

Evaluasi Performa Algoritma K-Nearest Neighbor untuk Klasifikasi Golongan Darah ABO dan Faktor Rhesus

^{1*}Enggar Ardi Prastiawan, ²Danar Putra Pamungkas, ³Made Ayu Dusea Widyadara

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

E-mail: ¹enggarardi25@gmail.com, ²danar@unpkediri.ac.id, ³madedara@unpkediri.ac.id

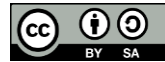
Penulis Korespondens : Enggar Ardi Prastiawan

Abstrak—Klasifikasi golongan darah secara cepat dan akurat sangat krusial dalam prosedur medis seperti transfusi darah. Penelitian ini mengevaluasi performa algoritma *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan citra golongan darah ABO dan faktor Rhesus menggunakan representasi fitur piksel langsung (*flattening*). Citra masukan melalui tahapan pra-pemrosesan meliputi *cropping*, *background removal*, konversi *grayscale*, histogram *equalization*, dan *Gaussian blur*, sebelum dibagi menjadi 80% data latih dan 20% data uji. Hasil pengujian hyperparameter tuning menunjukkan bahwa konfigurasi parameter terbaik dicapai pada nilai $K = 3$ dengan Akurasi Global sebesar 43% (42,86%). Rendahnya performa ini disebabkan oleh hilangnya informasi hubungan spasial tekstur citra akibat metode flattening piksel langsung (dimensi 4096), yang memicu penumpukan (*overlapping*) nilai intensitas piksel antara sampel positif aglutinasi dan negatif di ruang multidimensi. Selain itu, keterbatasan sampel data memicu bias yang kuat ke arah kelas mayoritas (kelas *O_Positive* mencapai *Recall* 100%), namun gagal total (performa 0.00%) pada kelas minoritas langka seperti *AB_Negative*, *AB_Positive*, dan *O_Negative*. Karena tingkat sensitivitas yang rendah ini sangat berisiko dalam prosedur medis nyata, penelitian selanjutnya direkomendasikan untuk mengganti representasi piksel langsung dengan metode ekstraksi fitur tekstur yang spesifik seperti GLCM atau LBP, serta menerapkan teknik penyeimbangan data latih seperti augmentasi citra atau SMOTE.

Kata Kunci—Klasifikasi Golongan Darah, *K-Nearest Neighbor*, *Flattening*, *Imbalanced Dataset*.

Abstract—A fast and accurate blood type classification is crucial in medical procedures such as blood transfusions. This study evaluates the performance of the *K-Nearest Neighbor* (KNN) algorithm for classifying ABO and Rhesus blood type images using direct pixel feature representation (*flattening*). Input images underwent preprocessing stages including *cropping*, *background removal*, *grayscale conversion*, *histogram equalization*, and *Gaussian blur*, before being split into 80% training data and 20% testing data. The hyperparameter tuning results showed that the best parameter configuration was achieved at $K = 3$ with a Global Accuracy of 43% (42.86%). This low performance is caused by the loss of spatial correlation in image texture due to the direct pixel flattening method (4096 dimensions), which triggers an overlap of pixel intensity values between agglutination-positive and negative samples in the multidimensional space. Furthermore, the limited data samples led to a strong bias toward the majority class (the *O_Positive* class achieved 100% Recall) but failed completely (0.00% performance) on rare minority classes such as *AB_Negative*, *AB_Positive*, and *O_Negative*. Since this low sensitivity level poses a high risk in real medical procedures, future research is recommended to replace direct pixel representation with specific texture feature extraction methods such as GLCM or LBP, as well as applying training data balancing techniques like image augmentation or SMOTE.

Keywords—Blood Type Classification, *K-Nearest Neighbor*, *Flattening*, *Imbalanced Dataset*.



I. PENDAHULUAN

Golongan darah ABO dan faktor Rhesus (Rh) memainkan peran penting dalam dunia medis, khususnya untuk transfusi darah, manajemen klinis wanita hamil, dan prosedur donor darah[1]. Kesalahan dalam menentukan kompatibilitas golongan darah dapat memicu reaksi hemolitik imun akut yang dapat berakibat fatal bagi keselamatan pasien[2]. Dengan kemajuan teknologi kecerdasan buatan, pengolahan citra digital kini banyak diterapkan untuk membantu mengotomatisasi pengujian laboratorium medis, termasuk visualisasi aglutinasi darah[3]. Melalui penggunaan algoritma pembelajaran mesin seperti *K-Nearest Neighbor* (KNN), klasifikasi visual sampel darah dapat dilakukan secara digital dengan mengekstrak fitur warna dan tekstur dari gambar skala abu-abu untuk menentukan klasifikasi golongan darah[4].

