



Implementasi vision transformer untuk klasifikasi jenis jerawat pada citra wajah berbasis web

Tasya Evrillia Widiarto¹, Imam Sanjaya², Zaenal Alamsyah³

Email: ¹tasya.evrillia_ti22@nusaputra.ac.id, ²imam.sanjaya@nusaputra.ac.id, ³zaenal.alamsyah@nusaputra.ac.id

^{1,2,3}Program Studi Teknik Informatika, Teknik Komputer dan Desain, Universitas Nusa Putra

Diterima: 28 Juni 2026 | Direvisi: 24 Agustus 2026 | Disetujui: 30 Agustus 2026

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Jerawat atau Acne Vulgaris tergolong gangguan kulit yang lazim dijumpai dan dapat berdampak pada kesehatan fisik maupun rasa percaya diri penderitanya. Karena penentuan jenis jerawat secara manual masih membutuhkan keahlian khusus, pendekatan kecerdasan buatan diperlukan untuk membantu proses klasifikasi tersebut. Penelitian ini menerapkan arsitektur Vision Transformer (ViT-B/16) untuk mengelompokkan jenis jerawat pada citra wajah ke dalam empat kategori, yakni normal, papula, pustula, dan nodul. Sebanyak 4.000 citra digunakan sebagai dataset dan diolah melalui pendekatan transfer learning dengan bobot awal hasil pra-pelatihan pada ImageNet. Model dilatih sepanjang 20 epoch memakai optimizer Adam dan learning rate 0,001. Pengujian memperlihatkan bahwa model dapat mencapai akurasi 86% pada data uji. Selanjutnya, model tersebut diintegrasikan ke dalam sebuah aplikasi berbasis web guna mempermudah identifikasi jenis jerawat. Temuan ini menegaskan bahwa Vision Transformer dapat mengklasifikasikan jenis jerawat secara akurat dan berpotensi menjadi alat bantu awal dalam mendeteksi kondisi kulit wajah secara otomatis.

Kata kunci: *Acne Vulgaris, Citra Wajah, Klasifikasi Jerawat, Vision Transformer*

Implementation of vision transformer for web-based acne type classification from facial images

Abstract

Acne Vulgaris is a widespread skin condition that can affect not only physical skin health but also a person's self-confidence. Because manually distinguishing acne types still demands specialized expertise, an artificial-intelligence-based approach is needed to support this classification task. This research applies a Vision Transformer (ViT-B/16) architecture to categorize acne lesions from facial images into four groups: normal, papule, pustule, and nodule. A total of 4,000 images were used as the dataset and processed through a transfer-learning strategy initialized with pre-trained ImageNet weights. The model was trained across 20 epochs with the Adam optimizer and a learning rate of 0.001. Testing showed that the model reached an accuracy of 86%. The resulting model was then embedded into a web-based application to streamline the acne identification workflow. These outcomes confirm that Vision Transformer can classify acne types reliably and holds promise as an automated early-screening tool for facial skin conditions.

Keywords: *Acne Vulgaris, Acne Classification, Facial Image, Vision Transformer*

1. PENDAHULUAN

Wajah menjadi bagian tubuh yang paling banyak disorot dalam pergaulan sosial sehari-hari, sehingga kondisi kulitnya kerap menjadi perhatian utama [1]. Salah satu masalah kulit wajah yang paling umum dijumpai adalah jerawat atau Acne Vulgaris, sebuah peradangan pada unit pilosebacea yang muncul akibat produksi sebum berlebih, sumbatan pori karena hiperkeratinisasi folikel, dan aktivitas bakteri pemicu inflamasi [2]. Prevalensi gangguan ini di Indonesia tergolong tinggi, diperkirakan mencapai

80–85% dan paling sering dialami kalangan remaja hingga dewasa muda [3]. Kelompok Studi Dermatologi Kosmetika Indonesia (KSDKI) bahkan mencatat Acne Vulgaris sebagai salah satu dari tiga penyakit kulit dengan kasus terbanyak yang ditangani klinik dan rumah sakit setiap tahun [4]. Selain memengaruhi kondisi fisik kulit, jerawat juga kerap menurunkan kepercayaan diri, memicu kecemasan sosial, hingga mengganggu produktivitas penderitanya [5]. Persoalan ini bertambah rumit karena masyarakat awam umumnya sulit membedakan jenis-jenis jerawat seperti papula, pustula, dan nodul yang masing-masing memerlukan penanganan berbeda, sehingga banyak yang akhirnya melakukan penanganan mandiri secara keliru dan memilih produk skincare yang justru dapat memperparah peradangan, sementara konsultasi langsung ke dokter spesialis kulit masih terkendala biaya dan jarak [6].

