# Rethinking Multi-view Mammogram Representation Learning via Counterfactual Reasoning with Kolmogorov-Arnold Theorem

Guoli Wang[1,2], Benzheng Wei[1(✉)], and Shuo Li[2]

[1] Shandong University of Traditional Chinese Medicine, Shandong 250301, China
wbz99@sina.com
[2] Case Western Reserve University, Cleveland, OH 44106, USA
shuo.li11@case.edu

**Abstract.** Cross-view interference caused by impure shared information for multiview mammogram representation. Existing methods are accustomed to assuming that purely complementary shared information are provided between multiple views, ignoring the negative side of the shared information. To address this issue, we propose the first Dual-view Mammography Causal Graph (DMCG) to model multi-view representation by capturing direct and mediation effects. Based on DMCG, we propose MammoCRKAN, the first counterfactual reasoning paradigm integrating the Kolmogorov-Arnold theorem for decoupling interfering information. MammoCRKAN comprises two key modules: the Spherical Sample Module (SSM), which enhances the direct effect of tumor features by aligning consistent geometric representations, and the Kolmogorov–Arnold Aggregate Module (KAAM), which decomposes complex joint causality into univariate effects to mitigate negative side of mediation effects. Moreover, We find that heterogeneous channel allocations across views outperform fixed matching channels. Extensive experiments on four publicly available mammogram datasets demonstrate the effectiveness of MammoCRKAN. Code is available at https://guoli-w.github.io/MammoCRKAN.

**Keywords:** Multi-View Representation Learning · Counterfactual Reasoning · Kolmogorov-Arnold Theorem · Mammographic Image Analysis.

## 1 Introduction

Existing multi-view mammographic representation learning methods suffer from cross-view interference. These methods [23,2,18,15,6,7,5,19] accustomed to assume that the cephalocaudal (CC) and mediolateral oblique (MLO) views contribute purely complementary shared information, thereby overlooking potential redundancy and confusion between shared features. As shown in Fig. 1(a), axillary features in the MLO view may introduce redundant interference into the CC view, while the uniform tumor density in the CC view can reduce the model's sensitivity to abnormalities in the MLO view, resulting in similarity

interference. Consequently, eliminating cross-view interference while preserving beneficial shared information remains a critical challenge.

Impure shared information disrupts multi-view representations. Existing methods focus on fusion strategies at various stages (early [23,6], mid [15,18,19], late [15,7]) to maximize shared information, yet they overlook feature confusion. Moreover, some approaches [23,15,7] increase the number of learnable views to capture more shared information, neglecting feature mismatch. Despite these efforts, directly decoupling interfering information from shared information remains challenging (Fig. 1(b)).

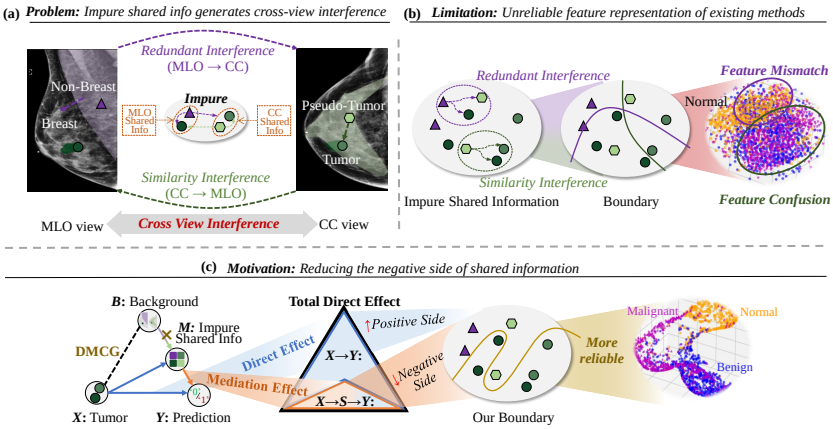


Fig. 1: (a) Impure shared info between MLO and CC views can cause cross-view interference. (b) Impure shared info leads to unreliable representations in existing multi-view methods. (c) DMCG models multi-view representations under causal perspective, demonstrating that enhancing positive direct effects and mitigating negative mediation effects improves representation performance.

Counterfactual reasoning can decouple interfering information from impure shared information by analyzing causal relationships [4,24,22]. Both counterfactual reasoning [26,1] and intervention [11,13] probe variable correlations via causal inference, with intervention controlling confounders and counterfactual reasoning optimizing the Total Direct Effect (TDE) to reveal contradictory effects [20]. However, modeling high-dimensional, nonlinear causal relationships in multi-view mammogram representations is challenging. The Kolmogorov–Arnold (K-A) theorem [21] mitigates this by decomposing complex mappings into resolvable univariate functions, thereby reducing modeling difficulty and enhancing the feasibility of counterfactual reasoning.

