

SpecF2M: A Spectral-Aware Multi-task Network Estimating Axial Length and Refractive Error from Pediatric Fundus Photographs

Mengxian He^{1*}[0009–0005–7585–374X], Xinyue Liu²[0009–0003–6294–9932]^{*},
Yunyun Sun³, Hao Wei¹[0000–0002–5719–8826], Mingqing
Zhang¹[0000–0002–7214–0569], Shunyi Zhang³, Lichun Wang⁴, Wu
Yuan¹[0000–0001–9405–519X]^{**}

¹ The Chinese University of Hong Kong, Sha Tin, New Territories, Hong Kong
wyuan@cuhk.edu.hk

² Guangdong Laboratory of Machine Perception and Intelligent Computing,
Shenzhen MSU-BIT University, China

³ Capital Medical University, China

⁴ Zhengzhou Second Hospital, China

Abstract. Spherical Equivalent Refraction (SER) and Axial Length (AL) are core indicators for pediatric myopia screening, yet their measurements require dedicated biometry and cycloplegic refraction. Fundus photography offers an accessible imaging modality, as myopia-related posterior-pole changes are visible in 45° fundus images. However, these cues are often low-contrast, spatially diffuse, and multi-scale. Moreover, AL, Sphere (SPH), and Cylinder (CYL) share partially overlapping but non-identical anatomical correlates. We propose SpecF2M, a spectral-aware multi-task network for estimating AL and SER components from pediatric fundus photographs. SpecF2M integrates a deterministic anatomy-guided enhancement module, a hybrid spatial-spectral backbone combining MixCNN and Hybrid Spectral Learning (HSL) blocks, and an expert-routing head for component-level estimation of AL, SPH, and CYL. On a pediatric cohort of 4,359 eligible child visits and 6,966 fundus images, SpecF2M outperforms controlled CNN/ViT baselines for AL and SPH estimation, achieving MAEs of 0.5347 mm and 0.7062 D, respectively. Component-level analysis further reveals asymmetric task coupling, where CYL exhibits weaker association with fundus-derived myopic patterns than AL/SPH. These results support fundus-based, screening-oriented estimation of pediatric myopia indicators, while external validation remains necessary before deployment.

Keywords: Axial Length · Spherical Equivalent Refraction · Myopia.

^{*} Contributed equally to this work.

^{**} Corresponding author

1 Introduction

Myopia is a global public health burden, with a high and increasing prevalence among children [15]. Early-onset myopia tends to progress rapidly, substantially increasing the risk of vision-threatening complications [9]. Among school-aged children, axial myopia is the most prevalent subtype, characterized by excessive elongation of the ocular axis, which shifts the retinal plane posterior to the focal plane and results in a progressive myopic refractive error [12].

Accordingly, in clinical practice, Spherical Equivalent Refraction (SER) and Axial Length (AL) are critical quantitative indicators to assess myopia severity and progression risk. SER is computed as $SER = SPH + 0.5 \times CYL$ [3], where SPH (Sphere) denotes the spherical refractive component, and CYL (Cylinder) denotes the astigmatic component. AL characterizes the structural growth of the eye along the anterior-posterior axis. Consequently, myopia progression in children is largely driven by AL elongation [5], whereas CYL remains relatively stable, suggesting that SER changes are mainly attributable to the SPH [6].

Therefore, measurements of AL and SER are crucial for screening pediatric myopia and estimating its risk of progression. However, obtaining SER requires cycloplegic refraction, which is time-consuming and often uncomfortable for children, whereas AL measurement depends on dedicated devices (e.g., Lenstar900). These required workflows, equipment, manpower, and high compliance limit accessibility and scalability, hindering large-scale screening and management of pediatric myopia.

