GeoCAN: Nonlinear Causal-Geometric Learning for Echocardiography Quality Assessment

Yiran Li[1], Kai Zheng[1] (✉), and Shuo Li[2]

[1] Qingdao Institute of Software and College of Computer Science and Technology, China University of Petroleum, Qingdao, Shandong, China
kaizheng@upc.edu.cn

[2] the Department of Computer and Data Science and the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA

Abstract. Causal learning is crucial for echocardiography quality assessment (EQA), demonstrating strong potential for learning fundamental quality representations. However, confounded quality degradations in EQA lead to spurious correlations and feature entanglement, obscuring causal relationships in quality modeling. In this paper, we propose a Geometric Causal Aggregation Network (GeoCAN) that embeds nonlinear causal–geometric learning into EQA to explicitly model relational causality and improve robustness under complex degradations. Specifically, Causal-Geometric Learning (CGL) deconfounds quality factors from visual degradations, thereby enhancing the discriminability of causal representations and suppressing spurious correlations. Nonlinear Relational Aggregation (NRA) disentangles quality representations from relational structures, explicitly separating relational cues from local visual features and alleviating feature entanglement. Extensive experiments on two public datasets and a private clinical dataset demonstrate that GeoCAN consistently achieves state-of-the-art performance, validating nonlinear causal–geometric learning as a principled solution for robust and reliable EQA. The code is available at https://github.com/Yiran0325/Echo-GeoCAN.

Keywords: Echocardiography · Quality Assessment · Causal Learning · Geometric Learning.

1 Introduction

Causal learning is crucial for echocardiography quality assessment (EQA), enabling the learning of fundamental quality representations [1, 6, 13]. With the increasing adoption of deep learning in EQA, confounded quality degradation mechanisms and the high cost of dense quality annotation have become major bottlenecks [4, 15]. Causal learning models the dependencies between quality-determining factors and imaging conditions, enabling more reliable EQA. Existing preliminary studies [2, 19] adopt causal learning strategies largely inspired by natural image analysis, where paradigms are developed under relatively stable imaging assumptions. However, EQA exhibits complex causal interactions

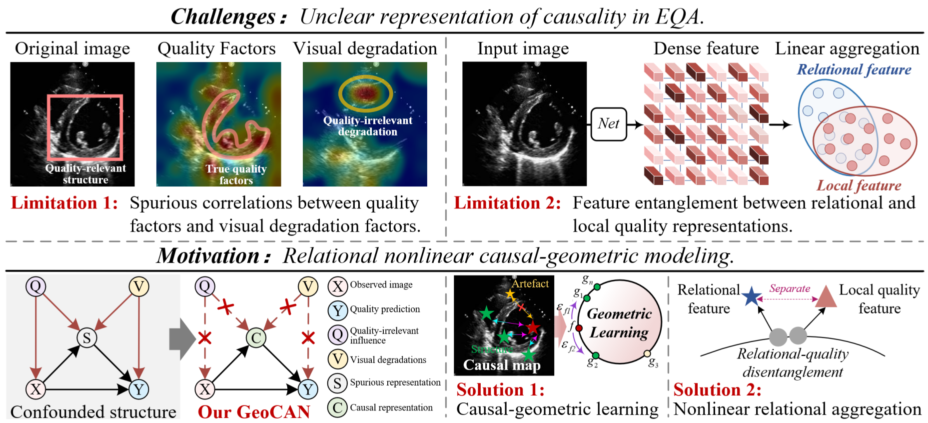

Fig. 1. Causal learning in echocardiography quality assessment (EQA) is hindered by spurious correlations and feature entanglement, motivating nonlinear causal–geometric learning to suppress quality-irrelevant confounders Q and degradation-related confounders V, clarifying the causal path $X \to Y$.

arising from external and experience-dependent conditions [10, 17]. This raises a fundamental question: How to effectively represent causal relationships in EQA quality modeling?

