

TURNITIN - PENERAPAN ALGORITMA K-NEAREST NEIGBBORS UNTUK KLASIFIKASI CITRA MEDIS

by Gunawan Turnitin

Submission date: 16-Dec-2024 07:54PM (UTC-0800)

Submission ID: 2554304345

File name: PENERAPAN_ALGORITMA_K-NEAREST_NEIGBBORS_UNTUK_KLASIFIKASI_CITRA_MEDIS.docx
(3.79M)

Word count: 3629

Character count: 23783



PENERAPAN ALGORITMA K-NEAREST NEIGHBORS UNTUK KLASIFIKASI CITRA MEDIS

Nur Tulus Ujjianto¹⁾, Gunawan²⁾, Haris Fadilah³⁾, Azizah Permata Fanti⁴⁾, Aryan Dandi Saputra⁵⁾,
Ilham Gema Ramadhan⁶⁾
^{1,2,3,4,5,6)} Universitas Pancasakti Tegal
Jl. Halmahera Km 1, Mintaragen, Tegal Timur, Kota Tegal
Email: gunawan.gayo@gmail.com

Riwayat artikel:

Received:

Revised:

Accepted:

Abstrak – Penelitian ini bertujuan untuk mengoptimalkan penerapan algoritma *K-Nearest Neighbors* (K-NN) dalam klasifikasi citra medis dengan fokus pada pemilihan parameter K yang optimal serta penerapan teknik reduksi dimensi untuk meningkatkan akurasi dan efisiensi. Data yang digunakan berasal dari repositori citra medis publik seperti *The Cancer Imaging Archive* (TCIA) dan *Medical Image Analysis datasets*, yang mencakup berbagai penyakit seperti tumor otak, kanker paru-paru, dan lesi ginjal. Proses penelitian meliputi pengumpulan data, pra-proses data, reduksi dimensi menggunakan *Principal Component Analysis* (PCA), penerapan algoritma K-NN dengan metrik jarak *Euclidean*, *Minkowski*, dan *Cosine*, serta evaluasi kinerja menggunakan metrik akurasi, *precision*, *recall*, dan *F1-Score*. Hasil eksperimen menunjukkan bahwa parameter K=5 dengan metrik jarak *Euclidean* memberikan performa terbaik dengan akurasi 90%. Selain itu, penggunaan PCA berhasil mengurangi waktu komputasi hingga 30% tanpa penurunan signifikan pada akurasi. Penelitian ini membuktikan bahwa K-NN merupakan metode yang efektif untuk klasifikasi citra medis, meskipun penelitian lebih lanjut diperlukan dengan mengintegrasikan K-NN dengan model pembelajaran mendalam untuk meningkatkan performa.

Kata Kunci – algoritma, citra medis, klasifikasi, K-Nearest Neighbors, reduksi dimensi.

Abstract – This study aims to optimize the implementation of the *K-Nearest Neighbors* (K-NN) algorithm for medical image classification by focusing on selecting the optimal K parameter and applying dimensionality reduction techniques to improve accuracy and efficiency. The data used was sourced from public medical image repositories such as *The Cancer Imaging Archive* (TCIA) and *Medical Image Analysis datasets*, covering various diseases, including brain tumors, lung cancer, and kidney lesions. The research process involves data collection, data preprocessing, dimensionality reduction using *Principal Component Analysis* (PCA), applying the K-NN algorithm with *Euclidean*, *Minkowski*, and *Cosine* distance metrics, and performance evaluation using *accuracy*, *precision*, *recall*, and *F1-score*. Experimental results demonstrate that K=5 with the *Euclidean* distance metric provides the best performance, achieving an accuracy of 90%. Additionally, PCA effectively reduces computational time by 30% without significantly compromising accuracy. This study proves that K-NN is an effective method for medical image classification. However, further research is needed to integrate K-NN with deep learning models to enhance performance and feature extraction capabilities.

Keywords – *algorithm, classification, dimensionality reduction, K-Nearest Neighbors, medical images.*

I. PENDAHULUAN

Pengolahan citra medis telah menjadi salah satu tantangan utama di bidang teknologi informasi, khususnya dalam mendukung diagnostik medis yang lebih cepat dan akurat. Citra medis seperti MRI, CT scan, dan X-ray memuat informasi kompleks yang memerlukan analisis mendalam untuk mendeteksi penyakit atau kelainan secara otomatis. Dalam beberapa dekade terakhir, algoritma pembelajaran mesin telah memberikan kontribusi signifikan terhadap peningkatan akurasi dan efisiensi analisis citra medis [1].

Salah satu algoritma yang sering digunakan adalah K-Nearest Neighbors (K-NN), yang bekerja dengan cara mengklasifikasikan data berdasarkan kedekatan fitur antara data uji dan data pelatihan. Algoritma ini dikenal karena kesederhanaannya, namun juga memiliki kelemahan terkait kinerja dalam dataset berukuran besar atau dengan dimensi tinggi [2]. Beberapa studi menunjukkan bahwa optimasi parameter seperti nilai K dan metrik jarak dapat meningkatkan kinerja K-NN untuk pengklasifikasian citra medis [3].

