

3D Classification of Paramagnetic Rim Lesions in Multiple Sclerosis via Asymmetric QSM–FLAIR Modeling

Veronica Pignedoli^{1,✉}, Giacomo Boffa^{2,3}, Nicoletta Noceti¹, Matilde Inglese^{2,3},
Francesca Odone¹, Matteo Moro¹

¹ MaLGa, DIBRIS, University of Genova, Genova, Italy

² DINOGMI, University of Genova, Genova, Italy

³ IRCCS Azienda Ospedaliera Metropolitana, Genova, Italy

✉ veronica.pignedoli@edu.unige.it

Abstract. Paramagnetic rim lesions (Rim^+) identified on susceptibility-sensitive MRI have recently emerged as a specific biomarker of chronic active inflammation in Multiple Sclerosis (MS) and are associated with long-term disability progression. However, susceptibility imaging and expert interpretation remain limited to specialized centers, visual assessment is time-consuming and variable, and the low prevalence of Rim^+ lesions poses severe class imbalance challenges for automated analysis. We propose FRODO, a 3D Fusion framework for Rim lesion classification using multimodal Deep-learning neuroimaging, designed for lesion-level $\text{Rim}^+/\text{Rim}^-$ classification from Quantitative Susceptibility Mapping (QSM) and FLAIR MRI. The architecture explicitly models modality asymmetry by treating QSM as the primary susceptibility-driven signal and conditioning it with FLAIR-derived structural context. To improve robustness under limited data, we employ self-supervised multimodal pretraining followed by supervised fine-tuning with contrastive regularization. The method was evaluated on a clinically acquired cohort of 88 people with MS with expert lesion annotations as reference standard. Results highlight improved performance compared to prior architectures, supporting the effectiveness of asymmetric multimodal modeling for automated chronic active lesion identification.

Keywords: Multiple Sclerosis · Lesion Classification · Multimodal alignment · 3D Convolutional Neural Network

1 Introduction

Clinical problem formulation. Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system characterized by focal demyelinating lesions with heterogeneous pathological evolution. Among these, chronic active lesions (Rim^+), also referred to as paramagnetic rim lesions (PRL), are defined by a rim of iron-laden microglia and macrophages surrounding a demyelinated core

and are associated with increased axonal damage and worse long-term disability progression [14, 1, 7]. In this context, lesions without this feature are referred to as Rim^- . The clinical relevance of Rim^+ lesions has recently been reinforced by the 2024 revision of the McDonald diagnostic criteria, which recognizes advanced imaging biomarkers as supportive evidence to improve diagnostic specificity [18]. In routine clinical practice, structural MRI sequences such as Fluid Attenuated Inversion Recovery (FLAIR) and T1-weighted (T1w) imaging are commonly acquired and are sensitive to lesion detection, but provide limited specificity regarding the underlying tissue substrate. Instead, Quantitative Susceptibility Mapping (QSM), a quantitative MRI technique that estimates magnetic susceptibility (χ), allows for in vivo assessment of iron accumulation and has emerged as a key modality for the identification of Rim^+ lesions [22]. However, visual Rim^+ identification in QSM is time-consuming, expertise-dependent, and susceptible to confounding factors such as veins and imaging artifacts [9]. These limitations call for automated classification procedures. In this work, we address the problem of automatic lesion-level ($\text{Rim}^+/\text{Rim}^-$) classification from lesion-centered MRI patches extracted from co-registered QSM and FLAIR volumes.

State of the art. Early computational approaches relied on susceptibility phase imaging. Barquero et al. [2] proposed RimNet, a 3D patch-based convolutional neural network operating on phase and FLAIR images through parallel VGG-inspired branches. Lou et al. [16] introduced the Automated Paramagnetic Rim Lesion (APRL) classifier, which leverages handcrafted radiomic features combined with SMOTE-based imbalance handling [5]. Hu et al. [11] proposed AL-PaCA, an end-to-end 3D model for Rim^+ classification using EPI sequences. Shifting towards Quantitative Susceptibility Mapping (QSM), Zhang et al. [24] proposed QSMRim-Net, an imbalance-aware architecture that integrates convolutional features from QSM and FLAIR with a large set of lesion-level radiomic descriptors derived from QSM, coupled with DeepSMOTE-based feature synthesis. QSMRim-Net represents the most comprehensive QSM-based framework currently available. However, its reliance on handcrafted radiomic features and latent-space oversampling introduces additional computational complexity and may limit generalization. Zhang et al. [25] introduced DeDA (Deep Directed Accumulator) to explicitly injects domain-specific geometric priors into neural networks to improve QSM-based Rim^+ identification.

