

Galaxy clusters in the LoTSS-DR3: Catalogues and detection pipeline for diffuse radio emission

C. Stuardi^{1,*}, G. Di Gennaro¹, A. Botteon¹, F. Braga², C. Gheller¹, F. Vazza², M. Balboni^{2,3},
N. Biava^{1,6}, A. Bonafede^{2,1}, M. Brüggen⁴, G. Brunetti¹, R. Cassano¹, M. Cianfaglione^{2,1}, V. Cuciti^{2,1},
F. De Gasperin¹, F. Gastaldello³, M. J. Hardcastle⁵, M. Hoeft⁶, H. J. A. Rottgering⁷, N. Sanvitale¹,
T. W. Shimwell^{8,7}, and R. J. van Weeren⁷

¹ INAF – Istituto di Radioastronomia (IRA), Via Gobetti 101, 40129 Bologna, Italy

² Dipartimento di Fisica e Astronomia, Università di Bologna, Via P. Gobetti 92/3, 40129 Bologna, Italy

³ INAF – IASF Milano, via A. Corti 12, 20133 Milano, Italy

⁴ Hamburger Sternwarte, University of Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany

⁵ Department of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

⁶ Thüringer Landessternwarte, Sternwarte 5, 07778 Tautenburg, Germany

⁷ Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

⁸ ASTRON, Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands

Received 13 February 2026 / Accepted 25 April 2026

ABSTRACT

Context. The third data release (DR3) of the LOFAR Two-metre Sky Survey (LoTSS) provides an unprecedented view of the northern sky at 144 MHz, containing more than 13 million radio sources. While compact sources can be efficiently identified with automated software packages, the detection of diffuse radio emission associated with galaxy clusters still requires dedicated processing and visual inspection. Given the scale of current and forthcoming radio surveys, automated approaches based on artificial intelligence are becoming essential to the identification of the most interesting targets.

Aims. We aim to develop an automated pipeline to construct a catalogue of galaxy clusters hosting diffuse radio emission from LoTSS-DR3 20" images. The pipeline is designed to provide both the probability that a cluster hosts diffuse radio emission and an interpretable image of its shape and morphology.

Methods. We employed Radio U-Net, a convolutional neural network optimised for image segmentation (i.e. pixel-level identification) of diffuse radio emission. To associate detected emission with individual clusters, we combined the network output with positional, mass, and redshift information from four X-ray- and Sunyaev–Zeldovich-selected cluster catalogues, resulting in a merged sample of 3822 clusters covered by the LoTSS-DR3.

Results. We produced a pixel-level segmentation map of the full LoTSS-DR3 and a quantitative indicator (\mathcal{R} value) for the presence of diffuse radio emission in each cluster. This enables the selection of sub-samples with specific properties for targeted follow-up or statistical studies. As a demonstration of the first application, we identified a sub-sample of 357 clusters selected at the highest network accuracy (76%), and we showed some examples of newly detected systems. For the second, using a larger statistical sample, we verified that the detection fraction of diffuse radio sources in the four catalogues increases with the mass and decreases with redshift of the clusters.

Conclusions. This work establishes a flexible and scalable framework for deep learning-based searches for diffuse radio emission in new-generation wide-area radio surveys.

Key words. techniques: image processing – catalogs – galaxies: clusters: intracluster medium

1. Introduction

Galaxy clusters represent the largest gravitationally bound structures in the Universe and form at the nodes of the cosmic web, a vast network of filaments on megaparsec scales through which matter accretes. The thermal plasma that fills galaxy clusters, known as the intracluster medium (ICM), is permeated by weak magnetic fields (typically 0.1–10 μG , see Govoni & Feretti 2004, for a review). During structure formation, mergers and accretion processes drive shocks and turbulence that can accelerate particles to relativistic energies (Brüggen et al. 2012; Brunetti & Jones 2014), producing synchrotron radiation detectable at radio wavelengths and thereby

revealing the presence of magnetic fields and relativistic electrons in the ICM.

Diffuse radio sources in galaxy clusters – such as giant radio halos, mini-halos, radio relics, and revived fossil plasma sources – are key tracers of these non-thermal components (van Weeren et al. 2019). Historically, these sources have been identified in surveys and then studied through targeted observations of individual systems (e.g. Feretti et al. 2012). In recent years, the advent of new-generation low-frequency interferometers such as the Low Frequency Array (LOFAR; van Haarlem et al. 2013), MeerKAT (Jonas & MeerKAT Team 2016), the Australian Kilometre Array Pathfinder (ASKAP; Hotan et al. 2021), and the Murchison Widefield Array (MWA; Tingay et al. 2013) has dramatically improved our capability to detect faint and extended radio emission. Diffuse radio sources

* Corresponding author: ccstuardi@gmail.com

typically have a surface brightness of $S_{1.4\text{GHz}} < 1 \mu\text{Jy arcsec}^{-2}$ at 1.4 GHz and become brighter at lower observing frequencies (ν) due to their steep radio spectrum ($\alpha > 1$, with $S_\nu \propto \nu^{-\alpha}$).

In particular, LOFAR was used to perform a systematic search for diffuse cluster radio emission at frequencies below 200 MHz in a mass-selected sample of 309 clusters (Botteon et al. 2022). This work has provided the largest collection to date, listing 138 clusters with diffuse radio emission (radio relics, radio halos, or sources with uncertain classification) and allowed subsequent statistical studies aimed at understanding the origin of the detected sources (Cuciti et al. 2023; Cassano et al. 2023; Jones et al. 2023; Bruno et al. 2023). Other recent collections include the MeerKAT Galaxy Cluster Legacy Survey (Knowles et al. 2022; Kolokythas et al. 2025), with 62 galaxy clusters hosting diffuse radio emission, and a pilot collection of 37 clusters in the Evolutionary Map of the Universe (EMU), performed with ASKAP (Duchesne et al. 2024). With the upcoming Square Kilometre Array (SKA; Dewdney et al. 2009), the number of known diffuse sources is expected to increase by orders of magnitude (e.g. Nuza et al. 2012; Cassano et al. 2015). More quantitatively, SKA1-Low is predicted to detect up to ~ 2600 radio halos out to $z \sim 0.6$ (Cassano et al. AASKAII), while the approximate number of detected radio relics will be ~ 1000 (Pal et al. AASKAII) in the southern hemisphere. This increase will require robust automated methodologies for the detection of these sources.

Traditional approaches to identifying diffuse radio emission in galaxy clusters have typically relied on manual or highly customised data reduction procedures and visual inspection (see e.g. Giovannini et al. 2020; Botteon et al. 2022; Knowles et al. 2022). However, given the scale of modern surveys, these methods are no longer feasible, owing both to the increase in the number of sources and to the computational cost and effort required for extensive data reprocessing. Robust identification often requires cross-matching with multi-wavelength data (optical, X-ray, and Sunyaev–Zeldovich observations) to confirm associations with galaxy clusters, while spectro-polarimetric analyses are necessary in many cases to firmly establish the nature of the detected sources. While human validation will likely remain essential, the development of automated, robust, and efficient detection techniques is crucial to fully exploit the scientific potential of upcoming high-volume datasets.

In recent years, deep learning (DL) techniques have emerged as powerful tools for astronomical image analysis, enabling automated source detection, classification, and segmentation (here, segmentation refers to the pixel-by-pixel identification of regions associated with specific astrophysical emission). Convolutional neural networks have been successfully applied to radio astronomical images, in particular to solve the task of source detection (e.g. Lukic et al. 2019; Riggi et al. 2023; Sortino et al. 2023; Cornu et al. 2024) and for the classification of radio galaxies (e.g. Riggi et al. 2024; Gupta et al. 2024; Lao et al. 2025). LOFAR images of radio galaxies have also been analysed using self-supervised learning (Mostert et al. 2021, 2023; Baron Perez et al. 2025).

The application of DL techniques to the detection and segmentation of diffuse radio emission is still relatively unexplored, primarily due to the limited number of known sources. Supervised networks require large training datasets, which, in the absence of sufficiently extensive observational samples, can be supplemented with simulated data. Early applications of DL to the recognition of diffuse radio emission in cosmological simulations were presented by Gheller et al. (2018), while Gheller & Vazza (2022) developed DL methods for the denois-

ing of radio images. The Radio U-Net model (Stuardi et al. 2024) has been successfully developed and optimised for the segmentation of diffuse cluster emission, and it was recently applied to LOFAR images for the first time, demonstrating a high performance in recovering extended low-surface-brightness structures at the detection limit (Stuardi et al. 2025). The application of vision transformers to the segmentation of diffuse radio emission in LOFAR images has been explored in Sanvitale et al. (2025), and other recent works applying DL to the detection of diffuse radio emission in galaxy clusters include Mishra et al. (2025) and Etsebeth et al. (2026).

In this paper, we apply Radio U-Net to the third data release of the LOFAR Two-metre Sky Survey (LoTSS-DR3; Shimwell et al. 2026), which covers 88% of the northern sky at 144 MHz. Our goal is to develop an automated DL-based pipeline capable of processing the full LoTSS-DR3 dataset and cross-matching the detected diffuse radio emission with four widely used galaxy cluster catalogues. The resulting merged catalogue provides a quantitative indicator of the presence of diffuse radio emission in each cluster, while the corresponding segmentation maps identify the location and morphology of the emission. At this stage, we do not make any attempt to automatically distinguishing different types of diffuse radio emission (e.g. radio relics, radio halos, revived fossil plasma), leaving source classification and the construction of science-ready catalogues as a task for potential users (e.g. Di Gennaro et al., in prep.). This approach is designed to provide a flexible and extensible framework for the systematic exploration of diffuse radio emission in galaxy clusters with current and future low-frequency radio surveys.

The paper is structured as follows: In Sect. 2 we introduce the four galaxy cluster catalogues used for cross-matching and explain our strategy to create a merged catalogue. In Sect. 3 we describe survey data, the machine learning algorithm, and the detection pipeline. In Sect. 4 we give an overview of the resulting catalogues and segmentation maps, while results are discussed in Sect. 5. We conclude and summarise our work in Sect. 6. Additional materials are collected in Appendices A, B, and C.

2. A merged ICM-selected catalogue of galaxy clusters

In order to associate diffuse radio emission with individual galaxy clusters, physical information such as cluster position, redshift, and mass is required. In particular, cluster mass is a key parameter in models of diffuse radio source formation, and it is essential for statistical studies of their occurrence (e.g. Cassano et al. 2007; Nuza et al. 2012; Cuciti et al. 2015). For this work, we made use of reliable mass estimates provided by cluster catalogues based on either Sunyaev–Zeldovich detections or X-ray measurements of the thermal emission from the ICM. We used the following four catalogues:

- The second Planck Sunyaev–Zeldovich cluster catalogue (PSZ2, Planck Collaboration XXVII 2016), with redshift updates from optical follow-ups (Buddendiek et al. 2015; Burenin 2017; Burenin et al. 2018; Barrena et al. 2018; Streblyanska et al. 2018, 2019; Aguado-Barahona et al. 2019; Boada et al. 2019; Zohren et al. 2019; Bahk & Hwang 2024). This catalogue contains 1334¹ clusters in the full sky and spans the redshift range $0.01 < z < 0.972$.

¹ The catalogue contains additional 319 clusters with no redshift information and therefore were not included in the final count.

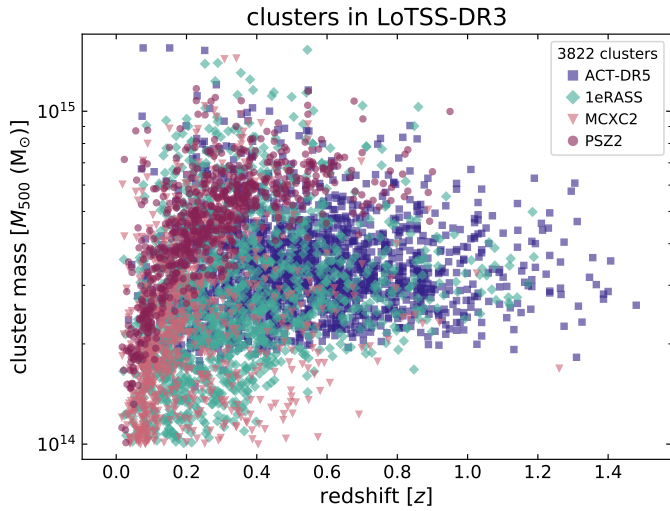


Fig. 1. Mass versus redshift plot of the 3822 galaxy clusters in the LoTSS-DR3 area after the merging of the four catalogues (ACT-DR5, 1eRASS, MCXC2, PSZ2). Following our selection criteria, only clusters with a mass above $10^{14} M_{\odot}$ were considered. In Appendix A, we show the plot of the individual catalogues before the merging.

- The fifth data release of the Atacama Cosmology Telescope cluster catalogue (ACT-DR5; Hilton et al. 2021). This catalogue spans mostly in the Southern Hemisphere (Dec $< +20^{\circ}$, for a total of $13\,168\text{ deg}^2$) and contains 4,195 optically confirmed Sunyaev–Zeldovich-selected galaxy clusters. The catalogue is 90% complete for masses $M_{500c} > 3.8 \times 10^{14} M_{\odot}$, and covers a redshift range of $0.04 < z < 1.91$.
- The second Meta-Catalogue of X-ray detected Clusters of galaxies, compiled from publicly available ROSAT All-Sky Survey and serendipitous X-ray cluster catalogues (MCXC2; Sadibekova et al. 2024), which contains 2221 clusters in the redshift range $0.003 < z < 1.261$. This catalogue updates the information on redshift, X-ray luminosity in the 0.2–2.4 keV range, and false detections provided by Piffaretti et al. (2011) in the first release, and additionally includes cluster members from the REFLEX-II (Böhringer et al. 2013), MACS (Mann & Ebeling 2012; Repp & Ebeling 2018), and RXGCC (Xu et al. 2022) catalogues.
- The first SRG/eROSITA All-Sky Survey cluster catalogue (1eRASS; Bulbul et al. 2024). This catalogue contains 12,247 optically confirmed galaxy groups and clusters detected in the 0.2–2.3 keV range located in the western Galactic half of the sky ($13\,116\text{ deg}^2$). The redshift range is $0.003 < z < 1.322$.

