

Klasifikasi Citra Lesi Kulit Menggunakan *Convolutional Neural Network*

Hans Aristo^{1,*}, Tri Ismardiko Widyawan²

^{1,2} Fakultas Ilmu Komputer, Program Studi Teknologi Informatika, Universitas Esa Unggul, Jakarta, Indonesia
Email: ^{1,*}hanzaristo@gmail.com, ²ismardiko@esaunggul.ac.id

^{*)} Email Penulis Utama

Abstrak—Kanker kulit merupakan salah satu penyakit dengan insidensi tinggi yang memerlukan deteksi dini untuk meningkatkan keberhasilan pengobatan. Penelitian ini bertujuan mengembangkan sistem klasifikasi otomatis lesi kulit menjadi jinak (*benign*) atau ganas (*malignant*) menggunakan arsitektur *Convolutional Neural Network* (CNN) pada citra dermatoskopi serta membandingkan performanya dengan *Artificial Neural Network* (ANN) dan EfficientNetB0 sebagai arsitektur CNN berbasis *transfer learning*. Evaluasi model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil eksperimen menunjukkan superioritas arsitektur berbasis CNN, dengan EfficientNetB0 sebagai model *pretrained* mencapai performa terbaik yaitu akurasi 0.98, presisi 0.97, *recall* 0.91, dan *F1-score* 0.94, dengan waktu pelatihan selama 22 menit 34 detik sebagai konsekuensi dari kompleksitas arsitektur yang lebih tinggi. Keunggulan EfficientNetB0 terletak pada penerapan *transfer learning* yang memungkinkan ekstraksi fitur visual kompleks secara optimal.

Kata Kunci: Klasifikasi Citra Lesi Kulit, Citra Medis, *Deep Learning*, CNN.

Abstract—*Skin cancer is among the most prevalent cancers and may lead to severe outcomes if not diagnosed early. Timely detection of skin lesions is vital for improving patient recovery prospects. This study focuses on designing an automated classification system using a Convolutional Neural Network (CNN) architecture to distinguish skin lesions as benign or malignant based on dermatoscopic image, and to compare its performance with an Artificial Neural Network (ANN) approach and EfficientNetB0 as a transfer learning-based CNN architecture. The dataset employed is DERM12345, comprising 12.345 high-resolution annotated dermatoscopic images. Model performance is assessed through accuracy, precision, recall, and F1-score metrics. Experimental findings demonstrate the superiority of CNN-based architectures, with EfficientNetB0 as the pretrained model achieving the best performance: 98% accuracy, 0.97 precision, 0.91 recall, and 0.94 F1-score, requiring 22 minutes and 34 seconds of training time as a consequence of its higher architectural complexity. The advantage of EfficientNetB0 lies in its transfer learning capability, enabling optimal extraction of complex visual features.*

Keywords: *Skin Lesion Classification, Medical Imaging, Deep Learning, CNN.*

1. PENDAHULUAN

Insidensi kanker kulit global menunjukkan tren peningkatan yang konsisten dalam beberapa waktu terakhir, dengan sekitar 6,64 juta kasus baru dilaporkan pada 2021 [1]. Kanker kulit sering kali diawali oleh lesi prakanker yang berpotensi berkembang menjadi kanker ganas [2]. Deteksi dini terhadap perubahan pada kulit dapat meningkatkan keberhasilan pengobatan serta mengurangi dampak negatif yang ditimbulkan oleh gejala kanker [3].

Citra digital medis adalah citra dari tubuh manusia yang dihasilkan melalui teknologi pencitraan [4]. Dalam praktik klinis, diagnosis lesi kulit bergantung pada interpretasi visual dokter, dermoskopi, dan biopsi. Namun, diagnosis manual memiliki tantangannya sendiri, sehingga diadopsikan *Computer-assisted Diagnosis* (CAD) untuk membantu meningkatkan akurasi deteksi kanker [5]. Kemajuan pesat dalam bidang *deep learning* telah mendorong perkembangan berbagai pendekatan analisis citra medis, khususnya dalam pengembangan sistem CAD untuk deteksi kanker [6].

Pendekatan *deep learning* telah menunjukkan tingkat akurasi diagnostik yang kompetitif dalam evaluasi penyakit kulit tertentu dan berpotensi mendukung proses pengambilan keputusan klinis [7]. Hasil tersebut menunjukkan bahwa sistem berbasis *deep learning* dapat berfungsi sebagai *decision-support system* yang membantu meningkatkan keandalan serta membantu dalam proses konfirmasi diagnosis pada evaluasi lesi kulit.

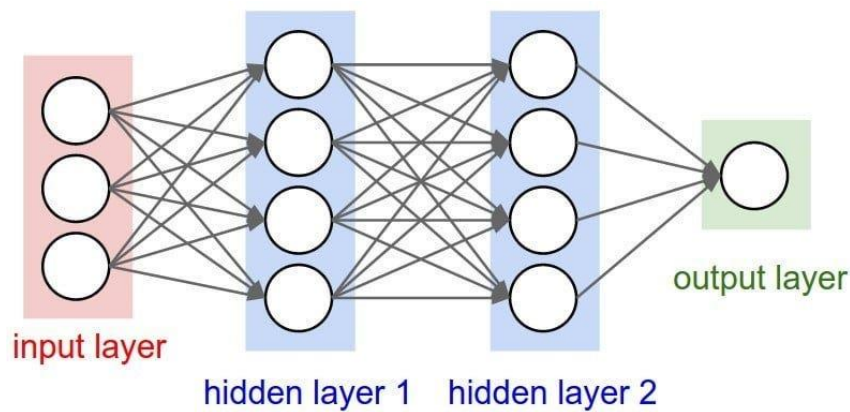
Salah satu arsitektur *deep learning* yang paling efektif untuk tugas klasifikasi gambar medis adalah *Convolutional Neural Network* (CNN), yang secara khusus dipakai untuk mengenali pola visual yang kompleks dan membedakan berbagai jenis lesi kulit secara otomatis [8]. Meskipun demikian, sebagian besar penelitian sebelumnya berfokus pada peningkatan performa model tunggal atau pengembangan arsitektur yang semakin kompleks, tanpa mengevaluasi secara sistematis bagaimana perbedaan tingkat kompleksitas arsitektur memengaruhi performa klasifikasi pada dataset modern berskala besar. Selain itu, perbandingan langsung antara *Artificial Neural Network* (ANN) berbasis *fully-connected* layer sebagai *baseline* non-spasial, *custom* CNN, dan model *pretrained* berbasis *transfer learning* masih terbatas dalam konteks dataset terbaru.

