

ABSTRAK

Pradana, Fahreza Rizky. 2026. Komparasi CNN dengan *Vision-Language* Model untuk Diagnosis Penyakit Kulit Tropis: Studi pada *Subset Dataset Derm1M*. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: Agung Nilogiri, S.T., M.Kom.; Qurrota A'yun, M.Pd.

Penyakit kulit tropis masih menjadi salah satu masalah kesehatan di Indonesia, sementara keterbatasan jumlah dokter spesialis kulit menyebabkan akses terhadap diagnosis dini belum merata. Perkembangan kecerdasan buatan membuka peluang pemanfaatan model berbasis citra sebagai pendukung diagnosis, namun penelitian yang membandingkan pendekatan *Convolutional Neural Network* (CNN) dan *Vision-Language* Model (VLM) sekaligus mengevaluasi kemampuan *interpretabilitasnya* masih terbatas. Penelitian ini bertujuan membandingkan kinerja EfficientNet-B4 sebagai model CNN, LLaVA-7B *Zero-Shot*, dan LLaVA-7B yang di-*fine-tuning* menggunakan *Quantized Low-Rank Adaptation* (QLoRA) pada *subset dataset* Derm1M. *Dataset* yang digunakan berupa *subset* Derm1M yang terdiri atas 5.500 citra dari lima kelas penyakit kulit dengan distribusi seimbang. Evaluasi kinerja dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, sedangkan kualitas penjelasan visual dianalisis menggunakan Grad-CAM berdasarkan metrik *Average Drop*, *Average Increase*, dan *Structural Similarity Index Measure* (SSIM). Hasil penelitian menunjukkan bahwa model LLaVA-7B yang di-*fine-tuning* menggunakan QLoRA memperoleh kinerja klasifikasi terbaik dengan akurasi sebesar 77,45% dan *macro* F1-score sebesar 77,62%, melampaui EfficientNet-B4 yang mencapai akurasi 67,45% serta LLaVA-7B *Zero-Shot* dengan akurasi 29,82%. Namun demikian, hasil analisis Grad-CAM menunjukkan bahwa EfficientNet-B4 menghasilkan penjelasan visual yang lebih baik. Model ini memperoleh nilai *Average Increase* tertinggi serta mencatat konsistensi spasial terbaik dengan nilai rata-rata SSIM sebesar 53%, dibandingkan LLaVA-7B QLoRA yang memperoleh nilai 26%. Penelitian ini menyimpulkan bahwa model dengan akurasi klasifikasi tertinggi belum tentu menghasilkan kualitas penjelasan visual terbaik, sehingga terdapat *trade-off* antara kinerja prediksi dan *interpretabilitas* yang perlu dipertimbangkan dalam penerapan kecerdasan buatan pada diagnosis penyakit kulit.

Kata Kunci: penyakit kulit tropis, EfficientNet-B4, LLaVA-7B, *Vision-Language* Model, QLoRA, Grad-CAM.

ABSTRACT

Pradana, Fahreza Rizky. 2026. *Comparison of CNN and Vision-Language Models for Tropical Skin Disease Diagnosis: A Study on the Derm1M Dataset Subset*. Undergraduate Thesis. Bachelor's Program in Informatics Engineering. University of Muhammadiyah Jember.

Advisors: Agung Nilogiri, S.T., M.Kom.; Qurrota A'yun, M.Pd.

Tropical skin diseases remain a significant health concern in Indonesia, while the limited availability of dermatologists has resulted in unequal access to early diagnosis. Recent advances in artificial intelligence have enabled the use of image-based models to support medical diagnosis. However, studies comparing Convolutional Neural Networks (CNNs) and Vision-Language Models (VLMs) while simultaneously evaluating their interpretability remain limited. This study compares the kinerjance of EfficientNet-B4 as a CNN model, LLaVA-7B Zero-Shot, and LLaVA-7B fine-tuned using Quantized Low-Rank Adaptation (QLoRA) on a subset of the Derm1M dataset. The dataset consists of 5,500 skin lesion images evenly distributed across five disease classes. Classification kinerjance was evaluated using accuracy, precision, recall, and macro F1-score, while visual explanation quality was assessed using Grad-CAM with Average Drop, Average Increase, and Structural Similarity Index Measure (SSIM) metrics. The results show that the QLoRA fine-tuned LLaVA-7B achieved the best classification kinerjance, with an accuracy of 77.45% and a macro F1-score of 77.62%, outperforming EfficientNet-B4, which achieved 67.45% accuracy, and LLaVA-7B Zero-Shot, which achieved 29.82% accuracy. However, Grad-CAM analysis indicates that EfficientNet-B4 produced better visual explanations. It achieved the highest Average Increase and the highest spatial consistency, with an average SSIM of 0.533, compared with 0.261 for the QLoRA fine-tuned LLaVA-7B. These findings indicate that the model with the highest classification accuracy does not necessarily provide the best visual interpretability, highlighting the trade-off between predictive kinerjance and interpretability in AI-assisted skin disease diagnosis.

Keywords: tropical skin diseases, EfficientNet-B4, LLaVA-7B, Vision-Language Model, QLoRA, Grad-CAM.