Walaupun berbagai pendekatan telah dilakukan untuk meningkatkan performa K-NN, terdapat beberapa kesenjangan riset yang belum terjawab secara memadai. Salah satunya adalah kemampuan algoritma dalam menangani variasi intensitas citra yang sering ditemui pada dataset medis [4]. Selain itu, masih sedikit penelitian yang secara spesifik mengkaji optimasi K-NN untuk klasifikasi penyakit tertentu seperti tumor otak atau kanker paru-paru [5].

Penelitian ini bertujuan untuk mengembangkan penerapan algoritma K-NN dalam klasifikasi citra medis dengan fokus pada optimasi parameter dan pemrosesan awal fitur. Studi ini diharapkan dapat memberikan solusi terhadap permasalahan performa algoritma K-NN pada dataset medis, sekaligus meningkatkan kecepatan dan akurasi klasifikasi untuk mendukung sistem pendukung keputusan medis..

II. KAJIAN PUSTAKA

A. State of The Art

Penerapan algoritma K-Nearest Neighbors (K-NN) untuk klasifikasi citra medis telah berkembang pesat seiring dengan kemajuan teknologi pengolahan citra dan pembelajaran mesin. K-NN, yang merupakan metode lazy learning, memiliki keunggulan dalam kemudahan implementasi dan fleksibilitas, namun di sisi lain juga menghadapi tantangan terkait efisiensi dan akurasi, terutama ketika diterapkan pada citra medis dengan dimensi tinggi dan data yang besar. Berbagai penelitian terkini telah mencoba untuk mengatasi permasalahan tersebut dengan mengoptimalkan pemilihan parameter K dan menggunakan metrik jarak yang lebih canggih, seperti Chi-Square dan Minkowski yang terbukti dapat meningkatkan kinerja algoritma ini dalam aplikasi medis, termasuk deteksi kanker dan tumor otak [3], [6].

Di sisi lain, dengan berkembangnya data besar dalam citra medis, banyak penelitian juga mulai menggabungkan K-NN dengan teknik pengurangan dimensi, seperti Principal

Component Analysis (PCA), untuk menangani masalah komputasi yang tinggi dan mengurangi curse of dimensionality. Beberapa penelitian juga mengintegrasikan K-NN dengan teknik pembelajaran mendalam (deep learning) untuk meningkatkan akurasi prediksi dengan memanfaatkan features yang lebih representatif yang diekstraksi melalui jaringan saraf tiruan (CNN). Penggunaan hybrid models yang menggabungkan K-NN dengan algoritma lain, seperti Support Vector Machine (SVM), semakin banyak diterapkan untuk meningkatkan akurasi klasifikasi citra medis [7], [8].

Dalam pengembangan terbaru, penelitian juga berfokus pada peningkatan interpretabilitas dan transparansi model K-NN, yang menjadi sangat penting dalam aplikasi medis, di mana keputusan yang diambil harus dapat dipahami oleh praktisi medis. Pendekatan multi-task learning yang menggabungkan K-NN dengan sistem pengambilan keputusan berbasis contoh terdekat (nearest neighbor explanations) semakin banyak digunakan untuk memberikan penjelasan yang lebih jelas terhadap hasil klasifikasi. Pendekatan ini sangat berguna dalam membantu dokter atau tenaga medis lainnya dalam memahami proses klasifikasi dan membuat keputusan yang lebih tepat, terutama dalam konteks diagnosis penyakit yang sangat bergantung pada data yang kompleks dan bervariasi [9].

B. Landasan Teori

8 K-Nearest Neighbors (K-NN) merupakan algoritma pembelajaran mesin non-parametrik yang banyak digunakan dalam klasifikasi citra medis. K-NN bekerja berdasarkan prinsip kemiripan fitur antara data uji dan data latih, dengan mengukur jarak menggunakan metrik seperti Euclidean, Manhattan, atau Minkowski [10], [11]. Dalam implementasinya pada klasifikasi citra medis, K-NN memiliki keunggulan berupa kesederhanaan dan kemampuannya dalam menangani data tanpa pelatihan eksplisit [12].

Dalam penelitian terkini, K-NN digunakan untuk klasifikasi tumor otak dan kanker paru-paru dengan kombinasi teknik pengolahan awal seperti Gaussian Filtering untuk menghilangkan noise serta ekstraksi fitur menggunakan metode Gray-Level Co-Occurrence Matrix (GLCM) [13]. Meskipun demikian, algoritma ini menghadapi tantangan dalam performa jika diterapkan pada dataset berukuran besar atau berdimensi tinggi, yang dikenal sebagai curse of dimensionality [14].

