes). In order to address this, we expand the search to a square aperture of width $2r_{\text{search}}$ centered at $(x_c, y_c)$ if $\mathbb{L}(x_c, y_c) = 0$:

$$r_{\text{search}} = 25 \text{ pixels} \tag{5.11}$$

corresponds to a location in our $150 \times 150$ pixel cuts that easily encompasses the anticipated extent of the core emission. We take into consideration any non-zero labels $l_j\}$ found in this aperture.

$$\mathbb{L}_{\text{search}} = \mathbb{L}\big(x_c - r_{\text{search}} : x_c + r_{\text{search}}\ y_c - r_{\text{search}} : y_c + r_{\text{search}}\big) \tag{5.12}$$

then choose the label that covers the greatest number of pixels in that area—that is, the mode of the label distribution—as the center label. The selection of bright bent radio galaxies is consistent with the majority-label approach's assumption that the genuine target source is usually the most extensive structure in the core region.

Bent radio AGN may be divided into many components by the thresholding process and feature long jets and tails that curve away from the host galaxy. We further search an expanded radius $r_{\text{extend}}$ around the chosen center place in order to capture not just the core but also these adjacent regions. We used $r_{\text{extend}} = 10$–$20$ pixels in our algorithm, which was once more empirically adjusted to cover common WAT/NAT extents in FIRST while avoiding far-off irrelevant sources. We gather all distinct non-zero labels inside this square or circular area with radius $r_{\text{extend}}$.

$$\{\ell_1, \ldots, \ell_M\} = \left\{ \mathbb{L}(x, y) \,\middle|\, \sqrt{(x - x_c)^2 + (y - y_c)^2} \le r_{\text{extend}},\ \mathbb{L}(x, y) > 0 \right\} \tag{5.13}$$

where the number of unique segments connected to the central source and its immediate lobes is denoted by $M$.

The mask $\mathbf{M}_L$ is then created as a binary picture that contains every pixel that corresponds to any of these gathered labels:

$$M_L(x, y) = \begin{cases} 1, & \mathbb{L}(x, y) \in \{\ell_1, \ldots, \ell_M\} \\ 0, & \text{otherwise} \end{cases} \tag{5.14}$$

By construction, $\mathbf{M}_L$ eliminates segments that are farther out in the field and are probably noise, sidelobe artifacts, or unrelated sources, while isolating the major segmented components connected to the center bent radio galaxy.

We optionally add a binary morphological dilation to $\mathbf{M}_L$ to guarantee that the mask fully covers the source and somewhat overestimates its spatial extent. This is done in the implementation using `scipy.ndimage.binary_dilation`, which fills in minor gaps and slightly expands the bounds of the detected regions by replacing each "on" pixel ($M_L(x, y) = 1$) with a small neighborhood (usually a $3 \times 3$ structural element):

Table 5.1: Distribution of WATs and NATs in our labeled dataset $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$, created following the Sasmal-ML catalog demonstrated in Table 10.1.

| Type | Batch | WATs | NATs | Batch total | Type total |
|---|---|---|---|---|---|
| | 0 | 39 | 25 | 64 | |
| | 1 | 39 | 25 | 64 | |
| | 2 | 39 | 25 | 64 | |
| | 3 | 39 | 25 | 64 | |
| Train | 4 | 39 | 25 | 64 | 576 |
| | 5 | 38 | 26 | 64 | |
| | 6 | 38 | 26 | 64 | |
| | 7 | 38 | 26 | 64 | |
| | 8 | 38 | 26 | 64 | |
| Test | 9 | 38 | 25 | 63 | 63 |
| **Total** | 10 | 385 | 254 | 639 | 639 |

$$\mathbf{M}_L^{(\mathrm{dil})} = \mathcal{D}_{d'}\big(\mathbf{M}_L\big) \tag{5.15}$$

When dilation with a structural element size determined by $d'$ is indicated by $\mathcal{D}_{d'}$. While just slightly expanding the mask's size, this dilation aids in recovering the faint margins of the jets and tails that could fall just below the segmentation threshold in a few isolated pixels.

The original picture header is kept for provenance while the (optionally) dilated binary mask $\mathbf{M}_L$ is recorded as a FITS file. Lastly, the previously mentioned `PyBDSF` mask $\mathbf{M}_{L1}$ is merged with this `Photutils`-based mask. An element-wise (Hadamard) product yields the final combined mask $\mathbf{M}_{L2}$.

$$\mathbf{M}_{L2}(x, y) = \mathbf{M}_{L1}(x, y)\, M_L(x, y) \tag{5.16}$$

such that only pixels recognized by both `PyBDSF` and `Photutils` as belonging to the source are kept (see Fig. 5.1). While maintaining the core and extended structures of the bent radio galaxies that are most important for the ensuing morphological classification, this conservative intersection significantly lowers the amount of false sources brought forward to the machine-learning pipeline.

### Conversion to PNG

The downloaded images and the generated masks were in FITS format, which is not suitable for efficient use in ML. Therefore, we converted the images and masks to PNG, which is widely used for image processing. For this conversion, the images were first normalized following Equation 5.1. The normalized images and masks were then converted to PNG using `Pillow`. In Fig. 5.1, the leftmost panel shows the original normalized image in the PNG format, and the masks are also shown as PNGs in the next three panels.

RGC has the function `fits_to_png` to convert the FITS files to PNG format automatically. It takes the FITS file name as input and converts the image to PNG format and returns an

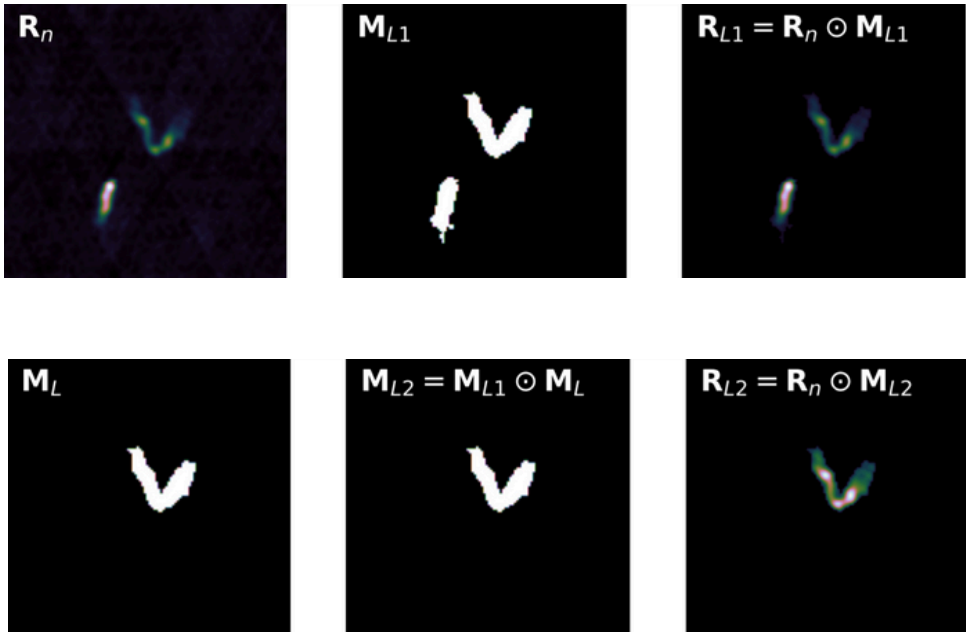



Figure 5.1: An example of the process of masking an image as described in Section 5.4 and 5.5. $\mathbf{R}_n$ is the normalized original image, $\mathbf{M}_{L1}$ its mask created using `PyBDSF` and $\mathbf{M}_L$ its mask produced through `Photutils`.

image in the form of a `PIL.Image` object. We also provide a function `fits_to_png_bulk` to convert all FITS files in a given directory to PNG format. The function takes the directory path and output directory path as input and converts all FITS files in the directory to PNG format and saves them in the output directory.

### Applying the masks

We remove background from the images by applying the binary masks (values either 0 or 1) generated in the previous steps. By multiplying a normalized image $\mathbf{R}_n$ element-wise with its binary mask, the background is masked, keeping only the source emissions, i.e. the masked image

$$\mathbf{R}_{L1} = \mathbf{R}_n \odot \mathbf{M}_{L1} \quad \mathbf{R}_{L2} = \mathbf{R}_n \odot \mathbf{M}_{L2} \tag{5.17}$$

which are shown in Fig. 5.1. RGC provides two functions named `mask_image` and `mask_image_bulk` in order to automate this process. The `mask_image` function takes `PIL.Image` objects of the image and the mask as inputs, applies the mask on the image, and returns the masked image as a `PIL.Image` object. The `mask_image_bulk` function processes an entire dataset by taking the directory paths of images and masks as input, applying the masks to all images, and saving the masked images in the specified output directory.

## 5.6 Sasmal-ML: a bent RAGN catalog and dataset for ML

The 703 images, masked, processed and converted following the aforementioned steps, were visually examined by three authors. The complete annotations are provided in our Sasmal-ML catalog provided in the GitHub repository of RGC as a `TSV` (tab separated values) file. Sixty-three rows of these catalog are shown in Appendix 10.1.

Sasmal-ML contains 703 rows and 16 columns, as described in Table 10.1 of the appendix. Among the 703 sources, 639 were finally selected for our datasets $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$. The reasons behind discarding 64 sources can be found in Sasmal-ML. Specifically, 31 sources were discarded because of their small (compact) size, 9 were discarded because of the confusion about their class (WAT or NAT), and 24 sources were discarded because of a combination of reasons, for example, low signal-to-noise ratio, low angular extent, and faint extended emission. The remaining 639 sources were given the label 'SMD' in the column named 'Hossain.' The last two columns of the catalog are related to the attention patterns of our DL model trained and tested on the data, to be discussed in results and analysis.

These 639 sources were used to create our batched dataset for ML. It contains 385 WATs and 254 NATs distributed in 9 training batches, and 1 test batch, as shown in Table 5.1. Each training batch contains around 39 WATs and 25 NATs, and the test batch contains 38 WATs and 25 NATs. The overall class imbalance is approximately 66%. The two datasets $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$ are given as a single `TAR.GZ` (tape archive, GNU zip) file in the GitHub repository of RGC.

# Chapter 6

# Architecture of the model

RGC combines different existing DL architectures to create a new framework for classifying RAGNs. It was first described by [7], or H23, where semi-supervised learning (SSL) methods were used to classify RAGNs into FRI and FRII types (see their Sections 3 and 4 for a detailed background). Here we use a refined version of RGC for classifying RAGNs into WATs and NATs and publish its Version 1.0 as a Python package for the first time.

Early examples of using SSL for FR classification used the contrastive learning method SimCLR [64], the self-supervised method BYOL [1], and the semi-supervised method FixMatch. H23 used SimCLR and BYOL in combination with an $E(2)$-equivariant steerable CNN (E2CNN; [59, 2]), and found BYOL to outperform SimCLR. Therefore, in this work, we include only BYOL with E2CNN. Both BYOL and E2CNN have been described sufficiently in the relevant papers, and in H23. Hence, below we focus mainly on the experimental setup of RGC 1.0 relevant for this work.

The RGC model has two stages: an unsupervised (or, equivalently, self-supervised) feature extraction, and a supervised fine-tuning. In the first stage, we pre-train RGC using BYOL on the unlabeled dataset $\mathbf{R}_U$. Here BYOL contains an E2CNN as an encoder, which preserves invariance under different transformations. In the second stage, we fine-tune the pre-trained model on the labeled dataset $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$ using supervised learning. So the first stage is task-agnostic, and the second stage task-specific where classification is performed as a downstream task. In Fig. 6.1, we show the first stage on the left, and the second stage on the right.

## 6.1 Self-supervised pre-training

The first stage of the RGC model is depicted on the left panel of Fig. 6.1. It shows how BYOL learns a representation $y_\theta$ that can be used for downstream tasks. Its online (top row) and target (bottom row) networks are similar, but asymmetric, and they have different weights $\theta$ and $\xi$, respectively. The online network has three stages: an encoder $f_\theta$, a projector $g_\theta$ and a predictor $q_\theta$. In our case E2CNN was used as the encoder $f_\theta$. BYOL uses its target network as regression 'targets' for training the online network. The process is shown from the left to the right of this panel. Given an image $x$, BYOL produces its two views $v$ and $v'$ by applying

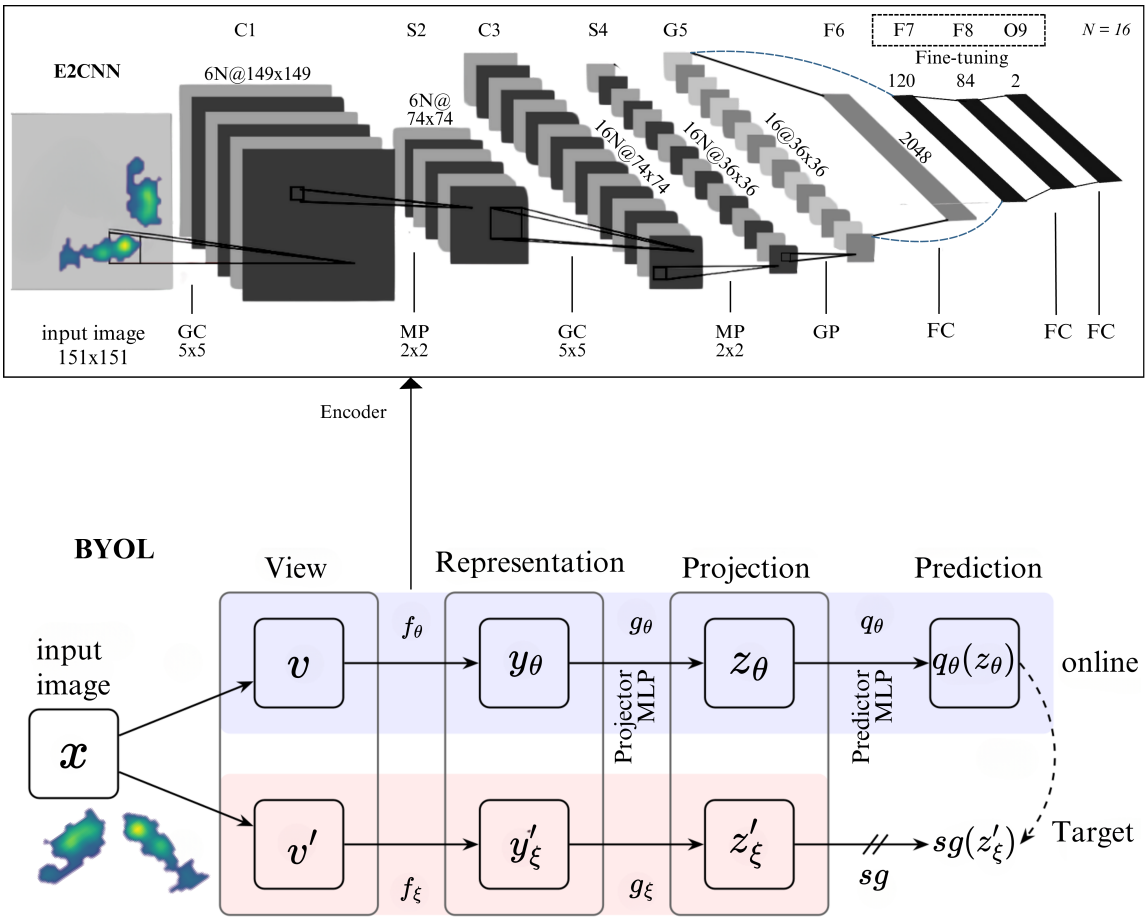



Figure 6.1: *Bottom*:The architecture of BYOL used in our self-supervised pre-training, adapted from [1]. The encoder within BYOL in our case is E2CNN which has been used for our supervised downstream fine-tuning. *Top*: The architecture of E2CNN as used by [2], adapted from the LeNet-5 architecture of [3]. The operations GC, MP, GP and FC stand for group convolution, max-pooling, group-pooling and fully-connected, respectively.

different augmentations $t$ and $t'$. From $v$, the online network creates a representation $y_\theta$ and a projection $z_\theta$; the target network produces the corresponding $y'_\xi$ and $z'_\xi$ from $v'$. Then the online network produces a prediction $q_\theta(z_\theta)$. Subsequently, $l_2$-normalization is applied to the online prediction and the target projection. The difference between these two $l_2$-normalized vectors is the loss function of BYOL, defined as (Equation 2 of [1])

$$\mathcal{L}_{\texttt{BYOL}} = \|\overline{q_\theta}(z_\theta) - \overline{z}'_\xi\|_2^2 \tag{6.1}$$

which is a mean-squared error (MSE). At each training step, a stochastic optimization is performed to minimize the loss with respect to $\theta$, but not $\xi$, which is shown by the stop-gradient (`sg`) in the left panel of Fig. 6.1.

