

The Impact of Longitudinal Mammogram Alignment on Breast Cancer Risk Assessment

Solveig Thrun^{a,*}, Stine Hansen^b, Zijun Sun^c, Nele Blum^d, Suaiba A. Salahuddin^a, Xin Wang^e,
Kristoffer Wickstrøm^a, Elisabeth Wetzter^a, Robert Jenssen^{a,f,g}, Maik Stille^d, Michael
Kampffmeyer^{a,f}

^a*Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, 9037, Norway*

^b*SPKI The Norwegian Centre for Clinical Artificial Intelligence, University Hospital of North Norway, Tromsø, 9019, Norway*

^c*Department of Computer Science and Engineering, University of Bologna, 40126, Bologna, Italy*

^d*Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering IMTE, 23562, Lübeck, Germany*

^e*Department of Radiology, Netherlands Cancer Institute (NKI), Amsterdam, 1066 CX, The Netherlands*

^f*Norwegian Computing Center, Oslo, 0373, Norway*

^g*Pioneer Centre for AI, University of Copenhagen, Copenhagen, 1350, Denmark*

Abstract

Regular mammography screening is crucial for early breast cancer detection. By leveraging deep learning-based risk models, screening intervals can be personalized, especially for high-risk individuals. While recent methods increasingly incorporate longitudinal information from prior mammograms, accurate spatial alignment across time points remains a key challenge. Misalignment can obscure meaningful tissue changes and degrade model performance. In this study, we provide insights into various alignment strategies, image-based registration, feature-level (representation space) alignment with and without regularization, and implicit alignment methods, for their effectiveness in longitudinal deep learning-based risk modeling. Using two large-scale mammography datasets, we assess each method across key metrics, including predictive accuracy, precision, recall, and deformation field quality.

Our results show that image-based registration consistently outperforms the more recently favored feature-based and implicit approaches across all metrics, enabling more accurate, temporally consistent predictions and generating smooth, anatomically plausible deformation fields. Although regularizing the deformation field improves deformation quality, it reduces the risk prediction performance of feature-level alignment. Applying image-based deformation fields within the feature space yields the best risk prediction performance.

These findings underscore the importance of image-based deformation fields for spatial alignment in longitudinal risk modeling, offering improved prediction accuracy and robustness. This approach has strong potential to enhance personalized screening and enable earlier interventions for high-risk individuals. The code is available at https://github.com/sot176/Mammogram_Alignment_Study_Risk_Prediction.git, allowing full reproducibility of the results.

Keywords: Breast Cancer Risk Prediction, Longitudinal Image Registration, Mammography Alignment, Deformation Field Evaluation

1. Introduction

Breast cancer remains the most prevalent cancer type among women worldwide, posing a significant public health challenge across diverse populations (Bray et al., 2024). Mammography is the current gold standard for breast cancer screening and is typically performed in two standard views, craniocaudal (CC) and mediolateral oblique (MLO), to provide comprehensive visualization of breast tissue (Bhushan et al., 2021). Widespread implementation of regular screening programs has led to a substantial reduction in breast cancer mortality, primarily through early detection and intervention (Zielonke et al., 2020). However, despite this progress, there remains significant room for improvement, particularly for individuals at higher risk (Rubio et al., 2024).

One of the key challenges in the detection of breast cancer is breast density, which not only increases the risk of developing cancer, but also reduces the sensitivity of mammography (Bodewes et al., 2022). Dense breast tissue can obscure malignant lesions, complicating early detection and increasing the likelihood of missed diagnoses during routine screening (Wang et al., 2014). As a result, women with high breast density face a dual burden: an increased risk of cancer and a reduced screening effectiveness. Although an increase in screening frequency could mitigate these risks, such measures are often constrained by the capacity of the healthcare system (Elkin et al., 2012), creating a trade-off between the benefits of early detection and the practical limitations of routine imaging. These challenges underscore the need for a more personalized approach to screening. In this context, personalized risk prediction models offer a promising solution, enabling tailored screening intervals that improve results while optimizing resource allocation (Eriksson et al., 2023).

Recent advances in deep learning have shown considerable promise in improving breast cancer risk prediction using mammography images (Hussain et al., 2024). However, while radiologists routinely compare current mammograms with previous exams to distinguish between benign and malignant changes and improve diagnostic precision (Akwo et al., 2024; Varela et al., 2005), early deep learning models primarily analyzed current images in isolation, overlooking valuable temporal information (Yala et al., 2021; Donnelly et al., 2024). Incorporating longitudinal data, such as prior screenings, can capture subtle tissue changes over time, which are often critical for accurate risk assessment (Jiang et al., 2023). This highlights the growing recognition that the temporal context is essential for robust breast cancer risk prediction, especially in cases where changes evolve gradually over multiple years.

Based on this insight, recent studies have shown that models that incorporate longitudinal mammography outperform those based solely on single-time-point images (Lee et al., 2023; Wang et al., 2023; Dadsetan et al., 2022; Karaman et al., 2024; Wang et al., 2024; Sun et al., 2025). By integrating both prior and current mammograms, these models more effectively capture temporal dynamics, thereby improving the accuracy of risk assessments. However, precise alignment of mammograms across time points is critical for reliably evaluating tissue changes. Misalignment can obscure or distort relevant features, leading to inaccurate predictions. This challenge is compounded by the non-rigid nature of breast tissue (Guo et al., 2006) and variations in breast positioning and compression during image acquisition. To address these issues, image registration, both through explicit and implicit techniques, plays a vital role in aligning sequential mammograms and improving the performance of longitudinal risk models.

*Corresponding author

Email address: solveig.thrun@uit.no (Solveig Thrun)

Various alignment strategies have been proposed for mammography, yet it remains unclear which approach is most effective for downstream tasks such as breast cancer risk prediction. Direct comparisons between these methods are scarce, and the relationship between alignment quality and predictive performance has not been systematically explored. In general, alignment methods can be categorized as explicit or implicit. Explicit alignment can be applied at the image level (input space) or the feature level (representation space), offering anatomically precise transformations, but potentially introducing artifacts or interpolation errors. In contrast, implicit alignment is learned jointly with feature extraction and may yield more robust representations, though often at the expense of anatomical precision. While each approach has its advantages and limitations, their comparative impact on predictive performance remains underexplored. Notably, there are currently no deep learning-based registration models developed for explicit longitudinal alignment of mammograms directly in the image space. Traditional registration methods, while effective, are often computationally intensive and time-consuming. To overcome these limitations, we introduce MammoRegNet, a deep learning model specifically designed for accurate and efficient registration of longitudinal mammograms, leveraging advances in explicit alignment techniques from other imaging domains.

This work presents the first comprehensive evaluation of both alignment paradigms, analyzing their trade-offs and exploring whether a hybrid strategy can optimize spatial alignment and predictive accuracy. Importantly, all alignment approaches are evaluated within a unified risk prediction framework to ensure a consistent and fair comparison of their impact on breast cancer risk prediction.

The main contributions of this study are summarized as follows:

- We present the first systematic evaluation of explicit (image-based and feature-based) and implicit alignment strategies for longitudinal breast cancer risk prediction, within a unified risk prediction framework to ensure consistency across comparisons.
- We demonstrate that spatial alignment, especially through image-based methods, consistently improves predictive accuracy and robustness over time.
- We show that enhancing deformation quality via regularization in feature alignment reduces risk prediction performance, revealing a key trade-off.

A preliminary version of this work appeared in (Thrun et al., 2025). In this extended study, we significantly enhance our previous work through the following key improvements:

1. We expand the scope of the analysis to include implicit alignment methods, enabling a more comprehensive comparison across alignment strategies. Our earlier work focused exclusively on explicit approaches.
2. We adopt higher-resolution mammography images (1664×2048 vs. 512×1024) to align with recent state-of-the-art studies (Yala et al., 2021; Donnelly et al., 2024; Karaman et al., 2024). Although image-based alignment previously showed superior performance at lower resolutions, we find that applying image-based deformation fields in feature space yields improved results at higher resolutions, likely due to richer and more semantically meaningful representations.
3. We enhance model performance by incorporating a temporal self-attention block that effectively captures dependencies across time between mammographic feature representations.

4. We validate our findings using the OA-BreaCR framework (Wang et al., 2024), demonstrating consistency of alignment effects across different longitudinal risk prediction models.
5. We include a more extensive literature review to better situate our work within the broader research landscape.
6. We provide a substantially expanded methodology section, with detailed descriptions of alignment strategies, risk prediction models, and a thorough evaluation of their effects on predictive performance.

The paper is structured as follows. Section 2 provides an overview of relevant related work, with a focus on breast cancer risk prediction and image registration. Section 3 presents a detailed description of the risk prediction pipeline incorporating various alignment methods. In Section 4, we describe the experimental setup and the evaluation metrics used to assess the performance of the different alignment strategies. Sections 5 and 6 present and discuss the results, respectively. Finally, Section 7 summarizes the key findings of the study and provides a future outlook.

2. Related Work

Deep learning has substantially advanced the fields of breast cancer risk prediction and image registration, driving improvements in early detection and longitudinal analysis. In the following, we first review recent developments in deep learning-based breast cancer risk prediction (Section 2.1), with an emphasis on models that incorporate temporal information across multiple screenings. We then discuss deep learning approaches to medical image registration (Section 2.2), highlighting emerging efforts to adapt these methods to the unique challenges of mammography, particularly in longitudinal settings.

2.1. Breast Cancer Risk Prediction

Breast cancer risk prediction is a well-established field, with early models relying on radiomic-based characteristics such as mammographic density and parenchymal texture extracted through hand-crafted algorithms (Anandarajah et al., 2022; Keller et al., 2012; Tan et al., 2016). Recent advances have increasingly focused on deep learning approaches that automatically learn predictive patterns from mammograms (Yala et al., 2021; Guo et al., 2006; Lee et al., 2023; Wang et al., 2024; Karaman et al., 2024; Yala et al., 2019; Dadsetan et al., 2022; Michel et al., 2023). Current models typically fall into two categories: single-timepoint models and those that incorporate temporal information from sequential screenings. They may use imaging data alone or combine them with clinical risk factors, such as age, family history, or hormonal status. In both types, extracted features, whether made by hand or learned, tend to focus on critical indicators such as tissue density, mass morphology, and textural characteristics (Damiani et al., 2023; Hussain et al., 2024).

Mirai (Yala et al., 2021) is a deep learning model designed for breast cancer risk prediction at multiple future timepoints, using standard mammographic views and, when available, non-imaging risk factors. To improve interpretability, AsymMirai (Donnelly et al., 2024) captures bilateral asymmetries in mammograms to approximate Mirai’s predictions while providing more transparent and explainable outputs.

A key limitation of many existing models, including Mirai and AsymMirai, is their reliance on a single mammographic timepoint, which ignores the temporal dynamics of breast tissue

evolution. Recently, several approaches have emerged that incorporate longitudinal imaging data to improve predictive accuracy. These methods follow two strategies: explicit alignment of images across timepoints or implicit modeling of temporal changes without direct spatial alignment.

Some approaches rely on recurrent neural networks, with an Long Short-Term Memory (LSTM)-based (Hochreiter and Schmidhuber, 1997) model capturing temporal dependencies across exams and highlights the importance of modeling change patterns for effective risk stratification (Li et al., 2023). Other methods employ transformer-based (Vaswani et al., 2017) architectures. PRIME+ (Lee et al., 2023) integrates prior mammograms using a transformer decoder and outperforms single-timepoint models in both short- and long-term risk prediction. Similarly, TRINet (Yeoh et al., 2024) combines sequential mammograms with time-aware attention mechanisms to predict short-term risk. Most recently, LoMaR (Karaman et al., 2024) extended Mirai by introducing a transformer-based model that processes multiple prior mammograms, implicitly modeling temporal changes without relying on non-imaging risk factors. Similarly, VMRA-MaR (Sun et al., 2025) presents a framework for predicting breast cancer risk that takes advantage of both longitudinal imaging data and bilateral asymmetry in mammograms.

While many models incorporate temporal information through feature fusion or sequential modeling, some take a more explicit approach by aligning either mammographic images or extracted feature maps to better capture localized changes over time. LRP-NET (Dadsetan et al., 2022) tracks spatiotemporal breast tissue changes across four prior exams by applying affine image registration before training. The results demonstrate the benefits of spatial alignment in longitudinal modeling, showing improved performance over single-timepoint baselines. In contrast, OA-BreaCR (Wang et al., 2024) spatially aligns feature maps extracted from sequential mammograms using an attention-based mechanism. By modeling temporal changes at the feature level and incorporating ordinal learning to account for time-to-event outcomes, OA-BreaCR enhances risk assessment while maintaining interpretability, supporting more personalized screening strategies.

Together, these studies highlight the growing importance of temporal modeling in the prediction of breast cancer risk, since sequential mammograms reveal dynamic patterns that are often missed by single-timepoint analysis. However, the alignment of temporal information, whether at the image or feature level, can significantly impact the model’s ability to capture meaningful changes. Therefore, a systematic evaluation of different alignment strategies is essential to understand their influence on downstream risk prediction performance. Although prior work has demonstrated the benefits of incorporating longitudinal information into breast cancer risk models, the role of alignment has remained underexplored, particularly in the context of mammography. Our work addresses this gap by providing insights into various alignment methods, including image-level registration, feature-level alignment, and implicit approaches, and evaluating their effects on risk prediction performance.

2.2. Registration

Image registration plays a critical role in longitudinal medical imaging by enabling the spatial alignment of scans acquired at different time points. In the context of breast cancer screening, explicit alignment, whether at the image or feature level, is essential for accurately tracking tissue changes across visits, such as evolving asymmetries or subtle structural shifts, which may signal early disease development (Im et al., 2024). By improving spatial correspondence over time, explicit registration enhances the interpretability and reliability of risk prediction models that leverage sequential imaging data. However, most deep learning-based image registration

methods have primarily been developed and evaluated on volumetric imaging modalities such as magnetic resonance imaging (MRI) and computed tomography (CT), with a strong focus on anatomical regions like the brain and abdomen (Chen et al., 2025).