Two primary limitations hinder accurate representation of causal relationships in EQA (Fig. 1): 1) Spurious correlations between quality factors and visual degradation factors. Quality representations in EQA are easily confounded by visual degradations that are not causally related to image quality. Non-decisive visual variations are incorporated into quality representations, leading to spurious correlations. It obscures the truly quality-determining factors, causing EQA to rely on unstable visual cues and thereby reducing robustness. 2) Feature entanglement between relational representations and local quality representations. Image quality arises from spatial–semantic interactions among visual factors, yet aggregation of relational and local features compresses structural information and induces feature entanglement. Thus, models struggle to disentangle relational structure variations from local variations, limiting their ability to capture degradation-related spatial–semantic interactions.

To sum up the above limitations, quality representation requires explicitly modeling the relational structure between quality-visual interactions. Cardiac anatomy exhibits inherent topological consistency, providing a stable relational prior that remains preserved despite variations in imaging quality. Therefore, geometric learning offers a principled framework for encoding topological relationships, enabling dependencies among quality-visual factors to be explicitly modeled instead of implicitly entangled in Euclidean representations, which is crucial for representing causality in EQA.

In this paper, we propose a novel Geometric Causal Aggregation Network (GeoCAN) for echocardiography quality assessment (Fig. 2). It addresses the above limitations via two key innovations: 1) Causal-Geometric Learning (CGL) deconfounds quality factors from visual degradations. It represents visual factors as nodes in a structured relational space and models their causal interactions that separate quality-determining factors from co-occurring visual factors. Causal interactions exhibit an inherently asymmetric geometric structure, where dominant quality factors become compactly organized in a latent hyperbolic space while co-occurring visual cues are naturally separated, improving the discriminability and stability of causal representations. 2) Nonlinear Relational Aggregation (NRA) disentangles quality representation from relational structures. It integrates relational structures into quality representations through nonlinear mappings, enabling explicit disentanglement from local visual features and distinct modulation across relational patterns. This enables the model to distinguish the independent effects of relational and visual variations on quality assessment, improving robustness under complex visual degradations.

Our contributions are summarized as follows: 1) For the first time, we propose GeoCAN, which is the first nonlinear causal-geometric learning framework for echocardiography quality assessment. 2) Our CGL is the first causal–geometric framework that integrates latent hyperbolic learning into causal modeling to deconfound quality factors from visual degradations, enhancing discriminability. 3) Our NRA explicitly disentangles relational structures from quality representations via nonlinear aggregation, enabling differentiated relational modulation and improved quality representation under complex degradation conditions.

2 Methodology

2.1 Causal-Geometric Learning (CGL)

The CGL deconfounds quality factors from visual degradations by jointly modeling causal-geometric interaction, enabling robust quality representation. Given an input image x, a feature extractor produces a set of visual factors $F = \left\{F^{(i)}\right\}_{i=1}^{K} \in \mathbb{R}^{K\times h\times w}$, each factor map encodes spatial activation responses associated with specific visual patterns. CGL estimates directed conditional dependencies between factor pairs to construct a causal map. Let $\mathbf{X} \in \mathbb{R}^{K\times h\times w}$ denote the stacked factor tensor composed of $\{F^{(i)}\}_{i=1}^{K}$. For any pair $F^{(i)}$ and $F^{(j)}$, the causal map from $F^{(j)}$ to $F^{(i)}$ is defined as

$$C_{ij} \triangleq P\left(F^{(i)} \mid F^{(j)}\right) = \frac{\mathrm{GLM}_{\alpha,\beta}\left(F^{(i)} \times F^{(j)}\right)}{\mathrm{GLM}_{\alpha,\beta}\left(F^{(j)}\right) + \varepsilon}, \tag{1}$$

where $F^{(i)} \times F^{(j)}$ is element-wise spatial interaction between two feature factors and ε is a stabilization constant. The nonlinear aggregation operator $\mathrm{GLM}_{\alpha,\beta}(\cdot)$ is implemented using a generalized Lehmer mean [14]:

$$\mathrm{GLM}_{\alpha,\beta}\left(F_k\right) = \frac{1}{\ln\alpha} \ln\left(\frac{\sum_{s=1}^{S} \alpha^{(\beta+1)F_k(s)}}{\sum_{s=1}^{S} \alpha^{\beta F_k(s)}}\right) \tag{2}$$

