

ABSTRAK

Melanoma adalah kanker kulit mematikan yang membutuhkan deteksi dini dengan presisi tinggi. Diagnosis visual menggunakan citra dermoskopi seringkali subjektif dan memakan waktu, sehingga arsitektur *Vision Transformer* (ViT) hadir sebagai solusi klasifikasi cerdas. Namun, implementasi ViT standar pada dataset medis berskala kecil rentan mengalami kegagalan generalisasi akibat ketiadaan bias induktif (*inductive bias*). Hal ini memicu *overfitting*, di mana model hanya menghafal citra mentah alih-alih mengekstraksi fitur patologis melanoma sesungguhnya. Selain itu, ketidakseimbangan kelas pada data medis sering menghasilkan tingginya angka *false negative* atau pasien kanker diprediksi sehat yang berisiko fatal. Oleh karena itu, diperlukan skenario optimasi pelatihan yang terfokus untuk mengatasi kelemahan mendasar ViT pada dataset terbatas guna menjamin keamanan implementasi klinis.

Penelitian ini mengusulkan strategi *Optimized Fine-Tuning* pada arsitektur *Vision Transformer* untuk klasifikasi melanoma. Akuisisi data menggunakan citra dermoskopi dari ISIC Archive dengan kelas *benign* dan *malignant*. Pra-pemrosesan data mencakup interpolasi ukuran dan normalisasi z-score. Pengembangan sistem dilakukan melalui evaluasi komparatif tiga skenario: *Linear Probing* (menahan bobot *backbone*), *Standard Fine-Tuning* (membuka seluruh lapisan tanpa regulasi), dan *Optimized Fine-Tuning* sebagai usulan utama. Skenario optimasi ini menerapkan kombinasi augmentasi berlapis (*RandAugment*, *Mixup*, *CutMix*) untuk memanipulasi ruang fitur secara dinamis. Guna mencegah penghafalan data, diterapkan regularisasi ketat melalui *Dropout*, pengoptimal AdamW dengan *Weight Decay*, dan *Label Smoothing*. Sistem ini juga mengintegrasikan fungsi kerugian *Weighted Cross-Entropy* dengan memberikan penalti lebih besar pada kelas minoritas untuk mengeliminasi bias prediksi.

Hasil pengujian menunjukkan bahwa skenario *Optimized Fine-Tuning* efektif mengatasi *overfitting* dan memberikan kinerja klasifikasi paling unggul. Skenario ini mencapai akurasi 92,57 persen, presisi 92,25 persen, dan F1-Score 92,57 persen pada data uji. Lebih lanjut, model secara kritis mampu meningkatkan sensitivitas (*recall*) menjadi 91,33 persen, yang berdampak langsung pada penurunan drastis jumlah kesalahan *false negative* menjadi hanya 26 kasus. Hasil empiris ini membuktikan bahwa kelemahan bawaan *Vision Transformer* pada data terbatas dapat diatasi secara tuntas melalui intervensi augmentasi dan regularisasi terstruktur. Kontribusi utama penelitian ini adalah terciptanya kerangka pelatihan klasifikasi yang tangguh untuk dataset medis berskala kecil. Sistem ini tidak sekadar mengejar tingginya akurasi matematis, melainkan mengedepankan kelayakan operasional klinis dengan menekan sekecil mungkin risiko fatal akibat kesalahan diagnosis.

Kata Kunci: *Vision Transformer*, Melanoma, *Fine-Tuning*, Augmentasi Data, Regularisasi

ABSTRACT

Melanoma is a deadly skin cancer that requires early detection with high precision. Visual diagnosis using dermoscopic images is often subjective and time-consuming; hence, the Vision Transformer (ViT) architecture emerges as an intelligent classification solution. However, the standard implementation of ViT on small-scale medical datasets is prone to generalization failure due to the lack of inductive bias. This triggers overfitting, where the model merely memorizes raw images instead of extracting the actual pathological features of melanoma. Furthermore, class imbalance in medical data often results in a high false negative rate (cancer patients predicted as healthy), which carries fatal risks. Therefore, a focused training optimization scenario is required to overcome the fundamental weaknesses of ViT on limited datasets to ensure the safety of clinical implementation.

This research proposes an Optimized Fine-Tuning strategy on the Vision Transformer architecture for melanoma classification. Data acquisition utilized dermoscopic images from the ISIC Archive, categorized into benign and malignant classes. Data preprocessing included size interpolation and z-score normalization. System development was conducted through a comparative evaluation of three scenarios: Linear Probing (freezing backbone weights), Standard Fine-Tuning (unfreezing all layers without regulation), and Optimized Fine-Tuning as the main proposition. This optimization scenario applied a combination of layered augmentation (RandAugment, Mixup, CutMix) to dynamically manipulate the feature space. To prevent data memorization, strict regularization was implemented through Dropout, the AdamW optimizer with Weight Decay, and Label Smoothing. The system also integrated a Weighted Cross-Entropy loss function by assigning a larger penalty to the minority class to eliminate prediction bias.

The test results demonstrated that the Optimized Fine-Tuning scenario effectively overcame overfitting and delivered the most superior classification performance. This scenario achieved an accuracy of 92.57 percent, a precision of 92.25 percent, and an F1-Score of 92.57 percent on the testing data. Furthermore, the model critically improved sensitivity (recall) to 91.33 percent, which directly impacted the drastic reduction of false negative errors to only 26 cases. These empirical results prove that the inherent weaknesses of the Vision Transformer on limited data can be thoroughly resolved through structured augmentation and regularization interventions. The main contribution of this research is the creation of a robust classification training framework for small-scale medical datasets. This system does not merely pursue high mathematical accuracy but prioritizes clinical operational feasibility by minimizing the fatal risks caused by misdiagnosis.

Keywords: *Vision Transformer, Melanoma, Fine-Tuning, Data Augmentation, Regularization*