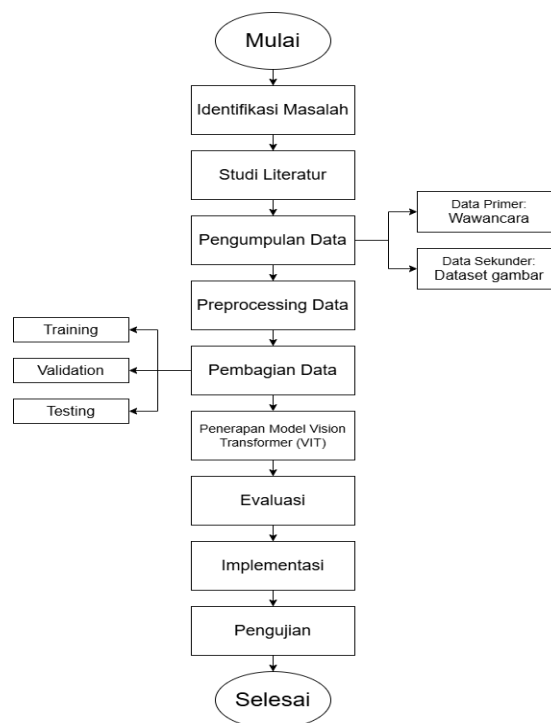
Kemajuan teknologi Computer Vision membuka jalan untuk mengatasi kendala tersebut melalui klasifikasi citra wajah yang otomatis dan objektif [7]. Berbagai penelitian terdahulu telah mengeksplorasi pendekatan ini dengan arsitektur Convolutional Neural Network (CNN) yang dikenal andal dalam mengekstraksi fitur visual secara berlapis [8]. Misalnya, klasifikasi jerawat berbasis CNN pernah mencapai akurasi hingga 96,57% [9], sementara ekstraksi fitur tekstur secara manual dengan GLCM dan Backpropagation hanya menghasilkan akurasi 56,67% akibat terbatasnya data dan tingginya kemiripan visual antarkelas [10]. Di sisi lain, pendekatan ensemble CNN berbasis Android pernah diujicobakan dengan akurasi 95,25% serta waktu inferensi yang relatif singkat [11], sedangkan CNN tipe VGG16 yang digunakan untuk menilai keparahan jerawat hanya mencapai akurasi sekitar 69% [12]. Penerapan CNN berbasis transfer learning juga terbukti efektif pada domain klasifikasi citra lainnya, seperti klasifikasi penyakit daun tomat menggunakan arsitektur VGG-19 yang mencapai akurasi hingga 97% [13]. Meski capaian tersebut cukup menjanjikan, CNN pada dasarnya memproses fitur gambar hanya dalam jangkauan lokal sesuai ukuran kernel konvolusinya, sehingga kurang mampu menangkap hubungan spasial antarwilayah wajah yang berjauhan [14] — suatu aspek yang justru penting untuk memahami pola penyebaran jerawat secara menyeluruh.

Kendala tersebut mendorong pemanfaatan arsitektur Vision Transformer (ViT) yang semula dirancang untuk pengolahan bahasa alami, namun kini banyak diadopsi untuk berbagai tugas visi komputer [15]. Melalui mekanisme self-attention, ViT mengolah citra sebagai kumpulan patch yang saling terhubung secara global sehingga mampu menangkap keterkaitan antarwilayah citra tanpa terbatas jarak spasial sebagaimana pada CNN [16]. Keunggulan ini juga terbukti pada domain klasifikasi citra lain, misalnya pada identifikasi 15 varian citra ikan koi, di mana ViT berhasil meningkatkan akurasi klasifikasi dibandingkan pendekatan CNN pada penelitian sebelumnya [17]. Pada klasifikasi kondisi kulit wajah berbasis web, penerapan ViT bahkan mencatatkan akurasi hingga 97,30% [18], meskipun fokusnya masih terbatas pada pengelompokan jenis kulit secara umum (berminyak, kering, normal) dan belum menjangkau secara spesifik jenis-jenis lesi jerawat inflamasi seperti papula, pustula, dan nodul.

Menyikapi kesenjangan tersebut, penelitian ini mengadopsi model Vision Transformer, khususnya varian ViT-Base/16-224-in21k, untuk mengklasifikasikan citra wajah ke dalam empat kategori, yaitu normal, papula, pustula, dan nodul. Berbeda dari penelitian-penelitian sebelumnya yang lebih menitikberatkan pada klasifikasi jenis kulit wajah secara umum, penelitian ini secara khusus menyasar identifikasi jenis jerawat inflamasi sekaligus mengintegrasikannya ke dalam sistem berbasis web. Pemilihan model ini didasari keunggulannya dalam memanfaatkan representasi fitur visual hasil pra-pelatihan pada ImageNet-21k melalui pendekatan transfer learning, sehingga proses fine-tuning pada data jerawat dapat berlangsung lebih efektif [18]. Resolusi input 224×224 piksel yang dipilih juga dinilai mampu menyeimbangkan kebutuhan detail visual dengan efisiensi komputasi yang diperlukan [19]. Guna menjamin kemudahan akses bagi pengguna, model yang dikembangkan kemudian diintegrasikan ke dalam sistem berbasis web yang bersifat cross-platform sehingga dapat diakses tanpa instalasi tambahan, baik melalui smartphone maupun komputer [20]. Penelitian ini secara umum bertujuan untuk: (1) menerapkan model Vision Transformer dalam mengklasifikasikan jenis jerawat pada citra wajah secara otomatis, (2) menganalisis performa model tersebut dalam mengenali karakteristik visual tiap kelas jerawat, dan (3) mengembangkan sistem berbasis web yang terintegrasi dengan model tersebut sebagai alat bantu identifikasi awal yang praktis dan mudah diakses masyarakat.

2. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan kuantitatif dengan desain eksperimen untuk membangun dan menguji model Vision Transformer (ViT-B/16) pada tugas klasifikasi jenis jerawat dari citra wajah. Rangkaian penelitian mencakup pengumpulan dan preprocessing dataset, pembagian data, pelatihan model, evaluasi performa dengan sejumlah metrik pengujian, serta implementasi model ke dalam aplikasi berbasis web guna mendukung identifikasi awal jenis jerawat secara otomatis.



Gambar 1. Tahapan Penelitian

2.1. Identifikasi Masalah

Tahap ini dilakukan untuk merumuskan fokus persoalan yang hendak dijawab dalam penelitian [21]. Pengamatan awal menunjukkan bahwa masyarakat masih kerap kesulitan membedakan jenis jerawat seperti papula, pustula, dan nodul, sehingga rawan melakukan penanganan yang kurang tepat. Di samping itu, diagnosis yang akurat umumnya membutuhkan pendampingan dokter spesialis kulit yang tidak selalu mudah dijangkau [22]. Atas dasar itu, dibutuhkan sebuah sistem berbasis web yang dapat membantu identifikasi awal jenis jerawat secara otomatis menggunakan model Vision Transformer (ViT).