Among these, convolutional neural network (CNN)-based models have been foundational in medical image registration. VoxelMorph (Balakrishnan et al., 2019) pioneered deep learning-based deformable registration by learning spatial transformations directly from image pairs in an unsupervised manner. Building on this foundation, models such as AMNet (Che et al., 2023) and PC-Reg (Yin et al., 2023) introduced architectural innovations to enhance performance, AMNet through its adaptive multi-level design, and PC-Reg with a pyramidal prediction–correction approach to handle large deformations. More recently, transformer-based architectures like TransMorph (Chen et al., 2022) and TransMatch (Chen et al., 2024) have been proposed for unsupervised deformable registration, with the latter incorporating a dual-stream feature matching strategy. NICE-Trans (Meng et al., 2023) further advances this trend by introducing a non-iterative coarse-to-fine transformer network that jointly learns both affine and deformable transformations, enabling efficient and accurate alignment without multi-stage optimization.

While deep learning-based image registration has made significant strides in volumetric imaging domains, its application to mammography remains relatively limited. A few studies have explored registering the two standard mammographic views, CC and MLO, within a single screening episode using CNN-based methods to enhance feature correspondence (Walton et al., 2022; Li et al., 2020). In contrast, longitudinal mammography registration, which involves aligning images from the same view acquired at different timepoints, has largely relied on conventional image processing techniques. Traditional registration algorithms have, for instance, been used to subtract prior images from recent ones, thereby improving the visibility of new abnormalities (Loizidou et al., 2021, 2022).

Although conventional registration methods have been applied to longitudinal mammography to align images from different timepoints, these approaches are often time-consuming and computationally expensive. We present MammoRegNet, a deep learning-based image registration approach tailored for longitudinal mammography. Building on recent advances in image registration for brain MRI, MammoRegNet efficiently aligns mammograms from the same view (e.g. CC or MLO) acquired at different timepoints. This enables high-quality spatial correspondence across longitudinal exams. By applying MammoRegNet, we provide new insights into how precise image-level alignment over time can enhance risk model performance, highlighting the potential of deep learning-based registration to improve breast cancer screening outcomes.

3. Methods

This section outlines our unified framework that facilitates the analysis of different alignment methods and their impact on risk prediction performance. We start by describing the core tasks in Section 3.1, and then introduce the baseline risk prediction model in Section 3.2 followed by the implicit alignment approach in Section 3.3. Finally, Section 3.4 details two explicit alignment strategies: feature-level alignment (Section 3.4.1) and image-level alignment (Section 3.4.2).

3.1. Problem definition

The risk prediction task involves estimating the probability that an individual will develop breast cancer within a specified time frame. In this study, we focus on predicting the likelihood of breast cancer occurrence within a five-year period following a given mammogram. Accurate

alignment is essential in this context, as it ensures consistent localization of anatomical structures over time and facilitates the interpretation of temporal changes. This consistency enables the model to better capture subtle longitudinal patterns, such as progressive tissue changes or emerging abnormalities, which may signal an increased risk of cancer.

To effectively leverage temporal information from longitudinal mammography scans, it is crucial to model how breast tissue evolves over time. One approach is implicit alignment, which relies on deep learning models to learn temporal dependencies directly from the data, without explicitly aligning the images or their corresponding features. These models capture the underlying temporal relationships between scans by learning patterns of progression or change across time points. However, to provide a more structured temporal integration, we can also incorporate explicit alignment, where the goal is to align a current mammogram $\mathbf{I}^{\text{cur}} \in \mathbb{R}^{H \times W}$ or its corresponding feature map $\mathbf{f}^{\text{cur}} \in \mathbb{R}^{C \times h \times w}$ with a prior mammogram $\mathbf{I}^{\text{pri}} \in \mathbb{R}^{H \times W}$ or its feature map $\mathbf{f}^{\text{pri}} \in \mathbb{R}^{C \times h \times w}$. Here, H and W denote the spatial resolution of the input images, while $h < H$ and $w < W$ indicate the reduced spatial resolution of the feature maps due to downsampling in the backbone network. C denotes the number of feature channels. In this case, we aim to compute a deformation field $\phi \in \mathbb{R}^{2 \times H \times W}$ that minimizes a cost function E , which measures the dissimilarity between the current image and the spatially transformed prior image. The formulation is given by the following optimization problem

$$\widehat{\phi} = \arg \min_{\phi} E(\mathbf{I}^{\text{cur}}, \mathbf{I}^{\text{pri}} \circ \phi) + \lambda R(\phi). \quad (1)$$

Here, E quantifies the image dissimilarity, while $R(\phi)$ is a regularization term that enforces desirable properties such as spatial smoothness and anatomical plausibility. The hyperparameter λ balances image similarity with regularization to prevent unrealistic deformations (e.g., folding). This formulation can also be extended to feature-level alignment, where feature maps are registered by minimizing the same cost function.

To evaluate the impact of alignment on risk prediction, we compare four strategies within a unified framework (Figure 1):

- No alignment (Section 3.2): Serves as a baseline where no temporal registration is performed between mammograms.
- Implicit alignment (Section 3.3): Learns spatial correspondences as part of the risk prediction process without an explicit registration step.
- Feature-level alignment (Section 3.4.1): Inspired by recent state-of-the-art methods (Wang et al., 2024), this approach aligns feature maps to enhance semantic consistency but may lose critical spatial information.
- Image-level alignment (Section 3.4.2): Operates directly on pixel space, preserving spatial and textural details important for detecting subtle longitudinal changes.

This setup enables a systematic assessment of how different alignment strategies influence the performance of longitudinal breast cancer risk prediction.

3.2. Risk Prediction Baseline

Our no alignment risk prediction baseline (Figure 1a) builds on previous works (Yala et al., 2021; Wang et al., 2024), integrating three key components: multilevel joint learning, a Temporal Self-Attention Block, and a Cumulative Probability Layer for time-dependent risk estimation. Leveraging a shared backbone allows us to evaluate the impact of different alignment

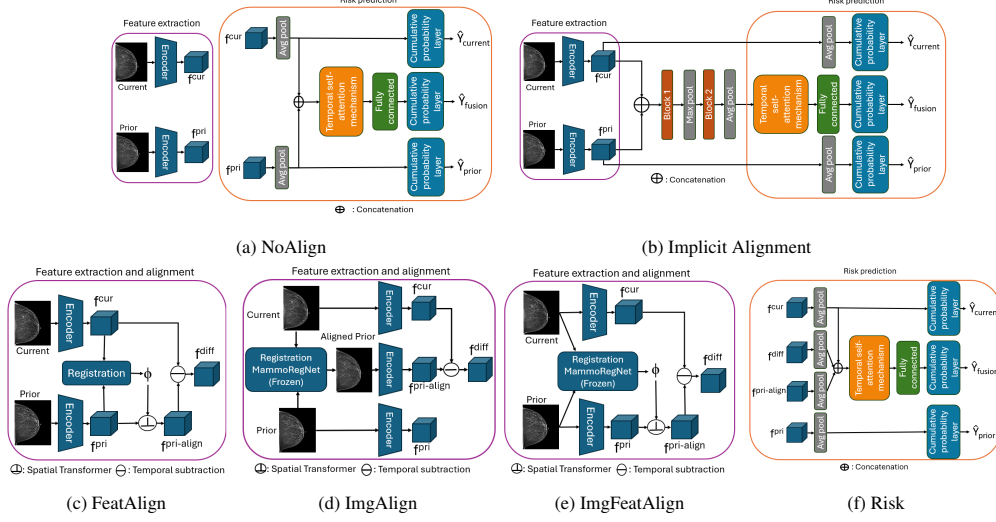


Figure 1: Overview of longitudinal risk prediction methods: (a) Direct feature extraction without alignment, (b) Implicit Alignment, (c) Feature-level alignment, (d) Image-level alignment with MammoRegNet, (e) Applying MammoRegNet’s deformation field in feature space, and (f) Risk prediction using alignment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

mechanisms on this baseline to assess their influence on risk prediction performance. A detailed visualization of the architecture can be found in the Appendix in Figure A.11a.

Multilevel Joint Learning: The model uses multilevel joint learning (Wang et al., 2024) to capture temporal and feature-level dependencies between current and prior mammographic images. By processing feature representations from both current and prior images, the model effectively learns to track risk progression over time.

Feature Extraction and Risk Prediction: The model uses a shared image encoder to extract feature maps from both current and prior mammograms, producing two feature maps: current f^{cur} and prior f^{pri} . Breast cancer risk is predicted at three levels:

- **Fused Feature Prediction:** Combines unaligned prior and current features, processed by temporal self-attention mechanism.
- **Current Feature Prediction:** Relies solely on current features.
- **Prior Feature Prediction:** Utilizes unaligned prior features.

Each prediction head is passed through a cumulative probability layer (Yala et al., 2021), producing a five-dimensional risk vector, each element representing the estimated occurrence of cancer in the next five years.

Temporal Self-Attention Block: Building on prior research that employs Transformer encoders and attention layers (Klanecek et al., 2024; Lee et al., 2023), the temporal self-attention block (Li et al., 2023) captures temporal dependencies between current and prior mammographic feature representations. It leverages a multi-head self-attention mechanism (Vaswani et al., 2017) to emphasize significant changes across time points by attending to temporal sequences of features. This mechanism enables the model to capture informative temporal changes, which helps to identify patterns that can signal an increased risk of breast cancer. Additionally, the block

includes a feedforward network and residual connections, which allow the model to focus on features predicting future risk while maintaining both stability and expressiveness. Overall, this architecture improves the model’s ability to detect subtle longitudinal changes, which are essential for identifying evolving risk indicators over time.

Cumulative Probability Layer: The cumulative probability layer (Yala et al., 2021; Mikhael et al., 2023; Lee et al., 2023) estimates the cumulative probability of an event over time, a formulation commonly used in survival analysis and time-to-event modeling. It consists of two fully connected layers: one that maps feature representations to hazard values at each time step, and another that computes a baseline hazard. A ReLU activation (Glorot et al., 2011) ensures the hazard values remain non-negative. The final cumulative probability at time t is obtained by summing the hazards from previous time steps, along with the baseline hazard, yielding a continuous cumulative risk representation over time, according to

$$P(t | \mathbf{f}) = h_{\text{base}}(\mathbf{f}) + \sum_{j=1}^t h_j(\mathbf{f}), \quad (2)$$

where $P(t | \mathbf{f})$ is the cumulative probability at time t , given the input feature map $\mathbf{f}$, $h_{\text{base}}(\mathbf{f})$ is the baseline hazard, and $h_j(\mathbf{f})$ is the hazard value at each time step j . This formulation allows the model to capture the sequential accumulation of risk over time while preserving temporal dependencies.

Loss function and Censoring: The model is trained for time-dependent risk estimation using binary cross-entropy (BCE) loss. The ground-truth $\mathbf{y}$ is a binary vector of length T_{max} , where T_{max} represents the maximum observation period (5 years). For each time step t , $\mathbf{y}(t)$ is 1 if the patient is diagnosed with cancer within t years and 0 otherwise. However, real-world clinical data often contain censored observations (i.e., patients lost to follow-up before an event occurs). To address this, we follow (Yala et al., 2021; Wang et al., 2024) and use a masking function $\delta(t)$ to ensure the loss computation considers only valid observations. The masking function is defined as follows

$$\delta(t) = \begin{cases} 1, & \text{if the follow-up period} \geq \min(t, T_{\text{max}}) \\ & \text{or if the patient is diagnosed within } n \text{ years.} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

This masking mechanism ensures that only valid observations contribute to the loss calculation, preventing bias due to incomplete follow-up periods.

Thus, the overall binary cross-entropy loss $\mathcal{L}_{\text{BCE}}$ is computed as

$$\mathcal{L}_{\text{BCE}} = \sum_{t=1}^{T_{\text{max}}} \delta(t) \cdot \ell_{\text{BCE}}(t), \quad (4)$$

where $\ell_{\text{BCE}}(t)$ represents the binary cross-entropy loss at time step t and is given by

$$\ell_{\text{BCE}}(t) = - [\mathbf{y}(t) \cdot \log(P(t)) + (1 - \mathbf{y}(t)) \cdot \log(1 - P(t))] , \quad (5)$$

with $\mathbf{y}(t)$ being the ground truth, and $P(t)$ being the predicted probability of cancer diagnosis at time t . This formulation ensures that loss contributions come only from valid, observed time points.

3.3. Implicit Alignment

The implicit alignment pipeline (Figure 1b) utilizes an encoder to extract feature representations from both the current and prior images. These extracted features are concatenated and passed through a convolutional feature extractor consisting of two convolutional blocks. A detailed visualization of the architecture is provided in the Appendix in Figure A.11b. Each block includes a convolutional layer with a 3×3 kernel, followed by batch normalization and ReLU activation. A max-pooling operation is applied after the first block to reduce the spatial resolution. Before being processed by the temporal self-attention block, global average pooling is applied to further reduce the spatial dimensions. The temporal self-attention mechanism enables the model to capture dependencies over time and detect subtle changes between the current and prior images. The attention-enhanced features are then passed through a cumulative probability layer to estimate the cancer risk over a five-year horizon. Consistent with our baseline risk prediction setup, additional cumulative probability layers are applied separately to the current and prior feature representations.

3.4. Explicit Alignment

In this section, we explore explicit alignment strategies for improving risk prediction, focusing on two distinct approaches: feature-level alignment and image-level alignment. Both methods aim to enhance the model’s ability to capture critical temporal changes, but they do so at different levels of representation, each offering unique advantages.

3.4.1. Feature-Level Alignment

The feature registration model builds upon the current state-of-the-art approach (Wang et al., 2024), which focuses on feature-level alignment between current and prior feature maps through the estimation and application of a dense deformation field. Figure 1c illustrates feature-level alignment, while Figure 1f shows risk prediction using the aligned feature maps. The model consists of two main components: an Alignment Block and a Spatial Transformer Block, inspired by previous works such as (Balakrishnan et al., 2019; Meng et al., 2023).

