MamaDino: Mammography-Aware Multi-view Attentional DINO for 3-Year Breast Cancer Risk Prediction

Ruggiero Santeramo[1] (✉)[0000−0001−7107−235X], Igor Zubarev[1][0009−0008−6705−0949], and Florian Jug[1] (✉)[0000−0002−8499−5812]

Fondazione Human Technopole, V.le Rita Levi-Montalcini 1, 20157, Milano, Italy
{ruggiero.santeramo, igor.zubarev, florian.jug}@fht.org

Abstract. Breast cancer screening programmes increasingly seek to move from one-size-fits-all intervals to risk-adapted, personalised strategies. Deep learning has enabled image-based risk models that improve short- to medium-term (1–5 year) prediction over traditional risk models, but leading approaches such as Mirai typically rely on convolutional backbones, high-resolution inputs and relatively simple multi-view fusion, with limited explicit modelling of contralateral asymmetry. We hypothesised that combining complementary inductive biases (convolutional and transformer-based) with explicit contralateral modelling could match state-of-the-art 3-year risk prediction even at substantially lower input resolution. We present MamaDino, a Mammography-Aware Multi-view Attentional DINO-based model that fuses frozen self-supervised DINOv3 (ViT-S) features with a trainable CNN encoder at 512×512 resolution and aggregates left–right information via a BilateralMixer to output a 3-year breast cancer risk score. The 512×512 input size was chosen for computational feasibility, not as a preferred clinical resolution. We train on 53,883 women from OPTIMAM and evaluate on matched 3-year case-control cohorts: an in-distribution test set from four UK screening sites and an external out-of-distribution (OOD) set from an unseen site. At breast level, MamaDino was comparable to Mirai on in-distribution and OOD tests while using $\sim$ 13× fewer input pixels. Adding the BilateralMixer yields an AUC of 0.736 (*vs*. Mirai 0.713) in-distribution and 0.677 (*vs*. 0.666) out-of-distribution with overlapping confidence intervals (CIs), with consistent performance across age, ethnicity, scanner, tumour type and grade. These results indicate that explicit contralateral reasoning plus complementary inductive biases can recover Mirai-level performance despite lower-resolution mammograms.

Keywords: Breast Cancer · Deep Learning · Risk Prediction · DINOv3 · Vision Transformers · Hybrid Networks

1 Introduction

Breast cancer is one of the most commonly diagnosed cancers worldwide and remains a leading cause of cancer-related mortality among women despite advances

in treatment and widespread mammography screening [22]. While reducing mortality, most screening programmes still apply uniform intervals (e.g., every 2-or 3 years) that ignore substantial heterogeneity in individual risk. Improving early detection and stratification is therefore central to personalised screening intervals and preventive strategies [19, 2].

Risk-adapted screening has traditionally relied on clinical and demographic scores (e.g., Gail [9], Claus [4], Tyrer–Cuzick/IBIS [23]), which offer only moderate discrimination and make limited use of mammographic information [19]. Hand-crafted density and texture features improve upon purely clinical models [3], but still underutilise the morphologic signal in full-field digital mammograms (FFDM). In contrast, deep learning (DL) models operating directly on FFDM, including Mirai [26], achieve strong short- to medium-term prediction and have been validated across cohorts [25, 18, 13, 8, 16]. Recent work under the "Better CAD, Better Risk" paradigm suggests that high-performing detection systems can be repurposed into accurate fixed-horizon risk models [20].

From a modelling perspective, many top-performing image-based risk models rely on high-resolution inputs (often >1M input pixels per view) and predominantly CNN backbones. Meanwhile, bilateral reasoning is clinically central, radiologists routinely compare the left and right breasts, and has been encoded via contralateral attention or left-right contrastive designs for detection and risk prediction [1, 17, 24, 27, 7]. In parallel, hybrid CNN-transformer architectures aim to combine local texture biases with global context [5, 6, 11], and self-supervised vision transformers such as DINOv3 [21] have emerged as strong general-purpose encoders with promising transfer to medical imaging and mammography, including at reduced resolution [10, 14, 15]. Despite these advances, most prior work uses single-stream backbones or lacks explicit bilateral modelling in the risk head.

