

Klasifikasi Penyakit Tumor Otak Menggunakan *K-Nearest Neighbour* Berbasis *Grey Level Coocurrence Matrix*

Kartika Candra Kirana¹, Ahmad Mursyidun Nidhom², Aufar Faiq Fadhlullah³, Galih Carlos Putra Siregar⁴, Hilham Bagus Begananda⁵

1. Universitas Negeri Malang, Indonesia | kartika.candra.ft@um.ac.id
2. Universitas Negeri Malang, Indonesia | nidhom.ft@um.ac.id
3. Universitas Negeri Malang, Indonesia | faigaufar@gmail.com
4. Universitas Negeri Malang, Indonesia | galihcarlosps@gmail.com
5. Universitas Negeri Malang, Indonesia | hilham.bagus.2005356@students.um.ac.id

Abstrak

Diagnosis tumor otak dilakukan melalui tes dan prosedur medis seperti pemeriksaan fisik, MRI atau CT scan, dan biopsi. Segmentasi citra tumor otak dapat memberikan informasi penting bagi dokter dalam merencanakan perawatan dan mengevaluasi tindak lanjut yang dapat dilakukan dalam penanganan pasien tumor otak. Dibutuhkan biaya yang besar dalam diagnosa tumor otak menggunakan MRI. Sistem yang dapat secara otomatis mendeteksi tumor otak diharapkan dapat mengurangi biaya diagnosa. Pada penelitian ini, diusulkan metode ekstraksi fitur GLCM yang dikombinasi dengan metode klasifikasi K-NN untuk mengklasifikasikan MRI otak tumor dan otak sehat. Kontribusi penelitian ini adalah mengevaluasi jumlah iterasi terbaik pada arsitektur yang diusulkan pada big data. Pada fase pertama dilakukan pra-pemrosesan citra. Pra-pemrosesan yang diusulkan terdiri dari: (1) normalisasi ukuran (*resize*), (2) mengubah ke bentuk keabuan (*grayscale*), menghilangkan/mengurangi derau (*noise removal*) menggunakan median filter, dan (4) mempertajam kualitas citra (*image sharpening*). Pada fase kedua, segmentasi citra dilakukan untuk mengambil foreground. Region of interest diperoleh menggunakan segmentasi Otsu Thresholding. Pada fase ketiga GLCM diekstraksi sebagai fitur utama. Pada fase keempat, klasifikasi dikelola menggunakan algoritma K-Nearest Neighbor (KNN). Performa sistem diuji coba kepada 1150 citra pada dataset Kaggle 'Preet Viradiya'. Hasil uji coba menunjukkan tingkat akurasi terbaik mencapai 81% dengan *k* terbaik sebanyak 3. Penelitian ini dapat disimpulkan bahwa usulan arsitektur penggunaan GLCM dan KNN memiliki performa yang unggul.

Kata Kunci

Tumor Otak, KNN, GLCM, Otsu, thresholding, MRI

1. Pendahuluan

Penyakit tumor otak merupakan kondisi yang sangat berbahaya disebabkan oleh pertumbuhan sel yang tidak normal di jaringan otak. Ada dua jenis tumor otak, yaitu tumor otak primer yang berasal dari sel-sel otak dan jaringan sekitarnya. Contohnya seperti Glioma yang tercatat terjadi pada sekitar 6 dari setiap 100.000 pasien setiap tahunnya (Suta, Hartati and Divayana, 2019). Sedangkan dalam tumor otak sekunder atau metastasis, tumor menyebar ke otak dari bagian tubuh lain. Gejala dari tumor otak tergantung pada ukuran dan lokasinya, bisa timbul secara perlahan atau tiba-tiba, seperti sakit kepala, mual, muntah, kehilangan keseimbangan, dan kelemahan pada bagian tubuh tertentu.

Diagnosis tumor otak dilakukan melalui tes dan prosedur medis seperti pemeriksaan fisik, MRI atau CT scan, dan biopsi. Segmentasi citra tumor otak dapat memberikan informasi penting bagi dokter dalam merencanakan perawatan dan mengevaluasi tindak lanjut yang dapat dilakukan dalam penanganan pasien tumor otak (Malarvizhi *et al.*, 2022). Proses segmentasi manual pada citra tumor otak dapat memakan waktu yang cukup lama, tergantung pada ukuran dan tingkat kompleksitas citra yang akan diproses. Durasi yang dibutuhkan bisa bervariasi dari beberapa menit hingga beberapa jam, bahkan hari, tergantung pada jumlah citra yang harus diproses dan keterampilan operator yang melakukan segmentasi manual. Selain itu, segmentasi manual membutuhkan ketelitian dan konsistensi yang tinggi, sehingga operator harus memperhatikan setiap detail dari citra yang sedang diolah. Karena kerumitan yang tinggi inilah segmentasi gambar otak MRI masih menjadi kesulitan utama dalam komunitas medis. Namun, saat ini belum ada algoritma yang secara standar diterima yang dapat mengidentifikasi tumor otak secara akurat. Oleh karena itu, untuk menemukan tumor otak MRI secara akurat, diperlukan studi yang lebih lanjut.

