Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

BE-WISER: Class-Aware Weak Slice-Level Supervision for Breast MRI

Adarsh Bhandary Panambur

adarsh.bhandary.panambur@fau.de

Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg https://orcid.org/0009-0004-8067-1150

Tri-Thien Nguyen

Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg

Saahil Islam

Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg

Michael Uder

Radiological Institute, University Clinic Erlangen, Erlangen, Germany

Sebastian Bickelhaupt

Radiological Institute, University Clinic Erlangen, Erlangen, Germany

Andreas Maier

Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg https://orcid.org/0009-0006-7583-9144

Siming Bayer

Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg

Research Article

Keywords: Breast cancer, Breast MRI, Weak supervision, Multiple-instance learning, Lesion localization

Posted Date: June 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9855612/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

BE-WISER: Class-Aware Weak Slice-Level Supervision for Breast MRI

Adarsh Bhandary Panambur[1*], Tri-Thien Nguyen[1,2], Saahil Islam[1], Michael Uder[2], Sebastian Bickelhaupt[2], Andreas Maier[1], Siming Bayer[1,3]

[1*]Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, Erlangen, 91058, Germany .
[2]Radiological Institute, University Clinic Erlangen, Erlangen, Germany .
[3]Siemens Healthineers, Erlangen, Germany .

*Corresponding author(s). E-mail(s): adarsh.bhandary.panambur@fau.de;

Abstract

Purpose: Breast MRI is sensitive for cancer assessment, but deep learning models often require dense lesion annotations. This study evaluates whether weak slice-level supervision can be extended from binary lesion localization to class-aware slice semantics.

Methods: We propose BE-WISER, a weakly supervised framework that builds on the recently proposed BE-WISE approach by extending binary slice-level lesion localization to three-class slice semantics. BE-WISER models each slice as normal, benign, or malignant and combines a Swin Transformer slice encoder, attention-based multiple-instance learning for breast-level classification, and a class-aware slice head for weak lesion localization. Gaussian targets from a single radiologist-indicated lesion slice provide weak supervision, and radiologist review was used to assess the clinical plausibility of predicted high-probability slices.

Results: On the independent ODELIA test set, BE-WISER with focal loss achieved the best breast-level performance, with an AUC of 0.864 and an ODELIA score of 0.699. This improved over BE-WISE with focal loss by 1.2 and 2.6 percentage points, respectively. For localization, BE-WISER achieved the lowest mean slice distance of 2.505 slices and the highest Hit@1 of 0.632. Radiologist review indicated that high-probability slices corresponded well to lesion-relevant slices.

Conclusion: BE-WISER demonstrates that pathology-aware slice semantics improve weakly supervised breast MRI diagnosis and lesion localization while preserving clinically interpretable slice-level evidence.

Keywords: Breast cancer, Breast MRI, Weak supervision, Multiple-instance learning, Lesion localization

1 Introduction

Breast cancer remains a major global healthcare burden and is among the leading causes of cancer incidence and mortality in women worldwide [1]. With millions of new cases reported annually, breast cancer screening programs have rapidly evolved over the past decades, with mammography established as the standard modality for population-wide screening. However, mammographic sensitivity is reduced in women with dense breast tissue, a challenge that is particularly relevant in populations with higher breast-density distributions [2]. In such cases, lesion characterization becomes increasingly difficult due to the overlapping appearance of dense fibroglandular tissue and malignant findings on mammograms. Breast magnetic resonance imaging (MRI) has therefore gained increasing importance in breast cancer diagnosis and management owing to its superior soft-tissue contrast and improved sensitivity for lesion detection in dense breasts [3]. Unlike mammography, breast MRI provides multi-parametric information through multiple imaging sequences, including pre-contrast, post-contrast, and subtraction imaging, enabling enhanced visualization of lesion morphology and contrast uptake behavior [4]. Nevertheless, breast MRI remains a costly and time-intensive imaging modality, generating large volumetric datasets that require careful interpretation by experienced radiologists.

