



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

MammoFlow: Multiview Mammogram Synthesis with Anatomically Consistent Flow Matching

Yuexi Du¹, Leya Barrientos², Laura Sheiman², John Lewin²,
Hemant Tagare^{1,2}, and Nicha C. Dvornek^{1,2}

¹ Department of Biomedical Engineering, Yale University,
² Department of Radiology & Biomedical Imaging, Yale University,
{yuexi.du, nicha.chitphakdithai}@yale.edu

Abstract. Multiview mammography relies on paired craniocaudal (CC) and mediolateral oblique (MLO) views to provide complementary projections of a 3D breast volume, enabling precise anomaly localization. However, acquiring high-quality, balanced datasets remains challenging for deep learning applications. We propose a novel method to synthesize multiview mammograms by leveraging the inherent geometric relationship between CC and MLO views. To enforce an implicit 3D consistency prior during generation, we develop an alignment module that searches a 2D affine transformation subspace to establish optimal anatomical correspondence. Leveraging this alignment, we introduce a pixel-space self-consistency loss based on the Earth Mover’s Distance (EMD) between the 1D anteroposterior (AP) axis tissue distributions of the generated images. Integrated into a pretrained flow matching model, MammoFlow forces synthesized pairs to share physically plausible tissue distributions from the chest wall to the nipple. To our knowledge, this is the first work to guide multiview mammogram generation using implicit geometric tissue correspondence. Our method demonstrates superior image quality, passes expert radiologist evaluation, and generates physically consistent pairs that improve downstream classification AUC by 5%.³

Keywords: Multiview Mammography · Image Synthesis · Anatomic Consistency

1 Introduction

Breast cancer remains a leading cause of cancer-related mortality among women worldwide [20]. As the gold standard for screening, full-field digital mammography (FFDM) is widely adopted due to its accessibility and sensitivity. During screening, radiologists rely on paired craniocaudal (CC) and mediolateral oblique (MLO) views to precisely localize anomalies, benefiting from the complementary perspectives these projections provide (Fig. 1(a)). This dual-view foundation is also critical for modern computer-aided diagnosis systems [24,25,16,8,4]. While data-driven deep learning methods have achieved remarkable success in FFDM

³ Code and pretrained model: <https://github.com/XYPB/MammoFlow>.

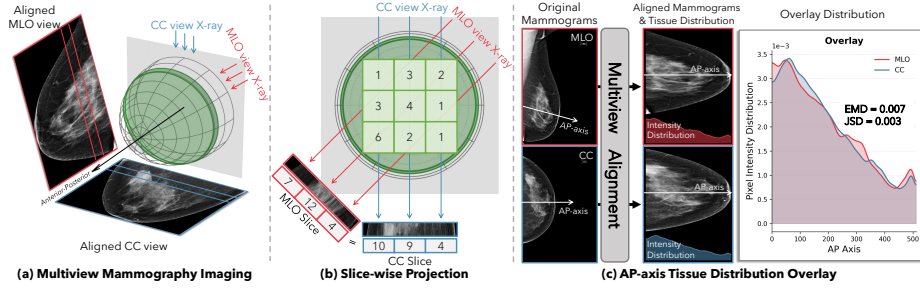


Fig. 1. Multiview Mammography (a) Multiview mammograms project a shared 3D volume. (b) Total tissue intensity is conserved across corresponding anteroposterior (AP) slices. (c) Aligned AP-axis distributions demonstrate high cross-view correlation.

analysis [3], acquiring high-quality, paired datasets remains hindered by privacy concerns and annotation costs, and the number of cancer cases comprises a small fraction of the curated samples. Consequently, limited and highly imbalanced data restrict the further development of deep learning applications in mammography.

