

Analisis Performa Algoritma K-Nearest Neighbor dalam Klasifikasi Penyakit Tumor Otak

Komang Gede Bagus Devit Aditiya^{a1}, I Wayan Santiyasa^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹aditiya.2208561073@student.unud.ac.id
²santiyasa@unud.ac.id

Abstract

Brain tumor disease poses a significant health challenge globally, including in Indonesia. Detecting brain tumors early is crucial for effective treatment. In this study, we investigated the performance of the K-Nearest Neighbor (KNN) algorithm in classifying brain tumor disease using brain image data. Our findings reveal that the choice of K value significantly impacts the KNN algorithm's performance. The highest accuracy of 81% was achieved with K=3, while the lowest accuracy of 66% occurred at K=7. On average, across all scenarios, the accuracy was 72.8%. These results underscore the importance of selecting the appropriate K value for optimal classification accuracy in brain tumor disease using the KNN algorithm.

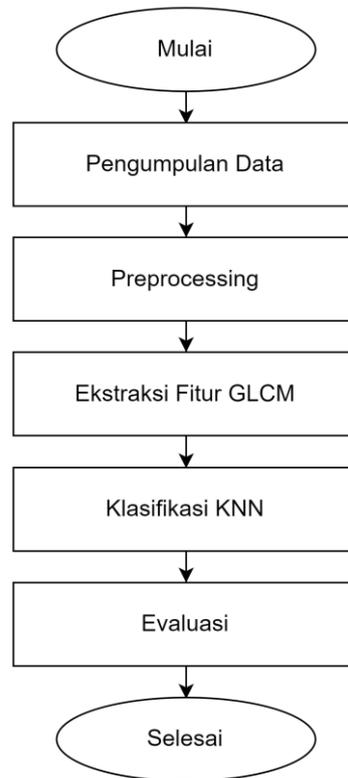
Keywords: Brain Tumor, Classifier, K-Nearest Neighbor, Grayscale, Accuracy

1. Pendahuluan

Tumor otak merupakan penyakit serius dan semakin mengkhawatirkan di seluruh dunia, termasuk Indonesia. Tumor otak bisa berasal dari pertumbuhan abnormal di otak itu sendiri (tumor primer) atau akibat penyebaran tumor (metastasis) ke bagian tubuh lain. Berbeda dengan tumor lainnya, tumor otak memiliki ciri khas yang unik. Meskipun tumor otak mungkin jinak dari sudut pandang histologis, tumor otak sering kali menjadi ganas karena lokasinya di dekat struktur vital atau di rongga tertutup yang sulit dijangkau. Tumor otak dikenal sebagai tumor paling ganas kedua setelah tumor darah (leukemia) [1], dan jumlah penderita tumor otak di Indonesia terus meningkat dalam beberapa tahun terakhir. Oleh karena itu, deteksi dini sangat penting dan perlu dilakukan sebelum tumor otak berkembang menjadi stadium lebih lanjut. Penting bagi tenaga medis dan masyarakat untuk terus mewaspadaai gejala dan perubahan yang terjadi. Menghadapi tantangan ini, penting untuk mengembangkan metode klasifikasi yang efektif untuk diagnosis dan pengobatan tumor otak. Beberapa algoritma klasifikasi yang dapat digunakan antara lain K-nearest neighbours, support vector machine, dan convolutional neural network. Salah satu metode yang sering digunakan dalam klasifikasi adalah K-Nearest Neighbors (K-NN). Metode KNN merupakan metode klasifikasi data yang mencocokkan data uji dengan data latih berdasarkan jarak mayoritas terdekat. KNN merupakan metode klasifikasi sederhana yang mampu menampung data pelatihan dalam jumlah besar [2]. Penelitian sebelumnya tentang klasifikasi tumor otak pada tahun 2020 berjudul "Klasifikasi tumor otak menggunakan metode support vector machine pada gambar resonansi magnetik" [3]. Penelitian ini menggunakan metode support vector machine (SVM) untuk mengklasifikasikan data gambar. Data set yang digunakan adalah data gambar dengan format .jpg yang berjumlah 100 buah data. Penelitian ini mencapai akurasi maksimum sebesar 76% dan akurasi rata-rata sebesar 70%. Dengan menguji keakuratan penelitian ini, penulis berharap dapat menggunakan metode berbeda untuk mengklasifikasikan tumor otak. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada pendekatan algoritmik yang diterapkan. Penelitian ini menggunakan metode K tetangga terdekat dan menggunakan ekstraksi fitur grey level co-occurrence matriks (GLCM) untuk dataset yang berbeda. Hasil akhir dari penelitian ini adalah tingkat akurasi metode K-NN dalam mengklasifikasikan data citra ke dalam label tumor dan non tumor.

