Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 19817-19830
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Klasifikasi Keparahan Acne Vulgaris Menggunakan LBP, HSV, dan Random Forest

Arjun Septian Amarta[1], M. Akbar Khadafi A[2], Yudha Septiana[3], Hizkia Siallagan[4]
[1-4]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang
[1]akbarkhadafi473@gmail.com, [2]arjunsptn39@gmail.com, [3]hizkiasiallagan001@gmail.com ,[4]yudhaseptiyana458@gmail.com ,[5]dosen02526@unpam.ac.id

Abstrak

Acne vulgaris merupakan gangguan inflamasi kulit yang banyak dialami remaja dan dewasa muda serta dapat menimbulkan dampak fisik, psikologis, dan sosial. Penilaian tingkat keparahan secara manual masih bergantung pada pengamatan dokter sehingga berpotensi dipengaruhi subjektivitas, variasi pencahayaan, dan perbedaan pengalaman pemeriksa. Penelitian ini bertujuan mengembangkan sistem klasifikasi otomatis tingkat keparahan acne vulgaris melalui pengolahan citra digital dengan menggabungkan fitur tekstur Local Binary Pattern (LBP), fitur warna Hue, Saturation, Value (HSV), dan algoritma Random Forest. Dataset terdiri atas 1.200 citra wajah yang dibagi seimbang ke dalam empat kelas, yaitu Grade I, Grade II, Grade III, dan Grade IV. Tahap penelitian meliputi perubahan ukuran citra menjadi 256×256 piksel, normalisasi, pengurangan derau menggunakan Gaussian blur, konversi citra ke grayscale dan HSV, ekstraksi 59 fitur LBP dan 102 fitur HSV, penggabungan menjadi 161 fitur, serta klasifikasi menggunakan 100 pohon keputusan. Evaluasi dilakukan menggunakan pembagian data terstratifikasi dan validasi silang lima lipatan. Hasil pengujian menunjukkan akurasi 87,67%, precision 0,88, recall 0,88, dan F1-score 0,87. Kinerja tertinggi diperoleh pada Grade I, sedangkan kesalahan paling banyak terjadi pada kelas yang berdekatan secara visual, terutama Grade III dan Grade IV. Analisis kepentingan fitur menunjukkan bahwa Hue, Saturation, dan pola tepi LBP memberikan kontribusi dominan. Temuan ini menunjukkan bahwa kombinasi tekstur dan warna mampu menghasilkan klasifikasi yang kompetitif, efisien, dan berpotensi digunakan sebagai alat skrining awal, tetapi tetap memerlukan validasi klinis serta pengujian pada citra dengan variasi perangkat, pencahayaan, dan warna kulit yang lebih luas.

Kata Kunci: *Acne Vulgaris, Klasifikasi, Local Binary Pattern, HSV, Random Forest, Pengolahan Citra*

1. Latar Belakang

Acne vulgaris merupakan salah satu penyakit kulit yang paling sering ditemukan, terutama pada remaja dan dewasa muda. Kondisi ini ditandai dengan munculnya komedo, papula, pustula, hingga nodul yang dapat menyebabkan peradangan kronis pada kulit wajah. Selain berdampak secara fisik, acne juga memengaruhi kondisi psikologis penderita seperti menurunnya kepercayaan diri.

Proses klasifikasi tingkat keparahan acne hingga saat ini masih dilakukan secara manual oleh dokter kulit, sehingga berpotensi menimbulkan subjektivitas dan perbedaan hasil diagnosis. Keterbatasan jumlah tenaga medis juga menjadi kendala dalam pelayanan yang cepat dan merata.

Perkembangan kecerdasan buatan dan pengolahan citra digital membuka peluang untuk mengembangkan sistem klasifikasi otomatis berbasis citra. Local Binary Pattern (LBP) efektif dalam mengekstraksi fitur tekstur kulit, sedangkan ruang warna HSV mampu merepresentasikan karakteristik warna dan kemerahan akibat inflamasi. Kombinasi kedua fitur tersebut dengan algoritma Random Forest diharapkan dapat menghasilkan sistem klasifikasi acne yang akurat dan andal.

2. Metode Penelitian

Penelitian ini menggunakan metode eksperimental dengan pendekatan kuantitatif. Dataset berupa citra wajah penderita acne vulgaris dikumpulkan dari sumber publik yang telah tervalidasi.

2.1 Dataset

Dataset yang digunakan terdiri dari 1200 citra wajah dengan acne vulgaris yang diperoleh dari repositori publik yang telah tervalidasi oleh dokter spesialis kulit. Dataset dibagi merata menjadi empat kelas tingkat keparahan dengan masing-masing 300 citra per kelas: Grade I (ringan), Grade II (sedang), Grade III (parah), dan Grade IV (sangat parah). Setiap citra telah di-resize menjadi 256×256 piksel dan dinormalisasi untuk memastikan konsistensi data.

2.2 Preprocessing

Tahap preprocessing meliputi: (1) Resize citra ke dimensi 256×256 piksel menggunakan interpolasi bilinear; (2) Normalisasi intensitas piksel ke rentang [0, 1]; (3) Gaussian blur dengan kernel 3×3 untuk mengurangi noise; (4) Konversi citra RGB ke grayscale untuk ekstraksi LBP dan ke ruang warna HSV untuk ekstraksi fitur warna.

2.3 Ekstraksi Fitur LBP

Local Binary Pattern (LBP) diterapkan pada citra grayscale dengan parameter $P = 8$ (jumlah piksel tetangga) dan $R = 1$ (radius). Untuk setiap piksel pusat dengan intensitas g_c, nilai LBP dihitung dengan:

$$LBP_{P,R}(x_c,y_c)= \sum_{p=0}^{P-1} s(g_p - g_c)\cdot 2^p \qquad (1)$$

dimana g_p adalah intensitas piksel tetangga

ke-p, dan fungsi $s(x)$ didefinisikan sebagai:

$$, \quad (x) = \begin{cases} 1 & \text{jika } x \geq 0 \\ 0, & \end{cases}$$

$s(2)$ jika x < 0

Digunakan varian Uniform LBP yang menghasilkan histogram dengan 59 bin. Histogram dinormalisasi untuk mendapatkan vektor fitur yang invariant terhadap perubahan pencahayaan.

2.4 Ekstraksi Fitur HSV

Citra RGB ditransformasi ke ruang warna HSV. Untuk setiap channel (H, S, V), diekstraksi statistik deskriptif (mean dan standar deviasi) dan histogram dengan 32 bin. Total fitur HSV yang diekstraksi adalah 102 dimensi (6 statistik + 96 histogram). Komponen Hue menangkap informasi warna dan kemerahan, Saturation menangkap intensitas warna, dan Value menangkap kecerahan.

2.5 Penggabungan Fitur dan Klasifikasi

Vektor fitur LBP (59 dimensi) dan HSV (102 dimensi) digabungkan menjadi vektor komposit 161 dimensi. Fitur dinormalisasi menggunakan StandardScaler untuk memastikan mean 0 dan standar deviasi 1. Dataset dibagi menjadi 75% data training dan 25% data testing dengan stratified sampling.

Model Random Forest diimplementasikan dengan 100 decision trees dan kedalaman maksimum 20. Kriteria Gini Impurity digunakan untuk menentukan split optimal:

$$Gini(D)= 1 - \sum p^2_i \qquad (3)$$

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

i=1 dimana$C$ adalah jumlah kelas dan p_i adalah

proporsi sampel dari kelas i dalam dataset D.