Salah satu algoritma pencarian buta adalah *K-Nearest Neighbour (KNN)*. *KNN* merupakan algoritma pencarian berbasis clustering yang cocok digunakan tanpa memiliki informasi label pada dataset. Penggunaan *K-Nearest Neighbor (KNN)* menunjukkan bahwa penggunaan *KNN* mendekati akurasi hasil yang diperoleh dengan diagnosis anatomopathologist.

(Boudjella *et al.*, 2022). Raju, Baskar dan Tiwari (2022) membandingkan penggunaan *KNN* dan SVM. Uji coba pada 40 data menunjukkan bahwa *KNN* memiliki performa yang lebih baik dibandingkan SVM hingga 93,4% (Raju, Baskar and Tiwari, 2022). Wasule dan Sonar (2017) juga membandingkan penggunaan SVM dan *KNN* pada multi class. Hasil penelitian tumor otak tersebut menunjukkan bahwa penggunaan *KNN* memiliki performa yang unggul dan mencapai akurasi 86% untuk membedakan tumor otak ganas, tumor otak ringan, dan glioma (Wasule and Sonar, 2017). Havaei Jodoin dan Larochelle (2014) membandingkan penggunaan fitur vector yang paling sederhana yang dapat digunakan untuk membedakan sinyal otak. Hasil penelitian menunjukkan bahwa sinyal MR mentah dan posisi X, Y, dan Z dapat dimanfaatkan dalam mengklasifikasi tumor otak (Havaei, Jodoin and Larochelle, 2014). Machhale dkk (2015) mengevaluasi penggunaan *KNN* dalam klasifikasi tumor otak berbasis citra. Hasil penelitian menunjukkan bahwa kombinasi *KNN* dan SVM menghasilkan akurasi terbaik hingga 98% pada

50 data uji coba. Bagaimanapun usulan sistem belum diujicobakan pada *big data* (Machhale *et al.*, 2015). Shanjida Islam dan Mohiuddin (2022) mengusulkan kombinasi konsep konvolusi sebagai fitur yang digunakan dan algoritma KNN sebagai mesin klasifikasi. Penelitian ini menunjukkan performa mencapai 98%. Bagaimanapun performa belum diujicoba pada data skala besar (Shanjida, Islam and Mohiuddin, 2022).

Disisi lain, metode klasifikasi tidak dapat berjalan baik tanpa adanya dukungan fitur yang representatif. *Grey Level Coocurrence Matrix* (GLCM) cocok digunakan dalam merepresentasikan pola citra keabuan seperti MRI. Hussain dan Khunteta (2020) menunjukkan bahwa GLCM dapat merepresentasikan fitur otak secara semantic (Hussain and Khunteta, 2020). Sooch, Anand, dan Deorari (2022) mengevaluasi penggunaan 13 fitur pada Gray Level Co-Occurrence Matrix (GLCM) pada citra otak (Sooch, Anand and Deorari, 2022). Penelitian ini menunjukkan performa 13 fitur mencapai akurasi yang baik. Jany dan Jayakumar mengusulkan 2-Level Grey Level Co-Occurrence Matrix (2L-GLCM) untuk mendapatkan hasil segmentasi yang lebih baik (Jany Shabu and Jayakumar, 2020). Yohanes dkk (2021) mengkombinasikan GLCM dan fitur sinyal DWT untuk meningkatkan representasi tumor otak (Yohannes, Widhiarso and Pratama, 2021). Kelima penelitian terdahulu menunjukkan performa yang unggul GLCM dalam merepresentasikan tekstur citra keabuan.

Berdasarkan penelitian diatas, KNN telah terbukti unggul dalam data jumlah kecil. Disisi lain GLCM memiliki performa yang unggul pada data keabuan. Bagaimanapun penggunaan KNN berbasis GLCM dalam big data belum diuji coba. Kontribusi penelitian ini adalah mengujicobakan performa KNN berbasis GLCM pada *big data*. Metode yang diajukan menggabungkan beberapa metode terdahulu. Pada tahap awal, dilakukan pra-pemrosesan terhadap citra tumor otak melalui teknik normalisasi citra. Untuk mempercepat pemrosesan data dilakukan perubahan ukuran citra. Selanjutnya citra RGB di konversi menjadi grayscale. Median filter diperdaya untuk proses *blurring* yang tampak pada MRI. Peningkatan kualitas gambar juga dilakukan menggunakan teknik penajaman citra (*image sharpening*). Ekstraksi *region of interest* (ROI) dikembangkan menggunakan konsep thresholding. Selanjutnya, fitur Gray Level Co-occurrence Matrix (GLCM) digunakan untuk merepresentasikan tekstur tumor otak pada MRI. Dalam konteks penelitian ini, GLCM berguna untuk mengekstraksi tekstur dari citra MRI sebagai representasi karakteristik tumor otak. Pada tahap akhir, K-Nearest Neighbors (KNN) diset sebagai algoritma klasifikasi yang akan digunakan dalam penelitian ini dalam menentukan citra yang memiliki otak yang normal atau terdapat tumor. Perbandingan jumlah k dievaluasi dalam *big data*.