The architecture of the encoder $f_\theta$ is based on E2CNN, which is designed to preserve invariance under transformations such as rotation and reflection, the Euclidean group $E(2)$. As shown on the right panel of Fig. 6.1, the encoder has 6 layers: the convolution layers C1 and C3, the subsampling layers S2 and S4, and the group-pooling layer G5. It follows the architectures of LeNet-5 [3] and AlexNet [84], modified by [2]. The input image has dimensions of $151 \times 151$ because of zero-padding. After a group convolution on the images with kernel size $5 \times 5$, we get $6N$ channels with dimensions $149 \times 149$, where $N = 16$ is the number of equivariant groups. After a max-pooling operation with a $2 \times 2$ kernel, the dimensions reduce to $74 \times 74$. Another group convolution and max-pooling reduce the dimensions to $36 \times 36$ and increase the number of channels to $16N$. Finally, a group-pooling operation combines the equivariant groups to create only 16 channels, which are flattened and passed through a fully-connected layer to produce a vector of dimension 2048.

## 6.2 Supervised fine-tuning

The second stage of the RGC model involves fine-tuning the pre-trained encoder on a labeled dataset of bent RAGNs to classify them into WATs and NATs. The model is trained to minimize the cross-entropy loss between the predicted and true labels of the input data, defined as

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N_c} y_i \log(p_i) \tag{6.2}$$

where $y_i$ is the true label of the input data, $p_i$ is the predicted probability of the input data belonging to the true class, and $N_c$ is the number of classes. The model is trained to minimize this loss by updating the parameters of the encoder using backpropagation. As shown on the right panel of Fig. 6.1, the fine-tuning stage takes the feature maps of G5 as input and passes them through two FC layers to produce two vectors of dimensions 120 and 84, consecutively. The output layer O9, fully connected with F8, finally gives us two logits which are used to calculate the class probabilities through the softmax function.

Table 6.1: Comparison of the hyperparameters used in our self-supervised pre-training and supervised fine-tuning.

| Hyperparameters | BYOL Pre-training | E2CNN Fine-tuning |
|---|---|---|
| Optimizer | Adam | Adam |
| Learning Rate | $3\times10^{-4}$ (constant) | $3\times10^{-4}$ (initial) |
| Batch size | 16 | 16 |
| Weight decay | | $10^{-6}$ |
| Scheduler | | ReduceLROnPlateau |
| Epochs | 500 (no early stopping) | 500 (early stopping) |

## 6.3 Training details

In Table 6.1, the hyperparameters used in our pre-training and fine-tuning are given. Both were performed in `Timaeus`, a high-performance Supermicro workstation containing an NVIDIA Leadtek QuadroRTX A4000 GPU with 16-GB memory.[1] During pre-training, we have used the Adam optimizer with no weight decay. The learning rate was kept constant throughout the 500 epochs. Because there were no labeled data in this stage, we continued the epochs without any early stopping until convergence of the loss.

During the supervised fine-tuning, we again used Adam optimizer, but this time with a weight decay, given in Table 6.1. The learning rate was the same as pre-training initially, but it was changed through a 'ReduceLROnPlateau' scheduler with a 'factor' of 0.9 and a 'patience' of 2. The scheduler waits for 2 epochs to evaluate the changes and then reduces the learning rate by 10%. Because we have labeled data in this stage, we implemented 'early stopping' based on the validation loss. TensorBoard tracks extensive measures such as precision, recall, and F1 score to monitor performance and reduce overfitting. Model checkpoints are saved on the basis of validation correctness.

## 6.4 Class activation mapping

In order to determine the significance of each spatial position on an image for a given class prediction, we used Grad-CAM where the gradients of any target flowing into the final convolutional layer are utilized to produce a coarse localization map $L^c$ highlighting the important regions in the image. For an input image $x$ and target class $c$, the process starts with a forward pass through the network to obtain the logit $y^c$ and then calculate the gradient $\partial y^c/\partial A^k_{ij}$ where $A^k_{ij}$ is the activation of a specific position $(i, j)$ in a feature map $k$. We calculated the 'importance weight' for each feature map as

$$\alpha^c_k = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{6.3}$$

[1]More details here: `https://cassa.site/hpc/`

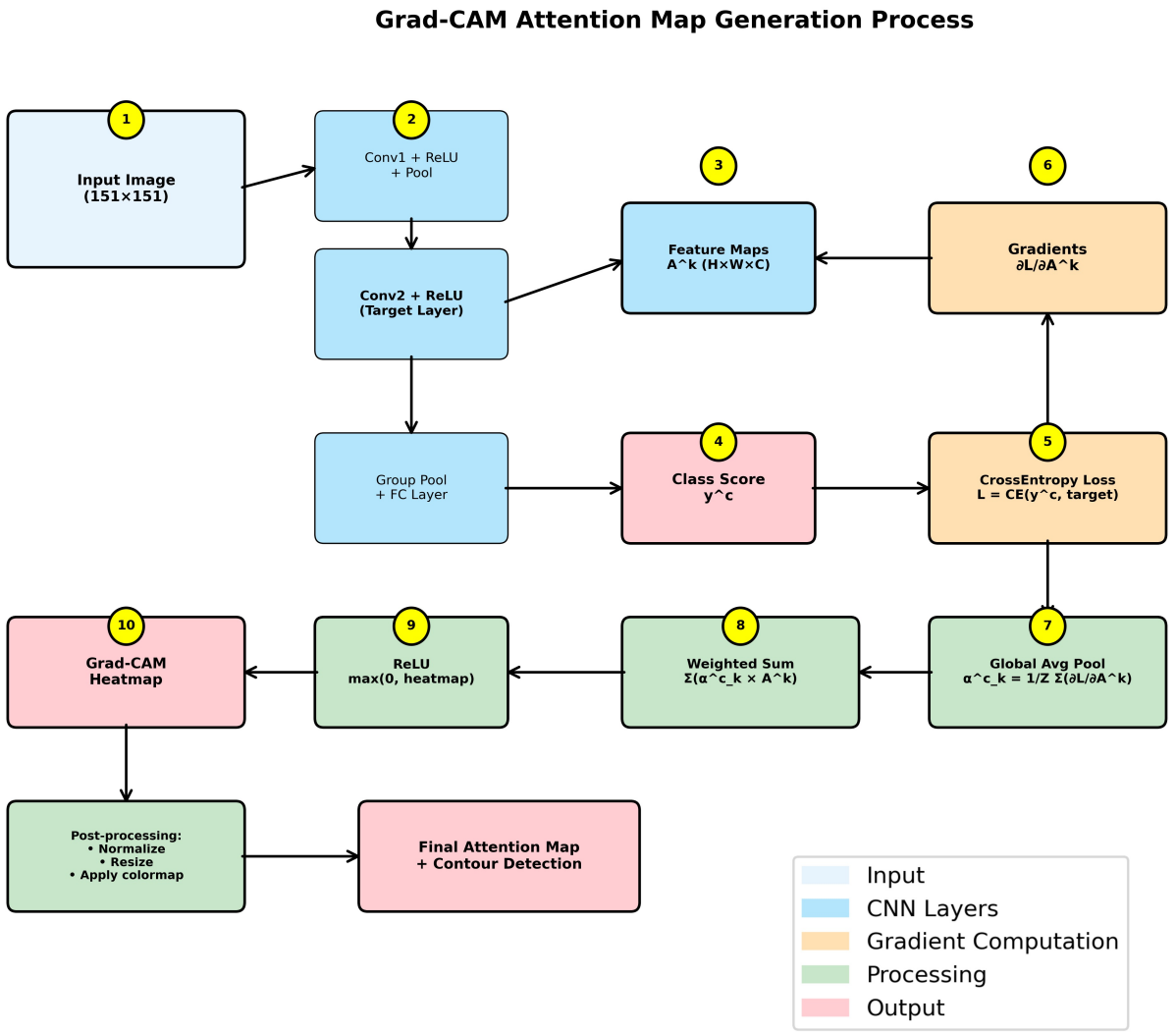



Figure 6.2: Grad-Cam implementation

where $Z$ is the total number of elements in the feature map. This parameter shows how much the feature map $k$ contributes to a class $c$. The feature maps are then multiplied with their corresponding importance weights and summed to produce

$$L^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \tag{6.4}$$

where the ReLU (Rectified Linear Unit) function ensures that only features with a positive influence on the class score are visualized. This procedure is used to produce a single 2D heatmap that highlights the regions of an image where the model paid the most attention. The heatmaps created from the dataset with spurious sources ($\mathbf{R}_{L1}$) is named $\mathbf{L}_{L1}$ and from the one without spurious sources ($\mathbf{R}_{L2}$ is named $\mathbf{L}_{L2}$, as shown in Appendix 10.2.

# Chapter 7

# Results & Analysis

The Radio Galaxy Classifier (RGC) is evaluated both quantitatively and qualitatively in this chapter for the purpose of classifying bent radio AGNs into wide-angle-tail (WAT) and narrow-angle-tail (NAT) morphologies. We examine the model using two closely similar test sets, $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$, which differ solely in how they handle unrelated and spurious radio emission in the field of view. While the second set, $\mathbf{R}_{L2}$, utilizes the improved masks from Chapter 5 to suppress sidelobes, background sources, and other pollutants, the first set, $\mathbf{R}_{L1}$, uses masks that contain all `PyBDSF`-identified components. The identical 63 labeled bent RAGNs (Table 5.1) are present in both sets, allowing for a controlled investigation of the impact of enhanced source isolation on downstream performance.

This is how we organize this chapter. For both masking procedures, we use confusion matrices to summarize the raw classification results in Section 7.1. In Section 7.2, we compare RGC to a set of fully supervised baselines and extract accuracy, precision, recall, and F1-score for WATs and NATs. These comparisons are covered in greater detail in Section 7.2, where the effects of semi-supervised pretraining and mask refining are highlighted. In Section 7.3, we use receiver operating characteristic (ROC) and precision–recall (PR) curves, together with the corresponding areas under the curve, to analyze the discriminative behavior of the classifier. The Expected Calibration Error (ECE) is then used in Section 7.4 to assess the estimated probability' dependability. Lastly, we summarize the key results and unresolved issues for bent radio galaxy classification and supplement these metrics with demographic and exploratory studies.

## 7.1 Confusion matrix

By comparing the predicted labels to the real labels on a held-out test set, a confusion matrix offers a condensed overview of a classifier's predictions. The confusion matrix for a binary classification issue like ours may be expressed as , where the positive class is assumed to be WATs and the negative class to be NATs.

$$\mathbf{C} = \begin{bmatrix} \mathrm{TP} & \mathrm{FN} \\ \mathrm{FP} & \mathrm{TN} \end{bmatrix} \tag{7.1}$$

, where each entry counts how many test samples fit into that category:

- *True Positives (TP)*: WATs correctly predicted as WATs;
- *True Negatives (TN)*: NATs correctly predicted as NATs;
- *False Positives (FP)*: NATs incorrectly predicted as WATs;
- *False Negatives (FN)*: WATs incorrectly predicted as NATs;

By construction, the total number of test samples $N$ meets

$$N = \mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}, \tag{7.2}$$

, and the distributions of true and predicted labels are recovered by the row and column sums of $\mathbf{C}$, respectively.

The RGC model produces a probability $p_i \in [0, 1]$ that the source is a WAT for every test sample $i$. By thresholding at 0.5, we transform these probabilities into hard class labels in order to construct the confusion matrices seen in Fig. 7.1:

$$\hat{y}_i = \begin{cases} 1, & p_i \geq 0.5 \\ 0, & p_i < 0.5 \end{cases} \tag{7.3}$$

where $\hat{y}_i = 1$ denotes a predicted WAT and $\hat{y}_i = 0$ a predicted NAT, and $y_i \in \{0, 1\}$ is the corresponding true label from the Sasmal et al. catalogue. The entries of the confusion matrix are then obtained by summing over the test set:

$$\mathrm{TP} = \sum_{i=1}^{N} \mathbb{1}\left(y_i = 1 \wedge \hat{y}_i = 1\right) \tag{7.4}$$

$$\mathrm{TN} = \sum_{i=1}^{N} \mathbb{1}\left(y_i = 0 \wedge \hat{y}_i = 0\right) \tag{7.5}$$

$$\mathrm{FP} = \sum_{i=1}^{N} \mathbb{1}\left(y_i = 0 \wedge \hat{y}_i = 1\right) \tag{7.6}$$

$$\mathrm{FN} = \sum_{i=1}^{N} \mathbb{1}\left(y_i = 1 \wedge \hat{y}_i = 0\right) \tag{7.7}$$

where $\mathbb{1}(\cdot)$ is the indicator function.

