Multi-Disease Diagnosis in Retinal Images via Patch-Level Reasoning and Selective Aggregation

Jianyang Xie[1], Yitian Zhao[2], Yanda Meng[3], Hao Zhang[4], Uazman Alam[1], and Yalin Zheng[1](✉)

[1]Eye and Vision Sciences Department, University of Liverpool, Liverpool, UK
[2]Ningbo Institute of Materials Technology and Engineering, CAS, China
[3] Bioengineering Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Saudi Arabia
[4] School of Software, Shandong University, Jinan, China
yalin.zheng@liverpool.ac.uk

Abstract. Diagnosing multiple diseases from retinal images is critical in real-world clinical applications, where patients often present with co-occurring conditions. However, existing approaches typically suffer from **co-occurrence-induced lesion-disease mismatching**: when diseases frequently co-occur, the model may spuriously associate lesions of one disease with another. Since all target diseases are present, this mismatch escapes supervision, leading to incorrect spatial reasoning and poor generalization when diseases appear independently during inference. To address this challenge, we propose **MD-PRSA**, a novel framework for **M**ulti-**D**isease diagnosis in retinal images via **P**atch-Level **R**easoning and **S**elective **A**ggregation. MD-PRSA employs a patch-based transformer that computes disease-specific predictions for each patch using learned queries, and then aggregates these predictions through a multiple instance learning (MIL) mechanism. This selective aggregation enforces a form of **spatial attribution constraint**, ensuring that only patches with strong disease-specific evidence contribute to the final decision. This design naturally suppresses mismatched patches caused by co-occurring conditions and provides more robust and interpretable diagnoses under weak supervision. We validate our method on two retinal image datasets with distinct imaging modalities: LID-FFA (fluorescein angiography) and ODIR-5K (color fundus photography). MD-PRSA achieves state-of-the-art results, improving mAUC by 0.84% on LID-FFA and 1.01% on ODIR-5K. The code have been released [1].

Keywords: Multi-disease Diagnosis · Patch-level Reasoning · Lesion-disease mismatching · Selective Aggregation.

1 Introduction

Retinal images are essential for computer-aided diagnosis of conditions such as glaucoma [5] and age-related macular degeneration [15]. Deep learning has

[1] https://github.com/davelailai/Multi-Disease-Diagnosis-in-Retinal-Images-via-Patch-Level-Reasoning-and-Selective-Aggregation.git

achieved remarkable success in detecting individual diseases [12, 24, 7, 22], but real-world cases often involve multiple co-occurring diseases [10], posing new diagnostic challenges. Existing multi-disease retinal diagnosis methods typically follow two paradigms: disease-localization, which identifies spatial regions associated with each disease [1, 20, 6], and disease-specific feature modeling, which extracts discriminative features for independent classification [3, 13, 23, 14].

Despite their effectiveness, both approaches struggle in the presence of frequent disease co-occurrence, often leading to a critical issue we term co-occurrence-induced lesion-disease mismatching. Where models may incorrectly attribute lesions from one disease to another. For instance, if disease j frequently co-occurs with disease i during training, the model may learn to associate features from j's lesions with the presence of i. As a result, during inference, the presence of disease j alone could spuriously trigger predictions for disease i, even if i is absent. This arises because standard models optimize only for image-level label correctness without supervision guiding which spatial regions are causally responsible. As a result, models can learn to attend to incorrect patches while still predicting the correct diseases, reinforcing spurious correlations over time.

To address this challenge, we propose MD-PRSA, a novel framework for multi-disease retinal diagnosis via **Patch-level Reasoning** and **Selective Aggregation**. Patch-level predictions are generated using a transformer-based decoder and then selectively aggregated through a Multiple Instance Learning (MIL) mechanism. This selective aggregation imposes a spatial attribution constraint, guiding the model to focus on causally relevant regions and avoid spurious correlations. Unlike standard transformer attention (e.g., Q2L [14]), which distributes weights across all patches without causal alignment [18], MIL enforces a task-driven competition among patches, effectively achieving implicit disease localization. To our knowledge, MD-PRSA is among the first frameworks to explicitly couple disease-specific patch reasoning with selective aggregation for mitigating lesion–disease mismatching in multi-disease retinal diagnosis.

