CheXtriev: Anatomy-Centered Representation for Case-Based Retrieval of Chest Radiographs

Naren Akash R J, Arihanth Tadanki, and Jayanthi Sivaswamy

Center for Visual Information Technology,
International Institute of Information Technology Hyderabad, India
naren.akash@research.iiit.ac.in
https://cvit.iiit.ac.in/mip/projects/chextriev

Abstract. We present CheXtriev, a graph-based, anatomy-aware framework for chest radiograph retrieval. Unlike prior methods focussed on global features, our method leverages graph transformers to extract informative features from specific anatomical regions. Furthermore, it captures spatial context and the interplay between anatomical location and findings. This contextualization, grounded in evidence-based anatomy, results in a richer anatomy-aware representation and leads to more accurate, effective and efficient retrieval, particularly for less prevalent findings. CheXtriv outperforms state-of-the-art global and local approaches by 18% to 26% in retrieval accuracy and 11% to 23% in ranking quality. The code is available at https://github.com/cvit-mip/chextriev.

Keywords: Case-based Retrieval · Graph Transformer · Radiographs

1 Introduction

In clinical practice, expertise accumulates with experiences [1]. Clinicians often use case-based reasoning [2] to understand and diagnose patients by drawing parallels between past cases and current presentations. This analogy-driven approach hinges on effectively retrieving relevant past cases, making medical image retrieval (MIR) a cornerstone of evidence-based clinical decision-making. MIR facilitates the identification of similar past cases, allowing clinicians to: (i) refine diagnosis by suggesting investigative features or proposing alternative diagnoses, (ii) optimize treatment planning based on past outcomes, and (iii) enhance explanation and failure recovery by identifying clinically relevant past errors and their corrective actions [3]. As such, robust MIR systems hold significant potential to improve patient outcomes and advance medical care.

Despite its promise, MIR presents distinct challenges [4]. Unlike general-purpose image retrieval, which focuses on global image regions, MIR has to contend with images that exhibit similar global features across patients, with subtle, fine-grained markers serving as critical disease indicators. This is particularly

N. Akash and A. Tadanki contributed equally.

evident in chest radiographs, the most common yet challenging imaging modality. Its interpretation is prone to errors due to the two-dimensional projection of three-dimensional structures, leading to the superimposition of anatomies, low-contrast features, and multiple subtle, non-specific abnormalities. Studies suggest error rates as high as 30% in patients with abnormal findings and with up to 56% disagreement among radiologists [5]. MIR, especially for cases with multiple findings, remains understudied despite major strides in computer vision for multi-label settings [6].

Automated case-based retrieval of chest radiographs necessitates learning *discriminative yet informative representations* that can help bridge the gap between low-level visual features and high-level clinical interpretations. Hashing, widely used for computational and storage efficiency, maps high-dimensional features to compact hash codes. Advances in deep learning offer end-to-end training of deep hashing models such as DRH [7], OSDH [8] and SH-EBM [9] for multi-label chest radiographs, automatically learning both features and hash codes. However, they often lose crucial information related to disease classification and the regions of interest. This has been addressed recently with attention-based triplet hashing (ATH) [10] to implicitly focus on specific regions. ATH still relies on global image representation, potentially limiting its ability to capture fine-grained clinical information. Other emerging cross-modal approaches, such as MMDL [11] and X-TRA [12], incorporate text reports. Existing MIR approaches use image representations which do not fully leverage radiologists' systematic approach [13] to read chest radiographs, where the anatomical location of observed findings is often used for differential diagnosis. While focusing on anatomy has shown promise in classification (AnaXNet [14]), change detection (CheXRelNet [15]), and report generation (RGRG [16]), to our best knowledge, no work has attempted to incorporate anatomical reasoning into the retrieval process itself.

In this paper, we focus on MIR for chest radiographs aiming to find relevant cases solely from visual content. Our approach addresses the limitations of current multi-label retrieval methods by leveraging anatomy-aware representation learning. Unlike existing methods that rely on global features, we explicitly target the subtle, yet critical informative details from specific anatomical regions and fuse them to learn a richer representation that incorporates the interplay between anatomical location and radiological findings. This contextualization, grounded in evidence-based anatomy[17], leads to more accurate, effective and efficient retrieval compared to methods using solely global features. Furthermore, we incorporate anatomy-aware saliency to highlight regions contributing to the retrieval decision to build model understanding and transparency.

