



Evaluasi Efektifitas Optimizer Adam dan SGD pada Klasifikasi Citra Dermoskopi dengan MobileNetV4

Ahmad Naufal*, Nur Rachmat

Fakultas Ilmu Komputer dan Rekayasa, Program Studi Informatika, Universitas Multi Data Palembang, Palembang, Indonesia

Email: ^{1,*}madnaufal12@gmail.com, ²nur.rachmat@mdp.ac.id

Email Penulis Korespondensi: madnaufal12@gmail.com

Abstrak—Penyakit kulit merupakan salah satu masalah kesehatan yang umum dijumpai dan memerlukan diagnosis yang cepat serta akurat. Keterbatasan tenaga ahli dermatologi dan tingginya subjektivitas diagnosis konvensional mendorong pengembangan sistem klasifikasi otomatis berbasis kecerdasan buatan. Penelitian ini bertujuan membandingkan performa optimizer Adam dan Stochastic Gradient Descent (SGD) pada arsitektur MobileNetV4 untuk klasifikasi delapan kelas penyakit kulit menggunakan dataset ISIC 2019. Dataset yang digunakan terdiri dari 23.257 citra dermoskopi valid setelah melalui tahap preprocessing yang mencakup penghapusan citra duplikat, penghilangan artefak rambut menggunakan metode blackhat morphology, dan strategi sampling asimetris di mana kelas mayoritas dibatasi maksimal 2.000 citra sementara kelas minoritas diperbanyak melalui augmentasi hingga mencapai jumlah target, guna menangani ketidakseimbangan kelas ekstrem dengan rasio hingga 53:1. Model dilatih menggunakan strategi pelatihan tiga fase dengan pendekatan gradual unfreezing pada backbone MobileNetV4 yang telah diinisialisasi dengan bobot pretrained ImageNet. Seluruh konfigurasi pelatihan dibuat identik untuk kedua optimizer kecuali algoritma optimasi dan learning rate-nya, guna memastikan perbandingan yang adil. Hasil evaluasi pada data uji menunjukkan bahwa optimizer Adam menghasilkan akurasi sebesar 71,07% dengan macro F1-score 0,72, sedangkan SGD menghasilkan akurasi 58,06% dengan macro F1-score 0,57. Adam mengungguli SGD pada seluruh delapan kelas penyakit kulit yang diklasifikasikan. Perbedaan performa sebesar 13,01% mengindikasikan bahwa mekanisme adaptive learning rate yang dimiliki Adam lebih efektif untuk dataset dermoskopi dengan distribusi kelas yang tidak seimbang dibandingkan SGD. Meskipun demikian, perlu dicatat bahwa Adam memerlukan memori komputasi yang lebih besar dibandingkan SGD akibat penyimpanan estimasi momen pertama dan kedua per parameter, sehingga trade-off efisiensi komputasi perlu dipertimbangkan dalam implementasi pada perangkat dengan sumber daya terbatas. Penelitian ini memberikan kontribusi empiris dalam pemilihan optimizer yang optimal untuk klasifikasi lesi kulit berbasis arsitektur ringan.

Kata Kunci: Adam Optimizer; Dermoskopi; ISIC 2019; Klasifikasi Penyakit Kulit; MobileNetV4; SGD; Transfer Learning

Abstract—Skin disease is one of the most common health problems and requires fast and accurate diagnosis. The limited availability of dermatology specialists and the high subjectivity of conventional diagnosis have driven the development of artificial intelligence-based automatic classification systems. This study aims to compare the performance of the Adam and Stochastic Gradient Descent (SGD) optimizers on the MobileNetV4 architecture for classifying eight classes of skin diseases using the ISIC 2019 dataset. The dataset consists of 23,257 valid dermoscopic images after preprocessing, which includes duplicate image removal, hair artifact elimination using the blackhat morphology method, and an asymmetric sampling strategy in which majority classes were capped at a maximum of 2,000 images while minority classes were augmented to reach the target count, in order to address extreme class imbalance with a ratio of up to 53:1. The model was trained using a three-phase training strategy with gradual unfreezing of the MobileNetV4 backbone initialized with pretrained ImageNet weights. All training configurations were made identical for both optimizers except for the optimization algorithm and learning rate, ensuring a fair comparison. Evaluation results on the test set show that the Adam optimizer achieved an accuracy of 71.07% with a macro F1-score of 0.72, while SGD achieved an accuracy of 58.06% with a macro F1-score of 0.57. Adam outperformed SGD across all eight skin disease classes. The performance difference of 13.01% indicates that Adam's adaptive learning rate mechanism is more effective for dermoscopic datasets with imbalanced class distributions compared to SGD. Nevertheless, it should be noted that Adam requires greater computational memory than SGD due to the storage of first and second moment estimates per parameter, and therefore the computational efficiency trade-off should be considered when deploying the model on resource-constrained devices. This study provides empirical contribution in selecting the optimal optimizer for skin lesion classification based on lightweight architectures.

Keywords: Adam Optimizer; Dermoscopy; ISIC 2019; MobileNetV4; SGD; Skin Disease Classification; Transfer Learning

1. PENDAHULUAN

Penyakit kulit merupakan salah satu persoalan kesehatan yang paling umum dijumpai lintas usia, dengan implikasi serius terhadap kualitas hidup hingga risiko mortalitas pada kasus ganas. Laporan Global Cancer Observatory tahun 2022 mencatat 331.722 baru melanoma dengan 58.667 kematian di seluruh dunia [1]. Selain melanoma, klasifikasi klinis juga harus berhadapan dengan berbagai variasi lesi seperti Basal Cell Carcinoma, Actinic Keratosis, dan Melanocytic Nevus yang menuntut ketelitian tinggi [2][3]. Proses diagnosis di lapangan masih terhambat oleh subjektivitas pemeriksa, keterbatasan distribusi dokter spesialis, dan tingginya kemiripan visual antar kelas lesi (visual overlap) [4]. Kondisi ini mendorong pengembangan sistem klasifikasi otomatis berbasis Convolutional Neural Network (CNN) yang telah terbukti menjadi pendekatan mutakhir dalam analisis citra medis [5].

