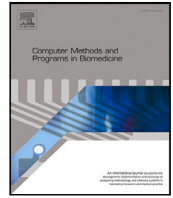




Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: <https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine>

Wavelet-based selection-and-recalibration network for Parkinson's disease screening in OCT images

Jingqi Huang^{a,1}, Xiaoqing Zhang^{b,a,1}, Richu Jin^a, Tao Xu^{c,d}, Zi Jin^{e,f}, Meixiao Shen^{e,f}, Fan Lv^{d,e,f}, Jiangfan Chen^{c,d}, Jiang Liu^{a,c,g,*}^a Research Institute of Trustworthy Autonomous Systems and Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China^b Center for High Performance Computing and Shenzhen Key Laboratory of Intelligent Bioinformatics, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China^c The State Key Laboratory of Ophthalmology, Optometry and Vision Science, Wenzhou Medical University, Wenzhou, Zhejiang, China^d The Ouyang Laboratory; The Affiliated Eye Hospital, Wenzhou Medical University, 270 Xueyuan Road, Wenzhou, Zhejiang, China^e National Engineering Research Center of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, China^f National Clinical Research Center for Ocular Diseases, Eye Hospital, Wenzhou Medical University, Wenzhou, China^g Singapore Eye Research Institute, 169856, Singapore

ARTICLE INFO

Dataset link: <https://github.com/Jingqi-H/Wav-eSRNet>

Keywords:

Deep learning

Parkinson's disease screening

Discrete wavelet transform

ABSTRACT

Background and Objective: Parkinson's disease (PD) is one of the most prevalent neurodegenerative brain diseases worldwide. Therefore, accurate PD screening is crucial for early clinical intervention and treatment. Recent clinical research indicates that changes in pathology, such as the texture and thickness of the retinal layers, can serve as biomarkers for clinical PD diagnosis based on optical coherence tomography (OCT) images. However, the pathological manifestations of PD in the retinal layers are subtle compared to the more salient lesions associated with retinal diseases.

Methods: Inspired by textural edge feature extraction in frequency domain learning, we aim to explore a potential approach to enhance the distinction between the feature distributions in retinal layers of PD cases and healthy controls. In this paper, we introduce a simple yet novel wavelet-based selection and recalibration module to effectively enhance the feature representations of the deep neural network by aggregating the unique clinical properties, such as the retinal layers in each frequency band. We combine this module with the residual block to form a deep network named Wavelet-based Selection and Recalibration Network (WaveSRNet) for automatic PD screening.

Results: The extensive experiments on a clinical PD-OCT dataset and two publicly available datasets demonstrate that our approach outperforms state-of-the-art methods. Visualization analysis and ablation studies are conducted to enhance the explainability of WaveSRNet in the decision-making process.

Conclusions: Our results suggest the potential role of the retina as an assessment tool for PD. Visual analysis shows that PD-related elements include not only certain retinal layers but also the location of the fovea in OCT images.

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by prominent motor dysfunction and nonmotor symptoms. With an aging global population, the number of PD patients is increasing rapidly, making it one of the leading causes of disability [1]. Currently, neuroimaging is incorporated into diagnostic criteria and is used to aid early and differential diagnosis [2]. However, these techniques are primarily employed to observe structural lesions in advanced

PD, such as subdural hemorrhage through magnetic resonance imaging, or the nigrostriatal dopaminergic system through positron emission computerized tomography. Additionally, neuroimaging examinations are time-consuming and costly.

Recent studies [3] suggest that the eye is an extension of the brain, making it reasonable to search for ocular manifestations of brain-related diseases. Some works have focused on the potential for

* Corresponding author.

E-mail address: liuj@sustech.edu.cn (J. Liu).¹ Jingqi Huang and Xiaoqing Zhang contributed equally to this work.<https://doi.org/10.1016/j.cmpb.2024.108368>

Received 30 January 2024; Received in revised form 30 July 2024; Accepted 7 August 2024

Available online 12 August 2024

0169-2607/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

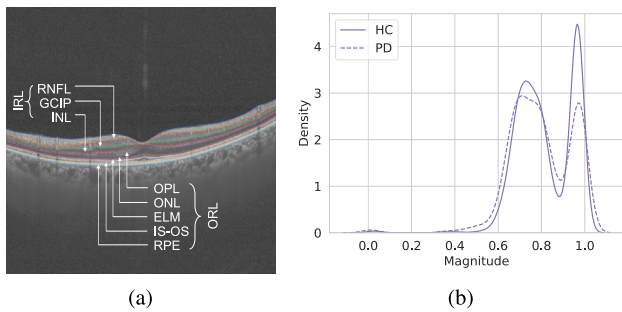


Fig. 1. (a) A B-scan with nine manually delineated retinal surfaces separating the following inner retinal layers (IRL) and outer retinal layers (ORL): the retinal nerve fiber layer (RNFL); the ganglion cell layer combined with the inner plexiform layer (GCIP); the inner nuclear layer (INL); the outer plexiform layer (OPL); the outer nuclear layer (ONL); the external limiting membrane (ELM); the inner segment (IS); the outer segment (OS); and the retinal pigment epithelium (RPE). Finally, above the RNFL is the vitreous, and below the RPE is the choroid. (b) Kernel density estimation feature plot for different categories: health control (HC) and PD among the whole OCT data.

screening brain-related diseases (e.g., Alzheimer's disease [4] and brain frailty [5]) using deep learning (DL) from retinal images. Additionally, retinal changes in PD have been shown in post-mortem histopathological studies [6]. This concept is supported by clinical studies indicating that the characteristic pathological changes in the ocular features of PD patients occur in the retinal layers [7]. These structural features, regarded as clinical prior, can be non-invasively imaged by optical coherence tomography (OCT), making OCT a potential imaging modality for detecting early PD. In this paper, we aim to build a computer-aided diagnosis (CAD) tool based on OCT images for automatic PD screening. This tool will assist clinicians in diagnosing and further validate the link between retinal layers and brain disease PD using computational methods.

Unlike retinal diseases with prominent lesions, such as fluid accumulation in diabetic macular edema [8], the unique pathology of PD patients in OCT is subtle changes in the retinal layers that are more complex and challenging to detect directly through DL approaches. For instance, compared to health control (HC), PD patients often exhibit focal thickening in the retinal nerve fiber layer (RNFL) [9], ganglion cell and inner plexiform layer (GCIP) [10], outer plexiform layer (OPL) [11] or outer nuclear layer (ONL) [12] (as shown in Fig. 1(a)). However, Fig. 1(b) shows that the distributions between HC and PD images are similar, indicating that the retinal changes in OCT images are not readily discernible. This presents a challenge for networks to learn discriminative features for PD screening.

Inspired by the frequency domain learning in textural edge feature extraction [13], we apply Discrete Wavelet Transform (DWT) to decompose 2D OCT data into an approximation and three detail coefficients that represent the low- and high-frequency components separately. We hypothesize that features from different sub-bands provide complementary and discriminative information. Fig. 2(a) suggests that the approximation coefficient is similar to the image, capturing the main structural features, while the remaining three coefficients emphasize edges and contours in the OCT image. Regarding Fig. 2(b), it could be seen that the relative difference in the density feature distribution for each details coefficient in the frequency space is more obvious compared to the approximation coefficient. Consequently, the retinal structures of RNFL, GCIP, OPL, and ONL exhibit more distinct outlines and textures in the coefficients than in the original images. However, directly applying DWT to DL approaches fails to discuss the importance of different multispectral information for specific diseases or rank them within the fused wavelet features.

According to the above systematical analysis, the following two important problems need to be solved when applying DWT to CNNs in PD screening automatically. (1) How can we select frequency information that captures rich retinal structural features, regarded as clinical

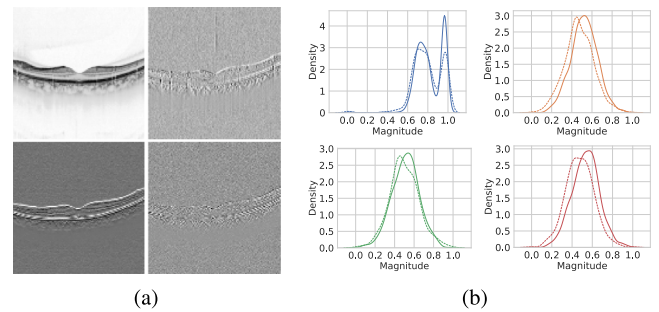


Fig. 2. (a) is an example of extracting DWT (with Haar filter) coefficients, followed by LL, LH, HL, HH (an approximation, horizontal detail, vertical detail, and diagonal detail subbands). (b) is kernel density estimation feature plot for the four frequency coefficients (solid line for HC and dashed line for PD) among the whole OCT images.

priors, without introducing noise in the high-frequency components? (2) How can we design a method to enhance meaningful features while suppressing weak ones for all structural and minutiae features in OCT images?

To better compress multi-spectral information, we introduce DWT decomposition for the latent feature maps from the different layers of CNNs to explore the various information between low- and high-frequency sub-bands. Note that the proposed module based on DWT consists of two main components: *selection* and *recalibration* to resolve the aforementioned problems, respectively. The selection operator automatically selects the information among four coefficients, reducing noise and without loss of information. It is followed by a recalibration operator, which controls the contribution of properties by adjusting the feature response while maintaining their structural details. The weights calculated by the above operators finally suppress irrelevant regions in input and highlight salient features useful for a specific task as well. Our proposed Selection and Recalibration Module (SRM) is easily integrated into standard CNN architectures such as the ResNet model with smaller computational overhead while increasing the prediction accuracy. Thus, in this paper, we propose a novel yet effective wavelet-based selection and recalibration network (WaveSRNet) to automatically screen PD in OCT images. To the best of our knowledge, this is the first work for PD screening using DL techniques with DWT in OCT images. The main contributions of this work are summarized as follows:

1. We propose a network based on frequency domain learning for automatic PD screening. It automatically focuses on specific retina structures (RNFL, GCIP, OPL, ONL) by incorporating prior clinical knowledge, addressing the model's limitation of the difficult identification in PD because of inconspicuous changes in the retinal structure.
2. We design a module using DWT to accurately screen PD by integrating low and high-frequency domain information and leveraging the distinctive properties of each frequency band.
3. This paper explains the inherent decision-making process of WaveSRNet through visualization analyses, enhancing the explainability of our method.
4. The results of PD screening further suggest the potential role of the retina as an assessment tool for neurological dysfunction in PD patients. Additionally, experiments have validated the effectiveness and versatility of our model on two public retinal disease recognition datasets.

The remainder of this paper is organized as follows. Section 2 briefly introduces the related works. Section 3 presents the proposed network. Section 4 details the experimental setup. Section 5 presents the performance of our method and analyses the results. Section 6 summarizes limitations and discusses future work. Finally, Section 7 provides the conclusion.