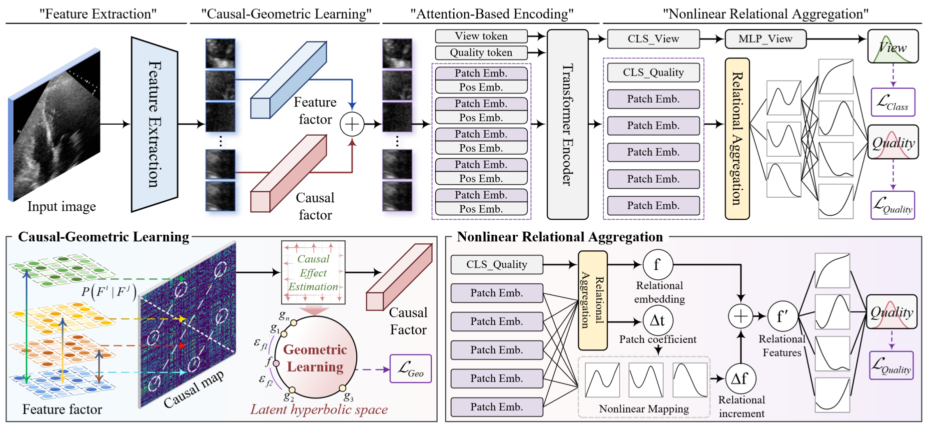

Fig. 2. Our GeoCAN models relational causality for echocardiography quality assessment. a) CGL deconfounds quality factors from visual degradations. b) NRA disentangles quality representation from relational structures.

where $F_k \in \mathbb{R}^{h\times w}$ is the spatial response map of the k-th visual factor, and S is the number of spatial elements. The learnable parameters α and β enable adaptive distribution sensitivity, allowing GLM to shift from min- to mean- to max-like pooling while remaining bounded and differentiable. By estimating conditional interactions for all factor pairs, CGL constructs a directed causal interaction matrix $\mathbf{C} = [C_{ij}]_{i,j=1}^{K} \in \mathbb{R}^{K\times K}$ to encode asymmetric dispositions:

$$\mathbf{g} = \mathrm{Norm}\left(\sigma\left(\frac{\mathbf{C}-\mathbf{C}^\top}{\tau}\right)\mathbf{1} - \sigma\left(\frac{\mathbf{C}-\mathbf{C}^\top}{\tau}\right)^\top \mathbf{1}\right), \tag{3}$$

where $\sigma(\cdot)$ is the sigmoid function, τ controls sensitivity, $\mathbf{1}$ is the all-ones vector, and $\mathrm{Norm}(\cdot)$ is per-sample standardization. The causal effect features are obtained via $\mathbf{X}_{\mathrm{cgl}} = \mathbf{X} \odot (1 + r\tanh(\mathbf{g}))$, where r controls the modulation strength and $\odot$ is element-wise multiplication.

To preserve the causal topology of $\mathbf{C}$, factor nodes are embedded into a hyperbolic manifold. Let z_i denote the embedding of the i-th factor derived from $\mathbf{X}_{\mathrm{cgl}}$, which is projected to hyperbolic space via the exponential map $h_i = \exp_0^c(z_i)$, where $\exp_0^c(\cdot)$ is the exponential map at the origin under curvature $c > 0$. Multi-hop structural interactions implied by $\mathbf{C}$ are approximated through a spectral expansion over singular values:

$$\mathrm{hop}(i,j) = \mathbf{U}\,\mathrm{diag}(\mathbf{S}')\,\mathbf{U}^\top, \tag{4}$$

where $\mathbf{S}'$ is derived from a λ-weighted power-series expansion capturing even-hop components (2-, 4-, . . .) [18]. Finally, dominant causal relations are enforced to be closer than suppressed ones in hyperbolic space via a ranking objective:

$$\mathcal{L}_{geo} = \sum_{(i,j^+,j^-)} \log\left(1 + \exp\left(\mathrm{hop}(i,j^-) - \mathrm{hop}(i,j^+)\right)\right). \tag{5}$$

Summarized advantage: Our CGL proposes a relational causal-geometric representation that suppresses spurious correlations. By deconfounding quality factors from visual degradations, CGL preserves hierarchical topology and yields stable, discriminative, and robust quality representations.

