

Pengolahan Data Skin Face Menggunakan Tools Rapidminer dengan Metode Data Mining

Noferina Harefa ^{a,1}, Kasih Setia Gaho ^{a,2}, Khairi Darmawan ^{a,3}, Fariz Raditya ^{a,4}, Dikdik Permana Wigandi ^{a,5}

^a Universitas Bina Sarana Informatika, Jakarta, 10450, Indonesia

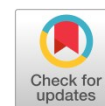
¹ noferinaharefa@gmail.com; ² kasihsetia@gmail.com; ³ farizraditya2802@gmail.com; ⁴ khairidarmawan25@gmail.com;

⁵ dikdik.dkn@bsi.ac.id

Diterima 13 Juni 2026; Direvisi 27 Oktober 2026; Disetujui 28 Oktober 2026

ABSTRAK

Perawatan kulit yang tepat memerlukan identifikasi tipe kulit yang akurat untuk membantu menentukan produk dan metode perawatan yang sesuai. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi tipe kulit wajah menggunakan pendekatan kecerdasan buatan berbasis pengolahan citra digital dan machine learning. Dataset yang digunakan terdiri dari 2.813 data dengan kategori tipe kulit meliputi normal, kering (dry), berminyak (oily), dan berjerawat (acne). Proses penelitian meliputi pengumpulan data, pra-pemrosesan citra, ekstraksi fitur, pelatihan model, serta evaluasi kinerja sistem. Sistem dirancang untuk menganalisis karakteristik kulit wajah secara otomatis dan menghasilkan prediksi tipe kulit secara cepat dan objektif. Hasil penelitian menunjukkan bahwa metode yang diterapkan mampu melakukan klasifikasi tipe kulit dengan baik sehingga dapat digunakan sebagai alat bantu dalam analisis kondisi kulit wajah. Sistem yang dikembangkan diharapkan dapat mendukung layanan konsultasi kecantikan digital dan membantu pengguna dalam memilih perawatan kulit yang sesuai dengan kondisi kulitnya.

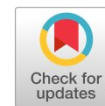


KATA KUNCI

tipe kulit wajah,
pengolahan citra digital,
machine learning,
klasifikasi,
kecerdasan buatan.

ABSTRACT

Proper skin care requires accurate skin type identification to determine suitable skincare products and treatment methods. This study aims to develop a facial skin type classification system using an artificial intelligence approach based on digital image processing and machine learning. The dataset used consists of 2,813 samples categorized into four skin types: normal, dry, oily, and acne-prone skin. The research process includes data collection, image preprocessing, feature extraction, model training, and system performance evaluation. The proposed system is designed to automatically analyze facial skin characteristics and provide fast and objective skin type predictions. The results indicate that the implemented method is capable of classifying skin types effectively and can be utilized as a supporting tool for facial skin condition analysis. The developed system is expected to support digital beauty consultation services and assist users in selecting skincare treatments that match their skin conditions.



KEYWORD

facial skin type,
digital image processing,
machine learning,
classification,
artificial intelligence.



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

1. Pendahuluan

Kulit merupakan organ terbesar pada tubuh manusia yang berfungsi sebagai pelindung terhadap berbagai faktor eksternal, seperti paparan sinar ultraviolet, polusi, mikroorganisme, serta perubahan lingkungan. Kondisi kulit yang sehat berperan penting dalam menjaga kesehatan dan meningkatkan

kepercayaan diri seseorang. Oleh karena itu, identifikasi tipe kulit menjadi langkah awal yang penting dalam menentukan produk dan metode perawatan kulit yang sesuai. Tipe kulit wajah umumnya diklasifikasikan menjadi beberapa kategori, seperti normal, kering (*dry*), berminyak (*oily*), dan berjerawat (*acne-prone*), yang masing-masing memiliki karakteristik dan kebutuhan perawatan yang berbeda. [1], [2]

Perkembangan industri kecantikan dan perawatan kulit dalam beberapa tahun terakhir menunjukkan peningkatan yang signifikan. Banyaknya produk perawatan kulit yang tersedia di pasaran membuat pengguna sering mengalami kesulitan dalam memilih produk yang sesuai dengan kondisi kulitnya. Kesalahan dalam pemilihan produk dapat menyebabkan iritasi, alergi, atau bahkan memperburuk kondisi kulit yang ada. Oleh karena itu, diperlukan suatu metode yang mampu mengidentifikasi tipe kulit secara cepat, akurat, dan mudah diakses oleh masyarakat. [3], [4]

Kemajuan teknologi kecerdasan buatan (*Artificial Intelligence / AI*), khususnya pada bidang *Machine Learning* dan *Deep Learning*, telah membuka peluang baru dalam pengembangan sistem analisis citra digital. Teknologi ini memungkinkan komputer untuk mengenali pola visual pada gambar dan melakukan klasifikasi secara otomatis. Berbagai penelitian menunjukkan bahwa metode berbasis *Convolutional Neural Network* (CNN) mampu memberikan performa yang baik dalam tugas klasifikasi citra, termasuk pada bidang dermatologi dan analisis kondisi kulit. [5], [6]

