

INTISARI

Gangguan retina seperti Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), dan Drusen merupakan kondisi yang sering menyebabkan penurunan penglihatan yang memerlukan deteksi dini dan akurat. *Optical Coherence Tomography* (OCT) adalah salah satu modalitas pencitraan yang luas digunakan dalam diagnosis kelainan retina karena mampu menampilkan struktur lapisan retina secara detail. Pendekatan dengan kecerdasan buatan berupa *deep learning* telah banyak dikembangkan untuk klasifikasi citra OCT secara otomatis. Namun, perbedaan arsitektur model dan strategi pelatihan dapat menyebabkan variasi dalam kinerja dan kemampuan generalisasi, sehingga diperlukan analisis untuk memahami pengaruhnya terhadap klasifikasi citra OCT retina.

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja Convolutional Neural Network (CNN) dan Vision Transformer (ViT) dalam klasifikasi citra OCT retina. Arsitektur ResNet-50 digunakan sebagai representasi CNN, sedangkan ViT-Base16 digunakan sebagai representasi Vision Transformer. Dua strategi pelatihan diterapkan, yaitu *training from scratch* dan *transfer learning* dengan skema *full fine-tuning*.

Pada hasil eksperimen, model ResNet-50 yang dilakukan *fine-tuning* mencapai akurasi sebesar 79%, lebih tinggi dibandingkan konfigurasi *training from scratch* yang hanya mencapai 73%. Sementara itu, model ViT-Base16 dengan *fine-tuning* menghasilkan akurasi tertinggi, yaitu sekitar 82%, sedangkan pendekatan *training from scratch* hanya mencapai 68%. Temuan ini menunjukkan potensi Vision Transformer melalui performa ViT-Base16 yang melampaui ResNet-50 pada konfigurasi *transfer learning*, dan *fine-tuning* memberikan peningkatan kinerja yang signifikan dibandingkan pelatihan dari awal pada kedua arsitektur.

Kata kunci : *Optical Coherence Tomography*, *Vision Transformer*, *Convolutional Neural Network*, *Transfer Learning*, Klasifikasi Citra Medis

ABSTRACT

Retinal disorders such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen are conditions that commonly could be the causes of visual impairment which require early and accurate detection. Optical Coherence Tomography (OCT) is widely used for retinal disease diagnosis due to its ability to provide detailed visualization of retinal layer structures. Deep learning-based approaches have been increasingly applied to automate OCT image classification, with variations in model architecture and training strategies may lead to different levels of performance and generalization. Therefore, further investigation is needed to understand how different deep learning architectures and training strategy perform retinal OCT image classification.

This study aims to analyze and compare the performance of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for retinal OCT multi-class diseases image classification. ResNet-50 is employed as the representative CNN architecture, while ViT-Base16 represents the Vision Transformer model. Two training strategies are evaluated on both model, which is training from scratch and transfer learning with full fine-tuning, to assess their impact on model performance.

The experimental results show that transfer learning significantly improves classification performance for both architectures. Fine-tuned ResNet-50 achieves an accuracy of 79%, outperforming the training-from-scratch configuration with an accuracy of 73%. Meanwhile, the fine-tuned ViT-Base16 model achieves the highest accuracy of approximately 82%, whereas the training-from-scratch approach reaches only 68%. These findings demonstrate the ViT potential over CNN as ViT-Base outperforms ResNet-50 under the transfer learning configuration, and that full fine-tuning provides a substantial advantage over training from scratch for both architectures.

Keywords : Optical Coherence Tomography, Vision Transformer, Convolutional Neural Network, Transfer Learning, Medical Image Classification