

ABSTRAK

Lesi kulit berpigmen, khususnya melanoma, merupakan salah satu jenis kanker kulit yang berbahaya dan sulit dibedakan dari lesi jinak. Deteksi dini menjadi faktor krusial dalam menurunkan risiko mortalitas. Namun, interpretasi citra *dermoskop* dalam praktik klinis menghadapi tantangan berupa variasi bentuk, warna, serta keberadaan artefak citra, yang dapat memengaruhi konsistensi penilaian antar-pengamat. Pendekatan *deep learning*, khususnya *Convolutional Neural Network* (CNN), telah banyak digunakan dalam klasifikasi lesi kulit berpigmen karena mampu mencapai akurasi yang tinggi. Meskipun demikian, model-model tersebut umumnya bersifat *black-box*, sehingga sulit dipahami oleh pengguna klinis dan berpotensi menurunkan tingkat kepercayaan dalam pengambilan keputusan medis.

Sejumlah penelitian sebelumnya telah mengintegrasikan teknik *Interpretable Machine Learning* (IML) atau *Explainable Artificial Intelligence* (XAI) untuk meningkatkan transparansi model *deep learning*. Namun, pendekatan-pendekatan tersebut masih memiliki keterbatasan, antara lain: (1) sebagian besar hanya berfokus pada klasifikasi biner, (2) interpretasi yang disediakan bersifat parsial dan teknis, (3) belum menggabungkan lebih dari satu mekanisme interpretasi secara terintegrasi, serta (4) tidak melibatkan evaluasi langsung oleh pengguna klinis untuk menilai dampak fitur interpretatif terhadap tingkat kepercayaan dan keberterimaan sistem. Celah inilah yang mendorong penelitian ini untuk mengusulkan pendekatan IML yang tidak hanya meningkatkan kinerja klasifikasi, tetapi juga tervalidasi secara klinis dari sisi pengguna akhir.

Penelitian ini mengembangkan model klasifikasi lesi kulit berpigmen berbasis *Interpretable Machine Learning* (IML) menggunakan dataset ISIC 2019 yang terdiri dari delapan kelas lesi kulit. Model dibangun dengan arsitektur *pre-trained Inception-v3* yang dioptimasi melalui penerapan *data augmentation*, *Bayesian hyperparameter tuning*, serta penambahan *attention module* sebagai bagian dari arsitektur model.

Hasil eksperimen menunjukkan bahwa model yang diusulkan mencapai akurasi sebesar 97,38%, nilai AUC sebesar 0,985, dan sensitivitas sebesar 98,50%, yang mengindikasikan kemampuan yang sangat baik dalam mendeteksi lesi ganas seperti melanoma. Untuk meningkatkan transparansi model, dua keluaran interpretatif disediakan, yaitu *attention map* berbasis Grad-CAM yang menyoroti area penting pada citra, serta tiga citra *dermoskop* paling mirip dari kelas prediksi menggunakan pendekatan *Content-Based Image Retrieval* (CBIR).

Evaluasi interpretabilitas dilakukan melalui studi pengguna yang melibatkan 33 dokter spesialis kulit. Dalam penelitian ini, dokter tidak digunakan sebagai *gold standard* diagnostik, melainkan sebagai pengguna akhir (*end-user*) untuk mengevaluasi persepsi kepercayaan dan keberterimaan sistem bantu diagnosis yang diusulkan. Responden diminta menilai dua versi model, yaitu model awal tanpa fitur interpretatif dan model akhir dengan fitur interpretatif, menggunakan kuesioner delapan pertanyaan. Hasil uji statistik *Wilcoxon Signed-Rank* menunjukkan adanya peningkatan kepercayaan pengguna yang signifikan terhadap model interpretatif.

Penelitian ini berkontribusi dalam pengembangan sistem bantu diagnosis berbasis *deep learning* yang tidak hanya unggul dari sisi kinerja klasifikasi, khususnya sensitivitas, tetapi juga lebih transparan dan memiliki tingkat akseptabilitas yang lebih baik dalam konteks praktik klinis.



