

Graph-of-Differences: Anatomy-Structured Difference Alignment for Medical Image Re-Identification

Nichula Wasalathilaka^{1*}, Abhijit Das^{3*}, Imran Razzak^{3,4}, and Dwarikanath Mahapatra²

¹ University of Peradeniya, Sri Lanka

² Khalifa University, Abu Dhabi, UAE

³ Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE

⁴ MedOS, Abu Dhabi, UAE

{abhijit.das, imran.razzak}@mbzuai.ac.ae

Abstract. Medical image re-identification (MedReID) enables longitudinal patient linkage but remains vulnerable to shortcut learning and produces decisions that clinicians cannot audit against named anatomy. We propose **Graph-of-Differences (GoD)**, which grounds identity comparisons in explicit anatomical structure. Each image is represented as an anatomy graph whose nodes correspond to named anatomical regions; given an image pair, soft node correspondence is established, and differences are computed over matched anatomy. A graph-level difference alignment objective ties these anatomy-matched differences to the global backbone difference, ensuring the retrieval signal is anchored in homologous structures rather than arbitrary spatial tokens. Explanations are defined over named graph nodes and quantitatively audited via node insertion/deletion tests—replacing unstable pixel heatmaps with verifiable, structure-level evidence. On internal benchmarks, GoD improves Rank-1 by **+7.1 pp** on fundus and **+3.1 pp** on CXR over a strong frozen-backbone baseline, with further gains on zero-shot external transfers confirming that anatomy grounding improves both accuracy and generalization. Code released at <https://github.com/GenMI-Lab/GoD.git>.

Keywords: Medical image re-identification · Anatomy graphs · Difference alignment · Auditability

1 Introduction

Medical images carry persistent, patient-specific biometric signals that survive de-identification established across chest X-rays [14,11], CT [7], and histopathology [5]. Medical image re-identification (MedReID) exploits these signals for longitudinal record linkage, cohort deduplication, and privacy risk auditing [16]. As these systems approach clinical deployment, a fundamental question emerges

*Equal contribution. Contact: dwarikanath.mahapatra@ku.ac.ae

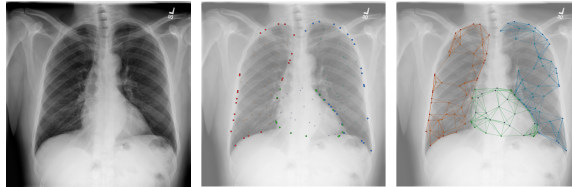


Fig. 1: **CXR anatomy graph construction.** (a) Input radiograph. (b) Nodes sampled from CheXmask masks: **left lung boundary**, **left interior**, **right lung boundary**, **right interior**, **heart boundary**, **heart interior**. (c) Graph $G(x) = (\mathcal{V}, \mathcal{E}, H)$ with k -NN ($k = 6$) and lung-symmetry edges; node features pooled from the frozen map $F(x)$ (Eq. 1).

that accuracy metrics alone cannot answer: *which anatomical structures is the model actually comparing, and can a clinician verify this?*

The gap accuracy cannot close. State-of-the-art MedReID (e.g., MaMI [16]) aligns inter-image feature differences in token space powerful, but with a structural limitation: differences are formed without any enforced correspondence to named anatomical regions. A difference vector may reflect the optic disc in one image and a vessel junction in the other, or worse, a scanner artefact or exposure shift [6]. Post-hoc saliency maps cannot reliably reveal this: pixel-level heatmaps are unstable under benign perturbations and unverifiable against named anatomy [13]. This is not an accuracy problem, it is an *auditability* problem where a mislinked record can cause incorrect treatment or a privacy breach.

