Vol. 11, No. 2, Oktober 2026, Page. 554-565
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Aplikasi Klasifikasi Jenis Jerawat Menggunakan Algoritma *Convolutional Neural Network* (CNN) Berbasis *Transfer Learning* untuk Rekomendasi Penanganan dan Perawatan Kulit

Siti Muyassaroh[*1], Tri Widodo[2]

[1,2]Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta, Indonesia
Email: [1]sitimuyassaroh011@gmail.com, [2]triwidodo@uty.ac.id

Abstrak

Jerawat (*acne vulgaris*) merupakan masalah kulit yang umum dan dapat memengaruhi kondisi psikologis penderitanya. Identifikasi jenis jerawat yang masih dilakukan secara visual dan mandiri bersifat subjektif serta berisiko menimbulkan pemilihan perawatan yang kurang tepat. Penelitian ini bertujuan mengembangkan aplikasi web untuk mengklasifikasikan lima jenis jerawat, yaitu *blackheads*, *cyst*, *papules*, *pustules*, dan *whiteheads*, menggunakan *Convolutional Neural Network* (CNN) berbasis *transfer learning* dengan arsitektur EfficientNetB1, dilengkapi rekomendasi penanganan dan perawatan kulit. Dataset publik dari Kaggle dan Roboflow dibagi menjadi 3.066 data latih, 928 data validasi, dan 927 data uji. Citra diubah ukurannya menjadi 224×224 piksel, kemudian model dilatih menggunakan *optimizer* Adam dengan variasi jumlah *epoch* dan *learning rate*. Model diintegrasikan ke dalam aplikasi web yang menerima citra dari galeri perangkat maupun hasil pengambilan gambar langsung melalui kamera. Pengujian menggunakan tiga seed menghasilkan rata-rata *accuracy* sebesar 88,35 ± 2,30%, *precision* makro 89,44 ± 2,42%, *recall* makro 86,68 ± 3,32%, dan *F1-score* makro 87,78 ± 3,06%. Untuk keperluan integrasi ke dalam aplikasi web yang membutuhkan satu model definitif, model dari iterasi *seed* 42 dipilih karena menunjukkan performa evaluasi tertinggi dengan *accuracy* mencapai 90,83%. Kesalahan klasifikasi paling sering terjadi antara *papules* dan *pustules* yang memiliki kemiripan visual. Aplikasi menampilkan jenis jerawat, *confidence score*, dan rekomendasi perawatan secara otomatis sehingga berpotensi membantu identifikasi awal yang lebih cepat dan objektif.

Kata kunci: *acne vulgaris*, aplikasi web, *convolutional neural network*, EfficientNetB1, klasifikasi jerawat, *transfer learning*.

Abstract

Acne vulgaris is a common skin condition that can affect the psychological well-being of those who suffer from it. Identifying acne types through visual self-assessment is subjective and may lead to inappropriate skincare choices. This study aims to develop a web application that classifies five acne types, namely blackheads, cyst, papules, pustules, and whiteheads, using a Convolutional Neural Network (CNN) based on EfficientNetB1 transfer learning, complemented by treatment and skincare recommendations. Public datasets from Kaggle and Roboflow were split into 3,066 training, 928 validation, and 927 test images. Images were resized to 224×224 pixels, and the model was trained with the Adam optimizer while varying the number of epochs and the learning rate. The model was integrated into a web application that accepts images from the device gallery or photos captured directly with the camera. Evaluation using three random seeds yielded a mean accuracy of 88.35 ± 2.30%, macro-precision of 89.44 ± 2.42%, macro-recall of 86.68 ± 3.32%, and macro-F1-score of 87.78 ± 3.06%. For integration into the web application, which requires a single definitive model, the model trained with seed 42 was selected because it achieved the highest evaluation performance with an accuracy of 90.83%. Misclassifications occurred most often between papules and pustules because of their visual similarity. The application displays the predicted acne type, confidence score, and care recommendations, and can support faster and more objective initial identification.

Keywords: *acne vulgaris, classification, convolutional neural network, Efficientnetb1, transfer learning, web application.*

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)

1. PENDAHULUAN

Kesehatan kulit wajah merupakan aspek penting yang memengaruhi kepercayaan diri dan interaksi sosial seseorang, dan salah satu permasalahan kulit wajah yang paling sering dijumpai adalah jerawat

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

(Acne Vulgaris)[1]. Jerawat diperkirakan memengaruhi sekitar 9,4% populasi dunia[2], [3]. Selain memengaruhi penampilan, jerawat berkaitan dengan risiko stigmatisasi, depresi, dan kecemasan[3],[4]. Jerawat memiliki beberapa jenis, antara lain komedo (blackheads dan whiteheads), papula, pustula, dan kista, yang dibedakan menjadi lesi non-inflamasi dan inflamasi dengan karakteristik yang berbeda. Pengenalan jenis lesi yang tepat diperlukan agar penanganan efektif[2], [4]. Namun, masih banyak penderita yang kesulitan membedakan jenis jerawat yang dialami dan menentukan perawatan yang sesuai[5].

Penilaian jenis dan tingkat keparahan jerawat oleh klinisi pun masih bersifat subjektif dan bergantung pada pengalaman[6],[7], sementara akses ke dokter spesialis kulit terbatas, terutama di daerah terpencil[7]. Kondisi ini menegaskan kebutuhan akan sistem klasifikasi otomatis yang lebih objektif dan praktis berbasis Convolutional Neural Network (CNN)[7], [8]. Namun, performa terbaik CNN umumnya dicapai ketika model dilatih pada data berlabel dalam jumlah besar[9], sedangkan ketersediaan citra jerawat berlabel relatif terbatas[1], [10]. Transfer learning, yaitu inisialisasi model dengan bobot yang telah dilatih pada dataset besar, dapat membantu meningkatkan performa dan mengurangi risiko overfitting pada dataset target yang kecil[9]. Pada penelitian ini, pendekatan tersebut diterapkan dengan arsitektur EfficientNetB1.

