

Reconsidering Explicit Longitudinal Mammography Alignment for Enhanced Breast Cancer Risk Prediction

Solveig Thrun¹[0009–0006–7349–9449], Stine Hansen²[0000–0002–0962–6160], Zijun Sun³[0009–0004–1848–6966], Nele Blum⁴[0009–0008–2622–181X], Suaiba A. Salahuddin¹[0000–0003–4407–0222], Kristoffer Wickstrøm¹[0000–0003–1395–7154], Elisabeth Wetzter¹[0000–0002–0544–8272], Robert Jenssen^{1,5,6}[0000–0002–7496–8474], Maik Stille⁴[0009–0009–7451–0779], and Michael Kampffmeyer^{1,5}[0000–0002–7699–0405]

¹ Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway solveig.thrun@uit.no

² SPKI The Norwegian Centre for Clinical Artificial Intelligence, University Hospital of North Norway, Tromsø, Norway

³ Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

⁴ Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering IMTE, Lübeck, Germany

⁵ Norwegian Computing Center, Oslo, Norway

⁶ Pioneer Centre for AI, University of Copenhagen, Copenhagen, Denmark

Abstract. Regular mammography screening is essential for early breast cancer detection and deep learning-based risk prediction methods have sparked interest to adjust screening intervals for high-risk groups. While early methods focused only on current mammograms, recent approaches leverage the temporal aspect of screenings to track breast tissue changes over time, requiring spatial alignment across different time points. Two main strategies for this have emerged: explicit feature alignment through deformable registration and implicit learned alignment using techniques like transformers, with the former providing more control over the alignment. However, the optimal approach for explicit alignment in mammography remains underexplored. In this study, we provide insights into where explicit alignment should occur (input space vs. representation space) and if alignment and risk prediction should be jointly optimized. We demonstrate that jointly learning explicit alignment in representation space while optimizing risk estimation performance, as done in the current state-of-the-art approach, results in a trade-off between alignment quality and predictive performance and show that image-level alignment is superior to representation-level alignment, leading to better deformation field quality and enhanced risk prediction accuracy. The code is available at https://github.com/sot176/Longitudinal_Mammogram_Alignment.git.

Keywords: Breast cancer risk prediction · Longitudinal mammogram · Image registration

1 Introduction

Breast cancer is among the most prevalent cancers globally [3] and is primarily detected through regular mammography screening, capturing cranio-caudal (CC) and medio-lateral oblique (MLO) views [2]. However, screening frequency is limited by healthcare capacity, creating a trade-off between early detection and resource availability [8]. A promising solution is developing more accurate risk prediction models for personalized screening, optimizing follow-up intervals based on individual risk. Although current mammograms can be used for this purpose, radiologists typically compare them with prior mammograms to distinguish benign from malignant changes, enhancing diagnostic accuracy [1,23].

Similarly, recent deep learning-based studies have demonstrated that longitudinal mammography has the potential to improve risk prediction, surpassing models based solely on current images [5,15,18,24,25], emphasizing the importance of analyzing temporal tissue changes. However, effective use of longitudinal mammograms requires precise spatial alignment or registration, as misalignment can obscure abnormalities and distort risk assessments. This is challenging due to the non-rigid nature of breast tissue [10] and variations in breast compression during examinations.

While various alignment strategies exist, direct comparisons are scarce, highlighting the need for deeper insights. Alignment methods fall into two categories: explicit [5,20,24] and implicit [15,18]. For longitudinal mammography, explicit alignment is preferred as it directly corrects spatial misalignment, ensuring anatomical consistency over time [4]. This approach preserves crucial information, reduces artifacts, and aids in tracking tissue changes. Therefore, we focus on explicit alignment, in form of affine or deformable registration in either the feature space or input space. Early methods performed registration in the input space before feature extraction using conventional registration techniques [5,20], which is computationally expensive and time-consuming, as it typically requires per-image optimization. Recent state-of-the-art approaches therefore align feature maps directly in the feature space [24]. However, this can degrade performance due to the large receptive field of the encoders and downsampling. Due to the absence of direct comparisons of alignment mechanisms, there is a lack of guidance on explicit alignment strategies for longitudinal mammography.