Fig. 7.1 displays the confusion matrices produced by assessing the RGC model on the two test sets $\mathbf{R}_{L1}$ (left) and $\mathbf{R}_{L2}$ (right). In order to immediately interpret the diagonal elements as empirical estimates of the class-wise recall, each matrix is normalized by the total number of true samples in that class:

$$\mathrm{Recall}_{\mathrm{WAT}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad \mathrm{Recall}_{\mathrm{NAT}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{7.8}$$

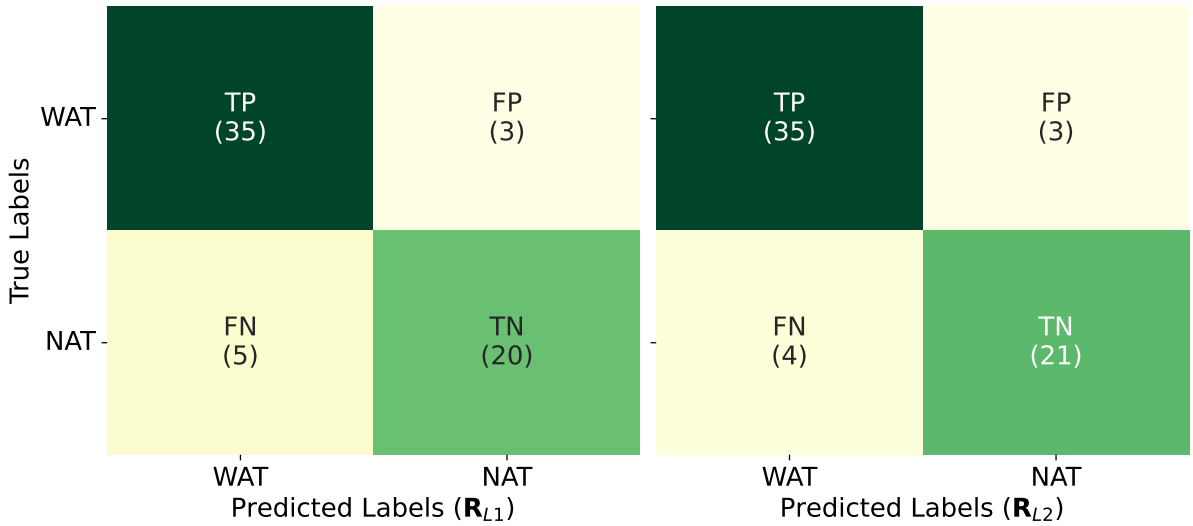



Figure 7.1: Confusion matrices for the classification of bent RAGNs into WATs and NATs, showing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for the RGC model trained on the datasets $\mathbf{R}_{L1}$ (left) and $\mathbf{R}_{L2}$ (right).

The confusion matrix for the revised, spurious-free dataset $\mathbf{R}_{L2}$ reveals that the model correctly identified 21 out of 25 NATs as NATs (TN = 21) and 35 out of 38 WATs as WATs (TP = 35). 3 WATs classified as NATs (FN = 3) and 4 NATs labeled as WATs (FP = 4) are examples of misclassifications. While correctly detecting 20 of the 25 NATs (TN = 20) and incorrectly categorizing 5 NATs as WATs (FP = 5), the model retains the same performance for WATs (TP = 35, FN = 3) when spurious sources are present in the images ($\mathbf{R}_{L1}$). Consequently, the total quantity of test samples is still

$$N = 38_{\text{WAT}} + 25_{\text{NAT}} = 63 \tag{7.9}$$

for both datasets, allowing a direct comparison.

Qualitatively, these confusion matrices show that RGC has a high recall for WATs under both masking schemes and that extra false positives for NATs represent the majority of the deterioration caused by spurious sources. To put it another way, the model is significantly more likely to perceive unrelated emission or artifacts in the field as extra WAT-like structure, which increases the confusion between NATs and WATs. At the level of raw counts, the overall differences between $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$ are small, but they hint at the slight but steady gains in accuracy and F1-score covered in Section 7.2 and the improved calibration looked at in Section 7.4.

## 7.2 Classification metrics

It is frequently helpful to summarize classifier performance using scalar metrics that capture many elements of behavior, even if the confusion matrix offers a straight count of accurate and

Table 7.1: Performance metrics of the RGC model on two datasets, with ($\mathbf{R}_{L1}$) and without ($\mathbf{R}_{L2}$) spurious sources, along with five additional fully supervised models trained and tested on $\mathbf{R}_{L1}$ [4]. See Section 7.2 and Section 7.2 for details.

| Model | Acc. [%] | WAT | | | NAT | | |
|---|---|---|---|---|---|---|---|
| | | Prec. | Rec. | F1 | Prec. | Rec. | F1 |
| RGC ($\mathbf{R}_{L2}$) | 88.88 | 0.89 | 0.92 | 0.90 | 0.87 | 0.84 | 0.85 |
| RGC ($\mathbf{R}_{L1}$) | 87.30 | 0.87 | 0.92 | 0.89 | 0.86 | 0.80 | 0.83 |
| ConvNeXT ($\mathbf{R}_{L1}$) | 84.12 | 0.85 | 0.89 | 0.87 | 0.83 | 0.76 | 0.79 |
| SWIN-B ($\mathbf{R}_{L1}$) | 80.95 | 0.80 | 0.92 | 0.85 | 0.84 | 0.64 | 0.73 |
| ResNet-50 ($\mathbf{R}_{L1}$) | 77.77 | 0.79 | 0.87 | 0.82 | 0.76 | 0.64 | 0.70 |
| ViT-B-16 ($\mathbf{R}_{L1}$) | 76.19 | 0.76 | 0.89 | 0.82 | 0.77 | 0.73 | 0.74 |
| VGG-16 ($\mathbf{R}_{L1}$) | 74.60 | 0.73 | 0.92 | 0.81 | 0.80 | 0.48 | 0.60 |

wrong predictions. In this study, we concentrate on four common metrics—accuracy, precision, recall, and F1-score—that are obtained from the confusion matrix entries. These values may be expressed in terms of TP, TN, FP, and FN, which are specified in Section 7.1, for a binary problem where WATs are the positive class and NATs are the negative class.

The percentage of properly categorized samples out of all test samples is the model's *accuracy*:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{\text{TP} + \text{TN}}{N} \tag{7.10}$$

Although accuracy is a useful overall summary, it can be deceptive in situations when classes are unbalanced since the majority class may dominate.

We calculate *precision* and *recall* independently for WATs and NATs to provide a more class-sensitive image. The accuracy for WATs (the positive class) is described as

$$\text{Precision}_{\text{WAT}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7.11}$$

which gauges the reliability of a positive prediction by calculating the percentage of predicted WATs that are really WATs. The matching recall, often referred to as the sensitivity or true positive rate, is

$$\text{Recall}_{\text{WAT}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7.12}$$

This calculates the percentage of real WATs that the classifier successfully retrieves.

In a similar vein, we may define accuracy and recall by symmetry and interpret NATs as the "positive" class:

$$\text{Precision}_{\text{NAT}} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{7.13}$$

$$\text{Recall}_{\text{NAT}} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{7.14}$$

These "NAT as positive" measures, which are given in Table 7.1 as the accuracy and recall

for NATs, are really calculated by simply switching the roles of the two classes in the confusion matrix.

The *F1-score* uses the harmonic mean of accuracy and recall to create a single figure. For WATs, this is

$$\mathrm{F1}_{\mathrm{WAT}} = 2 \times \frac{\mathrm{Precision}_{\mathrm{WAT}} \times \mathrm{Recall}_{\mathrm{WAT}}}{\mathrm{Precision}_{\mathrm{WAT}} + \mathrm{Recall}_{\mathrm{WAT}}} \tag{7.15}$$

An analogous expression holds for NATs:

$$\mathrm{F1}_{\mathrm{NAT}} = 2 \times \frac{\mathrm{Precision}_{\mathrm{NAT}} \times \mathrm{Recall}_{\mathrm{NAT}}}{\mathrm{Precision}_{\mathrm{NAT}} + \mathrm{Recall}_{\mathrm{NAT}}} \tag{7.16}$$

The F1-score is only high when both values are concurrently high since the harmonic mean penalizes significant differences between accuracy and recall, making it a valuable indicator of balanced performance.

We calculate these measures for WATs and NATs independently using the confusion matrices for the RGC model on the two test sets $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$. Table 7.1 summarizes the results and provides the same metrics for a collection of fully supervised baseline models for comparison. The first two rows of the table show that when trained and evaluated on the refined, spurious-free dataset $\mathbf{R}_{L2}$, our semi-supervised RGC model achieves a slightly higher overall accuracy of 88.88% as opposed to 87.30% for the original dataset $\mathbf{R}_{L1}$ that contains spurious sources.

The class-wise measurements show a similar pattern. For WATs, RGC achieves 0.89 precision, 0.92 recall, and 0.90 F1-score on $\mathbf{R}_{L2}$, whereas the equivalent values are 0.87, 0.92, and 0.89 on $\mathbf{R}_{L1}$. This shows that the model can reliably recover most WATs in both scenarios, but when spurious sources are eliminated, there is a little decrease in false positives. The gain is particularly noticeable for NATs: the F1-score, accuracy, and recall rise from $(0.86, 0.80, 0.83)$ on $\mathbf{R}_{L1}$ to $(0.87, 0.84, 0.85)$ on $\mathbf{R}_{L2}$. Put otherwise, the improved masks lessen the likelihood that NATs would be mistakenly classified as WATs by improving the model's ability to differentiate NATs from WAT-like background artifacts.

Additionally, Table 7.1 shows that for every model, the WAT measures are consistently higher than the NAT metrics, which reflects the class imbalance in the training data, where WATs are more common than NATs. This imbalance encourages the decision boundary to be optimized in favor of the WAT class, which results in somewhat improved recall and precision for WATs. Despite the imbalance, the RGC model on $\mathbf{R}_{L2}$ achieves F1-scores of 0.90 and 0.85 for WATs and NATs, respectively, indicating that it can continue to perform well on both classes.

In conclusion, the classification metrics demonstrate that: (i) RGC performs well and fairly evenly on WATs and NATs; and (ii) eliminating spurious and unrelated sources using the refined masks $\mathbf{R}_{L2}$ produces consistent, if slight, improvements in both overall accuracy and class-wise F1-scores. We contextualize these findings by comparing RGC to fully supervised baselines in the following section (Section 7.2), and in Section 7.3, we go beyond point metrics to investigate the complete score distributions using ROC and precision–recall curves.

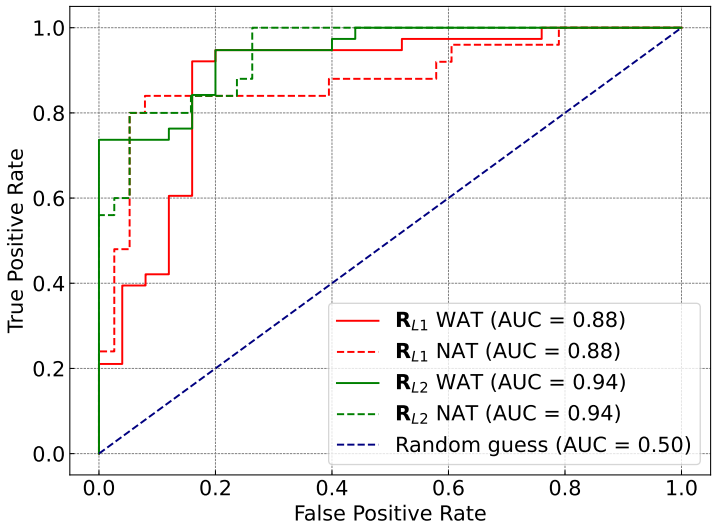



Figure 7.2: *R*OC curves and AUC scores for the RGC model trained and tested on two datasets $\mathbf{R}_{L1}$ (red) and $\mathbf{R}_{L2}$ (green), for both WATs (solid) and NATs (dashed), in comparison to the random guess (blue).

## 7.3 Discriminative ability

Classifier performance at a set decision threshold (in this case, $p = 0.5$ for allocating a source to the WAT class) is characterized by the confusion matrix and point metrics like accuracy and F1-score. For every source $i$, however, the RGC model produces a continuous score $p_i \in [0, 1]$, which may be understood as the model's assessed likelihood that the source is a WAT. In order to transform these scores into hard labels, the decision threshold $\tau$ might be changed.

$$\hat{y}_i(\tau) = \begin{cases} 1, & p_i \geq \tau \\ 0, & p_i < \tau \end{cases} \tag{7.17}$$

We are able to track the evolution of the trade-off between properly and wrongly categorized samples at every potential operational point. The precision–recall (PR) curve and the receiver operating characteristic (ROC) curve summarize this threshold-dependent behavior.

We define the *true positive rate* (TPR) and *false positive rate* (FPR) for WATs (considered as the positive class) for a particular threshold $\tau$ as

For a given threshold $\tau$, the defined *true positive rate* (TPR) and *false positive rate* (FPR) for WATs (treated as the positive class) as

$$\mathrm{TPR}(\tau) = \frac{\mathrm{TP}(\tau)}{\mathrm{TP}(\tau) + \mathrm{FN}(\tau)} \tag{7.18}$$

$$\mathrm{FPR}(\tau) = \frac{\mathrm{FP}(\tau)}{\mathrm{FP}(\tau) + \mathrm{TN}(\tau)} \tag{7.19}$$

where $\mathrm{TP}(\tau)$, $\mathrm{FP}(\tau)$, $\mathrm{TN}(\tau)$, and $\mathrm{FN}(\tau)$ are computed from the thresholded predictions $\hat{y}_i(\tau)$ as in Section 7.1.

Plotting $\mathrm{TPR}(\tau)$ vs $\mathrm{FPR}(\tau)$ as $\tau$ is swept from 1 (only extremely confident WAT predictions)

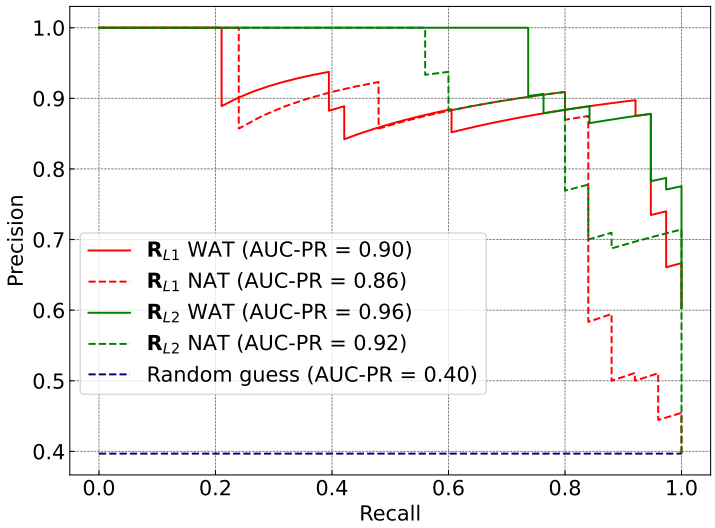



Figure 7.3: *P*recision-Recall (PR) curves and the corresponding AUC-PR scores for the same datasets and sources.

down to 0 (all sources classed as WATs) yields the ROC curve. A perfect classifier would reach the top-left corner (FPR $= 0,\ $ TPR $= 1$), but a classifier that performs no better than random guessing would lie along the diagonal TPR = FPR.

We calculate the *area under the ROC curve* (AUC) in order to reduce the ROC curve to a single scalar summary:

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR})\, d(\mathrm{FPR}) \tag{7.20}$$

where the discrete thresholds utilized in practice are summed over to approximate the integral numerically. The AUC ranges from 0 to 1, where a random classifier is represented by AUC $= 0.5$ and a completely discriminative classifier by AUC $= 1$. The AUC may be intuitively understood as the likelihood that a randomly selected positive example (WAT) would receive a higher score from the classifier than a randomly selected negative example (NAT).