2.6 Evaluasi Kinerja

Metrik evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score. Untuk klasifikasi multi-class, metrik dihitung menggunakan weighted average:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

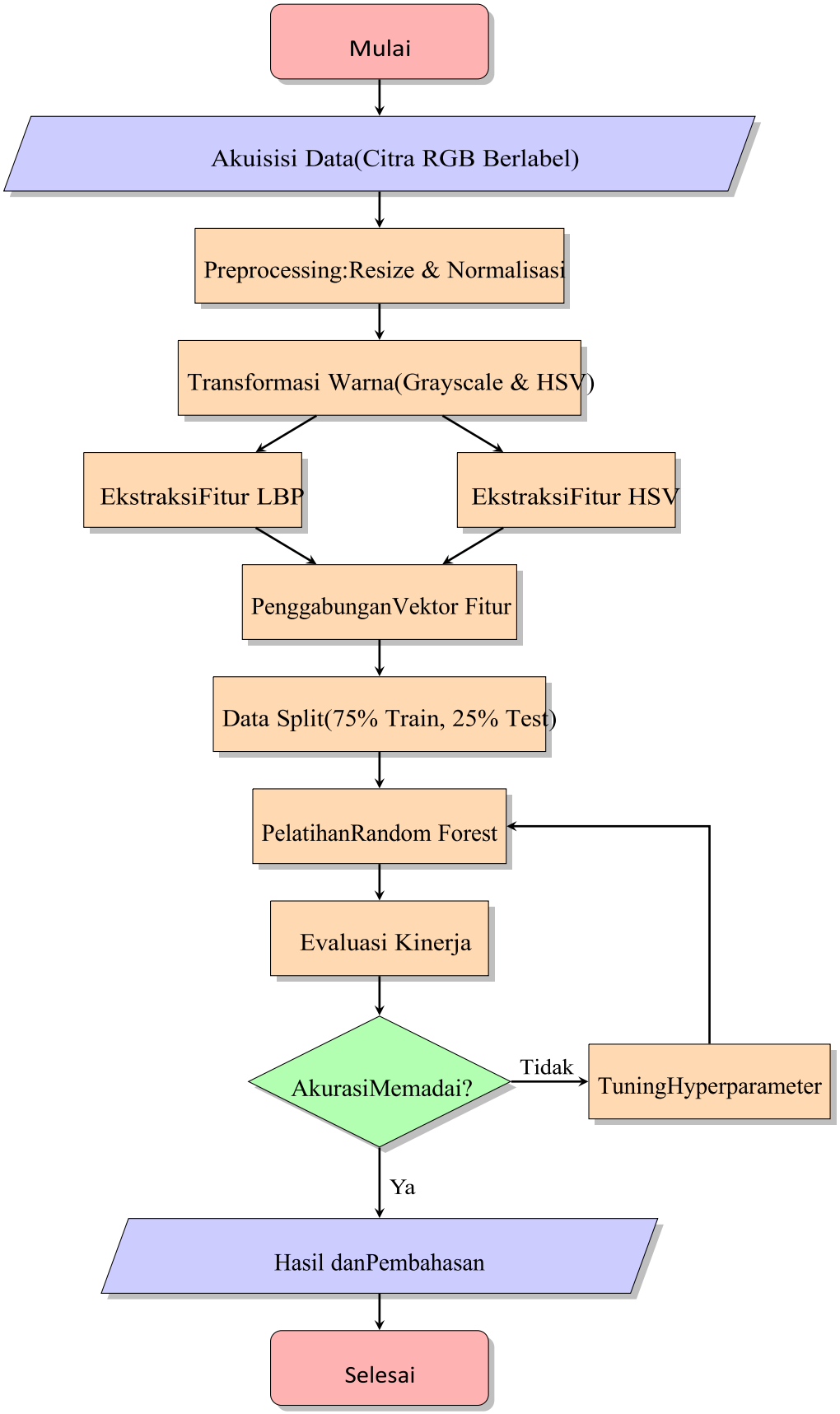
$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}} \quad (7)$$

Cross-validation 5-fold dilakukan pada data training untuk memvalidasi konsistensi kinerja model.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 1. : *Diagram Alir (Flowchart) Proses Data Processing dan Klasifikasi*

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3. Hasil dan Pembahasan

3.1 Kinerja Klasifikasi

Tabel 1 menunjukkan hasil evaluasi model Random Forest pada data testing (300 citra).

Model mencapai akurasi keseluruhan 87.67% dengan rata-rata F1-score 0.87, menunjukkan keseimbangan yang baik antara precision dan recall.

Table 1: Hasil Evaluasi Per Kelas

Kelas	Precision	Recall	F1-Score
Grade I	0.91	0.91	0.91
Grade II	0.88	0.87	0.87
Grade III	0.82	0.84	0.83
Grade IV	0.91	0.89	0.90
Rata-rata	0.88	0.88	0.87

Grade I (acne ringan) menunjukkan kinerja terbaik dengan F1-score 0.91, diikuti Grade IV (sangat parah) dengan 0.90. Grade II dan III memiliki F1-score lebih rendah (0.87 dan 0.83) karena kesamaan karakteristik visual yang membuat pembedaan lebih challenging.

3.2 Confusion Matrix

Tabel 2 menampilkan confusion matrix yang menunjukkan distribusi prediksi model. Diagonal utama merepresentasikan prediksi yang benar, sedangkan nilai di luar diagonal menunjukkan misklasifikasi.

Table 2: Confusion Matrix

Aktual/Pred	GI	GII	GIII	GIV
Grade I	68	5	2	0
Grade II	6	65	4	0
Grade III	1	4	63	7
Grade IV	0	0	8	67

Tidak ada sampel Grade I yang diprediksi sebagai Grade IV atau sebaliknya, menunjukkan model memiliki kemampuan baik dalam membedakan ekstrem tingkat keparahan. Misklasifikasi paling sering terjadi antara Grade III dan IV (15 kasus), yang dapat dijelaskan oleh kesamaan visual pada zona transisi kedua grade tersebut.

3.3 Feature Importance

Analisis feature importance menunjukkan bahwa mean Hue memiliki kontribusi tertinggi (0.087), sejalan dengan observasi klinis bahwa kemerahan akibat inflamasi adalah indikator visual utama tingkat keparahan acne. Histogram Hue pada rentang warna merah dan bin-bin LBP yang merepresentasikan edge pattern juga menunjukkan importance tinggi.

Komponen Saturation berkontribusi sekitar 15%, mengindikasikan bahwa intensitas warna berperan dalam menentukan tingkat keparahan. Channel Value memiliki kontribusi relatif lebih rendah (8%), menunjukkan bahwa variasi kecerahan kurang diskriminatif dibanding informasi warna untuk task klasifikasi acne.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.4 Cross-Validation

Hasil cross-validation 5-fold pada data training menunjukkan akurasi rata-rata 86.5% (±2.1%), mengindikasikan model stabil dan tidak overfitting. Standar deviasi yang relatif kecil menunjukkan konsistensi kinerja model pada berbagai subset data.

3.5 Perbandingan dengan Penelitian Terdahulu

Hasil akurasi 87,67% pada penelitian ini berada pada tingkat yang kompetitif untuk pendekatan berbasis fitur buatan dengan jumlah data yang relatif terbatas. Lim et al. mengembangkan penilaian acne berbasis convolutional neural network dan menunjukkan bahwa citra wajah dapat digunakan untuk melakukan pengelompokan tingkat keparahan secara otomatis [5]. Liu et al. kemudian mengusulkan AcneGrader berbasis ensemble deep learning untuk memperbaiki kestabilan prediksi antargrade [6]. Pendekatan-pendekatan tersebut memperlihatkan kecenderungan penggunaan fitur yang dipelajari secara otomatis, sedangkan penelitian ini mempertahankan fitur tekstur LBP dan warna HSV yang lebih sederhana, mudah dijelaskan, dan tidak memerlukan proses pelatihan jaringan saraf yang kompleks.

Perbandingan angka akurasi antarpublikasi harus dilakukan secara hati-hati karena setiap penelitian menggunakan skema penilaian, komposisi kelas, kualitas citra, dan prosedur validasi yang berbeda. Huynh et al. menggunakan citra telepon pintar untuk mendeteksi lesi dan menentukan tingkat keparahan berdasarkan Investigator's Global Assessment [7], sedangkan Wang et al. membangun aplikasi seluler dengan mekanisme klasifikasi lesi dan penilaian tingkat keparahan [10]. Gao et al. menilai sistem deteksi lesi dan grading pada skenario layanan kesehatan daring dan luring, sehingga variasi sumber citra menjadi bagian penting dari evaluasi [14]. Penelitian ini menggunakan empat grade yang seimbang dan pembagian data terstratifikasi; karena itu, hasilnya terutama menggambarkan kemampuan model pada kondisi dataset yang telah dinormalisasi.