Penelitian terdahulu telah menerapkan berbagai arsitektur deep learning untuk klasifikasi jerawat. [1] menggunakan MobileNet untuk mengklasifikasikan empat jenis jerawat berdasarkan 213 citra yang dikumpulkan dari Kaggle, Bing, dan Google, dengan akurasi sebesar 96,57%. [4] mengembangkan Dual Integrated Deep CNN untuk mengenali tujuh kategori jerawat menggunakan 420 citra yang diperluas menjadi 3.360 citra melalui augmentasi, dengan akurasi sebesar 97,53%. [5] membandingkan lima arsitektur deep learning untuk klasifikasi lima kategori jerawat pada 360 citra dan memperoleh akurasi validasi sebesar 92% menggunakan ResNet50V2. [6] mengusulkan DLI-Net yang menggabungkan DeepLabV3 untuk segmentasi dan Modified InceptionV3 untuk klasifikasi tiga kategori kondisi kulit, dengan akurasi 97% dan F1-score 0,973. Sementara itu, [7] menerapkan EfficientNetV2-S untuk klasifikasi jerawat pada wajah. Penelitian [4] juga menguji model mereka menggunakan dataset umum ACNE04 untuk mengklasifikasikan tingkat keparahan jerawat ke dalam empat kategori. Perbedaan jumlah citra, kategori yang diklasifikasikan, arsitektur, serta skema evaluasi perlu dipertimbangkan dalam membandingkan hasil penelitian tersebut karena akurasi yang tinggi tidak selalu menunjukkan performa yang setara pada karakteristik data yang berbeda.

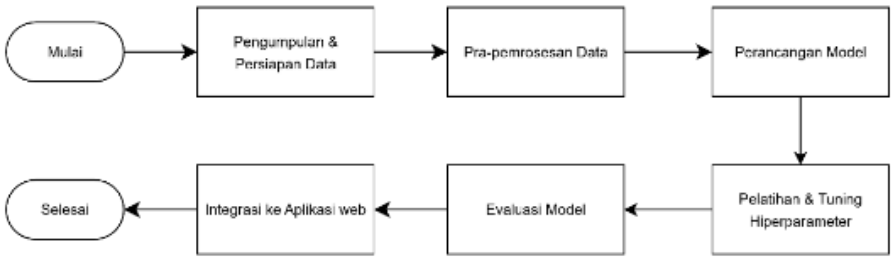
Dari kajian tersebut terlihat dua celah yang belum dijawab penelitian sebelumnya. Pertama, sebagian besar studi[1], [4], [6], [7] berhenti pada klasifikasi dan tidak menautkan keluaran model ke rekomendasi yang dapat ditelusuri ke pedoman klinis, sedangkan [5] menyajikan rekomendasi dalam aplikasi tetapi tidak menjelaskan dasar klinisnya. Kedua, hasil umumnya dilaporkan dari satu kali pelatihan, sehingga kestabilan model terhadap inisialisasi bobot dan faktor acak lainnya belum diketahui. Celah ini penting karena aplikasi yang menampilkan rekomendasi kepada pengguna awam berisiko mendorong pemilihan perawatan yang keliru bila hasil model tidak stabil, bias terhadap kelas mayoritas, atau rekomendasinya tidak berdasar.

Berdasarkan celah tersebut, penelitian ini menerapkan CNN berbasis transfer learning dengan arsitektur EfficientNetB1 untuk mengklasifikasikan lima jenis jerawat (blackheads, cyst, papules, pustules, dan whiteheads). EfficientNetB1 dipilih karena menggunakan compound scaling yang menyeimbangkan kedalaman, lebar, dan resolusi jaringan sehingga mencapai akurasi yang baik dengan jumlah parameter yang relatif kecil. Efisiensi ini sesuai dengan tujuan penelitian, yaitu model yang dijalankan pada aplikasi web, dan keluarga EfficientNet telah menunjukkan hasil yang baik pada klasifikasi jerawat[7]. Penelitian ini tidak membandingkan seluruh arsitektur secara sistematis; perbandingan terbatas dengan Vanilla CNN dan MobileNetV2 disajikan pada bagian Hasil. Kontribusi penelitian ini meliputi pemetaan hasil klasifikasi ke rekomendasi perawatan yang mengacu pada pedoman klinis [3], [11], [12] , evaluasi kestabilan model menggunakan tiga variasi seed dengan pembagian data yang identik, yang hasilnya dilaporkan dalam bentuk rata-rata dan simpangan baku, serta perbandingan performa dengan model baseline Vanilla CNN. Evaluasi kinerja dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Model terpilih kemudian diintegrasikan ke dalam aplikasi web yang mendukung input melalui unggah gambar dan kamera. Sistem ini diposisikan sebagai alat bantu skrining awal dan tidak menggantikan diagnosis oleh tenaga medis profesional.

Vol. 11, No. 2, Oktober 2026, Page. 554-565
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

2. METODE PENELITIAN

Tahapan penelitian ini dirancang secara terstruktur dan sistematis untuk membangun serta mengevaluasi model klasifikasi lima jenis jerawat menggunakan arsitektur EfficientNetB1 berbasis transfer learning, yang selanjutnya diintegrasikan ke dalam aplikasi berbasis web. Alur penelitian secara umum digambarkan dalam bentuk diagram alir (flowchart) pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan dan Persiapan Data

Dataset sekunder berupa citra jerawat diambil dari Kaggle[13] dan Roboflow[14], dengan lima kelas yaitu blackheads, cyst, papules, pustules, dan whiteheads. Pemakaian citra terbuka dari repositori daring juga lazim pada penelitian klasifikasi jerawat dan kulit wajah sebelumnya[1], [7], [10], [15]. Untuk menangani ketidakseimbangan kelas, pendekatan oversampling diterapkan pada tingkat data. Dataset eksternal dari Roboflow[14] ditambahkan khusus pada kelas whiteheads dalam subset data latih untuk melengkapi representasi visual kelas tersebut. Sementara itu, data validasi dan data uji tetap mengikuti pembagian dataset sumber. Meskipun penambahan data ini bertujuan memperkaya variasi citra whiteheads pada proses pelatihan, perbedaan karakteristik visual antar-sumber data berpotensi memengaruhi pola yang dipelajari model. Oleh karena itu, nilai precision 100% pada kelas whiteheads perlu diinterpretasikan secara hati-hati karena belum dapat dipastikan sejauh mana hasil tersebut mencerminkan pengenalan karakteristik lesi atau dipengaruhi oleh perbedaan karakteristik visual antar-sumber data.