To address the lack of insights into the choice of alignment, this study aims to evaluate explicit alignment strategies for longitudinal mammography and their impact on risk prediction. We examine whether alignment should occur at the image or feature level, demonstrating that feature alignment may not be optimal despite its use in current state-of-the-art approaches. We further demonstrate a clear objective function mismatch for feature-level approaches, where the risk prediction and quality alignment objectives typically disagree. Further, to facilitate efficient image level alignment, we take inspiration from recent advances in registration and transfer them to the mammography domain, introducing MammoRegNet, which to our knowledge is the first deep learning model for high-quality longitudinal mammogram registration.

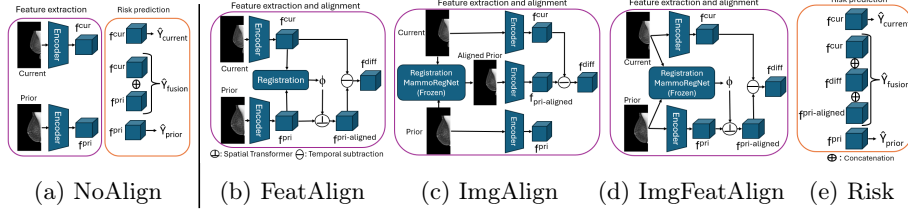


Fig. 1: Overview of longitudinal risk prediction methods: (a) Direct feature extraction without alignment, (b) Feature-level alignment, (c) Image-level alignment with MammoRegNet, (d) Applying MammoRegNet’s deformation field in feature space, and (e) Risk prediction using alignment methods (b), (c), and (d).

Our **main contributions** are:

1. We conduct a comprehensive evaluation of explicit alignment strategies and their effect on risk prediction performance using the publicly available EM-BED [14] and CSAW-CC [6] datasets.
2. We provide insights into the trade-offs between feature-level alignment quality and risk prediction performance.
3. We demonstrate that image-level alignment enhances both deformation field quality and risk prediction performance.
4. We introduce MammoRegNet, the first deep learning-based image registration model for aligning longitudinal mammography images.

2 Methods

Risk prediction estimates the probability of developing breast cancer within a specified time frame, focusing on a 5-year period following a specific mammogram. This process requires the alignment of longitudinal mammography images. Given a current image $\mathbf{I}^{\text{cur}} \in \mathbb{R}^{H \times W}$ or a feature map $\mathbf{f}^{\text{cur}} \in \mathbb{R}^{C \times H \times W}$ and a prior image $\mathbf{I}^{\text{pri}} \in \mathbb{R}^{H \times W}$ or a feature map $\mathbf{f}^{\text{pri}} \in \mathbb{R}^{C \times H \times W}$, the goal of explicit alignment is to determine a deformation field $\phi \in \mathbb{R}^{2 \times H \times W}$ that maximizes similarity between the transformed prior and current images (or feature maps). Effective alignment is crucial for precise anatomical registration and a smooth deformation field. To assess its impact on risk prediction, we compare different alignment strategies within a consistent framework: Fig. 1: (a) No explicit alignment (Sec. 2.1), (b) feature-level alignment (Sec. 2.2), as done in the current state-of-the-art approach [24], enhances detection of subtle patterns by focusing on high-level semantics, though it may miss important spatial relationships, and (c) image-level alignment (Sec. 2.3), retaining spatial and textural details crucial for identifying subtle risk indicators. This study examines how image-level and feature-level alignment strategies affect risk prediction in mammography.

2.1 Risk Prediction Baseline

Our risk prediction baseline (Fig. 1a) builds on [24] and [26], incorporating multilevel joint learning and a cumulative probability layer for time-dependent risk estimation. It uses an image encoder to generate two feature maps: current \mathbf{f}^{cur} and prior \mathbf{f}^{pri} . The model predicts breast cancer risk at three levels: (1) Fused Feature Prediction, using concatenated unaligned prior and current features; (2) Current Feature Prediction, using only current features; and (3) Prior Feature Prediction, using unaligned prior features. Following [24], each level outputs a six-dimensional risk vector: the first five elements estimate cancer occurrence over the next five years, and the last indicates the probability of remaining cancer-free. Only fused feature predictions determine the final risk assessment. The risk loss function employs binary cross-entropy at each prediction level, with masking to focus on observed time points and ignore unobserved future data.