Standard data augmentation, such as affine transformations and tumor artifacting [23], is commonly applied to alleviate data scarcity [18]. However, these augmentations operate based on GT images and cannot introduce true morphological diversity. Generative image synthesis has thus emerged as a promising alternative to create plausible mammograms for training downstream models [9]. Yet, existing approaches [1,6] primarily focus on single-view generation, neglecting the inherent multiview nature of clinical practice. While view-to-view translation frameworks like CA3D-Diff [12] address this gap, they still require a ground-truth reference view. A recent study, Mammo-RGB [7], attempts simultaneous dual-view generation by stacking views in color channels; however, it ignores explicit 3D anatomical relationships, which often result in cross-view artifacts.

In this work, we propose MammoFlow, a novel multiview mammogram synthesis method guided by the explicit anatomical correlation between CC and MLO views. As illustrated in Fig. 1(b), the 2D projection along the anteroposterior (AP) axis in both views originates from the same 3D breast volume; thus, the sum of their tissue intensities along this axis should be largely conserved. Inspired by this projection prior, we introduce an alignment module that searches a subspace of 2D affine transformations to spatially align the CC and MLO projections (Fig. 1(c)). Leveraging this alignment, we design a novel self-consistency loss based on the Earth Mover’s Distance (EMD). By formulating EMD as a differentiable geometric constraint, we minimize the distributional divergence of tissue along the AP axis during training. Integrating this module into a large-scale rectified flow model [5] allows our method to generate high-fidelity, dual-view mammograms that maintain 3D anatomical consistency. Extensive quantitative analysis and expert radiologist evaluations demonstrate the superiority of our generated images, which further show a non-trivial improvement in downstream classification performance when utilized as additional training data.

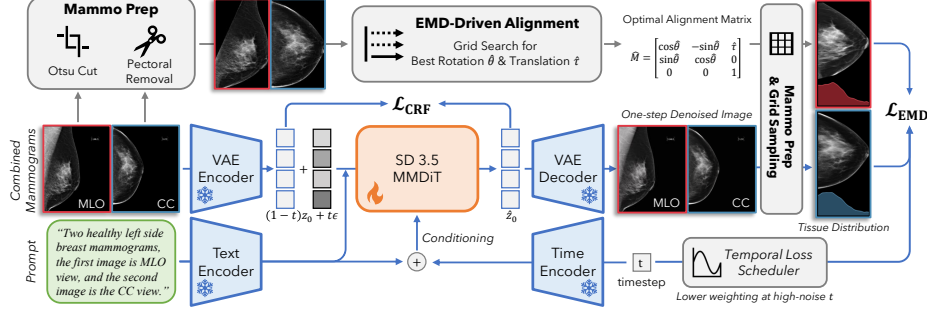


Fig. 2. Proposed Multiview Mammogram Synthesis Pipeline We propose to use the tissue correspondence relationship of the ground-truth images to guide the rectified flow (RF) training process. Besides the Conditional RF loss $\mathcal{L}_{CRF}$, we further optimize the Earth-Mover’s Distance (EMD) loss over the denoised images. We use the Temporal Loss Scheduler to reduce the influence of a blurry reconstructed image during training.

2 Methods

We first introduce the Conditional Flow Matching (CFM) image synthesis backbone of MammoFlow, followed by the novel EMD-driven view alignment module and the multiview EMD constraint loss. An overview of MammoFlow is in Fig. 2. During training, each raw CC/MLO pair and its text prompt are used for the CRF objective, and the real paired images provide the precomputed alignment used to evaluate $\mathcal{L}_{EMD}$ on the one-step reconstructed pair. During inference, the trained model takes only Gaussian noise and a prompt to generate novel paired views; no ground-truth image or alignment is required.