For the systems with available redshift information, we computed the corresponding R_{500} starting from the original mass given in the catalogue (M_{500}^2) and assuming a standard flat Λ CDM cosmology with $H_0 = 70\text{ km s}^{-1}\text{ Mpc}^{-1}$ and $\Omega_0 = 0.3$. The same cosmology was also used to convert the physical size of R_{500} into an angular scale. Duplicates across catalogues were merged when the separation of the two pairs matched within their $0.5R_{500}$ and their redshifts were consistent within the sum in quadrature of the two errors (i.e. $|z_1 - z_2| < \sqrt{(\Delta z_1)^2 + (\Delta z_2)^2}$). When no redshift error was provided in the catalogue, we

² M_{500} is defined as the total mass enclosed within a radius (R_{500}) where the mean density of the cluster is 500 times the critical density of the Universe at the cluster redshift.

Table 1. Galaxy clusters in the four catalogues.

Catalogue	$N_{\text{LoTSS-DR3}}$	N_{merged}	$N_{\text{merged } \mathcal{R}}$
PSZ2	686	686	626
ACT-DR5	1699	1617	1305
MCXC2	874	564	514
1eRASS	1324	955	922
Total	4853	3822	3367

Notes. Column 1: Catalogue name and total count. Column 2: Number of galaxy clusters within the LoTSS-DR3 area with $M_{500} > 10^{14} M_{\odot}$. Column 3: Number of galaxy clusters in the merged catalogue created as described in Sect. 2. Column 4: Number of galaxy clusters from the merged catalogue having an estimate of the \mathcal{R} value derived from the segmentation map (i.e. observed in a mosaic of quality 0 or 1, see Sect. 4).

adopted a photometric uncertainty³ of $\Delta z = 0.01 \times (1 + z)$. The grouping procedure followed the order of catalogues listed above so that all PSZ2 clusters are retained in the final catalogue, while ACT-DR5 clusters are included only if not already matched to PSZ2 entries, and similarly for MCXC2 and 1eRASS. All alternative names are kept in the merged catalogue. We note that some residual duplications may still be present, due to slight discrepancies in redshift estimates or positional information among the original catalogues.

From this large ICM-selected catalogue, we selected only the systems with mass $M_{500} \geq 10^{14} M_{\odot}$ that fall in the LoTSS-DR3 footprint. The final merged catalogue contains 3822 galaxy clusters, and the number of clusters belonging to each catalogue is listed in Table 1. Their mass-redshift distribution is shown in Fig. 1 while their position is shown in Fig. 2, overlaid on the Global Sky Model at 144 MHz derived by Zheng et al. (2017) with PyGDSM⁴. The four catalogues are not uniformly represented in the sample and are also unevenly distributed in the LoTSS-DR3 sky area, with the largest ACT-DR5 sample covering only the low-declination region by construction of the ACT survey.

3. Methods

In this section, we describe the methods that we have used to provide information on the presence of diffuse radio emission for each cluster in the merged catalogue created in Sect. 2. In particular, we first describe the LoTSS-DR3 (Sect. 3.1) and the main characteristics of the Radio U-Net architecture (Sect. 3.2), together with its implementation to process LoTSS-DR3 data and quality assessment of resulting segmentation maps (Sect. 3.3). In Sect. 3.4, we introduce the detection pipeline implemented to generate the final catalogue starting from the merged catalogue of galaxy clusters.

3.1. The LOFAR Two-metre Sky Survey third data release

The LoTSS is a wide-area 120–168 MHz survey aiming to cover the entire northern sky. Three partial LoTSS data releases

³ This is a rough estimate based on the fact that standard photometric redshift surveys report an uncertainty of $\sim 0.03\text{--}0.05 \times (1 + z)$ for individual galaxies (e.g. Beck et al. 2016; Wen & Han 2024), and that the cluster redshift is determined from $\sim 15\text{--}25$ member galaxies, reducing the uncertainty by a factor of $\sim 4\text{--}5$.

⁴ <https://github.com/telegraphic/pygdsms?tab=readme-ov-file>

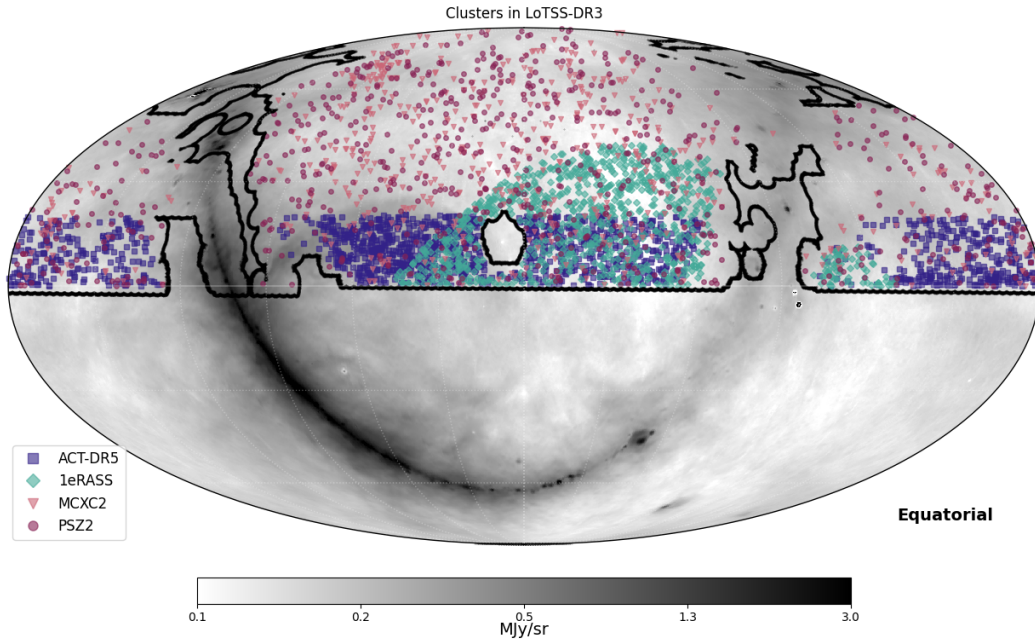


Fig. 2. Position of the 3822 galaxy clusters in the merged catalogue described in Sect. 2 overlaid on the Global Sky Model at 144 MHz derived by Zheng et al. (2017). The colour and shape of the markers represent different galaxy cluster catalogues, as explained by the legend. The black contours show the LoTSS-DR3 footprint.

have been published in the past years: a preliminary data release (Shimwell et al. 2017), and two full-quality releases, DR1 and DR2 (Shimwell et al. 2019, 2022). These releases provide Stokes I , Q , U , and V images, associated catalogues, and calibrated uv-data. The DR2 covered approximately 27% of the northern sky.

The recent DR3 (Shimwell et al. 2026) extends the sky coverage to 88% of the northern hemisphere, collecting 12 950 hours of observations over a period of 10.5 years. The final 120–168 MHz continuum images have angular resolution of $6''$ above declination 10° , increasing to $9''$ between 0° and 10° . The median rms sensitivity of the mosaics is $92 \mu\text{Jy beam}^{-1}$, improving to $68 \mu\text{Jy beam}^{-1}$ at high elevation but degrading to $183 \mu\text{Jy beam}^{-1}$ near the celestial equator, primarily due to projection effects that reduce the effective station collecting area. Low-resolution images with a common $20''$ resolution across the whole area are also released.

Upon completion, LoTSS will consist of 3168 individual pointings. The LoTSS-DR3 includes all fields observed before the suspension of LOFAR operations in September 2024 for the LOFAR 2.0 upgrade, excluding most of those within 10° of extremely bright radio sources (Cassiopeia A, Cygnus A, Virgo A, Taurus A, and Hercules A) which produce severe contamination in their surrounding fields. In total, 2551 pointings are included in DR3.

In many cases, the LoTSS-DR2 images are directly reused in DR3 (85% of DR2 pointings), whereas in others the images have been reprocessed using updated calibration pipelines or extended with additional observations. As with DR2, mosaics are constructed from a weighted combination of all datasets contributing to a given sky position, with the weighting determined by the local rms noise and the LOFAR primary beam response. Although DR3 adopts a revised mosaicking strategy designed to optimise the combination of overlapping observations, in this work, we used mosaics created with the same strategy as in DR2

(i.e. having the same size and name as individual LoTSS pointings) since the analysis was developed before the implementation of the new approach. Since mosaics can be formed at any position on the sky, this choice does not affect the quality of the data.

While LOFAR is well suited to detect extended and diffuse radio emission, thanks to its excellent uv-coverage at the shortest baselines, the automated calibration and imaging strategy adopted for the LoTSS is not optimised for such emission. As outlined by Shimwell et al. (2022), it is unavoidable that extended low-surface-brightness emission is occasionally not recovered by the deconvolution algorithm during calibration, leading to a partial suppression of its flux if the unrecovered flux is significant compared to the total in the field. Furthermore, if the emission is not completely deconvolved in the final image, its apparent brightness can be different from the real one (see section 3.4 of Shimwell et al. 2022, for a quantification of this effect). The latter effect is mitigated in low-resolution images at $20''$, where diffuse radio emission is generally deconvolved.

When measuring the flux densities of diffuse radio sources, dedicated pipelines are required for accurate source extraction and self-calibration of individual targets (van Weeren et al. 2021). This approach was adopted, for example, for the sample of 309 galaxy clusters from the PSZ2 catalogue within the LoTSS-DR2 area (hereafter, the LoTSS-DR2/PSZ2 sample), presented by Botteon et al. (2022). Applying such a procedure to a single galaxy cluster observed in four pointings may require 24–36 hours on a machine with 512 GB of RAM and 64 AMD cores, with the total processing time scaling linearly with the number of clusters. Consequently, the selection of the most suitable targets in LoTSS-DR3 for further processing with this dedicated pipeline – which will be the subject of future works – represents a key motivation for the present study.

3.2. Radio U-Net

Radio U-Net⁵ is a convolutional neural network developed to identify diffuse radio emission in interferometric images (Stuardi et al. 2024). The network has been implemented using Python and exploiting the KERAS software package (Chollet 2015), distributed as part of the Tensorflow framework (Abadi et al. 2015), version 2.3.0. Here we only give an overview of the network design and implementation. We refer the reader to Stuardi et al. (2024) for a full description of the network architecture, training dataset, and procedure as well as validation tests.

The model follows the U-Net design (Ronneberger et al. 2015), consisting of a contracting path and a symmetric expanding path. In the contracting path, the input image is progressively compressed through successive convolution, activation, normalisation, and pooling layers, allowing the network to learn increasingly abstract features of the image. In particular, in our implementation, at each of the four contracting levels two convolutional layers are used for feature extraction, followed by a ReLu (Rectified Linear Unit) activation function to introduce non-linearity (Agarap 2018). Batch normalisation is used after each convolutional layer to improve convergence, and the dimensional reduction is obtained with a max pooling function. The expanding path reconstructs the image to its original resolution in four levels, using transpose convolutions, followed by ReLu activation, and upsampling layers.

The final softmax layer produces pixel-by-pixel information about the presence of the diffuse radio emission. The output image has the same size as the input image, but pixel values range from 0 to 1, with higher values indicating the presence of diffuse radio emission. No thresholding is applied by the network to the segmented maps to preserve the full probabilistic information. However, a threshold can be introduced to create a binary map for visualisation and qualitative assessment purposes (e.g. in Sect. 3.3). This type of pixel-level classification is named semantic segmentation, and it is particularly effective for the detection of complex and multi-scale objects, providing information on both their shape and position in large images (see Csurka et al. 2022, for a review).

The network was trained using synthetic radio images derived from cosmological magneto-hydrodynamical simulations, where diffuse radio emission was modelled from shock-accelerated electrons and projected into mock lightcones (Vazza et al. 2019). More details about the creation of these synthetic sky images are given in Gheller & Vazza (2022), Sect. 4 of Stuardi et al. (2024), and references therein. These simulations do not include radio emission generated by turbulent re-acceleration, as expected for radio halos or bridges (Brunetti 2011; Brunetti & Vazza 2020). Nevertheless, the resulting images reproduce a range of morphologies and surface brightness distributions that overlap with those observed in radio halos and bridges, and previous works have shown that such sources can still be effectively detected by our approach (Stuardi et al. 2024; Sanvitale et al. 2025).

Synthetic sky images of diffuse radio emission were then processed to mimic LOFAR HBA observations at 6'' resolution, including Gaussian noise and imperfect deconvolution, so as to reproduce typical observational artefacts. They do not include other real radio sources. These mock images were used as input to the network, while a binary reference masks derived from the underlying sky images were employed to train the network to identify regions containing diffuse radio emission

above 10^{-8} Jy pixel⁻¹. This value is a factor $\sim 10^{-3}$ lower than the median noise level to train the network to recognise emission structures in low signal-to-noise conditions. Input images are also transformed in logarithmic scale and normalised between 0 and 1, using 10^{-10} and 10^{-5} Jy pixel⁻¹ as minimum and maximum data normalisation values. Pixels outside this range were clipped to the boundaries. The use of min-max normalisation does not lead to a significant loss of information, because it is applied after a logarithmic transformation of the input images, which enhances low-surface-brightness emission, while normalisation bounds are chosen to encompass the typical range relevant for diffuse radio emission. Synthetic sky models and images used for training are publicly available⁶.

To handle large input sizes (2000×2000 pixels), the images were divided into 192×192 pixel tiles. The training set comprises 100 images (i.e. 10 000 tiles), but a fraction of them (5%, that is, 500 tiles) is kept as a validation set to monitor the performance of the network during training. This choice represents a trade-off between ensuring a reliable estimate of the model performance and maximising the amount of data available for training, which is crucial for DL models. The tiles are all independently processed during training and validation. During the inference phase (i.e. when the network is applied to new images), tiles are instead extracted with partial overlap. Although each tile is still processed independently, only the central regions are retained and subsequently reassembled, in order to minimise boundary effects and avoid discontinuities at tile edges.

