

Analisis Performa Per-Kelas Klasifikasi Tipe Kulit Wajah Berbasis ResNet50

^{1*}Alfin Yusuf Isnawan Prayoga, ²Danar Putra Pamungkas, ³Made Ayu Dusea Widyadara

^{1,2,3}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

Jl. KH. Achmad Dahlan No. 76, Kelurahan Tamanan, Kecamatan Mojoroto, Kota Kediri, Jawa Timur 64112, Indonesia

E-mail: ¹yprayoga938@gmail.com, ²danar@unpkediri.ac.id, ³madedara@unpkediri.ac.id

Penulis Korespondens : Alfin Yusuf Isnawan Prayoga

Abstrak—Identifikasi tipe kulit wajah secara akurat merupakan kunci efektivitas penggunaan produk skincare. Penelitian ini menganalisis performa per-kelas model ResNet50 pada klasifikasi empat tipe kulit wajah, yaitu acne, dry, normal, dan oily, menggunakan dataset gabungan sebanyak 4.985 citra dari dua sumber Kaggle. Model dilatih menggunakan pendekatan transfer learning dengan bobot pretrained IMAGENET1K_V2, optimizer SGD, dan teknik augmentasi berupa Random Horizontal Flip dan Color Jitter. Hasil pengujian menunjukkan bahwa model mencapai akurasi keseluruhan sebesar 92%, namun terdapat ketimpangan performa antar kelas yang signifikan: kelas acne memperoleh recall 100%, sedangkan kelas dry hanya mencapai 84%, dengan gap sebesar 16 poin persentase. Analisis menunjukkan bahwa ketimpangan tersebut dipengaruhi oleh dua faktor, yaitu distribusi data yang tidak seimbang dan kemiripan karakteristik visual antar tipe kulit. Temuan ini mengindikasikan bahwa akurasi global tidak cukup sebagai satu-satunya indikator evaluasi model klasifikasi kulit wajah.

Kata Kunci— class imbalance, klasifikasi kulit wajah, performa per-kelas, ResNet50, transfer learning

Abstract— Accurate facial skin type identification is essential for effective skincare product selection. This study analyzes the per-class performance of ResNet50 in classifying four facial skin types: acne, dry, normal, and oily, using a combined dataset of 4,985 images from two Kaggle sources. The model was trained using a transfer learning approach with IMAGENET1K_V2 pretrained weights, SGD optimizer, and data augmentation techniques including Random Horizontal Flip and Color Jitter. Experimental results show that the model achieves an overall accuracy of 92%, yet significant per-class performance disparities exist: the acne class achieves 100% recall, while the dry class reaches only 84%, resulting in a 16 percentage-point gap. Analysis reveals that this disparity is driven by two factors: imbalanced data distribution and inherent visual similarity between skin type classes. These findings indicate that global accuracy alone is insufficient as the sole evaluation metric for facial skin classification models.

Keywords— class imbalance, facial skin classification, per-class performance, ResNet50, transfer learning

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Industri perawatan kulit (skincare) terus mengalami perkembangan seiring meningkatnya perhatian masyarakat terhadap kesehatan dan penampilan kulit. Berbagai produk perawatan kulit tersedia untuk mengatasi beragam kondisi kulit, seperti kulit berminyak, kering, normal, maupun berjerawat. Namun, efektivitas penggunaan produk

skincare sangat bergantung pada kesesuaian antara produk yang digunakan dengan tipe kulit pengguna [1]. Pada praktiknya, identifikasi tipe kulit masih sering dilakukan secara manual berdasarkan pengamatan visual atau pengalaman pribadi, sehingga tidak jarang terjadi kesalahan dalam menentukan tipe kulit [2]. Kondisi ini dapat dialami oleh berbagai kalangan, baik pria maupun wanita, terutama karena karakteristik antar tipe kulit sering kali memiliki kemiripan yang sulit dibedakan tanpa pengetahuan yang memadai.