Our contributions. Motivated by these considerations, we propose FRODO, a 3D end-to-end multimodal Fusion framework for Rim lesion classification using multimodal Deep-learning neuroimaging, operating directly on lesion-centered QSM and FLAIR patches for $\text{Rim}^+/\text{Rim}^-$ lesion classification. The architecture explicitly models modality asymmetry: QSM encodes the primary susceptibility-driven signature, while FLAIR provides structural context through spatial feature-wise linear modulation (FiLM) [19]. To improve robustness under extreme class imbalance, limited sample size and no available benchmarks, we incorporate self-supervised cross-modal pretraining and supervised contrastive regularization. The proposed method is evaluated on a cohort of 88 people with MS acquired in a routine clinical setting, reflecting heterogeneous imaging conditions

and real-world Rim⁺ prevalence. Our main contributions are: (i) an end-to-end multimodal lesion-level classifier based on QSM and FLAIR (<https://github.com/veronicapignedoli/FRODO>); (ii) an asymmetric conditioning strategy that preserves susceptibility-specific discriminative features while integrating structural context; and (iii) a comprehensive evaluation on a clinically representative dataset with statistical comparison against prior approaches.

2 Method

Our method explicitly models the asymmetric physiological roles of the two MRI modalities: QSM encodes the primary susceptibility-driven rim signature, while FLAIR provides structural context through spatial feature-wise linear modulation (FiLM) [19]. An overview is shown in Figure 1. The method follows a two-stage training strategy: first, self-supervised pretraining with a cross-modality objective; second, supervised fine-tuning for binary PRL classification.

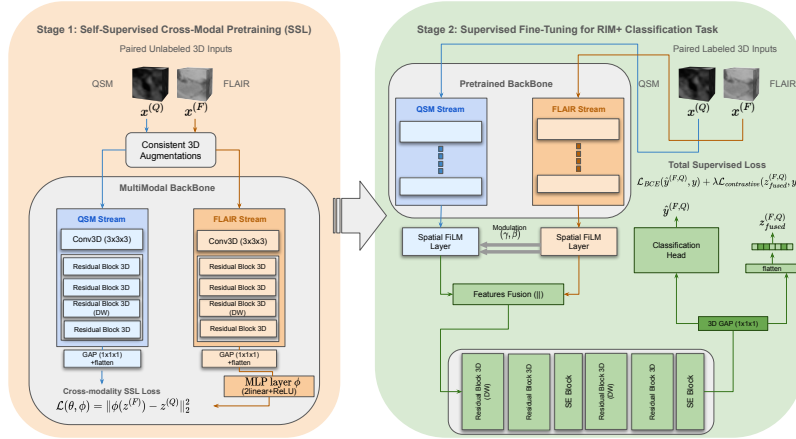


Fig. 1. Overview of the proposed multi-modal 3D classification framework for Rim⁺/Rim⁻ prediction. The network is first pretrained using a self-supervised contrastive objective and then fine-tuned with supervised binary classification loss.

2.1 Proposed architecture details

Two streams backbone. Given paired patches $x^{(Q)}$ (QSM) and $x^{(F)}$ (FLAIR), the network comprises two separate convolutional encoders built using 3D pre-activation residual blocks. Each encoder begins with a $3 \times 3 \times 3$ convolution, followed by two ResidualBlock3D modules with stride 1, a downsampling ResidualBlock3D with stride (2, 2, 2), and a final ResidualBlock3D with stride

1. Each residual block follows a pre-activation design consisting of Group Normalization and ReLU, followed by a $3 \times 3 \times 3$ convolution, a second Group Normalization and ReLU, another convolution, and a skip connection.