2 Method

Case-based chest radiograph retrieval systems identify images from a collection which are similar to a given query image. This is achieved by analyzing the

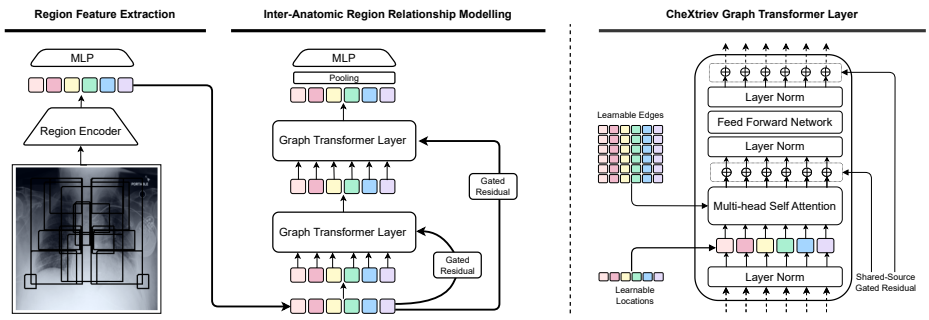

Fig. 1. A schematic illustration of CheXtriev.

visual content of the radiographs as well as focusing on the presence and location of anatomical abnormalities. In this work, we propose a system called CheXtriev, which first extracts informative features from anatomically defined regions and constructs a graph where nodes represent regions and edges capture spatial relationships and potential co-occurrences between findings. The graph leverages learnable location and edge embeddings to capture spatial context and relationships between regions, allowing a graph transformer architecture with global edge-aware attention and gated residuals to learn robust global-local context representations for accurate, effective and efficient retrieval. An overview of our proposed pipeline is shown in Figure 1.

2.1 Thoracic Radioanatomic Region Representation Extraction

Radiologists systematically analyse chest radiographs by sequentially assessing anatomical structures and interfaces to precisely detect abnormalities. Inspired by this, we identify eighteen essential anatomical regions in frontal radiographs. These regions include the right and left lungs, encompassing their apical, upper, mid, and lower zones, as well as the right and left hilar structures, costophrenic angles, mediastinum (and its upper portion), cardiac silhouette, and trachea. We obtain the anatomical region representations $R = \{r_i\}_{i=1}^{N} \in \mathbb{R}^{N \times D_R}$, where N is the number of regions and each r_i of dimension D_R is extracted by a fixed, pretrained ResNet50 feature extractor after the last global average pooling layer.

2.2 Modeling Inter-Region Relationships with Graph Transformer

The complex diagnostic reasoning process of a radiologist relies on anatomical interdependence, spatial pattern recognition, and disease co-occurrence. We design a novel graph transformer [18] framework, to capture both global and local contexts and learn latent relationships between different regions.

Feature projection and location embeddings. We project each of the region's features $r_i \in \mathbb{R}^{D_R}$ with a two-layer perceptron (MLP) with one ReLU nonlinearity. We *learn* N location embedding vectors $E^L \in \mathbb{R}^{D_R \times N}$ to distinguish

between one anatomical region from the other. Combining the features with corresponding embedding vectors provides a rich representation $\tilde{r}_i = \text{MLP}(r_i) + E_i^L$.