Our solution. We propose **Graph-of-Differences (GoD)**, which enforces identity comparisons over *named anatomical structures* the same optic disc, the same lung boundary, the same cardiac contour—rather than arbitrary spatial tokens. Each image is represented as an anatomy graph whose nodes are named anatomical regions constructed by modality-specific parsers. Soft node-to-node correspondence is established across an image pair, differences are computed exclusively over matched anatomy, and a *graph-level difference alignment* objective ties these structured differences to the global backbone signal. Because every comparison is grounded in an identified anatomical node, explanations are attributions over *named structures* with quantitative faithfulness evidence. This per-pair, difference-over-matched-anatomy formulation distinguishes GoD from prior graph-based MedReID such as the Graph Matching Network of Manesco et al. [12], which encodes each chest radiograph as a *single-image* graph embedding over anatomical-landmark nodes: it neither establishes per-pair correspondence between named structures, nor computes differences over matched nodes, nor audits node-level attributions for faithfulness. Our contributions are—

1. **Anatomy-structured difference alignment.** A graph-level objective computes differences over soft-corresponded anatomical nodes and aligns them with the global difference signal, grounding identity decisions in homologous anatomy and mitigating shortcut learning.

2. **Structure-level interpretability with faithfulness evidence.** Node-level attributions over named anatomical structures are quantitatively audited via insertion/deletion curves, replacing unverifiable pixel heatmaps; full clinical auditability further requires radiologist user studies (future work).
3. **Consistent gains across benchmarks.** On fundus (ODIR, Messidor-2) and CXR (NIH14, CheXpert), GoD improves Rank-1 by **+7.1 pp** and **+3.1 pp** over MaMI internally, with largest relative gains on zero-shot external splits.

SOTA Methods like MaMI asks: *Can a single model re-identify patients across diverse modalities?* It achieves this via token-space difference alignment with a modality-adaptive backbone (ComPA), but without enforced anatomical correspondence—leaving identity decisions unverifiable against named anatomy.

GoD asks: *Can we guarantee that identity decisions are grounded in homologous anatomy, and can a clinician verify this?* This requires an architectural commitment—an explicit anatomy graph, enforced node correspondence, and a structured difference objective—that cannot be retrofitted post-hoc. GoD uses MaMI’s encoder as a *frozen* prior, so all gains arise purely from anatomy-structured difference alignment.

2 Method

Task. Given query x_a and a gallery, MedReID retrieves same-patient images by ranking via cosine similarity $s(x_a, x_b) = \cos(z(x_a), z(x_b))$, where $z(\cdot)$ is a learned retrieval embedding [11,14,16].

Backbone. We adopt the pretrained MaMI encoder [16] (ViT-B[3] + ComPA) as a *frozen* Siamese feature extractor, producing a global descriptor $g(x) \in \mathbb{R}^{768}$ and spatial feature map $F(x) \in \mathbb{R}^{768 \times 14 \times 14}$. Freezing isolates the contribution of anatomy-structured difference alignment from backbone fine-tuning; only the graph encoder, fusion module, and projection head $\phi(\cdot)$ are trained.

GoD pipeline (Fig. 1). For each image: (i) extract backbone features, (ii) build an anatomy graph with backbone-pooled node features, (iii) fuse graph descriptor $g_G(x)$ with $g(x)$ to form $z(x)$. Per pair: establish soft node correspondence, compute node-wise differences over matched anatomy, and align the graph difference with the backbone difference via $\mathcal{L}_{\text{diff}}$ grounding identity decisions in homologous structures and enabling node-level auditing.

2.1 Anatomy Graph Construction

Each image is represented as $G(x) = (\mathcal{V}, \mathcal{E}, H)$: nodes $\mathcal{V}$ are named anatomical regions, edges $\mathcal{E}$ sparse geometric relations, and features $H = \{h_i\}$ are masked-average-pooled from $F(x)$, inheriting pretrained semantics while remaining spatially localised:

$$h_i = \frac{\sum_{u,v} M_i(u, v) F(x)_{u,v}}{\sum_{u,v} M_i(u, v) + \epsilon}. \quad (1)$$

For **fundus** ($N \leq 34$), nodes cover optic-disc/macula landmarks, vessel junctions, and polar anchor nodes, canonicalised to a nasal-side optic disc orientation. For **CXR** ($N = 128$), CheXmask [4] parses left lung, right lung, and heart