2.1 Conditional Flow Matching

Our multiview mammogram synthesis framework builds upon the conditional rectified flow (CRF) formulation [14,13], which constructs straight paths between the data distribution and a standard normal distribution. Given a combined multiview mammogram $x = [x_{MLO}, x_{CC}]$ and its corresponding text condition y (Fig. 2), we extract the clean latent representation z_0 using a pre-trained variational autoencoder (VAE) [10]. The linear probability path connecting z_0 to standard Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$ is defined as:

$$z_t = (1 - t)z_0 + t\epsilon, \quad t \in [0, 1] \quad (1)$$

where t represents the continuous timestep. The denoising model, parameterized as $\mathbf{v}_\theta(z_t, t, y)$, is trained to predict the velocity of this deterministic flow [14,5]. Since the target velocity is $\frac{\partial z_t}{\partial t} = \epsilon - z_0$, the CRF objective minimizes the expected squared difference between the predicted velocity and this target trajectory:

$$\mathcal{L}_{CRF} = \mathbb{E}_{z_0, \epsilon, t, y} [\|\mathbf{v}_\theta(z_t, t, y) - (\epsilon - z_0)\|^2] \quad (2)$$

During inference, the model generates novel samples by solving the corresponding ordinary differential equation (ODE). Because the defined flow is linear, the predicted velocity directly points toward the clean data, allowing us to approximate the one-step denoised latent $\hat{z}_0$ at any timestep t :

$$\hat{z}_0 = z_t - t\mathbf{v}_\theta(z_t, t, y) \quad (3)$$

2.2 EMD-Driven Alignment

Unlike CC views, MLO projections are captured at an angle and routinely include the pectoral muscle to visualize the chest wall. Consequently, the anatomical anteroposterior (AP) axis in the MLO view is inherently tilted relative to the image grid. Furthermore, view-specific, non-rigid breast compression introduces additional morphological variations. These characteristics make the direct comparison of spatial intensity values across raw mammograms geometrically inconsistent.

To resolve this, we propose an EMD-driven alignment module (Fig. 2). We first preprocess the raw images by removing text labels via Otsu thresholding, cropping the image to the breast region. For the MLO view, we apply the Hough transform to detect the pectoral muscle boundary and mask this region with zero intensity to prevent it from skewing the tissue distribution.

Next, we align the two views such that their AP axes are parallel to the horizontal image axis. Based on standard imaging protocols, we can assume that the CC view’s AP axis is already horizontally aligned [21]. Therefore, we restrict our geometric alignment to a 2D affine subspace consisting of a rotation angle θ and a horizontal translation τ applied to the MLO view. Since we trimmed the excess region and resized the MLO image after the affine transform, the stretch term is also considered implicitly. We denote this affine transformation as $\mathcal{A}(x_{MLO}; \theta, \tau)$. To evaluate the alignment quality, we measure the divergence between the 1D AP-axis tissue distributions. Let $\Phi(\cdot)$ denote the differentiable 1D projection operator, which sums the raw pixel intensities vertically at each position along the AP axis and normalizes by total image intensity. To account for non-rigid tissue deformation, $\Phi(\cdot)$ includes a horizontal 1D Gaussian smoothing filter, providing spatial relaxation. We utilize the Earth Mover’s Distance (EMD) to quantify the tissue correlation, which is defined as:

$$\text{EMD}(p, q) = \sum_{i=1}^W |P[i] - Q[i]| \quad (4)$$

where P and Q are the cumulative distribution functions of the normalized tissue distributions p and q , respectively, and W is the image width. Unlike Jensen-Shannon divergence (JSD), EMD explicitly penalizes the physical displacement of tissue along the AP axis, ensuring continuous and stable training gradients.

Because there are only two degrees of freedom (θ, τ) , we formulate the alignment as a discrete optimization problem, using grid search to find the optimal parameters $\hat{\theta}$ and $\hat{\tau}$ that minimize the distributional distance between the views:

$$(\hat{\theta}, \hat{\tau}) = \arg \min_{\theta, \tau} \text{EMD}(\Phi(x_{CC}), \Phi(\mathcal{A}(x_{MLO}; \theta, \tau))) \quad (5)$$

By heuristically bounding the search ranges for θ and τ based on physical imaging protocols, we ensure the search space is highly compressed, resulting in a computationally lightweight alignment process.