Keunggulan pendekatan LBP-HSV-Random Forest terletak pada kebutuhan komputasi yang lebih ringan dan interpretasi fitur yang lebih langsung. Fitur Hue dapat dikaitkan dengan pola kemerahan, Saturation dengan intensitas warna, sedangkan LBP menggambarkan perubahan tekstur lokal. Pada model deep learning, representasi visual dibentuk melalui banyak lapisan dan sering kali lebih sulit dijelaskan tanpa metode interpretabilitas tambahan. Meskipun demikian, penelitian berbasis deep learning memiliki keunggulan dalam mempelajari bentuk lesi, hubungan spasial, dan variasi pola yang tidak selalu dapat ditangkap oleh histogram tekstur dan warna. Oleh sebab itu, hasil penelitian ini tidak dimaksudkan untuk menyatakan bahwa fitur buatan selalu lebih unggul, tetapi menunjukkan bahwa pendekatan klasik masih relevan ketika ketersediaan data dan sumber daya komputasi terbatas.

Studi Li et al. menekankan pentingnya menggabungkan identifikasi lesi dengan penilaian grade agar keluaran sistem lebih mudah ditelusuri oleh dokter [11]. Kim et al. juga menunjukkan bahwa bantuan algoritma dalam menghitung lesi dapat meningkatkan ketelitian pembaca klinis [12]. Berbeda dengan kedua studi tersebut, keluaran penelitian ini langsung berupa kelas keparahan tanpa menunjukkan lokasi setiap lesi. Perbedaan ini menjadi batas interpretasi penting: model mampu mengenali pola global citra, tetapi belum memberikan bukti visual mengenai lesi mana yang paling memengaruhi keputusan. Dengan demikian, pengembangan berikutnya dapat menambahkan modul deteksi atau segmentasi lesi untuk melengkapi klasifikasi global. Tinjauan sistematis terbaru juga menempatkan deteksi lesi, penghitungan lesi, dan grading sebagai tiga arah utama pengembangan kecerdasan buatan untuk acne [15].

3.6 Implikasi Klinis

Sistem yang dikembangkan berpotensi berfungsi sebagai alat skrining awal untuk membantu pengelompokan tingkat keparahan acne vulgaris secara lebih konsisten. Dalam penggunaan praktis, citra wajah dapat diproses untuk menghasilkan prediksi Grade I sampai Grade IV, kemudian hasil tersebut digunakan sebagai informasi pendukung sebelum pemeriksaan langsung oleh tenaga kesehatan. Recall 0,91 pada Grade I menunjukkan bahwa sebagian besar citra acne ringan dapat dikenali, sedangkan precision 0,91 pada Grade IV menunjukkan bahwa prediksi sangat parah relatif dapat dipercaya pada data uji. Namun, keluaran ini tidak dapat diperlakukan sebagai diagnosis tunggal karena penentuan terapi memerlukan anamnesis, pemeriksaan kondisi kulit secara langsung, evaluasi jaringan parut, lokasi lesi, faktor hormonal, penggunaan obat, dan kondisi individual pasien [2].

Nilai utama sistem bukan menggantikan dokter, melainkan mengurangi variasi penilaian awal dan membantu dokumentasi perkembangan. Kajian mengenai skala acne menunjukkan bahwa beragam sistem penilaian memiliki fokus, kelebihan, dan keterbatasan yang berbeda [3]. Penelitian tentang konsistensi antarpenilai juga

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

memperlihatkan bahwa hasil grading dapat berubah bergantung pada skala dan pengalaman pemeriksa [4]. Dalam konteks tersebut, model komputasional dapat digunakan untuk memberikan pengukuran tambahan yang dilakukan dengan prosedur sama pada setiap citra. Agar manfaat ini tercapai, proses pengambilan gambar harus distandarkan, termasuk jarak kamera, arah wajah, kondisi pencahayaan, latar belakang, dan penggunaan kosmetik.

Implementasi di layanan primer atau teledermatologi juga perlu mempertimbangkan mekanisme rujukan. Prediksi Grade III dan Grade IV sebaiknya memunculkan rekomendasi pemeriksaan tenaga medis, bukan langsung menghasilkan saran obat. Kesalahan klasifikasi pada dua kelas tersebut dapat berdampak pada keterlambatan penanganan atau kecemasan pengguna. Sistem perlu menyajikan tingkat keyakinan prediksi dan pesan bahwa hasil merupakan estimasi berbasis citra. Pedoman pengembangan kecerdasan buatan dermatologi menekankan perlunya pelaporan data, proses validasi, potensi bias, dan batas penggunaan secara transparan [16].

Untuk pemantauan berkala, sistem dapat membantu membandingkan perubahan tingkat keparahan selama periode tertentu apabila citra diambil dalam kondisi yang seragam. Fungsi ini relevan untuk pencatatan, edukasi pasien, dan pemantauan respons secara visual, tetapi interpretasi akhir tetap harus mempertimbangkan penilaian klinis. Sistem juga perlu dirancang agar tidak mendorong pengguna melakukan swamedikasi berdasarkan satu hasil prediksi. Dengan posisi sebagai clinical decision support, keluaran model menjadi salah satu sumber informasi, bukan pengganti keputusan profesional.

3.7 Interpretasi Kontribusi Fitur Tekstur LBP

Local Binary Pattern merupakan deskriptor tekstur yang membandingkan intensitas piksel pusat dengan piksel di sekelilingnya. Pola biner yang dihasilkan kemudian diringkas ke dalam histogram sehingga informasi perubahan lokal pada permukaan kulit dapat direpresentasikan dalam bentuk numerik. Pendekatan ini sesuai dengan tujuan penelitian karena acne vulgaris tidak hanya ditandai oleh perubahan warna, tetapi juga oleh ketidakteraturan permukaan yang muncul akibat komedo, papula, pustula, nodul, bekas peradangan, dan bayangan halus pada area lesi. Ojala et al. memperkenalkan dasar LBP sebagai ukuran tekstur berbasis distribusi lokal [17], sedangkan penerapan lanjut pada pengenalan wajah menunjukkan kemampuan LBP dalam menggambarkan struktur mikro pada citra wajah [18].

Penggunaan P = 8 dan R = 1 pada penelitian ini menghasilkan pengamatan tekstur pada lingkungan piksel yang sangat lokal. Konfigurasi tersebut cocok untuk menangkap tepi kecil dan perubahan intensitas berukuran halus, tetapi dapat kehilangan informasi bentuk lesi yang lebih besar. Histogram Uniform LBP sebanyak 59 bin mengurangi kompleksitas representasi dengan mengelompokkan pola-pola yang dianggap penting. Efisiensi ini membantu Random Forest membangun pemisahan kelas tanpa jumlah fitur yang terlalu besar. Meskipun demikian, pola tekstur yang terbentuk dapat berasal dari sumber selain acne, misalnya pori-pori, rambut wajah, kerutan, bayangan, atau artefak kompresi. Oleh karena itu, kontribusi LBP perlu dipahami sebagai representasi tekstur kulit secara umum, bukan detektor lesi yang spesifik.

F1-score Grade I yang tinggi dapat dijelaskan karena kulit dengan acne ringan cenderung memiliki ketidakteraturan tekstur yang lebih terbatas dibandingkan kelas yang lebih berat. Ketika jumlah dan ukuran lesi meningkat, distribusi pola tepi juga berubah. Namun, Grade III dan Grade IV dapat memiliki pola lokal yang saling tumpang tindih karena keduanya sama-sama menunjukkan lesi inflamasi menonjol. Perbedaan klinis antara kedua grade tidak selalu hanya terletak pada tekstur satu area kecil, tetapi juga pada jumlah lesi, penyebaran, keterlibatan nodul, dan proporsi area wajah. Histogram LBP global dapat merangkum banyak pola, tetapi menghilangkan lokasi spasial dari pola tersebut.

