

Feature Space Guidance for Breast Cancer Classification in DCE-MRI

Benjamin Hamm^{1,4}, Yannick Kirchhoff^{1,2,3}, Maximilian Rokuss^{1,2,3,5}, Moritz Langenberg^{1,2}, Constantin Ulrich¹, Tassilo Wald^{1,2,5}, Jeremias Traub^{1,5}, Karol Gotkowski^{1,5}, and Klaus Maier-Hein^{1,4,6}

¹ German Cancer Research Center (DKFZ) Heidelberg, Division of Medical Image Computing, Germany

² Faculty of Mathematics and Computer Science, Heidelberg University, Germany

³ HIDSS4Health – Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany

⁴ Medical Faculty, Heidelberg University, Germany

⁵ Helmholtz Imaging, German Cancer Research Center (DKFZ), Heidelberg, Germany

⁶ Pattern Analysis and Learning Group, Department of Radiation Oncology, Heidelberg University Hospital, Germany
`benjamin.hamm@dkfz-heidelberg.de`

Abstract. Dynamic contrast enhanced breast MRI (DCE-MRI) is a powerful clinical tool for breast cancer detection, providing high resolution anatomical detail together with rich temporal contrast information. However, high dimensional 4D inputs, small lesions, and heterogeneous acquisition protocols across clinical sites hinder robust automated classification of healthy, benign, and malignant cases. To address these challenges, we propose a framework that dynamically analyzes latent representations to adapt to protocol-specific characteristics. Spatial variability is mitigated by reducing confounding background uptake and compensating for misalignment caused by deformable soft tissue. Additionally, relationships in the latent space across phases are leveraged to select the most informative temporal features, improving robustness to protocol-specific temporal variability. Finally, task specific discriminative features are promoted through large scale supervised lesion segmentation pretraining, which substantially enhances downstream finetuning. Evaluated under leave-one-center-out validation on the ODELIA dataset and the held-out AML cohort, the proposed framework substantially outperforms finetuned radiology foundation models and prior methods, improving mean AUROC by nearly 8 points and balanced accuracy by 4 points over the strongest baseline. Additionally, our method achieved first place in the MICCAI ODELIA Breast MRI Challenge 2025, further demonstrating its effectiveness for robust breast cancer classification. We publicly release our codebase under <https://github.com/MIC-DKFZ/CURIAtor>.

Keywords: Breast Cancer Classification · Adaptive Temporal Modeling · Dynamic Contrast Enhanced MRI

1 Introduction

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide [29]. Early and accurate diagnosis is therefore critical for improving patient outcomes and guiding personalized treatment [25]. Among available imaging modalities, dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has emerged as a powerful tool for breast cancer assessment, as it provides both high-resolution anatomical information and rich temporal signatures of tissue perfusion and permeability [19]. These functional characteristics enable improved sensitivity compared to conventional mammography and ultrasound, particularly in dense breast tissue [23]. Approaching this as a classification of healthy, benign, and malignant cases is attractive because case-level labels are easier to obtain than dense segmentation annotations, making the task formulation well suited for scalable screening [27].

Despite its clinical value, reliable automated classification of DCE-MRI remains a challenging problem. A typical examination consists of large three-dimensional volumes acquired at multiple time points following contrast injection, resulting in high-dimensional, heterogeneous data [14]. Lesions often occupy only a small fraction of the imaged volume, resulting in extreme foreground sparsity that weakens discriminative signal extraction. Moreover, acquisition protocols vary substantially across scanners and institutions, with differences in temporal resolution, number of phases, and contrast timing. This variability induces distribution shifts that degrade the performance and reliability of supervised learning systems trained on limited or homogeneous datasets [1].