Penerapan AI dalam bidang dermatologi dan kesehatan kulit terus berkembang karena mampu membantu proses analisis yang sebelumnya hanya dapat dilakukan oleh tenaga ahli. Sistem berbasis pengolahan citra digital memungkinkan identifikasi karakteristik kulit melalui foto wajah yang diambil menggunakan kamera digital atau perangkat seluler. Pendekatan ini dinilai lebih praktis, efisien, dan memiliki potensi untuk meningkatkan akses masyarakat terhadap layanan konsultasi kulit secara digital. [7], [8]

Beberapa penelitian terdahulu telah mengembangkan sistem klasifikasi tipe kulit menggunakan metode *Deep Learning*. Penelitian yang dilakukan oleh Kumala dkk. membandingkan penggunaan CNN, MobileNetV3, dan Segment Anything Model (SAM) untuk klasifikasi tipe kulit wajah dan menunjukkan bahwa model berbasis *Deep Learning* mampu menghasilkan performa klasifikasi yang baik pada berbagai kategori tipe kulit. [9], [10]

Selain itu, penelitian oleh Ahmad dkk. mengembangkan metode klasifikasi tipe kulit manusia menggunakan pengolahan citra dan *Deep Learning*. Hasil penelitian tersebut menunjukkan bahwa karakteristik visual pada citra wajah dapat dimanfaatkan untuk membedakan tipe kulit secara otomatis sehingga membantu pengguna dalam menentukan perawatan kulit yang sesuai.

Meskipun demikian, pengembangan sistem klasifikasi kulit masih menghadapi berbagai tantangan, seperti variasi pencahayaan saat pengambilan gambar, perbedaan warna kulit, kualitas citra, serta keberagaman dataset yang digunakan. Representasi berbagai jenis kulit dalam dataset menjadi faktor penting untuk memastikan model mampu bekerja secara konsisten pada berbagai kelompok pengguna.

Dalam penelitian ini dikembangkan sebuah sistem klasifikasi tipe kulit wajah menggunakan teknik *Machine Learning* berbasis citra digital. Dataset yang digunakan terdiri dari 2.813 data citra kulit dengan kategori normal, kering, berminyak, dan berjerawat. Sistem dirancang untuk melakukan proses prapemrosesan data, pelatihan model, dan klasifikasi otomatis terhadap citra wajah yang diberikan

pengguna. Dengan adanya sistem ini diharapkan proses identifikasi tipe kulit dapat dilakukan secara lebih cepat, objektif, dan mudah diakses dibandingkan metode konvensional.

Penelitian ini bertujuan untuk membangun model klasifikasi tipe kulit wajah yang mampu membantu pengguna dalam mengenali kondisi kulitnya secara otomatis melalui analisis citra digital. Hasil penelitian diharapkan dapat menjadi dasar pengembangan aplikasi konsultasi kecantikan digital yang lebih efektif serta mendukung pemanfaatan teknologi kecerdasan buatan dalam bidang kesehatan dan perawatan kulit.

2. Tinjauan Pusaka

2.1 Kulit Wajah dan Tipe Kulit

Kulit merupakan organ terluar tubuh manusia yang berfungsi sebagai pelindung terhadap berbagai faktor lingkungan, seperti sinar matahari, polusi, debu, dan mikroorganisme. Kondisi kulit yang sehat sangat penting untuk menjaga kesehatan serta meningkatkan rasa percaya diri seseorang.

Secara umum, tipe kulit wajah dibedakan menjadi empat kategori utama, yaitu **normal**, **kering (dry)**, **berminyak (oily)**, dan **berjerawat (acne)**. Kulit normal memiliki keseimbangan kadar minyak dan kelembapan yang baik. Kulit kering ditandai dengan kurangnya kadar air dan minyak sehingga terasa kasar. Kulit berminyak menghasilkan sebum berlebih yang menyebabkan wajah tampak mengilap, sedangkan kulit berjerawat ditandai dengan munculnya jerawat akibat penyumbatan pori-pori dan peradangan.

2.2 Pengolahan Citra Digital

Pengolahan citra digital adalah teknik yang digunakan untuk memperoleh, memperbaiki, dan menganalisis gambar digital sehingga informasi yang terkandung di dalamnya dapat dikenali oleh sistem komputer. Teknologi ini banyak digunakan dalam berbagai bidang, seperti kesehatan, keamanan, industri, dan kecerdasan buatan.

Pada penelitian klasifikasi tipe kulit, pengolahan citra digunakan untuk mengidentifikasi karakteristik kulit wajah berdasarkan gambar yang diambil menggunakan kamera. Melalui proses tertentu, sistem dapat mengenali pola dan fitur yang membedakan satu tipe kulit dengan tipe kulit lainnya.

2.3 Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan yang memungkinkan komputer mempelajari pola dari data dan membuat keputusan secara otomatis tanpa harus diprogram secara rinci. Metode ini bekerja dengan memanfaatkan data pelatihan untuk membangun model yang mampu melakukan prediksi atau klasifikasi terhadap data baru.