UNIVERSITAS
GADJAH MADA

Klasifikasi Lesi Kulit Berpigmen Pada Citra Dermoskop Melalui Integrasi Attention Mechanism Pada Pre-Trained Inception-V3

Erwin Setyo Nugroho, Hanung Adi Nugroho, Igi Ardiyanto

Universitas Gadjah Mada, 2026 | Diunduh dari <http://etd.repository.ugm.ac.id/>

Kata kunci—Lesi kulit berpigmen, melanoma, Interpretable Machine Learning, Convolution Neural Network, Inception-v3, Grad-CAM, Content-Based Image Retrieval, klasifikasi citra.

ABSTRACT

Pigmented skin lesions, particularly melanoma, represent one of the most dangerous types of skin cancer and are often difficult to distinguish from benign lesions. Early detection is crucial for reducing mortality risk. However, interpretation of dermoscopic images in clinical practice faces significant challenges due to variations in lesion shape, color, and the presence of image artifacts, which may affect inter-observer consistency. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have been widely adopted for pigmented skin lesion classification due to their high classification accuracy. Nevertheless, such models are generally regarded as black-box systems, making them difficult to interpret by clinicians and potentially reducing trust in clinical decision-making.

Several previous studies have integrated Interpretable Machine Learning (IML) or Explainable Artificial Intelligence (XAI) techniques to improve the transparency of deep learning models. However, these approaches still exhibit notable limitations, including: (1) a predominant focus on binary classification tasks, (2) interpretability mechanisms that are partial and technically oriented, (3) the lack of integrated multi-mechanism interpretability strategies, and (4) the absence of direct evaluation by clinical users to assess the impact of interpretability features on user trust and system acceptability. These gaps motivate the present study to propose an IML-based approach that not only improves classification performance but is also clinically validated from an end-user perspective.

This study develops an Interpretable Machine Learning-based model for pigmented skin lesion classification using the ISIC 2019 dataset, which consists of eight lesion classes. The model is built upon a pre-trained Inception-v3 architecture and optimized through data augmentation, Bayesian hyperparameter tuning, and the incorporation of an attention module as an integral component of the model architecture.

Experimental results demonstrate that the proposed model achieves an accuracy of 97.38%, an AUC value of 0.985, and a sensitivity of 98.50%, indicating strong capability in correctly identifying malignant lesions such as melanoma. To enhance model transparency, two interpretability outputs are provided: attention maps generated using Grad-CAM to highlight diagnostically relevant regions in the images, and three most similar dermoscopic images from the predicted class retrieved using a Content-Based Image Retrieval (CBIR) approach.

Interpretability evaluation was conducted through a user study involving 33 board-certified dermatologists. In this study, dermatologists were not used as a diagnostic gold standard, but rather as end-users to evaluate trust perception and system acceptability of the proposed clinical decision support system. Participants assessed two model versions, namely an initial model without interpretability features and a final model with interpretability features, using an eight-item questionnaire. Statistical analysis using the Wilcoxon Signed-Rank test revealed a significant increase in user trust toward the interpretable model.

This research contributes to the development of deep learning-based clinical decision support systems that are not only high-performing in classification accuracy, particularly sensitivity, but also more transparent and better accepted in real clinical practice.

Keywords—Pigmented skin lesions, melanoma, Interpretable Machine Learning, Convo-



UNIVERSITAS
GADJAH MADA

Klasifikasi Lesi Kulit Berpigmen Pada Citra Dermoskop Melalui Integrasi Attention Mechanism Pada Pre-Trained Inception-V3

Erwin Setyo Nugroho, Hanung Adi Nugroho, Igi Ardiyanto

Universitas Gadjah Mada, 2026 | Diunduh dari <http://etd.repository.ugm.ac.id/>

lutional Neural Network, Inception-v3, Grad-CAM, Content-Based Image Retrieval, image classification.