Spatial FiLM conditioning. After two-separate stream encoders, features extracted pass through the **Spatial FiLM layer**. Let $z^{(Q)}$ and $z^{(F)}$ denote intermediate QSM and FLAIR feature maps. Spatially varying modulation fields γ and β are generated from $z^{(F)}$ via $1 \times 1 \times 1$ convolutions. The modulated QSM representation is:

$$\tilde{z}^{(Q)} = (1 + \gamma) \odot z^{(Q)} + \beta, \quad (1)$$

where $\odot$ denotes element-wise multiplication. This spatial conditioning enables localized contextual adaptation while preserving QSM as the dominant modality; the network uses structural information from FLAIR modality to scale and shift the primary susceptibility signal. A $1 \times 1 \times 1$ convolution follows feature concatenation before deeper processing.

Channel recalibration and classification. Squeeze-and-excitation (SE) blocks [12] are applied at deeper stages to enhance channel selectivity, as shown in Figure 1. The **SE Block** first squeezes spatial dimensions via 3D Global Average Pooling (GAP), then computes channel-wise weights through a $1 \times 1 \times 1$ convolutional bottleneck with reduction and ReLU; finally scales the input tensor element-wisely using Sigmoid activation. After the second **SE Block**, the final feature map is aggregated via global average pooling and passed to a **Classification Head** composed by two fully connected layers with ReLU activation and dropout. A sigmoid activation produces the probability of a lesion being Rim^+ .

2.2 Training Strategy

Self-supervised multimodal pretraining. To improve representation robustness under limited data availability and severe class imbalance, the backbone is first initialized through a self-supervised multimodal pretraining phase [26, 20]. Pretraining is based on cross-modal alignment between intermediate QSM and FLAIR embeddings. Let $z^{(Q)}$ and $z^{(F)}$ denote modality-specific feature representations extracted at the bottleneck stage. We enforce modality-consistent representations through a regression objective $\mathcal{L}_{\text{SSL}} = \|\phi(z^{(F)}) - z^{(Q)}\|_2^2$, where $\phi(\cdot)$ is a lightweight multilayer perceptron mapping QSM features to the FLAIR embedding space. We minimize standard MSE loss between cross-modal embeddings, treating the QSM representation as a target, and enforce that the QSM features are predictable from the corresponding FLAIR features computed on the same spatially transformed patch. During pretraining, random 3D spatial augmentations (axis flips and 90° rotations) are applied consistently across modalities improving robustness [3]. After pretraining, the projection head is discarded, the pretrained weights are loaded for the backbone encoders up to the FiLM layers and the model is fine-tuned end-to-end for supervised classification.

Supervised fine-tuning. During supervised training, the model is optimized using a composite objective. For lesion classification, we employ binary cross-entropy

(BCE) as the primary loss. To improve class separability under severe imbalance, we additionally incorporate a supervised contrastive loss $\mathcal{L}_{\text{contrastive}}$ applied to the global lesion embedding [13]. The contrastive term follows the normalized temperature-scaled cross-entropy (NT-Xent) formulation [6, 17], promoting intra-class compactness and inter-class separation in the embedding space.

The overall training objective is defined as $\mathcal{L}_{\text{Sup}} = \mathcal{L}_{\text{BCE}} + \lambda \mathcal{L}_{\text{contrastive}}$, where λ controls the relative contribution of the contrastive component. The model outputs a probability $p \in [0, 1]$ representing the likelihood of a lesion being Rim^+ . A decision threshold t^* was used to obtain binary predictions: lesions with $p \geq t^*$ were classified as Rim^+ , whereas lesions with $p < t^*$ were classified as Rim^- .

