

KLASIFIKASI KANKER USUS BESAR MENGGUNAKAN METODE *K-NEAREST NEIGHBOR (K-NN)*

Classification of Colon Cancer Using K-Nearesr Neighbor (K-NN) Method

Ocky Tiaramukti¹, Ratri Dwi Atmaja, S.T., M.T.², R Yunendah Nur Fuadah, S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom Bandung
¹ockytiaramukti.ot@gmail.com, ²ratriidwiatmaja@telkomuniversity.ac.id, ³yunendah@gmail.com

Abstrak

Kanker usus besar atau kanker kolorektal adalah jenis kanker yang menyerang usus besar atau bagian terakhir pada sistem pencernaan manusia. Ada beberapa jenis kanker yang menyerang usus besar manusia yaitu *Lymphoma*, *Sarcoma*, dan *Carcinoma*. Tugas akhir ini bertujuan untuk menghasilkan suatu sistem yang dapat mendeteksi dan mengklasifikasi citra usus besar ke dalam tipe kanker *Lymphoma*, kanker *Carcinoma*, atau normal.

Sistem yang dirancang pada tugas akhir ini menggunakan 198 data patologi jaringan kanker usus besar. Perancangan sistem ini mengklasifikasikan kanker usus besar mulai dari *preprocessing*, ekstraksi ciri dengan menggunakan *Principal Analysis Component (PCA)* dan melakukan klasifikasi menggunakan metode *K-NN*. Pengujian dilakukan dengan mencoba beberapa pengaturan parameter inputan *K-NN*.

Hasil dari penelitian ini berupa sistem pengolahan citra yang mampu mendeteksi Penyakit kanker usus besar dan mengklasifikasikan jenisnya. Pada penelitian ini pencapaian akurasi tertinggi yang diperoleh menggunakan klasifikasi *K-NN* dengan $K=1$ dan jenis *distance minkowski*. Hasil yang diperoleh pada pengujian sistem deteksi kanker usus besar adalah 68.52%.

Kata kunci : *kanker usus besar, PCA, K-Nearest Neighbor algorithm*

Abstract

Colon cancer or colorectal cancer is a type of cancer that attacks the colon or the last part of human digestive system. There are several types of cancer that attack the human's colon is Lymphoma, Sarcoma, and Carcinoma. The aim of this final project is to produce a system that can detect and classify images into type of colon cancer Lymphoma, cancer Carcinoma, or normal.

The system designed in this final project uses 198 data of colon cancer tissue pathology. This system will classify colon cancer start from image preprocessing, feature extraction using Principal Componen Analysis (PCA) and classification using K-Nearest Neighbor (K-NN) method. Tests will be done by trying some K-NN input parameter setting

The result of this research is an image processing system that can detect and classify type of colon cancer. In this study the highest accuracy obtained using K-NN classification with $K=1$ and using minkowski distance. The accuracy result of this system is 68,52%.

Keywords : *colorectal cancer, PCA, K-Nearest Neighbor algorithm*

1. Pendahuluan

Usus merupakan salah satu organ pada sistem pencernaan yang bermula dari akhir lambung sampai anus. Organ usus terdiri dari dua bagian, yaitu usus halus dan usus besar yang masing-masing bagian memiliki perbedaan fungsi[4]. Di dalam usus besar atau kolon berada antara usus buntu dan *rectum*[9]. Organ pencernaan ini tidak lepas dari kelainan atau penyakit. Salah satu jenis penyakit pada usus besar adalah kanker usus besar. *Lymphoma* dan *Carcinoma* adalah jenis kanker yang terjadi pada kanker usus besar.

Pendeteksian dan pengklasifikasian pada penyakit kanker lebih awal akan mendapat penanganan lebih cepat sehingga terdapat kemungkinan untuk sembuh dan tidak selalu berujung pada kematian. Saat ini pengklasifikasian kanker usus besar dilakukan secara manual, yaitu dengan mengecek sel pada usus besar kemudian diletakkan di atas *preparat* dan diamati melalui mikroskop. Pendiagnosaan dengan cara tersebut sangat berhubungan dengan kualitas penglihatan masing-masing dokter. Kesalahan atau ketidakfokusan akan mempengaruhi hasil diagnosa.

