Concept-induced Graph Perception Model for Interpretable Diagnosis

Lei Zhao[1][0000-0002-3397-4042], Changjian Chen[1], Bin Pu[2][0009-0007-8771-6501],
Xiaoming Qi[3][0000-0002-3238-2002], Fengfeng Peng[1][0009-0003-0938-1970],
Chunlian Wang[4,*], Kenli Li[1,*], and Guanghua Tan[1,5][0000-0001-6001-2351]

[1] The College of Computer Science and Electronic Engineering, Hunan university,
Changsha, China
lkl@hnu.edu.com
[2] Department of Electronic and Computer Engineering, The Hong Kong University
of Science and Technology, Hong Kong, China
[3] Department of Electrical and Computer Engineering, National University of
Singapore, Singapore
[4] Department of Ultrasound, Xiangtan Central Hospital, Xiangtan, China
[5] Department of Ultrasound, The Third Hospital of Changha, Changsha, China

Abstract. Due to the high stakes in medical decision-making, there is a compelling demand for interpretable deep learning methods in medical image analysis. Concept-based interpretable models, which predict human-understandable concepts (e.g., plaque or telangiectasia in skin images) prior to making the final prediction (e.g., skin disease type), provide valuable insights into the decision-making processes of the model. However, existing concept-based models often overlook the intricate relationships between image sub-regions and treat concepts in isolation, leading to unreliable diagnostic decisions. To overcome these limitations, we propose a Concept-induced Graph Perception (CGP) Model for interpretable diagnosis. CGP probes concept-specific visual features from various image sub-regions and learns the interdependencies between these concepts through neighborhood structural learning and global contextual reasoning, ultimately generating diagnostic predictions based on the weighted importance of different concepts. Experimental results on three public medical datasets demonstrate that CGP mitigates the trade-off between task accuracy and interpretability, while maintaining robustness to real-world concept distortions.

Keywords: Explainable diagnosis · Concept-based interpretable model· Graph reasoning.

1 Introduction

Black-box deep learning methods have shown great promise in medical image analysis, offering the potential to revolutionize healthcare diagnostics and treatments, such as pneumonia detection [24, 21] and thyroid nodule diagnosis [26, 27]. Despite the encouraging performance, their opaque nature raises concerns about interpretability and trust [2, 12]. Therefore, it is crucial to develop interpretable approaches in medical image analysis to enhance understanding of the reasoning behind predictions [8, 18].

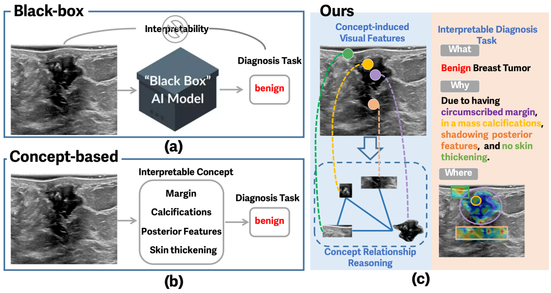


Fig. 1. (a) Black-box model diagnosis procedure. (b) Concept-based model diagnosis pipeline. (c) Our CGP model mimics expert diagnosis for more accurate and interpretable predictions.

Recently, concept-based approaches have been growing in popularity within interpretable deep learning, as they establish causal relationships between a set of human-understandable concepts and the final model decisions [1, 28]. A concept refers to a feature that is intuitively understandable by humans within the results generated by a model or a feature directly defined by the user [18, 17]. Specifically, unlike end-to-end black-box models in Fig. 1(a) that directly predict diagnostic outcomes from medical images, concept-based models first extract clinically relevant intermediate concepts from the input images and then predict the final diagnostic category based on these concepts. As shown in Fig. 1(b), concept-based approaches rely on features (such as lesion margin and calcification) to differentiate between malignant and benign tumors. For example, concept activation vectors (CAVs) [10] are used to project image representations into a concept subspace and subsequently verify whether the aggregated image representations contain clinical concepts. Concept Bottleneck Models (CBMs) [14], one of the most representative approaches, operate by first generating concepts from the input and then predicting the final label based on the identified concepts. Concept Embedding Models (CEMs) [5] improve CBMs by introducing positive and negative semantics to leverage high-level features in task prediction.