Existing approaches attempt to cope with the temporal and volumetric complexity of DCE-MRI through aggressive input simplifications. A common strategy is to predefine a fixed subset of temporal phases and applying this same configuration to all cases irrespective of protocol differences or lesion specific enhancement patterns [15,1,28,31]. While computationally convenient, such fixed phase selection implicitly assumes consistent temporal sampling across centers and does not account for variations in the number of acquired phases or irregular inter scan intervals. Other methods reduce dimensionality by applying maximum intensity projections (MIP) to individual 3D volumes or subtraction volumes independently, thereby collapsing the volumetric spatial information into multi-channel 2D representations [13,2]. Although this strategy reduces memory requirements and simplifies modeling, it introduces irreversible information loss by compressing depth and contrast uptake dynamics into a single intensity representation. In particular, subtle enhancement differences may be suppressed, while surrounding regions that enhance strongly at any time point can dominate the projection, potentially obscuring small lesions and diagnostically relevant kinetic patterns. These limitations motivate the need for a method that explicitly structures spatial and temporal invariance rather than relying on static dimensionality reduction, as existing approaches remain vulnerable to protocol-specific biases that contribute to degraded cross-center generalization in multi-site settings [1].

To overcome these challenges, we introduce a framework that dynamically

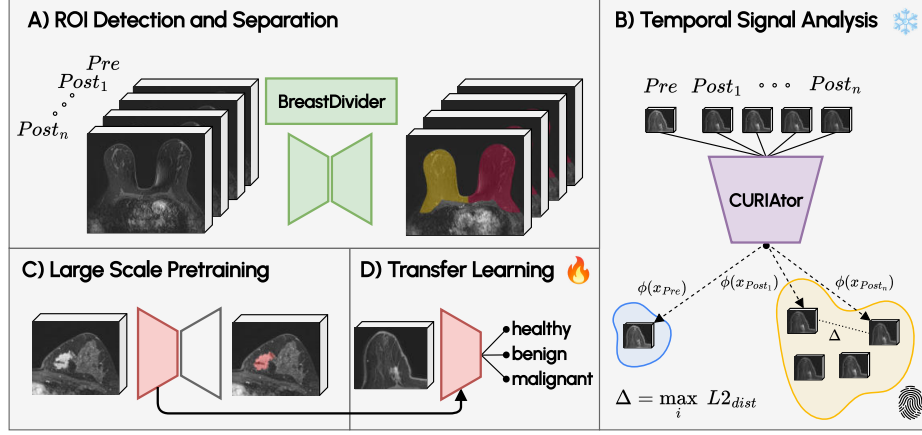


Fig. 1: Overview of the proposed foundation model-driven framework for breast cancer classification from DCE-MRI. The framework includes A) the *Breast-Divider* model for ROI detection and breast separation, B) adaptive temporal signal analysis with the *CURIA*tor module selects the most informative features, C) large-scale anatomically informed pretraining, and D) transfer learning for final lesion classification.

adapts to the latent characteristics of any given protocol for robust breast cancer classification from DCE-MRI. Rather than simplifying the input through static phase selection or projection, our approach explicitly organizes data modeling into successive levels of abstraction that address the three dominant sources of variability in clinical DCE-MRI.

First, spatial variability is addressed through anatomically informed localization that reduces confounding background uptake and compensates for misalignment caused by deformable soft tissue.

Second, temporal variability is addressed through adaptive relational modeling of enhancement dynamics. The Curia radiology foundation model (FM) [7] first embeds each temporal volume into a high level feature space, after which the representation constructs relationships across phases and dynamically selects the most informative temporal features. This enables the model to accommodate differences in phase number, timing, and sampling across centers without relying on a fixed acquisition scheme or predefined phase selection.

Third, to compensate for the weak supervisory signal inherent to breast-level classification, we leverage large-scale anatomically grounded pretraining, as pure classification, unlike segmentation, provides no explicit spatial guidance toward lesion regions. Pretraining on anatomically structured tasks therefore shapes feature representations before finetuning, improving sensitivity to small pathological patterns despite the absence of pixel level supervision.

To the best of our knowledge, we are the first to introduce a temporally adaptive component rather than relying on a static handling of dynamic acquisitions.

Furthermore, we are the first to integrate spatial alignment, temporal alignment, and supervised pretraining into a unified three stage hierarchy, forming a single coherent framework for robust breast cancer classification.