Dalam penelitian ini, *Machine Learning* digunakan untuk mengenali pola pada citra wajah dan mengelompokkan gambar ke dalam kategori tipe kulit yang sesuai. Semakin banyak dan beragam data yang digunakan, maka semakin baik kemampuan model dalam melakukan klasifikasi.

2.4 Deep Learning

Deep Learning adalah pengembangan dari *Machine Learning* yang menggunakan jaringan saraf tiruan dengan banyak lapisan untuk mempelajari data yang kompleks. Teknologi ini memiliki kemampuan yang sangat baik dalam mengenali pola visual pada gambar sehingga banyak digunakan dalam bidang pengolahan citra.

Keunggulan *Deep Learning* terletak pada kemampuannya melakukan ekstraksi fitur secara otomatis tanpa memerlukan proses identifikasi fitur secara manual. Dengan demikian, sistem dapat mempelajari karakteristik citra secara lebih efektif dan menghasilkan tingkat akurasi yang lebih tinggi.

2.5 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan salah satu metode *Deep Learning* yang dirancang khusus untuk mengolah data berbentuk citra. CNN bekerja dengan melakukan proses konvolusi untuk mengekstraksi fitur-fitur penting dari gambar, seperti tekstur, warna, dan pola tertentu.

Pada klasifikasi tipe kulit wajah, CNN mampu mengenali perbedaan karakteristik antara kulit normal, kering, berminyak, dan berjerawat berdasarkan pola visual yang terdapat pada citra wajah. Oleh karena itu, CNN menjadi salah satu metode yang banyak digunakan dalam penelitian klasifikasi citra.

2.6 Sistem Klasifikasi Tipe Kulit Wajah

Sistem klasifikasi tipe kulit wajah merupakan sistem yang dirancang untuk mengidentifikasi jenis kulit secara otomatis berdasarkan citra wajah yang diberikan pengguna. Sistem ini memanfaatkan teknologi pengolahan citra dan kecerdasan buatan untuk menganalisis kondisi kulit dan menghasilkan prediksi tipe kulit.

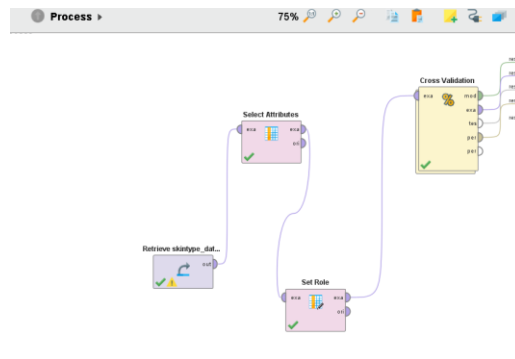
Proses klasifikasi dimulai dari pengambilan citra wajah, kemudian dilakukan prapemrosesan untuk meningkatkan kualitas gambar. Selanjutnya citra diproses oleh model klasifikasi untuk menentukan kategori tipe kulit yang paling sesuai. Hasil klasifikasi dapat digunakan sebagai informasi awal dalam menentukan produk dan perawatan kulit yang tepat.

2.7 Kerangka Pemikiran

Penelitian ini mengembangkan sistem klasifikasi tipe kulit wajah berbasis citra digital menggunakan metode *Machine Learning*. Data citra wajah yang diperoleh akan melalui tahap prapemrosesan, pelatihan model, dan proses klasifikasi. Sistem kemudian mengelompokkan citra ke dalam salah satu kategori tipe kulit, yaitu *Normal*, *Dry*, *Oily*, atau *Acne*. Hasil klasifikasi diharapkan dapat membantu pengguna mengenali kondisi kulit secara cepat dan objektif sehingga dapat menentukan perawatan yang sesuai.

3. Pengolahan data

\



Gambar 1. Import Library

Gambar 1 tersebut menunjukkan alur proses pengolahan data dan evaluasi model klasifikasi tipe kulit (*Skin Type Classification*) menggunakan RapidMiner. Setiap operator memiliki fungsi sebagai berikut:

1. *Retrieve skintype_data*
 - Digunakan untuk mengambil atau memanggil dataset tipe kulit yang telah disimpan pada repository RapidMiner.
 - Dataset ini menjadi sumber data utama yang akan diproses.
2. *Select Attributes*
 - Berfungsi untuk memilih atribut atau variabel yang akan digunakan dalam proses klasifikasi.
 - Atribut yang tidak diperlukan dapat dihapus agar model lebih fokus pada data yang relevan.
3. *Set Role*
 - Digunakan untuk menentukan peran atribut dalam dataset.
 - Biasanya atribut target (misalnya *Skin Type*: Normal, Dry, Oily, Acne) ditetapkan sebagai *Label*, sedangkan atribut lainnya menjadi fitur prediktor.
4. *Cross Validation*
 - Digunakan untuk menguji dan mengevaluasi performa model klasifikasi.
 - Dataset dibagi menjadi beberapa bagian (*fold*), kemudian dilakukan proses pelatihan (*training*) dan pengujian (*testing*) secara bergantian.
 - Metode ini membantu mengukur kemampuan model dalam memprediksi data baru dan mengurangi risiko *overfitting*.
5. *Results (res)*
 - Menampilkan hasil evaluasi model seperti:
 - Accuracy
 - Precision
 - Recall
 - Confusion Matrix
 - Performance Vector
 - Hasil ini digunakan untuk mengetahui tingkat keberhasilan model dalam mengklasifikasikan tipe kulit.

