

Komparasi Akurasi dan Interpretabilitas CNN, Transformer, dan ConvNeXt pada Lesi Kulit

Comparing the Accuracy and Interpretability of CNN, Transformer, and ConvNeXt for Skin Lesion Classification

¹M. Denny Richardo*

¹Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

¹Jl. Srijaya Negara, Bukit Besar, Kec. Ilir Barat I, Kota Palembang 30128, Sumatera Selatan, Indonesia

*e-mail: mdennyrichardo@gmail.com

(received: 2 May 2026, revised: 30 June 2026, accepted: 28 July 2026)

Abstrak

Deteksi dini kanker kulit berperan penting dalam meningkatkan peluang keberhasilan pengobatan, khususnya pada kasus melanoma. Berbagai penelitian telah memanfaatkan model deep learning untuk klasifikasi lesi kulit, namun sebagian besar masih berfokus pada satu jenis arsitektur serta lebih menekankan peningkatan akurasi dibandingkan interpretabilitas model. Penelitian ini bertujuan melakukan analisis komparatif terhadap tiga paradigma arsitektur deep learning, yaitu EfficientNet-B4 sebagai representasi Convolutional Neural Network (CNN), Swin Transformer sebagai representasi Vision Transformer (ViT), dan ConvNeXt sebagai representasi Modernized CNN pada klasifikasi tujuh jenis lesi kulit menggunakan dataset HAM10000. Seluruh model dilatih menggunakan konfigurasi pelatihan yang seragam dengan penerapan analytical class weighting untuk mengatasi ketidakseimbangan kelas. Evaluasi dilakukan menggunakan Accuracy, Precision, Recall, Macro F1-Score, Confusion Matrix, serta Explainable AI melalui Grad-CAM dan Attention Maps untuk menilai transparansi proses pengambilan keputusan model. Hasil eksperimen menunjukkan bahwa ConvNeXt memberikan performa terbaik dengan akurasi sebesar 91,62% dan Macro F1-Score sebesar 0,8897, serta menghasilkan nilai Recall yang lebih tinggi pada kelas Melanoma dibandingkan EfficientNet-B4 dan Swin Transformer. Visualisasi Explainable AI juga menunjukkan bahwa ConvNeXt mampu memfokuskan perhatian pada area lesi yang lebih relevan secara klinis dibandingkan kedua model lainnya. Temuan ini menunjukkan bahwa ConvNeXt memberikan keseimbangan yang lebih baik antara performa klasifikasi dan interpretabilitas sehingga berpotensi menjadi pendekatan yang andal untuk mendukung sistem diagnosis lesi kulit berbasis kecerdasan buatan.

Kata Kunci: ConvNeXt, Explainable AI, Grad-CAM, kanker kulit, swin transformer.

Abstract

Early detection of skin cancer plays an important role in improving treatment outcomes, particularly in cases of melanoma. Numerous studies have employed deep learning models for skin lesion classification; however, most have focused on a single architectural paradigm and placed greater emphasis on improving classification accuracy than on model interpretability. This study presents a comparative analysis of three deep learning architectural paradigms: EfficientNet-B4 as a representative Convolutional Neural Network (CNN), Swin Transformer as a representative Vision Transformer (ViT), and ConvNeXt as a representative modernized CNN for the classification of seven types of skin lesions using the HAM10000 dataset. All models were trained using consistent training configurations, with analytical class weighting applied to address class imbalance. Model performance was evaluated using accuracy, precision, recall, Macro F1-score, and confusion matrices, while Explainable AI (XAI) techniques, including Grad-CAM and attention maps, were employed to assess the transparency of the models' decision-making processes. The experimental results show that ConvNeXt achieved the best overall performance, with an accuracy of 91.62% and a Macro F1-score of 0.8897. It also achieved higher recall for the melanoma class than EfficientNet-B4

and Swin Transformer. The XAI visualizations further demonstrate that ConvNeXt was able to focus its attention on lesion regions that were more clinically relevant compared with the other two models. These findings indicate that ConvNeXt provides a better balance between classification performance and model interpretability, suggesting its potential as a reliable approach for supporting artificial intelligence-based skin lesion diagnosis systems.

Keywords: ConvNeXt, Explainable AI, Grad-CAM, skin cancer, swin transformer.

1 Pendahuluan

Kanker kulit tetap menjadi masalah kesehatan masyarakat yang semakin meningkat di seluruh dunia. Pada saat yang sama, diagnosis dini sangat penting untuk memungkinkan keberhasilan pengobatan. Namun, lesi kulit menunjukkan tingkat kemiripan antar-kelas yang tinggi (seperti melanoma agresif vs nevus jinak), sehingga diagnosis manual oleh tenaga medis menjadi sulit. Beberapa penelitian tentang otomatisasi deteksi penyakit kulit secara bertahap menunjukkan kemajuan di Indonesia, yaitu melalui metode berbasis arsitektur CNN untuk meningkatkan akurasi diagnostik awal [1]. Sistem cerdas semacam ini dapat membantu dokter kulit dalam melakukan skrining yang cepat dan akurat.

Meskipun arsitektur CNN konvensional telah memberikan kontribusi besar, model tersebut memiliki keterbatasan inheren dalam menangkap korelasi spasial global karena sifat operasinya yang berbasis lokal (piksel tetangga). Hal ini mendorong munculnya paradigma baru melalui Vision Transformer (ViT). Arsitektur seperti Swin Transformer memperkenalkan mekanisme shifted windows yang memungkinkan model untuk memahami konteks gambar secara menyeluruh, yang terbukti sangat efektif pada berbagai tugas visi komputer yang kompleks [2]. Namun, di sisi lain, muncul sebuah perdebatan akademik mengenai apakah arsitektur transformer murni benar-benar lebih unggul daripada CNN untuk data medis. Hal ini memicu lahirnya ConvNeXt, sebuah arsitektur yang memodernisasi desain konvolusi dengan mengadopsi prinsip-prinsip transformer guna mencapai performa yang lebih kompetitif tanpa kehilangan efisiensi spasial [3]. Masalah krusial lainnya yang sering diabaikan adalah ketidakseimbangan kelas (class imbalance) pada dataset medis seperti HAM10000, serta sifat model deep learning yang bersifat black-box. Tanpa penjelasan visual yang jelas, hasil diagnosis mesin sulit diterima sepenuhnya oleh praktisi medis. Oleh karena itu, diperlukan integrasi teknik Explainable AI (XAI) untuk memberikan transparansi pada keputusan model [4].