In contrast, 45° posterior-pole fundus photography (45° fundus images) is widely used in screening settings [22], offering a lower-cost, objective, and repeatable examination. Prior studies have shown that pediatric myopia progression is accompanied by characteristic posterior-pole changes that are visible on 45° fundus images [24]. As illustrated in Fig. 1, elongation of the ocular axis mechanically stretches the posterior globe, which gives rise to a set of common manifestations including (1) Optic Disc (OD) deformation or tilt [13], (2) the progression of Peripapillary Atrophy (PPA) around OD [13], (3) Fundus Tessellation (FT) due to retina-choroid thinning with increased choroidal vessel visibility [17], and (4) straighter, elongated Fundus Vessel (FV) with reduced tortuosity [7]. Together, the above changes combine global shape deformation with localized edge/texture alterations. These multi-scale features align naturally with spectral representations, which can capture myopic fundus changes more explicitly.

Recent studies have shown that SER and AL can be estimated from fundus images. Early work demonstrated DL-based refractive-error estimation and found SPH to be more predictable than CYL [21]. Other studies estimated AL from fundus photographs or UWF images and showed that AL-related cues are spatially diffuse and scale-dependent [1,11,23]. Joint AL/SER modeling has also been explored using statistical or deep-learning formulations [16,14,25]. However, existing methods mainly rely on coarse joint prediction or conventional CNN features, leaving component-aware coupling, anatomy-guided enhancement, and spectral-aware representation insufficiently explored.

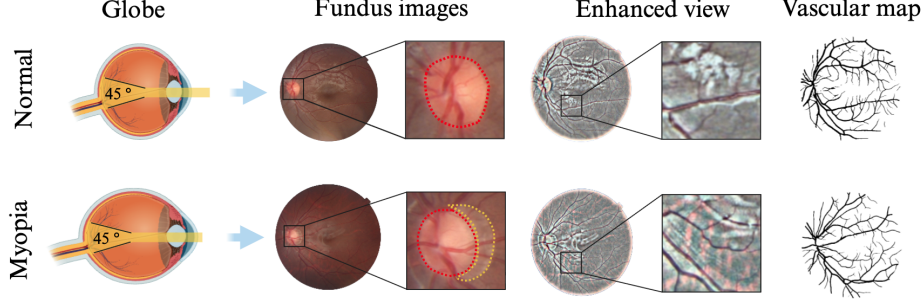


Fig. 1. Examples of a normal eye (AL 20.73 mm, SER +1.375 D) and a myopic eye (AL 26.11 mm, SER −6.375 D), showing anatomical regions related to OD, PPA, FT, and FV.

SpecF2M addresses these gaps from three perspectives:

1. **Expert-routing multi-task learning.** We formulate AL, SPH, and CYL estimation as a component-level multi-task problem and use expert routing to balance shared and task-specific representations.
2. **Anatomy-guided enhancement strategy.** We introduce a training-free anatomy-guided enhancement strategy to highlight low-contrast myopic structures, including OD/PPA, tessellation, and vessel-related patterns.
3. **Spectral-aware representation learning.** We design a spectral-aware backbone that combines local multi-resolution wavelet modeling with global Fourier-domain propagation.

To our knowledge, *SpecF2M* is the first framework to jointly estimate AL, SPH, and CYL from 45° pediatric fundus photographs, providing both screening-oriented prediction and insight into asymmetric AL/SPH/CYL coupling.

2 Problem formulation

Let $\mathcal{D} = \{(x_i, y_i^{AL}, y_i^{SPH}, y_i^{CYL})\}_{i=1}^N$ denote a pediatric fundus dataset, where x_i is a fundus image and y_i^{AL} , y_i^{SPH} , and y_i^{CYL} are the paired Axial Length (AL), Sphere (SPH), and Cylinder (CYL), respectively. We learn a multi-output regression model:

$$(\hat{y}_i^{AL}, \hat{y}_i^{SPH}, \hat{y}_i^{CYL}) = g_{\phi, \psi, \gamma}(h_{\theta}(x_i)),$$

where $h_{\theta}(\cdot)$ is a shared backbone and $g_{\phi, \psi, \gamma}(\cdot)$ denotes task-specific heads. The predicted SER is computed as $\text{SER}_{\text{pred}} = \hat{y}_{\text{SPH}} + 0.5 \times \hat{y}_{\text{CYL}}$. We explicitly predict SPH and CYL rather than using SER as an auxiliary loss because SER is linearly determined by SPH/CYL and may mask component-specific errors when their residuals have opposite signs.