2. Metode Penelitian

Pada tahapan ini akan dijelaskan urutan langkah-langkah atau proses yang akan penulis lakukan selama melakukan penelitian. Input penelitian ini berupa data gambar tumor otak, kemudian data tersebut dibagi menjadi data latih dan data uji, serta dibagi menjadi dua label yaitu gambar dengan tumor otak dan gambar tanpa tumor. Data citra akan menggunakan Gray Level Co-occurrence Matrix (GLCM) untuk mengekstraksi fitur. Output yang akan dihasilkan merupakan evaluasi algoritma KNN dengan ekstraksi fitur GLCM. Proses penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu “Brain Tumor Classification” yang diperoleh dari website kaggle.com. Datanya berupa file berformat “.jpg” dengan pixel yang bervariasi. Data yang digunakan sebanyak 270 data gambar, dan data tersebut dibagi menjadi 2 label yaitu 135 data “glioma_tumor” dan 135 data “no_tumor”. Kumpulan data dibagi menjadi data latih dan data uji. Data pelatihan menggunakan 90% dataset, sedangkan data pengujian menggunakan 10% sisanya.

2.2. Preprocessing

Pada tahap ini, data citra yang telah terkumpul akan diubah menjadi *grayscale* untuk menyederhanakan representasi warna. Selain itu, semua data akan disesuaikan ukurannya (resize) menjadi 200 x 200 piksel. Hal ini dilakukan untuk memastikan konsistensi ukuran dan memudahkan proses analisis, sehingga memungkinkan perbandingan yang lebih adil antara data yang berbeda.

2.3. Ekstraksi Fitur GLCM

GLCM (Gray Level Co-occurrence Matrix) adalah sebuah array yang mencatat seberapa sering pasangan piksel dengan tingkat keabuan tertentu muncul pada jarak (d) dan sudut (θ) tertentu. Biasanya, sudut yang dipertimbangkan adalah 0° , 45° , 90° , 135° , 180° , dan seterusnya, sementara jarak antar piksel umumnya diatur menjadi 1 piksel atau 2 piksel [4]. Dalam penelitian ini, empat atribut turunan dari GLCM digunakan dengan persamaan berikut:

a. Kontras

Kontras menggambarkan variabilitas intensitas antara nilai piksel yang paling terang dan paling gelap dalam pasangan piksel yang berdekatan. Perhitungan nilai kontras dapat dilakukan menggunakan persamaan nomor 1.

$$CON = \sum_{i,j} i - j^2 p_{i,j} \quad (1)$$

Keterangan:

i = baris ke- i dalam tingkat keabuan

j = kolom ke- j dalam tingkat keabuan

$p_{i,j}$ = probabilitas keabuan pada baris ke- i dan kolom ke- j

b. Entropi

Entropi mencerminkan tingkat ketidakteraturan dari distribusi nilai keabuan dalam citra. Ketika entropi memiliki nilai yang tinggi, ini menunjukkan bahwa struktur citra memiliki transisi nilai keabuan yang lebih teratur. Sebaliknya, nilai entropi yang rendah menandakan bahwa citra memiliki struktur yang kurang teratur atau lebih bervariasi dalam distribusi nilai keabuannya. Nilai Entropi dapat dilakukan menggunakan persamaan nomor 2.

$$EN = \sum_{i,j} p_{i,j} (-\ln(p_{i,j})) \quad (2)$$

Keterangan:

$p_{i,j}$ = Peluang keabuan baris ke - i , kolom ke - j

$\ln p_{i,j}$ = Logaritma natural dari $p_{i,j}$

c. Homogenitas

Homogenitas digunakan untuk mengevaluasi sejauh mana variasi intensitas dalam citra bersifat homogen atau seragam. Perhitungan nilai homogenitas (H) dapat dilakukan menggunakan persamaan nomor 3.