2.2 Feature-Level Alignment

To evaluate feature-level alignment, we follow the current state-of-the-art approach [24], using two 3×3 convolutional layers, batch normalization [13], and ReLU activation [9]. This block processes the concatenated feature maps to estimate a deformation field ϕ . A spatial transformer block applies this deformation to align the prior feature maps $\mathbf{f}^{\text{pri-aligned}}$. The alignment is trained using $\mathcal{L}_{\text{feat}}$. To further encourage smooth alignment, as commonly done in image-level alignment approaches, we further constrain their loss with a (optional) regularization of the displacement field and a Jacobian Determinant (JD) loss to penalize negative Jacobian Determinants resulting in

$$\mathcal{L}_{\text{feat}} = \alpha \left(\|\mathbf{f}^{\text{pri-aligned}} - \mathbf{f}^{\text{cur}}\|_2^2 \right) + \beta \left(\sum_{\mathbf{p} \in \Omega} \|\nabla \phi(\mathbf{p})\|^2 + \lambda \text{JD}(\phi) \right), \quad (1)$$

where α and β are weighting factors, Ω represents the set of spatial locations \mathbf{p} , ∇ denotes the spatial derivative of ϕ at \mathbf{p} , and λ is the regularization parameter. Fig. 1c illustrates feature-level alignment, while Fig. 1e shows risk prediction using the aligned feature maps. This approach enhances the risk prediction baseline by incorporating four key feature maps, as in [24]: current \mathbf{f}^{cur} , prior \mathbf{f}^{pri} , aligned prior $\mathbf{f}^{\text{pri-aligned}}$, and difference features $\mathbf{f}^{\text{diff}} = \mathbf{f}^{\text{cur}} - \mathbf{f}^{\text{pri-aligned}}$, which capture temporal changes. Following [26], positional encoding is applied to \mathbf{f}^{diff} to account for time gaps between screenings. While Current and Prior Feature predictions remain unchanged from the risk prediction baseline, Fused Feature Prediction integrates aligned prior, current, and difference features.

2.3 Image-Level Alignment

Traditional registration methods for longitudinal mammograms [5,20] are often computationally expensive, requiring per-image optimization, and are less accurate than deep learning approaches. As an alternative to feature-level methods,

we develop MammoRegNet, a deep learning-based registration network adapted from the Non-Iterative Coarse-to-Fine Transformer (NICE-Trans) network [21]. Originally designed for 3D brain MRI registration, NICE-Trans was modified for 2D mammography by replacing 3D convolutions with 2D ones and converting the Swin Transformer and Spatial Transformer blocks to 2D. MammoRegNet features an encoder with two parallel paths for current and prior images, each with five convolutional modules. The concatenated features are processed by a decoder with four SwinTrans modules and a convolutional module. Each SwinTrans module includes a convolutional layer and four SwinTrans blocks [19], with patch-expanding blocks to enhance spatial resolution. The decoder performs registration in a coarse-to-fine manner, starting with affine registration followed by four deformable refinements. Following [21], MammoRegNet is trained using

$$\mathcal{L}_{\text{img}} = (1 - \text{NCC}_{\text{affine}}) + (1 - \text{NCC}_{\text{final}}) + \gamma \left(\sum_{\mathbf{p} \in \Omega} \|\nabla \phi(\mathbf{p})\|^2 + \lambda \text{JD}(\phi) \right). \quad (2)$$

This unsupervised loss function includes image similarity measured by the normalized cross-correlation (NCC) between the current mammogram and both the affine- and final-transformed prior mammograms and a regularization term. Fig. 1c shows the image-level alignment pipeline, with risk prediction in Fig. 1e. We also evaluate applying MammoRegNet’s deformation field in feature space (Fig. 1d), with risk prediction outlined in Fig. 1e.