3 Experiments

3.1 Dataset and Preprocessing

The dataset comprises MRI scans from 88 people with relapsing-remitting MS acquired at San Martino Hospital (Genova, Italy) using a 3T MAGNETOM Prisma scanner (Siemens). The protocol included T1-weighted (T1w), FLAIR, and QSM sequences. Lesion masks were manually segmented and lesion status ($\text{Rim}^+/\text{Rim}^-$) was independently assigned by two expert clinicians. In cases of disagreement, lesions were jointly reviewed to reach a consensus. The study was conducted in accordance with the Declaration of Helsinki principles and approved by the local Institutional Ethics Committee (protocol ID 10346). Written informed consent was obtained from all participants. QSM volumes were acquired at $0.65 \times 0.65 \times 0.65$ mm resolution, while FLAIR volumes were acquired at $1.0 \times 1.0 \times 1.0$ mm resolution. All modalities were rigidly registered to subject-specific T1w space ($1.0 \times 1.0 \times 1.0$ mm resolution) using ANTs [23]. N4 bias field correction was performed to mitigate low-frequency intensity inhomogeneities. Lesion-centered 3D patches were then extracted from instance-level masks, where each connected component represented an individual lesion. Consistent with [21] and clinical expertise, lesions smaller than 110 voxels ($\approx 0.1, mL$) were excluded to ensure reliable morphological characterization. After preprocessing and size-based exclusion, the final dataset consisted of 1247 lesion-centered patches, of which 90 (7.22%) were RIM^+ , reflecting their low prevalence as reported in the literature [21]. For each lesion ℓ , the center of mass of its binary mask M_ℓ was used to extract a fixed $64 \times 64 \times 64$ voxel patch from both modalities, using zero-padding when necessary. Intensity normalization was performed independently per modality: FLAIR volumes were scaled using robust percentile normalization (0.5–99.5%) per participant, while QSM volumes were clipped to $[-300, 300]$ and linearly rescaled to preserve susceptibility contrast. To constrain spatial support to lesion-centered context, we applied adaptive 3D morphological dilation to the ground-truth mask and suppressed all voxels outside the dilated region. The dilation radius was derived from lesion volume (scaled by $k = 0.75$). The resulting mask was used to gate the patch, suppressing unrelated background tissue while preserving perilesional information.

3.2 Experimental Setup

We adopted a five-fold evaluation protocol at the participant level to prevent lesion-level data leakage and assess generalizability across different people. Participants were manually stratified to ensure balanced distribution of total lesion burden and number of RIM-positive individuals across folds. Each fold contained 16–18 participants, including 6–8 RIM-positive participants. For each evaluation run, one fold was held out as an independent test set, while the remaining four folds constituted the development set. Within the development set, lesions were randomly split into training (80%) and validation (20%) subsets. The training subset was used for model optimization, while the validation subset was used for early stopping and threshold t^* selection. This procedure was repeated five times, such that each participant served as test data exactly once. Final performance was computed by aggregating results across the five test folds. To address class imbalance, a `WeightedRandomSampler` was used to generate approximately class-balanced mini-batches by sampling lesions with probability inversely proportional to their class frequency.

Implementation Details. All experiments were implemented in PyTorch and executed on an NVIDIA H200 NVL GPU (MIG-enabled, up to 35 GB memory) using CUDA 13.0. Mixed-precision (bfloat16) training was employed. Pretraining is performed using AdamW [15] (learning rate 5×10^{-5} , weight decay 0.05) with batch size 10, mixed precision (bfloat16), early stopping (patience 20). During supervised training, the balancing coefficient λ is learned via a non-negative parametrization and the temperature parameter was set to $\tau = 0.1$. Optimization was performed using AdamW with gradient clipping. Early stopping was driven by validation loss during self-supervised pretraining, and by validation PR-AUC during supervised fine-tuning.

Evaluation. Performance was assessed at the lesion level under the fixed 5-fold evaluation protocol. Primary threshold-independent evaluation metrics were Precision–Recall AUC (PR-AUC) and ROC-AUC; for each fold, lesion-wise probabilities were generated on held-out test participants, and these metrics were computed from the full score distributions. Threshold-dependent metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV) and F1-score were computed using an optimized threshold t^* selected on the validation split by maximizing the F1-score via grid search over $t \in [0, 1]$. Summary metrics were obtained by averaging fold-wise metrics (mean $\pm$ standard deviation). To ensure a fair comparison, the exact same QSM–FLAIR input patch pairs, data splits, and evaluation protocols were used across all compared models. In addition to lesion-level metrics, we also computed person-level classification metrics.

3.3 Results

We evaluate the proposed architecture against QSMRim-Net [24] and a 3D ResNet-18 implemented in MONAI [10, 3] for binary classification. In addition, we conduct ablation studies to isolate the contribution of the contrastive component and the pretraining strategy. Specifically, we report: (i) training from