Penelitian ini bertujuan untuk mengisi celah tersebut dengan melakukan evaluasi komparatif terhadap tiga pendekatan berbeda: ANN sebagai baseline tradisional, custom CNN sebagai model spasial yang dirancang khusus, dan CNN berbasis pretrained melalui *transfer learning*. Dataset yang digunakan adalah DERM12345, dataset yang dirilis pada tahun 2024 dan mencakup lebih dari 12.000 citra dermatoskopi beranotasi lengkap beserta metadata pasien. Penggunaan dataset terbaru ini memungkinkan evaluasi performa model pada distribusi data dengan variasi kelas dan karakteristik visual yang lebih beragam, sehingga memberikan tantangan klasifikasi yang lebih realistis dibandingkan beberapa dataset terdahulu.

2. METODE PENELITIAN

2.1 Artificial Neural Network (ANN)

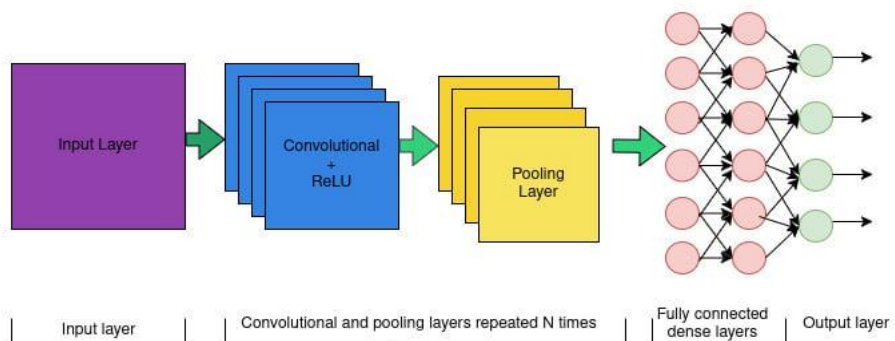
Secara hierarkis, ANN dibangun dari tiga komponen utama: lapisan input, serangkaian lapisan tersembunyi, dan lapisan output, yang masing-masing memuat neuron-neuron yang saling berinteraksi melalui bobot dan bias [9]. Pembelajaran terjadi melalui siklus pelatihan menggunakan algoritma *backpropagation* serta teknik optimasi untuk memperkecil nilai *loss function* [10].



Gambar 1. Model Artificial Neural Network

2.2 Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan turunan khusus dari *Artificial Neural Network* yang dioptimalkan untuk tugas pengenalan pola visual. Struktur CNN terdiri dari serangkaian lapisan hierarkis, meliputi lapisan konvolusi, lapisan *pooling*, dan lapisan terhubung penuh. Lapisan konvolusi berperan dalam mengidentifikasi dan mengekstraksi pola lokal pada citra, sementara lapisan *pooling* berfungsi untuk menurunkan resolusi spasial guna mengoptimalkan beban komputasi yang diperlukan. Setelah melalui beberapa iterasi ekstraksi dan reduksi, representasi fitur kemudian diratakan dan diumpungkan ke lapisan penentu keputusan untuk klasifikasi berdasarkan pola yang telah dipelajari [11].



Gambar 2. Model Convolutional Neural Network

2.3 Transfer Learning

Pendekatan pembelajaran transfer (*transfer learning*) merupakan teknik yang mengadopsi pengetahuan dari model yang telah dilatih sebelumnya menggunakan dataset berskala besar, untuk kemudian diterapkan dalam menyelesaikan tugas klasifikasi. Teknik ini sangat efektif terutama ketika ketersediaan data pelatihan terbatas, seperti dalam klasifikasi gambar medis, karena mampu meningkatkan akurasi [12].

Model populer seperti EfficientNetB0, ResNet, VGG, dan MobileNetV2 sering digunakan sebagai *feature extractor* atau *fine-tuning* pada lapisan akhir agar sesuai dengan target klasifikasi baru. Dalam studi ini, teknik pembelajaran transfer diterapkan dengan mengadaptasi model yang telah dilatih sebelumnya (*pretrained model*) bernama EfficientNetB0 untuk meningkatkan performa dalam proses klasifikasi lesi kulit.

2.4 Spesifikasi Perangkat Penelitian

Dalam penelitian ini, proses pengembangan dan pengujian dilakukan dengan menggunakan spesifikasi perangkat yang telah ditentukan:

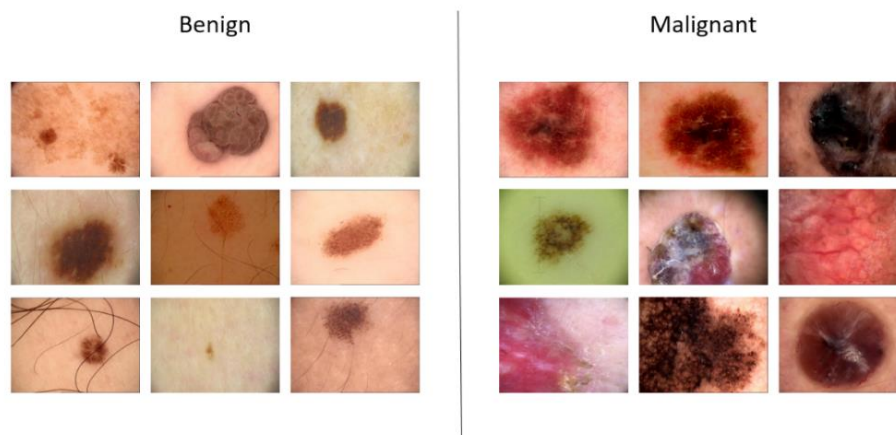
Tabel 1. Spesifikasi Perangkat Keras dan Perangkat Lunak

Perangkat	Spesifikasi
Processor	Intel Core i5-9400F
RAM	16GB DDR4
GPU	A100GPU
Penyimpanan	SSD 512GB
Sistem Operasi	Windows 11 Pro 64-bit
Environment	Conda dan Google Colab

2.5 Objek Penelitian

Data yang digunakan dalam penelitian ini merupakan berbagai citra dermatoskopi lesi kulit dari dataset DERM12345 yang dipublikasikan oleh Abdurrahim Yilmaz. Dataset ini memiliki struktur yang beragam, terdiri dari 5 super kelas, 15 kelas utama, 38 subkelas, dan mencakup 12.345 gambar dermatoskopi beresolusi tinggi [13].