We first propose a Dual-view Mammogram Causal Graph (DMCG) to model the TDE in multi-view representation learning. As shown in Fig. 1(c), TDE comprises both Direct Effect (DE) and Mediation Effect (ME). The tumor in dual-view $X$ serves as the independent variable, influencing dependent variable

$Y$ via the direct causal path $X \rightarrow Y$. Meanwhile, impure shared information acts as mediator $\mathcal{M}$, contributing to the ME. Since $\mathcal{M}$ consists of background $B$ and $X$, interference is naturally present. To mitigate the negative side of the ME, we first propose MammoCRKAN, a novel counterfactual reasoning paradigm for multi-view mammogram representation that incorporates the K-A theorem. MammoCRKAN consists of two key modules: Spherical Sample Module (SSM) and Kolmogorov-Arnold Aggregate Module (KAAM). SSM enhances the DE of the tumor itself, while KAAM reduces the negative side of the ME caused by impure shared information. **Our main contributions are:**

1. DMCG is the first causal graph to model cross-view interference in multi-view mammogram representation.
2. MammoCRKAN is the first paradigm to mitigate the negative side of view shared information via counterfactual reasoning with K-A theorem in multi-view mammogram representation learning.
3. MammoCRKAN is also the first to apply the K-A theorem to decompose complex joint causality into univariate functions, simplifying causal modeling and enhancing counterfactual reasoning in medical image analysis.

## 2 Method

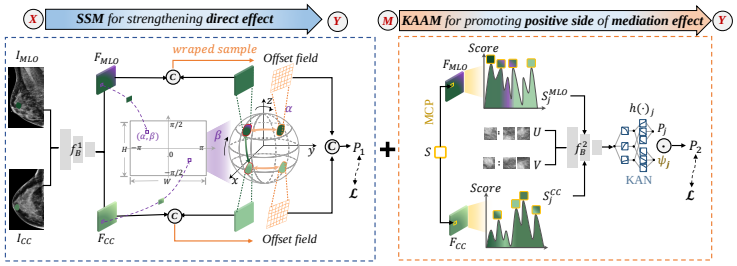


Fig. 2: MammoCRKAN consists of SSM and KAAM. SSM strengthening direct effect. KAAM promoting positive side of mediation effect.

### 2.1 Multi-view mammogram representation problem formulation

Let $I_i^{\text{MLO}}, I_i^{\text{CC}} \in \mathcal{I}_i$ be the $i$-th breast (left or right) with MLO and CC views. $Y_i \in \mathcal{Y} = \{0, 1, 3\}$ denotes the label ($0$ : Normal, $1$ : Benigns, $2$ : Malignant), and $Y_i^{\text{MLO}} = Y_i^{\text{CC}}$. Our goal is to train a Deep Neural Network (DNN) $f(\cdot)$ based on the $\mathcal{D}$ that can output accurate predictions for unseen examples.

To mitigate the negative side of mediator $\mathcal{M}$ (impure shared information), a commonly used strategy is TDE, which aims to strengthen the DE along independent variable $X \rightarrow$ dependent variable $Y$. In the context of multi-view

learning, for $I_{\mathrm{MLO}}$ and $I_{\mathrm{CC}}$ belonging to the ipsilateral breast, DMCG defines $\mathrm{TDE}(I_{\mathrm{MLO}}, I_{\mathrm{CC}})$ as the probability $P$ that the target tumor $x \in \{I_i^{\mathrm{MLO}}, I_i^{\mathrm{CC}}\}$ belongs to malignant, as shown in Eq. (1):

$$\mathrm{TDE}(\mathcal{I}) = \underbrace{P(Y_{\mathcal{M}} = 1 \mid X = x, B = b)}_{\text{Term 1: } P_1 \sim (x,b)} - \underbrace{P(Y_{\mathcal{M}} = 1 \mid X = x_0, B = b)}_{\text{Term 2: } P_2 \sim (b)}, \quad (1)$$

where $X = x_0$ represents the operation of masking the features of the target tumor $x$. The subscript $\mathcal{M}$ denotes the influence of the impure shared information $\mathcal{M}$ from two views on the prediction. The goal of TDE is to preserve the positive side of mediation effect through the mediator pathway $[X \to \mathcal{M} \to Y]$ while mitigating the negative side introduced by the mediator pathway. Specifically, Term 1 in Eq. (1) represents the prediction based on $\mathcal{M}$ formed by both $x$ and $b$, while Term 2 represents the prediction from $\mathcal{M}$ constructed solely from $b$ when $x$ is masked. Term 2 quantifies the interference from $b$, as the model predicts a "tumor" even without $x$, capturing the scenario where high breast density leads to similarity interference due to low distinguishability between tumor and background. The subtraction operator removes this negative mediation effect.