2.2. Studi Literatur

Tahap ini ditempuh dengan menelusuri berbagai referensi ilmiah seperti buku, jurnal, artikel, dan penelitian terdahulu yang relevan dengan topik yang dikaji [23]. Fokus kajian meliputi karakteristik jenis-jenis jerawat, teknologi computer vision, arsitektur Vision Transformer (ViT), serta perkembangan sistem berbasis web. Tahap ini bertujuan membangun landasan teori, memperkuat pemilihan metode penelitian, dan memastikan penelitian ini memiliki relevansi serta nilai kebaruan dibandingkan penelitian-penelitian sebelumnya [12].

2.3. Pengumpulan Data

2.3.1 Data Primer

Data primer dihimpun melalui wawancara semi-terstruktur yang dilakukan langsung bersama dokter spesialis kulit di Klinik dr. Metz, Kota Sukabumi. Wawancara ini bertujuan menggali informasi mendalam mengenai karakteristik visual masing-masing kelas yang digunakan dalam penelitian, yaitu normal, papula, pustula, dan nodul, sekaligus memvalidasi rekomendasi kandungan skincare yang akan ditampilkan pada sistem [24]. Hasil wawancara tersebut kemudian dijadikan validasi ahli (expert validation) dalam menyusun kelas klasifikasi dan rekomendasi perawatan kulit pada website yang dikembangkan.

2.3.2 Data Sekunder

Data sekunder pada penelitian ini berupa kumpulan citra wajah yang diambil dari dua platform publik, yakni Kaggle dan Roboflow. Dari proses pengumpulan tersebut diperoleh empat kelas data, yaitu 2.357 citra normal, 1.322 citra papula, 1.423 citra pustula, dan 1.000 citra nodul, sehingga total keseluruhan mencapai 6.102 citra.

Mengingat jumlah data antarkelas tidak seimbang, dilakukan penyeimbangan menggunakan metode Random Undersampling (RUS) sehingga jumlah data tiap kelas disesuaikan menjadi 1.000 citra, dan total dataset yang dipakai dalam penelitian menjadi 4.000 citra. Dataset yang telah seimbang ini selanjutnya melalui tahap preprocessing sebelum digunakan pada proses pelatihan dan pengujian model Vision Transformer (ViT).

2.4. Data Preprocessing

Tahap preprocessing ditempuh untuk menyiapkan dataset agar sesuai dengan kebutuhan model Vision Transformer (ViT) sebelum pelatihan dimulai [12]. Tujuannya adalah menyamakan format citra sehingga seluruh data dapat dimanfaatkan secara optimal pada tahap pelatihan. Selain itu, preprocessing juga membantu meningkatkan kualitas data sekaligus meminimalkan perbedaan karakteristik antar citra yang berpotensi memengaruhi hasil pelatihan.

Pada penelitian ini, preprocessing diawali dengan menyesuaikan ukuran seluruh citra menjadi 224×224 piksel memakai `transforms.Resize` agar sesuai dengan kebutuhan input arsitektur Vision Transformer [24]. Citra kemudian diubah menjadi bentuk tensor melalui fungsi `ToTensor` agar dapat diolah oleh framework deep learning. Pada tahap ini pula, rentang nilai piksel yang semula 0–255 dikonversi menjadi 0–1 untuk mempermudah proses komputasi. Tahap terakhir berupa normalisasi memakai nilai mean dan standard deviation dari ImageNet, yang bertujuan menyeragamkan sebaran data, menjaga stabilitas pelatihan, dan membantu model mempelajari karakteristik visual tiap kelas jerawat secara lebih optimal [12].

2.5. Pembagian Data

Setelah preprocessing rampung, dataset dipecah menjadi tiga bagian, yakni data pelatihan (training), data validasi (validation), dan data pengujian (testing) dengan perbandingan 70:15:15. Data pelatihan berfungsi melatih model mengenali karakteristik visual tiap kelas, data validasi digunakan untuk memantau performa model selama pelatihan sekaligus menekan risiko overfitting, sedangkan data pengujian dipakai untuk menilai kemampuan model terhadap data yang belum pernah dilihat pada tahap pelatihan maupun validasi, sehingga hasil yang diperoleh mencerminkan performa model secara lebih objektif.

2.6. Penerapan Model Vision Transformer

Pemodelan pada penelitian ini memakai arsitektur Vision Transformer (ViT-B/16) yang sebelumnya telah melalui tahap pre-training pada dataset ImageNet. Model ini diterapkan untuk mengelompokkan citra wajah ke dalam empat kategori, yaitu normal, papula, pustula, dan nodul. Pendekatan transfer learning dimanfaatkan agar pengetahuan yang telah dipelajari model sebelumnya dapat digunakan kembali sehingga proses pelatihan berjalan lebih efektif dan efisien.

Pelatihan model menggunakan data latih yang telah melewati tahap preprocessing dan pembagian data. Model dilatih selama 20 epoch dengan fungsi Cross Entropy Loss sebagai loss function dan algoritma Adam sebagai optimizer pada learning rate 0,001. Sepanjang pelatihan, bobot model diperbarui secara bertahap pada tiap iterasi sehingga model semakin mengenali karakteristik visual masing-masing kelas.

Selain pelatihan, validasi juga dilakukan pada setiap epoch memakai data validasi guna memantau kemampuan model mengklasifikasikan data yang belum pernah dipelajari sebelumnya. Tahap ini berfungsi mengevaluasi performa model sekaligus memastikan model mampu melakukan generalisasi dengan baik sebelum diuji pada data uji.