Oleh karena itu, pada tugas akhir ini dibuat sebuah perancangan sistem yang mampu mendeteksi penyakit kanker usus besar dengan memanfaatkan citra hasil pemeriksaan ahli medis. Penelitian mengenai klasifikasi kanker usus besar berbasis pengolahan citra digital telah dilakukan oleh beberapa peneliti sebelumnya. Pada penelitian ini dilakukan perancangan sistem klasifikasi kanker usus besar menggunakan ekstraksi ciri *Principal Component Analysis (PCA)* sebagai metode ekstraksi fitur dan metode *K-Nearest Neighbor (K-NN)* sebagai metode klasifikasi

1. Dasar Teori

1.1 Usus Besar

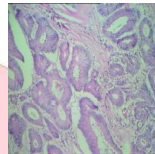
Usus besar adalah pencernaan terakhir pada manusia yang bekerja menghancurkan makanan menjadi asam amino, glukosa, lemak, air, vitamin, dan mineral dan bahan sisa hasil pencernaan terakhir yang dibuang dijadikan feses. Fungsi utama usus besar ialah untuk mengabsorpsi air kembali dan untuk mengeluarkan mucus yang berfungsi untuk melumasi dan membantu mengeluarkan feses beserta gas-gas[9].

1.2 Kanker Usus Besar

Kanker usus besar atau disebut juga kanker kolorektal merupakan salah satu jenis kanker ganas yang terjadi pada permukaan usus besar (kolon) dan rectum (bagian usus paling bawah sampai anus/dubur)[10].

1.2.1 Carcinoma

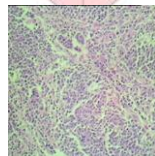
Carcinoma adalah segala jenis tumor (kanker) yang tumbuh dari sel di lapisan permukaan penutup atau membran pembatas dari organ[7].



Gambar 2.1 Sel kanker usus *carcinoma*

1.2.2 Lymphoma[1]

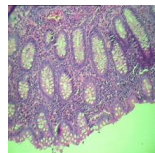
Lymphoma adalah kanker pada sel-sel limfatik dari system kekebalan tubuh. Struktur sel kanker *lymphoma* memiliki ukuran sel yang bervariasi dan menunjukkan 40% plasma sel mengalami diferensiasi. Secara morfologi, kromatin inti kasar dan padat, dan *nucleolus* tidak jelas.



Gambar 2.2 Sel kanker usus *lymphoma*

1.2.3 Normal

Berikut ini adalah gambar dari sel usus normal yang diamati dari mikroskop;



Gambar 2.3 Sel usus normal

2.3 Pengolahan Citra

Pengolahan citra adalah pemrosesan citra, khususnya dengan menggunakan komputer, menjadi citra yang kualitasnya lebih baik. Secara umum tahapan pengolahan citra digital meliputi akuisisi citra, peningkatan kualitas citra, segmentasi citra, representasi dan uraian, pengenalan dan interpretasi[8].

2.3.1 Grayscale atau citra keabuan

Citra digital keabuan (*grayscale*) setiap pikselnya mempunyai warna gradasi dari putih sampai hitam. Rentang tersebut berarti bahwa setiap piksel dapat diwakili oleh 8 bit, atau 1 byte. Rentang warna untuk keabuan sangat cocok digunakan untuk pengolahan *file* gambar.

2.3.2 Edge Detection[3]

Pendeteksian tepi merupakan langkah pertama untuk melingkupi informasi di dalam citra. Tujuan operasi pendeteksian tepi adalah untuk meningkatkan penampakan garis batas suatu daerah atau objek di dalam citra.