Meskipun metode komputasi menawarkan efisiensi waktu, tantangan utama dalam mengimplementasikan klasifikasi berbasis citra digital terletak pada akurasi dan stabilitas kinerja algoritma ketika dihadapkan pada variasi data yang dinamis[5]. Dalam praktiknya, kualitas gambar yang diambil oleh perangkat keras seringkali terganggu oleh pencahayaan, kontras rendah, dan dataset yang tidak seimbang, yang dapat memengaruhi objektivitas model[6]. Dalam percobaan ini, ketersediaan sampel data relatif terbatas dan tidak merata misalnya, kelas A_Positif dan A_Negatif mendominasi, sedangkan kelas AB_Positif memiliki jumlah sampel yang sangat minim[7]. Kesenjangan data ini menciptakan kesenjangan penelitian yang dapat memicu bias dalam algoritma KNN, sehingga memerlukan evaluasi mendalam untuk menentukan sejauh mana algoritma ini mampu mempertahankan kinerja terbaiknya dalam mengenali karakteristik unik dari setiap golongan darah[8].

Berdasarkan identifikasi masalah, formulasi masalah yang diajukan dalam penelitian ini adalah bagaimana tahapan pra-pemrosesan citra digital seperti pengubahan ukuran, peningkatan kontras dengan ekualisasi histogram, dan pengurangan *noise* menggunakan *Gaussian blur* memengaruhi kinerja klasifikasi algoritma *K-Nearest Neighbor* (KNN)[9]. Selain itu, penelitian ini juga menguji efektivitas dan konsistensi kinerja algoritma KNN dalam mengklasifikasikan citra golongan darah ABO dan faktor Rhesus ketika diuji menggunakan metrik evaluasi akurasi, presisi, *recall*, dan *f1-score* di tengah kondisi data yang tidak seimbang (*imbalanced dataset*)[10].

Tujuan penelitian ini adalah untuk mengevaluasi kinerja dan ketahanan algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan citra skala abu-abu golongan darah berdasarkan sistem ABO dan Rhesus[11]. Secara spesifik, penelitian ini bertujuan untuk mengukur tingkat keberhasilan model melalui analisis metrik *accuracy_score*, *precision_score*, *recall_score*, dan *f1_score* serta pemetaan melalui matriks kebingungan[12]. Melalui hasil evaluasi ini, diharapkan penelitian ini dapat memberikan kontribusi berupa rekomendasi ilmiah mengenai

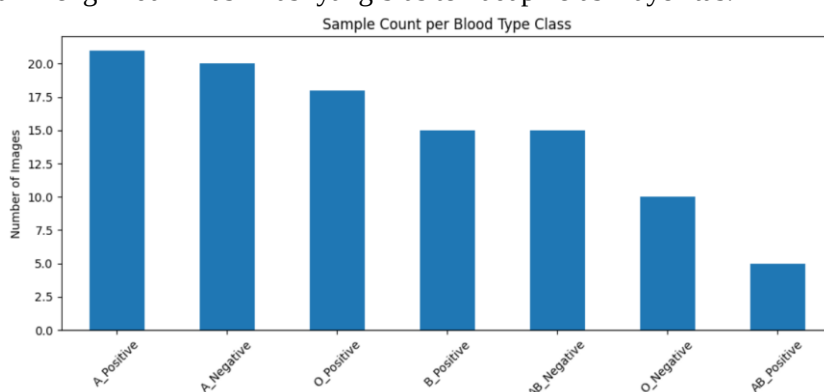
optimasi algoritma KNN dan penanganan data citra medis untuk pengembangan sistem identifikasi golongan darah yang lebih akurat, responsif, dan andal di masa mendatang[13].

II. METODE

A. Pengumpulan Data

Proses pengumpulan data dalam penelitian ini dilakukan dengan mengumpulkan data citra golongan darah medis yang disimpan secara lokal di direktori utama bernama *grayscale_dataset*. Dari hasil akuisisi data, diperoleh total 104 sampel citra digital skala abu-abu yang mewakili berbagai karakteristik variasi visual dalam aglutinasi darah. Semua sampel citra dikelompokkan menjadi 7 kelas label target yang berbeda berdasarkan kombinasi sistem golongan darah ABO dan faktor Rhesus (Rh).