2 Methods

Our method follows a multi-level representation hierarchy that progressively abstracts anatomical, temporal, and task-specific information from heterogeneous DCE-MRI data. A hierarchical overview of the proposed framework is shown in Fig. 1.

2.1 BreastDivider Model

First, we propose BreastDivider, a nnU-Net-based [11] 3D segmentation model trained to localize and separate left and right breast regions in clinical MRI volumes. It is trained on a curated large-scale dataset comprising 17,956 scans aggregated from multiple public cohorts spanning more than 55 centers [26,21,18,9,8,6,20,24]. The dataset covers a broad spectrum of acquisition protocols, including T1, contrast enhanced T1, T2, FLAIR, and diffusion weighted imaging, thereby capturing substantial inter center and inter sequence variability. The dataset was constructed using an iterative active learning strategy, starting from the Duke cohort [26] and progressively expanding. In successive training rounds, only high quality and informative examples were retained for subsequent optimization. A sample was deemed reliable when all five cross-validation folds achieved an agreement exceeding 0.95 Dice overlap. Training on this diverse and anatomically curated resource enables BreastDivider to achieve highly accurate localization performance (Dice: 99.08 ± 0.48) and to serve as a robust preprocessing component for downstream tasks.

2.2 CURIator Module

DCE-MRI acquisitions are inherently redundant, with diagnostically relevant information concentrated in a few physiologically distinct phases, particularly around the early post contrast period [16,17]. Clinical evidence from abbreviated and ultrafast breast MRI protocols shows that combining pre contrast imaging, the first post contrast phase, and a delayed phase capturing complementary enhancement can achieve performance close to full dynamic acquisitions [5]. Motivated by this observation, we design the *CURIator* module to operationalize this principle. Given the established importance of the pre contrast and first post contrast volumes, the module adaptively identifies a third phase that provides the most complementary temporal signal for the specific acquisition. Concretely, we select the most **dissimilar** post contrast phase. Drawing on principles from diversity-based active learning, samples that are far apart in feature space tend to provide the most new information relative to already observed data [3].

Formally, let $\mathcal{V} = \{V^n\}_{n=1}^N$ be the set of N temporal volumes of a single subject, where each volume $V^n \in \mathbb{R}^{D \times H \times W}$ represents one acquisition. The BreastDivider mask $M \in \{0, 1\}^{D \times H \times W}$ constrains all subsequent processing to the anatomically relevant region. All volumes are first cropped to the minimal three-dimensional bounding box enclosing the mask foreground, eliminating the large surrounding background that would otherwise contribute noise and increase computational cost. Each normalized slice V_d^n is independently processed by the pretrained Curia radiology FM [7], yielding slice-level features $X_d^n \in \mathbb{R}^{P \times C}$, where P is the number of spatial tokens and C is the embedding dimension. To improve robustness to local slice-to-slice misalignment, tokens are aggregated across slices using average pooling with a kernel size of 8, resulting in a volume-level embedding $E^n \in \mathbb{R}^{M \times C}$, where $M = D_r \cdot P$ and D_r denotes the pooled depth. Anatomical relevance is further enforced by projecting the ROI mask onto the token grid, yielding fractional overlap weights $w_{d,p} \in [0, 1]$ that down-weight contributions from background tokens in all subsequent distance computations. The pairwise dissimilarity between volumes i and j is computed as a masked, spatially weighted distance. For each slice d , the token-level squared distances are aggregated under the ROI weights as:

$$D_{d_r}^{ij} = \sqrt{\frac{\sum_p w_{d_r,p} \delta_{d_r,p}^{ij}}{\sum_p w_{d_r,p}}}, \quad (1)$$