Proses pada gambar menunjukkan tahapan persiapan data, penentuan label, dan evaluasi model klasifikasi tipe kulit menggunakan metode *Cross Validation di RapidMiner* untuk memperoleh performa model yang akurat dan dapat dipercaya.

3.1 Uji Performansi

PerformanceVector (Performance) ExampleSet (Set Role) Tree (Decision Tree)

Table View Plot View

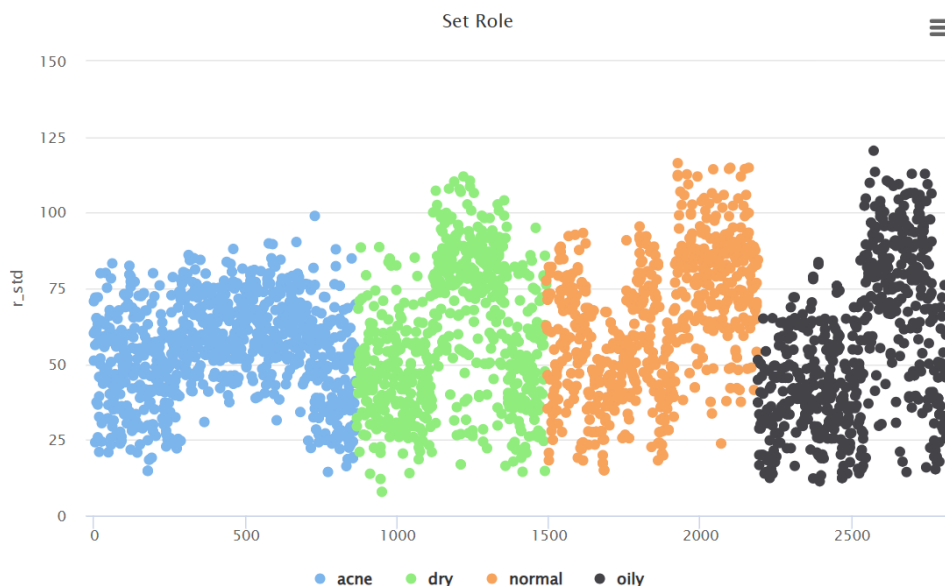
kappa: 0.020 +/- 0.011 (micro average: 0.020)

	true acne	true dry	true normal	true oily	class precision
pred. acne	867	621	675	605	31.32%
pred. dry	0	0	1	0	0.00%
pred. normal	0	3	21	1	84.00%
pred. oily	1	1	1	16	84.21%
class recall	99.88%	0.00%	3.01%	2.57%	

Gambar 2. Performance

Berdasarkan hasil evaluasi menggunakan metode *Cross Validation*, model *Decision Tree* menghasilkan nilai Kappa sebesar 0,020 yang menunjukkan tingkat kesesuaian prediksi yang sangat rendah. *Confusion Matrix* memperlihatkan bahwa model cenderung mengklasifikasikan sebagian besar data ke dalam kelas *Acne* sehingga kemampuan identifikasi kelas *Dry*, *Normal*, dan *Oily* masih sangat rendah. Oleh karena itu, diperlukan perbaikan pada tahap prapemrosesan data, penyeimbangan dataset, pemilihan atribut, atau penggunaan algoritma klasifikasi lain untuk meningkatkan performa model.

3.2 Uji Set Role



Gambar 3. Set Role

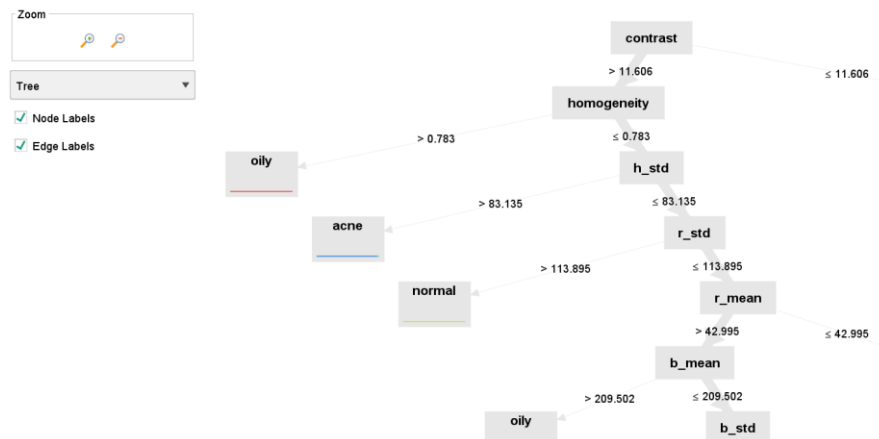
Gambar 3 adalah *Scatter Plot* (Diagram Tebar) yang menunjukkan sebaran nilai bernama *r_std* untuk 4 jenis kulit berbeda.