Proses penegakan diagnosis penyakit kulit di lapangan secara konsisten masih dihadapkan pada sejumlah tantangan klinis dan operasional. Kendala mendasar yang kerap dijumpai meliputi tingginya unsur subjektivitas pemeriksa, terbatasnya distribusi dokter spesialis di wilayah perifer, serta masalah efisiensi waktu pemeriksaan. Tingkat kesulitan ini semakin tereskalasi oleh faktor visual lesi kulit itu sendiri, yakni spektrum variasi morfologi yang sangat luas pada jenis penyakit yang sama, maupun tingginya kemiripan visual (visual overlap) antar penyakit kulit yang secara patologis tidak berhubungan [4]. Guna menjembatani keterbatasan diagnosis konvensional tersebut, intervensi teknologi komputasi medis menjadi sebuah urgensi untuk memfasilitasi proses diagnostik yang lebih objektif dan meminimalisasi



tingkat kesalahan (diagnostic error). Dalam perkembangannya, prosedur otomatisasi medis ini sangat bergantung pada kapabilitas Machine Learning berbasis Kecerdasan Buatan. Sebagai respons atas kompleksitas citra dermatologis, arsitektur Convolutional Neural Network (CNN) telah terbukti secara nyata dan diakui secara luas sebagai salah satu pendekatan paling mutakhir yang diimplementasikan dalam analisis citra medis [5].

Sejumlah penelitian terdahulu telah berupaya mengotomatisasi klasifikasi citra penyakit kulit menggunakan ragam arsitektur CNN konvensional. Agustina dkk. mengimplementasikan arsitektur VGG-16 dengan optimizer SGD dan berhasil mencapai akurasi sebesar 99,70%, namun penerapannya mengharuskan reduksi resolusi citra secara ekstrem hingga 64×64 piksel demi menyalahi beban komputasi yang tinggi [6]. Di sisi lain, Gusti dkk. menerapkan ResNet-50 dengan optimizer Adam dan memperoleh akurasi 93,88%, tetapi implementasinya menimbulkan ketergantungan pada pemrosesan berbasis server web akibat beban arsitektur yang dalam (deep architecture) [7]. Pola serupa ditemukan pada penelitian-penelitian lain yang menggunakan ResNet-50 maupun VGG-16. Meskipun menghasilkan metrik akurasi yang menjanjikan, penerapannya sering kali terhambat oleh tingginya kebutuhan spesifikasi perangkat keras.

Untuk menutupi celah inefisiensi tersebut, penelitian ini mengusulkan penggunaan arsitektur MobileNetV4 sebagai fondasi model klasifikasi. Dibandingkan pendahulunya MobileNetV3, MobileNetV4 mengintegrasikan blok Universal Inverted Bottleneck (UIB) yang menggantikan blok Inverted Residual konvensional, menghasilkan model berstatus Pareto-optimal dengan jumlah parameter lebih efisien pada tingkat akurasi yang setara atau lebih tinggi [8]. Dibandingkan arsitektur ringan lainnya seperti EfficientNet-B0 dan ShuffleNetV2, MobileNetV4 dirancang secara khusus untuk ekosistem perangkat mobile modern dengan optimasi hardware-aware yang lebih menyeluruh [8]. Keunggulan ini telah dibuktikan pada berbagai domain klasifikasi citra, termasuk deteksi serangan biometrik yang mencapai akurasi 99,69% menggunakan MobileNetV4-Conv-Small [9], serta klasifikasi penyakit daun tanaman dengan MobileNetV4 yang menunjukkan keunggulan signifikan dibandingkan varian MobileNet sebelumnya [10]. Dengan demikian, MobileNetV4 hadir sebagai jawaban atas limitasi arsitektur konvensional yang selama ini mendominasi literatur klasifikasi lesi kulit.

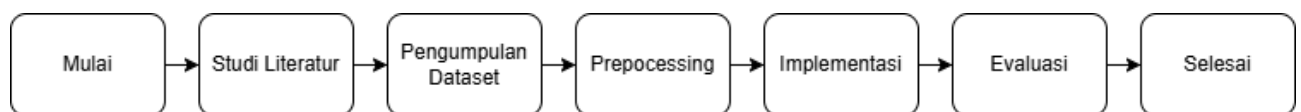
Di samping pemilihan arsitektur, kemampuan generalisasi model Deep Learning sangat dipengaruhi oleh algoritma optimasi (optimizer) yang digunakan. Tinjauan literatur menunjukkan bahwa Adam dan SGD telah banyak dibandingkan pada berbagai domain klasifikasi citra, namun hasil yang diperoleh sangat kontekstual dan bergantung pada karakteristik dataset serta arsitektur yang digunakan [11]. Pada domain klasifikasi lesi kulit dermoskopi, karakteristik spesifik seperti ketidakseimbangan kelas ekstrem (rasio hingga 53:1), variasi morfologi tinggi antar kelas, dan ukuran dataset yang relatif terbatas menciptakan kondisi optimasi yang berbeda secara fundamental dibandingkan dataset klasifikasi umum. Penelitian yang ada baru sebatas membandingkan Adam dan SGD pada domain klasifikasi penyakit tanaman dengan MobileNetV4 [10], sementara kajian serupa pada domain dermoskopi dengan dataset ISIC 2019 yang memiliki karakteristik class imbalance ekstrem belum pernah dilakukan. Oleh karena itu, terdapat kesenjangan penelitian yang perlu diisi: apakah mekanisme adaptive learning rate Adam lebih unggul secara konsisten dibandingkan SGD ketika berhadapan dengan tantangan spesifik dataset dermoskopi multikelas yang tidak seimbang pada arsitektur ringan MobileNetV4?

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan kinerja optimizer Adam dan SGD yang diterapkan pada arsitektur MobileNetV4 dalam mengklasifikasikan citra penyakit kulit menggunakan dataset ISIC 2019 yang mencakup 8 kelas lesi. Evaluasi dilakukan secara komprehensif menggunakan metrik Accuracy, Loss, Precision, Recall, dan F1-Score dengan menjaga seluruh hyperparameter lainnya seperti learning rate, ukuran batch, dan jumlah epoch sebagai variabel konstan, sehingga perbedaan kinerja dapat sepenuhnya diatribusikan pada perbedaan algoritma optimasi. Hasil penelitian ini diharapkan memberikan kontribusi empiris bagi pengembangan sistem deteksi dini penyakit kulit berbasis kecerdasan buatan yang efisien secara komputasi, sekaligus menjadi rujukan ilmiah dalam pemilihan optimizer optimal untuk arsitektur model berbobot ringan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam pelaksanaan penelitian ini, diterapkan tahapan metodologi yang terstruktur untuk meminimalisasi kesalahan dan memastikan proses evaluasi model klasifikasi berjalan dengan sistematis. Secara garis besar, langkah-langkah penelitian yang dilakukan mulai dari studi literatur hingga pengujian model ditunjukkan pada diagram alir di Gambar 1.



Gambar 1. Tahap Penelitian

Tahapan penelitian ini dimulai dengan melakukan pengkajian literatur secara mendalam yang bersumber dari jurnal nasional terakreditasi, jurnal internasional bereputasi, serta buku teks akademis. Fokus kajian literatur mencakup eksplorasi arsitektur berbobot ringan MobileNetV4, karakteristik matematis dari *optimizer Adam* dan *Stochastic Gradient Descent (SGD)*, serta penelitian-penelitian terdahulu yang berkaitan dengan deteksi dan klasifikasi citra penyakit kulit.