Penyakit tumor otak merupakan kondisi yang sangat berbahaya disebabkan oleh pertumbuhan sel yang tidak normal di jaringan otak. Ada dua jenis tumor otak, yaitu tumor otak primer yang berasal dari sel-sel otak dan jaringan sekitarnya. Contohnya seperti Glioma yang tercatat terjadi pada sekitar 6 dari setiap 100.000 pasien setiap tahunnya (Suta, Hartati and Divayana, 2019). Sedangkan dalam tumor otak sekunder atau metastasis, tumor menyebar ke otak dari bagian tubuh lain. Gejala dari tumor otak tergantung pada ukuran dan lokasinya, bisa

timbul secara perlahan atau tiba-tiba, seperti sakit kepala, mual, muntah, kehilangan keseimbangan, dan kelemahan pada bagian tubuh tertentu.

Grey Level Coocurrence Matrix (GLCM) cocok digunakan dalam merepresentasikan pola citra keabuan seperti MRI. Hussain dan Khunteta (2020) menunjukkan bahwa GLCM dapat merepresentasikan fitur otak secara semantic. Penelitian tersebut mencapai akurasi antara 91 persen hingga 97 persen pada enam citra uji coba (Hussain and Khunteta, 2020). Sooch, Anand, dan Deorari (2022) mengevaluasi penggunaan 13 fitur pada Gray Level Co-Occurrence Matrix (GLCM) pada citra otak (Sooch, Anand and Deorari, 2022). Penelitian ini menunjukkan performa 13 fitur mencapai akurasi yang baik. Jany dan Jayakumar mengusulkan 2-Level Grey Level Co-Occurrence Matrix (2L-GLCM) untuk mendapatkan hasil segmentasi yang lebih baik (Jany Shabu and Jayakumar, 2020). Yohanes dkk (2021) mengkombinasikan GLCM dan fitur sinyal DWT untuk meningkatkan representasi tumor otak (Yohannes, Widhiarso and Pratama, 2021). Kelima penelitian terdahulu menunjukkan performa yang unggul GLCM dalam merepresentasikan tekstur citra keabuan.

A B Malarvizi dkk (2022) memanipulasi gambar otak menggunakan median filter untuk proses *blurring*. Selanjutnya, gambar tersebut disegmentasi menggunakan teknik pengambangan batas. Algoritma pengelompokan K-means difungsikan untuk memisahkan area yang terdapat tumor dari area normal. Setelah itu, fitur-fitur pada gambar tersebut diekstraksi menggunakan GLCM. Dalam tahap klasifikasi, SVM digunakan sebagai *classifier* untuk membedakan tumor otak jinak dan ganas. Hasil akurasi dari usulan metode mencapai 80% (Malarvizhi *et al.*, 2022).

Penggunaan *K-Nearest Neighbor* (KNN) menunjukkan bahwa penggunaan KNN mendekati akurasi hasil yang diperoleh dengan diagnosis anatomopathologist (Boudjella *et al.*, 2022). Raju, Baskar dan Tiwari (2022) membandingkan penggunaan KNN dan SVM. Uji coba pada 40 data menunjukkan bahwa KNN memiliki performa yang lebih baik dibandingkan SVM hingga 93,4% (Raju, Baskar and Tiwari, 2022).

Wasule dan Sonar (2017) juga membandingkan penggunaan SVM dan KNN pada multi class. Hasil penelitian tumor otak tersebut menunjukkan bahwa penggunaan KNN memiliki performa yang unggul dan mencapai akurasi 86% untuk membedakan tumor otak ganas, tumor otak ringan, dan glioma (Wasule and Sonar, 2017).

Havaei Jodoin dan Laroche (2014) membandingkan penggunaan fitur vector yang paling sederhana yang dapat digunakan untuk membedakan sinyal otak. Hasil penelitian menunjukkan bahwa sinyal MR mentah dan posisi X, Y, dan Z dapat dimanfaatkan dalam mengklasifikasi tumor otak (Havaei, Jodoin and Laroche, 2014).

Machhale dkk (2015) mengevaluasi penggunaan KNN dalam klasifikasi tumor otak berbasis citra. Hasil penelitian menunjukkan bahwa kombinasi KNN dan SVM menghasilkan akurasi terbaik hingga 98% pada 50 data uji coba. Bagaimanapun usulan sistem belum diujicobakan pada *big data* (Machhale *et al.*, 2015).

Shanjida Islam dan Mohiuddin (2022) mengusulkan kombinasi konsep konvulusi sebagai fitur yang digunakan dan algoritma KNN sebagai mesin klasifikasi. Penelitian ini menunjukkan performa mencapai 98%. Bagaimanapun performa belum diujicoba pada data skala besar (Shanjida, Islam and Mohiuddin, 2022).

Ainani, Tri, dan Atar (2020) mengklasifikasi citra MRI dalam dua kelas, yaitu: tumor dan non-tumor. Penelitian tersebut juga mengkomparasikan tingkat akurasi klasifikasi dengan menggunakan kernel SVM yang berbeda, seperti Linear, RBF, dan Polynomial. Hasil terbaik mencapai 76% dengan menggunakan kernel Linear dan RBF. Bagaimanapun akurasi pada 50 data belum mencapai 80% (Ainani Shabrina Febrianti, Tri Arief Sardjono and Atar Fuady Babgei, 2020).

Chandra, Hafiz, dan Wijang Widhiarso mengklasifikasikan pneumonia pada citra ronsen dada. Metode klasifikasi yang digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN), sedangkan fitur yang diekstrak adalah Gray Level Co-Occurrence (GLCM). Tahap preprocessing yang digunakan meliputi *cropping*, *resizing*, *contrast stretching*, dan *thresholding*. Hasil penelitian menunjukkan akurasi terbaik mencapai 66,20% pada k sebesar 5 (Wijaya, Irsyad and Widhiarso, 2020).