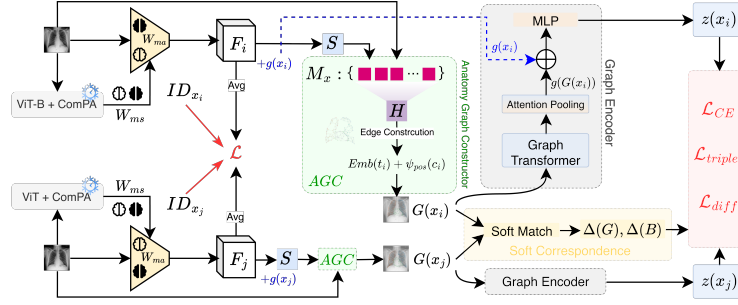


Fig. 2: **Overview of Graph-of-Differences (GoD).** A frozen MaMI encoder (ViT-B + ComPA) yields a global descriptor $g(x)$ and feature map $F(x)$. The Anatomy Graph Constructor (AGC) pools node features from $F(x)$ (Eq. 1) and builds graphs with k -NN and symmetry edges. A shared Graph Encoder yields $g_G(x)$, fused with $g(x)$ into the embedding $z(x)$. Per pair, soft correspondence computes anatomy-matched differences Δ_G , aligned with Δ_B via $\mathcal{L}_{\text{diff}}$ (Eq. 7).

masks; we sample 24 boundary + 24 interior nodes per lung and 16+16 for the heart, with farthest-point sampling[15] ensuring uniform spatial coverage. Each node connects to its $k=6$ nearest neighbours[20] (plus self-loops); CXR graphs additionally include left-right lung symmetry edges.

2.2 Graph Encoder and Retrieval Embedding

Graph transformer. A lightweight masked graph transformer[17,18] ($L=2$, $H_{\text{att}}=4$) encodes each node as:

$$\tilde{h}_i = h_i + \text{Emb}(t_i) + \phi_{\text{pos}}(c_i), \quad (2)$$

combining appearance, learnable type embedding, and coordinate MLP. Attention is masked to k -NN neighbourhoods[20], enforcing anatomically local message passing and yielding contextualised embeddings $H' = \{h'_i\}_{i=1}^N$.

Pooling and fusion. A learnable attention vector p pools H' into a graph descriptor[9] $g_G(x) \in \mathbb{R}^{768}$:

$$\alpha_i = \text{softmax}\left(\frac{p^\top h'_i}{\sqrt{d}}\right), \quad g_G(x) = \sum_i \alpha_i h'_i. \quad (3)$$

$g(x)$ and $g_G(x)$ carry complementary holistic and anatomy-structured signals; we concatenate and project to the final ℓ_2 -normalised embedding:

$$z(x) = \phi([g(x) \| g_G(x)]) / \|\cdot\|_2, \quad [g \| g_G] \in \mathbb{R}^{1536}, \quad z(x) \in \mathbb{R}^{512}. \quad (4)$$

2.3 Graph-of-Differences and Training

Soft correspondence. Given node embeddings $H_a \in \mathbb{R}^{N_a \times D}$, $H_b \in \mathbb{R}^{N_b \times D}$ for a pair (x_a, x_b) , we match nodes via temperature-scaled similarity and row-wise

Table 1: Evaluation data. Multi-site and cross-dataset settings stress-test robustness to acquisition variability where shortcut-exploiting models degrade most.

Modality	Split	Dataset	Source	Patients	Images
Fundus	Internal	ODIR-2019 [1]	Multi-site, multi-device	5,000	10,000
	External	Messidor-2[2]	Macula-centred, paired	874	1,748
CXR	Internal	NIH ChestX-ray14[19]	14-label, NLP-mined	30,805	112,120
	External	CheXpert[10]	Frontal + lateral	65,240	224,316

softmax, yielding matched embeddings $\tilde{H}_b = PH_b$:

$$S = \frac{1}{\tau} \text{norm}(H_a) \text{norm}(H_b)^\top, \quad P_{ij} = \text{softmax}_j(S_{ij}). \quad (5)$$

Unmatched nodes receive low correspondence mass, making this robust to parser imperfection and unequal node counts.