2.2 Nonlinear Relational Aggregation (NRA)

The NRA disentangles quality representation from redundant relational structures by distilling causality-preserving factors into compact and discriminative quality embeddings. The $\mathbf{X}_{\text{cgl}}$ are subsequently encoded by a Transformer to capture contextual dependencies, yielding relationally enriched embeddings $\{t_i\}_{i=1}^{N}$.

To enhance nonlinear relational expressiveness, each token is transformed through a dimension-wise functional decomposition:

$$\mathcal{F}(t) = \sum_{j=1}^{D} \sum_{m=1}^{M} w_{jm}\, \phi_{jm}(t_j), \tag{6}$$

where t_j is the j-th channel of token t, and $\phi_{jm}(\cdot)$ are learnable piecewise spline basis functions [12]. The nonlinear residual aggregation is defined as

$$f' = f + \beta \sum_{i=1}^{N} \psi_i\, \mathcal{F}(t_i), \tag{7}$$

where ψ_i is a relational importance weight, and β is a learnable scaling parameter. This residual enhancement selectively amplifies relational cues while suppressing redundant structural dependencies. Finally, the aggregated embedding is mapped to the quality prediction space via a scalar nonlinear functional operator: $\hat{y} = \sum_{m=1}^{M} u_m\, \psi_m(f')$, $\psi_m(\cdot)$ denotes learnable piecewise univariate functions.

Summarized advantage: Our NRA proposes a structure-aware nonlinear aggregation that alleviates feature entanglement. It disentangles quality information from relational structures, yielding compact and discriminative embeddings for robust quality prediction.

3 Experiments

3.1 Experimental Setup

Datasets. We evaluate GeoCAN on two public datasets and one private dataset. 1) CACTUS [5] contains 37736 echocardiographic images spanning six views with 11 annotated quality grades ranging from 0 to 10, and is split into training, validation, and test sets in a 4:1:1 ratio. 2) CLINAv-GKD [11] comprises 10 echocardiographic views with five quality grades, 7163 images are used for training and validation, and 1350 images for testing. 3) Echo-Q is a private clinical dataset collected from routine echocardiography using Philips systems,

Table 1. Quantitative comparison on public datasets (CACTUS and CLINav-GKD) demonstrates superior accuracy and correlation performance of GeoCAN.

	CACTUS					**CLINav-GKD**				
Methods	QA	CA	PLCC	SRCC	KRCC	QA	CA	PLCC	SRCC	KRCC
KonCept[8]	0.8056	-	0.9599	0.9639	0.8992	0.9519	-	0.9475	0.9469	0.9356
IQT[3]	0.8586	-	0.9827	0.9811	0.9154	0.9385	-	0.9274	0.9270	0.9145
TRIQ[16]	0.8381	**0.9811**	0.9809	0.9785	0.9478	0.9467	**0.9933**	0.9693	0.9697	0.9405
LAR-IQA[9]	0.8201	-	0.9623	0.9577	0.9229	0.8665	-	0.8953	0.8910	0.8618
Causal-IQA[19]	0.7425	0.9581	0.8636	0.8582	0.7306	0.8330	0.9393	0.8627	0.8661	0.7703
TreS[7]	0.8346	0.9701	0.9692	0.9663	0.9236	0.9015	0.8874	0.9158	0.9161	0.8756
GeoCAN	**0.8740**	0.9809	**0.9850**	**0.9833**	**0.9587**	**0.9696**	0.9926	**0.9897**	**0.9899**	**0.9831**

comprising de-identified images annotated by three expert sonographers under rigorous quality control, with seven views, five quality grades, 16166 images for training/validation, and 3253 images for testing. **Implementation Details.** GeoCAN is implemented in PyTorch and trained on an NVIDIA RTX 5090 GPU. Ultrasound images are cropped to retain the fan-shaped acoustic region and resized to 224×224. We use a batch size of 16 and train for 50 epochs using Adam with a learning rate of 1×10^{-4}. Performance is evaluated using Quality Accuracy (QA), Classification Accuracy (CA), Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Correlation Coefficient (SRCC) and Kendall Rank Correlation Coefficient (KRCC).