Dalam implementasinya, karakteristik dataset ini menunjukkan jumlah sampel yang tidak seimbang antar kelas (dataset tidak seimbang). Distribusi frekuensi sampel di setiap kelas target dicatat secara detail sebagai berikut: kelas A_Positif memiliki jumlah sampel terbanyak, yaitu 21 gambar, diikuti secara berurutan oleh kelas A_Negatif dengan 20 gambar, kelas O_Positif dengan 18 gambar, kelas AB_Negatif dengan 15 gambar, kelas B_Positif dengan 15 gambar, dan kelas O_Negatif dengan 10 gambar. Sementara itu, kelas AB_Positif adalah kelompok dengan representasi data paling sedikit, yaitu hanya memiliki 5 sampel gambar. Informasi tentang distribusi data yang tidak seimbang ini merupakan parameter penting dalam merancang strategi pemisahan data dan mengevaluasi kinerja model untuk menghindari klasifikasi yang bias terhadap kelas mayoritas.



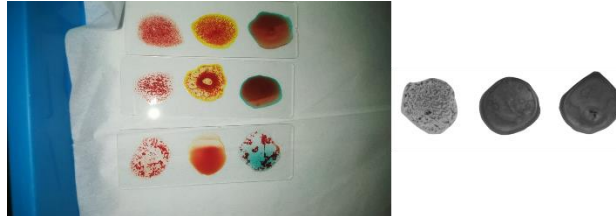
Gambar 1. Jumlah Dataset

B. Pra-Pemrosesan Citra

Sebelum ekstraksi fitur dan pemodelan menggunakan algoritma *K-Nearest Neighbor* (KNN), semua gambar golongan darah digital menjalani serangkaian langkah pra-pemrosesan gambar terstruktur. Proses ini dimulai dengan pemotongan gambar otomatis untuk mengisolasi wilayah pengamatan utama (*Region of Interest* atau ROI), sehingga menghilangkan area yang tidak relevan di luar sampel darah. Selanjutnya, teknik segmentasi berbasis ambang batas diterapkan untuk melakukan penghapusan latar belakang. Langkah ini bertujuan untuk memisahkan objek aglutinasi darah dari latar belakang kartu uji atau swab, yang dapat mengganggu objektivitas model.

Setelah latar belakang berhasil dihilangkan, citra dibaca dalam format skala keabuan (*grayscale*) menggunakan fungsi `cv2.IMREAD_GRAYSCALE` dan dimensinya diseragamkan menjadi ukuran matriks 64×64 piksel. Untuk mengoptimalkan kualitas

visual citra, metode Histogram Equalization diterapkan guna meratakan distribusi intensitas warna, yang kemudian diikuti oleh pengaplikasian filter Gaussian Blur dengan ukuran kernel 3×3 untuk mereduksi derau (*noise*) halus. Pada tahap akhir, seluruh nilai intensitas piksel dinormalisasi ke dalam rentang skala 0.0 hingga 1.0 dengan membaginya dengan nilai 255.0. Matriks citra 2D tersebut kemudian diratakan (*flattening*) menjadi sebuah vektor tunggal 1D berukuran 4096 dimensi fitur yang siap digunakan sebagai input pada tahap pelatihan dan pengujian model.



Gambar 2. Hasil Pra-Pemrosesan Data

C. Pembagian Data dan Penskalaan (Data Splitting dan Scaling)

Setelah pra-pemrosesan, 7 label golongan darah diubah menjadi angka (0-6) menggunakan Label *Encoding*. Dataset kemudian dibagi secara *stratified* menjadi 80% data latih (83 sampel) dan 20% data uji (21 sampel) untuk menjaga proporsi kelas dan mencegah bias. Sebelum masuk ke algoritma K-Nearest Neighbor (KNN), fitur piksel 1D (4096 dimensi) distandarisasi menggunakan *Standard Scaler* ($mean=0$, $varians=1$). Penskalaan ini wajib untuk KNN yang berbasis jarak (seperti *Euclidean*) guna mencegah dominasi nilai fitur tertentu, sehingga prediksi model lebih objektif dan akurat.