where $\delta_{d_r,p}^{ij} = \frac{1}{C} \sum_c (E_{m,c}^i - E_{m,c}^j)^2$ is the normalized squared Euclidean distance at token position $m = (d_r, p)$. Slice-level distances are averaged over the cropped depth to yield a scalar volume-level dissimilarity $D_{ij} = \frac{1}{D_r} \sum_{d_r} D_{d_r}^{ij}$. Collecting all pairwise values yields a symmetric distance matrix $\mathbf{D} \in \mathbb{R}^{N \times N}$, which defines a similarity manifold over the full acquisition sequence. Taking the first post-contrast volume V^{Post_1} as a fixed anchor *CURIA*tor selects the most informative remaining phase by maximizing its dissimilarity to V^{Post_1} : $n^* = \arg \max_{n \geq 3} D_{2n}$. The resulting triplet $\{V^{\text{Pre}}, V^{\text{Post}_1}, V^{n^*}\}$ constitutes a protocol-invariant temporal representation of the acquisition. Because the selection is driven entirely by the learned embedding geometry of Curia, the module requires no explicit phase labels, no predefined kinetic descriptors, and no assumptions about sampling regularity. The representation is therefore robust to variable phase counts, irregular inter-scan intervals, and scanner-specific acquisition differences across clinical centers. Sample volume level similarities are visualized in Fig. 2.

2.3 Transfer Learning Framework

Since self supervised learning often struggles to capture small pathologies and subtle abnormal regions [30], we adopt an anatomically informed transfer learning strategy to initialize and adapt our classification models across heterogeneous breast MRI datasets. We first perform breast tumor segmentation pretraining on the MAMA-MIA cohort by splitting bilateral examinations into individual breast

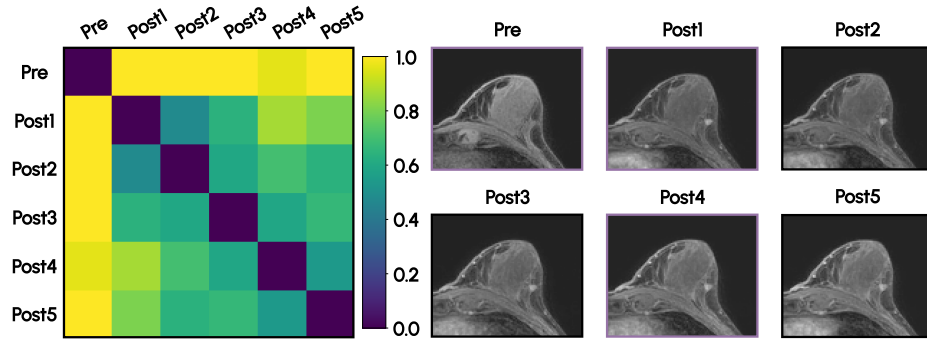


Fig. 2: Pairwise distance matrices $\mathbf{D}$ computed by the *CURIA*tor module for UMCU 8300274987 Left (ODELIA). Each matrix encodes the volume-level similarity between temporal acquisitions within the anatomically constrained ROI. Lower values indicate higher similarity, while higher values reflect stronger contrast-induced differences.

regions, resulting in more than 17,000 three dimensional breast volumes derived from 3,012 breast instances acquired across multiple imaging sequences. All inputs are treated as single channel images, enabling the encoder to learn sequence agnostic representations and to capture the full range of intensity and contrast variability present across acquisition protocols. This is achieved by training nnU-Net for 2000 epochs using the *ResEncL* configuration [12]. The pretrained encoder is then used as a universal initialization for the downstream classification task.

3 Experiments and Results

We finetune the classification model from the pretrained weights for 100 epochs using SGD with cross-entropy loss and weight decay of 3×10^{-5} . The learning rate follows a polynomial decay schedule with a 10-epoch linear warmup and a peak value of 1×10^{-3} . We evaluate our framework with focus on cross center generalization for the classification of healthy (H), benign (B), and malignant (M) breast tumors, as summarized in Table 1. We perform five fold cross validation on the ODELIA dataset, where each fold corresponds to a distinct clinical center to enforce strict separation across acquisition sites, scanners, and imaging protocols and thus provide a realistic assessment of distribution shift, and additionally evaluate on the held out AMBL dataset. Subgroup analyses are precluded due to the lack of harmonized cohort metadata. Both datasets are dominated by healthy cases (ODELIA: 683 H, 137 B, 202 M; and AMBL: 190 H, 91 B, 111 M), with AMBL comparatively more balanced across the three classes. Further acquisition details, such as scanner parameters, can be found in the respective original publications [20,8]. We benchmark our approach against finetuned FMs, namely Merlin [4], VISTA3D [10], CT-FM [22], and Curia [7]. We leave native

