

Pengembangan Aplikasi Mobile Pendeteksi Penyakit Kulit Menggunakan Model *Mobile-Net V3* Dengan Metode *Transfer Learning*

Gabriel Dwi Putra Krisma Rusan¹⁾, Nahumi Nugrahaningsih²⁾, Agus Sehatman Saragih³⁾

¹⁾²⁾³⁾Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya
Kampus Tunjung Nyaho Jalan Yos Sudarso, Palangka Raya, Kalimantan Tengah, Indonesia

¹⁾gabrielwiputra12@mhs.eng.upr.ac.id

²⁾nahumi@it.upr.ac.id

³⁾assaragih@it.upr.ac.id

Abstrak

Penyakit kulit merupakan salah satu masalah kesehatan dengan prevalensi tinggi di Indonesia, mencapai 4,60%–12,95% dari populasi atau sekitar 12,7 juta–35,9 juta orang dari 277,5 juta penduduk Indonesia. Keterbatasan jumlah dokter spesialis kulit serta rendahnya akses layanan kesehatan di daerah menyebabkan banyak kasus terlambat terdiagnosis. Keterlambatan ini dapat menimbulkan komplikasi serius, meningkatkan angka morbiditas, dan menurunkan kualitas hidup penderita. Oleh karena itu, diperlukan solusi berbasis teknologi yang dapat membantu masyarakat melakukan deteksi dini penyakit kulit secara mandiri, cepat, dan akurat.

Penelitian ini mengusulkan pengembangan aplikasi mobile pendeteksi penyakit kulit berbasis *transfer learning* dengan arsitektur MobileNetV3. *Dataset* yang digunakan merupakan gabungan dari DermNet, HAM10000, dan Mendeley Skin Disease Classification dengan total 28 kelas penyakit. Tahapan penelitian meliputi *preprocessing* data, pelatihan dan *fine-tuning* model MobileNetV3, serta evaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Model terbaik kemudian diintegrasikan ke dalam aplikasi berbasis Flutter yang menampilkan hasil diagnosis dan informasi edukatif kepada pengguna.

Hasil pengujian menunjukkan bahwa model MobileNetV3 dengan *transfer learning* menghasilkan akurasi sebesar 81%, presisi 79%, *recall* 79% dan *F1-score* 79%. Aplikasi yang dikembangkan mampu melakukan deteksi penyakit kulit dengan cepat dan efisien pada perangkat mobile. Dengan demikian, penelitian ini diharapkan dapat menjadi solusi *screening* awal yang praktis dan meningkatkan kesadaran masyarakat terhadap pentingnya diagnosis dini penyakit kulit di Indonesia.

Kata kunci: penyakit kulit, MobileNetV3, *transfer learning*, aplikasi mobile.

Abstract

Skin diseases are among the most prevalent health problems in Indonesia, affecting approximately 4.60%–12.95% of the population, or around 12.7–35.9 million people out of 277.5 million citizens. The limited number of dermatologists and the lack of access to specialized healthcare services in rural areas often lead to delayed diagnosis. Such delays can result in severe complications, increase morbidity rates, and reduce patients' quality of life. Therefore, a technology-based solution is needed to assist the public in conducting early, fast, and accurate skin disease detection independently.

This study proposes the development of a mobile skin disease detection application based on transfer learning using the MobileNetV3 architecture. The dataset used combines DermNet, HAM10000, and Mendeley Skin Disease Classification datasets, consisting of 28 classes of skin diseases. The research stages include data preprocessing, training and fine-tuning of the MobileNetV3 model, and evaluation using accuracy, precision, recall, and F1-score metrics. The

best-performing model was then integrated into a Flutter-based mobile application that displays diagnostic results and educational information for users.

The experimental results show that the MobileNetV3 model with transfer learning achieved an accuracy of 81%, precision of 79%, recall of 79%, and an F1-score of 79%. The developed application can perform skin disease detection quickly and efficiently on mobile devices. Thus, this study is expected to serve as a practical early-screening solution and raise public awareness of the importance of early diagnosis of skin diseases in Indonesia.

Keywords: skin disease, MobileNetV3, transfer learning, mobile app.

1. PENDAHULUAN

Penyakit kulit merupakan salah satu masalah kesehatan dengan prevalensi tinggi di Indonesia, yaitu 4,60%–12,95% dari populasi atau sekitar 12,7–35,9 juta orang. Berbagai jenis penyakit kulit seperti melanoma, psoriasis, dan dermatitis atopik dapat menurunkan kualitas hidup bahkan berpotensi fatal jika terlambat ditangani. Namun, keterbatasan jumlah dokter spesialis kulit yang hanya sekitar 1 untuk setiap 110.000 penduduk sehingga menghambat diagnosis dini, terutama di daerah rural [1].

Seiring meningkatnya penetrasi smartphone yang telah mencapai 79,5% pada tahun 2024 [2], terdapat peluang besar untuk mengimplementasikan kecerdasan buatan (*Artificial Intelligence*, AI) dalam mendukung skrining dini penyakit kulit berbasis citra. Berbagai penelitian menunjukkan bahwa model *deep learning* mampu menandingi akurasi dokter spesialis dalam mengklasifikasi citra penyakit kulit [3]. Namun, pengembangan sistem tersebut masih menghadapi tantangan keterbatasan data medis dan kebutuhan komputasi tinggi.

Transfer learning merupakan pendekatan yang memungkinkan pemanfaatan model pralatih (*pre-trained model*) dari *dataset* besar seperti ImageNet untuk diterapkan pada domain baru, seperti klasifikasi citra penyakit kulit, melalui proses penyesuaian (*fine-tuning*) [4]. Pendekatan ini mempercepat pelatihan dan meningkatkan akurasi meskipun data terbatas. Salah satu arsitektur populer untuk perangkat mobile adalah MobileNet, jaringan konvolusional ringan yang dirancang untuk efisiensi komputasi tanpa mengorbankan akurasi. Versi terbarunya, MobileNetV3, menggabungkan *Neural Architecture Search* (NAS) dan aktivasi *hard-swish* untuk meningkatkan performa dan efisiensi inferensi [5].

Berdasarkan hal tersebut, penelitian ini mengembangkan aplikasi mobile deteksi penyakit kulit berbasis *transfer learning* dengan arsitektur MobileNetV3. Model dilatih menggunakan gabungan tiga *dataset* besar yakni DermNet, HAM10000, dan Mendeley Skin Disease Classification yang mencakup 28 kelas penyakit kulit. Pendekatan ini diharapkan menghasilkan sistem deteksi penyakit kulit yang akurat, ringan, dan efisien, sehingga dapat digunakan secara langsung oleh masyarakat Indonesia sebagai alat bantu skrining awal.