Kami melakukan studi komparatif yang ekstensif terhadap tiga paradigma arsitektur SOTA, yaitu EfficientNet-B4 (CNN tradisional), Swin Transformer (Transformer murni) dan ConvNeXt (CNN modern). Kebaruan penelitian terletak pada pendekatan hibrida yang menggabungkan optimalisasi performa melalui analytical class weighting dengan validasi kualitatif berbasis XAI. Studi ini tidak hanya mengevaluasi arsitektur mana yang mencapai akurasi tertinggi, tetapi juga arsitektur mana yang memiliki lokalisasi fitur paling akurat secara klinis menggunakan Grad-CAM dan Attention Maps. Hasil penelitian ini diharapkan dapat memberikan referensi dalam pemilihan arsitektur deep learning yang memiliki keseimbangan antara performa klasifikasi dan interpretabilitas untuk mendukung pengembangan sistem diagnosis lesi kulit berbasis kecerdasan buatan.

2 Tinjauan Literatur

Penelitian mengenai klasifikasi lesi kulit telah didominasi oleh penggunaan Convolutional Neural Networks (CNN) dalam beberapa tahun terakhir. Arsitektur EfficientNet-B4 menjadi populer karena efisiensi parameternya melalui mekanisme compound scaling yang menyeimbangkan kedalaman, lebar, dan resolusi jaringan [5]. Beberapa studi awal menunjukkan bahwa CNN mampu mencapai akurasi tinggi pada dataset HAM10000, namun sering kali mengalami penurunan performa ketika berhadapan dengan masalah ketidakseimbangan kelas (class imbalance) yang ekstrem [6]. Strategi augmentasi data dan penggunaan weighted cross-entropy telah diusulkan sebagai solusi untuk meningkatkan sensitivitas terhadap kelas minoritas seperti Melanoma dan Dermatofibroma [7].

Meskipun demikian, sifat operasi konvolusi yang terbatas pada reseptor lokal tetap menjadi tantangan dalam menangkap karakteristik tekstur lesi yang tersebar secara global [8]. Sebagai alternatif terhadap keterbatasan CNN, Vision Transformer (ViT) memperkenalkan mekanisme self-attention yang memungkinkan model untuk menangkap hubungan antar-piksel di seluruh citra tanpa batasan spasial lokal [9]. Pengembangan lebih lanjut melalui Swin Transformer memperkenalkan

<http://sistemasi.ftik.unisi.ac.id>

konsep hierarki dan shifted windows, yang secara signifikan mengurangi kompleksitas komputasi sekaligus meningkatkan kemampuan model dalam mengenali fitur pada berbagai skala [10]. Beberapa penelitian terbaru (2023-2025) menunjukkan bahwa Swin Transformer memiliki keunggulan dalam mendeteksi tepi lesi yang tidak beraturan, yang merupakan indikator kunci keganasan pada kanker kulit [11], [12]. Namun, efektivitas ViT pada dataset medis berukuran menengah masih menjadi subjek perdebatan dibandingkan dengan pendekatan konvolusi yang telah matang.

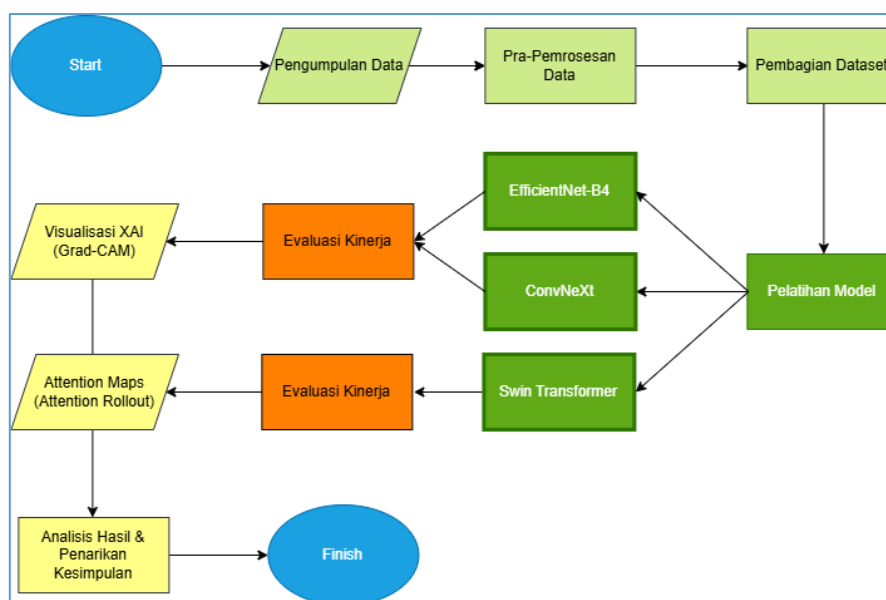
Di tengah popularitas Transformer, arsitektur ConvNeXt muncul sebagai upaya untuk membuktikan bahwa desain konvolusi yang dimodernisasi dapat menandingi performa ViT. Dengan mengadopsi teknik large kernel size, inverted bottlenecks, dan penggunaan aktivasi GeLU, ConvNeXt merepresentasikan evolusi terbaru dari CNN yang lebih tangguh. Studi komparatif menunjukkan bahwa ConvNeXt sering kali memberikan stabilitas pelatihan yang lebih baik pada citra medis dengan noise tinggi dibandingkan arsitektur berbasis attention murni. Keunggulan ini menjadikan ConvNeXt sebagai kandidat kuat untuk klasifikasi kanker kulit yang memerlukan presisi tinggi pada fitur tekstur dan bentuk secara simultan. Kebutuhan akan model yang dapat diinterpretasikan dalam dunia medis mendorong integrasi teknik Explainable AI (XAI). Grad-CAM (Gradient-weighted Class Activation Mapping) telah menjadi standar de facto untuk memvisualisasikan area penting dalam citra yang mendasari prediksi model CNN. Sementara itu, pada model berbasis Transformer, penggunaan Attention Maps memberikan gambaran tentang bagaimana bobot perhatian didistribusikan pada seluruh bagian lesi. Penggunaan XAI secara sinkron pada berbagai arsitektur model tidak hanya berfungsi sebagai alat validasi bagi dokter, tetapi juga sebagai metode untuk mendeteksi adanya bias pada data latih.

Berdasarkan tinjauan literatur tersebut, dapat diidentifikasi bahwa masih terdapat beberapa kesenjangan penelitian (research gap). Sebagian besar penelitian terdahulu hanya berfokus pada pengembangan atau evaluasi satu jenis arsitektur deep learning, baik Convolutional Neural Network (CNN) maupun Vision Transformer (ViT), sehingga belum memberikan perbandingan yang komprehensif antara CNN konvensional, Vision Transformer, dan Modernized CNN dalam kondisi eksperimen yang seragam. Selain itu, evaluasi yang dilakukan umumnya masih berorientasi pada peningkatan performa klasifikasi berdasarkan metrik kuantitatif, sedangkan analisis mengenai interpretabilitas model melalui Explainable AI (XAI) masih relatif terbatas. Oleh karena itu, penelitian ini difokuskan pada analisis komparatif EfficientNet-B4, Swin Transformer, dan ConvNeXt menggunakan dataset HAM10000 dengan konfigurasi pelatihan yang sama, serta mengintegrasikan evaluasi kuantitatif dan Explainable AI untuk memperoleh model yang tidak hanya memiliki performa klasifikasi yang tinggi, tetapi juga menghasilkan proses pengambilan keputusan yang lebih transparan dan dapat dipercaya pada aplikasi diagnosis lesi kulit.