2.7. Evaluasi

Evaluasi dilakukan untuk mengukur sejauh mana Vision Transformer (ViT) mampu mengklasifikasikan citra wajah ke dalam empat kelas, yaitu normal, papula, pustula, dan nodul [25]. Pengujian memanfaatkan data testing yang sama sekali tidak dilibatkan pada tahap pelatihan maupun validasi, sehingga hasilnya dapat mencerminkan performa model secara objektif. Proses evaluasi diawali dengan analisis confusion matrix yang membandingkan hasil prediksi model dengan label sebenarnya, guna mengetahui jumlah prediksi benar maupun salah pada tiap kelas. Dalam confusion matrix, True Positive (TP) merujuk pada data yang berhasil diklasifikasikan secara tepat ke kelas yang sesuai, sedangkan True Negative (TN) merujuk pada data yang berhasil dikenali bukan sebagai kelas tersebut. Sebaliknya, False Positive (FP) terjadi bila model keliru mengklasifikasikan data ke suatu kelas, sedangkan False Negative (FN) terjadi bila model gagal mengenali data yang sesungguhnya termasuk ke dalam kelas tersebut. Struktur confusion matrix yang dipakai pada penelitian ini disajikan pada Tabel 1.

Tabel 1. Confusion Matrix

Confusion Matrix		Kelas Sebenarnya	
		Positif	Negatif
Prediksi	Positif	True Positif	False Positif
	Negatif	False Negatif	True Negatif

Berpedoman pada nilai confusion matrix tersebut, performa model kemudian dinilai memakai metrik accuracy, precision, recall, dan F1-score. Accuracy menggambarkan ketepatan prediksi secara keseluruhan, sementara precision, recall, dan F1-score dipakai untuk menilai kemampuan model dalam mengklasifikasikan setiap kelas secara terpisah.

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

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

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

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

2.8. Implementasi

Implementasi sistem ditempuh dengan menanamkan model Vision Transformer (ViT) yang telah melalui pelatihan dan pengujian ke dalam website berbasis Flask yang dibangun memakai bahasa pemrograman Python. Sistem ini dirancang agar pengguna dapat memperoleh identifikasi awal jenis jerawat berdasarkan citra wajah yang mereka unggah. Alurnya dimulai ketika pengguna mengakses website dan mengunggah citra wajah melalui halaman yang disediakan. Citra tersebut kemudian diproses oleh model Vision Transformer untuk mengekstraksi karakteristik visual dan mengklasifikasikannya ke salah satu dari empat kelas — Normal, Papula, Pustula, atau Nodul. Hasil klasifikasi lalu dikirim kembali ke antarmuka website dan ditampilkan kepada pengguna beserta rekomendasi kandungan skincare yang sesuai dengan kondisi kulit yang terdeteksi. Melalui integrasi model klasifikasi dan website ini, pengguna dapat memperoleh gambaran awal kondisi kulit wajahnya secara lebih praktis hanya melalui satu platform yang dapat diakses lewat peramban web.

3. HASIL DAN PEMBAHASAN

3.1. Preprocessing Data

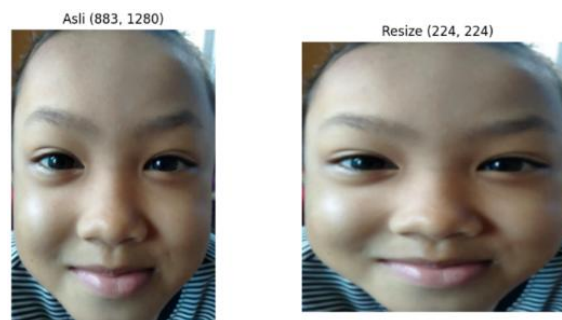
Setelah proses penyeimbangan selesai, dataset dibagi menjadi data latih, data validasi, dan data uji dengan rasio 70:15:15 agar tahap pelatihan, validasi, dan evaluasi model dapat berlangsung secara objektif. Sebaran dataset setelah undersampling dan pembagian data ditampilkan pada Tabel 2.

Tabel 2. Jumlah Data Set

Kelas	Training	Validation	Testing
Normal	700	150	150
Papula	700	150	150
Pustula	700	150	150
Nodul	700	150	150
Total	4000		

Berdasarkan Tabel 2, jumlah data pada tiap kelas sudah setara sehingga distribusi dataset menjadi lebih seimbang. Kondisi ini diharapkan membantu model mempelajari karakteristik visual tiap kelas secara merata sekaligus mengurangi kecenderungan model memihak pada satu kelas tertentu. Sebelum dipakai dalam pelatihan model Vision Transformer (ViT), seluruh citra terlebih dahulu melewati tahap preprocessing yang meliputi resize, konversi ke tensor, dan normalisasi, dengan tujuan menyeragamkan format data masukan sesuai kebutuhan model.

Tahap pertama adalah resize, yaitu menyesuaikan ukuran seluruh citra menjadi 224×224 piksel memakai fungsi `Resize(224,224)`, karena arsitektur Vision Transformer (ViT-B/16) mensyaratkan ukuran masukan yang seragam. Penyeragaman dimensi citra ini membuat proses pelatihan berlangsung lebih konsisten tanpa terpengaruh perbedaan ukuran gambar pada dataset. Contoh hasil resize dapat dilihat pada Gambar 2.



Gambar 2. Hasil Resize Citra Menjadi 224×224 Piksel

Setelah dimensi citra diseragamkan, citra dikonversi ke bentuk tensor melalui fungsi `ToTensor()` agar dapat diproses oleh framework PyTorch. Pada tahap ini, rentang nilai piksel yang semula 0–255 turut diubah menjadi 0–1 supaya lebih sesuai untuk komputasi model deep learning.

Tahap normalisasi terakhir memakai parameter mean dan standard deviation milik dataset ImageNet, yaitu mean = [0,485; 0,456; 0,406] dan standard deviation = [0,229; 0,224; 0,225], mengingat model Vision Transformer yang dipakai merupakan model pretrained hasil pelatihan pada ImageNet. Normalisasi ini membantu menyelaraskan sebaran nilai piksel agar pelatihan berjalan lebih stabil dan model dapat mempelajari pola data secara lebih efektif.