Training relied on categorical cross-entropy as the loss function and RMSprop as the optimiser (Tieleman & Hinton 2012; Goodfellow et al. 2016). Categorical cross-entropy measures how well the predicted probabilities match the true class (0 or 1 in our case), strongly penalising confident wrong predictions. The loss function is minimised during the training to obtain accurate predictions. Hyperparameters (e.g. learning rate 10^{-4} , batch size 50, 200 epochs) were selected following earlier optimisation, where the training was monitored using the validation loss to assess convergence and prevent overfitting (Stuardi et al. 2024).

Once trained, the model can process full-size simulated images (i.e. 2000×2000 pixels) in less than a second on a single NVIDIA Tensor Core GPU. This efficiency makes Radio U-Net particularly well-suited for large-scale surveys. In Stuardi et al. (2024), the network was applied to archival LoTSS-DR2 cut-outs of a test sample of galaxy clusters. The input images were first converted to a logarithmic scale and normalised to the range 10^{-7} – 10^{-2} Jy beam⁻¹. With respect to previous work, only the data ingestion stage of the pipeline (responsible for reading the input images) was modified, in order to handle images of varying sizes and to process LoTSS mosaics with non-uniform fields of view.

The entire LoTSS-DR3 at 20'' resolution, for a total of 2551 mosaics of size about 3000×3000 pixels, was processed with Radio U-Net in four hours, by distributing parallel jobs on the Leonardo⁷ supercomputer at the CINECA Italian Supercomputing Centre and offloading data reading into GPUs. The resulting dataset, consisting of thousands of FITS files matching the size of the original LoTSS pointings, occupies 110 GB.

3.3. Quality assessment of segmented images

From an initial visual inspection of the segmented images, we noticed that some pointings were incorrectly segmented because

⁶ <https://owncloud.ia2.inaf.it/index.php/s/IbFP1CCcPUresrr>

⁷ <https://www.hpc.cineca.it/systems/hardware/leonardo/>

⁵ <https://github.com/ICSC-Spoke3/Radio-U-Net>

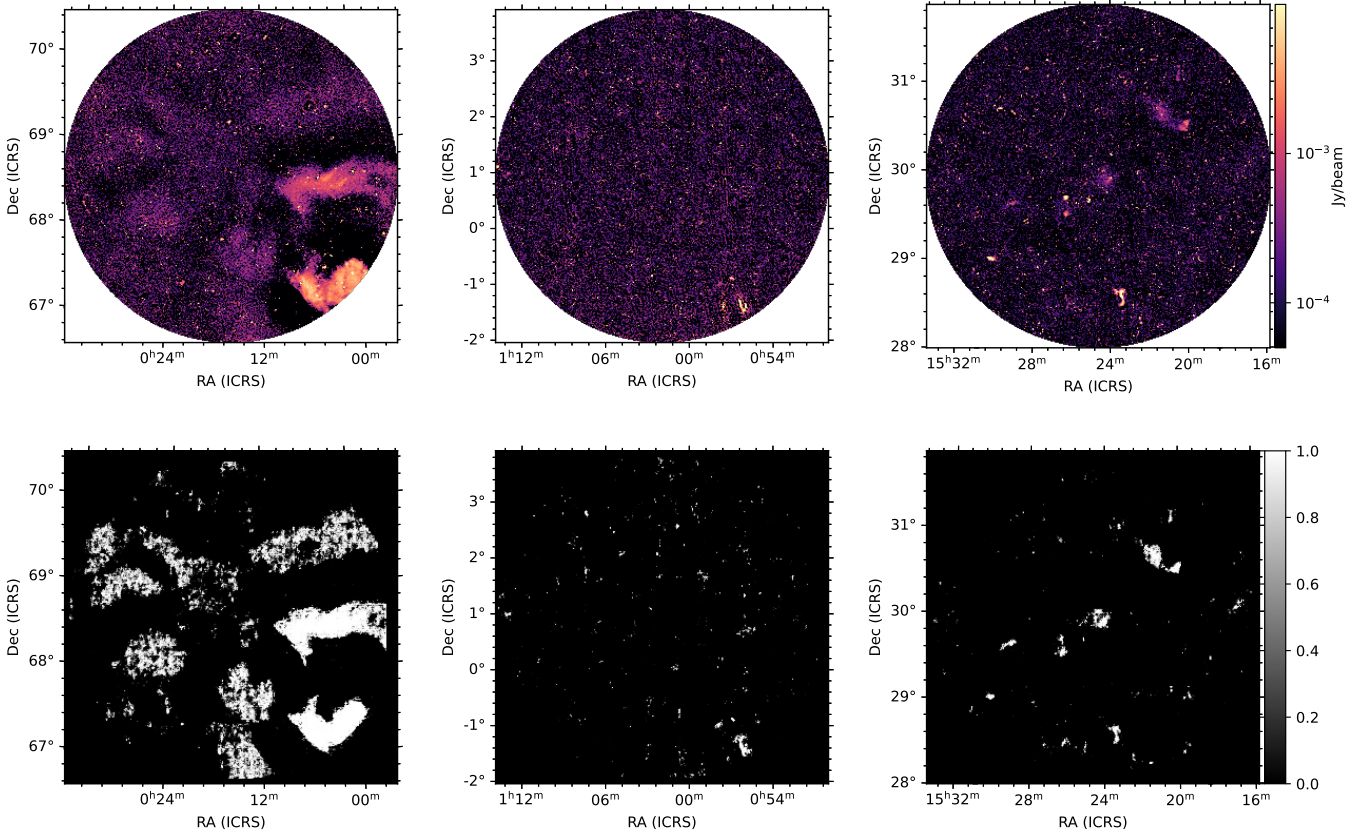


Fig. 3. Representative example of quality 2 (top and bottom left panels), quality 1 (top and bottom central panels), and quality 0 (top and bottom right panels) mosaics. The top row shows the LoTSS-DR3 mosaics with a common colour-scale. The bottom row shows the corresponding segmented maps produced by Radio U-Net. While quality class 2 mosaics show clear large-scale patterns, quality class 1 mosaics are characterised by a scattered distribution of erroneously segmented pixels.

of the presence of Galactic diffuse radio emission. This effect occurred mostly at Galactic latitude $|l| < 15^\circ$ but even up to $l = 75^\circ$ in the region of the North Polar Spur. An example of a pointing affected by this problem is shown in Fig. 3, left panel, with the corresponding segmentation map in the bottom panel. Galactic diffuse radio emission, which is extended over large angular scales and is observed in the form of large-scale ripples in the LoTSS survey, is recognised and segmented by the network. Galactic emission was not present in the original training dataset, which was designed to reproduce LoTSS-DR2 data quality and tested on sky regions where the Galactic foreground is negligible. As a consequence, the network is not able to exclude this emission from the segmentation maps, with the risk of misidentifying genuine diffuse radio emission of extragalactic origin.

In addition, we found that other pointings – mostly at low declination ($\delta < 15^\circ$) and at Galactic latitude $15^\circ < |l| < 30^\circ$ – show some signs of noisy segmentation, such as an increased number of scattered pixels with value approaching unity. A representative example is shown in Fig. 3, central panel, with the corresponding segmentation map in the bottom panel. This is likely due once again to the fact that the network cannot perfectly generalise to sky regions not represented in the training set. Still, in these pointings, the network correctly detects known diffuse radio emission of extragalactic origin. We therefore decided to retain these intermediate quality pointings in our analysis and provide segmentation maps for them, but with a distinct quality classification.

We note that segmentation quality is not directly related to the noise level of the mosaics, since large-scale Galactic ripples typically do not significantly affect the rms estimate, which is computed on small scales. Moreover, there is no exact correspondence between the presence of Galactic emission in LoTSS images and the surface brightness of the Galaxy derived from single-dish maps since the minimum baseline in regular LoTSS processing is 100 m. These considerations motivate a visual classification rather than cuts based on image noise or Galactic emission. The classification was performed on segmentation maps clipped at a threshold of 0.5, chosen as an intermediate value between 0 and 1 for visualisation purposes only.

Three classes were defined as follows:

- Class 0: Mosaics with segmented images of the same quality as LoTSS DR2 mosaics, where Radio U-Net was originally assessed.
- Class 1: Mosaics with correctly segmented images but generally showing a higher noise level (i.e. $>150 \mu\text{Jy beam}^{-1}$), resulting in an increased number of scattered pixels with a value larger than 0.5. While quality class 1 mosaics can still be used to search for diffuse radio emission, we recommend exercising caution when inspecting these segmented maps, as they are more prone to misidentifying imaging artefacts as diffuse structures.
- Class 2: Mosaics with clear segmentation errors easily visible as coherent stripes of pixels with values larger than 0.5. In order to avoid contamination of genuine diffuse cluster emission by Galactic foregrounds, we decided to exclude these pointings from further analysis.

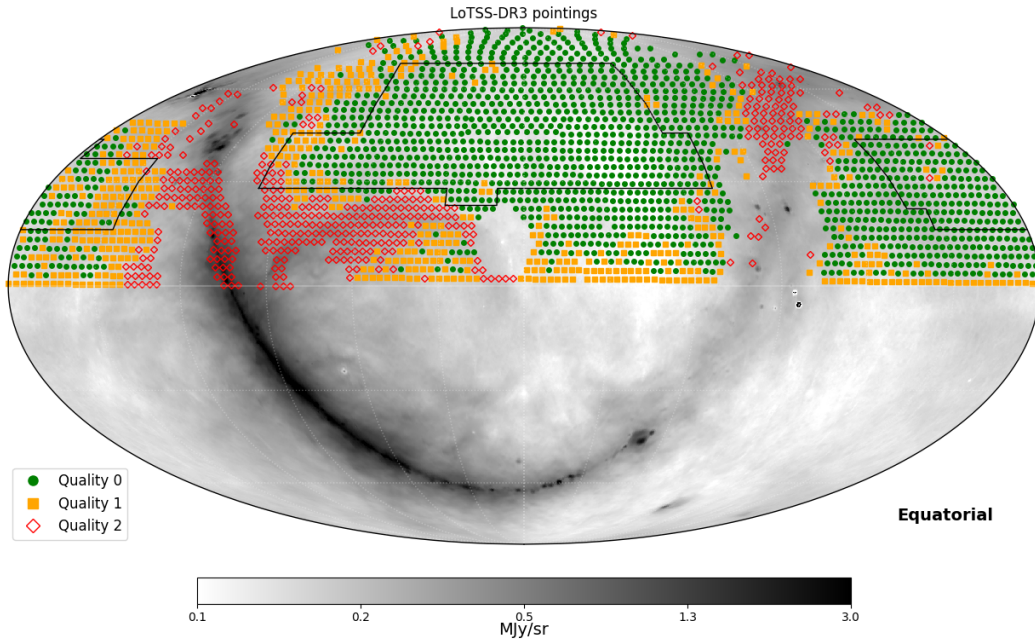


Fig. 4. Position of the 2551 pointings of the LoTSS-DR3 overlaid on the Global Sky Model at 144 MHz derived by Zheng et al. (2017). The colour and shape of the markers represent the quality class given in Sect. 3.3, as explained by the legend. In particular, red empty diamonds are quality 2 pointings, which were excluded by this analysis mainly due to the contamination of the Milky Way emission. Black lines show the LoTSS-DR2 footprint.

Among a total of 2551 LoTSS-DR3 mosaics, 1545 (60%) are of quality 0, 579 (23%) of quality 1, and 427 (17%) of quality 2. The position of the mosaics coloured by the three quality classes is shown in Fig. 4, overlaid on the Global Sky Model at 144 MHz. Their distribution is a direct consequence of the fact that we are extrapolating Radio U-Net results to sky regions characterised by substantially higher noise levels, and the network exhibits limited generalisation capability under these conditions. A possible solution to this problem is discussed in Sect. 5.3, but its implementation is out of the scope of this work.

We note that excluding the Galactic region (i.e. quality 2 mosaics) from the subsequent analysis does not significantly affect the detection pipeline based on the merged catalogue of galaxy clusters, as these catalogues already avoid most of it (see Fig. 2). In total, 455 out of 3822 galaxy clusters in the merged catalogue lie in regions classified with quality 2, of which 312 belong to the ACT-DR5 catalogue (a summary is in Table 1).

3.4. The detection pipeline

While the segmented images produced by Radio U-Net can be directly exploited for blind searches of diffuse radio emission across the entire LoTSS-DR3, additional physical information is required to assess whether the detected emission is associated with a galaxy cluster. In particular, cluster position, redshift, and mass are needed and can be retrieved from the merged catalogue created in Sect. 2. Excluding the 455 galaxy clusters lying in quality 2 mosaics, we applied the detection pipeline on the remaining 3367 galaxy clusters. The number of galaxy clusters considered in the subsequent analysis is reported in Table 1, subdivided among the four original catalogues ($N_{\text{merged } \mathcal{R}}$).

Using the position and angular size of each of the 3367 galaxy clusters, we computed for every system a single quantity that traces the presence of diffuse radio emission within the clus-

ter’s projected area. Following Stuardi et al. (2024), we defined the \mathcal{R} value as the ratio between the sum of pixel values in the segmented image within a circle of radius $2.2R_{500}$ centred on the cluster and the total number of pixels enclosed by that circle ($N_{2.2R_{500}}$):

$$\mathcal{R} = \frac{\sum_{m,n}^{2.2R_{500}} s_{m,n}}{N_{2.2R_{500}}}, \quad (1)$$

where $s_{m,n}$ is the output of the network at the (m,n) pixel. By design, \mathcal{R} approaches zero when no source is detected and reaches one when a diffuse radio emission completely fills the circle. The value of $2.2R_{500}$ was chosen, as it is the maximum distance at which radio relics were detected in the LoTSS-DR2/PSZ2 sample (Jones et al. 2023). This makes \mathcal{R} an effective discriminator between clusters with and without diffuse radio emission.