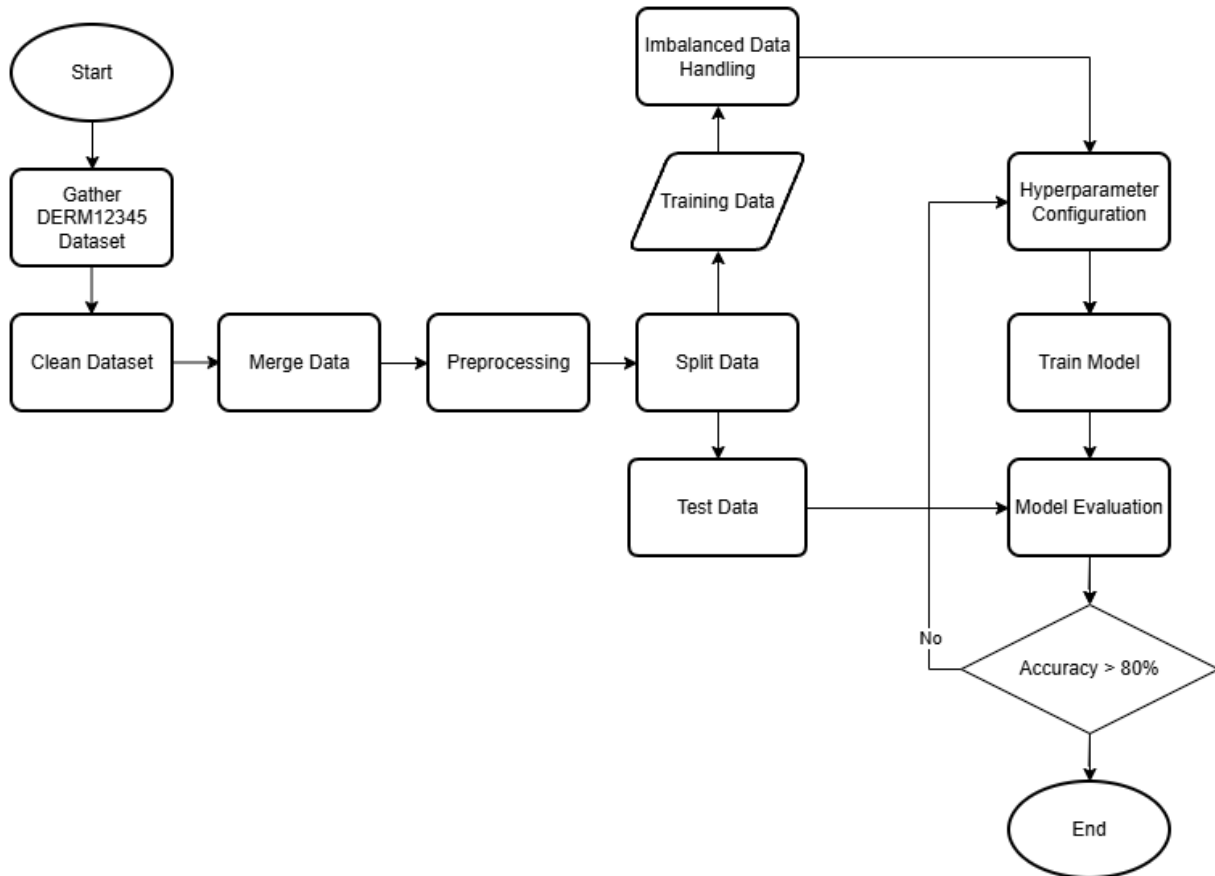
Setiap citra dalam dataset memiliki variasi warna, tekstur, pola pigmentasi, dan ukuran lesi yang berbeda-beda, sehingga menuntut model untuk mampu melakukan ekstraksi fitur yang robust terhadap noise dan variasi pencahayaan. Variasi ini menjadi tantangan utama dalam proses klasifikasi dan menjadikan dataset DERM12345 relevan untuk pengembangan model *deep learning* di bidang dermatologi.



Gambar 3. Lesi Kulit *Malignant* dan *Benign* pada Dataset DERM12345

2.6 Tahapan Penelitian

Flowchart pada gambar berikut memberikan gambaran menyeluruh mengenai aliran kerja penelitian, mulai dari tahap awal pengumpulan data hingga evaluasi akhir model. Dengan adanya representasi alur ini, setiap proses dapat dipantau dan dievaluasi untuk memastikan penelitian berjalan sesuai rencana:



Gambar 4. Flowchart Penelitian

2.7 Preprocessing Data

Pada tahap ini, pada dataset DERM12345 terlebih dahulu dilakukan *data cleaning* dengan menghapus kelas *indeterminate* dihapus dari *field* diagnosa karena tidak relevan pada pengujian model dan menimbulkan *noise*. Kemudian, dataset dipisahkan menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji menggunakan metode *train test split* dengan teknik *stratified sampling* agar proporsi kelas *benign* dan *malignant* tetap seimbang. Selain itu, untuk mengatasi ketidakseimbangan data pada kelas minoritas, dilakukan *oversampling* pada data latih sehingga model mendapatkan distribusi kelas yang lebih merata dan dapat belajar secara optimal.

Seluruh citra dalam dataset kemudian diubah ukurannya menjadi 224x224 pixel dengan 3 channel warna (RGB) sehingga menghasilkan input shape sebesar (150.528,) pada ANN dan (224, 224, 3) yang kompatibel dengan arsitektur CNN. Nilai pixel pada citra dinormalisasi menggunakan teknik *rescaling* ke dalam interval [0, 1] dengan membagi nilai pixel dengan 255. Teknik ini digunakan untuk mempercepat proses pelatihan dan memastikan setiap fitur memiliki skala yang seragam, sehingga mencegah dominasi fitur tertentu akibat perbedaan skala nilai pixel.

2.8 Design Eksperimen

Dalam penelitian ini, dilakukan eksperimen terhadap tiga arsitektur model yaitu *Artificial Neural Network* (ANN), *Convolutional Neural Network* (CNN), dan *pretrained model* EfficientNetB0 untuk melakukan klasifikasi lesi kulit. Ketiga model tersebut akan diuji secara adil, yaitu dengan menggunakan pengaturan hyperparameter yang terbaik untuk masing-masing arsitektur serta skenario pelatihan yang identik, guna memastikan bahwa perbedaan performa yang dihasilkan murni berasal dari arsitektur model, bukan dari perbedaan perlakuan selama pelatihan.