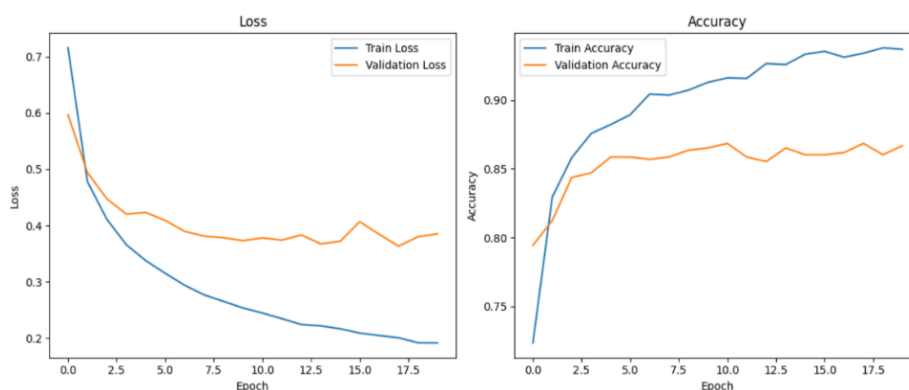
3.2. Pemodelan

Model klasifikasi jenis jerawat pada penelitian ini dibangun di atas arsitektur Vision Transformer (ViT-B/16) yang telah dipelajari (pretrained) pada dataset ImageNet. Pemanfaatan model pretrained ini bertujuan menggunakan kembali representasi fitur visual yang sudah dipelajari sebelumnya, sehingga pelatihan pada dataset jerawat dapat berlangsung lebih efektif. Model dilatih untuk mengklasifikasikan citra wajah ke dalam empat kelas, yaitu normal, papula, pustula, dan nodul. Parameter yang dipakai selama pelatihan disajikan pada Tabel 3.

Tabel 3. Parameter Pelatihan

No	Parameter	Nilai
1	Arsitektur	Model <i>Vision Transformer</i> (ViT-B/16)
2	<i>Optimizer</i>	Adam
3	<i>Loss Function</i>	CrossEntropyLoss
4	Metrik	<i>Accuracy, Loss</i>
5	<i>Epoch</i>	20
6	Batch Size	16
7	<i>Learning Rate</i>	0,001

Pelatihan model memakai data latih yang telah melewati tahap preprocessing. Sepanjang proses ini, model mempelajari karakteristik visual masing-masing kelas dari citra wajah yang menjadi masukan. Performanya dipantau pada setiap epoch memakai metrik accuracy dan loss, baik pada data pelatihan maupun data validasi.



Gambar 3. Grafik Loss dan Accuracy

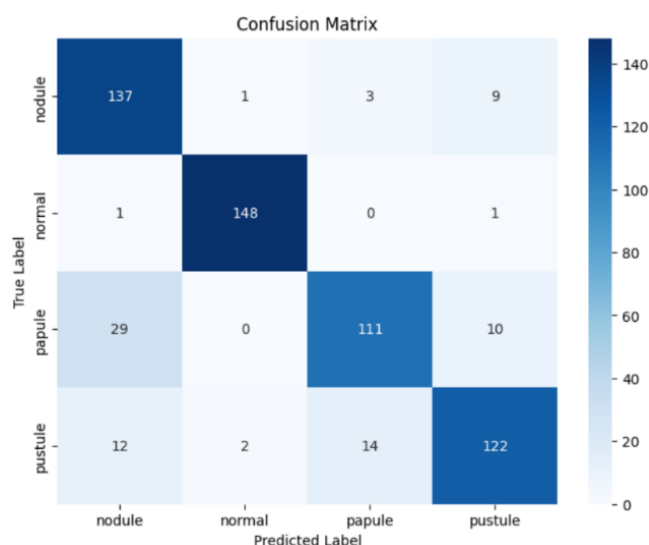
Grafik pada Gambar 3 memperlihatkan tren peningkatan accuracy pada data pelatihan maupun validasi sepanjang proses pelatihan berlangsung, sementara nilai loss pada kedua himpunan data menurun secara bertahap. Hal ini mengindikasikan model mampu mempelajari pola dan karakteristik visual dataset secara efektif.

Kurva validasi juga terlihat relatif stabil hingga akhir pelatihan, yang menandakan model tidak hanya menguasai data pelatihan tetapi juga cukup andal mengenali pola pada data baru yang belum pernah dipelajari. Model yang telah dilatih ini kemudian dilanjutkan ke tahap evaluasi guna mengukur performa klasifikasinya memakai data pengujian.

3.3. Evaluasi Model

3.3.1. Confusion Matrix

Evaluasi model memakai data pengujian yang tidak dilibatkan sama sekali pada tahap pelatihan maupun validasi. Salah satu metode yang digunakan adalah Confusion Matrix, yang menggambarkan secara rinci kemampuan model mengklasifikasikan tiap kelas berdasarkan perbandingan antara label aktual dan label hasil prediksi.



Gambar 4. Confusion Matrix

Gambar 4 memperlihatkan bahwa mayoritas citra berhasil diklasifikasikan dengan tepat oleh model. Kelas normal mencatat hasil terbaik dengan 148 prediksi benar dari 150 data uji, sedangkan kelas nodul, pustula, dan papula masing-masing memperoleh 137, 122, dan 111 prediksi yang sesuai dengan label sebenarnya.

Meski demikian, sejumlah kesalahan klasifikasi masih dijumpai, terutama pada kelas papula yang kerap tertukar dengan nodul atau pustula. Hal ini diduga terjadi karena kemiripan karakteristik visual antarjenis jerawat inflamasi, sehingga model mengalami kesulitan memisahkan sampel-sampel dengan pola visual yang serupa.

3.3.2. Hasil Pengukuran Performa

Guna memperoleh gambaran performa yang lebih menyeluruh, dilakukan pula pengukuran dengan metrik precision, recall, F1-score, dan accuracy sebagaimana disajikan pada Tabel 4.