Edge-aware graph attention. To model anatomical relationships, we construct a graph with *learnable* continuous edges to capture inter-region dependencies. Each edge in the graph connects two regions, denoted as j and i (neighbour and reference region, respectively). We calculate C-way multi-head attention that incorporates both region features and learnable edge information as follows. Reference region features h_i and neighbouring region features h_j are transformed into query $q_{c,i}^{(l)} \in \mathbb{R}^{D_T}$ and key $k_{c,j}^{(l)} \in \mathbb{R}^{D_T}$ vectors using trainable weights $W_{c,q}^{(l)}$, $W_{c,k}^{(l)}$ and biases $b_{c,q}^{(l)}$ and $b_{c,k}^{(l)}$. $q_{c,i}^{(l)} = W_{c,q}^{(l)} h_i^{(l)} + b_{c,q}^{(l)}$; $k_{c,j}^{(l)} = W_{c,k}^{(l)} h_j^{(l)} + b_{c,k}^{(l)}$. Here, l corresponds to l^{th} layer. Edge features e_{ij} are learnt and encoded and added into the key vector as additional information for each layer, allowing the model to capture context-specific latent relationships beyond inherent node features: $e_{c,ij} = W_{c,e} e_{ij} + b_{c,e}$ and $\alpha_{c,ij}^{(l)} = \frac{\left\langle q_{c,i}^{(l)}, k_{c,j}^{(l)} + e_{c,ij} \right\rangle}{\sum_{u \in \mathcal{N}(i)} \left\langle q_{c,i}^{(l)}, k_{c,u}^{(l)} + e_{c,iu} \right\rangle}$, where $\langle q, k \rangle = \exp\left(\frac{q^T k}{\sqrt{D_T}}\right)$ is exponential scale cosine similarity function and D_T is the hidden size of each head. We calculate a weighted sum of neighbouring region features, incorporating both node and edge information, and make a message aggregation to the reference region as: $v_{c,j}^{(l)} = W_{c,v}^{(l)} h_j^{(l)} + b_{c,v}^{(l)}$; $\hat{h}_i^{(l+1)} = \|_{c=1}^{C} \sum_{j \in \mathcal{N}(i)} \alpha_{c,ij}^{(l)} (v_{c,j}^{(l)} + e_{c,ij})$.

Multi-level feature aggregation. We introduce shared-source gated residual connections to *selectively* aggregate global-local context-aware features as: $g_i^{(l)} = W_r^{(l)} h_i^{(l)} + b_r^{(l)}$; $\beta_i^{(l)} = \text{sigmoid}(W_g^{(l)})[\hat{h}_i^{(l)}; r_i^{(l)}; \hat{h}_i^{(l+1)} - r_i^{(l)}])$; $h_i^{(l+1)} = \text{ReLU}(\text{LayerNorm}((1 - \beta_i^{(l)}) \hat{h}_i^{(l+1)} + \beta_i^{(l+1)} h_i^{(0)}))$. The aggregated, hierarchical representation incorporates more comprehensive, fine-grained information at different levels of granularity to better detect abnormalities among different regions.

Classifier. We average the multi-head output on the last layer L of graph transformer and remove non-linearities: $\hat{h}_i^{(L+1)} = \frac{1}{C} \sum_{c=1}^{C} \sum_{j \in \mathcal{N}(i)} \alpha_{c,ij}^{(L)} (v_{c,j}^{(L)} + e_{c,ij})$; $h_i^{(L+1)} = (1 - \beta_i^{(L)}) \hat{h}_i^{(L+1)} + \beta_i^{(L)} g_i^{(L)}$. We feed the mean-pooled features into shared linear layers for multi-label classification using class-wise binary cross-entropy loss. We perform normalization on the final layer features to obtain dense vectors that capture anatomy-aware representations for similarity search.

2.3 Efficient Retrieval and Visual Interpretability for Similarity

To efficiently retrieve relevant chest radiographs, we adopt a fast and exact k-nearest neighbour retrieval model using FAISS [19]. It builds a flat index for our encoded, dense data embeddings, enabling efficient inner product similarity search with the query vector. This non-parametric approach effectively identifies radiographs with the highest cosine similarity from the pre-built vector database.

In order to visually interpret the results, we adopt an occlusion-based method [20] and derive similarity-based saliency maps. Instead of using small, sliding occlusions, we directly occlude pre-defined anatomical regions within the retrieved image. We then calculate the cosine similarity between the query image and each occluded version. Regions causing significant drops in cosine similarity are deemed critical for overall similarity, while those with minimal change are considered less important. This analysis is visualized as a saliency map, highlighting the most influential regions for similarity between the query and retrieved image.