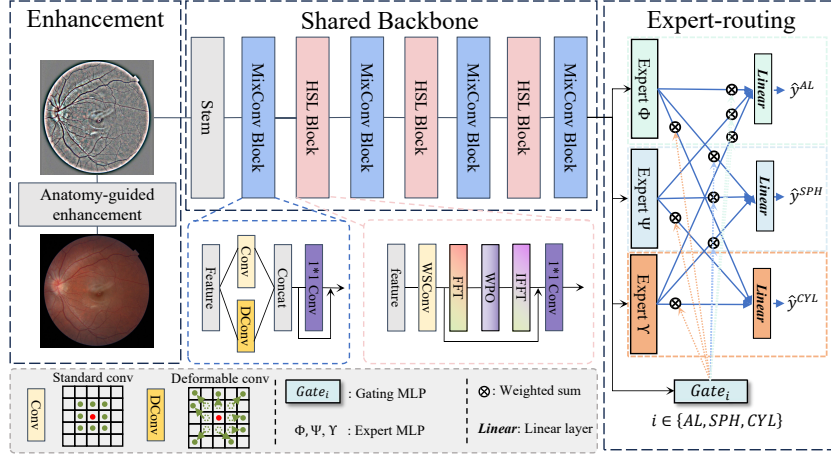


Fig. 2. The overview of proposed *SpecF2M*.

3 Methods

As shown in Fig. 2, *SpecF2M* contains three modules: anatomy-guided enhancement, a shared MixCNN–HSL backbone, and expert-routing regression heads. The enhanced image is first transformed into spatial–spectral features, which are then routed to expert-specific embeddings for AL, SPH, and CYL estimation.

3.1 Anatomy-guided enhancement strategy

As shown on the left of Fig. 2, we propose an anatomy-guided enhancement strategy for improving the discriminability of low-contrast, spatially diffuse myopic anatomical structures. Given a color fundus image $I \in \mathbb{R}^{H \times W \times 3}$, we define a deterministic enhancement operator $\mathcal{A}(\cdot)$:

$$\tilde{I} = \mathcal{A}(I) = \mathcal{F}_{\text{roi}} \left(\mathcal{F}_{\text{v-sup}} \left(\mathcal{F}_{\text{hf}} \left(\mathcal{F}_{\text{lum}}(I) \right) \right) \right). \quad (1)$$

Each component of $\mathcal{A}(\cdot)$ uses fixed hyperparameters and introduces no additional learnable parameters.

(1) Luminance enhancement ($\mathcal{F}_{\text{lum}}$). We first resize and convert I to LAB space $I^{\text{lab}} = \Phi_{\text{LAB}}(I)$, and apply CLAHE only on the luminance channel:

$$I_1 = \mathcal{F}_{\text{lum}}(I) = \Phi_{\text{LAB}}^{-1}([\text{CLAHE}(L), A, B]). \quad (2)$$

(2) High-frequency emphasis ($\mathcal{F}_{\text{hf}}$). To reduce low-frequency redundancy and highlight micro-structures, we apply an adaptive high-pass sharpening:

$$I_2 = \mathcal{F}_{\text{hf}}(I_1) = I_1 + \alpha(I_1 - G_{\sigma}(I_1)), \quad (3)$$

where $G_{\sigma}(\cdot)$ denotes Gaussian smoothing and α controls the sharpening strength.