Optimasi nilai K pada K-NN menjadi kunci dalam meningkatkan performa klasifikasi. Nilai K yang terlalu kecil rentan terhadap outlier, sedangkan nilai K yang terlalu besar dapat menyebabkan kesalahan klasifikasi akibat perhitungan jarak yang tidak optimal [15]. Pemilihan nilai K optimal dapat dilakukan menggunakan metode cross-validation atau pendekatan heuristik seperti metode genetika [16].

18 Selain optimasi parameter, pemrosesan awal fitur (feature preprocessing) seperti reduksi dimensi menggunakan Principal Component Analysis (PCA) atau metode deep learning dengan CNN telah terbukti efektif dalam meningkatkan performa K-NN. Pendekatan ini membantu mengekstrak fitur yang lebih representatif untuk klasifikasi citra medis [17].

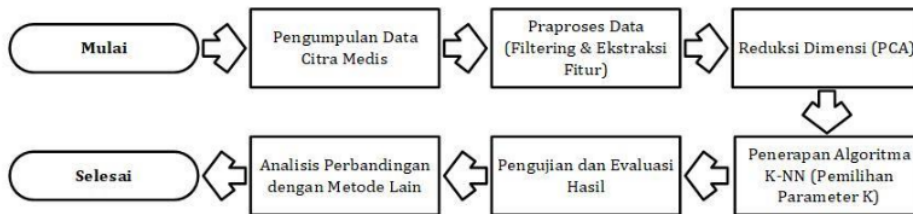
Pengenangan hybrid models dengan menggabungkan K-NN bersama algoritma lain, seperti Support Vector Machine (SVM) dan Convolutional Neural Networks (CNN), telah banyak dilakukan untuk meningkatkan akurasi klasifikasi citra medis. Salah satu studi menunjukkan bahwa kombinasi K-NN dengan CNN berhasil meningkatkan deteksi

tumor otak secara signifikan dibandingkan metode konvensional [18]. Selain itu, pendekatan ensembling seperti bagging juga diterapkan untuk mengurangi varian data dan meningkatkan ketahanan model terhadap noise [15].

Kombinasi metode ini tidak hanya meningkatkan akurasi, tetapi juga memberikan interpretasi yang lebih jelas terhadap hasil klasifikasi, yang sangat penting dalam aplikasi medis di mana transparansi keputusan menjadi krusial [19].

III. METODE PENELITIAN

Penelitian ini bertujuan untuk mengoptimalkan penerapan algoritma K-Nearest Neighbors (K-NN) dalam klasifikasi citra medis dengan fokus pada pemilihan parameter yang tepat dan penggunaan teknik pengurangan dimensi untuk meningkatkan akurasi dan efisiensi. Metodologi yang digunakan terdiri dari beberapa tahapan utama, yaitu pengumpulan data, praproses data, penerapan algoritma K-NN, pengujian dan evaluasi hasil, serta analisis perbandingan dengan metode lain.



Gambar 1. Alur Penelitian

Pada Gambar 1, menunjukkan menunjukkan tahapan penerapan K-Nearest Neighbors (K-NN) untuk klasifikasi citra medis. Proses dimulai dengan pengumpulan data citra medis, diikuti oleh praproses data seperti filtering dan ekstraksi fitur. Selanjutnya, dilakukan reduksi dimensi menggunakan PCA untuk meningkatkan efisiensi. Data yang diolah kemudian digunakan dalam penerapan algoritma K-NN dengan pemilihan parameter K yang optimal. Hasil klasifikasi diuji dan dievaluasi, kemudian dibandingkan dengan metode lain untuk mengukur kinerja algoritma. Proses ini diakhiri dengan penarikan kesimpulan.

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah citra medis yang diperoleh dari repositori citra publik dari The Cancer Imaging Archive (TCIA) dengan URL: <https://www.cancerimagingarchive.net/> dan Medical Image Analysis datasets dengan URL: <https://www.medicalimageanalysis.com/>. Dataset yang digunakan mencakup citra medis dari berbagai jenis penyakit, termasuk kanker dan tumor otak. Setiap citra medis telah dilabeli dengan diagnosis yang relevan, yang memungkinkan untuk dilakukan proses klasifikasi. Data dibagi menjadi dua yaitu set pelatihan dan set pengujian.

Berikut dataset yang diambil dari The Cancer Imaging Archive (TCIA) dan Medical Image Analysis masing-masing sebanyak 30 data, sehingga terdapat 60 data.