2. TINJAUAN PUSTAKA

2.1 Penyakit Kulit

Penyakit kulit merupakan suatu gangguan kesehatan yang disebabkan oleh pertumbuhan sel kulit yang tidak normal serta pengaruh dari lingkungan dengan tingkat kejadian yang tinggi diberbagai kelompok usia secara global [6][7].

2.2 MobileNet

MobileNet adalah arsitektur *Convolutional Neural Network* (CNN) yang ringan, efisien dalam penggunaan sumber daya komputasi, dan memiliki kinerja cepat, sehingga cocok untuk diimplementasikan pada perangkat mobile [8][9].

2.2.1 MobileNetV3

MobileNetV3 menggabungkan *Neural Architecture Search* (NAS) dan optimasi manual untuk menghasilkan jaringan yang efisien dan akurat di perangkat mobile. Arsitektur ini menggunakan aktivasi *hard-swish* dan *squeeze-and-excitation blocks* untuk meningkatkan representasi fitur, serta mengoptimalkan layer awal dan akhir guna menekan *latency* tanpa mengurangi akurasi.

Tersedia dua varian, yaitu *Large* untuk akurasi tinggi dan *Small* untuk perangkat terbatas. Melalui *platform-aware NAS*, MobileNetV3 mencapai keseimbangan optimal antara kecepatan dan akurasi [5].

2.3 Transfer Learning

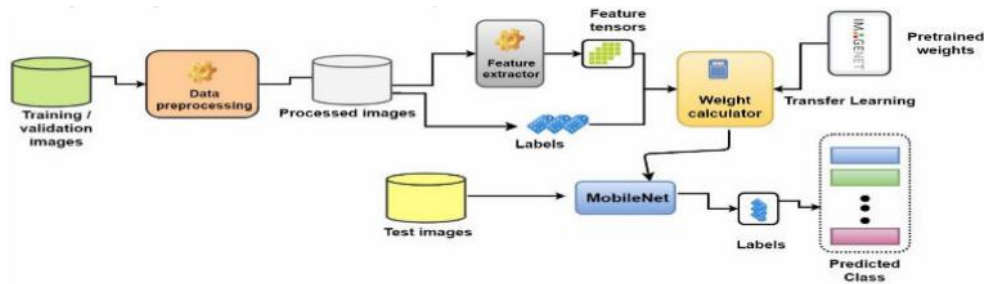
Transfer Learning merupakan suatu metode yang memanfaatkan model yang telah dilatih sebelumnya untuk mengatasi keterbatasan data dengan mengambil bagian tertentu dari model tersebut. Metode ini memungkinkan pengurangan jumlah parameter dan penggunaan pengetahuan yang telah diperoleh dari tugas sebelumnya untuk membantu dalam pengenalan atau pelatihan model baru [10][11].

2.4 Hierarchical Model

Hierarchical model merupakan pendekatan klasifikasi di mana proses prediksi diatur secara bertingkat berdasarkan struktur kelas, sehingga setiap level model menangani subset yang lebih spesifik dari data. Misalnya, dalam sistem e-commerce, kategori produk sering disusun dalam struktur pohon; model *Deep Hierarchical Classification* (DHC) menyertakan representasi hierarkis dan strategi *representation sharing* antar lapisan serta memperkenalkan fungsi *combined loss* untuk menghukum ketidakkonsistenan prediksi hierarkis, terbukti meningkatkan akurasi dibanding pendekatan klasifikasi datar [12].

3. METODE PENELITIAN

3.1 Alur Pengembangan Model



Gambar 1. Diagram Alur Pengembangan Model

Gambar 1 menunjukkan alur utama penelitian dalam pengembangan model deteksi penyakit kulit berbasis *transfer learning* dengan arsitektur MobileNet. Proses dimulai dari pengumpulan dan *preprocessing* data (penyesuaian ukuran, normalisasi, dan augmentasi), lalu citra diproses menggunakan *feature extractor* MobileNet ber-bobot *pretrained* dari ImageNet. Hasil ekstraksi fitur digunakan untuk melatih dan menyesuaikan bobot model (*fine-tuning*) agar mampu mengenali pola spesifik penyakit kulit. Model terlatih kemudian diuji dengan citra uji untuk menghasilkan prediksi kelas penyakit kulit secara otomatis.

3.1.1 Sumber Data

Penelitian ini menggunakan tiga *dataset* citra penyakit kulit yang tersedia secara publik, yaitu DermNet, HAM10000, dan Mendeley Skin Disease Classification *Dataset*. Setiap *dataset* memiliki karakteristik berbeda, baik dari segi jumlah kelas penyakit, jumlah citra per kelas, maupun teknik pengambilan gambar (klinis maupun dermatoskopik). Untuk perincian tiap *dataset* dapat dilihat pada Tabel 1 di bawah.

Tabel 1. Sumber dan detail data

No	Nama <i>Dataset</i>	Asal	Jumlah Data	Jumlah Kelas
1	Dermnet	Internet (Kaggle)	19500 gambar	23
2	HAM10000	Internet (Kaggle)	10015 gambar	7
3	Mendeley Skin Disease Classification	Internet (Mendeley)	9548 gambar	5

Penggunaan ketiga *dataset* tersebut bertujuan untuk memperkaya variasi data dan cakupan kelas penyakit kulit dalam penelitian ini.

3.1.2 Preprocessing Data

Penyaringan Data

Pada tahap *preprocessing*, dilakukan penyaringan data untuk memastikan kualitas dan konsistensi *dataset* sebelum pelatihan. Proses ini mencakup penghapusan gambar yang tidak relevan, rusak, atau duplikat, serta penyesuaian label yang salah agar sesuai dengan kategori sebenarnya. Pemeriksaan dilakukan secara manual dan semiotomatis menggunakan skrip deteksi duplikasi dan validasi label, sehingga dihasilkan *dataset* yang bersih, konsisten, dan siap digunakan untuk pelatihan model.

Penggabungan *Dataset*

Setelah dilakukan penyaringan data pada tiap *dataset*, ketiga *dataset* kemudian digabungkan menjadi satu *dataset* gabungan. Penggabungan *dataset* bisa dilakukan secara manual maupun semi otomatis dengan menggunakan skrip penggabungan *dataset*.