Table 1: Cross-center performance comparison on the ODELIA[20] and AMBL[8] datasets. Results are reported in terms of AUROC and balanced accuracy (BA) for each clinical center and on AMBL, as well as their mean. Bootstrap 95% CIs in brackets.

Dataset	CAM	MHA	RUMC	UKA	UMCU	AMBL	Mean
AUROC [%] ↑							
Curia[7]	53.5 [48.5, 58.9]	50.5 [42.0, 60.3]	37.6 [16.6, 59.4]	51.4 [44.5, 58.5]	53.7 [47.1, 59.9]	52.2 [47.7, 56.8]	49.8
VISTA3D[10]	54.1 [48.8, 59.6]	62.0 [53.3, 70.1]	57.1 [33.0, 82.0]	49.5 [43.1, 55.7]	61.6 [55.1, 68.0]	51.5 [47.6, 55.7]	56.0
CT-FM[22]	52.0 [47.2, 57.5]	49.1 [40.5, 57.3]	35.9 [25.6, 46.9]	48.6 [42.0, 55.6]	61.4 [54.6, 67.6]	55.4 [51.3, 59.7]	50.4
Merlin [4]	64.0 [59.1, 68.6]	59.0 [49.7, 68.9]	67.7 [52.6, 83.0]	48.4 [42.1, 55.2]	55.2 [48.3, 61.5]	56.0 [52.6, 60.1]	58.4
MIP[13,2]	72.4 [65.8, 79.4]	67.6 [60.5, 74.9]	67.7 [60.6, 74.7]	70.3 [62.7, 77.5]	76.6 [69.4, 83.0]	49.8 [47.9, 52.0]	67.4
Ours	82.7 [78.1, 87.5]	79.0 [70.4, 86.5]	72.8 [60.4, 86.0]	73.5 [67.2, 79.2]	71.7 [66.6, 76.7]	71.1 [69.3, 72.7]	75.1
Balanced Accuracy [%] ↑							
Curia [7]	33.6 [32.2, 35.2]	33.0 [32.3, 33.3]	22.6 [3.10, 42.4]	36.2 [29.8, 43.1]	36.1 [30.1, 41.7]	33.2 [32.8, 33.3]	32.4
VISTA3D[10]	33.2 [33.0, 33.3]	43.2 [37.9, 48.1]	46.1 [26.7, 64.8]	34.2 [26.8, 41.8]	46.7 [40.9, 52.4]	33.8 [33.0, 34.8]	39.5
CT-FM[22]	36.6 [31.8, 41.6]	37.4 [31.4, 43.3]	32.4 [30.1, 50.0]	33.6 [25.8, 41.7]	42.3 [36.8, 47.6]	37.3 [34.7, 40.3]	36.6
Merlin[4]	48.0 [44.2, 52.0]	44.5 [39.0, 49.7]	38.7 [30.5, 65.2]	30.9 [24.3, 37.5]	43.2 [37.9, 48.1]	39.9 [37.0, 42.9]	40.9
MIP[13,2]	63.3 [58.0, 68.7]	61.5 [56.4, 66.7]	56.7 [52.7, 61.4]	61.3 [56.3, 66.6]	69.3 [64.0, 74.7]	50.1 [49.5, 50.8]	60.4
Ours	72.5 [69.7, 75.1]	64.5 [59.4, 70.5]	62.2 [52.0, 71.2]	61.7 [57.5, 65.8]	63.2 [58.0, 68.9]	62.5 [61.1, 63.9]	64.4

4D model architectures as future work. All non-CURIA baselines were fine-tuned end-to-end with input resolution and CURIAator-selected inputs matched to our method; We used AdamW with a cosine schedule (100 epochs, batch 32, 50 steps/epoch). For CURIA: Frozen 2D embeddings, slice-wise mean aggregation and linear probe. Furthermore, we compare against existing approaches based on MIP. Finally, we conduct ablation studies to assess the contributions of the transfer learning strategy, the ROI prediction, and the *CURIAator* module by comparing adaptive temporal selection against static input configurations.