Temuan feature importance yang menunjukkan beberapa bin LBP berkontribusi tinggi mengindikasikan bahwa model memanfaatkan pola tepi dan transisi intensitas tertentu dalam pengambilan keputusan. Namun, nilai kepentingan pada Random Forest tidak menjelaskan arah hubungan secara langsung. Sebuah bin yang penting dapat meningkat pada satu grade dan menurun pada grade lain, atau berinteraksi dengan Hue dan Saturation. Analisis lanjutan dapat menggunakan partial dependence, permutation importance, atau SHAP untuk melihat pengaruh fitur secara lebih rinci. Visualisasi patch citra yang menghasilkan pola LBP dominan juga akan membantu menghubungkan fitur komputasional dengan karakteristik dermatologis.

Pengembangan berikutnya dapat menggunakan multi-scale LBP dengan beberapa nilai radius untuk menangkap lesi berukuran kecil dan besar secara bersamaan. Varian rotation-invariant atau local ternary pattern dapat diuji untuk mengurangi sensitivitas terhadap orientasi dan noise. Spatial pyramid juga dapat diterapkan agar histogram dihitung pada beberapa wilayah wajah, sehingga informasi lokasi tidak sepenuhnya hilang. Perbaikan tersebut tetap mempertahankan kelebihan pendekatan berbasis fitur buatan, yaitu kebutuhan komputasi rendah dan interpretasi yang relatif mudah, sambil meningkatkan kemampuan membedakan kelas berdekatan.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.8 Interpretasi Kontribusi Fitur Warna HSV

Ruang warna HSV memisahkan informasi warna dasar, kejenuhan, dan kecerahan ke dalam tiga komponen yang berbeda. Pemisahan ini memberikan keuntungan dibandingkan penggunaan RGB secara langsung karena perubahan intensitas pencahayaan lebih banyak tercermin pada Value, sedangkan Hue dan Saturation dapat mempertahankan sebagian informasi kromatik. Pada citra acne, kemerahan akibat inflamasi, perubahan warna pascainflamasi, dan perbedaan antara kulit normal dengan area lesi merupakan petunjuk penting. Hasil penelitian menunjukkan mean Hue memiliki nilai kepentingan tertinggi, sedangkan Saturation juga memberikan kontribusi yang bermakna. Temuan tersebut konsisten dengan asumsi bahwa karakteristik warna menjadi pembeda utama antargrade.

Kontribusi Hue perlu ditafsirkan dengan memperhatikan sifat sirkular komponen tersebut. Nilai Hue yang berdekatan pada batas awal dan akhir skala dapat merepresentasikan warna yang serupa, sehingga penggunaan rata-rata sederhana berpotensi menghasilkan interpretasi yang kurang stabil. Histogram Hue sebanyak 32 bin membantu mengurangi masalah tersebut karena distribusi warna dipertahankan, tetapi pemilihan batas bin dapat memengaruhi fitur. Pada kondisi pencahayaan hangat, dingin, atau bercampur, warna kulit dan lesi dapat bergeser. Kamera yang berbeda juga menggunakan white balance, sensor, dan pemrosesan internal yang berbeda. Karena itu, kepentingan Hue pada dataset penelitian belum menjamin kestabilan pada perangkat lain.

Saturation menggambarkan tingkat kejenuhan warna dan dapat membantu membedakan area inflamasi yang tampak lebih merah atau lebih kuat dibandingkan kulit di sekitarnya. Kontribusi sekitar 15% menunjukkan bahwa informasi intensitas warna melengkapi Hue. Namun, Saturation dapat berubah karena pantulan cahaya, kulit berminyak, penggunaan kosmetik, dan kompresi citra. Nilai yang terlalu rendah pada area sangat terang dapat membuat Hue kurang bermakna, sedangkan nilai tinggi pada bayangan berwarna dapat menyesatkan model. Oleh sebab itu, proses pengambilan citra dan preprocessing menjadi bagian penting agar fitur warna mencerminkan kondisi kulit, bukan kondisi kamera.

Channel Value memiliki kontribusi lebih rendah dibandingkan Hue dan Saturation. Hal ini dapat diartikan bahwa tingkat kecerahan global tidak cukup spesifik untuk membedakan grade acne. Meskipun demikian, Value tetap berguna karena bayangan dari lesi menonjol dan kontras antara lesi dengan kulit sekitar dapat memengaruhi pola visual. Rendahnya importance tidak berarti fitur harus langsung dihapus karena Random Forest dapat memanfaatkan interaksi nonlinier antara Value dengan LBP atau komponen warna lain. Evaluasi ablation diperlukan untuk memastikan apakah penghapusan Value meningkatkan, menurunkan, atau tidak mengubah kinerja.

Untuk meningkatkan ketahanan fitur warna, penelitian selanjutnya dapat menerapkan color constancy, koreksi white balance, atau normalisasi berbasis area kulit sehat. Kalibrasi menggunakan kartu warna juga dapat dipertimbangkan pada pengumpulan dataset klinis. Selain HSV, ruang warna CIELAB dapat diuji karena memisahkan luminance dan komponen kromatik dalam representasi yang dirancang mendekati persepsi manusia. Namun, penambahan ruang warna harus diikuti seleksi fitur agar dimensi tidak meningkat tanpa manfaat yang jelas.

3.9 Komplementaritas Fitur LBP dan HSV

Penggabungan 59 fitur LBP dan 102 fitur HSV membentuk vektor komposit 161 dimensi. Prinsip utama penggabungan tersebut adalah bahwa tekstur dan warna membawa informasi yang berbeda tetapi saling melengkapi. Dua citra dapat memiliki tingkat kemerahan yang serupa namun tekstur berbeda, atau memiliki tekstur menonjol dengan warna yang tidak terlalu merah. Model yang hanya menggunakan satu kelompok fitur berisiko mengabaikan sebagian karakteristik lesi. Kombinasi keduanya memungkinkan Random Forest membentuk aturan pemisahan yang melibatkan pola tekstur, distribusi warna, dan interaksi antarfitur.

Komplementaritas ini dapat menjelaskan kinerja yang seimbang antara precision dan recall. Precision 0,88 menunjukkan sebagian besar prediksi pada suatu grade sesuai dengan label sebenarnya, sedangkan recall 0,88 menunjukkan sebagian besar sampel pada setiap grade dapat dikenali. Keseimbangan tersebut penting karena model tidak hanya menghindari false positive, tetapi juga menjaga kemampuan menemukan kelas yang benar. Namun, weighted average dapat menyembunyikan variasi antarkelas. Dataset pada penelitian ini seimbang, sehingga pengaruh setiap grade relatif serupa, tetapi evaluasi pada data nyata yang tidak seimbang dapat menghasilkan pola berbeda.

Random Forest sesuai untuk vektor gabungan karena mampu memodelkan hubungan nonlinier tanpa asumsi distribusi tertentu. Setiap pohon membuat pemisahan berdasarkan subset sampel dan fitur, kemudian hasil akhir ditentukan melalui agregasi banyak pohon [19]. Proses tersebut mengurangi ketergantungan pada satu aturan dan dapat menangani interaksi, misalnya kombinasi Hue tertentu dengan bin LBP yang menunjukkan pola tepi.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Keberadaan 100 pohon memberikan keragaman model, sedangkan kedalaman maksimum 20 membatasi kompleksitas setiap pohon agar tidak terus menyesuaikan noise.

Walaupun fitur telah dinormalisasi menggunakan StandardScaler, algoritma berbasis pohon secara teoritis tidak bergantung pada skala karena split ditentukan oleh urutan nilai. Normalisasi tetap bermanfaat untuk menjaga konsistensi pipeline, memudahkan perbandingan dengan algoritma lain, dan memastikan data tersimpan dalam rentang yang terkendali. Pada eksperimen lanjutan, pengaruh normalisasi dapat diuji secara eksplisit agar tahapan yang tidak meningkatkan kinerja dapat disederhanakan. Dokumentasi pipeline juga penting untuk mencegah data leakage; scaler harus dihitung dari data training dan diterapkan ke data testing menggunakan parameter yang sama.