2.3.3 Ekualisasi Histogram[2]

Ekualisasi histogram merupakan suatu cara yang bertujuan untuk memperoleh histogram yang intensitasnya terdistribusi secara seragam pada citra. Pendekatan yang dilakukan adalah untuk mendapatkan aras keabuan yang lebih luas pada daerah yang memiliki banyak piksel dan mempersempit aras keabuan pada daerah yang berpiksel sedikit.

2.4 Ekstraksi Ciri

Ekstraksi ciri yaitu pengambilan ciri atau sifat tertentu yang dimiliki oleh suatu objek citra sehingga dapat membedakan citra yang satu dengan citra lainnya.

2.4.1 Principal Component Analysis (PCA)

PCA adalah metode dengan melakukan pemetaan atau transformasi set data dari dimensi lama ke dimensi baru (yang relatif berdimensi lebih rendah) dengan memanfaatkan teknik aljabar linear tanpa memerlukan masukan parameter tertentu dalam memberikan keluaran hasil pemetaannya. Prinsip dasar dan algoritma PCA adalah memproyeksikan *image* ke dalam bidang ruang *eigen*-nya[3].

2.5 Proses Klasifikasi

Proses klasifikasi pada tugas akhir ini menggunakan metode *K-Nearest Neighbor* (K-NN). Metode K-NN adalah metode yang melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan obyek tersebut. Tujuan dari algoritma ini adalah mengklasifikasikan obyek berdasarkan atribut dan sampel data latih.

Ada beberapa teknik yang digunakan K-NN untuk menghitung jarak tetangga terdekatnya, yaitu[5]:

a) *Euclidean Distance*

Euclidean distance menghitung akar kuadrat perbedaan dua vector, rumusnya yaitu :

$$d(x, y) = \sqrt{\sum (X_i - Y_i)^2} \quad (2.10)$$

b) *City Block*

City Block distance atau *Manhattan distance* yaitu menghitung jarak tetangga dengan cara menghitung perbedaan nilai absolut dari dua vektor. Rumusnya yaitu :

$$d(x, y) = \sum |X_i - Y_i| \quad (2.11)$$

c) *Cosine*

Cosine distance dirumuskan sebagai berikut :

$$d_i = 1 - \cos(A, B) \quad (2.12)$$

d) *Minkowski*

Minkowski distance dirumuskan sebagai berikut :

$$d(x, y) = \|x - y\|_q = \left(\sum |x - y|^q \right)^{\frac{1}{q}} \quad (2.13)$$

dimana $q \geq 1$ adalah parameter yang bisa dideteksi. Apabila $q=1$, maka jarak tersebut menjadi jarak *Manhattan* (*cityblock*). Sedangkan untuk $q=2$, jarak tersebut menjadi jarak *Euclidean*.

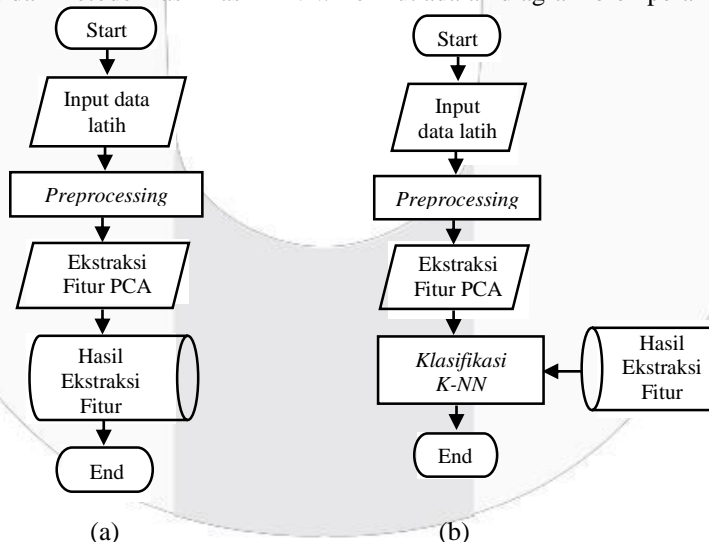
3. Perancangan dan Simulasi Sistem

3.1 Gambaran Umum Perancangan Sistem

Sistem yang dirancang merupakan sistem yang mampu mengklasifikasikan citra usus besar menjadi jenis kanker *Carcinoma*, kanker *Lymphoma*, atau normal.