Tabel 4. Classification Report

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Nodule	0,77	0,91	0,83	150
Normal	0,98	0,99	0,98	150
Papule	0,87	0,74	0,80	150
Pustule	0,86	0,81	0,84	150

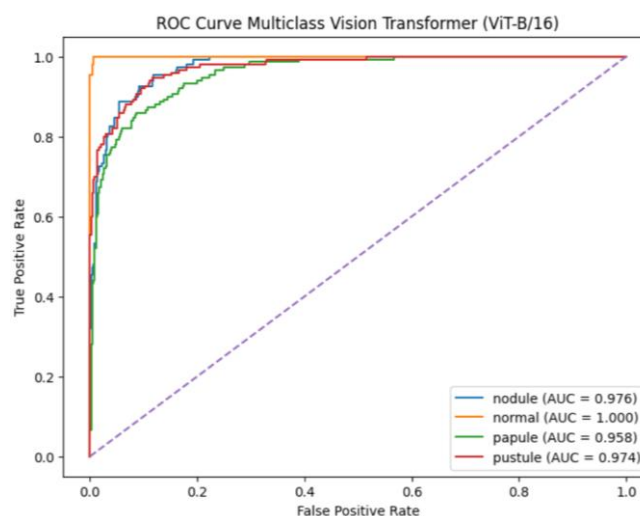
Akurasi	0,86	600
----------------	-------------	------------

Hasil evaluasi menunjukkan model memperoleh accuracy 86%, precision 87%, recall 86%, dan F1-score 86%, yang menegaskan bahwa Vision Transformer mampu mengklasifikasikan keempat kelas jerawat dengan kinerja yang baik.

Kelas normal mencatat F1-score tertinggi, yakni 98%, yang menandakan model mengenali karakteristik visual kulit normal secara konsisten. Sebaliknya, kelas papula memperoleh F1-score terendah sebesar 80%, sejalan dengan temuan pada Confusion Matrix di mana sejumlah citra papula masih terklasifikasi sebagai jenis jerawat inflamasi lain. Secara umum, hasil evaluasi ini menunjukkan model cukup andal dalam membedakan karakteristik visual antarkelas yang diteliti.

3.3.3. ROC-AUC

Selain Confusion Matrix dan Classification Report, evaluasi model juga memakai kurva Receiver Operating Characteristic (ROC) dan nilai Area Under Curve (AUC) untuk mengukur kemampuan model membedakan tiap kelas berdasarkan probabilitas prediksi yang dihasilkan.



Gambar 5. Kurva ROC-AUC

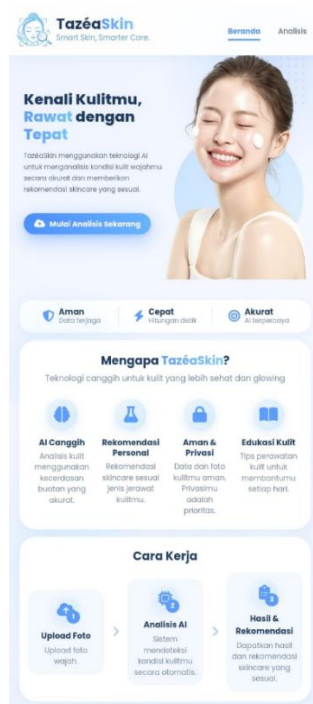
Gambar 5 menunjukkan seluruh kurva ROC berada di atas garis diagonal dan cenderung mendekati sudut kiri atas, yang mengindikasikan model mampu membedakan karakteristik visual antarkelas dengan baik. Nilai AUC pada masing-masing kelas turut menunjukkan bahwa probabilitas prediksi yang dihasilkan model cukup konsisten dalam mengidentifikasi kelas normal, papula, pustula, dan nodul.

Temuan ini selaras dengan hasil evaluasi Confusion Matrix dan Classification Report yang menegaskan kinerja baik Vision Transformer dalam mengklasifikasikan jenis jerawat. Dengan demikian, evaluasi ROC-AUC semakin memperkuat bukti bahwa model memiliki kemampuan diskriminasi yang memadai antarkelas jerawat pada citra wajah.

3.4 Implementasi Sistem

Model Vision Transformer (ViT) yang telah melalui pelatihan dan evaluasi selanjutnya ditanamkan ke dalam website berbasis Flask yang dibangun dengan bahasa pemrograman Python. Implementasi ini bertujuan menghadirkan model klasifikasi dalam bentuk aplikasi yang dapat langsung dipakai pengguna. Melalui website tersebut, pengguna dapat mengunggah citra wajah dan memperoleh hasil identifikasi jenis jerawat secara otomatis, lengkap dengan rekomendasi kandungan skincare sesuai hasil klasifikasi.

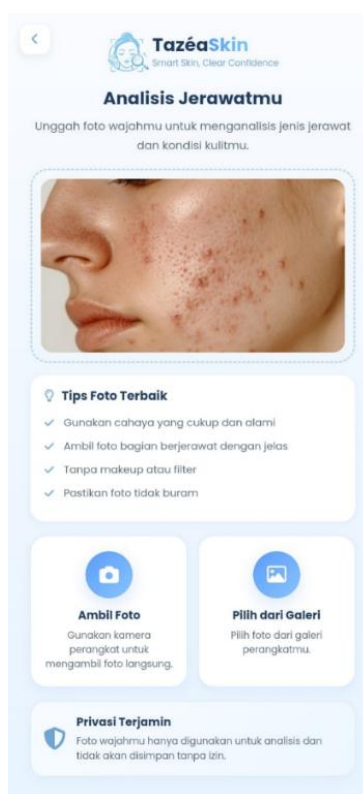
3.4.1. Halaman Utama



Gambar 6. Halaman Utama

Halaman utama merupakan tampilan pertama yang muncul saat pengguna membuka website TazéaSkin. Halaman ini memuat informasi tujuan pengembangan sistem, fitur-fitur yang tersedia, serta tahapan penggunaan website, sekaligus berfungsi memperkenalkan sistem kepada pengguna sebelum masuk ke proses klasifikasi. Tampilannya yang ringkas dan informatif membantu pengguna memahami fungsi website beserta langkah-langkah untuk memperoleh hasil identifikasi jenis jerawat.