Ketidaktahuan tipe kulit berdampak langsung pada pemilihan produk skincare yang tidak tepat, yang berpotensi memperburuk kondisi kulit pengguna [3]. Teknologi kecerdasan buatan berbasis pengolahan citra wajah menawarkan solusi menjanjikan: sistem berbasis deep learning dapat menganalisis foto wajah dan mengklasifikasikan tipe kulit hanya dalam hitungan detik melalui kamera smartphone, tanpa biaya konsultasi dermatologi. Convolutional Neural Network (CNN), khususnya arsitektur ResNet50, menjadi pilihan utama berbagai penelitian di bidang ini karena terbukti efektif dalam berbagai tugas klasifikasi citra [4].

hampir seluruh penelitian hanya melaporkan akurasi global sebagai tolak ukur utama, tanpa menganalisis disparitas performa antar kelas secara mendalam. Padahal, akurasi global yang tinggi tidak menjamin setiap kelas terdeteksi dengan keandalan setara [5], [6]. Dalam konteks klasifikasi kulit, hal ini sangat problematik karena misklasifikasi pada kelas tertentu memiliki konsekuensi berbeda terhadap kondisi kulit pengguna. Kulit kering yang salah teridentifikasi sebagai normal tidak akan mendapat hidrasi cukup, sementara kulit berjerawat yang salah teridentifikasi tidak akan mendapat bahan aktif anti-acne yang dibutuhkan.

Penelitian ini hadir untuk mengisi gap tersebut dengan melakukan analisis mendalam terhadap ketimpangan performa per-kelas pada model ResNet50 untuk klasifikasi empat tipe kulit wajah (acne, dry, normal, oily). Kontribusi penelitian ini meliputi: (1) analisis kuantitatif ketimpangan performa per-kelas yang belum dilakukan secara mendalam pada penelitian sebelumnya, (2) identifikasi dua faktor penyebab ketimpangan beserta bukti empiris dari penelitian pembandingan, dan (3) rekomendasi praktis berbasis bukti agar standar evaluasi model AI skincare tidak hanya berhenti pada akurasi global.

II TINJAUAN PUSTAKA

Bab ini memaparkan penelitian terdahulu serta landasan teori yang mendukung penelitian. Pembahasan meliputi penelitian terkait klasifikasi tipe kulit wajah, konsep arsitektur ResNet50 dan transfer learning, serta permasalahan class imbalance yang sering ditemui pada tugas klasifikasi citra.

A. Penelitian Terkait Klasifikasi Tipe Kulit Wajah

Sylvagno dan Rochadiani (2025) menggunakan ResNet50 untuk klasifikasi empat tipe kulit wajah dan memperoleh akurasi $90,94\% \pm 0,60\%$. Penelitian tersebut melaporkan adanya class imbalance dan menemukan bahwa proses balancing data justru menurunkan akurasi model [5]. Novita et al. (2025) mengembangkan sistem berbasis InceptionV3 dengan akurasi 91,7% dan menemukan perbedaan confidence score yang cukup besar antar kelas [7]. Herimanto et al. (2024) menunjukkan bahwa distribusi data yang tidak seimbang menyebabkan kelas minoritas memperoleh performa terendah pada model

MobileNetV3 [6]. Sementara itu, Putri et al. (2025) menggunakan Vision Transformer (ViT) pada dataset yang telah diseimbangkan, namun masih menemukan misklasifikasi yang tinggi pada kelas tertentu akibat kemiripan karakteristik visual antar tipe kulit [8]. Berdasarkan penelitian-penelitian tersebut, klasifikasi tipe kulit wajah masih menghadapi dua tantangan utama, yaitu ketidakseimbangan distribusi data dan kemiripan karakteristik visual antar kelas. Oleh karena itu, penelitian ini berfokus pada analisis performa per-kelas menggunakan ResNet50 untuk mengevaluasi kemampuan model dalam mengenali setiap tipe kulit wajah secara lebih mendalam.

B. Arsitektur ResNet50 dan Transfer Learning

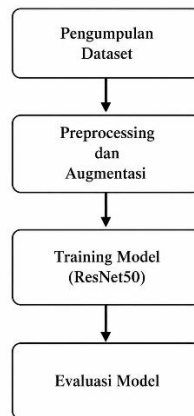
ResNet50 merupakan arsitektur deep Convolutional Neural Network (CNN) yang diperkenalkan oleh He et al. (2016) melalui konsep residual learning dengan penggunaan skip connection untuk mengatasi permasalahan vanishing gradient pada jaringan yang dalam [9]. Model ini tersedia dalam bentuk pretrained pada dataset ImageNet yang terdiri dari lebih dari satu juta citra dan 1.000 kelas objek. Pada penelitian ini, lapisan fully connected terakhir dimodifikasi menjadi empat kelas sesuai dengan jumlah tipe kulit wajah yang digunakan, kemudian seluruh parameter model dilatih ulang melalui proses fine-tuning.