Alignment Block: The Alignment Block, adapted from (Wang et al., 2024), processes the concatenated current and prior feature maps through two convolutional layers with a kernel size of 3×3 , followed by batch normalization (Ioffe and Szegedy, 2015) and ReLU activation. This block is responsible for predicting a dense deformation field ϕ . The predicted deformation field is then applied to the prior feature map $\mathbf{f}^{\text{pri}}$ using the Spatial Transformer Block, producing the aligned prior feature map $\mathbf{f}^{\text{pri-align}}$.

Spatial Transformer Block: The Spatial Transformer Block (Balakrishnan et al., 2019; Meng et al., 2023) uses bilinear sampling to deform the prior feature map according to the predicted deformation field. This ensures a smooth, differentiable alignment between the current and prior feature maps.

Loss Function and Regularization: In contrast to the current state-of-the-art approach (Wang et al., 2024), we introduce an additional regularization loss, as commonly done in image-level alignment approaches, on top of the feature alignment loss. Specifically, we apply a L_2 regularization on the displacement field and a Jacobian Determinant (JD) loss to penalize negative Jacobian Determinants. These regularization terms encourage smoother and more anatomically plausible deformations, helping to further improve alignment accuracy, which was not part of the current state-of-the-art approach. The general feature alignment loss function $\mathcal{L}_{\text{feat}}$ is defined as

$$\mathcal{L}_{\text{feat}} = \alpha \left(\|\mathbf{f}^{\text{pri-align}} - \mathbf{f}^{\text{cur}}\|_2^2 \right) + \beta \left(\sum_{\mathbf{p} \in \Omega} \|\nabla \phi(\mathbf{p})\|^2 + \lambda \text{JD}(\phi) \right), \quad (6)$$

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. The regularization term is composed of two components: L_2 regularization of the displacement field, $\sum_{\mathbf{p} \in \Omega} \|\nabla \phi(\mathbf{p})\|^2$, and Jacobian determinant loss, $\text{JD}(\phi)$. This regularization encourages the model to learn a smooth and physically plausible transformation, rather than simply minimizing the differences at the feature level. The regularization parameter λ controls the trade-off between alignment accuracy and the physical plausibility of the transformation.

This approach improves the risk prediction baseline by incorporating four key feature maps, as in (Wang et al., 2024): current $\mathbf{f}^{\text{cur}}$, prior $\mathbf{f}^{\text{pri}}$, aligned prior $\mathbf{f}^{\text{pri-align}}$, and difference features.

Temporal Information Encoding: To capture meaningful temporal variations, we adopt the approach from (Loizidou et al., 2021; Wang et al., 2024) by generating difference features through temporal subtraction $\mathbf{f}^{\text{diff}} = \mathbf{f}^{\text{cur}} - \mathbf{f}^{\text{pri-align}}$, capturing meaningful temporal variations. As in (Wang et al., 2023), positional encoding is applied to $\mathbf{f}^{\text{diff}}$ to account for the varying time intervals between consecutive screenings. The positional encoding, which uses sinusoidal-based embeddings generated by applying sine- and cosine functions with logarithmically scaled frequencies, captures the temporal differences between the current and prior feature maps. The time gap between screenings is normalized and interpolated, mapping continuous time values to discrete position embeddings. Dropout is applied for regularization, and the resulting positional encodings are added to the input difference features.

Risk Prediction with Aligned Features: A detailed visualization of the risk prediction architecture using alignment at the feature level is provided in the Appendix (Figure A.11c). In this model, risk prediction is again performed at three levels: Current Feature Prediction, Prior Feature Prediction, and Fused Feature Prediction. While the Current and Prior Feature Predictions remain unchanged from the baseline, the Fused Feature Prediction involves concatenating the aligned prior features $\mathbf{f}^{\text{pri-align}}$, the current features $\mathbf{f}^{\text{cur}}$, and the difference features $\mathbf{f}^{\text{diff}}$. The inclusion of difference features and positional encoding enhances the model’s ability to incorporate temporal information, ultimately improving its performance in predicting breast cancer risk.

3.4.2. Image-Level Alignment

Traditional registration methods for longitudinal mammograms (Loizidou et al., 2022; Dadsetan et al., 2022) are often computationally expensive, require per-image optimization, and are less accurate compared to deep learning-based approaches. As an alternative to feature-level alignment, we propose MammoRegNet, a deep learning-based registration network inspired by the Non-Iterative Coarse-to-Fine Transformer (NICE-Trans) model (Meng et al., 2023). Figure 2 shows the architecture of MammoRegNet. Originally designed for 3D brain MRI registration, NICE-Trans was adapted for 2D mammography by replacing 3D convolutions with 2D operations and adjusting the Swin Transformer and Spatial Transformer blocks to work in the 2D domain.

MammoRegNet: MammoRegNet performs both affine and deformable registration within a single unified network, leveraging an encoder-decoder structure.

The encoder consists of two parallel paths, each designed to independently learn features from the current (fixed) and prior (moving) mammograms. Both paths consist of five convolu-

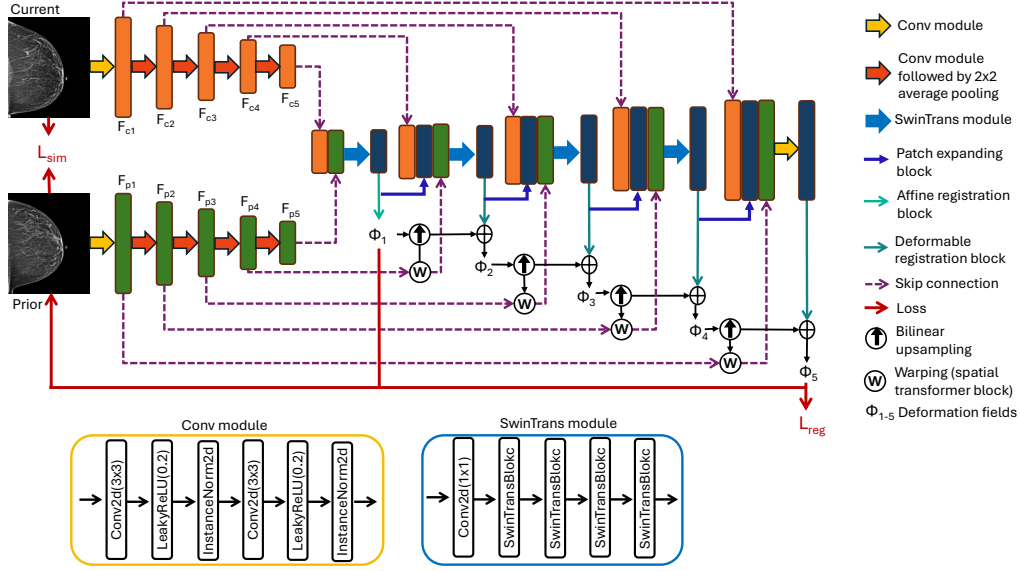


Figure 2: Architecture of the proposed MammoRegNet.

tional modules with shared weights. Each convolutional module follows a consistent structure: a convolutional layer with a kernel size of 3×3 , followed by a Leaky ReLU activation function with a negative slope of 0.2, an instance normalization layer, another convolutional layer with a kernel size of 3×3 , a second Leaky ReLU activation with a slope of 0.2, and a final instance normalization layer. To progressively reduce spatial resolution while preserving essential features, average pooling with a 2×2 kernel is applied after the second, third, fourth, and fifth convolutional modules.

The features of the current and prior mammograms are then concatenated and passed to the decoder, which comprises four SwinTrans modules and one convolutional module. Each SwinTrans module includes a convolutional layer and four SwinTrans blocks (Liu et al., 2021), with patch-expanding blocks between adjacent modules to enhance spatial resolution. The decoder performs the registration in a coarse-to-fine manner consisting of five steps, starting with an affine registration followed by four steps of deformable refinements. At each step, the output of the decoder module is input to either an affine or deformable registration block, which maps the features to a displacement field. This iterative process results in progressively finer registration with each decoder step, effectively enabling joint affine and deformable registration.

Loss Function: As in (Meng et al., 2023), MammoRegNet is trained using an unsupervised loss function that consists of two main components: an image similarity term and a regularization term. The image similarity is measured using normalized cross-correlation (NCC) between the current mammogram and the transformed prior mammograms at both the affine and final stages of transformation. The loss function is formulated as

$$\mathcal{L}_{\text{img}} = (1 - \text{NCC}_{\text{affine}}) + (1 - \text{NCC}_{\text{final}}) + \gamma \left(\sum_{\mathbf{p} \in \Omega} \|\nabla \phi(\mathbf{p})\|^2 + \lambda \text{JD}(\phi) \right), \quad (7)$$

where γ is a weighting factor, Ω represents the set of spatial locations $\mathbf{p}$, ∇ denotes the spatial

derivative of ϕ at $\mathbf{p}$, and λ is the regularization parameter.

The first term in the loss function, $(1 - \text{NCC}_{\text{affine}})$, encourages alignment between the affine-transformed prior mammogram and the current mammogram. The second term, $(1 - \text{NCC}_{\text{final}})$, ensures that the final transformation closely aligns the previous mammogram with the current one. The regularization term is similar to the regularization in Equation 6 and aims to prevent physically implausible deformations by encouraging smoothness in the transformation.

Temporal Alignment and Risk Prediction: Figures 1d and 1e present two alignment strategies integrated into our unified risk prediction framework. More detailed visualizations of these strategies are provided in Appendix A.11d and A.11e.

In the image-level alignment setup (Figure 1d), the current, prior, and aligned prior mammograms are first encoded to extract features. Temporal subtraction is then performed to obtain difference features. All resulting feature maps are passed to the risk prediction module, following a process similar to that of feature-level alignment.

In the image-based feature alignment strategy (Figure 1e), we further explore applying MammoRegNet’s deformation field directly within the feature space, rather than at the image level.

The risk prediction module for both alignment strategies is summarized in Figure 1f.

4. Experimental Setup

In this section, we present a comprehensive overview of the experimental setup used to train and evaluate our models. We begin by describing the datasets (Section 4.1), followed by a detailed explanation of how cancer and cancer-free patients are selected, and how the time-to-cancer is calculated for each mammogram (Section 4.2). Next, we outline the pre-processing steps applied to the mammography images (Section 4.3) and the evaluation metrics used to assess model performance (Section 4.4). Finally, we provide an overview of the implementation details for the models and training processes (Section 4.5).

4.1. Datasets

In this study, we evaluate the performance of our various risk prediction methods using two large, publicly available mammography datasets. This approach ensures reproducibility and encourages further research in the field.

EMory BrEast Imaging Dataset (EMBED)¹: The EMBED (Jeong et al., 2023) dataset is a comprehensive and racially diverse mammography dataset that includes 3.4 million screening and diagnostic images collected from 110,000 patients between 2013 and 2020, using Hologic, GE, and Fujifilm systems. The publicly released portion of the dataset represents 20% of the total 2D and C-view collection, comprising 364,564 mammograms from 23,256 patients. In addition to mammography images, EMBED includes clinical information, such as BI-RADS tissue density scores (Sickles, 2013).

Cohort of Screen-Aged Women Case Control (CSAW-CC)²: The CSAW-CC (Dembrower et al., 2020) dataset is another mammography dataset specifically designed to support AI research aimed at improving breast cancer screening, diagnostics, and prognostics. It consists of mammography images from breast cancer screenings conducted at Karolinska University Hospital in Stockholm, Sweden, between November 2008 and December 2015. All images were captured

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

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

using Hologic systems. The dataset includes 873 cases of first-time breast cancer diagnosed among women aged 40–74 years, along with a random selection of 7,850 healthy controls from the same time period. In addition to mammography images, the CSAW-CC dataset provides clinical information, including breast percentage density estimates generated using LIBRA software (Keller et al., 2012).

For both datasets, we included only patients with at least five years of follow-up data.

4.2. Label Generation

In this section, we describe the process used to generate labels for the Time-to-Cancer variable, which represents the time interval between the date of mammogram screening and the date of cancer diagnosis.

Following the approach in OA-BreaCR (Wang et al., 2024), positive cases in the EMBED dataset were identified based on the BI-RADS score and the pathological severity score. Specifically, positive cases included mammograms with a BI-RADS score of 6 (indicating known biopsy-proven cancer) or a pathological severity score of 0 or 1 (corresponding to invasive and non-invasive cancers). Negative cases were selected from mammograms with BI-RADS scores of 1 or 2 (indicating negative or benign findings), as well as images of BI-RADS 0 that were later reclassified as BI-RADS 1 or 2 in subsequent diagnostic studies.

The Time-to-Cancer label for each patient was calculated by determining the temporal interval between the date of the mammogram and the date of cancer diagnosis. The diagnosis date was identified by grouping the data by patient ID and selecting the most recent exam year among the positive cases. The Time-to-Cancer in years was then computed as the difference between the year of cancer diagnosis and the year of the mammogram examination.

Following the approach in LoMaR (Karaman et al., 2024), positive cases in the CSAW-CC dataset were identified based on radiological timing (`rad_timing`), where images with a `rad_timing` of 1 correspond to screen-detected cancers, and those with a `rad_timing` of 2 correspond to interval cancers. The negative group consisted of images without any recorded `rad_timing`. To determine the cancer year for each patient, the dataset was grouped by patient ID, and the record with the latest exam year was selected. The cancer year was then assigned based on the `rad_timing` of the cancer diagnosis for this exam:

- For screen-detected cancers (`rad_timing` = 1), the cancer year was set to the year of the mammogram exam.
- For interval cancers (`rad_timing` = 2), the cancer year was set to one year after the exam year. However, if the most recent exam year was 2016, the cancer year was adjusted to 2016 to accommodate specific follow-up protocols, as 2016 marked the final year of the CSAW-CC study.