3 Experiments

3.1 Data

We use the MIMIC-CXR-JPG v2.0.0 [21] database for all our experiments and follow the official MIMIC-CXR [22] data splits. We select the frontal chest radiograph images (PA or AP view) with valid bounding box coordinates for all eighteen anatomical regions. These annotations are provided by the Chest ImaGenome v1.0.0 [23] dataset, which is constructed from MIMIC-CXR. This resulted in 226,473 training, 1,863 validation, and 3,191 testing images. Similar to AnaXNet [14], we focus on nine specific findings: lung opacity (LO), pleural effusion (PE), atelectasis (AT), enlarged cardiac silhouette (ECS), pulmonary edema/hazy opacity (PE/HO), pneumothorax (PTX), consolidation (CONS), fluid overload/heart failure (FO/HF) and pneumonia (PN). We consider a radiograph positive for a finding if any region is labelled positive for that finding. We evaluate the retrieval performance by querying each test radiograph with findings against the remaining test set as the database.

3.2 Experimental Settings

In our experiments, we resize all chest radiographs to 224×224. The graph transformer consists of 2 layers with 8 multi-head attention heads and the model dimension is set to $D_T = 64$. We utilize the Adam optimizer with a learning rate of $1e-4$ in PyTorch Lightning. We incorporate early stopping based on validation loss with patience of 4 evaluations. Training is limited to a maximum of 30 epochs with a batch size of 16 and gradient accumulation occurs every 8 epochs. Further, we use a learning rate schedule with a reduction factor of 0.1 and set gradient clipping to 0.5. The training process was distributed across four NVIDIA GeForce RTX-2080 Ti GPUs.

Metrics. We use three metrics [10] to assess the precision and quality of retrieval: (i) Average Precision (AP) measures the average position of relevant cases within the retrieved list, thus evaluating the ranking quality, (ii) Hit Ratio (HR) indicates the proportion of retrieved cases that are relevant, thus reflecting the retrieval effectiveness, (iii) Reciprocal Rank (RR) is the reciprocal of the rank of the first relevant case in the returned list, assessing the retrieval efficiency.

Table 1. Impact of design choices in CheXtriev on retrieval and ranking performance. Here, IRM and MLF denote inter-anatomic region modelling using graph transformers and multi-level features with gated residuals, respectively.

Variants	IRM	Edge Connectivity	Location	MLF	mAP	mHR	mRR
V0	×	×	×	×	51.8	39.7	54.0
V1	GT	Shared Binary	×	Global	52.9	39.9	55.2
V2	GT	Shared Uniform	×	Global	53.9	40.8	56.3
V3	GT	Shared Uniform	Learnable	Global	54.0	**40.9**	56.4
V4	GT	Unique $R_i - R_j$	Learnable	Local	51.8	39.8	53.7
V5	GT	Unique $R_i - R_j$	×	Global	54.6	**40.9**	57.1
V6	GT	Unique $R_i - R_j$	Learnable	Global	**55.1**	40.7	**57.4**

4 Results and Discussion

Various variants of the proposed solution, namely, CheXtriev, were constructed to assess the contributions of various components. First, we present the evaluation results of these variants. Then, we highlight the strengths of CheXtriev, such as learning from anatomy-aware features via a local approach using the results of benchmarking experiments.

CheXtriev variants. The performance results of the variants are presented in Table 1. The baseline variant (V0) solely relies on the mean pooling of extracted region features and achieves 51.8% mAP, 39.7% mHR, and 54.0% mRR. The consistent performance boost observed from V1 to V6 (in global MLF), ranges from +2.1% (in mAP for V1) to +6.4% (for V6); this underscores the advantage of graph transformers with fully connected unique learnable edges over uniform edge-sharing schemes and naive handcrafted adjacency. This result also suggests that the retrieval task benefits from considering all region pairs, potentially capturing intricate latent relationships between regions. It can be seen that local gated residual connections (V4) lead to a significant drop in performance (6.38% lower mAP and 6.89% lower mRR) relative to a global one (V6). This emphasizes the value of global gated residual connections with selective refinement for learning multi-level features. The fact that V6 and V3 outperform V5 and V2, respectively, suggests that the learnable location embedding improves model performance by capturing crucial spatial context for accurate ranking.

Assessment of anatomy-aware feature extraction. A trend that can be observed from Table 1 is that relative to the baseline V0 which lacks inter-region relationship modelling, all graph transformer (GT) variants have a modest but consistent improvement in identifying relevant cases (similar mHR). This suggests that, in general, anatomy-aware feature extraction is more effective (than a strategy that does not consider anatomical relationships) at retrieving relevant cases, and the design of CheXtriev (V6) gives it an additional edge to precisely rank and retrieve these cases (higher mAP and mRR).