2. Related work

2.1. OCT images with CNNs

As a high-resolution retinal imaging method, OCT is a non-invasive imaging technique that visualizes retinal cross-sections [14]. In recent years, retinal OCT image segmentation or classification tasks via CNNs have been widely used in retina-related diseases. For instance, segmentation for medical analysis is rapidly evolving in retinal layer segmentation [15], fluid segmentation [16], choroid tissue segmentation [17], hyperreflective foci segmentation [18], and hybrid segmentation [19]. As for classification, Yoo et al. [20] detect rare diseases using the concept of few-shot learning with OCT and solve the problem of a small amount of OCT data on rare diseases by the generative adversarial network. [21] demonstrate the relationship between neovascular age-related macular degeneration and volumes for all segmented tissues on OCT images. Das et al. [22] propose a method for automated classification of macular pathologies by fusing information from multiple image scales to capture the inter-scale variations among the pathological manifestations. These papers consider the different characteristics such as the position and volumes of the lesion of the OCT image to improve the performance. Motivated by these, the retinal changes of the pathological structure in PD are considered in our method, which is particularly critical in clinical practice on PD screening.

2.2. Attention mechanism

Many methods based on CNNs have achieved optimal classification performance in recent years. However, convolution as a local operation fails to focus on the target while suppressing the less useful ones. Thus, many methods have started to focus on the attention mechanism. The existing attention mechanism models are divided into different types: channel attention (pay attention to what), spatial attention (pay attention to where), and hybrid combined channel & spatial attention. First, different channels in different feature maps usually represent different properties [23]. The channel attention learns the importance of each channel and finally assigns different weight coefficients to each channel, which is a properties selection process. Squeeze-and-excitation (SE) attention [24] is presented for this purpose. Second, spatial attention is seen as an adaptive spatial region selection mechanism. Wang et al. [25] show that the response at a certain pixel is the sum of the feature weights at all other points and Non-local is proposed to pass information back and forth over long distances. Third, the convolutional block attention module (CBAM) [26] adaptively selects both important objects and regions, combining the advantages of channel attention and spatial attention.

Besides, some studies have also begun to explain attention from different perspectives, which capture more features through low- and high-frequency components in the frequency domain. For example, Wang et al. [13] embed the wavelet attention module in the UNet to the segmentation task rather than the classification. Zhao et al. [27] introduce a wavelet-attention CNN through DWT but they only capture the information in the high-frequency domain. Salman et al. [28] design interdependent channel attention with multiple different wavelet filters, which fails to consider the relation between low- and high-frequency information. The Dual Wavelet Attention Network (DWAN) [29] uses the information of each frequency band after DWT to obtain spatial information, without considering the noise interference of high-frequency bands. The Frequency Channel Attention Network (FcNet) introduced by [30] applies a discrete cosine transform to channel attention, but the best performance reported by the paper requires the manual selection of well-performing frequency components one by one. Based on these observations, we propose a method for CNNs to learn useful features in the frequency domain space to represent clinical characteristics and to implicitly suppress irrelevant regions in an input image while highlighting salient features useful for a specific task.

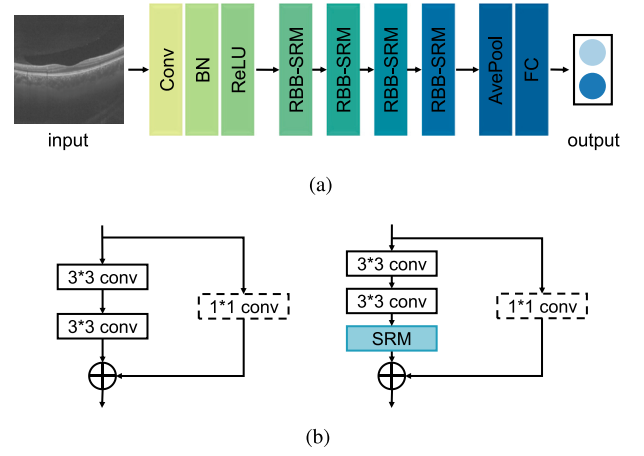


Fig. 3. (a) is an example of the architecture of the WaveSRNet. On the left of the (b) illustrate the existing residual building block (RBB) and on the right is our designed SRM embedded in RBB. SRM uses the statistical features of low- and high-frequency sub-bands as additional information, enhancing the representation of retinal structures.

3. Methodology

In this section, the DWT and attention mechanism are first reviewed. We prove that the Global Average Pooling (GAP) is a special case of DWT. Then, we propose the wavelet-based selection-and-recalibration network (WaveSRNet) for PD screening on OCT images. The proposed SRM is seamlessly integrated into CNNs architectures such as ResNets and trained end-to-end, as shown in Fig. 3(a). The details of the proposed module are shown in (b). The left side depicts the residual building block (RBB) of the original ResNet, while the right side presents our configuration of the SRM integrated into the RBB. SRM consists of two main components: *selection* and *recalibration*. With the adaptive selection, WaveSRNet embeds the useful frequency sub-bands into latent information while neglecting the noise component. With a channel-wise recalibration operator, WaveSRNet leverages the frequency-based latent information to reweight the importance of each property. In these ways, it has the potential to learn more representative and discriminative features among OCT images about clinical prior, thus achieving better performance of PD screening.

3.1. Revisit of 2D DWT and attention mechanism

2D DWT. Suppose $\varphi(x)$ is a scale function and $\psi(x)$ is a wavelet function. For an image $f(x, y) \in \mathbb{R}^{M \times N}$, the approximation and detail coefficients are defined as

$$W_{\varphi}(0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot \varphi_{0,m,n}(x, y), \quad (1)$$

$$W_{\psi}^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot \psi_{j,m,n}^i(x, y), \quad (2)$$

where $i = \{H, V, D\}$ are three high-frequency components, which save the horizontal, vertical, and diagonal details of $f(x, y)$, respectively. The scale function of orthogonal wavelet basis satisfies as $\varphi(x, y) = \varphi(x)\varphi(y)$. Thus,

$$W_{\varphi}(0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot \varphi_{0,m,n}(x) \cdot \varphi_{0,m,n}(y). \quad (3)$$

For the Haar wavelet transform, $\varphi_{0,m,n}(x) \cdot \varphi_{0,m,n}(y) = 1$ in its shifting windows, is equal to the average pooling in the sub-domain. In addition, the $\sum_{m,n \in 0,1,2,\dots,2^{j-1}} W_{\varphi}(j_0, m, n)$ is formulated as

$$\sum_{m,n \in 0,1,2,\dots,2^{j-1}} W_{\varphi}(0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y). \quad (4)$$

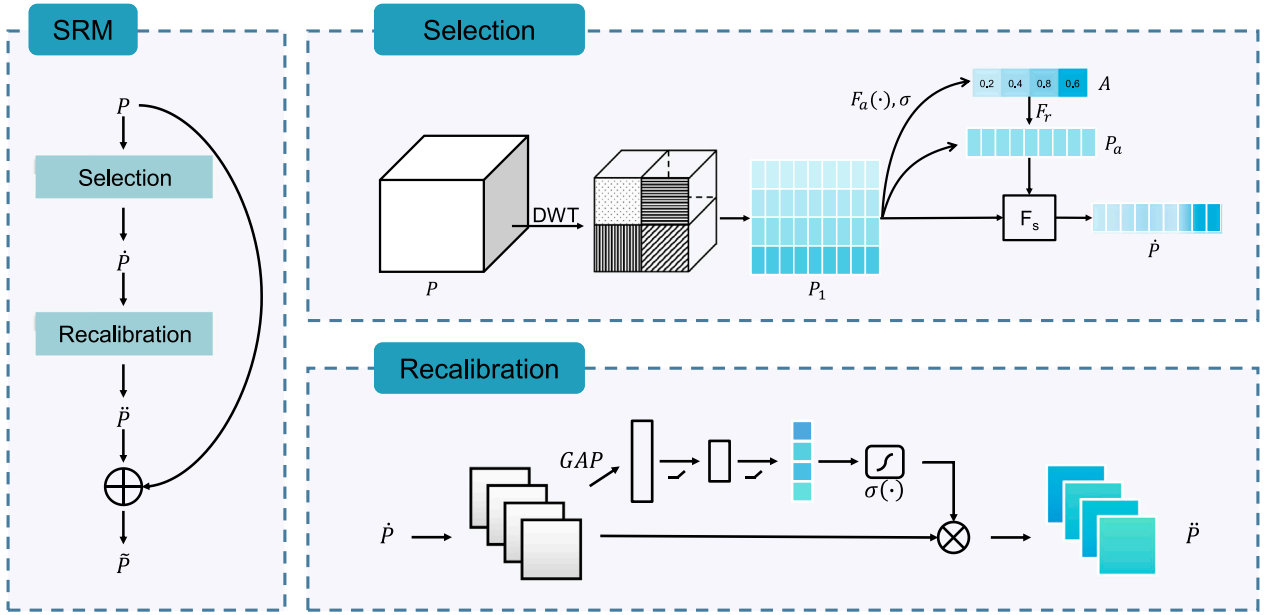


Fig. 4. Architectural configuration of the SRM. The module has two sequential sub-modules: *selection* and *recalibration*. The intermediate feature map \mathbf{P} is decomposed by DWT and adaptively refined through our module (SRM) at every residual block of CNNs.

If a GAP is applied over the M and N to learn a global feature in $f(x, y)$, then it can be formulated as $\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N f(x, y)$. Thus, it is clear that the sum of low-frequency subband feature coefficients is equivalent to GAP:

$$\sum_{m,n \in 0,1,2,\dots,2^j-1} W_\varphi(0, m, n) = \text{GAP}(f(x, y)) \sqrt{MN}. \quad (5)$$

Attention mechanism. There have been several attempts to incorporate attention mechanisms to improve the performance of CNNs in classification tasks in medical image analysis. Specifically, channel attention such as [24] first calculates the importance of each channel, and then applies it to the feature map, thereby improving the expressive ability of the network in feature representation. Let us represent an intermediate input feature map as $\mathbf{P} \in \mathbb{R}^{C \times H \times W}$ that passes through the feature extractor, where C is the number of channels, H and W indicate spatial dimensions. Taking the channel-wise squeeze-and-excitation block as an example, we place this block on \mathbf{P} to re-weight it to $\tilde{\mathbf{P}}$. In this way, the attention mechanism is written as:

$$\alpha_{\text{att}} = \sigma(\mathbf{F}_{\text{MLP}}(\mathbf{F}_{\text{compression}}(\mathbf{P}))), \quad (6)$$

where $\alpha \in \mathbb{R}^C$ is a vector of attention weight, σ is the sigmoid function to rescale activations to $[0, 1]$, \mathbf{F}_{MLP} represents the multi-layer perceptron (MLP) with one hidden layer, and $\mathbf{F}_{\text{compression}} : \mathbb{R}^{C \times H \times W} \rightarrow \mathbb{R}^C$ is a compression operation such as GAP. The final output of the block is obtained by rescaling \mathbf{P} with an operation:

$$\tilde{\mathbf{P}} = \mathbf{P} \otimes \alpha_{\text{att}}, \quad (7)$$

where \otimes denotes element-wise multiplication.