(3) Vessel-prior soft suppression ($\mathcal{F}_{\text{v-sup}}$). Since contrast enhancement can amplify vascular linear patterns, we derive a continuous soft vessel mask $M_v \in (0, 1)^{H \times W}$ from the green channel G via multi-orientation black-hat filtering: $V = \sum_{\theta \in \Theta} \text{BH}(G; SE_\theta)$, $M_v = \sigma\left(\frac{V - \mu(V)}{s(V)}\right)$, and suppress vessel textures in a soft manner:

$$I_3 = \mathcal{F}_{\text{v-sup}}(I_2) = I_2 - \lambda(M_v \odot I_2), \quad (4)$$

which increases the relative saliency of diffuse tessellation and structural details.

(4) OD/PPA ROI enhancement with soft fusion ($\mathcal{F}_{\text{roi}}$). We obtain an optic-disc ROI mask $M_{od} = \mathcal{R}(I_3)$ by deterministic morphology and geometric fitting, enhance the ROI by a lightweight operator $\mathcal{E}(\cdot)$ (contrast redistribution + mild sharpening), and softly fuse it back to avoid boundary artifacts:

$$I_{od} = \mathcal{E}(I_3 \odot M_{od}), \quad W = G_\tau(M_{od}), \quad \tilde{I} = \mathcal{F}_{\text{roi}}(I_3) = (1 - W) \odot I_3 + W \odot I_{od}, \quad (5)$$

where $G_\tau(\cdot)$ is the ROI mask to form a continuous blending map $W \in [0, 1]$.

3.2 Spectral-aware representation learning

Myopic fundus changes are inherently multi-scale and spectrally structured, motivating a spectral-aware representation. Inspired by [4, 18], we construct a shared backbone by interleaving MixCNN blocks with HSL blocks. This design progressively injects spectral information while preserving spatial information.

In the MixCNN block, we adopt a parallel design that combines the standard convolution and deformable convolution kernels to learn complementary spatial features. Given an input feature map $\mathbf{X} \in \mathbb{R}^{B \times C \times H \times W}$, the standard branch produces $\mathbf{F}_{\text{std}} = \text{Conv}(\mathbf{X})$, which focuses on learning stable local textures and regular structural patterns. In parallel, the deformable branch applies adaptive sampling to account for geometric deformations: $\mathbf{F}_{\text{def}} = \text{DConv}(\mathbf{X})$, where the sampling offsets are learned to better align with irregular anatomical structures. The outputs are concatenated along the channel dimension and fused by a 1×1 convolution $\mathbf{F}_{\text{fuse}} = \text{Conv}_{1 \times 1}([\mathbf{F}_{\text{std}}, \mathbf{F}_{\text{def}}])$, providing a robust spatial representation for subsequent HSL blocks.

The HSL blocks further extend the representation into the spectral domain to learn multi-scale wavelet components and global frequency interactions. We construct a hybrid spectral learning block by cascading a modified [4] Wavelet-Selective CNN layer (WSConv) and a Fourier-domain global Wave Propagation Operator (WPO) [18]. We first apply an ℓ -level 2D wavelet transform (WT) per channel to obtain multi-resolution subbands. Let $\mathcal{W}(\cdot)$ and $\mathcal{W}^{-1}(\cdot)$ denote WT and inverse WT (IWT), respectively. For level $i = 1, \dots, \ell$, we decompose the current low-frequency band as

$$(\mathbf{X}_{LL}^{(i)}, \mathbf{X}_H^{(i)}) = \mathcal{W}(\mathbf{X}_{LL}^{(i-1)}), \quad \mathbf{X}_{LL}^{(0)} = \mathbf{X}, \quad (6)$$

where $\mathbf{X}_H^{(i)}$ stacks the high-frequency subbands $\{LH, HL, HH\}$. We then perform subband-specific convolutions:

$$\mathbf{Y}_{LL}^{(i)} = \text{Conv}_{LL}^{(i)}(\mathbf{X}_{LL}^{(i)}), \quad \mathbf{Y}_H^{(i)} = \text{Conv}_H^{(i)}(\mathbf{X}_H^{(i)}). \quad (7)$$