Split Data

Setelah *dataset* digabungkan menjadi satu, tahap selanjutnya adalah *split* atau membagi data menjadi 3, yaitu data latih, validasi dan test dengan rasio pembagian berturut-turut 70/15/15. Pembagian dilakukan secara *stratified* agar proporsi tiap kelas tetap terjaga di setiap *subset*.

Augmentasi Data

Untuk meningkatkan keragaman data latih dan mencegah *overfitting*, penelitian ini menerapkan augmentasi sederhana pada tahap *preprocessing* yang hanya dilakukan pada data latih. Transformasi yang digunakan meliputi *random resized crop* (skala 0.7–1.0), *resize* ke 224×224 piksel sesuai standar MobileNetV3 [13], *random horizontal flip* dengan probabilitas 0.5, serta normalisasi menggunakan fungsi *preprocess_input()*. Pendekatan ini dipilih karena cukup memperkaya variasi spasial tanpa mengubah struktur klinis lesi, sehingga hasil pelatihan tetap realistis dan relevan dengan kondisi medis sebenarnya.

3.1.3 Pengembangan Model

Arsitektur dan Tahapan *Transfer Learning*

Penelitian ini menggunakan arsitektur MobileNetV3-Large sebagai backbone karena efisien, berukuran kecil, dan memiliki akurasi tinggi untuk klasifikasi citra pada perangkat mobile. Dengan pendekatan *transfer learning*, bobot *pre-trained* dari ImageNet disesuaikan melalui *fine-tuning* pada beberapa layer akhir agar mampu mengenali pola spesifik penyakit kulit. Sebagian layer backbone dibuka (*unfreeze*) dan dilatih ulang bersama *classification head* dengan *learning rate* kecil untuk menghasilkan model yang adaptif dan akurat.

Training Model

Pelatihan model menggunakan optimizer Adam dengan *learning rate* 0.0001 dan *batch size* 32 untuk menjaga stabilitas dan generalisasi [14]. Fungsi *loss* yang digunakan adalah *categorical crossentropy* dengan *class weight* guna mengatasi ketidakseimbangan kelas. Beberapa *callback* diterapkan, seperti *Early Stopping*, *ReduceLROnPlateau*, *Model Checkpoint*, dan *CSV Logger* untuk memantau serta mengoptimalkan proses pelatihan hingga maksimal 100 *epoch*.

Sebelum dilakukan *training* pada model, perlu menambahkan parameter *class_weight*. *Class weight* berfungsi untuk memitigasi risiko ketidakseimbangan data sebagai akibat dari penggabungan *dataset* pada tahap sebelumnya. Pemberian *class weight* bertujuan untuk memberikan bobot lebih besar pada kelas-kelas dengan jumlah data lebih sedikit, sehingga model tidak hanya fokus pada kelas mayoritas, melainkan juga mampu mempelajari karakteristik kelas minoritas. Persamaan (1) berikut merupakan rumus untuk menentukan *class weight* untuk masing-masing kelas pada *dataset*.

$$w_j = \frac{N}{k \times n_j} \quad (1)$$

Keterangan:

w_j = bobot untuk kelas ke- j

N = total jumlah sampel pada *dataset* latih
k = jumlah kelas
 n_j = jumlah sampel pada kelas ke-j

3.1.4 Evaluasi Model

Evaluasi efektivitas model dilakukan berdasarkan hasil pengujian pada data uji terpisah. Metrik evaluasi utama meliputi:

Akurasi

Akurasi merupakan presentase prediksi yang benar pada seluruh kelas. Akurasi dihitung dengan menggunakan persamaan (2) di bawah ini.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Keterangan:

Akurasi = prediksi yang benar pada seluruh kelas
TP (*True Positive*) = jumlah sampel positif yang berhasil diprediksi benar
TN (*True Negative*) = jumlah sampel negatif yang berhasil diprediksi benar
FP (*False Positive*) = jumlah sampel negatif yang salah diprediksi sebagai positif
FN (*False Negative*) = jumlah sampel positif yang salah diprediksi sebagai negatif

Presisi

Presisi mengukur seberapa tepat prediksi positif yang dibuat model. Semakin tinggi presisi, semakin sedikit kesalahan model dalam mengklasifikasikan data negatif sebagai positif. Presisi dihitung dengan menggunakan persamaan (3) di bawah ini.

$$Presisi = \frac{TP}{TP+FP} \quad (3)$$

Keterangan:

Presisi = ketepatan prediksi positif yang dibuat model
TP (*True Positive*) = jumlah sampel positif yang berhasil diprediksi benar
FP (*False Positive*) = jumlah sampel negatif yang salah diprediksi sebagai positif

Recall

Recall mengukur kemampuan model dalam menemukan semua data positif yang benar. Semakin tinggi *recall*, semakin sedikit kasus positif yang terlewatkan. *Recall* dihitung dengan menggunakan persamaan (4) di bawah ini.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Keterangan:

Recall = tingkat kelengkapan prediksi positif
TP (*True Positive*) = jumlah sampel positif yang berhasil diprediksi benar
FN (*False Negative*) = jumlah sampel positif yang salah diprediksi sebagai negatif

F1-Score

F1-Score merupakan rata-rata harmonis antara presisi dan *recall*. Metrik ini penting ketika data bersifat tidak seimbang, karena memberikan gambaran yang lebih adil antara kesalahan *false positive* dan *false negative*. *F1-Score* dihitung dengan menggunakan persamaan (5) di bawah ini.

$$F1 = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (5)$$

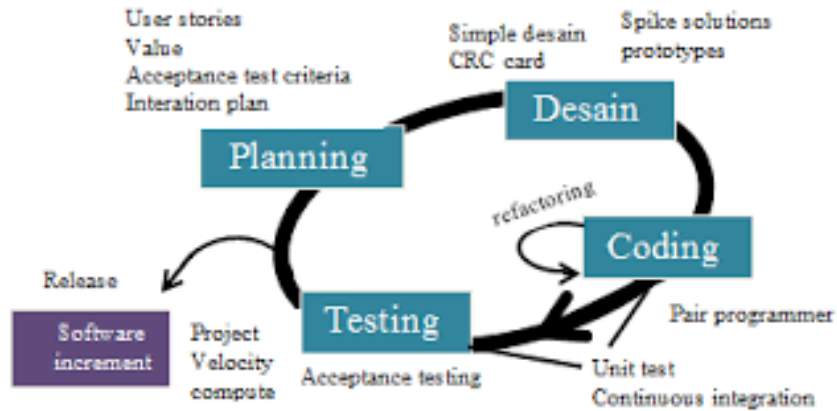
Keterangan:

F1 = rata-rata harmonis antara presisi dan *recall*

Presisi = ketepatan prediksi positif yang dibuat model
Recall = tingkat kelengkapan prediksi positif

3.2 Metodologi Pengembangan Perangkat Lunak

Metodologi pengembangan perangkat lunak yang digunakan dalam penelitian ini adalah *Extreme Programming (XP)*. XP dipilih karena mampu beradaptasi terhadap perubahan kebutuhan pengguna, mendukung iterasi cepat, dan menghasilkan perangkat lunak yang berkualitas tinggi dalam waktu pengembangan yang relatif singkat. *Extreme Programming* memiliki beberapa tahapan seperti yang ditampilkan pada Gambar 2 di bawah ini.



Gambar 2. Extreme Programming

Pada tahap *planning*, menentukan tujuan jangka pendek dan panjang yang ingin dicapai. Semua fitur yang diinginkan akan disusun dalam bentuk cerita pengguna (*user stories*). Setiap cerita pengguna dipecah menjadi tugas-tugas yang lebih kecil dan ditargetkan untuk diselesaikan dalam iterasi yang singkat. Pada tahap desain, fokus utama adalah menciptakan solusi sederhana dan langsung untuk setiap cerita pengguna yang telah ditentukan. Dalam XP, desain dilakukan secara iteratif dan berkelanjutan, dengan tujuan menghasilkan sistem yang dapat berkembang dengan mudah seiring waktu. Proses *coding* mencakup penerapan model yang telah dirancang menjadi antarmuka pengguna dengan menggunakan bahasa pemrograman seperti Dart dan Python secara terstruktur. Pada tahap *testing*, aplikasi yang telah dikembangkan akan diuji coba. Pengujian ini dilakukan oleh pengguna sistem dan difokuskan pada pemeriksaan fitur serta kinerja keseluruhan sistem.

4. PEMBAHASAN

4.1 Preprocessing Data

Setelah dilakukan penyaringan data pada ketiga *dataset*, maka dari kurang lebih 35000 data tersaring menjadi 16.819 data yang bersih. Rincian penyaringan dapat dilihat pada Tabel 2.

Tabel 2. Rincian hasil penyaringan data tiap *dataset*

No	Nama <i>dataset</i>	Jumlah awal gambar	Jumlah akhir gambar
1	Dermnet	19500	14368
2	HAM10000	10015	2110
3	Mendeley Skin Disease Image	9548	341

Setelah menyaring data pada masing-masing *dataset*, langkah selanjutnya adalah menggabungkan semuanya menjadi satu *dataset* gabungan. Tahap penggabungan menghasilkan total 28 kelas. Rincian hasil penggabungan *dataset* dapat dilihat pada Tabel 3.

Tabel 3. Daftar kelas gabungan dan asal *dataset*-nya

Kelas Gabungan (28 Kelas Final)	Nama Kelas Asli di Mendeley	Nama Kelas Asli di HAM10000	Nama Kelas Asli di DermNet	Total Data
Acne	Acne	–	Acne and Rosacea Photos	400
Actinic keratoses	–	Actinic keratoses (AKIEC)	Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	616
Alopecia	–	–	Alopecia Photos	289
Atopic Dermatitis	–	–	Atopic Dermatitis Photos	236
Basal Cell Carcinoma	–	Basal cell carcinoma (BCC)	Seborrheic Keratoses and other Benign Tumors	862
Benign keratosis-like lesions	–	Benign keratosis-like (BKL)	Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	1956
Bullous	–	–	Bullous Disease Photos	422
Cellulitis Impetigo	–	–	Cellulitis, Impetigo and other Bacterial Infections Photos	360
Dermatofibroma	–	Dermatofibroma (DF)	–	115
Eczema	–	–	Eczema Photos	1416
Exanthems	–	–	Exanthems and Drug Eruptions Photos	483
Herpes HPV	–	–	Herpes HPV Infections Photos	1014
Light Diseases	–	–	Light Diseases Photos	458
Lupus	–	–	Lupus and other Connective Tissue Diseases Photos	514
Melanoma	–	Melanoma (MEL)	Melanoma Photos	384
Melanocytic Nevi	–	Melanocytic nevi (NV)	–	1100
Nail Fungus	–	–	Nail Fungus Photos	216
Pigmentation Disorders	Pigmentation Disorders	–	–	97
Poison Ivy	–	–	Poison Ivy Photos	273
Psoriasis	Nail Psoriasis	–	Psoriasis Photos	937
SJS-TEN	SJS-TEN	–	–	113
Scabies Lyme	–	–	Scabies, Lyme Disease and other Infestations and Bites Photos	422
Systemic Disease	–	–	Systemic Disease Photos	532
Tinea Candidiasis	–	–	Tinea, Candidiasis and other Fungal Infections Photos	1381
Urticaria Hives	–	–	Urticaria Hives Photos	265
Vascular Lesions	–	Vascular lesions (VASC)	Vascular Tumors, Hemangiomas, and other Vascular Lesions Photos	662
Vasculitis	–	–	Vasculitis Photos	514
Wart Molluscum	–	–	Warts, Molluscum and other Viral Infections Photos	782

Tabel 3 memperlihatkan proses standarisasi nama kelas dari masing-masing *dataset* asal menjadi 28 kelas gabungan. Beberapa kelas berasal dari lebih dari satu *dataset*, misalnya Basal Cell Carcinoma yang pada HAM10000 dinamai “Basal Cell Carcinoma (BCC)” sedangkan pada DermNet masuk dalam kategori “Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions”. Setelah digabung, keduanya diseragamkan menjadi kelas “Basal Cell Carcinoma”. Contoh lain adalah Acne, yang pada Mendeley disebut “Acne” dan pada DermNet disebut “Acne

and Rosacea Photos”, kemudian distandarkan menjadi kelas “Acne” pada *dataset* gabungan. Dengan adanya standarisasi ini, model dapat dilatih menggunakan label yang konsisten sehingga mengurangi kebingungan akibat perbedaan penamaan antar *dataset*, serta memastikan seluruh variasi data yang merepresentasikan penyakit yang sama masuk ke kelas yang tepat.

