

Optimasi Model *Deep Learning* untuk Klasifikasi Penyakit Kulit Multikelas Menggunakan CNN dan *Vision Transformer*

Deep Learning Model Optimization for Multiclass Skin Disease Classification Using CNN and Vision Transformer

Fachri Ayudi Fitrony*, Muhamad Arief Muhaimin

Universitas Nahdlatul Ulama Pasuruan. Jl. Warung Dowo Utara, Pohjentrek, Kabupaten Pasuruan, Indonesia.

*Email: fahri@unupasuruan.ac.id

Abstrak – Penyakit kulit merupakan tantangan kesehatan global yang signifikan, memerlukan diagnosis akurat dan tepat waktu untuk pengobatan yang efektif. Dalam penelitian ini, kami membahas secara mendalam ranah klasifikasi penyakit kulit yang rumit, dengan memanfaatkan teknik pembelajaran mendalam tingkat lanjut. Penyakit kulit, yang dapat disebabkan oleh peradangan, infeksi, tumor, atau kondisi autoimun, dapat muncul sebagai ruam, bintik-bintik, benjolan, atau sisik, atau tetap tanpa gejala sampai menjadi parah. Pendekatan diagnostik konvensional seperti inspeksi visual dan palpasi seringkali kurang akurat. Kecerdasan buatan (AI) meningkatkan ketepatan diagnostik dengan menganalisis sejumlah besar gambar kulit untuk mendeteksi pola halus yang mungkin tidak dikenali oleh dokter. Penelitian ini menyajikan model diagnostik penyakit kulit multikelas yang dikembangkan menggunakan dataset publik (Kaggle). Dataset dikategorikan menjadi tujuh kelas penyakit untuk pelatihan model. Sebanyak 25 model pra-terlatih, termasuk jaringan saraf konvolusional (CNN) dan transformator visi (ViT), disempurnakan. Tiga model teratas digabungkan menjadi sebuah ensemble menggunakan model *Transfer Learning* dari CNN. Untuk memastikan keberhasilannya, model tersebut diuji melalui lima eksperimen acak dan divalidasi menggunakan teknik holdout. Model ensemble yang diusulkan, Swin-ViT-EfficientNetB4, mencapai akurasi pengujian tertinggi sebesar 97,5%, menunjukkan potensi yang kuat untuk diagnosis penyakit kulit yang akurat sejak dini.

Kata kunci: Penyakit Kulit Multikelas; Deep Learning; CNN; Ensemble Learning; Swin-ViT-EfficientNetB4

Abstract - Skin diseases are a significant global health challenge, requiring accurate diagnosis and timely treatment for effective treatment. In this study, we delve deeply into the complex realm of skin disease classification, leveraging advanced deep learning techniques. Skin diseases, which can be caused by inflammation, infection, tumors, or autoimmune conditions, can present as rashes, spots, bumps, or scales, or remain symptomatic without progressing to severe disease. Diagnostic approaches such as visual inspection and palpation often lack accuracy. Artificial intelligence (AI) improves diagnostic accuracy by analyzing large numbers of skin images to detect subtle patterns that may go unnoticed by clinicians. This study presents a multiclass skin disease diagnostic model developed using a public dataset (Kaggle). The dataset was suggested to be seven disease classes for model training. A total of 25 pre-trained models, including convolutional neural networks (CNNs) and vision transformers (ViTs), were fine-tuned. The top three models were combined into an ensemble using a transfer learning model. To ensure its efficacy, the model was tested through five randomized experiments and validated using a holdout technique. The proposed ensemble model, Swin-ViT-EfficientNetB4, achieved the highest testing accuracy of 97.5%, demonstrating strong potential for accurate and early diagnosis of skin diseases.

Keywords: Multiclass Skin Disease; Deep Learning; CNN; Ensemble Learning; Swin-ViT-EfficientNetB4

1. PENDAHULUAN

Kulit adalah organ terbesar dalam tubuh manusia dan memainkan banyak peran, termasuk perlindungan, pengaturan suhu tubuh, dan produksi vitamin D. Paparan sinar ultraviolet (UV) yang berkepanjangan tidak hanya menyebabkan perubahan seluler pada kulit tetapi juga merupakan penyebab utama penyakit kulit. Selain paparan UV, penuaan adalah faktor penting lainnya, sementara pemicu lainnya termasuk luka bakar, radiasi, warisan genetik,

keracunan arsenik kronis, dan sistem kekebalan tubuh yang tidak seimbang (Chiu et al., 2025). Kondisi kulit juga dapat disebabkan oleh sejumlah hal, termasuk kolonisasi bakteri, reaksi alergi, ketidakseimbangan mikroba, infeksi jamur, atau pigmentasi abnormal. Karena beberapa kondisi kulit dapat berkembang menjadi keganasan dan bahkan menjadi kronis, pengobatan yang tepat waktu sangat penting untuk membatasi pertumbuhan dan penyebarannya (Deshmukh et al., 2023). Penyakit kulit dapat muncul dengan tingkat keparahan yang beragam, mulai dari gejala ringan hingga kondisi kronis. Bentuknya pun bermacam-macam, seperti peradangan, infeksi, tumor (neoplastik), hingga gangguan autoimun. Merujuk pada laporan Skin Cancer Asia tahun 2009 (Kim & Del Rosso, 2009). Kanker kulit saat ini merupakan jenis kanker yang paling umum, dan meskipun *Malignant Melanoma* (MM) hanya menyumbang sekitar 1% dari semua kanker kulit, namun merupakan penyebab utama sebagian besar kematian terkait kanker kulit. Tahi lalat pada kulit sering menimbulkan kekhawatiran karena ketidakpastian diagnosis. Meskipun sebagian besar tahi lalat tidak berbahaya, beberapa lesi yang tampak seperti tahi lalat sebenarnya mungkin adalah kanker kulit. Khususnya *Malignant Melanoma* (MM) dan *Basal cell carcinoma* (BCC) (Cha, 2022). Pemeriksaan juga diperlukan untuk penyakit neoplastik lain yang dicurigai secara klinis yang mungkin menyerupai MM untuk memfasilitasi diagnosis dini dan pengobatan.