The Time-to-Cancer label was then calculated for each image as the difference between the cancer year and the exam year.

4.3. Pre-processing

The EMBED and CSAW-CC datasets originally consist of images in DICOM format, which are converted into 16-bit grayscale PNG images for consistency and compatibility. Following the pre-processing procedure outlined in (Yala et al., 2021), we enhanced image quality and removed background artifacts—such as textual annotations commonly present in mammography

Table 1: Time-to-cancer label distribution and dataset splits. “Positive” denotes cases diagnosed within a given number of years. “Negative” includes those diagnosed after 5 years (BC >5Y) or never diagnosed.

Dataset	Split	Positive						Negative	Total
		1Y	2Y	3Y	4Y	5Y	All	>5Y	
EMBED	Train	539	190	163	90	67	1049	25076	26125
	Val	216	60	57	39	26	398	9998	10396
	Test	310	141	97	62	46	656	15021	15677
	Total	1065	391	317	191	139	2103	50095	52198
CSAW-CC	Train	960	480	336	324	184	2284	31376	33660
	Val	396	188	160	132	72	948	12376	13324
	Test	568	312	200	204	112	1396	18804	20200
	Total	1924	980	696	660	368	4628	62556	67184

Table 2: Density distribution across the EMBED and CSAW-CC datasets.

Split	EMBED (BI-RADS)				CSAW-CC		
	A	B	C	D	Low	Medium	High
Train	2822	11225	10608	1110	24932	8404	324
Val	1216	4288	4256	540	9755	3460	109
Test	1885	6547	6492	608	15145	4963	92
Total	5923	22060	21356	2258	49832	16827	525

images—by applying a contour detection procedure. All contours in the image are detected, and only the largest contour, corresponding to the breast tissue, is retained. This contour is then used to create a mask that isolates the breast region from the background. The isolated breast image is resized to a size of 1664×2048 pixels while preserving the aspect ratio. Finally, the resized image is normalized and converted to a 16-bit format.

Following (Wang et al., 2024), we included only patients with at least five years of follow-up data for risk prediction. The datasets are randomly split at the patient level into training, validation, and test sets in a 5:2:3 ratio. Table1 presents the distribution of the Time-to-Cancer label and the dataset split for both the EMBED and CSAW-CC datasets.

Unlike the EMBED dataset, which includes BI-RADS density categories, the CSAW-CC dataset provides breast percent density estimates generated using the LIBRA software. These estimates are categorized into three equally sized groups: low, medium, and high density levels. Table 2 shows the density distribution for both the EMBED and CSAW-CC datasets.

For training the registration models (MammoRegNet), we randomly selected 1,000 patients from the risk prediction dataset, using two images per laterality and view combination. This selection method helps minimize training time and accommodate memory limitations. The dataset is then randomly split at the patient level into training, validation, and test sets, following the same 5:2:3 ratio.

4.4. Evaluation Metrics

We evaluate registration performance and risk prediction accuracy using a set of complementary metrics.

Image Registration Metrics: For the evaluation of registration performance, we follow the methods outlined in (Ha et al., 2020; Meng et al., 2023; de Vos et al., 2019).

To assess image-wise similarity, we use the Normalized Cross-Correlation between the fixed image $\mathbf{I}^{\text{cur}}$ and the moving image $\mathbf{I}^{\text{pri}}$, computed over all pixels $\mathbf{p} \in \Omega$, where Ω denotes the image

domain:

$$\text{NCC} = \frac{\sum_{\mathbf{p} \in \Omega} (\mathbf{I}^{\text{cur}}(\mathbf{p}) - \mu_I^{\text{cur}})(\mathbf{I}^{\text{pri}}(\mathbf{p}) - \mu_I^{\text{pri}})}{\sqrt{\sum_{\mathbf{p} \in \Omega} (\mathbf{I}^{\text{cur}}(\mathbf{p}) - \mu_I^{\text{cur}})^2 \sum_{\mathbf{p} \in \Omega} (\mathbf{I}^{\text{pri}}(\mathbf{p}) - \mu_I^{\text{pri}})^2}}, \quad (8)$$

where $\mathbf{I}^{\text{cur}}(\mathbf{p})$ and $\mathbf{I}^{\text{pri}}(\mathbf{p})$ denote the intensity values at pixel $\mathbf{p}$, and μ_I^{cur} and μ_I^{pri} are the mean intensities of $\mathbf{I}^{\text{cur}}$ and $\mathbf{I}^{\text{pri}}$ over Ω , respectively.

To evaluate the quality of the deformation field in both image- and feature-level alignment, we compute the Jacobian determinant, which characterizes local geometric transformations. Let $\mathbf{p} = (x, y) \in \mathbb{R}^2$ denote a spatial coordinate in the image domain where the deformation is evaluated.

The vector displacement field is defined as

$$\boldsymbol{\phi}(\mathbf{p}) = \begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} \in \mathbb{R}^2, \quad (9)$$

where $u(x, y)$ and $v(x, y)$ are horizontal and vertical displacement components, respectively. The Jacobian matrix $\mathbf{J}(\boldsymbol{\phi}) \in \mathbb{R}^{2 \times 2}$ at point $\mathbf{p}$ is defined as

$$\mathbf{J}(\boldsymbol{\phi})(\mathbf{p}) = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}, \quad (10)$$

which contains the partial derivatives describing how $\boldsymbol{\phi}$ changes with respect to $\mathbf{p}$.

The Jacobian determinant given by

$$\det(\mathbf{J}(\boldsymbol{\phi})(\mathbf{p})) = \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} \frac{\partial v}{\partial x}, \quad (11)$$

quantifies the local volume change induced by the deformation at $\mathbf{p}$. Values of $\det(\mathbf{J}(\boldsymbol{\phi})(\mathbf{p})) \leq 0$ indicate foldings or non-invertible regions in the deformation field, which are undesirable in medical image registration. A negative determinant corresponds to a locally non-invertible transformation (e.g., folding), which is anatomically implausible. A value of 1 indicates no volume change, values greater than 1 indicate expansion, and values between 0 and 1 indicate shrinkage. We report the Percentage of Negative Jacobian Determinants (NJD) (Ha et al., 2020), where lower values suggest more plausible deformations. In addition, we computed the standard deviation of the Jacobian determinant to quantify spatial variability throughout the deformation field: Lower values imply smoother and more consistent deformations.

Risk Prediction Metrics: We assess model performance using the concordance index (C-index) (Uno et al., 2011) and the Area Under the Receiver Operating Characteristic Curve (AUC) (Hanley and McNeil, 1982) for predicting breast cancer risk over a 1-5 year period. Both metrics are widely used in medical imaging and survival analysis (Yala et al., 2021; Lee et al., 2023; Wang et al., 2024; Mikhael et al., 2023). The AUC captures overall classification performance, while the C-index evaluates the model's ability to correctly rank risk predictions over time.

To ensure robust performance estimates, 95% confidence intervals (CIs) for all metrics are computed via bootstrapping with 1,000 resamples.

4.5. Implementation Details

For the risk prediction task, we adopt the pre-trained Mirai encoder (Yala et al., 2021) as our backbone. The Mirai encoder is based on a ResNet-18 architecture (He et al., 2016), modified to exclude the global average pooling and final fully connected layers. This modification allows us to retain fine-grained spatial feature representations essential for downstream processing. Importantly, the encoder is kept frozen during training to preserve the pre-trained representations. In the feature-level alignment setting, we jointly optimize the risk prediction and alignment tasks using a combined loss function, with the weighting coefficient $\alpha = 0.1$ (see Equation 6). For image-level alignment, the registration model is frozen, and only the risk prediction loss is computed.

Risk prediction models are trained using the Adam optimizer (Kingma and Ba, 2017) with a learning rate of 1×10^{-5} and weight decay set to 1×10^{-6} . We use a batch size of 20 and train for 40 epochs. To improve generalization and ensure stable convergence, early stopping and learning rate decay are applied based on the validation C-index. The learning rate is halved if no improvement is observed over 5 consecutive epochs, and training is terminated if the validation metric does not improve within 15 epochs. To prevent overfitting, we applied data augmentations (e.g., RandomAffine, ColorJitter, RandomGamma, and RandomCrop) during training.

For the MammoRegNet architecture, we also use the Adam optimizer (Kingma and Ba, 2017) with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-6} . The batch size is set to 6, and training is conducted for 100 epochs. Each epoch consists of 1,500 randomly selected image pairs to ensure robust learning. In the loss function (Equation 7), we use $\lambda = 1 \times 10^{-5}$ and $\gamma = 1$ to balance the contributions of different loss components. No learning rate scheduler is applied for MammoRegNet.

All models are implemented in PyTorch 2.0.1 (Paszke et al., 2019) and trained on an AMD Instinct MI210 with 64 GB of memory. Training times ranged from 8 to 31 hours, depending on the complexity of the model and the size of the data set.

5. Results

In this section, we present both quantitative and qualitative evaluations of our proposed alignment strategies for longitudinal breast cancer risk prediction. Section 5.1 examines the impact of different alignment methods on the overall performance of risk prediction. Section 5.2 provides an in-depth analysis using the EMBED dataset, covering Receiver Operating Characteristic – Area Under the Curve (ROC-AUC) curves (Section 5.2.1), precision scores (Section 5.2.2), recall scores (Section 5.2.3), short-term performance across breast density categories (Section 5.2.4), and long-term performance across density categories (Section 5.2.5). Finally, Section 5.3 compares the quality of the deformation field between the image-based and feature-based registration methods.

5.1. Effect of Different Alignment Methods on Risk Prediction

In this study, we evaluated the impact of various alignment strategies, no alignment, implicit alignment, and explicit alignment at both the feature and image levels, on the performance of breast cancer risk prediction. As part of the image-level alignment, we introduce MammoRegNet, a deep learning-based registration model for longitudinal mammography images. Although this section focuses on risk prediction, detailed results of MammoRegNet registration performance are presented in Appendix B.

Table 3: 1–5 year breast cancer risk prediction using different alignment methods. C-index and AUC values with 95% confidence intervals.

Dataset	Method	C-index (%) $\uparrow$	Follow-up year AUC (%) $\uparrow$				
			1-yr	2-yr	3-yr	4-yr	5-yr
EMBED	NoAlign	64.0 (61.7-66.7)	64.9 (62.1-67.9)	63.8 (61.1-66.5)	63.7 (61.2-66.3)	62.2 (59.6-64.9)	55.7 (51.4-60.0)
	Implicit	70.9 (68.6-73.3)	72.5 (69.3-75.5)	71.2 (68.6-73.7)	69.3 (66.6-71.8)	70.4 (67.9-72.9)	65.7 (62.0-69.7)
	FeatAlign	72.2 (69.5-75.5)	72.4 (69.5-75.6)	72.2 (69.7-74.8)	72.0 (69.7-74.6)	71.4 (69.1-74.0)	68.5 (64.8-72.0)
	FeatAlignReg	70.6 (67.8-73.2)	71.2 (68.3-74.3)	71.3 (68.6-74.0)	70.7 (68.2-73.5)	70.7 (68.3-73.3)	65.7 (61.7-69.6)
	ImgAlign	72.3 (69.6-74.8)	73.6 (70.6-76.5)	71.7 (69.0-74.2)	72.3 (69.8-74.5)	72.4 (70.0-74.7)	69.7 (66.2-73.4)
	ImgFeatAlign	74.7 (72.3-77.0)	75.0 (72.1-77.7)	75.5 (73.1-77.9)	75.3 (73.1-77.4)	75.9 (73.6-78.0)	72.5 (68.9-75.7)
CSAW-CC	NoAlign	65.9 (64.0-67.8)	66.1 (63.8-68.3)	66.2 (64.3-68.2)	65.7 (63.8-67.6)	65.1 (63.2-67.0)	66.8 (64.5-68.9)
	Implicit	67.6 (65.8-69.7)	68.2 (65.7-70.6)	68.5 (66.5-70.5)	68.3 (66.3-70.2)	67.5 (65.4-69.5)	68.7 (66.3-71.1)
	FeatAlign	69.1 (67.0-71.1)	70.1 (67.9-72.4)	69.9 (67.9-71.8)	70.0 (68.1-71.9)	69.5 (67.7-71.4)	71.6 (69.4-73.8)
	FeatAlignReg	68.4 (66.4-70.4)	68.9 (66.7-71.2)	69.2 (67.4-71.1)	69.8 (68.0-71.6)	69.1 (67.3-71.0)	72.0 (69.9-74.2)
	ImgAlign	70.2 (68.1-72.1)	71.2 (68.9-73.4)	71.0 (69.2-73.1)	71.7 (69.9-73.4)	71.1 (69.1-72.9)	73.9 (71.7-76.0)
	ImgFeatAlign	70.4 (68.2-72.3)	72.0 (69.6-74.2)	71.5 (69.6-73.4)	72.6 (70.8-74.5)	72.0 (70.0-74.1)	75.2 (73.1-77.5)

Table 3 summarizes the results of the 1- to 5-year breast cancer risk prediction, reporting both the C-index and the AUC (with 95% confidence intervals) for each alignment approach. These metrics are essential to evaluate the model’s ability to capture temporal dynamics and accurately predict both short-term and long-term risk outcomes. To assess generalizability, we also compare alignment strategies within the OA-BreaCR framework (Wang et al., 2024), as shown in Table 4.

Analysis of the results presented in Table 3 reveals key performance trends across alignment strategies and time intervals. Across both datasets, ImgFeatAlign consistently achieves the highest C-index, reflecting a strong predictive capability. In contrast, NoAlign yields the lowest scores, highlighting the importance of incorporating alignment in longitudinal models. Among all methods, ImgFeatAlign stands out not only for its superior C-index but also for its stable AUC performance across all time horizons. ImgAlign also performs well, but are consistently outperformed by ImgFeatAlign. The Implicit method offers moderate performance but worse than all explicit alignment approaches. NoAlign, meanwhile, shows the steepest AUC decline over time, particularly within the EMBED dataset.