This motivates a key question: can explicit contralateral reasoning and complementary inductive biases improve short-term risk prediction compared to single-backbone baselines (CNN-only or Transformer-only) without explicit contralateral fusion, and to what extent does performance depend on input resolution versus structured multi-view fusion? We study 3-year fixed-horizon risk prediction from routine screening FFDM and test the hypothesis that a resolution-efficient hybrid model can match high-resolution CNN approaches such as Mirai while using substantially fewer pixels. Concretely, we combine a frozen DINOv3 prior with a trainable CNN texture encoder and a bilateral mixing head at 512×512 for 3-year risk prediction.

2 Methods

2.1 Dataset and Cohort Selection

Data were sourced from the OPTIMAM Mammography Image Database [12] and comprised women attending routine FFDM screening at four UK services (Jarvis, Leicester, Imperial, St George's) between 2010 and 2021. Exams were

standard two-view bilateral screenings (L-CC, R-CC, L-MLO, R-MLO) using *for presentation* images (i.e., vendor-processed for clinical viewing, as opposed to *for processing* images closer to raw detector outputs). Women aged 40-80 years were included; exams with missing views or implants were excluded. All data were anonymised.

To standardise inputs across sites and manufacturers, all DICOM mammograms were converted to 8-bit PNG. Right-breast views were flipped to left-align laterality. Images were zero-padded on the right and bottom to the dataset's maximum height/width to preserve aspect ratios and enable batching, then resized to each model's input: 512×512 for MamaDino and 1664×2048 for Mirai. We evaluated the released Mirai model using its intended inference preprocessing; because clinical risk factors were unavailable, we provided images only and used Mirai's patient-level 3-year risk score for all comparisons. After filtering, the training cohort comprised **53,883** women. Approximately 500 developed biopsy-confirmed cancer during the screening interval (CI), 1,500 had benign findings (B), 11,000 had malignant or suspicious recalls (M), and the remainder were normal (N). Many women contributed multiple exams (mean 2.57 episodes/patient); splits were patient-wise, i.e. all episodes from a given woman were assigned to a single train/val/test partition. Supervision at *episode level* used the most severe bilateral outcome, with OPTIMAM lesion annotations providing breast-level labels. During training, CI/M/B episodes and their screening exams in the preceding 3 years (CIP, MP) were labelled positive at episode and breast level using lesion laterality; N episodes and contralateral normal breasts in unilateral M episodes were labelled negative.

Most imaging was acquired on Hologic systems (96%), with the remaining 4% from GE, Siemens and Philips. Pixel intensities were normalised using the training-set mean. Self-reported ethnicity was not used as a model input.

Test cohorts. For a baseline screening exam at time t_0, we define a patient-level binary target $y \in \{0, 1\}$. Patients are labelled $y = 1$ if either as (i) a biopsy-conf irmed *interval cancer* diagnosed in $(t_0, t_0 + 3\text{y})$, or (ii) a *screen-detected malignancy* at the first subsequent routine screen (which, for this subset, occurred after three years). Patients are labelled $y = 0$ if there was no cancer diagnosis prior to, or at, the first subsequent 3-year screen. We then constructed two matched case–control test sets. The internal held-out cohort (in-distribution; disjoint from train/val) used 1:2 matching on site, age and manufacturer (Hologic, Siemens, Philips) and included 525 cases and 1,050 controls. The external cohort (out-of-distribution) was drawn from the unseen Oxford site and used 1:4 matching on age and manufacturer (Hologic, GE), yielding 376 cases and 1,504 controls; GE constituted $\sim 1/3$ of exams and ethnicity was largely unavailable.