Deep learning-based computer-aided diagnosis systems have been increasingly investigated for breast MRI analysis. However, reliable lesion detection and segmentation often require voxel-level expert annotations, which are expensive, time-consuming, and difficult to scale across large multi-sequence MRI datasets. Breast-level classification offers a more practical alternative by predicting lesion presence and pathology from labels such as BIRADS scores [5]. In this context, weakly supervised learning is methodologically well suited, and multiple-instance learning (MIL) provides a natural framework for aggregating slice-level evidence under breast-level supervision while retaining attention-based interpretability [6–8]. Transformer-based encoders further support this direction by improving representation learning across slices and slice sequences, as demonstrated by the Swin Transformer and Medical Slice Transformer in 3D medical imaging, including breast MRI [9, 10]. Explainable anomaly detection in breast MRI has also shown that clinically meaningful tumor localization can be achieved without segmented lesion masks, highlighting the value of spatially grounded explanations beyond examination-level prediction [11]. Building on these ideas, BE-WISE combined a Swin-based slice encoder with attention-based MIL and Gaussian-smoothed supervision from a single radiologist-indicated slice to jointly learn breast-level diagnosis and weak slice-level lesion localization [6]. However, existing

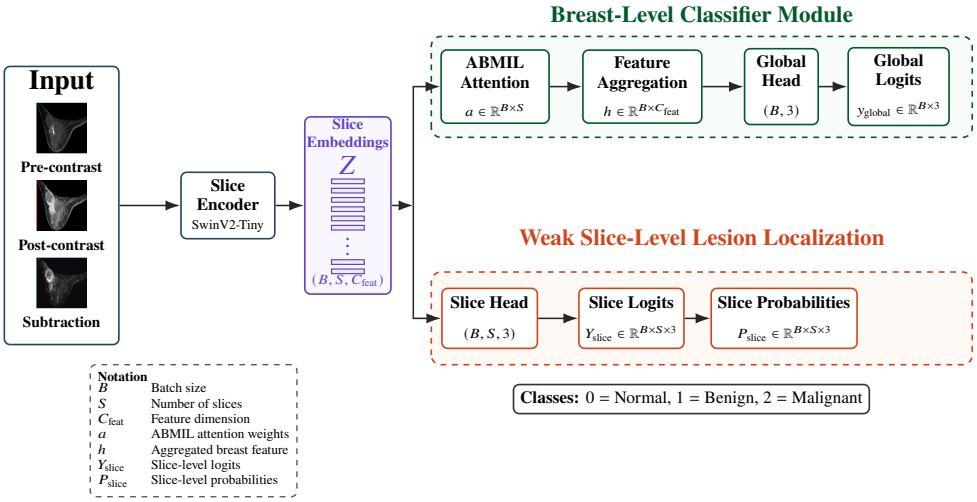


Fig. 1: Overview of the proposed BE-WISER framework. Multi-sequence breast MRI are encoded into slice embeddings **Z**, which are used for attention-based MIL breast-level classification and three-class slice-level prediction for weak lesion localization.

weakly supervised breast MRI pipelines mainly aggregate slice evidence for global classification or recover localization without assigning diagnostically explicit semantics to individual slices [6, 10, 11]. Although BE-WISE enables weak slice-level lesion localization, its binary objective does not distinguish benign from malignant slice evidence [6]. BE-WISER addresses this by introducing three-class slice-semantic supervision, explicitly modeling normal, benign, and malignant evidence to improve both diagnosis and localization.

The main contributions of this paper are : 1. We propose BE-WISER, a weakly supervised breast MRI framework that extends binary slice-aware supervision to three-class slice-semantic learning by modeling slices as normal, benign, or malignant; 2. We show that three-class slice semantics improve both breast-level diagnosis and slice-level lesion localization compared with breast-level baselines and the previous BE-WISE framework [6]; 3. We evaluate the clinical plausibility of the resulting slice-level evidence using radiologist review and investigate annotation efficiency in a reduced-label setting.

2 Methods

Figure 1 summarizes the proposed BE-WISER architecture. Multi-sequence breast MRI slices are encoded using a shared Swin Transformer, producing slice embeddings that are used by two parallel branches. The breast-level branch aggregates slice embeddings with attention-based MIL for global classification, while the slice-level branch learns weak lesion localization. BE-WISER extends the original binary BE-WISE formulation to three-class slice semantics.