2.1.1 Pengumpulan Dataset

Penelitian ini menggunakan dataset ISIC 2019 (*International Skin Imaging Collaboration 2019*) yang merupakan gabungan dari dataset HAM10000 [3] dan BCN20000 [2], dengan total 25.331 citra dermoskopi yang mencakup delapan kelas penyakit kulit. Distribusi jumlah citra per kelas ditunjukkan pada Tabel 1.

Tabel 1. Dataset ISIC 2019

Kelas	Singkatan	Jumlah
Melanoma	MEL	4.522
Melanocytic Nevi	NV	12.875
Basal Cell Carcinoma	BCC	3.323
Actinic Keratos	AK	867
Benign Keratos	BKL	2.624
Dermatofibroma	DF	239
Vascular Lesion	VASC	253
Squamous Cell Carcinoma	SCC	628
Total		25.331

Dataset ini menunjukkan ketidakseimbangan kelas (*class imbalance*) yang signifikan, dengan rasio kelas mayoritas terhadap minoritas mencapai 53:1 (NV:DF). Kondisi ini memerlukan penanganan khusus pada tahap preprocessing dan pelatihan model [3].

2.1.2 Preprocessing

Seluruh citra melalui serangkaian tahap preprocessing sebelum dimasukkan ke dalam model. Tahap pertama adalah penghilangan artefak rambut (*hair removal*) menggunakan metode *blackhat morphology* dengan kernel berukuran 17×17 , diikuti rekonstruksi area menggunakan algoritma Inpainting Telea [12]. Metode *blackhat morphology* bekerja dengan mendeteksi struktur gelap berukuran kecil (rambut) terhadap latar belakang terang, kemudian menghapusnya melalui operasi morfologi *closing* yang diikuti pengurangan citra asli. Rekonstruksi area yang terhapus dilakukan menggunakan algoritma Inpainting Telea yang mengisi piksel kosong berdasarkan informasi piksel tetangga secara iteratif [12]. Perlu dicatat bahwa operasi morfologi dengan kernel besar berpotensi mengaburkan tekstur halus lesi pada area yang berdekatan dengan artefak rambut. Untuk meminimalkan risiko ini, ukuran kernel 17×17 dipilih sebagai keseimbangan antara efektivitas penghapusan rambut dan preservasi detail tekstur lesi — kernel yang lebih kecil tidak efektif menghapus rambut tebal, sementara kernel yang lebih besar berisiko menghilangkan fitur lesi yang krusial untuk klasifikasi.

Tahap kedua adalah pembersihan data, yang mencakup penghapusan citra *downsampled* dan deteksi duplikasi menggunakan hash MD5. Citra *downsampled* dibuang karena merupakan data yang telah diproses sebelumnya untuk kepentingan penelitian tertentu sehingga berpotensi menurunkan kualitas fitur yang dipelajari model [2].

Tahap ketiga adalah penanganan *class imbalance*. Dataset ISIC 2019 memiliki ketidakseimbangan kelas yang ekstrem dengan rasio hingga 53:1 antara kelas NV dan DF. Untuk mengatasinya, diterapkan strategi sampling asimetris: kelas mayoritas (NV, MEL, BCC, BKL) dibatasi maksimal 2.000 citra, sementara kelas minoritas (VASC, DF, AK, SCC) digunakan seluruhnya dan diperbanyak melalui augmentasi hingga mencapai jumlah target. Selain itu, diterapkan *class weight* berbasis *balanced* pada fungsi *loss* agar model memberikan perhatian proporsional terhadap setiap kelas. Distribusi data sebelum dan sesudah preprocessing ditunjukkan pada Tabel 2.

Dataset kemudian dibagi dengan rasio 80:10:10 untuk data latih, validasi, dan uji menggunakan *random seed* 42 guna memastikan *reproducibility*. Augmentasi data diterapkan secara *online* pada data latih menggunakan transformasi *random crop*, *horizontal* dan *vertical flip*, rotasi, *color jitter*, serta *random erasing*, sedangkan data validasi dan uji hanya dinormalisasi menggunakan nilai *mean* dan *standard deviation* ImageNet.

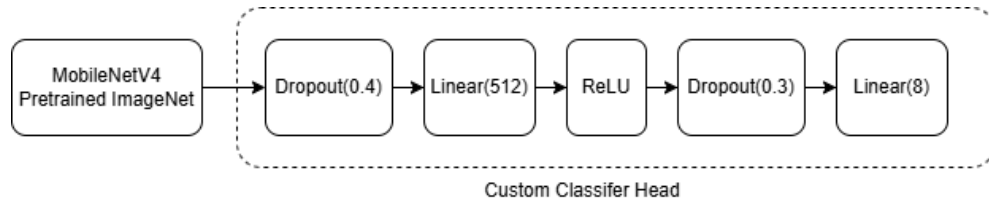
Tabel 2. Hasil Distribusi Data

Kelas	Sebelum	Sesudah (Train)	Sesudah (Val)	Sesudah (Test)
MEL	4.522	1.600	200	200
NV	12.875	1.600	200	200
BCC	3.323	1.600	200	200
AK	867	700	109	109
BKL	2.624	1.600	200	200
DF	239	400	30	24
VASC	253	400	32	26
SCC	628	700	79	63
Total	25.331	8.600	1.050	1.022

2.1.3 Arsitektur Model

Arsitektur yang digunakan sebagai *backbone* dalam penelitian ini adalah MobileNetV4 varian *Conv Small* dengan bobot *pretrained* ImageNet. MobileNetV4 dipilih karena dirancang khusus untuk efisiensi komputasi tanpa mengorbankan

akurasi, sehingga cocok untuk klasifikasi citra medis pada perangkat dengan sumber daya terbatas [8]. Penggunaan bobot *pretrained* ImageNet memungkinkan model memanfaatkan fitur visual umum yang telah dipelajari dari jutaan citra, sehingga mempercepat konvergensi dan meningkatkan performa pada dataset yang relatif kecil [13].



Gambar 2. Visualisasi Arsitektur

Bagian *classifier head* bawaan MobileNetV4 dimodifikasi untuk menyesuaikan jumlah kelas pada penelitian ini. Arsitektur *classifier head* yang digunakan ditunjukkan pada Gambar 2 dan terdiri dari lapisan *Dropout*(0.4), *Linear*($n \rightarrow 512$), *ReLU*, *Dropout*(0.3), dan *Linear*($512 \rightarrow 8$). Penambahan dua lapisan *Dropout* berfungsi sebagai regularisasi untuk mencegah *overfitting*, sedangkan lapisan *Linear* terakhir menghasilkan prediksi probabilitas untuk delapan kelas penyakit kulit. Pelatihan dilakukan menggunakan strategi *gradual unfreezing* dalam tiga fase, sebagaimana ditunjukkan pada Tabel 3. Strategi ini diterapkan untuk melindungi representasi fitur *pretrained* dari pembaruan yang terlalu agresif di awal pelatihan.