2.2 Model Architecture

We design a hybrid fusion network that couples a frozen DINOv3 vision transformer with a trainable SE-ResNeXt101 to leverage complementary inductive

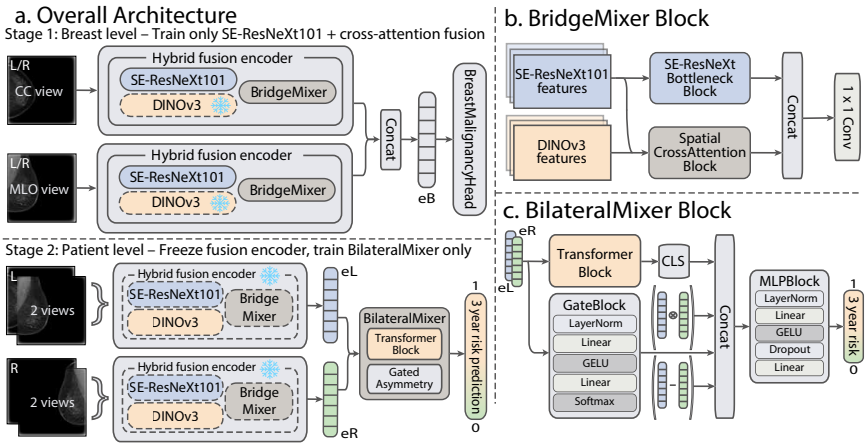

Fig. 1. MamaDino architecture: MamaDino architecture and staged training. Each CC/MLO view is processed through a frozen DINOv3 branch and a trainable view-specific SE-ResNeXt101 branch; BridgeMixer/cross-attention forms left/right breast embeddings. Stage 1 trains SE-ResNeXt101 plus cross-attention with DINOv3 frozen; Stage 2 freezes this fusion encoder and trains only BilateralMixer for patient-level 3-year risk prediction.

biases. For each breast $b \in \{\mathrm{L}, \mathrm{R}\}$, we stack the two standard views (CC and MLO) into $x^{(b)} \in \mathbb{R}^{2\times 3\times H\times W}$ (Fig. 1).

Backbone encoders and cross-attentional fusion. The transformer branch is a frozen DINOv3 ViT-S/16+ that outputs patch tokens per view. In parallel, two trainable view-specific (CC and MLO) SE-ResNeXt101 encoders (same architecture, untied weights) extract convolutional features. We resize both streams to a common 16×16 grid, project to a shared dimension d (here d=512) with 1×1 convolutions, and apply multi-head cross-attention where CNN tokens query DINO tokens, followed by residual feed-forward layers. This allows the convolutional stream to integrate globally coherent, semantically rich transformer features (Fig. 1.a).

BridgeMixer block and breast embedding. On top of the SE-ResNeXt backbone, a SE-based BridgeMixer projects features through a bottleneck to d channels, concatenates this projection with the cross-attention output for each view, and compresses the result back to d channels via a 1×1 convolution. This yields fused view-specific feature maps $F^{\text{view}} \in \mathbb{R}^{d\times 16\times 16}$ for CC and MLO. We concatenate the two views along the channel dimension to obtain a joint breast representation $F^{(b)} \in \mathbb{R}^{2d\times 16\times 16}$, apply adaptive average pooling to 2×2, and flatten the result into an $8d$-dimensional per-breast embedding (Fig. 1.b).

Breast-level classifier and patient-level baseline. For each breast, the embedding is passed through a two-layer MLP with LayerNorm, GELU and

dropout to produce a logit $z^{(b)}$, trained with a breast-level binary focal loss. At inference we obtain probabilities $p^{(\mathrm{L})}$ and $p^{(\mathrm{R})}$ and define patient risk as $p_{\text{patient}} = \max\big(p^{(\mathrm{L})}, p^{(\mathrm{R})}\big)$, assuming malignancy in either breast determines patient risk; this max aggregation provides the baseline for the bilateral fusion head (Fig. 1.a).

The BilateralMixer: Patient-level prediction. A key component is the BilateralMixer, a patient-level fusion module that combines left and right breast embeddings using a shallow transformer, a learned asymmetry gate, and symmetric feature composition to model bilateral context, asymmetry, and concordance (Fig. 1.c).