In Stuardi et al. (2024), using the galaxy cluster catalogue presented by Botteon et al. (2022) as a benchmark, we showed that more than 80% of PSZ2 clusters in LoTSS-DR2 hosting diffuse radio emission have $\mathcal{R} > 0.015$. Adopting this threshold to separate clusters with diffuse emission (DE) from those without (NDE) results in an overall classification accuracy of 73%, defined as the fraction of correctly classified clusters over the total sample. We note that galaxy clusters hosting radio sources classified as ‘uncertain’ by Botteon et al. (2022) – that is, sources whose emission is significantly affected by imaging artefacts or whose morphology, size, and/or position are inconsistent with standard halo or relic classifications – are included in the DE category. Although not clearly classified, these systems do host diffuse radio emission, making the \mathcal{R} value sensitive also to revived fossil plasma sources or radio phoenixes (see Bruno et al. 2025, for a recent work on these sources).

For each cluster in the merged catalogue, we annotated the survey pointings in which it was observed, its angular distance

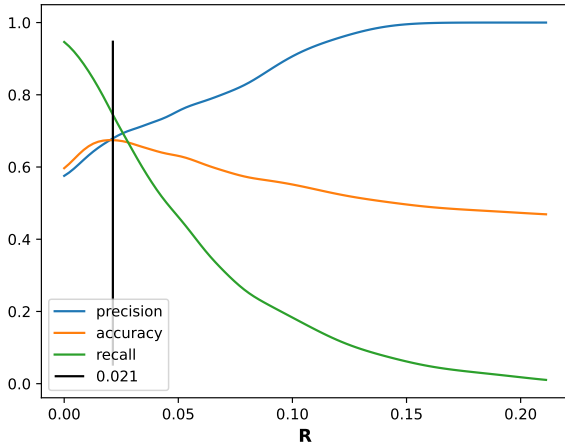


Fig. 5. Binary classification metrics for the test sample of 246 LoTSS-DR2/PSZ2 galaxy clusters (Botteon et al. 2022) as a function of the value of \mathcal{R} used as a threshold, considering its uncertainties.

from the pointing centre, and the local noise at the cluster position in each individual pointing prior to mosaicking. This information is useful for any subsequent reprocessing of individual targets, where a careful selection of the relevant pointings may be required. It is also important in light of the different mosaicking strategies adopted in the LoTSS DR2 and DR3. In the DR2 scheme (Shimwell et al. 2022), mosaics retain the same footprint as individual pointings and may therefore overlap. In contrast, DR3 mosaics are generated on a HEALPix grid and do not overlap. Although a given cluster should appear identical in multiple mosaics, the segmentation quality of the mosaics can vary, since each mosaic covers a different region of the sky.

When a cluster was observed in multiple pointings, the \mathcal{R} value was measured in the quality 0 mosaic where the cluster lies closest to the centre. If no quality 0 mosaic was available or if the cluster (within its $2.2R_{500}$) fell outside the mosaic, the nearest quality 1 mosaic was used. This quality is stored as a quality class for the \mathcal{R} value (0 or 1). Of the 3367 clusters outside of quality 2 regions, 2216 have quality class 0 and 1151 have quality class 1. In rare cases where a cluster lies at the boundary between two mosaics, the segmented mosaics were first merged before estimating \mathcal{R} . This occurred only seven times, and for these, the mosaics combined to compute \mathcal{R} are listed in the ‘Comment’ column of the catalogue. We also annotated cases in which the \mathcal{R} value was computed from a mosaic whose corresponding pointing is not present in the LoTSS-DR3 but instead includes only the edge regions of neighbouring pointings. This situation arises from the use of the DR2 mosaicking strategy and affects only 28 clusters located near the edges of the DR3 footprint. Furthermore, in the ‘Overlap’ column, we list any clusters whose projected positions overlap within $2.2R_{500}$ on the sky plane, even when they lie at different redshifts. Since \mathcal{R} does not account for overlaps, the computed value integrates over all overlapping systems.

Statistical uncertainties on the \mathcal{R} values were derived on the basis of the results obtained in Stuardi et al. (2024), as detailed in Appendix B. Uncertainties increase with the redshift of the galaxy cluster. To define a threshold in \mathcal{R} for distinguishing DE from NDE clusters, we followed Stuardi et al. (2024), computing classical evaluation metrics (accuracy, precision and recall) for a test sample of 246 LoTSS-DR2/PSZ2 galaxy clusters from Botteon et al. (2022) as functions of the chosen \mathcal{R} value. Unlike

the previous analysis, we now include the uncertainty in \mathcal{R} for each cluster.

Each cluster i in a sample of N galaxy clusters is associated with an observed value \mathcal{R}_i and measurement uncertainty σ_i . We assumed the true value, $\mathcal{R}_i^{\text{true}}$, follows a normal distribution. For a classification threshold, \mathcal{R}_t , the probability that object i belongs to class DE (i.e. that its true value exceeds the threshold) is

$$p_i(\mathcal{R}_t) = P(\mathcal{R}_i^{\text{true}} > \mathcal{R}_t) = 1 - \Phi\left(\frac{\mathcal{R}_t - \mathcal{R}_i}{\sigma_i}\right),$$

where Φ is the cumulative distribution function of the standard normal distribution.

Let $y_i \in \{0, 1\}$ be the true class labels of the 246 galaxy clusters (NDE and DE) chosen as a test sample from Botteon et al. (2022). The expected counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for threshold \mathcal{R}_t are

$$\mathbb{E}[\text{TP}(\mathcal{R}_t)] = \sum_i y_i p_i(\mathcal{R}_t),$$

$$\mathbb{E}[\text{FP}(\mathcal{R}_t)] = \sum_i (1 - y_i) p_i(\mathcal{R}_t),$$

$$\mathbb{E}[\text{TN}(\mathcal{R}_t)] = \sum_i (1 - y_i) (1 - p_i(\mathcal{R}_t)),$$

$$\mathbb{E}[\text{FN}(\mathcal{R}_t)] = \sum_i y_i (1 - p_i(\mathcal{R}_t)).$$

From these, we computed the expected accuracy (fraction of correct classifications to the total number of objects), precision (the ratio of the TP to the total number of objects identified as positive), and recall (the ratio of the TP to the total number of DE objects):

$$\mathbb{E}[\text{Accuracy}(\mathcal{R}_t)] = \frac{\mathbb{E}[\text{TP}(\mathcal{R}_t)] + \mathbb{E}[\text{TN}(\mathcal{R}_t)]}{N},$$

$$\mathbb{E}[\text{Precision}(\mathcal{R}_t)] = \frac{\mathbb{E}[\text{TP}(\mathcal{R}_t)]}{\mathbb{E}[\text{TP}(\mathcal{R}_t)] + \mathbb{E}[\text{FP}(\mathcal{R}_t)]},$$

$$\mathbb{E}[\text{Recall}(\mathcal{R}_t)] = \frac{\mathbb{E}[\text{TP}(\mathcal{R}_t)]}{\sum_i y_i}.$$

While precision highlights the purity of DE predictions, the recall focuses on the network’s ability to identify all DE galaxy clusters. These definitions are the same used by Stuardi et al. (2024), with the only variation of considering the uncertainty on the \mathcal{R} value in the computation.

The optimal threshold \mathcal{R}_t^* is chosen as the value of \mathcal{R}_t that maximises the selected performance metric. Once \mathcal{R}_t^* is determined, clusters are classified deterministically: as DE if $\mathcal{R}_i > \mathcal{R}_t^*$ and as NDE otherwise. As shown in Fig. 5, the classification accuracy peaks at $\mathcal{R}_t^* = 0.021$ where the accuracy is 67%, the precision is 68%, and the recall is 74%. Choosing a higher precision threshold helps build a sample with fewer false detections of DE, whereas prioritising higher recall reduces the risk of missing DE galaxy clusters, at the cost of increasing FPs.

To highlight potentially uncertain classifications, it is possible to introduce a binary classification flag. A classification is considered reliable if its measured \mathcal{R} value lies more than one standard deviation from the threshold:

$$|\mathcal{R}_i - \mathcal{R}_t^*| > \sigma_i. \quad (2)$$

This approach ensures that both threshold selection and individual measurement uncertainties are explicitly accounted for in the final classification.

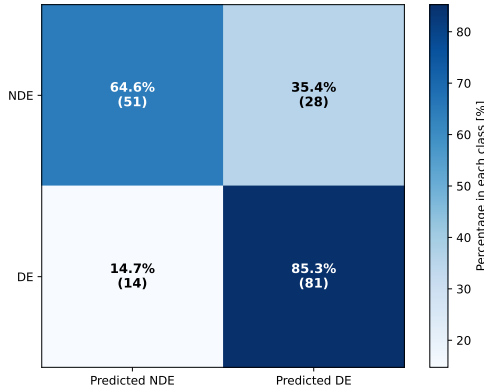


Fig. 6. Confusion matrix obtained for the test sample of 174 galaxy clusters with a good classification flag using $\mathcal{R}_t^* = 0.021$ as a detection threshold. The number in brackets is the total number in each box.

In the case that $\mathcal{R}_t^* = 0.021$ is chosen as a threshold, we verified that classification metrics improve using the same test sample from Botteon et al. (2022) but excluding clusters with an unreliable classification flag. Although this further selection reduced the test sample to 174 clusters, it allows us to achieve 76% accuracy, 74% precision, and 85% recall in the classification—meaning that 85% of clusters hosting genuine diffuse radio emission are correctly identified as DE. Out of the correctly identified DE galaxy clusters, 54 host radio halos or candidate radio halos, 22 host at least a radio relic or a candidate, and 18 host only a diffuse radio emission classified as uncertain by Botteon et al. (2022).

The confusion matrix obtained with the detection pipeline on this test sample of 174 galaxy clusters is shown in Fig. 6. Notice, however, that this residual sample is slightly biased, counting for 95 DE clusters and 79 NDE clusters. We note that a relatively large fraction of NDE systems is classified as DE by the network (35%). This may have two main causes. First, some NDE systems may host extended radio emission associated with radio galaxies, which the network identifies as DE. Second, the NDE classification based on visual inspection may be uncertain in borderline cases, where human judgement is also intrinsically limited, and faint diffuse radio emission may be overlooked. In some cases, the network may, in fact, detect low brightness emission that is close to the noise level and difficult to identify.

The detection pipeline thus provides for each galaxy cluster a quantitative indicator – that is, the \mathcal{R} value – which can be used, together with its uncertainty, to identify clusters hosting diffuse radio emission. The choice of $\mathcal{R}_t^* = 0.021$ as a threshold to separate clusters with and without diffuse radio emission maximises the accuracy of this classification. However, since the above performance metrics are computed on the LoTSS-DR2/PSZ2 galaxy cluster sample, which is mostly located in high-quality LoTSS regions, we note that, although for clusters with quality class 1 the \mathcal{R} value still provide an indication of the presence of diffuse radio emission, the classification performance based on the \mathcal{R} value may be less reliable. Furthermore, it is not possible to verify how the classification metrics depend on the original galaxy cluster catalogue since, to date, the only reference sample is based on the PSZ2 catalogue.

4. Results

In this section, we first describe the main results of this work (Sect. 4.1), which consist of the following:

- Segmentation maps of the LoTSS-DR3 produced with Radio U-Net, associated with the full list of 2551 LoTSS-DR3 pointings and the quality classes associate to each mosaic (having the same shape and name of individual pointings as for the DR2 mosaicking scheme), as described in Sect. 3.3;
- A merged catalogue of 3822 galaxy clusters within the LoTSS-DR3 footprint, of which 3367 have an estimated \mathcal{R} value obtained from Radio U-Net results.

Finally, we illustrate how these results can be used to select a sub-sample of galaxy clusters with a high probability of hosting diffuse radio emission and provide some examples (Sect. 4.2).

4.1. Segmentation maps and a merged catalogue of galaxy clusters with \mathcal{R} values

Example of segmentation maps produced by Radio U-Net are shown in Fig. 3. These maps provide pixel-level information about the presence of diffuse radio emission as recognised by the network. The resolution of these images is the same as the original LoTSS images at 20'', having a pixel size of 4.5''. Values approaching 1 are recognised by the network as part of diffuse radio emission. Generally, the boundaries of these regions are very sharp, with values decreasing below 0.5 within 4–5 pixels.

As extensively discussed in Stuardi et al. (2024), Radio U-Net tends to misidentify radio emission from extended, low-surface-brightness active galactic nuclei (AGNs) as diffuse radio emission related to the ICM. This behaviour arises because AGN-related emission was not included in the original training set. A similar misidentification can also occur for nearby galaxies, whose radio emission may appear approximately circular and thus resemble that of radio halos. Finally, as discussed in Sect. 3.3, pointings characterised by the presence of Galactic diffuse radio emission or by higher noise levels are more prone to misidentification by the network (see two examples in Fig. 3).

To provide a standardised access and visualisation of the segmentation maps, we have created a Hierarchical Progressive Surveys (HiPS, Fernique et al. 2015) map of all quality classes 1 and 0 mosaics, displaying only pixels with values above 0.9. This threshold was chosen to optimise the visualisation of large fields of view, highlighting only regions detected with the highest probability. This map is available on the dedicated page of the survey website⁸ and can be visualised on Aladin⁹ for the inspection of the full DR3 area (excluding Galactic regions). We also provide a table listing the 2551 LoTSS-DR3 pointings with the associated mosaic's quality class, while segmentation maps of single mosaics are made available upon request.