During the inference process, the first probability $P_1$ in TDE can be directly obtained by training a DNN with $I_{\mathrm{MLO}}$ and $I_{\mathrm{CC}}$. Unfortunately, it is challenging to obtain the second probability $P_2$ by masking the features of $x$, as the location of the target tumor is unknown. Considering that subtraction operators often introduce negative values, we transform the $P_2$ into its complementary form, which represents the probability for the target tumor $x$ when masking its background $b$, as shown as Eq. (2):

$$\mathrm{TDE}(\mathcal{I}) = \underbrace{P(Y_{\mathcal{M}} = 1 \mid X = x, B = b)}_{\text{Term 1: } P_1 \sim (x,b)} + \underbrace{\psi P(Y_{\mathcal{M}} = 1 \mid X = x, B = b_0)}_{\text{Term 2: } P_2 \sim (x)}, \quad (2)$$

where $B = b_0$ represents the operation of masking the features from background $b$ in the impure shared information $\mathcal{M}$, and $\psi$ is a parameter related to the data.

This transformation shifts the negative mediation effect caused by $\mathcal{M}$ (solely $b$) into an enhancement of the direct effect from $x$, preventing cross-view interference from impure shared information. As Eq. (2) shows, strengthening Term 1 (DE) and promoting Term 2 (positive ME) enhances the TDE. Accordingly, MammoCRKAN employs SSM for Term 1 ($X \to Y$) and KAAM for Term 2 ($X \to \mathcal{M} \to Y$).

### 2.2 SSM for strengthening direct effect from $X$ in $[X \to Y]$

To obtain a more stable direct effect (Term 1 of Eq. (2)), SSM is proposed to learn consistent geometric features of the target tumor across views. Specifically, by feeding both the original $I_{\mathrm{MLO}}$ and $I_{\mathrm{CC}}$ into the first backbone $f_b^1(\cdot)$ simultaneously, coarse-grained feature maps $F_{\mathrm{MLO}} \in \mathbb{R}^{H \times W \times C}$, and $F_{\mathrm{CC}} \in \mathbb{R}^{H \times W \times C}$ can be obtained. Subsequently, SSM resamples these inputs to introduce spherical geometric properties and learn consistent features in a spherical coordinate system.

First, each pixel in each view is transformed into the azimuth angle $\alpha$ and the polar angle $\beta$ on the sphere according to: $\alpha = \frac{x-(W-1)/2}{W} \cdot 2\pi, \quad \beta = -\frac{y-(H-1)/2}{H} \cdot \pi$, where $\alpha \in [-\pi, \pi]$, $\beta \in [-\pi/2, \pi/2]$, $x$ and $y$ are the pixel indices of the $F_{\mathrm{MLO/CC}}$, and the $(\alpha, \beta) \in \mathbb{R}^{H \times W \times 2}$ is the spherical coordinate grid.

Directly using the grid $(\alpha, \beta)$ may yield insufficient geometric information. SSM designs $G(\alpha, \beta) = (\sin\alpha, \cos\alpha, \sin\beta, \cos\beta, \cos\alpha\cos\beta)$. to learn 5D spherical geometric properties. These spherical geometric properties are embedded into $F_{\mathrm{MLO/CC}}$, enhancing the consistency of cross-view target tumor geometric features. To ensure that sampling points more accurately align geometric properties across different views, SSM generates an offset field $(\Delta\alpha, \Delta\beta)$ through two convolutional layers: $(\Delta\alpha, \Delta\beta) = f_{\mathrm{Convs}}(F_{\mathrm{MLO/CC}} \cdot G(\alpha, \beta))$. Then, SSM using `Gridsample`, to interpolate the $F_{\mathrm{MLO/CC}}$ and resample based on the offset field, aligning tumor consistency geometric features across views, as shown as Eq. (3):

$$F^{o}_{\mathrm{MLO/CC}} = F_{\mathrm{MLO/CC}}(\alpha_0 + \Delta\alpha, \beta_0 + \Delta\beta), \tag{3}$$

where $(\alpha_0, \beta_0)$ are the standard grid points in the initial spherical coordinate system. $F^{o}_{\mathrm{MLO}}, F^{o}_{\mathrm{CC}}$ are the two-view feature map that had aligned consistent geometry features. Finally, by $f_{\mathrm{cls}}(x) = (w^{T}_{\mathrm{cls}} \cdot x)$ classifying direct prediction logits $P_1 = f_{\mathrm{cls}}(\mathrm{Concat}(F^{o}_{\mathrm{MLO}}, F^{o}_{\mathrm{CC}})$.

**Summary of Advantage:** SSM aligns consistent geometric features of tumor across cross-view samples, reinforcing the direct effect.

### 2.3 KAAM for promoting positive side of mediation effect from $\mathcal{M}$ in $[X \rightarrow \mathcal{M} \rightarrow Y]$

To obtain the predicted probabilities $P_2$ (Term 2 in Eq. (2)), KAAM leverages the K-A theorem [21] and the Maximum Coverage Problem (MCP) [8]. It decomposes the total mediated effect into univariate effects centered on the target tumor $x$. MCP selects $M$ patches from two views to maximize $x$'s coverage, and KAAM aggregates their predictions to approximate $P_2$.