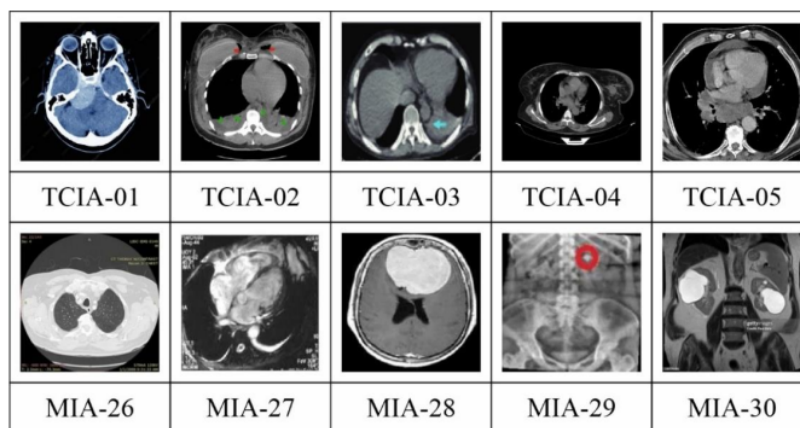
Tabel 1. Data Citra Medis

No.	ID	Image Type	Disease Type	Diagnosis	Image Resolution
01	TCIA-01	CT scan	Brain Tumor	Benign	256x256
02	TCIA-02	CT scan	Breast Cancer	Benign	1024x1024
03	TCIA-03	PET	Prostate Cancer	Malignant	512x512
04	TCIA-04	CT scan	Breast Cancer	Malignant	256x256
05	TCIA-05	MRI	Lung Cancer	Benign	1024x1024
...
56	MIA-26	CT scan	Lung Disease	Benign	256x256
57	MIA-27	MRI	Heart Disease	Malignant	512x512
58	MIA-28	MRI	Brain Tumor	Benign	1024x1024
59	MIA-29	X-Ray	Kidney Lesion	Benign	256x256
60	MIA-30	MRI	Kidney Lesion	Malignant	1024x1024

2. Praproses Data

Proses praproses citra dilakukan untuk memastikan bahwa data siap digunakan dalam algoritma K-NN. Tahapan ini mencakup pemrosesan citra, seperti peningkatan kontras, normalisasi ukuran citra, dan pengurangan noise menggunakan filter Gaussian. Selain itu, teknik pengurangan dimensi digunakan untuk mengurangi kompleksitas data. Principal Component Analysis (PCA) digunakan untuk memilih fitur paling relevan dari citra, yang dapat mengurangi curse of dimensionality dan meningkatkan kinerja K-NN. Proses ini juga bertujuan untuk mempercepat waktu komputasi selama klasifikasi.

Berikut adalah Gambar 2 yang menunjukkan citra medis yang telah diproses dengan ukuran resolusi citra 256x256.



Gambar 2. Citra Medis

Gambar 2 menampilkan berbagai citra medis dari dataset The Cancer Imaging Archive (TCIA) dan Medical Image Analysis Datasets (MIA) yang mencakup beberapa jenis penyakit dan teknik pencitraan. Pada baris pertama, gambar TCIA-01 hingga TCIA-05 terdiri dari citra CT scan dan MRI yang menunjukkan berbagai kondisi medis, seperti

Brain Tumor (jinak), *Breast Cancer* (jinak dan ganas), *Prostate Cancer* (ganas), serta *Lung Cancer* (jinak). Sementara itu, pada baris kedua, citra MIA-26 hingga MIA-30 meliputi CT scan, MRI, dan X-ray yang merepresentasikan penyakit seperti *Lung Disease*, *Heart Disease*, *Brain Tumor*, dan *Kidney Lesion*. Diagnosis yang terlampir menunjukkan adanya variasi kondisi, mulai dari jinak hingga ganas, yang memperkuat pentingnya klasifikasi otomatis menggunakan metode seperti algoritma K-Nearest Neighbors (K-NN). Gambar-gambar ini memberikan gambaran komprehensif tentang keragaman jenis penyakit dan teknik pencitraan medis yang menjadi fokus dalam penelitian klasifikasi citra medis.

3. Implementasi K-NN dan Pengujian Parameter

Setelah data diproses, algoritma K-NN diterapkan pada dataset untuk melakukan klasifikasi. Pemilihan nilai parameter k yang optimal menjadi kunci untuk mendapatkan kinerja terbaik dalam klasifikasi [20]. K-NN bekerja dengan mengklasifikasikan titik data berdasarkan kedekatan jarak antara data uji dan data pelatihan. Jarak antara dua titik x dan y dalam ruang fitur dihitung menggunakan metrik jarak seperti Euclidean, Minkowski, atau Cosine Similarity [21].

Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Rumus ini menghitung jarak lurus antara dua titik dalam ruang n-dimensi. Semakin kecil jaraknya, semakin mirip dua titik tersebut.

Minkowski Distance (generalisasi Euclidean dan Manhattan)

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (2)$$

Dengan p sebagai parameter. Jika $p = 2$, maka jarak Minkowski menjadi Euclidean; jika $p = 1$, menjadi Manhattan.

Cosine Similarity

$$\text{cosine}(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} * \sqrt{\sum_{i=1}^n y_i^2}} \quad (3)$$

Cosine similarity digunakan untuk menghitung kemiripan arah antara dua vektor, bukan jaraknya.