In summury, our contributions are summarized as follows: (1) We introduce MD-PRSA, combining transformer-based patch-level reasoning with MIL-based selective aggregation for accurate and spatially grounded multi-disease diagnosis. (2) Our approach enforces a spatial attribution constraint, encouraging the model to focus on a minimal subset of causally relevant patches, mitigating co-occurrence-induced lesion-disease mismatching. (3) Extensive experiments on two retinal datasets with diverse imaging modalities validate the effectiveness and generalizability of our approach.

2 Method

2.1 Preliminary

Given a retinal image $\mathbf{X}$ potentially exhibiting multiple co-existing diseases, the goal of multi-disease diagnosis is to predict the presence of each disease. Let L denote the total number of disease categories, and let $\mathbf{Y} = [y_1, \ldots, y_L]$ represent

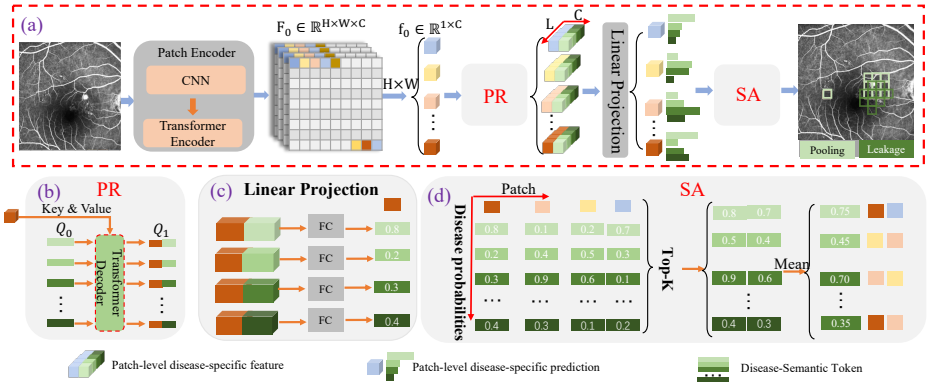

Fig. 1: Framework of the proposed MD-PRSA. As shown in (a), the input image is split into patches and processed by the Patch-Level Disease Reasoning (PR) module (b) using disease-specific queries Q_0 to extract disease-aware features. Patch-level predictions are obtained via a linear head (c). The Selective Aggregation (SA) module (d) selects the Top-K informative patches per disease to generate final image-level predictions.

the corresponding ground-truth labels, where $y_l \in \{0, 1\}$ indicates the absence or presence of the l-th disease. Specifically, $y_l = 1$ if image $\mathbf{X}$ contains the l-th disease, and $y_l = 0$ otherwise.

2.2 Overview of the Proposed Method

The architecture of our method is shown in Figure 1 and consists of three components: **Patch Encoder**: The input image is split into non-overlapping patches, processed by a CNN backbone and transformer encoder to capture local and global features. **Patch-Level Disease Reasoning**: A transformer decoder with disease-specific queries models interactions between disease semantics and patch features, enabling patch-wise disease prediction. **Selective Aggregation**: Top-K patch-level probabilities are aggregated to produce the final diagnosis, enforcing a spatial attribution constraint that suppresses spurious correlations.

2.3 Patch Encoder

Given an input retinal image $\mathbf{X} \in R^{H_0 \times W_0 \times C_0}$, we first extract spatial features using a hybrid encoder composed of a CNN backbone followed by a transformer encoder layer. The CNN captures local visual patterns, producing a feature map $\mathbf{F}_0 \in R^{H \times W \times C}$, where each feature corresponds to a local receptive field (i.e., a patch) of the original image. The transformer encoder then enriches these local features with global contextual dependencies. To facilitate patch-wise processing, the feature map $\mathbf{F}_0$ is reshaped into a sequence of patch tokens, $\mathbf{F}_0 \in R^{HW \times 1 \times C}$, where each token $f_0 \in R^{1 \times C}$ represents a patch embedding.