$$H = \sum_{i,j} \frac{p_{i,j}}{1+|i-j|} \quad (3)$$

Keterangan:

i = baris ke- i dalam tingkat keabuan

j = kolom ke- j dalam tingkat keabuan

d. Energi

Energi merupakan metrik yang mengindikasikan seberapa fokusnya pasangan piksel dengan tingkat keabuan tertentu di dalam matriks. Perhitungan energi (E) dapat dilakukan dengan menggunakan persamaan yang dinyatakan dalam nomor 4.

$$E = \sum_{i,j} (p_{i,j})^2 \quad (4)$$

i = baris ke- i dalam tingkat keabuan

j = kolom ke- j dalam tingkat keabuan

$p_{i,j}$ = probabilitas keabuan pada baris ke- i dan kolom ke- j

2.4. Klasifikasi K-NN

K-Nearest Neighbor (K-NN) merupakan algoritma klasifikasi yang memanfaatkan kedekatan antara data uji dengan data lainnya. Pada algoritma K-NN, ketika data mempunyai dimensi q maka dapat diukur jarak antara data tersebut dengan data lainnya. Jarak ini digunakan sebagai indikator kedekatan atau kemiripan antara data uji dan data latih. Nilai K pada K-NN mengacu pada jumlah data yang dipertimbangkan selama proses klasifikasi yang paling dekat dengan data uji [5]. Berikut adalah langkah-langkah dalam menghitung metode K-NN:

- a. Tetapkan nilai K ke jumlah tetangga terdekat yang akan dipertimbangkan.
- b. Hitung jarak Euclidean kuadrat antara data yang akan diklasifikasikan (contoh kueri) dan setiap sampel data yang ada.
- c. Urutkan objek berdasarkan jarak Euclidean, dari yang terkecil hingga yang terbesar.
- d. Kumpulkan kategori dari tetangga terdekat.
- e. Dengan menggunakan kelas mayoritas dari tetangga terdekat, nilai terhitung dari sebuah instance query dapat diprediksi.

2.5. Evaluasi

Pada penelitian ini, evaluasi akan dihasilkan setelah melakukan pengujian dengan menggunakan *Confusion Matrix*.

Tabel 1. Confusion Matrix

| Label | Terklarifikasi Positif | Terklarifikasi Negatif |
|---------|------------------------|------------------------|
| Positif | TP | FN |
| Negatif | FP | TN |

Keterangan:

- True Positive (TP) : Jumlah total data positif yang diklasifikasikan dengan benar.
- True Negative (TN) : Jumlah total data negatif yang diklasifikasikan dengan benar.
- False Positive (FP) : Jumlah total data positif yang seharusnya negatif, tetapi salah diklasifikasikan sebagai positif.
- False Negative (FN) : Jumlah total data negatif yang seharusnya positif, tetapi salah diklasifikasikan sebagai negatif.

Dari Confusion Matrix, nilai Akurasi dapat dihitung menggunakan persamaan berikut.

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (5)$$

3. Hasil dan Diskusi

Penelitian ini bertujuan untuk mengklasifikasi tumor otak menggunakan algoritma K-nearest neighbour (KNN) dengan ekstraksi fitur Gray Level Co-Occurrence Matrix (GLCM) yang diimplementasikan dalam bahasa pemrograman Python. Sebelum melakukan proses klasifikasi, data citra dari tumor otak akan diekstraksi fitur menggunakan GLCM. Fitur yang digunakan dari data citra, yaitu kontras, entropi, homogenitas, dan energy. Proses klasifikasi tumor otak dilakukan dengan mencari label terdekat berdasarkan jarak antara data uji dengan data latih yang ada, kemudian menentukan kelas mayoritas dari tetangga terdekat. Selain itu, penelitian ini bertujuan untuk mengevaluasi dampak pemilihan fitur parameter K pada algoritma K-Nearest Neighbor terhadap hasil klasifikasi. Nilai K yang dieksperimen pada penelitian ini antara lain $k=1$, $k=3$, $k=5$, $k=7$, dan $k=9$.