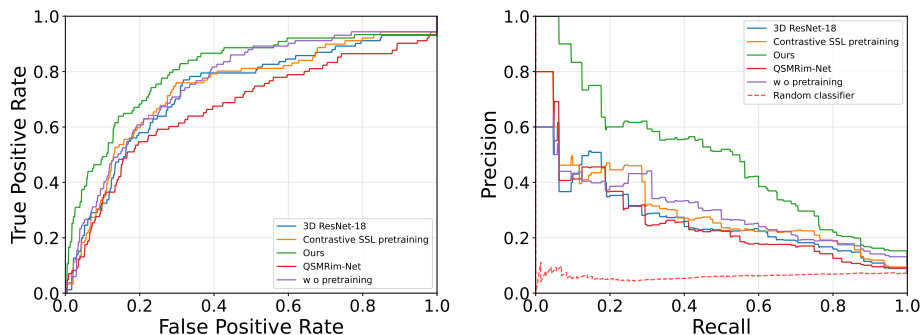


Fig. 2. Lesion-wise ROC and PR curves for the five models. The displayed are 5-fold evaluation protocol mean curves, obtaining by interpolating each fold onto a common FPR grid and averaging TPR point-wise. The PR curve also reports a simulated random-classifier baseline with positive-class prevalence of 7.22%. No confidence bands are reported for readability.

scratch using the composite objective $\mathcal{L}_{\text{BCE}} + \lambda \mathcal{L}_{\text{contrastive}}$; and (ii) self-supervised contrastive pretraining with NT-Xent objective [4] followed by supervised fine-tuning with $\mathcal{L}_{\text{BCE}}$. All models were trained using a `WeightedRandomSampler` to account for class imbalance. QSMRim-Net [24] was excluded from this setting, as it incorporates DeepSMOTE within its framework to handle imbalance. We therefore evaluated QSMRim-Net re-training the model and tested on our dataset with the same protocol evaluation, using the authors’ official GitHub implementation. During SSL contrastive pretraining, the temperature parameter was set to $\tau_{\text{SSL}} = 0.1$.

Table 1. Area-under-curve metrics at the lesion level (averaged across AUCs test folds). Values are reported as mean $\pm$ standard deviation.

Method	ROC AUC $\uparrow$	PR AUC $\uparrow$
QSMRim-Net [24]	0.658 $\pm$ 0.091	0.156 $\pm$ 0.100
3D ResNet-18 [3]	0.726 $\pm$ 0.052	0.234 $\pm$ 0.026
w/o pretraining + $\mathcal{L}_{\text{contrastive}}$	0.789 $\pm$ 0.067	0.271 $\pm$ 0.125
Contrastive SSL pretraining	0.800 $\pm$ 0.031	0.256 $\pm$ 0.129
FRODO (ours)	0.869 $\pm$ 0.049	0.457 $\pm$ 0.141

Lesion-level evaluation. The proposed model achieves the highest discriminative performance across all metrics, see Fig. 2 and Table 1, with ROC AUC 0.869 ± 0.049 and PR AUC 0.457 ± 0.141 , substantially outperforming both QSMRim-Net and the ResNet baseline. Under strong class imbalance (7.22% RIM prevalence), PR AUC provides a more faithful estimate of minority-class discrimination. The improvement in PR AUC indicates that the learned repre-

sensation better separates positive from negative samples in regions of low recall, suggesting improved feature compactness and class-conditional separability. Table 2 reports the performances obtained by the optimized threshold. The benefit of our method is primarily driven by increased precision ($\text{PPV } 0.437 \pm 0.099$) while maintaining competitive sensitivity (0.447 ± 0.221) and high specificity (0.956 ± 0.021), indicating a reduction in false-positive activations without sacrificing recall. The higher F1-score (0.460 ± 0.155) further confirms a more balanced precision–recall trade-off. These results suggest that multimodal conditioning combined with contrastive objectives induces a more structured embedding space. Ablation analyses further support this interpretation. Removing SSL pretraining leads to consistent degradation in both ROC AUC and PR AUC, with a particularly pronounced drop in PR performance. This behavior indicates that cross-modal pretraining enhances minority-class feature alignment prior to supervised optimization. Overall, the combination of spatial conditioning and contrastive pretraining yields a more discriminative and stable lesion-level representation compared to purely supervised baselines.

Table 2. Lesion-level performance comparison across test folds. Values are reported as mean $\pm$ standard deviation.