Specifically, $F_{\mathrm{MLO}}$ and $F_{\mathrm{CC}}$ are normalized by category and added to $F^{*}_{\mathrm{MLO}}$ and $F^{*}_{\mathrm{CC}}$. MCP selects $M$ candidate region coordinates $\{S^{\mathrm{MLO}}_{j}\}^{M}_{j=1}$ and $\{S^{\mathrm{CC}}_{j}\}^{M}_{j=1}$ with the highest weights in each view by iteratively searching on $F_{\mathrm{MLO}}$ and $F_{\mathrm{CC}}$, using predefined rectangular bounding boxes ($512 \times 512$ in this experiment) and a designed content scoring Function $f_{\mathrm{score}}(S, F^{*}_{\mathrm{MLO/CC}}) = \sum_{(i,j)\in S} F^{*}_{\mathrm{MLO/CC}}[i, j]$.

$$S^{\mathrm{MLO/CC}}_{t+1} = \mathrm{argmax}_{S^t} f_{\mathrm{score}}(S^t, F^{*}_{\mathrm{MLO/CC}}), t \in [0, M-1], \tag{4}$$

where $t$ represents the number of times each candidate coordinate is generated. During each selection, the weight reduction for the selected regions is controlled by introducing the cover rate parameter $r$. Candidate regions in both views vary according to: $F^{*}_{\mathrm{t+1}}[i, j] = F^{*}_{t}[i, j] \cdot \big(1 - r \cdot S_t[i, j]\big), t \in [0, M-1]$. The $r$ allows some weight overlap between different regions, enabling information sharing between regions while maintaining contextual information. Based on these coordinates, MCP selects the corresponding patches $\{U_j\}^{M}_{j=1} \in I_{\mathrm{MLO}}$, $\{V_j\}^{M}_{j=1} \in I_{\mathrm{CC}}$.

Since patches are retrieved from coarse-grained feature maps, the target tumor may appear in several patches with significant variation. KAAM decomposes the weight $\psi$ from Eq. (2) for each patch based on its prediction. Specifically, feeding $\text{add}(U_j, V_j)$ into the second backbone $f_B^2(\cdot)$ yields univariate representations $\{F_j\}_{j=1}^M$, which are inherently discrete. KAAM then applies a B-spline function $h(\cdot)$ [14] to map each $F_j$ into a continuous smooth space $h_j(F_j)$ and compute its logits $\{p_j\}_{j=1}^M$ (Eq. (5)).

$$p_j = \sigma\left(\sum_{l=1}^{M} h_l(F_l)\right) \cdot h_j(F_j), \tag{5}$$

where $\sigma(\cdot)$ represents the Sigmoid function. Based on $p_j$, weight $\psi$ can be defined as: $\psi_j = \frac{\exp(P_j/\tau)}{\sum_{l=1}^{M}\exp(P_l/\tau)}, \forall j \in [M]$. The $\tau$ represents a temperature parameter that controls the degree of emphasis on a particular patch. Then, the predicted logits $P_2$ of Eq. (2) for different univariate effect representation can be obtained by aggregating the logits $p_j$ via $P_2 = \sum_{j=1}^{m} \psi_j \cdot p_j$. Given the predictions $P_1$ and $P_2$, the loss function $\mathcal{L}$ of MammoCRKAN is defined as $\mathcal{L} = \ell(P_1, Y) + \ell(P_2, Y)$. $\ell(\cdot,\cdot)$ is Focal loss [10].

**Summary of Advantage:** KAAM decomposes the mediation effect into a set of univariate effects, thereby reducing negative side of mediation effect while enhancing positive side.

## 3 Experimental Results

### 3.1 Experimental setup

**Datasets.** We use 4 public datasets, considering only ipsilateral MLO-CC image pairs. **(a)** *INBreast* [16] contains 200 pairs from 115 patients (normal, benign, and malignant), with 280 pairs for training and 116 for testing. **(b)** *VinDr-Mammo* [17] includes 10,243 pairs from Vietnam, pre-split at a 4:1 patient-level ratio (8,195 train, 2,046 test). **(c)** *CBIS-DDSM* [9] comprises 1,324 pairs (1,054 train, 270 test). **(d)** *CMMD* [3] features 5,202 images from 1,775 patients (benign or malignant), with 1,820 pairs for training, 260 for validation, and 520 for testing. **Implementation Details.** MammoCRKAN is implemented in PyTorch and runs on an A100 GPU (40GB). We use a batch size of 8, an initial learning rate of $1 \times 10^{-4}$ with AdamW, weight decay of $1 \times 10^{-4}$, and a cosine annealing scheduler (min LR: $1 \times 10^{-8}$). The model is trained for 80 epochs with early stopping. Images are resized to $2944 \times 1920$ and horizontally flipped to maintain right-oriented breasts, following [2]. Performance is measured via AUC, accuracy, precision, and F1 score.