Table 2. Comparison of top-5 retrieval performance on MIMIC-CXR dataset against global baselines (CNN, ATH) and a local variant (AnaXNet). (.)* indicates $p < 0.05$.

	Global CNN			ATH [10]			AnaXNet [14]			CheXtriev		
Findings	AP	HR	RR	AP	HR	RR	AP	HR	RR	AP	HR	RR
LO	90.4*	85.9*	92.2*	88.2*	82.2*	90.9*	88.8*	82.9*	91.2*	**91.7**	**87.6**	**93.3**
PE	68.6*	54.7*	72.6	62.1*	46.3*	66.1*	63.7*	46.2*	67.2*	**71.4**	**59.4**	**74.7**
AT	60.3*	44.1*	64.2*	54.3*	36.8*	57.8*	59.1*	42.3*	62.8*	**64.4**	**49.1**	**67.6**
ECS	64.7*	48.3*	67.9*	61.3*	45.4*	64.9*	64.7*	48.5*	68.4*	**69.4**	**56.0**	**73.7**
PE/HO	56.5*	40.2*	59.2*	51.8*	34.7*	54.7*	53.5*	34.9*	57.4*	**62.5**	**47.8**	**65.8**
PTX	22.4*	8.5*	22.7*	8.9*	4.0*	9.1*	31.1	**12.7**	31.8	**31.5**	12.5	**32.4**
CONS	24.2*	13.2	25.2*	23.3*	11.8*	23.8*	27.6	12.5	28.3	**31.6**	**14.6**	**32.7**
FO/HF	14.8*	7.3*	14.8*	15.8*	7.0*	16.2*	17.8*	7.2*	17.9*	**28.8**	**11.8**	**29.3**
PN	39.0*	23.0*	41.3*	37.1*	21.9*	38.8*	41.4*	23.6*	43.4*	**44.7**	**27.1**	**47.2**
Mean	49.0	36.1	51.1	44.8	32.2	46.9	49.7	34.5	52.0	**55.1**	**40.7**	**57.4**
wMean	67.0	55.2	69.6	63.1	50.5	66.1	65.7	52.2	68.6	**70.8**	**59.5**	**73.4**

Assessment of global vs. local approaches. We compare the performance results of CheXtriev, against both global and local methods. Table 2 presents these results for top five retrieved images. Two global baselines are considered, one based on ResNet-50 extracted dense features (column 1) and ATH [10], a SOTA retrieval approach (column 2). A student's t-test was done to establish the statistical significance of the difference in results between CheXtriev and the baselines, and the significant values ($p < 0.05$) are marked with an asterisk. Relative to both the global baselines, CheXtriev has higher AP values across all nine investigated findings, with these ranging from 91.7% (LO) to 28.8% (FO/HF). This trend also holds for HR and RR metrics. Both macro-mean and weighted mean metrics are reported in the last two rows of Table 2 to draw insights into the overall retrieval efficacy from an unbalanced database. Macro-mean assigns equal weightage to all findings, which may introduce a bias to frequently occurring classes. Hence, the weighted mean assigns weights to classes inversely proportional to their frequency. It can be observed that the weighted mean is higher than the macro mean for all metrics. The mAP improvement over the baselines ranges from at least 12% (first column) to a notable 23% (second column). The noteworthy point is that the improvement is quite significant for classes with lower prevalence, such as FO/HF (+82.3% to +94.6%), PTX (+40.6% to +253.9%), CONS (+30.6% to +35.6%), PN (+14.6% to +20.5%), demonstrating CheXtriev's ability to learn more discriminative and powerful visual representations compared to global methods. These results point to CheXtriev's ability to accurately rank and efficiently retrieve relevant cases, even for less frequent classes.