Dataset yang digunakan telah memuat citra hasil augmentasi dari sumbernya sehingga augmentasi tambahan tidak diterapkan dalam penelitian ini. Total dataset terdiri atas 4.921 citra RGB, yang terbagi menjadi 3.066 data latih, 928 data validasi, dan 927 data uji (sekitar 62%:19%:19%) sesuai pembagian pada dataset sumber. Dataset tersebut telah memiliki tiga subset terpisah, yaitu train, validation, dan test, yang masing-masing terdiri atas lima folder kelas jerawat. Penelitian ini menggunakan pembagian subset (train, validation, test) bawaan dari repositori sumber guna menjaga konsistensi dengan distribusi asli dataset. Mengingat dataset merupakan hasil agregasi publik yang telah memuat augmentasi bawaan, modifikasi atau pemfilteran ulang pada tingkat citra tidak dilakukan untuk mempertahankan keutuhan data awal. Contoh citra setiap kelas ditampilkan pada Gambar 2.



Gambar 2. Contoh Sampel Citra Jenis Jerawat dalam Dataset

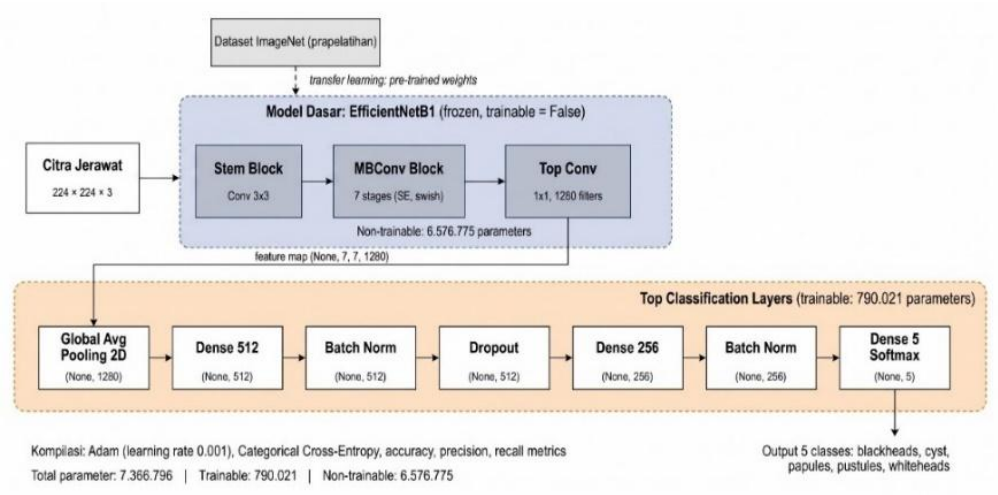
2.2 Pra-pemrosesan Data

Tahap pra-pemrosesan (preprocessing) dilakukan untuk menyelaraskan karakteristik citra masukan sebelum diproses oleh arsitektur Convolutional Neural Network (CNN). Seluruh citra diubah menjadi 224×224 piksel dengan layer Resizing, ukuran yang juga dipakai pada penelitian klasifikasi jerawat dan

penyakit kulit[1], [7], [16]. Augmentasi tambahan tidak diterapkan karena sebagian besar data sumber sudah berupa hasil augmentasi.

2.3 Perancangan Model

Model memakai EfficientNetB1 berbobot pre-trained ImageNet sebagai ekstraktor fitur. Base model dibekukan (trainable = False) sehingga hanya lapisan di atasnya yang dilatih[9]. Di atasnya disusun Global Average Pooling 2D yang mengubah feature map (7, 7, 1280) menjadi vektor 1280[8], lalu dua blok Dense (512 dan 256 unit), masing-masing dengan Batch Normalization dan Dropout untuk menstabilkan pelatihan dan menekan overfitting[7], [8]. Output layer berupa Dense 5 unit dengan aktivasi Softmax pada persamaan (1)[8]. Arsitektur ini (Gambar 3) memiliki 7.366.796 parameter, terdiri atas 790.021 parameter trainable dan 6.576.775 non-trainable.



Gambar 3. Arsitektur Model

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \qquad (1)$$

Dengan z_i skor keluaran (logit) kelas ke-i dan $K = 5$ jumlah kelas.

Sebagai pembanding, dirancang pula baseline Vanilla CNN yang dilatih dari awal tanpa bobot prapelatihan, terdiri atas tiga blok Conv2D dan Max Pooling, lalu Flatten dan Dense layer.

2.4 Pelatihan dan Tuning Parameter

Pelatihan model dilakukan menggunakan arsitektur EfficientNetB1 dengan variasi jumlah epoch 30, 33, 35, dan 40 pada learning rate 0,0001. Selanjutnya, dilakukan perbandingan learning rate 0,0001, 0,0005 dan 0,001 pada konfigurasi 30 epoch dengan ukuran batch 32 dan seed 42. Pemilihan konfigurasi terbaik didasarkan pada akurasi validasi dan selisih akurasi antara data latih dan validasi, sedangkan data uji tidak digunakan sebagai dasar pemilihan konfigurasi.

Selama pelatihan, mekanisme EarlyStopping dan ModelCheckpoint disiapkan untuk memantau validation loss. Pada seluruh skenario yang diuji, pelatihan mencapai batas maksimum epoch tanpa memicu penghentian dini, sehingga ModelCheckpoint berperan utama dalam menyimpan bobot model pada epoch dengan validation loss terendah. Evaluasi kestabilan model dilakukan menggunakan tiga variasi seed, yaitu 12, 42, dan 72, dengan pembagian data yang sama. Hasil evaluasi ketiga seed dilaporkan dalam bentuk rata-rata dan simpangan baku. Model dengan konfigurasi terpilih, yaitu 30 epoch dan learning rate 0,001, menggunakan seed 42 untuk integrasi ke dalam aplikasi web. Evaluasi dengan tiga variasi seed dilakukan secara terpisah untuk mengukur kestabilan performa model.