Therefore, GAP is a simple and efficient way to compression and is equal to the sum of low-frequency sub-band feature coefficients proved by Eq. (5). In this paper, since the pathological features of PD are manifested in small changes in the retina, our method fully utilizes the information from high-frequency sub-bands to assist the network in learning features, instead of directly using GAP to compress global spatial information, which can lead to information loss. Meanwhile, due to the diversity of features, a recalibration operator is proposed to emphasize or suppress features.

3.2. Selection-and-recalibration module

Based on the theoretical analysis above, we can see that using GAP in the network preserves only the lowest frequency information and ignores the other frequency information which encodes the useful information in representing structures on OCT images. Thus, we propose a novel module in the wavelet domain named SRM which mainly uses low- and high-frequency features to construct the module.

Given an input tensor \mathbf{P} , SRM generates channel-wise selection weights $\tilde{\mathbf{P}}$ and recalibration weight $\tilde{\mathbf{P}}$ as illustrated in Fig. 4. The overall process is summarized as:

$$\tilde{\mathbf{P}} = \mathbf{M}_s(\mathbf{P}), \quad (8)$$

$$\tilde{\mathbf{P}} = \mathbf{M}_r(\tilde{\mathbf{P}}), \quad (9)$$

$$\tilde{\mathbf{P}} = \tilde{\mathbf{P}} \otimes \mathbf{P}, \quad (10)$$

where \mathbf{M}_s and \mathbf{M}_r denote selection and recalibration sub-modules respectively, and $\tilde{\mathbf{P}}$ denotes the final refined output. During multiplication \otimes , the weight values are broadcasted (copied) accordingly: channel attention values are broadcasted along the spatial dimension. The following describes the details of each sub-module.

3.2.1. Selection for choosing frequency components

In clinical practices, the basis of screening PD on OCT images for doctors is the previously introduced retinal (e.g., RNFL, GCIP, OPL, and ONL) whose changing of structures is weak. Thus, the spatial distribution of these features is critical in PD screening. Given this motivation, we use 2D DWT to capture both the image's spatial-frequency components and obtain representation in texture [31]. Here, we investigate how to select frequency components after DWT decomposition. Recently, most of the selection strategies in existing wavelet-based methods are manual. For instance, Qin et al. [30] propose three kinds of criteria, where two-step selection criteria choose the top-k highest performance frequency components based on the results from the individual evaluation of each frequency component. On the contrary, we select frequency components automatically by considering their inconsistency in a sequence of frequency domains. The most important thing is that our selection operation is trainable and non-manual, which greatly improves efficiency.

Wavelet feature extraction and embedding. For the input tensor \mathbf{P} , the 2D DWT decomposes the tensor into four subbands which are written as:

$$\mathbf{LL}, \mathbf{LH}, \mathbf{HL}, \mathbf{HH} = \text{2D DWT}(\mathbf{P}), \quad (11)$$

where the four outputs are the approximation (\mathbf{LL}), horizontal detail (\mathbf{LH}), vertical detail (\mathbf{HL}), and diagonal detail (\mathbf{HH}) subbands, respectively. The value of each map represents the wavelet coefficients. To embed the different information into latent space, we employ a compression operator:

$$\mathbf{P}_{\text{LL}} = \frac{1}{H/2 \times W/2} \sum_{i=0}^{H/2} \sum_{j=0}^{W/2} \mathbf{LL}_{ij}, \quad (12)$$

and the same as the remaining coefficients. In this way, our network incorporates low- and high-frequency features, alleviating information loss due to GAP compression. Then, these features are stacked to build a feature map:

$$\mathbf{P}_1 = [\mathbf{P}_{\text{LL}}, \mathbf{P}_{\text{LH}}, \mathbf{P}_{\text{HL}}, \mathbf{P}_{\text{HH}}], \quad (13)$$

where $\mathbf{P}_1 \in \mathbb{R}^{4 \times C}$ represents the compressed features of multiple sub-bands combined in frequency.

Wavelet feature selection. Recently, feature selection of different instances improves model performance towards better generalization by using the properties of the data distribution directly [32]. Thus, we design a wavelet feature selection sub-module to reduce the focus on their out-of-distribution features by considering the inconsistency between the four subbands, avoiding noise interference when introducing high-frequency information. In specific, the ranking scores among representations of four sub-bands \mathbf{P}_1 are driven by the classifier:

$$\mathbf{A} = \sigma(\mathbf{F}_a(\mathbf{P}_1; \phi_a)), \quad (14)$$

where \mathbf{F}_a represents the linear classifier with its parameters ϕ_a . The ranking score is then applied to \mathbf{P}_1 through the operator $\mathbf{F}_r(\cdot)$ to obtain the compressed vector corresponding to the maximum score. The out-of-distribution feature is formalized as:

$$\mathbf{p}_a = \mathbf{F}_r(\mathbf{P}_1, \mathbf{A}). \quad (15)$$

Finally, the selection operation of features is modeled as:

$$\mathbf{P}_2 = \mathbf{F}_s(\mathbf{P}_1, \mathbf{p}_a), \quad (16)$$

where the \mathbf{F}_s includes two steps: the first step is to eliminate noise interference, and the second step is to fuse the remaining features. The feature is transformed into:

$$\mathbf{P}_2 = \text{ReLU}(\mathbf{P}_1 - \mathbf{p}_a), \quad (17)$$

$$\tilde{\mathbf{P}} = \sigma(\mathbf{W}_1^1(\text{ReLU}(\mathbf{W}_2^1 \mathbf{P}_2)), \quad (18)$$

where $\mathbf{W}_1^1 \in \mathbb{R}^{C \times \frac{C}{r}}$ and $\mathbf{W}_2^1 \in \mathbb{R}^{\frac{C}{r} \times C}$ are weights of two fully-connected layers. The parameter r indicates the reduction ratio. In this way, we obtain an aggregated description of different information on each sub-band, and this description automatically weakens the noise and focuses on the important retinal layers in the OCT.

3.2.2. Recalibration for wavelet features

The features are considerable diversity of representation among four subbands. Inspired by the success of the recalibration [33], we apply recalibration to leverage aggregated features, enhancing discriminative performance. This approach enables the network to use global information to selectively emphasize informative features and suppress less useful ones. To achieve this, we adopt a simple feed-forward network and a sigmoid function to convert $\tilde{\mathbf{P}}$ into recalibration weights by this sub-module. The formula is as follows:

$$\alpha = \sigma(\mathbf{W}_1^2(\text{ReLU}(\mathbf{W}_2^2 \tilde{\mathbf{P}} + \mathbf{b}_2^2)) + \mathbf{b}_1^2). \quad (19)$$

Table 1

Data source description. We provide the summary of the PD-OCT dataset, including the number of eyes and sex, as well as the average age, axial length, and best-corrected visual acuity.

Characteristics	HC	PD	P value
Eye, n	88	37	–
Age (years)	61.61 ± 5.94	65.73 ± 6.90	0.0149
Sex (Male/Female)	17/32	15/11	1.0000
Axial length (mm)	23.48 ± 1.04	23.18 ± 0.87	0.3708
Best corrected visual acuity (logMAR)	0.98 ± 0.92	1.01 ± 0.92	0.1787

where $\mathbf{W}_1^2 \in \mathbb{R}^{C \times \frac{C}{r}}$ and $\mathbf{W}_2^2 \in \mathbb{R}^{\frac{C}{r} \times C}$ are linear transformations and \mathbf{b}^2 are bias terms. Then, the selected features are recalibrated by the weight α , so the output is obtained by:

$$\tilde{\mathbf{P}} = \tilde{\mathbf{P}} \otimes \alpha. \quad (20)$$

where $\tilde{\mathbf{P}} \in \mathbb{R}^C$ is the output of the recalibration \mathbf{M}_r .

Finally, we obtain the final output of the SRM by rescaling \mathbf{P} with the $\tilde{\mathbf{P}}$:

$$\tilde{\mathbf{P}} = \mathbf{P} \otimes \mathbf{h}(\tilde{\mathbf{P}}), \quad (21)$$

where $\mathbf{h}(\cdot)$ is a repeat operation to align the feature space.

In clinical research, different retinal layers in OCT have different contributions to the screening of PD, and we propose that different layers of structures, textures, and details information represented by frequency sub-bands are distinguished. Therefore, our SRM learns such prior knowledge through DWT and obtains new weights through the selection and recalibration sub-module, capturing spatial frequency information and feature representation while reducing noise interference caused by high frequencies.

4. Dataset and experiment setting

4.1. Datasets

A clinical PD-OCT dataset is collected by a cooperative hospital from 2020 to 2023, in China. The Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) motor part total and motor III are used for clinical assessment. Table 1 describes the main characteristics of the dataset used for this study. For all OCT images selected in this dataset, we collect the statistics of important information, such as the number of eyes and sex, the average age, axial length, and best-corrected visual acuity, etc. All the OCT images are captured along the eye's optical axis by an SVision VG200, where the area of images is 12 mm × 2.7 mm centered on the macula. The dataset collection of this paper is conducted according to the tenets of the Helsinki Declaration. Due to the retrospective nature and fully anonymized usage of the dataset, the medical ethics committee exempts us from informing the patients.

Our PD-OCT dataset consists of 26 PD subjects and 49 HC subjects. Among these, 25 subjects have OCT images from only one eye, while the remaining 50 have images from both eyes. Each eye includes 18 retinal OCT scans with a resolution of 1561 × 2048 pixels. In total, there are 2250 OCT images, categorized into HC (negative) and PD (positive) with 1584 and 666 images, respectively. The dataset is split into 60% for training, 20% for validation, and 20% for testing.

Predictive models may introduce or reinforce biases implicit in the data. In this paper, we apply fairness methods in data collection and in-processing to mitigate the sample and label bias. (1) Data collection. All subjects underwent a series of standard ophthalmologic examinations, including refraction, best-corrected visual acuity test, orthoptic examination, slit-lamp biomicroscopy, and IOP measurements. There are no significant differences regarding age, sex, axial length, and best-corrected visual acuity between the two groups (HC and PD), as shown in Table 1. Exclusion criteria for all eligible individuals included having any history of diabetes, glaucoma, Alzheimer's disease or other known

Table 2

Comparison of different state-of-the-art methods on the PD-OCT dataset.