2.8.1 Hyperparameter

Sebagai konfigurasi umum, beberapa hyperparameter diterapkan secara seragam pada seluruh model guna menjamin efisiensi pelatihan dan mengendalikan *overfitting*. *Optimizer* Adam dipilih karena kemampuannya

dalam menyesuaikan *learning rate* secara adaptif, sehingga mempercepat pencapaian konvergensi. Nilai *dropout rate* 0,5 digunakan untuk meningkatkan regularisasi dengan cara mengacak sebagian neuron selama fase pelatihan. Proses pelatihan dibatasi hingga 20 epoch maksimal, dengan penerapan *early stopping* yang memiliki patience sebesar 5—mekanisme ini akan menghentikan pelatihan secara otomatis apabila tidak terjadi peningkatan performa pada data validasi selama lima epoch berturut-turut, sebagai upaya pencegahan *overfitting*.

Tabel 2. Fixed Hyperparameter

No	Hyperparameter	Output
1	Optimizer	Adam
2	Dropout Rate	0.5
3	Early Stopping	Patience = 5
4	Epoch	20

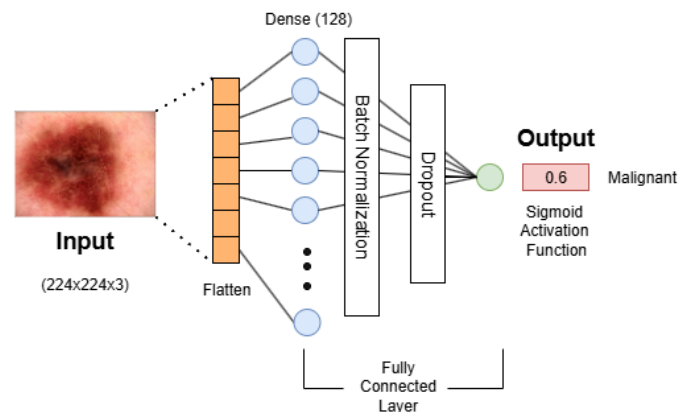
Beberapa hyperparameter berikut disesuaikan melalui proses optimisasi menggunakan metode grid search untuk mencari kombinasi terbaik yang menghasilkan performa model tertinggi.

Tabel 3. Tuning Hyperparameter

No	Hyperparameter	Output
1	Batch Size	[16, 32, 64]
2	Learning Rate	[0.001, 0.0001, 0.00001]

2.8.2 Komponen Arsitektur Artificial Neural Network

Penelitian ini mengimplementasikan struktur *Artificial Neural Network* yang dibangun untuk menangani tugas klasifikasi, diwujudkan melalui serangkaian lapisan neuron yang saling berinterkoneksi. Setiap lapisan memiliki fungsi dan peran tersendiri dalam mentransformasi data dari bentuk awal hingga menghasilkan keluaran akhir yang merepresentasikan hasil prediksi. Struktur jaringan ini terdiri dari beberapa lapisan utama, mulai dari lapisan input hingga lapisan output, dengan aktivasi dan teknik regularisasi yang disesuaikan untuk meningkatkan performa model. Adapun visualisasi dari arsitektur ANN yang digunakan dalam penelitian ini:



Gambar 5. Arsitektur Model *Artificial Neural Network*

Berikut adalah komponen-komponen dari arsitektur *Artificial Neural Network* yang digunakan pada penelitian ini:

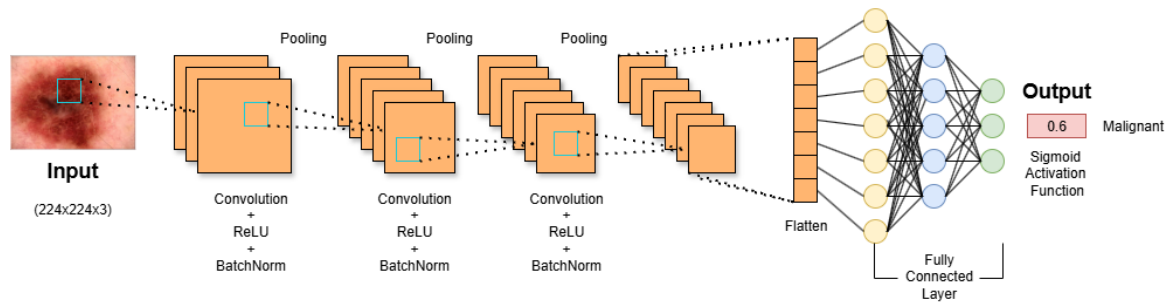
Tabel 4. Arsitektur Model *Artificial Neural Network*

Layer	Activation	Output
Flatten	-	150.528
Dense1	ReLU	128
BatchNorm1	-	128
Dropout1 (0.5)	-	128
Dense2	-	128

2.8.3 Komponen Arsitektur Custom Convolutional Neural Network

Custom CNN merujuk pada arsitektur *Convolutional Neural Network* yang dirancang secara khusus dengan beberapa lapisan konvolusi, *batch normalization*, dan *pooling*, serta dilatih dari awal tanpa menggunakan bobot *pretrained*. *Convolutional Neural Network* yang digunakan dalam penelitian ini dirancang untuk mengekstraksi fitur visual secara bertahap melalui proses konvolusi, normalisasi, dan penyederhanaan dimensi, sebelum akhirnya melakukan klasifikasi pada lapisan *fully-connected*.

Setiap lapisan memiliki fungsi spesifik dalam mengolah citra mulai dari bentuk mentah hingga menjadi representasi fitur yang lebih abstrak dan informatif. Proses ini dimulai dari lapisan konvolusi yang mendeteksi pola dasar seperti tepi dan tekstur, dilanjutkan dengan *pooling* untuk mereduksi dimensi dan mencegah *overfitting*, serta normalisasi untuk menstabilkan proses pelatihan. Setelah fitur-fitur utama berhasil diekstraksi, data kemudian diratakan dan diproses melalui beberapa lapisan *dense* untuk menghasilkan keluaran akhir berupa nilai probabilitas prediksi. Adapun visualisasi dari arsitektur CNN yang digunakan pada gambar berikut:



Gambar 6. Arsitektur Model Artificial Neural Network

Berikut adalah komponen-komponen dari arsitektur *Custom Convolutional Neural Network* yang digunakan pada penelitian ini:

Tabel 5. Arsitektur Model Custom Convolutional Neural Network

Layer	Kernel Size	Stride	Activation	Output
Conv1	(3, 3)	(1, 1)	ReLU	(224, 224, 3)
BatchNorm1	-	-	-	(224, 224, 32)
MaxPool1	(2, 2)	(2, 2)	-	(112, 112, 32)
Conv2	(3, 3)	(1, 1)	ReLU	(112, 112, 64)
BatchNorm2	-	-	-	(112, 112, 64)
MaxPool2	(2, 2)	(2, 2)	-	(56, 56, 64)
Conv3	(3, 3)	(1, 1)	-	(56, 56, 128)
BatchNorm3	-	-	-	(56, 56, 128)
MaxPool3	(2, 2)	(2, 2)	-	(28, 28, 128)
Flatten	-	-	-	(100352,)
Dropout1 (0.5)	-	-	-	(100352,)
Dense1	-	-	ReLU	(128,)
Dense2	-	-	Sigmoid	(1,)