However, the existing studies have two limitations in the application of concept learning to clinical diagnosis: (1) Most existing methods rely on image-level features to learn concept information. In contrast, experts first identify symptoms by analyzing semantic concepts within different sub-regions of a medical image before making a diagnosis. Existing methods overlook the intricate semantics at the sub-region level, leading to unreliable concept detection. (2) Existing methods map input data into isolated concept embeddings or scalar concept scores for downstream diagnostic tasks. In clinical practice, however, experts make a diagnosis based on the interdependence of various concepts. These methods overlook critical interrelationships and are insufficient to provide in-depth reasoning for model diagnostic inference.

To address these issues, we propose an interpretable model called the Concept-induced Graph Perception (CGP) model that simulates expert diagnosis in Fig. 1(c). CGP first introduces a Concept-induced Adaptive Perception (CAP) module, which performs concept modeling and enables region-level feature perception under image-level label supervision. Building on this, CGP presents a Dual-scale Concept Graph Bottleneck (DCGB), which assesses different concept weights based on neighborhood structural learning and global contextual reasoning, and then makes a final diagnostic judgment. Specifically, our contributions are as follows:

(1) We propose CGP, an interpretable model for diagnosis. It takes a step toward mimicking the reasoning process of black-box models and provides logical, concept-level explanations for final diagnostic decisions. We provide both qualitative and quantitative results to demonstrate its state-of-the-art efficacy and reliability.

(2) We propose CAP, which employs human-interpretable concepts to guide and regularize the extraction of sub-region visual features through attention. CAP provides fine-grained visual-semantic information for subsequent relationship reasoning, while enhancing concept detection accuracy.

(3) We design DCGB to simulate the idea of diagnosis by clinical experts. It builds dual-scale graph reasoning to identify the importance of each concept and form final diagnostic predictions based on different concept weights.

2 Method

Fig. 2 illustrates the overall framework of our CGP model for interpretable computer-aided diagnosis. CGP incorporates the CAP (Section 2.1) module for concept modeling and the DCGB (Section 2.2) for assessing concept weights via structural and contextual reasoning, ultimately generating accurate diagnostic judgments.

2.1 CAP for Sub-region Perception by Image-level Label

CAP simulates the expert diagnosis process and performs concept modeling, enabling the perception of region-level feature space under the supervision of image-level labels. It involves 3 steps: *Concept Modeling*, *Perception Guidance*, and *Optimization Stage*.

In *Concept Modeling*, CAP identifies the attributes related to diseases and models these attributes in a form that the model can comprehend. Based on the image attributes of interest to clinical experts, a set of k concept classes is used to represent the attributes. In the concept learning process, CAP proposes using CLIP [22] pre-trained models to enable the model to understand the defined

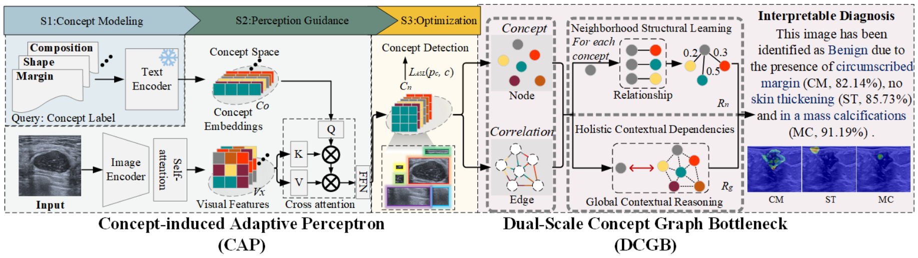


Fig. 2. The overall pipeline of our proposed framework simulates the clinical decision-making process through two main modules: CAP and DCGB. CAP extracts human-interpretable concept features from diverse sub-regions using image-level prompts. These visual-semantic features are structured into a concept graph, where DCGB performs dual-scale reasoning to model both neighborhood structural and global contextual dependencies. The final diagnosis is made by aggregating the weighted contributions of each concept, enabling accurate predictions with faithful visual and textual explanations.

concept text c, as embeddings $C_0 = [c_1, c_2, \ldots, c_k] \in \mathbb{R}^{k \times d}$. The constructed concept could be treated as the concept space for clinical diagnosis.