Let $\mathbf{e}_L, \mathbf{e}_R \in \mathbb{R}^d$ denote the per-breast embeddings. BilateralMixer fuses them in three steps. *(1) Symmetric relational encoding:* we form the token sequence $[\mathrm{CLS}, \mathbf{e}_L, \mathbf{e}_R]$ and apply a compact transformer encoder; the CLS token aggregates bilateral information via self-attention, producing a patient embedding $\mathbf{c}$. The design is symmetric and does not privilege laterality. *(2) Gated asymmetry weighting:* an MLP predicts attention weights (α_L, α_R) from both embeddings and their absolute difference,

$$[\alpha_L, \alpha_R] = \mathrm{softmax}\big(f_\theta([\mathbf{e}_L, \mathbf{e}_R, |\mathbf{e}_L - \mathbf{e}_R|])\big),$$

allowing the model to upweight the more suspicious side while retaining bilateral context. *(3) Symmetric feature composition:* we concatenate

$$\mathbf{z} = \big[\mathbf{c},\ \alpha_L \mathbf{e}_L + \alpha_R \mathbf{e}_R,\ |\mathbf{e}_L - \mathbf{e}_R|,\ \mathbf{e}_L \odot \mathbf{e}_R\big],$$

where $|\mathbf{e}_L - \mathbf{e}_R|$ captures order-invariant asymmetry and $\mathbf{e}_L \odot \mathbf{e}_R$ captures concordance. An MLP-head then maps $\mathbf{z}$ to the patient-level malignancy logit.

2.3 Training

Per-Channel Augmentation We construct a pseudo-RGB input by applying three complementary intensity transforms to the same image. During training, channels receive independent stochastic brightness and contrast jitter, while a third channel applies CLAHE (12×12 tiles) to enhance low-contrast structures. At inference, this mapping is deterministic (no stochastic augmentation) and serves only as a fixed three-channel preprocessing step. An ablation across three input resolutions showed consistent improvements over naive channel replication.

Model Training In stage one (Fig. 1.a), we trained the hybrid fusion encoder using 512×512 mammograms with per-channel augmentations; validation and test images used only deterministic resizing, per-channel preprocessing, and normalisation. The DINOv3 ViT-S/16+ branch (self-supervised, pretrained on natural images) was kept frozen, while the ImageNet-initialised SE-ResNeXt101 branch and cross-attention fusion modules were trained to predict breast-level malignancy. Optimisation used AdamW with binary focal loss and early stopping on breast-level AUC. In stage two (Fig. 1.a), we froze the fusion encoder

Table 1. Overall AUC for 3-year risk prediction on in-distribution and out-of-distribution test sets, comparing single-stream baselines, Mirai and MamaDino variants (breast and patient level). Mirai is run with intended preprocessing; images-only inputs; reported metric uses patient-level 3-year risk score. We additionally report *resolution ablations* (training/testing at 224×224 and 320×320) and a *concatenation ablation* that replaces the BilateralMixer with a simple left/right embedding concatenation followed by a patient-level head.

		In-Distribution Test Set		**OOD Test Set**	
Model	Resolution	AUC	(95% CI)	AUC	(95% CI)
DINO-only (ablation)*	512×512	0.621	–	–	–
SEResNeXt-only (ablation) *	512×512	0.668	–	–	–
MamaDino (res. ablation)*	224×224	0.638	–	–	–
MamaDino (res. ablation)*	320×320	0.649	–	–	–
MamaDino † (concat. ablation)	512×512	0.731	–	–	–
Mirai† [26]	1664×2048	0.713	(0.686–0.740)	0.676	(0.646-0.707)
MamaDino * (ours)	512×512	0.727	(0.701–0.754)	0.666	(0.635-0.696)
MamaDino † (ours)	512×512	**0.736**	(0.710–0.762)	**0.677**	(0.647-0.707)

* Predictions were at breast-level, and patient risk corresponded to the max of the two breast scores.
† Predictions were at patient-level.

and trained only the BilateralMixer (Fig. 1.c) for patient-level prediction, again using AdamW, focal loss, and early stopping on patient-level AUC. Both stages ran on a single NVIDIA H200: stage 1 used batch size 32, learning rate 1×10^{-5}, up to 50 epochs; stage 2 used batch size 64, learning rate 5×10^{-7}, up to 25 epochs.