Given that enhancement attenuates LL content, we compute a lightweight gate g from pooled LH responses via an MLP and sigmoid activation. We subsequently use the LH-guided gate to modulate the LL response and add it back as a residual. Following the linearity of IWT, WSCnv aggregates outputs across levels via a recursive reconstruction:

$$\mathbf{Z}^{(\ell+1)} = \mathbf{0}, \mathbf{Z}^{(i)} = \mathcal{W}^{-1}(\mathbf{Y}_{LL}^{(i)} + \mathbf{Z}^{(i+1)}, \mathbf{Y}_H^{(i)}), \quad (8)$$

where $i = \ell, \dots, 1$, and produces the feature: $\mathbf{U}_0 = \text{WSCnv}(\mathbf{X})$.

We next propagate $\mathbf{U}_0$ globally with WPO. Let $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ denote the 2D FFT/IFFT applied per channel. WPO evolves $\mathbf{U}_0$ over the internal time t under the underdamped wave dynamics in the frequency domain:

$$\mathbf{U}_t = \mathcal{F}^{-1} \left\{ e^{-\frac{\alpha}{2}t} \left[\mathcal{F}(\mathbf{U}_0) \cos(\omega_d t) + \frac{\sin(\omega_d t)}{\omega_d} \left(\mathcal{F}(\mathbf{V}_0) + \frac{\alpha}{2} \mathcal{F}(\mathbf{U}_0) \right) \right] \right\}, \quad (9)$$

where α is the damping coefficient, v is the wave speed (encoded in ω_d), and $\mathbf{V}_0$ is the initial velocity. In practice, we take the real part after IFFT: $\tilde{\mathbf{U}} = \Re(\mathbf{U}_t)$, and set $\alpha = 0.1$ and $v = 1$ as the starting point, and set t according to the depth of the HSL block. Finally, we fuse the globally propagated feature with the WSCnv output via a skip connection and a 1×1 convolution: $\mathbf{Y} = \text{Conv}_{1 \times 1}(\tilde{\mathbf{U}} + \mathbf{U}_0)$.

In this work, HSL is treated as an integrated spectral block that combines local multi-resolution subband modeling and global frequency-domain propagation; our empirical claim is the effectiveness of this coupled design rather than complete causal attribution of each individual spectral operator.

3.3 Expert-routing multi-task learning

We formulate AL together with SPH and CYL as a SER-component-level multi-task setup.

Given an input enhanced fundus image $\tilde{I}$, the preceding modules output a shared feature representation $\mathbf{Y}$. On top of $\mathbf{Y}$, we adopt the experts routing scheme to balance shared representation learning and task-specific disentanglement for three regression tasks: $\mathcal{T} = \{\text{AL}, \text{SPH}, \text{CYL}\}$.

We instantiate three experts $\{g_\phi, g_\psi, g_\gamma\} \in g_{\phi, \psi, \gamma}$ that transform the shared feature into expert-specific embeddings: $\mathbf{e}_k = g_k(\mathbf{Y})$, $k \in \{\phi, \psi, \gamma\}$. where $\mathbf{e}_k \in \mathbb{R}^d$ denotes the expert feature. All experts take the same input $\mathbf{Y}$ but are parameterized independently, allowing diversity.

For each task $t \in \mathcal{T}$, a dedicated gate G_t predicts routing weights over experts: $\boldsymbol{\alpha}_t = \text{softmax}(G_t(\mathbf{Y})) \in \mathbb{R}^3$, $\sum_k \alpha_{t,k} = 1$. The routed task feature is obtained via a convex combination of expert embeddings: $\mathbf{z}_t = \sum_{k \in \{\phi, \psi, \gamma\}} \alpha_{t,k} \mathbf{e}_k$.