Method	Accuracy	F1-score	Sensitivity
QSMRim-Net [24]	0.855 ± 0.058	0.235 ± 0.072	0.338 ± 0.180
3D ResNet-18 [3]	0.898 ± 0.028	0.211 ± 0.122	0.189 ± 0.136
w/o pretraining + $\mathcal{L}_{\text{contrastive}}$	0.864 ± 0.077	0.325 ± 0.063	0.444 ± 0.107
Contrastive SSL pretraining	0.877 ± 0.056	0.319 ± 0.081	0.400 ± 0.128
FRODO (ours)	0.921 ± 0.012	0.460 ± 0.155	0.447 ± 0.221
Method	Specificity	PPV	
QSMRim-Net [24]	0.893 ± 0.072	0.220 ± 0.063	
3D ResNet-18 [3]	0.946 ± 0.035	0.239 ± 0.184	
w/o pretraining + $\mathcal{L}_{\text{contrastive}}$	0.895 ± 0.085	0.253 ± 0.095	
Contrastive SSL pretraining	0.914 ± 0.066	0.265 ± 0.129	
FRODO (ours)	0.956 ± 0.021	0.437 ± 0.099	

Person-level evaluation. Following the considerations reported in [24] and adapting them to our dataset, a person was considered *positive* if at least one Rim^+ lesion was present (≥ 1 Rim^+ lesion). The baseline QSMRim-Net achieved a patient-level sensitivity of 0.743, with a low specificity of 0.442 and an overall accuracy of 0.563. Our proposed model achieved a more balanced performance profile, maintaining a competitive sensitivity of 0.657 while delivering a substantially improved specificity of 0.673 and an accuracy of 0.667. Our method showed a stronger correlation between the predicted and ground-truth Rim^+ lesion count per participant (Pearson $r = 0.755$), substantially outperforming QSMRim-Net ($r = 0.439$).

Statistical Tests. DeLong’s test [8] for correlated ROC curves was applied to out-of-fold lesion-level predictions, concatenated over the 5 test sets. The proposed model significantly outperformed QSMRimNet in terms of ROC-AUC (0.839 vs 0.658; $\Delta = 0.181$; $p - value < 0.001$), with improvements observed across all test folds. For PR-AUC, statistical significance was assessed using a bootstrapping approach with 1000 replicates applied to the out-of-fold predictions. The 95% confidence interval (CI) for the PR-AUC of our proposed model was [0.2618, 0.5170], whereas the corresponding 95% CI for QSMRim-Net was [0.0812, 0.2640]. Notably, the two intervals did not overlap, indicating that the improvement in PR-AUC achieved by our framework is statistically significant.

4 Conclusion

We presented FRODO, an end-to-end multimodal framework for lesion-level classification of PRL in Multiple Sclerosis using QSM and FLAIR MRI. The proposed architecture models modality asymmetry by prioritizing susceptibility-driven information from QSM while structurally conditioning it with complementary FLAIR context. Evaluation on a clinically representative cohort highlight improved lesion- and person-level discrimination compared to baseline architectures, supporting robust automated Rim⁺/Rim[−] classification without reliance on handcrafted radiomic features. Future work will focus on validation across larger multi-center cohorts with heterogeneous acquisition protocols, as well as integration into end-to-end lesion detection and longitudinal monitoring pipelines; strategies to address data scarcity and severe class imbalance will be investigated, including generative augmentation of Rim⁺ lesions and leveraging larger unlabeled datasets to strengthen the self-supervised pretraining stage.

Disclosure of Interests. The authors declare no competing interests.

References

1. Absinta, M., Sati, P., Masuzzo, F., Nair, G., Sethi, V., Kolb, H., Ohayon, J., Wu, T., Cortese, I.C., Reich, D.S.: Association of chronic active multiple sclerosis lesions with disability in vivo. *JAMA neurology* **76**(12), 1474–1483 (2019)
2. Barquero, G., La Rosa, F., Kebiri, H., Lu, P.J., Rahmanzadeh, R., Weigel, M., Fartaria, M.J., Kober, T., Théaudin, M., Du Pasquier, R., et al.: Rimnet: A deep 3d multimodal mri architecture for paramagnetic rim lesion assessment in multiple sclerosis. *NeuroImage: Clinical* **28**, 102412 (2020)
3. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701* (2022)
4. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in neural information processing systems* **33**, 12546–12558 (2020)
5. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **16**, 321–357 (2002)