2.4 Patch-Level Disease Reasoning

To model disease presence at the patch level, we employ a transformer decoder that learns disease-specific representations using semantic queries, as illustrated in Figure 1(b). Unlike prior methods that aggregate patch features into a global representation [14, 17], our approach preserves spatial granularity by generating disease-aware features for each patch.

Specifically, we apply cross-attention between learnable disease queries and patch embeddings, enabling each patch to be interpreted in the semantic context of each disease. The decoder consists of a self-attention layer, a cross-attention layer, and a feed-forward network (FFN). Let $Q_0 \in R^{L\times C}$ denote the learnable disease queries, where L is the number of diseases. The decoding is defined as:

$$\begin{aligned} &\text{Self-Attention:} \quad Q \leftarrow \text{MultiHeadAttn}(Q_0, Q_0, Q_0) \\ &\text{Cross-Attention:} \quad Q \leftarrow \text{MultiHeadAttn}(Q, \mathbf{F}_0, \mathbf{F}_0) \\ &\text{Feed Forward:} \quad Q \leftarrow \text{FFN}(Q) \end{aligned} \tag{1}$$

where the disease-to-patch cross-attention allows each disease query to selectively attend to discriminative patches.

The resulting representation $Q \in R^{HW\times L\times C}$ provides disease-specific features for each patch, enabling joint reasoning over local visual content and disease semantics. This design promotes attention to disease-relevant regions and supports fine-grained, interpretable diagnosis.

2.5 Selective Aggregation for Diagnosis

HHaving obtained disease-specific patch-level features, we aggregate them for image-level diagnosis using a Multiple Instance Learning (MIL) formulation. Each retinal image $\mathbf{X}$ is treated as a bag of $N = H \times W$ instances (i.e., patches): $\mathbf{X} = \{x^1, x^2, \ldots, x^N\}$, where each patch x^n is associated with a disease-specific feature vector $q^n \in R^{L\times C}$. Under the MIL assumption, an image is positive for disease l if at least one patch indicates its presence:

$$y_l = \begin{cases} 1, & \text{if } \sum_{n=1}^{N} y_l^n > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

For each disease l, we first compute patch-level probabilities. Each disease-specific feature $q_l^n \in R^C$ is projected to a scalar logit via a linear layer followed by a sigmoid:

$$p_l^n = \text{Sigmoid}\left(\mathbf{W}_l^\top q_l^n + b_l\right) \tag{3}$$

where $\mathbf{W}_l \in R^C$ and $b_l \in R$ are learnable parameters. Stacking all patch-level predictions for disease l yields a score vector $\mathbf{P}_l = [p_l^1, \ldots, p_l^N]^\top \in R^N$.

To obtain the image-level prediction, we apply Top-k selective aggregation by averaging the k highest patch-level scores:

$$p_l = \frac{1}{k}\sum_{j=1}^{k} p_l^{n_j}, \quad n_j = \arg\max_k(\mathbf{P}_l). \tag{4}$$

This selective aggregation enforces a spatial attribution constraint, emphasizing the most discriminative regions while suppressing spurious responses caused by disease co-occurrence. Moreover, Top-k pooling provides a smooth interpolation between max- and average-pooling, balancing sensitivity and robustness.

3 Experiments

3.1 Datasets and Evaluation

We evaluate the proposed method on two retinal datasets with different imaging modalities: ODIR-5K [11] and LID-FFA [21].

ODIR-5K contains 10,000 binocular fundus images from 5,000 patients labeled across eight disease categories: Normal (N), diabetic retinopathy (D), glaucoma (G), cataract (C), age-related macular degeneration (A), hypertension (H), myopia (M), and other abnormalities (O). The dataset is split into training (3,500 patients), on-site test (1,000), and off-site test (500). For training, we further use an 80/20 split for training/validation, and evaluate on both test sets.