Each routed feature $\mathbf{z}_t$ is fed into a lightweight linear regressor to produce the estimation: $\hat{y}_t = \mathbf{w}_t^\top \mathbf{z}_t + b_t, t \in \mathcal{T}$, yielding $(\hat{y}_i^{AL}, \hat{y}_i^{SPH}, \hat{y}_i^{CYL})$.

We optimize a weighted multi-task regression objective:

$$\mathcal{L} = \sum_{t \in \mathcal{T}} \lambda_t \ell(\hat{y}_t, y_t), \quad (10)$$

where $\ell(\cdot)$ is a regression loss and λ_t balances task contributions.

4 Experiments

The data were collected from a pediatric myopia cohort across seven primary schools, including 4,359 eligible child visits. After excluding images with unclear optic disc, strong reflection, or artifacts, 6,966 fundus images with paired AL and cycloplegic refraction measurements were included. The mean age was 8.22 ± 1.09 years; mean AL, SPH, and CYL were 23.88 ± 1.07 mm, -0.19 ± 1.85 D, and -0.83 ± 0.95 D, respectively. Cycloplegia was induced with 1% cyclopentolate hydrochloride eye drops before autorefraction. The study was approved by TRECKY2019-058. Images and clinical measurements were de-identified. Written informed consent was provided by parents or legal guardians, and verbal assent was obtained from children when applicable.

4.1 Evaluation setup

Evaluation. We split the data by subject ID to avoid inter-eye leakage, keeping both eyes of the same subject in the same split. Training, validation, and test sets followed an 80/10/10 ratio and were stratified by myopia level. Performance was evaluated using Mean Absolute Error (MAE) and R^2 for SPH, CYL, and AL.

Baseline. We compared a range of representative baseline methods covering the ViT and CNN families in both single-task and standard hard multi-task settings and reported the best performance.

Implementation. SpecF2M was trained on an NVIDIA RTX 4090D GPU for 150 epochs with a batch size of 32, a learning rate of 10^{-4} , and a weight decay of 3×10^{-5} . We used MSE loss with task weights AL:SPH:CYL = 0.5:1:1 and geometric augmentations including random rotation, flipping, and mirroring. WConv used a Daubechies-1 wavelet with a 2-level decomposition. Each expert MLP was a one-layer 512-d MLP with ReLU and dropout 0.2.

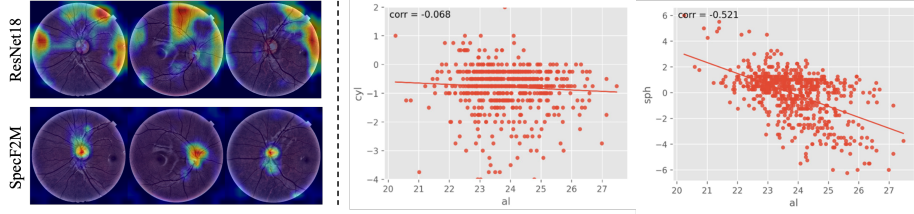
4.2 Main results

Table 1 compares SpecF2M with controlled CNN/ViT baselines and ablations. SpecF2M achieves the best AL and SPH estimation performance, with MAE/ R^2 of 0.5347/0.6038 for AL and 0.7062/0.7251 for SPH.

Grad-CAM visualizations in Fig. 3 show that SpecF2M produces more compact responses around the OD and adjacent vascular structures than ResNet18.

Table 1. Comparison with baselines and ablation study of *SpecF2M*.