Best cross-institutional generalization. Our framework consistently outperforms radiology FMs across all centers, suggesting that general-purpose models, while effective for generic feature extraction, are less suited for finetuning on tasks with localized targets. Moreover, the improvements over MIP-based approaches underscore the limitations of aggressively collapsing depth dimensions.

Superior adaption to protocol variability. Table 2 reports ablations isolating each component’s contribution. Removing pretraining clearly degrades performance, confirming the value of transfer learning for robust representation initialization. Replacing the BreastDivider ROI prediction with whole-volume processing also sharply reduces performance, underscoring the benefit of anatomically constrained feature extraction. Finally, the *CURIAator* selection consis-

tently outperforms static sequence configurations, showing that dynamic temporal weighting provides significant discriminative signal.

Table 2: Ablation study evaluating the contribution of each component of the proposed framework. Performance is reported in terms of mean AUROC and balanced accuracy (BA) across centers. $\Delta\%$ denotes the relative performance change with respect to the final framework using *CURIA*tor.

	ROI	Pretraining	<i>CURIA</i> tor	Mean AUROC	Mean BA
Ablation [$\Delta\%$]					
No Pretraining	✓	–	✓	-19.19	-15.61
Whole Volume	–	✓	✓	-19.89	-15.00
Input Sequences [$\Delta\%$]					
$Pre + Post_{n/2} + Post_n + T2$	✓	✓	–	-11.33	-3.16
$Pre + Post_{n/2} + Post_n$	✓	✓	–	-11.32	-3.15
$Pre + Post_1 + Post_n$	✓	✓	–	-11.65	-3.22
$Pre + Post_1 + Post_2$	✓	✓	–	-4.86	-1.66
Ours [%]	✓	✓	✓	75.13	64.42

4 Discussion and Conclusion

In this work, we propose a comprehensive framework for robust breast cancer classification from DCE-MRI. Our approach addresses three central challenges of clinical DCE-MRI analysis: high dimensional 4D inputs, small lesion extent relative to background anatomy, and substantial inter center protocol variability. To the best of our knowledge, we are the first to introduce a native 4D temporally adaptive component rather than relying on static phase selection or collapsing volumes using MIP. We further integrate spatial alignment, temporal alignment, and supervised lesion pretraining into a unified three stage hierarchy, forming a single coherent architecture to address the outlined challenges. Extensive cross center evaluation on ODELIA and AMBL demonstrates consistent and substantial improvements over finetuned radiology foundation models and MIP based baselines under identical training conditions. Our method achieves a mean AUROC of 75.13 and a mean BA of 64.42 across centers, outperforming the strongest MIP baseline by nearly 8 AUROC points and 4 BA points. Replacing *CURIA*tor with static phase combinations reduces performance by up to 11.65 AUROC points and over 3 BA points. The observed performance gains indicate that robust generalization under real world distribution shift stems not from increasing model capacity or self supervised pretraining scale alone, but from explicitly structured, anatomy aware and temporally adaptive representation design. In addition, our method achieved first place in the MICCAI ODELIA Breast MRI Challenge 2025, further demonstrating its effectiveness for robust clinical breast cancer classification.

Acknowledgments. This work was supported by the Helmholtz Association under the joint research program “HIDSS4Health – Helmholtz Information and Data Science School for Health” and under the Helmholtz Foundation Model Initiative (HFMI), project “The Human Radiome Project” (THRP). This work was partially funded by Helmholtz Imaging, a platform of the Helmholtz Information & Data Science Incubator, by “NUM 2.0“ (FKZ: 01KX2121), by “NUM 3.0” (FKZ: 01KX2524), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number 402688427, and by the Helmholtz AI project “Effective Privacy-Preserving Adaptation of Foundation Models for Medical Tasks” (PAFMIM; ZT-I-PF-5-227). Maximilian Rokuss was supported by the Google PhD Fellowship Program.