Although both FeatAlign and FeatAlignReg rely on feature-level alignment, their predictive performance differs. FeatAlign employs a feature-based registration model to align features across time and outperforms FeatAlignReg, which introduces a regularization term to enforce smoother deformation fields. This added constraint likely contributes to the performance drop, highlighting a trade-off between deformation field quality and risk prediction accuracy. Specifically, improving deformation quality through regularization leads to reduced predictive performance. These findings are consistent across both datasets.

Notably, alignment methods that incorporate a deformation field from image-based deep learning registration, namely ImgAlign and ImgFeatAlign, consistently outperform all others.

Table 4: 1–5 year breast cancer risk prediction using different alignment methods. C-index and AUC values with 95% confidence intervals. Results are also shown for the OA-BreaCR framework (Wang et al., 2024) to assess consistency across different risk prediction models.

Dataset	Method	C-index (%) $\uparrow$	Follow-up year AUC (%) $\uparrow$				
			1-yr	2-yr	3-yr	4-yr	5-yr
EMBED	FeatAlign	70.4	70.3	69.9	71.0	71.1	70.9
		(68.3-72.5)	(67.1-73.2)	(67.2-72.4)	(68.5-73.4)	(68.9-73.4)	(68.6-73.3)
	FeatAlignReg	67.9	67.1	67.6	67.9	67.1	67.8
		(65.5-70.3)	(63.9-70.4)	(64.8-70.4)	(65.0-70.5)	(64.2-69.8)	(65.0-70.5)
	ImgAlign	71.0	72.1	71.4	71.6	71.9	72.3
		(68.8-73.2)	(69.8-74.8)	(68.6-74.2)	(68.9-74.2)	(69.4-74.2)	(70.0-74.6)
	ImgFeatAlign	72.3	72.9	72.3	72.4	72.3	73.2
		(70.3-74.2)	(70.0-75.7)	(67.4-72.7)	(70.1-74.9)	(70.0-74.7)	(70.8-75.6)
CSAW-CC	FeatAlign	59.1	61.4	59.5	60.9	63.1	63.4
		(57.3-60.9)	(59.0-63.7)	(57.4-61.6)	(58.9-62.9)	(61.1-65.2)	(61.4-65.5)
	FeatAlignReg	58.1	58.0	57.6	58.8	61.0	60.7
		(56.4-59.8)	(55.7-60.2)	(55.6-59.5)	(57.0-60.7)	(59.0-62.9)	(58.7-62.7)
	ImgAlign	60.8	63.3	61.2	63.8	65.7	66.3
		(59.1-62.6)	(60.9-65.6)	(59.2-63.2)	(61.8-65.8)	(63.6-67.6)	(64.2-68.2)
	ImgFeatAlign	61.6	63.5	61.7	65.0	67.0	67.7
		(61.2-65.8)	(59.1-63.7)	(59.6-63.8)	(63.0-66.9)	(65.0-68.9)	(65.6-69.9)

These approaches benefit from spatial deformation fields that enhance temporal alignment, and this added spatial coherence appears to be critical to their success. Among them, ImgFeatAlign achieves the best overall results. Its strong performance is likely due to the integration of image-derived deformation fields directly into the feature space, enabling more accurate spatial-temporal alignment.

Furthermore, ImgFeatAlign exhibits minimal performance degradation over time, indicating its robustness for longitudinal prediction. In general, these results underscore the value of advanced alignment strategies, especially those incorporating spatial or regularization-based constraints, in modeling breast cancer risk over time and improving the reliability of temporal prediction models.

Table 4 presents the results when the same alignment methods are integrated into the OA-BreaCR framework (Wang et al., 2024). We observe consistent trends: ImgAlign and ImgFeatAlign, which incorporate image-based deformation fields from deep learning registration, again outperform alignment approaches that align feature maps solely by minimizing the L2 distance.

In the OA-BreaCR setting, ImgAlign feeds the aligned images directly into the encoder for risk prediction, without applying the attention-based alignment used in our primary framework. In contrast, ImgFeatAlign uses the unaligned images as encoder input, while leveraging the deformation fields from MammoRegNet to perform alignment at the feature level. Despite these architectural differences, both variants demonstrate superior predictive performance, underscoring the generalizability of spatial alignment benefits across risk modeling approaches.

To further analyze the performance of alignment methods within our own risk prediction framework, Figure 3 presents the C-index across different breast density categories for both datasets. This stratified evaluation provides additional insight into model robustness under varying imaging conditions. As shown, ImgFeatAlign consistently achieves the highest C-index values across nearly all density levels, highlighting its ability to generalize across diverse imaging scenarios. In contrast, NoAlign performs the worst, exhibiting substantial C-index declines and higher variability, particularly in higher density categories, highlighting the limitations of unaligned approaches in longitudinal prediction tasks. Among feature-based alignment methods,

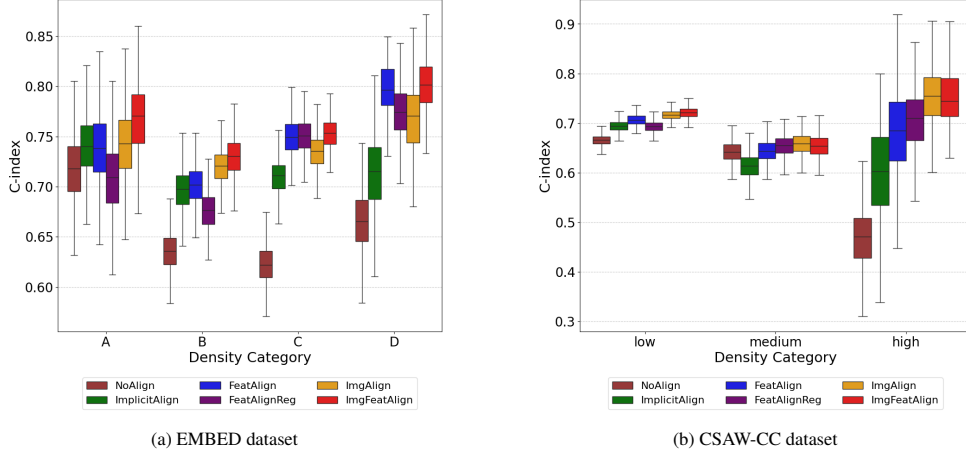


Figure 3: Analysis of C-index by density category for the different alignment methods across both datasets.

FeatAlign generally outperforms FeatAlignReg, again highlighting that the added regularization term in FeatAlignReg, which promotes smoother and more stable transformations, leads to reduced risk prediction performance. Implicit methods perform comparably well in lower breast density categories, but struggle in higher density categories.

In comparison, image-based methods, especially ImgFeatAlign, show improved performance as breast density increases. This indicates that incorporating spatial transformations at the image level is particularly advantageous for modeling complex structural variations, leading to more accurate risk predictions in dense breast tissue.

5.2. In-Depth Analysis on the EMBED dataset

While both datasets are used to evaluate overall model performance, the following analyses, including ROC curves, precision trends, and recall trends, focus specifically on the EMBED dataset. Compared to the CSAW-CC dataset, EMBED offers greater diversity, incorporating images from multiple imaging systems and a broader patient population. This diversity enhances the robustness and generalizability of the findings, making it a more suitable benchmark for evaluating alignment methods in real-world scenarios.

5.2.1. ROC-AUC Curves

Figure 4 presents the ROC AUC scores on the EMBED dataset over a five-year follow-up period, comparing the performance of various alignment methods. ImgFeatAlign consistently achieves the highest AUC across all years, demonstrating both superior predictive performance and temporal stability, with only minimal degradation in later years. ImgAlign ranks second, further reinforcing the effectiveness of image-based alignment strategies.

In contrast, NoAlign yields the lowest AUC scores, with a pronounced decline over time, underscoring the importance of spatial alignment for preserving predictive accuracy in longitudinal settings. Feature-level alignment methods show more variability: FeatAlign consistently outperforms FeatAlignReg, indicating that while regularization may enhance deformation smoothness, it adversely affects predictive performance. Nonetheless, both methods remain inferior to their image-based counterparts.

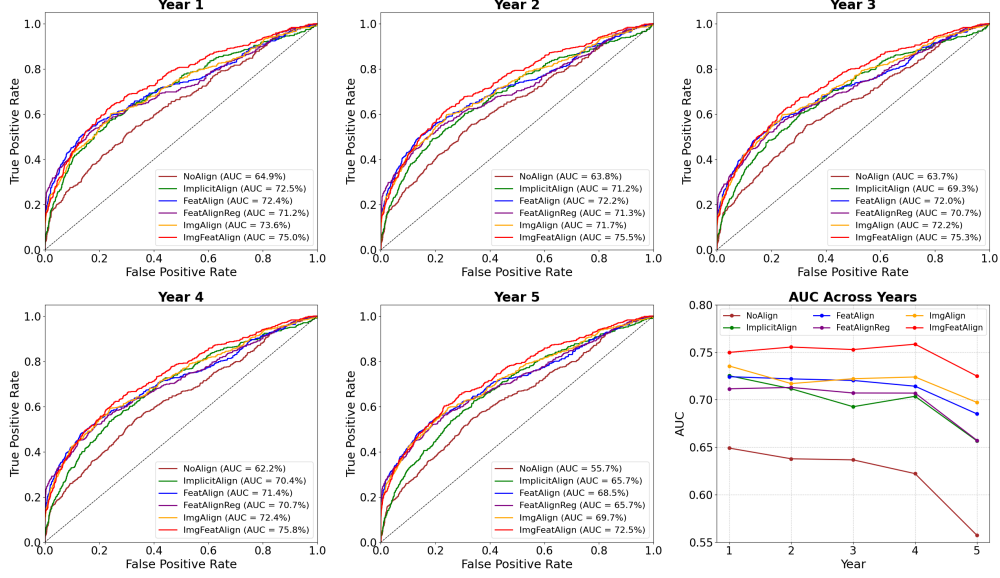


Figure 4: AUC-ROC curves for risk predictions on the EMBED dataset across 1-5 year outcomes: comparison of different alignment methods.

ImplicitAlign performs similarly to FeatAlignReg, but lags behind both FeatAlign and the image-based methods (ImgAlign and ImgFeatAlign), suggesting that implicit and feature-based approaches are less effective at capturing temporal dynamics and correcting spatial misalignments.

The superior performance of ImgFeatAlign and ImgAlign can be attributed to their use of deformation fields derived from image-level registration, which enables more precise spatial alignment over time. This spatial consistency is critical for improving longitudinal prediction. Conversely, the poor performance of NoAlign highlights the detrimental effect of ignoring alignment, as it impairs the model’s ability to learn temporally meaningful and discriminative features. These findings reinforce the central role of explicit alignment, particularly image-based, in enhancing the robustness and accuracy of temporal breast cancer risk prediction.

5.2.2. Precision

Figure 5 presents the precision of different alignment methods in predicting the time until cancer diagnosis. Across all prediction horizons, explicit alignment methods, FeatAlign, FeatAlignReg, ImgAlign, and ImgFeatAlign, consistently outperform both implicit alignment (ImplicitAlign) and the unaligned baseline (NoAlign), highlighting the advantage of explicit alignment in capturing discriminative patterns for accurate risk prediction.

Among these, FeatAlignReg achieves the highest precision across all years, indicating that the incorporation of regularization to enforce anatomically plausible deformation fields does not compromise and may even enhance precision. FeatAlign and ImgFeatAlign exhibit comparable performance, both maintaining strong and stable precision. ImgAlign, while still outperforming the implicit and unaligned approaches, lags slightly behind ImgFeatAlign.

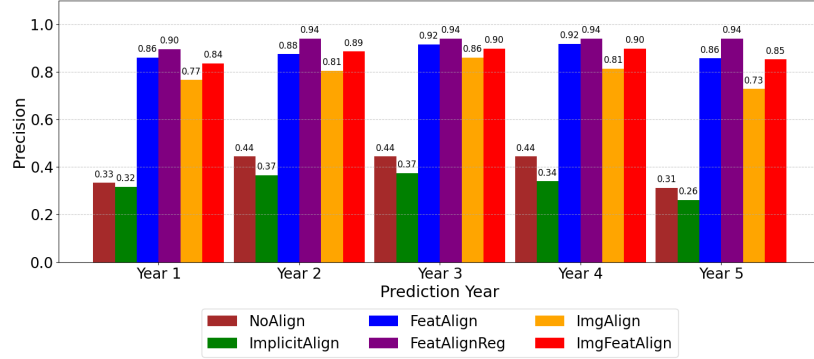


Figure 5: Comparison of alignment methods in predicting years to cancer development, measured by precision.

In contrast, ImplicitAlign shows the weakest precision, with NoAlign performing marginally better but still significantly below the explicit alignment methods.

High precision is essential in cancer risk prediction, as it reduces the rate of false positives, minimizing unnecessary follow-ups, avoiding patient anxiety, and conserving healthcare resources. The superior precision of explicit alignment strategies reinforces their potential utility in clinical settings where specificity is as critical as sensitivity in managing risk and guiding interventions.

5.2.3. Recall

Figure 6 illustrates the recall performance of various alignment strategies in predicting the time until cancer diagnosis. Explicit image-level alignment methods, such as ImgAlign and ImgFeatAlign, consistently outperform implicit alignment (ImplicitAlign), feature-level methods (FeatAlign and FeatAlignReg), and the unaligned baseline (NoAlign) across all prediction horizons. This demonstrates the effectiveness of image-level alignment in capturing spatial and temporal patterns critical for identifying high-risk patients.

Within the feature-level methods, introducing a regularization term to enforce anatomically plausible deformation fields (FeatAlignReg) leads to a decline in recall, suggesting a trade-off between anatomical plausibility and predictive performance.

Among all methods, ImgFeatAlign achieves the highest recall across all years, with ImgAlign also showing strong performance. In contrast, NoAlign consistently yields the lowest recall, highlighting the limitations of models that omit alignment.