3 Metode Penelitian

3.1. Alur Penelitian

Desain penelitian komparatif ini terdiri dari serangkaian tahapan komputasi yang terstruktur dan menyeluruh, sebagaimana diilustrasikan pada diagram alur (Gambar 1). Pendekatan yang digunakan mengadopsi evaluasi dua arah, analisis performa kuantitatif untuk mengukur metrik keandalan dan analisis visual kualitatif untuk memastikan transparansi arsitektur deep learning. Eksperimen diawali dengan akuisisi dataset klinis, diikuti oleh rekayasa data melalui pra-pemrosesan dan penanganan ketidakseimbangan kelas. Penelitian ini berfokus pada eksperimen paralel yang melatih tiga filosofi arsitektur State-of-the-Art secara independen dengan parameter yang seragam. Standardisasi alur kerja diterapkan secara ketat untuk memastikan komparasi metrik dan peta atensi yang dihasilkan valid, reliabel, dan dapat direplikasi dalam skenario klinis yang nyata.



Gambar 1 Alur penelitian

3.2. Pengumpulan Data

Dataset yang menjadi subjek utama dalam eksperimen ini adalah HAM10000 (Human Against Machine with 10000 training images), sebuah repositori citra dermoskopi berskala besar yang dikurasi secara profesional. Dataset ini mencakup 10.015 citra resolusi tinggi yang merepresentasikan tujuh kategori lesi kulit dengan tingkat prevalensi tinggi dalam diagnosis dermatologi, yaitu Actinic keratoses (akiec), Basal cell carcinoma (bcc), Benign keratosis-like lesions (bkl), Dermatofibroma (df), Melanoma (mel), Melanocytic nevi (nv), dan Vascular lesions (vasc). Satu tahapan krusial dalam proses pengumpulan data ini adalah integrasi file metadata berekstensi .csv yang melampirkan atribut demografis dan klinis untuk setiap sampel. Penggunaan file metadata ini bersifat esensial dalam alur kerja sistem karena berfungsi sebagai jembatan pemetaan yang menghubungkan identitas unik citra (Image ID) dengan ground truth diagnosis penyakit aslinya. Tanpa sinkronisasi metadata ini, proses labeling otomatis tidak dapat dilakukan secara presisi, sehingga berisiko menyebabkan poisoning data selama fase pelatihan model. Berikut jumlah dan contoh dataset yang digunakan dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1 Distribusi kelas pada dataset HAM10000

Label	Nama Kelas (diagnosa)	Jumlah Citra	Persentase
nv	Melanocytic nevi	6.705	66.95%
mel	Melanoma	1.113	11.11%
bkl	Benign keratosis-like lesions	1.099	10.97%
bcc	Basal cell carcinoma	514	5.13%
akiec	Actinic keratoses	327	3.27%
vasc	Vascular lesions	142	1.42%
df	Dermatofibroma	115	1.15%
Total		10.015	100.00%

Terlihat adanya kesenjangan yang sangat ekstrem antara kelas mayoritas (nv) dan kelas minoritas (df). Hal inilah yang mendasari penggunaan teknik Analytical Class Weighting agar model tidak mengalami bias saat proses pembelajaran.

3.3. Pra-pemrosesan Data

Karakteristik citra medis mentah yang rentan terhadap distorsi visual dan dominasi kelas tertentu menuntut adanya tahapan pra-pemrosesan yang berlapis. Distribusi kelas pada dataset HAM10000 sangat bias ke arah kelas *Melanocytic nevi* (nv) yang mendominasi lebih dari 60% total populasi data. Untuk mengatasi kendala ini, empat langkah rekayasa data diterapkan sebelum citra diumpankan ke dalam model:

1. Standardisasi Resolusi Spasial (Resizing): Seluruh citra diekstraksi dan direduksi resolusinya menjadi dimensi spasial 224x224 piksel. Proses ini menggunakan metode interpolasi *bilinear* untuk meminimalisir hilangnya detail fitur morfologis lesi, sekaligus menyesuaikan tensor input standar yang disyaratkan oleh arsitektur EfficientNet, Swin Transformer, dan ConvNeXt.
2. Normalisasi Berbasis ImageNet: Nilai intensitas piksel (RGB) dinormalisasi menggunakan distribusi statistik (*mean* dan *standard deviation*) yang diturunkan dari dataset ImageNet. Hal ini bertujuan untuk menstabilkan dan mempercepat proses konvergensi algoritma optimasi selama fase *backpropagation*.
3. Augmentasi Data Geometris: Untuk memaksa model mempelajari invarian fitur dari suatu lesi, rotasi acak, *horizontal flip*, dan *vertical flip* diaplikasikan. Teknik ini mensimulasikan berbagai sudut pengambilan gambar dermoskop, sehingga arsitektur tidak hanya menghafal orientasi spesifik dari data latih.
4. Analytical Class Weighting: Sebagai intervensi algoritmik terhadap *class imbalance*, penelitian ini menerapkan pembobotan kelas sintetis pada fungsi *Cross-Entropy Loss*. Bobot ini dihitung secara terbalik (*inversely proportional*) terhadap frekuensi kemunculan setiap kelas di *training set*. Mekanisme ini memberikan penalti error yang jauh lebih besar (*heavy loss penalty*) apabila model salah memprediksi kelas minoritas yang krusial seperti Melanoma, memaksa model untuk tidak sekadar memprediksi kelas mayoritas demi mengejar akurasi semu.

3.4. Pembagian Dataset

Untuk menjamin pengujian yang berkeadilan, total 10.015 populasi data dipecah ke dalam subset Training Set (untuk pembaruan bobot model) dan Test Set (untuk evaluasi akhir yang tidak pernah dilihat model sebelumnya) dengan rasio standar 80:20. Pembagian ini tidak dilakukan secara acak sederhana, melainkan menggunakan teknik Stratified Sampling. Pendekatan stratifikasi ini menjamin bahwa persentase distribusi dari ketujuh kelas lesi di dalam subset pengujian sama persis dengan proporsi pada dataset populasi asli. Hal ini krusial untuk mencegah terjadinya anomali statistik di mana kelas dengan populasi sangat kecil (seperti Dermatofibroma) tidak terwakili secara memadai saat fase pengujian arsitektur.