D. Algoritma Klasifikasi dan Optimasi

Klasifikasi golongan darah pada penelitian ini menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan metrik Euclidean Distance untuk mengukur jarak antar-vektor fitur (4096 dimensi). Label kelas diprediksi berdasarkan suara terbanyak (*majority voting*) dari K tetangga. Untuk mendapatkan performa yang optimal, dilakukan *hyperparameter tuning* dengan menguji nilai K dari 1 hingga 15 secara iteratif. Setelah nilai K terbaik ditemukan, model diekspor menjadi berkas biner (blood_knn_model.pkl) menggunakan pustaka joblib agar dapat langsung digunakan (*deployment*) pada data baru tanpa perlu dilatih ulang.

E. Metrik Evaluasi

Untuk mengukur, menguji, dan memvalidasi performa algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan citra golongan darah secara objektif, penelitian ini menggunakan instrumen evaluasi berupa *Confusion Matrix* yang dilanjutkan dengan perhitungan empat metrik statistik utama, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Karena dataset memiliki karakteristik *multilabel* atau banyak kelas (7 kelas golongan darah), evaluasi dilakukan dengan menghitung nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) untuk setiap kelas secara spesifik.

1. Accuracy

Akurasi merupakan metrik yang digunakan untuk mengukur rasio total prediksi benar (baik positif maupun negatif) terhadap keseluruhan sampel data uji yang ada.

Metrik ini memberikan gambaran umum mengenai ketepatan model dalam menebak seluruh kelas golongan darah, yang dirumuskan sebagai:

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

2. Precision

Presisi mengukur tingkat keakuratan antara data yang diprediksi positif oleh model dengan data yang sebenarnya bernilai positif. Di dalam kasus ini, presisi menunjukkan seberapa besar kemampuan model KNN dalam meminimalkan kesalahan deteksi (*false alarm*), seperti memprediksi suatu citra sebagai golongan darah A+ padahal sebenarnya adalah golongan darah lain. Rumus presisi dihitung melalui persamaan:

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

3. Recall

Recall atau sensitivitas mengukur keberhasilan model dalam menemukan kembali (*retrieve*) seluruh sampel yang benar-benar termasuk ke dalam kelas positif. Dalam konteks klasifikasi medis golongan darah, *recall* sangat krusial karena mengindikasikan seberapa mampu algoritma mengidentifikasi semua sampel golongan darah tertentu tanpa ada yang terlewat (meminimalkan nilai *False Negative*). Persamaan untuk menghitung *recall* adalah:

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

4. F1-Score

F1-Score merupakan metrik evaluasi yang menghitung nilai rata-rata harmonik (*harmonic mean*) dari komponen presisi dan *recall*. Metrik ini menjadi indikator performa yang sangat penting dan lebih objektif dalam penelitian ini, mengingat dataset citra golongan darah yang digunakan berada dalam kondisi tidak seimbang (*imbalanced dataset*), di mana nilai akurasi saja tidak cukup untuk merepresentasikan ketangguhan model pada kelas minoritas (seperti kelas *AB_Positive*). Perumusan untuk *F1-Score* dinyatakan sebagai:

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

Melalui pengujian komprehensif berbasis keempat metrik di atas serta visualisasi multilabel confusion matrix, efektivitas dan stabilitas algoritma KNN dalam memetakan citra aglutinasi darah dapat dianalisis secara mendalam dan valid.

III. HASIL DAN PEMBAHASAN

A. Pra-pemrosesan

Proses pra-pemrosesan diawali dengan tahap *cropping* (pemotongan) dan background removal untuk mengisolasi area melingkar tempat sampel darah bereaksi pada kartu golongan darah, sehingga algoritma tidak terganggu oleh informasi latar belakang kertas. Selanjutnya, citra dikonversi menjadi skala keabuan (*grayscale*) untuk memangkas dimensi saluran warna dari RGB (3 saluran) menjadi matriks tunggal yang lebih efisien secara komputasi. Mengingat variasi pencahayaan saat pengambilan gambar di laboratorium sangat tinggi, penerapan *histogram equalization* terbukti berhasil menonjolkan perbedaan kontras secara signifikan antara gumpalan sel darah dengan cairan serum di sekitarnya. Rangkaian ini ditutup dengan penerapan *Gaussian blur* untuk menghaluskan piksel-piksel tajam yang berpotensi dianggap sebagai bintik aglutinasi palsu oleh komputer.