Difference alignment. Node-wise differences over matched anatomy and their mean-pooled aggregate are:

$$\delta_i = h_i(x_a) - \tilde{h}_i(x_b), \quad \Delta_G = \frac{1}{N_a} \sum_i \delta_i. \quad (6)$$

We align Δ_G with the global backbone difference $\Delta_B = g(x_a) - g(x_b)$ via a learnable projection W :

$$\mathcal{L}_{\text{diff}} = 1 - \cos(W\Delta_G, \Delta_B), \quad (7)$$

enforcing that anatomy-matched differences explain the global retrieval signal. Because $\mathcal{L}_{\text{diff}}$ operates directly on $\{\delta_i\}$, the insertion/deletion audit targets exactly the component the attributions describe; propagating attributions through the fusion into $z(x)$ would require learned credit assignment and is deferred (Sec. 5) rather than approximated. The full objective is[8]:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{triplet}} + \lambda \mathcal{L}_{\text{diff}}, \quad \lambda=0.1, \quad m=0.3. \quad (8)$$

3 Experiments

Datasets and Evaluation. We evaluate on fundus and CXR under **internal** (held-out split) and **external** (zero-shot transfer) protocols with patient-disjoint splits, reporting cumulative-matching-characteristic **Rank-1 retrieval accuracy (R1)**, i.e. the fraction of queries whose top-ranked gallery image is a true same-patient match, and **mean Average Precision (mAP)**. External gains cannot be attributed to overfitting and directly test whether anatomy grounding improves generalisation under real-world acquisition variability.

Implementation Details. The MaMI backbone (ViT-B + ComPA) is *frozen* throughout; only the graph encoder, fusion module, projection head, and identity classifier are trained via AdamW with cosine annealing over 30 epochs. All hyperparameters are reported in Table 6. Pairs are mined with a BatchHard

Table 2: MedReID performance across modalities.

Method	CXR				Fundus			
	Internal (NIH14)		External (CheXpert)		Internal (ODIR)		External (Messidor-2)	
	R1↑	mAP↑	R1↑	mAP↑	R1↑	mAP↑	R1↑	mAP↑
Backbone-only	0.8723	0.7867	0.3084	0.3788	0.6396	0.3631	0.6499	0.7432
Backbone + Graph (w/o $\mathcal{L}_{diff}$)	0.8659	0.7686	0.3068	0.3710	0.6357	0.3610	0.6327	0.7336
Full (Graph-of-Differences)	0.9033	0.8294	0.3359	0.3926	0.7111	0.3894	0.6579	0.7492

Table 3: **Component ablation.** (A) backbone-only; (B) +graph; (C) +correspondence; (D) full GoD. Mean $\pm$ std over three seeds.

	Components			Fundus (ODIR)		CXR (NIH14)	
	Graph	Corr.	$\mathcal{L}_{diff}$	R1↑	mAP↑	R1↑	mAP↑
(A)	–	–	–	0.6396 $\pm$ 0.0098	0.3631 $\pm$ 0.0032	0.8723 $\pm$ 0.0005	0.7867 $\pm$ 0.0007
(B)	✓	–	–	0.6357 $\pm$ 0.0020	0.3610 $\pm$ 0.0018	0.8659 $\pm$ 0.0041	0.7686 $\pm$ 0.0074
(C)	✓	✓	–	0.6569 $\pm$ 0.0096	0.3652 $\pm$ 0.0119	0.8797 $\pm$ 0.0068	0.7908 $\pm$ 0.0054
(D)	✓	✓	✓	0.7111 $\pm$ 0.0013	0.3894 $\pm$ 0.0010	0.9033 $\pm$ 0.0066	0.8294 $\pm$ 0.0113

sampler (batch size 64); inputs are resized to 224×224 with ImageNet normalisation and label smoothing 0.1. For external evaluation, models trained on the internal split are applied zero-shot to the external dataset with identical pre-processing and no fine-tuning. All experiments are run on 2 NVIDIA Quadro GV100 GPUs (32 GB each), Intel Xeon Gold 5218R CPUs (2 sockets, 80 cores, 80 threads), running Ubuntu 22.04.5 LTS.