Simultaneously, we examine the *precision–recall* (PR) curve, which is especially useful for datasets that are unbalanced, meaning that one class is more uncommon than the other. The accuracy and recall for WATs for each threshold $\tau$ are provided by

$$\mathrm{Precision}(\tau) = \frac{\mathrm{TP}(\tau)}{\mathrm{TP}(\tau) + \mathrm{FP}(\tau)} \tag{7.21}$$

$$\mathrm{Recall}(\tau) = \frac{\mathrm{TP}(\tau)}{\mathrm{TP}(\tau) + \mathrm{FN}(\tau)} \tag{7.22}$$

Plotting Precision($\tau$) versus Recall($\tau$) as $\tau$ fluctuates yields the PR curve. The PR curve is more directly focused on the quality of the positive predictions and is consequently sensitive to the prevalence of the positive class, in contrast to the ROC curve, which handles the positive and negative classes symmetrically. The accuracy of a classifier that makes random label estimates based on the class prior would be equivalent to the percentage of positives in the dataset.

The *area under the PR curve* (AUC-PR) is defined as

$$\text{AUC-PR} = \int_0^1 \text{Precision}(\text{Recall})\, d(\text{Recall}) \tag{7.23}$$

mathematically approximated in practice once more. While numbers around the positive-class percentage show performance similar to random guessing, values near 1 show that the classifier retains both high accuracy and strong recall across a wide variety of thresholds.

The ROC curves and AUC scores for the RGC model that was trained and evaluated on the two datasets $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$ are presented separately for WATs (solid lines) and NATs (dashed lines) in Figure 7.2. Strong discriminative capacity is demonstrated in both situations by the ROC curves, which approach the top-left corner of the figure and lie well above the diagonal corresponding to a random classifier. This is quantified by the AUC scores: the total AUC for the dataset with spurious sources ($\mathbf{R}_{L1}$) is $\sim 0.88$, whereas it rises to $\sim 0.94$ for the refined dataset without spurious sources ($\mathbf{R}_{L2}$). This improvement reflects the fact that eliminating unrelated emission helps the model assign more separated scores to WATs and NATs, and it is consistent with the improved accuracy and F1-scores reported in Section 7.2.

ROC curves by themselves can exaggerate performance since they are relatively insensitive to changes in the amount of false positives for the minority class due to the moderate imbalance in favor of WATs in our dataset. The matching PR curves and AUC-PR scores for WATs and NATs on $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$ are shown in Figure 7.3 in order to address this. The accuracy of a random classifier that predicts WATs with the same frequency as in the data is shown in these plots by the horizontal baseline at $\simeq 0.40$. For all classes and both datasets, the PR curves for RGC are noticeably higher than this baseline across all recall levels, indicating that the model generates highly informative source rankings.

RGC obtains AUC-PR $\approx 0.96$ for WATs and $\approx 0.92$ for NATs for the spurious-free dataset $\mathbf{R}_{L2}$, demonstrating outstanding discriminative performance despite the class imbalance. These values are somewhat lower but still significantly higher than the random baseline when spurious sources ($\mathbf{R}_{L1}$) are present, indicating that the model is resilient to some degree of contamination. The ROC AUC may be comparable for both classes, but the AUC-PR shows that the model finds it somewhat easier to maintain high precision at high recall for WATs than for NATs, reflecting both the data imbalance and the more heterogeneous appearance of NATs. This is a noteworthy difference between WATs and NATs that is not fully captured by the AUC alone.

## 7.4 Model calibration

A probabilistic classifier should be *well calibrated*, meaning that its projected probabilities should represent the actual frequency of the related occurrences, in addition to being accurate and discriminative. For instance, we would ideally anticipate that around 80% of the sources for which the model gives a WAT probability of 0.8 are indeed WATs. When projected scores are utilized downstream for scientific decision-making, such as when building probabilistic catalogs, propagating categorization uncertainty into population studies, or prioritizing sources for follow-

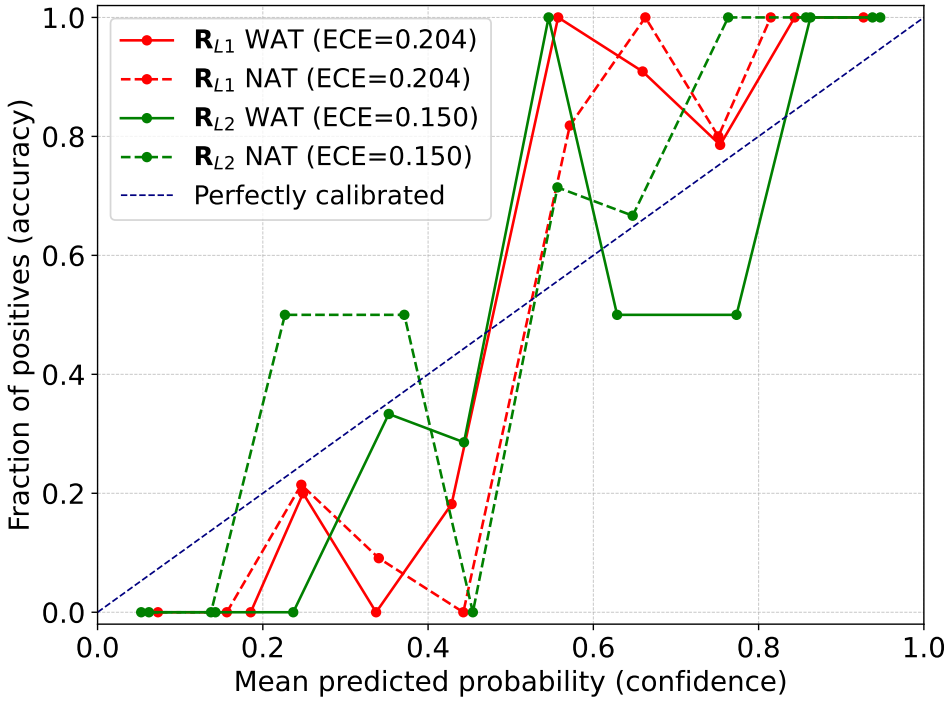



Figure 7.4: ECE curves (described in Section 7.4) for our RGC model trained and tested on two datasets for classifying WATs and NATs. The colors and linestyles are the same as Fig. 7.2.

up, calibration is crucial

We use the *Expected Calibration Error* (ECE) proposed by [78] to measure the calibration of RGC. Take a collection of $N$ test samples, each of which has a true label $y_i \in 0, 1\}$ and a projected probability $p_i \in [0, 1]$ of being a WAT. First, we divide the interval $[0, 1]$ into equal-width $M$ disjoint bins

$$\{\mathbf{B}_m\}_{m=1}^{M}$$

$$B_m = \left(\frac{m-1}{M}, \frac{m}{M}\right) \qquad m = 1 \ldots M \tag{7.24}$$

and assign each sample $i$ to the bin $B_m$ for which $p_i \in B_m$. For each bin $B_m$, we then compute:

- the *empirical accuracy* (or fraction of positives)

$$\mathrm{acc}(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \mathbb{1}\,(y_i = 1) \tag{7.25}$$

- the *average confidence* (mean predicted probability)

$$\mathrm{conf}(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} p_i \tag{7.26}$$

where $|B_m|$ denotes the number of samples in bin $B_m$, and $\mathbb{1}(\cdot)$ is the indicator function.

A perfectly calibrated model would satisfy

$$\mathrm{acc}(B_m) = \mathrm{conf}(B_m) \tag{7.27}$$

for every bin $m$, in order for the mean projected probability in each confidence interval to match the empirical proportion of true WATs. There will be departures from this ideal in real life. By calculating a weighted average of the absolute differences and utilizing the bin frequencies as weights, the ECE combines these variances into a single scalar [Equation 3 of 78]:

$$\mathrm{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{N} \left|\mathrm{acc}(B_m) - \mathrm{conf}(B_m)\right| \tag{7.28}$$

By construction, ECE $\in [0, 1]$, with bigger values denoting a wider discrepancy between anticipated and empirical probabilities and ECE $= 0$ denoting perfect calibration. In accordance with, we balance resolution against statistical noise inside each bin by using a reasonable number of bins (e.g., $M = 10$).

The *reliability diagram* or calibration curve, which depicts $\mathrm{acc}(B_m)$ as a function of $\mathrm{conf}(B_m)$ for all bins, provides an additional visual diagnostic. The diagonal line acc = conf represents the points for a correctly calibrated model; systematic deviations above or below this line signify under-confidence or over-confidence, respectively.

The calibration curves for our RGC model trained and tested on the two datasets $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$ are displayed in Figure 7.4, where WATs and NATs are plotted independently using the same color and linestyle scheme as in Fig. 7.2. The curves in each of the four scenarios are very near to the diagonal, suggesting that the model's projected probabilities are generally in good agreement with the observed frequencies. Quantitatively, the ECE for both WATs and NATs in the revised, spurious-free dataset $\mathbf{R}_{L2}$ is $\simeq 0.150$, but it rises to $\simeq 0.204$ for the dataset with spurious sources ($\mathbf{R}_{L1}$).

In addition to improving the classification metrics (Section 7.2) and discriminative capabilities (Section 7.3), these values demonstrate that eliminating spurious and irrelevant radio emission through the revised masks also results in better-calibrated probability estimations. It makes sense that when bright sidelobes or background sources that are not part of the bent radio galaxy are present in the input photos, the model may become overconfident in predictions that are influenced by these artifacts, which would increase the disparity between accuracy and confidence. The model's predictions more accurately represent the real morphological information in the jets and tails, and the confidence scores become more reliable by removing such pollutants from $\mathbf{R}_{L2}$.

## 7.5 Findings and Challenges

The findings in this chapter highlight a number of unresolved issues and lead to numerous important conclusions on the behavior of the RGC model.

Our first major performance discovery is that bent radio AGNs are successfully and fairly

evenly classified into WATs and NATs by the semi-supervised RGC. RGC outperforms all fully supervised baselines trained on $\mathbf{R}_{L1}$ [4] on the revised, spurious-free dataset $\mathbf{R}_{L2}$, achieving an accuracy of 88.88% with F1-scores of 0.90 for WATs and 0.85 for NATs (Table 7.1). This illustrates the efficacy of combining a small, carefully selected labeled sample with self-supervised pretraining on a large unlabeled set. Given the small amount of the labeled dataset and the structural intricacy of bending radio sources, the gain is particularly noteworthy.

Second, we discover that nearly all assessment criteria consistently benefit from enhanced source separation using modified masks. Removing unrelated emission results in higher accuracy and F1-scores, especially for NATs; better separation of WAT and NAT score distributions (higher ROC AUC and AUC-PR); and lower Expected Calibration Error, indicating more reliable probabilities, according to a comparison of $\mathbf{R}_{L1}$ (with spurious sources) and $\mathbf{R}_{L2}$ (without). Together, our findings demonstrate that meticulous picture pre-processing and masking are essential for establishing reliable machine-learning models for radio morphology categorization, rather than being only aesthetic.

Third, with ROC AUCs of $\sim$ 0.88–0.94 and AUC-PR values up to $\sim$ 0.96 for WATs and $\sim$ 0.92 for NATs on $\mathbf{R}_{L2}$, the ROC and PR studies show that RGC is highly discriminative across a broad range of operating thresholds. This implies that RGC may be applied flexibly in various scientific situations by changing the decision threshold to give priority to either purity (high accuracy) or completeness (high recall). For example, a lower threshold can be used to favor stronger NAT recall at the price of some additional false positives in follow-up campaigns when missing uncommon NATs is more expensive than adding a few extra WATs.

Even with these positive results, there are still a number of obstacles. The ongoing performance difference between WATs and NATs is one of the challenges. Because of the training set's class imbalance and NATs' larger morphological variety, WATs outperform NATs in all models in terms of accuracy, recall, and F1-scores. It is still more difficult to accurately categorize NATs with extremely twisted or dispersed tails or those entrenched in intricate cluster or group contexts.

The comparatively limited quantity of the labeled dataset presents a second difficulty. Rare or unusual configurations are still underrepresented, which might result in systemic failure modes that only show up in bigger survey applications, even if semi-supervised learning somewhat mitigates this. This restriction would be lessened and more detailed taxonomy beyond the binary WAT/NAT distinction would be possible by scaling up the labeled sample, either through further expert annotation or by utilizing citizen-science initiatives.

Interpretability and failure analysis present a third difficulty. Even while Grad-CAM and exploratory plots show that the revised masks direct the model's focus toward physically significant jets and tails, certain misclassifications are still hard to understand, particularly when the radio morphology is unclear. In these situations, it is difficult to distinguish between "label ambiguity" and "model error." Integrating machine-learning classifiers into astrophysical inference pipelines will need the development of more methodical methods for measuring label uncertainty and connecting model scores to physical factors (e.g., environment richness, host-galaxy features).

As mentioned in introduction, the automated classification of bent RAGNs is important for both computational and astrophysical reasons. Studying tail bending across different redshifts provides key insights into the evolution of galaxies in dense environments. However, a limitation of binary classifiers is that tail bending, in reality, forms a continuum, with $\theta_{OA}$ ranging smoothly from 0° to 180°. The histogram of $\theta_{OA}$ in [24, Figure 8] illustrates this continuous distribution, and [44] cautions against relying solely on bending angles when classifying a source as a WAT. A DL framework like RGC mitigates this limitation by not depending exclusively on $\theta_{OA}$, as evident from the Grad-CAM localization maps in Appendix 10.2. Some apparent misclassifications may therefore correspond to genuinely ambiguous sources. Further work is required to identify which physical features of a bent RAGN image become most salient for an AI model's attention. Consequently, this study contributes toward developing a physics-informed foundation model for RAGNs, enabling AI systems to reason about a wide range of their physical properties.

The implications of class imbalance, discussed in the previous subsection, provide another perspective for exploring the astrophysics of RAGNs. The higher abundance and more accurate classification of WATs are consistent with expectations that they are typically associated with more massive, centrally located hosts such as BCGs, as noted in [44] and the references therein. In contrast, the lower abundance and poorer classification performance of NATs are in line with the view that these sources trace faster-moving satellite galaxies embedded in denser ICM environments [85, 86].

## 7.6 Astrophysical implications

As mentioned in Section introduction, the automated classification of bent RAGNs is important for both computational and astrophysical reasons. Studying tail bending across different redshifts provides key insights into the evolution of galaxies in dense environments. However, a limitation of binary classifiers is that tail bending, in reality, forms a continuum, with $\theta_{OA}$ ranging smoothly from 0° to 180°. The histogram of $\theta_{OA}$ in [24, Figure 8] illustrates this continuous distribution, and [44] cautions against relying solely on bending angles when classifying a source as a WAT. A DL framework like RGC mitigates this limitation by not depending exclusively on $\theta_{OA}$, as evident from the Grad-CAM localization maps in Appendix 10.2. Some apparent misclassifications may therefore correspond to genuinely ambiguous sources. Further work is required to identify which physical features of a bent RAGN image become most salient for an AI model's attention. Consequently, this study contributes toward developing a physics-informed foundation model for RAGNs, enabling AI systems to reason about a wide range of their physical properties.

The implications of class imbalance, discussed in the previous subsection, provide another perspective for exploring the astrophysics of RAGNs. The higher abundance and more accurate classification of WATs are consistent with expectations that they are typically associated with more massive, centrally located hosts such as BCGs, as noted in [44] and the references therein. In contrast, the lower abundance and poorer classification performance of NATs are in line

with the view that these sources trace faster-moving satellite galaxies embedded in denser ICM environments [85, 86].