2.5 Evaluasi Model

Evaluasi model dilakukan pada data testing untuk mengukur kemampuan generalisasi model dalam mengklasifikasikan lima jenis jerawat dengan membandingkan nilai prediksi (predicted label) terhadap

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

nilai aktualnya (actual label)[1], [7]. Evaluasi ini memanfaatkan confusion matrix yang memuat True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) sebagai dasar perhitungan metrik kuantitatif.

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

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

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

$$\text{Loss} = -\sum_{i=1}^{C} y_i \cdot \log(\hat{y}_i) \quad (5)$$

Performa model diukur dengan accuracy (2), precision (3), dan recall (4), sedangkan galat probabilitas prediksi diukur dengan categorical cross-entropy loss.

2.6 Integrasi ke Aplikasi Web

Model terlatih disimpan lalu dimuat oleh aplikasi web. Platform web dipilih karena dapat diakses lewat browser tanpa instalasi[17]. Pengguna dapat mengunggah gambar dari galeri atau mengambil foto langsung melalui kamera. Citra disesuaikan ke ukuran input model, lalu inferensi menghasilkan jenis jerawat beserta confidence score. Hasilnya dipetakan secara otomatis ke rekomendasi penanganan dan perawatan kulit, seperti pada aplikasi rekomendasi berbasis Android di[5].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pelatihan Model

Pelatihan dilakukan menggunakan EfficientNetB1 dengan base model yang dibekukan. Pengujian mencakup variasi epoch 30, 33, 35, dan 40 pada learning rate (LR) 0,0001, serta perbandingan LR 0,0001, 0,0005, dan 0,001 pada 30 epoch. Seluruh eksperimen menggunakan batch size 32 dan seed 42. Pemilihan konfigurasi mempertimbangkan akurasi validasi dan selisih akurasi data latih dengan validasi.

Selama pelatihan, EarlyStopping memantau validation loss dengan patience sebesar 8 epoch. Seluruh eksperimen mencapai batas maksimum epoch yang ditetapkan. ModelCheckpoint digunakan untuk menyimpan bobot model dengan validation loss terbaik.

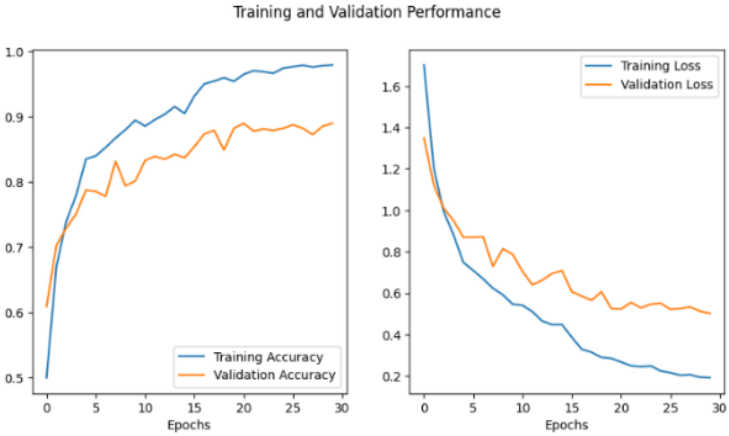
Tabel 1. Hasil Tuning Hiperparameter

Skenario Eksperimen	Akurasi Training	Akurasi Validasi	Akurasi Testing
Variasi Epoch (LR = 0,0001)			
30 Epoch	92,99%	85,02%	88,03%
33 Epoch	94,29%	86,10%	86,95%
35 Epoch	95,11%	85,88%	87,27%
40 Epoch	95,92%	86,64%	86,84%
Variasi Learning Rate (30 Epoch)			
0,0001	92,99%	85,02%	88,03%

0,0005	97,85%	88,15%	89,86%
0,001	97,98%	89,01%	90,83%

Berdasarkan Tabel 1, penambahan epoch pada LR 0,0001 meningkatkan akurasi training, tetapi peningkatan akurasi validasi relatif terbatas dan tidak berlangsung secara monoton. Selisih akurasi training-validasi juga melebar dari 7,97 menjadi 9,28 poin persentase. Sementara itu, penggunaan LR 0,001 pada 30 epoch menghasilkan akurasi validasi tertinggi, yaitu 89,01%. Oleh karena itu, konfigurasi tersebut ditetapkan sebagai konfigurasi terbaik di antara skenario yang diuji, meskipun belum dapat dinyatakan sebagai konfigurasi optimum secara global.

Setelah konfigurasi terbaik diperoleh, dilakukan eksperimen fine-tuning dengan membuka 30 lapisan terakhir EfficientNetB1, sementara lapisan BatchNormalization tetap dibekukan. Fine-tuning dilakukan menggunakan learning rate 0,00001 dengan batas maksimum 15 epoch. Hasil evaluasi menunjukkan bahwa setelah proses fine-tuning, model memperoleh accuracy sebesar 90,08%, precision 91,11%, recall 89,54%, dan test loss 0,4670. Dibandingkan dengan model sebelum fine-tuning yang memperoleh accuracy 90,83% dan test loss 0,4795, terjadi penurunan accuracy sebesar 0,75 poin persentase, sedangkan test loss mengalami penurunan. Hasil ini menunjukkan bahwa fine-tuning tidak meningkatkan seluruh metrik evaluasi secara bersamaan. Oleh karena itu, efektivitas fine-tuning perlu ditinjau berdasarkan keseluruhan metrik dan karakteristik kesalahan klasifikasi yang dihasilkan.



Gambar 4. Performa Model

Gambar 4 menunjukkan kurva akurasi dan loss dari konfigurasi terpilih. Akurasi training dan validasi meningkat pada awal pelatihan, kemudian peningkatannya melambat. Pada akhir pelatihan, akurasi training mencapai 97,98% dan akurasi validasi 89,01%, sedangkan loss training dan validasi masing-masing mencapai sekitar 0,2 dan 0,5. Perbedaan antara kedua nilai akurasi tersebut menunjukkan bahwa performa model pada data latih lebih tinggi dibandingkan pada data validasi.

