Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Visual Foundation Models with Adapter Learning for Parkinson's Disease Recognition based on OCT Images

Jianwei Zhang
Wenzhou Medical University
Tongyao Wang
Wenzhou Medical University
Dandan Shao
Wenzhou Medical University
Yuting Yang
Wenzhou Medical University
Jiasheng He
Wenzhou Medical University
Mingzhe Zheng
Wenzhou Medical University
Zhuoyan Yang
Wenzhou Medical University
Jianyong Wang
Wenzhou Medical University
Zi Jin
Wenzhou Medical University
Meixiao Shen
Wenzhou Medical University
Fan Lv
Wenzhou Medical University
Xiong Zhang
Wenzhou Medical University
Jiang-Fan Chen
Wenzhou Medical University
Tao Xu
xutao@wmu.edu.cn

Wenzhou Medical University

Research Article

Keywords: Parkinson's disease recognition, parameter-efficient finetuning, optical coherence tomography image, visual foundation models, adapter

Posted Date: August 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8533438/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Visual Foundation Models with Adapter Learning for Parkinson's Disease Recognition based on OCT Images

Jianwei Zhang[1†], Tongyao Wang[1†], Dandan Shao[1†], Yuting Yang[1], Jiasheng He[1], Mingzhe Zheng[1], Zhuoyan Yang[2], Jianyong Wang[2], Zi Jin[1,3], Meixiao Shen[1,3], Fan Lv[1,3], Xiong Zhang[2], Jiang-Fan Chen[1,3*], Tao Xu[1,3*]

[1]State Key Laboratory of Eye Health, Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China.
[2]Department of Neurology and Institute of Geriatric Neurology, the Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University, Wenzhou, Zhejiang, China.
[3]Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Eye Hospital, Wenzhou Medical University, Wenzhou, Zhejiang, China.

*Corresponding author(s). E-mail(s): xutao@wmu.edu.cn;
†These authors contributed equally to this work.

Abstract

Parkinson's disease (PD) is the leading neurodegenerative disorder in the world. Clinical studies have verified that the pathology changes in the retina structure can be potential biomarkers for PD diagnosis based on optical coherence tomography (OCT) images. Scholars have gradually applied deep neural networks (DNNs) to automatic PD recognition and have achieved good results. However, building an accurate DNN-based PD recognition model requires a large number of OCT images and expensive computational resources. Moreover, recent pre-trained visual foundation models (VFMs) have achieved promising success in a variety of visual recognition tasks with the aid of parameter-efficient finetuning techniques by tuning only a few trainable parameters, rather than traditional full finetuning techniques by updating all pre-trained parameters. Therefore, we transfer the pre-trained VFM with adapter learning paradigm (VFM-Adapter)

techniques to automatic PD recognition based on OCT images. To be specific, by freezing pretrained parameters of VFMs, we introduce three lightweight adapters: global adapter, local adapter, and token-wise adapter, to effectively leverage the transferability of feature representations from VFMs to OCT images for equipping VFMs with strong PD recognition capability. The results on the clinical OCT-PD dataset show that our proposed VFM-Adapter can achieve competitive or even better PD recognition performance through comparisons to state-of-the-art DNNs with substantially fewer tunable parameters.

Keywords: Parkinson's disease recognition, parameter-efficient finetuning, optical coherence tomography image, visual foundation models, adapter

1 Introduction

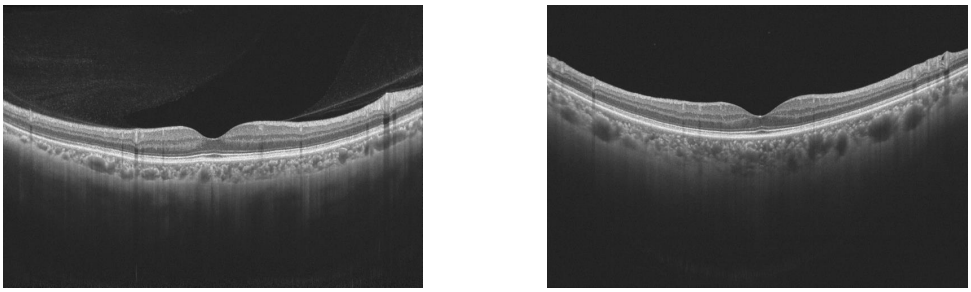
Parkinson's disease (PD) is one of the most common neurodegenerative disorders worldwide, and its clinical symptoms include motor and non-motor symptoms [1, 2]. Pathological changes of PD are mainly caused by progressive loss of dopaminergic neurons in the substantia nigra and the pathological accumulation of α-synuclein in cells [3]. As an age-related disorder, the incidence of PD is rapidly increasing with global population aging [4]. In the past years, several neuroimaging techniques have been widely employed to investigate dopaminergic degeneration and pathological progression, such as magnetic resonance imaging (MRI). However, MRI techniques have limitations for effectively monitoring extranigral pathologies in PD due to its long course. Additionally, MRI examination is expensive and time-consuming, which restricts the accessibility of patients. As a result, many patients with PD might miss the critical stage for effective intervention and treatment, imposing heavy burdens on individuals and society.