Disclosure of Interests. Maximilian Rokuss was supported by the Google PhD Fellowship Program. Other than that the authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alanazi, M., Alsharif, B.: Domain shift in breast dce-mri tumor segmentation: A balanced lococv study on the mama-mia dataset. *Diagnostics* **16**(2), 362 (2026)
2. Antropova, N., Abe, H., Giger, M.L.: Use of clinical mri maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. *Journal of Medical Imaging* **5**(1), 014503–014503 (2018)
3. Ash, J.T., Zhang, C., Krishnamurthy, A., et al.: Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671* (2019)
4. Blankemeier, L., Cohen, J.P., Kumar, A., Van Veen, D., Gardezi, S.J.S., Paschali, M., Chen, Z., Delbrouck, J.B., Reis, E., Truys, C., et al.: Merlin: A vision language foundation model for 3d computed tomography. *Research Square* pp. rs–3 (2024)
5. Cao, Y., Huang, Y., Chen, X., et al.: Optimizing ultrafast dynamic contrast-enhanced mri scan duration in the differentiation of benign and malignant breast lesions. *Insights into imaging* **15**(1), 112 (2024)
6. Comstock, C., Gatsonis, C., Newstead, G., Snyder, B., Gareen, I., Bergin, J., Rahbar, H., Sung, J., Jacobs, C., Harvey, J., et al.: Abbreviated breast mri and digital tomosynthesis mammography in screening women with dense breasts (ea1141). *The Cancer Imaging Archive: Little Rock, AR, USA* (2023)
7. Dancette, C., Khlaut, J., Saporta, A., Philippe, H., Ferreres, E., Callard, B., Danielou, T., Alberge, L., Machado, L., Tordjman, D., et al.: Curia: A multi-modal foundation model for radiology. *arXiv preprint arXiv:2509.06830* (2025)
8. Daniels, D., Last, D., Cohen, K., Mardor, Y., Sklair-Levy, M.: Standard and delayed contrast-enhanced mri of malignant and benign breast lesions with histological and clinical supporting data (advanced-mri-breast-lesions)(version 2)[data set]. *the cancer imaging archive* (2024)
9. Garrucho, L., Kushibar, K., Reidel, C.A., Joshi, S., Osuala, R., Tsirikoglou, A., Bobowicz, M., Del Riego, J., Catanese, A., Gwoździewicz, K., et al.: A large-scale multicenter breast cancer dce-mri benchmark dataset with expert segmentations. *Scientific data* **12**(1), 453 (2025)
10. He, Y., Guo, P., Tang, Y., Myronenko, A., Nath, V., Xu, Z., Yang, D., Zhao, C., Simon, B., Belue, M., et al.: Vista3d: A unified segmentation foundation model for 3d medical imaging. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 20863–20873 (2025)

11. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
12. Isensee, F., Wald, T., et al.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 488–498. Springer (2024)
13. Jing, X., Wielema, M., Cornelissen, L.J., van Gent, M., Iwema, W.M., Zheng, S., Sijens, P.E., Oudkerk, M., Dorrius, M.D., van Ooijen, P.M.: Using deep learning to safely exclude lesions with only ultrafast breast mri to shorten acquisition and reading time. *European radiology* **32**(12), 8706–8715 (2022)
14. Kataoka, M., Honda, M., Sagawa, H., Ohashi, A., Sakaguchi, R., Hashimoto, H., Iima, M., Takada, M., Nakamoto, Y.: Ultrafast dynamic contrast-enhanced mri of the breast: from theory to practice. *Journal of Magnetic Resonance Imaging* **60**(2), 401–416 (2024)
15. Khaled, R., Vidal, J., Vilanova, J.C., Marti, R.: A u-net ensemble for breast lesion segmentation in dce mri. *Computers in biology and medicine* **140**, 105093 (2022)
16. Kuhl, C.K.: Abbreviated breast mri: state of the art. *Radiology* **310**(3), e221822 (2024)
17. Kuhl, C.K., Schrading, S., Strobel, K., Schild, H.H., Hilgers, R.D., Bieling, H.B.: Abbreviated breast magnetic resonance imaging (mri): first postcontrast subtracted images and maximum-intensity projection—a novel approach to breast cancer screening with mri. *Journal of Clinical Oncology* **32**(22), 2304–2310 (2014)
18. Li, W., Newitt, D., Gibbs, J., Wilmes, L., Jones, E., Arasu, V., Strand, F., Onishi, N., Nguyen, A., Kornak, J., et al.: I-spy 2 breast dynamic contrast enhanced mri trial (ispy2)(version 1). *Cancer Imaging Arch* (2022)
19. Mann, R.M., et al.: Breast mri: state of the art. *Radiology* **292**(3), 520–536 (2019)
20. Müller-Franzes, G., Sánchez, L.E., Payne, N., et al.: A european multi-center breast cancer mri dataset. *arXiv preprint arXiv:2506.00474* (2025)
21. Newitt, D., Hylton, N., et al.: Multi-center breast dce-mri data and segmentations from patients in the i-spy 1/acrin 6657 trials. *Cancer Imaging Arch* **10**(7) (2016)
22. Pai, S., Hadzic, I., Bontempi, D., Bressen, K., Kann, B.H., Fedorov, A., Mak, R.H., Aerts, H.J.: Vision foundation models for computed tomography. *arXiv preprint arXiv:2501.09001* (2025)
23. Riedl, C.C., Luft, N., et al.: Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density. *Journal of Clinical Oncology* **33**(10), 1128–1135 (2015)
24. Rokuss, M., Hamm, B., Kirchhoff, Y., Maier-Hein, K.: Divide and conquer: A large-scale dataset and model for left-right breast mri segmentation. *arXiv preprint arXiv:2507.13830* (2025)
25. Saadatmand, S., Bretveld, R., Siesling, S., Tilanus-Linthorst, M.M.: Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients. *Bmj* **351** (2015)
26. Saha, A., Harowicz, M.R., Grimm, L.J., Weng, J., Cain, E.H., Kim, C.E., Ghate, S.V., Walsh, R., Mazurowski, M.A.: Dynamic contrast-enhanced magnetic resonance images of breast cancer patients with tumor locations [data set]. *The Cancer Imaging Archive* **10** (2021)
27. Saldanha, O.L., Zhu, J., Müller-Franzes, G., Carrero, Z.I., Payne, N.R., Escudero Sánchez, L., Varoutas, P.C., Kyathanahally, S., Laleh, N.G., Pfeiffer, K., et al.: Swarm learning with weak supervision enables automatic breast cancer detection in magnetic resonance imaging. *Communications medicine* **5**(1), 38 (2025)

28. Satake, H., Kinoshita, F., Ishigaki, S., Kato, K., Jo, Y., Shimada, S., Masuda, N., Naganawa, S.: Predictive performance of radiomic features extracted from breast mr imaging in postoperative upgrading of ductal carcinoma in situ to invasive carcinoma. *Magnetic Resonance in Medical Sciences* **24**(4), mp-2023 (2025)
29. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **71**(3), 209–249 (2021)
30. Wald, T., Ulrich, C., Lukyanenko, S., Goncharov, A., Paderno, A., Miller, M., Maerkisch, L., Jaeger, P., Maier-Hein, K.: Revisiting mae pre-training for 3d medical image segmentation. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 5186–5196 (2025)
31. Yin, H.l., Jiang, Y., Xu, Z., Jia, H.h., Lin, G.w.: Combined diagnosis of multi-parametric mri-based deep learning models facilitates differentiating triple-negative breast cancer from fibroadenoma magnetic resonance bi-rads 4 lesions. *Journal of Cancer Research and Clinical Oncology* **149**(6), 2575–2584 (2023)