Meskipun secara visual tahapan pra-pemrosesan ini berhasil menghasilkan citra berukuran 64×64 piksel yang bersih, tajam, dan seragam, performa akhir sistem masih tertahan di angka 42,86%. Hal ini menunjukkan bahwa visualisasi yang optimal bagi mata manusia tidak serta-merta berbanding lurus dengan kemudahan mesin dalam mengenali pola apabila tidak ditunjang oleh metode ekstraksi fitur matematis yang tepat, sebagaimana yang akan dibahas pada poin berikutnya.

B. Data Splitting

Setelah melalui rangkaian manipulasi citra pada fase pra-pemrosesan, tahapan krusial berikutnya sebelum proses pelatihan model adalah pembagian dataset (*data splitting*) menjadi data latih (*training set*) dan data uji (*testing set*). Pada eksperimen ini, dataset dibagi dengan proporsi rasio 80% untuk data latih (83 sampel) dan 20% untuk data uji (21 sampel). Konfigurasi pembagian ini dipilih untuk memastikan algoritma memiliki kecukupan referensi objek saat fase pembelajaran jarak sekaligus menyisakan porsi data yang valid untuk menguji kemampuan generalisasi model KNN.

Meskipun pembagian data diimplementasikan menggunakan metode acak yang proporsional (*Stratified Shuffle Split*) guna mempertahankan sebaran persentase kelas dari dataset asli, keterbatasan jumlah total data (104 citra) menciptakan tantangan besar saat disebar ke dalam 7 kelas golongan darah yang berbeda. Karakteristik dataset yang tidak seimbang (*imbalanced dataset*) menyebabkan subset data uji memiliki jumlah sampel yang sangat kritis pada beberapa kelas tertentu, seperti kelas *AB_Positive* yang hanya diwakili oleh 1 sampel citra, serta kelas *O_Negative* yang hanya memiliki 2 sampel citra.

Kondisi sebaran data yang sangat minim pada subset pengujian ini menjadi faktor kunci yang mempengaruhi stabilitas metrik evaluasi model. Ketika sebuah kelas hanya memiliki 1 atau 2 sampel uji, satu saja kesalahan prediksi akan langsung menjatuhkan nilai sensitivitas (*Recall*) kelas tersebut sebesar 50% hingga 100%. Pembagian data ini membuktikan bahwa meskipun secara teoretis rasio 80:20 adalah standar yang ideal, efektivitasnya sangat bergantung pada kuantitas total sampel. Tanpa jumlah data awal yang masif pada tiap-tiap kelas, pemisahan data secara acak tetap berisiko menghasilkan ruang pengujian yang kurang representatif bagi performa model secara objektif.

C. Hasil Optimasi Parameter K

Tahap awal pengujian model dilakukan untuk mencari nilai parameter tetangga terdekat (*K*) yang paling optimal pada algoritma *K-Nearest Neighbor* (KNN). Evaluasi dilakukan secara iteratif menggunakan subset data uji dengan menguji variasi nilai *K* dari rentang 1

hingga 15, di mana fungsi pengukur kedekatan antar-vektor fitur dihitung menggunakan metode Euclidean Distance.

```
K=1 -> Accuracy=0.3333
K=2 -> Accuracy=0.2857
K=3 -> Accuracy=0.4286
K=4 -> Accuracy=0.3333
K=5 -> Accuracy=0.3810
K=6 -> Accuracy=0.4286
K=7 -> Accuracy=0.4286
K=8 -> Accuracy=0.3810
K=9 -> Accuracy=0.3333
K=10 -> Accuracy=0.2857
K=11 -> Accuracy=0.3333
K=12 -> Accuracy=0.2857
K=13 -> Accuracy=0.2857
K=14 -> Accuracy=0.2857
K=15 -> Accuracy=0.2857

Best K : 3
Best Accuracy : 0.42857142857142855
```

Gambar 3. Hasil nilai K

Berdasarkan pergerakan nilai akurasi yang dihasilkan oleh program, performa model KNN berada pada tingkat yang cukup rendah dan mengalami fluktuasi yang konstan. Tingkat akurasi tertinggi (*optimum K*) berhasil diperoleh pada saat nilai $K = 3$ dengan raihan akurasi sebesar 42,86% (atau 0.42857). Konfigurasi parameter lainnya menunjukkan performa yang lebih rendah, di mana nilai $K = 1$ hanya menghasilkan akurasi 33,33%, sementara peningkatan nilai K yang lebih besar (dari $K = 5$ hingga $K = 15$) membuat akurasi tertahan secara mendatar di angka 38,10%. Berdasarkan hasil eksplorasi parameter ini, arsitektur model KNN dikunci pada konfigurasi $K = 3$ untuk pengujian performa lebih lanjut.