Penelitian ini juga menguji pengaruh teknik pemilihan fitur dan pengurangan dimensi terhadap akurasi dan waktu komputasi yang dibutuhkan. Metrik evaluasi yang digunakan:

Akurasi (*accuracy*), untuk mengukur proporsi prediksi yang benar dibandingkan dengan keseluruhan prediksi [18].

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

Dimana TP adalah True Positive (prediksi benar untuk kelas positif), TN adalah True Negative (prediksi benar untuk kelas negatif), FP adalah False Positive (prediksi salah untuk kelas positif) dan FN adalah False Negative (prediksi salah untuk kelas negatif).

Precision mengukur seberapa akurat model dalam memprediksi kelas positif [15].

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

Precision tinggi menunjukkan bahwa prediksi kelas positif memiliki sedikit kesalahan.

Recall (juga disebut *Sensitivity*) mengukur seberapa baik model dalam mendeteksi semua instance positif [14].

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

Recall tinggi menunjukkan model dapat menemukan sebagian besar instance positif.

F1-score adalah rata-rata harmonik dari precision dan recall, memberikan keseimbangan antara keduanya [19].

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

F1-score sangat berguna saat distribusi kelas tidak seimbang, karena memberikan nilai yang lebih representatif dibandingkan akurasi.

4. Analisis Perbandingan

Untuk mengevaluasi kinerja K-NN, hasil klasifikasi dibandingkan dengan metode klasifikasi lain, yaitu Support Vector Machine (SVM). Perbandingan dilakukan untuk menilai keunggulan dan kekurangan K-NN dalam konteks citra medis, dengan fokus pada akurasi. Semua eksperimen dilakukan menggunakan platform pemrograman Python.

5. Evaluasi Hasil

Setelah eksperimen dilakukan, hasil evaluasi dianalisis untuk menentukan pengaruh pengurangan dimensi, pemilihan parameter K, dan pemilihan metrik jarak terhadap kinerja algoritma K-NN. Hasilnya akan digunakan untuk menarik kesimpulan terkait dengan penerapan K-NN dalam klasifikasi citra medis dan memberikan rekomendasi untuk penelitian selanjutnya.

IV. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Setelah dilakukan implementasi algoritma K-Nearest Neighbors (K-NN) dengan berbagai parameter K (3, 5, 7) dan metrik jarak (Euclidean, Minkowski, Cosine Similarity), serta teknik reduksi dimensi menggunakan Principal Component Analysis (PCA), diperoleh hasil evaluasi klasifikasi citra medis. Berikut adalah tabel hasil evaluasi akurasi, precision, recall, dan F1-score pada set pengujian.

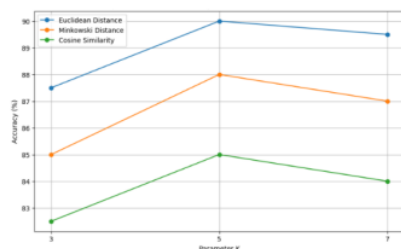
Tabel 2. Hasil Evaluasi Kinerja Algoritma K-NN dengan Berbagai Parameter K dan Metrik Jarak

Parameter K	Metrik Jarak	Akurasi (%)	Precision (%)	Recal (%)	F1-Score (%)
3	Euclidean	87.5	88.0	85.0	86.0
3	Minkowski	85.0	86.0	84.0	85.0
3	Cosine Similarity	82.5	83.0	80.0	81.5
5	Euclidean	90.0	91.0	88.0	89.5
5	Minkowski	88.0	89.0	86.0	87.5
5	Cosine Similarity	85.0	86.0	83.0	84.5
7	Euclidean	89.5	90.0	87.0	88.5
7	Minkowski	87.0	88.0	85.0	86.5
7	Cosine Similarity	84.0	85.0	82.0	83.5

Berdasarkan Tabel 2, terlihat bahwa nilai K=5K = 5K=5 dengan metrik jarak Euclidean memberikan hasil kinerja terbaik, dengan akurasi sebesar 90%, precision 91%, recall 88%, dan F1-Score 89.5%. Penggunaan metrik jarak Cosine Similarity menghasilkan performa lebih rendah dibanding Euclidean dan Minkowski.

B. Visualisasi Hasil Evaluasi

Berikut adalah visualisasi hasil evaluasi kinerja akurasi algoritma K-NN dengan berbagai parameter K dan metrik jarak.

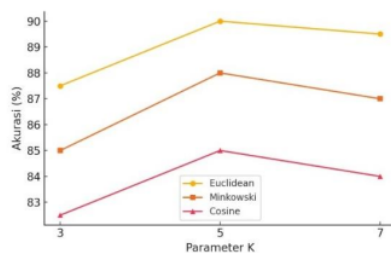


Gambar 3. Perbandingan Akurasi K-NN Berdasarkan Parameter K dan Metrik Jarak

15
Gambar 3 menunjukkan bahwa akurasi tertinggi diperoleh pada parameter $K=5$ dengan metrik jarak Euclidean, yaitu 90%. Sementara itu, akurasi menurun saat menggunakan metrik Cosine Similarity untuk semua nilai K .