While segmentation maps can be used to readily identify the presence of diffuse radio emission, a comparison with multi-wavelength data is required to associate the emission detected by the network with individual objects. For this reason, and building on the results presented by Stuardi et al. (2024), we cross-matched the segmentation maps produced by Radio U-Net with the position of galaxy clusters contained in the merged catalogue created in Sect. 2. Therefore, as explained in Sect. 3.4, for each of the 3822 galaxy clusters in our merged catalogue, we found the reference pointing and, for those in quality class 0 and 1 pointings, we computed the \mathcal{R} value from the segmentation map. The merged catalogue thus provides the following information:

- Name of the cluster;
- Coordinates in equatorial (J2000) coordinate systems;
- Redshift;

⁸ https://lofar-surveys.org/clusters_dr3.html

⁹ <https://aladin.cds.unistra.fr/hips/>

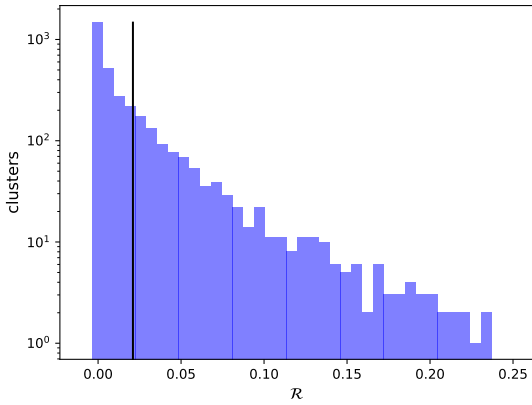


Fig. 7. Distribution of the \mathcal{R} values computed for the 3367 galaxy clusters in the merged catalogue outside quality 2 regions. The vertical line represents the value $\mathcal{R}_t^* = 0.021$.

- M_{500} and R_{500} ;
- Alternative name derived from the cross-match between the four original catalogues;
- Name of the mosaic at the position of the cluster in the DR3 mosaicking scheme;
- The names of all pointings used to create the mosaic at the position of the cluster (i.e. names of mosaics in the DR2 mosaicking scheme);
- The distance of the cluster from the pointing centre of each pointing;
- Noise at the cluster location in each pointing before mosaicking;
- Quality class of each mosaic in the DR2 mosaicking scheme;
- Comments on the pointing used for the computation of \mathcal{R} ;
- Name of the mosaic used to compute \mathcal{R} . For quality class 2 pointings, this is not listed since the \mathcal{R} value is not computed;
- The value of \mathcal{R} and its uncertainty, given on the basis of the cluster redshift as explained in Appendix B;
- List of galaxy clusters which overlap with the system within their $2.2R_{500}$.

The table is available on a dedicated page of the survey website¹⁰. The information provided in this table can be used to build a sub-sample of galaxy clusters for further studies (an example is provided in the following section).

It is also useful to derive global statistics for this large sample of galaxy clusters. The distribution of \mathcal{R} values across the merged sample –restricted to the 3367 clusters located in regions where the segmentation maps are classified as quality 0 or 1– is shown in Fig. 7. Using a threshold of $\mathcal{R} > 0.021$, we identify 918 galaxy clusters as candidates hosting diffuse radio emission, or DE. In this case, we do not exclude clusters based on the uncertainty of their \mathcal{R} value, in order to retain the largest possible sample size, at the expense of a lower classification accuracy.

However, this number must be corrected for the classification performance of our detection pipeline. As discussed in Sect. 3.4, the threshold $\mathcal{R}_t^* = 0.021$ corresponds to a precision of 68% and a recall of 74% when tested on the LoTSS-DR2/PSZ2 sample of 246 galaxy clusters. Under the assumption that these performance metrics are applicable to the LoTSS-DR3 data, we estimated the intrinsic fraction of galaxy clusters hosting diffuse radio emission, f_{DE} , as

Table 2. Occurrence of galaxy clusters and detection fraction of the four galaxy cluster catalogues.

Catalogue	$N_{\text{LoTSS-DR3}}$	N_0	N_1	N_2	$N_{\mathcal{R}>0.021}$	f_{DE}
PSZ2	686	459	167	60	344	50%
ACT-DR5	1699	696	684	319	327	22%
MCXC2	874	611	188	75	283	32%
1eRASS	1324	867	405	52	286	21%

Notes. Column 1: Catalogue name. Column 2: Number of galaxy clusters within the LoTSS-DR3 area with $M_{500} > 10^{14} M_{\odot}$. Column 3: Number of galaxy clusters observed in quality class 0 mosaics. Column 4: Number of galaxy clusters observed in quality class 1 mosaics. Column 5: Number of galaxy clusters observed in quality class 2 mosaics and therefore excluded from the analysis. Column 6: Number of galaxy clusters with \mathcal{R} value larger than 0.021. Column 7: Intrinsic fraction of clusters with diffuse radio emission over the total number of clusters analysed for each catalogue, computed following Eq. (3).

$$f_{\text{DE}} = \frac{\text{Precision}(\mathcal{R}_t^* = 0.021)}{\text{Recall}(\mathcal{R}_t^* = 0.021)} \frac{N_{\mathcal{R}>0.021}}{N_0 + N_1} \simeq 0.92 \frac{N_{\mathcal{R}>0.021}}{N_0 + N_1}, \quad (3)$$

which yields $f_{\text{DE}} \simeq 25\%$ for the merged catalogue. This estimate should be treated with caution, as both the adopted threshold and the corresponding precision and recall are derived from a test sample based on LoTSS-DR2/PSZ2 galaxy clusters. Given that the associated uncertainties are dominated by systematic effects related to sample selection and transferability of the classifier performance, we do not attempt to provide a formal error estimate for f_{DE} .

In order to derive statistical information on the detection rate within each catalogue, it is convenient to associate each cluster with its original galaxy cluster catalogue. This association is performed using the alternative names listed in the merged catalogue. Table 2 reports the detection rate for each of the four galaxy cluster catalogues. The total number of clusters exceeds that of the merged catalogue, as several clusters appear in more than one catalogue.

The intrinsic fraction of galaxy clusters hosting diffuse radio emission is highest in the PSZ2 catalogue (50%). This is expected, as this catalogue preferentially includes more massive and lower-redshift systems compared to the other three catalogues, as shown in Fig. 1. Both theoretical models for the origin of radio halos (Cassano et al. 2007) and cosmological simulations investigating the formation of radio relics (Lee et al. 2024) predict that the fraction of galaxy clusters hosting diffuse radio sources increases with increasing cluster mass and decreasing redshift, and this is confirmed by observational results in the case of radio halos (Cassano et al. 2023).

We test this expectation further by dividing each cluster catalogue into mass and redshift bins and computing the intrinsic fraction of galaxy clusters hosting diffuse radio emission in each bin. The results for the PSZ2 catalogue are shown in Fig. 8, while the corresponding plots for the other three catalogues are presented in Fig. A.2. Apart from statistical fluctuations in bins containing a small number of clusters (with the total number of clusters in each bin shown in brackets), we recover a clear trend of increasing detection fraction towards higher masses and lower redshifts for all catalogues.

Botteon et al. (2022) found that, in the LoTSS-DR2 sample, the detection rate of galaxy clusters hosting a radio halo is $30 \pm 11\%$, while it is $10 \pm 6\%$ for clusters hosting one or more

¹⁰ https://lofar-surveys.org/clusters_dr3.html

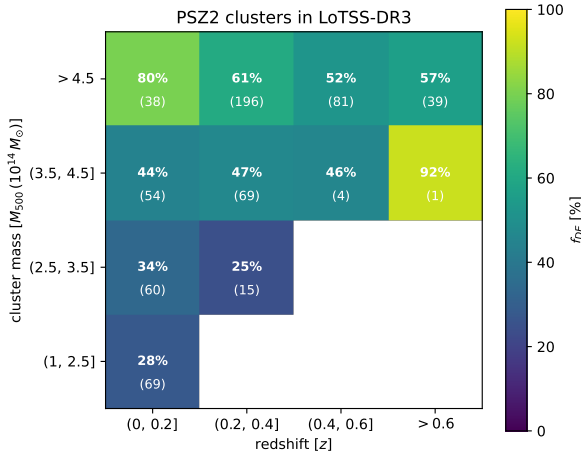


Fig. 8. Intrinsic fraction of clusters with diffuse radio emission over the total number of clusters in each mass and redshift bin in the PSZ2 cluster catalogue computed following Eq. (3). The total number of clusters in each bin (N_0+N_1) is also shown in brackets.

radio relics. They also reported that $51 \pm 14\%$ of the clusters do not show evidence of diffuse synchrotron emission in the ICM. This is closely consistent with the intrinsic detection fraction of 50% that we have obtained for the PSZ2 sample, where radio halos, relics and other diffuse radio sources of uncertain classification are accounted for. Furthermore, they predicted the detection of 251 ± 92 galaxy clusters hosting a radio halo and 83 ± 50 clusters hosting at least one relic from the PSZ2 catalogue in the full LoTSS survey. Following our approach we have found 344 galaxy clusters hosting diffuse radio emission in the LoTSS-DR3/PSZ2 sample. Although a direct comparison among these numbers should be treated with caution, the overall fraction of detected systems is broadly consistent with these predictions.

4.2. Curated sub-sample of galaxy clusters detected by the network

Finally, we have used the merged catalogue to select a sub-sample of galaxy clusters having the highest classification accuracy provided by our method. We have selected all clusters in the merged catalogue with (i) quality class 0 (ii) $\mathcal{R} > 0.021$, and (iii) a reliable classification flag with respect to this threshold, meaning that $\mathcal{R} - 0.021 > \sigma$, where σ is the uncertainty on the \mathcal{R} value discussed in Appendix B. This set contains 357 galaxy clusters. The table and the images of this sub-sample are available on a dedicated page of the survey website¹¹.

For this sub-sample, we performed a visual inspection of the LoTSS-DR3 images at both high and low angular resolution, aimed at identifying clear false detections through a comparison with the segmentation maps produced by Radio U-Net. Clear false detections were identified when the regions segmented by the network overlapped with extended AGN emission, nearby galaxies, or other foreground or background galaxy clusters projected along the line of sight. We found a total of 19 clear false detections.

We subsequently searched the literature to identify previous studies of each cluster. In total, we identified 109 galaxy clusters with literature confirmation of the presence of diffuse radio emission, in the form of radio halos, mini-halos, radio relics,

phenixes, remnant radio sources, or diffuse radio sources with uncertain classification.

Out of the 357 clusters, 104 clusters belong to the PSZ2 catalogue and lie within the LoTSS-DR2 area already studied by Botteon et al. (2022). However, 31 of these systems were not classified as hosts of diffuse radio emission in this previous work, either because of the low quality of the available data (9 clusters) or because the detected emission was considered by Botteon et al. (2022) to be unrelated to the ICM (22 clusters, but in three of these cases the emission was later reclassified as a remnant radio source by Bruno et al. 2025). Seven clusters belong instead to the sample of non-PSZ2 systems found to host diffuse radio emission by Hoang et al. (2022). Interestingly, both inter-cluster radio bridges discovered with LOFAR – the Abell 399–401 system (Govoni et al. 2019) and Abell 1758 (Botteon et al. 2020) – are included in our curated sub-sample.

Among the remaining 229 galaxy clusters lacking previous studies in the literature, several show radio morphologies that can be readily associated with standard classes of diffuse radio emission, such as radio halos or radio relics. Examples of these cases are shown in Fig. C.2. Follow-up studies are already ongoing on a few of these systems, such as the high-redshift galaxy cluster PSZ2 G209.79+10.23 ($z=0.677$), possibly hosting a radio halo and double radio relics (Di Gennaro et al., in prep.), or the possible double radio relic system in ACT-CL J0245.1+1843 (Pfeifer et al., in prep.). Another peculiar case is represented by three galaxy clusters in the curated sub-sample that belong to a single interacting system around PSZ2 G343.33+83.19, in which the network detected diffuse radio emission connecting the three clusters (see Fig. C.1). For most of the remaining candidate new diffuse radio sources, dedicated source extraction and follow-up analyses of individual objects are required to confirm the association of the diffuse radio emission with the ICM and to determine its physical nature.

In summary, the curated sub-sample comprises 357 galaxy clusters, of which 19 are clear false detections and 109 have been previously confirmed to host some form of diffuse radio emission in the literature. The remaining 229 clusters will be the subject of dedicated post-processing and follow-up investigations.

We note that this sub-sample does not represent a complete census of all galaxy clusters with diffuse radio emission identified by the network, but it was selected in order to ensure the highest accuracy achievable by the network (76%), according to the performance metrics discussed in Sect. 3.4. Some clusters may have been excluded either because they are located in regions classified as quality class 1 or 2, or because their \mathcal{R} value lies close to the adopted threshold. This is, for example, the case for PSZ2 G135.76–62.03 (Abell 0168; Hoang et al., in prep.), PSZ2 G075.71+13.51 (Abell 2319; Drabant et al., in prep.), and PSZ2 G200.82+27.42 and ACT-CL J0220.9+0652 (Pfeifer et al., in prep.), which are detected with high significance but fall within quality class 1 regions, and for which dedicated follow-up studies are already in progress.

5. Discussion

In this section, we discuss the results obtained in this paper. In particular, we focus on the use of the catalogues produced in this work and on possible future improvements of the pipeline.

5.1. Towards statistical studies of galaxy clusters

This work is meant to provide automatic diffuse radio source detection in the largest possible sample of galaxy clusters and

¹¹ https://lofar-surveys.org/clusters_dr3.html

to guide blind detection of new diffuse radio sources. These kinds of automated methodologies are becoming increasingly important with new-generation wide-area surveys, which will allow statistical studies based on hundreds of sources. With our method, we obtained a trend of an increasing detection rate of diffuse radio sources with increasing cluster mass and decreasing redshift (Sect. 4.1), which is particularly encouraging, as it reproduces behaviours observed in statistical samples (Cassano et al. 2023). However, we recognise that the creation of a statistically meaningful sub-sample of galaxy clusters (such as those based on the LoTSS-DR2/PSZ2 sample; e.g. Bruno et al. 2023; Cassano et al. 2023; Cuciti et al. 2023; Jones et al. 2023), requires additional efforts. This includes, for example, cleaning the present catalogues from spurious detections, comparing detections with multi-wavelength datasets, and carefully measuring the flux of diffuse radio emission.

In order to create a subsample of galaxy clusters for scientific purposes, a decision should be made on the \mathcal{R} value threshold to be used as a parameter for detection: higher \mathcal{R} values will lead to purer samples, with a smaller fraction of DE objects but detected with higher significance, while lower \mathcal{R} values will lead to more complete sample but with a higher fraction of false detection. In Sect. 3.4 we selected a sub-sample with the highest possible accuracy which is 76%. In this curated sub-sample, the precision is 74%, as derived from the test carried out on the LoTSS-DR2/PSZ2 sample, meaning that the 74% of clusters classified as DE by the network are true detections. While the implementation of more sophisticated automated pipelines is desirable, this is the best achievable compromise between automation and accuracy at present.