We also compare CheXtriev against AnaXNet, a SOTA local approach designed for classification tasks. CheXtriev exhibited significant improvements in mean AP, particularly for findings associated with known *blind spots* [24,25] in chest radiographs, such as lung apices, hilar structures and inferior lung bases (for example, FO/HF +61.80%, PE/HO +16.82%, CONS +14.49% higher). This can

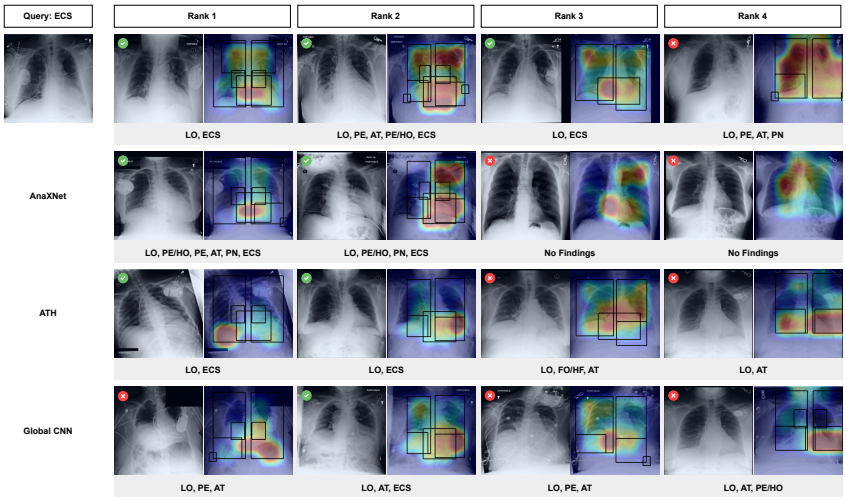

Fig. 2. Retrieval performance and saliency map analysis for a sample query image with enlarged cardiac silhouette (ECS). Each row displays the top 4 retrieved images and their corresponding occlusion-based saliency maps generated by CheXtriev, AnaXNet, ATH and Global CNN (top to bottom). A retrieved image is considered correct if it matches the specific finding of interest, in this case ECS, as the query image.

be attributed to CheXtriev's global edge-aware attention mechanism and fully connected learnable continuous edges, which enable it to capture anatomical relationships and latent finding co-occurrences more effectively than AnaXNet's naive handcrafted binary edges and limited expressiveness from neighborhood aggregation. Additionally, CheXtriev's location embeddings and gated residual connections have aided in learning superior visual representations, building upon the strengths of local approaches to capture well the subtle abnormalities in areas prone to human error during interpretation. The results also indicate classification-optimized features may not be the most effective for retrieval.

Interpretability analysis. We analyze saliency maps [20] to assess which anatomical regions influenced retrieval for each query image (see Figure 2). The first three retrieved images by CheXtriev all exhibit ECS, with the saliency maps highlighting a focus on the cardiac silhouette region. However, in the saliency map for the fourth retrieved image, the model's attention diffused throughout the lung region, indicating an incorrect retrieval. Some images retrieved by AnaXNet lack any findings for ECS. Even though ATH and CNN manage to retrieve correct cases, the corresponding saliency maps highlight irrelevant regions. This suggests other models might be overly sensitive to visual changes in irrelevant parts of the image, potentially due to weaknesses in encoding spatial information and context within the image. We also note that these saliency maps may not be meaningful for methods that do not use anatomy-aware features, underscoring the additional benefits of our approach.

5 Conclusion

In this work, we propose CheXtriev, a graph-based radiograph retrieval model inspired by the systematic approach radiologists use to interpret radiographs and grounded in evidence-based anatomy. Key novel features of our approach were explicitly targeting informative details from specific anatomical regions, modelling the interplay between anatomical location and findings and fusing them into a richer anatomy-aware representation. Superior results over existing methods underscore the benefit of this contextualization in achieving more accurate, effective, and efficient case retrieval, particularly for the less prevalent findings. A preliminary analysis with anatomy-aware saliency maps indicates that it may be possible to use them, albeit with some further extension to cover multiple findings, for interpretability of the retrieved results. Overall, CheXtriev offers a promising approach for medical image retrieval tasks, particularly in chest radiography, where subtle anatomical variations hold significant diagnostic value.

Acknowledgments. We thank Dr. L.T. Kishore for his insights on how radiologists analyze and interpret chest radiographs in clinical practice.

Disclosure of Interests. There are no conflicts of interest to declare.