4 Results

Re-Identification Performance. Table 2 reports Rank-1 and mAP across all four benchmarks. GoD outperforms the frozen-backbone baseline on every split: **+3.1 pp R1** / **+4.3 pp mAP** on CXR and **+7.1 pp R1** / **+2.6 pp mAP** on fundus (internal), with gains arising entirely from the graph encoder, fusion module, and $\mathcal{L}_{diff}$. The gain is orthogonal to backbone tuning: on a fine-tuned MaMI encoder GoD still adds **+2.5 pp** (CXR) and **+5.4 pp** (fundus) R1 (Table 4), with the best absolute result on the fine-tuned backbone. Adding a graph branch *without* correspondence or alignment yields marginal or negative changes, confirming gains come from *aligning differences over matched anatomy*, not graph structure alone. Gains hold on zero-shot external transfers (**+2.8 pp** CXR, **+0.8 pp** fundus R1): on Messidor-2, GoD reaches R1 = 0.6579 (95% CI [0.6491, 0.6664]) versus the backbone at 0.6499 ([0.6398, 0.6604]), indicating identity-stable anatomical signals rather than dataset-specific shortcuts.

4.1 Ablation Study

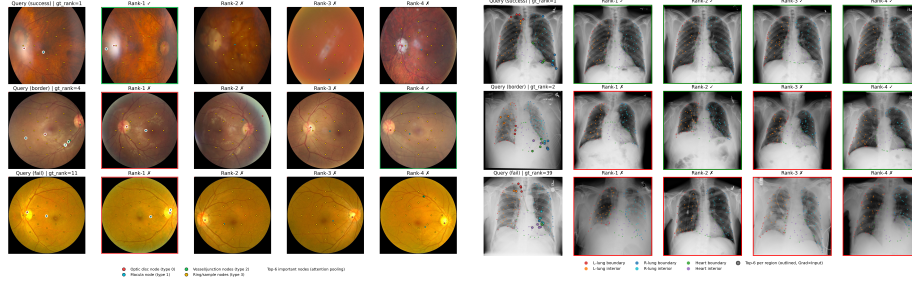
Table 3 progressively adds components over three seeds. A graph branch alone (B) slightly hurts performance; adding soft correspondence (C) recovers and surpasses backbone-only; the full model with $\mathcal{L}_{diff}$ (D) yields the largest gains

Table 4: **GoD is orthogonal to backbone fine-tuning** ($R1\uparrow$). GoD improves both frozen and fine-tuned MaMI; best combines both. Frozen: three-seed means (Table 3).

Configuration	CXR	Fundus
Backbone, frozen	0.8723	0.6396
Backbone, fine-tuned	0.8841	0.6712
GoD, frozen	0.9033	0.7111
GoD, fine-tuned	0.9089	0.7253

Table 5: **Parser-robustness stress test** ($R1\uparrow$). GoD stays above the no-graph baseline through 50% node dropout and 10-px dilation.

Perturbation	CXR	Fundus
Clean (default)	0.9033	0.7111
10% nodes dropped	0.8987	0.7044
30% nodes dropped	0.8821	0.6852
50% nodes dropped	0.8543	0.6517
5-px dilation	0.8951	0.7038
10-px dilation	0.8782	0.6859
No graph (frozen)	0.8723	0.6396



(a) Fundus (ODIR). Nodes: optic disc, macula, vessel junction, ring/anchor. (b) CXR (NIH14). Nodes: left/right lung and heart boundary/interior (CheXmask). Highlighted = top-attended nodes in g_G . Highlighted = top- k GradXInput on g_G .

Fig. 3: **Anatomy-aware retrieval explanations.** Each row: one query and top-4 retrievals; green/red borders = correct/incorrect. Successful retrievals concentrate attribution on stable structures; failures are diffuse.

(+7.1 pp fundus, +3.1 pp CXR $R1$). The monotonic improvement validates that difference alignment—not graph expressivity—is the key driver.

4.2 Explainability and Auditability

Qualitative explanations (Fig. 3). In successful retrievals, high-attribution nodes concentrate on anatomically stable structures—optic disc and vessel junctions (fundus); lung boundaries and cardiac contours (CXR). Borderline and failure cases show diffuse or peripheral attributions, consistent with confounders such as exposure shifts or device artefacts. Because attributions are over *named nodes*, a clinician can directly interrogate which structures drove the decision—something pixel heatmaps cannot support.