For comparison, we trained three baselines with the same preprocessing and optimisation: (i) DINO-only with a linear classifier on frozen features, (ii) SE-ResNeXt-only at 512×512, and (iii) hybrid breast-level fusion with max pooling across breasts at inference. Subgroup AUCs were computed for pre-specified strata (age, ethnicity, tumour type/grade, manufacturer) and reported where sample sizes were adequate.

3 Results

Prediction of 3-year Cancer Risk. We compare models using AUC (95% CI via DeLong's method) on matched in-distribution and OOD cohorts. In Tab. 1, we report 3-year fixed-horizon discrimination on the matched in-distribution and external OOD cohorts. On the in-distribution set, single-stream baselines (DINO-only, SE-ResNeXt-only) underperformed all hybrid variants. Mirai achieved AUCs of 0.713 (in-distribution) and 0.666 (OOD) at 1664×2048. At 512×512, breast-level MamaDino (max aggregation) reached 0.727 (in-distribution) and 0.666 (OOD), matching Mirai externally while using ∼13× fewer pixels. Adding the BilateralMixer improved patient-level AUC to 0.736 (in-distribution) and 0.677 (OOD), and was comparable to Mirai; external confidence intervals overlapped, indicating robustness under manufacturer shift.

Table 2. Subgroup AUCs for 3-year breast cancer risk prediction of MamaDino and Mirai on in-distribution and OOD (Oxford) test cohorts, stratified by age, ethnicity, scanner, cancer type and tumour grade (95% CI).

Test Cohorts	In-Distribution Test			OOD Test Set		
	N (cases)	MamaDino	Mirai	**N (cases)**	MamaDino	Mirai
Age (y)						
< 60	798 (266)	**0.746** (0.710-0.783)	0.727 (0.690-0.764)	910 (182)	0.668 (0.626-0.710)	**0.693** (0.651-0.735)
$60 - 65$	390 (130)	**0.705** (0.650-0.760)	0.656 (0.597-0.715)	524 (105)	**0.675** (0.618-0.732)	0.657 (0.596-0.717)
65+	387 (129)	**0.747** (0.696-0.797)	0.743 (0.691-0.794)	446 (89)	**0.691** (0.624-0.758)	0.672 (0.606-0.737)
Median*	59 [54–64]			60 [55–64]		
Ethnicity						
White	929 (306)	0.723 (0.689-0.758)	**0.726** (0.691-0.760)	–	–	–
Asian	131 (43)	**0.801** (0.723-0.880)	0.660 (0.556-0.764)	–	–	–
Black/African	48 (18)	0.770 (0.617-0.924)	**0.839** (0.728-0.950)	–	–	–
Scanner						
GE Med. S.	39 (13)	**0.725** (0.555-0.895)	0.627 (0.443-0.811)	605 (121)	**0.669** (0.618-0.721)	0.622 (0.566-0.677)
Hologic, Inc.	1509 (503)	**0.740** (0.713-0.766)	0.722 (0.694-0.749)	1275 (255)	0.679 (0.642-0.716)	**0.705** (0.669-0.741)
Siemens	27 (9)	**0.463** (0.217-0.709)	0.457 (0.222-0.691)	–	–	–
Cancer type**						
DCIS	(102)	0.696 (0.636-0.757)	**0.711** (0.650-0.771)	(62)	0.659 (0.586-0.731)	**0.689** (0.617-0.761)
Invasive	(400)	**0.748** (0.718-0.778)	0.713 (0.682-0.745)	(314)	**0.680** (0.647-0.714)	0.673 (0.640-0.707)
Tumour grade**						
G1	(78)	**0.755** (0.684-0.826)	0.686 (0.613-0.758)	(79)	**0.683** (0.621-0.745)	0.679 (0.614-0.745)
G2	(215)	**0.787** (0.751-0.824)	0.756 (0.717-0.796)	(143)	**0.695** (0.645-0.745)	0.660 (0.609-0.711)
G3	(92)	**0.641** (0.572-0.709)	0.633 (0.561-0.705)	(88)	0.664 (0.600-0.728)	**0.686** (0.623-0.749)

* Data are age medians, with IQRs in square brackets.
** Cancer data from 3-year follow-up.