2.1 Dataset and preprocessing

Compared with our previous work, this study uses the latest ODELIA Breast MRI Challenge 2025 dataset release, which includes additional multicenter examinations from a newly incorporated institution, increasing both cohort size and institutional diversity [12]. The patient-independent cohort comprised 926 training, 230 validation, and 292 test examinations from six institutions. Class distributions were preserved across splits and included normal, benign, and malignant breast examinations, with an expected imbalance toward normal cases reflecting clinical prevalence [12]. Each examination was represented as a volumetric dynamic contrast-enhanced MRI (DCE-MRI) slice sequence. Following the BreastDivider framework [13], left and right breasts were automatically segmented and cropped from original volumetric MRI scans. Model inputs comprised three channels: pre-contrast, latest available post-contrast, and early subtraction images. The latest post-contrast phase was selected to accommodate site-specific variation in acquisition protocols, while subtraction images were computed after geometric alignment by subtracting the pre-contrast from the earliest post-contrast phase. Each cropped breast volume was resampled to 36 axial slices with an in-plane resolution of 224×224, resulting in an input tensor of shape $(S, 3, H, W)$, where $S = 36$ and $H = W = 224$. For lesion-bearing examinations, i.e., benign and malignant cases, a board-certified radiologist with more than five years of experience reviewed the ODELIA reference labels and selected the slice most representative of the abnormality using the BE-WISE annotation tool[1]. Intensities were min–max normalized per channel and examination, and random horizontal and vertical flips were applied during training.

2.2 BE-WISE and BE-WISER

2.2.1 Slice encoder and breast-level classifier

Given a preprocessed breast MRI volume $\mathbf{X} \in \mathbb{R}^{B\times S\times 3\times H\times W}$, each axial slice is independently encoded using a shared ImageNet-pretrained SwinV2-Tiny backbone with its classification head removed, following BE-WISE [6]. This produces slice embeddings $\mathbf{Z} \in \mathbb{R}^{B\times S\times C_{\text{feat}}}$, where B denotes the batch size, $S = 36$ denotes the number of slices, and $C_{\text{feat}} = 768$ denotes the feature dimension. We denote the embedding of the i-th slice by $\mathbf{z}_i \in \mathbb{R}^{C_{\text{feat}}}$.

The breast-level classifier follows an attention-based MIL formulation. The slice embeddings $\mathbf{Z}$ are transformed into attention weights $\mathbf{a} \in \mathbb{R}^{B\times S}$, normalized across slices, and used to obtain a weighted breast-level representation:

$$\mathbf{h} = \sum_{i=1}^{S} a_i \mathbf{z}_i, \qquad \mathbf{h} \in \mathbb{R}^{B\times C_{\text{feat}}}. \tag{1}$$

The aggregated representation $\mathbf{h}$ is passed through layer normalization, dropout, and a linear global head to produce breast-level logits $\mathbf{y}_{\text{global}} \in \mathbb{R}^{B\times 3}$ for normal, benign,

[1] The BE-WISE annotation tool and resulting annotations will be made publicly available at https://github.com/adarshbhandaryp/BEWISE_Tool.

and malignant classification. The global classification loss $\mathcal{L}_{\text{global}}$ is implemented using either cross-entropy or focal loss, depending on the experimental setting.

2.2.2 Weak slice-level lesion localization: from BE-WISE to BE-WISER

BE-WISE [6] performs weak slice-level lesion localization using a binary slice-level lesion predictor. For benign and malignant examinations, weak supervision is generated from the radiologist-indicated lesion slice s_0 using a Gaussian weak target:

$$t_i = \exp\left(-\frac{(i - s_0)^2}{2\sigma^2}\right), \tag{2}$$

where i denotes the slice index and σ controls the spatial extent of supervision. Normal examinations are assigned zero lesion targets across all slices. The BE-WISE objective is:

$$\mathcal{L}_{\text{BE-WISE}} = \alpha\mathcal{L}_{\text{global}} + (1 - \alpha)\mathcal{L}_{\text{slice}}^{\text{binary}}. \tag{3}$$

BE-WISER retains the BE-WISE encoder and breast-level classifier, but reformulates weak slice supervision from binary lesion evidence to three-class slice semantics. The slice-level branch is applied directly to $\mathbf{Z}$ and consists of layer normalization, a linear projection, GELU activation, dropout, and a final linear layer. It produces per-slice logits $\mathbf{Y}_{\text{slice}} \in \mathbb{R}^{B\times S\times 3}$. Softmax normalization over the class dimension yields per-slice class probabilities:

$$\mathbf{P}_{\text{slice}} = \text{softmax}(\mathbf{Y}_{\text{slice}}), \qquad \mathbf{P}_{\text{slice}} \in \mathbb{R}^{B\times S\times 3}. \tag{4}$$

Thus, each slice receives a probability distribution over normal, benign, and malignant classes.