Berikut adalah ringkasan poin utamanya:

- Sumbu Y (r_std): Kemungkinan besar adalah indikator variasi warna atau tekstur kulit (*Standard Deviation of Red channel*).
- Sumbu X: Jumlah sampel data (total sekitar 2800 data).
- Tujuan Grafik: Menunjukkan bahwa setiap jenis kulit memiliki pola nilai yang berbeda-beda:
 - **acne (Biru)**: Nilainya paling stabil dan berkumpul di rentang menengah (**35–75**).
 - **dry (Hijau)**: Polanya bergelombang dan sempat melonjak hingga nilai **110**.
 - **normal (Oranye)**: Fluktuasinya tinggi, menyebar dari nilai rendah hingga melonjak ke **115**.
 - **oily (Hitam)**: Memiliki titik data tertinggi di antara semua kelompok, melonjak hingga mendekati **120**.

Maka, grafi ini membuktikan bahwa nilai r_std berbeda untuk setiap jenis kulit, sehingga variabel ini cocok digunakan untuk mendeteksi atau mengklasifikasikan jenis kulit (misalnya dalam sistem *Machine Learning*).

3.3 Uji *Decision Tree*



Gambar 4. Decision Tree

Gambar 4 menampilkan sebagian struktur pohon keputusan hasil pemodelan Decision Tree. Simpul akar pohon menggunakan atribut *contrast* dengan nilai pemisah 11,606. Data dengan nilai *contrast* lebih dari 11,606 diarahkan ke simpul *homogeneity*. Pada simpul ini, data dengan nilai *homogeneity* lebih dari 0,783 diklasifikasikan ke dalam kelas *oily*, sedangkan data dengan nilai kurang dari atau sama dengan 0,783 diarahkan ke simpul *h_std*.

Pada simpul *h_std*, data dengan nilai lebih dari 83,135 diklasifikasikan ke dalam kelas *acne*, sedangkan data dengan nilai kurang dari atau sama dengan 83,135 diteruskan ke simpul *r_std*. Data dengan nilai *r_std* lebih dari 113,895 diklasifikasikan ke dalam kelas *normal*, sedangkan data dengan nilai kurang dari atau sama dengan 113,895 diteruskan ke simpul *r_mean*. Pada simpul *r_mean*, data dengan nilai lebih dari 42,995 diarahkan ke simpul *b_mean*. Data dengan nilai *b_mean* lebih dari 209,502 diklasifikasikan ke dalam kelas *oily*, sedangkan data dengan nilai kurang dari atau sama dengan 209,502 diteruskan ke simpul *b_std*.

Cabang *contrast* dengan nilai kurang dari atau sama dengan 11,606, cabang *r_mean* dengan nilai kurang dari atau sama dengan 42,995, serta cabang lanjutan dari simpul *b_std* berlanjut ke simpul berikutnya yang tidak tampak pada potongan gambar. Secara keseluruhan, atribut yang muncul pada bagian pohon ini adalah *contrast*, *homogeneity*, *b_std*, *r_std*, *r_mean*, *b_mean*, dan *b_std*. Kelas yang dihasilkan pada simpul daun yang terlihat adalah *oily*, *acne*, dan *normal*

Contoh alur ringkas:

Jika *Contrast* > 11.606 DAN *Homogeneity* > 0.783 , maka sistem langsung memutuskan kulit tersebut adalah *oily*. Jika tidak memenuhi, sistem akan turun ke bawah untuk memeriksa nilai *h_std* guna mendeteksi apakah kulit tersebut *acne*, dan seterusnya.

Table 1 Hasil Klasifikasi dengan RapidMiner

Model	Classification Error	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naive Bayes	0,6	0,0	118,0	82996,0	56,9	127,1
Generalized Linear Model	0,6	0,0	112,0	62548,0	373,6	107,6
Logistic Regression	0,8	0,0	-136,0	101454,0	606,8	1328,0
Fast Large Margin	0,6	0,0	126,0	78330,0	761,5	719,1
Deep Learning	0,6	0,0	218,0	25446,0	783,5	231,1
Decision Tree	0,7	0,0	-74,0	24750,0	200,1	214,2
Random Forest	0,7	0,0	40,0	180862,0	611,8	1703,1
Gradient Boosted Trees	0,5	0,0	254,0	195323,0	963,0	995,6
Support Vector Machine	0,7	0,0	0,0	373894,0	2101,7	3878,2

Pengujian dilakukan terhadap sembilan algoritma klasifikasi pada platform RapidMiner Studio, yaitu Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees, dan Support Vector Machine. Setiap model dievaluasi berdasarkan *classification error*, standar deviasi, nilai *gains*, serta waktu komputasi yang meliputi total waktu, waktu pelatihan per 1.000 baris data, dan waktu penilaian (*scoring*) per 1.000 baris data.