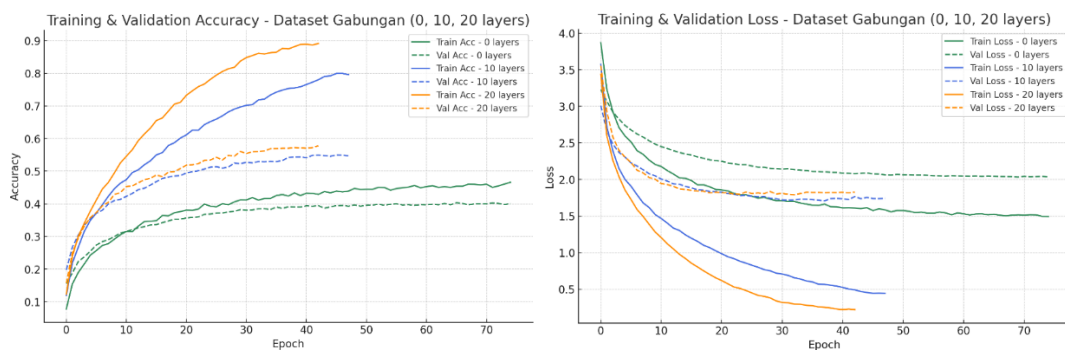
Setelah dilakukan penggabungan menjadi satu *dataset*, data dibagi menjadi tiga yaitu data latih, validasi dan test dengan rasio secara berurutan 70/15/15. Setelah itu data latih akan diaugmentasi secara *on-the-fly* sesuai dengan teknik augmentasi sebelumnya saat melatih model.

4.2 Training Model

Pelatihan model dilakukan dalam tiga skenario untuk menganalisis pengaruh jumlah *layer* backbone MobileNetV3-Large yang dibuka (*unfreeze*) terhadap performa model. Strategi ini bertujuan menemukan konfigurasi optimal antara *general feature extraction* (sedikit *layer* dibuka) dan *task-specific adaptation* (lebih banyak *layer* dibuka). Pada Skenario 1 (*unfreeze* 0 Layer), seluruh bobot backbone dipertahankan (*frozen*) dan hanya lapisan klasifikasi akhir yang dilatih ulang menggunakan *dataset* DermNet, dengan tujuan memanfaatkan representasi umum dari *ImageNet* tanpa penyesuaian mendalam. Skenario 2 (*unfreeze* 10 Layer) membuka sepuluh *layer* terakhir untuk dilatih ulang agar model memiliki kemampuan adaptasi terbatas terhadap pola khusus penyakit kulit sekaligus menjaga stabilitas bobot awal. Sementara itu, skenario 3 (*unfreeze* 20 Layer) melibatkan pelatihan ulang dua puluh *layer* terakhir, memberikan fleksibilitas lebih tinggi bagi model untuk beradaptasi dengan karakteristik data, namun berisiko lebih besar mengalami *overfitting* jika data pelatihan tidak cukup kuat.

4.3 Evaluasi

Gambar 3 dan Tabel 5 menunjukkan perbandingan performa model pada tiga skenario jumlah *layer* yang dibuka (*unfreeze*) selama pelatihan. Pada skenario *unfreeze* 0 layer, akurasi validasi hanya mencapai sekitar 43%, menunjukkan bahwa model hanya berfungsi sebagai *feature extractor* tanpa mampu menyesuaikan diri dengan pola baru pada *dataset*. Ketika 10 *layer* terakhir dibuka, performa meningkat signifikan dengan akurasi validasi sekitar 54% dan kurva pelatihan yang lebih stabil. Sementara itu, skenario *unfreeze* 20 *layer* menghasilkan performa terbaik dengan akurasi validasi 56,5% dan *macro F1-score* 57,4%, disertai selisih kecil antara kurva pelatihan dan validasi, meskipun masih terdapat sedikit fluktuasi yang mengindikasikan *overfitting* ringan. Secara keseluruhan, semakin banyak *layer* yang dibuka untuk *fine-tuning*, semakin baik kemampuan model dalam mengenali pola spesifik pada data penyakit kulit.



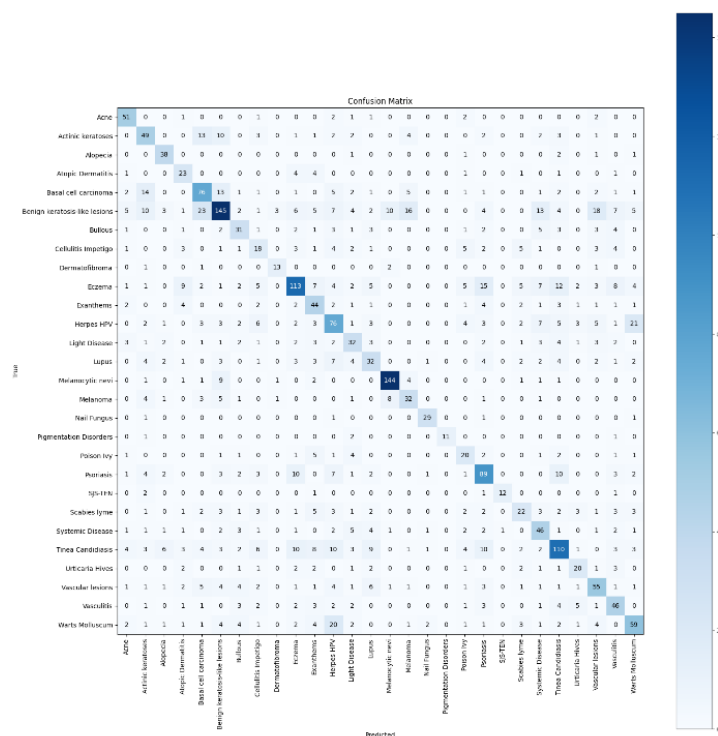
Gambar 3. Kurva akurasi dan *loss* pada data latih dan validasi model

Tabel 4. Perbandingan metrik tiap skenario model

Skenario (<i>Unfreeze</i>)	Accuracy	Macro Precision	Macro Recall	Macro F1
0 layer	43,1%	43,3%	50,3%	44,3%
10 layer	54,4%	55,2%	58,6%	56,1%
20 layer	56,5%	57,3%	58,7%	57,4%