For lesion-bearing examinations, the Gaussian weak target defines the weak lesion region around s_0. Slices within this region inherit the corresponding breast-level class, benign or malignant, whereas the remaining slices are supervised as normal. For normal examinations, all slices are assigned the normal class. The BE-WISER objective is:

$$\mathcal{L}_{\text{BE-WISER}} = \alpha\mathcal{L}_{\text{global}} + (1 - \alpha)\mathcal{L}_{\text{slice}}^{\text{3-class}}, \tag{5}$$

where $\mathcal{L}_{\text{slice}}^{\text{3-class}}$ denotes the class-aware slice-level cross-entropy loss applied to $\mathbf{Y}_{\text{slice}}$.

In contrast to BE-WISE, which learns lesion versus non-lesion slice evidence, BE-WISER converts weak slice-level lesion localization into explicit normal, benign, and malignant slice semantics. This preserves the interpretable MIL structure of BE-WISE while explicitly coupling local slice semantics with breast-level pathology prediction.

3 Experimental setup

3.1 Training details

All experiments were repeated across five random seeds using identical preprocessing, optimization, and evaluation settings for the breast-level baselines, BE-WISE, and

Table 1: Breast-level classification performance on the independent test set. BL denotes the breast-level baseline, and CE denotes cross-entropy. Values denote mean (standard deviation); best results are shown in bold.

Method	CE		Focal	
	AUC ↑	ODELIA ↑	AUC ↑	ODELIA ↑
BL	0.796 (0.076)	0.578 (0.117)	0.807 (0.037)	0.591 (0.069)
BE-WISE	0.849 (0.010)	0.669 (0.016)	0.852 (0.006)	0.673 (0.005)
BE-WISER	**0.856 (0.017)**	**0.692 (0.043)**	**0.864 (0.004)**	**0.699 (0.027)**

BE-WISER. Models were trained with AdamW using a learning rate of 5×10^{-6}, weight decay of 10^{-4}, batch size of 8, and weighted random sampling. Training was performed for up to 100 epochs with early stopping after 11 epochs. For focal loss, $\gamma \in \{0.5, 1.0, 1.5, 2.0\}$ was evaluated. For slice-aware supervision, Gaussian margins $\sigma \in \{1, 2, 3\}$ were explored. Final test results were obtained from the checkpoint with the highest validation AUC, and the independent test set was used only once for final evaluation. Performance was assessed using micro-averaged AUC and the ODELIA score, following the ODELIA Challenge protocol [12]. The ODELIA score combines AUC, sensitivity at 90% specificity, and specificity at 90% sensitivity. Slice-level localization was evaluated using MSD and Hit@K. MSD denotes the average absolute difference, in slice indices, between the predicted peak lesion slice and the radiologist-indicated lesion slice. Hit@K measures whether the predicted peak slice lies within $\pm K$ slices of the radiologist-indicated lesion slice, with $K \in \{1, 3, 5\}$.

3.2 Reduced-label learning with pseudo weak slice supervision

To evaluate annotation efficiency, we studied BE-WISER under a reduced-label protocol in which only 50% of the training examinations retained radiologist-provided slice annotations. The training cohort of 926 examinations was divided into 462 examinations with slice supervision and 464 examinations treated as slice-unlabeled during the first stage, while preserving patient independence as well as class and institutional distributions. We compared three reduced-label settings: slice supervision (S), no slice supervision (NS), and pseudo weak slice supervision (PS). In the first setting, denoted 50% S, BE-WISER was trained only with the 462 examinations containing radiologist-provided slice supervision. In the second setting, denoted 50% S + 50% NS, the remaining 464 training examinations were added using only breast-level labels, without slice-level supervision. In the third setting, denoted 50% S + 50% PS, a stage-one BE-WISER model trained on the annotated subset was used to generate pseudo weak slice supervision for the remaining positive training examinations. These pseudo annotations were then combined with the original radiologist-provided annotations for stage-two retraining on the full training cohort. Model selection was performed exclusively on the validation set, and the independent test set was reserved for final evaluation. This protocol assesses whether pseudo weak slice supervision can recover performance under limited radiologist annotation.