Subgroup analysis Tab. 2 summarises patient-level AUCs by age, ethnicity, manufacturer, cancer type and grade. MamaDino was broadly stable across age, and on OOD data matched or exceeded Mirai in older groups while Mirai was stronger under 60. Across ethnicity strata (in-distribution), performance was similar in White women, with larger but imprecise differences in Asian and Black groups due to small samples. By vendor, MamaDino slightly improved over Mirai on Hologic and was comparable on Siemens; on GE it was consistently superior in both cohorts. Both models discriminated invasive cancers better than DCIS; MamaDino improved for invasive disease on both cohorts, with comparable DCIS performance in-distribution and a small drop OOD. By tumour grade, MamaDino generally matched or exceeded Mirai, with the largest gains in low/intermediate grades and similar performance in high-grade disease.

Ablations Evaluating MamaDino across three input resolutions showed a consistent increase in discrimination with higher spatial resolution, suggesting that fine-scale mammographic structure remains informative for short-term risk prediction. We further compared the proposed BilateralMixer with a simpler patient-level baseline that concatenates left and right breast embeddings and applies a shallow prediction head; this concatenation ablation underperformed the full BilateralMixer, indicating that the gain arises from structured contralateral fusion rather than merely exposing both breasts to the classifier. Mirai was evaluated only at its native resolution; nevertheless, these results suggest that aggressive downsampling is unlikely to improve image-based risk models, and motivate a more systematic resolution study as future work.

4 Discussion

MamaDino achieved 3-year risk discrimination comparable to, and in some settings exceeding, MIRAI while operating at 512×512 resolution ($\sim$13$\times$ fewer input pixels). At breast level, the hybrid encoder with max aggregation matched MIRAI on the external cohort (AUC 0.666), and adding bilateral fusion improved patient-level discrimination in both in-distribution and OOD evaluation (AUC 0.736 and 0.677, respectively). Ablations indicate that these gains are not attributable to any single component. Single-stream variants were weaker (DINO-only 0.621; SE-ResNeXt-only 0.668), supporting the value of combining global self-supervised priors with mammography-specific texture modelling. Bilateral modelling also mattered, although improvement was small in isolation. Replacing the BilateralMixer with simple left/right concatenation reduced AUC (0.731 vs. 0.736), indicating BilateralMixer's role as one component in a stacked design rather than a standalone source of improvement. Resolution ablations showed a monotonic drop at 320×320 and 224×224 (0.649 and 0.638), suggesting robustness to reduced resolution but a clear benefit from retaining spatial detail.

These findings support the view that structured contralateral integration and complementary representation learning are central to short-term mammographic risk prediction, with strong performance achievable even under constrained input resolution. Lower resolution was used here primarily as a controlled stress-test of representation and fusion choices, rather than to argue against high-resolution mammography. Notably, MamaDino uses imaging alone, whereas MIRAI requires structured clinical covariates; comparable performance under this constraint suggests that clinically meaningful short-term risk signal is substantially encoded in mammographic phenotype, potentially enabling deployment where risk factors are missing or inconsistently recorded.

Across cohorts, discrimination was broadly stable under scanner vendor shift (notably Hologic vs. GE), consistent with self-supervised transformer priors improving domain robustness. As in prior work, invasive cancers were easier to predict than DCIS, implying weaker pre-diagnostic imaging signatures for in-situ disease, and may also be consistent with downsampling attenuating fine-scale signals such as microcalcifications.

Limitations include restriction to UK screening services and matched case–control sampling, limiting calibration assessment and generalisation to other screening intervals or populations. Siemens scanners and non-White groups were under-represented. We evaluated the released MIRAI model without retraining, as its original training procedure relies on structured clinical covariates and a multi-stage pipeline unavailable in OPTIMAM; consequently, resolution and architectural effects cannot be fully disentangled. Future work should include mammography-specific self-supervised pretraining, full or parameter-efficient DINOv3 fine-tuning, multinational prospective evaluation, and scaling the same design to high-resolution inputs to quantify the accuracy–compute trade-off.