C. Class Imbalance dalam Klasifikasi Citra Medis

Class imbalance merupakan tantangan fundamental dalam klasifikasi citra medis. Model cenderung bias terhadap kelas mayoritas karena fungsi loss memberikan penalti lebih besar untuk kelas yang lebih sering muncul, sehingga recall kelas minoritas secara sistematis lebih rendah. Berbagai pendekatan telah diajukan seperti SMOTE, undersampling, class-weighted loss, dan WeightedRandomSampler. Namun temuan Sylvagno dan Rochadiani (2025) menunjukkan bahwa teknik balancing sederhana justru menurunkan akurasi ke 50–60%, mengindikasikan perlunya pendekatan yang lebih spesifik sesuai karakteristik domain [5].

III. METODE

Penelitian ini menggunakan pendekatan deep learning dengan arsitektur ResNet50 untuk mengklasifikasikan tipe kulit wajah ke dalam empat kelas, yaitu acne, dry, normal, dan oily. Dataset yang digunakan berasal dari dua dataset publik Kaggle yang digabungkan menjadi satu dataset terpadu. Tahapan penelitian meliputi pengumpulan dataset, preprocessing dan augmentasi data, pelatihan model, serta evaluasi performa model. Alur penelitian yang digunakan ditunjukkan pada Gambar 1.



Gambar 1

Berdasarkan Gambar 1, penelitian diawali dengan pengumpulan dataset dan dilanjutkan dengan preprocessing serta augmentasi data. Data yang telah diproses kemudian digunakan untuk melatih model ResNet50 menggunakan pendekatan transfer learning. Selanjutnya, model dievaluasi menggunakan data testing untuk mengukur performa klasifikasi berdasarkan metrik accuracy, precision, recall, F1-score, dan confusion matrix.

A. Dataset dan Distribusi Kelas

Dataset yang digunakan dalam penelitian ini berasal dari dua dataset publik di Kaggle, yaitu *Dry*, *Oily*, and *Normal Skin Types* <https://www.kaggle.com/datasets/shakyadissanayake/oily-dry-and-normal-skin-types-dataset> dan *Acne Dataset* <https://www.kaggle.com/datasets/nayanchaure/acne-dataset>. Kedua dataset digabungkan menjadi empat kelas, yaitu *acne*, *dry*, *normal*, dan *oily*, dengan total 4.985 citra [5], [10]. Penelitian ini menggunakan pembagian dataset bawaan (*predefined split*) yang tersedia pada masing-masing dataset sumber, sehingga distribusi data antar kelas tidak sepenuhnya proporsional. Hasil pembagian terdiri atas 4.046 citra untuk training, 697 citra untuk validasi, dan 242 citra untuk testing. Distribusi data pada setiap kelas ditunjukkan pada Tabel 1.

Tabel 1 Distribusi Dataset per Kelas pada Setiap Subset

Kelas	Total Citra	Training	Validasi	Testing
Acne	1.833	1.290	435	108
Dry	758	652	71	35
Normal	1.274	1.104	111	59
Oily	1.120	1.000	80	40
Total	4.985	4.046	697	242

Berdasarkan Tabel 1, distribusi data pada setiap kelas tidak sepenuhnya seimbang. Kelas *acne* memiliki jumlah citra yang lebih banyak dibandingkan kelas lainnya, sedangkan kelas *dry* memiliki jumlah citra paling sedikit. Kondisi ini berpotensi menimbulkan bias model terhadap kelas mayoritas dan menyebabkan perbedaan performa klasifikasi antar

kelas. Oleh karena itu, selain menggunakan akurasi sebagai metrik evaluasi, penelitian ini juga menganalisis precision, recall, dan F1-score pada masing-masing kelas.

B. Preprocessing dan Augmentasi Data

Seluruh citra di-resize menjadi 224×224 piksel sesuai standar input ResNet50 dan dinormalisasi menggunakan parameter mean = 0.485, 0.456, 0.406 serta std = 0.229, 0.224, 0.225 dari dataset ImageNet [11]. Pada data training diterapkan augmentasi berupa Random Horizontal Flip dengan probabilitas 50% dan Color Jitter untuk meningkatkan keragaman data pelatihan serta membantu meningkatkan kemampuan generalisasi model dan mengurangi overfittin [12]. Augmentasi tidak diterapkan pada data validasi dan testing agar evaluasi model dilakukan pada data yang merepresentasikan kondisi sebenarnya.