3.4.2. Halaman Scan



Gambar 7. Halaman Scan

Halaman unggah citra menjadi sarana interaksi utama antara pengguna dan sistem, tempat pengguna memilih dan mengunggah citra wajah yang akan dianalisis model Vision Transformer. Setelah citra diunggah, sistem menjalankan serangkaian proses mulai dari pemuatan gambar, preprocessing sesuai kebutuhan model, ekstraksi fitur, hingga klasifikasi berdasarkan pola visual yang terdeteksi. Tahap ini penting karena kualitas citra yang diunggah turut menentukan hasil identifikasi sistem.

3.4.3. Halaman Hasil



Gambar 8. Halaman Hasil

Sistem menampilkan halaman hasil begitu proses klasifikasi rampung, berisi informasi jenis jerawat yang terdeteksi pada citra wajah berdasarkan probabilitas prediksi tertinggi dari model Vision Transformer. Selain hasil klasifikasi, website turut memberikan rekomendasi kandungan skincare yang disesuaikan dengan jenis jerawat yang teridentifikasi, sehingga pengguna memperoleh gambaran awal kondisi kulit wajah beserta pilihan perawatan yang relevan dengan hasil tersebut.

3.5 Pengujian Sistem

Pengujian sistem memakai metode Black Box Testing dengan pendekatan Functional Testing, yang difokuskan pada fungsi-fungsi utama website mulai dari navigasi halaman, unggah citra, pengambilan gambar lewat kamera, proses klasifikasi memakai model Vision Transformer, hingga penampilan hasil klasifikasi beserta rekomendasi skincare.

Tabel 5. Hasil Pengujian

No	Fitur	Skenario Pengujian	Yang Diharapkan	Hasil
1	Halaman Utama	Membuka website	Halaman utama berhasil ditampilkan dengan seluruh komponen antarmuka	✓
2	Navigasi	Memilih menu Analisis	Sistem mengarahkan pengguna ke halaman klasifikasi	✓
3	Tombol CTA	Menekan tombol "Mulai Analisis Sekarang"	Sistem mengarahkan pengguna ke halaman klasifikasi	✓
4	Halaman Scan	Membuka halaman scan	Halaman scan berhasil ditampilkan beserta form unggah citra dan petunjuk penggunaan	✓
5	Upload Citra	Mengunggah citra wajah	Sistem menerima dan menampilkan citra yang diunggah	✓
6	Kamera	Mengambil foto menggunakan kamera	Sistem berhasil mengakses kamera dan menampilkan hasil pengambilan gambar	✓
7	Proses Klasifikasi	Menjalankan analisis citra	Sistem melakukan klasifikasi jenis jerawat menggunakan model Vision Transformer	✓
8	Halaman Hasil	Menampilkan hasil klasifikasi	Sistem menampilkan jenis jerawat yang terdeteksi berdasarkan hasil prediksi model	✓
9	Tingkat Kepercayaan	Menampilkan nilai confidence	Sistem menampilkan persentase keyakinan hasil prediksi	✓
10	Rekomendasi Skincare	Menampilkan rekomendasi skincare	Sistem menampilkan rekomendasi kandungan skincare sesuai hasil klasifikasi	✓
11	Tips Perawatan	Menampilkan tips perawatan	Sistem menampilkan tips perawatan sesuai hasil klasifikasi	✓
12	Analisis Ulang	Menekan tombol "Analisis Ulang"	Sistem mengarahkan pengguna kembali ke halaman klasifikasi	✓

Hasil pengujian pada Tabel 5 menunjukkan seluruh fungsi yang diuji berjalan sesuai kebutuhan sistem: website mampu menerima citra dari pengguna, menjalankan klasifikasi jenis jerawat, menampilkan hasil prediksi beserta tingkat kepercayaannya, serta memberikan rekomendasi kandungan skincare dan tips perawatan yang sesuai. Dengan demikian, seluruh fitur utama website TazéaSkin terbukti berfungsi baik dan mendukung proses klasifikasi jenis jerawat secara otomatis

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa model Vision Transformer (ViT-B/16) berhasil diterapkan untuk mengklasifikasikan jenis jerawat pada citra wajah ke dalam empat kelas — normal, papula, pustula, dan nodul. Dengan pendekatan transfer learning dan dataset sebanyak 4.000 citra, model ini mencapai akurasi 86% pada data pengujian. Hasil evaluasi memperlihatkan kemampuan model yang baik dalam membedakan karakteristik visual tiap kelas, dengan performa terbaik pada kelas normal, sementara kesalahan klasifikasi masih muncul pada kelas papula akibat kemiripan karakteristik visualnya dengan jenis jerawat inflamasi lain.

Model yang dikembangkan juga berhasil diintegrasikan ke dalam aplikasi berbasis web memakai framework Flask. Hasil pengujian Black Box Testing menunjukkan seluruh fungsi utama sistem — mulai dari unggah citra, klasifikasi, hingga penyajian hasil identifikasi beserta rekomendasi kandungan skincare — berjalan sesuai harapan. Dengan demikian, sistem ini terbukti mampu mendukung identifikasi awal jenis jerawat secara otomatis melalui platform berbasis web.

