

**Title:** Multi-Head Self-Attention Mechanisms in Vision Transformers for Retinal Image Classification

**Author:** Oahidul Islam; Kowshik Kumer; Sumaia Akter; Md Mohi Uddin

**Abstract:** Diabetic Retinopathy (DR) is a serious condition that can lead to eyesight loss in those who have diabetes if not identified early. The traditional diagnosis approach relies on the manual scrutiny of retinal images by specialists, a process that can be both time-consuming and susceptible to errors. Artificial intelligence (AI) has the potential to revolutionize this field by providing automated, precise, and rapid identification of diabetic retinopathy (DR). In our study, we utilized Vision Transformers (ViTs) to achieve effective classification of retinal images. ViTs are advanced models that divide images into patches and utilize multi-head self-attention approaches to identify important features. The model demonstrated outstanding achievement, with a success rate of 96.13%, precision of 0.92, recall of 0.95, F1 score of 0.93, and ROC- AUC score of 0.969. The results of this study have significant implications for improving the detection of DR, enabling timely intervention, and potentially safeguarding the vision of a large number of individuals. Future research will focus on improving the resilience of the model by incorporating a wider range of datasets and optimizing its integration with medical procedures to ensure dependable and efficient performance in real-world scenarios.

**Keywords:** Vision Transformer, Retinal Image Classification, Deep Learning, Diabetic Retinopathy, Medical Diagnostics, Fundus Image, Classification

**DOI:** <https://doi.org/10.1109/COMPAS60761.2024.10795956>

<https://ieeexplore.ieee.org/document/10795956>