3.1. Hasil Ekstraksi Fitur GLCM

Pada tahap ekstraksi fitur menggunakan Gray Level Co-occurrence Matrix (GLCM), empat atribut utama yang digunakan adalah, kontras, entropi, homogenitas, dan energi. Masing-masing fitur ini dihitung berdasarkan distribusi spasial intensitas piksel dalam citra grayscale. Kontras mengukur variasi intensitas, entropi menilai ketidakteraturan, homogenitas menunjukkan keseragaman, dan energi mencerminkan konsentrasi pasangan piksel tertentu. Hasil ekstraksi fitur ini kemudian digunakan sebagai input dalam algoritma K-Nearest Neighbor (KNN) untuk klasifikasi tumor otak. Berikut adalah hasil dari ekstraksi fitur menggunakan GLCM.

| file | energy_0 | homogeneity_0 | entropy_0 | contrast_0 | energy_45 | homogeneity_45 | entropy_45 | contrast_45 | energy_90 | homogeneity_90 | entropy_90 | contrast_90 | energy_135 | homogeneity_135 | entropy_135 | contrast_135 | label |
|--|----------|---------------|-----------|------------|-----------|----------------|------------|-------------|-----------|----------------|------------|-------------|------------|-----------------|-------------|--------------|----------|
| 0 Dataset FX/Data LabHigioma_tumor(1).jpg | 0.011321 | 0.158942 | 0.512728 | 0.157942 | 0.010702 | 0.026107 | 0.526486 | 0.026007 | 0.011429 | 0.145660 | 0.505921 | 0.144660 | 0.010709 | 0.023613 | 0.526181 | 0.023513 | tumor |
| 1 Dataset FX/Data LabHigioma_tumor(10).jpg | 0.008094 | 0.243156 | 0.549517 | 0.242156 | 0.007531 | 0.035525 | 0.565919 | 0.035425 | 0.008145 | 0.190531 | 0.545342 | 0.189531 | 0.007536 | 0.021706 | 0.565988 | 0.021706 | tumor |
| 2 Dataset FX/Data LabHigioma_tumor(100).jpg | 0.002314 | 0.151736 | 0.681638 | 0.150736 | 0.001831 | 0.025629 | 0.699563 | 0.025529 | 0.002311 | 0.161990 | 0.678437 | 0.160990 | 0.001831 | 0.021781 | 0.701234 | 0.021781 | tumor |
| 3 Dataset FX/Data LabHigioma_tumor(101).jpg | 0.001644 | 0.208022 | 0.709067 | 0.207022 | 0.001324 | 0.031098 | 0.725966 | 0.030998 | 0.001632 | 0.199097 | 0.704715 | 0.198097 | 0.001324 | 0.037503 | 0.727992 | 0.037403 | tumor |
| 4 Dataset FX/Data LabHigioma_tumor(102).jpg | 0.001831 | 0.211256 | 0.700739 | 0.210256 | 0.001516 | 0.030291 | 0.718986 | 0.030191 | 0.001898 | 0.190781 | 0.697064 | 0.189781 | 0.001520 | 0.038188 | 0.729688 | 0.038088 | tumor |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 265 Dataset FX/Data Uji_no_tumor/image(3).jpg | 0.008466 | 0.185798 | 0.652098 | 0.184798 | 0.007914 | 0.032765 | 0.672287 | 0.032665 | 0.008543 | 0.184694 | 0.650092 | 0.183694 | 0.007983 | 0.032907 | 0.671427 | 0.032807 | no_tumor |
| 266 Dataset FX/Data Uji_no_tumor/image(6).jpg | 0.008681 | 0.147914 | 0.588961 | 0.146914 | 0.007715 | 0.020247 | 0.606783 | 0.020147 | 0.008222 | 0.076640 | 0.581509 | 0.075640 | 0.007721 | 0.019074 | 0.606590 | 0.018974 | no_tumor |
| 267 Dataset FX/Data Uji_no_tumor/image(7).jpg | 0.010982 | 0.198838 | 0.640703 | 0.198838 | 0.010417 | 0.029750 | 0.621130 | 0.029650 | 0.010955 | 0.148152 | 0.597446 | 0.147152 | 0.010421 | 0.030263 | 0.621142 | 0.030163 | no_tumor |
| 268 Dataset FX/Data Uji_no_tumor/image(8).jpg | 0.027083 | 0.259505 | 0.477446 | 0.258505 | 0.026588 | 0.047096 | 0.467764 | 0.046996 | 0.027173 | 0.189392 | 0.472408 | 0.188392 | 0.026583 | 0.033379 | 0.489239 | 0.033279 | no_tumor |
| 269 Dataset FX/Data Uji_no_tumor/image(9).jpg | 0.023219 | 0.319439 | 0.522789 | 0.318439 | 0.022681 | 0.040152 | 0.534352 | 0.040052 | 0.023354 | 0.233778 | 0.517316 | 0.232778 | 0.022697 | 0.049396 | 0.534186 | 0.049296 | no_tumor |