Methods	Params	Flops	ACC	AUC	F1	Kappa
Inception	21.79	2.84	0.6759	0.5803	0.3000	0.1073
ResNet	11.17	1.74	0.7616	0.7147	0.4550	0.3235
TinyViT	20.59	4.05	0.7222	0.6910	0.3407	0.1964
ResMLP	14.74	2.97	0.7407	0.7572	0.4105	0.2664
WavePooling	11.96	48.46	0.6829	0.4915	0.0284	0.0359
WCNN	9.88	2.75	0.7662	0.7297	0.3804	0.2859
DAWN	11.21	6.91	0.7639	0.8030	<u>0.4688</u>	0.3355
WaveCNet	11.17	2.26	0.7477	0.6981	0.4410	0.2956
RIRNet	21.76	3.60	0.7292	0.5646	0.3158	0.1917
RCRNet	11.18	1.75	0.7477	0.7062	0.3313	0.2292
FITNet	85.39	4.87	0.7153	0.6491	0.4581	0.2682
WaveSRNet	12.57	1.85	0.8056	<u>0.7955</u>	0.5532	0.4468

neurological disease, eye surgery or any other retinal pathology, and insufficient scan quality. (2) During training, we use data augmentation (such as random horizontal flip, rotation, color jitter, and cropping) to increase sample size and improve generalization and robustness.

4.2. Experiment detail

We use the PyTorch toolbox and RTX 2080 Ti GPU with 11 GB of memory to implement all our experiments. During training, we use the SGD optimizer with decay and momentum of 0.9 to train all the models from scratch. The weight decay is set to 1×10^{-4} always. The learning decay with an initial learning rate of 0.01 is adopted. We use a NVIDIA GPU for training and the batch size is set to 64. All our experiments are trained with a cross-entropy loss function:

$$\mathcal{L} = - \sum_{i=0}^K y_i \log(\hat{y}_i), \quad (22)$$

where y_i represents the ground truth on the i category (usually 0 or 1), and \hat{y}_i represents the probability predicted by the model.

In this study, we adopt three commonly used evaluation metrics including accuracy (ACC), area under curve (AUC), F1, and Kappa scores, where ACC is the primary metric for assessing the prediction accuracy of the models. ACC presents the closeness of the measurement results to the true value, F1 gives the harmonic mean of precision and sensitivity, and the higher the AUC score, the more powerful the classifier. These metrics are defined as follows:

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Sen} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{PR} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{F1} &= \frac{2 \times \text{PR} \times \text{Sen}}{\text{PR} + \text{Sen}}, \end{aligned} \quad (23)$$

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

5. Results and analysis

In this section, we conduct extensive experiments to report the results of the proposed approach compared to other state-of-the-art methods on our PD-OCT dataset. The adopted methods are categorized into two aspects: state-of-the-art methods (four strong baselines, four wavelet-based methods, three oct-related methods) and five attention-based methods.

Table 3Comparison of different state-of-the-art methods on the PD-OCT dataset. **SE**: channel attention; **CBAM**: convolutional block attention module; **Coord**: coordinate attention; **Fca**: frequency channel attention; **DWA**: dual wavelet attention. **SRM**: selection and recalibration module.

Methods	Params	Flops	ACC	AUC	F1	Kappa
ResNet18	11.17	1.74	<u>0.7616</u>	0.7147	<u>0.4550</u>	0.3235
ResNet18+SE	11.26	1.74	<u>0.7616</u>	0.8003	0.4372	0.3122
ResNet18+CBAM	11.26	1.74	0.7407	0.6819	0.2632	0.1815
ResNet18+Coord	11.31	1.74	0.7569	0.6472	0.3137	0.2350
ResNet18+Fca	11.26	1.74	0.6898	0.5821	0.2472	0.0926
ResNet18+DWA	11.21	6.91	0.7569	0.7536	0.3137	0.2350
ResNet18+SRM	12.57	1.85	0.8056	<u>0.7955</u>	0.5532	0.4468
ResNet50	23.51	4.04	0.7431	0.7167	0.3353	0.2242
ResNet50+SE	26.04	3.81	0.7083	0.5980	0.1600	0.0735
ResNet50+CBAM	26.04	4.04	0.7593	<u>0.7415</u>	<u>0.4091</u>	0.2917
ResNet50+Coord	26.04	4.07	0.7315	0.6397	0.2750	0.1724
ResNet50+Fca	26.02	4.04	0.6898	0.5980	0.3163	0.1364
ResNet50+DWA	24.16	17.23	0.7616	0.6713	0.4181	0.3005
ResNet50+SRM	63.78	7.55	0.7847	0.8410	0.6204	0.4703

5.1. Comparison with state-of-the-art methods

Strong baselines. We compare WaveSRNet with strong baselines for PD screening and report results on three evaluation measures in Table 2. Strong baselines include two CNN networks: Inception [34] and ResNet [35], a transformer-related model: TinyViT [36], and a multilayer perception-related model: ResMLP [37], where ResMLP uses 12-layer models. It is observed from the table that the relative best performances are 0.7616, 0.7147, and 0.4550 tested on ResNet among the baselines, and our method increases by 4.40%, 8.08%, and 9.82% compared with it. Considering the performance of PD screening, our method achieves relatively better accuracy at a small increase in the parameters and complexity. These prove that the WaveSRNet designed in this study extracts discriminative features on the structure of the retina, which has the potential to assist doctors in clinical diagnosis as a CAD tool.

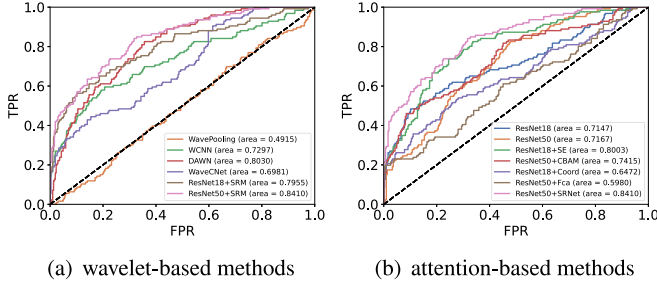
Wavelet-based methods. Different from our wavelet-based module, there are other wavelet-integrated modes in works of literature, including WavePooling [38], WaveCNet [39], WCNN with Level-4 DWT composition [40], and DAWN with 128 initial convolutional layers [41]. As can be seen from the experimental results in Table 2, the proposed method achieves relatively significant performance gains on our PD-OCT dataset, especially for the ACC. Although the number of parameters is increased somewhat, our approach shows at least 3.94% and 8.44% improvement in ACC and F1 compared with the wavelet-based model. Besides, the computation cost of our approaches is proven by evaluating FLOPs, achieving better PD screening performance/efficiency trade-offs. To more intuitively compare the performance of each model under the AUC metric, the results of various methods are drawn into ROC curves in this study, as shown in Fig. 5(a). Although the AUC value in the table ranks second, the backbone ResNet50 with SRM has increased by at least 3.8% compared with other wavelet-based methods. Thus, the model updates its weights to focus on the multi-layer structure in the retina through selection and recalibration, emphasizing useful features. And the DWT-based SRM well extracts the details and texture information in the retinal structure, which is beneficial for the model to learn a richer semantic representation.

OCT-related methods. In addition to general approaches, our method is also compared with OCT-related medical image classification methods applied in specific domains. These methods include RIRNet [42], RCRNet [43] and FITNet [44]. Table 2 provides the experimental results. We can observe that RIRNet produces the better ACC and FITNet produces the better Kappa among advanced methods related to OCT-based retinal disease classification. However, compared with our method, the performance of these methods is still a little

Table 4

Five-fold cross-validation among WaveSRNet and state-of-the-art methods on the PD-OCT dataset.

Methods	ACC	AUC	F1	Kappa
ResNet	0.7144 \pm 0.0325	0.7172 \pm 0.0150	0.4137 \pm 0.0577	0.2419 \pm 0.0260
TinyViT	0.7181 \pm 0.0312	0.7498 \pm 0.0258	0.4832 \pm 0.0549	0.2955 \pm 0.0462
ResNet+SE	0.7236 \pm 0.0271	0.7514 \pm 0.0084	0.4761 \pm 0.0706	0.2978 \pm 0.0469
ResNet+CABM	0.7269 \pm 0.0443	0.7463 \pm 0.0323	0.4842 \pm 0.0911	0.3140 \pm 0.0623
DAWN	0.7398 \pm 0.0543	0.6372 \pm 0.0603	0.2604 \pm 0.1323	0.1808 \pm 0.1067
RIRNet	0.7245 \pm 0.0157	0.6885 \pm 0.0350	0.3577 \pm 0.0770	0.2157 \pm 0.0391
RCRNet	0.7102 \pm 0.0233	0.7012 \pm 0.0118	0.3839 \pm 0.0840	0.2164 \pm 0.0371
FITNet	0.7315 \pm 0.0205	0.6956 \pm 0.1003	0.3113 \pm 0.1874	0.2049 \pm 0.1255
WaveSRNet	0.7449 \pm 0.0408	0.7516 \pm 0.0325	0.5676 \pm 0.0516	0.3920 \pm 0.0596

**Fig. 5.** ROC curves of the attention-based and wavelet-based methods on the PD screening task.

behind. We argue that this is due to the fundamental dissimilarity between the OCT image characteristics of PD and those assumed by these methods. For example, RIRNet designs a module according to the clinical diagnosis standard of nuclear cataracts (correlation between severity levels and averaged nucleus density) to enhance the representation of the feature map in both the upper and lower regions. FITNet enriches the feature representation by leveraging obvious differences in the position, shape, and size of the lesion structure across various scanning directions. However, the unique pathology of PD patients in OCT is subtle changes in the retinal layers (as shown in Fig. 1(a)) that are more complex and difficult to detect directly through DL approaches. Therefore, we propose leveraging both low- and high-frequency features to enhance the attention of subtle changes among retinal layers.

5.2. Comparison with attention-based methods

The comparison of different attention-based models on the PD-OCT dataset is shown in Table 3. Experimental results show that SE and CBAM show the second-best performance separately in different depths of CNNs by analyzing multiple metrics. Our proposed SRM outperforms channel, spatial, coord-based attention, and wavelet-based attention (Fca and DWA) on both depths, achieving at least 4.4% and 2.54% improvements respectively on PD screening. Moreover, the Fca-based attention method performs the worst (ACC 68.98%) in PD screening, which is even only 1.39% higher than the worst baseline Inception. This may be because FcaNet explores different combinations of frequency components by individual experiments and that uses manual selection strategies for other tasks cannot be adapted to our specific clinical diagnostic task. In contrast, our adaptive and trained selection submodule embedded in the SRM facilitates the adaptation of retinal structures across different sub-bands, assisting the model in learning discriminative representations. Finally, We choose the better results of attention embedded in CNNs of different depths to plot the receiver operating characteristic (ROC) curves, as shown in Fig. 5(b). Results show that our proposed methods, including SRM with two backbones, outperform the state-of-the-art attention-based methods. Thus, the features from SRM better reflect the variation between different retinal layers.