Di bidang visi komputer, klasifikasi penyakit kulit merupakan topik penelitian yang sangat penting. Karena variasi yang signifikan dalam pengalaman klinis dokter kulit dan sifat proses diagnostik yang memakan waktu, pengembangan sistem diagnostik berbantuan komputer yang andal untuk membantu dalam deteksi dan klasifikasi penyakit kulit sangat penting untuk meningkatkan kualitas perawatan kesehatan. Metode sebelumnya untuk klasifikasi penyakit kulit seringkali melibatkan ekstraksi fitur yang dibuat secara manual yang dikombinasikan dengan algoritma pembelajaran mesin (Ahmed et al., 2026). Penelitian terdahulu yang dilakukan oleh Houssein dkk (2024) dengan melakukan uji coba model DCNN yang diusulkan untuk mengklasifikasi penyakit kulit dengan dataset yang tidak seimbang serta mendapatkan akurasi training sebesar 94%. Akan tetapi model yang diusulkan terbilang cukup tradisional (Houssein et al., 2024). Penelitian kedua oleh Asaduzzaman dkk (2024) melakukan pendekatan pembelajaran mesin untuk meningkatkan kinerja sistem diagnosis berbantuan komputer (CADx) untuk mendiagnosis penyakit kulit. Dalam sistem CADx yang diusulkan, diskriminator jaringan adversarial generatif (GAN) digunakan untuk mengidentifikasi (dan menghapus) citra palsu. Menerapkan GAN (diskriminator) dan SMOTE, LDA, SVM, CNN, dan *ensemble* CNN-SVM menunjukkan akurasi masing-masing sebesar 76%, 83%, 87%, dan 94%. Perbedaan dengan penelitian yang diusulkan adalah *ensemble* dari penelitian terdahulu masih menggunakan CNN-SVM (Asaduzzaman et al., 2024). Maka dari penelitian terdahulu pengusul memberikan dasar bagi kerangka kerja *ensemble* yang kami usulkan, bertujuan untuk mengatasi tantangan ini melalui arsitektur terbaru dan strategi evaluasi yang kuat.

Dalam beberapa tahun terakhir, model *Vision Transformer* (ViT) telah muncul sebagai kemajuan signifikan dalam tugas visi komputer (K et al., 2025). Tidak seperti CNN, ViT menggunakan mekanisme perhatian yang memungkinkan pemrosesan data *input* yang lebih efektif, terutama dalam analisis citra skala besar. Selain itu, ViT menunjukkan kesadaran spasial, memungkinkan model untuk menangkap hubungan spasial dalam data, yang

merupakan kemampuan penting untuk klasifikasi penyakit kulit yang akurat (Ye et al., 2025).

Penelitian ini mengkaji penerapan metode *Deep Ensemble Learning* yang mengintegrasikan arsitektur *Convolutional Neural Networks* (CNN) dan *Vision Transformer* (ViT) untuk meningkatkan akurasi klasifikasi penyakit kulit multikelas. Tujuan utamanya adalah mengembangkan kerangka kerja (*framework*) pembelajaran mendalam yang handal serta memiliki kemampuan generalisasi yang tinggi dalam mengidentifikasi berbagai jenis penyakit kulit.

Tinjauan Pustaka

Beberapa penelitian sebelumnya telah mengeksplorasi penggunaan *Deep Learning* dalam ranah penyakit kulit untuk meningkatkan akurasi diagnosis sebagai berikut: penelitian terdahulu oleh Sao dan Kushwaha (2024) membahas tentang pengklasifikasian otomatis penyakit kulit menggunakan *Convolutional Neural Network* (CNN) dengan aplikasi MATLAB. Mereka menggunakan dataset gambar dermatologi untuk melatih model CNN yang bisa membedakan berbagai jenis penyakit kulit seperti eksim, psoriasis, melanoma, dan lain-lain. Hasil dari penelitian ini menunjukkan bahwa model CNN yang dikembangkan mampu mencapai tingkat akurasi sebesar 92% dalam mengklasifikasi berbagai jenis penyakit kulit dari gambar dermoskopi. Kekurangan dari penelitian ini terletak pada model yang sangat bergantung pada resolusi tinggi dan kualitas gambar yang konsisten. Penggunaan gambar berkualitas rendah dapat mengurangi akurasi (Sao & Kushwaha, 2024). Penelitian kedua oleh Rachmawati (2024). Penelitian ini membahas klasifikasi jenis penyakit kulit yang terdeteksi, seperti melanoma, nevus, dan keratosis seboroik. Penelitian ini mengusulkan metode pembelajaran mesin, yaitu Jaringan Saraf Konvolusional (*Convolutional Neural Network*), untuk mendiagnosis penyakit kulit secara otomatis. Penelitian ini menggunakan teknik augmentasi data dan penyeimbangan data yang diimplementasikan pada dataset citra lesi kulit ISIC 2018. Pengujian dengan teknik dan metode yang diusulkan menghasilkan akurasi sebesar 96%. Teknik augmentasi dan penyeimbangan data yang diusulkan dapat secara signifikan meningkatkan kinerja model. Akan tetapi model yang diusulkan memiliki keterbatasan model lama yaitu CNN (Rachmawati et al., 2024).