3.2 Results and Analysis

Comparison with State-of-the-Arts (Table 1, Table 2). We compare GeoCAN with KonCept [8], IQT [3], TRIQ [16], LAR-IQA [9], Causal-IQA [19], and TreS [7], on two public datasets (Table 1) and Echo-Q (Table 2). Competing models are retrained on our datasets using the training codes provided by their respective authors to ensure fair comparison. The quantitative results reveal three key observations: 1) Superior correlation performance. GeoCAN achieves the strongest correlation performance across all datasets, surpassing the strongest baseline by up to 4.26 points in KRCC, with consistent gains in PLCC and SRCC. These improvements reflect enhanced ordinal consistency and more robust modeling of relative quality differences. 2) Strong multi-task performance. GeoCAN improves QA by up to 2.29% over the strongest baseline and consistently maintains leading classification accuracy across all datasets, indicating enhanced multi-task learning (MTL) discriminability. 3) Stable clinical generalization. Competing methods degrade noticeably on the private Echo-Q dataset, whereas GeoCAN remains stable across correlation and classification metrics. This stability indicates reduced reliance on visual correlations and improved robustness.

Ablation Studies. (Table 3, Table 4). We conduct ablation studies to evaluate the contribution of each component in GeoCAN on both public (Table 3)

Table 2. Quantitative comparison on the private clinical Echo-Q dataset demonstrates superior accuracy and correlation performance of GeoCAN.

Methods	QA	CA	PLCC	SRCC	KRCC
KonCept[8]	0.6403	-	0.7352	0.7256	0.6731
IQT[3]	0.8478	-	0.8686	0.8684	0.8369
TRIQ[16]	0.8564	0.9917	0.8725	0.8705	0.8416
LAR-IQA[9]	0.8433	-	0.8935	0.8910	0.8618
Causal-IQA[19]	0.7546	0.8241	0.7763	0.7683	0.6671
TreS[7]	0.8539	-	0.8707	0.8694	0.8391
GeoCAN	**0.8721**	**0.9923**	**0.9027**	**0.8983**	**0.8770**

Table 3. Ablation study on public datasets (CACTUS and CLINav-GKD) demonstrates the effectiveness of the proposed components.

	CACTUS					**CLINav-GKD**				
Variants	QA	CA	PLCC	SRCC	KRCC	QA	CA	PLCC	SRCC	KRCC
w/o CGL	0.8621	0.9588	0.9695	0.9696	0.9427	0.9156	0.9696	0.9864	0.9467	0.8465
w/o NRA	0.8524	0.9679	0.9814	0.9796	0.9519	0.8778	0.9822	0.8915	0.8895	0.8664
w/o MTL	0.8652	-	0.8922	0.8684	0.8608	0.8652	-	0.8762	0.8756	0.8255
GeoCAN	**0.8740**	**0.9809**	**0.9850**	**0.9833**	**0.9587**	**0.9696**	**0.9926**	**0.9897**	**0.9899**	**0.9831**

and private (Table 4) datasets: 1) w/o CGL: extracted features are directly fed into the Transformer encoder without causal–geometric modeling; 2) w/o NRA: Transformer outputs are directly used for quality prediction without nonlinear relational aggregation; 3) w/o MTL: the view classification head is removed, and only the quality assessment task is optimized. The ablation results reveal three key findings: 1) CGL is essential for preserving ordinal quality structure. Removing CGL causes the largest drop in ranking metrics, with KRCC decreasing by up to 13.66 points on CLINav-GKD. This demonstrates that CGL is critical for maintaining stable pairwise quality ordering and capturing hierarchical quality relationships. 2) NRA mitigates feature entanglement in quality representation. Removing NRA leads to the largest declines in PLCC and SRCC, indicating reduced consistency between predicted and ground-truth quality scores. This supports the role of nonlinear relational aggregation in stabilizing quality representation. 3) MTL provides complementary supervision. Removing MTL leads to declines in both accuracy and correlation metrics. It demonstrates that joint view classification and quality assessment enable shared semantic learning, improving discriminability.

Table 4. Ablation study on private clinical Echo-Q dataset demonstrates the effectiveness of the proposed components.