Berdasarkan nilai *classification error*, Gradient Boosted Trees menunjukkan kinerja terbaik dengan tingkat kesalahan 0,5 atau akurasi sekitar 50%. Naive Bayes, Generalized Linear Model, Fast Large Margin, dan Deep Learning berada pada kelompok berikutnya dengan *classification error* sebesar 0,6. Decision Tree, Random Forest, dan Support Vector Machine menghasilkan *classification error* sebesar 0,7, sedangkan Logistic Regression memiliki kinerja terendah dengan *classification error* sebesar 0,8. Nilai standar deviasi seluruh model tercatat 0,0, yang menunjukkan bahwa variasi kinerja antarpengujian sangat kecil pada tingkat ketelitian satu angka desimal. Karena *classification error* juga ditampilkan dalam satu angka desimal, perbedaan kinerja di antara model yang memiliki nilai sama, misalnya keempat model dengan *classification error* 0,6, belum dapat dibedakan secara tegas.

Nilai *gains* memperkuat temuan tersebut. Gradient Boosted Trees memperoleh nilai *gains* tertinggi sebesar 254, diikuti oleh Deep Learning sebesar 218, Fast Large Margin sebesar 126, Naive Bayes sebesar 118, dan Generalized Linear Model sebesar 112. Sebaliknya, Logistic Regression dan Decision Tree menghasilkan nilai *gains* negatif, masing-masing sebesar -136 dan -74, sedangkan Support Vector Machine memperoleh nilai 0. Nilai *gains* yang negatif atau nol menunjukkan bahwa model tersebut belum memberikan keuntungan prediksi dibandingkan dengan tebakan acak berdasarkan proporsi kelas.

Dari sisi efisiensi komputasi, Decision Tree dan Deep Learning memiliki total waktu proses tercepat, masing-masing sebesar 24.750 ms dan 25.446 ms. Naive Bayes mencatat waktu pelatihan paling singkat, yaitu 56,9 ms per 1.000 baris, sedangkan Generalized Linear Model memiliki waktu *scoring* tercepat, yaitu 107,6 ms per 1.000 baris. Support Vector Machine merupakan model dengan beban komputasi terbesar, dengan total waktu 373.894 ms, waktu pelatihan 2.101,7 ms, dan waktu *scoring* 3.878,2 ms per 1.000 baris. Gradient Boosted Trees, meskipun unggul dalam akurasi, membutuhkan total waktu yang relatif tinggi, yaitu 195.323 ms, dengan waktu pelatihan 963,0 ms per 1.000 baris.

Secara keseluruhan, Gradient Boosted Trees merupakan model dengan kinerja klasifikasi terbaik, sedangkan Deep Learning menawarkan keseimbangan yang baik antara kinerja dan waktu komputasi. Namun, tingkat kesalahan seluruh model masih tergolong tinggi. Sebagai pembandingan, model yang selalu menebak kelas terbanyak (Normal, 28,1%) akan memiliki *classification error* sekitar 0,72. Model terbaik hanya menurunkan tingkat kesalahan menjadi sekitar 0,5, bahkan Logistic Regression berkinerja lebih rendah daripada pembandingan tersebut. Hal ini mengindikasikan bahwa fitur warna dan tekstur global yang digunakan belum cukup representatif untuk membedakan keempat jenis kulit wajah secara akurat.

4. Hasil dan Analisis

Table 2 Parameter dan Keterangan

Parameter	Keterangan
Total gambar	2.485 gambar
Jumlah kelas label	4 kelas (Normal, Kering, Berminyak, Berjerawat)
Format file	HDF5 / RapidMiner Table (.rmhdf5table)
Ukuran file	459 KB
Format gambar sumber	.jpg dan .jpeg

Berdasarkan Tabel 2 dataset yang digunakan dalam penelitian ini terdiri atas 2.485 gambar wajah. Seluruh gambar terbagi ke dalam empat kelas label jenis kulit, yaitu Normal, Kering, Berminyak, dan Berjerawat. Gambar sumber berformat .jpg dan .jpeg. Hasil ekstraksi fiturnya disimpan dalam format HDF5 RapidMiner Table (.rmhdf5table) dengan ukuran file sebesar 459 KB. Ukuran file yang relatif kecil ini menunjukkan bahwa data telah berbentuk fitur numerik hasil ekstraksi, bukan citra mentah, sehingga lebih efisien untuk diolah dan dianalisis.

Table 3 Distribusi Kelas Berdasarkan Jenis Kulit

Kelas	Jumlah Gambar	Proporsi (%)	Keterangan
Normal	698	28,1%	Terbanyak
Kering	625	25,1%	—
Berminyak	622	25,0%	—
Berjerawat	540	21,7%	Tersedikit
Total	2.485	100%	—

Dapat dilihat pada Tabel 3 distribusi data pada setiap kelas jenis kulit menunjukkan bahwa kelas Normal memiliki jumlah gambar terbanyak, yaitu 698 gambar (28,1%). Kelas Kering dan Berminyak berjumlah hampir sama, masing-masing 625 gambar (25,1%) dan 622 gambar (25,0%). Kelas Berjerawat memiliki jumlah gambar paling sedikit, yaitu 540 gambar (21,7%). Selisih antara kelas terbanyak dan tersedikit hanya 158 gambar, dengan rasio sekitar 1,3 : 1, sehingga distribusi kelas dapat dikategorikan relatif seimbang. Kondisi ini cukup baik untuk pelatihan model klasifikasi karena tidak ada kelas yang mendominasi secara ekstrem, sehingga risiko model cenderung memprediksi kelas mayoritas dapat diminimalkan. Meskipun demikian, kelas Berjerawat sebagai kelas minoritas tetap perlu diperhatikan saat mengevaluasi kinerja model, misalnya dengan melihat nilai *precision* dan *recall* pada setiap kelas, bukan hanya akurasi keseluruhan.