3.2 Perancangan Sistem

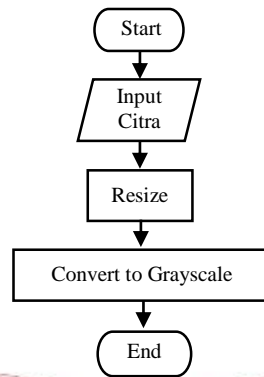
Pada penelitian ini akan dirancang sistem klasifikasi kanker usus besar menggunakan ekstraksi fitur *Principal Component Analysis* (PCA) dan metode klasifikasi K-NN. Berikut adalah diagram blok perancangan sistem.



Gambar 3.2 Diagram alir perancangan sistem klasifikasi *colon cancer*
(a) Data Latih (b) Data Uji

3.2.1 Preprocessing

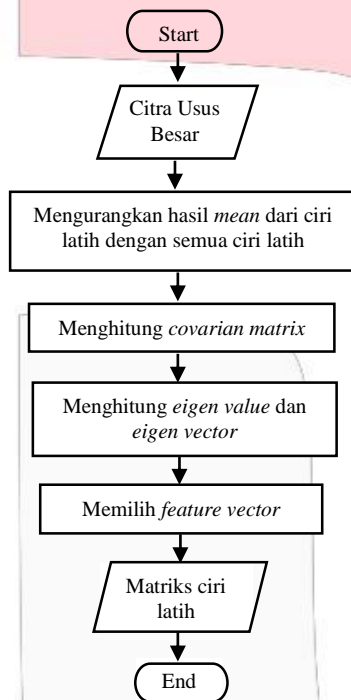
Preprocessing merupakan proses awal yang dilakukan pada suatu citra untuk meningkatkan kualitas citra. Selanjutnya dilakukan perubahan ukuran citra, lalu dikonversi menjadi citra *grayscale*. Berikut adalah diagram alir dari *preprocessing*.



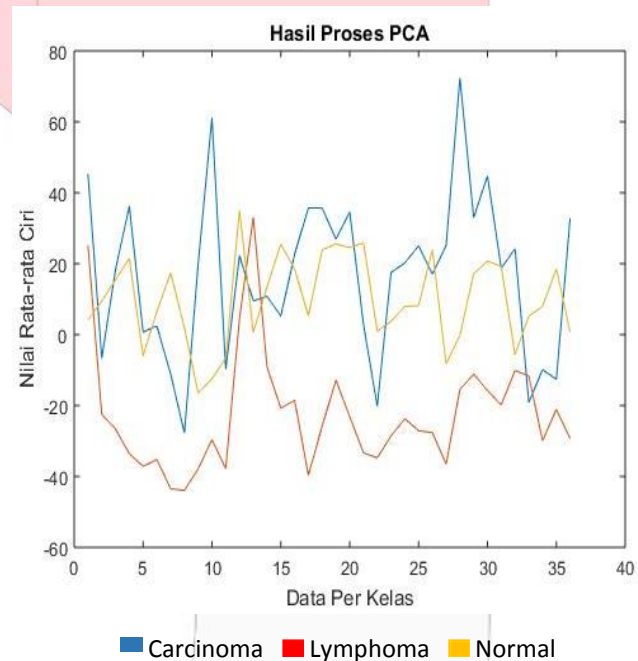
Gambar 3.4 Diagram Alir *Preprocessing*

3.2.2 Proses *Principal Component Analysis*

Principal Component Analysis adalah proses memproyeksikan *image* ke dalam biag ruang *eigen*-nya. Berikut adalah diagram alir PCA:



Gambar 3.5 Diagram Alir Proses PCA[6]



Gambar 3.6 Hasil Proses PCA

3.2.3 Perancangan Model dan Klasifikasi K-NN[5]

Setelah dilakukan proses ekstraksi ciri, selanjutnya akan dilakukan proses klasifikasi menggunakan metode *K-Nearest Network* (K-NN). Klasifikasi ini bertujuan untuk mengelompokkan 3 kelas yaitu kanker *Carcinoma*, kanker *Lymphoma*, dan normal. Pada tahap proses data latih hasil dari ekstraksi ciri akan disimpan dalam database, kemudian akan dihitung jarak terdekat berdasarkan jarak antara data yang akan diuji dengan hasil tahap proses data latih pada database sebelumnya. Berikut tahapan klasifikasi K-NN:

1. Menentukan jenis *distance* yang paling baik digunakan antara *Euclidean*, *cityblock*, *cosine*, atau *minkowski*.
2. Menentukan variabel K yang paling baik antara 1, 3, 5, 9
3. Menggunakan jenis *distance* yang paling baik sesuai hasil pengujian
4. Menghitung jarak titik data uji dengan titik data *database* terdekat
5. Menentukan jumlah titik dari masing-masing kelas *database* terdekat yang terbanyak tergantung nilai variabel K
6. Mengeluarkan hasil klasifikasi

3.3 Akurasi Sistem

Akurasi merupakan ukuran ketepatan sistem dalam mengidentifikasi masukan yang diberikan. Akurasi dari suatu sistem dapat dirumuskan sebagai berikut :

$$akurasi = \frac{Jumlah\ Data\ Benar}{Jumlah\ Citra\ Uji} \times 100\% \quad (3.1)$$

3.4 Waktu Komputasi

Waktu komputasi adalah waktu yang dibutuhkan sistem untuk melakukan suatu proses. Pada sistem ini, waktu komputasi dihitung dengan menggunakan *toolbox* yang ada pada Matlab, sehingga akan didapatkan waktu komputasi sistem.

$$Waktu\ Komputasi = Waktu\ Selesai - Waktu\ Mulai \quad (3.2)$$

4. Analisis dan Hasil Pengujian

Pengujian sistem dilakukan bertujuan untuk mengetahui performansi sistem berdasarkan parameter akurasi.

4.1 Analisis Hasil Pengujian

4.1.1 Analisis Hasil Pengujian Berdasarkan Komposisi Perbandingan Data Latih dan Data Uji

Pengujian pertama dilakukan dengan menerapkan komposisi perbandingan data latih dan data uji. Berikut adalah perbandingannya :

- 60 : 138 dengan tiap kelas 20 : 46
- 78 : 120 dengan tiap kelas 26 : 40
- 90 : 108 dengan tiap kelas 30 : 36
- 99 : 99 dengan tiap kelas 33 : 33

Tabel 4.1 Data Benar Pengujian Berdasarkan Perbandingan Data

<u>Data Latih</u> <u>Data Uji</u>	Jenis Distance	CARCINOMA	LYMPHOMA	NORMAL	AKURASI
60 : 138	Euclidean	22	8	11	29,71%
	Minkowski				
	q = 3	22	12	0	24,64%
	q = 4	22	12	0	24,64%
	q = 5	22	13	0	25,36%
	q = 6	22	19	0	29,71%
78 : 120	Euclidean	21	9	10	33,33%
	Minkowski				
	q = 3	5	15	1	17,5%
	q = 4	5	15	1	17,5%
	q = 5	5	20	1	21,67%
	q = 6	5	20	1	21,67%
90 : 108	Euclidean	8	8	28	40,74%
	Minkowski				
	q = 3	4	5	36	41,67%
	q = 4	4	5	36	41,67%
	q = 5	4	11	36	47,22%
	q = 6	4	11	36	47,22%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%
99 : 99	Euclidean	7	7	26	40,40%
	Minkowski				
	q = 3	4	5	31	40,40%
	q = 4	4	5	33	42,42%
	q = 5	4	9	33	46,46%
	q = 6	4	9	33	46,46%