Quantitative faithfulness audit (Fig. 4, Table 6). Counterfactual node deletion degrades $R1$ more than random removal ($\Delta AUC_{R1} = 0.0195$, CI [0.0046, 0.0342] on NIH14; 0.0223, CI [0.0107, 0.0348] on ODIR), and insertion restores performance faster than random at small m . This confirms attributions identify nodes the model *genuinely relies on*—not spurious gradients—distinguish-

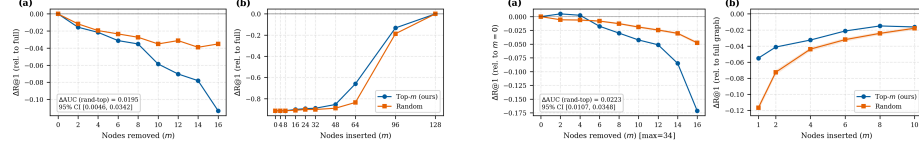


Fig. 4: **Node faithfulness audit.** Removing top- m attributed nodes degrades R1 more than random deletion (CXR: $\Delta\text{AUC}_{\text{R1}} = 0.0195$, CI [0.0046, 0.0342]; Fundus: 0.0223, CI [0.0107, 0.0348]); insertion restores R1 faster at small m . This confirms attributed nodes are causally relied upon, not spurious gradients.

Table 6: **Implementation details and ablations.** Node faithfulness uses Grad $\times$ Input (mean, 95 % bootstrap CI; higher is better). Graph granularity scales node count vs. default ($1.0\times$).

Optimisation						
Optimizer : AdamW (wd 10^{-4}) LR backbone : $10^{-3} \rightarrow 10^{-5}$ cos. Epochs : 30						
LR GoD : $5 \times 10^{-4} \rightarrow 10^{-5}$ cos.						
Losses & Graph						
$\lambda_{\text{tri}}, \lambda_{\text{diff}} : 1.0, 0.1$			Label smooth. $\epsilon : 0.1$		Margin m , scale $s : 0.3, 30$	
$\tau, k, L, H_{\text{att}} : 0.1, 6, 2, 4$			Nodes : Fundus ≤ 34 / CXR 128			
Node Faithfulness ($\Delta\text{Del-AUC}_{\text{R1}} \uparrow$)			Graph Granularity			
Dataset	Score [CI]	Ratio	CXR (NIH14)		Fundus (ODIR)	
			R1	mAP	R1	mAP
ODIR (Fundus)	0.0223 [.011, .035]	$0.5 \times$	0.9015	0.8297	0.6910	0.3842
NIH14 (CXR)	0.0195 [.005, .034]	$0.75 \times$	0.9064	0.8316	0.6983	0.3865
		$1.0 \times$	0.9105	0.8423	0.7124	0.3904
		$1.5 \times$	0.9128	0.8438	0.7159	0.3911
		$2.0 \times$	0.9037	0.8314	0.7118	0.3902

ing GoD from post-hoc saliency [13]. Graph granularity peaks at $1.5\times$ node scaling; sparser graphs lose discriminative structure, denser graphs dilute correspondence, validating our default sizes ($N \leq 34$ fundus; $N = 128$ CXR).

5 Discussion

The ablations show anatomy grounding pays off only when correspondence and difference alignment act together: a graph branch alone is inert, and gains come specifically from comparing *matched* nodes. The largest improvements appear on external, zero-shot benchmarks (CheXpert, Messidor-2), indicating identity-stable anatomical signals rather than dataset-specific shortcuts—the key property for cross-site deployment. The insertion/deletion tests confirm GoD’s attributions are *causally faithful*: high-attribution nodes are ones the model genuinely relies on, not spurious high-gradient locations. A radiologist can thus verify that a match was driven by *named anatomical nodes* rather than an unstable pixel

heatmap, and the same attributions identify which structures carry identity signal, enabling targeted de-identification that suppresses only those structures.

Limitations. GoD relies on modality-specific parsers; our stress test (Table 5) shows graceful degradation, staying above the no-graph baseline through 50% node dropout and 10-px dilation. It is limited to 2D, and node attributions are audited at the graph branch rather than propagated through the fusion into $z(x)$. Establishing clinical utility beyond this faithfulness evidence would require radiologist reader studies.