Untuk membuktikan manfaat penggabungan, penelitian selanjutnya perlu melakukan ablation study dengan tiga skenario: LBP saja, HSV saja, dan LBP-HSV. Perbandingan tidak hanya menggunakan akurasi, tetapi juga F1-score per kelas, confusion matrix, waktu ekstraksi, waktu pelatihan, dan ukuran model. Apabila kombinasi memberikan peningkatan konsisten, klaim komplementaritas menjadi lebih kuat. Jika peningkatan kecil, seleksi fitur dapat digunakan untuk menghapus dimensi yang redundan dan mempercepat sistem.

3.10 Pengaruh Preprocessing terhadap Kualitas Klasifikasi

Tahap preprocessing menentukan apakah fitur yang diekstraksi merepresentasikan kondisi kulit atau justru artefak citra. Prinsip pengolahan citra menempatkan resolusi, interpolasi, dan pengurangan derau sebagai faktor yang memengaruhi informasi yang dipertahankan [21]. Resize menjadi 256×256 piksel menyamakan ukuran input dan menekan kebutuhan komputasi. Keuntungan ini disertai risiko hilangnya detail kecil, terutama komedo atau papula berukuran terbatas. Interpolasi bilinear menghasilkan citra yang lebih halus, tetapi proses perataan dapat mengubah pola tepi yang kemudian dibaca oleh LBP. Oleh sebab itu, ukuran citra bukan sekadar parameter teknis, melainkan faktor yang memengaruhi informasi klinis yang dipertahankan.

Gaussian blur dengan kernel 3×3 digunakan untuk mengurangi noise frekuensi tinggi. Pada citra telepon pintar, noise dapat berasal dari sensor, kompresi, penajaman digital, atau kondisi cahaya rendah. Pengurangan noise dapat menstabilkan histogram tekstur, tetapi blur berlebihan dapat menghilangkan batas lesi. Kernel 3×3 merupakan pilihan konservatif yang menjaga banyak detail, namun parameter sigma dan perbandingan dengan median filter perlu dilaporkan agar proses dapat direplikasi. Evaluasi visual terhadap citra sebelum dan sesudah blur juga penting untuk memastikan lesi tetap terlihat.

Konversi ke grayscale untuk LBP menyederhanakan informasi warna dan memusatkan analisis pada perubahan intensitas. Langkah ini sesuai dengan karakter LBP, tetapi formula grayscale memberikan bobot berbeda pada channel RGB sehingga warna merah dapat menghasilkan intensitas tertentu yang kemudian memengaruhi tekstur. Sementara itu, konversi ke HSV mempertahankan informasi warna untuk kelompok fitur kedua. Pemisahan dua jalur preprocessing ini merupakan desain yang logis karena masing-masing fitur menerima representasi yang sesuai dengan tujuan ekstraksi.

Normalisasi intensitas ke rentang [0,1] membantu konsistensi numerik, terutama apabila citra berasal dari format atau kedalaman bit yang berbeda. Namun, normalisasi global tidak menghilangkan variasi pencahayaan lokal. Area dahi yang terkena pantulan cahaya dapat jauh lebih terang dibandingkan pipi, walaupun keduanya berada pada citra yang sama. Teknik seperti adaptive histogram equalization, illumination correction, atau pemisahan area wajah dapat diuji untuk mengurangi variasi tersebut. Penggunaannya harus hati-hati karena peningkatan kontras dapat membuat kemerahan atau tekstur tampak lebih kuat dari kondisi sebenarnya.

Pipeline preprocessing perlu diuji pada kondisi yang menyerupai penggunaan nyata. Citra dengan latar belakang berbeda, rotasi wajah, rambut menutupi kulit, kacamata, masker, kosmetik, dan kualitas kamera rendah dapat menurunkan kualitas fitur. Sebelum klasifikasi, sistem dapat menambahkan quality control untuk menolak citra yang terlalu gelap, kabur, terpotong, atau tidak menampilkan area wajah dengan memadai. Penolakan input yang buruk lebih aman daripada memaksakan prediksi dengan keyakinan palsu.

3.11 Analisis Kesalahan dan Ambiguitas Antargrade

Misklasifikasi paling banyak terjadi pada kelas yang berdekatan, khususnya Grade III dan Grade IV. Pola ini menunjukkan bahwa model mampu membedakan kondisi ekstrem, tetapi menghadapi kesulitan pada batas keputusan yang memiliki kemiripan visual. Dalam praktik grading, perbedaan tingkat keparahan dapat ditentukan oleh jumlah lesi inflamasi, keberadaan nodul, luas area terlibat, dan kombinasi beberapa jenis lesi [3]. Vektor histogram global tidak secara langsung menghitung lesi atau mengukur luas keterlibatan, sehingga dua citra dengan distribusi warna dan tekstur serupa dapat ditempatkan pada kelas yang salah.

Kesalahan antara Grade II dan Grade III juga dapat terjadi karena transisi klinis bersifat kontinu, sedangkan model dipaksa menghasilkan kelas diskrit. Citra yang berada di sekitar batas grade mungkin juga dinilai berbeda oleh dokter. Variabilitas antarpenilai pada beberapa sistem grading telah dilaporkan [4], sehingga sebagian

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

kesalahan model dapat dipengaruhi oleh ketidakpastian label, bukan hanya kelemahan algoritma. Untuk mengatasi hal ini, proses anotasi idealnya melibatkan lebih dari satu dokter, mencatat tingkat kesepakatan, dan menggunakan konsensus atau label probabilistik pada kasus ambigu.

Tidak adanya prediksi Grade I sebagai Grade IV atau sebaliknya merupakan indikasi bahwa model menangkap perbedaan global antara acne ringan dan sangat parah. Hal tersebut penting karena kesalahan ekstrem berpotensi paling merugikan. Akan tetapi, jumlah data uji pada setiap kelas perlu dipertimbangkan. Dengan 300 citra uji secara keseluruhan, setiap kelas diperkirakan memiliki sekitar 75 citra apabila pembagian tetap seimbang. Satu atau dua kesalahan tambahan dapat mengubah persentase per kelas secara terlihat, sehingga interval kepercayaan atau pengujian berulang akan memberikan gambaran ketidakpastian yang lebih baik.

Analisis kesalahan sebaiknya tidak berhenti pada confusion matrix. Setiap citra yang salah dapat dikelompokkan berdasarkan penyebab potensial, seperti pencahayaan buruk, warna kulit, occlusion, bekas luka, dominasi hiperpigmentasi, lesi kecil yang tersebar, atau lesi besar yang terlokalisasi. Pemeriksaan tersebut dapat menemukan pola kegagalan yang tidak terlihat dari metrik agregat. Misalnya, model mungkin konsisten salah pada kulit dengan refleksi tinggi atau pada citra yang menampilkan hanya sebagian wajah. Temuan seperti ini dapat digunakan untuk memperbaiki pengumpulan data dan preprocessing.

Penggunaan probabilitas kelas dapat membantu menangani ambiguitas. Apabila dua kelas memiliki probabilitas hampir sama, sistem dapat menampilkan status 'perlu verifikasi' daripada memilih satu grade dengan pasti. Pendekatan ordinal juga layak dipertimbangkan karena Grade I sampai Grade IV memiliki urutan. Model ordinal memberikan penalti lebih besar pada kesalahan yang jauh, misalnya Grade I diprediksi Grade IV, dibandingkan kesalahan satu tingkat. Strategi tersebut lebih sesuai dengan struktur klinis tingkat keparahan.