5.3. Five-fold cross validation

Table 4 lists the five-fold cross-validation results of our WaveSRNet and eight other state-of-the-art methods on the PD-OCT dataset. Here, the test set is consistent with the experiments above. Then, we split the remaining samples into five disjoint sets; at each time, four sets are used for training, and a set is used for validating. As shown in the table, we can summarize the following information. (1) In the k-fold experiment, the performance of all methods shows a slight decline compared to the original paper. But it remains within a reasonable range. (2) Our method achieves the highest Kappa score of 0.3920 among all methods. Moreover, considering that other medical image tasks [45,46] report smaller Kappa values, our score below 0.5 is still deemed acceptable. (3) Our method demonstrates superior performance, achieving scores of 0.7449, 0.7516, 0.5676, and 0.3920 in ACC, AUC, F1, and Kappa, respectively. Compared to the method with the second-highest Kappa value (ResNet+CABM), each indicator shows an improvement of 0.018, 0.0053, 0.0834, and 0.078, respectively.

5.4. Visualization and interpretation

To enhance the explainability of the proposed frequency learning approach in relation to the pathology of PD (retinal layers), this section provides some visualization experiments to explore the decision-making process of WaveSRNet further. We first use the feature map to observe the information captured by the method, then use the heat map to display the features that the model focuses on, and finally explore the influence of the feature on the prediction result through the shapley additive explanations (shap) value.

Visualization of feature maps among four frequency coefficients. The aim of this analysis is to illustrate that our DWT-based method effectively extracts information from the retinal structure, while also investigating the factors that influence the model's decision-making process. Fig. 6 visually depicts the frequency coefficient features obtained through DWT decomposition at each stage of the network, and we show the first nine feature maps. We make the following three observations about the role of the DWT: (1) The approximation (LL) and three-direction subbands (LH, HL, and HH) focus on specific regions of the retina, particularly at SR_1_2. (2) In addition to the approximation coefficient, the other three sub-bands effectively supplement low-frequency information, helping the network learn detailed information (e.g., edges, texture, patterns, parts of objects) of retinal structures from the specific area of the OCT image. (3) At greater depth, the features become more abstract, as seen in SR_4_2.

Grad-CAM for different attention methods. We visualize the heat maps generated by the advanced attention-based methods and the proposed SRM, finding what features are more important than others to the model. Fig. 7 provides two CAM examples of different categories. We can see that: (1) In addition to our method, other attention-based models focus on the retinal structure of negative samples, while the attention area of positive samples is more divergent. (2) Regardless of positive or negative samples, SRM noticed retinal structures in OCT. (3) The SRM focuses on more fine-grained retinal structures. For instance, it highlights the structures of GCIP on the negative sample while the

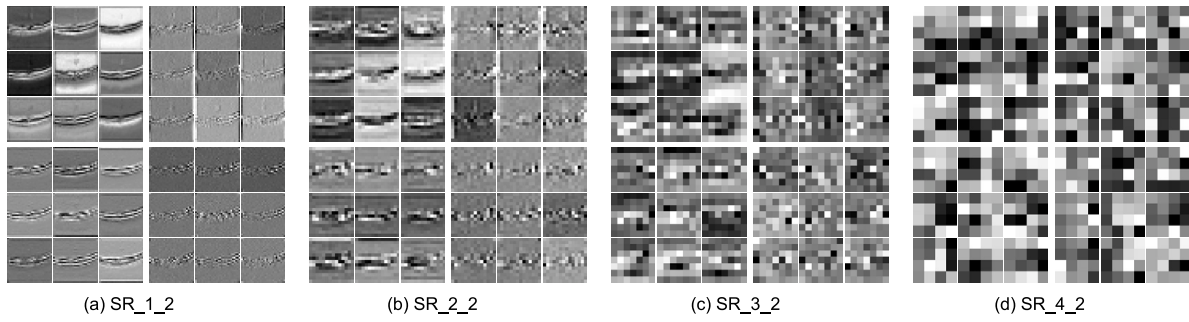


Fig. 6. Visualization of feature maps among four frequency coefficients in our fully trained WaveSRNet. Each set of feature maps is named according to the following scheme: $SR_{stageID_blockID}$. For instance, (a) SR_1_2 represents the latent feature map in SRM at Stage 1, block 2. Each stage displays 3×3 (nine-square grid) feature maps, divided into four components in the frequency-domain space: LL, LH, HL, and HH.

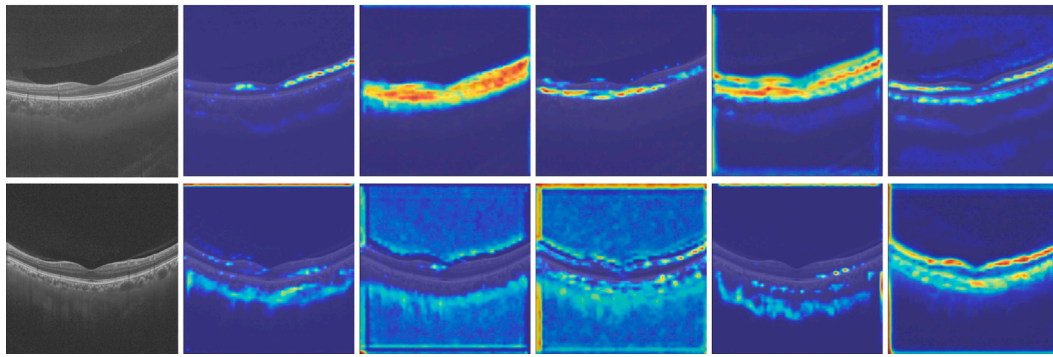


Fig. 7. The Grad-CAM visualization results of other state-of-the-art attention methods (SE, CBAM, Coord, Fca) and the proposed WaveSRNet. The top row shows an example of a negative sample, while the bottom row shows an example of a positive sample.

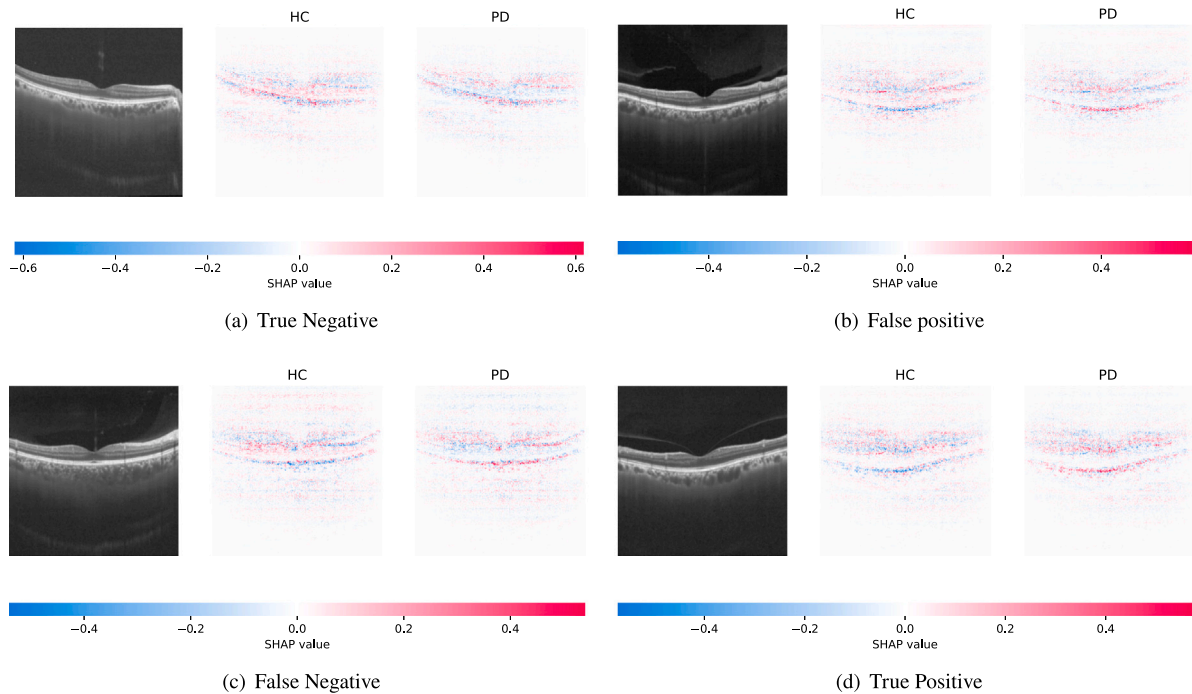


Fig. 8. Examples of PD screening on our method from the test set. For each sub-figure, the ensuing SHAP plot consists of the original image followed by 2 dim grayscale versions of itself.

Table 5

Performance comparisons of the proposed sub-modules and the baseline models on our PD-OCT dataset. **D**, DWT; **S**, Selection sub-module; **R**, Recalibration sub-module.

Methods	ACC	AUC	F1
ResNet18	0.7616	0.7147	0.4550
ResNet18+D	0.7755	0.7476	0.4921
ResNet18+DS	0.7986	0.8055	0.5837
ResNet18+DR	0.7963	<u>0.7963</u>	0.5926
ResNet18+DSR	0.8056	0.7955	0.5532
ResNet50	0.7431	<u>0.7167</u>	0.3353
ResNet50+D	0.7130	0.6506	0.3187
ResNet50+DS	0.7338	0.6294	0.2857
ResNet50+DR	<u>0.7477</u>	0.7152	<u>0.4108</u>
ResNet50+DSR	0.7847	0.8410	0.6204

positive is RNFL, GCIP, and ORL (including OPL, and ONL). Hence, it can be concluded that our designed sub-modules help CNNs learn the discriminative representation by devoting different contributions to eight retinal layers in each category, which agrees with clinical research and the different contributions of different layers on the retina, proving the effectiveness of our proposed method.