Clinical studies have found that PD-related pathology manifestations also existed on the eye's retina, providing an alternative means for PD diagnosis [5, 6], which has gained growing attention from academic and industrial communities. Recently, retinal imaging techniques have been widely applied to investigate PD's pathological hallmarks on the retina in a non-invasive manner, especially optical coherence tomography (OCT) imaging, as shown in Fig. 1. Ortuño-Lizarán et al. [7] found the degeneration of dopaminergic cells and their dendritic plexus of PD patients in the retina structure. Unlu et al. [8] investigated the significant difference of retinal layers between the PD group and healthy control (HC) group, such as inner nuclear layer, outer plexiform layer, retinal nerve fiber layer, inner plexiform layer, and outer nuclear layer. Murueta-Goyena et al. [9] found the association of retinal thickness with PD's progression of cognitive decline. Hasanov et al. [10] showed that pathologies of PD are related to the retina's functional and morphological changes. Tang et al. [11] found the relationship between OCT image features and PD. Jordan et al. [5] reported that morphological features and texture features of retinal layers extracted from OCT images can be adopted as the biomarkers for clinical PD diagnosis. Apart from clinical studies, Chen et al. [12] employed the Wavelet Transformer to segment retinal layers of PD patients

for analyzing retinal thickness easily. Richardson et al. [13] utilized a convolutional neural network (CNN) for automated PD recognition. Shi et al. [14] applied a retinal structure guidance-and-adaption network (RSGANet) to recognize PD. Huang et al. [15] proposed a wavelet-based selection-and-recalibration network (WaveSRNet) to screen PD. However, constructing a promising DNN-based PD recognition model often requires a number of OCT images and expensive computational overhead, presenting significant challenges for scholars.

Recently, scholars have achieved good progress by transferring vision foundation models (VFMs) to downstream tasks [16–20] due to their strong representation and generalization abilities, including object detection and video recognition. The classical full-finetuning technique is a widely-used paradigm for transferring VFMs to other learning tasks by updating all their pretrained parameters. However, it is difficult and expensive to fine-tune VFMs' all parameters effectively under limited computational resources and training data. To tackle this challenge, scholars have developed parameter-efficient finetuning (PEFT) techniques, which adapt VFMs to downstream tasks by freezing all their pre-trained parameters and only updating the negligible extra parameters. Existing PEFT techniques generally can be grouped into prompt learning and adapter learning. The former usually constructs task-related prompts as additional input information, allowing VFMs to produce expected outputs. Herzig et al. [21] applied a multi-task prompt learning method to tackle the video understanding task by freezing VFMs. Zhang et al. [22] developed the semantic-oriented visual prompt learning method for automated diabetic retinopathy grading. However, those prompt learning methods often require hand-crafted/professional engineering design and increase computational overhead as the number of visual prompts grows, restricting their generalization capability to different learning tasks. On the contrary, adapter learning infuses a few lightweight adapters into VFMs' specific layers, to better mine the helpful feature representations from pretrained VFMs by finetuning a few parameters. Jie et al. [23] introduced a convolution-based adapter for visual tasks. Zhao et al. [24] applied a token pyramid pooling-driven style adapter to transfer the pre-trained feature representation distributions to fundus image feature representation distributions. Moreover, considering that it is difficult to collect massive OCT images of SM and access expensive computational resources, it is impossible to train VFMs from scratch. We also found that previous works have not attempted to apply pretrained VFMs to recognize PD automatically from OCT images.

Motivated by the above systematic analysis, this paper aims to adapt the pre-trained VFM for automated PD recognition by investigating the knowledge transferability of VFM to OCT images with the adapter learning paradigm. By freezing pretrained parameters of VFMs and inserting a few lightweight adapters during finetuning with the aid of PEFT techniques, termed VFM-Adapter. To be specific, the VFM-Adapter comprises three different main adapters: a local adapter, a global adapter, and a token-wise adapter. Particularly, the local adapter and the global adapter are designed to capture subtle and salient feature representations, respectively. The token-wise adapter is devised to adjust the token-wise value distributions for better mining task-specific knowledge from VFMs. The experimental results also



(a) (b)

Fig. 1 Two representative OCT images of healthy (a) and Parkinson's disease (PD) (b).

show that our proposed VFM-Adapter obtains competitive or even better PD recognition performance than existing state-of-the-art DNNs with fewer tunable parameters. In general, the main contributions of this paper are summarized as follows:

- We transfer pretrained VFM to automated PD recognition with the aid of PEFT techniques based on OCT images named VFM-Adapter, which fully adapts task-specific representations from VFM to OCT images.
- In the proposed VFM-Adapter, we propose a local adapter and a global adapter to capture PD's significant silent and subtle token representations. Additionally, a token-wise adapter is developed to dynamically adjust the token-wise value distributions for better learning PD-specific knowledge.
- Extensive experiments on the clinical OCT-PD dataset demonstrate that our proposed method achieves competitive or even better PD recognition performance than existing state-of-the-art DNNs with fewer tunable parameters. Further analysis is applied to improve the decision-making interpretability process of our method.