3.12 Stabilitas Model dan Kemampuan Generalisasi

Hasil validasi silang lima lipatan sebesar 86,5% dengan deviasi ±2,1% menunjukkan bahwa kinerja model relatif konsisten pada beberapa pembagian data training. Nilai rata-rata yang dekat dengan akurasi data testing 87,67% memberikan indikasi awal bahwa model tidak hanya menyesuaikan satu subset. Namun, validasi silang internal belum membuktikan generalisasi ke populasi atau perangkat berbeda. Apabila citra dari pasien yang sama muncul pada lipatan berbeda, model dapat mengenali karakteristik individu dan menghasilkan estimasi terlalu optimistis. Pembagian data idealnya dilakukan pada tingkat pasien, bukan hanya tingkat citra.

Generalisasi dipengaruhi oleh keragaman dataset. Faktor penting meliputi usia, jenis kelamin, warna kulit, distribusi lesi, penggunaan kosmetik, jenis kamera, resolusi, pencahayaan, dan latar pengambilan gambar. Studi terbaru menunjukkan bahwa evaluasi pada skenario daring dan luring diperlukan karena perbedaan sumber citra dapat mengubah kinerja [14]. Penelitian dengan data kecil atau tidak seimbang juga menegaskan perlunya strategi khusus untuk menghindari dominasi kelas tertentu [13]. Dataset penelitian ini seimbang, yang mendukung pembelajaran setiap grade, tetapi belum menggambarkan distribusi pasien nyata.

Validasi eksternal merupakan tahap penting sebelum sistem digunakan. Model perlu diuji pada dataset dari institusi lain tanpa pelatihan ulang untuk menilai ketahanan representasi LBP dan HSV. Penurunan kinerja pada data eksternal dapat menunjukkan domain shift. Jika hal tersebut terjadi, pilihan perbaikan meliputi kalibrasi warna, penyesuaian preprocessing, retraining dengan data gabungan, atau domain adaptation. Kinerja eksternal harus dilaporkan bersama karakteristik data agar pengguna memahami cakupan sistem.

Kalibrasi probabilitas juga diperlukan. Model dapat memiliki akurasi tinggi tetapi menghasilkan probabilitas yang terlalu yakin. Brier score, expected calibration error, dan reliability diagram dapat digunakan untuk menilai kesesuaian antara confidence dan frekuensi prediksi benar. Dalam sistem klinis, confidence yang terkalibrasi membantu menentukan kapan model dapat memberikan hasil langsung dan kapan kasus harus dirujuk untuk pemeriksaan manual.

Pengujian robustness dapat dilakukan dengan menambahkan gangguan terkontrol seperti perubahan brightness, blur, rotasi kecil, kompresi JPEG, dan variasi white balance. Tujuannya bukan meningkatkan angka secara artifisial, melainkan mengetahui batas toleransi model. Jika perubahan kecil menyebabkan prediksi berubah drastis, sistem belum cukup stabil untuk penggunaan lapangan. Data augmentation pada tahap training dapat membantu, tetapi transformasi harus tetap realistis dan tidak mengubah label klinis.

3.13 Efisiensi Komputasi dan Relevansi Random Forest

Pendekatan Random Forest memiliki kebutuhan komputasi yang lebih ringan dibandingkan banyak arsitektur deep learning. Setelah fitur diekstraksi, proses training dilakukan pada matriks 1.200×161, sehingga dapat dijalankan pada komputer umum tanpa GPU. Keunggulan ini relevan untuk institusi pendidikan, layanan kesehatan dengan infrastruktur terbatas, atau pengembangan prototipe mobile. Model juga dapat disimpan dalam ukuran relatif kecil dan melakukan inferensi cepat setelah preprocessing selesai.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Random Forest menggabungkan banyak decision tree yang dilatih melalui bootstrap dan pemilihan subset fitur [19]. Keragaman pohon mengurangi varians dibandingkan satu pohon keputusan. Dengan 100 pohon, prediksi dihasilkan dari voting, sehingga kesalahan satu pohon tidak langsung menentukan keluaran. Kedalaman maksimum 20 memberikan kapasitas untuk memodelkan interaksi kompleks. Namun, parameter tersebut perlu dipilih melalui validasi, karena pohon terlalu dalam dapat mempelajari noise, sedangkan pohon terlalu dangkal dapat gagal menangkap hubungan penting.

Feature importance merupakan salah satu alasan pemilihan Random Forest karena memberikan gambaran awal fitur yang digunakan model. Akan tetapi, impurity-based importance dapat bias terhadap fitur dengan banyak kemungkinan split atau fitur yang saling berkorelasi. Histogram HSV memiliki banyak bin yang dapat berbagi informasi, sehingga kepentingan dapat terbagi atau terkonsentrasi secara tidak merata. Permutation importance dapat digunakan sebagai pembanding karena mengukur penurunan kinerja ketika satu fitur diacak. Jika hasil kedua metode konsisten, interpretasi menjadi lebih kuat.

Dibandingkan CNN, pipeline fitur buatan memisahkan tahap ekstraksi dan klasifikasi. Pemisahan ini memudahkan pengujian setiap komponen, tetapi membatasi kemampuan model mempelajari representasi baru. CNN dapat menangkap bentuk, lokasi, dan konteks lesi langsung dari piksel. Studi acne grading modern banyak menggunakan CNN, ensemble, deteksi objek, atau jaringan ringan untuk aplikasi ponsel [5]-[11]. Meskipun akurasi beberapa sistem lebih tinggi, kebutuhan data, komputasi, dan kompleksitas implementasi juga meningkat. Oleh karena itu, pemilihan metode harus mempertimbangkan tujuan sistem, bukan hanya akurasi tertinggi.

Optimasi lebih lanjut dapat dilakukan dengan membandingkan jumlah pohon, kedalaman, minimum samples split, maksimum fitur per split, dan class weight. Randomized search atau Bayesian optimization dapat digunakan, tetapi seluruh proses tuning harus berada di dalam validasi silang untuk mencegah bias. Waktu ekstraksi fitur dan inferensi juga perlu dicatat agar klaim efisiensi dapat dibuktikan secara kuantitatif. Implementasi berbasis pustaka seperti scikit-learn mendukung replikasi pipeline dan evaluasi standar [20].

3.14 Kesiapan Implementasi pada Aplikasi dan Teledermatologi

Sistem dapat diintegrasikan ke aplikasi desktop, web, atau mobile melalui beberapa tahap. Pengguna mengambil atau mengunggah citra, modul quality control memeriksa kelayakan gambar, preprocessing menghasilkan citra grayscale dan HSV, modul ekstraksi membentuk vektor 161 fitur, kemudian Random Forest memberikan grade dan confidence. Hasil dapat disertai informasi bahwa prediksi hanya untuk skrining awal. Alur ini relatif sederhana dan memungkinkan pemrosesan lokal pada perangkat apabila model dan library yang digunakan mendukung.

Standarisasi pengambilan citra merupakan syarat utama. Aplikasi dapat menyediakan bingkai wajah, panduan jarak, instruksi menghadap depan, dan indikator pencahayaan. Pengambilan tiga sudut—depan, kiri, dan kanan—dapat memberi cakupan lebih lengkap dibandingkan satu citra. Beberapa penelitian acne berbasis telepon pintar menggunakan beberapa sudut atau deteksi lesi untuk meningkatkan penilaian [7], [10], [12]. Pada penelitian ini, mekanisme penggabungan beberapa citra belum tersedia, sehingga aplikasi awal sebaiknya membatasi penggunaan sesuai format data training.

Sistem juga perlu menyimpan metadata minimum yang relevan, seperti waktu pengambilan, perangkat, dan kondisi cahaya, tanpa mengumpulkan data pribadi berlebihan. Untuk pemantauan, citra pada kunjungan berbeda dapat dibandingkan, tetapi perubahan perangkat atau pencahayaan dapat menyerupai perubahan klinis. Aplikasi sebaiknya mengingatkan pengguna menggunakan kondisi pengambilan yang sama. Riwayat prediksi dapat ditampilkan dalam grafik, tetapi interpretasinya harus berhati-hati karena perubahan satu grade dapat berada dalam batas ketidakpastian model.