In *Perception Guidance*, CAP extracts the visual features of medical image x and regularizes the features focusing on the sub-regions related to the diagnosis according to the concept space. The abundant visual features $C_n^{(1)}$ are extracted by a backbone network and a self-attention module. The process can be expressed as $C_n^{(1)} = SelfAttn(\tilde{V}_x, \tilde{V}_x, V_x) \in \mathbb{R}^{HW \times d}$, where $V_x \in \mathbb{R}^{HW \times d}$ represents the linear projection of the backbone network-extracted visual features $V_0 \in \mathbb{R}^{H \times W \times d_0}$, and the tilde denotes the original vectors modified by adding positional embeddings [25]. For the extracted feature sequence $C_n^{(1)}$, CAP guides perception progress by associating the concept space with the visual feature, focusing on concept-related visual features through cross-attention: $C_n^{(2)} = CrossAttn(\tilde{C}_{n-1}^{(1)}, \tilde{C}_n^{(1)}, C_n^{(1)}) \in \mathbb{R}^{k \times d}$. Since concepts correspond to different image regions, visual features with high responses to the learned concepts represent the relevant regions. As a result, the outputs $C_n = FFN(C_n^{(2)}) \in \mathbb{R}^{k \times d}$ focus on the most relevant image regions for each concept label, effectively injecting class-specific visual context into the queries.

In *Optimization Stage*, concept-based learning is formulated to guide the establishment of correlations. CAP utilizes the concept-guided visual-semantic extractor to predict clinical concept features from medical images and obtain concept probabilities p_c through a fully connected layer. The concept detection task is trained using a simplified asymmetric loss [15] $L_{ASL}(p_c, c)$, where p_c and c are the prediction and ground truth for the concept detection task.

2.2 DCGB for Graph Reasoning Interpretable Diagnosis

DCGB captures both neighborhood structural and global contextual graph reasoning to simulate the way medical experts assess relationships between symptoms before making a final diagnosis. Specifically, based on the concept visual features $C_n = [c_n^{(1)}, c_n^{(2)}, \dots, c_n^{(k)}]$, we construct a concept relationship graph $G = (A, C_n)$, where each node corresponds to a specific concept. The adjacency matrix $A \in \mathbb{R}^{k \times k}$ defines the connectivity, which is determined by a fully connected layer.

Neighborhood structural learning. To learn the relationships between concepts in their local neighborhoods, we first compute the similarity matrix $S \in \mathbb{R}^{k \times k}$ based on concept features, where the similarity $s_{ij} \in S$ between concepts i and j is calculated as the dot product of their respective feature vectors $c_n^{(i)}$ and $c_n^{(j)}$. After obtaining the similarity matrix S, we perform element-wise multiplication with the normalized adjacency matrix $\hat{D}^{-1}\hat{A}$, resulting in $\hat{S}$ where similarity scores for non-connected concept nodes are all zeros. To this end, each row of $\hat{S}$ is averaged to compute the importance of each concept within its relevant neighborhood concepts. Finally, a softmax function is applied to obtain the final weights R_n, which reflect the relative importance of each concept based on its connections to surrounding concepts, as formalized below:

$$S = C_n {C_n}^T \in \mathbb{R}^{k \times k}, \quad \hat{S} = S \circ (\hat{D}^{-1}\hat{A}) \in \mathbb{R}^{k \times k}, \quad R_n = \text{softmax}\left(\frac{1}{k}\hat{S}\mathbf{1}\right) \in \mathbb{R}^k, \tag{1}$$

where $\hat{A} = A + I$ adds self-loops, $\hat{D}$ is the diagonal degree matrix with $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$, and $\circ$ denotes element-wise multiplication.