3 Experiments and Results

3.1 Experimental Setting

Dataset and Pre-processing This study uses two large, publicly available mammography datasets to ensure reproducibility and encourage further research. The Emory Breast Imaging Dataset (EMBED)¹ [14] comprises 364,564 mammograms from 23,256 patients, collected from 2013 to 2020 using Hologic, GE, and Fujifilm systems. The Cohort of Screen-Aged Women Case Control (CSAW-CC)² [6] comprises images from 1,103 breast cancer diagnoses and 7,850 healthy controls, collected from 2008 to 2015 at Karolinska University Hospital, Sweden. Following [24], for the risk prediction dataset we included patients with at least five years of follow-up data for risk prediction. Images were resized to 512×1024 pixels while maintaining aspect ratio, and were divided into training, validation, and test sets (5:2:3 ratio). For registration training, we randomly selected 1,000 patients from the risk dataset, using two images per laterality and view combination to reduce training time and split it into training, validation, and test sets (5:2:3 ratio).

¹ <https://aws.amazon.com/marketplace/pp/prodview-unw4li5rkivs2#overview>

² <https://snd.se/en/catalogue/dataset/2021-204-1>

Table 1: 1–5 Year Breast Cancer Risk Prediction of different alignment methods: C-index and AUC with 95% CI in parentheses.

		C-index (%) \uparrow	Follow-up year AUC (%) \uparrow				
			1-year	2-year	3-year	4-year	5-year
EMBED	NoAlign	63.6 (61.2-66.0)	63.9 (61.0-66.8)	64.1 (61.5-66.7)	64.1 (61.7-66.6)	64.7 (62.1-67.2)	59.8 (55.8-63.8)
	FeatAlign	64.9 (62.4-67.3)	64.2 (61.1-67.2)	64.5 (61.8-67.1)	65.4 (62.9-67.7)	65.3 (62.8-67.7)	65.8 (61.7-69.4)
	FeatAlignReg	64.1 (61.6-66.6)	63.9 (60.8-66.9)	64.8 (62.0-67.5)	64.6 (62.1-67.0)	63.5 (61.0-65.8)	60.6 (56.7-64.4)
	ImgFeatAlign	66.1 (63.6-68.5)	65.9 (63.0-68.6)	66.2 (63.4-68.7)	67.0 (64.5-69.5)	65.9 (63.4-68.5)	64.2 (60.3-68.1)
	ImgAlign	66.9 (64.6-69.3)	67.4 (64.4-70.4)	67.5 (64.9-70.2)	68.1 (64.9-70.6)	66.6 (64.0-69.2)	65.6 (61.6-69.4)
CSAW-CC	NoAlign	54.1 (51.9-56.1)	54.2 (51.5-56.9)	55.5 (53.4-57.7)	55.6 (53.5-57.8)	54.7 (52.5-56.9)	55.7 (53.2-58.2)
	FeatAlign	57.0 (54.8-58.9)	57.3 (54.5-60.1)	58.2 (55.8-60.2)	58.5 (56.2-60.6)	58.5 (56.2-60.6)	60.5 (58.0-63.0)
	FeatAlignReg	56.0 (53.8-57.9)	56.3 (53.5-59.0)	57.3 (55.3-59.5)	57.8 (55.7-60.1)	57.6 (55.5-59.9)	60.4 (57.8-62.9)
	ImgFeatAlign	60.6 (58.7-62.5)	61.8 (59.2-64.5)	61.9 (59.8-63.9)	62.7 (60.7-64.8)	61.3 (59.2-63.4)	63.9 (61.4-66.5)
	ImgAlign	61.3 (59.3-63.4)	62.1 (59.3-64.9)	62.4 (60.4-64.4)	63.8 (61.7-65.9)	64.0 (61.9-66.1)	66.2 (63.7-68.7)

Evaluation Metrics Following [21], we evaluate registration performance using NCC to assess image-wise similarity and measure deformation field quality for both image-level and feature-level alignment by the percentage of Negative Jacobian Determinants (NJD) [11]. Risk prediction performance is assessed using the concordance index (C-index) and Area Under the Curve (AUC) for the 1–5 year risk prediction task, as in [18,24,26]. 95% confidence intervals (CI) for NCC, NJD, AUC, and C-index are estimated via bootstrapping with 1,000 iterations.