2.8.4 Komponen Arsitektur Pretrained Convolutional Neural Network

Arsitektur EfficientNetB0 yang digunakan dalam penelitian ini memanfaatkan model pralatih berbasis *compound scaling*, yaitu pendekatan yang secara seimbang menskalakan kedalaman, lebar, dan resolusi jaringan untuk memperoleh performa optimal dengan kompleksitas yang efisien. Model ini berperan sebagai *feature extractor* yang mampu menangkap pola visual secara lebih kaya melalui blok konvolusi yang telah dioptimalkan. Setelah melalui proses ekstraksi fitur oleh EfficientNetB0, keluaran jaringan kemudian disederhanakan melalui operasi *Global Average Pooling* sehingga menghasilkan representasi vektor yang kompak namun informatif. Tahap selanjutnya mencakup penerapan *dropout* sebagai mekanisme regularisasi untuk mengurangi risiko *overfitting*, sebelum data diteruskan menuju lapisan *dense* guna melakukan pemetaan akhir ke kelas target. Output model kemudian dinyatakan dalam bentuk probabilitas melalui fungsi aktivasi sigmoid.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Optimasi Hyperparameter

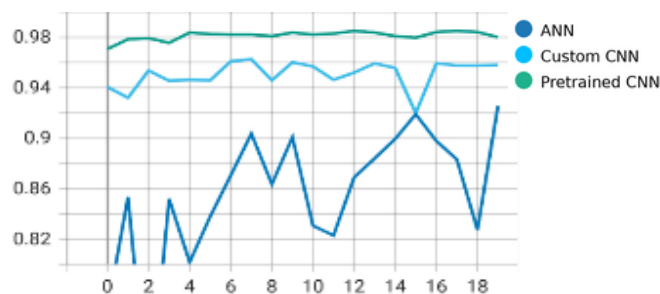
Ketepatan pemilihan hyperparameter merupakan faktor kritis, karena secara langsung berdampak besar pada kinerja akhir model. Konfigurasi yang tidak optimal berisiko menyebabkan model tidak mampu menangkap pola data (*underfitting*) atau menghafal *noise* pada data latih (*overfitting*).

Untuk mencapai tujuan ini, penelitian mengimplementasikan metode *grid search*. Teknik ini melakukan eksplorasi sistematis dengan menguji seluruh kemungkinan kombinasi nilai yang telah didefinisikan sebelumnya pada ruang pencarian. Adapun hyperparameter yang disetel dalam penelitian ini meliputi *learning rate* (dengan nilai 0.001, 0.0001, dan 0.00001) dan *batch size* (yakni 16, 32, dan 64). Berikut adalah kombinasi hyperparameter terbaik dari setiap model:

3.2 Hasil Optimasi Hyperparameter

Fase dalam alur penelitian ini adalah menguji tiap model-model yang telah dilatih dengan kombinasi hyperparameter yang terbaik untuk mengevaluasi kinerja dan ketepatannya dalam tugas klasifikasi. Guna mempermudah analisis, proses evaluasi dibantu dengan TensorBoard sebagai platform visualisasi. Dengan alat ini, hasil-hasil dari proses pelatihan dan validasi dapat divisualisasikan dalam bentuk grafik dan kurva kinerja, memungkinkan pengamatan serta perbandingan yang lebih intuitif dan mendalam.

Hasil visualisasi akurasi pada data pengujian terhadap jumlah epoch untuk masing-masing model ditampilkan pada Gambar 7. Grafik ini menunjukkan perkembangan performa model selama proses pelatihan, sehingga pola peningkatan, stabilitas, maupun potensi *overfitting* dapat diamati secara komparatif. Melalui visualisasi ini, perbedaan karakteristik pembelajaran antara ANN, custom CNN, dan model berbasis *transfer learning* dapat dianalisis secara lebih jelas berdasarkan tren akurasi yang dihasilkan pada setiap epoch:



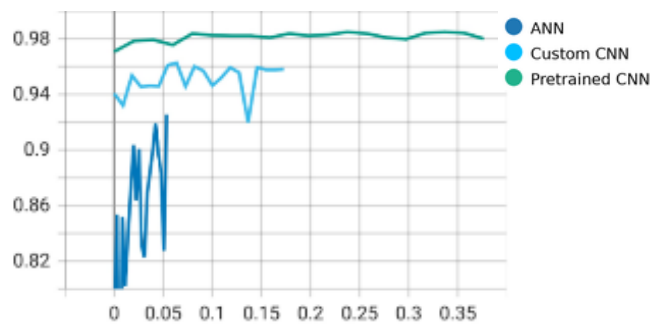
Gambar 7. Visualisasi Akurasi pada Test Dataset Terhadap Jumlah Epoch

Berdasarkan visualisasi akurasi validasi terhadap jumlah epoch, terlihat adanya perbedaan performa yang cukup signifikan antara ketiga model yang diuji. Model ANN menunjukkan akurasi yang relatif rendah dan fluktuatif sepanjang proses pelatihan. Pola ini menandakan bahwa ANN kesulitan dalam mencapai kestabilan pembelajaran, sehingga akurasi yang dihasilkan kurang konsisten dan berada di bawah dua model lainnya.

Berbeda dengan ANN, model custom CNN memperlihatkan kinerja yang lebih baik dengan akurasi yang lebih stabil pada grafik dan lebih tinggi di atas 0,94 pada test dataset. Walaupun masih terdapat variasi naik-turun di beberapa epoch, performa CNN secara keseluruhan menunjukkan kemampuan generalisasi yang lebih baik dibandingkan ANN, sehingga dapat dianggap lebih handal dalam mengklasifikasikan citra lesi kulit.

Sementara itu, model EfficientNetB0 berhasil mencapai performa tertinggi. Sejak epoch awal hingga akhir, akurasi validasi model ini konsisten mendekati 0,98 dan tetap stabil tanpa adanya fluktuasi yang signifikan. EfficientNetB0 memiliki kapabilitas untuk menangkap representasi data secara efektif, menjaga stabilitas akurasi pada level yang optimal selama seluruh iterasi pelatihan.