DeepSHAP of our WaveSRNet. Here, our aim is to explore the decision-making process utilized by the model in screening PD on OCT images. We achieve this by leveraging the DeepSHAP method [47], which allows us to discern the contribution of each relevant feature to PD screening with the WaveSRNet. Fig. 8 shows the shap maps of four examples of the proposed method. Beneath the plot is a scale that reads from negative to positive, color-coded from deep blue to bright red, helping to show the intensity of the SHAP value assigned to each pertinent pixel. We make the following two observations about shap maps. First, for the two samples (HC and PD) with correct predictions in Fig. 8(a) 8(d), there are obvious red pixels in the fovea area of the HC, while the OCT of PD appears in the RNFL and IS-OS. Second, for the samples in Fig. 8(b) 8(c) (incorrect prediction), the model has more difficulty determining which class the image belongs to because the regions of interest are more diffuse. These observations suggest that the significance of retinal structures in various areas of the OCT differs across categories. Besides, the pathological changes of PD patients on the retina are not only related to the layers but also to the areas with different diameters centered on the fovea of the macula, whose result is consistent with the clinical research [10].

5.5. Ablation studies

5.5.1. Importance of each sub-module

We here analyze the role of wavelet transform in the CNNs, and conduct experiments on the two sub-modules (**S** for selection and **R** for recalibration) based on DWT (**D**) to verify that the designed modules are effective in improving the performance of PD screening. As shown in Table 5, the results of the original CNNs (e.g., ResNet with different depths) are taken as the baseline, which is sourced from the official PyTorch. Here, we only take ResNet18 as an example to analyze. From the results, it is clear that methods with **D** achieve better performance compared with the baseline. And the DWT-based method without an **S** and an **R** sub-module performs well (77.55%) compared with ResNet18 (76.76%) and obtains about 1.39% improvement on ACC. Besides, we observe a similar trend with regard to the effect of **S** and **R**, finding that SRM (80.56% ACC) outperforms ResNet18+D (77.55% ACC) by 3.01%. Besides, although the results of embedding different sub-modules in ResNet50 are not well, the final method ResNet50+DSR is superior: the performance on AUC and F1 is substantially better than ResNet18+DSR. Thus, it suggests that the proposed approach shows better performance when embedding SRM into baselines.

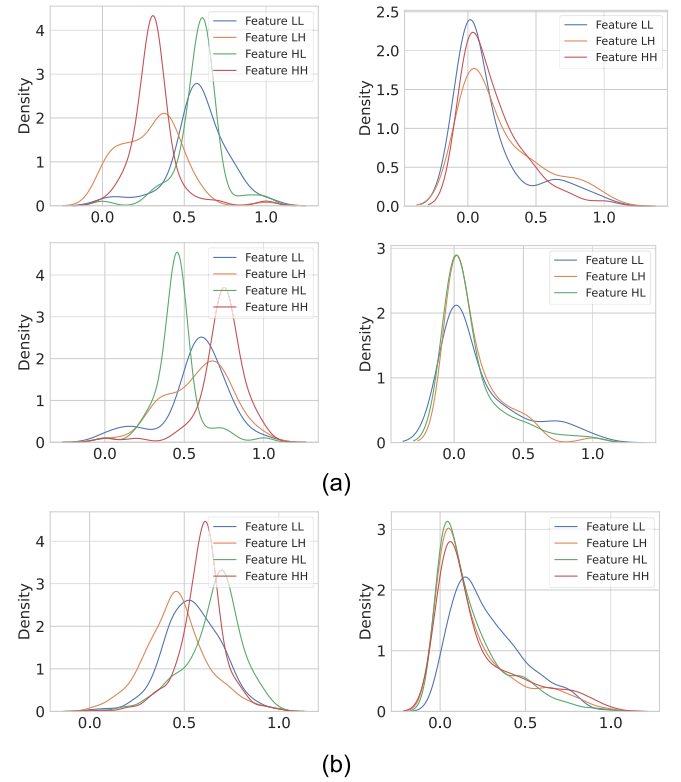


Fig. 9. Normalized density plots of the deep feature magnitudes are presented. In (a), the top row displays the HC sample and the bottom row shows the PD sample, both randomly selected from the test set. (b) illustrates the mean normalized histograms of feature magnitudes from the test set. For each sub-figure in (a) and (b), the left column represents the feature before the selection operation, while the right column represents the feature after the selection.

5.5.2. The effects of different combinations of frequency components

In our approach, one critical problem is to design which frequency components are used for the model. In order to achieve better performance in a specific task, some researchers [13] investigate the performance of each component individually and different components in combination with each other, which is time-consuming and labor-intensive. In this paper, our proposed WaveSRNet extracts information adaptively from different frequency coefficients for a specific task. Here, we compare our adaptive strategy of selection with manual combinations of four components while preserving the recalibration sub-module. As shown in Table 6, the results demonstrate two things. First, the screening performance of PD is robust to a range of selection strategies with the exception of **LL+LH**. Specifically, our adaptive selection strategy achieves relatively better ACC and second-best F1. Second, in the experimental results of the manual selection strategy, selecting only three of these sub-bands achieves slightly superior results, where the performance of the selection scheme without vertical or diagonal high-frequency information ranks second (79.63%). These results now provide evidence that the model introduces noise in high frequency while extracting low- and high-frequency information at the same time. Therefore, our method with adaptive and trained selection sub-module reduces the interference of useless information while offsetting part of high-frequency information.

To facilitate visual observation of the changes before and after wavelet feature selection, we plot the deep feature magnitudes of the four coefficients at Fig. 9. It shows the distribution of feature values across four frequency components for samples of different classes. The x-axis represents the feature values, and the y-axis represents the density estimate. By comparing the curves, we can visually observe the distribution differences of images in the frequency domain feature

Table 6

The performance of different selections on our PD-OCT dataset. ‘S’ means our adaptive selection operation.

Selection	ACC	AUC	F1
S	0.8056	0.7955	<u>0.5532</u>
LL+LH	0.7431	0.6050	0.3729
LL+HL	0.7940	0.7948	0.5340
LL+HH	0.7801	0.7783	0.4693
LH+HL	0.7917	0.7658	0.5000
LH+HH	0.7593	0.6971	0.4526
HL+HH	0.7870	<u>0.8082</u>	0.4588
LL+LH+HL	<u>0.7963</u>	0.7964	0.5644
LL+LH+HH	<u>0.7963</u>	0.7883	0.5217
LH+HL+HH	0.7940	0.7959	0.5189
LL+HL+HH	0.7894	0.8317	0.5517

Table 7

The performance on our PD-OCT dataset for SRM at different reduction ratios.

r	Params	ACC	AUC	F1
2	12.57	0.8056	<u>0.7955</u>	0.5532
4	12.22	0.7500	0.7287	0.4757
8	12.05	<u>0.7847</u>	0.7455	0.5419
16	11.96	<u>0.7778</u>	0.8151	<u>0.5429</u>

Table 8

cSE: channel squeeze and excitation; sSE: spatial squeeze and excitation; csSE: spatial and channel squeeze and excitation.

Methods	Params	ACC	AUC	F1
cSE	12.57	0.8056	<u>0.7955</u>	0.5532
sSE	11.87	0.7894	0.8204	<u>0.5134</u>
csSE	12.57	0.7662	0.7591	0.4521

space. (a) shows that the selection sub-module automatically removes the influence of one subband based on the ranking score A. This observation is consistent with findings in the previous experiment (Table 6): based on the results, eliminating a single wavelet coefficient has a relatively more favorable effect than using two wavelet coefficients. Regarding (b), the features of low- and high-frequency bands after selection have better consistency than those before selection. Consequently, our network obtains more helpful information by eliminating out-of-distribution or uncertainty features on OCT images to improve the performance at the PD screening task.

5.5.3. The impact of reduction ratio r

The reduction ratio r introduced in Eq. (18). To investigate the impact of this hyperparameter, we conduct experiments based on the baseline ResNet18 for a range of different r values. The comparison in Table 7 shows a difference in performance as r increases, where we get the best ACC when $r = 2$. Thus, we use 2 as a parameter for r for all our experiments.

5.5.4. Recalibration strategy

We use channel squeeze and excitation as our recalibration strategy because it is highly effective for image classification. The advantage of using such a strategy is that every intermediate layer has a totally receptive field of the input image. Here, we evaluate the effect of recalibration strategies. The results of the PD screening task are reported in Table 8. It demonstrates that SRM with channel squeeze and excitation significantly boosts the performance of the baseline architecture ResNet18 with almost the same number of parameters.

5.5.5. Position of SRM

We introduce and explore the impact of nine possible configurations of our SRM in CNN architectures such as ResNet18. We first add SRM to the intermediate stages: SR_1, SR_2, SR_3, SR_4, SR_12, SR_123, SR_234, and SR_1234. Besides, there is a variant that integrates SRM into existing architectures: InteSRM. As shown in Table 9, the usage of SRM at

Table 9

Effect of different positions of SRM with ResNet18 on PD-OCT dataset. **Baseline**: CNN architectures such as ResNet18; **SR_stageID**: add SRM to stageID; **InteSRM**, integrates SRM into RBB.

Positions	Params	ACC	AUC	F1
Baseline	11.17	0.7616	0.7147	0.4550
SR_1	11.18	0.7384	0.7240	0.3152
SR_2	11.21	0.7292	0.7307	0.3743
SR_3	11.30	0.7708	0.7764	0.4277
SR_4	11.70	0.7616	0.7233	0.4976
SR_12	11.21	<u>0.8032</u>	0.8028	0.5550
SR_123	11.35	0.7014	0.6113	0.2873
SR_234	11.86	0.7199	0.7264	0.3665
SR_1234	11.87	0.8009	0.7503	0.5000
InteSRM	12.57	0.8056	<u>0.7955</u>	<u>0.5532</u>

positions of SR_1, SR_2, SR_123, SR_234 leads to a drop in performance compared with the baseline, while SR_3, SR_4, SR_12, SR_1234, and InteSRM each perform similarly well. Among these, SR_12 achieves the relatively same performance as InteSRM, which shows that our proposed module achieves comparable classification performance with a small number of parameters increased compared with the ResNet18.

5.6. Validation

To carry out a more comprehensive analysis and validate the general performance of our proposed WaveSRNet, we conduct extensive experiments on two public OCT datasets (i.e., SD-OCT dataset and UCSD dataset) of medical images in distinctly different retinal diseases: dry age-related macular degeneration, diabetic macular edema, drusen and choroidal neovascularization.

First, the SD-OCT dataset [48] includes three categories: 15 normal subjects, 15 patients with dry age-related macular degeneration, and 15 patients with diabetic macular edema, acquired in Institutional Review Board-approved protocols at Duke University, Harvard University, and the University of Michigan. In the experiment, we adopt the same experiment settings as previous works used to ensure the classification results’ fairness [48]. Second, the University of California San Diego (UCSD) dataset [49] is comprised of 4 diagnosis categories (normal, drusen, diabetic macular edema, and choroidal neovascularization), leading to a multi-class classification task. The training set contains a total of 83,484 OCT images while the testing set contains 1000. A more detailed introduction of the dataset can be referred to [49]. We train all methods on 80% images from the training dataset, the remaining images are used for validating, and the testing set is used for final evaluation. It is noted that the training set contains challenging images with many various lesions while the testing set is very small and contains good quality images. Therefore, we discuss the results for the validation set in the results section. Due to the computational load and operational efficiency, the images from the above two datasets are resized to 232 and cropped randomly into 224×224 .