# Chapter 8

# Sustainability

Sustainable Development Goals (SDGs) provide a global framework for connecting scientific research with long–term social, economic, and environmental impact.[1] In the context of this thesis, sustainability mainly concerns how the proposed radio AGN classifier contributes to sustainable innovation in data–intensive science and how the computational workflow is designed to limit unnecessary energy use. This section briefly outlines how the work aligns with selected SDGs, with a focus on SDG 9 (Industry, Innovation and Infrastructure).

In order to categorize bent radio AGNs from survey data, the thesis first creates and assesses a semi-supervised deep learning pipeline. Large radio continuum surveys, which are a component of contemporary astronomy's scientific infrastructure, may be handled more effectively and scalably because to the work's automation of a process that has historically required in-depth expert visual examination. [2]Specifically, the RGC model shows that equivariant CNNs combined with self-supervised pretraining can minimize the requirement for huge labeled datasets while retaining competitive performance, which is crucial for future survey facilities that will generate enormous amounts of data.

Second, the thesis aims to the best standards for reusable and open research infrastructure. The curated datasets and RGC Python package are meant to be shared with the community so that other groups can replicate, expand, or modify the models to fit various surveys or similar categorization issues. This is in line with SDG 9 aims, which place a strong emphasis on developing technological innovation and scientific research capabilities, particularly in data-intensive sectors.

Third, the effort promotes capacity building at the intersection between astrophysics and computer science. This thesis reduces the barrier to entry for researchers and students from resource-constrained institutions who want to work with radio survey data and machine learning by documenting an entire pipeline, from data preprocessing and catalog construction to model training, evaluation, and interpretability. More equitable involvement in scientific infrastructure and innovation is supported by this kind of information transmission.

[1]See, e.g., the United Nations Sustainable Development Goals.

[2]Infrastructure and innovation are key components of sustainable development, according to SDG 9.

# Chapter 9

# Conclusion

## 9.1 Project Summary

With this paper, we publish two ML-friendly datasets ($\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$) of radio active galactic nuclei (RAGN) labeled as wide-angle-tail (WAT) and narrow-angle-tail (NAT) sources, and a semi-supervised model for the binary classification of WATs and NATs trained and tested on these labeled datasets and a publicly available unlabeled dataset ($\mathbf{R}_U$). A previous version of $\mathbf{R}_{L1}$ was published in [4] with some initial benchmarking results and a previous version of the RGC model was briefly described in [7]. All our datasets are created from the VLA images of RAGNs taken as part of the FIRST survey.

As described in Section 5 of this paper, $\mathbf{R}_U$ contains around 20,000 unlabeled sources, and $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$ contain 639 labeled sources preprocessed by us through `PyBDSF` and `Photutils`, respectively. The second labeled dataset $\mathbf{R}_{L2}$ does not contain any spurious source, unlike the first labeled dataset. Among the labeled sources, 66% are WATs. A catalog of the labeled sources is included in the RGC package, and a demonstration of the catalog is given in Appendix 10.1. This catalog was created using a catalog of more than 700 WATs and NATs classified through visual inspection by [23]. Although over 4000 WATs and NATs have been identified in the FIRST images (Table 1.1), we adopt the Sasmal catalog for its reliability as a fully visually inspected sample.

Our RGC model, described in A semi-supervised model for the VLA images of bent radio AGNs section consists of a self-supervised pre-training stage with BYOL on the unlabeled data, followed by a supervised fine-tuning stage with E2CNN on the two labeled datasets. The performance and calibration of the model are described in Section 7, and its transparency, limitations and astrophysical implications are described in Section **??**. Here we summarize the key results of this work.

1. The semi-supervised RGC model outperforms the baseline supervised models, achieving accuracies of 88.88% on $\mathbf{R}_{L2}$ and 87.30% on $\mathbf{R}_{L1}$ (Table 7.1).

2. The RGC model trained and tested on the dataset without spurious sources ($\mathbf{R}_{L2}$) outperforms the version trained on images with spurious sources across all measures—performance

metrics (Fig. 7.1 and Table 7.1), discriminative ability (ROC and PR curves in Figs. 7.2 and 7.3), and model calibration (ECE curves in Fig. 7.4).

3. Because of class imbalance, most models yield higher precision, recall, and F1-scores for WATs than for NATs. The RGC model trained on $\mathbf{R}_{L2}$, however, achieves high F1-scores for both (0.90 for WATs, 0.85 for NATs). The higher abundance of WATs are consistent with their association with the dominant galaxies within groups and clusters.

4. If spurious sources are present in an image, the RGC model distributes attention between the target and the contaminants, as seen in the Grad-CAM localization maps overplotted as contours in Appendix 10.2. The superior performance of $\mathbf{R}_{L2}$ results from the absence of such spurious sources, as reflected in the ROC, PR, and ECE transparency metrics.

5. The class imbalance between WATs and NATs has a significant effect on the attention of the RGC model. For WATs, the attention contours (Appendix 10.2) are typically centered on the peak and enclose the source almost symmetrically. In contrast, for NATs, the contours tend to trace mostly the outer edges of the source (see Appendix 10.2).

6. DL frameworks can advance RAGN research further when extended into foundation models trained on diverse physical properties, enabling multimodal characterization beyond binary classification and capturing the continuous morphology from WATs to NATs.

## 9.2 Future Work: Multimodal Self-Supervised Radio Galaxy Classification from multi-spectral images

The previous study showed that the reliability of radio galaxy classification models may be significantly increased by carefully curated datasets and explicit control of spurious correlations, an issue that has also been highlighted more broadly in recent machine-learning work on radio galaxies. We demonstrated that standard supervised deep networks can unintentionally latch onto survey artifacts and annotation-induced biases by creating "spurious" and "clean" variants of the dataset and cross-evaluating models trained on each, and that explainability tools such as Grad-CAM offer a helpful lens for identifying when a model relies on non-astrophysical cues [74]. In the era of large radio surveys such as LOFAR, ASKAP and MeerKAT, this constitutes a significant first step toward reliable, bias-aware radio morphology classification [87].

At the same time, the technique adopted in the previous paper exposes several limitations that motivate a second, more ambitious investigation. First, the models relied purely on supervised learning, necessitating labelled training instances at a time when manual annotation cannot keep pace with the volume and diversity of sources in modern surveys. It is not feasible to label all pertinent morphologies, including wide-angle tailed (WAT) sources and the full Fanaroff–Riley (FR0/FR1/FR2) family, across current and future datasets . Second, the training setup operated on relatively small, pre-cut postage stamps centred on known sources, presuming that source detection, deblending and centring have already been solved upstream; this can restrict what the network is able to learn about extended or complex structures .

Third, the analysis was limited to single-band radio imaging, although abundant complementary information is available at other wavelengths, particularly in the optical, where host-galaxy properties and environments are better constrained.

These limitations naturally suggest three directions for extension. The first is to lessen reliance on large labelled datasets by using contemporary self-supervised learning methods that can learn general-purpose representations from large collections of unlabelled radio images, before fine-tuning on a limited set of labels. The second is to relax the strict preprocessing pipeline by exposing the model to larger sky regions, allowing it to learn, at least partially, to localise and characterise radio galaxies in a more realistic setting, in line with recent work on foundation-style models for resolved radio sources. The third is to incorporate multi-spectral information, particularly the combination of radio and optical imaging, in order to investigate how features learned in one band transfer to or complement those from the other, and whether joint representations can enhance WAT and FR0/FR1/FR2 classification while remaining robust to the kinds of spurious correlations identified in the previous chapter.

By investigating self-supervised, multi-spectral, and more loosely preprocessed modeling strategies, and by assessing their performance on both new data and the curated dataset presented in Chapter 6, the present chapter builds directly on these insights. In doing so, our aim is to move beyond merely diagnosing spurious correlations toward designing training regimes and representations that are intrinsically more reliable and more scalable to forthcoming surveys.

The tremendous expansion of data quantities from current and future radio surveys has propelled the shift towards automated radio-galaxy morphology categorization. Supervised convolutional networks were used in early deep learning applications for FR I/FR II classification, showing that CNNs can achieve high accuracy on curated datasets but are still reliant on significant labeled training samples.

An increasing amount of research has emphasized that accuracy alone is insufficient to evaluate model dependability in practical survey situations. Using variational inference, Mohan et al. quantified predictive uncertainty for FR I/FR II classification on the MiraBest dataset, demonstrating that uncertainty calibration is an essential part of reliable morphology pipelines and that deep models can be confidently incorrect when exposed to distribution shifts. In a more recent research, Mohan et al. assessed several Bayesian neural network techniques for radio-galaxy classification, specifically examining distribution-shift detection, uncertainty calibration, and prediction performance across various radio and optical datasets . These findings highlight the need for techniques that can indicate when they come across out-of-distribution or systematically distinct sources and are resilient to dataset biases.

Concurrently, a number of studies have brought attention to the drawbacks of single-band inputs and tiny, pre-cut postal stamps. Research employing MiraBest and similar samples usually assumes that the network operates on fixed-size cuts with a single, known source and that source detection and centering are resolved upstream [88]. On the other hand, coupled localization and classification can better fit the operating requirements of wide-area surveys, as demonstrated by object-detection techniques that work on broader fields, such as deep-learning models applied directly to survey photos [66]. Furthermore, it has been shown that multi-

wavelength data—especially the combination of radio and optical imaging—improves source association and morphological categorization. For instance, combining FIRST with SDSS data [66].

# Chapter 10

# Appendix

# 10.1 Sasmal-ML catalog

Table 10.1: Information about the 63 sources from our Sasmal-ML catalog that are included in our Test Batch.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RI | SI | MI | FCG | RA | Dec | Label | Batch | d | T_isl | T_pix | Saikia | Asad | Hossain | Attention RL1 | Attention RL2 |
| 10 | 12 | 9 | J0046-0805 | 00 46 36.48 | -08 05 45.6 | WAT | test batch | 2 | 1.6 | 5 | ... | sig | SMD | g | g |
| 12 | 14 | 10 | J0057-0336 | 00 57 33.71 | -03 36 09.6 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 18 | 21 | 15 | J0229+0429 | 02 29 02.09 | +04 29 03.3 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 24 | 27 | 21 | J0312-0633 | 03 12 55.23 | -06 33 43.1 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 33 | 37 | 29 | J0715+4829 | 07 15 19.07 | +48 29 32.0 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 34 | 38 | 30 | J0717+4405 | 07 17 26.66 | +44 05 02.3 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 43 | 48 | 37 | J0744+4353 | 07 44 42.18 | +43 53 07.8 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 44 | 47 | 38 | J0744+1658 | 07 44 59.31 | +16 58 59.7 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 48 | 52 | 42 | J0750+5841 | 07 50 18.30 | +58 41 36.3 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 49 | 51 | 43 | J0750+1741 | 07 50 49.80 | +17 41 41.9 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 54 | 58 | 47 | J0811+1029 | 08 11 40.88 | +10 29 27.3 | WAT | test batch | 0 | 3 | 5 | | | SMD | c | g |
| 91 | 95 | 81 | J0902+1819 | 09 02 16.30 | +18 19 04.3 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 93 | 97 | 83 | J0904+5834 | 09 04 57.31 | +58 34 49.9 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 122 | 127 | 110 | J0943+0611 | 09 43 09.45 | +06 11 31.8 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 151 | 156 | 137 | J1031-0640 | 10 31 30.66 | -06 40 28.1 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 168 | 174 | 152 | J1105+5943 | 11 05 24.26 | +59 43 41.9 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 171 | 177 | 155 | J1114+3625 | 11 14 01.78 | +34 25 03.2 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 184 | 190 | 168 | J1130+2524 | 11 30 48.83 | +25 24 35.6 | WAT | test batch | 0 | 3 | 5 | | | SMD | c | c |
| 194 | 200 | 177 | J1145-0757 | 11 45 53.08 | -07 57 53.6 | WAT | test batch | 0 | 3 | 5 | | bad | SMD | g | g |
| 197 | 203 | 180 | J1148+2332 | 11 48 33.14 | +23 32 25.9 | WAT | test batch | 0 | 3 | 5 | | | SMD | s2 | g |
| 207 | 213 | 189 | J1201+3257 | 12 01 51.87 | +32 57 01.3 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 215 | 220 | 197 | J1213-0327 | 12 13 45.43 | -03 27 09.5 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 217 | 223 | 199 | J1219+0014 | 12 19 05.57 | +00 14 16.4 | WAT | test batch | 1 | 2.6 | 4 | ... | sig | SMD | g | g |
| 230 | 236 | 212 | J1247+4042 | 12 47 01.00 | +40 42 02.3 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 234 | 240 | 216 | J1250+0839 | 12 50 10.59 | +08 39 52.7 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 264 | 274 | 242 | J1345+6110 | 13 45 03.06 | +61 10 31.6 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 265 | 271 | 243 | J1345+0855 | 13 45 04.67 | +08 55 07.2 | WAT | test batch | 1 | 1.8 | 3.8 | | sig | SMD | s | g |
| 270 | 276 | 248 | J1351+0712 | 13 51 44.72 | +07 12 17.0 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 276 | 284 | 253 | J1401+5654 | 14 01 09.98 | +56 54 20.6 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 291 | 298 | 268 | J1416+2147 | 14 16 13.79 | +21 47 42.8 | WAT | test batch | 1 | 2 | 5 | | sig | SMD | s3 | g |
| 352 | 360 | 324 | J1601+3155 | 16 01 30.74 | +31 55 32.4 | WAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 355 | 363 | 325 | J1608+0141 | 16 08 03.55 | +01 41 54.34 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 357 | 365 | 326 | J1616+0926 | 16 16 53.17 | +09 26 35.7 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 390 | 399 | 356 | J2139+1008 | 21 39 46.25 | +10 08 28.0 | WAT | test batch | 0 | 3 | 5 | | | SMD | s2 | g |
| 402 | 411 | 367 | J2213-0854 | 22 13 12.43 | -08 54 37.1 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 409 | 418 | 374 | J2301+0037 | 23 01 35.06 | +00 37 13.1 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 411 | 420 | 376 | J2312-0919 | 23 12 12.16 | -09 19 28.6 | WAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 416 | 425 | 380 | J2334-0759 | 23 34 55.59 | -07 59 14.4 | WAT | test batch | 0 | 1 | 4.2 | ... | sig | SMD | g | cut |
| 428 | 7 | 390 | J0054+0302 | 00 54 43.83 | +03 02 10.5 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 447 | 27 | 404 | J0356+0013 | 03 56 50.87 | +00 13 23.2 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 451 | 31 | 408 | J0718+4820 | 07 18 25.52 | +48 20 20.6 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | lim |
| 454 | 34 | 411 | J0731+3251 | 07 31 13.85 | +32 51 15.1 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 474 | 56 | 426 | J0829+6322 | 08 29 10.55 | +63 22 28.5 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 476 | 58 | 428 | J0834+3945 | 08 34 30.09 | +39 45 09.5 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | n |
| 478 | 60 | 430 | J0846+2024 | 08 46 47.94 | +20 24 48.9 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | lim |
| 480 | 62 | 432 | J0851+1615 | 08 51 10.34 | +16 15 30.6 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 484 | 66 | 436 | J0902+3505 | 09 02 30.30 | +35 05 11.0 | NAT | test batch | 0 | 3 | 5 | | | SMD | n | g |
| 493 | 75 | 445 | J0933+1341 | 09 33 28.30 | +13 41 37.5 | NAT | test batch | 0 | 3 | 5 | | | SMD | lim | n |
| 509 | 92 | 459 | J1016+4708 | 10 16 06.64 | +47 08 06.6 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | lim |
| 514 | 97 | 464 | J1032+3152 | 10 32 15.87 | +31 52 35.7 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | lim |
| 536 | 119 | 481 | J1132+1344 | 11 32 40.23 | +13 44 26.2 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 541 | 124 | 485 | J1145+1529 | 11 45 22.20 | +15 29 43. | NAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 552 | 135 | 496 | J1217+0336 | 12 17 31.4 | +03 36 57.0 | NAT | test batch | 0 | 3 | 5 | | | SMD | n | lim |
| 555 | 138 | 499 | J1222-0449 | 12 22 23.69 | -04 49 35.4 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 600 | 183 | 542 | J1403+0610 | 14 03 13.29 | +06 10 08.3 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 602 | 185 | 544 | J1407+4257 | 14 07 13.04 | +42 57 35.7 | NAT | test batch | 0 | 3 | 5 | | | SMD | n | g |
| 625 | 208 | 565 | J1509+3327 | 15 09 57.37 | +33 27 15.0 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | g (s) |
| 635 | 218 | 575 | J1541+0301 | 15 41 41.70 | +03 01 03.0 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 668 | 251 | 605 | J1652+3055 | 16 52 05.38 | +30 55 17.0 | NAT | test batch | 0 | 3 | 5 | | | SMD | s | g |
| 670 | 253 | 607 | J1700+4442 | 17 00 47.39 | +44 42 42.0 | NAT | test batch | 0 | 3 | 5 | | | SMD | n | g |
| 680 | 264 | 617 | J1733+4229 | 17 33 43.25 | +42 29 12.1 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 682 | 266 | 619 | J2102-0145 | 21 02 14.61 | -01 45 09.5 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |
| 694 | 279 | 631 | J2239-0932 | 22 39 02.43 | -09 32 35.6 | NAT | test batch | 0 | 3 | 5 | | | SMD | g | g |