Penelitian selanjutnya adalah dari Auinunisa (2025). Penelitian ini berhasil merancang dan mengimplementasikan sistem klasifikasi berbasis citra digital untuk mendeteksi berbagai jenis penyakit kulit menggunakan algoritma *Convolutional Neural Network* (CNN) dengan arsitektur VGG19 (Ainunnisa et al., 2025).

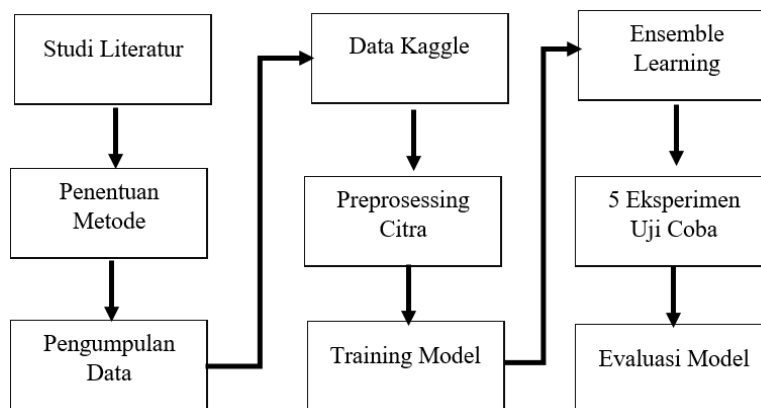
Penelitian oleh Maheswari (2025). Penelitian ini mengimplementasikan dan membandingkan tiga model pembelajaran mendalam yaitu *Convolutional Neural Network* (CNN), *Deep Neural Network* (DNN), dan EfficientNet untuk mendeteksi kanker kulit menggunakan dataset HAM10000. Proses penelitian meliputi pra-pemrosesan, pelatihan model, evaluasi kinerja, dan integrasi ke dalam aplikasi web interaktif berbasis Flask. Evaluasi menggunakan metrik akurasi, presisi, recall, skor F1, dan AUC. Hasil pengujian menunjukkan bahwa *EfficientNet* berkinerja terbaik dengan akurasi pengujian 78,44%, diikuti oleh CNN sebesar 69,76%, sedangkan DNN hanya mencapai 40,52% karena kehilangan informasi spasial. (Maheswari & Gunawan, 2025)

Terakhir penelitian oleh Vuran (2025). Tujuan dari penelitian ini adalah untuk mengembangkan dan mengevaluasi model diagnosis otomatis penyakit *Mpox* dari gambar penyakit kulit menggunakan arsitektur *deep learning* berbasis transformer. Hasil dari penelitian ini menunjukkan bahwa arsitektur transformer berbasis *SwinTransformer* mencapai tingkat keberhasilan tertinggi dengan akurasi sekitar 93,71% dalam klasifikasi multi-kelas penyakit kulit termasuk *Mpox*. Secara umum, semua arsitektur transformer yang diuji telah menunjukkan tingkat keberhasilan yang tinggi, dengan ViT, DINO, dan *SwinTransformer* masing-masing mencapai lebih dari 93%, sedangkan MAE mencapai sekitar 84,60%. Perbedaan dari penelitian ini terletak pada integrasi tiga model pra-latih *Swin Transformer*, *Vision Transformer* (ViT), dan *EfficientNetB4* melalui strategi voting (*soft voting*) untuk menciptakan model ensemble yang lebih kuat dan komprehensif. Pendekatan ini memadukan keunggulan CNN (*EfficientNetB4*) dalam mengekstrak fitur lokal dan transformator (Swin, ViT) untuk memahami konteks global, yang tidak selalu dilakukan secara bersamaan dalam penelitian lain. (Vuran et al., 2025).

2. DATA DAN METODOLOGI

Alur Penelitian

Gambar 1. di bawah mengilustrasikan alur kerja langkah demi langkah dari studi yang diusulkan. Proses pengembangan model dimulai dengan dataset Kaggle, yang menjalani serangkaian langkah pra-pemrosesan, termasuk kompilasi data, pengubahan ukuran gambar, dan klasifikasi ke dalam tujuh kategori penyakit kulit yang berbeda. Hal ini menghasilkan dataset terstruktur yang disebut data 7 kelas.



