Eye Disease Image classification using Vision Transformer

Jayakrishna Vuppalapati
Praveenkumar Kavali
Krishna Prithvi Battula
Tinku Rao Kotha

Timely and accurate identification and diagnosis of retinal diseases are essential to prevent eventual vision loss, which may be temporary or permanent. Here, we present a Vision Transformer model, called retinal ViT, which improves medical image processing using the self-attention mechanism. The key purpose of this study is to determine the transformer-based model to substitute convolutional modules used in CNN models while matching or surpassing their performance. Retinal ViT, differs from traditional CNN-based convolutional neural networks through its usage of self-attention to capture longer dependencies in retinal images. This approach builds on transformer's capacity to identify complicated patterns and connections over spatial axes, reportedly increasing the model's discriminating power across retinal conditions. Ultimately, the retinal ViT architecture is framed around a multi-class classification into which the predictor feeds and assigns a given retinal image to one of many illnesses based on presence. This classification consists of simplified activations via sigmoid functions to predict multiple tags concurrently. The classification psyches use the capabilities of deep neural network's capacity for expressive power to predict retinal disorders with high confidence. In our experiments, extensive analysis of the ViT's performance will be done on the most used public dataset, which contains a wide variety of retinal images. Ultimately, substantial comparisons with contemporary CNN models were done to illustrate comparative metrics for both models. In summary, the proposed retinal ViT approach will show promising hope for medical analysis of images with a signal of how the self-attention capabilities foster potential methodologies to improve transformational model performance for data perception and management functions. Thus, this study lays a foundation for future enhancements concerning self-attention for accurate and efficient detection and treatment of retinal disorders.