High recall is essential in cancer risk prediction, as it ensures that the majority of high-risk individuals are correctly identified, thereby minimizing missed diagnoses. This capability is critical for enabling early interventions and improving patient outcomes. The strong recall of explicit image-based alignment methods underscores their potential clinical value in supporting effective risk stratification and treatment planning.

5.2.4. Short term performance across density categories

Figure 7 presents the AUC for short-term risk prediction (i.e., within one year) across different breast density categories. Image-level alignment methods, particularly ImgFeatAlign, consistently achieve the highest performance across most density levels. ImgFeatAlign demonstrates

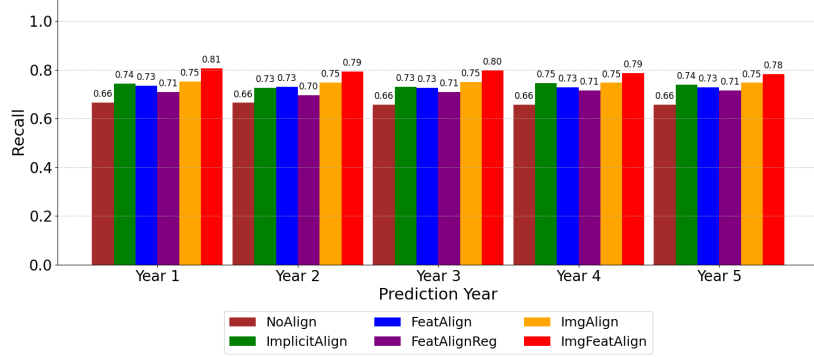


Figure 6: Comparison of alignment methods in predicting years to cancer development, measured by recall.

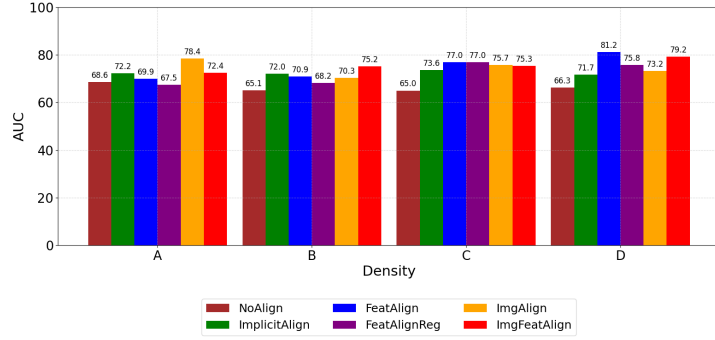


Figure 7: Analysis of AUC by density category for the different alignment methods for getting cancer in 1-Year (short term risk).

notable robustness in higher-density cases (e.g., Density D), where it outperforms all other methods. ImgAlign also performs competitively, especially in lower-density cases (e.g., Density A), where it achieves the highest AUC.

Feature-level alignment methods (FeatAlign and FeatAlignReg) show moderate performance, generally lagging behind image-based methods. FeatAlign outperforms FeatAlignReg in most density categories, but neither matches the performance of ImgFeatAlign. NoAlign consistently exhibits the lowest AUC, while ImplicitAlign shows marginally better results but remains behind the explicit image-level alignment approaches.

These results underscore the effectiveness of image-level alignment, particularly ImgFeatAlign, in delivering reliable short-term risk prediction across varying tissue densities. They further highlight the importance of robust spatial alignment strategies, especially in high-complexity scenarios such as dense breast tissue, where accurate risk modeling is more challenging.

5.2.5. Long term performance across density categories

Figure 8 presents AUC values for long-term risk prediction (i.e., over multiple years) stratified by breast density category. The results underscore the critical importance of alignment strategies in maintaining predictive performance over extended time horizons.

NoAlign consistently yields the lowest AUC across all density categories, with performance

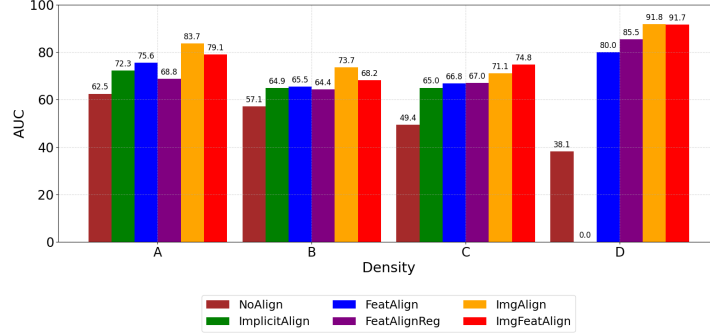


Figure 8: Analysis of AUC by density category for the different alignment methods for getting cancer in 5-Year (long term risk).

deteriorating further as breast density increases. This pattern highlights the limitations of models lacking spatial alignment, particularly in more complex anatomical contexts.

In contrast, image-based alignment methods—*ImgAlign* and *ImgFeatAlign*—achieve the highest AUCs across all categories, demonstrating strong robustness and generalization, even in challenging scenarios. Feature-level alignment methods (*FeatAlign* and *FeatAlignReg*) and *ImplicitAlign* perform moderately well in categories A, B, and C but fall short of image-based methods. Notably, in category D (very dense tissue), where the dataset size is particularly limited (12 images), *ImgFeatAlign* and *ImgAlign* maintain high performance, achieving AUCs of 91.7 and 91.8, respectively. In stark contrast, *ImplicitAlign* fails entirely in this category, with an AUC of 0, indicating a complete inability to discriminate between positive and negative cases.

Interestingly, while *FeatAlign* outperforms *FeatAlignReg* in lower-density categories, the reverse holds in higher-density settings. This suggests that the anatomically constrained deformations enforced by *FeatAlignReg* may better capture the complex tissue structures in dense breasts.

The limited sample size in density D amplifies the challenges faced by less robust methods, underscoring the importance of alignment strategies, particularly image-based, that can generalize effectively in data-scarce scenarios. These findings reinforce the superiority of *ImgFeatAlign* and *ImgAlign* for long-term cancer risk prediction, especially in anatomically complex or high-density cases.

5.3. Comparison of Deformation Field Quality

Table 5 reports the quality of deformation fields across methods, measured by the Percentage of Negative Jacobian Determinants (NJD) and the standard deviation of the Jacobian determinant. Image-based registration methods, *ImgAlign* and *ImgFeatAlign*, achieve the highest deformation quality, exhibiting the lowest NJD values and minimal standard deviation across both datasets. These results show that image-based approaches produce smooth, spatially coherent, and anatomically plausible deformation fields.

In contrast, *FeatAlign* (feature-based registration without regularization) shows poor deformation quality, with elevated NJD values and high variability, suggesting implausible and unstable transformations. Introducing regularization in *FeatAlignReg* significantly improves performance, reducing both NJD and standard deviation. However, it still falls slightly short of the quality achieved by image-based methods.

Table 5: Quantitative comparison of deformation field quality. NJD: percentage of negative Jacobian determinants. Std. dev. Jacobian: standard deviation of the Jacobian determinant.

Dataset	Method	NJD (%) ↓	Std. dev. Jacobian ↓
EMBED	FeatAlign	2.7096 (2.6808-1.7383)	0.4599 (0.4562-0.4640)
	FeatAlignReg	0.0385 (0.0171-0.0791)	0.0889 (0.0855-0.0931)
	ImgAlign / ImgFeatAlign	0.0013 (0.0012-0.0015)	0.1408 (0.1403-0.1413)
CSAW-CC	FeatAlign	3.6970 (3.6667-3.7270)	0.5965 (0.5897-0.6030)
	FeatAlignReg	0.0059 (0.0049-0.0071)	0.1329 (0.1323-0.1334)
	ImgAlign / ImgFeatAlign	0.0008 (0.0006-0.0009)	0.1402 (0.1397-0.1406)

Overall, these findings underscore the superiority of image-based registration in generating anatomically consistent deformation fields and highlight the critical role of regularization in improving the plausibility of feature-based alignment approaches.

Figures 9 and 10 illustrate the deformation fields generated by the three registration methods, showing the displacement vector fields (top row) and the corresponding Jacobian determinant maps (bottom row). FeatAlign produces highly irregular and noisy deformation fields, characterized by pixel-wise, unstructured transformations lacking global coherence. The absence of regularization results in regions with invalid, non-invertible deformations, as indicated by orange and red areas in the Jacobian maps—markers of negative Jacobian determinants and poor transformation quality.

The introduction of regularization in FeatAlignReg substantially improves the quality of the deformation. The resulting fields are smoother and more structured, with the elimination of invalid regions and preservation of invertibility. However, the transformations remain largely localized and do not achieve the same degree of global alignment observed in image-based methods.

In contrast, the deformation fields produced by ImgAlign and ImgFeatAlign, both based on the MammoRegNet model, exhibit globally coherent and smooth displacements, with no invalid regions. Their deformation vectors are spatially consistent and anatomically meaningful, reflecting global alignment of structures rather than isolated pixel-level adjustments. These transformations benefit from MammoRegNet’s joint affine and deformable components, which contribute to anatomically plausible and spatially stable registration.

Figures B.12 and B.13 further illustrate examples of longitudinal mammography alignment using MammoRegNet, highlighting the ability of image-based methods to produce robust and interpretable deformation fields across time.

6. Discussion

This study demonstrates the clear superiority of image-based alignment methods, particularly ImgFeatAlign, for longitudinal breast cancer risk prediction. Across multiple evaluation criteria, including predictive accuracy, precision, recall, and deformation field quality, ImgFeatAlign consistently outperforms feature-based approaches. These results underscore the critical role of precise spatial alignment in modeling temporal changes in mammography.

The strength of image-based methods lies in their ability to preserve fine-grained anatomical structures, enabling the generation of smooth and spatially consistent deformation fields. In con-

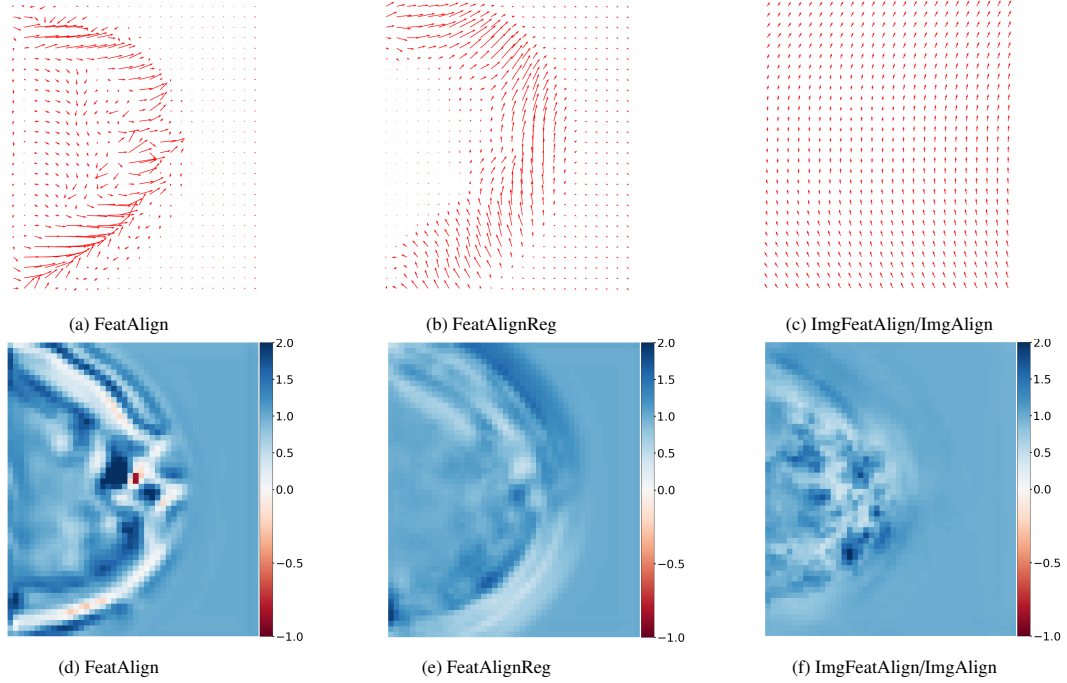


Figure 9: Comparison of deformation field quality. The top row shows displacement vectors, and the bottom row displays Jacobian determinant maps (white/blue: valid; orange/red: invalid non-invertible regions).

trast, feature-based approaches, which operate at a more abstract level, often compromise spatial fidelity. Although regularization techniques, as implemented in FeatAlignReg, enhance the plausibility of deformation fields, they also reduce risk prediction performance. This highlights a trade-off between the quality of the deformation field and the effectiveness of risk prediction.

A notable contribution of this work is the introduction of a hybrid strategy, ImgFeatAlign, which applies image-derived deformation fields within the feature space. This approach leverages the anatomical consistency of image-based registration while benefiting from the representational efficiency of feature-level modeling. By combining spatial precision and feature-level abstraction, the approach achieves robust and reliable performance across different patient populations and imaging conditions.

From a translational perspective, the ability to maintain accurate predictions at multiple time points is vital to identifying high-risk patients. This is particularly critical in challenging imaging conditions, where spatial misalignment can compromise model reliability. ImgFeatAlign exhibits strong generalization across datasets and breast density categories, supporting its potential for real-world clinical deployment.

In summary, these findings highlight that spatial alignment is not merely a preprocessing step, but a foundational component in building robust, trustworthy models for longitudinal risk prediction.