LID-FFA includes 5,435 fundus fluorescein angiography images from 500 patients with six pathological features: Leakage (L), Transmission and Pooling (TP), Staining (ST), Shading (SD), Non-Perfusion (NP), and Vessel Abnormality (VA). Multiple lesions may co-occur per image. We follow the official split of 350/50/100 patients for training/validation/testing.

Data Pre-processing. All images are resized to 512×512, randomly cropped to 448 × 448, and horizontally flipped. **Evaluation Metrics.** Performance is measured using macro F1-score (MAF1), mean average precision (mAP), and mean AUC (mAUC), with a threshold of 0.5.

3.2 Implementation Details

All networks were implemented using `mmpretrain` [4] and trained on an NVIDIA Tesla A100 GPU (80 GB). Optimization used stochastic gradient descent (SGD) with an initial learning rate of 0.005 and a one-cycle learning rate schedule [19]. Models were trained for 100 epochs with a batch size of 64. Asymmetric loss [16] is used for training.

3.3 Comparison with The State-Of-The-Art

We compared our method with representative baselines, including asymmetric loss [16] for class imbalance, GCN-based approaches [3, 23] for disease co-occurrence, Transformer-based methods [14, 17, 21] for disease-image interactions, and region localization methods [6] for identifying disease-relevant areas. All methods used a consistent asymmetric loss and a DenseNet121 [9] backbone pre-trained on ImageNet-1k. For our model and other Transformer baselines, the encoder and decoder had 1 and 2 layers, respectively. Three MIL aggregation strategies are evaluated: max pooling, average pooling, and Top-K average.

Table 1: Comparison with SOTA on LID-FFA and OIA-ODIR (%). Bold indicates best performance, underlined indicates improvement over other methods. $K = 10$ for LID-FFA, $K = 50$ for OIA-ODIR.

	LID-FFA			**OIA-ODIR (on-site)**			**OIA-ODIR (off-site)**		
Model	MAF1	mAP	mAUC	MAF1	mAP	mAUC	MAF1	mAP	mAUC
Base [8]	77.89	80.83	65.43	26.81	39.44	82.09	28.30	47.55	86.01
Asy [16]	83.56	87.94	82.50	46.10	51.78	85.96	53.03	57.82	87.07
GCN [13]	84.68	87.89	82.48	49.55	49.34	82.99	**56.89**	58.32	87.12
dyGCN [23]	85.02	87.98	83.51	44.73	48.04	86.07	50.44	58.33	87.69
ML [17]	85.27	87.73	84.32	47.01	42.91	86.48	48.41	56.80	88.12
Q2L [14]	84.80	87.36	84.11	47.06	49.52	86.22	51.49	57.82	87.30
Q2L Causal [21]	84.82	88.22	84.69	44.76	53.27	87.82	51.35	58.45	88.71
MCAR [6]	84.41	87.88	84.25	40.78	52.16	87.18	48.22	57.55	88.49
MD-PRSA_Avg	84.99	87.66	84.04	42.87	51.25	<u>88.96</u>	51.09	58.43	<u>88.96</u>
MD-PRSA_Max	<u>85.06</u>	**88.50**	<u>85.43</u>	48.15	51.81	<u>88.14</u>	53.40	57.74	<u>89.00</u>
MD-PRSA_Top-K	**85.89**	88.01	**85.53**	**50.40**	**54.96**	**88.83**	**55.67**	**62.32**	**89.72**

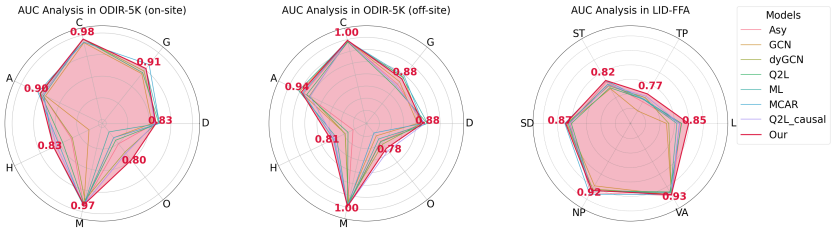

Fig. 2: AUC analysis for each disease.