Variants	QA	CA	PLCC	SRCC	KRCC
w/o CGL	0.8641	0.9763	0.8945	0.8933	0.8636
w/o NRA	0.8592	0.9814	0.8777	0.8764	0.8476
w/o MTL	0.8632	-	0.8794	0.8777	0.8498
GeoCAN	**0.8721**	**0.9923**	**0.9027**	**0.8983**	**0.8770**

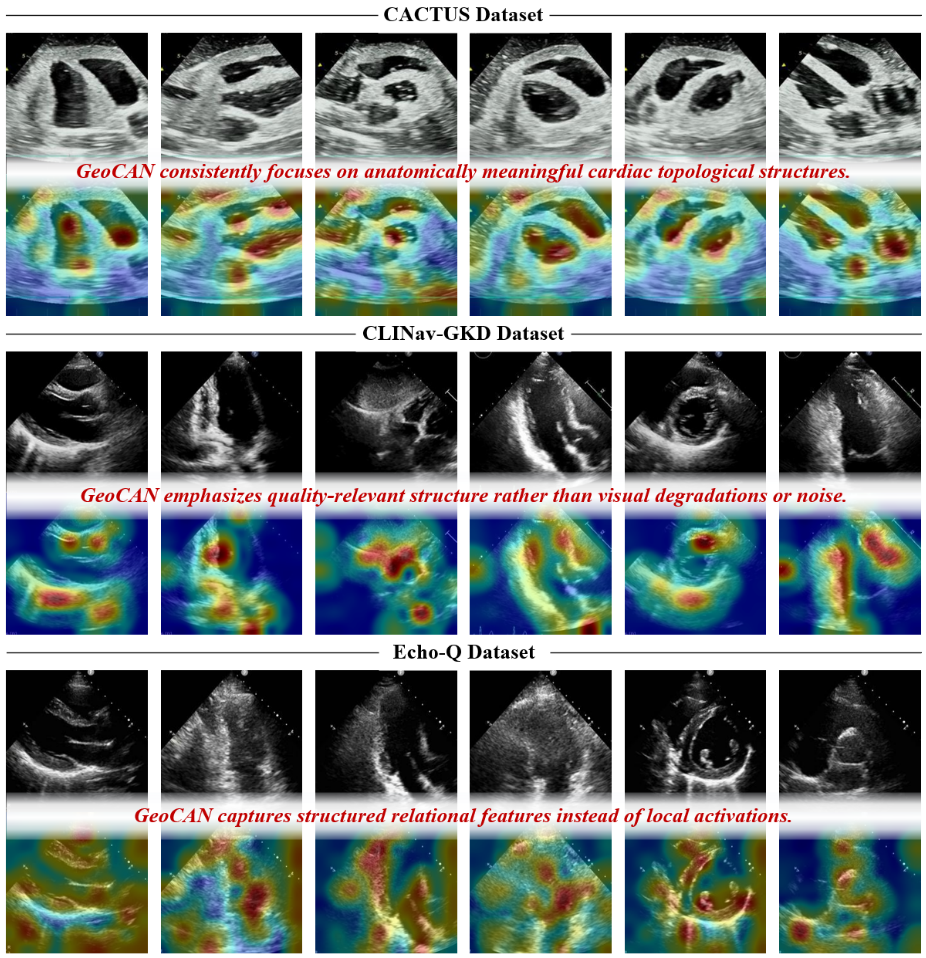

Fig. 3. Qualitative visualization of GeoCAN

Qualitative analysis. Qualitative results (Fig. 3) validate the effectiveness of GeoCAN. GeoCAN consistently focuses on anatomically meaningful cardiac topological structures, with attention concentrated on chamber boundaries and myocardial regions closely associated with image quality. It emphasizes quality-relevant structural cues rather than superficial degradations or visually correlated artifacts, indicating alignment with quality-determining factors instead of spurious visual correlations. The attention patterns exhibit relational structure beyond isolated local activations, supporting the role of nonlinear relational aggregation in mitigating feature entanglement and promoting topology-aware quality representation.

4 Conclusion

In this paper, we advance nonlinear causal–geometric learning for echocardiography quality assessment by proposing GeoCAN. CGL integrates latent hyperbolic geometry into causal modeling to deconfound quality factors from visual degradations and suppress spurious correlations. NRA disentangles relational structures from quality representations, enabling independent modeling of relational and visual variations while reducing feature entanglement. By unifying causal and geometric relational learning, GeoCAN provides a principled solution for topology-aware robust quality representation, offering a promising direction for reliable and interpretable clinical assessment.