### 3.2 Results and analysis

**Comparison with State-of-the-Arts (Table 1).** Our model is compared on 4 datasets with SOTA methods. Using only two unilateral views, our method

Table 1: MammoCRKAN achieves the best classification results on four mainstream datasets. △ shows how much MammoCRKAN exceeds SOTA. Growth values are in red. Second-best results are in blue. U: Unilateral; B: Bilateral.

| Method | Venue | Views | INBreast | | | | VinDr-Mammo | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | ACC (%) ↑ | Prec. (%) ↑ | AUC (%) ↑ | F1 (%) ↑ | ACC (%) ↑ | Prec. (%) ↑ | AUC (%) ↑ | F1 (%) ↑ |
| DMV-CNN [23] | TMI'19 | B4 | 66.51$_{\pm 2.28}$ | 54.94$_{\pm 1.64}$ | 72.19$_{\pm 2.38}$ | 53.54$_{\pm 1.99}$ | 77.70$_{\pm 0.21}$ | 64.64$_{\pm 0.26}$ | 81.48$_{\pm 0.77}$ | 61.87$_{\pm 0.53}$ |
| GMIC [19] | MedIA'21 | U1 | 63.36$_{\pm 0.61}$ | 59.17$_{\pm 3.51}$ | 76.60$_{\pm 0.37}$ | 44.03$_{\pm 2.14}$ | 64.98$_{\pm 0.52}$ | 47.28$_{\pm 0.18}$ | 64.98$_{\pm 0.52}$ | 26.22$_{\pm 0.21}$ |
| GLAM [12] | MIDL'21 | U1 | 55.93$_{\pm 2.44}$ | 46.61$_{\pm 2.14}$ | 66.96$_{\pm 1.48}$ | 45.61$_{\pm 0.85}$ | 64.99$_{\pm 0.50}$ | 46.96$_{\pm 1.20}$ | 71.37$_{\pm 0.69}$ | 25.24$_{\pm 1.06}$ |
| MVCCL [2] | MICCAI'22 | U2 | 56.03$_{\pm 0.91}$ | 54.97$_{\pm 1.72}$ | 78.18$_{\pm 1.56}$ | 57.30$_{\pm 0.88}$ | 80.36$_{\pm 0.49}$ | 71.68$_{\pm 0.13}$ | 82.39$_{\pm 0.62}$ | 73.84$_{\pm 1.25}$ |
| PHYSEnet [15] | Cancer'22 | U2 | 64.08$_{\pm 2.63}$ | 59.08$_{\pm 2.40}$ | 71.22$_{\pm 2.29}$ | 59.16$_{\pm 1.97}$ | 80.07$_{\pm 1.14}$ | 71.33$_{\pm 2.97}$ | 84.65$_{\pm 2.17}$ | 72.86$_{\pm 1.71}$ |
| LoraViT [27] | ISBI'23 | U1 | 65.81$_{\pm 0.50}$ | 60.63$_{\pm 0.17}$ | 75.88$_{\pm 0.10}$ | 57.45$_{\pm 0.36}$ | 71.27$_{\pm 0.25}$ | 65.81$_{\pm 0.29}$ | 77.26$_{\pm 0.02}$ | 61.72$_{\pm 0.60}$ |
| MV-Swin-T [18] | ISBI'24 | U2 | 58.05$_{\pm 1.07}$ | 56.40$_{\pm 0.70}$ | 70.28$_{\pm 4.76}$ | 55.24$_{\pm 1.04}$ | 58.66$_{\pm 0.90}$ | 41.73$_{\pm 0.16}$ | 60.23$_{\pm 0.31}$ | 47.12$_{\pm 0.70}$ |
| Bi-Mamba [25] | MICCAI'24 | U2 | 64.47$_{\pm 0.93}$ | 61.30$_{\pm 0.62}$ | 77.23$_{\pm 0.89}$ | 60.45$_{\pm 1.30}$ | 78.59$_{\pm 0.67}$ | 70.62$_{\pm 1.30}$ | 77.54$_{\pm 0.70}$ | 71.32$_{\pm 1.21}$ |
| **MammoCRKAN** | - | U2 | **69.01**$_{\pm 1.13}$ | **63.63**$_{\pm 0.68}$ | **81.10**$_{\pm 0.15}$ | **64.01**$_{\pm 1.28}$ | **81.11**$_{\pm 0.58}$ | **73.93**$_{\pm 0.11}$ | **84.71**$_{\pm 1.31}$ | **75.75**$_{\pm 0.60}$ |
| △ SOTA | - | - | **+ 2.50** | **+ 2.33** | **+ 2.92** | **+ 3.56** | **+ 0.75** | **+ 2.25** | **+ 0.06** | **+ 1.91** |