Table 1 reports performance on ODIR-5K and LID-FFA. Using the optimal Top-K (Section 3.5), our method achieves the highest mAUC: +0.84% on LID-FFA, +1.01% on ODIR-5K (on-site), and +1.01% on ODIR-5K (off-site). Top-K pooling consistently outperforms average and max pooling, reflecting the clinical intuition that diagnosis relies on a limited number of highly relevant regions, which enhances both performance and interpretability in multi-disease retinal analysis. Figure 2 shows the AUC comparison for each disease. Our method matches state-of-the-art performance on well-recognized conditions and demonstrates clear improvements on challenging diseases where prior models underperform. These results highlight the robustness and broad diagnostic capability of our framework in complex multi-disease scenarios.

3.4 Quantifying Gains from MD-PRSA

In multi-disease diagnosis, frequent co-occurrence of certain diseases can lead models to form spurious associations, erroneously attributing lesions from one disease to another. For example, if disease j often appears alongside disease i, the model may associate features of j with i, producing false positives for i when only j is present. To evaluate this, we adopt a *conditional evaluation setting*. For each disease pair (i, j), with i as the target and j as a frequently

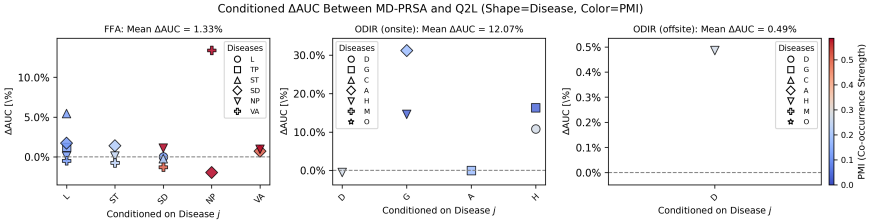

Fig. 3: Conditional ΔAUC across three datasets. Each subplot shows the change in AUC (%) for detecting disease i conditioned on the presence of co-occurring disease j (x-axis). Marker shapes indicate target diseases i, and color encodes the pointwise mutual information (PMI) [2] between each pair, reflecting co-occurrence strength. Only pairs with positive PMI are shown. The mean ΔAUC across these pairs is reported in each title. Higher values indicate that MD-PRSA better mitigates lesion-disease mismatching caused by co-occurrence.

Table 2: Ablation study

(a) Ablation of PR and SA

Dataset	w/o PR	w/o SA	Ours
LID-FFA	84.22	84.80	**85.53**
ODIR-5K (on-site)	87.45	87.82	**88.83**
ODIR-5K (off-site)	89.47	88.71	**89.72**

(b) Top-K analysis

K	5	10	20	50	100
LID-FFA	84.31	**85.53**	85.16	83.90	84.67
ODIR-5K (on-site)	86.26	86.17	87.43	**88.83**	87.16
ODIR-5K (off-site)	89.12	89.54	89.22	**89.72**	89.40

co-occurring disease, we compute the conditional AUC, $\mathrm{AUC}(i \mid j = 1)$, over test samples where $\mathrm{gt}(j) = 1$. This measures the model's ability to detect i under co-occurrence. We quantify improvement over a baseline (Q2L [14]) as:

$$\Delta\mathrm{AUC}(i \mid j = 1) = \mathrm{AUC}_{\mathrm{PRSA}}(i \mid j = 1) - \mathrm{AUC}_{\mathrm{Q2L}}(i \mid j = 1). \tag{5}$$

A positive $\Delta\mathrm{AUC}(i \mid j = 1)$ indicates more accurate detection of i despite the presence of j, reflecting reliance on disease-specific evidence rather than co-occurrence-induced shortcuts. As shown in Figure 3, MD-PRSA consistently improves across disease pairs, demonstrating that selective aggregation enforces disease-specific patch attribution, enhancing both accuracy and robustness in complex multi-disease scenarios.