Gambar 1. Tahapan Penelitian

Dataset Kaggle yang digunakan dalam penelitian ini relatif kecil dan menunjukkan ketidakseimbangan kelas yang signifikan. Penerapan validasi silang (*Cross Validation*) dalam kondisi seperti itu dapat menyebabkan hasil yang bias atau tidak stabil karena representasi kelas minoritas yang tidak memadai di setiap layer. Untuk memastikan kekokohan, kami melakukan lima eksperimen acak uji coba independen menggunakan pendekatan validasi *holdout*. Dalam setiap eksperimen, dataset dibagi secara acak menjadi subset pelatihan dan pengujian dengan rasio 80:20, distratifikasi berdasarkan distribusi kelas

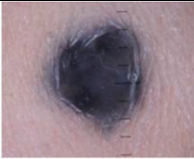
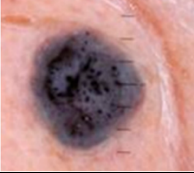
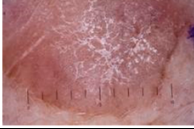
untuk mempertahankan representasi proporsional.

Proses ini diulang lima kali dengan *seed* acak yang berbeda, dan nilai akurasi dan deviasi standar yang dilaporkan di semua tabel mencerminkan rata-rata dan variabilitas di kelima percobaan ini. Untuk menghindari kebocoran data, tidak ada gambar yang dibagikan antara set pelatihan dan pengujian, sehingga keduanya benar-benar saling eksklusif. Pendekatan ini memungkinkan model untuk dilatih pada bagian data yang representatif sambil mempertahankan set independen untuk evaluasi yang tidak bias.

Dataset

Penelitian ini menggunakan dataset citra dermoskopi yang dikumpulkan dari website Kaggle dengan link <https://www.kaggle.com/datasets/aurellashilla/skin-disease>, setiap kasus dikonfirmasi oleh patologi. Semua pasien dalam dataset ini adalah orang Asia. Dataset diklasifikasikan menggunakan kerangka kerja ISIC dan mencakup melanoma (111 kasus), nevus melanositik (199 kasus), karsinoma sel basal (45 kasus), keratosis aktinik (82 kasus), keratosis jinak (134 kasus), dermatofibroma (42 kasus) dan lesi vaskular (53 kasus), dengan total 666 kasus. Informasi mengenai dataset Kaggle disediakan pada Tabel 1. dibawah. Distribusi label dalam dataset ini sangat tidak merata, dengan rasio jumlah label maksimum terhadap minimum mencapai 4,738. Ini menunjukkan ketidakseimbangan yang signifikan. Untuk mengatasi masalah ini, studi ini akan menggunakan teknik augmentasi data untuk mencapai distribusi yang lebih seimbang dalam dataset pelatihan. Berikut ini adalah deskripsi singkat tentang tujuh jenis lesi neoplastik kulit yang menarik secara klinis.

Tabel 1. Jumlah kasus untuk masing-masing dari 7 kategori dalam dataset.

Kelas	Jumlah	Label	Gambar
Malignant Melanoma	111	0	
Melanocytic Nevus	199	1	
Basal cell carcinoma	45	2	
Actinic Keratosis	82	3	

Benign keratosis	134	4	
Dermatofibroma	42	5	
Vascular	53	6	

Data Preprocessing

Dataset Kaggle dibagi menjadi set pelatihan dan pengujian, dengan pembagian 80:20. Dataset distratifikasi berdasarkan distribusi kelas untuk memastikan bahwa setiap kelas terwakili di kedua set tanpa tumpang tindih gambar. Semua gambar diubah ukurannya menjadi 224×224 piksel dan dinormalisasi untuk memfasilitasi komputasi dalam model pembelajaran mendalam. Untuk meningkatkan keragaman data pelatihan, teknik augmentasi data seperti pembalikan acak, rotasi, translasi, dan penskalaan diterapkan (Lubis & Yulianto, 2023). Augmentasi ini hanya diterapkan pada set pelatihan dan bukan pada set pengujian. Sebelum pelatihan model, prosedur augmentasi data statis diimplementasikan untuk menyeimbangkan distribusi label dalam dataset pelatihan. Misalnya, distribusi label awal adalah (89, 159, 36, 65, 107, 34, 42). Setelah menerapkan teknik augmentasi statis, distribusi disesuaikan menjadi (178, 159, 144, 130, 107, 170, 168). Selain augmentasi statis yang diterapkan sebelum pelatihan untuk menyeimbangkan distribusi kelas, strategi augmentasi data dinamis digunakan selama fase pelatihan untuk lebih mendiversifikasi data masukan dan meningkatkan kemampuan generalisasi model (Setiyani & Krissyano, 2026).

Transfer Learning

Transfer learning melibatkan penerapan model dari satu domain ke domain baru, mentransfer pengetahuan dari tugas sumber ke tugas target, menghilangkan kebutuhan untuk melatih model dari awal (Sinaga et al., 2024). Keuntungannya terletak pada pengurangan jumlah data pelatihan yang dibutuhkan. Dua puluh lima model pra-terlatih yang berbeda dipilih, termasuk arsitektur berbasis transformer seperti *Swin Transformer* (kecil, dasar, besar) dan *Vision Transformer* (b16, b32, l16, l32), serta model berbasis CNN seperti *EfficientNet* (B1–B5, V2B2, V2B3), *MobileNetV2*, *DenseNet* (121, 169, 201), *VGGNet16*, *Xception*, *ResNet50*, *InceptionV3*, *InceptionResNetV2*, dan *ConvNeXt* (Dasar, Kecil) (Nurkhasanah & Murinto, 2022).