Acknowledgments. This work was supported by the Young Scholars Fund of the Independent Innovation Research Program of China University of Petroleum under Grant 26CX06027A.

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

References

1. Carloni, G., Pachetti, E., Colantonio, S.: Causality-driven one-shot learning for prostate cancer grading from mri. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 2616–2624 (October 2023)
2. Chen, B., Solebo, A.L., Bao, W., Taylor, P.: Medical image quality assessment based on probability of necessity and sufficiency. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–4. IEEE (2025)
3. Cheon, M., Yoon, S.J., Kang, B., Lee, J.: Perceptual image quality assessment with transformers. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 433–442 (2021)
4. Christensen, M., Vukadinovic, M., Yuan, N., Ouyang, D.: Vision–language foundation model for echocardiogram interpretation. Nature Medicine **30**(5), 1481–1488 (2024)
5. Elmekki, H., Alagha, A., Sami, H., Spilkin, A., Mariel Zanuttini, A., Zakeri, E., Bentahar, J., Kadem, L., Xie, W.F., Pibarot, P., Mizouni, R., Otrok, H., Singh, S., Mourad, A.: Cactus: An open dataset and framework for automated cardiac assessment and classification of ultrasound images using deep transfer learning (2025)
6. Ghorbani, A., Ouyang, D., Abid, A., He, B., Chen, J.H., Harrington, R.A., Liang, D.H., Ashley, E.A., Zou, J.Y.: Deep learning interpretation of echocardiograms. NPJ digital medicine **3**(1), 10 (2020)
7. Golestaneh, S.A., Dadsetan, S., Kitani, K.M.: No-reference image quality assessment via transformers, relative ranking, and self-consistency. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3209–3218 (2022)
8. Hosu, V., Lin, H., Sziranyi, T., Saupe, D.: Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. IEEE Transactions on Image Processing **29**, 4041–4056 (2020)

9. Jamshidi Avanaki, N., Ghildyal, A., Barman, N., Zadtootaghaj, S.: Lar-iqa: A lightweight, accurate, and robust no-reference image quality assessment model. In: Computer Vision – ECCV 2024 Workshops. pp. 328–345. Cham (2025)
10. Jiang, H., Sun, Z., Jia, N., Li, M., Sun, Y., Luo, S., Song, S., Huang, G.: Cardiac copilot: Automatic probe guidance for echocardiography with world model. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 190–199. Cham (2024)
11. Li, Y., Fan, Y., Cui, X., Zhang, Y., Song, J., Hu, X., Chen, Y., Zheng, K., Cui, L., Liu, Z., et al.: Clinav-gkd: Vision-language latent hyperbolic geometric knowledge distillation for real-world 6-dof echocardiography probe navigation. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2417–2422. IEEE (2025)
12. 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)
13. Lopez-Paz, D., Nishihara, R., Chintala, S., Scholkopf, B., Bottou, L.: Discovering causal signals in images. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6979–6987 (2017)
14. Terziyan, V., Malyk, D., Golovianko, M., Branytskyi, V.: Hyper-flexible convolutional neural networks based on generalized lehmer and power means. Neural Networks **155**, 177–203 (2022)
15. Vukadinovic, M., Chiu, I.M., Tang, X., Yuan, N., Chen, T.Y., Cheng, P., Li, D., Cheng, S., He, B., Ouyang, D.: Comprehensive echocardiogram evaluation with view primed vision language ai. Nature pp. 1–3 (2025)
16. You, J., Korhonen, J.: Transformer for image quality assessment. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 1389–1393 (2021)
17. Yue, Y., Wang, Y., Jiang, H., Liu, P., Song, S., Huang, G.: Echoworld: Learning motion-aware world models for echocardiography probe guidance. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 25993–26003 (2025)
18. Zheng, K., Wang, J., Xu, J.: Graph-theoretic insights into bayesian personalized ranking for recommendation. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025)
19. Zhong, Y., Wu, X., Zhang, L., Yang, C., Jiang, T.: Causal-IQA: Towards the generalization of image quality assessment based on causal inference. In: Proceedings of the 41st International Conference on Machine Learning. pp. 61747–61762 (2024)