Table 2: Slice-level localization performance. MSD denotes mean slice distance, and Hit@K indicates localization within $\pm K$ slices of the radiologist-indicated lesion slice. Values denote mean (standard deviation); best results are shown in bold.

Method	MSD ↓	Hit@1 ↑	Hit@3 ↑	Hit@5 ↑
BE-WISE (CE)	2.838 (0.126)	0.554 (0.037)	0.735 (0.044)	0.804 (0.023)
BE-WISER (CE)	2.662 (0.270)	0.608 (0.047)	**0.784 (0.009)**	**0.848 (0.017)**
BE-WISE (Focal)	2.804 (0.310)	0.564 (0.037)	0.745 (0.045)	0.819 (0.017)
BE-WISER (Focal)	**2.505 (0.220)**	**0.632 (0.026)**	0.775 (0.034)	0.843 (0.037)

4 Results

4.1 Breast-level classification performance

Table 1 reports breast-level classification on the independent test set. BE-WISER consistently outperformed the breast-level baseline and BE-WISE. With CE, BE-WISER improved AUC by 6.0 percentage points and ODELIA score by 11.4 percentage points over the breast-level baseline. With focal loss, BE-WISER achieved the best results, improving AUC by 5.7 percentage points and ODELIA score by 10.8 percentage points over the focal baseline. Relative to BE-WISE, BE-WISER further increased ODELIA score by 2.3 percentage points with CE and 2.6 percentage points with focal loss, supporting the benefit of class-aware slice-semantic supervision. Paired statistical testing showed that BE-WISER achieved statistically significant improvements in ODELIA score over the baseline models ($p < 0.0001$).

4.2 Slice-level lesion localization performance

Table 2 reports slice-level localization on positive test examinations with valid radiologist annotations. BE-WISER consistently improves over BE-WISE, reducing MSD and increasing Hit@K across most settings. BE-WISER with focal loss achieves the most precise peak-slice alignment, with the lowest MSD of 2.505 slices and the highest Hit@1 of 0.632. BE-WISER with CE achieves the highest Hit@3 and Hit@5, indicating stronger agreement with the broader lesion region. Similar to the slice annotation, a radiologist identified the slice containing the lesion center independently, which aligned with the predicted high-probability slices. These findings suggest that class-aware slice semantics improve not only quantitative localization accuracy, but also the clinical plausibility of the model's slice-level explanations.

4.3 Reduced-label learning with pseudo weak slice supervision

Figure 2 summarizes ODELIA score performance across full-data and reduced-label experiments. In the full-data setting, performance increases from breast-level baselines to BE-WISE and BE-WISER, with BE-WISER using focal loss achieving the highest ODELIA score of 0.699. This represents a 10.8 percentage point gain over the focal breast-level baseline, supporting the value of class-aware weak slice-level lesion localization. In the reduced-label setting, BE-WISER trained with only 50%

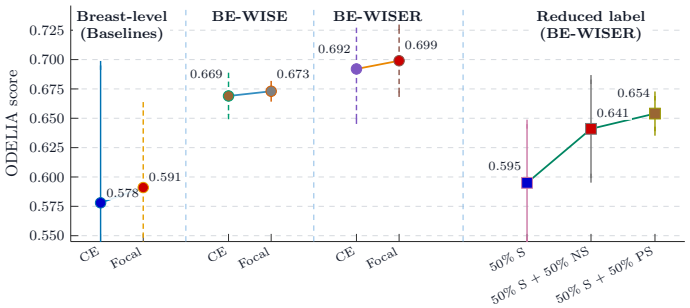
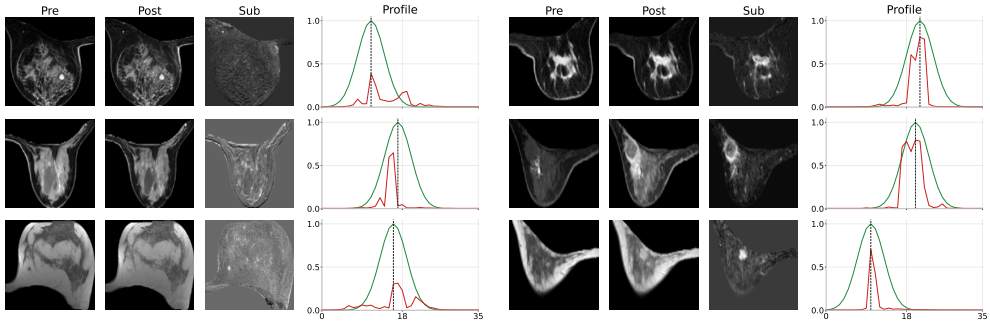