Selain metrik akurasi, informasi terkait waktu pelatihan juga diperoleh melalui TensorBoard dalam bentuk *relative time* pada Gambar 8, yaitu durasi yang dihitung sejak proses training dimulai hingga mencapai epoch terakhir. Informasi ini digunakan untuk menganalisis efisiensi komputasi masing-masing model, sehingga tidak hanya aspek ketepatan klasifikasi yang dibandingkan, tetapi juga performa dari sisi waktu pemrosesan.



Gambar 8. Visualisasi Waktu yang Diperlukan Untuk Melatih Model

Model ANN tercatat sebagai model dengan waktu pelatihan paling singkat, yaitu sekitar 3 menit untuk menyelesaikan 20 epoch. Tingkat kompleksitas arsitektur menjadi penyebab utama dari efisiensi waktu ini, karena secara komputasi proses *forward propagation* dan *backward propagation* dapat dieksekusi dengan lebih ringan.

CNN menunjukkan waktu pelatihan yang lebih tinggi, yakni sekitar 10 menit untuk 20 epoch. Hal ini disebabkan oleh adanya operasi konvolusi pada beberapa lapisan, yang secara signifikan meningkatkan jumlah perhitungan yang harus dilakukan. Meskipun lebih lambat daripada ANN, CNN menawarkan kemampuan ekstraksi fitur visual yang jauh lebih baik, sehingga membutuhkan proses komputasi yang lebih berat namun memberikan performa klasifikasi yang lebih kuat.

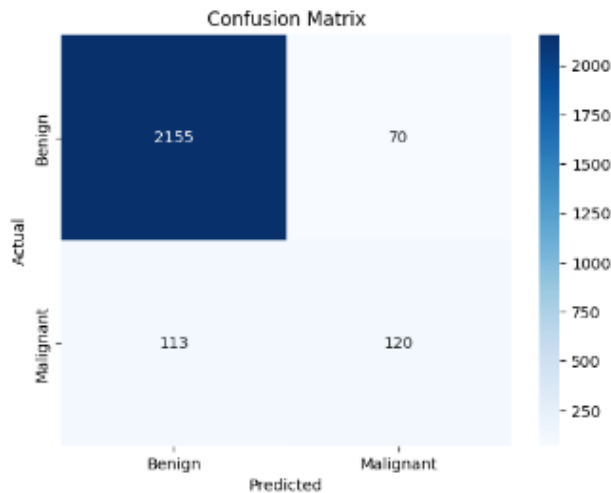
EfficientNetB0 memerlukan waktu pelatihan paling lama, dengan durasi sekitar 22 menit untuk menyelesaikan jumlah epoch yang sama. Lamanya waktu ini dapat dijelaskan melalui kompleksitas arsitekturnya yang lebih dalam dan lebih lebar, serta penggunaan teknik *compound scaling* yang sengaja dirancang untuk meningkatkan akurasi melalui peningkatan parameter dan operasi komputasi. Meskipun menuntut kapasitas komputasi yang lebih tinggi, EfficientNetB0 secara umum menghasilkan representasi ciri yang lebih kuat serta stabil.

Perbedaan durasi ini terutama disebabkan oleh kompleksitas arsitektur masing-masing model. ANN dengan struktur yang sederhana dapat menyelesaikan proses pelatihan lebih cepat, meskipun dengan akurasi yang lebih rendah. CNN yang memiliki lapisan konvolusi lebih banyak memerlukan waktu tambahan, namun mampu menghasilkan performa yang lebih baik. EfficientNetB0, sebagai model dengan arsitektur paling kompleks, membutuhkan durasi pelatihan terpanjang, sejalan dengan jumlah parameter yang besar serta proses komputasi yang lebih intensif.

3.3 Hasil Confusion Matrix

3.3.1 Confusion Matrix Artificial Neural Network

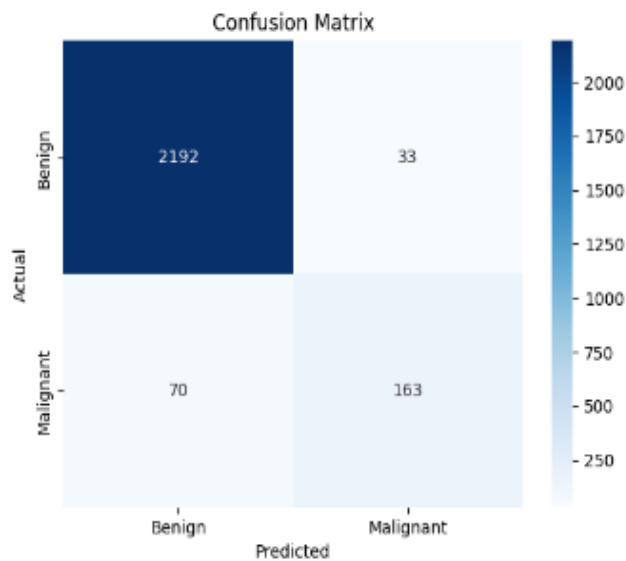
Model ANN menghasilkan *confusion matrix*, menunjukkan bahwa distribusi prediksi terhadap kelas sebenarnya belum konsisten. Beberapa kelas dapat diklasifikasikan dengan cukup baik, namun pada kelas minoritas terjadi peningkatan kesalahan prediksi yang signifikan.



Gambar 9. Confusion Matrix Artificial Neural Network

3.3.2 Confusion Matrix Custom Convolutional Neural Network

Model Custom CNN menunjukkan peningkatan performa signifikan dibandingkan model ANN, terutama dalam hal stabilitas prediksi dan kemampuan mengenali pola visual yang lebih kompleks. Berdasarkan hasil *confusion matrix*, distribusi prediksi pada kelas *benign* dan *malignant* menjadi lebih konsisten, menandakan bahwa model ini mampu meminimalkan kesalahan klasifikasi secara lebih efektif.