3.5. Implementasi dan Pelatihan Model

Eksperimen komputasi dijalankan secara paralel terhadap tiga arsitektur *deep learning* yang mengusung filosofi desain jaringan saraf yang berbeda:

1. EfficientNet-B4 (Baseline CNN): Arsitektur ini dipilih sebagai representasi puncak dari CNN konvensional. Menggunakan konsep *compound scaling*, model ini secara simultan menyeimbangkan kedalaman jaringan, lebar channel, dan resolusi citra menggunakan blok Mobile Inverted Bottleneck (MBConv).
2. Swin Transformer (Representasi ViT): Sebagai antitesis dari operasi konvolusi lokal, Swin Transformer mengandalkan mekanisme Shifted Windows Multi-Head Self-Attention (SW-MSA). Model ini memecah citra menjadi sekumpulan patches dan menghitung relasi atensi antarpiksel pada skala global, sehingga memungkinkan pertukaran informasi antarkomponen spasial yang berjauhan.
3. ConvNeXt (Representasi Modernized CNN): Model ini merepresentasikan evolusi arsitektur konvolusi yang meniru paradigma transformer. ConvNeXt mengimplementasikan *depthwise*

<http://sistemasi.ftik.unisi.ac.id>

convolution dengan ukuran kernel raksasa (7 x 7), mengganti aktivasi ReLU dengan GeLU, serta menggunakan *Layer Normalization* sebagai pengganti *Batch Normalization*. Arsitektur ini diuji untuk memvalidasi apakah struktur CNN modern dapat melampaui kemampuan *self-attention* transformer.

Mengingat durasi pelatihan yang relatif singkat (10-12 epoch) dan untuk mempercepat konvergensi model pada dataset medis yang kompleks, penelitian ini mengimplementasikan teknik Transfer Learning. Seluruh model (EfficientNet-B4, Swin Transformer, dan ConvNeXt) diinisialisasi menggunakan bobot pra-terlatih (pretrained weights) dari dataset ImageNet. Pendekatan ini memungkinkan model untuk memanfaatkan fitur dasar visual secara lebih efisien sebelum dilakukan proses fine-tuning secara spesifik pada dataset HAM10000. Seluruh model dilatih di atas kerangka kerja PyTorch (*timm library*) pada lingkungan komputasi Google Colab menggunakan akselerasi GPU. Parameter optimasi diatur secara seragam dengan *optimizer* AdamW, *learning rate* awal sebesar $1e-4$, *weight decay* untuk mencegah *overfitting*, serta *batch size* 16.

3.6. Evaluasi Kinerja Kuantitatif

Pengukuran efektivitas model tidak hanya bergantung pada metrik akurasi global, yang sering kali menghasilkan "paradoks akurasi" pada dataset yang tidak seimbang. Evaluasi diperdalam dengan mengkalkulasi matriks kebingungan (Confusion Matrix) untuk melacak presisi prediksi antar kelas. Kinerja puncak dinilai berdasarkan metrik Precision, Recall (sensitivitas terhadap diagnosis kelas ganas), serta metrik utama berupa Macro F1-Score. Penggunaan Macro F1-Score memastikan bahwa performa model pada kelas minoritas (seperti Vascular lesions) dianggap sama pentingnya dengan performa pada kelas mayoritas, yang merepresentasikan ketangguhan model secara klinis.

3.7. Analisis Interpretabel (Explainable AI / XAI)

Untuk mendekonstruksi sifat "kotak hitam" dari *deep learning* dan membangun kepercayaan (*trust*) bagi tenaga medis, teknik *Explainable AI* diaplikasikan pada fase inferensi akhir. Untuk rumpun model berbasis konvolusi (EfficientNet-B4 dan ConvNeXt), teknik *Gradient-weighted Class Activation Mapping* (Grad-CAM) digunakan. Algoritma ini mengekstrak informasi gradien yang mengalir ke lapisan konvolusi spasial terakhir untuk menghasilkan peta panas (*heatmap*) yang menyoroti area regional paling berpengaruh terhadap keputusan diagnosis. Sementara itu, untuk arsitektur Swin Transformer, ekstraksi peta atensi (*Attention Maps*) dilakukan dengan memvisualisasikan matriks bobot dari modul *self-attention* untuk memetakan area mana pada lesi kulit yang saling berinteraksi dan menjadi fokus utama jaringan selama proses klasifikasi.

4 Hasil dan Pembahasan

4.1. Evaluasi Kinerja Klasifikasi Kuantitatif

Evaluasi metrik kuantitatif merupakan tolok ukur utama untuk membandingkan kemampuan ekstraksi fitur antara CNN konvensional (EfficientNet-B4), arsitektur *vision transformer* (Swin Transformer), dan CNN modern (ConvNeXt). Kinerja model dievaluasi menggunakan *Classification Report* pada *test set*. Tabel 2 merangkum hasil perbandingan metrik Akurasi, *Macro F1-Score*, dan *Weighted F1-Score* dari ketiga arsitektur tersebut.

Tabel 2 Perbandingan metrik evaluasi kinerja model

Arsitektur Model	Akurasi	Macro F1-Score	Weighted F1-Score
EfficientNet-B4	82,83%	0,7215	0,8309
Swin Transformer	84,93%	0,7848	0,8562
ConvNeXt	91,62%	0,8897	0,9170

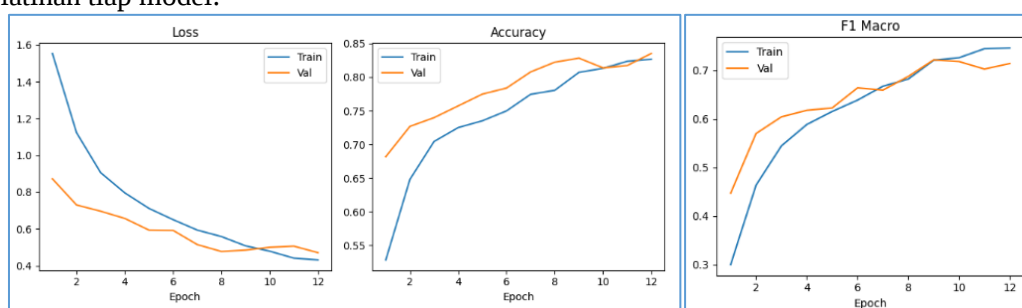
Berdasarkan Tabel 2, evolusi desain arsitektur memberikan dampak yang sangat signifikan terhadap kemampuan klasifikasi. EfficientNet-B4 sebagai *baseline* mencapai akurasi sebesar 82,83% dengan *Macro F1-Score* 0,7215. Implementasi Swin Transformer berhasil meningkatkan akurasi

menjadi 84,93% dengan *Macro F1-Score* 0,7848. ConvNeXt menunjukkan kinerja terbaik di antara ketiga model yang dievaluasi dengan rekor akurasi 91,62% dan *Macro F1-Score* yang sangat stabil di angka 0,8897.