Global contextual reasoning. While neighborhood structural learning captures fine-grained relationships within each concept node's immediate neighborhood, global relational reasoning is essential for understanding the holistic contextual dependencies among concepts. We assess the significance of each concept node within the context of the entire concept graph as follows:

$$\hat{C_n} = \hat{D}^{-1}\hat{A}C_n \in \mathbb{R}^{k \times d}, \quad R_g = \text{softmax}(\hat{C_n}p) \in \mathbb{R}^k, \tag{2}$$

where a learnable vector $p \in \mathbb{R}^d$ is shared across all concept nodes and optimized jointly with the entire model during training. Typically, the neighboring region of a concept can be treated as a subgraph, with the concept itself serving as the center of this subgraph. By aggregating neighbor information for each concept node, the new features $\hat{C_n}$ represent the information from its subgraph. A learnable projection vector p, shared across all concept nodes in the concept graph, is then applied to the contextual-aggregated feature matrix $\hat{C_n}$, yielding the global importance weight R_g. Finally, the importance of each concept node is determined by $R = R_n + R_g$.

Based on the optimization of CAP, DCGB subsequently computes the concept importance weights. A linear predictor f_d is then applied to these concept importance weights, mapping the concept subspace to disease prediction. The linear predictor is highly interpretable, as its decisions are based on clinical concept importance weights, which align with expert reasoning. The weight matrix of the linear predictor reflects each concept's contribution to the final decision, and experts can adjust it for more reliable diagnoses when errors or counterintuitive results arise. Hence, the entire optimization process is as follows:

$$L_{ASL}(p_c, c) + L_{ASL}(f_d(R), y) + L_{Cons}(R, p_c), \tag{3}$$

where p_c and c are the prediction and ground truth for the concept detection task, and y denote ground truth for the disease classification task. L_{Cons} represents the mean squared error loss, and L_{ASL} is a simplified asymmetric loss.

3 Experiments

3.1 Experimental Setup

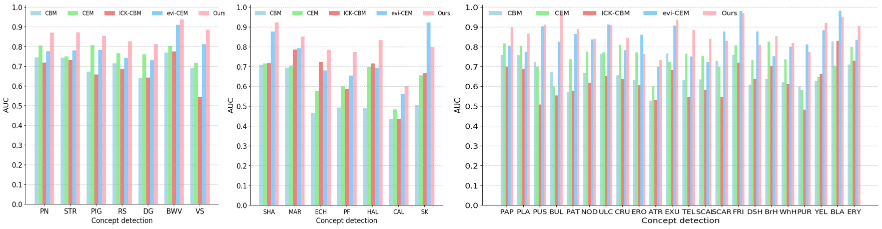
Datasets. Derm7pt [9] contains 1,011 dermoscopic images, including 20 distinct skin disease diagnoses and detailed annotations for 7 clinical concepts derived from the seven-point skin lesion malignancy checklist. We selected 827 images diagnosed as either Nevus or Melanoma. **Fitzpatrick 17k** [7] comprises 3,230 skin images categorized into Malignant, Benign, or Non-neoplastic classes. Consistent with previous studies, we selected 22 clinical concepts—each present in at least 50 images—from the 48 general medical concepts densely annotated by two dermatologists. **BrEaST** [20] is an ultrasound breast image dataset annotated with seven concepts derived from BI-RADS descriptors. It consists of 256 images across three diagnostic categories: Benign, Malignant, and Normal. For our study, we used 254 images classified as either Benign or Malignant, corresponding to abnormal breast conditions.

Implementation Details. We adopt the official PyTorch implementation for both the backbone and Transformer modules [25]. The model was trained for 100 epochs using the Adam optimizer [13] with a learning rate of 1×10^{-4}, true weight decay of 1×10^{-2}, and $(\beta_1, \beta_2) = (0.9, 0.9999)$. We adopt the 1-cycle learning rate schedule [23] and apply exponential moving average to model parameters with a decay factor of 0.9997. For regularization, we use Cutout [4] with a factor of 0.5 and true weight decay [16] of 1×10^{-2}. Input images are normalized with mean and standard deviation, and augmented with RandAugment [3].