3.5 Ablation Experiments

Effect of Patch-Level Reasoning and Selective Aggregation To isolate the effect of PR, we removed the reasoning module and applied SA directly on the initial patch features F_0, reshaped to $F_0 \in R^{HW \times L \times C}$. As shown in Table 2a, removing PR led to mAUC drops of 1.31% on LID-FFA and 1.38% on OIA-ODIR (on-site), confirming its importance for disease-specific representation learning. To evaluate SA, we replaced the Top-K aggregation with global average pooling

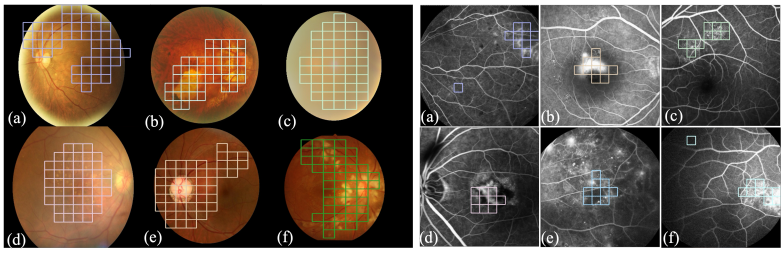

Fig. 4: ODIR-5K (left): Example key regions identified for (a) Diabetic Retinopathy (D), (b) Myopia (M), (c) Cataract (C), (d) Age-related Macular Degeneration (A), (e) Glaucoma (G), and (f) Hypertension (H). LID-FFA (right): Example key regions identified for (a) Leakage (L), (b) Transmission and Pooling (TP), (c) Staining (ST), (d) Shading (SD), (e) Non-Perfusion (NP), and (f) Vessel Abnormality (VA).

over $Q \in R^{HW \times L \times C}$ to obtain $Q^* \in R^{L \times C}$, followed by prediction. As reported in Table 2a, excluding SA reduced mAUC by 1.01% on ODIR-5K and 0.73% on LID-FFA, highlighting the benefit of selective aggregation for robust diagnosis.

Top-K in Selective Aggregation In the selective aggregation stage, the choice of Top-K patches plays a critical role in final diagnosis. We evaluate five configurations: $K = \{5, 10, 20, 50, 100\}$, with results shown in Table 2b. Performance varies with K and depends heavily on the dataset. This aligns with clinical intuition: in ODIR-5K (color fundus images), systemic diseases like diabetic retinopathy and glaucoma manifest across wide retinal regions, requiring larger K to capture distributed cues. In contrast, LID-FFA (fluorescein angiography) images target localized vascular abnormalities, where a smaller K better emphasizes key lesion areas. This suggests K should be tailored to the disease characteristics and imaging modality.

3.6 Visualization of Disease-Relevant Regions

Our results show that diagnosis based on selective aggregation achieves the best multi-disease performance. To ensure these regions correspond to clinically meaningful findings rather than artefacts, we visualized the key patches selected by our model for each disease (Figure 4). The visualizations highlight medically relevant areas, confirming the model's ability to capture important disease features and offering interpretable insights into its diagnostic decisions.

4 Discussion and Conclusion

In this work, we propose MD-PRSA, a novel framework for multi-disease retinal diagnosis that integrates patch-level disease reasoning with MIL-based selective

aggregation. By enforcing a spatial attribution constraint, MD-PRSA mitigates co-occurrence–induced lesion–disease mismatching and improves both diagnostic accuracy and interpretability. Extensive experiments on two retinal datasets with different imaging modalities demonstrate state-of-the-art performance. Future work will explore more expressive MIL aggregation strategies and disease-adaptive key region selection to further enhance performance.

Acknowledgements This work was partially supported by the UK Medical Research Council (MR/Z506175/1) and the UK Engineering and Physical Sciences Research Council (EP/X01441X/1).