Berdasarkan percobaan awal, kami menemukan bahwa penyempurnaan semua lapisan model yang telah dilatih sebelumnya menghasilkan kinerja yang lebih baik daripada membekukan lapisan awal. Oleh karena itu, dalam pendekatan kami, semua bobot jaringan diperbarui selama pelatihan untuk memaksimalkan adaptasi fitur terhadap dataset Kaggle.

Ensemble Arsitektur

Penelitian ini menggunakan teknik pembelajaran *ensemble* untuk meningkatkan kinerja prediksi model dengan menggabungkan model yang berbeda untuk mencapai hasil yang lebih baik. Metode *ensemble voting* digunakan, di mana *output* dari beberapa model independen digabungkan sebagai *input* untuk melatih model baru. Dalam pendekatan ini, beberapa model (sub-model) dilatih secara independen dan prediksinya digabungkan melalui proses voting (Ibrahim et al., 2025). Setiap model membuat prediksi pada setiap sampel *input*, dan prediksi *output* akhir adalah prediksi yang menerima mayoritas suara di antara semua model. Metode ini memanfaatkan kekuatan beberapa algoritma pembelajaran untuk meningkatkan kinerja prediksi dan mengurangi kemungkinan kesalahan dari model tunggal (Wona Maria Misela A et al., 2025).

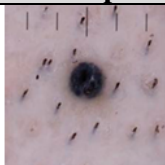
Penelitian ini melakukan analisis komparatif strategi *hard voting* dan *soft voting*. Dalam *hard voting*, setiap pengklasifikasi menghasilkan label kelas diskrit. Prediksi akhir ditentukan oleh suara mayoritas, di mana kelas yang menerima suara terbanyak dipilih. Dalam *soft voting*, setiap pengklasifikasi memberikan distribusi probabilitas atas semua kelas. Probabilitas ini dirata-ratakan, dan kelas dengan probabilitas gabungan tertinggi dipilih sebagai prediksi akhir. Dalam penelitian ini, semua pengklasifikasi diberi bobot yang sama. Untuk *ensemble* akhir, tiga pengklasifikasi yang lebih baik dipilih berdasarkan kinerja individual mereka di seluruh arsitektur *SwinTransformer*, *Vision Transformer*, dan berbasis CNN.

3. HASIL PENELITIAN

Penghapusan Skala Dermoskopik

Gambar asli berisi skala dermoskopik, yang dihilangkan dengan memotong gambar sebelum pelatihan. Baik gambar asli maupun gambar yang dipotong digunakan untuk pelatihan model. Seperti yang ditunjukkan pada Tabel 2., gambar yang dipotong menunjukkan kinerja yang lebih unggul dalam pelatihan dan pengujian dibandingkan dengan gambar yang belum diproses. Oleh karena itu, semua pelatihan dan pengujian selanjutnya dilakukan menggunakan gambar yang dipotong.

Tabel 2. Evaluasi kinerja Transformer Swin pada dataset asli dan yang telah dipotong.

Dataset	Citra Terpotong	Model	Train ACC	Test ACC
<i>Original</i>		<i>Swin Base Transformers</i>	0.991 ± 0.005	0.818 ± 0.046
<i>Cropped</i>			0.992 ± 0.005	0.845 ± 0.027

Pengaruh Model Pra-terlatih

Tabel 3. menyajikan kinerja pelatihan dan pengujian dari 25 model pra-terlatih di dua kategori arsitektur: CNN dan *Vision Transformer* (ViT), yang diterapkan pada klasifikasi

penyakit kulit tujuh kelas. Model CNN, seperti EfficientNetB1 dan B3, menunjukkan akurasi pelatihan yang kuat (hingga 0,993), tetapi akurasi pengujian yang moderat, dengan *EfficientNetB4* mencapai 0,816. Model ViT menunjukkan generalisasi yang sedikit lebih baik, dengan Vision l16 dan l32 Transformer mencapai akurasi pengujian 0,840. *Model Swin*, khususnya *Swin Large Transformer*, mengungguli yang lain dengan akurasi pengujian tertinggi 0,849, menunjukkan kemampuan superior dalam menangkap fitur lokal dan global. Secara keseluruhan, hasil menunjukkan bahwa arsitektur *Swin* menawarkan pendekatan yang lebih seimbang dan efektif untuk klasifikasi citra medis, menggabungkan kekuatan CNN dan transformer untuk meningkatkan keandalan diagnostik.

Metode Pemungutan Suara Keras dan Pemungutan Suara Lunak (*Ensemble*)

Tabel 4 membandingkan kinerja berbagai strategi pemungutan suara *ensemble*, pemungutan suara keras dan pemungutan suara lunak, di tiga kategori arsitektur: CNN, ViT, dan *ensemble* campuran (*Swin-ViT-EfficientNetB4*). Di bawah pemungutan suara keras, akurasi pengujian terbaik adalah 0,954 baik dari *ensemble ViT* saja maupun *ensemble* campuran. Di bawah pemungutan suara lunak, *ensemble* campuran mencapai akurasi pengujian tertinggi sebesar 0,961, mengungguli semua konfigurasi lainnya. Model CNN dan ViT berkinerja baik secara individual, tetapi model Swin menunjukkan akurasi pengujian yang lebih rendah, terutama dengan pemungutan suara keras (0,903).