C. Arsitektur Model dan Konfigurasi Training

Model yang digunakan adalah ResNet50 dengan bobot pretrained IMAGENET1K_V2 yang diunduh dari PyTorch torchvision. Lapisan fully connected (fc) terakhir yang semula memiliki output sebanyak 1.000 kelas untuk ImageNet dimodifikasi menjadi 4 kelas sesuai dengan jumlah tipe kulit yang diklasifikasikan. Modifikasi ini dilakukan dengan mengganti fc layer dengan lapisan Linear baru yang diinisialisasi secara acak. Seluruh parameter model, termasuk lapisan-lapisan pretrained, dilatih ulang (fine-tuned) selama proses training.

Tabel 2. Konfigurasi Pelatihan Model ResNet50

Parameter	Nilai	Keterangan
Backbone	ResNet50	Pretrained IMAGENET1K_V2
Optimizer	SGD	Stochastic Gradient Descent
Learning Rate Awal	0.1	LR awal sebelum scheduling
Momentum	0.9	Momentum SGD
LR Scheduler	StepLR	step_size=15, gamma=0.1
LR setelah Epoch 15	0.01	Turun 10× di epoch ke-15
Loss Function	CrossEntropyLoss	Standar untuk multi-class
Epochs	20	Dengan best model tracking
Batch Size	32	-
Input Size	224×224 piksel	Standar ResNet50
Augmentasi Training	H-Flip + Color Jitter	Tidak diterapkan di val/test

Best model tracking diterapkan selama training: model dengan akurasi validasi tertinggi disimpan sebagai checkpoint final untuk evaluasi, menghindari overfitting pada epoch-epoch akhir.

D. Metrik Evaluasi

Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data pengujian, sedangkan precision, recall, dan F1-score digunakan untuk mengevaluasi performa model pada setiap kelas [13].

Selain evaluasi secara keseluruhan, analisis dilakukan pada masing-masing kelas untuk mengidentifikasi perbedaan performa model dalam mengenali tipe kulit wajah. Pendekatan ini penting karena distribusi data antar kelas tidak sepenuhnya seimbang. Selain itu, confusion matrix digunakan untuk menganalisis pola misklasifikasi dan mengidentifikasi kelas-kelas yang sering tertukar. Hasil evaluasi tersebut kemudian digunakan untuk menganalisis performa model ResNet50 serta membandingkannya dengan temuan pada penelitian terdahulu.

IV. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil evaluasi model ResNet50 pada klasifikasi empat tipe kulit wajah, yaitu dry, normal, oily, dan acne. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, F1-score, serta confusion matrix untuk mengukur performa model. Selain itu, dilakukan analisis ketimpangan performa antar kelas dan faktor-faktor yang memengaruhinya. Hasil yang diperoleh kemudian dibandingkan dengan penelitian terdahulu untuk memberikan pemahaman yang lebih komprehensif mengenai kemampuan model dalam mengklasifikasikan tipe kulit wajah.

A. Performa Keseluruhan Model

Model ResNet50 yang dilatih pada penelitian ini mencapai akurasi sebesar 92% pada data testing yang terdiri atas 242 sampel. Performa tersebut lebih tinggi dibandingkan MobileNetV3 yang digunakan oleh Herimanto et al. (2024) dengan akurasi 84%, Vision Transformer (ViT) oleh Futri et al. (2025) dengan akurasi 78%, serta ResNet50 pada penelitian Sylvagno dan Rochadiani (2025) yang memperoleh akurasi rata-rata $90,94\% \pm 0,60\%$.

	precision	recall	f1-score	support
0	0.88	0.84	0.86	75
1	0.90	0.87	0.88	128
2	0.86	0.90	0.88	112
3	0.99	1.00	0.99	184
accuracy			0.92	499
macro avg	0.91	0.90	0.90	499
weighted avg	0.92	0.92	0.92	499