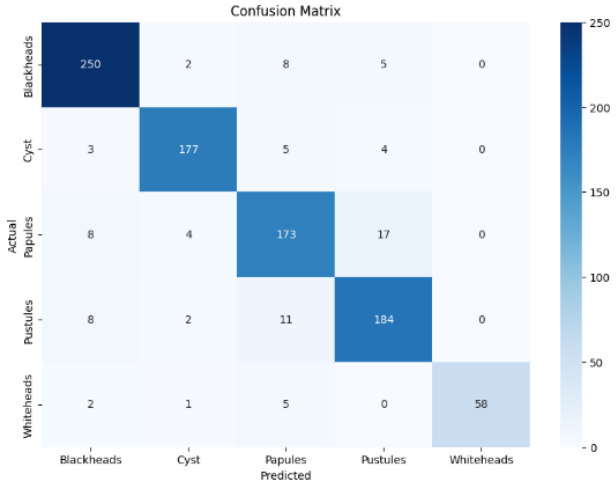
3.2 Hasil Evaluasi Model

Evaluasi dilakukan pada 927 citra uji yang tidak digunakan pada pelatihan maupun pemilihan konfigurasi. Untuk keperluan implementasi ke dalam aplikasi web, evaluasi detail difokuskan pada model dengan performa terbaik, yaitu model dengan konfigurasi LR 0,001 pada seed 42. Model ini memperoleh accuracy sebesar 90,83% (842 dari 927 citra terklasifikasi benar) dan loss 0,4795. Dengan jumlah data uji tersebut, selang kepercayaan 95% untuk akurasi berada pada kisaran 89,0%-92,7% (perkiraan). Precision, recall, dan F1-score dihitung per kelas dengan pendekatan one-vs-rest, lalu dirata-ratakan secara macro (tiap kelas berbobot sama) dan weighted (berbobot sesuai jumlah data uji). Hasilnya disajikan pada Tabel 2, sedangkan distribusi prediksi ditampilkan pada Gambar 5.

Tabel 2. Metrik Evaluasi per Kelas pada Data Uji

Kelas	Precision (%)	Recall (%)	F1-score (%)	Jumlah data uji
Blackheads	92,25	94,34	93,28	265
Cyst	95,16	93,65	94,40	189

Papules	85,64	85,64	85,64	202
Pustules	87,62	89,76	88,67	205
Whiteheads	100,00	87,88	93,55	66
Macro average	92,13	90,25	91,11	927
Weighted average	90,93	90,83	90,84	927



Gambar 5. Confusion Matrix Model pada Data Uji

Berdasarkan Gambar 5, prediksi benar tertinggi ada pada blackheads (250), diikuti pustules (184), cyst (177), papules (173), dan whiteheads (58). Kinerja terendah ada pada papules (precision dan recall 85,64%). Dari 85 citra yang salah, 28 (32,9%) merupakan kekeliruan timbal balik papules-pustules (17 papules terbaca pustules, 11 sebaliknya), sejalan dengan karakteristik klinisnya: pustule adalah papule berisi nanah yang pada citra beresolusi rendah tampak samar[1] Kekeliruan lain yang menonjol adalah blackheads-papules (8 dan 8) dan pustules terbaca blackheads (8).

Kelas whiteheads memiliki precision 100% namun recall 87,88% (8 dari 66 citra terbaca kelas lain, 5 di antaranya papules). Dengan hanya 66 data uji, estimasi kelas ini kurang stabil.

Untuk menguji kestabilan terhadap faktor acak (inisialisasi bobot dan dropout), pelatihan dijalankan pada tiga seed (12, 42, 72) dengan pembagian data yang identik; hasilnya pada Tabel 3.

Tabel 3. Kinerja Model pada Berbagai Inisialisasi Seed

Seed	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Test Loss
12	87,92	88,74	86,11	87,14	0,5424
42	90,83	92,13	90,25	91,11	0,4795
72	86,30	87,44	83,69	85,09	0,5462
Rata-rata	88,35 ± 2,30	89,44 ± 2,42	86,68 ± 3,32	87,78 ± 3,06	0,5227 ± 0,0375

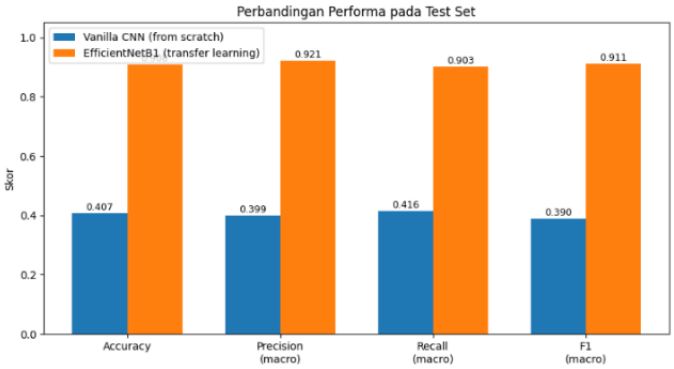
Rata-rata accuracy 88,35% (±2,30), precision makro 89,44%, recall makro 86,68%, dan F1-score makro 87,78%, dengan accuracy antar seed berkisar 86,30% - 90,83%. Model relatif konsisten, tetapi seed 42 yang dipakai pada Tabel 2 adalah yang terbaik, sehingga performa umum lebih tepat dirujuk ke nilai rata-rata.

Sebagai pembanding, Vanilla CNN dilatih dari awal tanpa bobot prapelatihan menggunakan tiga blok konvolusi dengan 32, 64, dan 128 filter, diikuti MaxPooling, Flatten, Dense 128 neuron, dan Dropout 0,5. Pelatihan menggunakan optimizer Adam dengan learning rate 0,001 dan maksimum 30 epoch, disertai EarlyStopping untuk menghentikan pelatihan ketika performa validasi tidak membaik. Vanilla CNN memperoleh accuracy sebesar 40,70% dan macro F1-score sebesar 39,00% pada data uji. Perbandingan dilakukan menggunakan seed 42 dan data uji yang sama. Hasil pada Gambar 6 menunjukkan bahwa EfficientNetB1 memperoleh performa yang lebih tinggi dibandingkan Vanilla CNN pada konfigurasi eksperimen yang digunakan.