Keunggulan ConvNeXt sangat krusial dalam konteks klinis, khususnya pada deteksi kelas minoritas yang mematikan, seperti Melanoma (mel). Berdasarkan laporan klasifikasi, tingkat sensitivitas (*Recall*) untuk kelas Melanoma mengalami peningkatan yang signifikan dari 0,6188 (61,88%) pada EfficientNet-B4, naik menjadi 0,7679 (76,79%) pada Swin Transformer, dan mencapai puncaknya pada 0,8125 (81,25%) pada ConvNeXt. Hal ini mengindikasikan kemampuan superior ConvNeXt dalam menunjukkan kemampuan yang lebih baik dalam mengidentifikasi kasus Melanoma sehingga meningkatkan nilai Recall dibandingkan model lainnya.

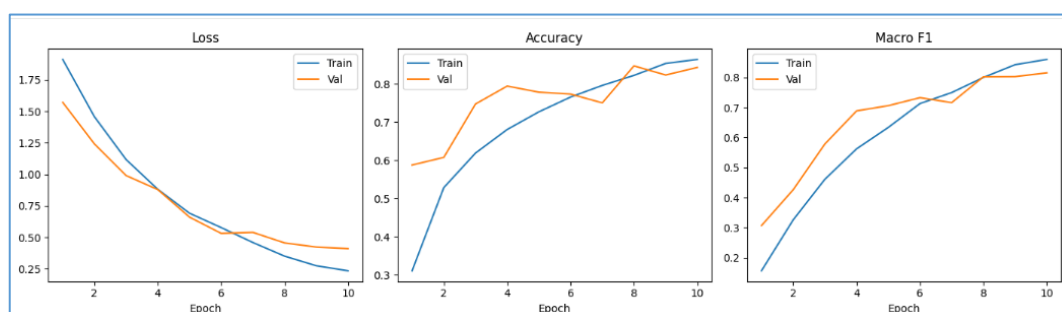
4.2. Analisis Stabilitas Pembelajaran (Learning Curves)

Analisis grafik historis pelatihan (*Loss*, Akurasi, dan *Macro F1*) digunakan untuk mengamati stabilitas konvergensi model selama proses *training* dan *validation*. Berdasarkan visualisasi grafik pelatihan, terdapat perbedaan karakteristik konvergensi yang mencolok antar model. Berikut gambar dari pelatihan tiap model.



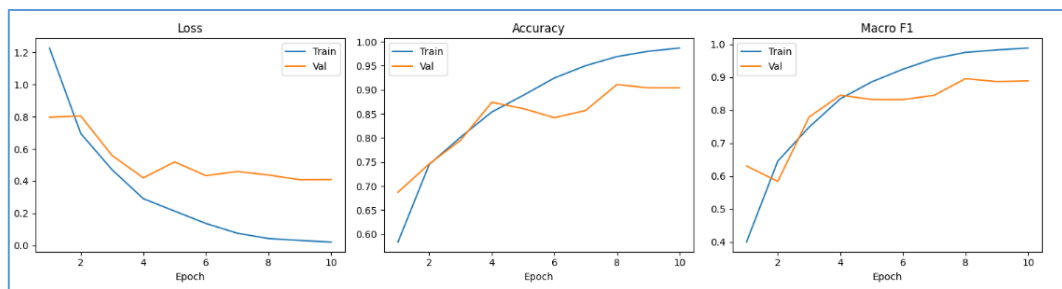
Gambar 2 Learning curves model EfficientNet-B4

Pada arsitektur EfficientNet-B4 (*gambar 2*), kurva akurasi validasi (*val accuracy*) terlihat mengalami fluktuasi dan mulai mendatar (*plateau*) di kisaran 0,80 hingga 0,85 seiring berjalannya iterasi. Di sisi lain, kurva *Loss* dan Akurasi dari Swin Transformer menunjukkan pergerakan yang lebih rapat antara data latih dan data validasi, mengindikasikan ketahanan model terhadap *overfitting* yang lebih baik dibandingkan CNN standar. Pada *gambar 3* Kurva menunjukkan konvergensi yang lebih lambat di awal namun sangat stabil pada tahap akhir. Jarak antara akurasi *training* dan *validation* sangat tipis, membuktikan ketahanan arsitektur *Transformer* dalam melakukan generalisasi fitur.



Gambar 3 Learning curves model swin transformer

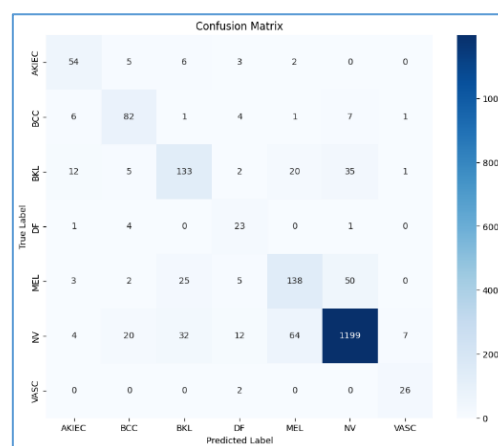
Kendati demikian, kurva konvergensi paling halus (*smooth*) dan stabil ditunjukkan oleh ConvNeXt (*gambar 4*). Kurva *Validation Accuracy* pada ConvNeXt secara konsisten mengikuti kurva *Training Accuracy* hingga mendekati batas atas 0,95 (95%) tanpa menunjukkan divergensi yang signifikan. Begitu pula dengan kurva *Macro F1* yang stabil, mengindikasikan bahwa ConvNeXt memiliki kemampuan mempelajari distribusi fitur seluruh kelas—baik mayoritas maupun minoritas—secara proporsional sejak *epoch* awal.



Gambar 4 Learning curves model ConvNeXt

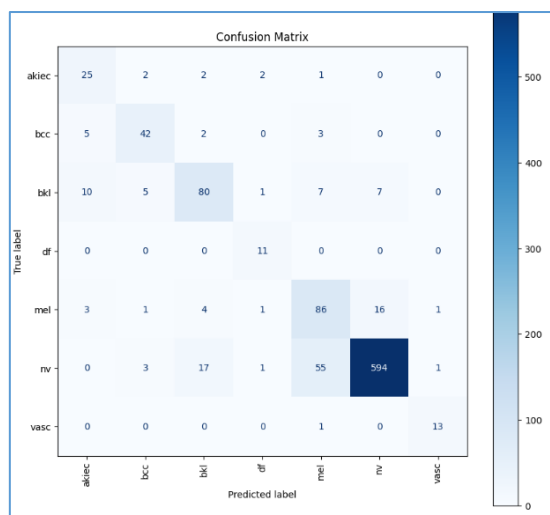
4.3. Analisis Kesalahan Prediksi (Confusion Matrix)

Untuk membedah lebih dalam mengenai distribusi kesalahan prediksi antar kelas, matriks kebingungan (*confusion matrix*) dari ketiga model dievaluasi. Fokus utama diletakkan pada kelas dengan tingkat kesulitan visual yang tinggi, yaitu pemisahan antara kelas Melanoma ganas (mel) dengan *Melanocytic nevi* jinak (nv). Berikut gambar dari confusion matrix tiap model.