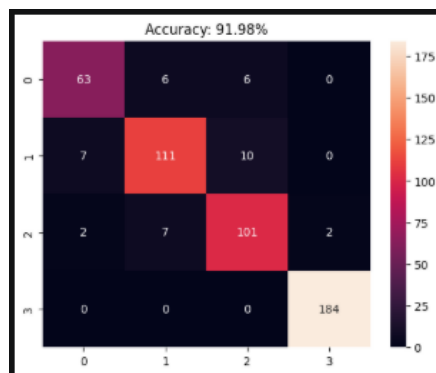
Gambar 2 Classification Report

Berdasarkan Gambar 2, evaluasi dilakukan terhadap empat kelas tipe kulit wajah, yaitu dry, normal, oily, dan acne. Hasil evaluasi menunjukkan bahwa kelas acne memiliki performa terbaik dengan precision 0,99, recall 1,00, dan F1-score 0,99. Sebaliknya, kelas dry memperoleh performa terendah dengan precision 0,88, recall 0,84, dan F1-score 0,86. Sementara itu, kelas normal dan oily menunjukkan performa yang relatif seimbang dengan F1-score masing-masing sebesar 0,88. Meskipun model mencapai akurasi

keseluruhan yang tinggi, hasil evaluasi per-kelas menunjukkan adanya variasi performa antar kelas yang akan dianalisis lebih lanjut pada subbab berikutnya.

B. Analisis Ketimpangan Performa Per-Kelas

Berdasarkan hasil evaluasi pada Gambar 2, terdapat perbedaan performa yang cukup jelas antar kelas. Kelas acne menunjukkan performa terbaik dengan recall sebesar 100%, sedangkan kelas dry memiliki recall terendah sebesar 84%. Gap recall sebesar 16 poin persentase tersebut menunjukkan bahwa kemampuan model dalam mengenali setiap tipe kulit wajah tidak sepenuhnya merata.



Gambar 3 Confusion Matrix

Gambar 3 memperlihatkan pola misklasifikasi pada setiap kelas. Seluruh sampel acne berhasil diklasifikasikan dengan benar tanpa misklasifikasi ke kelas lain. Sebaliknya, kelas dry masih mengalami kesalahan klasifikasi, di mana masing-masing 6 sampel salah diprediksi sebagai normal dan oily. Pada kelas normal terdapat 7 sampel yang salah diprediksi sebagai dry dan 10 sampel sebagai oily, sedangkan pada kelas oily jumlah misklasifikasi relatif lebih sedikit. Temuan ini menunjukkan bahwa kesalahan klasifikasi lebih sering terjadi pada kelas-kelas yang memiliki karakteristik visual yang mirip dibandingkan pada kelas acne yang memiliki ciri visual yang lebih khas.

Menariknya, kelas oily tidak menjadi kelas dengan performa terendah meskipun memiliki jumlah sampel yang lebih sedikit dibandingkan beberapa kelas lainnya. Hal ini mengindikasikan bahwa jumlah data bukan satu-satunya faktor yang memengaruhi performa klasifikasi. Oleh karena itu, diperlukan analisis lebih lanjut untuk mengidentifikasi faktor-faktor yang menyebabkan terjadinya ketimpangan performa antar kelas.

C. Faktor Penyebab Ketimpangan

1. Faktor Pertama: Distribusi Class Imbalance yang Ekstrem

Salah satu faktor yang diduga memengaruhi ketimpangan performa adalah distribusi data yang tidak seimbang. Berdasarkan Tabel 1, kelas acne memiliki 1.290 citra training, hampir dua kali lipat dibandingkan kelas dry yang hanya berjumlah 652 citra. Ketimpangan serupa juga terlihat pada data validasi dan testing. Dalam proses pelatihan, model memperoleh lebih banyak paparan terhadap kelas acne sehingga berpotensi membentuk representasi fitur yang lebih baik dibandingkan kelas dry. Hal ini tercermin

pada hasil evaluasi, di mana kelas acne memperoleh recall sebesar 100%, sedangkan kelas dry hanya mencapai 84%.

Temuan ini sejalan dengan Herimanto et al. (2024) yang melaporkan bahwa kelas dengan jumlah sampel paling sedikit memperoleh performa terendah. Selain itu, Sylvagno dan Rochadiani (2025) juga menemukan bahwa kelas dry merupakan salah satu kelas yang paling sering mengalami misklasifikasi.

2. *Faktor Kedua: Kemiripan Tekstur Visual yang Inheren*

Selain distribusi data, kemiripan karakteristik visual antar tipe kulit juga diduga memengaruhi performa model. Berdasarkan confusion matrix pada Gambar 3, sebagian besar kesalahan klasifikasi terjadi pada kelas dry, normal, dan oily, sedangkan seluruh sampel acne berhasil diklasifikasikan dengan benar.