Implementation Details We use a ResNet-18 [12], pre-trained on ImageNet [7], without the last fully connected layer. For feature-level alignment, risk prediction and alignment are jointly trained using a combined loss with $\alpha = 1e - 2$ in Eq. 1 as in [24]. For image-level alignment, only the risk loss is used, with the registration model frozen. We use the Adam optimizer [17] (learning rate = $1e - 4$, weight decay = $1e - 5$), a batch size of 12, and 100 epochs. Early stopping and learning rate decay are based on the validation C-index: the learning rate halves after 5 stagnant epochs, and training stops after 15. For MammoRegNet, we use the Adam optimizer (learning rate = $1e - 4$, weight decay = $1e - 6$), a batch size of 20, and 100 epochs, with 1,500 image pairs randomly selected per epoch. λ is set to $1e - 5$, and γ to 1 in Eq. 2. All models were implemented in PyTorch and trained on an NVIDIA GeForce RTX 3090, taking between 14 and 72 hours.

3.2 Results

Effect of Feature- vs. Image-Level Alignment on Risk Prediction We assess the impact of feature-level and image-level alignment on risk prediction

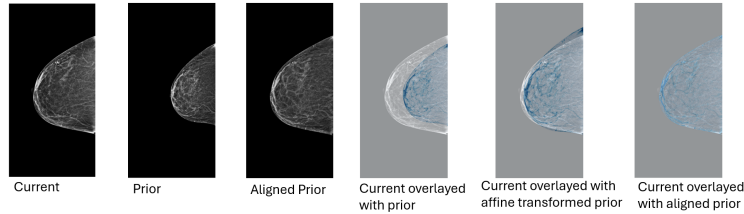


Fig. 2: Registration result of MammoRegNet.

Table 2: MammoRegNet registration performance (95% CI in parentheses).

	NCC before (%) \uparrow	NCC affine (%) \uparrow	NCC final (%) \uparrow	NJD (%) \downarrow
EMBED	74.5	83.1 (82.8-83.4)	95.5 (95.5-95.6)	0.0016 (0.0013-0.0020)
CSAW-CC	78.3	85.2 (84.8-85.7)	95.8 (95.5-96.0)	0.0129 (0.0088-0.0180)

performance (Table 1) by comparing NoAlign (Fig. 1a), FeatAlign (Fig. 1b) with $\beta = 0$, FeatRegAlign (Fig. 1b) with $\beta = 0.2$, ImgFeatAlign (Fig. 1d), and ImgAlign (Fig. 1c). Our results demonstrate that image-level deformation enhances risk prediction, with ImgFeatAlign outperforming feature-level methods and ImgAlign achieving the highest accuracy. Both datasets show consistent results, suggesting that aligning images before feature extraction better preserves context and enhances feature representativeness, underscoring the benefits of image-level alignment for robust risk prediction. Table 2 presents the performance of MammoRegNet and Fig. 2 shows a registration result. MammoRegNet achieves precise alignment across datasets and deformation scales, effectively handling both large and small deformations, thereby enhancing the accuracy and reliability of longitudinal analyses by correcting spatial inconsistencies.

Comparison of Deformation Field Quality While downstream risk prediction performance is one way to assess the quality of the alignment procedures, current approaches within the mammography domain often overlook the quality of the deformation fields, potentially providing an inaccurate picture of how the alignment mechanism benefits the overall performance. We therefore assess the deformation field quality both quantitatively and qualitatively, focusing on displacement vectors and negative Jacobian determinants. FeatAlign methods exhibit significant local deformations (Fig. 3a) and the highest percentage of negative Jacobian determinants, reaching 1.6273. Introducing regularization in FeatAlignReg reduces these local deformations (Fig. 3b) and the percentage of negative Jacobian determinants to 0.0025. However, this comes at the cost of a slight risk prediction performance drop (see Table 1). MammoRegNet, employed in ImgFeatAlign and ImgAlign, further enhances global smoothness by producing a continuous transformation (Fig. 3c), resulting in the fewest negative Jacobian determinants at 0.0016. These findings underscore the benefits

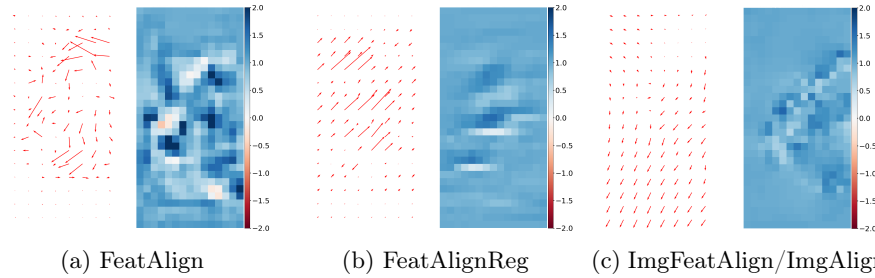


Fig. 3: Comparison of deformation field quality. Left: displacement vectors. Right: Jacobian determinants (white/blue: valid, red: invalid).

of input-level alignment in improving both risk prediction performance and the fidelity of learned deformation fields.