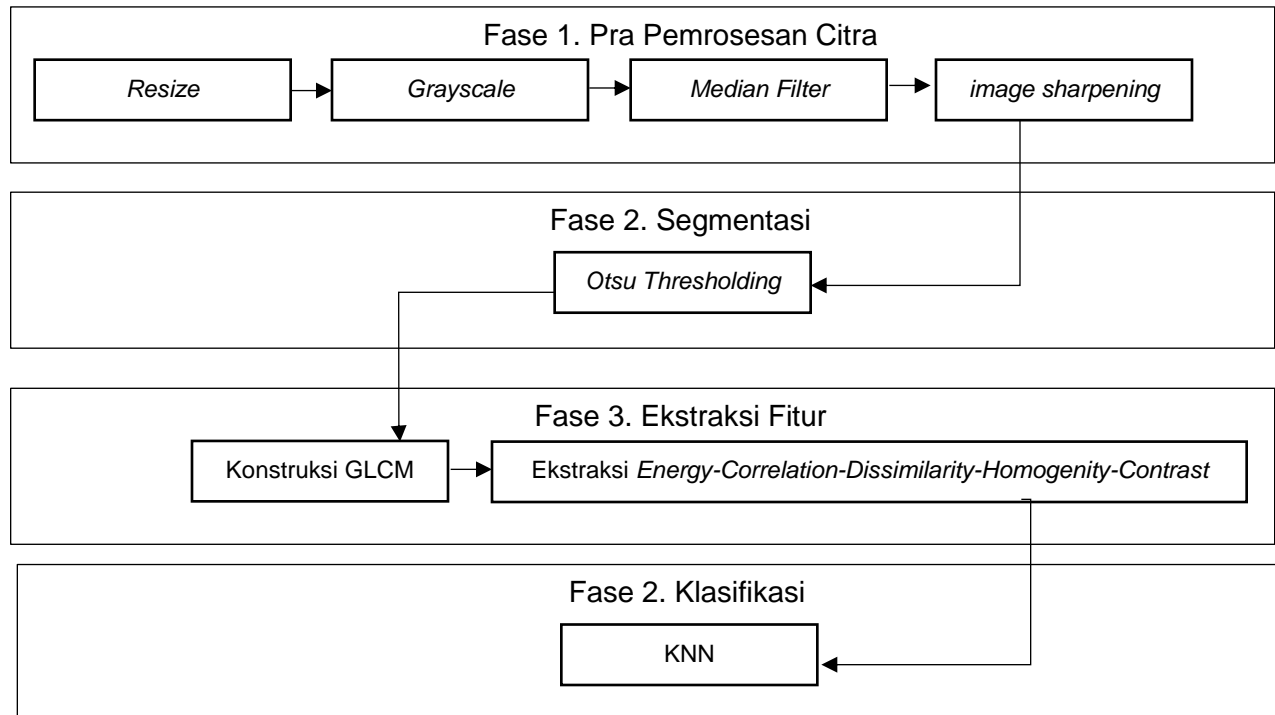
2. Metode

Metode yang akan digunakan dalam penelitian ini dimulai dengan akuisisi data citra yang diambil situs Kaggle.com yang diunggah oleh pengguna bernama Preet Viradiya (<https://www.kaggle.com/datasets/preetviradiya/brian-tumor-dataset>). Pada fase pertama dilakukan pra pemrosesan citra. Pra-pemrosesan yang diusulkan terdiri dari: (1) normalisasi ukuran (*resize*), (2) mengubah ke bentuk keabuan (*grayscale*), menghilangkan/mengurangi derau (*noise removal*) menggunakan median filter, dan (4) mempertajam kualitas citra (*image sharpening*). Pada fase kedua, segmentasi citra dilakukan untuk mengambil *foreground*. Region of interest diperoleh menggunakan segmentasi Otsu Thresholding. Pada fase ketiga GLCM diekstraksi sebagai fitur utama. Pada fase keempat, klasifikasi dikelola menggunakan algoritma K-Nearest Neighbor (KNN). Semua citra awal sebelum dilakukan pra-pemrosesan akan disimpan dalam folder yang telah dipilih dengan nama sesuai klasifikasi yang didapat untuk citra tersebut dan format file .png. Output yang didapatkan adalah dalam bentuk confusion matrix dan tabel prediksi yang memuat akurasi, presisi, recall, dan f1-score. Arsitektur KNN berbasis GLCM ditunjukkan pada Gambar 1.

1) Pra Pemrosesan

Tahap pra-pemrosesan yang digunakan pada penelitian ini pertama kali adalah *resize* yang mengubah ukuran tinggi dan/atau lebar dari sebuah citra dengan mempertahankan rasio aspek dari citra tersebut atau tidak. Tiap citra akan dinormalisasi ke ukuran 256x256 pixel. Kemudian tiap citra akan diubah ke bentuk keabuan (*grayscale*). Kemudian penerapan *noise removal* untuk menghilangkan suara pada citra. Digunakan teknik *Median Filtering* yang

merupakan teknik pemfilteran nonlinier dimana nilai pixel diganti dengan nilai median dari pixel tetangganya. Kemudian diterapkan *image sharpening* untuk mempertajam kualitas citra yang sebelumnya blur dengan meningkatkan kualitas tepi pada citra.