Tabel 3. Strategi Pelatihan 3 fase

Fase	Epoch	Backbone	Learning Rate Adam	Learning Rate SGD
1- Head only	10	<i>Freeze</i>	1×10^{-3}	1×10^{-2}
2- Partial Unfreeze	30	Sebagian terbuka	1×10^{-4}	1×10^{-3}
3- Full fine-tune	40	Seluruhnya terbuka	1×10^{-4}	1×10^{-3}

Pada Fase 1, seluruh parameter *backbone* dibekukan dan hanya *classifier head* yang dilatih. Pada Fase 2, tiga *stage* terakhir *backbone* dibuka secara bertahap disertai *linear warmup* selama 5 *epoch* pertama untuk menghindari kerusakan fitur *pretrained*. Pada Fase 3, seluruh parameter *backbone* dibuka untuk *full fine-tuning* dengan *learning rate* yang lebih kecil. Bobot terbaik berdasarkan akurasi validasi tertinggi disimpan di setiap fase menggunakan mekanisme *early stopping* dengan *patience* sebesar 10 *epoch*.

2.1.4 Konfigurasi Training

Proses pelatihan model dilakukan dalam tiga fase secara bertahap (*gradual unfreezing*) untuk melindungi representasi fitur *pretrained* dari pembaruan yang terlalu agresif. Pada Fase 1, seluruh parameter *backbone* dibekukan dan hanya *classifier head* yang dilatih selama 10 *epoch*. Pada Fase 2, tiga *stage* terakhir *backbone* dibuka dengan *linear warmup* selama 5 *epoch* pertama selama 30 *epoch*. Pada Fase 3, seluruh parameter *backbone* dibuka untuk *full fine-tuning* selama 40 *epoch*. Mekanisme *early stopping* dengan *patience* 10 *epoch* diterapkan pada setiap fase untuk menghentikan pelatihan apabila akurasi validasi tidak meningkat, sehingga total *epoch* maksimal adalah 80.

Optimizer merupakan algoritma yang digunakan untuk memperbarui parameter model selama pelatihan dengan tujuan meminimalkan fungsi loss [14]. Penelitian ini membandingkan dua optimizer yang diterapkan secara identik pada seluruh konfigurasi pelatihan kecuali algoritma dan *learning rate*-nya. Optimizer pertama adalah Adam (*Adaptive Moment Estimation*) yang menggunakan estimasi momen pertama dan kedua dari *gradient* untuk menyesuaikan *learning rate* secara adaptif per parameter [15]. Optimizer kedua adalah SGD (*Stochastic Gradient Descent*) dengan momentum yang memperbarui parameter menggunakan *gradient* seragam dengan tambahan momentum untuk menstabilkan arah pembaruan bobot [9]. Perbedaan *learning rate* antara Adam dan SGD merupakan penyesuaian terhadap karakteristik inheren masing-masing algoritma, di mana SGD membutuhkan *learning rate* yang lebih besar untuk mencapai konvergensi yang setara dengan Adam [11]. Seluruh konfigurasi *hyperparameter* disajikan pada Tabel 4.

Tabel 4. Konfigurasi *hyperparameter* pelatihan

Parameter	Adam	SGD	Keterangan
β_1, β_2	0.9, 0.999	–	Parameter Adam
Momentum	–	0.9	Parameter SGD
LR Fase 1	1×10^{-3}	1×10^{-2}	Rasio 1:10
LR Fase 2	1×10^{-4}	1×10^{-3}	Rasio 1:10
LR Fase 3	1×10^{-4}	1×10^{-3}	Rasio 1:10
<i>Weight decay</i>	5×10^{-4}	1×10^{-4}	Regularisasi
<i>Loss function</i>	CrossEntropyLoss + class weight + label smoothing 0.1		Identik
<i>Batch size</i>	64		Identik
<i>Scheduler</i>	CosineAnnealingLR		Identik
<i>Warmup</i>	Linear, 5 <i>epoch</i> (Fase 2 & 3)		Identik



Parameter	Adam	SGD	Keterangan
<i>Gradient clipping</i>	max_norm = 1.0		Identik
<i>Early stopping</i>	patience = 10 epoch		Identik
<i>Random seed</i>	42		<i>Reproducibility</i>

Fungsi *loss* yang digunakan adalah *CrossEntropyLoss* dengan *class weighting* berbasis *balanced* untuk menangani ketidakseimbangan kelas, serta *label smoothing* sebesar 0.1 untuk mencegah model terlalu yakin terhadap prediksi pada kelas-kelas yang memiliki kemiripan visual [16]. *Scheduler* *CosineAnnealingLR* diterapkan dalam setiap fase untuk membantu model keluar dari *local minima* secara lebih efektif [17], disertai *gradient clipping* dengan nilai *max_norm* = 1.0 yang diterapkan identik pada kedua optimizer untuk menjaga stabilitas pelatihan.

2.1.5 Metrik Evaluasi

Evaluasi performa kedua model dilakukan menggunakan empat metrik kuantitatif yang umum digunakan dalam klasifikasi citra medis [18], yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Mengingat distribusi kelas pada dataset ISIC 2019 yang tidak seimbang, perhitungan metrik dilakukan menggunakan pendekatan *macro average* agar setiap kelas mendapat bobot yang sama tanpa dipengaruhi oleh jumlah sampel masing-masing kelas [19]. Selain metrik kuantitatif, *Confusion Matrix* digunakan sebagai alat analisis visual untuk mengidentifikasi pola kesalahan prediksi antar kelas [18]. *Accuracy* mengukur proporsi prediksi yang benar dari seluruh data uji, sebagaimana ditunjukkan pada persamaan (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision mengukur proporsi prediksi positif yang benar-benar positif, sebagaimana ditunjukkan pada persamaan (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall mengukur proporsi data positif aktual yang berhasil diprediksi dengan benar, sebagaimana ditunjukkan pada persamaan (3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score merupakan rata-rata harmonik dari *Precision* dan *Recall* yang memberikan keseimbangan antara keduanya, sebagaimana ditunjukkan pada persamaan (4). Metrik ini sangat relevan untuk dataset tidak seimbang karena sensitif terhadap performa pada kelas minoritas [19].

