

ABSTRACT

Brain tumors represent a significant global health challenge, with current diagnostic techniques often being time-consuming and resource-intensive. Deep learning-based automation has emerged as a promising solution for facilitating early detection and improving diagnostic efficiency. Among various approaches, the Vision Transformer (ViT) has shown strong potential for medical image classification; however, its high computational cost remains a limitation. To address this, we propose a modified ViT architecture that reduces the number of encoder blocks while enhancing the classification head with additional layers. Experimental results demonstrate that the optimized configuration with nine encoder blocks achieves superior performance, with an accuracy of 99.39%, precision of 99.36%, recall of 99.33%, and an F1-score of 99.34%. These findings highlight that greater architectural complexity does not necessarily translate into improved classification capability. By integrating gamma correction, reducing encoder blocks, and refining the classification head, the proposed model maintains state-of-the-art accuracy while significantly reducing computational demands. This makes the approach particularly suitable for deployment in resource-constrained medical environments.

Keywords: *Brain tumor detection, Vision Transformer, Gamma correction*

ABSTRAK

Tumor otak merupakan tantangan kesehatan global yang signifikan, dengan teknik diagnostik saat ini seringkali memakan waktu dan memerlukan sumber daya yang besar. Otomatisasi berbasis deep learning telah muncul sebagai solusi menjanjikan untuk memfasilitasi deteksi dini dan meningkatkan efisiensi diagnostik. Di antara berbagai pendekatan, Vision Transformer (ViT) telah menunjukkan potensi yang kuat untuk klasifikasi gambar medis; namun, biaya komputasi yang tinggi tetap menjadi batasan. Untuk mengatasi hal ini, penelitian ini mengusulkan arsitektur ViT yang dimodifikasi dengan mengurangi jumlah blok encoder sambil meningkatkan kepala klasifikasi dengan lapisan tambahan. Hasil eksperimen menunjukkan bahwa konfigurasi yang dioptimalkan dengan sembilan blok encoder mencapai kinerja superior, dengan akurasi 99,39%, presisi 99,36%, recall 99,33%, dan skor F1 99,34%. Temuan ini menyoroti bahwa kompleksitas arsitektur yang lebih tinggi tidak selalu berarti kemampuan klasifikasi yang lebih baik. Dengan mengintegrasikan koreksi gamma, mengurangi blok encoder, dan menyempurnakan kepala klasifikasi, model yang diusulkan mempertahankan akurasi terdepan sambil secara signifikan mengurangi tuntutan komputasi. Hal ini membuat pendekatan ini sangat cocok untuk diterapkan di lingkungan medis dengan sumber daya terbatas.

Kata Kunci: Klasifikasi Tumor Otak, *Vision Transformer*, *Gamma correction*