Table 10 presents the classification results of our WaveSRNet, the state-of-the-art methods (strong baselines and wavelet-based methods), and attention-based methods on two public OCT datasets. It can be seen that the best classification results are achieved on the backbone of ResNet18 and ResNet50 respectively on the two datasets. With more parameters, better data fitting, and richer features, the ResNet50 backbone is more suitable for tasks with a large amount of data. Therefore, the UCSD dataset performs better on the ResNet50 backbone while the SD-OCT dataset on ResNet18. As for the first public SD-OCT dataset, the ResNet18+SRM obtains 96.45% in the ACC and significantly obtains at least 0.48% improvement on all evaluation measures through comparisons to others. As for the second public UCSD dataset, ResNet50+SRM gets the highest ACC with 97.28%, which outperforms attention-based methods over 0.75% at least, e.g., ResNet50+CBAM (96.53%) and outperforms attention-based methods over 0.96% at least, e.g., WaveCNet (96.32%). Overall, the classification results prove the effectiveness of our method in reweighting the relative importance of different layers of the retina adaptively on OCT images.

Table 10

Performance comparison of strong baselines and the state-of-the-art methods on two public datasets.

Methods	SD-OCT				UCSD			
	ACC	AUC	F1	Kappa	ACC	AUC	F1	Kappa
Inception	0.8341	0.9520	0.8290	0.7353	0.9651	0.9955	0.9508	0.9480
ResNet18	0.8470	0.9844	0.8748	0.7695	0.9596	0.9948	0.9421	0.9395
ResNet50	0.8857	0.9772	0.8813	0.8183	0.9598	0.9953	0.9424	0.9398
TinyViT	0.8680	0.9644	0.8555	0.7882	0.9622	0.9960	0.9454	0.9434
ResMLP	0.6602	0.6944	0.6558	0.4649	0.8894	0.9701	0.8395	0.8337
ResNet18+SE	0.9243	0.9908	0.9261	0.8830	0.9360	0.9926	0.9085	0.9031
ResNet18+CBAM	0.8873	0.9862	0.8880	0.8270	0.9532	0.9953	0.9339	0.9297
ResNet18+Coord	0.8210	0.9772	0.8440	0.7336	0.9113	0.9859	0.8765	0.8645
ResNet18+Fca	0.8180	0.9772	0.8239	0.7223	0.9508	0.9942	0.9328	0.9268
ResNet18+DWA	0.9098	0.9941	0.9060	0.8583	0.9262	0.9900	0.8979	0.8884
ResNet50+SE	0.8647	0.9628	0.8562	0.7872	0.9260	0.9920	0.9097	0.8917
ResNet50+CBAM	0.8631	0.9733	0.8692	0.7901	0.9653	0.9966	0.9517	0.9486
ResNet50+Coord	0.7295	0.8888	0.7032	0.5535	0.9626	0.9961	0.9474	0.9441
ResNet50+Fca	0.8164	0.9520	0.8069	0.7080	0.9325	0.9893	0.9073	0.8998
ResNet50+DWA	0.8535	0.9609	0.8440	0.7717	0.9639	0.9963	0.9490	0.9460
WavePooling	0.6586	0.8724	0.6276	0.4620	0.8921	0.9787	0.8592	0.8099
WCNN	0.8986	0.9828	0.9040	0.8429	0.9607	0.9952	0.9445	0.9414
DAWN	0.8052	0.9768	0.8230	0.7088	0.9547	0.9946	0.9355	0.9324
WaveCNet	0.9597	0.9985	0.9577	0.9368	0.9632	0.9967	0.9471	0.9453
RIRNet	0.8084	0.9437	0.7996	0.6887	0.9556	0.9951	0.9359	0.9334
RCRNet	0.8261	0.9234	0.8083	0.7225	0.9632	0.8423	0.7803	0.7599
FITNet	0.8937	0.9761	0.9001	0.8334	0.9525	0.9951	0.9336	0.9292
ResNet18+SRM	0.9645	0.9947	0.9652	0.9949	0.9673	0.9961	0.9537	0.9514
ResNet50+SRM	0.9339	0.9910	0.9266	0.8961	0.9728	0.9969	0.9609	0.9594

6. Discussion

This study addresses a method that utilizes the DWT to extract the low-frequency information of the structure in the retinal layers and to preserve the high-frequency features while eliminating the noise. Specifically, we propose a novel SRM to reweight features extracted from the frequency domain by transforming the neglected frequency domain information into clinical prior knowledge. In this study, we present two key findings: (1) Our experimental results suggest the potential of the retina as an assessment tool for PD. (2) Visual analysis indicates that PD-related elements include not only specific layers of the retina but also the fovea's location in OCT images.

Although the WaveSRNet achieves the best classification results through comparisons to advanced methods and we also use visualization analysis methods to verify the effectiveness of our method, there are some limitations in this paper, which are summarized as follows:

1. Our network works on each B-scan individually.
2. We only focus on the content of structural information, neglecting the position characteristics.
3. We lack a joint analysis of retinal structure and function due to data sources.
4. Our approach may be affected by the traits of individuals, such as ethnicity.

First, this paper takes each image as a sample, thus the representation of retinal layers learned by the network is limited to 2D, while 3D-based images can generate and reconstruct the volumetric measurements of different structures such as the entire macula, single-layer or multi-layer retinal complex. Second, our approach focuses on content information of structures among the retinal layers, without considering the position. Existing clinical studies have proved that changes in the position of each layer of the retina are related to PD, such as quadrants, angles, or different range divisions centered on the fovea [10,50]. Third, according to the clinical research, we lack an analysis of changes in retinal structure and function, which jointly reflect the progression of PD. Fourth, our data is collected in China, thus ethnicity is a sensitive attribute in our approach.

To address the limitations mentioned above, our future work will develop a specific DL method for both 2D and 3D OCT images. We

will explore more information regarding the structure and function of retinal layers and the position for the relative position and orientation of pathologic lesions. We will dedicate ourselves to enhancing screening performance, finding that the changes in layers are related to the thickness and volume in multiple directions, and exploring the correlation between retinal structures and function.

In conclusion, we propose that using retinal images for PD screening carries the following three clinical implications: (1) Due to its advantages such as low cost compared with MRI in most countries, lack of radiation hazard, and no requirement for premedication, OCT is good accessibility in primary hospitals for large-scale screening of PD high-risk groups. (2) The WaveSRNet has the potential to be a CAD tool for screening PD, advancing the clinical precision diagnosis, and scientific research development. (3) The investigations into the relationship between pathological processes in the retina and pathological processes in the brain could provide novel insights into the pathogenesis of PD.

7. Conclusions

In this paper, we propose a novel WaveSRNet for PD screening in OCT images, which is inspired by the findings of clinical research and the effectiveness of DWT for extracting retinal structural details. It makes full use of low- and high-frequency information to recombine the discriminative wavelet features of the retina by selection and recalibration. Besides, our proposed SRM can be integrated into any CNN architecture seamlessly with negligible overheads and is end-to-end trainable along with base CNNs because of its characteristics of lightweight. Our experiments on screening PD and two public OCT datasets demonstrate that simply inserting structural details on the region of the retina into standard CNN architectures such as ResNet boosts the network's performance. Furthermore, we empirically validate the relative importance of SRM in different layers of the retina for PD screening through ablation studies and explore the network's internal behavior through visual interpretation analyses. Thus, the WaveSRNet has the potential to be a CAD tool for screening PD, advancing clinical precision diagnosis, and scientific research development.

CRediT authorship contribution statement

Jingqi Huang: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation. **Xiaoqing Zhang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Richu Jin:** Supervision, Funding acquisition. **Tao Xu:** Funding acquisition, Data curation. **Zi Jin:** Funding acquisition, Data curation. **Meixiao Shen:** Funding acquisition, Data curation. **Fan Lv:** Funding acquisition, Data curation. **Jiangfan Chen:** Supervision, Funding acquisition, Data curation. **Jiang Liu:** Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The source codes and public data to reproduce the results reported in this study can be downloaded from: <https://github.com/Jingqi-H/WaveSRNet>.

The private PD-OCT dataset will be made available on reasonable request by contacting the corresponding author (J. Liu: liuj@sustech.edu.cn).

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Grant No. 82272086, No. 62101236), the Leading Goose Program of Zhejiang (2023C03079), and the Shenzhen Natural Science Fund (JCYJ20200109140820699 and the Stable Support Plan Program 20200925174052004).