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Keterangan: *TP* (*True Positive*) adalah jumlah prediksi positif yang benar, *TN* (*True Negative*) adalah jumlah prediksi negatif yang benar, *FP* (*False Positive*) adalah jumlah prediksi positif yang salah, dan *FN* (*False Negative*) adalah jumlah prediksi negatif yang salah.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pelatihan Model

Proses pelatihan kedua model dilakukan selama maksimal 80 epoch dengan strategi tiga fase menggunakan dataset ISIC 2019 yang telah melalui tahap preprocessing. Model Adam mencapai akurasi validasi terbaik sebesar 70,94% pada epoch ke-74, sedangkan model SGD mencapai akurasi validasi terbaik sebesar 57,82% pada epoch ke-76. Ringkasan hasil pelatihan kedua model disajikan pada Tabel 5.

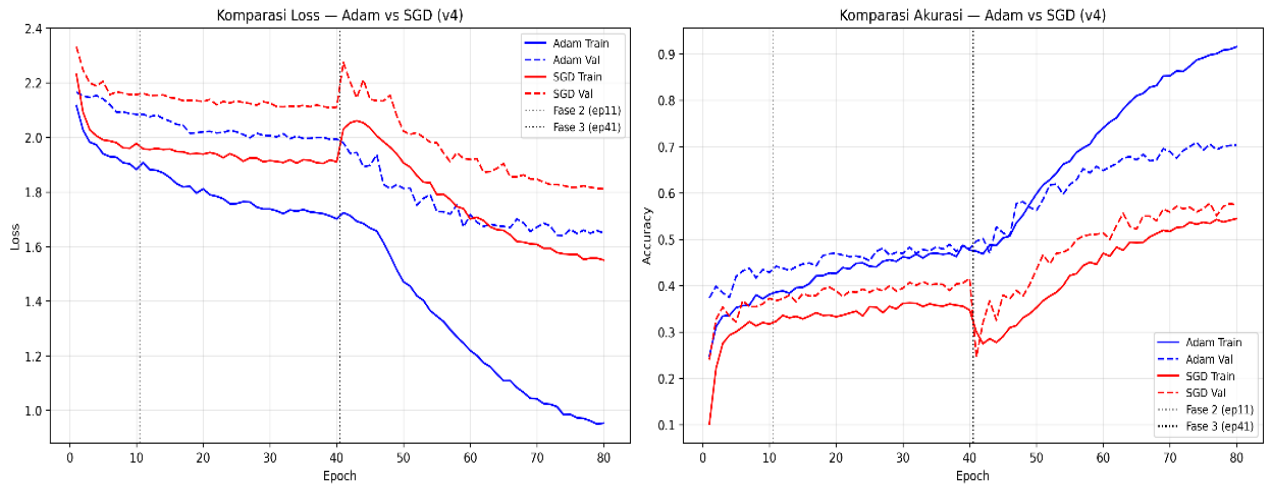
Tabel 5. Ringkasan hasil penelitian kedua *Optimizer*

Metrik	Adam	SGD
<i>Val Accuracy</i> Terbaik	70,94%	57,82%
<i>Epoch</i> Terbaik	74	76
<i>Train Loss</i> Akhir	0,9533	1,5507
<i>Val Loss</i> Akhir	1,6495	1,8113
<i>Gap Train-Val Loss</i>	0,6962	0,2606

Grafik perbandingan *loss* dan akurasi selama pelatihan disajikan pada Gambar 3. Adam menunjukkan penurunan *loss* yang konsisten dan lebih cepat dibandingkan SGD pada ketiga fase pelatihan, khususnya pada Fase 3 (*full fine-tuning*, epoch 41–80) di mana *loss* Adam turun secara signifikan dari 1,72 menuju 0,95. SGD menunjukkan penurunan *loss* yang lebih lambat dengan nilai *loss* akhir yang lebih tinggi (1,5507), mengindikasikan konvergensi yang belum optimal pada konfigurasi yang digunakan.



Perbandingan Adam vs SGD – MobileNetV4
(Fase: 10+30+40 epoch | Sampling asimetris | Filter duplikasi)



Gambar 3. Grafik perbandingan loss dan akurasi pelatihan

3.2 Analisis Kurva Training

Berdasarkan Gambar 3, kurva *train accuracy* Adam terus meningkat hingga 92%, sementara kurva *validation accuracy* stabil di sekitar 70%. Gap antara keduanya sebesar 0,6962 mengindikasikan adanya *mild overfitting* pada model Adam. Penerapan *dropout* (0,4 dan 0,3) pada *classifier head* serta *label smoothing* 0,1 telah membantu menekan *overfitting*, namun gap tersebut masih terjadi. Untuk penelitian selanjutnya, beberapa pendekatan teknis yang lebih tepat untuk mengatasi masalah generalisasi ini antara lain: (1) penerapan teknik augmentasi yang lebih agresif seperti *CutMix* atau *MixUp* untuk memperkaya variasi data latih, (2) peningkatan nilai *weight decay* atau penambahan regularisasi L1/L2 pada lapisan *backbone*, (3) penggunaan *stochastic depth* (DropPath) pada blok UIB MobileNetV4 sebagai regularisasi tambahan, dan (4) eksplorasi varian MobileNetV4 yang lebih besar seperti *Conv-Medium* untuk meningkatkan kapasitas representasi tanpa mengorbankan efisiensi secara berlebihan. Perlu ditekankan bahwa mengurangi jumlah epoch tidak menyelesaikan masalah generalisasi secara fundamental, melainkan hanya menghentikan pelatihan lebih awal sebelum *overfitting* memburuk.

Sebaliknya, gap *train-val* SGD yang lebih kecil (0,2606) menunjukkan generalisasi yang lebih seimbang, namun pada tingkat akurasi yang lebih rendah. Kondisi ini mengindikasikan SGD mengalami *underfitting* relatif, model tidak sepenuhnya mengekstraksi informasi diskriminatif dari data latih. Hal ini konsisten dengan karakteristik SGD yang membutuhkan lebih banyak iterasi dan penyeteraan *learning rate* yang lebih teliti untuk mencapai konvergensi optimal dibandingkan Adam [11].

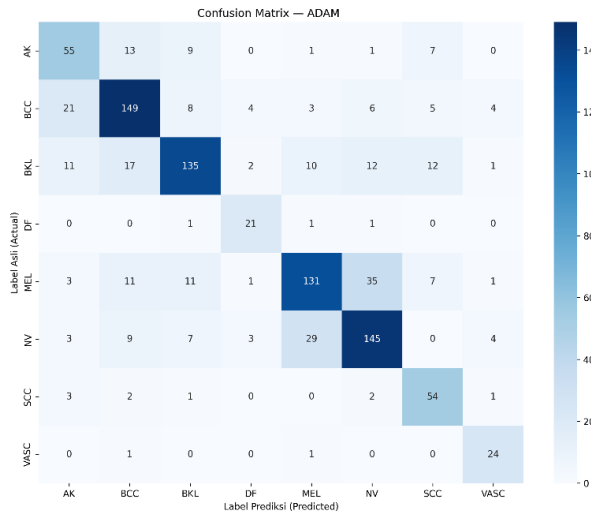
Pada Fase 3, strategi *gradual unfreezing* terbukti memberikan dampak positif pada kedua model — terlihat dari penurunan *loss* yang lebih tajam setelah epoch ke-41 pada kedua kurva. Temuan ini konsisten dengan hasil penelitian yang menunjukkan bahwa *fine-tuning* bertahap pada model *pretrained* menghasilkan adaptasi fitur yang lebih baik untuk domain yang berbeda dari data pre-training [13].