2.3 Multiview EMD Regularization

The EMD-driven alignment module provides the optimal transformation parameters $(\hat{\theta}, \hat{\tau})$ for the ground-truth image pair. We leverage this geometric prior to guide the training of the rectified flow model. Because the physical tissue constraint must be evaluated in the image domain, we first decode the predicted clean latent $\hat{z}_0$ to obtain reconstructed dual-view mammograms $\hat{x} = [\hat{x}_{CC}, \hat{x}_{MLO}]$.

To ensure the synthesized views maintain anatomically consistent 3D breast anatomy, we optimize the EMD between the generated views during training. Since the prediction $\hat{x}$ is spatially anchored to the ground-truth x conditioning, we can directly apply the precomputed optimal alignment parameters $(\hat{\theta}, \hat{\tau})$ to the generated MLO view. Similarly, we reuse the precomputed Otsu thresholding mask and pectoral region mask so that this process is differentiable. We implement this affine transformation $\mathcal{A}(\cdot)$ using differentiable grid sampling, allowing the gradient to flow back through the VAE decoder. The self-consistency EMD loss for the generated mammogram pair is defined as:

$$\mathcal{L}_{\text{EMD}} = \text{EMD}(\Phi(\hat{x}_{CC}), \Phi(\mathcal{A}(\hat{x}_{MLO}; \hat{\theta}, \hat{\tau}))) \quad (6)$$

Temporal Loss Scheduler. During CRF training, the model predicts the clean latent $\hat{z}_0$ in a single step from timestep t . While the deterministic nature of flow matching ensures the velocity vector points toward the original data, one-step predictions from high-noise timesteps (large t) often lack structural fidelity. Because the EMD loss relies on anatomical geometry, applying it to highly blurry predictions is uninformative and can destabilize training. To mitigate the influence of poorly structured $\hat{x}$ at high noise levels, we introduce a temporal loss scheduler. We apply a cosine decay schedule based on the continuous timestep $t \in [0, 1]$ to suppress the EMD penalty at large t and upweight it as t approaches 0. Our final optimization objective is thus formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{CRF}} + \frac{\lambda}{2} (\cos(\pi t) + 1) \mathcal{L}_{\text{EMD}}, \quad (7)$$

where λ is the weight of $\mathcal{L}_{\text{EMD}}$. The $\mathcal{L}_{\text{EMD}}$ is effective only during training, making it a plug-and-play module for general flow-matching models. Moreover, our method brings no extra cost during inference.

3 Experiments

3.1 Experimental Settings

Datasets. We evaluate our method on three public mammography datasets: (1) CSAW [19] contains 98k paired mammograms, including 740 cases with visible

Table 1. Image Generation Results. We report the FID and FrD for image quality and the relative difference (%) of mean EMD and JSD from the ground truth values to measure multiview correlation. * indicates the method requires a ground truth image as a reference. Best synthetic results are in bold.

Method	CSAW				VinDr				RSNA			
	FID	FrD	Δ EMD	Δ JSD	FID	FrD	Δ EMD	Δ JSD	FID	FrD	Δ EMD	Δ JSD
GT (Oracle)	8.89	3.02	-	-	8.36	5.42	-	-	9.76	12.0	-	-
Rand. Shuff.	-	-	72.14	123.54	-	-	61.35	83.95	-	-	64.92	83.91
CA3D-Diff*	63.3	8.72	14.12	16.89	70.5	6.72	173.83	498.86	64.5	8.73	37.20	90.98
Mammo-RGB	88.6	14.9	60.67	188.53	102.8	21.5	33.01	49.57	148.6	15.1	15.43	20.98
Vanilla (w/o $\mathcal{L}_{\text{EMD}}$)	73.4	17.7	10.82	68.52	70.5	48.8	18.37	54.61	66.3	16.4	7.75	28.86
MammoFlow	53.3	6.60	1.08	9.57	67.5	12.4	4.19	15.22	65.4	12.7	2.73	33.48

Table 2. Ablation Study. Performance of different multiview designs on CSAW. Best results are in bold; second best are underlined.