2 Materials and method

2.1 Materials

To validate the effectiveness of our method, we constructed an OCT image dataset named OCT-PD, which was collected from the Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University between October 2022 and June 2024. The study was approved by the Ethics Committees of the School of Optometry and Ophthalmology and the Eye Hospital of Wenzhou Medical University (No. 2023-130) and was conducted in accordance with the Declaration of Helsinki. The OCT-PD dataset contains 1,440 OCT images from 80 cases, including 720 images from 40 healthy controls and 720 images from 40 PD patients, with each case comprising 18 OCT images. All cases were independently diagnosed by three clinical experts with at least five years of experience, ensuring high-quality annotations. Fig. 1 shows representative OCT images of a healthy control case and a PD case. For model evaluation, five-fold cross-validation was performed at the case level, with four subsets used for training and one subset used for validation in each fold.

2.2 Method

2.2.1 General framework

Fig. 2 presents the general overview of the VFM-Adapter framework for automated PD recognition based on OCT images, by fully leveraging the feature representations transferability from pre-trained VFMs to OCT images with the help of PEFT techniques. In the proposed VFM-Adaption framework, we develop three different adapters: global adapter, local adapter, and token-wise adapter, as offered in Fig. 2(b)-(e). The global adapter is utilized to capture salient feature representations, and the local adapter is applied to learn small feature representations. Moreover, we employ the token-wise adapter to modify the token-wise value distributions for better capturing task-specific feature representations from pre-trained VFMs. As offered in Fig. 2(a), given the OCT images, this paper applies patching and embedding operations to produce token maps via a convolutional layer with kernel size 3×3; then, a series of pre-trained Transformer blocks with three adapters are applied to capture significant token representations; finally, a global average pooling (GAP) operator and a softmax classifier are applied to produce the PD recognition results. Additionally, this paper utilizes the classical cross-entropy loss L_{CE} to compute the predicted PD recognition errors of VFM-Adapter, which is written as follows:

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N} y_{ij} \log p_{ij}, \tag{1}$$

where $y_{ij} \in \{0, 1\}$, p_{ij}, and N denote the ground truth, predicted labels, the number of training samples in each mini-batch.

2.2.2 Global adapter

Fig. 2(b) offers a simple implementation of the global adapter, which consists of two fully-connected (FC) layers and a GELU activation function. Previous works have shown that this design mode can guide DNNs to effectively model long-range dependency among tokens, which also helps capture salient token representations. Here, we formulate the global adapter as follows:

$$F = FC(GELU(FC(X_1))), \tag{2}$$

where $X_1 \in R^{B\times C\times T}$ is the intermediate tokens and $F \in R^{B\times C\times T}$ indicates the augmented sequence token (B, C and T denote batch size, channel number, and token maps).

2.2.3 Local adapter

To learn small lesions effectively, we develop a local adapter with the aid of the convolution operator in a lightweight yet effective manner, as shown in Fig. 2(c). The inherent motivation behind it is that the convolution operator with a small kernel size is skilled at capturing small feature representations well, which has been proved