Gambar 1. Arsitektur Sistem

2) Segmentasi Otsu Thesholding

Segmentasi adalah proses membagi citra menjadi beberapa segmen yang umumnya digunakan untuk mendeteksi objek atau elemen yang terdapat pada citra untuk lebih mudah dianalisis. Salah satu teknik segmentasi adalah *threshold*, yang merupakan proses memisahkan pixel menjadi latar depan dan latar belakang. Teknik segmentasi yang digunakan adalah *Otsu Thresholding*. Ditemukan oleh Nobuyuki Otsu, *Otsu Thresholding* membaca nilai tiap pixel pada citra dan memproses histogramnya, lalu menggunakan algoritma untuk mencari nilai *threshold* yang optimal yang kemudian digunakan untuk membedakan latar depan dan latar belakang dalam bentuk pixel hitam dan putih (*binary*).

3) Gray Level Co-occurrence Matrix (GLCM)

Feature Extraction digunakan untuk merepresentasikan gambar dengan menangkap elemen-elemen penting pada citra, salah satu contohnya adalah *Grey Level Co-occurrence Matrix* (GLCM). GLCM digunakan untuk mengekstrak fitur dari citra yang terdiri dari *Energy*,

Correlation, Dissimilarity, Homogeneity, dan Contrast. Cara kerja GLCM adalah untuk setiap segmen, kemunculan pasangan nilai piksel dengan hubungan spasial tertentu disimpan. Matriks yang dihasilkan memberikan informasi mengenai frekuensi pasangan nilai piksel yang muncul pada jarak dan arah tertentu dalam citra. Matriks fitur keluaran ini yang kemudian akan digunakan untuk klasifikasi untuk pengklasifikasian tumor.

$$\text{Energi} = \sum \sum (p_i, j)^2, \quad \text{Pers. (1)}$$

$$\text{Korelasi} = \sum \sum (i-\mu)(j-\mu)^2, \quad \text{Pers. (2)}$$

$$\text{Dissimilarity} = \sum \sum (i-j), \quad \text{Pers. (3)}$$

$$\text{Homogenitas} = \sum \sum p(i,j)/1+(i-j)^2, \quad \text{Pers. (4)}$$

$$\text{Kontras} = \sum \sum (i-j)^2 P(i,j), \quad \text{Pers. (5)}$$

dimana matriks- i dan j , μ adalah median, dan P menunjukkan Probabilitas.

4) *K-Nearest Neighbor (K-NN)*

K-Nearest Neighbor (K-NN) adalah salah satu algoritma klasifikasi yang paling sederhana karena tidak memerlukan pengetahuan mengenai data terdahulu seperti algoritma *machine learning* lainnya. Cara kerja KNN adalah dengan menetapkan label kelas berdasarkan suara terbanyak (majority vote) atau label yang paling sering muncul di sekitar titik tertentu dari data itulah yang akan digunakan[6]. 'k' dalam KNN adalah parameter jumlah tetangga terdekat yang yang diperlukan dalam mayoritas dari proses pemungutan suara. Ada beberapa cara untuk menghitung jarak tetangga seperti *Euclidean Distance, Manhattan Distance, dan Hamming Distance* namun yang paling populer dan umum digunakan adalah *Euclidean Distance* yang merupakan perbedaan antara titik baru dan yang sudah ada, kemudian dikuadrat, dijumlahkan semuanya, lalu diakar kuadrat. Untuk persamaannya seperti pada Persamaan (6). Setelah K tetangga ditemukan, maka digunakan label kelas yang sudah ada untuk menemukan label kelas dari data baru melalui voting terbanyak.

$$D(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}, \quad \text{Pers. (6)}$$

dimana: $D(x, y)$ menunjukkan jarak antara x dan y , x_i menunjukkan data testing ke- i dan y_i merepresentasikan data training ke- i .

5) Evaluasi

Pada tahap evaluasi, 1150 data pada Kaggle ‘Preet Viradiya’ diuji coba. *Confusion Matrix* diperdagangkan untuk mengetahui perbandingan kelas target dan hasil kelas yang didapat menggunakan KNN berbasis GLCM. Tabel 2 x 2 digunakan untuk memetakan perbandingan kelas sebenarnya dari tumor dan tidak tumor dengan prediksi yang dihasilkan usulan sistem. Tabel *confusion matrix* ditunjukkan pada Tabel 1.