In-Depth Analysis of Weighting and Alignment Our findings, along with those in [24], show that feature-level approaches benefit from jointly optimizing risk prediction and alignment. Training them separately reduced risk prediction performance, resulting in a C-index of 62.2, compared to 64.9 (C-index, EMBED dataset). We further explore how different weightings in joint training affect performance and evaluate feature-level and image-level alignment across breast density categories. This distinction is crucial, as higher breast density increases cancer risk and complicates lesion detection. Unlike the EMBED dataset, which includes BI-RADS density categories [22], the CSAW-CC dataset provides breast percent density estimates from the LIBRA software [16], which we categorized into three equally sized groups, corresponding to low, medium and high density levels. Fig. 4a and 4b display the C-index by density category and alignment method for CSAW-CC and EMBED, respectively. ImgAlign performs best, especially in high-density cases, surpassing FeatAlign and FeatAlignReg, highlighting the advantages of image-level alignment for dense breast tissue. Further, to investigate the trade-off between risk prediction performance and deformation field quality, we evaluate the effect of the alignment weight α in Eq. 1 on the EMBED dataset in Fig. 4c and 4d. Increasing the alignment weight improves deformation field quality by reducing negative Jacobian determinants but decreases risk prediction performance emphasizing that these objectives are not necessarily well aligned.

4 Conclusion and Outlook

In this study, we provide insights into explicit alignment methods for longitudinal mammograms and their effects on risk prediction performance. Our findings indicate that image-level alignment not only improves the deformation field quality but also enhances risk prediction accuracy and that jointly training feature

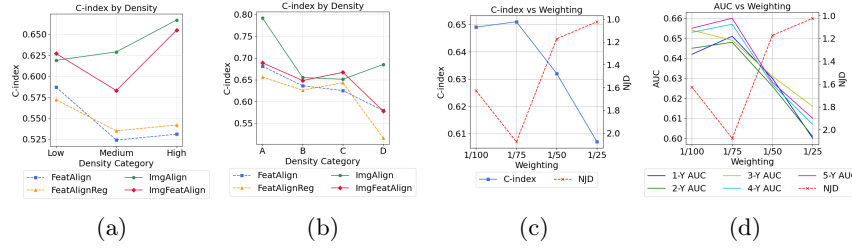


Fig. 4: Analysis of Weighting and Alignment: (a) C-index by density category (CSAW-CC), (b) C-index by density category (EMBED), (c) C-index by weighting (FeatAlign, EMBED), (d) AUC by weighting (FeatAlign, EMBED).

alignment and risk prediction results in a trade-off between alignment quality and risk prediction performance. Further, to facilitate efficient image-based alignment, we introduce MammoRegNet, the first deep learning-based model for aligning longitudinal mammograms, paving the way towards more accurate risk prediction models. In future work, we aim to further build on this framework that facilitates comparisons between alignment approaches, extending it to settings where implicit alignment approaches are considered and multiple prior images.

Acknowledgments. This work was supported by The Research Council of Norway (Visual Intelligence, grant no. 309439 as well as FRIPRO grant no. 315029 and IKT-PLUSS grant no. 303514).

Disclosure of Interests. The authors declare no competing interests.