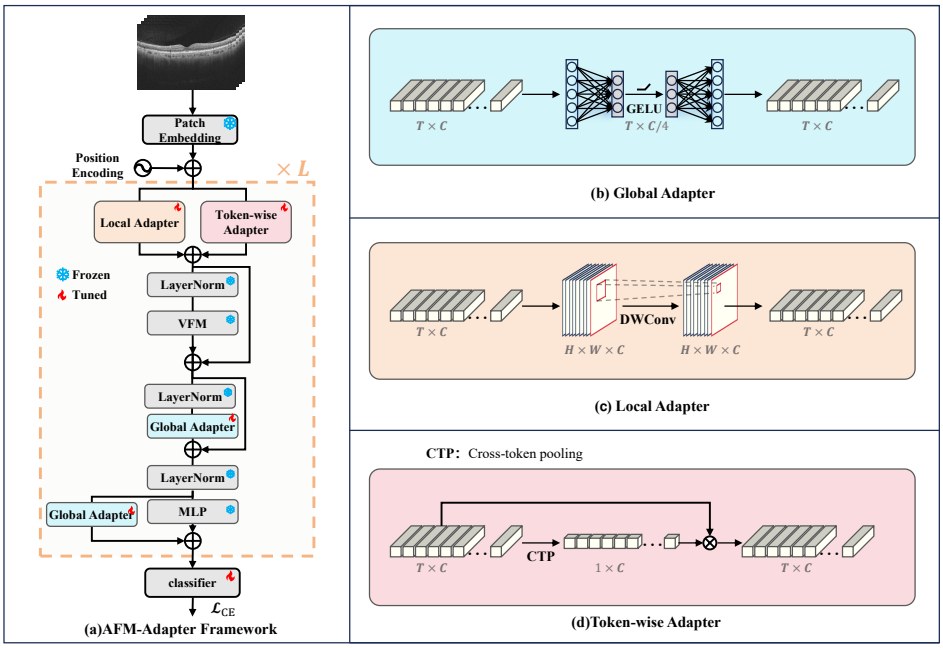


Fig. 2 The general overview of the VFM-Adapter framework for automated PD recognition (a) from OCT images. We employ local adapter, global adapter, and token-wise adapter to capture informative feature presentations and adjust the token-wise representation distributions.

by previous works. Specifically, given the input token sequence maps $X_2 \in R^{B\times C\times T}$, we first conduct a sequence-to-image $Seq2Img$ operation to facilitate subtle lesion representation capturing by conversing the 1D token sequence into 2D feature maps $I \in R^{B\times C\times H\times W}$ (H and W denote the feature map resolution), which is formulated as:

$$I = Seq2Img(X_2). \tag{3}$$

Next, we apply a depth-wise convolution ($DWConv$) operator with 3×3 kernel size to capture small lesions and enhance token representation diversities in a lightweight manner, and then utilize an image-to-sequence $Img2Seq$ operation to restore the original shape of input token sequence maps:

$$I^{'} = DWConv_{3\times3}(I), Z = Img2Seq(I'), \tag{4}$$

when $I^{'}$ and Z denote the learned 2D feature maps and encode token sequence maps.

2.2.4 Token-wise adapter

To better leverage the task-specific knowledge from VFMs, this paper proposed a token-wise adapter (TW-adapter) to dynamically adjust the token-wise representation distributions, which comprises two operations: cross-token pooling and token-wise adaption, as presented in Fig. 2(d). The cross-token pooling operation is applied

to aggregate token-wise statistics from different token maps, which is formulated as follows:

$$\mu_{b,c,i} = \frac{1}{T}\sum_{t=1}^{T} x_{b,c,t}, \tag{5}$$

where $\mu(b, c, i)$ denotes the token-wise average statistics context of the intermediate token $x_{b,c,i} \in R^{1\times T}$. Based on Eq. 5, we can obtain token-wise average statistics context map $P \in R^{B\times C\times 1}$

Next, we employ the token-wise adaption operation to convert token-wise average statistics into token-wise weights via token-independent weighting strategy, so as to better change the the token-wise representation distributions. Here, we formulate the token-wise adaption operator as follows:

$$Z = P \odot BN(F). \tag{6}$$

3 Experiments and result analysis

3.1 Evaluation Metrics

We utilize six commonly accepted evaluation metrics to assess the general performance of all methods: accuracy (ACC), sensitivity (Sen), precision (Pr), Specificity (Spe), F1, and area under the receiver operating characteristic curve (AUC)

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{7}$$

$$Sen = \frac{TP}{TP + FN}, \tag{8}$$

$$Pr = \frac{TP}{TP + FP}, \tag{9}$$

$$Spe = \frac{TN}{TN + FP}, \tag{10}$$

$$F1 = \frac{2 * Pr * Sen}{PR + Sen}, \tag{11}$$

where TP, FP, TN and FN denote the numbers of true positives, false positives, true negatives and false negatives, respectively.

3.2 Experimental settings

This paper implements our proposed method, competitive DNNs based on Pytorch. We utilize the stochastic gradient descent (SGD) method as the optimizer to update the parameters of pre-trained DNNs with default settings in the training. The key motivations to adopt this strategy for automated OCT-based SM recognition are as follows: previous works have verified the effectiveness of employing pre-trained DNNs to medical image analysis under the medical datasets with the aid of transfer learning

Table 1 PD recognition performance comparison of full finetuning and PEFT methods under different VFMs (Adp denotes adapter).