Tabel 1. Pengaturan jenis dan ukuran huruf dalam dokumen

		Prediksi	
		Tumor	Tidak Tumor
Groundtruth	Tumor	<i>TP</i>	<i>FN</i>
	Tidak Tumor	<i>FP</i>	<i>TN</i>

Pada Tabel 1, *TP* menunjukkan prediksi tumor yang benar, *TN* menunjukkan prediksi bukan tumor yang benar, *FN* menunjukkan prediksi tumor yang salah sedangkan *FP* menunjukkan prediksi tidak tumor yang salah. Setelah didapat hasil *confusion matrix*, selanjutnya dilakukan perhitungan akurasi presisi, *recall* dan *F1-Score* yang ditunjukkan secara berturut-turut pada Pers. (7)-(10). Akurasi dihitung untuk menunjukkan prediksi algoritma secara keseluruhan. Presisi digunakan untuk memantau prediksi tumor yang benar dari yang telah diprediksi. Recall dihitung untuk melihat hasil prediksi tumor terhadap data sebenarnya pada kelas tumor. Sementara itu, *F1 score* adalah rerata *recall* dan *presisi*.

$$\text{Akurasi} = (TP+TN):(TP+FN+TN+FP) \quad \text{Pers. (7)}$$

$$\text{Presisi} = TP:(FP+TP) \quad \text{Pers. (8)}$$

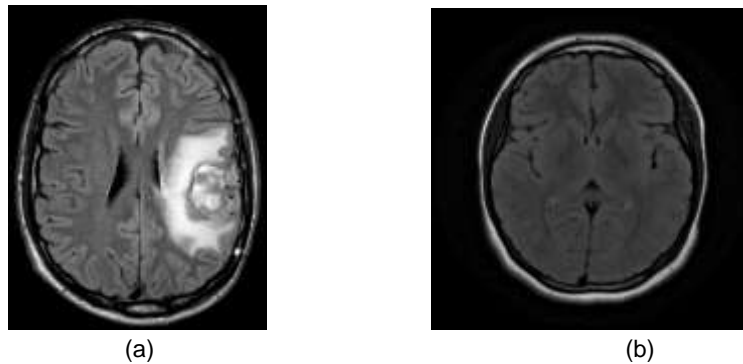
$$\text{Recall} = TP:(TP+TN) \quad \text{Pers. (9)}$$

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall}) : (\text{Precision} + \text{Recall}) \quad \text{Pers. (10)}$$

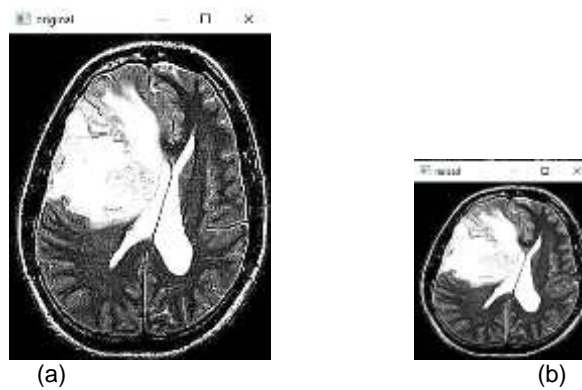
3. Hasil dan Pembahasan

Dataset yang digunakan pada penelitian ini dibagi menjadi dua kelas yakni tumor otak dan sehat dalam dua sub-folder berbeda. Dataset tersebut berisi 4,600 citra yang terbagi atas 2,513 citra otak tumor dan 2,087 otak sehat. Berikut merupakan contoh gambar dari kedua kelas tersebut. Sampel MRI ditunjukkan pada Gambar 2.

Resize adalah mengubah dan menormalisasikan ukuran citra menggunakan library OpenCV versi 4.7.0. Ukuran citra diubah menjadi ukuran 256x256 piksel. Contoh dari *resize* adalah sebagai berikut dimana citra awal, yakni tumor (1427) berukuran 278x351 piksel. Sampel proses *resize* ditunjukkan pada Gambar 3.



Gambar 2. MRI (a) Tumor (b) Tidak Tumor



Gambar 3. Resize (a) Sebelum (b) Setelah

Implementasi *Grayscale* dilakukan dengan bantuan *library* OpenCV versi 4.7.0. Berikut merupakan contoh dari citra Tumor (1427) sesudah diimplementasi *grayscale*. Implementasi *noise removal* dilakukan untuk menghilangkan *noise salt and paper* pada citra dengan teknik *median filtering* dan menggunakan *library* OpenCV versi 4.7.0. Nilai ukuran *apertur linear* (atau *ksize* dalam OpenCV) yang digunakan adalah 5. Contoh dari citra Cancer (1427) setelah diimplementasi *noise removal* yang mengakibatkan citra menjadi lebih blur ditunjukkan pada Gambar 4.



Gambar 4. Median Filter

Implementasi *image sharpening* dilakukan untuk meningkatkan kualitas dan ketajaman citra sesudah dilakukan *median filtering* pada tahap sebelumnya. Digunakan *library* OpenCV 4.7.0 dan *library* NumPy 1.24.2. Nilai *depth* yang digunakan adalah -1 dan nilai *kernel* yang digunakan adalah $[0, -1, 0]$, $[-1, 5, -1]$, $[0, -1, 0]$ yang merupakan *kernel sharpening*. Contoh dari citra Cancer (1427) setelah diimplementasi *image sharpening* ditunjukkan pada Gambar 5.



Gambar 5. Image Sharpening

Implementasi *threshold* yang digunakan adalah *threshold otsu* dengan bantuan *library* OpenCV versi 4.7.0. Nilai *threshold* yang digunakan adalah 0, nilai maksimum yang digunakan adalah 255, dan teknik yang digunakan adalah *Binary + Otsu*. Gambar 6 merupakan contoh dari citra tumor (1427) setelah diimplementasi *thresholding otsu*.