$\mathcal{L}_{\text{EMD}}$	Temporal Loss Scheduler	Gaussian Smoothing	λ	FID	FrD	Δ EMD	Δ JSD
-	-	-	-	60.41	6.95	10.82	68.52
✓	-	✓	0.10	64.71	14.26	18.39	<u>15.06</u>
✓	cosine	-	0.10	74.67	51.72	10.82	193.54
✓	linear	✓	0.10	<u>54.54</u>	6.59	14.40	22.27
✓	cosine	✓	1.00	83.42	7.22	10.62	104.66
✓	cosine	✓	0.10	53.32	6.60	1.08	9.57
✓	cosine	✓	0.01	66.93	<u>12.52</u>	24.80	358.77

cancer based on segmentation masks (80%/20% train/test split). (2) **VinDr** [15] includes 20k paired mammograms with $\sim$ 1k cancer cases (official train/test split). (3) **RSNA** [2] provides 54k images and $\sim$ 1k cancer cases (85%/15% train/test split). All images are resized to 512×512 .

Implementation. We use a pre-trained Stable Diffusion 3.5 Medium [5] backbone. We freeze all components except the Multimodal Diffusion Transformer (MMDiT) at training. Multiview mammograms are spatially concatenated along the horizontal axis. We fine-tune the model in half-precision on a single NVIDIA H200 GPU using the AdamW optimizer, a constant learning rate of 1×10^{-6} , and a batch size of 16 for 40k steps. For the EMD-driven alignment, the search space for θ is $[-10^\circ, 10^\circ]$ with a 2° step size and for τ is $[-64, 64]$ pixels with an 8-pixel step size. The initial $\mathcal{L}_{\text{EMD}}$ weight is set to $\lambda = 0.1$ and applied with a cosine decay scheduler. Inference is performed using 100 steps, sampling 1k images for downstream evaluation. A fixed random seed of 42 is used throughout the experiment. We train the downstream classifier with a learning rate of 1×10^{-3} , a weight decay of 0.1, and a batch size of 32 for 25 epochs on each dataset.

Baselines. We compare our approach against **CA3D-Diff** [12], a view-to-view translation model requiring an oracle reference, and **Mammo-RGB** [7], a simultaneous multiview generation method reproduced by ourselves with the same backbone. We also include a **Vanilla** baseline without the proposed $\mathcal{L}_{\text{EMD}}$.

Metrics. Models are evaluated on two aspects: (1) *Image Quality*, measured via **FID** [17] for overall image quality and **FrD** [11] for radiomic feature consistency

Table 3. Cancer Classification. We add different amounts of synthetic cancer images for training breast cancer classification and report AUC-ROC on the original test set. Best results are in bold.

Training Data	CSAW	VinDr	RSNA
Real	.7452	.7775	.7676
Real + 1k Synth.	.7899	.8055	.7711
Real + 5k Synth.	.7904	.8196	.7822

Table 4. Reader Study. We report ratio (mean $\pm$ standard deviation) of identifying the multiview mammograms as real and paired.

Method	Authenticity	Pairing
GT	97.5 (± 2.5)	100.0 (± 0.0)
CA3D-Diff	2.5 (± 2.5)	82.5 (± 2.5)
Mammo-RGB	0.0 (± 0.0)	95.0 (± 5.0)
MammoFlow	37.5 (± 12.5)	97.5 (± 2.5)

relevant to diagnosis; (2) *Multiview Correspondence*, evaluated as the relative percentage difference in mean divergence between the generated and the ground truth (GT) AP-axis tissue distribution: $\Delta D = 100\% \times \frac{|\text{mean}(D_{syn}) - \text{mean}(D_{GT})|}{\text{mean}(D_{GT})}$, where D denotes either EMD or JSD. We apply affine alignment only when computing EMD/JSD for paired images; all other evaluations use raw synthetic images.