References

- [1] Valery L. Feigin, Amanuel Alemu Abajobir, Kalkidan Hassen Abate, Foad Abd-Allah, Abdishakur M. Abdulle, Semaw Ferede Abera, Gebre Yitayih Abyu, Muktar Beshir Ahmed, Amani Nidhal Aichour, Ibtiheh Aichour, et al., Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015, *Lancet Neurol.* 16 (11) (2017) 877–897.
- [2] Ronald B. Postuma, Daniela Berg, Matthew Stern, Werner Poewe, C. Warren Olanow, Wolfgang Oertel, José Obeso, Kenneth Marek, Irene Litvan, Anthony E. Lang, et al., MDS clinical diagnostic criteria for Parkinson's disease, *Mov. Disorders* 30 (12) (2015) 1591–1601.
- [3] Xinhui Xie, bin Chen, Yitian Zhao, Hao Wang, Shuxian Xu, Yan Hu, Han Rong, Jiang Liu, Initiative of eye-brain joint computing in neuropsychiatric disorders, *Chin. J. Psych.* 53 (6) (2020) 546–552.
- [4] Carol Y. Cheung, An Ran Ran, Shujun Wang, Victor T.T. Chan, Kaiser Sham, Saima Hilal, Narayanaswamy Venkatasubramanian, Ching-Yu Cheng, Charumathi Sabanayagam, Yih Chung Tham, et al., A deep learning model for detection of alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study, *Lancet Digit. Health* 4 (11) (2022) e806–e815.
- [5] Alexander Y. Lau, Vincent Mok, Jack Lee, Yuhua Fan, Jinsheng Zeng, Bonnie Lam, Adrian Wong, Chloe Kwok, Maria Lai, Benny Zee, Retinal image analytics detects white matter hyperintensities in healthy adults, *Ann. Clin. Transl. Neurol.* 6 (1) (2019) 98–105.
- [6] Carmen Harnois, T. Di Paolo, Decreased dopamine in the retinas of patients with Parkinson's disease, *Invest. Ophthalmol. Vis. Sci.* 31 (11) (1990) 2473–2475.
- [7] Jee Young Lee, Antonio Martin-Bastida, Ane Murueta-Goyena, Iigo Gabilondo, Nicolás Cuenca, Paola Piccini, Beomseok Jeon, Multimodal brain and retinal imaging of dopaminergic degeneration in Parkinson disease, *Nat. Rev. Neurol.* 18(4) (2022).
- [8] Joaquim de Moura, Gabriela Samagaio, Jorge Novo, Pablo Almuina, María Isabel Fernández, Marcos Ortega, Joint diabetic macular edema segmentation and characterization in OCT images, *J. Digit. Imaging* 33 (2020) 1335–1351.
- [9] Elena Garcia-Martin, Diego Rodriguez-Mena, Maria Satue, Carmen Almarcegui, Isabel Dolz, Raquel Alarcia, Maria Seral, Vicente Polo, Jose M. Larrosa, Luis E. Pablo, Electrophysiology and optical coherence tomography to evaluate Parkinson disease severity, *Invest. Ophthalmol. Vis. Sci.* 55 (2) (2014) 696–705.
- [10] Ane Murueta-Goyena, Rocío Del Pino, Marta Galdós, Begoña Arana, Marian Acera, Mar Carmona-Abellán, Tamara Fernández-Valle, Beatriz Tijero, Olaia Lucas-Jiménez, Natalia Ojeda, et al., Retinal thickness predicts the risk of cognitive decline in parkinson disease, *Ann. Neurol.* 89 (1) (2021) 165–176.
- [11] Elena Garcia-Martin, Jose M. Larrosa, Vicente Polo, Maria Satue, Marcia L. Marques, Raquel Alarcia, Maria Seral, Isabel Fuertes, Sofia Otin, Luis E. Pablo, Distribution of retinal layer atrophy in patients with Parkinson disease and association with disease severity and duration, *Am. J. Ophthalmol.* 157 (2) (2014) 470–478.
- [12] Nicolas M. Roth, Shiv Saidha, Hanna Zimmermann, Alexander U. Brandt, Justine Isensee, Agnieszka Benkhellouf-Rutkowska, Matthias Dornauer, Andrea A. Kühn, Thomas Müller, Peter A. Calabresi, et al., Photoreceptor layer thinning in idiopathic Parkinson's disease, *Mov. Disorders* 29 (9) (2014) 1163–1170.
- [13] Cong Wang, Meng Gan, Wavelet attention network for the segmentation of layer structures on OCT images, *Biomed. Opt. Exp.* 13 (12) (2022) 6167–6181.
- [14] Ignacio A. Viedma, David Alonso-Caneiro, Scott A. Read, Michael J. Collins, Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey, *Neurocomputing* (2022).
- [15] Xiaoming Liu, Tianyu Fu, Zhifang Pan, Dong Liu, Wei Hu, Jun Liu, Kai Zhang, Automated layer segmentation of retinal optical coherence tomography images using a deep feature enhanced structured random forests classifier, *IEEE J. Biomed. Health Inform.* 23 (4) (2018) 1404–1416.
- [16] Abdolreza Rashno, Dara D. Koozekanani, Keshab K. Parhi, Oct fluid segmentation using graph shortest path and convolutional neural network, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2018, pp. 3426–3429.
- [17] Xiaodan Sui, Yuanjie Zheng, Benzhen Wei, Hongsheng Bi, Jianfeng Wu, Xuemei Pan, Yilong Yin, Shaoting Zhang, Choroid segmentation from optical coherence tomography with graph-edge weights learned from deep convolutional neural networks, *Neurocomputing* 237 (2017) 332–341.
- [18] László Varga, Attila Kovács, Tamás Grósz, Géza Thury, Flora Hadarits, Rózsa Dégi, József Dombi, Automatic segmentation of hyperreflective foci in OCT images, *Comput. Methods Programs Biomed.* 178 (2019) 91–103.
- [19] David Alonso-Caneiro, Jason Kugelman, Jared Hamwood, Scott A. Read, Stephen J. Vincent, Fred K. Chen, Michael J. Collins, Automatic retinal and choroidal boundary segmentation in OCT images using patch-based supervised machine learning methods, in: Computer Vision—ACCV 2018 Workshops: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers 14, Springer, 2019, pp. 215–228.
- [20] Tae Keun Yoo, Joon Yul Choi, Hong Kyu Kim, Feasibility study to improve deep learning in OCT diagnosis of rare retinal diseases with few-shot classification, *Med. Biol. Eng. Comput.* 59 (2021) 401–415.
- [21] Gabriella Moraes, Dun Jack Fu, Marc Wilson, Hagar Khalid, Siegfried K. Wagner, Edward Korot, Daniel Ferraz, Livia Faes, Christopher J. Kelly, Terry Spitz, et al., Quantitative analysis of OCT for neovascular age-related macular degeneration using deep learning, *Ophthalmology* 128 (5) (2021) 693–705.
- [22] Vineeta Das, Samarendra Dandapat, Prabin Kumar Bora, Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images, *Biomed. Signal Process. Control* 54 (2019) 101605.
- [23] Long Chen, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, Tat-Seng Chua, Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5659–5667.
- [24] Jie Hu, Li Shen, Gang Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
- [25] Xiaolong Wang, Ross Girshick, Abhinav Gupta, Kaiming He, Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.
- [26] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon, Chm: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.
- [27] Xiangyu Zhao, Peng Huang, Xiangbo Shu, Wavelet-attention CNN for image classification, *Multimedia Syst.* 28 (3) (2022) 915–924.
- [28] Hadi Salman, Caleb Parks, Shi Yin Hong, Justin Zhan, WaveNets: Wavelet channel attention networks, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 1107–1113.
- [29] Yuting Yang, Licheng Jiao, Xu Liu, Fang Liu, Shuyuan Yang, Lingling Li, Puhua Chen, Xiufang Li, Zhongjian Huang, Dual wavelet attention networks for image classification, *IEEE Trans. Circuits Syst. Video Technol.* (2022).
- [30] Zequn Qin, Pengyi Zhang, Fei Wu, Xi Li, Fcanet: Frequency channel attention networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 783–792.
- [31] Yamin Ji, Laijun Sun, Yingsong Li, Dandan Ye, Detection of bruised potatoes using hyperspectral imaging technique based on discrete wavelet transform, *Infrared Phys. Technol.* 103 (2019) 103054.
- [32] Philip Chikontwe, Soo Jeong Nam, Heounjeong Go, Meejeong Kim, Hyun Jung Sung, Sang Hyun Park, Feature re-calibration based multiple instance learning for whole slide image classification, in: Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II, Springer, 2022, pp. 420–430.

- [33] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Andrea Vedaldi, Gather-excite: Exploiting feature context in convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 31 (2018).
- [34] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, Rethinking the inception architecture for computer vision, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
- [35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [36] Kan Wu, Jinnian Zhang, Houwen Peng, Mengchen Liu, Bin Xiao, Jianlong Fu, Lu Yuan, Tinyvit: Fast pretraining distillation for small vision transformers, in: *European Conference on Computer Vision*, Springer, 2022, pp. 68–85.
- [37] Hugo Touvron, Piotr Bojanowski, Mathilde Caron, Matthieu Cord, Alaaeldin El-Nouby, Edouard Grave, Gautier Izacard, Armand Joulin, Gabriel Synnaeve, Jakob Verbeek, et al., Resmlp: Feedforward networks for image classification with data-efficient training, *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (4) (2022) 5314–5321.
- [38] Travis Williams, Robert Li, Wavelet pooling for convolutional neural networks, in: *International Conference on Learning Representations*, 2018.
- [39] Qiufu Li, Linlin Shen, Sheng Guo, Zhihui Lai, Wavecnet: Wavelet integrated cnns to suppress aliasing effect for noise-robust image classification, *IEEE Trans. Image Process.* 30 (2021) 7074–7089.
- [40] Shin Fujieda, Kohei Takayama, Toshiya Hachisuka, Wavelet convolutional neural networks, 2018, *arXiv preprint arXiv:1805.08620*.
- [41] Maria Ximena Bastidas Rodriguez, Adrien Gruson, Luisa Polania, Shin Fujieda, Flavio Prieto, Kohei Takayama, Toshiya Hachisuka, Deep adaptive wavelet network, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3111–3119.
- [42] Xiaoqing Zhang, Zunjie Xiao, Huazhu Fu, Yan Hu, Jin Yuan, Yanwu Xu, Risa Higashita, Jiang Liu, Attention to region: Region-based integration-and-recalibration networks for nuclear cataract classification using AS-OCT images, *Med. Image Anal.* 80 (2022) 102499.
- [43] Xiaoqing Zhang, Zunjie Xiao, Bing Yang, Xiao Wu, Risa Higashita, Jiang Liu, Regional context-based recalibration network for cataract recognition in AS-OCT, *Pattern Recognit.* 147 (2024) 110069.
- [44] Shaobin Chen, Zhenquan Wu, Mingzhu Li, Yun Zhu, Hai Xie, Peng Yang, Cheng Zhao, Yongtao Zhang, Shaochong Zhang, Xinyu Zhao, et al., Fit-net: Feature interaction transformer network for pathologic myopia diagnosis, *IEEE Trans. Med. Imaging* 42 (9) (2023) 2524–2538.
- [45] Audrey Duran, Gaspard Dussert, Olivier Rouvière, Tristan Jaouen, Pierre-Marc Jodoin, Carole Lartizien, ProstAttention-net: A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans, *Med. Image Anal.* 77 (2022) 102347.
- [46] Mitko Veta, Yujing J. Heng, Nikolas Stathonikos, Babak Ehteshami Bejnordi, Francisco Beca, Thomas Wollmann, Karl Rohr, Manan A. Shah, Dayong Wang, Mikael Rousson, et al., Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge, *Med. Image Anal.* 54 (2019) 111–121.
- [47] Scott M. Lundberg, Su-In Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [48] Pratul P. Srinivasan, Leo A. Kim, Priyatham S. Mettu, Scott W. Cousins, Grant M. Comer, Joseph A. Izatt, Sina Farsiu, Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, *Biomed. Opt. Express* 5 (10) (2014) 3568–3577.
- [49] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C.S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, *Cell* 172 (5) (2018) 1122–1131.
- [50] Jee-Young Lee, Jeeyun Ahn, Sohee Oh, Joo Young Shin, Yu Kyeong Kim, Hyunwoo Nam, Beomseok Jeon, Retina thickness as a marker of neurodegeneration in prodromal lewy body disease, *Mov. Disorders* 35 (2) (2020) 349–354.