Method	Backbone	ACC(%)	Sen(%)	F1(%)	Spe(%)	Param(M)	Tun.(M)	FLOPs(G)
Full FT	ViT-B	69.86±9.80	76.94±13.21	71.74±9.52	62.78±17.00	85.65	85.65	11.21
	ViT-B+Adp	69.72±6.05	74.58±10.69	70.88±9.46	64.86±16.27	96.43	96.43	13.33
	Swin-T	72.01±5.28	65.83±7.38	70.07±5.72	78.19±8.02	86.77	86.77	15.17
	Swin-T+Adp	76.25±9.63	74.17±11.26	75.76±9.55	78.33±15.92	93.95	93.95	16.47
PEFT	ViT-B	60.56±5.75	86.53±5.21	71.33±12.67	34.58±10.19	85.65	<0.01	11.21
	ViT-B+Adp	68.82±6.18	86.81±6.13	79.74±8.79	50.83±12.58	96.43	10.78	13.33
	Swin-T	54.17±3.28	49.86±10.08	54.13±3.90	58.47±7.85	86.77	<0.01	15.17
	Swin-T+Adp	77.66±5.52	78.01±4.62	77.93±3.98	77.31±14.25	93.95	7.18	16.47

techniques. We set batch size, training epochs, and initial learning rate to 32, 100, and 0.0035. Here, we decrease the learning rate by a factor of 5 every 30 epochs. For data augmentation, we use commonly-used random flipping and rotating methods for augmenting training data, and the input image size is resized into 224 ×224. All experiments are run on a server with one TITAN V GPU.

3.3 Result analysis and discussion

3.3.1 PD recognition performance comparison of full finetuning and parameter-efficient finetuning techniques under different VFMs

Table 1 offers PD recognition results of full finetuning and parameter-efficient finetuning techniques across two different VFMs (Swin-T and ViT-B) based on the OCT-PD dataset. It can be seen that our VFM-Adapter consistently outperforms Swin-T across two finetuning techniques. Based on the full finetuning technique, VFM-Adapter significantly outperforms Swin-T by 4.24% in accuracy and 8.34% in sensitivity. Our proposed VFM-Adapter based on parameter-efficient finetuning technique with 7.18M tunable parameter number is far less than it based on full finetuning technique with 93.95M tunable parameter number. VFM-Adapter with parameter-efficient finetuning technique even achieves better PD recognition performance (77.66% of accuracy and 78.01% of sensitivity) than it with full finetuning technique (76.25% of accuracy and 74.17% of sensitivity). As for taking ViT-B as the VFM backbone, VFM-Adapter performs worse than ViT-B under full finetuning technique, but VFM-Adapter outperforms ViT-B based on parameter-efficient finetuning technique. The possible reason to explain the results is that our proposed adapters can not exploit task-specific knowledge from ViT-B under limited OCT images by adopting the full finetuning technique. Overall, the results demonstrate the effectiveness of VFM-Adapter in transferring the feature representations from VFMs to OCT-based PD recognition based on the parameter-efficient finetuning technique, aligning with our expectations.

Table 2 Performance comparisons of our VFM-Adapter and state-of-the-art deep neural networks based on the OCT-PD dataset.

Method	ACC(%)	Spe(%)	Pr(%)	Sen(%)	F1(%)
ResNet	74.65±7.24	69.86±15.30	73.73±8.54	79.44±14.25	75.48±7.85
SENet [25]	76.11±6.29	85.00±4.18	81.76±2.55	67.22±15.01	72.87±9.56
ECANet [26]	74.38±5.26	78.75±10.62	77.57±5.91	70.00±13.63	72.62±7.55
SKNet [27]	76.32±5.23	78.47±10.07	78.35±6.22	74.17±14.44	75.18±7.48
RIRNet [28]	75.62±4.63	76.81±12.38	77.42±5.55	74.44±16.00	74.63±6.79
GCANet [29]	69.37±4.50	69.17±10.51	69.95±6.39	69.58±10.86	69.15±5.76
PCANet [30]	75.00±7.40	73.33±17.05	76.35±12.20	76.67±9.72	75.57±6.71
EfficientNet	76.32±5.92	69.58±11.78	73.85±5.75	83.06±13.14	77.43±6.92
ConvNeXt	68.40±5.40	75.42±2.04	71.03±3.26	61.39±11.19	65.55±7.57
ViT [31]	69.86±9.80	62.78±17.00	68.43±10.38	76.94±13.21	71.74±9.52
Swin-T [32]	72.01±5.28	78.19±8.02	75.55±6.25	65.83±7.38	70.07±5.72
ResMLP [33]	65.14±6.84	77.64±11.63	70.95±10.48	52.64±14.12	59.32±10.85
MedMamba [34]	76.74±5.58	73.06±3.62	74.76±3.05	80.42±12.19	77.22±6.88
Full FT	76.25±9.63	78.33±15.92	78.54±11.65	75.76±9.55	72.92±9.04
Linear-Head	54.17±3.67	49.86±11.27	54.13±4.36	58.47±8.78	55.86±4.47
VFM-Adapter	77.66±5.52	77.31±14.25	79.58±11.49	78.01±4.62	77.93±3.98