Sebagai model pembanding berbasis transfer learning, MobileNetV2 dilatih menggunakan bobot pralatih ImageNet dengan backbone yang dibekukan dan lapisan klasifikasi untuk lima kelas jerawat.

Vol. 11, No. 2, Oktober 2026, Page. 554-565
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Pelatihan menggunakan ukuran citra 224 × 224 piksel, batch size 32, learning rate awal 0,001, dan maksimum 30 epoch. Proses pelatihan menggunakan seed 42, dengan EarlyStopping, ReduceLROnPlateau, dan ModelCheckpoint untuk mengontrol pelatihan serta menyimpan model terbaik berdasarkan validation loss. Pada data uji yang sama sebanyak 927 citra, MobileNetV2 memperoleh accuracy 70,87%, macro precision 69,90%, macro recall 70,00%, dan macro F1-score 69,80%.



Gambar 6. Perbandingan Performa Model Baseline Vanilla CNN dan EfficientNetB1 pada Data Uji

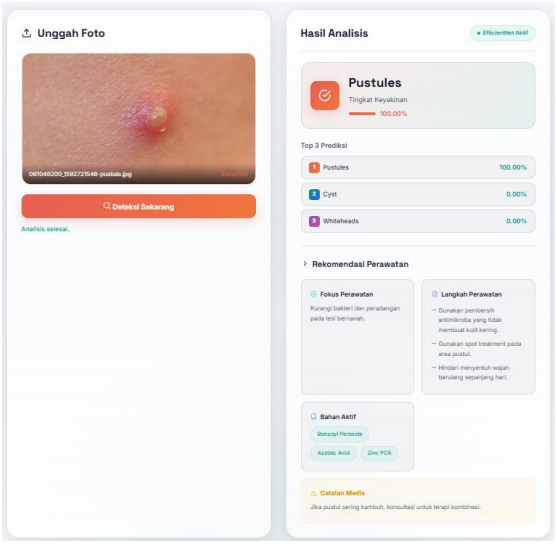
3.3 Implementasi Aplikasi Web

Model terbaik (30 epoch, LR 0,001) disimpan lalu dimuat aplikasi web tanpa pelatihan ulang. Pengguna memilih citra melalui unggah berkas atau kamera pada halaman masukan (Gambar 7).



Gambar 7. Halaman Masukan

Citra diubah ke 224×224 piksel, lalu model menghasilkan probabilitas lima kelas. Sistem menampilkan tiga kelas teratas (top-3) beserta confidence score kelas utama, dan kelas utama dipetakan ke rekomendasi penanganan dan perawatan kulit pada halaman hasil (Gambar 8).



Gambar 8. Halaman Hasil

Top-3 membantu pengguna melihat kelas alternatif saat keyakinan model rendah, terutama untuk pasangan yang mirip seperti papules dan pustules. Pada contoh Gambar 8, citra pustules diprediksi benar dengan confidence 100%.

Rekomendasi pada aplikasi diadaptasi dari tiga pedoman klinis. Bahan aktif dan terapi mengacu pada pedoman[3] dan [11], yaitu benzoyl peroxide dan retinoid topikal (rekomendasi kuat), serta asam salisilat dan azelaic acid (rekomendasi bersyarat)[3]. Kelas cyst diarahkan ke dokter kulit karena isotretinoin oral hanya untuk jerawat berat dengan resep[3]. Saran perawatan umum, seperti membersihkan wajah secara lembut dan tidak memencet jerawat, mengacu pada [12].

Tabel 4. Pemetaan Kelas ke Rekomendasi pada Aplikasi

Kelas	Tipe (EuroGuiDerm)	Rekomendasi yang ditampilkan	Sumber
Blackheads, Whiteheads	Komedonal	Retinoid topikal (adapalene 0,1% tersedia bebas); asam salisilat sebagai opsi tambahan; tabir surya harian	Retinoid: kuat; asam salisilat: bersyarat; tabir surya bersama retinoid: keterangan pedoman[3] , [11]
Papules, Pustules	Papulopustular	Benzoyl peroxide, retinoid topikal, atau kombinasi keduanya	Kuat, certainty sedang[3], [11]
Cyst	Nodular	Konsultasi dokter kulit; tidak ada saran pengobatan mandiri	Isotretinoin untuk jerawat berat[3], [11]

Pemetaan kelas pada Tabel 4 mengikuti pengelompokan tipe jerawat pada [11]. Informasi yang ditampilkan bersifat awal dan bukan pengganti diagnosis atau resep dokter.

Pengujian aplikasi dilakukan dengan cara pengujian fungsional (black-Box), memeriksa apakah setiap fitur berjalan sesuai skenario tanpa melihat kode program (Tabel 5).

Tabel 5. Pengujian Fungsional (black-Box)

No.	Skenario	Hasil yang diharapkan	Hasil
1	Unggah berkas JPG/PNG valid	Citra tampil, klasifikasi berjalan	Sesuai
2	Ambil foto melalui kamera	Kamera aktif, foto diproses	Sesuai
3	Izin kamera ditolak	Pesan kesalahan tampil	Sesuai
4	Tekan analisis tanpa memilih citra	Pesan peringatan tampil	Sesuai
5	Unggah berkas bukan gambar	Berkas ditolak dengan pesan	Sesuai
6	Hasil klasifikasi	Kelas, top-3, confidence, dan rekomendasi tampil	Sesuai

3.4 Pembahasan

Hasil pengujian menunjukkan bahwa EfficientNetB1 berbasis transfer learning mampu mengklasifikasikan lima jenis jerawat dengan baik. Konfigurasi 30 epoch dan learning rate 0,001 menghasilkan akurasi testing sebesar 90,83%. Penggunaan transfer learning sesuai untuk dataset dengan jumlah data terbatas karena memanfaatkan bobot model yang telah dilatih sebelumnya[9]. Hasil ini juga sejalan dengan penelitian sebelumnya yang menunjukkan efektivitas CNN dan transfer learning dalam klasifikasi jerawat[5], [7], serta penerapan deep learning pada klasifikasi citra[8].