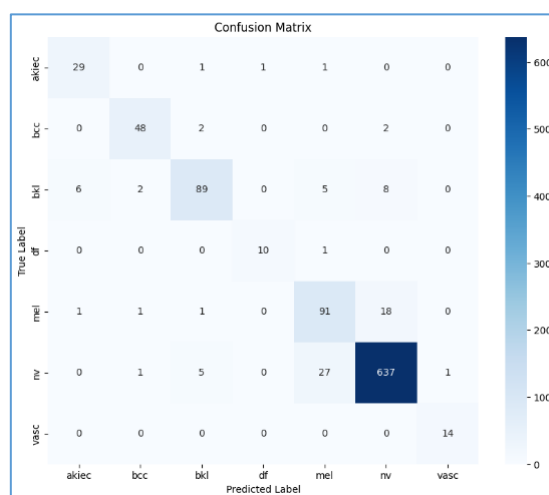
Gambar 5 Confusion matrix model EfficientNet-B4

Confusion matrix pada arsitektur EfficientNet-B4 (*gambar 5*) menunjukkan karakteristik tipikal dari model CNN konvensional saat berhadapan dengan dataset yang sangat tidak seimbang (*imbalanced*). Meskipun model mampu mengidentifikasi kelas mayoritas *Melanocytic nevi* (nv) dengan sangat baik, yaitu mencapai 1.199 prediksi benar, terdapat kebocoran data yang signifikan pada kelas-kelas minoritas lainnya. Fenomena ini terlihat jelas pada kelas Melanoma (mel), di mana sebanyak 136 sampel ganas justru terdeteksi secara keliru sebagai Nevus yang bersifat jinak. Kesalahan prediksi ini merupakan indikator bahwa EfficientNet-B4 cenderung mengalami bias terhadap kelas mayoritas, meskipun teknik *class weighting* telah diupayakan dalam fase pelatihan. Selain itu, terdapat 35 sampel *Benign keratosis-like lesions* (bkl) yang juga salah diklasifikasikan sebagai Nevus, menunjukkan bahwa model kesulitan membedakan tekstur morfologi yang serupa antara kedua jenis lesi tersebut. Pada kelas *Basal cell carcinoma* (bcc), model hanya berhasil mengidentifikasi 82 sampel dengan benar, sementara sisanya tersebar pada kelas *akiec*, *bkl*, dan *mel*. Model ini juga menunjukkan keterbatasan dalam mengenali *Dermatofibroma* (df), dengan hanya 23 prediksi yang tepat dari total sampel. Kelemahan ini dipertegas oleh angka *Recall* untuk kelas *mel* yang hanya menyentuh 61,88%, sebuah angka yang cukup berisiko dalam skenario klinis nyata. Ketidakmampuan model dalam memisahkan fitur kritis antara lesi ganas dan jinak sering kali disebabkan oleh operasi konvolusi kernel kecil (3×3) yang kurang mampu menangkap konteks spasial yang lebih luas. Secara keseluruhan, matriks EfficientNet-B4 menggambarkan model yang masih didominasi oleh fitur dominan dari kelas mayoritas, sehingga mengurangi tingkat kepercayaan (*reliability*) diagnosa untuk penyakit kulit yang bersifat langka namun mematikan.



Gambar 6 Confusion matrix model swin transformer

Implementasi Swin Transformer membawa perubahan signifikan dalam pola distribusi prediksi melalui mekanisme *self-attention* yang dimilikinya. Berdasarkan confusion matrix pada gambar 6, model ini menunjukkan peningkatan ketajaman dalam membedakan kelas *Melanoma* dibandingkan dengan EfficientNet-B4. Angka misklasifikasi dari kelas *mel* ke *nv* berhasil ditekan dari 136 sampel menjadi 89 sampel, sebuah progres yang menunjukkan bahwa hubungan fitur global membantu model mengenali pola keganasan dengan lebih baik. Pada kelas mayoritas *Nevus* (*nv*), Swin Transformer mencatatkan 637 prediksi benar dengan tingkat presisi yang sangat tinggi mencapai 96,27%. Namun, model ini masih menyisakan tantangan pada kelas *akiec*, di mana hanya 25 sampel yang terprediksi dengan benar sementara beberapa sampel lainnya masih "tercecer" ke kelas *bcc* dan *mel*. Kemampuan model dalam mengidentifikasi *Vascular lesions* (*vasc*) tetap stabil dengan 14 prediksi tepat, menunjukkan bahwa fitur warna yang khas pada lesi ini dapat ditangkap dengan baik oleh mekanisme jendela (*shifted windows*). Menariknya, Swin Transformer memberikan hasil *Recall* yang jauh lebih baik untuk kelas *Melanoma* yaitu sebesar 76,79%, naik signifikan dari hasil model CNN sebelumnya. Hal ini membuktikan bahwa arsitektur berbasis *attention maps* lebih unggul dalam memetakan dependensi jangka panjang pada citra medis yang memiliki variasi tekstur yang halus. Walaupun demikian, masih terlihat adanya tumpang tindih pada kelas *Benign keratosis-like lesions* (*bkl*), di mana 30 sampel masih salah terprediksi sebagai *mel*. Fluktuasi pada kurva akurasi validasi dalam grafik pelatihan juga tercermin dalam matriks ini, yang menunjukkan bahwa meskipun model cerdas, stabilitasnya dalam membedakan antar-kelas minoritas belum mencapai tahap optimal. Secara sistematis, Swin Transformer mulai menggeser dominasi bias kelas mayoritas menuju klasifikasi yang lebih terdistribusi secara adil di seluruh kategori penyakit.