Downstream Tasks. To evaluate practical utility, we fine-tune our model on malignant cases to synthesize cancerous data distributions, where the prompt controls the malignancy. We then train an EfficientNet-B2 [22] multiview binary cancer classifier (concatenating view features into an MLP) by augmenting the real dataset with 1k and 5k synthetic malignant pairs. Additionally, two expert radiologists performed a blinded reader study on 80 multiview mammograms (20 pairs per method). For each pair, they assessed authenticity (real vs. synthetic) and anatomical pairing correctness across views.

3.2 Results

Quantitative Evaluation. Main results are in Tab. 1. Computing FID and FrD between two subsets of the GT test set establishes an oracle lower bound. To validate our multiview alignment metrics, we evaluate randomly shuffled GT pairs, observing a significant increase in both ΔEMD and ΔJSD ($p < 0.0001$, Mann-Whitney U test), confirming these metrics accurately capture geometric misalignment. Our MammoFlow consistently outperforms other multiview generation methods in both image quality and multiview correspondence. Notably, no significant difference was found comparing the EMD distribution of our synthetic pairs against the ground truth on the CSAW dataset ($p = 0.3089$, Mann-Whitney U test), confirming the high geometric fidelity of our generations. The addition of $\mathcal{L}_{\text{EMD}}$ not only improves cross-view tissue correlation but also improves overall radiomic fidelity compared to the vanilla baseline. While CA3D-Diff achieves slightly lower FrD scores on the VinDr and RSNA datasets, this is expected as it relies on an oracle ground truth reference image. However, CA3D-Diff still exhibits inferior multiview correspondence because it assumes a fixed 45° rigid rotation between views, ignoring the patient-specific adjustments made in clinical practice. Mammo-RGB fails to match our alignment performance, as stacking views in color channels inherently ignores explicit spatial correlation.

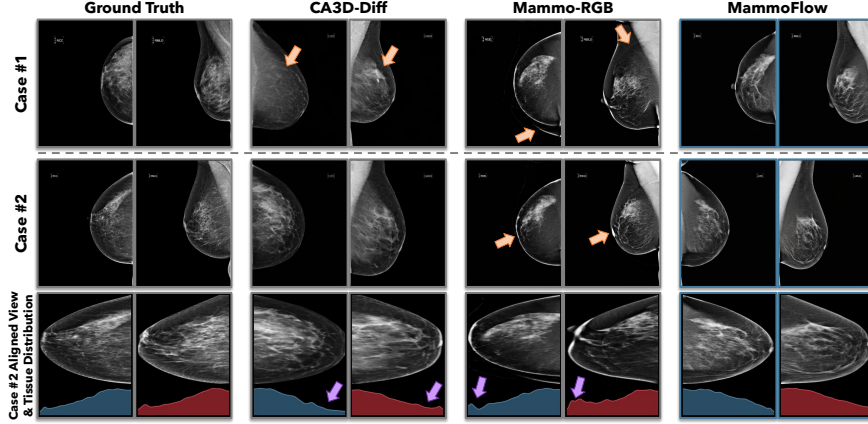


Fig. 3. Qualitative Results. We visualize two random CC/MLO pairs generated from Gaussian noise with the same prompt for each method. The first column shows GT images. Rows 1–2 show synthetic cases, and the final row plots the aligned view and AP-axis tissue distribution for case 2. The orange arrow highlights artifacts and mismatched anatomical structure. The purple arrow highlights mismatched tissue distribution.