After this selection, in many cases, a detection can still be confirmed only by the comparison with multi-wavelength datasets, for example, to discriminate between diffuse radio emission connected to AGN and related to the ICM, for which the comparison with optical datasets is necessary. This is even more important for classification purposes, for example, regarding the position of the diffuse radio emission with respect to the X-ray emission of the cluster to discriminate between relics and halos. Additionally, radio spectral index information should be used to discriminate between different acceleration mechanisms at work in different sources. This will be available in a wide area thanks, for example, to LOFAR observation in the low (54 MHz) and high (144 MHz) bands. While the classification of sources could remain uncertain even after these comparisons, these are clearly important steps to be taken for subsequent studies.

Furthermore, as mentioned in Sect. 3.1, the automated calibration and imaging strategy of LoTSS is not optimised for the diffuse radio emission of galaxy clusters. For the purpose of this work – which is also minimising re-processing – we have directly used LoTSS images at 20'' as input for the pipeline, where diffuse radio emission could be incompletely deconvolved or affected by the presence of residual calibration and imaging artefacts. When a sub-sample is selected, dedicated pipelines should be used for source extraction and self-calibration of each target (van Weeren et al. 2021; Botteon et al. 2022), which allows a more robust characterisation of their morphology and measurement of their fluxes. However, as already noted in Stuardi et al. (2024) and further explored in Stuardi et al. (2025), in some cases the network is able to detect diffuse radio emission that is not detected in extracted images, probably because they are at the detection limit, but emerges in deeper observations or lower-resolution images. Hence, putting together machine learning information and tailored post-processing of selected candidates is still essential.

5.2. Small angular size systems

The accuracy metrics reported for the current catalogues are all based on the LoTSS-DR2/PSZ2 sample, which contains more low-redshift and high-mass systems with respect to other galaxy cluster catalogues (see Appendix A, Fig. A.1). As explored in Appendix B, the variability of the \mathcal{R} value increases for small angular size systems, i.e. high-redshift and low-mass clusters, and consequently, the accuracy metrics could vary in different samples. Hence, the computed \mathcal{R} value for high-redshift and low-mass systems should be taken with caution.

In particular, the detection of diffuse radio sources at high redshifts is more difficult because of redshift dimming and increasing inverse Compton energy losses, which make the emission fainter and harder to separate from contaminating sources. Moreover, the majority of high-redshift clusters have fairly low masses ($M_{500} \lesssim 5 \times 10^{14} M_{\odot}$), where the presence of diffuse radio emission is less likely (Di Gennaro et al. 2021; Cassano et al. 2023; Di Gennaro et al. 2025). In many cases, only increasing the angular resolution of LOFAR observations would help firmly establish the presence of the diffuse radio emission detected by Radio U-Net. Increasing the angular resolution to the sub-/arcsecond regime is now possible with LOFAR's international baselines. This has been helpful in confirming the presence of diffuse radio emission in clusters (HLavacek-Larrondo et al. 2025; Di Gennaro et al. 2026). In the future, dedicated studies of a large sample of high-redshift clusters may help better train the network in the high-redshift regime.

5.3. Machine learning methodology improvements

Strategies to enhance the performance of DL methods applied to the detection of diffuse and extended radio sources are currently being explored by several of the authors. A first approach focuses on improving the network architecture itself, in particular by incorporating attention mechanisms. In this context, attention layers allow the network to dynamically weight different regions of the input image, enabling it to focus on spatial features that are more relevant for the task at hand while suppressing less informative or contaminating structures. The effectiveness of this approach has already been demonstrated by the TUNA network (Sanvitale et al. 2025). Additional improvements may be achieved by fine-tuning the models on real observational data, a procedure already explored in Stuardi et al. (2024) and currently being further tested on TUNA.

A second strategy targets improvements to the training data, aiming at producing more realistic and comprehensive mock observations. This includes incorporating radio galaxies either by extracting them directly from real observations or by using new generations of cosmological simulations in which AGN activity and star formation are modelled self-consistently. Furthermore, training data can be enhanced by explicitly accounting for Galactic foreground emission and other large-scale diffuse components, which may otherwise be misclassified as cluster-related emission. This would improve the generalisation capabilities of the network.

An additional research direction under investigation involves a multi-wavelength approach, in which radio, X-ray, and optical observations are jointly used as network inputs. This strategy has already proven effective in the context of galaxy cluster identification and classification in X-ray surveys, where machine learning techniques have been successfully employed to distinguish between relaxed and disturbed systems or to identify galaxy clusters (Kosiba et al. 2020; Sadikov et al. 2025). Extending this

approach to radio studies may help the network better isolate emission physically associated with the ICM.

Beyond supervised learning, self-supervised and semi-supervised methods are also being explored as a way to mitigate the limited availability of labelled data. Recent studies (e.g. Mostert et al. 2021, 2023; Slijepcevic et al. 2024; Baron Perez et al. 2025) have shown that such approaches can significantly improve performance and robustness, particularly for rare and complex radio sources. Self-supervised methodologies will be even more important in the SKA era, when new sources with previously unknown morphologies could be discovered.

6. Conclusions

We have presented the first catalogue of diffuse radio emission in galaxy clusters extracted from LoTSS-DR3 based on DL. By combining the unprecedented sky coverage of LoTSS with the Radio U-Net convolutional neural network, we have developed an automated pipeline capable of identifying extended low-surface-brightness emission related to the ICM.

Our analysis processed 2551 LoTSS-DR3 mosaics and cross-matched the resulting segmentation maps with 3822 clusters from four complementary catalogues (PSZ2, ACT-DR5, MCXC2, and 1eRASS). The method provides both a quantitative indicator for the presence of diffuse radio emission in each cluster (R value) and a pixel-level segmentation map of the entire survey. The detection fraction of diffuse radio sources in the four catalogues increases in high-mass and low-redshift systems, as predicted by current models and simulations.

From the same catalogue, we also curated a sub-sample of galaxy clusters with a reliable classification flag containing 357 galaxy clusters showing significant evidence of diffuse radio emission. Dedicated pipelines should be used for source extraction and self-calibration of each new target in this sub-sample (van Weeren et al. 2021; Botteon et al. 2022), to allow for a more robust characterisation of their morphology and measurement of their fluxes.

Although our approach does not yet distinguish between different types of diffuse radio emission (e.g. radio halos, relics, or phoenixes and remnant plasma sources), it establishes a framework for automatic research of diffuse radio emission sources in wide-area radio surveys. The same procedure can also be applied to different radio surveys by tailoring the training dataset to the specific telescope.

Limitations regarding our method remain, primarily related to the network’s ability to generalise in the presence of Galactic diffuse radio emission, imaging artefacts, and diffuse radio emission associated with extended AGN, which can lead to false detections. In future developments, we will focus on retraining the Radio U-Net using more realistic simulations and an expanded set of real observational data.

The catalogue released with this paper provides a valuable resource for the community, enabling follow-up studies of individual systems, statistical analyses, and cross-correlation with multi-wavelength datasets. The upcoming SKA and LOFAR2.0 surveys will dramatically expand the available data volume, and methodologies such as the one presented in this work will be essential for exploiting their full scientific potential.

Data availability

The processed data products and catalogues generated in this work are publicly available on a dedicated page of the

LOFAR Surveys website at https://lofar-surveys.org/clusters_dr3.html. These include three FITS tables with information on: (i) the full sample of 3822 galaxy clusters in the merged catalogue, (ii) the 357 clusters in the curated sub-sample, and (iii) all LoTSS-DR3 mosaics. A HiPS file showing the segmentation map for the entire LoTSS-DR3 (excluding quality 2 mosaics) is also available at the same location, while segmentation maps of individual mosaics are provided upon reasonable request. The catalogues are available at the CDS via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/710/A219> and on Zenodo <https://zenodo.org/records/20022649>.

The LoTSS-DR3 data underlying this article are publicly available from the LOFAR Surveys webpage at (https://lofar-surveys.org/dr3_release.html). The Radio U-Net code used in this work is available at <https://github.com/ICSC-Spoke3/Radio-U-Net>. The simulations used to train the network are available at <https://owncloud.ia2.inaf.it/index.php/s/IbFPlCCpUresrr> and are described by Stuardi et al. (2024).

Acknowledgements. This paper is supported by the Fondazione ICSC, Spoke 3 Astrophysics and Cosmos Observations – National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) Project ID CN_00000013 “Italian Research Center for High-Performance Computing, Big Data and Quantum Computing” funded by MUR Missione 4 Componente 2 Investimento 1.4: Potenziamento strutture di ricerca e creazione di “campioni nazionali di R&S (M4C2-19)” – Next Generation EU (NGEU), and it’s also supported by (Programma Operativo Nazionale, PON), “Tematiche di Ricerca Green e dell’Innovazione”. The results of this work have been produced on the Leonardo Supercomputer at CINECA (Bologna, Italy) in the framework of the IS CRA project IsC2 COS4ML (P.I. C.Stuardi) and also supported by the INAF Mini Grant (P.I. C.Stuardi) awarded by the “Bando Ricerca Astrofisica Fondamentale 2024”. We also acknowledge the usage of online storage tools kindly provided by the INAF Astronomical Archive (IA2) initiative (<http://www.ia2.inaf.it>). LOFAR (van Haarlem et al. 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the LOFAR ERIC under a joint scientific policy. The LOFAR resources have benefited from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Université d’Orléans, France; BMFT, MKW-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; The Science and Technology Facilities Council, UK; Ministry of Science and Higher Education, Poland; The Istituto Nazionale di Astrofisica (INAF), Italy. This research made use of the Dutch national e-infrastructure with support of the SURF Cooperative (e-infra 180169) and the LOFAR e-infra group. The Jülich LOFAR Long Term Archive and the German LOFAR network are both coordinated and operated by the Jülich Supercomputing Centre (JSC), and computing resources on the supercomputer JUWELS at JSC were provided by the Gauss Centre for Supercomputing e.V. (grant CHTB00) through the John von Neumann Institute for Computing (NIC). This research made use of the University of Hertfordshire high-performance computing facility and the LOFAR-UK computing facility located at the University of Hertfordshire and supported by STFC [ST/P000096/1], and of the Italian LOFAR-IT computing infrastructure supported and operated by INAF, including the resources within the PLEIADI special “LOFAR” project by USC-C of INAF, and by the Physics Department of Turin university (under an agreement with Consorzio Interuniversitario per la Fisica Spaziale) at the C3S Supercomputing Centre, Italy. This research is part of the project LOFAR Data Valorization (LDV) [project numbers 2020.031, 2022.033, and 2024.047] of the research programme Computing Time on National Computer Facilities using SPIDER that is (co-)funded by the Dutch Research Council (NWO), hosted by SURF through the call for proposals of Computing Time on National Computer Facilities. A.Bon. and M.B. acknowledge support from the ERC CoG BELOVED, N.101169773. F.dG., G.D.G., N.B. and M.C. acknowledge support from the ERC CoG ULU, N. 101086378. M.J.H. thanks the UK STFC for support [ST/V000624/1, ST/Y001249/1]. F.G. acknowledges the financial contribution from the INAF GO grant 1.05.24.02.10 Extended Radio Emission in Galaxy Clusters at deep focus with MeerKAT. F.V. acknowledges funding under the European Union’s Horizon Europe program through the ERC Synergy Grant COSMOMAG (Project Id. 101224803).