Berdasarkan tabel 4.1 hasil pengujian perbandingan komposisi data latih dan data uji, komposisi 90:108 atau 30 data latih dan 36 data uji untuk masing-masing kelas menggunakan jenis *distance minkowski* dengan nilai $q = 5$ menunjukkan performansi terbaik di antara perbandingan komposisi lainnya, yaitu dengan nilai akurasi sebesar 47,22%. Hal ini dikarenakan data latih dianggap cukup untuk melatih data uji yang berjumlah 36 sehingga pola latih data tidak terlalu ketat.

4.1.2 Analisis Hasil Pengujian Berdasarkan *Resize*

Pengujian ini merupakan pengujian performansi yang dilakukan dengan melakukan *resize* pada preprocessing dengan mengubah ukuran piksel citra menjadi 560×560, 450×450, 360×360, 270×270, 200×200 piksel.

Tabel 4.2 Data Benar Pada Pengujian Berdasarkan *Resize*

RESIZE	DISTANCE	CARCINOMA	LYMPHOMA	NORMAL	AKURASI
768 × 576	Euclidean	8	8	28	40,%
	Minkowski				
	q = 3	4	5	36	41,67%
	q = 4	4	5	36	41,67%
	q = 5	4	11	36	47,22%
	q = 6	4	11	36	47,22%
560 × 560	Euclidean	25	16	14	50,93%
	Minkowski				
	q = 3	8	16	26	46,30%
	q = 4	8	16	29	49,07%
	q = 5	8	16	30	50%
	q = 6	12	16	28	51,85%
450 × 450	Euclidean	11	10	32	49,07%
	Minkowski				
	q = 3	6	9	36	47,22%
	q = 4	8	9	34	47,22%
	q = 5	13	9	18	37,04%
	q = 6	29	9	12	40,74%
360 × 360	Euclidean	27	7	4	35,19%
	Minkowski				
	q = 3	19	13	29	56,48%
	q = 4	20	13	24	52,78%
	q = 5	24	23	32	54,63%
	q = 6	23	13	21	52,78%
270 × 270	Euclidean	18	28	19	60,19%
	Minkowski				
	q = 3	22	28	24	68,52%
	q = 4	11	27	28	61,11%
	q = 5	11	27	29	62,04%
	q = 6	11	27	26	59,26%
200 × 200	Euclidean	25	13	11	45,37%
	Minkowski				
	q = 3	9	11	33	49,07%
	q = 4	13	13	33	54,63%
	q = 5	13	16	33	57,41%
	q = 6	13	19	32	59,41%
	q = 7	13	25	29	62,04%

Tabel 4.2 menunjukkan bahwa hasil tertinggi adalah ketika citra di *resize* dengan ukuran 270 × 270 piksel dan menggunakan distance *minkowski* pada klasifikasi K-NN. Nilai akurasi tertinggi yang diperoleh yaitu 68,52%. Hasil pengujian ini membuktikan bahwa *resize* pada *pre-processing* menentukan hasil ciri yang diperoleh. Pengaruh *resize* berkaitan dengan bagaimana ciri yang diambil dapat mencakup keseluruhan informasi yang terdapat dalam setiap piksel citra.

4.1.3 Analisis Hasil Pengujian Berdasarkan *Layer* Citra

Pengujian ketiga dilakukan dengan citra yang dikonversi ke dalam satu *layer*, yaitu RGB mejadi *grayscale*, RGB menjadi *red layer*, RGB menjadi *green layer*, dan RGB menjadi *blue layer*. Berdasarkan tabel 4.3, perubahan *layer* pada citra ini memperoleh hasil tertinggi ketika citra dikonversi menjadi citra *grayscale* menggunakan distance *minkowski* pada klasifikasi K-NN. Nilai akurasi yang diperoleh yaitu 68,52%.