Gambar 4 menunjukkan *confusion matrix* pada skenario terbaik yaitu skenario 20 *layer* menunjukkan bahwa sebagian besar kelas dapat dikenali dengan baik, terutama kelas dengan

karakteristik yang cukup khas seperti Acne, Urticaria Hives, dan Melanoma. Namun, masih terdapat sejumlah kelas yang sulit dibedakan satu sama lain. Misalnya, Psoriasis sering tertukar dengan Eczema maupun Exanthems, sedangkan Actinic keratoses kadang salah diprediksi sebagai Basal cell carcinoma atau Benign keratosis-like lesions. Pada kelompok infeksi, Tinea Candidiasis masih kerap salah diprediksi sebagai Cellulitis Impetigo, dan Scabies lyme kadang terklasifikasi sebagai Poison Ivy karena kemiripan manifestasi kulitnya. Sementara itu, kelas langka seperti Dermatofibroma dan Systemic Disease menunjukkan jumlah prediksi benar yang lebih rendah, yang dapat dipengaruhi keterbatasan jumlah data. Secara umum, meskipun performa terbaik hanya mencapai *macro F1* sebesar 57,4%, model sudah menunjukkan kemampuan yang cukup baik dalam mengenali sebagian besar kelas, dengan tantangan utama pada kelas yang memiliki tumpang tindih visual atau jumlah data yang terbatas.



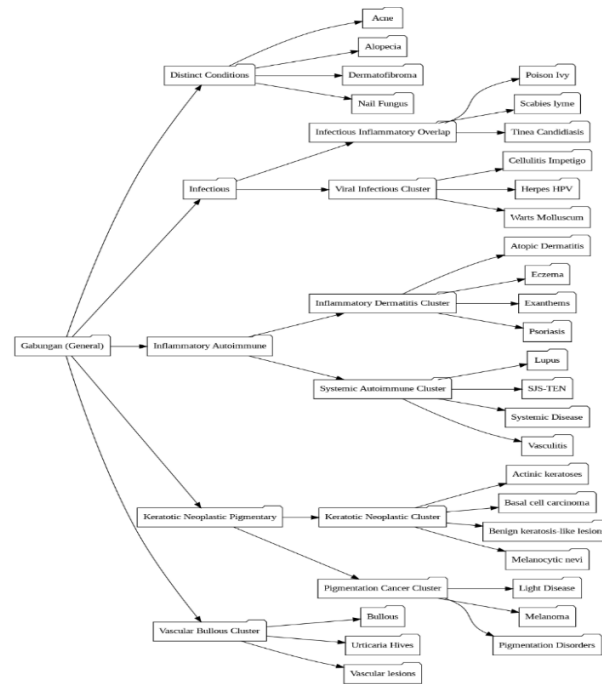
Gambar 4. Confusion matrix model pada data test

4.4 Implementasi Model dengan Pendekatan Hierarkis

Penelitian ini menggunakan pendekatan hierarkis untuk mengatasi keterbatasan model *flat classification* yang hanya mencapai akurasi 55–56% dan *macro F1-score* 54–57%. Dalam pendekatan ini, penyakit kulit dikelompokkan berdasarkan etiologi dan manifestasi klinis menjadi beberapa kategori seperti Infectious, Inflammatory Autoimmune, Keratotic Neoplastic Pigmentary, Vascular Bullous Disorders, dan Distinct Conditions [15]. Setiap node hierarki dilatih dengan model MobileNetV3 terpisah yang bertugas membedakan subkategori di dalamnya, sehingga proses inferensi dilakukan bertahap dari model umum hingga spesifik. Strategi ini meniru cara kerja dokter kulit dalam mendiagnosis penyakit dan membantu mengurangi kesalahan klasifikasi lintas-domain, menghasilkan sistem yang lebih terstruktur dan akurat [16].

4.4.1 Preprocessing Data

Preprocessing data tidak jauh berbeda dengan sebelumnya, yang berbeda hanya pada saat sebelum dilakukan pembagian data, struktur *dataset* diatur ulang dengan pendekatan hierarkis yang mana pengelompokan kelas penyakit kulitnya diambil dari *website* milik DermNet [15]. Pendekatan hierarkis menghasilkan struktur dataset baru seperti pada Gambar 5 berikut.



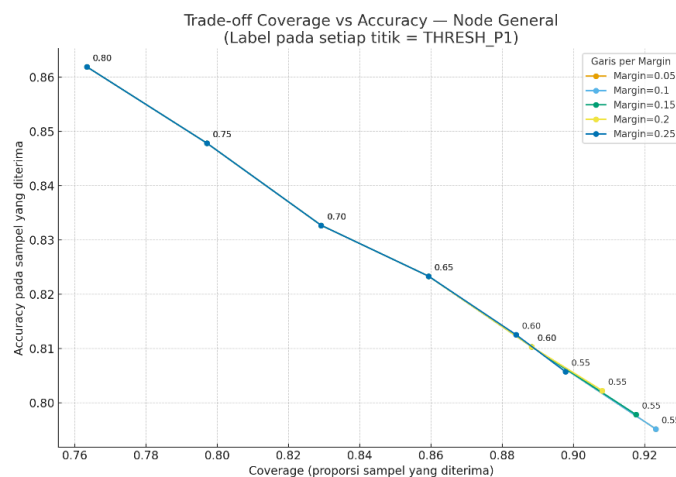
Gambar 5. Stuktur *dataset* dengan pendekatan hierarkis

4.4.2 Training

Training pada pendekatan hierarkis menggunakan 3 skenario yang sama dengan pendekatan sebelumnya.

4.4.3 Evaluasi

Pipeline inferensi hierarkis dirancang menyerupai proses diagnosis dokter kulit, dimulai dari *General Node* yang membedakan lima kategori utama, dilanjutkan ke *Cluster* dan *Subcluster Node* hingga menghasilkan prediksi satu dari 28 kelas beserta *confidence score*-nya. Pendekatan bertahap ini membagi beban klasifikasi secara terstruktur untuk mengurangi kesalahan lintas-domain. Seperti terlihat pada Gambar 6, dengan penyetulan ambang (*threshold*) $p_1 = 0,80$ dan *margin* 0,25, sistem mencapai akurasi 86,2% pada sampel terpilih, meningkatkan reliabilitas serta mencegah kesalahan rute akibat prediksi ambigu.