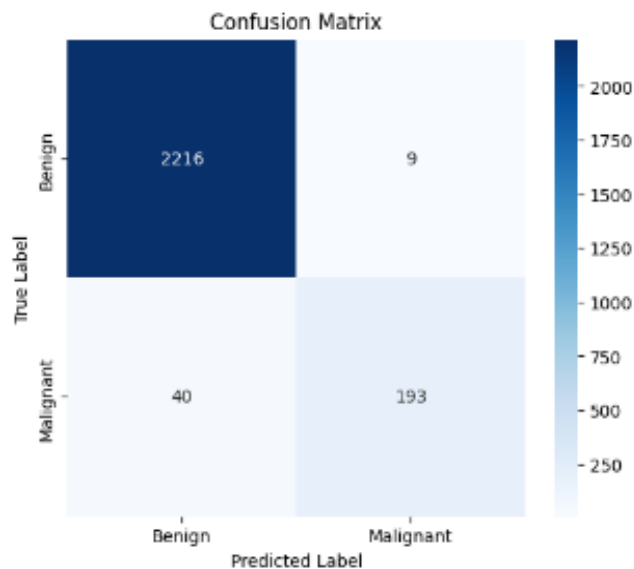
Gambar 10. Confusion Matrix Custom Convolutional Neural Network

Confusion matrix pada model ini memperlihatkan peningkatan jumlah prediksi benar di seluruh kelas, menandakan kemampuan generalisasi yang lebih baik. Hal ini membuktikan bahwa arsitektur CNN yang dirancang khusus mampu menangani kompleksitas pola pada dataset citra dengan lebih baik dibandingkan jaringan syaraf tiruan tradisional.

3.3.3 Confusion Matrix Pretrained Convolutional Neural Network (EfficientNetB0)

Dalam pengujian, EfficientNetB0 mencapai tingkat akurasi tertinggi dibandingkan dengan dua model lainnya, seperti yang terlihat dari hasil *confusion matrix*. Temuan ini menunjukkan bahwa penerapan transfer learning dari EfficientNetB0 berhasil memanfaatkan pengetahuan dari pelatihan pada dataset berskala besar untuk mengidentifikasi pola-pola kompleks dalam dataset penelitian ini.

Confusion matrix mengungkap bahwa sebagian besar sampel berhasil diklasifikasikan secara tepat, dengan jumlah kesalahan prediksi yang minimal. Hasil ini membuktikan keunggulan EfficientNetB0 dalam kemampuan generalisasi serta efisiensi ekstraksi fitur visual, yang didukung oleh desain arsitekturnya yang mengoptimalkan keseimbangan antara kedalaman, lebar, dan resolusi jaringan.



Gambar 11. *Confusion Matrix Pretrained Convolutional Neural Network (EfficientNetB0)*

Secara umum, temuan ini mengonfirmasi bahwa penggunaan arsitektur *pretrained* CNN seperti EfficientNetB0 berhasil meningkatkan kemampuan ekstraksi fitur secara substansial, sehingga menghasilkan kinerja klasifikasi yang lebih stabil dan akurat untuk kedua kelas.

Dengan demikian, pendekatan berbasis *pretrained model* EfficientNetB0 dapat dianggap sebagai strategi yang paling efektif untuk meningkatkan performa klasifikasi citra pada penelitian ini.

3.4 Analisis Komparatif

Analisis komparatif dilakukan untuk menilai perbedaan performa antara tiga model yang digunakan, yaitu ANN, CNN, dan EfficientNetB0. Ketiganya memiliki karakteristik arsitektur dan kompleksitas yang berbeda, sehingga menghasilkan variasi performa yang signifikan.

Dari sisi arsitektur ANN, model ini tidak memiliki kemampuan ekstraksi fitur spasial dan hanya mengandalkan nilai pixel yang telah diratakan. Akibatnya, ANN kurang efektif dalam mengenali pola tekstural dan struktural pada citra lesi kulit. Hal ini tercermin dari nilai *accuracy* sebesar 0.93 dan *F1-score* sebesar 0.76, dengan performa yang masih lemah pada kelas *malignant*.

Berbeda dengan itu, CNN memiliki tiga blok konvolusi dengan kernel 3×3 yang diikuti *batch normalization* dan *max pooling*. Struktur berlapis ini memungkinkan model mengekstraksi fitur spasial secara bertahap, mulai dari tepi hingga pola tekstur halus. Evaluasi menunjukkan peningkatan performa yang signifikan dibandingkan ANN, dengan *accuracy* mencapai 0.96 dan *F1-score* sebesar 0.87. CNN mampu melakukan klasifikasi dengan lebih stabil, meskipun masih terdapat kesalahan pada kelas *malignant* akibat kemiripan visual antarlesi.

EfficientNetB0 menampilkan hasil terbaik di antara ketiganya. Sebagai model *pretrained*, bobot awalnya sudah mengenali fitur umum seperti bentuk dan tekstur dari dataset besar. Setelah diadaptasi ke dataset ini, model mampu menyesuaikan fitur spesifik pada lesi kulit. Hasilnya, *accuracy* mencapai 0.98, *precision* 0.97, dan *F1-score* 0.94, menjadikannya model paling unggul baik dari sisi akurasi maupun stabilitas pembelajaran. Selain melihat peningkatan performa, analisis komparatif juga perlu mempertimbangkan *trade-off* antara akurasi dan waktu pelatihan. ANN merupakan model dengan waktu latih tercepat, yaitu 3 menit 12 detik, namun memiliki performa paling rendah. Di sisi lain, EfficientNetB0 membutuhkan waktu pelatihan paling panjang, mencapai 22 menit 34 detik, tetapi menghasilkan akurasi dan *F1-score* tertinggi. Dengan demikian, pemilihan model dapat disesuaikan dengan kebutuhan sistem.

Perbedaan performa antar model tidak hanya dipengaruhi oleh kompleksitas arsitektur, tetapi juga oleh karakteristik data citra yang digunakan. Lesi kulit memiliki variasi warna, tekstur, ukuran, dan batas tepi yang sangat beragam, sehingga model yang mampu mempertahankan informasi spasial cenderung menghasilkan performa lebih baik. ANN yang menggunakan representasi pixel yang diratakan menjadi 1 dimensi, sehingga kehilangan hubungan spasial antar pixel dan menyebabkan informasi struktur visual menjadi terbatas. Sebaliknya, CNN mampu mengekstraksi fitur lokal melalui operasi konvolusi sehingga lebih sensitif terhadap pola tekstur dan bentuk lesi.

Selain itu, performa EfficientNetB0 yang lebih tinggi dipengaruhi oleh proses *transfer learning*, di mana model telah mempelajari representasi visual umum dari dataset berskala besar sebelum proses *fine-tuning*. Inisialisasi bobot yang lebih informatif membantu model mencapai konvergensi lebih stabil serta mengurangi kebutuhan pembelajaran fitur dasar dari awal. Faktor lain yang turut memengaruhi performa adalah distribusi kelas pada dataset, di mana variasi visual antar lesi *benign* dan *malignant* yang saling menyerupai dapat meningkatkan tingkat kesalahan prediksi, khususnya pada model dengan kapasitas representasi yang lebih rendah.