D. Hasil Evaluasi Klasifikasi Model KNN

Setelah mengunci parameter terbaik pada $K = 3$, model diuji menggunakan 21 sampel data uji (*testing set*) yang mewakili 7 kelas variasi golongan darah ABO dan Rhesus. Berdasarkan laporan klasifikasi (*classification report*) yang diekstrak dari sistem, model KNN mencatatkan nilai Akurasi Global sebesar 43% (0.43), dengan rata-rata makro (*Macro Avg*) sebesar 34% dan rata-rata berbobot (*Weighted Avg*) sebesar 42%.

Secara rinci, performa klasifikasi algoritma KNN ($K = 3$) untuk masing-masing kelas target dijabarkan pada tabel berikut:

Tabel 1. Tabel Evaluasi

Table Performa Evaluasi				
Kelas Target	Precision	Recall	F1-Score	Jumlah sample (support)
A_Negative	0.50	0.50	0.50	4
A_Positive	0.40	0.50	0.44	4
AB_Negative	0.00	0.00	0.00	3
AB_Positive	0.00	0.00	0.00	1

B_Positive	0.50	0.33	0.40	3
O_Negative	0.00	0.00	0.00	2
O_Positive	0.57	1.00	0.73	4
Macro Avg	0.28	0.33	0.30	21
Weighted Avg	0.35	0.43	0.37	21

E. Analisis Matrix Confusi (*Confusion Matrix*)

Guna membedah penyebaran prediksi dan memetakan kesalahan klasifikasi (*misclassification*) secara detail, hasil pengujian direpresentasikan ke dalam *Confusion Matrix*. Dari total 21 sampel data uji, model KNN berhasil memprediksi dengan benar sebanyak 15 sampel citra, sementara 6 sampel lainnya mengalami salah klasifikasi.

	precision	recall	f1-score	support
AB_Negative	0.20	0.33	0.25	3
AB_Positive	0.00	0.00	0.00	1
A_Negative	0.67	1.00	0.80	4
A_Positive	0.67	0.50	0.57	4
B_Positive	0.00	0.00	0.00	3
O_Negative	0.00	0.00	0.00	2
O_Positive	0.33	0.50	0.40	4
accuracy			0.43	21
macro avg	0.27	0.33	0.29	21
weighted avg	0.35	0.43	0.37	21

Gambar 4. Analisis Confusion Matrix

Berdasarkan visualisasi matriks konfusi, rincian penyebaran data adalah sebagai berikut:

1. *O_Positive*: Menjadi kelas dengan tingkat pengenalan paling tinggi, di mana seluruh sampel asli (4 sampel) berhasil dideteksi seluruhnya dengan benar (*Recall* = 1.00). Namun, kelas ini menerima 3 sampel usupan salah prediksi dari kelas lain (*A_Negative*, *AB_Negative*, dan *B_Positive*), sehingga nilai *Precision*-nya turun menjadi 0.57.
2. *A_Positive*: Dari 4 sampel asli yang diuji, model berhasil menebak 2 sampel dengan benar, sementara 2 sampel lainnya salah dideteksi sebagai *A_Negative* dan *B_Positive*.
3. *A_Negative*: Dari 4 sampel asli, hanya 2 sampel yang berhasil ditebak dengan benar, sedangkan 2 lainnya tersasar ke kelas *A_Positive* dan *O_Positive*.
4. *B_Positive*: Dari 3 sampel asli, hanya 1 sampel yang berhasil diprediksi dengan benar, sementara 2 sampel lainnya salah diidentifikasi masuk ke kelas *A_Positive* dan *O_Positive*.
5. *AB_Negative*, *AB_Positive*, dan *O_Negative*: Menjadi titik kelemahan fatal model di mana nilai metriknya 0.00. Dari seluruh total sampel asli pada ketiga kelas ini (3 sampel *AB_Negative*, 1 sampel *AB_Positive*, dan 2 sampel *O_Negative*), tidak ada satu pun sampel yang berhasil diprediksi dengan benar oleh model KNN. Sampel-sampel tersebut sepenuhnya salah diklasifikasikan menyebar ke kelas mayoritas seperti *A_Negative*, *A_Positive*, dan *B_Positive*.