Gambar 2. Hasil Ekstraksi Fitur dengan GLCM

3.2. Pengujian Model

3.2.1. Pengujian dengan Nilai K Bernilai 1

Pengujian yang menggunakan parameter K=1, hasilnya menunjukkan rata-rata akurasi sebesar 77%, dengan 21 data terklasifikasi dengan benar dan 6 data terklasifikasi secara tidak tepat. Detail hasil pengujian dapat dilihat pada tabel 2.

Tabel 2. Hasil Pengujian Nilai K = 1

| Kelas | Terklarifikasi Positif | Terklarifikasi Negatif |
|---------|------------------------|------------------------|
| Positif | 10 | 4 |
| Negatif | 2 | 11 |

3.2.2. Pengujian dengan Nilai K Bernilai 3

Pengujian yang menggunakan parameter K=3, hasilnya menunjukkan rata-rata akurasi sebesar 81%, dengan 22 data terklasifikasi dengan benar dan 5 data terklasifikasi secara tidak tepat. Detail hasil pengujian dapat dilihat pada tabel 3.

Tabel 3. Hasil Pengujian Nilai K = 3

| Kelas | Terklarifikasi Positif | Terklarifikasi Negatif |
|---------|------------------------|------------------------|
| Positif | 11 | 3 |
| Negatif | 2 | 11 |

3.2.3. Pengujian dengan Nilai K Bernilai 5

Pengujian yang menggunakan parameter K=5, hasilnya menunjukkan rata-rata akurasi sebesar 70%, dengan 19 data terklasifikasi dengan benar dan 8 data terklasifikasi secara tidak tepat. Detail hasil pengujian dapat dilihat pada tabel 4.

Tabel 4. Hasil Pengujian Nilai K = 5

| Kelas | Terklarifikasi Positif | Terklarifikasi Negatif |
|---------|------------------------|------------------------|
| Positif | 9 | 5 |
| Negatif | 3 | 10 |

3.2.4. Pengujian dengan Nilai K Bernilai 7

Pengujian yang menggunakan parameter K=7, hasilnya menunjukkan rata-rata akurasi sebesar 66%, dengan 18 data terklasifikasi dengan benar dan 9 data terklasifikasi secara tidak tepat. Detail hasil pengujian dapat dilihat pada tabel 5.

Tabel 5. Hasil Pengujian Nilai K = 7

| Kelas | Terklarifikasi Positif | Terklarifikasi Negatif |
|---------|------------------------|------------------------|
| Positif | 8 | 6 |
| Negatif | 3 | 10 |

3.2.5. Pengujian dengan Nilai K Bernilai 9

Pada pengujian yang menggunakan parameter K=9, hasilnya menunjukkan rata-rata akurasi sebesar 70%, dengan 19 data terklasifikasi dengan benar dan 8 data terklasifikasi secara tidak tepat. Detail hasil pengujian dapat dilihat pada tabel 6.

Tabel 6. Hasil Pengujian Nilai K = 9

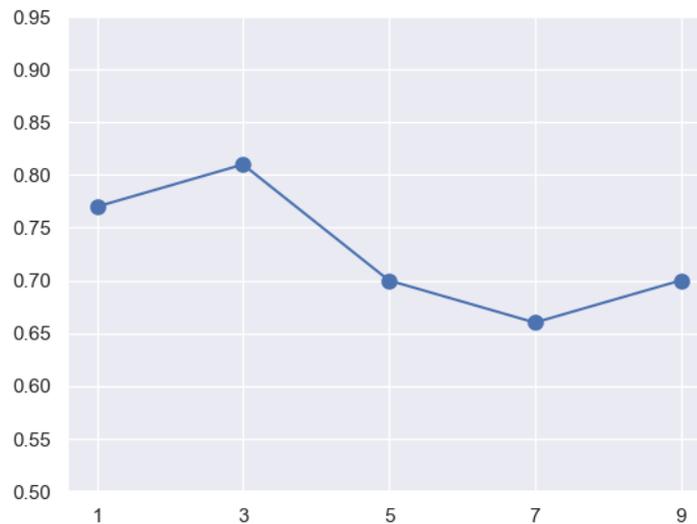
| Kelas | Terklarifikasi Positif | Terklarifikasi Negatif |
|---------|------------------------|------------------------|
| Positif | 9 | 5 |
| Negatif | 3 | 10 |

