

OPTIMASI *K-NEAREST NEIGHBOR* UNTUK SISTEM KLASIFIKASI KONDISI KATARAK

Yunendah Nur Fuadah¹, Rita Magdalena²
Nor Kumalasari Caecar Pratiwi³, Steven Palondongan⁴

^{1, 2, 3} Fakultas Teknik Elektro, Universitas Telkom, Bandung, Indonesia

⁴ Astra Graphia Information Technology (AGIT), Bandung, Indonesia

¹yunendah@telkomuniversity.ac.id,²ritamagdalen@telkomuniversity.ac.id,

³caecarnkcp@telkomuniversity.ac.id,⁴stevenpalondongan277@gmail.com

Abstrak

Katarak adalah salah satu faktor penyebab kebutaan yang kasusnya banyak ditemukan di Indonesia. Berdasarkan data dari Kementerian Kesehatan Republik Indonesia, angka kejadian kebutaan akibat katarak menduduki peringkat yang lebih tinggi dari glaukoma. Untuk menekan peningkatan jumlah kebutaan akibat katarak, sebuah sistem deteksi dini kondisi katarak dapat menjadi sebuah solusi. Pada penelitian ini, dilakukan optimasi metode *Gray Level Co-Occurrence Matrix* (GLCM) untuk mengekstrak informasi dari masukan berupa citra mata. Pada proses ekstraksinya, parameter yang diukur adalah jarak antar piksel (d), sudut (θ), dan fitur statistik. Informasi tersebut kemudian diklasifikasikan dengan menggunakan *K-Nearest Neighbor* (K-NN) kedalam tiga kondisi, yaitu kondisi normal, katarak imatur, dan katarak matur. Pengujian yang dilakukan adalah menentukan nilai k dan persamaan perhitungan jarak yaitu *Euclidean*, *Minkowski*, *Chebychev* dan *City Block*. Berdasarkan hasil pengujian diperoleh akurasi pengujian sebesar 93,33 % dengan menggunakan parameter yang paling optimal yaitu $d = 2$, $\theta = 135^\circ$, nilai $k = 5$ dengan persamaan jarak *Euclidean* dan *Minkowski*.

Kata Kunci: Katarak, GLCM, Fitur Statistik, k-NN

Abstract

Cataract is one of the dominant factors that may cause blindness in Indonesia. Based on the data from the Ministry of Health of Republic Indonesia, the number of blindness incident caused by cataract were higher than glaucoma. To reduce the increasing number of incident, an early detection system may give a better solution. In this research, we optimize the use of the *Gray Level Co-occurrence Matrix* (GLCM) to extract the information from an eye image. The extraction process, the calculated parameters were the distance between pixels (d), angle (θ), and statistical features. Those features then classified using *K-Nearest Neighbor* (K-NN) into three conditions, normal, immature cataract, and mature cataract. The parameters tested at this step are k values and distance calculation equations used, namely *Euclidean*, *Minkowski*, *Chebychev* and *City Block*. Based on the test results obtained testing accuracy of 93,33% by using the most optimal parameters, namely $d = 2$, $\theta = 135^\circ$, $k = 5$ with *Euclidean* and *Minkowski* distance equation.

Key Words: Cataract, GLCM, Statistical Feature, k-NN

1. Pendahuluan

Katarak merupakan kondisi kekeruhan pada lensa mata yang disebabkan oleh pengapuran pada lensa mata sehingga mengakibatkan daya penglihatan mata berkurang. Penyakit katarak dapat menyebabkan terjadinya gangguan rabun jauh (miopi) dan dapat menyebabkan kebutaan secara permanen apabila terlambat dalam mendapatkan penanganan medis. Berdasarkan hasil data survei dari Kementerian Kesehatan Republik Indonesia, penyebab utama kebutaan di Indonesia adalah penyakit katarak (0,78%),

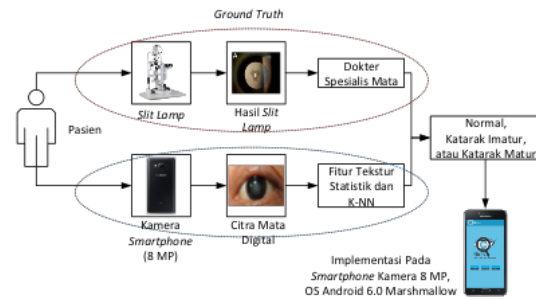
disusul penyakit glaukoma (0,40%) dan penyakit lainnya [1]. Selanjutnya, berdasarkan data terbaru dari riset dasar kesehatan pada tahun 2013, prevalensi katarak di Indonesia untuk semua umur adalah 1,8%. Menurut ketentuan WHO kebutaan yang cukup tinggi atau lebih dari 1% seperti di Indonesia bukan lagi menjadi masalah medis saja, melainkan menjadi masalah sosial yang perlu mendapatkan perhatian bersama untuk mendapatkan solusi yang paling efektif [2].