Gambar 7 Confusion matrix model ConvNeXt

ConvNeXt (*gambar 7*) menunjukkan performa confusion matrix yang paling mendekati ideal, merepresentasikan keunggulan integrasi antara efisiensi konvolusi dan filosofi arsitektur transformer. Diagonal utama pada matriks ini terlihat sangat dominan dan tegas, mengindikasikan bahwa mayoritas sampel pada setiap kelas berhasil diprediksi dengan tepat. Perbaikan paling radikal terlihat pada kelas *Melanoma* (mel), di mana angka kesalahan prediksi yang jatuh ke kelas *Nevus* (nv) berkurang drastis menjadi hanya 18 sampel. Keberhasilan ini sangat krusial karena meningkatkan *Recall* Melanoma ke angka 81,25%, menunjukkan peningkatan dibandingkan dua model yang dievaluasi. Pada kelas mayoritas nv, model ini mencapai performa hampir sempurna dengan 637 prediksi benar dan hanya sedikit sekali sampel yang bocor ke kelas lain. ConvNeXt juga menunjukkan kemampuan luar biasa dalam mengidentifikasi kelas minoritas *Vascular lesions* (vasc) dengan akurasi 100% (14 dari 14 sampel). Untuk kelas *Basal cell carcinoma* (bcc), model secara konsisten memprediksi 48 sampel dengan benar, menunjukkan stabilitas fitur yang sangat kuat. Selain itu, kelas *Dermatofibroma* (df) yang biasanya sulit dikenali, berhasil diidentifikasi sebanyak 10 dari 11 sampel, menghasilkan *F1-score* kelas tersebut sebesar 0,9091. Penggunaan *large kernel* (7×7) pada ConvNeXt terbukti mampu mengekstraksi fitur batas lesi yang tidak teratur dengan jauh lebih presisi dibandingkan dua model sebelumnya. Matriks ini secara keseluruhan mencerminkan mengapa ConvNeXt mencapai akurasi total 91,62%, di mana hampir tidak ada kegagalan identifikasi yang bersifat sistematis pada kategori diagnosa tertentu. Stabilitas yang ditunjukkan pada kurva *Macro F1* dalam grafik pelatihan tervalidasi sepenuhnya melalui distribusi angka-angka pada confusion matrix ini, menjadikannya model yang paling dapat diandalkan secara medis.

4.4. Validasi Klinis melalui Explainable AI (XAI)

Tingginya akurasi harus dibarengi dengan transparansi algoritmik. Pada tahap ini, peta atensi visual diekstraksi dari ketujuh kelas lesi (akiec, bcc, bkl, df, mel, nv, vasc) menggunakan Grad-CAM untuk model berbasis konvolusi dan *Attention Maps* untuk model berbasis *transformer*. Berikut hasil visualisasi tiap kelas dari tiap model yang diujikan bisa dilihat pada gambar 8 di bawah ini.

Hasil Visualisasi XAI (Grad-CAM & Attention Maps)							
Class of Skin Cancer	Akiec / Actinic keratoses	Bcc / Basal cell carcinoma	Bkl / Benign keratosis-like lesions	Df / Dermatofibroma	Mel / Melanoma	Nv / Melanocytic nevi	Vasc / Vascular lesions
Original Image							
EfficientNet-B4							
Swin Transformer							
ConvNeXt							

Gambar 8 Visualisasi XAI (Grad-CAM & attention maps)

Hasil visualisasi membuktikan bahwa EfficientNet-B4 cenderung memberikan lokalisasi gradien yang tersebar (*diffuse*) dan beberapa kali melebar hingga ke latar belakang kulit sehat. Swin Transformer memberikan visualisasi *attention* yang lebih terfokus pada pusat lesi, namun karakteristik operasinya yang berbasis *grid/window* kadang membuat tepian (*border*) kanker tidak tersorot secara penuh. Sebaliknya, ConvNeXt menampilkan heatmap Grad-CAM yang sangat presisi, memvalidasi area morfologis lesi berdasarkan kriteria klinis dermatologi yang dikenal sebagai kriteria ABCDE

(Asymmetry, Border irregularity, Color variation, Diameter, dan Evolving). Pada sampel Melanoma, peta atensi ConvNeXt tidak hanya terpusat pada bagian tengah, melainkan secara spesifik menyoroti Asymmetry (ketidaksimetrisan bentuk lesi) dan Border irregularity (batas tepi lesi yang kabur atau tidak beraturan), yang merupakan indikator patologis utama dari tingkat keganasan kanker kulit. Hal ini memvalidasi secara visual bahwa akurasi kuantitatif 91,62% yang dicapai oleh ConvNeXt dihasilkan dari pengambilan keputusan berbasis fitur biologis dan parameter medis yang benar, bukan sekadar menghafal artefak atau noise dari latar belakang gambar.

5 Kesimpulan

Penelitian ini melakukan komparasi terhadap tiga paradigma arsitektur deep learning, yaitu EfficientNet-B4 sebagai representasi Convolutional Neural Network (CNN), Swin Transformer sebagai representasi Vision Transformer, dan ConvNeXt sebagai representasi Modernized CNN dalam klasifikasi tujuh jenis lesi kulit menggunakan dataset HAM10000. Berdasarkan hasil eksperimen, ConvNeXt menunjukkan performa terbaik dengan akurasi sebesar **91,62%**, Precision sebesar **0,9026**, Recall sebesar **0,8852**, dan Macro F1-Score sebesar **0,8897**, sehingga memberikan kinerja yang lebih baik dibandingkan EfficientNet-B4 maupun Swin Transformer pada konfigurasi pelatihan yang sama. Temuan ini mengindikasikan bahwa pendekatan modernisasi arsitektur konvolusi mampu memberikan keseimbangan yang baik antara kemampuan ekstraksi fitur lokal dan representasi fitur global pada klasifikasi citra lesi kulit.

Selain menghasilkan performa klasifikasi yang lebih tinggi, ConvNeXt juga menunjukkan kemampuan yang lebih baik dalam mengenali kelas minoritas, khususnya Melanoma, yang tercermin dari nilai Recall yang lebih tinggi dibandingkan kedua model lainnya. Hasil analisis kurva pembelajaran menunjukkan proses konvergensi yang stabil dengan perbedaan yang relatif kecil antara data pelatihan dan validasi, sehingga mengindikasikan kemampuan generalisasi model yang baik pada dataset yang memiliki ketidakseimbangan kelas.

Evaluasi menggunakan Explainable AI melalui Grad-CAM dan Attention Maps menunjukkan bahwa ConvNeXt secara konsisten memfokuskan perhatian pada area lesi yang relevan dibandingkan EfficientNet-B4 dan Swin Transformer. Temuan tersebut mendukung interpretabilitas keputusan model dan menunjukkan bahwa peningkatan performa klasifikasi juga diikuti oleh visualisasi perhatian model yang lebih representatif terhadap area lesi. Dengan demikian, ConvNeXt berpotensi menjadi salah satu alternatif arsitektur deep learning yang andal untuk mendukung pengembangan sistem diagnosis lesi kulit berbasis kecerdasan buatan yang tidak hanya memiliki performa klasifikasi yang baik, tetapi juga mampu memberikan tingkat transparansi yang lebih tinggi dalam proses pengambilan keputusan.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan yang dapat menjadi peluang pengembangan pada penelitian selanjutnya. Penggunaan dataset yang lebih beragam dari berbagai institusi maupun penerapan teknik sintesis data, seperti **Generative Adversarial Networks (GAN)**, berpotensi meningkatkan kemampuan generalisasi model, terutama pada kelas minoritas. Selain itu, implementasi model ke dalam aplikasi berbasis web atau perangkat bergerak dapat memperluas pemanfaatannya sebagai sistem pendukung diagnosis awal. Pengembangan berikutnya juga dapat diarahkan pada optimasi hyperparameter yang lebih komprehensif serta eksplorasi metode Explainable AI lainnya, seperti **SHAP (SHapley Additive exPlanations)**, untuk memberikan interpretasi yang lebih mendalam terhadap proses pengambilan keputusan model.