Tabel 3. Hasil Pelatihan dan Pengujian dari 25 Model yang Telah Dilatih Sebelumnya.

Arsitektur	Jumlah Model	Model	Train ACC	Test ACC
CNN	11	EfficientNetB1	0.993±0.004	0.793±0.050
		EfficientNetB2	0.960±0.009	0.748±0.015
		EfficientNetB3	0.989±0.007	0.815±0.027
		EfficientNetB4	0.993±0.004	0.816±0.016
		EfficientNetV2B2	0.965±0.028	0.797±0.036
		EfficientNetV2B3	0.985±0.007	0.806±0.015
		MobileNetV2	0.944±0.056	0.749±0.033
		DenseNet121	0.954±0.016	0.736±0.069
		DenseNet169	0.932±0.038	0.650±0.055
		VGG16	0.967±0.025	0.777±0.034
		ResNet	0.912±0.019	0.787±0.025
ViT	4	Vision b16 Transformer	0.995±0.003	0.824±0.022
		Vision b32 Transformer	0.996±0.004	0.821±0.032
		Vision l16 Transformer	0.995±0.003	0.840±0.019
		Vision l32 Transformer	0.997±0.005	0.840±0.012

Tabel 4. Hasil Perbandingan Berbagai Metode *Ensemble* Pemungutan Suara.

<i>Voting Method</i>	Arsitektur	Model	Train ACC	Test ACC
<i>Hard Voting</i>	CNN	EfficientNetB3- EfficientNetB4 EfficientNetV2B3	0.999±0.001	0.948±0.017

	Vit	ViTb16-ViTl16-ViTl32	1.000±0.000	0.954±0.013
	<i>Ensemble</i>	Swin-ViT-EfficientNetB4	0.999±0.001	0.954±0.020
<i>Soft Voting</i>	CNN	EfficientNetB3- EfficientNetB4 EfficientNetV2B3	0.999±0.001	0.952±0.010
	ViT	ViTb16-ViTl16-ViTl32	1.000±0.000	0.955±0.013
	<i>Ensemble</i>	Swin-ViT-EfficientNetB4	0.999±0.001	0.961±0.020

Temuan-temuan ini menegaskan efektivitas dari *soft voting ensemble* yang mengintegrasikan berbagai jenis model. Dengan menggabungkan CNN (yang kuat dalam ekstraksi fitur lokal) dengan transformer (yang efektif dalam menangkap konteks global), model ansambel ini mencapai generalisasi dan ketangguhan yang unggul. Pendekatan ini memitigasi keterbatasan dari masing-masing arsitektur individu dan meningkatkan akurasi diagnostik dalam klasifikasi penyakit kulit. Hasil tersebut mendukung penggunaan pembelajaran *ensemble* hibrida (*hybrid ensemble learning*) sebagai strategi yang menjanjikan untuk meningkatkan keandalan dalam analisis citra medis.

Hasil penelitian ini menunjukkan bahwa model *ensemble learning* secara signifikan mengungguli model individual dalam hal presisi dan stabilitas untuk klasifikasi penyakit kulit. Pada kumpulan data Kaggle (*Public*), model yang menggabungkan *Swin Transformer*, *Vision Transformer* (ViT), dan *EfficientNetB4* menunjukkan performa yang luar biasa dan berhasil meningkatkan akurasi klasifikasi gambar.

Peningkatan ini disebabkan oleh kekuatan yang saling melengkapi dari arsitektur yang diintegrasikan. *Convolutional Neural Networks* (CNN), seperti *EfficientNetB4*, sangat efektif dalam menangkap fitur spasial lokal melalui lapisan konvolusi hierarkis, sehingga sangat cocok untuk citra medis yang kaya akan tekstur. Sebaliknya, *Vision Transformers* (ViT) dan *Swin Transformers* unggul dalam memodelkan ketergantungan jarak jauh dan informasi kontekstual global dengan memanfaatkan mekanisme *self-attention*.

4. PEMBAHASAN

Hasil penelitian ini menunjukkan bahwa model *ensemble* secara signifikan mengungguli model individu dalam hal presisi dan stabilitas untuk klasifikasi penyakit kulit. Dalam dataset publik, model yang menggabungkan *Swin Transformer*, *Vision Transformer* (ViT), dan *EfficientNetB4* menunjukkan performa yang sangat baik serta berhasil meningkatkan akurasi klasifikasi gambar.

Peningkatan ini disebabkan oleh kekuatan yang saling melengkapi dari arsitektur yang terintegrasi tersebut:

- Convolutional Neural Networks* (CNN): Seperti *EfficientNetB4*, sangat efektif dalam menangkap fitur spasial lokal melalui lapisan konvensional hierarkis, sehingga sangat cocok untuk citra medis yang kaya akan tekstur.

- b. *Vision Transformers (ViT) & Swin Transformers*: Unggul dalam memodelkan dependensi jarak jauh dan informasi kontekstual global dengan memanfaatkan mekanisme *self-attention*.

Pendekatan *hybrid ensemble* ini mendapat keuntungan dari ekstraksi fitur lokal sekaligus pembelajaran representasi global, yang menghasilkan generalisasi dan ketangguhan (*robustness*) yang lebih baik. Dengan menggabungkan model berbasis CNN dan *Transformer* melalui metode *soft voting*, model *ensemble* ini mampu memitigasi keterbatasan dari masing-masing arsitektur individu dan mencapai akurasi diagnostik yang unggul. Temuan ini menegaskan potensi kerangka kerja *deep learning hybrid* dalam memajukan klasifikasi penyakit kulit otomatis, terutama pada dataset klinis yang beragam dan kompleks.