Test-time Intervention for Faithfulness. We employ test-time intervention on concepts to assess faithfulness. During inference time, we first obtain the concept weights from the DCGB module, then intervene on specific concepts by adjusting their weights, and observe the resulting changes in the final model decisions. Fig. 4(a) presents two test-time intervention cases. In the first case, we increase the predicted weight of the atypical pigment network (APN), resulting in a corrected diagnosis from melanoma to nevus. In the second case, we set the

Table 1. Quantitative comparisons of disease diagnosis with state-of-the-art concept-based methods. The subscript "d" and "c" represent the performance of disease diagnosis and concept detection, respectively.

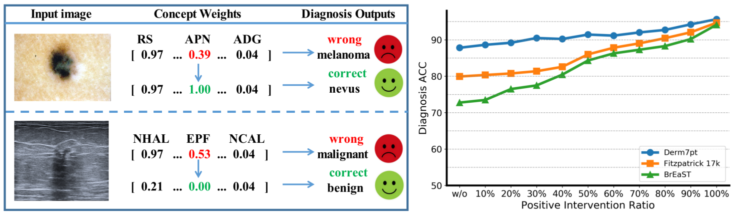
Method	Derm7pt				Fitzpatrick 17k				BrEaST			
	Disease Diagnosis		Concept Detection		Disease Diagnosis		Concept Detection		Disease Diagnosis		Concept Detection	
	ACC_d	AUC_d	ACC_c	AUC_c	ACC_d	AUC_d	ACC_c	AUC_c	ACC_d	AUC_d	ACC_c	AUC_c
CBM [14]	82.19	81.92	74.88	71.13	78.05	72.48	81.20	67.68	**74.91**	75.41	69.33	54.23
CEM [5]	81.56	76.39	79.40	77.29	75.36	75.08	88.69	73.47	74.51	72.20	79.41	63.41
ICK-CBM [19]	84.69	81.08	77.52	68.00	79.50	72.34	87.24	62.81	50.98	60.94	78.99	66.21
evi-CEM [6]	77.81	72.32	78.23	79.06	78.47	73.15	90.40	83.86	72.55	73.85	79.41	74.08
Ours	**87.81**	**88.47**	**82.55**	**86.61**	**79.92**	**81.20**	**92.73**	**86.93**	72.75	**89.06**	**81.93**	**79.52**



(a) Derm7pt (b) BrEaST (c) Fitzpatrick17k

Fig. 3. The fine-grained results of the concept detection task on the Derm7pt, BrEaST and Fitzpatrick17k datasets.

incorrectly predicted concept weight for empty posterior features (EPF) to 0, which aligns the model's decision with the dermatologists' findings. When positively intervening by adjusting the concept weights for wrong predictions, task accuracy consistently improved across all datasets in Fig. 4(b) as the proportion of interventions increased. This suggests that the predicted concepts faithfully explain the model's decision-making, which makes our approach highly adaptable and editable for real-world medical image diagnosis applications.



(a) Intervention cases **(b) Intervention results**

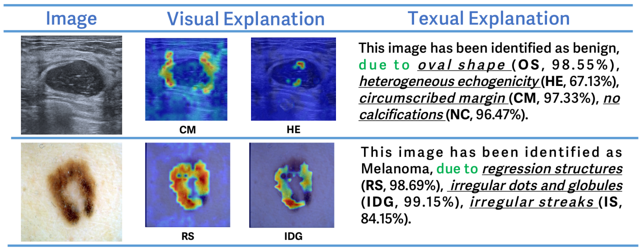
Fig. 4. Test-time interventions on concepts to correct model predictions.

Training with Uncertain Concept Labels for Robustness. In dermatological diagnosis, we evaluate the model's robustness for diagnostic tasks using uncertain training concept labels generated by the dermatosis foundation model

Table 2. Comparison of the diagnostic task on the Fitzpatrick 17k dataset, trained with uncertain concept labels for robustness analysis.