F. Hasil Analisis Performa Model KNN

Rendahnya akurasi puncak model KNN yang tertahan di angka 42,86% ($K = 3$) disebabkan oleh ketidakmampuan metode representasi piksel langsung (*flattening* 4096 dimensi) dalam mengenali pola aglutinasi darah. Proses perataan matriks citra 2D menjadi vektor 1D menghilangkan informasi spasial penting mengenai susunan tekstur gumpalan darah yang acak. Akibatnya, nilai intensitas piksel mentah antara sampel darah yang menggumpal (positif) dan homogen (negatif) saling tumpang tindih secara masif di ruang multidimensi, sehingga algoritma KNN mengalami kebingungan (*underfitting*) meskipun citra telah melalui tahap pra-pemrosesan (*histogram equalization* dan *Gaussian blur*).

Di sisi lain, model juga mengalami kegagalan fatal akibat kondisi data yang tidak seimbang (*imbalanced dataset*). Algoritma KNN menunjukkan bias yang sangat kuat ke kelas mayoritas (seperti *O_Positive* dengan *Recall* 100%), namun gagal total dengan performa 0.00% pada kelas minoritas seperti *AB_Negative*, *AB_Positive*, dan *O_Negative*. Karena jumlah sampel data latih yang sangat sedikit dan miripnya nilai keabuan citra, radius pencarian jarak *Euclidean* pada kelas minoritas selalu didominasi dan terserap oleh voting kelas mayoritas di sekitarnya.

Dengan akurasi global hanya 43% serta kegagalan deteksi pada golongan darah langka dan Rhesus negatif, kombinasi sistem ini sangat berbahaya dan tidak layak digunakan dalam implementasi medis nyata seperti transfusi darah. Untuk penelitian selanjutnya, sangat direkomendasikan beralih ke metode ekstraksi fitur tekstur yang spesifik seperti GLCM atau LBP, serta menerapkan teknik penyeimbangan data seperti Data Augmentation atau SMOTE guna mengangkat sensitivitas model pada kelas minoritas.

IV. KESIMPULAN

Berdasarkan hasil eksperimen klasifikasi citra golongan darah ABO dan faktor Rhesus menggunakan algoritma *K-Nearest Neighbor* (KNN), dapat disimpulkan bahwa performa model yang dihasilkan masih tergolong rendah dengan tingkat Akurasi Global yang tertahan di angka 43% pada nilai parameter optimal $K = 3$. Rendahnya akurasi ini disebabkan oleh keterbatasan metode representasi piksel langsung (*flattening* matriks menjadi vektor 1D berdimensi 4096) yang menghilangkan informasi spasial susunan tekstur, sehingga nilai intensitas piksel antara sampel darah yang menggumpal (positif) dan homogen (negatif) saling tumpang tindih (*overlapping*) secara masif di ruang multidimensi. Selain itu, kondisi data yang tidak seimbang (*imbalanced dataset*) memicu bias pemodelan yang sangat kuat ke arah kelas mayoritas (seperti kelas *O_Positive* dengan *Recall* 100%), namun gagal total dengan mencatatkan nilai performa 0.00% pada kelas-kelas minoritas langka meliputi *AB_Negative*, *AB_Positive*, dan *O_Negative*. Mengingat tingkat sensitivitas yang rendah ini sangat berisiko menimbulkan salah diagnosis dalam prosedur transfusi medis nyata, maka untuk pengembangan penelitian selanjutnya sangat disarankan untuk mengganti ekstraksi fitur piksel langsung dengan metode ekstraksi fitur tekstur yang spesifik seperti GLCM atau LBP, serta menerapkan teknik penyeimbangan data latih menggunakan augmentasi citra atau algoritma SMOTE.

UCAPAN TERIMA KASIH

Terima kasih sebesar-besarnya kepada Universitas Nusantara PGRI Kediri dan dosen pembimbing atas fasilitas laboratorium serta bimbingannya.