Referensi

- [1] Tschandl, P., Rosendahl, C. & Kittler, H. *The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions*. *SCI Data* 5, 180161 (2018). <https://doi.org/10.1038/sdata.2018.161>
- [2] M. S. Islam, A. C. Chamgoué, and G. P. Dubb, "Benchmarking State-of-the-Art Vision Transformer Architectures for the Automated Classification of Pigmented Skin Lesions," *Computers and Electronics in Medicine*, Vol. 3, No. 1, pp. 42–47, Jan. 2026, DOI: 10.69882/adba.cem.2026015.
- [3] C. Supriyanto et al., "A Bibliometric Review of Deep Learning Approaches in Skin Cancer Research," Mar. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. DOI: <http://sistemasi.ftik.unisi.ac.id>

- 10.3390/computation13030078.
- [4] X. Yu and X. Lei, “Application of the Multi-Omics Liquid Biopsy Method M2P-HCC in Early Liver Cancer Screening for High-Risk Individuals with Hepatitis B-Related Liver Cancer,” *Diagnostics*, Vol. 13, No. 15, Aug. 2023, DOI: 10.3390/diagnostics13152484.
 - [5] S. Reddy, A. Shaheed, and R. Patel, “Artificial Intelligence in Dermoscopy: Enhancing Diagnosis to Distinguish Benign and Malignant Skin Lesions,” *Cureus*, Feb. 2024, DOI: 10.7759/cureus.54656.
 - [6] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, “A GAN-based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification,” *IEEE Access*, Vol. 12, pp. 16498–16513, 2024, DOI: 10.1109/ACCESS.2024.3360215.
 - [7] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, “Multiclass Skin Cancer Classification using EfficientNets – a First Step Towards Preventing Skin Cancer,” Dec. 01, 2022, Elsevier Masson s.r.l. DOI: 10.1016/j.neuri.2021.100034.
 - [8] I. A. Pardosi, R. Yunis, and A. Halim, “Skin Lesion Diagnosis through Deep Learning and Hybrid Texture Feature Augmentation,” *Teknika*, Vol. 14, No. 2, pp. 264–269, Jul. 2025, DOI: 10.34148/teknika.v14i2.1253.
 - [9] S. Aburass, O. Dorgham, J. al Shaqsi, M. Abu Rumman, and O. Al-Kadi, “Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases,” Dec. 01, 2025, Springer Nature. DOI: 10.1007/s10278-025-01481-y.
 - [10] A. Kumar and Jyoti, “A Comprehensive Review of Transformer Models in Brain Tumor Analysis,” *Expert Syst Appl*, Vol. 300, p. 130509, Mar. 2026, DOI: 10.1016/j.eswa.2025.130509.
 - [11] H. Fang, Y. Sun, S. Zhang, X. Teng, X. Li, and X. Yu, “HSPF-Net: Hybrid CNN-Transformer with Serial-Parallel Fusion for Skin Lesion Segmentation,” *Digit Signal Process*, Vol. 172, p. 105891, Mar. 2026, DOI: 10.1016/j.dsp.2026.105891.
 - [12] M. Waqar et al., “Explainable Clinical Diagnosis through Unexploited Yet Optimized Fine-Tuned ConvNeXt Models for Accurate Monkeypox Disease Classification,” *SLAS Technol*, Vol. 33, Aug. 2025, DOI: 10.1016/j.slast.2025.100336.
 - [13] Roy, A.L., Al Maruf, A., Sony, R.I. et al. *Fostering Trust and Interpretability in Skin Cancer Classification: a Hybrid Framework of Deep Learning and Machine Learning with Explainable AI*. *J Ambient Intell Human Comput* 16, 879–894 (2025). <https://doi.org/10.1007/s12652-025-04984-2>.
 - [14] M. M. Islam, N. Akter, M. Assaduzzaman, M. M. H. Shimul, and R. K. R. Sarker, “Explainable AI for Breast Cancer Detection: Biglycan Biomarker Classification with Transfer Learning,” *Intell Based Med*, Vol. 13, Mar. 2026, DOI: 10.1016/j.ibmed.2025.100340.
 - [15] I. Pacal, B. Ozdemir, J. Zeynalov, H. Gasimov, and N. Pacal, “A Novel CNN-ViT-based Deep Learning Model for Early Skin Cancer Diagnosis,” *Biomed Signal Process Control*, Vol. 104, Jun. 2025, DOI: 10.1016/j.bspc.2025.107627.
 - [16] M. Denny Richardo and H. Satria, “Comparative Analysis of Explainable AI Models for Pneumonia Detection in Chest X-rays using Grad-CAM,” *Sistem Informasi dan Komputer*, Vol. 14, No. 04, pp. 475–483, Okt. 2025, DOI: 10.32736/sisfokom.v4i4.2450.
 - [17] Y. Gulzar, B. I. Ya’u, M. Alkanan, and C. W. Onn, “ScNet: a Lightweight CNN with Depthwise and SE Modules for Skin Lesion Classification,” *Comput Methods Biomech Biomed Eng Imaging Vis*, Vol. 13, No. 1, 2025, DOI: 10.1080/21681163.2025.2576198.
 - [18] S. Nurmaini et al., “Deep Learning for Improving the Effectiveness of Routine Prenatal Screening for Major Congenital Heart Diseases,” *J Clin Med*, Vol. 11, No. 21, Nov. 2022, DOI: 10.3390/jcm11216454.
 - [19] G. H. Dagnaw, M. el Mouhtadi, and M. Mustapha, “Skin Cancer Classification using Vision Transformers and Explainable Artificial Intelligence,” *J Med Artif Intell*, Vol. 7, No. June, 2024, DOI: 10.21037/jmai-24-6.
 - [20] S. Cherukuri and S. D. Vara Prasad, “EnsembleSkinNet: a Transfer Learning-based Framework for Efficient Skin Cancer Detection with Explainable AI Integration,” *Front Oncol*, Vol. 15, 2025, DOI: 10.3389/fonc.2025.1699960.