Temuan studi ini mengungkapkan potensi besar metode *ensemble deep learning* dalam aplikasi klinis, khususnya dalam meningkatkan diagnosis dini dan hasil pengobatan untuk penyakit kulit. Namun, untuk lebih meningkatkan nilai praktis dari model-model ini, penelitian di masa depan harus fokus pada penyelesaian masalah generalisasi yang ada saat ini dan terus mengeksplorasi strategi *ensemble* yang lebih optimal guna meningkatkan performa diagnostik model di berbagai dataset yang berbeda, sehingga model tersebut dapat memainkan peran yang lebih besar dalam lingkungan klinis yang nyata.

Perbandingan dari penelitian terdahulu oleh Tasar (2023) tentang klasifikasi penyakit kulit menggunakan metode *deep transfer learning*. Penelitian menguji beberapa model CNN yang sudah dilatih sebelumnya untuk mendeteksi jenis penyakit kulit dari gambar dermoskopi dengan akurasi hingga 94%. Tujuannya membantu deteksi kanker kulit lebih cepat dan akurat. Menggunakan arsitektur *CNN VGG16*, *ResNet50*, *DenseNet121*, *MobileNet*, dan *Xception*. Hasil perbandingan untuk klasifikasi penyakit kulit tujuh kelas, termasuk penulis, tahun, metode, validasi, dataset, dan akurasi keseluruhan pada set pengujian (Tasar, 2023). Model *ensemble* yang diusulkan, yang mengintegrasikan *Swin Transformer*, *Vision Transformer*, dan *EfficientNetB4* dengan strategi pemungutan suara lunak, secara signifikan mengungguli pendekatan berbasis CNN tradisional. Dibandingkan dengan penelitian sebelumnya menggunakan dataset dan validasi *holdout* yang sama, model *ensemble* yang diusulkan mencapai akurasi pengujian terbaik sebesar 98,5%, menunjukkan kinerja klasifikasi yang unggul. Hal ini menyoroti efektivitas penggabungan berbagai arsitektur untuk meningkatkan akurasi diagnostik dalam tugas klasifikasi penyakit kulit multikelas.

5. KESIMPULAN

Penelitian ini memperkenalkan kerangka kerja pembelajaran mendalam *ensemble* yang mengintegrasikan *Swin Transformer*, *Vision Transformer*, dan *EfficientNetB4* untuk mengatasi tantangan klasifikasi penyakit kulit multikelas. Dengan menggabungkan kekuatan komplementer dari jaringan saraf konvolusional dan arsitektur berbasis transformer, model yang diusulkan mencapai akurasi diagnostik yang tinggi, mencapai 98,5% pada dataset Kaggle.

Hasil ini melampaui beberapa pendekatan terkini. Misalnya, Liu et al. (2024) mengembangkan *SkinNet* menggunakan metode penumpukan *ensemble* dan melaporkan

akurasi 86,7% dan AUC 0,96 pada dataset HAM10000 (Liu et al., 2024). Demikian pula, Thwin dan Park (2024) menggunakan *ensemble* VGG16, *Inception* V3, dan ResNet50, mencapai akurasi hingga 97% pada dataset ISIC yang seimbang (Thwin, 2024).

Dibandingkan dengan metode-metode ini, model kami menunjukkan kinerja yang kompetitif atau superior, terutama dalam skenario klasifikasi multikelas. Lebih lanjut, integrasi model berbasis transformer dan berbasis CNN memungkinkan strategi ekstraksi fitur hibrida yang meningkatkan generalisasi dan ketahanan. Secara keseluruhan, penelitian ini meletakkan dasar yang kuat untuk pengembangan sistem diagnostik yang cerdas, andal, dan mudah diakses untuk analisis penyakit kulit. Kontribusi ini menyediakan solusi praktis dan terukur untuk diagnosis dermatologis berbantuan komputer, khususnya dalam pengaturan klinis di mana deteksi dini dan akurat sangat penting. Dalam penelitian selanjutnya, kami berencana untuk melakukan validasi lintas dataset untuk menilai kemampuan model dalam melakukan generalisasi di berbagai kelompok demografis dan pengaturan klinis.