Dampak negatif dari kebutaan yang diakibatkan oleh katarak diantaranya akan mengganggu produktivitas

dan mobilitas penderitanya, hal ini akan berdampak terhadap sosio-ekonomi bagi masyarakat, yang mengakibatkan menurunnya kualitas hidup masyarakat. Salah satu solusi untuk menanggulangi kebutaan disebabkan oleh katarak adalah dengan melakukan pemeriksaan kondisi mata sedini mungkin apabila terasa mengalami gangguan. Pemeriksaan mata untuk deteksi katarak dilakukan oleh dokter spesialis mata dengan menggunakan *slit lamp*. Terbatasnya jumlah tenaga dokter spesialis mata dan *slit lamp* di beberapa rumah sakit kecil dan puskesmas-puskesmas di Indonesia menjadi hambatan yang menyebabkan terlambatnya penanganan medis untuk mendeteksi katarak. Oleh karena itu, dikembangkan sistem deteksi katarak berbasis pengolahan citra digital sebagai alat bantu dalam mendeteksi katarak.

Terdapat beberapa penelitian terkait pengembangan aplikasi deteksi katarak berbasis pengolahan citra digital. Fokus penelitian sistem deteksi katarak fokus kepada pencarian metode ekstraksi ciri yang paling optimal, sedangkan untuk proses klasifikasi yang paling sering digunakan adalah *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (K-NN). SVM sering digunakan karena termasuk kedalam pengklasifikasi diskriminatif yang mampu memisahkan kumpulan data kedalam kelas tertentu dengan memanfaatkan *hyperplane* [3, 4]. Adapun beberapa penelitian yang menggunakan SVM sebagai pengklasifikasi adalah sebagai berikut. Sistem klasifikasi katarak serius dan non serius menggunakan metode analisis tekstur statistik dan SVM dengan akurasi 86,6% [5], sistem deteksi katarak untuk kondisi normal, katarak dan *post cataract* menggunakan metoda *Edge pixel Count* dan SVM dengan akurasi 88,39% [6], sistem deteksi katarak untuk kondisi katarak insipien dan katarak lanjut menggunakan metode *Principal Component Analysis* (PCA) dan SVM dengan akurasi 89% [7]. Adapun sistem deteksi katarak yang berbasis K-NN diantaranya adalah sistem klasifikasi kondisi mata normal dan katarak menggunakan metode analisis tekstur statistik dan K-NN dengan akurasi 94,5% [8].

Pada beberapa penelitian terkait tentang klasifikasi kondisi katarak tersebut terdapat dua metode klasifikasi yang telah digunakan, yaitu *Support Vector Machine* dan *K-Nearest Neighbor*, kedua metode klasifikasi tersebut dapat mengklasifikasikan kondisi katarak untuk dua kondisi. Berdasarkan hasil akurasi yang diperoleh dengan menggunakan metode *K-Nearest Neighbor* diperoleh nilai akurasi yang lebih tinggi dibandingkan metode *Support Vector Machine*. Terdapat beberapa kelebihan dari metode K-NN diantaranya K-NN merupakan metode klasifikasi yang sederhana, tidak terpengaruh *noise* pada tahap klasifikasi data, dan efektif untuk mengklasifikasikan data dalam jumlah besar [9]. Pada penelitian ini sistem deteksi katarak yang dikembangkan dapat mengklasifikasikan kondisi mata



Gambar 1. Pemeriksaan Pasien dengan *Slit Lamp* (atas) dan Kamera *Smartphone* (bawah).

kedalam tiga kondisi yaitu normal, katarak imatur dan katarak matur dimana pada penelitian sebelumnya hanya dua kondisi saja yaitu normal dan katarak, metode yang digunakan pada penelitian ini adalah *Gray Level Co-Occurrence Matrix* (GLCM) pada tahap ekstraksi ciri dan *K-Nearest Neighbor* (K-NN) pada tahap klasifikasi.

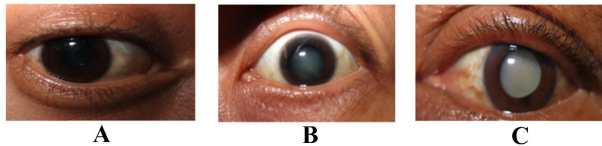
Pada penelitian ini dilakukan optimasi metode GLCM dan K-NN dalam klasifikasi kondisi katarak. Pada metode GLCM, sebagai ekstraksi ciri, parameter yang memengaruhi yaitu jarak antar piksel (d), sudut analisis (θ), dan fitur statistik. Nilai parameter GLCM yang akan digunakan pada penelitian ini berdasarkan hasil optimasi GLCM pada penelitian sebelumnya [10]. Jarak antar piksel (d) yang digunakan untuk pengujian adalah 1 sampai dengan 2, sudut analisis (θ) 0° , 45° , 90° , 135° , dan empat fitur statistik. Sedangkan untuk metode klasifikasi K-NN dilakukan analisis terkait pengaruh nilai k dan persamaan perhitungan jarak yang digunakan (*Euclidean*, *Minkowski*, *Chebyshev* dan *City Block*) terhadap akurasi sistem yang dihasilkan.