3.3.2 Comparisons to state-of-the-art deep neural networks

Table 2 presents PD recognition performance comparisons of our VFM-Adapter and state-of-the-art deep neural networks the OCT-PD dataset. These deep neural networks include classical CNNs, Transformers, ResMLP, Mamba and VFMs. CNNs used for comparison as follows: squeeze-and-excitation network (SENet) [25], efficient-channel network (ECANet) [26], selective kernel network (SKNet) [27], region-based integration-and-recalibration network (RIRNet) [28], gated channel attention network (GCANet) [29], pyramid pixel context adaption network (PCANet) [30].We summarize as follows: 1) Compared to SKNet, RIRNet, PCANet, VFM-Adapter obtains absolute accuracy gains of 1.34%, 2.04%, and 2.66%, respectively. 2) VFM-Adapter achieves competitive PD recognition performance compared to recent vision Transformers, with accuracy improvements of 7.80% over ViT, 5.65% over Swin-T, and 12.52% over ResMLP. 3) Compared to finetuning methods, VFM-Adapter shows consistent improvements with an absolute accuracy gain of 1.41% over Full FT and 23.49% over Linear-Head.

3.3.3 Effects of Different Adapters in VFM-Adapter

In this section, we conduct ablation experiments to investigate the effects of three adapters in VFM-Adapter, as listed in Table 3. Here, we take the Swin-T as the VFM backbone. We observe that each adapter brings PD recognition performance, showing they have unique roles. Their combination further enhances the PD recognition results,

Table 3 Performance comparisons of different adapters in VFM-Adapter based on the OCT-PD dataset

Method	ACC	Spe	Sen	F1
VFM	61.88±6.27	49.86±27.79	73.89±18.21	65.48±5.20
global adapter	73.75±6.60	72.92±12.30	74.58±8.08	74.00±6.27
local adapter	74.17±8.61	70.14±12.45	78.19±11.12	75.11±8.07
TW-adapter	72.29±8.82	68.06±18.19	76.53±8.71	73.63±6.81
global adapter+TW-adapter	73.47±7.60	77.50±15.63	69.44±12.53	72.20±8.17
global adapter+local adapter	77.22±5.31	71.81±12.32	82.64±10.29	78.31±5.09
local adapter+TW-adapter	64.93±7.40	57.92±28.00	71.94±16.90	66.96±4.89
VFM-Adapter	77.66±5.52	77.31±14.25	78.01±4.62	77.93±3.98

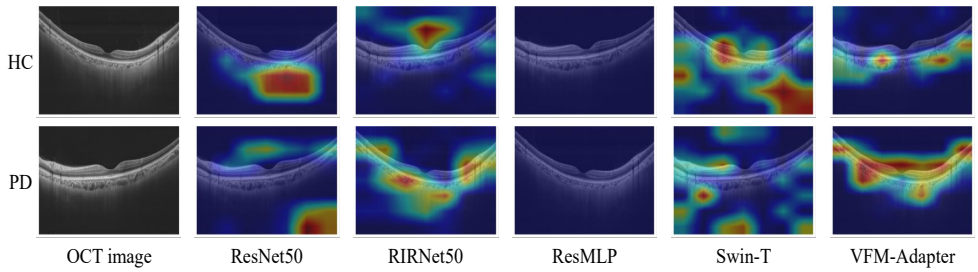


Fig. 3 Grad-CAM visualizations of VFM-Adapter and other five representative deep neural networks (ResNet50, RIRNet50, Swin-T, ResMLP, and VFM) on two representative OCT images of HC and PD.

indicating our method is capable of fully leveraging the feature representations from VFMs to OCT-based PD recognition.

3.3.4 Visualization analysis

Fig. 3 shows the heatmap visualization comparisons of our VFM-Adapter and other five representative deep neural networks (ResNet50, RIRNet50, Swin-T, ResMLP, and VFM) via Grad-CAM techniques to illustrate how these deep neural networks focus on informative lesions of PD from OCT images. It can be seen that our VFM-Adapter pays more attention to salient and subtle lesions through comparisons to other comparable deep neural networks with the help of three proposed adapters.

4 Conclusion

We develop a VFM-Adapter to recognize PD automatically from OCT images with the aid of the PEFT technique. To effectively utilize the representational capability of VFM, we design a global adapter and a local adapter to highlight salient and subtle feature representations, respectively. A token-wise adapter is employed to modify token-wise value distributions, enabling better transfer of task-specific knowledge from VFMs to OCT images. The comprehensive experiments on the OCT-PD dataset manifest the superiority of VFM-Adapter over advanced DNNs and other state-of-the-art

PEFT methods. In the future, we plan to investigate the generalization of our method on the OCT image dataset of PD from other medical centers.

Acknowledgements

Not applicable.

Authors' Contributions

J.W.Z.: Formal analysis, Investigation, Methodology, Writing – original draft. T.Y.W.: Data curation, Formal analysis, Investigation, Methodology, Software, Visualization. D.D.S.: Data curation, Investigation, Methodology, Validation. Y.T.Y.: Data curation, Investigation, Validation. J.S.H.: Investigation, Validation. M.Z.Z.: Investigation, Visualization. Z.Y.Y.: Investigation, Visualization. J.Y.W.: Investigation. Z.J.: Formal analysis, Software. M.X.S.: Formal analysis, Software. F.L.: Project administration, Resources. X.Z.: Resources. J.-F.C.: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. T.X.: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. All authors reviewed the manuscript.