PUSTAKA

- Ahmed, S., Sakib, K. M. S., & Hany, U. (2026). Hybrid deep learning architecture for skin disease classification. *Franklin Open*, 15(03), 426–451. <https://doi.org/10.1016/j.fraope.2026.100556>
- Asaduzzaman, A., Thompson, C. C., & Uddin, M. J. (2024). Machine Learning Approaches for Skin Neoplasm Diagnosis. *ACS Omega*. <https://doi.org/10.1021/acsomega.4c03640>
- Cha, J. (2022). *Features Selection and Extreme Learning Machine*.
- Chiu, T. M., Chi, I. C., Li, Y. C., & Tseng, M. H. (2025). Deep Ensemble Learning for Multiclass Skin Lesion Classification. *Bioengineering*, 12(9), 1–16. <https://doi.org/10.3390/bioengineering12090934>
- Deshmukh, A. A., Wanjale, K., Jadhav, T. A., Khankal, D. V., Diwate, A. D., & Athawale, S. V. (2023). Multi-Class Skin Diseases Classification Using Hybrid Deep Convolutional Neural Network. *International Journal of Intelligent Systems and Applications in Engineering*, 11(10s), 11–22.
- Houssein, E. H., Abdelkareem, D. A., Hu, G., Hameed, M. A., Ibrahim, I. A., & Younan, M. (2024). An effective multiclass skin cancer classification approach based on deep convolutional neural network. *Cluster Computing*, 27(9), 12799–12819. <https://doi.org/10.1007/s10586-024-04540-1>
- Ibrahim, A. T., Abdullahi, M., Kana, A. F. D., Mohammed, M. T., & Hassan, I. H. (2025). Categorical classification of skin cancer using a weighted ensemble of transfer learning with test time augmentation. *Data Science and Management*, 8(2), 174–184. <https://doi.org/10.1016/j.dsm.2024.10.002>
- K, M. R., Krishna, G. S., Supriya, K., & Sorgile, M. (2025). LesionAid: vision transformers-based skin lesion generation and classification – A practical review. *Multimedia Tools and Applications*, 84(33), 41405–41426. <https://doi.org/10.1007/s11042-025-20797-z>
- Kim, G. K., & Del Rosso, J. Q. (2009). Skin cancer in Asians part 1: Nonmelanoma skin cancer. *Journal of Clinical and Aesthetic Dermatology*, 2(8), 39–42.
- Liu, X., Yu, Z., Tan, L., Yan, Y., & Shi, G. (2024). Enhancing Skin Lesion Diagnosis with Ensemble Learning. *2024 4th International Conference on Computer Science and Blockchain, CCSB 2024*, 94–98. <https://doi.org/10.1109/CCSB63463.2024.10735675>
- Lubis, C., & Yulianto, D. (2023). *Menggunakan Convolutional Neural Network (Cnn)*

- Dengan Arsitektur Vgg16*. 8(1), 135–140.
- Maheswari, S., & Gunawan, D. (2025). Deteksi Dini Kanker Kulit Menggunakan Cnn, Dnn, Dan Efficientnet: Pendekatan Deep Learning Berbasis Web. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2), 932–944. <https://doi.org/10.36341/rabit.v10i2.6417>
- Nurkhasanah, N., & Murinto, M. (2022). Klasifikasi Penyakit Kulit Wajah Menggunakan Metode Convolutional Neural Network. *Sainteks*, 18(2), 183. <https://doi.org/10.30595/sainteks.v18i2.13188>
- Pendi, D., Doni, K., & Convolutional, M. (2025). *Jurnal Teknologi Terpadu NETWORK MENGGUNAKAN ARSITEKTUR VGG19*. 11(2), 87–93.
- Rachmawati, S., Minarno, A. E., & Azhar, Y. (2024). Klasifikasi Lesi Kulit Menggunakan Convolutional Neural Network. *Jurnal Repositor*, 4(3), 363–372. <https://doi.org/10.22219/repositor.v4i3.32327>
- Sao, J., & Kushwaha, S. (2024). Automated Multi-Class Skin Disease Classification Via Convolutional Neural Network in Matlab Environment. *ShodhKosh: Journal of Visual and Performing Arts*, 5(6), 1930–1937. <https://doi.org/10.29121/shodhkosh.v5.i6.2024.5107>
- Setiyani, S. H., & Krissyano, Y. N. (2026). *Penerapan Convolutional Neural Network (CNN) Untuk Deteksi dan Klasifikasi Penyakit Kulit Pada Citra Dermatoskop*. 8–16.
- Sinaga, J. T., Faudyta, H. A., & Subhiyacto, E. R. (2024). Klasifikasi Kanker Kulit menggunakan Convolutional Neural Network dengan Optimasi Arsitektur. *Building of Informatics, Technology and Science (BITS)*, 6(3), 446–457. <https://doi.org/10.47065/bits.v6i3.6141>
- Taşar, B. (2023). SkinCancerNet: Automated Classification of Skin Lesion Using Deep Transfer Learning Method. *Traitement Du Signal*, 40(1), 285–295. <https://doi.org/10.18280/ts.400128>
- Thwin, S. M. (2018). Applied Sciences. *Early Writings on India*, 124–134. <https://doi.org/10.4324/9781315232140-14>
- Vuran, S., Ucan, M., Akin, M., & Kaya, M. (2025). Multi-Classification of Skin Lesion Images Including Mpox Disease Using Transformer-Based Deep Learning Architectures. *Diagnostics*, 15(3). <https://doi.org/10.3390/diagnostics15030374>
- Wona Maria Misela A, Rahayu Wanti, & Wirantasa Umar. (2025). Klasifikasi Dan Deteksi Penyakit Kulit Melalui Pengolahan Citra Dengan Metode Cnn. *Jurnal Riset Dan Aplikasi Mahasiswa Informatika (JRAMI)*, 06(01), 43–51.
- Ye, C., Li, J., & Shuai, Q. (2025). *Evaluating the Performance and Clinical Applications of Multiclass Deep Learning Models for Skin Cancer Pathology Diagnosis (ISIC): A Comparative Analysis of CNN, ViT, and VLM*. February 2025, 92–103. <https://doi.org/10.1145/3731763.3731793>