Table 4 Dataset berdasarkan Kelompok Fitur

Kelompok Fitur	Nama Kolom	Deskripsi
Warna RGB	r_mean, r_std	Rata-rata dan standar deviasi kanal warna Merah (Red)
	g_mean, g_std	Rata-rata dan standar deviasi kanal warna Hijau (Green)
	b_mean, b_std	Rata-rata dan standar deviasi kanal warna Biru (Blue)
Warna HSV	h_mean, h_std	Rata-rata dan standar deviasi kanal Hue (warna dominan)
	s_mean, s_std	Rata-rata dan standar deviasi kanal Saturation (kejernihan warna)
	v_mean, v_std	Rata-rata dan standar deviasi kanal Value (kecerahan)
Tekstur GLCM	energy	Energi tekstur dari Gray-Level Co-occurrence Matrix
	homogeneity	Homogenitas tekstur permukaan kulit
	correlation	Korelasi antar piksel pada tekstur
	contrast	Kontras tekstur gambar
Lain-lain	brightness	Keccerahan keseluruhan gambar
	acne	Indikator kondisi jerawat pada gambar
Identitas	filename	Nama file gambar sumber (.jpg/.jpeg)

label	Kelas jenis kulit (target klasifikasi)
-------	--

Dataset yang digunakan dalam penelitian ini terdiri atas 18 fitur numerik hasil ekstraksi citra yang berperan sebagai variabel prediktor, serta dua atribut identitas. Fitur-fitur tersebut dikelompokkan ke dalam empat kategori, yaitu fitur warna pada ruang warna RGB, fitur warna pada ruang warna HSV, fitur tekstur berbasis *Gray-Level Co-occurrence Matrix* (GLCM), dan fitur tambahan yang berkaitan dengan karakteristik visual kulit wajah.

Fitur warna RGB diekstraksi dari tiga kanal warna, yaitu merah (*red*), hijau (*green*), dan biru (*blue*). Setiap kanal direpresentasikan oleh dua parameter statistik, yakni nilai rata-rata dan standar deviasi, sehingga diperoleh enam atribut, yaitu *r_mean*, *r_std*, *g_mean*, *g_std*, *b_mean*, dan *b_std*. Nilai rata-rata menggambarkan intensitas umum masing-masing kanal warna pada permukaan kulit, sedangkan standar deviasi menunjukkan tingkat variasi warna pada keseluruhan area citra. Kedua parameter ini digunakan untuk merepresentasikan perbedaan warna kulit antarkelas, seperti munculnya area kemerahan pada kulit berjerawat atau warna yang cenderung kusam pada kulit kering.

Fitur warna HSV diperoleh melalui konversi citra dari ruang warna RGB ke ruang warna *Hue*, *Saturation*, *Value*. Sama halnya dengan fitur RGB, setiap kanal HSV direpresentasikan oleh nilai rata-rata dan standar deviasi, sehingga diperoleh enam atribut, yaitu *h_mean*, *h_std*, *s_mean*, *s_std*, *v_mean*, dan *v_std*. Kanal *hue* merepresentasikan corak warna dominan, kanal *saturation* merepresentasikan tingkat kejenuhan warna, dan kanal *value* merepresentasikan tingkat kecerahan. Ruang warna HSV dipilih sebagai pelengkap RGB karena mampu memisahkan informasi kromatik dari informasi intensitas cahaya, sehingga lebih stabil terhadap variasi pencahayaan pada saat pengambilan gambar. Karakteristik tersebut relevan untuk mengenali kulit berminyak yang umumnya memantulkan cahaya lebih kuat dibandingkan jenis kulit lainnya.

Fitur tekstur diekstraksi menggunakan metode GLCM, yaitu matriks yang memetakan frekuensi kemunculan pasangan nilai keabuan piksel pada jarak dan orientasi tertentu. Dari matriks tersebut diturunkan empat atribut tekstur. Atribut *energy* mengukur tingkat keseragaman pola tekstur, sedangkan *homogeneity* menggambarkan tingkat kehalusan permukaan kulit. Atribut *correlation* menunjukkan derajat keterkaitan nilai keabuan antarpiksel yang bertetangga, sedangkan *contrast* mengukur besarnya perbedaan intensitas antarpiksel yang berdekatan. Keempat atribut ini digunakan untuk membedakan permukaan kulit yang halus dan rata dari permukaan kulit yang kasar, bersisik, maupun yang memiliki bintik jerawat.