Pada teledermatologi, model dapat membantu prioritas antrean. Citra dengan prediksi berat, confidence rendah, atau kualitas buruk dapat ditandai untuk ditinjau lebih dahulu. Namun, triase otomatis perlu dirancang dengan sensitivitas yang memadai dan mekanisme override oleh tenaga kesehatan. Sistem tidak boleh menolak konsultasi hanya karena model memprediksi ringan. Integrasi dengan rekam medis harus mengikuti kebijakan keamanan, autentikasi, audit akses, dan persetujuan pasien.

Pengujian usability juga dibutuhkan. Sistem yang akurat dapat gagal digunakan apabila instruksi tidak jelas, waktu proses lama, atau hasil membingungkan. Evaluasi melibatkan pasien, dokter, perawat, dan petugas administrasi untuk menilai kemudahan, kepercayaan, serta pengaruh terhadap alur kerja. Uji prospektif pada lingkungan layanan nyata akan memberikan informasi yang tidak diperoleh dari eksperimen dataset statis.

3.15 Aspek Etika, Privasi, dan Potensi Bias

Citra wajah merupakan data biometrik yang sensitif karena dapat mengidentifikasi individu. Pengumpulan dan penggunaan dataset harus didasarkan pada persetujuan yang jelas, tujuan terbatas, dan perlindungan akses.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Identitas, nama file, metadata lokasi, dan informasi perangkat yang tidak diperlukan perlu dihapus atau dipisahkan. Penyimpanan harus menggunakan enkripsi dan kontrol hak akses. Apabila citra dikirim ke server, pengguna perlu diberi informasi mengenai lokasi pemrosesan, lama penyimpanan, dan pihak yang dapat mengakses data.

Bias dapat muncul apabila dataset tidak mewakili variasi warna kulit dan kondisi populasi. Fitur HSV sangat berkaitan dengan warna sehingga model mungkin belajar karakteristik kelompok tertentu. Perbedaan kemerahan juga tidak selalu terlihat sama pada semua warna kulit. Jika sebagian besar data berasal dari satu kelompok, kinerja agregat 87,67% dapat menyembunyikan penurunan pada kelompok lain. Evaluasi harus dilaporkan secara terpisah menurut karakteristik demografis yang relevan apabila data tersedia dan penggunaan variabel tersebut mendapat persetujuan etis.

Pedoman CLEAR Derm menekankan transparansi tentang sumber data, proses anotasi, pembagian dataset, metrik, dan validasi untuk sistem kecerdasan buatan berbasis citra dermatologi [16]. Prinsip tersebut relevan meskipun model penelitian ini menggunakan machine learning klasik. Pengembang perlu mendokumentasikan parameter LBP, bin HSV, preprocessing, Random Forest, versi perangkat lunak, dan seed eksperimen. Dokumentasi memungkinkan penelitian direplikasi dan memudahkan audit ketika hasil tidak sesuai.

Keadilan juga berkaitan dengan akses. Sistem ringan dapat memperluas skrining pada wilayah dengan keterbatasan dokter kulit, tetapi penggunaan telepon pintar dan internet tidak merata. Aplikasi harus dirancang agar dapat bekerja pada perangkat menengah, menyediakan instruksi sederhana, dan tidak memerlukan unggahan berulang berukuran besar. Pilihan pemrosesan lokal dapat mengurangi biaya data sekaligus melindungi privasi, meskipun pembaruan model menjadi lebih kompleks.

Tanggung jawab keputusan harus dinyatakan dengan jelas. Pengguna perlu mengetahui bahwa model dapat salah dan tidak menilai seluruh aspek klinis. Tenaga kesehatan perlu dapat melihat confidence, kualitas input, dan riwayat perubahan, bukan hanya label akhir. Mekanisme pelaporan kesalahan dan pembaruan model juga diperlukan. Pengembangan teknologi harus mengikuti evaluasi manfaat dan risiko, bukan hanya peningkatan akurasi.

3.16 Keterbatasan Penelitian

Keterbatasan pertama adalah informasi mengenai sumber dan karakteristik dataset belum dijelaskan secara rinci. Walaupun dataset disebut berasal dari repositori publik dan tervalidasi dokter, artikel perlu menyebutkan nama repositori, kriteria inklusi, jumlah pasien unik, distribusi usia, jenis kelamin, warna kulit, serta apakah satu pasien menyumbang lebih dari satu citra. Informasi tersebut menentukan risiko bias dan data leakage. Tanpa pemisahan berbasis pasien, hasil validasi dapat lebih tinggi karena citra individu yang sama memiliki karakteristik serupa.

Keterbatasan kedua adalah label keparahan disajikan sebagai Grade I sampai Grade IV tanpa penjelasan lengkap mengenai sistem grading dan kriteria setiap kelas. Literatur menunjukkan terdapat berbagai skala dengan definisi berbeda [3]. Agar model dapat digunakan dan dibandingkan, prosedur anotasi harus menyebutkan skala, jumlah dokter, mekanisme konsensus, dan tingkat kesepakatan. Kasus yang berada pada batas kelas juga perlu didokumentasikan.

Keterbatasan ketiga adalah analisis hanya menggunakan metrik agregat dan hasil per kelas. Confusion matrix disebutkan, tetapi nilai lengkap perlu ditampilkan agar pembaca dapat menghitung jenis kesalahan. Interval kepercayaan, ROC one-versus-rest, area under curve, specificity, dan calibration belum tersedia. Metrik tambahan bukan untuk memperbanyak angka, tetapi untuk memahami perilaku model pada berbagai kebutuhan.

Keterbatasan keempat adalah belum adanya eksperimen pembanding yang menggunakan fitur tunggal atau algoritma lain pada data yang sama. Tanpa ablation, kontribusi nyata LBP, HSV, dan penggabungannya belum dapat dipisahkan. Perbandingan dengan SVM, logistic regression, k-nearest neighbors, gradient boosting, atau CNN ringan dapat menunjukkan apakah Random Forest merupakan pilihan optimal. Waktu komputasi dan ukuran model juga belum dilaporkan.

Keterbatasan kelima adalah model belum diuji secara eksternal dan prospektif. Akurasi internal tidak cukup untuk menyimpulkan kesiapan klinis. Kondisi nyata memiliki variasi perangkat, pencahayaan, posisi wajah, dan kualitas citra yang lebih besar. Validasi oleh dokter spesialis pada populasi independen, uji pengguna, serta evaluasi dampak terhadap keputusan klinis diperlukan sebelum sistem digunakan sebagai alat pelayanan.

Keterbatasan terakhir berkaitan dengan interpretabilitas. Feature importance memberikan gambaran kontribusi global, tetapi tidak menunjukkan bagian citra yang memengaruhi satu prediksi. Sistem belum mampu menandai lesi, menghitung jumlah lesi, atau menjelaskan alasan suatu citra dimasukkan ke grade tertentu. Penggabungan klasifikasi global dengan deteksi lesi akan meningkatkan keterlacakan hasil.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.17 Arah Pengembangan Penelitian Selanjutnya

Penelitian selanjutnya sebaiknya dimulai dengan penguatan dataset. Pengumpulan citra perlu melibatkan beberapa institusi dan perangkat dengan protokol yang terdokumentasi. Setiap pasien dapat direkam dari beberapa sudut dengan pencahayaan standar dan kondisi penggunaan nyata. Label diberikan oleh minimal dua dokter, kemudian kasus berbeda pendapat diselesaikan melalui konsensus. Pembagian training, validation, dan testing dilakukan berdasarkan pasien untuk mencegah kebocoran.

Eksperimen fitur dapat diperluas melalui multi-scale LBP, CIELAB, color constancy, dan fitur bentuk. Seleksi fitur dapat digunakan untuk mengurangi redundansi 161 dimensi. Ablation study memberikan bukti kelompok fitur mana yang paling berguna. Selain Random Forest, gradient boosting dan support vector machine dapat dibandingkan menggunakan prosedur validasi yang sama. Tujuan perbandingan bukan sekadar mencari angka tertinggi, tetapi menilai trade-off antara akurasi, interpretabilitas, kecepatan, dan ukuran model.