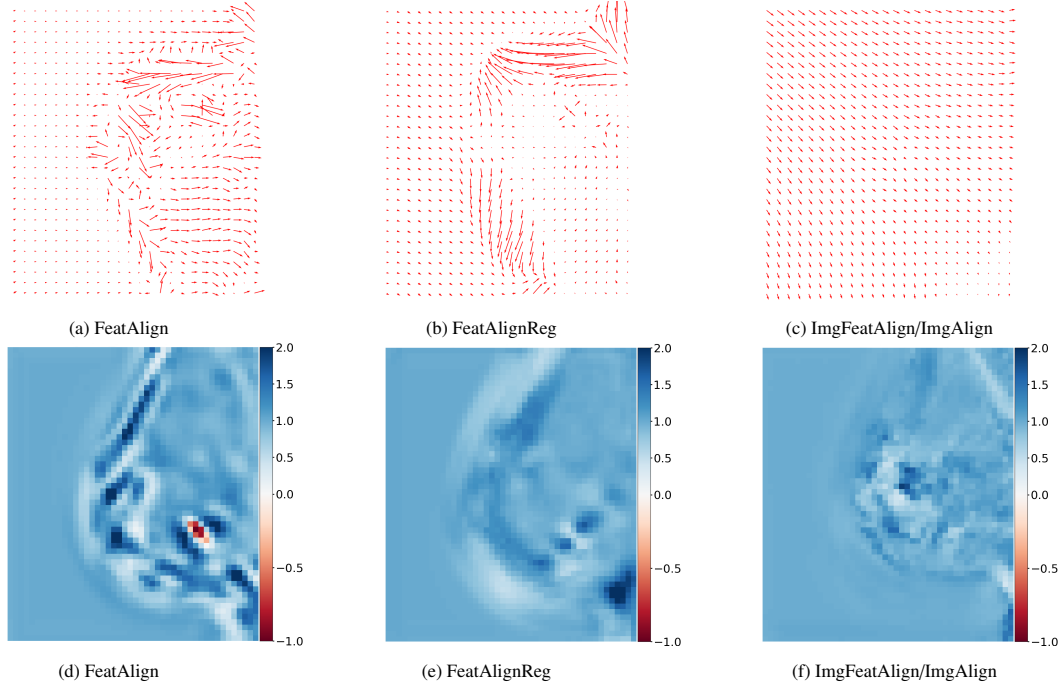


Figure 10: Comparison of deformation field quality. The top row shows displacement vectors, and the bottom row displays Jacobian determinant maps (white/blue: valid; orange/red: invalid non-invertible regions).

7. Conclusion and Outlook

In summary, this study underscores the critical role of accurate spatial alignment in longitudinal prediction of breast cancer risk. Among the evaluated methods, image-based approaches, particularly `ImgFeatAlign`, consistently outperform others by preserving anatomical detail and enabling temporally consistent modeling. By incorporating deformation fields from image-based registration into the feature space, `ImgFeatAlign` achieves an effective balance between spatial precision and high-level feature abstraction, resulting in improved predictive accuracy. These findings have significant clinical implications, as robust longitudinal modeling can support personalized screening strategies and enable timely intervention for people at elevated risk.

Looking ahead, future research could extend the proposed alignment strategies to a wider range of longitudinal imaging applications, including modeling disease progression and predicting treatment response. Integrating aligned imaging features with clinical, genetic, and demographic data may further enhance model interpretability and improve risk stratification. Furthermore, incorporating multiview mammography data, specifically CC and MLO views, into a unified prediction framework holds promise for further improving the accuracy and robustness of breast cancer risk assessment. These directions offer promising avenues for advancing precision medicine through more comprehensive and reliable imaging-based prognostics.

CRediT authorship contribution statement

Solveig Thrun: Conceptualization, Software, Methodology, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Stine Hansen:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Zijun Sun:** Methodology, Writing - review & editing. **Nele Blum:** Methodology, Writing - review & editing, Supervision. **Suaiba A. Salahuddin:** Methodology, Writing - review & editing. **Xin Wang:** Methodology, Writing - review & editing. **Kristoffer Wickstrøm:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Elisabeth Wetzer:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Robert Jenssen:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition. **Maik Stille:** Methodology, Writing - review & editing, Supervision. **Michael Kampffmeyer:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The source code used in this study is publicly available on GitHub, with the link provided in the abstract. All datasets used in this work are publicly accessible and links to request access are provided in the Experimental Setup section.

Acknowledgements

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

Appendix A. Schematic Illustrations of Longitudinal Risk Prediction Architectures

Figure A.11 provides a detailed visualization of the various alignment strategies evaluated within the unified risk prediction framework. Each subfigure illustrates a distinct architectural design, including no alignment, implicit alignment, feature-level alignment (FeatAlign, FeatAlignReg), image-level alignment (ImgAlign) and a combined Image-Feature Alignment (ImgFeatAlign). These visualizations highlight how temporal and modality-specific information is integrated or aligned between different model variants to improve temporal consistency and predictive performance.

Appendix B. MammoRegNet

Table B.6 summarizes the performance of the image-based registration model, MammoRegNet, on the EMBED and CSAW-CC datasets. The results demonstrate a marked improvement in image similarity, measured by Normalized Cross-Correlation, across all registration stages, from the initial input through affine and final deformable registration. This progression highlights the model’s ability to achieve accurate spatial alignment of mammography images. Moreover, the Percentage of Negative Jacobian Determinants remains exceptionally low across both datasets, indicating that MammoRegNet produces smooth and anatomically plausible deformation fields. Together, these findings underscore the model’s effectiveness in delivering high-quality, reliable mammography image registration.

Figures B.12 and B.13 demonstrate the registration performance of MammoRegNet across a variety of cases, including examples from patients with and without cancer. For the cancerous images, the location of the malignancy is indicated by a radiologist-annotated region of interest (ROI). In each example, the model successfully aligns the prior mammogram to the current view using a combination of affine and deformable transformations. The aligned prior mammograms (c) closely match the anatomical structures of the current mammograms (a), demonstrating MammoRegNet’s precision in spatial alignment. Overlay visualizations further reveal the alignment process: substantial misalignment is evident in the initial overlay (d), partially corrected with affine registration (e), and fully refined through deformable registration (f). Notably, MammoRegNet performs consistently across both MLO and CC views and for both left and right breasts, confirming its robustness across diverse clinical imaging scenarios. These results highlight MammoRegNet’s capacity to produce precise, anatomically consistent alignments, enabling high-fidelity image registration for longitudinal mammography analysis.

Table B.6: MammoRegNet registration performance (95% CI in parentheses).

Dataset	NCC before (%) $\uparrow$	NCC affine (%) $\uparrow$	NCC final (%) $\uparrow$	NJD (%) $\downarrow$
EMBED	70.7	78.7	92.2	0.0013
	(70.1–71.3)	(78.3–79.1)	(92.0–92.4)	(0.0012–0.0015)
CSAW-CC	76.0	81.6	92.8	0.0008
	(75.3–76.6)	(81.1–82.0)	(92.5–93.0)	(0.0006–0.0009)

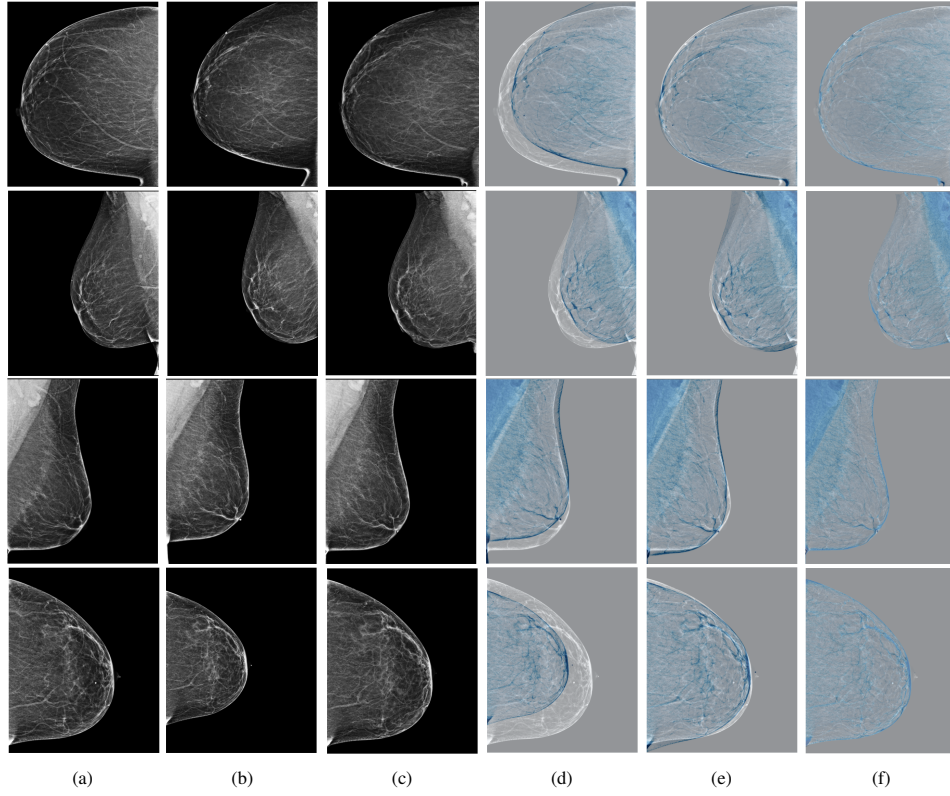


Figure B.12: Registration result of MammoRegNet. Visual results of the MammoRegNet registration performance for patients without cancer. (a) Current mammogram, (b) prior mammogram, (c) aligned prior mammogram, (d) current mammogram overlaid with the prior mammogram, (e) current mammogram overlaid with the affine-transformed prior mammogram, and (f) current mammogram overlaid with the final transformed prior mammogram.

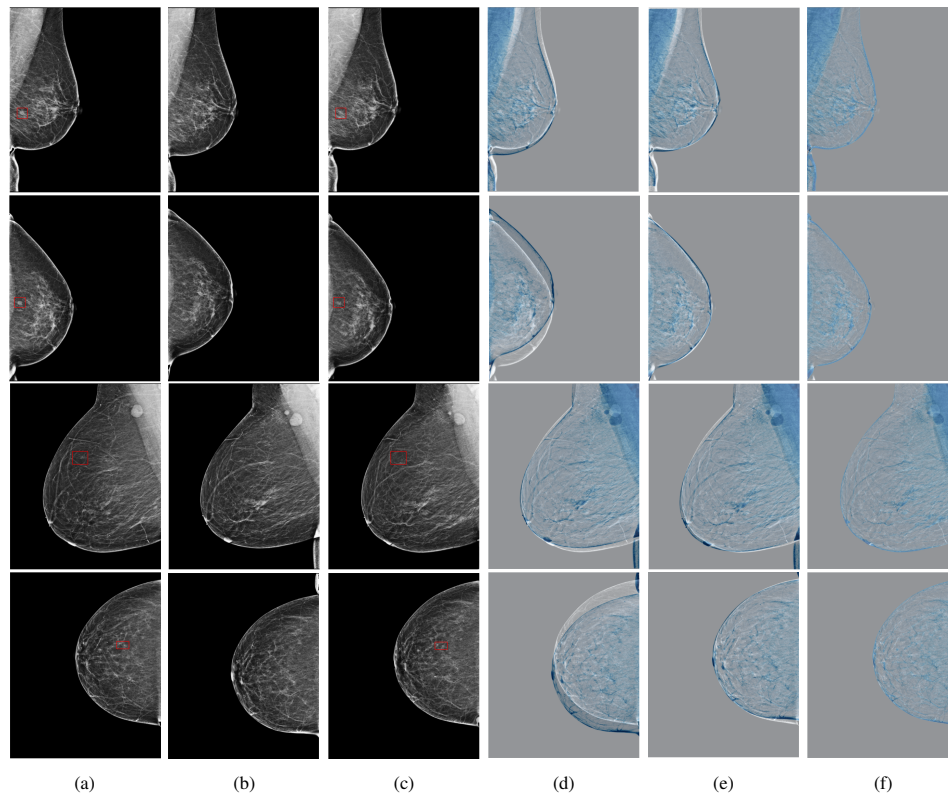


Figure B.13: Registration result of MammoRegNet. Visual results of the MammoRegNet registration performance for patients with cancer. The tumor region is highlighted by a bounding box on the current mammogram and the aligned prior mammogram. (a) Current mammogram, (b) prior mammogram, (c) aligned prior mammogram, (d) current mammogram overlaid with the prior mammogram, (e) current mammogram overlaid with the affine-transformed prior mammogram, and (f) current mammogram overlaid with the final transformed prior mammogram.

References

- Akwo, J.D., Trieu, P.D.Y., Barron, M.L., Reynolds, T., Lewis, S.J., 2024. Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance. *Clin. Radiol.* 79, 12. doi: <https://doi.org/10.1016/j.crad.2024.09.007>.
- Anandarajah, A., Chen, Y., Colditz, G.A., Hardi, A., Stoll, C., Jiang, S., 2022. Studies of parenchymal texture added to mammographic breast density and risk of breast cancer: a systematic review of the methods used in the literature. *Breast Cancer Res.* 24(1). doi: <https://doi.org/10.1186/s13058-022-01600-5>.
- Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* 38, pp. 1788-1800. doi: <https://doi.org/10.1109/TMI.2019.2897538>.
- Bhushan, A., Gonsalves, A., Menon, J.U., 2021. Current state of breast cancer diagnosis, treatment, and theranostics. *Pharmaceutics* 13, 723. doi: <https://doi.org/10.3390/pharmaceutics13050723>.