Selain ketiga kelompok fitur tersebut, dataset juga memuat dua atribut tambahan, yaitu *brightness* dan *acne*. Atribut *brightness* menyatakan tingkat kecerahan keseluruhan citra, sedangkan atribut *acne* merupakan indikator kondisi jerawat pada citra. Kedua atribut ini memberikan informasi yang lebih spesifik mengenai kondisi permukaan kulit, terutama dalam membedakan kelas Berjerawat dari kelas lainnya.

Dua atribut lainnya dalam dataset berfungsi sebagai identitas data. Atribut *filename* menyimpan nama file citra sumber dengan format .jpg atau .jpeg, sehingga setiap baris data dapat ditelusuri kembali ke citra aslinya. Atribut *label* berisi kelas jenis kulit wajah, yaitu Normal, Kering, Berminyak, dan Berjerawat, yang ditetapkan sebagai variabel target dalam proses klasifikasi. Pada tahap pemodelan, atribut *filename*

tidak diikutsertakan sebagai fitur masukan karena tidak mengandung informasi visual yang relevan terhadap jenis kulit.

5. Kesimpulan

Dataset ini dirancang untuk melatih model *machine learning* berbasis klasifikasi yang mampu mengenali jenis kulit wajah secara otomatis dari citra. Fitur-fitur di dalamnya telah diekstraksi dari gambar asli, meliputi fitur warna pada ruang warna RGB dan HSV serta fitur tekstur berbasis *Gray Level Co-occurrence Matrix* (GLCM). Dengan demikian, dataset ini dapat langsung digunakan pada berbagai algoritma klasifikasi seperti Random Forest, *Support Vector Machine* (SVM), maupun *Neural Network*. Distribusi kelas dalam dataset juga relatif seimbang, tanpa ada kelas yang mendominasi secara ekstrem, sehingga risiko model menjadi bias terhadap kelas tertentu dapat diminimalkan. Dataset disimpan dalam format RapidMiner HDF5 Table (.rmhdf5table) yang kompatibel langsung dengan platform RapidMiner Studio, sehingga proses pemodelan dan analisis dapat dilakukan tanpa konversi data tambahan.

5.1 Saran

Untuk meningkatkan akurasi model, langkah selanjutnya adalah melakukan normalisasi fitur serta bereksperimen dengan algoritma klasifikasi lain, seperti Random Forest, Naive Bayes, dan *k-Nearest Neighbor* (k-NN). Selain itu, fitur citra perlu diperkaya, misalnya dengan menambahkan fitur tekstur multi-arah dan fitur lokal pada area kulit, agar representasi data lebih menggambarkan karakteristik setiap jenis kulit wajah. Seluruh hasil dan proses yang telah dilakukan akan menjadi dasar dalam penulisan artikel ilmiah serta bahan presentasi kelompok pada pertemuan berikutnya

Daftar Pustaka

- [1] S. Saiwaeo, S. Arwatchananukul, and L. Mungmai, "Heliyon Human skin type classification using image processing and deep learning approaches," *Heliyon*, vol. 9, no. 11, p. e21176, 2023, doi: 10.1016/j.heliyon.2023.e21176.
- [2] Nisrina Nur Kumala, "Performance Comparison of Facial Skin Type Classification Using the Segment Anything Model", doi: <https://doi.org/10.57152/malcom.v6i2.2604>.
- [3] N. Alipour, T. Burke, and J. Courtney, "Skin Type Diversity in Skin Lesion Datasets : A Review," *Curr. Dermatol. Rep.*, pp. 198–210, 2024, doi: 10.1007/s13671-024-00440-0.
- [4] P. Reinders, M. Augustin, P. Med, A. Fleyder, and M. Otten, "Exploring Acceptability , Barriers , and Facilitators for Digital Health in Dermatology : Qualitative Focus Groups With Dermatologists , Nurses , and Patients Corresponding Author :," vol. 7, pp. 1–16, 2024, doi: 10.2196/57172.
- [5] T. G. Debelee, "Skin Lesion Classification and Detection Using Machine Learning Techniques : A Systematic Review," 2023.
- [6] B. H. S. Aljohani, W. M. S. Yafooz, and A. Alsaeedi, "AI-driven dermatological diagnostics : A systematic review of machine and deep learning methods for skin disease classification," *Array*, vol. 30, no. December 2025, p. 100797, 2026, doi: 10.1016/j.array.2026.100797.
- [7] K. P. Venkatesh, M. M. Raza, G. Nickel, S. Wang, and J. C. Kvedar, "Deep learning models across the range of skin disease," pp. 2–3, 2024, doi: 10.1038/s41746-024-01033-8.

- [8] M. Groh, L. Soenksen, F. Lau, R. Han, A. Kim, and O. Badri, "with the Fitzpatrick 17k Dataset".
- [9] A. Lucchese *et al.*, "Correlation between Oral Lichen Planus and Viral Infections Other Than HCV : A Systematic Review," 2022.
- [10] A. Esteva, K. Chou, S. Yeung, N. Naik, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision," *npj Digit. Med.*, pp. 1–9, 2021, doi: 10.1038/s41746-020-00376-2.