Hasil tertinggi pada pengujian ini juga merupakan hasil tertinggi pada pengujian sebelumnya. Tidak ada kenaikan akurasi dari pengujian sebelumnya. Konversi citra menjadi *red layer*, *green layer*, dan *blue layer* memberikan pengaruh pada hasil akurasi, namun hasil akurasinya lebih rendah jika dibandingkan konversi citra menjadi citra *grayscale*.

Tabel 4.3 Data Benar Pengujian Berdasarkan *Layer* pada Citra

		CARCINOMA	LYMPHOMA	NORMAL	AKURASI
Grayscale	Euclidean	18	28	19	60,19%
	Minkowski				
	q = 3	22	28	24	68,52%
	q = 4	11	27	28	61,11%
	q = 5	11	27	29	62,04%
	q = 6	11	27	26	59,26%
q = 7	10	28	18	51,85%	
Red	Euclidean	13	25	22	55,56%
	Minkowski				
	q = 3	22	15	2	36,11%
	q = 4	12	23	3	35,19%
	q = 5	11	30	4	41,67%
	q = 6	12	29	4	41,67%
q = 7	12	28	4	40,74%	
Green	Euclidean	24	25	14	55,83%
	Minkowski				
	q = 3	15	23	32	64,81%
	q = 4	29	28	12	57,50%
	q = 5	25	28	12	54,17%
	q = 6	25	26	12	52,50%
q = 7	25	26	12	52,50%	
Blue	Euclidean	24	6	9	34,17%
	Minkowski				
	q = 3	21	4	3	25,92%
	q = 4	32	3	7	38,89%
	q = 5	31	3	7	37,96%
	q = 6	32	3	6	37,96%
q = 7	33	3	3	36,11%	

4.1.4 Analisis Hasil Pengujian Berdasarkan Jenis *Distance*

Pengujian keempat dilakukan dengan mengubah jenis *distance* pada metode klasifikasi K-NN untuk melihat jenis *distance* yang memberikan hasil akurasi terbaik. *Distance* yang diujikan adalah *cityblock* dan *cosine*.

Tabel 4.4 Data Benar pada Pengujian Berdasarkan Jenis *Distance*

	CARCINOMA	LYMPHOMA	NORMAL	AKURASI
Euclidean	18	28	19	60,19%
Minkowski				
q=3	22	28	24	68,52%
q=4	11	27	28	61,11%
q=5	11	27	29	62,04%
q=6	11	27	26	59,26%
q=7	10	28	18	51,85%
Cityblock	23	34	1	53,70%
Cosine	26	33	6	60,19%

Hasil pengujian berdasarkan Tabel 4.4 menunjukkan bahwa jenis *distance minkowski* dengan nilai $q = 3$ masih menjadi hasil tertinggi, yaitu dengan nilai akurasi 68,52%. Jenis *distance* lain memberikan akurasi kurang dari jenis *distance minkowski*. Berdasarkan hasil pengujian ini, jenis *distance* pada metode klasifikasi K-NN memberikan pengaruh untuk performansi sistem.

4.1.5 Analisis Hasil Pengujian Berdasarkan Nilai Variabel K

Pengujian keempat dilakukan dengan mengubah variabel K untuk melihat variabel K yang menghasilkan akurasi terbaik dengan pada proses klasifikasi K-NN. Nilai variabel K yang digunakan adalah 1, 3, 5, 7, dan 9. *Distance* yang digunakan dalam pengujian ini adalah jenis *minkowski* dengan nilai $q = 3$.

Berdasarkan Tabel 4.5, nilai akurasi tertinggi pada pengujian ini adalah saat nilai variabel $K = 1$ dan menggunakan jenis *distance minkowski*, yaitu 68,52%. Hasil akurasi ini sama dengan hasil tertinggi pada pengujian sebelumnya. Nilai variabel K memberikan pengaruh pada performansi sistem dalam tingkat akurasi. Ketika nilai K diubah, nilai akurasi berubah tetapi hasilnya cenderung menurun jika dibandingkan hasil tertinggi pada pengujian sebelumnya.