| Method | Venue | Views | DDSM-CBIS | | | | CMMD | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | ACC (%) ↑ | Prec. (%) ↑ | AUC (%) ↑ | F1 (%) ↑ | ACC (%) ↑ | Prec. (%) ↑ | AUC (%) ↑ | F1 (%) ↑ |
| DMV-CNN [23] | TMI'19 | B4 | 86.59$_{\pm 1.69}$ | 83.82$_{\pm 1.96}$ | 85.40$_{\pm 0.85}$ | 64.52$_{\pm 3.62}$ | 74.92$_{\pm 1.70}$ | 80.40$_{\pm 0.75}$ | 74.37$_{\pm 0.82}$ | 82.99$_{\pm 1.45}$ |
| GMIC [19] | MedIA'21 | U1 | 88.28$_{\pm 2.01}$ | 87.86$_{\pm 2.50}$ | 87.27$_{\pm 0.44}$ | 72.50$_{\pm 2.07}$ | 78.50$_{\pm 0.92}$ | 82.68$_{\pm 0.79}$ | 70.13$_{\pm 1.55}$ | 86.55$_{\pm 0.81}$ |
| GLAM [12] | MIDL'21 | U1 | 79.40$_{\pm 0.91}$ | 81.38$_{\pm 0.86}$ | 85.70$_{\pm 0.98}$ | 69.53$_{\pm 0.95}$ | 75.64$_{\pm 1.71}$ | 84.47$_{\pm 0.84}$ | 64.70$_{\pm 1.89}$ | 85.73$_{\pm 1.41}$ |
| MVCCL [2] | MICCAI'22 | U2 | 79.32$_{\pm 1.47}$ | 86.82$_{\pm 1.07}$ | 86.04$_{\pm 0.59}$ | 73.80$_{\pm 0.53}$ | 74.91$_{\pm 2.48}$ | 83.52$_{\pm 0.55}$ | 72.79$_{\pm 1.40}$ | 84.69$_{\pm 2.31}$ |
| PHYSEnet [15] | Cancer'22 | U2 | 82.92$_{\pm 0.46}$ | 91.75$_{\pm 0.47}$ | 92.84$_{\pm 0.42}$ | 82.46$_{\pm 0.41}$ | 74.75$_{\pm 0.91}$ | 81.26$_{\pm 0.95}$ | 70.24$_{\pm 2.38}$ | 84.32$_{\pm 0.95}$ |
| LoraViT [27] | ISBI'23 | U1 | 75.71$_{\pm 0.71}$ | 85.27$_{\pm 1.29}$ | 90.47$_{\pm 0.74}$ | 76.00$_{\pm 0.59}$ | 76.01$_{\pm 2.59}$ | 82.75$_{\pm 1.49}$ | 69.39$_{\pm 0.81}$ | 81.63$_{\pm 1.17}$ |
| MV-Swin-T [18] | ISBI'24 | U2 | 76.60$_{\pm 1.98}$ | 88.95$_{\pm 0.70}$ | 80.28$_{\pm 0.64}$ | 73.92$_{\pm 0.17}$ | 77.10$_{\pm 1.10}$ | 80.49$_{\pm 2.49}$ | 66.12$_{\pm 0.65}$ | 86.87$_{\pm 0.31}$ |
| Bi-Mamba [25] | MICCAI'24 | U2 | 79.91$_{\pm 0.74}$ | 88.00$_{\pm 0.63}$ | 91.07$_{\pm 0.55}$ | 77.98$_{\pm 0.61}$ | 77.99$_{\pm 0.34}$ | 84.42$_{\pm 1.08}$ | 71.02$_{\pm 0.76}$ | 84.05$_{\pm 0.84}$ |
| **MammoCRKAN** | - | U2 | **91.70**$_{\pm 0.27}$ | **92.50**$_{\pm 0.11}$ | **94.97**$_{\pm 1.08}$ | **84.43**$_{\pm 0.63}$ | **80.13**$_{\pm 0.70}$ | **85.85**$_{\pm 0.55}$ | **78.30**$_{\pm 1.03}$ | **87.94**$_{\pm 0.14}$ |
| △ SOTA | - | - | **+ 8.78** | **+ 0.75** | **+ 2.13** | **+ 1.97** | **+ 1.63** | **+ 1.38** | **+ 5.51** | **+ 1.07** |

outperforms four-view method, and significantly improves all metrics over other two-view methods, highlighting the importance of eliminating crosstalk interference. It also excels in differentiating challenging intra-class benign and malignant tumors compared to single-view models.