Ablation Study. Tab. 2 highlights the contribution of each component. Unconstrained paired generation (w/o $\mathcal{L}_{EMD}$) degrades both structural anatomy and image quality. Applying the EMD loss with a constant weight distorts breast structures due to harmful gradient signals at high-noise timesteps. Removing the Gaussian spatial relaxation limits the ability to handle non-rigid tissue deformation. Among temporal schedulers, cosine decay performs best by mainly penalizing structural divergence only at low-noise timesteps. $\mathcal{L}_{EMD}$ weight of $\lambda = 0.1$ achieves the best balance between visual fidelity and geometric alignment.

Downstream Classification. As reported in Tab. 3, incorporating our synthetic images improves the breast cancer classification AUC by up to 5% on CSAW, demonstrating the high diagnostic value of our generated anatomies.

Reader Study. Tab. 4 shows experts can easily identify GT images. Among the generative models, MammoFlow was the most deceptive, with 37.5% of images rated as real. Furthermore, our generated pairs achieved a 97.5% pairing success rate, surpassing the oracle-guided CA3D-Diff framework.

Qualitative Results. Visual comparisons in Fig. 3 confirm that our method demonstrates the most anatomically consistent multiview generation. Baselines frequently synthesize artifacts (*e.g.*, contours of the other view in Mammo-RGB samples) or localized structures (*e.g.*, high-density tissue) in one view that are absent in the other (orange arrows). The plotted AP-axis distributions further demonstrate this: while our synthesized distributions closely mirror the native mass distributions of the GT data, other methods show obvious tissue density mismatches along the projection axis (purple arrows).

4 Discussion and Conclusion

We presented MammoFlow, a multiview mammogram synthesis framework leveraging a flow-matching model and implicit 3D tissue guidance. By embedding an EMD-driven alignment module, our method minimizes cross-view spatial divergence, generating anatomically consistent pairs that significantly improve downstream classification.

Limitations include increased training GPU memory ($\sim 20\%$) due to pixel-space constraints requiring VAE backpropagation. Furthermore, our 2D affine alignment approximates complex 3D compressions, and classical pectoral masking struggles with very dense anatomies. Future work will optimize computational efficiency and model fully non-rigid 3D deformations across views.

Acknowledgments This work was supported by NIH grant R21EB032950.

Disclosure of Interests The authors have no competing interests in this work and other related research.

References

1. Montoya-del Angel, R., Sam-Millan, K., Vilanova, J.C., Martí, R.: Mam-e: Mammographic synthetic image generation with diffusion models. *Sensors* **24**(7), 2076 (2024)
2. Carr, C., FelipeKitamura, MD, P., Partridge, G., inversion, Kalpathy-Cramer, J., Mongan, J., Andriole, K., Lavender, Vazirabad, M., Riopel, M., Ball, R., Dane, S., Chen, Y.: Rsn screening mammography breast cancer detection. <https://kaggle.com/competitions/rsna-breast-cancer-detection> (2022), kaggle
3. Chang, Y.W., Ryu, J.K., An, J.K., Choi, N., Park, Y.M., Ko, K.H., Han, K.: Artificial intelligence for breast cancer screening in mammography (ai-stream): preliminary analysis of a prospective multicenter cohort study. *Nature Communications* **16**(1), 2248 (2025)
4. Du, Y., Chen, L., Dvornek, N.C.: Geometry-guided local alignment for multi-view visual language pre-training in mammography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 299–310. Springer (2025)
5. Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., et al.: Scaling rectified flow transformers for high-resolution image synthesis. In: Forty-first international conference on machine learning (2024)
6. Garrucho, L., Kushibar, K., Osuala, R., Diaz, O., Catanese, A., Del Riego, J., Bobowicz, M., Strand, F., Igual, L., Lekadir, K.: High-resolution synthesis of high-density breast mammograms: Application to improved fairness in deep learning based mass detection. *Frontiers in oncology* **12**, 1044496 (2023)
7. Garza-Abdala, J.A., Fumagal-González, G.A., Avendano, D., Cardona, S., Hussain, S., de Avila-Armenta, E., Toscano-Martínez, J.H., Gurmendi, D.S., Pedro-Pérez, A.A., Tamez-Pena, J.G.: Mammorgb: Dual-view mammogram synthesis using denoising diffusion probabilistic models. *arXiv preprint arXiv:2511.22759* (2025)
8. Ghosh, S., Poynton, C.B., Visweswaran, S., Batmanghelich, K.: Mammo-clip: A vision language foundation model to enhance data efficiency and robustness in mammography. In: International conference on medical image computing and computer-assisted intervention. pp. 632–642. Springer (2024)