Hal ini menunjukkan bahwa kelas acne memiliki ciri visual yang lebih khas, sedangkan perbedaan antara dry, normal, dan oily cenderung lebih halus sehingga lebih sulit dibedakan oleh model. Temuan serupa dilaporkan oleh Futri et al. (2025), yang menemukan bahwa kelas normal tetap menjadi kelas dengan performa terendah meskipun menggunakan dataset yang telah diseimbangkan. Hasil tersebut mengindikasikan bahwa kemiripan visual merupakan tantangan yang bersifat inheren dalam klasifikasi tipe kulit wajah.

Tabel 3 Perbandingan Pola Ketimpangan Per-Kelas Antar Penelitian

Penelitian	Arsitektur	Kelas Seimbang?	Kelas Terlemah	Nilai	Kelas Terkuat	Nilai
Sylvagno & Rochadiani (2025)	ResNet50	Tidak	Dry (salah ke Acne: 19 kasus)	Confusion matrix	Acne (127 benar)	Confusion matrix
Herimanto et al. (2024)	MobileNetV3	Tidak	Combination	F1=0.57	Acne	F1=0.95
Futri et al. (2025)	ViT	Ya (seimbang)	Normal	F1=0.72	Combination	F1=0.86
Penelitian Ini	ResNet50	Tidak	Dry	Recall=0.84	Acne	Recall=1.00

Berdasarkan Tabel 3, terlihat bahwa kelas dengan karakteristik visual yang lebih khas cenderung memperoleh performa lebih baik dibandingkan kelas yang memiliki kemiripan visual dengan kelas lain. Temuan ini menunjukkan bahwa evaluasi model tidak cukup dilakukan hanya berdasarkan akurasi keseluruhan, tetapi juga perlu mempertimbangkan performa pada setiap kelas.

D. *Implikasi untuk Evaluasi dan Pengembangan Sistem AI Skincare*

Temuan penelitian ini menunjukkan bahwa akurasi keseluruhan tidak selalu mencerminkan kemampuan model dalam mengenali setiap tipe kulit wajah secara merata. Meskipun model ResNet50 mencapai akurasi sebesar 92%, hasil evaluasi per-kelas

menunjukkan adanya perbedaan performa yang cukup signifikan antara kelas acne dan dry. Oleh karena itu, penggunaan akurasi sebagai satu-satunya indikator performa berpotensi memberikan gambaran yang kurang lengkap mengenai kualitas model.

Dalam pengembangan sistem AI untuk identifikasi tipe kulit wajah, evaluasi perlu dilakukan tidak hanya pada tingkat keseluruhan, tetapi juga pada masing-masing kelas. Penggunaan metrik per-kelas seperti precision, recall, dan F1-score, serta analisis confusion matrix, dapat memberikan gambaran yang lebih rinci mengenai kemampuan model dan pola misklasifikasi yang terjadi. Pendekatan tersebut membantu mengidentifikasi kelemahan model sekaligus menjadi dasar pengembangan metode yang lebih efektif pada penelitian selanjutnya.

E. Strategi Perbaikan yang Direkomendasikan

Berdasarkan hasil analisis, beberapa strategi dapat dipertimbangkan untuk mengurangi ketimpangan performa antar kelas pada penelitian selanjutnya. Pendekatan penanganan class imbalance seperti class-weighted loss, WeightedRandomSampler, maupun augmentasi yang lebih terarah pada kelas minoritas berpotensi meningkatkan representasi data selama proses pelatihan. Selain itu, pengembangan augmentasi khusus pada kelas dry dapat dilakukan untuk memperkaya variasi karakteristik visual yang dipelajari model.

Di samping itu, eksplorasi arsitektur deep learning yang lebih modern, seperti EfficientNet atau ConvNeXt, dapat menjadi alternatif untuk meningkatkan kemampuan model dalam membedakan karakteristik visual yang relatif mirip antar tipe kulit wajah. Namun, efektivitas strategi-strategi tersebut masih perlu dibuktikan melalui eksperimen lebih lanjut untuk memperoleh performa yang lebih merata pada seluruh kelas.

V. KESIMPULAN

Penelitian ini menganalisis performa per-kelas model ResNet50 pada klasifikasi empat tipe kulit wajah, yaitu acne, dry, normal, dan oily. Hasil pengujian menunjukkan bahwa model mencapai akurasi 92%, lebih tinggi dibandingkan beberapa penelitian pembandingan. Namun, analisis per-kelas mengungkap adanya ketimpangan performa, di mana kelas acne memperoleh recall 100%, sedangkan kelas dry hanya 84%, sehingga terdapat gap sebesar 16 poin persentase.