Funding

The Science & Technology Initiative projects, the STI 2030-Major Project (2021ZD0203400); The Key Research and Development Program of Zhejiang Province (2023C03079); The Start-up Fund from the Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health, OJQDSP2022007); The Program Project from the State Key Laboratory of Eye Health, Wenzhou Medical University (J01-20190101); The Zhejiang Provincial Natural Science Foundation of China (LMS26C090001); The Wenzhou Municipal Science and Technology Bureau (GY20250239).

Author Information

Jianwei Zhang, Tongyao Wang and Dandan Shao contributed equally to this work as co-first authors.

Availability of Data and Materials

This paper contains all the data.

Ethics Approval

This study was approved by the Ethics Committees of the School of Optometry and Ophthalmology and the Eye Hospital of Wenzhou Medical University (Approval No. 2023-130) and conducted in accordance with the ethical principles of the Declaration

of Helsinki. Written informed consent was obtained from all participants prior to enrollment.

Clinical Trial Number

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

References

[1] Thakur, M., Dhanalakshmi, S., Kuresan, H., Senthil, R., Narayanamoorthi, R., Lai, K.W.: Automated restricted boltzmann machine classifier for early diagnosis of parkinson's disease using digitized spiral drawings. Journal of Ambient Intelligence and Humanized Computing **14**(1), 175–189 (2023)

[2] Tang, C., Zhang, X., Miao, H., Wang, X., Liu, J.: Retinal imaging techniques for neurodegenerative diseases: Parkinson's disease and beyond. MedComm–Future Medicine **4**(3), 70034 (2025)

[3] Lee, J.-Y., Martin-Bastida, A., Murueta-Goyena, A., Gabilondo, I., Cuenca, N., Piccini, P., Jeon, B.: Multimodal brain and retinal imaging of dopaminergic degeneration in parkinson disease. Nature Reviews Neurology **18**(4), 203–220 (2022)

[4] Soni, R., Mathur, K., Shah, J.: An update on new-age potential biomarkers for parkinson's disease. Ageing Research Reviews, 102208 (2024)

[5] Alves, J.N., Westner, B.U., Højlund, A., Weil, R.S., Dalal, S.S.: Structural and functional changes in the retina in parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry **94**(6), 448–456 (2023)

[6] Xie, X., Chen, b., Zhao, Y., Wang, H., Xu, S., Hu, Y., Rong, H., Liu, J.: Initiative of eye-brain joint computing in neuropsychiatric disorders. Chinese Journal of Psychiatry **53**(6), 546–552 (2020)

[7] Ortuño-Lizarán, I., Sánchez-Sáez, X., Lax, P., Serrano, G.E., Beach, T.G., Adler, C.H., Cuenca, N.: Dopaminergic retinal cell loss and visual dysfunction in parkinson disease. Annals of neurology **88**(5), 893–906 (2020)

[8] Unlu, M., Gulmez Sevim, D., Gultekin, M., Karaca, C.: Correlations among multifocal electroretinography and optical coherence tomography findings in patients with parkinson's disease. Neurological Sciences **39**, 533–541 (2018)

[9] Murueta-Goyena, A., Romero-Bascones, D., Teijeira-Portas, S., Urcola, J.A., Ruiz-Martínez, J., Del Pino, R., Acera, M., Petzold, A., Wagner, S.K., Keane, P.A., *et al.*: Association of retinal neurodegeneration with the progression of cognitive decline in parkinson's disease. npj Parkinson's Disease **10**(1), 26 (2024)

[10] Hasanov, S., Demirkilinc Biler, E., Acarer, A., Akkın, C., Colakoglu, Z., Uretmen, O.: Functional and morphological assessment of ocular structures and follow-up of patients with early-stage parkinson's disease. International ophthalmology **39**, 1255–1262 (2019)

[11] Tang, C., Zhang, Z., Xu, T., Jin, Z., Shen, M., Zhang, X., Lu, F., Chen, J., Wang, X., Zhang, X., Liu, J.: Quantitative evaluation of image features in the retinal outer plexiform layer of parkinson's disease by oct. Chinese Journal of Optometry Ophthalmology and Visual Science **27**(4), 249–254 (2025)

[12] Chen, Y., Zhang, X., Wang, T., Tang, C., Ye, H., Liu, J.: Waveformer: A wavelet transformer for parkinson disease's retinal layer segmentation in oct. In: 2024 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2024). IEEE

[13] Richardson, A., Kundu, A., Henao, R., Lee, T., Scott, B.L., Grewal, D.S., Fekrat, S.: Multimodal retinal imaging classification for parkinson's disease using a convolutional neural network. Translational Vision Science & Technology **13**(8), 23–23 (2024)