6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PmLR (2020)
7. Dal-Bianco, A., Oh, J., Sati, P., Absinta, M.: Chronic active lesions in multiple sclerosis: classification, terminology, and clinical significance. *Therapeutic Advances in Neurological Disorders* **17**, 17562864241306684 (2024)
8. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* pp. 837–845 (1988)
9. Elliott, C., Rudko, D.A., Arnold, D.L., Fetco, D., Elkady, A.M., Araujo, D., Zhu, B., Gafson, A., Tian, Z., Belachew, S., et al.: Lesion-level correspondence and longitudinal properties of paramagnetic rim and slowly expanding lesions in multiple sclerosis. *Multiple Sclerosis Journal* **29**(6), 680–690 (2023)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. Hu, F., Ren, Z., Chen, L., Valcarcel, A.M., Dworkin, J., Renner, B., Daboul, L., O'Donnell, C.M., Verter, E.D., Manning, A.R., et al.: Automated segmentation of multiple sclerosis lesions, paramagnetic rims, and central vein sign on mri provides reliable diagnostic biomarkers. *Imaging Neuroscience* **3**, IMAG-a (2025)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
13. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. *Advances in neural information processing systems* **33**, 18661–18673 (2020)
14. Kuhlmann, T., Ludwin, S., Prat, A., Antel, J., Brück, W., Lassmann, H.: An updated histological classification system for multiple sclerosis lesions. *Acta neuropathologica* **133**(1), 13–24 (2017)
15. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
16. Lou, C., Sati, P., Absinta, M., Clark, K., Dworkin, J.D., Valcarcel, A.M., Schindler, M.K., Reich, D.S., Sweeney, E.M., Shinohara, R.T.: Fully automated detection of paramagnetic rims in multiple sclerosis lesions on 3t susceptibility-based mr imaging. *NeuroImage: Clinical* **32**, 102796 (2021)
17. Mildenerger, D., Hager, P., Rueckert, D., Menten, M.J.: A tale of two classes: adapting supervised contrastive learning to binary imbalanced datasets. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 10305–10314 (2025)
18. Montalban, X., Lebrun-Frénay, C., Oh, J., Arrambide, G., Moccia, M., Amato, M.P., Amezcu, L., Banwell, B., Bar-Or, A., Barkhof, F., et al.: Diagnosis of multiple sclerosis: 2024 revisions of the mcdonald criteria. *The Lancet Neurology* **24**(10), 850–865 (2025)
19. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
20. Taleb, A., Loetzsch, W., Danz, N., Severin, J., Gaertner, T., Bergner, B., Lippert, C.: 3d self-supervised methods for medical imaging. *Advances in neural information processing systems* **33**, 18158–18172 (2020)
22. Tazza, F., Boffa, G., Schiavi, S., Lapucci, C., Piredda, G.F., Cipriano, E., Zacà, D., Roccatagliata, L., Hilbert, T., Kober, T., et al.: Multiparametric characterization

- and spatial distribution of different ms lesion phenotypes. *American Journal of Neuroradiology* **45**(8), 1166–1174 (2024)
22. Tranfa, M., Pontillo, G., Petracca, M., Brunetti, A., Tedeschi, E., Palma, G., Cocozza, S.: Quantitative mri in multiple sclerosis: from theory to application. *American Journal of Neuroradiology* **43**(12), 1688–1695 (2022)
 23. Tustison, N.J., Cook, P.A., Holbrook, A.J., Johnson, H.J., Muschelli, J., Devenyi, G.A., Duda, J.T., Das, S.R., Cullen, N.C., Gillen, D.L., Yassa, M.A., Stone, J.R., Gee, J.C., Avants, B.B.: The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports* **11**(1), 9068 (Apr 2021). <https://doi.org/10.1038/s41598-021-87564-6>, <https://doi.org/10.1038/s41598-021-87564-6>
 24. Zhang, H., Nguyen, T.D., Zhang, J., Marcille, M., Spincemaille, P., Wang, Y., Gauthier, S.A., Sweeney, E.M.: Qsmrim-net: Imbalance-aware learning for identification of chronic active multiple sclerosis lesions on quantitative susceptibility maps. *NeuroImage: Clinical* **34**, 102979 (2022)
 25. Zhang, H., Wang, R., Hu, R., Zhang, J., Li, J.: Deda: Deep directed accumulator. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 765–775. Springer (2023)
 26. Zong, Y., Mac Aodha, O., Hospedales, T.M.: Self-supervised multimodal learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **47**(7), 5299–5318 (2024)