Ketimpangan tersebut diduga dipengaruhi oleh dua faktor utama, yaitu distribusi data yang tidak seimbang dan kemiripan karakteristik visual antar tipe kulit. Temuan ini menunjukkan bahwa akurasi keseluruhan tidak selalu merepresentasikan kemampuan model secara merata, sehingga evaluasi perlu mempertimbangkan metrik per-kelas seperti precision, recall, F1-score, serta analisis confusion matrix.

Untuk penelitian selanjutnya, dapat diterapkan teknik penanganan class imbalance seperti class-weighted loss atau WeightedRandomSampler, serta pengembangan augmentasi dan arsitektur deep learning yang lebih optimal guna meningkatkan performa model pada seluruh kelas tipe kulit wajah.

DAFTAR PUSTAKA

- [1] N. Suhandi, "Literatur Riview Identifikasi Dan Klasifikasi Skincare Berdasarkan Permasalahan Kulit Wajah," *J. Tata Rias*, vol. 14, no. 2, hal. 69–83, 2024.
- [2] S. D. Kusumaningrum, "Ekstraksi Ciri untuk Klasifikasi Tipe Kulit Wajah Menggunakan Citra Foto," 2024, *Universitas Islam Indonesia*.
- [3] P. Aisyah, "Gambaran Sel Epitel Kulit Wajah Pengguna Whitening Cream Non-BPOM Pada Mahasiswa Institut Teknologi Sains Dan Kesehatan Insan Cendekia Medika Jombang," 2025, *ITSKes Insan Cendekia Medika Jombang*.
- [4] F. Mursyid *dkk.*, "DETEKSI ACNE VULGARIS DAN JENIS KULIT PADA CITRA WAJAH BERBASIS YOLOV7 DAN RESNET50," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 3, hal. 120–130, 2025.
- [5] G. S. Sylvagno dan T. H. Rochadiani, "Skincare Recommendation System Based on Facial Skin Type with Real-Time Weather Integration," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 2, hal. 230–237, 2025.
- [6] Herimanto, A. S. Dharma, J. Amalia, D. Largo, dan C. A. P. Sihite, "Performance Analysis of MobileNetV3-based Convolutional Neural," vol. 5, no. 158, hal. 701–709, 2024, [Daring]. Tersedia pada: <http://jurnal.iaii.or.id>
- [7] Nur, Ade, dan Ahmad, "Design of Facial Skin Type Detection Application Using CNN with Inceptionv3 Model and Google Cloud Platform," *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 2, hal. 1239–1248, 2025, doi: 10.35314/v171yr28.
- [8] N. Effendi, D. Handoko, F. Azim, dan F. Farida, "Jurnal Computer Science and Information Technology (CoSciTech) things," vol. 5, no. 2, hal. 358–366, 2024.
- [9] K. He, X. Zhang, S. Ren, dan J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, hal. 770–778.
- [10] S. D. Kusumaningrum dan I. Muhimmah, "ANALISIS FAKTOR DAN METODE UNTUK MENENTUKAN TIPE KULIT WAJAH," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 4, hal. 753–762, 2023, [Daring]. Tersedia pada: <https://jtiik.ub.ac.id/index.php/jtiik/article/view/6955/pdf>
- [11] D. A. Agustina, "Klasifikasi citra jenis kulit wajah dengan algoritma convolutional neural network (cnn) resnet-50," *J. Ris. Sist. Inf.*, vol. 1, no. 3, hal. 1–7, 2024.
- [12] M. M. Nugraha, M. D. F. Saputra, D. A. Fauzan, dan E. Y. Puspaningrum, "Augmentasi Data Pengenalan Citra Batik Yogyakarta Menggunakan Pendekatan Random Crop, Rotate, Dan Mixup," in *Prosiding Seminar Nasional Informatika Bela Negara (SANTIKA)*, 2025, hal. 152–158.
- [13] S. Clara, D. Laksmi Prianto, R. Al Habsi, E. Friscila Lumbantobing, dan N. Chamidah, "Implementasi Seleksi Fitur Pada Algoritma Klasifikasi Machine Learning Untuk Prediksi Penghasilan Pada Adult Income Dataset," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, vol. 2, no. 1, hal. 741–747, 2021.