Eksperimen fine-tuning dilakukan dengan membuka 30 lapisan terakhir EfficientNetB1, sementara lapisan BatchNormalization tetap dibekukan, menggunakan learning rate 0,00001 dan maksimum 15 epoch. Model setelah fine-tuning memperoleh accuracy 90,08% dan test loss 0,4670, dibandingkan model sebelum fine-tuning dengan accuracy 90,83% dan test loss 0,4795, sehingga accuracy turun 0,75 poin persentase meskipun test loss menurun.

Baseline Vanilla CNN yang dilatih dari awal hanya mencapai akurasi uji 40,70% dan macro F1-score 39,00%, jauh di bawah EfficientNetB1. Selisih ini menunjukkan bahwa bobot prapelatihan dapat membantu meningkatkan performa klasifikasi pada dataset yang digunakan.

Sebagai pembanding tambahan, MobileNetV2 diuji menggunakan bobot pralatih ImageNet dengan backbone yang dibekukan dan lapisan klasifikasi untuk lima kelas jerawat. Pelatihan dilakukan dengan ukuran citra 224 × 224 piksel, batch size 32, learning rate awal 0,001, dan maksimum 30 epoch. Pada data uji yang sama, MobileNetV2 memperoleh akurasi 70,87% dan macro F1-score 69,80%. Hasil tersebut lebih tinggi dibandingkan Vanilla CNN, tetapi masih di bawah EfficientNetB1 yang memperoleh akurasi 90,83%. Perbandingan ini menunjukkan bahwa performa model transfer learning dapat berbeda bergantung pada arsitektur dan konfigurasi yang digunakan.

Untuk mengevaluasi pengaruh penambahan data, dataset Kaggle dikombinasikan dengan data Roboflow sehingga menghasilkan 4.921 citra, terdiri atas 3.066 data latih, 928 data validasi, dan 927 data uji. Dibandingkan model yang dilatih menggunakan dataset Kaggle saja, macro F1-score meningkat dari 90,68% menjadi 91,11%. Recall kelas blackheads meningkat dari 91,70% menjadi 94,34%, cyst dari 92,59% menjadi 93,65%, dan pustules dari 84,39% menjadi 89,76%. Hasil ini menunjukkan peningkatan performa pada beberapa kelas, meskipun pengaruh penambahan data belum merata pada seluruh kelas.

Meskipun demikian, kesalahan klasifikasi paling banyak terjadi antara kelas papules dan pustules, yaitu sebanyak 28 dari total 85 kesalahan. Hal ini berkaitan dengan kemiripan karakteristik visual kedua jenis lesi sehingga sulit dibedakan melalui citra[1]. Namun, kedua kelas tersebut dipetakan ke kategori rekomendasi papulopustular yang sama dalam sistem, sehingga kesalahan klasifikasi di antara keduanya tidak mengubah kategori rekomendasi yang ditampilkan. Meski demikian, model masih perlu ditingkatkan agar mampu membedakan kedua jenis lesi secara lebih konsisten. Evaluasi model tidak hanya mengandalkan accuracy, tetapi juga menggunakan precision, recall, dan F1-score[18] untuk memberikan gambaran performa klasifikasi yang lebih menyeluruh, terutama pada data dengan distribusi kelas yang tidak seimbang.

Kesalahan klasifikasi pada kelas cyst perlu mendapat perhatian khusus karena berpotensi memengaruhi rekomendasi penanganan yang diterima pengguna. Dari 189 citra cyst pada data uji, sebanyak 12 citra salah diklasifikasikan sebagai kelas lain. Dalam sistem ini, rekomendasi konsultasi dokter diberikan berdasarkan hasil klasifikasi cyst. Oleh karena itu, apabila citra cyst diprediksi sebagai kelas lain, pengguna berpotensi tidak memperoleh anjuran konsultasi dokter yang sesuai. Temuan ini menunjukkan bahwa kesalahan klasifikasi tidak hanya memengaruhi metrik performa model, tetapi juga dapat berdampak pada relevansi rekomendasi yang ditampilkan.

Secara fungsional, aplikasi telah dapat menerima dan memproses citra jerawat yang diambil langsung melalui kamera ponsel. Namun, akurasi inferensi pada citra nyata (real-world) tersebut belum diukur secara metrik dan masih dapat dipengaruhi oleh kondisi pencahayaan, sudut pengambilan gambar, dan kualitas citra. Perbedaan karakteristik foto dengan citra pada dataset juga berpotensi memengaruhi ketepatan prediksi. Oleh karena itu, pengujian pada kondisi pengambilan gambar yang lebih beragam masih diperlukan untuk mengetahui konsistensi performa model dalam penggunaan nyata.

Sistem ini merupakan alat bantu skrining awal, bukan alat diagnosis, sehingga hasilnya tidak boleh menjadi dasar keputusan pengobatan tanpa konsultasi dokter. Penelitian ini memiliki beberapa keterbatasan: model hanya mengenali lima kelas morfologi, tidak menilai tingkat keparahan, dan tetap memilih salah satu kelas meskipun citra bukan jerawat; dataset berasal dari repositori publik [13], [14] yang didistribusikan dengan pembagian (split) bawaan tanpa pemfilteran duplikasi, sehingga potensi kebocoran data (data leakage) antarsubset tidak dapat dikesampingkan sepenuhnya; ketiadaan informasi warna kulit sehingga kinerja pada kelompok kulit yang kurang terwakili belum diketahui; confidence score belum dikalibrasi sehingga nilai tinggi tidak menjamin ketepatan; dan rekomendasi perawatan belum divalidasi dermatolog. Dari sisi privasi, citra wajah merupakan data pribadi. Citra yang diunggah disimpan untuk pengembangan model lanjutan hanya apabila pengguna memberikan persetujuan, dengan akses terbatas dan tanpa metadata lokasi, serta tidak dibagikan kepada pihak lain.

