Self-Supervised Vision Transformers in Diabetic Retinopathy: OCT Analysis Without Labeled Data

Hilal Hacıoğlu[1], Gizem Elif Atlı[2], Furkan Kırık[2], Hakan Özdemir[2]

[1]Bezmialem Vakif University, Faculty of Medicine, Istanbul, Türkiye

[2]Bezmialem Vakif University, Faculty of Medicine, Department of Ophthalmology, Istanbul, Türkiye

Introduction: Artificial intelligence (AI) models for detecting diabetic retinopathy (DRP) in optical coherence tomography (OCT) images often require large labeled datasets, a time-consuming and labor-intensive process. Self-supervised Vision Transformers (SS-ViT) address this by training directly on unlabeled OCT images, minimizing manual labeling. SS-ViT leverages transformer-based attention to localize subtle pathologies (e.g., microaneurysms) by focusing on discriminative spatial features in unlabeled datasets. This study evaluates SS-ViT's efficacy in analyzing OCT findings from DRP patients.

Methods: Standardized 19-slice OCT images from 43 DRP patients were processed using an open-access Vision Transformer (ViT) model developed by Facebook AI, trained via the DINO (self-distillation with no labels) method. The model generated attention maps for each slice, highlighting discriminative features. Two ophthalmologists evaluated eight pathologies (e.g., cystoid macular edema, vitreomacular traction) and false-positive artifacts. Sensitivity per pathology was calculated as the ratio of correctly flagged images to total positive cases.

Results: The model analyzed 815 OCT images. Sensitivities were: cystoid macular edema 84.6% (269/318), cystoid degeneration 26.5% (13/62), serous macular detachment 33.3% (10/30), posterior hyaloid 72.1% (88/122), vitreomacular traction 64.5% (20/31), hyperreflective spots 81.9% (86/105), exudates 93.0% (80/86). False-positive rate was 93.0% (80/86). Overall sensitivity across all pathologies was 75.2%. Specificity was not assessed due to non-lesion-specific attention maps.

Discussion and Conclusion: To our knowledge, this is the first study evaluating SS-ViT on DRP OCT images. While high sensitivity for exudates (93.0%) and cystoid macular edema (84.6%) suggests clinical potential, poor performance in detecting cystoid degeneration (26.5%) and serous macular detachment (33.3%) highlights algorithmic limitations. The high false-positive rate (93.0%) necessitates clinician verification for reliable deployment. Future work should refine the model to improve accuracy across all pathologies.

Keywords: AI, Diabetic Retinopathy, Optical Coherence Tomography, Self-supervised Vision Transformers