This table (Table 10.1) contains the 63 sources used in our Test Batch. The complete annotated catalog can be found in our GitHub package (`https://github.com/cassaiub/rgc/`). The 16 numbered columns of this catalog are described below.

1. RI: RGC Index of 703 sources used throughout this paper for all bent radio active galactic nuclei (RAGN).
2. SI: original Sasmal Index of these RAGNs used in the catalog of [23].
3. MI: Masking Index of the 639 sources which were finally included in our labeled datasets; the preprocessing masks are stored using these indices.
4. FCG stands for FIRST Component Group or FIRST Catalog Galaxy, which is a common prefix for the IAU-compliant names of these objects.
5. RA: right ascension coordinate of a source.
6. Dec: declination coordinate of a source.
7. Label: the label given by [23] which is our ground truth.
8. Batch: the batch type in our batched labeled dataset $\mathbf{R}_{L1}$ and $\mathbf{R}_{L2}$.
9. $d$: the dilation used in `PyBDSF` preprocessing.
10. $T_{isl}$: the island threshold used in `PyBDSF` preprocessing.
11. $T_{pix}$: the pixel threshold used in `PyBDSF` preprocessing.
12. Saikia: the annotations by Saikia, hidden in this table with '...'. The complete annotations can be found in Github.
13. Asad: the annotations by Asad. Here 'sig' means more sigma clipping was needed during preprocessing, 'bad' means the source should be removed.
14. Hossain: the annotations by Hossain.
15. Attention RL1: comments about the quality of the attentions of our model on the dataset $\mathbf{R}_{L1}$. Here 'g' indicates good attention, 's' indicates the presence of a spurious source, 's2' and 's3' two and three spurious sources, respectively, 'c' indicates confusion about whether a component is spurious or part of the target source, 'n' indicates no attention, 'lim' stands for limited attention that did not cover or delimit the source wholly.
16. Attention RL2: comments about the quality of the attentions of our model on the dataset $\mathbf{R}_{L2}$. The letters have the same meaning as the previous column, except 'cut' which means part of the source was cropped while trying to remove spurious sources.

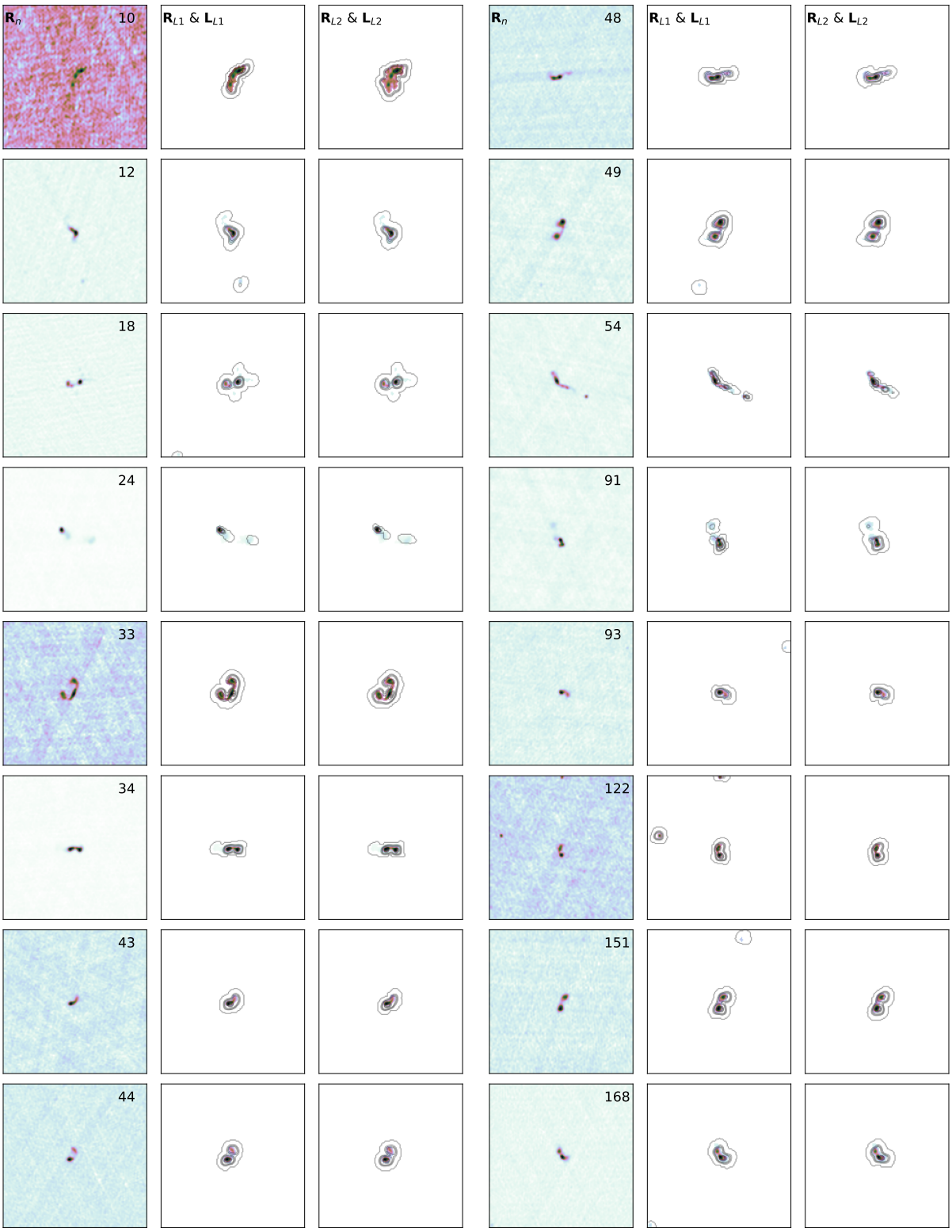
## 10.2 Grad-CAM localization maps

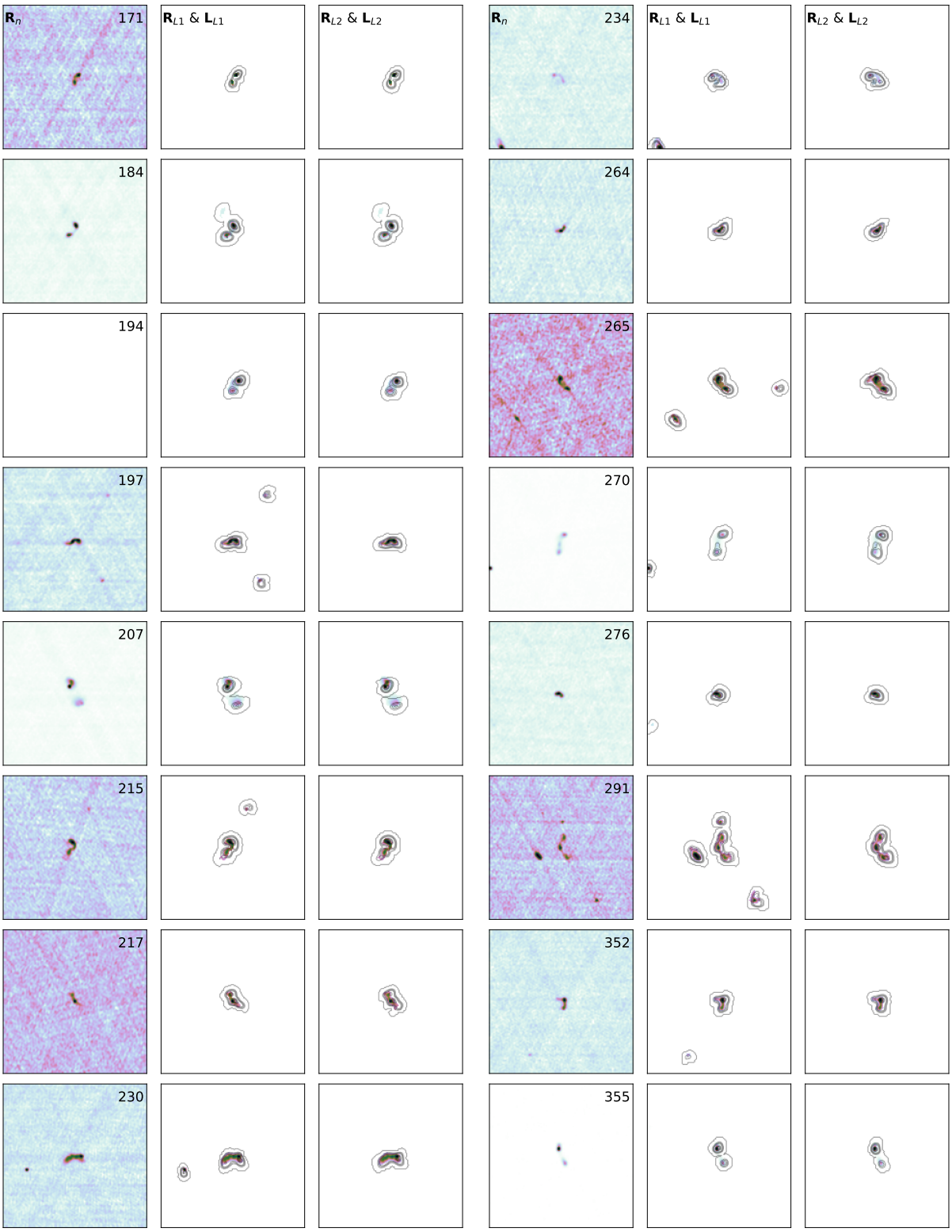
In the following pages we provide the original normalized images $\mathbf{R}_n$, the masked images $\mathbf{R}_{L1}$ preprocessed using `PyBDSF`, and the masked images $\mathbf{R}_{L2}$ preprocessed using `Photutils` for the 63 RAGNs in the test batch of our batched datasets.

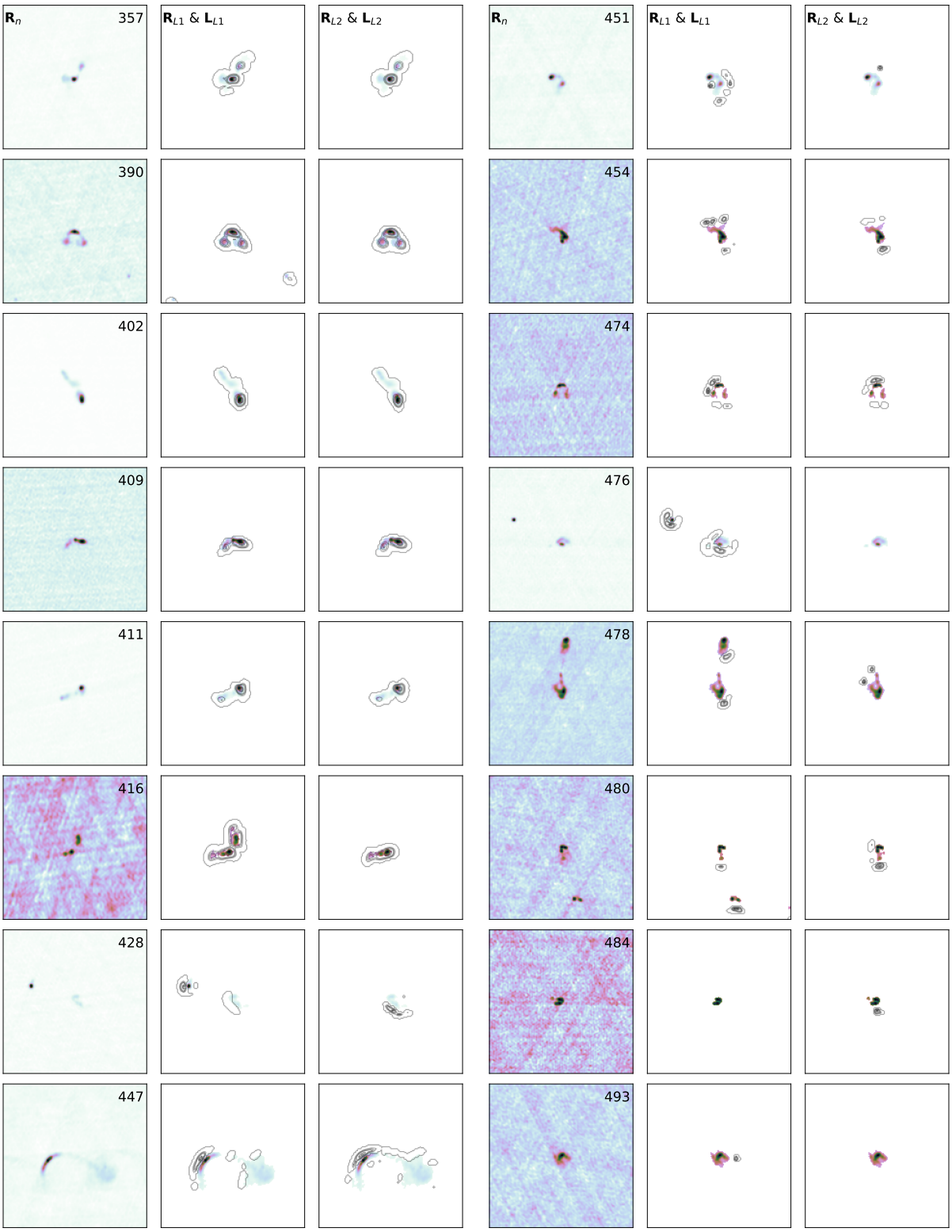
The corresponding localization maps produced by Grad-CAM ($\mathbf{L}_{L1}$ and $\mathbf{L}_{L2}$) are overplotted on the masked images as contours. The contours have been calculated from the heatmaps produced following the mechanism described in Section 6.4.