Table 2: Ablation studies on the CBIS-DDSM and CMMD testing demonstrate the significant improvements of the proposed innovations.

| CBIS-DDSM | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method | Baseline | SSM | MCP | KAN | ACC (%) ↑ | Prec. (%) ↑ | AUC (%) ↑ | F1 (%) ↑ |
| I | ✓ | | | | 85.69 | 83.11 | 84.37 | 76.79 |
| II | ✓ | ✓ | | | 87.36 **(+1.67)** | 86.58 **(+3.47)** | 87.55 **(+3.18)** | 79.77 **(+2.98)** |
| III | ✓ | ✓ | ✓ | | 89.94 **(+2.58)** | 90.48 **(+3.90)** | 89.12 **(+1.57)** | 82.55 **(+2.78)** |
| IV | ✓ | ✓ | ✓ | ✓ | **91.70 (+1.76)** | **92.50 (+2.02)** | **94.97 (+5.85)** | **84.43 (+1.88)** |
| CMMD | | | | | | | | |
| Method | Baseline | SSM | MCP | KAN | ACC (%) ↑ | Prec. (%) ↑ | AUC (%) ↑ | F1 (%) ↑ |
| I | ✓ | | | | 74.35 | 78.70 | 70.92 | 80.68 |
| II | ✓ | ✓ | | | 77.41 **(+3.06)** | 80.32 **(+1.62)** | 73.53 **(+2.61)** | 83.78 **(+3.10)** |
| III | ✓ | ✓ | ✓ | | 79.65 **(+2.24)** | 83.13 **(+2.81)** | 75.80 **(+2.27)** | 85.70 **(+1.92)** |
| IV | ✓ | ✓ | ✓ | ✓ | **80.13 (+0.48)** | **85.85 (+2.72)** | **78.30 (+2.50)** | **87.94 (+2.24)** |

**Component ablation studies (Table 2). (I):** Use a dual-view ResNet-50 baseline where both views share a feature extractor and fuse features before classification. **(II):** Feed extracted features into SSM to enhance the direct effect along $X \to Y$. **(III):** Add $P_1$ directly to $p_j$ in each block without decomposing $\psi$ (Eq. 2). **(IV):** Apply KAAM (MCP + KAN) to boost the positive mediation effect and reduce its negative side along $X \to \mathcal{M} \to Y$.

**Parameter ablation studies (Fig. 3, Table 3).** The mediator $\mathcal{M}$ delivers optimal shared information (comprising $B$ and $X$) at a cover rate of $r = 0.3$, maximizing the positive mediation effect. When $r = 0$, negative effects increase, and at $r = 1.0$, incomplete contextual information undermines the positive effect. Both Fig. 3 and Table 3 confirm that impure shared information induces cross-view interference and weakens $X$'s positive contribution in the mediation effect.
**Visualization (Fig. 4).** Heterogeneous cross-view channel configurations outperform forced alignment in the spherical geometric space. As shown in Fig. 4(a), the optimal performance is achieved with channel combinations of {MLO: 256, CC: 128}. Increasing feature channels only adds redundancy and interference. Moreover, Fig. 4(b) demonstrates that MammoCRKAN yields more reliable representations both within and between tumor classes.

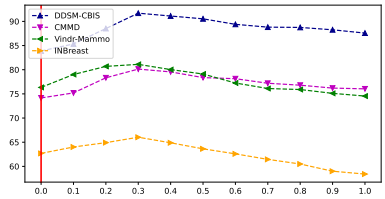


Fig. 3: Comparison of different cover rate $r$ of KAAM. Best results at 0.3.

Table 3: Effects of different region size $S$ and number $M$ of KAAM.

| $S$ | AUC (%) | $M$ | AUC (%) |
|---|---|---|---|
| 64×64 | 89.31 | 1 | 89.30 |
| 128×128 | 90.11 | 3 | 90.30 |
| 256×256 | 92.95 | **6** | **94.97** |
| **512×512** | **94.97** | 9 | 89.98 |
| 1024×1024 | 89.76 | 12 | 89.23 |

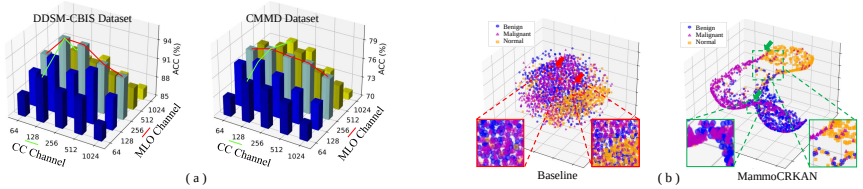


Fig. 4: (a) Heterogeneous channel allocations for MLO and CC views outperform fixed matching channels. (b)MammoCRKAN demonstrates superior representation both within and across classes.

## 4 Conclusion

In conclusion, our work tackles cross-view interference in multi-view mammogram representation by introducing the first DMCG. Building on DMCG, we propose MammoCRKAN—the first counterfactual reasoning paradigm integrating the Kolmogorov–Arnold theorem to decouple interfering information. By

enhancing direct tumor feature effects via the SSM and mitigating negative side of mediation effect via the KAAM, MammoCRKAN significantly improves representation quality. Extensive experiments validate its effectiveness.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China (No.62372280, No.62402297), the Demonstration Projects of Science and Technology for the People of Qingdao City (No.23-2-8-smjk-2-nsh), the Natural Science Foundation of Shandong Province (2024MF139, 2023QF094, ZR2022QG051), the Special fund of Qilu Health and Health Leading Talents Training Project.