DAFTAR PUSTAKA

- [1] R. Ruliati, S. Shofiyah, and P. Aisyah, "Pemeriksaan Golongan Darah Sistem Abo Pada Anak Pra Sekolah Di Tk Bina Insani Candimulyo Jombang," *Jurnal Bhakti Civitas Akademika*, vol. 8, no. 1, pp. 13–19, 2025, doi: 10.56586/jbca.v8i1.421.
- [2] R. Ariani, A. KY Astuti, R. Puspita, H. Pratomo Setyo, and D. D Nanar, "PENINGKATAN PEMAHAMAN IMUNOHEMATOLOGI BAGI PETUGAS BANK DARAH RUMAH SAKIT MELALUI PELATIHAN PELAYANAN DARAH," *Jurnal Pengabdian Multidisiplin dan Pemberdayaan Masyarakat*, vol. 2, no. 1, pp. 32–37, Jun. 2025, doi: 10.59485/abdikestrada.v2i1.106.
- [3] R. Kusumawati, "KECERDASAN BUATAN MANUSIA (ARTIFICIAL INTELLIGENCE); TEKNOLOGI IMPIAN MASA DEPAN," *ULUL ALBAB Jurnal Studi Islam*, vol. 9, no. 2, pp. 257–274, Dec. 2018, doi: 10.18860/ua.v9i2.6218.
- [4] W. R. Hasibuan, I. P. Sari, and M. Basri, "Klasifikasi Kerusakan (Cacat) pada Biji Kopi Arabika Menggunakan Algoritma KNN (K-Nearest Neighbor)," *Blend Sains Jurnal Teknik*, vol. 3, no. 4, pp. 452–459, May 2025, doi: 10.56211/blendsains.v3i4.781.
- [5] D. Apriadi and M. I. Sarif, "Optimasi Algoritma Pengolahan Citra Digital Menggunakan Metode Pembelajaran Mesin untuk Deteksi Penyakit," *Journal of Data Science and Informatics Engineering*, vol. 1, no. 1, pp. 31–37, Dec. 2025, doi: 10.64803/jodsie.v1i1.16.
- [6] N. Musfira, F. Septina, and T. Agnizarridlo, "Peran Pendidikan Klinik terhadap Pencapaian Kompetensi Radiologi Kedokteran Gigi pada Mahasiswa Profesi: Sebuah Tinjauan Literatur," *E-Prodenta Journal of Dentistry*, vol. 6, no. 1, pp. 596–610, Jun. 2022, doi: 10.21776/ub.eprodenta.2022.006.01.8.
- [7] D. Handoko and D. Meilantika, "Perbandingan Estimator Kovarians Sampel, Shrinkage, dan Robust dengan Data Sintetis Gaussian," *Journal of Technology and Data Science*, vol. 3, no. 2, pp. 258–269, Dec. 2025, doi: 10.59025/e4a8e8de.
- [8] N. A. Jafar, "Smart Classification: Perbandingan Algoritma Gaussian Naive Bayes dan K-Nearest Neighbor untuk Prediksi Kinerja Akademik Mahasiswa," *Telekontran : Jurnal Ilmiah Telekomunikasi, Kendali dan Elektronika Terapan*, vol. 14, no. 1, pp. 1–11, Apr. 2026, doi: 10.34010/telekontran.v14i1.18682.
- [9] D. Wicaksono, D. P. Almeyda, I. M. M. Putra, and L. Malihatuningrum, "Analisis Perbandingan Metode Pra Pemrosesan Citra untuk Deteksi Tepi Canny pada Citra Berbagai Kondisi Jalan menggunakan Bahasa Pemrograman Python," *JURNAL TEKNOLOGI DAN ILMU KOMPUTER PRIMA (JUTIKOMP)*, vol. 7, no. 1, pp. 17–31, Apr. 2024, doi: 10.34012/jutikomp.v7i1.3872.
- [10] A. K. Wardhani, R. I. Sudra, E. Nugraha, and A. N. Putri, "Optimasi Nilai K pada Algoritma K-Nearest Neighbor untuk Klasifikasi Kesehatan Janin," *Journal of Computer Science and Technology (JCS-TECH)*, vol. 5, no. 1, pp. 44–51, May 2025, doi: 10.54840/jcstech.v5i1.360.
- [11] Achhmad Agam and Supatman, "Analisa Citra Warna Darah Reject Berdasarkan Fitur Histogram Menggunakan KNN," *Elkom: Jurnal Elektronika dan Komputer*, vol. 18, no. 2, pp. 316–324, Dec. 2025, doi: 10.51903/elkom.v18i2.3332.
- [12] Z. Jannah, R. Kurniawan, and S. Anwar, "STUDI ALGORITMA NEURAL NETWORK DALAM KLASIFIKASI SENTIMEN PENGGUNA SHOPEE: PENINGKATAN AKURASI MODEL," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6113.
- [13] N. T. Ujianto, Gunawan, H. Fadillah, A. P. Fanti, A. D. Saputra, and I. G. Ramadhan, "Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 1, pp. 33–43, Feb. 2025, doi: 10.24246/itexplore.v4i1.2025.pp33-43.