Figures

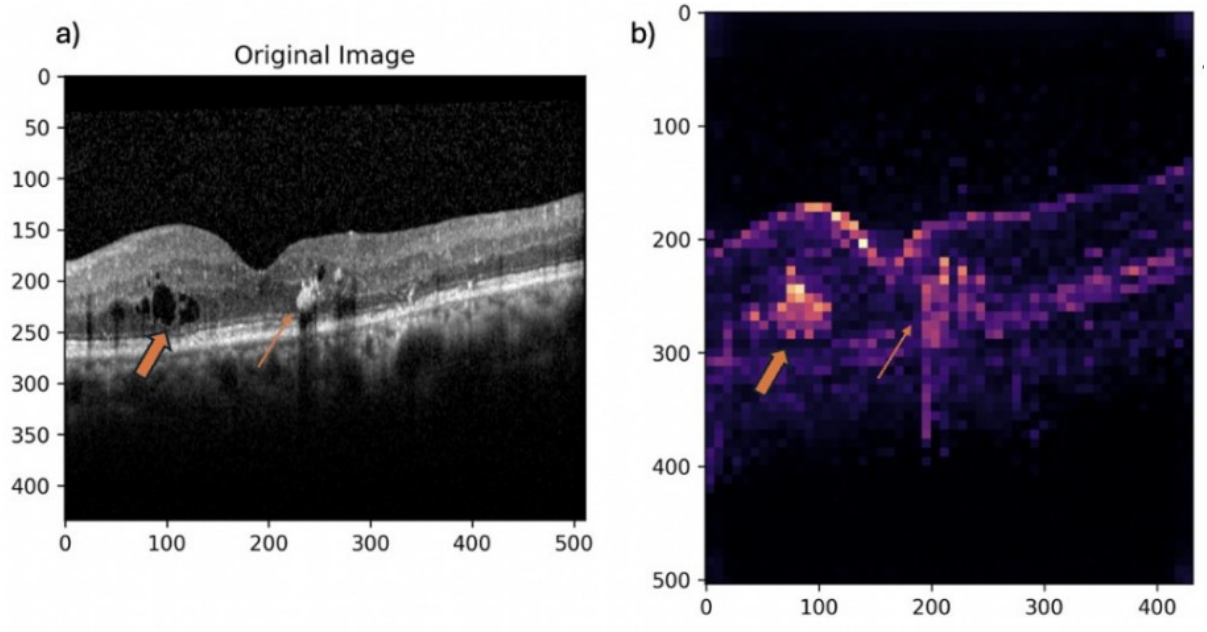

Figure 1: Cystoid Macular Edema and Exudates on OCT Image and SS-ViT Model Detection

Panel (a): Cystoid macular edema (thick arrow) and exudates (thin arrow) on OCT image.
Panel (b): Signal visualization of cystoid macular edema (thick arrow) and exudates (thin arrow) detected by the SS-ViT model.

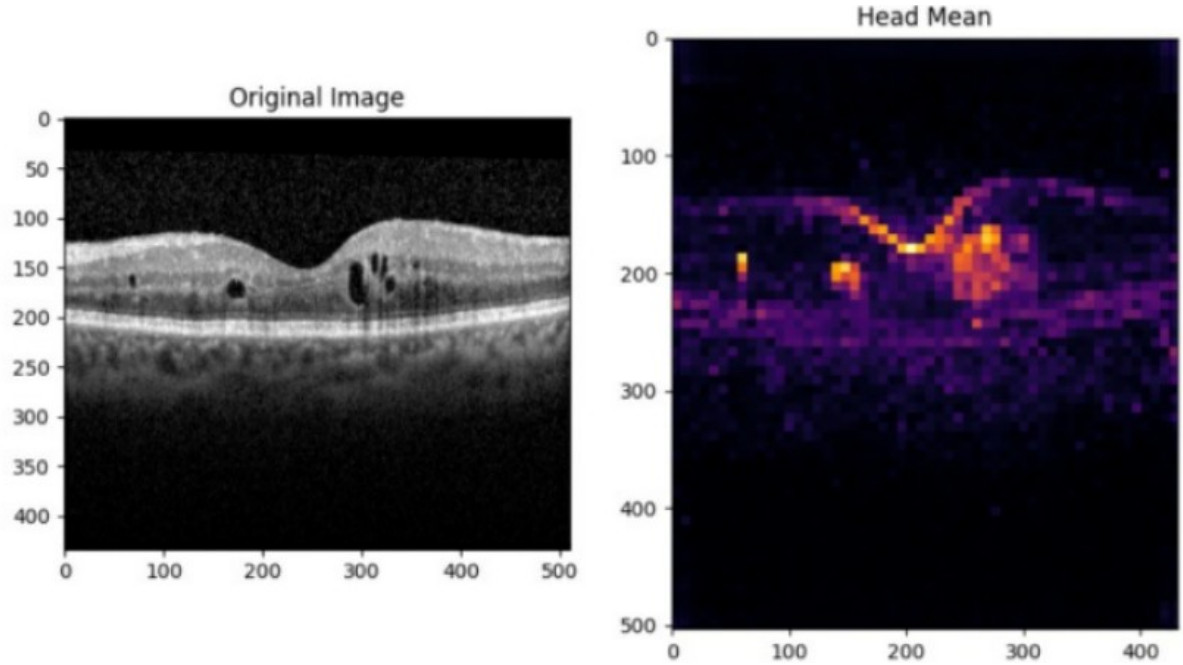

Figure 2: Diabetic Macula Edema Detection with SS-ViT Model

Left panel: Corresponding OCT (Optical Coherence Tomography) image.
Right panel: SS-ViT-generated image highlighting enhanced signal intensity in edematous tissues associated with diabetic retinopathy.