Each page has 6 columns and 8 rows. The first 3 columns show the original and masked images of 8 sources, and the last three columns show 8 more sources, resulting in 16 sources per page, except the last page where we show 15 sources.

The detailed labels are given only on the first row of each page, which are the same for the lower rows of the page. The integer on the upper right corner of every image on the first columns indicate the RGC index (RI), as given in the first column of Table 10.1 in the previous appendix.


$\mathbf{R}_n$
10
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
$\mathbf{R}_n$
48
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
12
49
18
54
24
91
33
93
34
122
43
151
44
168


$\mathbf{R}_n$
171
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
$\mathbf{R}_n$
234
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
184
264
194
265
197
270
207
276
215
291
217
352
230
355


$\mathbf{R}_n$
357
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
$\mathbf{R}_n$
451
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
390
454
402
474
409
476
411
478
416
480
428
484
447
493

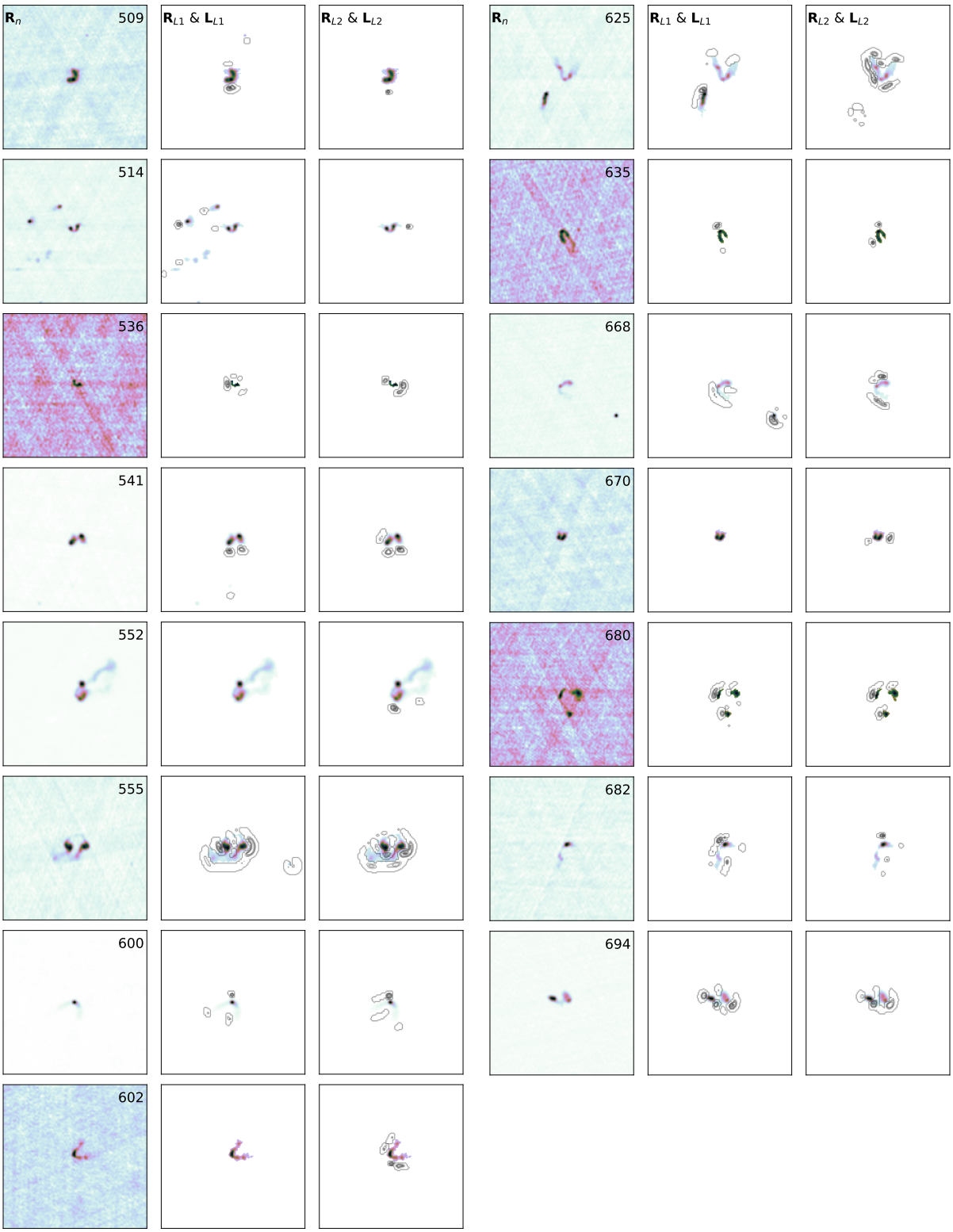

$\mathbf{R}_n$
509
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
$\mathbf{R}_n$
625
$\mathbf{R}_{L1}$ & $\mathbf{L}_{L1}$
$\mathbf{R}_{L2}$ & $\mathbf{L}_{L2}$
514
635
536
668
541
670
552
680
555
682
600
694
602

Table 10.2: Notation glossary

| Symbol | Definition | First use |
|---|---|---|
| $\mathbf{R}$ | Original FITS images of RAGNs | Eq. (1) |
| $\mathbf{R}_n$ | Normalized image | Eq. (1) |
| $B(x, y)$ | Background intensity | Eq. (2) |
| $\mathbf{M}_{L1}$ | Binary mask (`PyBDSF`) | Eq. (4) |
| $T$ | Global detection threshold (`Photutils`) | Eq. (5) |
| $\mathbf{M}_L$ | Binary mask (`Photutils`) | Eq. (7) |
| $\mathcal{L}_{\texttt{BYOL}}$ | BYOL loss function | Eq. (9) |
| $\mathcal{L}_{CE}$ | Cross-entropy loss function | Eq. (10) |
| $\alpha_k^c$ | Importance weight of feature map $k$ for class $c$ | Eq. (11) |
| $L^c$ | Grad-CAM localization map for class $c$ | Eq. (12) |
| $\theta_{OA}$ | Opening angle of the radio jets | Section 2 |
| $\theta_{LAS}$ | Largest angular size of an RAGN | Section 3.1 |
| $\mathbf{R}_U$ | Unlabeled dataset from RGZ | Section 3.1 |
| $T_{\text{isl}}$ | Island threshold for source detection | Section 3.2.2 |
| $T_{\text{pix}}$ | Peak threshold for source detection | Section 3.2.2 |
| $\mathbf{M}_{L2}$ | Final combined mask | Section 3.2.3 |
| $\mathbf{R}_{L1}$ | Masked image including spurious sources (`PyBDSF`) | Section 3.2.5 |
| $\mathbf{R}_{L2}$ | Masked image without spurious sources | Section 3.2.5 |
| $\mathbf{L}_{L1}$ | Grad-CAM localization map for $\mathbf{R}_{L1}$ | Section 4.4 |
| $\mathbf{L}_{L2}$ | Grad-CAM localization map for $\mathbf{R}_{L2}$ | Section 4.4 |
| $r_{il}$ | Distance between the peaks of an image and its localization map | Section 6.2 |

# Bibliography


[1] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent a new approach to self-supervised learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[2] Anna M. M. Scaife and Fiona Porter. Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks. , 503(2):2369–2379, May 2021.

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[4] Mir Sazzat Hossain, Khan Muhammad Bin Asad, Payaswini Saikia, Adrita Khan, Md Akil Raihan Iftee, Rakibul Hasan Rajib, Arshad Momen, Md Ashraful Amin, Amin Ahsan Ali, and AKM Mahbubur Rahman. Rgc-bent: A novel dataset for bent radio galaxy classification. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 2868–2873, 2025.

[5] Dipankar Maitra. A glimpse into the life and works of Radha Gobinda Chandra, the village astronomer from the Indian sub-continent. In *110th Annual Meeting of the American Association of Variable Star Observers (AAVSO)*, November 2021.

[6] Ramesh Kapoor. Comet 1P/Halley in 1910: tales from India. *Journal of Astronomical History and Heritage*, 26(2):411–440, June 2023.

[7] Mir Sazzat Hossain, Sugandha Roy, K.M.B. Asad, Arshad Momen, Amin Ahsan Ali, M Ashraful Amin, and A. K. M. Mahbubur Rahman. Morphological classification of radio galaxies using semi-supervised group equivariant cnns. *Procedia Computer Science*, 222:601–612, 2023. International Neural Network Society Workshop on Deep Learning Innovations and Applications (INNS DLIA 2023).

[8] Christopher J. Conselice, Aaron Wilkinson, Kenneth Duncan, and Alice Mortlock. The Evolution of Galaxy Number Density at z ¡ 8 and Its Implications. , 830(2):83, October 2016.

[9] P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, H. Böhringer, A. Bogomolov, T. Boller, K. Borm, W. Bornemann, H. Bräuninger, M. Brüggen, H. Brunner, M. Brusa, E. Bulbul, M. Buntov, V. Burwitz, W. Burkert, N. Clerc, E. Churazov, D. Coutinho, T. Dauser, K. Dennerl, V. Doroshenko, J. Eder, V. Emberger, T. Eraerds, A. Finoguenov, M. Freyberg, P. Friedrich, S. Friedrich, M. Fürmetz, A. Georgakakis, M. Gilfanov, S. Granato, C. Grossberger, A. Gueguen, P. Gureev, F. Haberl, O. Hälker, G. Hartner, G. Hasinger, H. Huber, L. Ji, A. v. Kienlin, W. Kink, F. Korotkov, I. Kreykenbohm, G. Lamer, I. Lomakin, I. Lapshov, T. Liu, C. Maitra, N. Meidinger, B. Menz, A. Merloni, T. Mernik, B. Mican, J. Mohr, S. Müller, K. Nandra, V. Nazarov, F. Pacaud, M. Pavlinsky, E. Perinati, E. Pfeffermann, D. Pietschner, M. E. Ramos-Ceja, A. Rau, J. Reiffers, T. H. Reiprich, J. Robrade, M. Salvato, J. Sanders, A. Santangelo, M. Sasaki, H. Scheuerle, C. Schmid, J. Schmitt, A. Schwope, A. Shirshakov, M. Steinmetz, I. Stewart, L. Strüder, R. Sunyaev, C. Tenzer, L. Tiedemann, J. Trümper, V. Voron, P. Weber, J. Wilms, and V. Yaroshenko. The eROSITA X-ray telescope on SRG. , 647:A1, March 2021.

[10] et al. Merloni. The SRG/eROSITA all-sky survey. First X-ray catalogues and data release of the western Galactic hemisphere. , 682:A34, February 2024.

[11] E. Bulbul, A. Liu, M. Kluge, X. Zhang, J. S. Sanders, Y. E. Bahar, V. Ghirardini, E. Artis, R. Seppi, C. Garrel, M. E. Ramos-Ceja, J. Comparat, F. Balzer, K. Böckmann, M. Brüggen, N. Clerc, K. Dennerl, K. Dolag, M. Freyberg, S. Grandis, D. Gruen, F. Kleinebreil, S. Krippendorf, G. Lamer, A. Merloni, K. Migkas, K. Nandra, F. Pacaud, P. Predehl, T. H. Reiprich, T. Schrabback, A. Veronica, J. Weller, and S. Zelmer. The SRG/eROSITA All-Sky Survey. The first catalog of galaxy clusters and groups in the Western Galactic Hemisphere. , 685:A106, May 2024.

[12] Alam et al. The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. , 219(1):12, July 2015.

[13] A. Kurcz, M. Bilicki, A. Solarz, M. Krupa, A. Pollo, and K. Małek. Towards automatic classification of all WISE sources. , 592:A25, July 2016.

[14] Robert H. Becker, Richard L. White, and David J. Helfand. The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. , 450:559, September 1995.

[15] Ray P. Norris, Joshua Marvil, J. D. Collier, Anna D. Kapińska, Andrew N. O'Brien, L. Rudnick, Heinz Andernach, Jacobo Asorey, Michael J. I. Brown, Marcus Brüggen, Evan Crawford, Jayanne English, Syed Faisal ur Rahman, Miroslav D. Filipović, Yjan Gordon, Gülay Gürkan, Catherine Hale, Andrew M. Hopkins, Minh T. Huynh, Kim HyeongHan, M. James Jee, Bärbel S. Koribalski, Emil Lenc, Kieran Luken, David Parkinson, Isabella Prandoni, Wasim Raja, Thomas H. Reiprich, Christopher J. Riseley, Stanislav S. Shabala, Jaimie R. Sheil, Tessa Vernstrom, Matthew T. Whiting, James R. Allison, C. S. Anderson, Lewis Ball, Martin Bell, John Bunton, T. J. Galvin, Neeraj Gupta, Aidan Hotan,

Colin Jacka, Peter J. Macgregor, Elizabeth K. Mahony, Umberto Maio, Vanessa Moss, M. Pandey-Pommier, and Maxim A. Voronkov. The Evolutionary Map of the Universe pilot survey. , 38:e046, September 2021.

[16] T. W. Shimwell, M. J. Hardcastle, C. Tasse, P. N. Best, H. J. A. Röttgering, W. L. Williams, A. Botteon, A. Drabent, A. Mechev, A. Shulevski, R. J. van Weeren, L. Bester, M. Brüggen, G. Brunetti, J. R. Callingham, K. T. Chyży, J. E. Conway, T. J. Dijkema, K. Duncan, F. de Gasperin, C. L. Hale, M. Haverkorn, B. Hugo, N. Jackson, M. Mevius, G. K. Miley, L. K. Morabito, R. Morganti, A. Offringa, J. B. R. Oonk, D. Rafferty, J. Sabater, D. J. B. Smith, D. J. Schwarz, O. Smirnov, S. P. O'Sullivan, H. Vedantham, G. J. White, J. G. Albert, L. Alegre, B. Asabere, D. J. Bacon, A. Bonafede, E. Bonnassieux, M. Brienza, M. Bilicki, M. Bonato, G. Calistro Rivera, R. Cassano, R. Cochrane, J. H. Croston, V. Cuciti, D. Dallacasa, A. Danezi, R. J. Dettmar, G. Di Gennaro, H. W. Edler, T. A. Enßlin, K. L. Emig, T. M. O. Franzen, C. García-Vergara, Y. G. Grange, G. Gürkan, M. Hajduk, G. Heald, V. Heesen, D. N. Hoang, M. Hoeft, C. Horellou, M. Iacobelli, M. Jamrozy, V. Jelić, R. Kondapally, P. Kukreti, M. Kunert-Bajraszewska, M. Magliocchetti, V. Mahatma, K. Małek, S. Mandal, F. Massaro, Z. Meyer-Zhao, B. Mingo, R. I. J. Mostert, D. G. Nair, S. J. Nakoneczny, B. Nikiel-Wroczyński, E. Orrú, U. Pajdosz-Śmierciak, T. Pasini, I. Prandoni, H. E. van Piggelen, K. Rajpurohit, E. Retana-Montenegro, C. J. Riseley, A. Rowlinson, A. Saxena, C. Schrijvers, F. Sweijen, T. M. Siewert, R. Timmerman, M. Vaccari, J. Vink, J. L. West, A. Wołowska, X. Zhang, and J. Zheng. The LOFAR Two-metre Sky Survey. V. Second data release. , 659:A1, March 2022.