[14] Shi, H., Wei, J., Jin, R., Peng, J., Wang, X., Hu, Y., Zhang, X., Liu, J.: Retinal structure guidance-and-adaption network for early parkinson's disease recognition based on oct images. Computerized Medical Imaging and Graphics, 102463 (2024)

[15] Huang, J., Zhang, X., Jin, R., Xu, T., Jin, Z., Shen, M., Lv, F., Chen, J., Liu, J.: Wavelet-based selection-and-recalibration network for parkinson's disease screening in oct images. Computer Methods and Programs in Biomedicine **256**, 108368 (2024)

[16] Zhang, Y., Ma, X., Huang, K., Li, M., Heng, P.-A.: Semantic-oriented visual prompt learning for diabetic retinopathy grading on fundus images. IEEE Transactions on Medical Imaging **43**(8), 2960–2969 (2024)

[17] Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, J., Luo, P.: Adaptformer: Adapting vision transformers for scalable visual recognition. Advances in Neural Information Processing Systems **35**, 16664–16678 (2022)

[18] Zhang, X., Zhao, J., Li, Y., Wu, H., Zhou, X., Liu, J.: Efficient pyramid channel attention network for pathological myopia recognition with pretraining-and-finetuning. Artificial Intelligence in Medicine **154**, 102926 (2024)

[19] Zhao, J., Zhang, X., Zhang, J., Sun, H., Nie, Q., Xiao, Z., Xiao, L., Zhang, F., Hu, Y., Liu, J.: Token pyramid pooling-driven style adapter learning with dual-view balanced loss for imbalanced diabetic retinopathy grading. Pattern Recognition, 112194 (2025)

[20] Jiang, H., Gao, M., Liu, Z., Tang, C., Zhang, X., Jiang, S., Yuan, W., Liu, J.: Glanceseg: Real-time microaneurysm lesion segmentation with gaze-map-guided foundation model for early detection of diabetic retinopathy. IEEE Journal of Biomedical and Health Informatics, 1–14 (2024)

[21] Herzig, R., Abramovich, O., Ben Avraham, E., Arbelle, A., Karlinsky, L., Shamir, A., Darrell, T., Globerson, A.: Promptonomyvit: Multi-task prompt learning improves video transformers using synthetic scene data. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 6803–6815 (2024)

[22] Zhang, Y., Ma, X., Huang, K., Li, M., Heng, P.-A.: Semantic-oriented visual prompt learning for diabetic retinopathy grading on fundus images. IEEE Transactions on Medical Imaging (2024)

[23] Jie, S., Deng, Z.-H., Chen, S., Jin, Z.: Convolutional bypasses are better vision transformer adapters. In: ECAI 2024, pp. 202–209. IOS Press, ??? (2024)

[24] Zhao, J., Zhang, X., Zhang, J., Sun, H., Nie, Q., Xiao, Z., Xiao, L., Zhang, F., Hu, Y., Liu, J.: Token pyramid pooling-driven style adapter learning with dual-view balanced loss for imbalanced diabetic retinopathy grading. Pattern Recognition, 112194 (2025)

[25] Jie, Hu, Li, Shen, Samuel, Albanie, Gang, Sun, Enhua, Wu: Squeeze-and-excitation networks. TPAMI (2019)

[26] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11531–11539 (2020). https://doi.org/10.1109/CVPR42600.2020.01155

[27] Li, X., Wang, W., Hu, X., Yang, J.: Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 510–519 (2019)

[28] Zhang, X., Xiao, Z., Fu, H., Hu, Y., Yuan, J., Xu, Y., Higashita, R., Liu, J.: Attention to region: Region-based integration-and-recalibration networks for nuclear cataract classification using as-oct images. Medical Image Analysis **80**, 102499

(2022)

[29] Xiao, Z., Zhang, X., Higashita, R., Hu, Y., Yuan, J., Chen, W., Liu, J.: Gated channel attention network for cataract classification on as-oct image. In: Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28, pp. 357–368 (2021). Springer

[30] Zhang, X., Xiao, Z., Wu, X., Chen, Y., Zhao, J., Hu, Y., Liu, J.: Pyramid pixel context adaption network for medical image classification with supervised contrastive learning. IEEE Transactions on Neural Networks and Learning Systems **36**(4), 6802–6815 (2025)

[31] Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[32] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

[33] Touvron, H., Bojanowski, P., Caron, M., Cord, M., El-Nouby, A., Grave, E., Izacard, G., Joulin, A., Synnaeve, G., Verbeek, J., et al.: Resmlp: Feedforward networks for image classification with data-efficient training. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)

[34] Yue, Y., Li, Z.: Medmamba: Vision mamba for medical image classification. arXiv preprint arXiv:2403.03849 (2024)