. **Tabel 6.** Hasil Komparatif Performa dan Waktu Pelatihan Model

Model	Accuracy	Precision	Recall	F1-score	Waktu Latih
ANN	0.93	0.79	0.74	0.76	3m 12s
Custom CNN	0.96	0.90	0.84	0.87	10m 22s
Pretrained CNN	0.98	0.97	0.91	0.94	22m 34s

Jika dibandingkan dengan hasil yang dilaporkan oleh Furkan Yuceyalcin, Abdurrahim Yilmaz, dan Burak Temelkuran pada skenario klasifikasi biner, beberapa model *pretrained* berbasis *transfer learning* yang dioptimalkan untuk domain medis menunjukkan performa *F1-score* yang tinggi, seperti MedImageInsights (97,52%), MedSigLIP (96,43%), dan Derm Foundation (96,04%) [14]. Meskipun metrik yang dilaporkan pada studi tersebut berada pada kisaran yang sangat kompetitif, capaian *F1-score* EfficientNetB0 sebesar 0,94 dalam penelitian ini tetap menunjukkan performa yang kuat pada tugas klasifikasi keganasan biner.

4. KESIMPULAN

Penelitian ini melakukan evaluasi komparatif terhadap tiga arsitektur jaringan saraf, yaitu *Artificial Neural Network* (ANN), *custom Convolutional Neural Network* (CNN), dan model pretrained EfficientNetB0, dalam tugas klasifikasi citra lesi kulit menjadi kategori *benign* dan *malignant*. Hasil penelitian menunjukkan bahwa arsitektur yang mampu mempertahankan informasi spasial citra memberikan performa klasifikasi yang lebih baik dibandingkan pendekatan berbasis *fully-connected layer*.

Model *custom CNN* menunjukkan peningkatan kinerja yang konsisten dibandingkan ANN, sementara EfficientNetB0 sebagai model berbasis *transfer learning* menghasilkan performa tertinggi secara keseluruhan. Temuan ini mengindikasikan bahwa kompleksitas arsitektur dan pemanfaatan representasi fitur pralatih berkontribusi signifikan terhadap peningkatan kemampuan generalisasi model dalam analisis citra dermatoskopi.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Evaluasi dilakukan pada satu dataset dengan skenario klasifikasi biner, sehingga generalisasi hasil terhadap dataset lain atau skenario multi-kelas masih perlu dikaji lebih lanjut. Selain itu, penelitian ini belum melakukan uji signifikansi statistik untuk memastikan perbedaan performa antar model secara inferensial. Variasi distribusi kelas dan kemiripan visual antar lesi juga berpotensi memengaruhi hasil prediksi.

Penelitian selanjutnya dapat diarahkan pada evaluasi lintas dataset untuk menguji robustnes model, penerapan pendekatan *explainable artificial intelligence* (XAI) guna meningkatkan interpretabilitas klinis, serta integrasi data klinis tambahan seperti metadata pasien untuk membangun sistem CAD yang lebih komprehensif dan aplikatif dalam praktik dermatologi.

UCAPAN TERIMAKASIH

Penulis menyampaikan terima kasih kepada Universitas Esa Unggul, khususnya Program Studi Teknologi Informatika, atas dukungan dan fasilitas selama penelitian ini. Penghargaan juga ditujukan kepada keluarga saya, dosen, serta teman-teman seangkatan atas doa, semangat maupun dukungan yang tak ternilai sepanjang proses penulisan karya ilmiah ini.

DAFTAR PUSTAKA

- [1] L. Zhou, Y. Zhong, L. Han, Y. Xie, and M. Wan, "Global, regional, and national trends in the burden of melanoma and non-melanoma skin cancer: insights from the global burden of disease study 1990–2021," *Sci. Rep.*, vol. 15, no. 1, pp. 1–14, 2025, doi: 10.1038/s41598-025-90485-3.
- [2] N. T. Meibodi, Y. Nahidi, Z. Javidi, A. R. Taheri, M. Afzalaghace, and S. Jahanfakhr, "A clinicopathologic study of precancerous skin lesions," *Iran. J. Dermatology*, vol. 15, no. 61, pp. 89–94, 2012.
- [3] P. Karp, K. Karp, M. Kądziela, R. Zajdel, and A. Żebrowska, "The Importance of Early Detection and Prevention of Atypical Skin Lesions and Other Melanoma Risk Factors in a Younger Population," *Cancers (Basel)*, vol. 16, no. 24, pp. 1–12, 2024, doi: 10.3390/cancers16244264.
- [4] B. Sugandi, "Teknologi Citra untuk Peningkatan Kualitas Hidup yang Lebih Baik," *J. Integr.*, vol. 10, no. 1, pp. 21–27, 2018.
- [5] J. J. Fenton *et al.*, "Effectiveness of computer-aided detection in community mammography practice," *J. Natl. Cancer Inst.*, vol. 103, no. 15, pp. 1152–1161, 2011, doi: 10.1093/jnci/djr206.
- [6] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [7] N. Nadour, T. Duguet, S. Zahedi, H. Figoni, and R. Liard, "Diagnostic accuracy of artificial intelligence compared to family physicians and dermatologists for skin conditions: a systematic review and meta-analysis," *BMC Prim. Care*, vol. 26, no. 1, 2025, doi: 10.1186/s12875-025-03073-9.
- [8] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.
- [9] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer (Long Beach, Calif.)*, vol. 29, no. 3, pp. 31–44, 1996, doi: 10.1109/2.485891.
- [10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.
- [11] M. Ardifa Rizqi Ramadhan, Titan Apriyana Nadine Ananta, Afigo Azus Zakkyfriza, In'am Vaviansyah H, and Yahya Nour Fauzan, "Perbandingan Jumlah Layer Pada Convolutional Neural Network Untuk Meningkatkan Akurasi Dalam Klasifikasi Gambar," *Merkurius J. Ris. Sist. Inf. dan Tek. Inform.*, vol. 2, no. 5, pp. 211–217, 2024, doi: 10.61132/mercurius.v2i5.301.
- [12] A. Farahani, B. Pourshojae, K. Rasheed, and H. R. Arabnia, "A Concise Review of Transfer Learning," *Proc. - 2020 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2020*, pp. 344–351, 2020, doi: 10.1109/CSCI51800.2020.00065.
- [13] A. Yilmaz, S. P. Yasar, G. Gencoglan, and B. Temelkuran, "DERM12345: A Large, Multisource Dermatoscopic Skin Lesion Dataset with 40 Subclasses," *Sci. Data*, vol. 11, no. 1, pp. 1–12, 2024, doi: 10.1038/s41597-024-04104-3.
- [14] F. Yuceyalcin, A. Yilmaz, and B. Temelkuran, "A Hierarchical Benchmark of Foundation," pp. 1–16.