References

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, paper and Software available from <https://www.tensorflow.org/>
- Agarap, A. F. 2018, arXiv e-prints [arXiv:1803.08375]
- Aguado-Barahona, A., Barrena, R., Streblyanska, A., et al. 2019, *A&A*, **631**, A148
- Bahk, H., & Hwang, H. S. 2024, *ApJS*, **272**, 7
- Baron Perez, N., Brügger, M., Kasieczka, G., & Lucie-Smith, L. 2025, *A&A*, **699**, A302
- Barrena, R., Streblyanska, A., Ferragamo, A., et al. 2018, *A&A*, **616**, A42
- Beck, R., Dobos, L., Budavári, T., Szalay, A. S., & Csabai, I. 2016, *MNRAS*, **460**, 1371
- Boada, S., Hughes, J. P., Menanteau, F., et al. 2019, *ApJ*, **871**, 188
- Böhringer, H., Chon, G., Collins, C. A., et al. 2013, *A&A*, **555**, A30
- Botteon, A., van Weeren, R. J., Brunetti, G., et al. 2020, *MNRAS*, **499**, L11
- Botteon, A., Shimwell, T. W., Cassano, R., et al. 2022, *A&A*, **660**, A78
- Brügger, M., Bykov, A., Ryu, D., & Röttgering, H. 2012, *Space Sci. Rev.*, **166**, 187
- Brunetti, G. 2011, *JApA*, **32**, 437
- Brunetti, G., & Jones, T. W. 2014, *Int. J. Mod. Phys. D*, **23**, 1430007
- Brunetti, G., & Vazza, F. 2020, *Phys. Rev. Lett.*, **124**, 051101
- Bruno, L., Brunetti, G., Botteon, A., et al. 2023, *A&A*, **672**, A41
- Bruno, L., Botteon, A., Dallacasa, D., et al. 2025, *A&A*, **704**, A245
- Buddendiek, A., Schrabback, T., Greer, C. H., et al. 2015, *MNRAS*, **450**, 4248
- Bulbul, E., Liu, A., Kluge, M., et al. 2024, *A&A*, **685**, A106
- Burenin, R. A. 2017, *Astron. Lett.*, **43**, 507
- Burenin, R. A., Bikmaev, I. F., Khamitov, I. M., et al. 2018, *Astron. Lett.*, **44**, 297
- Cassano, R., Brunetti, G., Setti, G., Govoni, F., & Dolag, K. 2007, *MNRAS*, **378**, 1565
- Cassano, R., Bernardi, G., Brunetti, G., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 73
- Cassano, R., Cuciti, V., Brunetti, G., et al. 2023, *A&A*, **672**, A43
- Chollet, F., et al. 2015, Keras, <https://keras.io>
- Cornu, D., Salomé, P., Semelin, B., et al. 2024, *A&A*, **690**, A211
- Csurka, G., Volpi, R., & Chidlovskii, B. 2022, *Found. Trends Comput. Graphics Vision*, **14**, 1
- Cuciti, V., Cassano, R., Brunetti, G., et al. 2015, *A&A*, **580**, A97
- Cuciti, V., Cassano, R., Sereno, M., et al. 2023, *A&A*, **680**, A30
- Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, *IEEE Proc.*, **97**, 1482
- Di Gennaro, G., van Weeren, R. J., Brunetti, G., et al. 2021, *Nat. Astron.*, **5**, 268
- Di Gennaro, G., Brügger, M., Moravec, E., et al. 2025, *A&A*, **695**, A215
- Di Gennaro, G., Timmerman, R., Hoeft, M., et al. 2026, *A&A*, **709**, A202
- Duchesne, S. W., Botteon, A., Koribalski, B. S., et al. 2024, *PASA*, **41**
- Etsebeth, V., Lochner, M., Kolokythas, K., Knowles, K., & Tolley, E. 2026, *RAS Techn. Instrum.*, submitted, [arXiv:2602.15930]
- Feretti, L., Giovannini, G., Govoni, F., & Murgia, M. 2012, *A&A Rev.*, **20**, 54
- Fernique, P., Allen, M. G., Boch, T., et al. 2015, *A&A*, **578**, A114
- Gheller, C., & Vazza, F. 2022, *MNRAS*, **509**, 990
- Gheller, C., Vazza, F., & Bonafede, A. 2018, *MNRAS*, **480**, 3749
- Giovannini, G., Cau, M., Bonafede, A., et al. 2020, *A&A*, **640**, A108
- Goodfellow, I., Bengio, Y., & Courville, A. 2016, *Deep Learning* (Cambridge, MA: MIT Press)
- Govoni, F., & Feretti, L. 2004, *Int. J. Mod. Phys. D*, **13**, 1549
- Govoni, F., Orrù, E., Bonafede, A., et al. 2019, *Science*, **364**, 981
- Gupta, N., Hayder, Z., Norris, R. P., Huynh, M., & Petersson, L. 2024, *PASA*, **41**
- Hilton, M., Sifón, C., Naess, S., et al. 2021, *ApJS*, **253**, 3
- HLavacek-Larrondo, J., Timmerman, R., Pfrommer, C., et al. 2025, *ApJ*, **987**, L40
- Hoang, D. N., Brügger, M., Botteon, A., et al. 2022, *A&A*, **665**, A60
- Hotan, A. W., Bunton, J. D., Chippendale, A. P., et al. 2021, *PASA*, **38**
- Jonas, J., & MeerKAT Team, 2016, *MeerKAT Science: On the Pathway to the SKA*, 1
- Jones, A., de Gasperin, F., Cuciti, V., et al. 2023, *A&A*, **680**, A31
- Knowles, K., Cotton, W. D., Rudnick, L., et al. 2022, *A&A*, **657**, A56
- Kolokythas, K., Venturi, T., Knowles, K., et al. 2025, *MNRAS*, **543**, 1638
- Kosiba, M., Lieu, M., Altieri, B., et al. 2020, *MNRAS*, **496**, 4141
- Lao, B., Andernach, H., Yang, X., et al. 2025, *ApJS*, **276**, 46
- Lee, W., Pillepich, A., ZuHone, J., et al. 2024, *A&A*, **686**, A55
- Lukic, V., de Gasperin, F., & Brügger, M. 2019, *Galaxies*, **8**, 3
- Mann, A. W., & Ebeling, H. 2012, *MNRAS*, **420**, 2120
- Mishra, A. K., Tolley, E., Krishna, S. P., & Kneib, J.-P. 2025, *MNRAS*, **538**, 2905
- Mostert, R. I. J., Duncan, K. J., Röttgering, H. J. A., et al. 2021, *A&A*, **645**, A89
- Mostert, R. I. J., Morganti, R., Brienza, M., et al. 2023, *A&A*, **674**, A208
- Nuza, S. E., Hoeft, M., van Weeren, R. J., Gottlöber, S., & Yepes, G. 2012, *MNRAS*, **420**, 2006
- Piffaretti, R., Arnaud, M., Pratt, G. W., Pointecouteau, E., & Melin, J. B. 2011, *A&A*, **534**, A109
- Planck Collaboration XXVII. 2016, *A&A*, **594**, A27
- Repp, A., & Ebeling, H. 2018, *MNRAS*, **479**, 844
- Riggi, S., Magro, D., Sortino, R., et al. 2023, *Astron. Comput.*, **42**, 100682
- Riggi, S., Ceconello, T., Palazzo, S., et al. 2024, *PASA*, **41**
- Ronneberger, O., Fischer, P., & Brox, T. 2015, in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015* (Springer), Lect. Notes Comput. Sci., 9351, 234
- Sadibekova, T., Arnaud, M., Pratt, G. W., Tarrío, P., & Melin, J. B. 2024, *A&A*, **688**, A187
- Sadikov, M., Hlavacek-Larrondo, J., Perreault-Levasseur, L., et al. 2025, *ApJ*, **986**, 24
- Sanvitale, N., Gheller, C., Vazza, F., et al. 2025, *MNRAS*, **541**, 3479
- Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, *A&A*, **598**, A104
- Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, *A&A*, **622**, A1
- Shimwell, T. W., Hardcastle, M. J., Tasse, C., et al. 2022, *A&A*, **659**, A1
- Shimwell, T. W., Hardcastle, M. J., Tasse, C., et al. 2026, *A&A*, **707**, A198
- Slijepcevic, I. V., Scaife, A. M. M., Walmsley, M., et al. 2024, *RAS Tech. Instrum.*, **3**, 19
- Sortino, R., Magro, D., Fiameni, G., et al. 2023, *Exp. Astron.*, **56**, 293
- Streblyanska, A., Barrena, R., Rubiño-Martín, J. A., et al. 2018, *A&A*, **617**, A71
- Streblyanska, A., Aguado-Barahona, A., Ferragamo, A., et al. 2019, *A&A*, **628**, A13
- Stuardi, C., Gheller, C., Vazza, F., & Botteon, A. 2024, *MNRAS*, **533**, 3194
- Stuardi, C., Botteon, A., Sereno, M., et al. 2025, *A&A*, **695**, L16
- Tieleman, T., & Hinton, G. 2012, Lecture 6.5 - RMSProp: Divide the Gradient by a Running Average of its Recent Magnitude, Coursera: Neural Networks for Machine Learning, lecture notes
- Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, *PASA*, **30**
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, **556**, A2
- van Weeren, R. J., de Gasperin, F., Akamatsu, H., et al. 2019, *Space Sci. Rev.*, **215**, 16
- van Weeren, R. J., Shimwell, T. W., Botteon, A., et al. 2021, *A&A*, **651**, A115
- Vazza, F., Etori, S., Roncarelli, M., et al. 2019, *A&A*, **627**, A5
- Wen, Z. L., & Han, J. L. 2024, *ApJS*, **272**, 39
- Xu, W., Ramos-Ceja, M. E., Pacaud, F., Reiprich, T. H., & Erben, T. 2022, *A&A*, **658**, A59
- Zheng, H., Tegmark, M., Dillon, J. S., et al. 2017, *MNRAS*, **464**, 3486
- Zohren, H., Schrabback, T., van der Burg, R. F. J., et al. 2019, *MNRAS*, **488**, 2523

Appendix A: Single cluster catalogues and detection fraction

In this Appendix, we show the mass-redshift diagram of the single cluster catalogues used to generate the merged catalogue in Sec. 2 (Fig. A.1). We also show the intrinsic fraction of clusters detected with diffuse radio emission in the ACT-DR5, MCXC2 and 1eRASS catalogues, separated in mass and redshift bins as explained in Sec. 4.1 (Fig. A.2).

Appendix B: Uncertainty on the \mathcal{R} value

In this work, we slightly modified the detection pipeline compared to Stuardi et al. (2024) by applying Radio U-Net to the entire LoTSS-DR3 mosaics, rather than to individual cutouts centred on each cluster. The \mathcal{R} values computed for the 246 LoTSS-DR2/PSZ2 galaxy clusters used as a test sample by Stuardi et al. (2024) can therefore serve as a benchmark for the updated pipeline. Comparing the \mathcal{R} values obtained with the two methods also allowed us to verify the robustness of the \mathcal{R} parameter and to derive an appropriate statistical uncertainty.

In Fig. B.1, we plot the \mathcal{R} values calculated in this work against those from Stuardi et al. (2024), denoted as \mathcal{R}_{S24} , for the 246 clusters classified as DE or NDE by Botteon et al. (2022). As expected, the two measurements follow a clear linear trend, and most clusters remain in the same class (identified using the threshold $\mathcal{R}_{S24} = 0.015$ here). However, in some cases, significant differences in \mathcal{R} are evident.

We find that these discrepancies are statistically more pronounced for cluster with small projected angular size (see Fig. B.2). The main cause of the discrepancy is the inference uncertainty associated with the model. Indeed, Radio U-Net is used here as a deterministic model. However, when it is applied for inference on images with different tiling schemes and varying noise properties, it will produce variability in the predictions for the same sources. This effect is expected to be more severe at high redshift where the diffuse radio emission is intrinsically fainter and difficult to detect. Small changes in the input can lead to large differences in the network's prediction and, since the \mathcal{R} parameter is, by definition, derived directly from the extent of the diffuse radio emission, and this emission at higher redshift is expected to span only a small number of pixels, a few FP or FN detections or shifts in the predicted values can produce significant differences in \mathcal{R} .

Another aspect to consider is the absence of high-redshift sources in the simulation set used to train the network, where diffuse radio emission was projected into mock lightcones up to $z = 0.15$ (Vazza et al. 2019). This could reduce the inference performances of the network for $z > 0.15$ (see also Fig. B.3). Another contributing factor is the choice of tile size for the segmentation: the tiles are designed to contain only a small fraction of the diffuse radio emission at $z < 0.15$. When the radius used to compute \mathcal{R} is smaller than the tile size, the \mathcal{R} value becomes sensitive to the relative position of the tile and the diffuse radio emission. Stuardi et al. (2024) performed segmentation on images centred on the cluster positions, whereas in this work, the segmentation was carried out on entire pointings, meaning the tiles were randomly centred with respect to the clusters. This effect is illustrated in Fig. B.2, which also shows that cluster size and redshift are correlated, suggesting that both the under-representation of high- z systems in the training set and the choice of tile size could contribute to increased uncertainty in the \mathcal{R} values for small angular size systems.

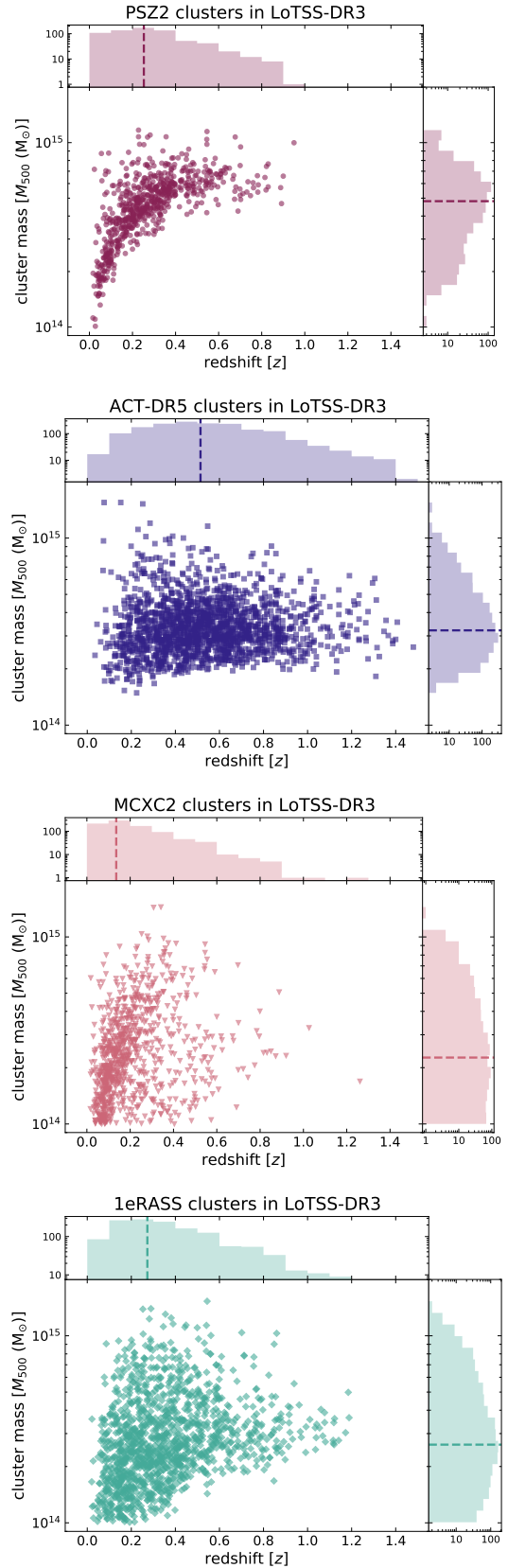


Fig. A.1. Mass versus redshift plots of the single selected cluster catalogues in LoTSS-DR3. Dashed lines in the horizontal and vertical histograms show the median values of the redshift and mass, respectively.