Method	SPH		CYL		AL	
	MAE	R^2	MAE	R^2	MAE	R^2
ResNet18 [8]	0.8216	0.6445	0.5777	0.3995	0.8140	0.6029
DenseNet121 [10]	0.7678	0.6607	0.5673	0.4030	0.5903	0.5469
VGG16-BN [19]	0.7715	0.6662	0.5527	0.4216	0.5646	0.5689
ViT-S/16 [20]	0.9101	0.5132	0.6228	0.1222	0.7618	0.2498
ViT-B/16 [2]	0.7710	0.6647	0.5178	0.4887	0.5787	0.5669
<i>SpecF2M</i> (Our)	0.7062	0.7251	0.5508	0.4632	0.5347	0.6038
w/o Enhanced	0.7547	0.6704	0.5433	0.4676	0.5703	0.5457
w/o HSL block	0.8233	0.6341	0.5596	0.4505	0.6053	0.5134
w/o Expert-routing	0.8674	0.5601	0.5295	0.3634	0.6202	0.4958

**Fig. 3.** The Grad-CAM and task correlation analysis.

Ablations show that enhancement, HSL, and expert routing consistently benefit AL and SPH, whereas CYL changes are less stable. This is consistent with the correlation analysis in Fig. 3: AL is correlated with SPH ($\text{corr}=-0.521$) but nearly uncorrelated with CYL ($\text{corr}=-0.068$), suggesting weaker fundus-derived coupling for CYL.

5 Discussion and Conclusion

We propose SpecF2M for screening-oriented joint estimation of pediatric AL and refraction components from widely available 45° fundus photographs, highlighting its potential as an accessible image-based decision-support approach for pediatric myopia assessment. While *SpecF2M* consistently outperforms controlled CNN/ViT baselines and ablations at AL and SPH estimation, the component-wise results further reveal an asymmetric coupling: AL and SPH benefit more reliably from multi-task learning, supporting stronger shared anatomical correlates and cross-task synergy, whereas CYL remains noisier and more variable, indicating higher stochasticity from the estimation based on fundus patterns.

Acknowledgements

This work was supported in part by the Research Grants Council (RGC) of Hong Kong SAR (GRF14213125, GRF14201824, GRF14216222), and the Innovation and Technology Fund (ITF) of Hong Kong SAR (ITS/252/23).

Disclosure of Interests

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

References

1. Dong, L., Hu, X.Y., Yan, Y.N., Zhang, Q., Zhou, N., Shao, L., Wang, Y.X., Xu, J., Lan, Y.J., Li, Y., et al.: Deep learning-based estimation of axial length and subfoveal choroidal thickness from color fundus photographs. *Frontiers in cell and developmental biology* **9**, 653692 (2021)
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)
3. Enaholo, E.S., Musa, M.J., Zeppieri, M.: The spherical equivalent (2023)
4. Finder, S.E., Amoyal, R., Treister, E., Freifeld, O.: Wavelet convolutions for large receptive fields. In: *European conference on computer vision*. pp. 363–380. Springer (2024)
5. Flitcroft, D.I., He, M., Jonas, J.B., Jong, M., Naidoo, K., Ohno-Matsui, K., Rahi, J., Resnikoff, S., Vitale, S., Yannuzzi, L.: Imi-defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. *Investigative ophthalmology & visual science* **60**(3), M20–M30 (2019)
6. Harvey, E.M., Miller, J.M., Twelker, J.D., Sherrill, D.L.: Longitudinal change and stability of refractive, keratometric, and internal astigmatism in childhood. *Investigative Ophthalmology & Visual Science* **56**(1), 190–198 (2015)
7. He, H.L., Liu, Y.X., Liu, H., Zhang, X., Song, H., Xu, T.Z., Fang, Y., Ma, Y., Ren, H.Y., Ling, S.G., et al.: Deep learning-enabled vasculometry depicts phased lesion patterns in high myopia progression. *Asia-Pacific Journal of Ophthalmology* **13**(4), 100086 (2024)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
9. Hu, Y., Ding, X., Guo, X., Chen, Y., Zhang, J., He, M.: Association of age at myopia onset with risk of high myopia in adulthood in a 12-year follow-up of a chinese cohort. *JAMA ophthalmology* **138**(11), 1129–1134 (2020)
10. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4700–4708 (2017)
11. Jeong, Y., Lee, B., Han, J.H., Oh, J.: Ocular axial length prediction based on visual interpretation of retinal fundus images via deep neural network. *IEEE Journal of Selected Topics in Quantum Electronics* **27**(4), 1–7 (2020)