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

References

1. Agu, N.N., Wu, J.T., Chao, H., Lourentzou, I., Sharma, A., Moradi, M., Yan, P., Hendler, J.: Anaxnet: anatomy aware multi-label finding classification in chest x-ray. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. pp. 804–813. Springer (2021)
2. Bouma, G.: Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL **30**, 31–40 (2009)
3. Chen, Z.M., Wei, X.S., Wang, P., Guo, Y.: Multi-label image recognition with graph convolutional networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5177–5186 (2019)
4. Contributors, M.: Openmmlab's pre-training toolbox and benchmark. https://github.com/open-mmlab/mmpretrain (2023)
5. Diaz-Pinto, A., Colomer, A., Naranjo, V., Morales, S., Xu, Y., Frangi, A.F.: Retinal image synthesis and semi-supervised learning for glaucoma assessment. IEEE transactions on medical imaging **38**(9), 2211–2218 (2019)
6. Gao, B.B., Zhou, H.Y.: Learning to discover multi-class attentional regions for multi-label image recognition. IEEE Transactions on Image Processing **30**, 5920–5932 (2021)
7. He, A., Li, T., Li, N., Wang, K., Fu, H.: Cabnet: Category attention block for imbalanced diabetic retinopathy grading. IEEE Transactions on Medical Imaging **40**(1), 143–153 (2021). https://doi.org/10.1109/TMI.2020.3023463
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
10. Ju, L., Yu, Z., Wang, L., Zhao, X., Wang, X., Bonnington, P., Ge, Z.: Hierarchical knowledge guided learning for real-world retinal disease recognition. IEEE Transactions on Medical Imaging (2023)

11. Li, N., Li, T., Hu, C., Wang, K., Kang, H.: A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In: Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Virtual Event, November 15–16, 2020, Revised Selected Papers 3. pp. 177–193. Springer (2021)
12. Li, X., Jia, M., Islam, M.T., Yu, L., Xing, L.: Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis. IEEE Transactions on Medical Imaging **39**(12), 4023–4033 (2020)
13. Lin, J., Cai, Q., Lin, M.: Multi-label classification of fundus images with graph convolutional network and self-supervised learning. IEEE Signal Processing Letters **28**, 454–458 (2021). https://doi.org/10.1109/LSP.2021.3057548
14. 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)
15. Maqsood, S., Damaševičius, R., Shah, F.M., Maskeliunas, R.: Detection of macula and recognition of aged-related macular degeneration in retinal fundus images. Computing and Informatics **40**(5), 957–987 (2021)
16. Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M., Zelnik-Manor, L.: Asymmetric loss for multi-label classification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 82–91 (2021)
17. Ridnik, T., Sharir, G., Ben-Cohen, A., Ben-Baruch, E., Noy, A.: Ml-decoder: Scalable and versatile classification head. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 32–41 (2023)
18. Serrano, S., Smith, N.A.: Is attention interpretable? arXiv preprint arXiv:1906.03731 (2019)
19. Smith, L.N., Topin, N.: Super-convergence: Very fast training of neural networks using large learning rates. In: Artificial intelligence and machine learning for multi-domain operations applications. vol. 11006, pp. 369–386. SPIE (2019)
20. Tang, P., Yan, X., Nan, Y., Xiang, S., Krammer, S., Lasser, T.: Fusionm4net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification. Medical Image Analysis **76**, 102307 (2022)
21. Xie, J., Chen, X., Zhao, Yitiannand Meng, Y., Zhao, H., Nguyen, A., Li, X., Zheng, Y.: Multi-disease detection in retinal images guided by disease causal estimation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 743–753 (2024)
22. Xie, J., Yi, Q., Wu, Y., Zheng, Y., Liu, Y., Macerollo, A., Fu, H., Xu, Y., Zhang, J., Behera, A., et al.: Deep segmentation of octa for evaluation and association of changes of retinal microvasculature with alzheimer's disease and mild cognitive impairment. British Journal of Ophthalmology (2023)
23. Ye, J., He, J., Peng, X., Wu, W., Qiao, Y.: Attention-driven dynamic graph convolutional network for multi-label image recognition. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16. pp. 649–665. Springer (2020)
24. Zhou, Y., Wang, B., Huang, L., Cui, S., Shao, L.: A benchmark for studying diabetic retinopathy: Segmentation, grading, and transferability. IEEE Transactions on Medical Imaging **40**(3), 818–828 (2021). https://doi.org/10.1109/TMI.2020.3037771