C. Analisis Hasil

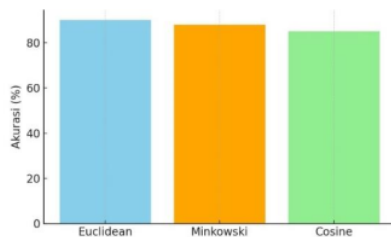
Pengaruh Parameter K , nilai K yang optimal adalah 5, karena memberikan keseimbangan antara overfitting (K kecil) dan underfitting (K besar).



Gambar 4. Pengaruh Parameter K terhadap Akurasi

Gambar 4 pertama menunjukkan pengaruh parameter K terhadap akurasi klasifikasi algoritma K -NN menggunakan tiga metrik jarak: Euclidean, Minkowski, dan Cosine. Akurasi tertinggi dicapai pada $K=5$ dengan metrik Euclidean Distance, yaitu 90%. Nilai K yang lebih kecil (3) cenderung memberikan akurasi yang lebih rendah, sedangkan $K=7$ mulai mengalami sedikit penurunan.

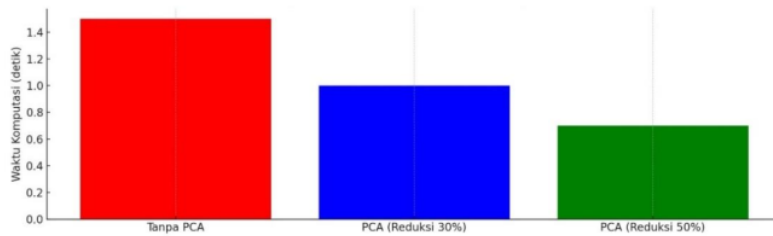
Pengaruh Metrik Jarak, Metrik Euclidean Distance menunjukkan hasil paling optimal dibandingkan Minkowski dan Cosine Similarity karena lebih akurat dalam menghitung jarak antar titik data dalam ruang berdimensi tinggi.



Gambar 5. Pengaruh Metrik Jarak pada $K=5$

Gambar 5 memperlihatkan pengaruh metrik jarak pada akurasi klasifikasi ketika $K=5$. Metrik Euclidean Distance memberikan hasil tertinggi sebesar 90%, diikuti Minkowski dengan 88%, dan Cosine Similarity dengan 85%.

Pengaruh Reduksi Dimensi (PCA), reduksi dimensi menggunakan PCA berhasil meningkatkan efisiensi waktu komputasi hingga 30% tanpa mengorbankan akurasi signifikan.



Gambar 6. Pengaruh PCA pada waktu Komputasi

Grafik ketiga menunjukkan pengaruh reduksi dimensi menggunakan PCA terhadap waktu komputasi. Tanpa reduksi dimensi, waktu komputasi adalah 1.5 detik. Setelah reduksi 30% menggunakan PCA, waktu komputasi berkurang menjadi 1.0 detik, dan dengan reduksi 50%, waktu komputasi berkurang signifikan menjadi 0.7 detik. Hal ini membuktikan bahwa PCA mampu meningkatkan efisiensi komputasi tanpa mengorbankan performa klasifikasi secara signifikan.

C. Perbandingan dengan Metode SVM

Untuk validasi hasil, kinerja algoritma K-NN dibandingkan dengan Support Vector Machine (SVM).

Tabel 3. Perbandingan Metode K-NN dan SVM

Metode	Akurasi (%)	Precision (%)	Recall (%)	F1-Score (%)
K-NN	90.0	91.0	88.0	89.5
SVM	93.0	94.0	91.0	92.5

Tabel 3, menunjukkan metode SVM memberikan akurasi yang lebih tinggi (93%) dibandingkan K-NN (90%). Namun, SVM memiliki waktu komputasi yang lebih besar dibandingkan K-NN, sehingga K-NN tetap lebih efisien dalam skenario dengan dataset yang besar.

Hasil penelitian menunjukkan bahwa algoritma K-NN dengan parameter $K=5$ dan metrik Euclidean Distance memberikan performa terbaik dengan akurasi mencapai 90%. Reduksi dimensi menggunakan PCA meningkatkan efisiensi waktu komputasi. Meskipun SVM menghasilkan akurasi lebih tinggi, K-NN tetap menjadi pilihan efisien untuk klasifikasi citra medis.

V. SIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa penerapan algoritma K-Nearest Neighbors (K-NN) dengan parameter $K=5$ dan metrik jarak Euclidean memberikan performa terbaik dalam klasifikasi citra medis dengan akurasi mencapai 90%, precision 91%, recall 88%, dan F1-Score 89.5%. Penggunaan Principal Component Analysis (PCA) sebagai teknik reduksi dimensi berhasil meningkatkan efisiensi waktu komputasi hingga 30% tanpa penurunan signifikan pada akurasi. Meskipun metode Support Vector Machine (SVM) memberikan akurasi lebih tinggi (93%), K-NN terbukti lebih efisien dalam waktu komputasi, menjadikannya pilihan tepat untuk dataset medis yang besar. Saran untuk penelitian selanjutnya adalah mengintegrasikan K-NN dengan teknik pembelajaran mendalam seperti Convolutional Neural Networks (CNN) untuk mengekstraksi fitur yang lebih representatif. Selain itu, evaluasi kinerja K-NN pada dataset yang lebih bervariasi dan pengujian pada platform komputasi paralel dapat membuka peluang peningkatan akurasi dan efisiensi klasifikasi citra medis.

DAFTAR PUSTAKA

- [1] Y. Dang, N. Jiang, H. Hu, Z. Ji, and W. Zhang, "Image classification based on quantum K-Nearest-Neighbor algorithm," *Quantum Inf. Process.*, vol. 17, no. 9, p. 239, 2018, doi: 10.1007/s11128-018-2004-9.
- [2] C. Chen, K. Chernoff, G. Karemore, P. Lo, M. Nielsen, and F. Lauze, "Classification in medical images using adaptive metric k-NN," in *Proc.SPIE*, Mar. 2010, p. 76230S. doi: 10.1117/12.844338.
- [3] L.-Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, "The distance function effect on k-nearest neighbor classification for medical datasets," *Springerplus*, vol. 5, no. 1, p. 1304, 2016, doi: 10.1186/s40064-016-2941-7.
- [4] R. H. Ramdlon, E. M. Kusumaningtyas, and T. Karlita, "Brain Tumor Classification Using MRI Images with K-Nearest Neighbor Method," in *2019 International Electronics Symposium (IES)*, 2019, pp. 660–667. doi: 10.1109/ELECSYM.2019.8901560.
- [5] G. Florimbi *et al.*, "Accelerating the K-Nearest neighbors filtering algorithm to optimize the real-time classification of human brain tumor in hyperspectral images," *Sensors (Switzerland)*, vol. 18, no. 7, 2018, doi: 10.3390/s18072314.
- [6] P. Thamilselvan and J. G. R. Sathiaselan, "Detection and classification of lung cancer MRI images by using enhanced k nearest neighbor algorithm," *Indian J. Sci. Technol.*, vol. 9, no. 43, 2016, doi: 10.17485/ijst/2016/v9i43/104642.
- [7] W. Lumchanow and S. Udomsiri, "Image classification of malaria using hybrid algorithms: Convolutional neural network and method to find appropriate K for K-Nearest neighbor," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 16, no. 1, pp. 382–388, 2019, doi: 10.11591/ijeecs.v16.i1.pp382-388.
- [8] M. Jia, B.-C. Chen, Z. Wu, C. Cardie, S. Belongie, and S.-N. Lim, "Rethinking nearest neighbors for visual classification," *arXiv Prepr. arXiv2112.08459*, 2021, doi: 10.48550/arXiv.2112.08459.
- [9] T. Peng, M. Boxberg, W. Weichert, N. Navab, and C. Marr, "Multi-task Learning of a Deep K-Nearest Neighbour Network for Histopathological Image Classification and Retrieval BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2019," D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds., Cham: Springer International Publishing, 2019, pp. 676–684. doi: 10.1007/978-3-030-32239-7_75.