5 Conclusion

We introduced MamaDino, a hybrid model that fuses frozen DINOv3 features, trainable SE-ResNeXt encoders, and a bilateral fusion head for 3-year breast cancer risk prediction. At a compute-constrained 512×512 resolution, MamaDino was comparable, with overlapping CIs, to Mirai's discrimination on in-distribution and OOD UK cohorts, with broadly stable performance across demographics, tumour characteristics, and scanner vendors. Importantly, our use of lower resolution is intended to *isolate and validate* the value of complementary inductive biases and explicit contralateral modelling under a constrained pixel budget, rather than to argue against high-resolution imaging. Overall, MamaDino supports the thesis that complementary inductive biases and explicit contralateral modelling shift emphasis from single-backbone processing toward structured multi-view and bilateral fusion as key components of personalised risk prediction.

Acknowledgments. The images and data used in this publication are derived from the OPTIMAM imaging database [12], we would like to acknowledge the OPTIMAM project team and staff at the Royal Surrey NHS Foundation Trust who developed the OPTIMAM database, and Cancer Research UK who funded the creation and maintenance of the database.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alterson, R., Plewes, D.B.: Bilateral symmetry analysis of breast mri. Physics in Medicine & Biology **48**(20), 3431 (2003)
2. Brentnall, A.R., Atakpa, E.C., Hill, H., Santeramo, R., Damiani, C., Cuzick, J., Montana, G., Duffy, S.W.: An optimization framework to guide the choice of thresholds for risk-based cancer screening. NPJ Digital Medicine **6**(1), 223 (2023)
3. Brentnall, A.R., Harkness, E.F., Astley, S.M., Donnelly, L.S., Stavrinos, P., Sampson, S., Fox, L., Sergeant, J.C., Harvie, M.N., Wilson, M., et al.: Mammographic density adds accuracy to both the tyrer-cuzick and gail breast cancer risk models in a prospective uk screening cohort. Breast Cancer Research **17**(1), 147 (2015)
4. Claus, E.B., Risch, N., Thompson, W.D.: Genetic analysis of breast cancer in the cancer and steroid hormone study. American journal of human genetics **48**(2), 232 (1991)
5. Dai, Z., Liu, H., Le, Q.V., Tan, M.: CoAtNet: Marrying convolution and attention for all data sizes. arXiv preprint arXiv:2106.04803 (2021). https://doi.org/10.48550/arXiv.2106.04803, `https://arxiv.org/abs/2106.04803`
6. d'Ascoli, S., Touvron, H., Leavitt, M., Morcos, A., Biroli, G., Sagun, L.: ConViT: Improving vision transformers with soft convolutional inductive biases. arXiv preprint arXiv:2103.10697 (2021). https://doi.org/10.48550/arXiv.2103.10697, `https://arxiv.org/abs/2103.10697`