DAFTAR PUSTAKA

- [1] F. H. Rizki Khaerunisa, "Experience dan Persepsi Perempuan Terhadap Dampak Kesehatan Kulit dalam Keputusan Penggunaan Produk Skincare," *SOSIOGLOBAL J. Pemikir. dan Penelit. Sociol.*, vol. 9, no. 2, pp. 113–132, 2025.
- [2] T. D. Siahaan *et al.*, "Hubungan Antara Kejadian Acne Vulgaris dengan Harga Diri Remaja," *J. Mutiara Ners*, vol. 3, no. 1, pp. 15–21, 2020.
- [3] H. T. Sibero *et al.*, "Prevalensi dan Gambaran Epidemiologi Akne Vulgaris di Provinsi Lampung," *JK Unila (Jurnal Kedokt. Univ. Lampung)*, vol. 3, 2019.
- [4] V. F. Sundoro, T. Djannatun, and E. D. Maharsi, "Hubungan Personal Hygiene Wajah Terhadap Keparahan Acne Vulgaris Pada Remaja SMA Negeri 3 Jakarta," *Cerdika J. Ilm. Indones.*, vol. 4, no. September, pp. 753–765, 2024.
- [5] D. T. Aryani and W. Riyaningrum, "Hubungan Acne Vulgaris (AV) dengan Kepercayaan Diri pada Mahasiswa Muhammadiyah Purwokerto Angkatan 2021," *J. Kesehat. Tambusai*, vol. 3, no. September, pp. 434–441, 2022.
- [6] H. Mariatul Qibthiyah, Aris Fadillah, "Swamedikasi Acne Vulgaris di Kalangan Mahasiswa," *J. Kesehat. Tambusai*, vol. 5, pp. 12666–12676, 2024.
- [7] T. Susilowati and N. Pratiwi, "Sistem Deteksi Otomatis Jenis Jerawat Berbasis Convolution Neural Network (CNN) dan Framework Flask," *J. Artif. Intell. Digit. Bus.*, vol. 5, no. 1, pp. 4413–4420, 2026.
- [8] M. Nurkhasanah, "Klasifikasi Penyakit Kulit Wajah Menggunakan Metode Convolutional Neural Network," *SAINTEKS*, vol. 18, no. 2, pp. 183–190, 2021.
- [9] F. Ramadhana, G. A. Mahardika, N. Day, and E. Y. Puspaningrum, "Klasifikasi Penyakit Daun Cabai Menggunakan Vision Transformer," *Semin. Nas. Inform. Bela Negara*, vol. 5, pp. 131–135, 2025.
- [10] J. A. Figo, N. Yudistira, and A. W. Widodo, "Deteksi Covid-19 dari Citra X-ray menggunakan Vision Transformer," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1116–1125, 2023.
- [11] A. Hasbi and B. Tumanggor, "Implementation of the Vision Transformer (ViT) Method in Soil Type Classification," *J. Ilm. Sist. Inf.*, vol. 5, no. 1, pp. 574–583, 2026.
- [12] S. S. Futri, Dila Aura, Ivana Lucia Kharisma, "Implementasi Model Vision Transformer pada Klasifikasi Jenis Kulit Wajah Berbasis Website," *J. CoSciTech (Comput. Sci. Inf. Technol.)*, vol. 6, no. 2, pp. 214–229, 2025.
- [13] K. Citra, P. Daun, T. Menggunakan, M. Convolutional, and N. Network, "Klasifikasi Citra Penyakit Daun Tomat Menggunakan Metode Convolutional Neural Network (CNN) Dengan Arsitektur VGG-19," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 6, no. 3, pp. 405–413, 2025.
- [14] S. M. Reyvan Revolusioner Ar, Agusriyati, "Pendeteksian Dini Stunting pada Balita Menggunakan Vision Transformer (ViT)," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 3, 2025.
- [15] A. H. R. Mochammad Fajar Fadhillah, "Implementasi Convolutional Neural Network (CNN) Untuk Klasifikasi Tanaman Hias Berbasis Aplikasi Web Laravel," *J. IPSIKOM*, vol. 13, no. 2, 2025.
- [16] Y. F. Achmad *et al.*, "Identifikasi Jenis Jerawat Berdasarkan Tekstur Menggunakan GLCM dan Backpropagation," *SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 20, no. 2, 2021.
- [17] H. Aji, A. Kautsar, W. Bismi, D. Novianti, and M. Qommarudin, "Deteksi Jenis Jerawat Berbasis Android menggunakan Ensemble Deep Learning dengan Optimasi Layer Parsial," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 7, no. 2, pp. 260–267, 2025.
- [18] L. Watef and F. Mahanto, "Deteksi dan Visualisasi Berbasis Computer Vision untuk Analisis Gambar Dermatologis dalam Penilaian Keparahan Jerawat," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 6, no. 1, pp. 45–53, 2024.
- [19] R. A. Kumalasanti, C. Rooyen, and M. Christy, "Deteksi Ekspresi Wajah Menggunakan Visual Geometry Group - 16 Layer Convolutional Neural Network," *Konstelasi Konvergensi Tek. dan Sist. Inf.*, vol. 101, pp. 128–135, 2026.
- [20] M. Nurhidayanti, "Penerapan Deep Learning dalam Pengenalan Wajah untuk Sistem Keamanan," *J. Inform. Indones.*, vol. 1, no. 1, pp. 29–37, 2025.
- [21] R. S. Ulya Qistina, Febriyani, Zainab al-Kubra, "Rumusan Masalah Sebagai Kompas Ilmiah: Menavigasi Identifikasi, Batasan, dan Pertanyaan Penelitian," *J. Rumusan*, 2026.
- [22] R. L. Hasanah and M. Hasan, "Deteksi Lesi Acne Vulgaris Pada Citra Jerawat Wajah Menggunakan Metode K-Means Clustering," *Indones. J. Softw. Eng.*, vol. 8, no. 1, pp. 46–51, 2022.
- [23] C. Dewi, A. Suyitno, and E. Pujiastuti, "Studi Literatur : Model Pembelajaran Blended Learning dalam Meningkatkan Kemampuan Berpikir Kreatif dan Rasa Ingin Tahu Siswa dalam Pembelajaran Matematika," *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 272–281, 2022.
- [24] G. Ekayanda and M. Rahardi, "Analysis of Deep Learning Algorithms Using ConvNeXt and Vision Transformer for Brain Tumor Disease," *J. Appl. Informatics Comput.*, vol. 9, no. 6, 2025.
- [25] X. Zhai *et al.*, "An Image Is Worth 16X16 Words: Transformers For Image Recognition At Scale," *ICLR 2021 (International Conf. Learn. Represent.)*, 2021.