Method	CBM [14]		CEM [5]		ICK-CBM [19]		evi-CEM [6]		Ours	
	ACC_d	AUC_d	ACC_d	AUC_d	ACC_d	AUC_d	ACC_d	AUC_d	ACC_d	AUC_d
Uncertain label	79.50	72.56	77.43	72.48	79.09	72.84	**79.71**	73.57	79.19	**80.85**

[11], while retaining the original test data. Our CGP outperforms other methods, achieving an AUC of 80.85% for the diagnostic task, while competing methods achieve AUCs ranging from 72.48% to 73.57% in Table 2. Unlike other baseline models that combine embedded concept features or scalar concept scores linearly for downstream tasks, our CGP model leverages graph reasoning to capture the holistic relationships between concepts, ensuring robustness under various concept distortions and ambiguous labels.



This image has been identified as benign, due to *oval shape* (**OS**, 98.55%), *heterogeneous echogenicity* (**HE**, 67.13%), *circumscribed margin* (**CM**, 97.33%), *no calcifications* (**NC**, 96.47%).

This image has been identified as Melanoma, due to *regression structures* (**RS**, 98.69%), *irregular dots and globules* (**IDG**, 99.15%), *irregular streaks* (**IS**, 84.15%).

Fig. 5. Visual and textual explanations for our model's understandability across different datasets.

Diagnosis Explanations for Understandability. Understandability is key in explainable medical AI, ensuring model decisions are transparent and comprehensible to healthcare professionals. Fig. 5 provides detailed examples of the explanations. We present local visual and textual explanations regarding the decision-making process of our CGP model. For visual explanations, we generate concept activation maps by directly leveraging the attention weights from the Transformer's key-query interactions between image tokens (representing spatial regions) and concept tokens (learned embeddings). These weights naturally show each image region's contribution to a concept, allowing direct visualization of concept-specific attention maps without post-hoc gradients. Additionally, we include textual summaries of the final disease diagnosis results and the confidence scores for all concepts to establish causal relationships between explanations and model decisions.

4 Conclusion

In this paper, we propose a concept-based interpretable model named CGP to simulate the clinical expert's decision-making process, consisting of CAP and DCGB. In CGP, CAP employs human-interpretable concepts to guide the extraction of sub-region visual features. DCGB incorporates dual-scale concept relationship graph reasoning to guide accurate disease prediction. Extensive experiments demonstrate that CGP offers high interpretability, improved robustness, and superior performance. It provides both visual and textual explanations to establish causal relationships between explanations and model decisions. Additionally, CGP corrects errors through concept interventions in collaboration with human experts, improving transparency in medical AI for clinical applications.

Acknowledgments. This work was supported by the National Key R&D Program of China (Grant No. 2022YFF0606302), the National Natural Science Foundation of China (Grant No. 62272159), and in part by the Natural Science Foundation of Hunan Province (Grant No. 2025JJ60408).

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

References

1. Barbiero, P., Ciravegna, G., Giannini, F., Zarlenga, M.E., Magister, L.C., Tonda, A., Lió, P., Precioso, F., Jamnik, M., Marra, G.: Interpretable neural-symbolic concept reasoning. In: International Conference on Machine Learning. pp. 1801–1825. PMLR (2023)
2. Bie, Y., Luo, L., Chen, H.: Mica: Towards explainable skin lesion diagnosis via multi-level image-concept alignment. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 837–845 (2024)
3. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 702–703 (2020)
4. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
5. Espinosa Zarlenga, M., Barbiero, P., Ciravegna, G., Marra, G., Giannini, F., Diligenti, M., Shams, Z., Precioso, F., Melacci, S., Weller, A., et al.: Concept embedding models: Beyond the accuracy-explainability trade-off. Advances in Neural Information Processing Systems **35**, 21400–21413 (2022)
6. Gao, Y., Gao, Z., Gao, X., Liu, Y., Wang, B., Zhuang, X.: Evidential concept embedding models: Towards reliable concept explanations for skin disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 308–317. Springer (2024)
7. Groh, M., Harris, C., Soenksen, L., Lau, F., Han, R., Kim, A., Koochek, A., Badri, O.: Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1820–1828 (2021)