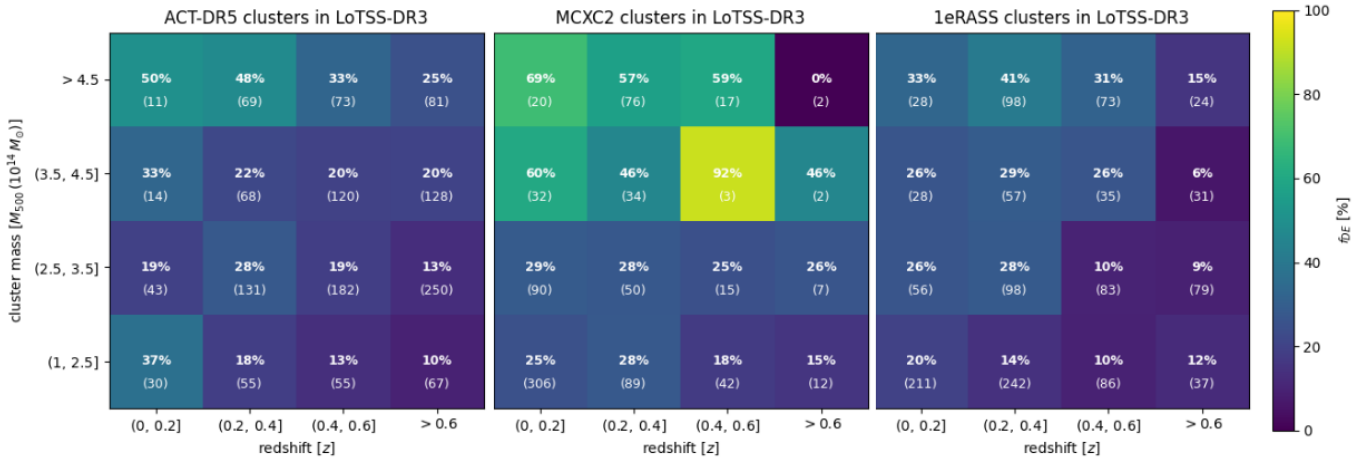


Fig. A.2. Intrinsic fraction of clusters with diffuse radio emission in the ACT-DR5, MCXC2 and 1eRASS catalogues computed following Eq. 3 and plotted for different mass and redshift bins. The total number of clusters in each bin (in quality 0 and quality 1 mosaics) is also shown in brackets.

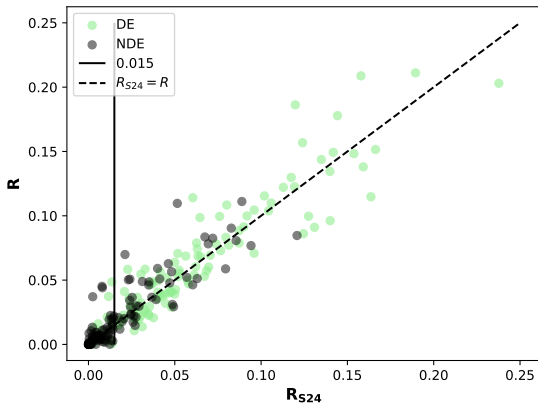


Fig. B.1. Comparison of the \mathcal{R} value computed by Stuardi et al. (2024) with the one computed in this work for the test sample of LoTSS-DR2/PSZ2 clusters. The vertical line marks the threshold used for classification by Stuardi et al. (2024).

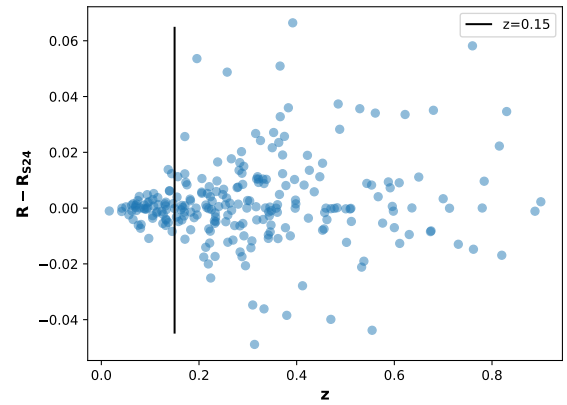


Fig. B.3. Difference between the \mathcal{R} value computed in this work and in that of Stuardi et al. (2024) for the test sample of LoTSS-DR2/PSZ2 clusters as a function of the clusters redshift.

Table B.1. Uncertainties on the \mathcal{R} value computed in each redshift bin.

z_{bin}	$\sigma_{z_{bin}}$
$z \leq 0.1394$	0.004
$z \in (0.1394, 0.23225]$	0.012
$z \in (0.23225, 0.321677]$	0.015
$z \in (0.321677, 0.4561]$	0.019
$z > 0.4561$	0.019

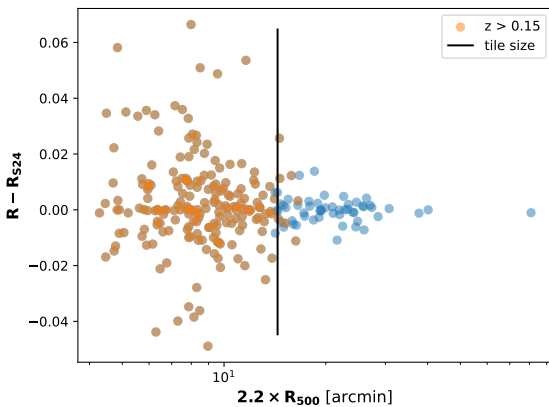


Fig. B.2. Difference between the \mathcal{R} value computed in this work and in that of Stuardi et al. (2024) for the test sample of LoTSS-DR2/PSZ2 clusters as a function of the radius used to compute \mathcal{R} . The size of the tile used by Radio U-Net is shown by the vertical line.

To account for this, we computed a statistical uncertainty on \mathcal{R} as a function of cluster redshift. The test sample was divided into five redshift bins, each containing 49 clusters, and for each bin we calculated the standard deviation of the absolute differences between the two \mathcal{R} estimates:

$$\sigma_{z_{bin}} = \sigma(|\mathcal{R} - \mathcal{R}_{S24}|_{z_{bin}}), \quad (\text{B.1})$$

where the five bins are listed in Tab. B.1 with corresponding uncertainties.

These uncertainties are assigned to the corresponding \mathcal{R} values in the cluster catalogues produced in this work.

Appendix C: Examples of galaxy clusters from the curated sub-sample

Fig. C.1 shows an example of a peculiar system in the curated sub-sample where diffuse radio emission was detected by the network in a region connecting three nearby galaxy clusters. In Fig. C.2, we show instead some examples of galaxy clusters in the curated sub-sample without previous studies in the literature and showing diffuse radio emission, which, due to its morphology and position with respect to the cluster, could be attributed to the presence of radio halos or radio relics.

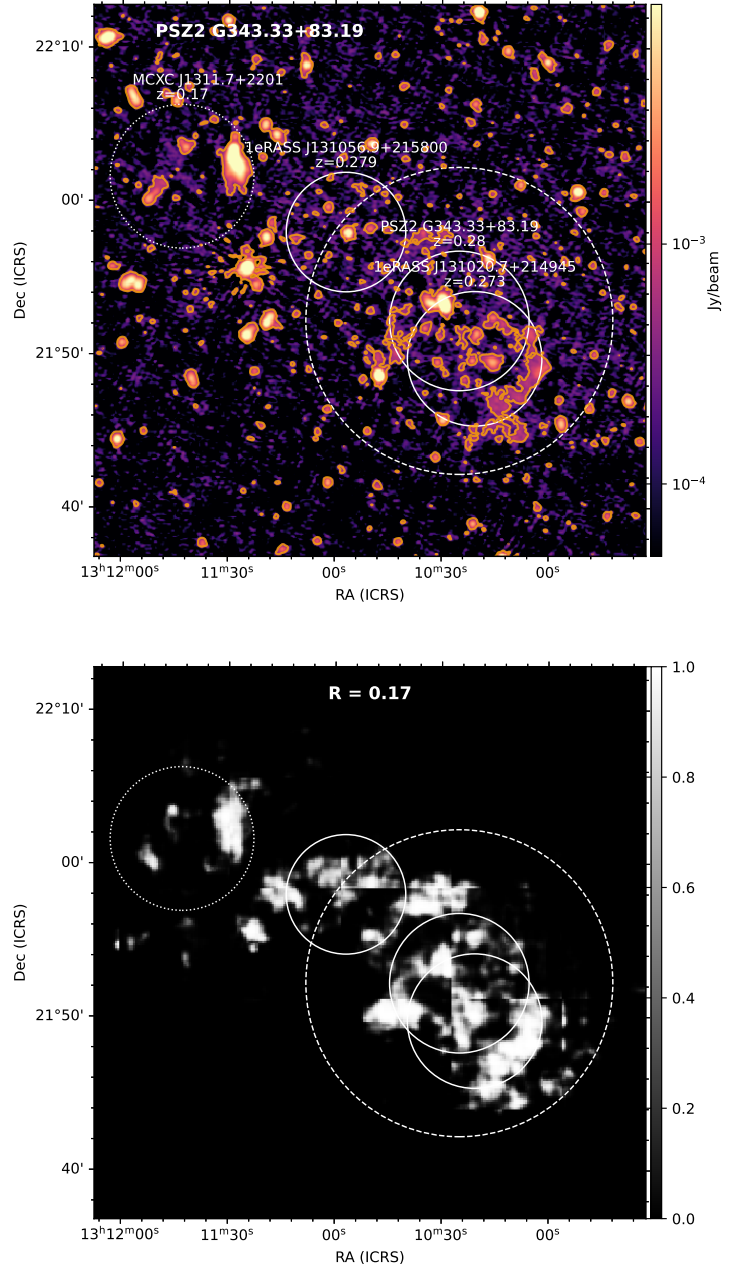


Fig. C.1. Example of a peculiar system within the curated sub-sample. The LoTSS-DR3 image at $20''$ resolution is shown in the top panel, with contours at the 3σ level (0.39 mJy/beam). The three solid white circles show the R_{500} of the three clusters found at similar redshifts. The dashed circle shows the $2.2R_{500}$ radius of the main cluster (PSZ2 G343.33+83.19), while the dotted circle shows another cluster in the foreground. The segmentation map is shown in the bottom panel.

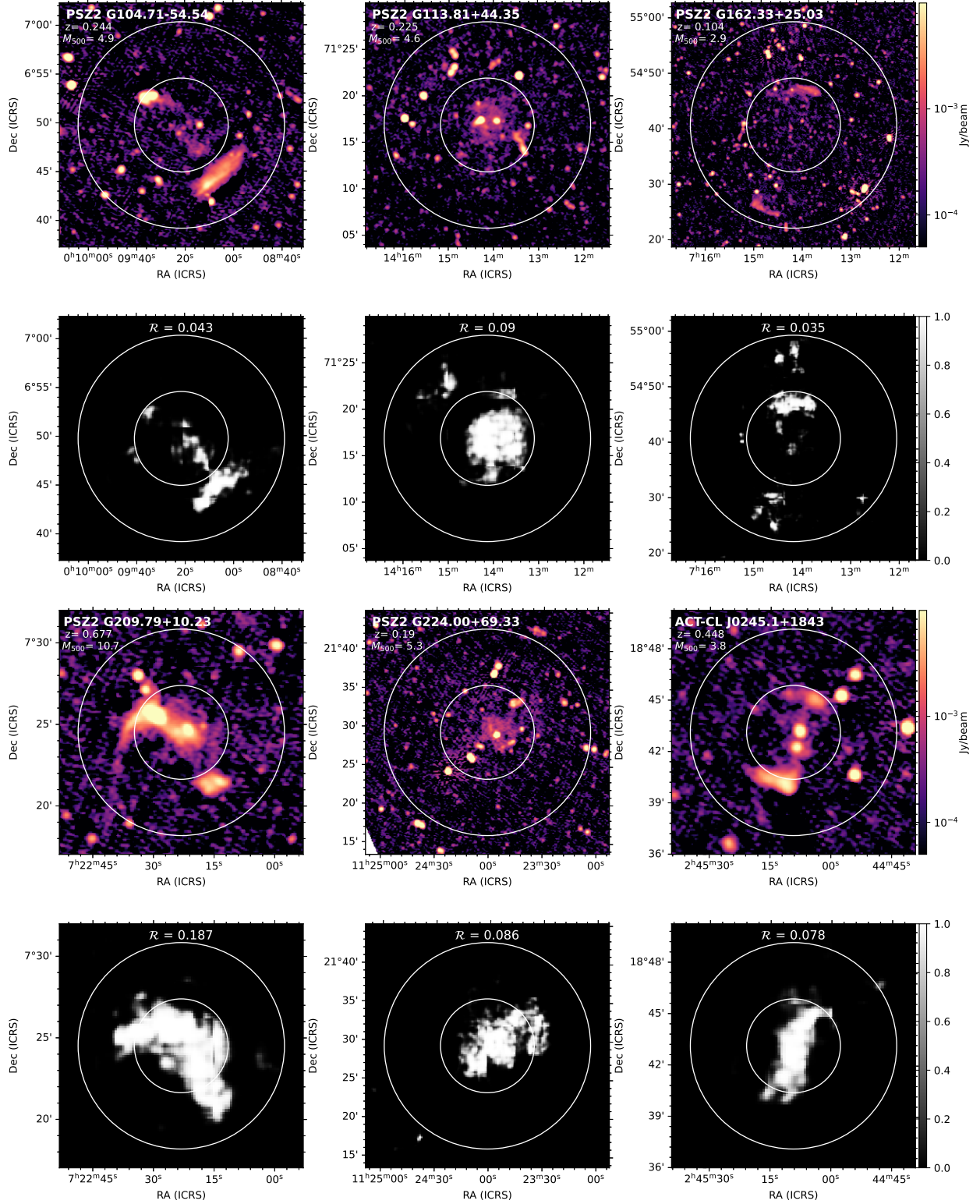


Fig. C.2. Example of six galaxy clusters in the curated sub-sample showing diffuse radio emission which, due to its morphology and position with respect to the cluster, could be attributed to the presence of radio halos or radio relics. The LoTSS-DR3 image at 20'' resolution is shown in the first and third rows. The inner white circle represents the R_{500} of the cluster, while the outer circle represents $2.2R_{500}$. Redshift and M_{500} (in units of $10^{14}M_{\odot}$) are shown in the image. The \mathcal{R} value of each cluster is shown over the segmented maps shown in the second and fourth rows.