3.3 Hasil Evaluasi pada Data Uji

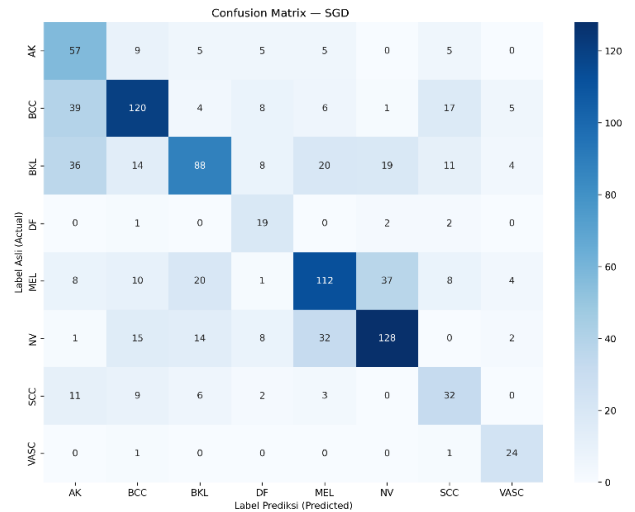
Evaluasi akhir dilakukan pada data uji yang sepenuhnya terpisah dari data latih dan validasi, dengan total 1.022 citra uji dari delapan kelas. Hasil *classification report* lengkap disajikan pada Tabel 6, sedangkan *confusion matrix* masing-masing model disajikan pada Gambar 4 (Adam) dan Gambar 5 (SGD).

Tabel 6. Classification Report Adam dan SGD pada Data Uji

Kelas	Precision Adam	Recall Adam	F1 Adam	Precision SGD	Recall SGD	F1 SGD	Support
AK	0,57	0,64	0,60	0,38	0,66	0,48	109
BCC	0,74	0,74	0,74	0,67	0,60	0,63	200
BKL	0,78	0,68	0,73	0,64	0,44	0,52	200
DF	0,68	0,88	0,76	0,37	0,79	0,51	24
MEL	0,74	0,66	0,70	0,63	0,56	0,59	200
NV	0,72	0,72	0,72	0,68	0,64	0,66	200
SCC	0,64	0,86	0,73	0,42	0,51	0,46	63
VASC	0,69	0,92	0,79	0,62	0,92	0,74	26
<i>Accuracy</i>			0,71			0,58	1.022
<i>Macro</i>	0,69	0,76	0,72	0,55	0,64	0,57	1.022



Gambar 4. Confusion Matrix Adam



Gambar 5. Confusion Matrix SGD

Perhitungan metrik evaluasi menggunakan pendekatan *macro average* agar setiap kelas mendapat bobot yang sama tanpa dipengaruhi jumlah sampel masing-masing kelas, mengingat distribusi data uji yang tidak seimbang.

3.4 Perbandingan Performa Adam dan SGD

Optimizer Adam menghasilkan akurasi pengujian sebesar 71,07% dengan *macro F1-score* 0,72, sedangkan SGD menghasilkan akurasi 58,06% dengan *macro F1-score* 0,57. Perbedaan akurasi sebesar 13,01% ini konsisten dengan karakteristik inheren kedua algoritma optimasi. Adam menggunakan estimasi momen pertama dan kedua dari *gradient* untuk menyesuaikan *learning rate* secara adaptif per parameter [15], sehingga lebih efektif dalam menangani distribusi *gradient* yang beragam pada dataset dengan ketidakseimbangan kelas tinggi seperti ISIC 2019. SGD dengan *momentum* seragam memerlukan konfigurasi *learning rate* yang lebih presisi untuk mencapai konvergensi yang setara [20], sehingga pada konfigurasi penelitian ini Adam terbukti lebih unggul untuk kasus klasifikasi lesi kulit dermoskopi delapan kelas. Meskipun Adam terbukti lebih unggul, perlu dicatat bahwa akurasi 71,07% pada dataset dermoskopi delapan kelas masih berada di bawah ambang yang umumnya diharapkan untuk aplikasi klinis. Sebagai perbandingan, penelitian dengan arsitektur yang lebih dalam seperti ResNet-50 mencapai akurasi 93,88% [7], meskipun dengan beban komputasi yang jauh lebih tinggi dan dataset yang berbeda. Rendahnya akurasi pada penelitian ini kemungkinan disebabkan oleh beberapa faktor: (1) keterbatasan kapasitas representasi MobileNetV4-Conv-Small sebagai arsitektur ringan yang secara inheren memiliki lebih sedikit parameter dibandingkan arsitektur dalam, (2) kompleksitas tugas klasifikasi delapan kelas dengan class imbalance ekstrem, dan (3) adanya mild overfitting pada model Adam yang membatasi kemampuan generalisasi. Temuan ini mengindikasikan bahwa trade-off antara efisiensi komputasi dan akurasi perlu dipertimbangkan secara eksplisit dalam pemilihan arsitektur untuk aplikasi medis.

3.5 Perbandingan Penelitian Terdahulu

Hasil penelitian ini perlu dikontekstualisasikan terhadap penelitian-penelitian terdahulu pada domain yang serupa untuk mengukur posisi kontribusinya secara objektif. Tabel 7 merangkum perbandingan hasil penelitian ini dengan beberapa penelitian sejenis.

Tabel 7. Perbandingan Penelitian Terdahulu

Penelitian	Arsitektur	Dataset	Kelas	Akurasi Terbaik	Optimizer Terbaik
Saputra & Al Rivan [21]	ResNet-152	ISIC 2019	2 (biner)	87,85%	MBGD
Anugrah & Rachmat [10]	MobileNetV4-Conv-Medium	Daun Jagung	4	95,25%	SGD
Prabhat & Mahato [11]	CNN Umum	CIFAR/ImageNet	Umum	71,59%	Adam+LAWN
Penelitian diusulkan	MobileNetV4-Conv-Small	ISIC 2019	8	71,07%	Adam



Saputra dan Al Rivian [21] membandingkan arsitektur ResNet-152 dan AlexNet pada dataset ISIC 2019 yang sama untuk klasifikasi delapan kelas kanker kulit yang dikelompokkan menjadi dua kategori benign dan malignant. Penelitian tersebut memperoleh akurasi terbaik sebesar 87,85% menggunakan ResNet-152 dengan optimizer MBGD. Meskipun akurasi tersebut lebih tinggi dari penelitian ini (71,07%), perbedaan mendasar terletak pada kompleksitas tugas klasifikasi: penelitian tersebut menyederhanakan masalah menjadi klasifikasi biner (benign vs malignant), sementara penelitian ini mempertahankan klasifikasi delapan kelas secara penuh yang secara inheren jauh lebih sulit akibat tingginya visual overlap antar kelas. Selain itu, ResNet-152 memiliki kapasitas representasi yang jauh lebih besar dibandingkan MobileNetV4-Conv-Small yang digunakan penelitian ini, dengan konsekuensi beban komputasi yang lebih tinggi dan tidak ideal untuk deployment pada perangkat bergerak.