References

1. Akwo, J., Trieu, P.Y., Barron, M., Reynolds, T., Lewis, S.: Access to prior screening mammograms affects the specificity but not sensitivity of radiologists’ performance. *Clinical Radiology* **79**(12) (2024)
2. Bhushan, A., Gonsalves, A., Menon, J.U.: Current state of breast cancer diagnosis, treatment, and theranostics. *Pharmaceutics* **13**(5) (2021)
3. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **74**(3), 229–263 (2024)
4. Chen, M., Tustison, N.J., Jena, R., Gee, J.C.: *Image Registration: Fundamentals and Recent Advances Based on Deep Learning*. Springer (2023)
5. Dadsetan, S., Arefan, D., Berg, W.A., Zuley, M.L., Sumkin, J.H., Wu, S.: Deep learning of longitudinal mammogram examinations for breast cancer risk prediction. *Pattern Recognition* **132** (2022)
6. Dembrower, K., Lindholm, P., Strand, F.: A multi-million mammography image dataset and population-based screening cohort for the training and evaluation of deep neural networks—the cohort of screen-aged women (CSAW). *Journal of Digital Imaging* **33**(2) (2020)

7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
8. Elkin, E.B., Snow, J.G., Leoce, N.M., Atoria, C.L., Schrag, D.: Mammography capacity and appointment wait times: barriers to breast cancer screening. *Cancer Causes & Control* **23**, 45–50 (2012)
9. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (2011)
10. Guo, Y., Sivaramakrishna, R., Lu, C.C., Suri, J.S., Laxminarayan, S.: Breast image registration techniques: a survey. *Medical and Biological Engineering and Computing* **44** (2006)
11. Ha, I.Y., Wilms, M., Heinrich, M.: Semantically guided large deformation estimation with deep networks. *Sensors* **20**(5) (2020)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)
13. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on International Conference on Machine Learning (2015)
14. Jeong, J.J., Vey, B.L., Bhimireddy, A., Kim, T., Santos, T., Correa, R., Dutt, R., Mosunjac, M., Oprea-Ilie, G., Smith, G., Woo, M., McAdams, C.R., Newell, M.S., Banerjee, I., Gichoya, J., Trivedi, H.: The EMory BrEast imaging Dataset (EM-BED): A racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images. *Radiology: Artificial Intelligence* **5**(1) (2023)
15. Karaman, B.K., Dodelzon, K., Akar, G.B., Sabuncu, M.R.: Longitudinal mammogram risk prediction. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer-Verlag (2024)
16. Keller, B.M., Nathan, D.L., Wang, Y., Zheng, Y., Gee, J.C., Conant, E.F., Kontos, D.: Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. *Medical Physics* **39**(8) (2012)
17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2017)
18. Lee, H., Kim, J., Park, E., Kim, M., Kim, T., Kooi, T.: Enhancing breast cancer risk prediction by incorporating prior images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Springer-Verlag (2023)
19. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows . In: IEEE International Conference on Computer Vision. IEEE Computer Society (2021)
20. Loizidou, K., Skouroumouni, G., Nikolaou, C., Pitris, C.: Automatic breast mass segmentation and classification using subtraction of temporally sequential digital mammograms. *IEEE Journal of Translational Engineering in Health and Medicine* **10** (2022)
21. Meng, M., Bi, L., Fulham, M., Feng, D., Kim, J.: Non-iterative coarse-to-fine transformer networks for joint affine and deformable image registration. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Springer-Verlag (2023)
22. Sickles, E.A.: ACR BI-RADS® atlas, breast imaging reporting and data system. American College of Radiology. (2013)

23. Varela, C., Karssemeijer, N., Hendriks, J.H., Holland, R.: Use of prior mammograms in the classification of benign and malignant masses. *European Journal of Radiology* **56**(2) (2005)
24. Wang, X., Tan, T., Gao, Y., Marcus, E., Han, L., Portaluri, A., Zhang, T., Lu, C., Liang, X., Beets-Tan, R., Teuwen, J., Mann, R.: Ordinal Learning: Longitudinal Attention Alignment Model for Predicting Time to Future Breast Cancer Events from Mammograms . In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. Springer-Verlag (2024)
25. Wang, X., Tan, T., Gao, Y., Su, R., Zhang, T., Han, L., Teuwen, J., D’Angelo, A., Drukker, C.A., Schmidt, M.K., Beets-Tan, R., Karssemeijer, N., Mann, R.: Predicting up to 10 year breast cancer risk using longitudinal mammographic screening history. *medRxiv* (2023)
26. Yala, A., Mikhael, P.G., Strand, F., Lin, G., Smith, K., Wan, Y.L., Lamb, L., Hughes, K., Lehman, C., Barzilay, R.: Toward robust mammography-based models for breast cancer risk. *Science Translational Medicine* **13**(578) (2021)