4. KESIMPULAN

Berdasarkan hasil pengujian dan pembahasan yang telah dilakukan, arsitektur EfficientNetB1 dengan pendekatan transfer learning mampu mengklasifikasikan lima kategori morfologi jerawat. Pada evaluasi menggunakan tiga variasi seed, model memperoleh rata-rata accuracy sebesar 88,35% ± 2,30% dan macro F1-score sebesar 87,78% ± 3,06%. Sementara itu, model dengan seed 42 yang digunakan

dalam aplikasi web memperoleh accuracy sebesar 90,83% dan macro F1-score sebesar 91,11%. Hasil analisis kesalahan menunjukkan bahwa kemiripan visual, khususnya antara papules dan pustules, masih menjadi faktor penyebab kesalahan klasifikasi.

Penerapan model ke dalam platform web menyediakan fasilitas klasifikasi melalui fitur unggah gambar maupun pengambilan gambar secara langsung menggunakan kamera. Hasil klasifikasi disertai confidence score dan rekomendasi perawatan sebagai informasi awal bagi pengguna. Namun, performa model masih menunjukkan variasi antar-seed dan indikasi overfitting, dengan selisih akurasi pelatihan dan pengujian sebesar 7,15 poin persentase pada seed 42. Selain itu, rekomendasi yang diberikan belum divalidasi oleh dermatolog sehingga sistem tidak menggantikan diagnosis medis profesional. Pengembangan selanjutnya dapat difokuskan pada peningkatan konsistensi model, penambahan variasi data, dan validasi rekomendasi oleh tenaga medis.

5. DAFTAR PUSTAKA

[1] A. R. Andini and I. Yuadi, "Klasifikasi Jenis Jerawat Berdasarkan Convolutional Neural Network," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 301–308, Jan. 2025, doi: 10.57152/malcom.v5i1.1678.

[2] M. Vasam, S. Korutla, and R. A. Bohara, "Acne vulgaris: A review of the pathophysiology, treatment, and recent nanotechnology based advances," Dec. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.bbrep.2023.101578.

[3] R. V. Reynolds *et al.*, "Guidelines of care for the management of acne vulgaris," *J. Am. Acad. Dermatol.*, vol. 90, no. 5, pp. 1006.e1-1006.e30, May 2024, doi: 10.1016/j.jaad.2023.12.017.

[4] M. B. Islam, M. S. Junayed, A. Sadeghzadeh, N. Anjum, A. A. Jeny, and A. F. M. S. Shah, "Acne Vulgaris Detection and Classification: A Dual Integrated Deep CNN Model," *Informatica (Slovenia)*, vol. 47, no. 4, pp. 577–592, Dec. 2023, doi: 10.31449/inf.v47i4.4384.

[5] Prayitno and M. Abduh, "Optimalisasi Model Deep Learning CNN Berbasis Android untuk Identifikasi Jenis Jerawat dan Rekomendasi Perawatan Kulit Wajah," in *Seminar Keinsinyuran*, 2024.

[6] S. Sharmin *et al.*, "A Hybrid CNN Framework DLI-Net for Acne Detection with XAI," *J. Imaging*, vol. 11, no. 4, Apr. 2025, doi: 10.3390/jimaging11040115.

[7] A. Y. Pramono and K. Kusnawi, "Multi-Class Facial Acne Classification using the EfficientNetV2-S Deep Learning Model," *Aviation Electronics, Information Technology, Telecommunications, Electricals, and Controls (AVITEC)*, vol. 7, no. 3, pp. 299–309, Aug. 2025, doi: 10.28989/avitec.v7i3.3157.

[8] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif. Intell. Rev.*, vol. 57, no. 4, Apr. 2024, doi: 10.1007/s10462-024-10721-6.

[9] J. Plested, M. Phiri, and T. Gedeon, "Deep transfer learning for image classification: a survey," *Artif. Intell. Rev.*, vol. 59, no. 3, Jan. 2026, doi: 10.1007/s10462-026-11491-z.

[10] R. Rianto and D. R. Listianto, "Convolutional Neural Network untuk mengklasifikasi tingkat keparahan jerawat," *AITI: Jurnal Teknologi Informasi*, vol. 20, pp. 167–176, Aug. 2023.

[11] A. Nast *et al.*, "Update of the EuroGuiDerm evidence-based guideline for the treatment of acne—Short version," *Journal of the European Academy of Dermatology and Venereology*, vol. 40, no. 7, pp. 1162–1172, Jul. 2026, doi: 10.1111/jdv.70331.

[12] National Institute for Health and Care Excellence, "Acne vulgaris: management," London, 2021. [Online]. Available: www.nice.org.uk/guidance/ng198

[13] tiswan14, "Acne dataset Image ," Kaggle. Accessed: Sep. 29, 2026. [Online]. Available: https://www.kaggle.com/datasets/tiswan14/acne-dataset-image

[14] acne-label, "Whiteheads Kaggle," Roboflow Universe. Accessed: Sep. 29, 2026. [Online]. Available: https://universe.roboflow.com/acne-label/whiteheads-kaggle-37ey8/dataset/1

[15] A. P. Gani and E. Poerwandono, "Pengembangan Aplikasi Android Berbasis Machine Learning untuk Pendeteksi Jenis Kulit Wajah Menggunakan Deep learning," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 6, Dec. 2025.

[16] N. Tarisafitri and A. Aljabar, "Classification of Skin Diseases using Digital Image Processing with MobileNetV2 Architecture," *Nusantara Journal of Artificial Intelligence and Information Systems*, vol. 1, no. 1, pp. 35–44, Jun. 2025, doi: 10.47776/nuai.v1i1.1594.

[17] S. M. Saia, N. G. Nelson, S. N. Young, S. Parham, and M. Vandegrift, "Ten simple rules for researchers who want to develop web apps," *PLoS Comput. Biol.*, vol. 18, no. 1, Jan. 2022, doi: 10.1371/journal.pcbi.1009663.

[18] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, "A survey on imbalanced learning: latest research, applications and future directions," *Artif. Intell. Rev.*, vol. 57, no. 6, Jun. 2024, doi: 10.1007/s10462-024-10759-6.