Fig. 2: ODELIA score across breast-level baselines, BE-WISE, BE-WISER, and reduced-label BE-WISER. Circles denote full-supervision experiments and squares denote reduced-label experiments. S, NS, and PS indicate slice supervision, no slice supervision, and pseudo weak slice supervision, respectively. Markers show the mean over five seeds; error bars show standard deviation.



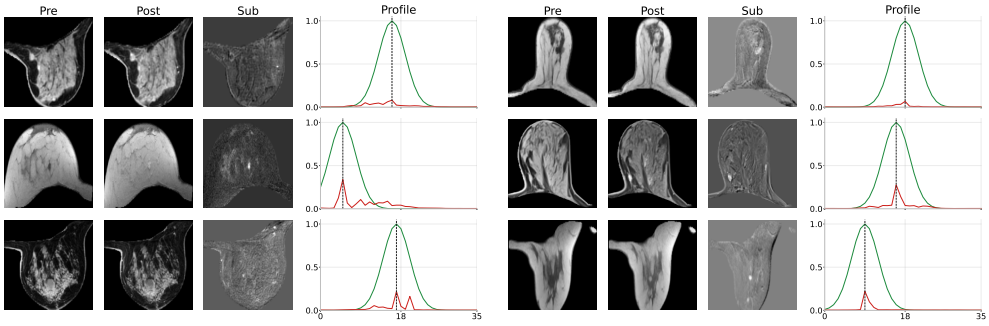
(a) Benign (b) Malignant

Fig. 3: Correctly localized benign and malignant examples. The green curve shows the Gaussian weak target, the red curve the predicted positive-class probability, and the dashed line the radiologist-indicated lesion slice.

radiologist-provided slice supervision reaches an ODELIA score of 0.595. Adding the remaining training cases with breast-level labels only improves the score to 0.641, and adding pseudo weak slice supervision further increases it to 0.654. Thus, pseudo supervision provides a 5.9 percentage point gain over the 50% slice-supervised setting and a 1.3 percentage point gain over using the remaining cases without slice supervision.

4.4 Qualitative slice-level analysis

Figure 3 shows representative benign and malignant cases in which the predicted slice probability profiles align with the radiologist-indicated lesion region. These examples illustrate that BE-WISER not only improves global classification but also produces spatially meaningful slice-level evidence. The alignment between predicted peaks and