6 Conclusion

We presented Graph-of-Differences (GoD), which grounds medical image re-identification in named anatomy by computing differences over soft-corresponded nodes and aligning them with the global backbone signal. GoD improves retrieval over frozen and fine-tuned baselines, generalises to zero-shot external transfers, and yields node-level explanations with quantitative faithfulness evidence. Future work includes 3D graph correspondence for volumetric CT/MRI, propagating attributions through the fusion into $z(x)$, and a shared anatomical ontology for cross-modality retrieval and attribution-guided de-identification.

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

References

1. Peking university international competition on ocular disease intelligent recognition (odir-2019). <https://odir2019.grand-challenge.org/> (2019), grand Challenge dataset and competition on ocular disease classification
2. Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., Charton, B., Klein, J.C.: Feedback on a publicly distributed image database: The MESSIDOR database. *Image Analysis & Stereology* **33**(3), 231–234 (August 2014). <https://doi.org/10.5566/ias.1155>
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: *International Conference on Learning Representations (ICLR)* (2021)
4. Gaggion, N., Mosquera, C., Aineseder, M., Mansilla, L., Milone, D., Ferrante, E.: Chexmask database: A large-scale dataset of anatomical segmentation masks for chest x-ray images. *PhysioNet* (January 2025). <https://doi.org/10.13026/3705-zg36>, version 1.0.0
5. Ganz, J., Ammeling, J., Jabari, S., Breininger, K., Aubreville, M.: Re-identification from histopathology images. *Medical Image Analysis* **99**, 103335 (2025). <https://doi.org/10.1016/j.media.2024.103335>
6. Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., Wichmann, F.A.: Shortcut learning in deep neural networks. *Nature Machine Intelligence* **2**(11), 665–673 (2020). <https://doi.org/10.1038/s42256-020-00257-z>

7. Heinrich, A.: Automatic personal identification using a single ct image. *European Radiology* **35**(5), 2422–2433 (2025)
8. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017). <https://doi.org/10.48550/arXiv.1703.07737>, <https://arxiv.org/abs/1703.07737>
9. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. *CoRR* **abs/1802.04712** (2018), <http://arxiv.org/abs/1802.04712>
10. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghighi, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D.A., Halabi, S.S., Sandberg, J.K., Jones, R., Larson, D.B., Langlotz, C.P., Patel, B.N., Lungren, M.P., Ng, A.Y.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. vol. 33, pp. 590–597 (2019). <https://doi.org/10.1609/aaai.v33i01.3301590>
11. Macpherson, M.S., Hutchinson, C.E., Horst, C., Goh, V., Montana, G.: Patient reidentification from chest radiographs: An interpretable deep metric learning approach and its applications. *Radiology: Artificial Intelligence* **5**(6) (2023). <https://doi.org/10.1148/ryai.230019>
12. Manesco, J.R.R., Jodas, D., Zanella, M.J.G., Santos, M.K., Papa, J.P.: Graph feature embeddings for patient re-identification from chest x-ray images. In: *2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. pp. 1–6. IEEE (2024)
13. Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., van Keulen, M., Seifert, C.: From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable ai. *ACM Computing Surveys* **55**(13s), 1–42 (2023). <https://doi.org/10.1145/3583558>
14. Packhäuser, K., Gündel, S., Münster, N., Syben, C., Christlein, V., Maier, A.: Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest x-ray data. *Scientific Reports* **12**(1) (2022). <https://doi.org/10.1038/s41598-022-19045-3>
15. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 30 (2017)
16. Tian, Y., Ji, K., Zhang, R., Jiang, Y., Li, C., Wang, X., Zhai, G.: Towards all-in-one medical image re-identification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 30774–30786 (2025). <https://doi.org/10.1109/CVPR52734.2025.02866>
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 30 (2017)
18. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: *International Conference on Learning Representations (ICLR)* (2018)
19. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2097–2106 (2017). <https://doi.org/10.1109/CVPR.2017.369>
20. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)* **38**(5), 1–12 (2019)