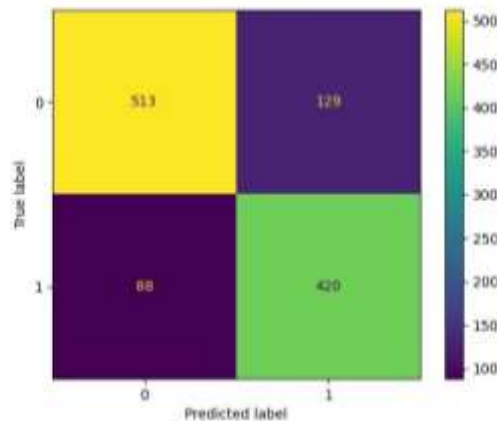
Gambar 6. Otsu Thresholding

Implementasi GLCM digunakan untuk menganalisis dan mengekstraksi fitur dalam bentuk matriks yang terdapat pada citra yang kemudian akan digunakan dalam klasifikasi. Fitur tekstur yang diambil adalah *energy*, *correlation*, *dissimilarity*, *homogeneity*, dan *contrast*. Digunakan jarak 2 dan sudut 0° , 45° , 90° , dan 135° . Digunakan bantuan *library* Pandas versi 1.5.3 dan *scikit-image* versi 0.20.0. Simulasi matriks dari ekstraksi fitur citra Cancer (1427) setelah diimplementasi GLCM dicontohkan pada Tabel 2.

Tabel 2. Pengaturan jenis dan ukuran huruf dalam dokumen

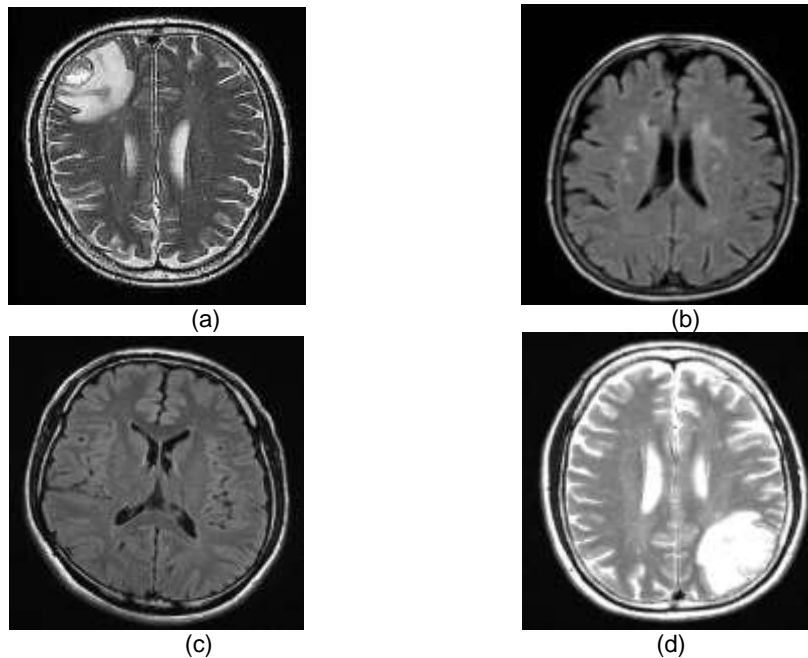
Derajat (°)	Energi	Korelasi	Dissimilarity	Homogenity	Kontras
0	0.722888269	0.765329425	23.5533034	0.907635525	6006.092366
45	0.727150055	0.784398119	21.63921569	0.915141636	5518
90	0.724818966	0.774003562	22.68270177	0.911049596	5784.088952
135	0.728681073	0.791196729	20.95686275	0.917817488	5344

Pada implementasi algoritma K-NN di penelitian ini, digunakan nilai $k = 3$ secara default dan secara pengujian memberikan hasil tertinggi dikarenakan $k = 1$ memberikan hasil yang *overfitting*, $k = 5$ memberikan hasil yang sama, dan nilai k lainnya memberikan hasil yang lebih rendah. Dataset yang berisi 4,600 citra kemudian dibagi menjadi 75% untuk *training* (3450 citra) dan 25% untuk *testing* (1150 citra), dimana data *training* terdiri atas 1,871 citra otak tumor dan 1,579 citra otak sehat, sedangkan data *testing* terdiri atas 642 citra otak tumor dan 508 citra otak sehat. Untuk *Confusion Matrix* yang didapatkan dari $k = 3$, dimana otak tumor dilambangkan dengan angka 0 dan otak sehat dilambangkan dengan angka 1, yang dapat dilihat dalam Gambar 7.



Gambar 7. Hasil Confusion Matrix Terbaik

Gambar 8 memperlihatkan contoh dari prediksi *True Positive*, *True Negative*, *False Positive*, dan *False Negative* oleh algoritma K-NN, dimana *Positive* adalah otak tumor dan *Negative* adalah otak sehat. Hasil perhitungan dari *evaluation metrics* untuk *Confusion Matrix* dalam Gambar 8 adalah seperti pada Gambar 9. Metode yang diajukan berhasil mendapatkan nilai akurasi sebesar 81%, presisi sebesar 81%, *recall* sebesar 81%, dan *f1-score* sebesar 81%.