Gambar 6. Kurva *trade-off coverage vs accuracy* pada *node general*

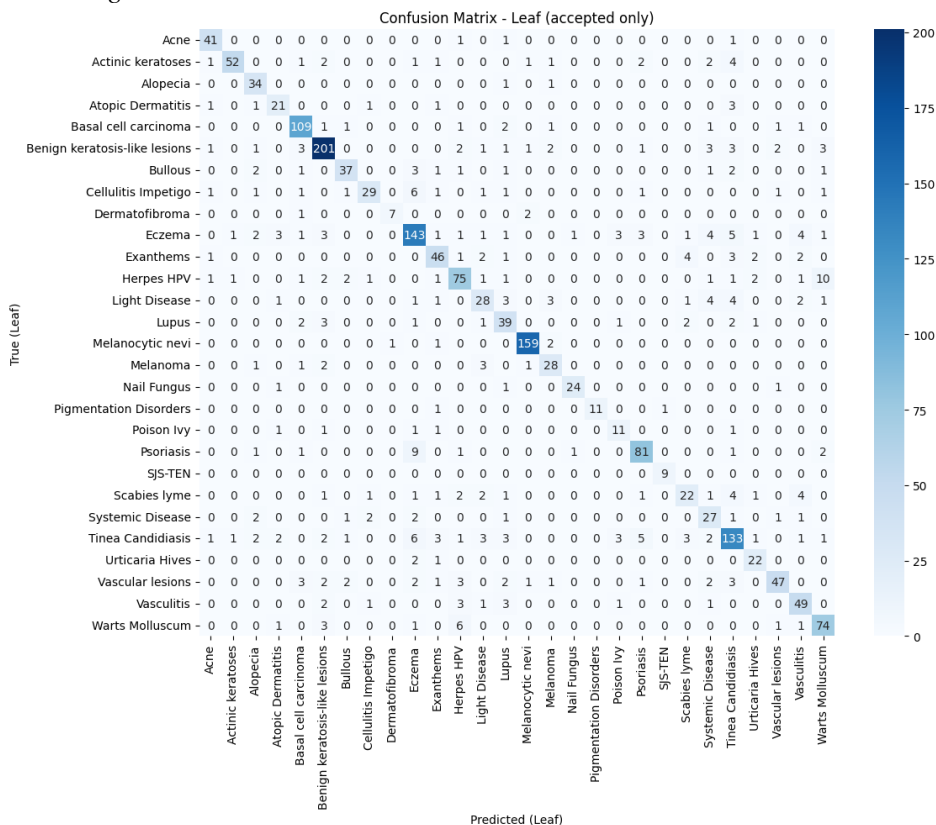
Tabel 6 memperlihatkan adanya peningkatan signifikan pada seluruh metrik ketika menggunakan pendekatan hierarkis dibandingkan klasifikasi flat. Akurasi naik sekitar 15,6%

(dari 56,5% menjadi 72,1%), dan nilai F1 makro meningkat dari 57,4% menjadi 70,0%. Selanjutnya, ketika ditambahkan mekanisme *tuned thresholds* pada Node General, performa pipeline semakin membaik. Meskipun jumlah sampel yang diloloskan berkurang menjadi 76,34% dari total (1926 citra), akurasi pada sampel yang lolos naik drastis hingga 81% dengan macro precision, recall, dan F1 masing-masing sebesar 79%. Hal ini menunjukkan bahwa strategi *threshold* berperan sebagai filter yang efektif: sistem menjadi lebih selektif dalam meloloskan prediksi, sehingga kesalahan lintas-domain dapat ditekan lebih jauh, sekaligus meningkatkan reliabilitas hasil inferensi.

Tabel 5. Perbandingan metrik antara pendekatan awal dan pendekatan hierarkis

Model	Accuracy	Macro Precision	Macro Recall	Macro F1	Support
Flat (28 kelas)	56,5%	57,3%	58,7%	57,4%	2523
Hierarkis	72,1%	70%	70%	70%	2523
Hierarkis + Tuned Thresholds pada Node General	81%	79%	79%	79%	1926 (76,34% dari total)

Seperti terlihat pada Gambar 7 di bawah, *confusion matrix* model hierarkis dengan *tuned thresholds* menunjukkan pola diagonal yang kuat meskipun hanya 76,34% citra yang diloloskan. Prediksi benar per kelas tetap tinggi, sementara kesalahan pada kelas yang sebelumnya tumpang tindih seperti Eczema, Atopic Dermatitis, dan Exanthems berkurang signifikan. Hal ini menunjukkan bahwa penerapan *threshold* berfungsi efektif sebagai filter selektif yang meningkatkan reliabilitas hasil. Dengan demikian, pendekatan hierarkis dengan *tuned thresholds* terbukti mampu menekan kesalahan lintas-domain dan memperkuat akurasi sistem untuk keperluan *screening*.



Gambar 7. Confusion matrix inferensi keseluruhan pendekatan hierarkis + *tuned threshold node general*

4.5 Implementasi Aplikasi Mobile

4.5.1 Testing

Setelah aplikasi berhasil dirancang dan dikembangkan, selanjutnya adalah melakukan pengujian pada aplikasi untuk memastikan bahwa semua fungsionalitas berjalan dengan baik. Salah satu metode yang sering digunakan untuk menguji aplikasi yang telah dibuat adalah *blackbox testing*. Testing dilakukan oleh seseorang yang memiliki masalah pada kulit nya dan telah melakukan pemeriksaan pada dokter kulit dan mendapatkan diagnosa yaitu penyakit Acne dari dokter. Hasil test dapat dilihat pada Tabel 4.33 dibawah ini.

Tabel 6. *Blackbox Testing*

No	Kondisi Awal	Hal yang Dilakukan	Hal yang Diinginkan	Hasil
1	Halaman Utama	Klik <i>button scan</i>	Masuk pada halaman ambil gambar penyakit kulit	<i>Valid</i>
2	Halaman Ambil Gambar Penyakit Kulit	Klik <i>button</i> tengah untuk mengambil gambar penyakit	Menampilkan hasil dari pengambilan gambar	<i>Valid</i>
3	Halaman Ambil Gambar Penyakit Kulit	Klik <i>button</i> tengah untuk mengambil gambar penyakit	Menampilkan <i>pop up</i> pesan bahwa gambar tidak terdeteksi dan melakukan pengambilan gambar ulang	<i>Valid</i>
4	Halaman Ambil Gambar Penyakit Kulit	Klik <i>button</i> galeri untuk mengambil gambar penyakit dari galeri	Menampilkan hasil dari pengambilan gambar	<i>Valid</i>
5	Halaman Ambil Gambar Penyakit Kulit	Klik <i>button</i> galeri untuk mengambil gambar penyakit dari galeri	Menampilkan <i>pop up</i> pesan bahwa gambar tidak terdeteksi dan melakukan pengambilan gambar ulang	<i>Valid</i>
6	Halaman Detail Diagnosis	Melihat hasil screening dan informasi terkait hasil prediksi	Menampilkan penyakit dan informasi yang sesuai dengan diagnosis dokter kulit	<i>Valid</i>
7	Halaman <i>Screening Log</i>	Klik <i>search bar</i>	Menampilkan pencaharian riwayat penyakit yang telah dideteksi sebelumnya	<i>Valid</i>
8	Halaman <i>Screening Log</i>	Klik <i>button</i> filter tanggal dan bulan untuk melihat riwayat penyakit kulit	Menampilkan data riwayat penyakit kulit sesuai dengan tanggal dan bulan yang dipilih	<i>Valid</i>