9. Heng, Y., Yinghua, M., Khan, F.G., Khan, A., Ali, F., AlZubi, A.A., Hui, Z.: Survey: application and analysis of generative adversarial networks in medical images. *Artificial Intelligence Review* **58**(2), 39 (2024)
10. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013)
11. Konz, N., Osuala, R., Verma, P., Chen, Y., Gu, H., Dong, H., Chen, Y., Marshall, A., Garrucho, L., Kushibar, K., Lang, D.M., Kim, G.S., Grimm, L.J., Lewin, J.M., Duncan, J.S., Schnabel, J.A., Diaz, O., Lekadir, K., Mazurowski, M.A.: Fréchet radiomic distance (frd): A versatile metric for comparing medical imaging datasets. *Medical Image Analysis* **110**, 103943 (2026). <https://doi.org/10.1016/j.media.2026.103943>, <https://www.sciencedirect.com/science/article/pii/S1361841526000125>
12. Li, X., Yang, K., Li, Q., Wang, Z.: Bidirectional mammogram view translation with column-aware and implicit 3d conditional diffusion. *arXiv preprint arXiv:2510.04947* (2025)
13. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747* (2022)
14. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003* (2022)
15. Nguyen, H.T., Nguyen, H.Q., Pham, H.H., Lam, K., Le, L.T., Dao, M., Vu, V.: Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Scientific Data* **10**(1), 277 (2023)
16. Petrini, D.G., Shimizu, C., Roela, R.A., Valente, G.V., Folgueira, M.A.A.K., Kim, H.Y.: Breast cancer diagnosis in two-view mammography using end-to-end trained efficientnet-based convolutional network. *Ieee access* **10**, 77723–77731 (2022)
17. Seitzer, M.: pytorch-fid: FID Score for PyTorch. <https://github.com/mseitzer/pytorch-fid> (August 2020), version 0.3.0
18. Sorkhei, M., Liu, Y., Azizpour, H., Azavedo, E., Dembrower, K., Ntola, D., Zouzos, A., Strand, F., Smith, K.: Csaw-m: An ordinal classification dataset for benchmarking mammographic masking of cancer. *arXiv preprint arXiv:2112.01330* (2021)
19. Strand, F.: CSAW-CC (mammography) – a dataset for AI research to improve screening, diagnostics and prognostics of breast cancer (2022). <https://doi.org/10.5878/45vm-t798>, <https://doi.org/10.5878/45vm-t798>
20. 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)
21. Sweeney, R.J.I., Lewis, S.J., Hogg, P., McEntee, M.F.: A review of mammographic positioning image quality criteria for the craniocaudal projection. *The British journal of radiology* **91**(1082), 20170611 (2018)
22. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International conference on machine learning*. pp. 6105–6114. PMLR (2019)
23. Walsh, R., Tardy, M.: A comparison of techniques for class imbalance in deep learning classification of breast cancer. *Diagnostics* **13**(1), 67 (2022)
24. Wang, Z., Xian, J., Liu, K., Li, X., Li, Q., Yang, X.: Dual-view correlation hybrid attention network for robust holistic mammogram classification. *arXiv preprint arXiv:2306.10676* (2023)
25. Yamazaki, A., Ishida, T.: Two-view mammogram synthesis from single-view data using generative adversarial networks. *Applied Sciences* **12**(23), 12206 (2022)