References

1. Kolodner, Janet L.: The Role of Experience in Development of Expertise. In: Annual AAAI Conference on Artificial Intelligence (1982).
2. Slade, Stephen: Case-based Reasoning: A Research Paradigm. In: AI Magazine, **12**(1), 42-42 (1991).
3. Kolodner, Janet L., et al.: Using Experience in Clinical Problem Solving: Introduction and Framework. IEEE Transactions on Systems, Man, and Cybernetics, **17**(3), 420-431 (1987).
4. Li, Zhongyu, et al.: Large-scale Retrieval for Medical Image Analytics: A Comprehensive Review. Medical image analysis **43**, 66-84, (2018).
5. Gefter, Warren B., et al.: Commonly Missed Findings on Chest Radiographs: Causes and Consequences. Chest, **163**(3), 650-661 (2023).
6. Rodrigues, Josiane, et al.: Deep Hashing for Multi-label Image Retrieval: A Survey. Artificial Intelligence Review, **53**, 5261-5307 (2020).
7. Conjeti, Sailesh, et al.: Hashing with Residual Networks for Image Retrieval. In: Medical Image Computing and Computer Assisted Intervention MICCAI, LNCS. Springer, Cham. (2017).
8. Chen, Zhixiang, et al.: Order-Sensitive Deep Hashing for Multimorbidity Medical Image Retrieval. In: Medical Image Computing and Computer Assisted Intervention – MICCAI, LNCS. Springer, Cham. (2018).
9. Huang, Peng, et al.: Energy-Based Supervised Hashing for Multimorbidity Image Retrieval. In: Medical Image Computing and Computer Assisted Intervention - MICCAI, LNCS. Springer, Cham. (2021).
10. Fang, Jiansheng, et al.: Deep Triplet Hashing Network for Case-based Medical Image Retrieval. Medical Image Analysis 69, 101981 (2021).

11. Yu, Yang, et al.: Multimodal Multitask Deep Learning for X-Ray Image Retrieval. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI, LNCS. Springer, Cham. (2021).
12. van Sonsbeek, Tom, et al.: X-TRA: Improving Chest X-ray Tasks with Cross-Modal Retrieval Augmentation. In: International Conference on Information Processing in Medical Imaging - IPMI, LNCS. Springer, Cham. (2023).
13. Raoof, Suhail, et al.: Interpretation of Plain Chest Roentgenogram. Chest, **141**(2), 545-558 (2012).
14. Agu, Nkechinyere N., et al.: AnaXNet: Anatomy Aware multi-label Finding Classification in Chest X-ray. In: Medical Image Computing and Computer Assisted Intervention - MICCAI, LNCS. Springer, Cham. (2021).
15. Karwande, Gaurang, et al.: CheXRelNet: An Anatomy-Aware Model for Tracking Longitudinal Relationships Between Chest X-rays. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI, LNCS. Springer, Cham. (2022).
16. Tanida, Tim, et al. Interactive and Explainable Region-guided Radiology Report Generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2023).
17. Yammine, Kaissar: Evidence-based Anatomy. Clinical Anatomy, **27**(6), 847-852, (2014).
18. Dwivedi, Vijay Prakash, et al.: A Generalization of Transformer Networks to Graphs. In: AAAI 2021 Workshop on Deep Learning on Graphs: Methods and Applications (2020).
19. Douze, Matthijs, et al.: The Faiss Library. arXiv preprint arXiv:2401.08281 (2024).
20. Hu, Brian, et al.: X-MIR: Explainable Medical Image Retrieval. In IEEE/CVF Winter Conference on Applications of Computer Vision (2022).
21. Johnson, Alistair, et al.: MIMIC-CXR-JPG, A Large Publicly Available Database of Labeled Chest Radiographs. arXiv preprint arXiv:1901.07042 (2019).
22. Johnson, Alistair, et al.: MIMIC-CXR, a De-identified Publicly Available Database of Chest Radiographs with Free-text Reports. Scientific Data, **6**(1), 317 (2019).
23. Wu, Joy T, et al.: Chest Imagenome Dataset for Clinical Reasoning. In: Advances in Neural Information Processing Systems (2021).
24. Ropp, Alan, et al. Did I Miss That: Subtle and Commonly Missed Findings on Chest Radiographs. Current Problems in Diagnostic Radiology, **44**(3), 277-289 (2015).
25. de Groot, Patricia M., et al. Pitfalls in Chest Radiographic Interpretation: Blind Spots. Seminars in Roentgenology, **50**(3), 197-209, (2015).