Hasil pengujian menunjukkan bahwa aplikasi DermaScan mampu mendeteksi penyakit kulit secara konsisten dengan hasil dokter, seperti pada kasus *Acne* dengan *confidence score* 0.92. Seluruh fitur utama yaitu pengambilan gambar, pratinjau, deteksi ulang, dan riwayat diagnosis berfungsi sempurna (100% valid pada delapan skenario *blackbox*). Dengan akurasi model di atas 81% pada skenario terbaik (hierarkis + *tuned thresholds*), aplikasi ini terbukti andal sebagai alat bantu skrining awal sebelum konsultasi ke dokter.

5. KESIMPULAN

Penelitian ini berhasil mengembangkan aplikasi mobile DermaScan berbasis *transfer learning* dengan arsitektur MobileNetV3, menggunakan gabungan tiga dataset publik (DermNet, HAM10000, dan Mendeley) yang mencakup 28 kelas penyakit kulit. Pendekatan *fine-tuning* dan hierarkis terbukti meningkatkan performa model dengan akurasi terbaik mencapai 81% serta distribusi prediksi yang konsisten. Implementasi ke dalam aplikasi Flutter berjalan baik, fitur kamera, galeri, hasil deteksi, dan riwayat diagnosis berfungsi sempurna, dengan hasil uji selaras dengan diagnosis dokter (*confidence score* 0.92 pada kasus *Acne*). Secara keseluruhan, DermaScan terbukti efektif sebagai alat bantu skrining dini penyakit kulit yang akurat, edukatif, dan mudah diakses masyarakat.

6. DAFTAR PUSTAKA

- [1] N. S. Rahayu, A. D. Puteri, and L. M. A. Isnaeni, "Hubungan Perilaku Masyarakat Dan Penggunaan Air Sungai Dengan Gangguan Penyakit Kulit Di Desa Kampung Pinang Wilayah Kerja Puskesmas Pantai Raja," *J. Imliah Ilmu Kesehat.*, vol. 1, no. 3, p. 2023, 2023.

-
- [2] I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, "Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application," *Electron.*, vol. 11, no. 9, pp. 1–19, 2022, doi: 10.3390/electronics11091294.
- [3] I. Abunadi and E. M. Senan, "Deep learning and machine learning techniques of diagnosis dermoscopy images for early detection of skin diseases," *Electron.*, vol. 10, no. 24, 2021, doi: 10.3390/electronics10243158.
- [4] J. M. *Honey Janorial and PoojaPatre3, "images using Transfer Learning Technique," pp. 888–895, 2020.
- [5] S. Qian, "2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2022," *6th Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2022 - Proc.*, no. Icbaie, pp. 490–497, 2022, doi: 10.3390/proceedings2021072002.
- [6] H. A. Pangestu and Kusriani, "Peningkatan kinerja arsitektur ResNet50 untuk menangani masalah overfitting dalam klasifikasi penyakit kulit," *TEMATIK*, vol. 11, no. 1, pp. 65–71, 2024, doi: 10.38204/tematik.v11i1.1876.
- [7] M. R. A. Ar Rasyid, K. Khotimah, and S. Suaedah, "Sistem pakar pra-diagnosa penyakit kulit dengan metode forward chaining dan CNN," *Jurnal Insan Peduli Informatika, Sistem Informasi dan Sains Data (JIPIS)*, pp. 23–33, 2025, doi: 10.xxxxx/xxxxx.
- [8] F. M. Utami, R. Magdalena, and S. Saidah, "Deteksi Jenis Kulit Wajah Menggunakan Convolutional Neural Network Arsitektur MobileNet (*Detection of Facial Skin Type Classification Using Convolutional Neural Network With MobileNet Architecture*)," 2022.
- [9] A. Fuadi, A. Suharso, and J. H. Ronggo Waluyo, "Perbandingan Arsitektur MobileNet dan NASNetMobile untuk Klasifikasi Penyakit pada Citra Daun Kentang," *Universitas Singaperbangsa Karawang*, 2022.
- [10] M. I. Wahid, A. Lawi, D. A. Muh, and A. Siddik, "Prosiding Seminar Nasional Teknik Elektro dan Informatika (SNTEI) 2022 – Teknik Informatika," *SNTEI 2022*, 2022.
- [11] M. H. Fadlun and U. Hayati, "Klasifikasi tumor otak menggunakan Convolutional Neural Network dan transfer learning," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, pp. 289–295, 2024.
- [12] D. Gao, W. Yang, H. Zhou, Y. Wei, Y. Hu, and H. Wang, "Deep hierarchical classification for category prediction in e-commerce system," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, vol. 2020-July, pp. 64–68, 2020, doi: 10.18653/v1/2020.ecnlp-1.10.
- [13] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. Le, and H. Adam, "Searching for MobileNetV3," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, 2019, doi: 10.1109/ICCV.2019.00148.
- [14] S. Kim, J. Lee, and H. Park, "Determination of optimal batch size of deep learning models with time series data," *Sustainability*, vol. 16, no. 14, p. 5936, 2024.
- [15] DermNetNZ, "Dermatology disease categories and classification," *DermNet New Zealand*, 2024. [Online]. Available: <https://dermnetnz.org>.
- [16] Z. Yu, T. Nguyen, Y. Gal, L. Ju, S. S. Chandra, L. Zhang, P. Bonnington, V. Mar, Z. Wang, and Z. Ge, "Skin lesion recognition with class-hierarchy regularized hyperbolic embeddings," in *Medical Image Computing & Computer Assisted Intervention – MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds., vol. 13433, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2022, pp. 594–603, doi: 10.1007/978-3-031-16437-8_57.
-