Gambar 8. MRI Contoh (a) TP (b) TN (c) FP (d) FN

```
[INFO] evaluating k-NN classifier...
      precision    recall  f1-score   support

Brain Tumor      0.85      0.80      0.83      642
Healthy          0.77      0.83      0.79      508

 accuracy              0.81      1150
 macro avg             0.81      0.81      0.81      1150
weighted avg             0.81      0.81      0.81      1150
```

Gambar 9. Matriks Evaluasi

4. Kesimpulan

Penelitian ini menerapkan proses klasifikasi pada citra otak sehat dan otak tumor. Metode yang diusulkan dalam penelitian ini diawali dengan tahap *pre-processing* yang terdiri atas normalisasi ukuran gambar (*resize*), skala keabuan (*greyscale*), pengurangan derau (*noise removal*), mempertajam gambar (*image sharpening*). Setelah tahap *pre-processing*, dilanjutkan ke tahap segmentasi menggunakan *Otsu Thresholding*. Kemudian dilakukan proses ekstraksi fitur menggunakan *Grey Level Co-occurrence Matrix* (GLCM) menjadi matrix yang menyimpan informasi fitur *energy*, *correlation*, *dissimilarity*, *homogeneity*, dan *contrast*. Fitur-fitur ini yang kemudian dimasukkan ke dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbor*. Hasil yang diperoleh adalah akurasi sebesar 81%, presisi sebesar 81%, dan *recall* sebesar 81%

yang mana dinilai cukup baik dan akurat namun masih bisa ditingkatkan lebih jauh lagi. Penelitian selanjutnya diharapkan dapat menerepakan metode segmentasi dan ekstraksi fitur yang lain serta metode klasifikasi yang lebih kompleks seperti sistem jaringan syaraf tiruan seperti *Convolutional Neural Network* (CNN) untuk memperoleh tingkat akurasi yang lebih tinggi.

Daftar Rujukan

- Ainani Shabrina Febrianti, Tri Arief Sardjono and Atar Fuady Babgei (2020) 'Klasifikasi Tumor Otak pada Citra Magnetic Resonance Image dengan Menggunakan Metode Support Vector Machine', *Jurnal Teknik ITS*, Vol. 9, No. 1(1).
- Boudjella, A. et al. (2022) 'Machine Learning on the Diagnosis of Brain Tumors', in *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*. IEEE, pp. 1–5. doi: 10.1109/ISPA54004.2022.9786301.
- Havaei, M., Jodoin, P.-M. and Larochelle, H. (2014) 'Efficient Interactive Brain Tumor Segmentation as Within-Brain kNN Classification', in *2014 22nd International Conference on Pattern Recognition*. IEEE, pp. 556–561. doi: 10.1109/ICPR.2014.106.
- Hussain, A. and Khunteta, A. (2020) 'Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features', in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 38–43. doi: 10.1109/ICIRCA48905.2020.9183385.
- Jany Shabu, S. L. and Jayakumar, C. (2020) 'Brain Tumor Classification with MRI Brain Images Using 2-Level GLCM Features and Sparse Representation based Segmentation', in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, pp. 793–799. doi: 10.1109/ICISS49785.2020.9315971.
- Machhale, K. et al. (2015) 'MRI brain cancer classification using hybrid classifier (SVM-KNN)', in *2015 International Conference on Industrial Instrumentation and Control (IIC)*. IEEE, pp. 60–65. doi: 10.1109/IIC.2015.7150592.
- Malarvizhi, A. B. et al. (2022) 'Brain tumour classification using machine learning algorithm', *Journal of Physics: Conference Series*, 2318(1). doi: 10.1088/1742-6596/2318/1/012042.
- Raju, C. H. S., Baskar, R. and Tiwari, S. K. (2022) 'Classification of Brain Tumor Images with Improved Accuracy Using KNN and Comparing with SVM', in *2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*. IEEE, pp. 1–8. doi: 10.1109/MACS56771.2022.10023148.
- Shanjida, S., Islam, M. S. and Mohiuddin, M. (2022) 'MRI-Image based Brain Tumor Detection and Classification using CNN-KNN', in *2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*. IEEE, pp. 900–905. doi: 10.1109/GlobConET53749.2022.9872168.
- Sooch, S. K., Anand, D. and Deorari, R. (2022) 'Brain Tumor detection with GLCM feature extraction and hybrid classification approach', in *2022 International Conference on Cyber Resilience (ICCR)*. IEEE, pp. 1–5. doi: 10.1109/ICCR56254.2022.9996065.
- Suta, I. B. L. M., Hartati, R. S. and Divayana, Y. (2019) 'Diagnosa Tumor Otak Berdasarkan Citra MRI (Magnetic Resonance Imaging)', *Majalah Ilmiah Teknologi Elektro*, 18(2). doi: 10.24843/mite.2019.v18i02.p01.
- Wasule, V. and Sonar, P. (2017) 'Classification of brain MRI using SVM and KNN classifier', in *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)*.

IEEE, pp. 218–223. doi: 10.1109/SSPS.2017.8071594.

Wijaya, C., Irsyad, H. and Widhiarso, W. (2020) 'Klasifikasi Pneumonia Menggunakan Metode K-Nearest Neighbor Dengan Ekstraksi Glcm', *Jurnal Algoritme*, 1(1), pp. 33–44. doi: 10.35957/algoritme.v1i1.431.

Yohannes, Widhiarso, W. and Pratama, I. (2021) 'Combination of DWT Variants and GLCM as a Feature for Brain Tumor Classification', in *2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, pp. 197–202. doi: 10.23919/EECSI53397.2021.9624249.