3.3. Hasil Klasifikasi

Berdasarkan pengujian yang sudah dilakukan, pemilihan nilai K berpengaruh pada performa algoritma KNN. Akurasi maksimum yang tercapai adalah 81% saat menggunakan nilai K=3, sementara akurasi minimum adalah 66% dengan nilai K=7. Rata-rata akurasi dari semua skenario adalah 72,8%, sebagaimana tercantum dalam Tabel 7. Terlihat bahwa semakin besar nilai K, akurasi cenderung menurun, menunjukkan bahwa nilai K memengaruhi akurasi. Oleh karena itu, disarankan untuk mencapai akurasi optimal dengan menggunakan nilai K=3.

Tabel 7. Rekapitan Hasil Pengujian

| No Pengujian | K | Akurasi |
|--------------|-------------|---------|
| 1 | Pengujian 1 | 1 77% |
| 2 | Pengujian 2 | 3 81% |
| 3 | Pengujian 3 | 5 70% |
| 4 | Pengujian 4 | 7 66% |
| 5 | Pengujian 5 | 9 70% |
| | Rata-rata | 72,8% |



Gambar 3. Grafik Akurasi Pengujian

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, penulis dapat mengimplementasikan algoritma K-Nearest Neighbor (KNN) untuk melakukan klasifikasi penyakit tumor otak menggunakan data citra otak. Hasil pengujian menunjukkan bahwa nilai K (jumlah tetangga terdekat) memiliki dampak yang sangat signifikan terhadap performa algoritma KNN dalam proses klasifikasi tumor otak. Penulis menemukan bahwa performa algoritma KNN terbaik dalam klasifikasi tumor otak diperoleh pada nilai K=3 dengan akurasi sebesar 81%. Sebaliknya, nilai akurasi terendah terjadi saat nilai K=7, dengan akurasi mencapai 66%. Performa algoritma KNN dengan nilai K lebih dari 3 cenderung menurun dalam melakukan klasifikasi tumor otak. Secara rata-rata, akurasi dari seluruh skenario pengujian adalah sebesar 72,8%. Penelitian ini menemukan pentingnya pemilihan nilai K yang tepat dalam melakukan klasifikasi menggunakan algoritma *K-Nearest Neighbor*.

Daftar Pustaka

- [1] S. Heranurweni, B. Destyningtias, and A. K. Nugroho, "Klasifikasi Pola Image pada Pasien Tumor Otak Berbasis Jaringan Syaraf Tiruan (Studi Kasus Penanganan Kuratif Pasien Tumor Otak)," *eLEKTRIKAL*, Vol. 10, No. 2, 2018.
- [2] A. Yuliani, A. Labellapansa, and A. Yulianti, "Klasifikasi Citra Daun Kelapa Sawit Yang Terkena Dampak Hama Menggunakan Metode K-Nearest Neighbor," *Seminar Nasional Informatika Medis*, 2019
- [3] A. S. Febrianti, T. A. Sardjono, and A. F. Babgei, "Klasifikasi Tumor Otak pada Citra Magnetic Resonance Image dengan Menggunakan Metode Support Vector Machine," *Jurnal Teknik ITS*, Vol. 9, No. 1, 2020.
- [4] D. P. Pamungkas, "Ekstraksi Citra Menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek (Orchidaceae)," *Innovation in Research of Informatics*, Vol. 1, No. 2, 2019.
- [5] A. A. D. Halim, S. Anraeni, "Analisis Klasifikasi Dataset Citra Penyakit Pneumonia Menggunakan Metode K-Nearest Neighbor (KNN)," *Indonesian Journal of Data and Science*, Vol. 2, No. 1, 2021.

Halaman ini sengaja dibiarkan kosong