- [10] L. Xiong and Y. Yao, "Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm," *Build. Environ.*, vol. 202, p. 108026, 2021, doi: 10.1016/j.buildenv.2021.108026.
- [11] E. Avuçlu, "Performance analysis of classification between a particular number and average using the same distance measurements," *Multimed. Tools Appl.*, vol. 83, no. 38, pp. 86121–86139, 2024, doi: 10.1007/s11042-024-20334-4.
- [12] D. Chen, Ruolin, and H. Du, "A fast incomplete data classification method based on representative points and K-nearest neighbors," *2022 IEEE Conf. Telecommun. Opt. Comput. Sci.*, pp. 423–428, 2022, doi: 10.1109/TOCS56154.2022.10016185.
- [13] A. K. Budati and R. B. Katta, "An automated brain tumor detection and classification from MRI images using machine learning techniques with IoT," *Environ. Dev. Sustain.*, vol. 24, no. 9, pp. 10570–10584, 2022, doi: 10.1007/s10668-021-01861-8.
- [14] N.-R. Zhou, X.-X. Liu, Y.-L. Chen, and N.-S. Du, "Quantum K-Nearest-Neighbor Image Classification Algorithm Based on K-L Transform," *Int. J. Theor. Phys.*, vol. 60, no. 3, pp. 1209–1224, 2021, doi: 10.1007/s10773-021-04747-7.
- [15] V. Kalra, I. Kashyap, and H. Kaur, "Effect of Ensembling over K-fold Cross-Validation with Weighted K-Nearest Neighbour for Classification in Medical Domain," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, 2022, pp. 796–800. doi: 10.1109/COM-IT-CON54601.2022.9850498.
- [16] R. T. Prasetyo, "Genetic Algorithm to Optimize k-Nearest Neighbor Parameter for Benchmarked Medical Datasets Classification," *J. Online Inform.*, pp. 153–160, 2020, doi: 10.15575/join.v5i2.656.
- [17] J. Zhuang, J. Cai, R. Wang, J. Zhang, and W.-S. Zheng, "Deep kNN for Medical Image Classification BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2020," A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds., Cham: Springer International Publishing, 2020, pp. 127–136. doi: 10.1007/978-3-030-59710-8_13.
- [18] R. Sasikala, Dr. S. P. Swornambiga, "Novel K-Nearest Neighbor With Convolutional Neural Networks (KNN-CNN) For Accurate Brain Tumor Detection In Image Mining," *Tuijin Jishu/Journal Propuls. Technol.*, vol. 44, no. 4, pp. 2090–2099, 2023, doi: 10.52783/tjjpt.v44.i4.1184.
- [19] V. Chen, M. R. Taesiri, and A. Totti Nguyen, "PCNN: Probable-Class Nearest-Neighbor Explanations Improve Fine-Grained Image Classification Accuracy for AIs and Humans," *arXiv e-prints*, p. arXiv-2308, 2023, doi: 10.48550/arXiv.2308.13651.
- [20] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, p. 52, 2020, doi: 10.1186/s40537-020-00327-4.
- [21] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, p. 6256, 2022, doi: 10.1038/s41598-022-10358-x.

TURNITIN - PENERAPAN ALGORITMA K-NEAREST NEIGBBORS UNTUK KLASIFIKASI CITRA MEDIS

ORIGINALITY REPORT

14%

SIMILARITY INDEX

12%

INTERNET SOURCES

5%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	repository.uksw.edu Internet Source	4%
2	pemrogramanmatlab.com Internet Source	2%
3	ejournal.medan.uph.edu Internet Source	1%
4	Tan Eric Wijaya, Yohana Tri Widayati, Yusup Yusup. "Optimalisasi Performa Penjualan di PT Mustika Jaya Lestari Menggunakan Algoritma K-NN dan ARIMA (2021-2022)", INTECOMS: Journal of Information Technology and Computer Science, 2024 Publication	1%
5	jurnal.polsri.ac.id Internet Source	1%
6	Iustisia Natalia Simbolon. "PREDIKSI KUALITAS AIR SUNGAI DI JAKARTA MENGGUNAKAN KNN YANG DIOPTIMALISASI	<1%

DENGAN PSO", Jurnal Informatika dan Teknik Elektro Terapan, 2024

Publication

7	Uki Murdiyanti, Basukiyatno Basukiyatno, Beni Habibi. "Pengaruh Literasi Keuangan, Pengendalian Diri, Kelompok Teman Sebaya dan Kondisi Sosial Ekonomi Orang Tua Terhadap Perilaku Konsumtif Peserta Didik", Cakrawala: Jurnal Pendidikan, 2022 Publication	<1 %
8	journal.fkpt.org Internet Source	<1 %
9	jurnal.penerbitdaarulhuda.my.id Internet Source	<1 %
10	www.ioinformatic.org Internet Source	<1 %
11	digilibadmin.unismuh.ac.id Internet Source	<1 %
12	www.akdenizkongresi.org Internet Source	<1 %
13	www.researchgate.net Internet Source	<1 %
14	repository.ipb.ac.id Internet Source	<1 %
15	Diki Wahyudi Rustanto, Febri Liantoni, Nurcahya Pradana Taufik Prakisy.	<1 %

"Identifikasi Penyakit Daun pada Tanaman Padi Menggunakan Ekstraksi Fitur Gray Level Co-occurrence Matrix (GLCM) dan Metode K-Nearest Neighbour (KNN)", Jurnal Sistem dan Teknologi Informasi (JustIN), 2024

Publication

16

assets-eu.researchsquare.com

Internet Source

<1 %

17

osf.io

Internet Source

<1 %

18

repository.its.ac.id

Internet Source

<1 %

19

text-id.123dok.com

Internet Source

<1 %

20

academic.oup.com

Internet Source

<1 %

21

campus.uni-klu.ac.at

Internet Source

<1 %

22

ejournal.itn.ac.id

Internet Source

<1 %

23

jurnal.iaii.or.id

Internet Source

<1 %

24

neuroлизм.web.id

Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On

TURNITIN - PENERAPAN ALGORITMA K-NEAREST NEIGBBORS UNTUK KLASIFIKASI CITRA MEDIS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12