(a) Benign misclassified as normal (b) Malignant misclassified as normal

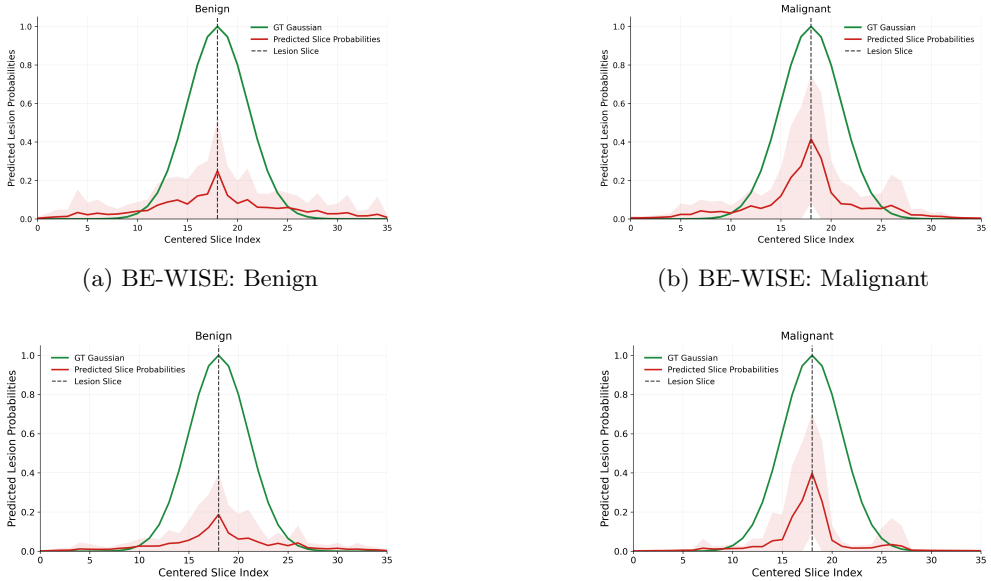
Fig. 4: Misclassified but localized benign and malignant examples. Although the breast-level prediction is normal, the positive-class probability still peaks near the radiologist-indicated lesion slice.

weak Gaussian targets supports the use of slice-semantic supervision as an interpretable localization signal. Figure 4 highlights cases that were misclassified at the breast level but still showed localized slice-level activations near the radiologist-indicated lesion slice. This behavior is clinically relevant because it suggests that the model may retain useful spatial evidence even when the breast-level prediction is incorrect. Such cases support the role of BE-WISER as an assistive system in which slice-level outputs can guide radiologist review rather than replace clinical decision-making.

Figure 5 compares the average slice probability profiles of BE-WISE and BE-WISER for benign and malignant test examinations after centering each case on the radiologist-indicated lesion slice. For visualization, the Gaussian weak targets were aligned and normalized before averaging, providing a common lesion-centered reference across cases. In all settings, predicted probabilities increase around the target region, indicating that both models learn a meaningful spatial prior from weak slice supervision. Secondary peaks in the averaged profiles may reflect cases with multifocal or spatially extended lesions, where clinically relevant enhancement is present beyond the single annotated representative slice. BE-WISE shows broader and more irregular off-target responses, consistent with noisier binary slice-level predictions [6]. In contrast, BE-WISER produces smoother and more concentrated profiles with reduced peripheral activation, suggesting that three-class slice-semantic supervision improves the stability and spatial coherence of lesion localization.

5 Discussion

This study shows that weakly supervised breast MRI classification benefits from moving beyond binary lesion localization toward more class-defined slice semantics. BE-WISER extends BE-WISE by assigning normal, benign, and malignant labels at the slice level, thereby linking local lesion evidence more directly to the breast-level diagnosis. Under matched preprocessing, optimization, and evaluation settings,



(a) BE-WISE: Benign

(b) BE-WISE: Malignant

(c) BE-WISER: Benign

(d) BE-WISER: Malignant

Fig. 5: Average benign and malignant slice probability profiles after centering on the radiologist-indicated lesion slice. Panels (a,b) show BE-WISE and panels (c,d) show BE-WISER. BE-WISER produces smoother, more lesion-centered profiles with reduced off-target fluctuations, especially in malignant cases.

this class-aware formulation improved both AUC and ODELIA score over breast-level baselines and BE-WISE. The localization analysis further supports this finding: BE-WISER reduced the distance to radiologist-indicated lesion slices and improved Hit@K performance, suggesting stronger spatial agreement with clinically relevant lesion regions. The reduced-label experiments indicate that BE-WISER can also improve annotation efficiency. With only 50% of radiologist-provided slice annotations, BE-WISER still retained competitive performance. Adding the remaining cases with breast-level labels improved the ODELIA score, and pseudo weak slice supervision further narrowed the gap to full slice-level guidance. This suggests that BE-WISER can generate useful lesion-centered pseudo annotations for positive cases, which is clinically relevant because dense lesion masks are costly and single-slice annotations may not always be available at scale. Qualitative results provide clinically important context beyond aggregate metrics. Correctly localized benign and malignant cases showed probability peaks near radiologist-indicated lesion regions, while some breast-level misclassifications still retained localized slice-level activation close to the annotated lesion slice. This indicates that classification and localization errors may not always coincide. In practice, such slice-level outputs could support radiologist review by highlighting suspicious regions, but should be interpreted as decision-support evidence rather than autonomous diagnostic output.