Anugrah dan Rachmat [10] mengevaluasi tiga varian MobileNetV4 (Conv-Small, Conv-Medium, Conv-Large) menggunakan optimizer Adam dan SGD untuk klasifikasi penyakit daun jagung empat kelas. Penelitian tersebut memperoleh akurasi terbaik 95,25% menggunakan MobileNetV4-Conv-Medium dengan optimizer SGD, dan secara konsisten SGD mengungguli Adam pada seluruh varian arsitektur. Temuan ini menarik untuk dikontraskan dengan hasil penelitian ini, di mana Adam justru mengungguli SGD secara signifikan (71,07% vs 58,06%). Perbedaan hasil ini mengkonfirmasi bahwa performa relatif Adam terhadap SGD sangat bergantung pada karakteristik spesifik dataset — dataset daun jagung dengan empat kelas seimbang (1.000 citra per kelas, tanpa class imbalance) lebih cocok untuk SGD, sementara dataset dermoskopi ISIC 2019 dengan ketidakseimbangan kelas ekstrem (rasio 53:1) dan delapan kelas terbukti lebih efektif ditangani oleh mekanisme adaptive learning rate milik Adam.

Pada aspek perbandingan optimizer, Prabhat dan Mahato [11] juga menunjukkan bahwa performa Adam dapat ditingkatkan mendekati SGD melalui teknik layer-wise weight normalization (LAWN), mengindikasikan bahwa gap performa Adam dan SGD bersifat kontekstual dan dapat direduksi melalui strategi pelatihan yang tepat. Secara keseluruhan, penelitian ini berkontribusi mengisi celah literatur yang spesifik: perbandingan empiris Adam dan SGD pada arsitektur MobileNetV4-Conv-Small dengan kondisi class imbalance ekstrem (53:1) pada dataset dermoskopi delapan kelas ISIC 2019, yang belum pernah dikaji secara eksplisit pada penelitian-penelitian sebelumnya.

3.6 Analisis Per Kelas

3.6.1 Kelas dengan Performa Terbaik

Kelas VASC mencatat *F1-score* tertinggi untuk Adam (0,79) dengan *recall* 92%. Hasil serupa diperoleh SGD dengan *recall* 92% dan *F1-score* 0,74. Tingginya *recall* kedua model pada kelas ini menunjukkan bahwa lesi vaskular memiliki karakteristik visual yang cukup distingtif sehingga mudah dikenali oleh model, meskipun jumlah sampel ujinya paling sedikit (26 sampel). Ini merupakan dampak positif dari penerapan strategi *class weighting* dan sampling asimetris pada tahap preprocessing.

3.6.2 Kelas dengan Performa Terendah

Kelas AK mencatat *F1-score* terendah untuk Adam (0,60) dan SGD (0,48). Rendahnya performa pada kelas ini dapat dikaitkan dengan jumlah sampel uji yang lebih sedikit (86 sampel) dibandingkan kelas mayoritas, serta kemiripan morfologi aktinik keratosis dengan kelas lainnya seperti BCC dan SCC yang sama-sama merupakan lesi epitelial [3].

3.7 Analisis Pola Kesalahan Klasifikasi

3.7.1 Kebingungan MEL dan NV

Analisis *confusion matrix* pada Gambar 4 menunjukkan pola kesalahan paling dominan pada Adam adalah kebingungan antara kelas MEL dan NV, di mana 35 dari 200 citra MEL diprediksi sebagai NV. Pola serupa ditemukan pada SGD dengan 37 kesalahan pada kombinasi yang sama. Fenomena ini bukan semata kegagalan model, melainkan mencerminkan tantangan diagnostik klinis yang diakui secara luas dalam literatur dermoskopi melanoma dan melanocytic nevi memiliki kemiripan visual yang tinggi, terutama pada tahap awal perkembangan lesi [3].

3.7.2 Kebingungan BCC dan AK

Pada *confusion matrix* SGD Gambar 5, kesalahan terbesar terjadi pada kelas BCC di mana 39 dari 200 citra BCC diprediksi sebagai AK. Adam hanya menghasilkan 21 kesalahan untuk pola yang sama, menunjukkan kemampuan diskriminasi yang lebih baik. Kemiripan morfologi antara karsinoma sel basal dan aktinik keratosis pada stadium awal memang menjadi tantangan diagnosis dermoskopi yang sulit bahkan bagi klinisi berpengalaman [2].

3.8 Implikasi Hasil Penelitian

Hasil penelitian ini memberikan bukti empiris bahwa pada arsitektur MobileNetV4 dengan dataset dermoskopi delapan kelas yang tidak seimbang, optimizer Adam menghasilkan performa klasifikasi yang lebih tinggi dibandingkan SGD. Temuan ini konsisten dengan penelitian terdahulu yang menunjukkan keunggulan Adam pada tugas klasifikasi citra medis [5].

Dengan akurasi 71,07% dan arsitektur yang ringan, MobileNetV4 dengan optimizer Adam berpotensi diimplementasikan sebagai sistem pendukung keputusan klinis berbasis perangkat bergerak untuk deteksi dini lesi kulit.



Hal ini sejalan dengan tujuan utama pengembangan arsitektur MobileNetV4 yang dirancang untuk mencapai keseimbangan optimal antara akurasi dan efisiensi komputasi pada perangkat mobile [8].