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

## References

1. Cai, L., Fang, H., Xu, N., Ren, B.: Counterfactual causal-effect intervention for interpretable medical visual question answering. IEEE Transactions on Medical Imaging (2024)
2. Chen, Y., Wang, H., Wang, C., Tian, Y., Liu, F., Liu, Y., Elliott, M.S., McCarthy, D.J., Frazer, H.M., Carneiro, G.: Multi-view local co-occurrence and global consistency learning improve mammogram classification generalisation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2022)
3. Cui, C., Li, L., Cai, H., Fan, Z., Zhang, L., Dan, T., Li, J., Wang, J.: The chinese mammography database (cmmd): An online mammography database with biopsy confirmed types for machine diagnosis of breast. The Cancer Imaging Archive **1** (2021)
4. Hoch, S.J.: Counterfactual reasoning and accuracy in predicting personal events. Journal of Experimental Psychology: Learning, Memory, and Cognition **11**(4), 719 (1985)
5. Jain, K., Bansal, A., Rangarajan, K., Arora, C.: Mmbcd: Multimodal breast cancer detection from mammograms with clinical history. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 144–154. Springer (2024)
6. Jain, K., Rangarajan, K., Arora, C.: Follow the radiologist: Clinically relevant multi-view cues for breast cancer detection from mammograms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 102–112. Springer (2024)
7. Karaman, B.K., Dodelzon, K., Akar, G.B., Sabuncu, M.R.: Longitudinal mammogram risk prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 437–446. Springer (2024)
8. Khuller, S., Moss, A., Naor, J.S.: The budgeted maximum coverage problem. Information processing letters **70**(1), 39–45 (1999)
9. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific data **4**(1), 1–9 (2017)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)

11. Liu, H., Li, Q., Nie, W., Xu, Z., Liu, A.: Causal intervention for brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 160–170. Springer (2024)
12. Liu, K., Shen, Y., Wu, N., Chledowski, J., Fernandez-Granda, C., Geras, K.J.: Weakly-supervised high-resolution segmentation of mammography images for breast cancer diagnosis. Proceedings of machine learning research **143**, 268–285 (2021)
13. Liu, R., Liu, H., Li, G., Hou, H., Yu, T., Yang, T.: Contextual debiasing for visual recognition with causal mechanisms. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12755–12765 (2022)
14. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024)
15. Lopez, E., Grassucci, E., Valleriani, M., Comminiello, D.: Multi-view hypercomplex learning for breast cancer screening. arXiv preprint arXiv:2204.05798 (2022)
16. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: Inbreast: toward a full-field digital mammographic database. Academic radiology **19**(2), 236–248 (2012)
17. 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)
18. Sarker, S., Sarker, P., Bebis, G., Tavakkoli, A.: Mv-swin-t: Mammogram classification with multi-view swin transformer. 2024 IEEE International Symposium on Biomedical Imaging (ISBI) pp. 1–5 (2024)
19. Shen, Y., Wu, N., Phang, J., Park, J., Liu, K., Tyagi, S., Heacock, L., Kim, S.G., Moy, L., Cho, K., et al.: An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. Medical image analysis **68**, 101908 (2021)
20. Tang, K., Niu, Y., Huang, J., Shi, J., Zhang, H.: Unbiased scene graph generation from biased training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3716–3725 (2020)
21. Tikhomirov, V.: On the representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition. In: Selected Works of AN Kolmogorov, pp. 383–387. Springer (1991)
22. Verma, S., Boonsanong, V., Hoang, M., Hines, K., Dickerson, J., Shah, C.: Counterfactual explanations and algorithmic recourses for machine learning: A review. ACM Computing Surveys **56**(12), 1–42 (2024)
23. Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzębski, S., Févry, T., Katsnelson, J., Kim, E., et al.: Deep neural networks improve radiologists' performance in breast cancer screening. IEEE transactions on medical imaging **39**(4), 1184–1194 (2019)
24. Xie, M.K., Xiao, J.H., Peng, P., Niu, G., Sugiyama, M., Huang, S.J.: Counterfactual reasoning for multi-label image classification via patching-based training. arXiv preprint arXiv:2404.06287 (2024)
25. Yang, Z., Zhang, J., Wang, G., Kalra, M.K., Yan, P.: Cardiovascular disease detection from multi-view chest x-rays with bi-mamba. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 134–144. Springer (2024)
26. Zhang, L., Zhai, X., Zhao, Z., Zong, Y., Wen, X., Zhao, B.: What if the tv was off? examining counterfactual reasoning abilities of multi-modal language models.

In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21853–21862 (2024)
27. Zhu, Y., Shen, Z., Zhao, Z., Wang, S., Wang, X., Zhao, X., Shen, D., Wang, Q.: Melo: Low-rank adaptation is better than fine-tuning for medical image diagnosis. 2024 IEEE International Symposium on Biomedical Imaging (ISBI) pp. 1–5 (2023)