8. Hossain, M.I., Zamzmi, G., Mouton, P.R., Salekin, M.S., Sun, Y., Goldgof, D.: Explainable ai for medical data: current methods, limitations, and future directions. ACM Computing Surveys **57**(6), 1–46 (2025)
9. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE journal of biomedical and health informatics **23**(2), 538–546 (2018)
10. Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., et al.: Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In: International conference on machine learning. pp. 2668–2677. PMLR (2018)
11. Kim, C., Gadgil, S.U., DeGrave, A.J., Omiye, J.A., Cai, Z.R., Daneshjou, R., Lee, S.I.: Transparent medical image ai via an image–text foundation model grounded in medical literature. Nature Medicine **30**(4), 1154–1165 (2024)
12. Kim, I., Kim, J., Choi, J., Kim, H.J.: Concept bottleneck with visual concept filtering for explainable medical image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 225–233. Springer (2023)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Koh, P.W., Nguyen, T., Tang, Y.S., Mussmann, S., Pierson, E., Kim, B., Liang, P.: Concept bottleneck models. In: International conference on machine learning. pp. 5338–5348. PMLR (2020)
15. Liu, S., Zhang, L., Yang, X., Su, H., Zhu, J.: Query2label: A simple transformer way to multi-label classification. arXiv preprint arXiv:2107.10834 (2021)
16. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations
17. Luo, Y., Lu, Z., Liu, L., Huang, Q.: Deep fusion of human-machine knowledge with attention mechanism for breast cancer diagnosis. Biomedical Signal Processing and Control **84**, 104784 (2023)
18. Marcinkevičs, R., Wolfertstetter, P.R., Klimiene, U., Chin-Cheong, K., Paschke, A., Zerres, J., Denzinger, M., Niederberger, D., Wellmann, S., Ozkan, E., et al.: Interpretable and intervenable ultrasonography-based machine learning models for pediatric appendicitis. Medical Image Analysis **91**, 103042 (2024)
19. Pang, W., Ke, X., Tsutsui, S., Wen, B.: Integrating clinical knowledge into concept bottleneck models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 243–253. Springer (2024)
20. Pawłowska, A., Ćwierz-Pieńkowska, A., Domalik, A., Jaguś, D., Kasprzak, P., Matkowski, R., Fura, Ł., Nowicki, A., Żołek, N.: Curated benchmark dataset for ultrasound based breast lesion analysis. Scientific Data **11**(1), 148 (2024)
21. Pu, B., Wang, L., Yang, J., He, G., Dong, X., Li, S., Tan, Y., Chen, M., Jin, Z., Li, K., et al.: M3-uda: a new benchmark for unsupervised domain adaptive fetal cardiac structure detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11621–11630 (2024)
22. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
23. Smith, L.N.: A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. arXiv preprint arXiv:1803.09820 (2018)

24. Sun, Z., Gu, Y., Liu, Y., Zhang, Z., Zhao, Z., Xu, Y.: Position-guided prompt learning for anomaly detection in chest x-rays. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 567–577. Springer (2024)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
26. Wang, J., Zheng, N., Wan, H., Yao, Q., Jia, S., Zhang, X., Fu, S., Ruan, J., He, G., Chen, X., et al.: Deep learning models for thyroid nodules diagnosis of fine-needle aspiration biopsy: a retrospective, prospective, multicentre study in china. The Lancet Digital Health **6**(7), e458–e469 (2024)
27. Wu, X., Tan, G., Luo, H., Chen, Z., Pu, B., Li, S., Li, K.: A knowledge-interpretable multi-task learning framework for automated thyroid nodule diagnosis in ultrasound videos. Medical Image Analysis **91**, 103039 (2024)
28. Zhang, Y., Tiňo, P., Leonardis, A., Tang, K.: A survey on neural network interpretability. IEEE Transactions on Emerging Topics in Computational Intelligence **5**(5), 726–742 (2021)