7. Donnelly, J., Moffett, L., Barnett, A.J., Trivedi, H., Schwartz, F., Lo, J., Rudin, C.: Asymmirai: interpretable mammography-based deep learning model for 1–5-year breast cancer risk prediction. Radiology **310**(3), e232780 (2024)
8. Ellis, S., Gomes, S., Trumble, M., Halling-Brown, M.D., Young, K.C., Chaudhry, N.S., Harris, P., Warren, L.M.: Deep learning for breast cancer risk prediction: application to a large representative uk screening cohort. Radiology: Artificial Intelligence **6**(4), e230431 (2024)
9. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., Mulvihill, J.J.: Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. JNCI: Journal of the National Cancer Institute **81**(24), 1879–1886 (1989)
10. Gao, Y., Li, H., Yuan, F., Wang, X., Gao, X.: Dino u-net: Exploiting high-fidelity dense features from foundation models for medical image segmentation. arXiv preprint arXiv:2508.20909 (2025)
11. Goyal, A., Bengio, Y.: Inductive biases for deep learning of higher-level cognition. Proceedings of the Royal Society A **478**(2266), 20210068 (2022)
12. Halling-Brown, M.D., Warren, L.M., Ward, D., Lewis, E., Mackenzie, A., Wallis, M.G., Wilkinson, L.S., Given-Wilson, R.M., McAvinchey, R., Young, K.C.: Optimam mammography image database: a large-scale resource of mammography images and clinical data. Radiology: Artificial Intelligence **3**(1), e200103 (2020)
13. Lauritzen, A.D., von Euler-Chelpin, M.C., Lynge, E., Vejborg, I., Nielsen, M., Karssemeijer, N., Lillholm, M.: Assessing breast cancer risk by combining ai for lesion detection and mammographic texture. Radiology **308**(2), e230227 (2023)
14. Liu, C., Chen, Y., Shi, H., Lu, J., Jian, B., Pan, J., Cai, L., Wang, J., Zhang, Y., Li, J., et al.: Does dinov3 set a new medical vision standard? arXiv preprint arXiv:2509.06467 (2025)
15. Manigrasso, F., Milazzo, R., Russo, A.S., Lamberti, F., Strand, F., Pagnani, A., Morra, L.: Mammography classification with multi-view deep learning techniques: Investigating graph and transformer-based architectures. Medical Image Analysis **99**, 103320 (2025)
16. Mendes, J., Oliveira, B., Araújo, C., Galrão, J., Mota, A.M., Garcia, N.C., Matela, N.: Deep learning in breast cancer risk prediction: a review of recent applications in full-field digital mammography. Frontiers in Oncology **15**, 1656842 (2025)
17. Mohamed, A., Fakhry, S., Basha, T.: Bilateral analysis boosts the performance of mammography-based deep learning models in breast cancer risk prediction. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1440–1443. IEEE (2022)
18. Omoleye, O.J., Woodard, A.E., Howard, F.M., Zhao, F., Yoshimatsu, T.F., Zheng, Y., Pearson, A.T., Levental, M., Aribisala, B.S., Kulkarni, K., et al.: External evaluation of a mammography-based deep learning model for predicting breast cancer in an ethnically diverse population. Radiology: Artificial Intelligence **5**(6), e220299 (2023)
19. Pashayan, N., Antoniou, A.C., Ivanus, U., Esserman, L.J., Easton, D.F., French, D., Sroczynski, G., Hall, P., Cuzick, J., Evans, D.G., et al.: Personalized early detection and prevention of breast cancer: Envision consensus statement. Nature reviews Clinical oncology **17**(11), 687–705 (2020)
20. Santeramo, R., Damiani, C., Wei, J., Montana, G., Brentnall, A.R.: Are better ai algorithms for breast cancer detection also better at predicting risk? a paired case–control study. Breast Cancer Research **26**(1), 25 (2024)

21. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
22. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **71**(3), 209–249 (2021)
23. Tyrer, J., Duffy, S.W., Cuzick, J.: A breast cancer prediction model incorporating familial and personal risk factors. Statistics in medicine **23**(7), 1111–1130 (2004)
24. Wang, X., Tan, T., Gao, Y., Han, L., Zhang, T., Lu, C., Beets-Tan, R., Su, R., Mann, R.: Disasymnet: Disentanglement of asymmetrical abnormality on bilateral mammograms using self-adversarial learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 57–67. Springer (2023)
25. Yala, A., Lehman, C., Schuster, T., Portnoi, T., Barzilay, R.: A deep learning mammography-based model for improved breast cancer risk prediction. Radiology **292**(1), 60–66 (2019)
26. Yala, A., Mikhael, P.G., Strand, F., Lin, G., Smith, K., Wan, Y.L., Lamb, L., Hughes, K., Lehman, C., Barzilay, R.: Toward robust mammography-based models for breast cancer risk. Science Translational Medicine **13**(578), eaba4373 (2021). https://doi.org/10.1126/scitranslmed.aba4373, `https://www.science.org/doi/10.1126/scitranslmed.aba4373`, this paper introduces the Mirai model for multi-timepoint breast cancer risk prediction.
27. Zhou, Z., Arefan, D., Zuley, M., Sumkin, J., Wu, S.: Sta-risk: A deep dive of spatio-temporal asymmetries for breast cancer risk prediction. arXiv preprint arXiv:2505.21699 (2025)