These findings should be interpreted in light of several methodological considerations. Slice-level supervision was derived from a single representative lesion slice, providing weak localization rather than exact location of the lesion. This low-burden annotation strategy may not fully capture multifocal, irregularly shaped mass enhancements or non-mass enhancing lesions. Gaussian target generation further approximates lesion spread along the slice axis and may oversimplify heterogeneous enhancement patterns. In addition, pseudo weak slice annotations may inherit errors from the stage-one model. External validation on other datasets and reader studies are therefore needed to assess generalization, localization reliability, and clinical utility.

Overall, BE-WISER provides a scalable compromise between breast-level supervision and dense lesion annotation. By introducing class-aware weak slice-level lesion localization, it improves diagnostic performance while preserving clinically interpretable slice-level evidence for breast MRI assessment.

6 Conclusion

We introduced BE-WISER, a weakly supervised breast MRI framework that extends binary slice-aware learning to class-aware slice-semantic supervision. By modeling normal, benign, and malignant evidence at the slice level, BE-WISER improved breast-level diagnosis and slice-level lesion localization compared with breast-level baselines and the previous BE-WISE formulation. In the reduced-label setting, pseudo weak slice supervision further improved performance when only 50% of radiologist-provided slice annotations were retained. These findings suggest that pathology-aware slice semantics provide an effective and annotation-efficient strategy for weakly supervised breast MRI analysis, supporting scalable diagnostic modeling with clinically interpretable slice-level evidence.

Acknowledgements The authors thank the ODELIA Breast MRI dataset organizers [12] and acknowledge the scientific support and HPC resources provided by NHR@FAU. The hardware is funded by the German Research Foundation (DFG).

Conflict of interest The authors declare no conflict of interest.

References

[1] Sung, H., *et al.*: 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) https://doi.org/10.3322/caac.21660

[2] Boyd, N.F., *et al.*: Mammographic density and the risk and detection of breast cancer. New England journal of medicine **356**(3), 227–236 (2007)

[3] Bakker, M.F., *et al.*: Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. New England Journal of Medicine **381**(22), 2091–2102 (2019) https://doi.org/10.1056/NEJMoa1903986

[4] Mann, R.M., *et al.*: Breast MRI: State of the Art. Radiology **292**(3), 520–536 (2019) https://doi.org/10.1148/radiol.2019182947

[5] Sickles, E.A., *et al.*: ACR BI-RADS Atlas: Breast Imaging Reporting and Data System, 5th edn. American College of Radiology, Reston, VA (2013)

[6] Panambur, A.B., *et al.*: Breast MRI evaluation with weakly-informed slice-level explanation. In: Bildverarbeitung Für die Medizin 2026. Informatik Aktuell, pp. 10–17. Springer, Wiesbaden, Germany (2026). https://doi.org/10.1007/978-3-658-51100-5_2

[7] Ilse, M., *et al.*: Attention-based deep multiple instance learning. In: Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136 (2018). https://proceedings.mlr.press/v80/ilse18a.html

[8] Lu, M.Y., *et al.*: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021) https://doi.org/10.1038/s41551-020-00682-w

[9] Liu, Z., *et al.*: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10012–10022 (2021). https://doi.org/10.1109/ICCV48922.2021.00986

[10] Müller-Franzes, G., *et al.*: Medical slice transformer for improved diagnosis and explainability on 3D medical images with DINOv2. Scientific Reports **15**(1), 23979 (2025) https://doi.org/10.1038/s41598-025-09041-8

[11] Oviedo, F., *et al.*: Cancer detection in breast MRI screening via explainable AI anomaly detection. Radiology **316**(1), 241629 (2025) https://doi.org/10.1148/radiol.241629

[12] Escudero Sánchez, L., *et al.*: ODELIA breast MRI challenge 2025. In: Med Image Comput Comput Assist Interv (MICCAI) 2025 (2025). https://doi.org/10.5281/zenodo.15075570 . https://doi.org/10.5281/zenodo.15075570

[13] Rokuss, M., et al.: Divide and conquer: A large-scale dataset and model for left–right breast MRI segmentation. arXiv preprint arXiv:2507.13830 (2025)