- Bodewes, F.T.H., van Asselt, A.A., Dorrius, M.D., Greuter, M.J.W., de Bock, G.H., 2022. Mammographic breast density and the risk of breast cancer: A systematic review and meta-analysis. *Breast* 66, pp. 62-68. doi: <https://doi.org/10.1016/j.breast.2022.09.007>.
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A., 2024. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 74, pp. 229-263. doi: <https://doi.org/10.3322/caac.21834>.
- Che, T., Wang, X., Zhao, K., Zhao, Y., Zeng, D., Li, Q., Zheng, Y., Yang, N., Wang, J., Li, S., 2023. AMNet: Adaptive multi-level network for deformable registration of 3D brain MR images. *Med. Image Anal.* 85. doi: <https://doi.org/10.1016/j.media.2023.102740>.
- Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y., 2022. TransMorph: Transformer for unsupervised medical image registration. *Med. Image Anal.* 82. doi: <https://doi.org/10.1016/j.media.2022.102615>.
- Chen, M., Tustison, N.J., Jena, R., Gee, J.C., 2023. Image registration: Fundamentals and recent advances based on deep learning. In: *Machine Learning for Brain Disorders*, pp. 435-458. doi: https://doi.org/10.1007/978-1-0716-3195-9_14.
- Chen, Z., Zheng, Y., Gee, J.C., 2024. TransMatch: A transformer-based multilevel dual-stream feature matching network for unsupervised deformable image registration. *IEEE Trans. Med. Imaging* 43, pp. 15-27. doi: <https://doi.org/10.1109/TMI.2023.3288136>.
- Chen, J., Liu, Y., Wei, S., Bian, Z., Subramanian, S., Carass, A., Prince, J.L., Du, Y., 2025. A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. *Med. Image Anal.* 100. doi: <https://doi.org/10.1016/j.media.2024.103385>.
- Dadsetan, S., Arefan, D., Berg, W.A., Zuley, M.L., Sumkin, J.H., Wu, S., 2022. Deep learning of longitudinal mammogram examinations for breast cancer risk prediction. *Pattern Recognit.* 132. doi: <https://doi.org/10.1016/j.patcog.2022.108919>.
- Damiani, C., Kalliatakis, G., Sreenivas, M., Al-Attar, M., Rose, J., Pudney, C., Lane, E.F., Cuzick, J., Montana, G., Brentnall, A.R., 2023. Evaluation of an AI model to assess future breast cancer risk. *Radiology* 307(5). doi: <https://doi.org/10.1148/radiol.222679>.
- Dembrower, K., Lindholm, P., Strand, F., 2020. 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). *J. Digit. Imaging* 33, pp. 408-413. doi: <https://doi.org/10.1007/s10278-019-00278-0>.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*.
- Donnelly, J., Moffett, L., Barnett, A.J., Trivedi, H., Schwartz, F., Lo, J., Rudin, C., 2024. Asym-Mirai: Interpretable mammography-based deep learning model for 1–5-year breast cancer risk prediction. *Radiology* 310(3). doi: <https://doi.org/10.1148/radiol.232780>.

- Elkin, E.B., Snow, J.G., Leoce, N.M., Atoria, C.L., Schrag, D., 2012. Mammography capacity and appointment wait times: barriers to breast cancer screening. *Cancer Causes Control* 23, pp. 45–50.
- Eriksson, M., Czene, K., Vachon, C., Conant, E.F., Hall, P., 2023. A clinical risk model for personalized screening and prevention of breast cancer. *Cancers* 15, 3246. doi: <https://doi.org/10.3390/cancers15123246>.
- Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. In: *Proc. Fourteenth Int. Conf. on Artificial Intelligence and Statistics*.
- Guo, Y., Sivaramakrishna, R., Lu, C.-C., Suri, J.S., Laxminarayan, S., 2006. Breast image registration techniques: a survey. *Med. Biol. Eng. Comput.* 44, pp. 15–26. doi: <https://doi.org/10.1007/s11517-005-0016-y>.
- Ha, I.Y., Wilms, M., Heinrich, M., 2020. Semantically guided large deformation estimation with deep networks. *Sensors* 20, 1392. doi: <https://doi.org/10.3390/s20051392>.
- Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, pp. 29–36.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770–778. doi: <https://doi.org/10.1109/CVPR.2016.90>.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, pp. 1735–1780. doi: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hussain, S., Ali, M., Naseem, U., Nezhadmoghadam, F., Jatoi, M.A., Gulliver, T.A., Tamez-Peña, J.G., 2024. Breast cancer risk prediction using machine learning: a systematic review. *Front. Oncol.* 14, 1343627. doi: <https://doi.org/10.3389/fonc.2024.1343627>.
- Im, J.E., Khalifa, M., Gregory, A.V., Erickson, B.J., Kline, T.L., 2024. A systematic review on the use of registration-based change tracking methods in longitudinal radiological images. *J. Imaging Inform. Med.* doi: <https://doi.org/10.1007/s10278-024-01333-1>.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proc. Int. Conf. Mach. Learn. (ICML)*.
- Jeong, 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., 2023. The EMory BrEast imaging Dataset (EMBED): a racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images. *Radiol. Artif. Intell.* 5, 1. doi: <https://doi.org/10.1148/ryai.220047>.
- Jiang S, Bennett DL, Rosner BA, Colditz GA. Longitudinal Analysis of Change in Mammographic Density in Each Breast and Its Association With Breast Cancer Risk. *JAMA Oncol.* 2023 Jun 1;9(6):808-814. doi: <https://doi.org/10.1001/jamaoncol.2023.0434>.
- Karaman, B.K., Dodelzon, K., Akar, G.B., Sabuncu, M.R., 2024. Longitudinal mammogram risk prediction. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 437–446. doi: https://doi.org/10.1007/978-3-031-72086-4_41.

- Keller, B.M., Nathan, D.L., Wang, Y., Zheng, Y., Gee, J.C., Conant, E.F., Kontos, D., 2012. 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. *Med. Phys.* 39, pp. 4903–4917. doi: <https://doi.org/10.1118/1.4736530>.
- Klanecek, Z., Wang, Y.K., Wagner, T., Cockmartin, L., Marshall, N., Schott, B., Deatsch, A., Studen, A., Jarm, K., Krajc, M., Vrhovec, M., Bosmans, H., Jeraj, R., 2024. Longitudinal interpretability of deep learning-based breast cancer risk prediction. *Phys. Med. Biol.* 70, 1. doi: <https://doi.org/10.1088/1361-6560/ad9db3>.
- Kingma, D.P., Ba, J., 2017. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980. <https://arxiv.org/abs/1412.6980>.
- Lee, H., Kim, J., Park, E., Kim, M., Kim, T., Kooi, T., 2023. Enhancing breast cancer risk prediction by incorporating prior images. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 389–398. doi: https://doi.org/10.1007/978-3-031-43904-9_38.
- Li, J., Wang, X., Qian, C., 2020. Mammography registration for unsupervised learning based on CC and MLO views. In: *Proc. 3rd Int. Conf. on Artificial Intelligence and Pattern Recognition*. doi: <https://doi.org/10.1145/3430199.3430238>.
- Li, H., Robinson, K., Lan, L., Baughan, N., Chan, C.W., Embury, M., Whitman, G.J., El-Zein, R., Bedrosian, I., Giger, M.L., 2023. Temporal machine learning analysis of prior mammograms for breast cancer risk prediction. *Cancers* 15(7), 2141. doi: <https://doi.org/10.3390/cancers15072141>.
- Li, T.Z., Xu, K., Gao, R., Tang, Y., Lasko, T.A., Maldonado, F., Sandler, K.L., Landman, B.A., 2023. Time-distance vision transformers in lung cancer diagnosis from longitudinal computed tomography. *Proc SPIE Int Soc Opt Eng.* 12464: 1246412. doi: <https://doi.org/10.1117/12.2653911>.
- Loizidou, K., Skouroumouni, G., Pitris, C., Nikolaou, C., 2021. Digital subtraction of temporally sequential mammograms for improved detection and classification of microcalcifications. *Eur. Radiol. Exp.* 5(1). doi: <https://doi.org/10.1186/s41747-021-00238-w>.
- Loizidou, K., Skouroumouni, G., Nikolaou, C., Pitris, C., 2022. Automatic breast mass segmentation and classification using subtraction of temporally sequential digital mammograms. *IEEE J. Transl. Eng. Health Med.* 10, 1–11. doi: <https://doi.org/10.1109/JTEHM.2022.3219891>.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin Transformer: hierarchical vision transformer using shifted windows. In: *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 9992–10002. doi: <https://doi.org/10.1109/ICCV48922.2021.00986>.
- Meng, M., Bi, L., Fulham, M., Feng, D., Kim, J., 2023. Non-iterative coarse-to-fine transformer networks for joint affine and deformable image registration. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 750–760. doi: https://doi.org/10.1007/978-3-031-43999-5_71.

- Michel, A., Ro, V., McGuinness, J.E., Mutasa, S., Terry, M.B., Tehranifar, P., May, B., Ha, R., Crew, K.D., 2023. Breast cancer risk prediction combining a convolutional neural network-based mammographic evaluation with clinical factors. *Breast Cancer Res. Treat.* 200, pp.237–245. doi: <https://doi.org/10.1007/s10549-023-06966-4>
- Mikhael, P.G., Wohlwend, J., Yala, A., Karstens, L., Xiang, J., Takigami, A.K., Bourgouin, P.P., Chan, P., Mrah, S., Amayri, W., Juan, Y.H., Yang, C.T., Wan, Y.L., Lin, G., Sequist, L.V., Fintelmann, F.J., Barzilay, R., 2023. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. *J. Clin. Oncol.* 41, pp. 2191–2200. doi: <https://doi.org/10.1200/JCO.22.01345>.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: an imperative style, high-performance deep learning library. In: *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 32, pp. 8024–8035.
- Rubio, I.T., Drukker, C.A., Esgueva, A., 2024. Risk-based breast cancer screening: what are the challenges? *Tumori J.* 0(0). doi: <https://doi.org/10.1177/03008916241306971>.
- Sickles, E.A., ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, American College of Radiology, 2013.
- Sun, Z., Thrun, S., Kampffmeyer, M., 2025. VMRA-MaR: An Asymmetry-Aware Temporal Framework for Longitudinal Breast Cancer Risk Prediction. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. arXiv preprint arXiv:2506.17412. <https://arxiv.org/abs/2506.17412>.
- Tan, M., Zheng, B., Leader, J.K., Gur, D., 2016. Association between changes in mammographic image features and risk for near-term breast cancer development. *IEEE Trans. Med. Imaging* 35(7). doi: <https://doi.org/10.1109/tmi.2016.2527619>.
- Thrun, S., Hansen, S., Sun, Z., Blum, N., Salahuddin, S.A., Wickstrøm, K., Wetzter, E., Jenssen, R., Stille, M., Kampffmeyer, M., 2025. Reconsidering Explicit Longitudinal Mammography Alignment for Enhanced Breast Cancer Risk Prediction. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. arXiv preprint arXiv:2506.19363. <https://arxiv.org/abs/2506.19363>.
- Uno, H., Cai, T., Pencina, M.J., D’Agostino, R.B., Wei, L.J., 2011. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* 30(10). doi: <https://doi.org/10.1002/sim.4154>.
- Varela, C., Karssemeijer, N., Hendriks, J.H.C.L., Holland, R., 2005. Use of prior mammograms in the classification of benign and malignant masses. *Eur. J. Radiol.* 56(2). doi: <https://doi.org/10.1016/j.ejrad.2005.04.007>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30. doi: <https://doi.org/10.5555/3295222.3295349>.

- de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* 52. doi: <https://doi.org/10.1016/j.media.2018.11.010>.
- Walton, W.C., Kim, S.-J., Mullen, L.A., 2022. Automated registration for dual-view X-ray mammography using convolutional neural networks. *IEEE Trans. Biomed. Eng.* 69(11). pp. 3538-3550. doi: <https://doi.org/10.1109/TBME.2022.3173182>.
- Wang, A.T., Vachon, C.M., Brandt, K.R., Ghosh, K., 2014. Breast density and breast cancer risk: a practical review. *Mayo Clin. Proc.* 89(4). pp. 548–557. doi: <https://doi.org/10.1016/j.jmayocp.2013.12.014>.
- 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., and Mann, R., 2023. Predicting up to 10 year breast cancer risk using longitudinal mammographic screening history. *medRxiv*. doi: <https://doi.org/10.1101/2023.06.28.23291994>.
- Wang, X., Tan, T., Gao, Y., Marcus, E., Han, L., Portaluri, A., Zhang, T., Lu, C., Liang, X., Beets-Tan, R., Teuwen, J., and Mann, R., 2024. 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)*, pp. 155–165. doi: https://doi.org/10.1007/978-3-031-72378-0_15.
- Yala, A., Lehman, C., Schuster, T., Portnoi, T., Barzilay, R., 2019. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology* 292(1). doi: <https://doi.org/10.1148/radiol.2019182716>.
- Yala, A., Mikhael, P.G., Strand, F., Lin, G., Smith, K., Wan, Y.-L., Lamb, L., Hughes, K., Lehman, C., and Barzilay, R., 2021. Toward robust mammography-based models for breast cancer risk. *Sci. Transl. Med.* 13(578). pp. 60-66. doi: <https://doi.org/10.1126/scitranslmed.aba4373>.
- Yeoh, H., Strand, F., Phan, R., Rahmat, K., and Tan, M., 2024. A New Time-decay Radiomics Integrated Network (TRINet) for Short-term Breast Cancer Risk Prediction, *arXiv*. doi: <https://doi.org/10.48550/arXiv.2412.03081>.
- Yin, W., Sonke, J.-J., Gavves, E., 2023. PC-Reg: A pyramidal prediction–correction approach for large deformation image registration. *Med. Image Anal.* 90. doi: <https://doi.org/10.1016/j.media.2023.102978>.
- Zielonke, N., Gini, A., Jansen, E.E.L., Anttila, A., Segnan, N., Ponti, A., Veerus, P., de Koning, H.J., van Ravesteijn, N.T., Heijnsdijk, E.A.M., Heinävaara, S., Sarkeala, T., Cañada, M., Pitter, J., Széles, G., Voko, Z., Minozzi, S., Senore, C., van Ballegooijen, M., Driesprong - de Kok, I., Lansdorp – Vogelaar, I., Ivanus, U., Jarm, K., Novak Mlakar, D., Primic-Žakelj, M., McKee, M., and Priaulx, J., 2020. Evidence for reducing cancer-specific mortality due to screening for breast cancer in Europe: A systematic review. *Eur. J. Cancer* 127, pp. 191–206. doi: <https://doi.org/10.1016/j.ejca.2019.12.010>.