4. KESIMPULAN

Penelitian ini telah berhasil membandingkan performa optimizer Adam dan SGD pada arsitektur MobileNetV4 untuk klasifikasi delapan kelas penyakit kulit menggunakan dataset ISIC 2019. Hasil eksperimen menunjukkan bahwa Adam menghasilkan akurasi 71,07% dengan macro F1-score 0,72, mengungguli SGD yang hanya mencapai akurasi 58,06% dengan macro F1-score 0,57 perbedaan 13,01% yang mengkonfirmasi bahwa mekanisme adaptive learning rate Adam lebih efektif dalam menangani dataset dermoskopi dengan ketidakseimbangan kelas ekstrem (rasio 53:1) dibandingkan SGD yang menggunakan pembaruan gradient seragam. Strategi pelatihan tiga fase dengan gradual unfreezing dan class weighting terbukti efektif, tercermin dari kemampuan model mengenali kelas minoritas seperti DF (recall 88%) dan VASC (recall 92%). Namun demikian, akurasi 71,07% masih berada di bawah ambang ideal untuk aplikasi klinis, dan gap train-val loss Adam sebesar 0,6962 mengindikasikan adanya mild overfitting yang belum sepenuhnya teratasi dengan dropout dan label smoothing yang diterapkan. Pola kesalahan dominan berupa kebingungan antara kelas MEL dan NV mencerminkan tantangan visual overlap yang inheren pada domain dermoskopi dan menjadi batasan utama model saat ini. Berdasarkan temuan spesifik ini, penelitian selanjutnya disarankan untuk mengeksplorasi teknik augmentasi CutMix dan MixUp untuk memperkaya variasi data latih guna mengatasi overfitting, penerapan focal loss sebagai pengganti class weighting untuk penanganan class imbalance yang lebih adaptif, serta pengujian varian MobileNetV4 yang lebih besar seperti Conv-Medium untuk meningkatkan kapasitas representasi pada kelas-kelas dengan visual overlap tinggi.

REFERENCES

- [1] J. Ferlay *dkk.*, “Global Cancer Observatory: Cancer Today,” Lyon, France, 2024. Diakses: 1 April 2026. [Daring]. Tersedia pada: <https://gco.iarc.who.int/today>
- [2] C. Hernández-Pérez *dkk.*, “BCN20000: Dermoscopic Lesions in the Wild,” *Sci. Data*, vol. 11, no. 1, Des 2024, doi: 10.1038/s41597-024-03387-w.
- [3] P. Tschandl, C. Rosendahl, dan H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Sci. Data*, vol. 5, Agu 2018, doi: 10.1038/sdata.2018.161.
- [4] A. N. Pratama, “Sistem Klasifikasi Penyakit Kulit pada Manusia Convolutional Neural Network (CNN) EfficientNet B2,” *Paradigma: Jurnal Filsafat, Sains, Teknologi, dan Sosial Budaya*, vol. 30, hlm. 30–37, 2024, doi: <https://doi.org/10.33503/paradigma.v30i2.439>.
- [5] R. Agarwal dan D. Godavarthi, “Skin Disease Classification Using CNN Algorithms,” *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 9, no. 1, Mei 2023, doi: 10.4108/eetpht.9.4039.
- [6] R. Agustina, R. Magdalena, dan N. K. C. Pratiwi, “Klasifikasi Kanker Kulit menggunakan Metode Convolutional Neural Network dengan Arsitektur VGG-16,” *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 10, no. 2, hlm. 446, Apr 2022, doi: 10.26760/elkomika.v10i2.446.
- [7] G. P. H. P. Gusti, E. Haerani, F. Syafria, F. Yanto, dan S. K. Gusti, “Implementasi Algoritma Convolutional Neural Network (Resnet-50) untuk Klasifikasi Kanker Kulit Benign dan Malignant,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, hlm. 984–992, Jun 2024, doi: 10.57152/malcom.v4i3.1398.
- [8] D. Qin *dkk.*, “MobileNetV4 : Universal Models for the Mobile Ecosystem,” dalam *Computer Vision – ECCV 2024*, Spring Nature Link, Nov 2024. doi: 10.1007/978-3-031-73661-2_5.
- [9] R. Suhartoyo, “Implementation of MobileNetV4 and Efficient Channel Attention in Anti-Spoofing Face Attack Detection,” *INOVTEK Polbeng-Seri Informatika*, vol. 11, no. 1, 2026, doi: 10.35314/kth2nc32.
- [10] M. N. Anugrah dan N. Rachmat, “Klasifikasi Penyakit Daun Tanaman Jagung Menggunakan Pendekatan Transfer Learning Arsitektur MobileNetV4,” *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT)*, vol. 4, no. 4, hlm. 2174–2182, 2025, doi: 10.55826/jtmit.v4i4.1393.
- [11] G. Prabhat dan A. Kumar Mahato, “Adam Vs. SGD: Achieving Comparable Generalization In Image Classification Through Adaptive Techniques,” *International Journal of Current Science (IJCS PUB)*, vol. 15, no. 2, hlm. 273–279, Apr 2025, Diakses: 15 Juni 2026. [Daring]. Tersedia pada: <https://www.rjpn.org/ijcspub/papers/IJCS25B1033.pdf>
- [12] A. Telea, “An Image Inpainting Technique Based on the Fast Marching Method,” *Journal of Graphics Tools*, vol. 9, no. 1, hlm. 25–36, 2004, doi: 10.1080/10867651.2004.10487596.
- [13] E. S. Nugroho, I. Ardiyanto, dan H. A. Nugroho, “Boosting the performance of pretrained CNN architecture on dermoscopic pigmented skin lesion classification,” *Skin Research and Technology*, vol. 29, no. 11, Nov 2023, doi: 10.1111/srt.13505.
- [14] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA: O’Reilly Media, Inc., 2022.
- [15] D. P. Kingma dan J. Ba, “Adam: A Method for Stochastic Optimization,” Jan 2014, [Daring]. Tersedia pada: <http://arxiv.org/abs/1412.6980>
- [16] A. T. Ibrahim, M. Abdullahi, A. F. D. Kana, M. T. Mohammed, dan I. H. Hassan, “Categorical classification of skin cancer using a weighted ensemble of transfer learning with test time augmentation,” *Data Science and Management*, vol. 8, no. 2, hlm. 174–184, Jun 2025, doi: 10.1016/j.dsm.2024.10.002.
- [17] B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, dan K. Lakshmana, “Skin lesion classification of dermoscopic images using machine learning and convolutional neural network,” *Sci. Rep.*, vol. 12, no. 1, Des 2022, doi: 10.1038/s41598-022-22644-9.
- [18] S. Sathyanarayanan, “Confusion Matrix-Based Performance Evaluation Metrics,” *African Journal of Biomedical Research*, hlm. 4023–4031, Nov 2024, doi: 10.53555/ajbr.v27i4s.4345.



- [19] M. K. Hasan, M. A. Ahamad, C. H. Yap, dan G. Yang, “A survey, review, and future trends of skin lesion segmentation and classification,” *Elsevier Ltd.* Maret 2023, doi: 10.1016/j.combiomed.2023.106624.
- [20] I. Goodfellow, Y. Bengio, dan A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016. Diakses: 9 April 2026. [Daring]. Tersedia pada: <https://www.deeplearningbook.org>
- [21] T. Saputra dan M. E. Al-Rivan, “Analisis Performa ResNet-152 dan AlexNet dalam Klasifikasi Jenis Kanker Kulit,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol 8, no. 1, 2023, doi: <http://dx.doi.org/10.30998/string.v8i1.16464>.