12. Jonas, J.B., Spaide, R.F., Ostrin, L.A., Logan, N.S., Flitcroft, I., Panda-Jonas, S.: Imi—nonpathological human ocular tissue changes with axial myopia. *Investigative Ophthalmology & Visual Science* **64**(6), 5–5 (2023)
13. Kim, T.W., Kim, M., Weinreb, R.N., Woo, S.J., Park, K.H., Hwang, J.M.: Optic disc change with incipient myopia of childhood. *Ophthalmology* **119**(1), 21–26 (2012)
14. Li, Y., Huang, Q., Zhong, C., Yang, D., Li, M., Welsh, A.H., Liu, A., Fu, B., Liu, C.C., Zhou, X.: Oucopula: Bi-channel multi-label copula-enhanced adapter-based cnn for myopia screening based on ou-uwf images. *arXiv preprint arXiv:2403.11974* (2024)
15. Morgan, I.G., French, A.N., Ashby, R.S., Guo, X., Ding, X., He, M., Rose, K.A.: The epidemics of myopia: aetiology and prevention. *Progress in retinal and eye research* **62**, 134–149 (2018)
16. Oh, R., Lee, E.K., Bae, K., Park, U.C., Yu, H.G., Yoon, C.K.: Deep learning-based prediction of axial length using ultra-widefield fundus photography. *Korean journal of ophthalmology: KJO* **37**(2), 95 (2023)
17. Ohno-Matsui, K., Kawasaki, R., Jonas, J.B., Cheung, C.M.G., Saw, S.M., Verhoeven, V.J., Klaver, C.C., Moriyama, M., Shinohara, K., Kawasaki, Y., et al.: International photographic classification and grading system for myopic maculopathy. *American journal of ophthalmology* **159**(5), 877–883 (2015)
18. Shu, Z., Wu, J., Yan, W., Liu, X., Zhang, H., Liu, C., Mao, Y., Chen, J.: Waveformer: Frequency-time decoupled vision modeling with wave equation. *arXiv preprint arXiv:2601.08602* (2026)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
20. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H.: Training data-efficient image transformers & distillation through attention. In: *International conference on machine learning*. pp. 10347–10357. PMLR (2021)
21. Varadarajan, A.V., Poplin, R., Blumer, K., Angermueller, C., Ledsam, J., Chopra, R., Keane, P.A., Corrado, G.S., Peng, L., Webster, D.R.: Deep learning for predicting refractive error from retinal fundus images. *Investigative ophthalmology & visual science* **59**(7), 2861–2868 (2018)
22. Williams, G.A., Scott, I.U., Haller, J.A., Maguire, A.M., Marcus, D., McDonald, H.R.: Single-field fundus photography for diabetic retinopathy screening: a report by the american academy of ophthalmology. *Ophthalmology* **111**(5), 1055–1062 (2004)
23. Yang, C.N., Chen, W.L., Yeh, H.H., Chu, H.S., Wu, J.H., Hsieh, Y.T.: Convolutional neural network-based prediction of axial length using color fundus photography. *Translational Vision Science & Technology* **13**(5), 23–23 (2024)
24. Yui, F., Bernabeu, M.O., Dhillon, B., Strang, N., MacGillivray, T.: Retinal changes from hyperopia to myopia: not all diopters are created equal. *Investigative Ophthalmology & Visual Science* **65**(5), 25–25 (2024)
25. Zhong, C., Li, Y., Yang, D., Li, M., Zhou, X., Fu, B., Liu, C.C., Welsh, A.H.: Cecnn: Copula-enhanced convolutional neural networks in joint prediction of refraction error and axial length based on ultra-widefield fundus images. *The Annals of Applied Statistics* **19**(2), 1292–1313 (2025)