Pendekatan deep learning dapat dikembangkan sebagai pembanding atau sistem hibrida. CNN ringan dapat mengekstraksi fitur otomatis, kemudian fitur tersebut digabungkan dengan LBP dan HSV. Deteksi objek dapat digunakan untuk menemukan komedo, papula, pustula, dan nodul sebelum menghitung tingkat keparahan, sebagaimana dilakukan pada berbagai studi acne modern [7], [8], [11], [12], serta dirangkum dalam tinjauan sistematis terbaru [15]. Model hibrida memungkinkan informasi lokal lesi dan karakteristik global wajah digunakan bersama.

Evaluasi perlu mencakup external validation, robustness, fairness, calibration, dan interpretabilitas. Hasil dilaporkan per kelompok perangkat dan warna kulit jika memungkinkan. Confidence threshold dapat digunakan untuk menentukan kasus yang memerlukan verifikasi dokter. Uji prospektif dapat membandingkan penilaian dokter tanpa sistem, dokter dengan sistem, dan sistem saja. Desain tersebut membantu mengetahui apakah model benar-benar meningkatkan konsistensi dan efisiensi.

Tahap akhir adalah pengembangan prototipe aplikasi dengan pemrosesan aman. Aplikasi menyediakan panduan pengambilan citra, pemeriksaan kualitas, prediksi grade, confidence, dan peringatan batas penggunaan. Penyimpanan data mengikuti persetujuan dan keamanan yang memadai. Uji usability menilai apakah pengguna memahami hasil dan apakah tenaga kesehatan merasa sistem membantu. Dengan tahapan tersebut, penelitian dapat berkembang dari eksperimen klasifikasi menjadi sistem pendukung yang bertanggung jawab, dapat direplikasi, dan relevan bagi kebutuhan layanan.

4. Kesimpulan

Penelitian ini mengembangkan klasifikasi otomatis tingkat keparahan acne vulgaris menggunakan kombinasi fitur tekstur LBP, fitur warna HSV, dan algoritma Random Forest. Model memperoleh akurasi 87,67%, precision 0,88, recall 0,88, serta F1-score 0,87. Hasil tersebut menunjukkan bahwa penggabungan tekstur dan warna mampu memberikan kinerja yang seimbang pada empat kelas keparahan. Analisis kepentingan fitur menunjukkan bahwa Hue, Saturation, dan pola tepi LBP menjadi unsur dominan dalam proses klasifikasi. Kinerja terbaik diperoleh pada Grade I, sedangkan kesalahan lebih banyak muncul pada Grade III dan Grade IV yang memiliki karakteristik visual berdekatan. Dengan kebutuhan komputasi yang relatif ringan, model berpotensi digunakan sebagai alat skrining dan pendukung dokumentasi, bukan sebagai pengganti pemeriksaan dokter. Penggunaan lebih lanjut memerlukan penjelasan sumber dataset, pembagian berbasis pasien, validasi eksternal, pengujian pada variasi warna kulit dan perangkat, analisis ablation, serta evaluasi interpretabilitas dan kalibrasi. Pengembangan deteksi lesi, sistem ordinal, dan aplikasi dengan kontrol kualitas citra dapat meningkatkan keterlacakan dan kesiapan implementasi.

Referensi

[1] Williams, H. C., Dellavalle, R. P., & Garner, S. (2012). Acne vulgaris. The Lancet, 379(9813), 361-372. https://doi.org/10.1016/S0140-6736(11)60321-8.

[2] Reynolds, R. V., Yeung, H., Cheng, C. E., et al. (2024). Guidelines of care for the management of acne vulgaris. Journal of the American Academy of Dermatology, 90(5), 1006.e1-1006.e30. https://doi.org/10.1016/j.jaad.2023.12.017.

[3] Bae, I. H., Kwak, J. H., Na, C. H., Kim, M. S., Shin, B. S., & Choi, H. (2024). A comprehensive review of the acne grading scale in 2023. Annals of Dermatology, 36(2), 65-73. https://doi.org/10.5021/ad.23.094.

[4] Yu, R., Fei, W., Ning, X., & Cui, Y. (2022). Inter-rater variability and consistency within four acne grading systems recommended in China, USA, and Europe. Journal of Cosmetic Dermatology, 21(11), 6156-6162. https://doi.org/10.1111/jocd.15178.

[5] Lim, Z. V., Akram, F., Ngo, C. P., et al. (2020). Automated grading of acne vulgaris by deep learning with convolutional neural networks. Skin Research and Technology, 26(2), 187-192. https://doi.org/10.1111/srt.12794.

[6] Liu, S., Fan, Y., Duan, M., et al. (2022). AcneGrader: An ensemble pruning of the deep learning base models to grade acne. Skin Research and Technology, 28(5), 677-688. https://doi.org/10.1111/srt.13166.

[7] Huynh, Q. T., Nguyen, P. H., Le, H. X., et al. (2022). Automatic acne object detection and acne severity grading using smartphone images and artificial intelligence. Diagnostics, 12(8), 1879. https://doi.org/10.3390/diagnostics12081879.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

[8] Wen, H., Yu, W., Wu, Y., Zhao, J., Liu, X., Kuang, Z., & Fan, R. (2022). Acne detection and severity evaluation with interpretable convolutional neural network models. Technology and Health Care, 30(S1), 143-153. https://doi.org/10.3233/THC-228014.

[9] Lin, Y., Jiang, J., Ma, Z., et al. (2022). KIEGLFN: A unified acne grading framework on face images. Computer Methods and Programs in Biomedicine, 221, 106911. https://doi.org/10.1016/j.cmpb.2022.106911.

[10] Wang, J., Luo, Y., Wang, Z., et al. (2023). A cell phone app for facial acne severity assessment. Applied Intelligence, 53(7), 7614-7633. https://doi.org/10.1007/s10489-022-03774-z.

[11] Li, J., Du, D., Zhang, J., et al. (2023). Development and validation of an artificial intelligence-powered acne grading system incorporating lesion identification. Frontiers in Medicine, 10, 1255704. https://doi.org/10.3389/fmed.2023.1255704.

[12] Kim, D. H., Sun, S., Cho, S. I., Kong, H. J., Lee, J. W., Lee, J. H., & Suh, D. H. (2023). Automated facial acne lesion detecting and counting algorithm for acne severity evaluation and its utility in assisting dermatologists. American Journal of Clinical Dermatology, 24(4), 649-659. https://doi.org/10.1007/s40257-023-00777-5.

[13] Bernhard, R., Bletterer, A., Le Caro, M., García Álvarez, E., Kostov, B., & Herrera Egea, D. (2024). Automatic acne severity grading with a small and imbalanced data set of low-resolution images. Dermatology and Therapy, 14(11), 2953-2969. https://doi.org/10.1007/s13555-024-01283-0.

[14] Gao, N., Wang, J., Zhao, Z., Chu, X., Lv, B., Han, G., Ni, Y., & Xie, G. (2025). Evaluation of an acne lesion detection and severity grading model for Chinese population in online and offline healthcare scenarios. Scientific Reports, 15(1), 1119. https://doi.org/10.1038/s41598-024-84670-z.

[15] Traini, D. O., Palmisano, G., Guerriero, C., & Peris, K. (2025). Artificial intelligence in the assessment and grading of acne vulgaris: A systematic review. Journal of Personalized Medicine, 15(6), 238. https://doi.org/10.3390/jpm15060238.

[16] Daneshjou, R., Barata, C., Betz-Stablein, B., et al. (2022). Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR Derm consensus guidelines. JAMA Dermatology, 158(1), 90-96. https://doi.org/10.1001/jamadermatol.2021.4915.

[17] Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1), 51-59. https://doi.org/10.1016/0031-3203(95)00067-4.

[18] Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12), 2037-2041. https://doi.org/10.1109/TPAMI.2006.244.

[19] Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324.

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

[21] Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing (4th ed.). Pearson.

DOI: https://doi.org/10.31004/riggs.v5i2.5271
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)