Tabel 4.5 Data Benar pada Pengujian Berdasarkan Nilai Variabel K

	CARCINOMA	LYMPHOMA	NORMAL	AKURASI
K=1	22	28	24	68,52%
K=3	25	26	13	59,26%
K=5	23	24	8	50,93%
K=7	22	24	15	56,48%
K=9	23	21	17	56,48%

5. Kesimpulan dan Saran

5.1 Kesimpulan

1. Sistem ini sudah mampu mendeteksi kanker usus besar dengan pembagian kelas yaitu normal, kanker *carcinoma*, dan kanker *lymphoma* dengan akurasi maksimal 68,52%.
2. Perbandingan komposisi data latih dan data uji yang menghasilkan akurasi terbaik adalah 30 : 36 setiap kelasnya.
3. Proses *resize* pada *preprocessing* yang menghasilkan nilai akurasi tertinggi adalah ketika citra berukuran 270×270 piksel.
4. Konversi citra RGB menjadi citra *grayscale* menghasilkan nilai akurasi tertinggi dibandingkan dengan konversi citra RGB menjadi citra satu *layer* (*Red*, *Green*, atau *Blue*).
5. Perubahan jenis *distance* pada klasifikasi metode K-NN mempengaruhi hasil akurasi pada sistem. Hasil akurasi tertinggi dihasilkan oleh jenis *distance minkowski*
6. Pengujian dengan mengubah nilai variabel K 1, 3, 5, 7, dan 9 dan menggunakan *distance minkowski* mendapatkan nilai akurasi paling tinggi yaitu 68,52% ketika nilai K=1 menggunakan *minkowski distance* dengan nilai q=3.

5.2 Saran

Saran yang diharapkan untuk perbaikan dan pengembangan penelitian ini yaitu :

1. Perancangan sistem dikembangkan dengan ekstraksi ciri dan klasifikasi yang bisa menghasilkan performansi lebih baik
2. Memperbanyak parameter ciri yang diambil dari data
3. Sumber data dari sumber data asli.

DAFTAR REFERENSI

- [1] A. R. Putra, *Klasifikasi Kanker Usus Besar Berbasis Pengolahan Citra Digital dengan Metode Radial Basis Function (RBF) 1*, no. x. Bandung: Institut Teknologi Telkom, 2012.
- [2] A. Kadir and A. Susanto, *Teori dan Aplikasi Pengolahan Citra*. Yogyakarta: ANDI, 2013.
- [3] E. Prasetyo, *Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI, 2012.
- [4] F. Geneser, *Buku Teks Histologi, Jilid 2*. Jakarta: Binarupa Aksara, 1994.
- [5] K. Chomboon, P. Chujai, P. Teerarassammee et al, *An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm*. Thailand: Suranaree University of Technology. 2015
- [6] M. H. Purnomo and A. Muntasa, *Konsep pengolahan citra digital dan ekstraksi fitur*. Yogyakarta: Graha Ilmu, 2010.
- [7] Peters M, *British Medical Association A-Z family medical encyclopedia*. London: Dorling Kindersley, 2008.
- [8] W. Sulisty, Y. R. Bech, and F. F. Y, *Analisis Penerapan Metode Median Filter Untuk Mengurangi Noise Pada Citra Digital*. Bali: Konferensi Nasional
- [9] Y. D. Kurniawan, *Klasifikasi Kanker Usus Besar Menggunakan Ekstraksi Ciri Grey Level Co-occurrence Matrix dengan Metode Levenberg-Marquardt Algorithm*. Bandung : Universitas Telkom. 2013
- [10] W. Desen and Zhizhong, *Kanker Usus Besar* [Online]. Available: <http://www.asiancancer.com/indonesian/cancer-topics/colon-cancer/>. [Accessed: 13 September 2016]