[17] P. N. Best and T. M. Heckman. On the fundamental dichotomy in the local radio-AGN population: accretion, evolution and host galaxy properties. , 421(2):1569–1582, April 2012.

[18] J. K. Banfield, O. I. Wong, K. W. Willett, R. P. Norris, L. Rudnick, S. S. Shabala, B. D. Simmons, C. Snyder, A. Garon, N. Seymour, E. Middelberg, H. Andernach, C. J. Lintott, K. Jacob, A. D. Kapińska, M. Y. Mao, K. L. Masters, M. J. Jarvis, K. Schawinski, E. Paget, R. Simpson, H. R. Klöckner, S. Bamford, T. Burchell, K. E. Chow, G. Cotter, L. Fortson, I. Heywood, T. W. Jones, S. Kaviraj, Á. R. López-Sánchez, W. P. Maksym, K. Polsterer, K. Borden, R. P. Hollow, and L. Whyte. Radio Galaxy Zoo: host galaxies and radio morphologies derived from visual inspection. , 453(3):2326–2340, November 2015.

[19] O. Ivy Wong, A. F. Garon, M. J. Alger, L. Rudnick, S. S. Shabala, K. W. Willett, J. K. Banfield, H. Andernach, R. P. Norris, J. Swan, M. J. Hardcastle, C. J. Lintott, S. V. White, N. Seymour, A. D. Kapińska, H. Tang, B. D. Simmons, and K. Schawinski. Radio Galaxy Zoo data release 1: 100185 radio source classifications from the FIRST and ATLAS surveys. , 536(4):3488–3506, February 2025.

[20] Ray P. Norris. Extragalactic radio continuum surveys and the transformation of radio astronomy. *Nature Astronomy*, 1:671–678, September 2017.

[21] R. Norris, K. Basu, M. Brown, E. Carretti, A. D. Kapinska, I. Prandoni, L. Rudnick, and N. Seymour. The SKA Mid-frequency All-sky Continuum Survey: Discovering the unexpected and transforming radio-astronomy. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 86, April 2015.

[22] Steven Ndung'u, Trienko Grobler, Stefan J. Wijnholds, Dimka Karastoyanova, and George Azzopardi. Advances on the morphological classification of radio galaxies: A review. , 97:101685, December 2023.

[23] Tapan K. Sasmal, Soumen Bera, Sabyasachi Pal, and Soumen Mondal. A New Catalog of Head-Tail Radio Galaxies from the VLA FIRST Survey. , 259(2):31, April 2022.

[24] Baoqiang Lao, Heinz Andernach, Xiaolong Yang, Xiang Zhang, Rushuang Zhao, Zhen Zhao, Yun Yu, Xiaohui Sun, and Sheng-Li Qin. Identification of 4876 Bent-tail Radio Galaxies in the FIRST Survey Using Deep Learning Combined with Visual Inspection. , 276(2):46, February 2025.

[25] Emmet Golden-Marx, E. L. Blanton, R. Paterno-Mahler, M. Brodwin, M. L. N. Ashby, E. Moravec, L. Shen, B. C. Lemaux, L. M. Lubin, R. R. Gal, and A. R. Tomczak. The High-redshift Clusters Occupied by Bent Radio AGN (COBRA) Survey: Radio Source Properties. , 907(2):65, February 2021.

[26] Sabyasachi Pal and Shobha Kumari. A new catalog of head–tail radio galaxies from LoTSS DR1. *Journal of Astrophysics and Astronomy*, 44(1):17, June 2023.

[27] S. Dehghan, M. Johnston-Hollitt, T. M. O. Franzen, R. P. Norris, and N. A. Miller. Bent-tailed Radio Sources in the Australia Telescope Large Area Survey of the Chandra Deep Field South. , 148(5):75, November 2014.

[28] Netai Bhukta, Sushanta K. Mondal, and Sabyasachi Pal. Tailed radio galaxies from the TIFR GMRT sky survey. , 516(1):372–390, October 2022.

[29] E. Vardoulaki, V. Backöfer, A. Finoguenov, F. Vazza, J. Comparat, G. Gozaliasl, I. H. Whittam, C. L. Hale, J. R. Weaver, A. M. Koekemoer, J. D. Collier, B. Frank, I. Heywood, S. Sekhar, A. R. Taylor, S. Pinjarkar, M. J. Hardcastle, T. Shimwell, M. Hoeft, S. V. White, F. An, F. Tabatabaei, Z. Randriamanakoto, and M. D. Filipovic. The jet paths of radio active galactic nuclei and their cluster weather. , 695:A178, March 2025.

[30] B. L. Fanaroff and J. M. Riley. The morphology of extragalactic radio sources of high and low luminosity. , 167:31P–36P, May 1974.

[31] B. Mingo, J. H. Croston, M. J. Hardcastle, P. N. Best, K. J. Duncan, R. Morganti, H. J. A. Rottgering, J. Sabater, T. W. Shimwell, W. L. Williams, M. Brienza, G. Gurkan, V. H. Mahatma, L. K. Morabito, I. Prandoni, M. Bondi, J. Ineson, and S. Mooney. Revisiting the Fanaroff-Riley dichotomy and radio-galaxy morphology with the LOFAR Two-Metre Sky Survey (LoTSS). , 488(2):2701–2721, September 2019.

[32] T. W. Shimwell and et al. The lofar two-metre sky survey. ii. first data release. *Astronomy & Astrophysics*, 622:A1, 2019.

[33] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2010.

[34] Taco S. Cohen and Max Welling. Group Equivariant Convolutional Networks. *arXiv e-prints*, page arXiv:1602.07576, February 2016.

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

[36] S. Sasmal, S. Paul, S. Hota, and et al. Cataloguing bent radio galaxies from vla first survey. *The Astronomical Journal*, 2022.

[37] Avery F. Garon, Lawrence Rudnick, O. Ivy Wong, Tom W. Jones, Jin-Ah Kim, Heinz Andernach, Stanislav S. Shabala, Anna D. Kapińska, Ray P. Norris, Francesco de Gasperin, Jean Tate, and Hongming Tang. Radio Galaxy Zoo: The Distortion of Radio Galaxies by Galaxy Clusters. , 157(3):126, March 2019.

[38] R. Paterno-Mahler, E. L. Blanton, J. D. Wing, and M. Brodwin. Cobra: Clusters occupied by bent radio agn. i. sample and x-ray properties. *The Astrophysical Journal*, 844:78, 2017.

[39] C. Megan Urry and Paolo Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. 107:803, September 1995.

[40] G. Miley. The structure and dynamics of radio galaxies and quasars. *Annual Review of Astronomy and Astrophysics*, 18:165–218, 1980.

[41] H. Miraghaei and P. N. Best. The host galaxies and large-scale environments of radio-loud agn. *Monthly Notices of the Royal Astronomical Society*, 466:4346–4368, 2017.

[42] R. D. Baldi and A. Capetti. Fr0 radio galaxies and the bulk of the radio-loud agn population. *Astronomy & Astrophysics*, 576:A38, 2015.

[43] F. N. Owen and L. Rudnick. Radio sources with wide-angle tails in Abell clusters of galaxies. , 205:L1–L4, April 1976.

[44] Christopher P. O'Dea and Stefi A. Baum. Wide-Angle-Tail (WAT) Radio Sources. *Galaxies*, 11(3):67, May 2023.

[45] Mitchell C. Begelman, Martin J. Rees, and Roger D. Blandford. Massive black hole binaries in active galactic nuclei. *Nature*, 279:770–773, 1979.

[46] James E. Gunn and J. Richard Gott. On the infall of matter into clusters of galaxies and some effects on their evolution. *The Astrophysical Journal*, 176:1–19, 1972.

[47] Torsten A. Enßlin, Peter L. Biermann, Ulrich Klein, and Sonja Kohle. Cluster radio relics as a tracer of shock waves of the large-scale structure formation. *Astronomy & Astrophysics*, 332:395–409, 1998.

[48] E. L. Blanton, M. D. Gregg, and D. J. Helfand. Bent-tail radio source environments: A probe of the intracluster medium. *The Astrophysical Journal*, 121:2915–2927, 2001.

[49] M. Y. Mao and et al. Morphology of bent-jet radio sources and cluster environment. *Monthly Notices of the Royal Astronomical Society*, 474:1507–1517, 2018.

[50] T. Venturi and et al. Merger dynamics and radio source morphology. *Astronomy & Astrophysics*, 554:A112, 2013.

[51] J. J. Condon, W. D. Cotton, E. B. Fomalont, and et al. Resolving radio sources in surveys. *The Astrophysical Journal*, 758:23, 2012.

[52] L. Saripalli and C. P. O'Dea. Morphological studies of bent radio galaxies in clusters. *The Astrophysical Journal*, 897:48, 2020.

[53] M. J. Hardcastle and et al. Deep lofar observations and radio agn morphologies. *Monthly Notices of the Royal Astronomical Society*, 488:1381–1399, 2019.

[54] T. Alger and et al. Machine learning in radio astronomy: Classification and source finding. *Monthly Notices of the Royal Astronomical Society*, 493:2590–2606, 2020.

[55] S. Wu and et al. Deep learning for radio galaxy morphology classification. In *Proceedings of Machine Learning Research*, pages 45–56, 2020.

[56] A. K. Aniyan and K. Thorat. Classifying radio galaxies with deep convolutional neural networks. *The Astrophysical Journal Supplement Series*, 230:20, 2017.

[57] K. L. Polsterer and et al. Automatic classification of radio galaxies using convolutional neural networks. *Astronomy and Computing*, 18:107–119, 2016.

[58] K. Hassani and A.H. Khasahmadi. Contrastive learning: A framework for semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[59] Maurice Weiler and Gabriele Cesa. *General E(2)-equivariant steerable CNNs.* Curran Associates Inc., Red Hook, NY, USA, 2019.

[60] R. Agarwal and et al. Uncertainty quantification in deep learning for radio astronomy. *The Astronomical Journal*, 157:13, 2019.

[61] 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*, 2(1):293–306, January 2023.

[62] T. Alger and et al. Combining machine learning and multi-wavelength data for robust radio galaxy classification. *Monthly Notices of the Royal Astronomical Society*, 480:2804–2816, 2018.

[63] M. S. Hossain and et al. Semi-supervised learning approaches to bent radio agn classification. *Monthly Notices of the Royal Astronomical Society*, 503:3120–3136, 2023.

[64] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

[65] Kathleen O'Neil and collaborators. A catalog of bent-tailed radio galaxies in galaxy clusters. *The Astronomical Journal*, 161:000–000, 2021.

[66] Y. Zhang, C. Wu, X. Zhang, and et al. Automatic detection and classification of radio galaxy images by deep learning. *Monthly Notices of the Royal Astronomical Society*, 511(3):3562–3574, 2022.

[67] J. Tang and et al. Impact of preprocessing on radio source classification. *Monthly Notices of the Royal Astronomical Society*, 494:4329–4342, 2020.

[68] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, May 2015.

[69] 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.

[70] J. Schlemper and et al. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53:197–207, 2019.

[71] S. Wu and et al. Deep residual networks for radio galaxy classification. *Monthly Notices of the Royal Astronomical Society*, 484:1292–1302, 2019.

[72] S. Slijepcevic and et al. Addressing class imbalance in radio galaxy morphology classification. *Monthly Notices of the Royal Astronomical Society*, 510:3825–3840, 2022.

[73] T. Y. Lin and et al. Focal loss for dense object detection. *IEEE International Conference on Computer Vision*, pages 2980–2988, 2017.

[74] 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)*, pages 618–626, 2017.

[75] S. Afshar and et al. Explainable deep learning models in medical image analysis. *Journal of Biomedical Informatics*, 108:103541, 2020.

[76] Larry Bradley et al. Photutils: an astropy package for detection and photometry of astronomical sources. *Astronomy and Computing*, 32:100402, 2020.

[77] D. J. Hand. Measuring classifier performance: A coherent alternative to the area under the roc curve. *Machine Learning*, 55:103–123, 2001.

[78] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 1321–1330. JMLR.org, 2017.

[79] Inigo V. Slijepcevic, Anna M. M. Scaife, Mike Walmsley, Micah Bowles, O. Ivy Wong, Stanislav S. Shabala, and Hongming Tang. Radio Galaxy Zoo: using semi-supervised learning to leverage large unlabelled data sets for radio galaxy classification under data set shift. , 514(2):2599–2613, August 2022.

[80] Inigo V. Slijepcevic, Anna M. M. Scaife, Mike Walmsley, Micah Bowles, O. Ivy Wong, Stanislav S. Shabala, and Sarah V. White. Radio galaxy zoo: towards building the first multipurpose foundation model for radio astronomy with self-supervised learning. *RAS Techniques and Instruments*, 3(1):19–32, January 2024.

[81] Larry Bradley, Brigitta Sipőcz, Thomas Robitaille, Erik Tollerud, Zè Vinícius, Christoph Deil, Kyle Barbary, Tom J Wilson, Ivo Busko, Axel Donath, Hans Moritz Günther, Mihai Cara, P. L. Lim, Sebastian Meßlinger, Zach Burnett, Simon Conseil, Michael Droettboom, Azalee Bostroem, E. M. Bray, Lars Andersen Bratholm, William Jamieson, Adam Ginsburg, Geert Barentsen, Matt Craig, Sergio Pascual, Shivangee Rathi, Marshall Perrin, and Brett M. Morris. astropy/photutils: 2.2.0, February 2025.

[82] Larry Bradley. Photutils: Python package for detection and photometry of astronomical sources. Zenodo, 2024. Version 1.x.

[83] Astropy Collaboration et al. The astropy project: community python software for astronomy. *The Astrophysical Journal*, 935:167, 2022.

[84] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[85] G. K. Miley, G. C. Perola, P. C. van der Kruit, and H. van der Laan. Active Galaxies with Radio Trails in Clusters. , 237(5353):269–272, June 1972.

[86] J. O. Burns. Stormy Weather in Galaxy Clusters. *Science*, 280:400, April 1998.

[87] Ray P. Norris et al. The evolutionary map of the universe pilot survey. *Publications of the Astronomical Society of Australia*, 38:e046, 2021.

[88] Devina Mohan and Anna M. M. Scaife. Quantifying uncertainty in deep learning approaches to radio galaxy classification. *Monthly Notices of the Royal Astronomical Society*, 511(3):3722–3739, 2022.