2. Metodologi

2.1 Gambaran Umum Penelitian

Berdasarkan Gambar 1, hasil pemeriksaan pasien menggunakan *slit lamp* menjadi acuan dalam penelitian ini untuk memperoleh data berupa citra mata dengan kondisi mata pasien yang sudah mendapatkan pemeriksaan secara medis. Penelitian ini, mengembangkan sistem deteksi katarak menggunakan kamera digital berbasis pengolahan citra untuk menentukan kondisi mata pasien apakah menderita katarak imatur, katarak matur atau normal. Hasil deteksi kondisi mata dari sistem yang dikembangkan akan dibandingkan dengan hasil deteksi kondisi mata berdasarkan pemeriksaan secara medis.

Gambar 2 merupakan citra mata digital untuk kondisi normal, katarak imatur dan katarak matur yang digunakan dalam penelitian ini. Katarak imatur merupakan suatu kondisi mata dimana kekeruhan terjadi pada sebagian lensa mata, kondisi katarak imatur ini dapat mengakibatkan rabun jauh (*miopi*). Katarak matur



Gambar 2. Citra Mata Digital (A). Normal; (B). Katarak Imatur ; (C). Katarak Matur.

merupakan katarak lanjutan dari katarak imatur, dimana pada kondisi ini kekeruhan terjadi pada seluruh lensa mata, apabila kondisi tersebut dibiarkan tanpa penanganan medis, maka pasien beresiko tinggi mengalami kebutaan. Oleh karena itu, pada penelitian ini dikembangkan sistem deteksi dini dan klasifikasi kondisi katarak pada citra mata. Data citra mata digital tersebut selanjutnya diolah oleh sistem klasifikasi katarak berbasis pengolahan citra digital menggunakan metode GLCM pada tahap ekstraksi ciri dan metode K-NN pada tahap klasifikasi. Setelah mendapatkan parameter-parameter terbaik dari metode GLCM dan K-NN, selanjutnya dilakukan implementasi pada *smartphone android*.

2.2 Diagram Alir Sistem

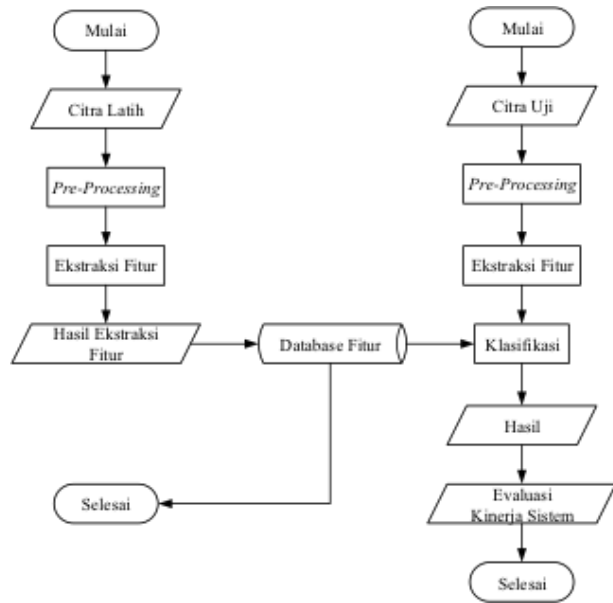
Pada penelitian ini dilakukan analisis performansi metode *K-Nearest Neighbor* (K-NN) untuk klasifikasi kondisi katarak. Pada Gambar 3 terdapat dua blok, tahap pelatihan sistem pada diagram blok bagian kiri dan tahap pengujian sistem pada bagian kanan. Pada tahap pelatihan data latih akan mengalami *preprocessing* kemudian ekstraksi fitur. Hasil dari tahap ekstraksi fitur data latih akan digunakan pada saat tahap klasifikasi data uji. Pada tahap pengujian, data uji akan mengalami tahapan yang sama seperti data latih yaitu *preprocessing* dan ekstraksi fitur. Hasil dari ekstraksi fitur data uji akan diklasifikasikan kedalam kondisi normal, katarak imatur dan katarak matur berdasarkan data latih.

2.2.1 Citra Mata Digital

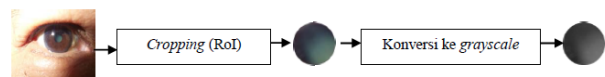
Data yang digunakan dalam penelitian ini merupakan citra mata digital yang terdiri dari citra mata untuk kondisi normal, katarak imatur dan katarak matur. Pengambilan data dilakukan pada pasien penderita katarak di RS Cicendo Bandung. Jumlah data total yang digunakan pada penelitian ini adalah 90 data citra mata digital, dengan rincian, 45 citra mata latih dan 45 citra mata uji, yang masing- masing terdiri dari 15 citra mata untuk setiap kondisi normal, katarak imatur dan katarak matur.

2.2.2 Preprocessing

Pada tahap *preprocessing* dilakukan proses *cropping* untuk mendapatkan daerah pupil yang



Gambar 3. Diagram Alir Sistem.



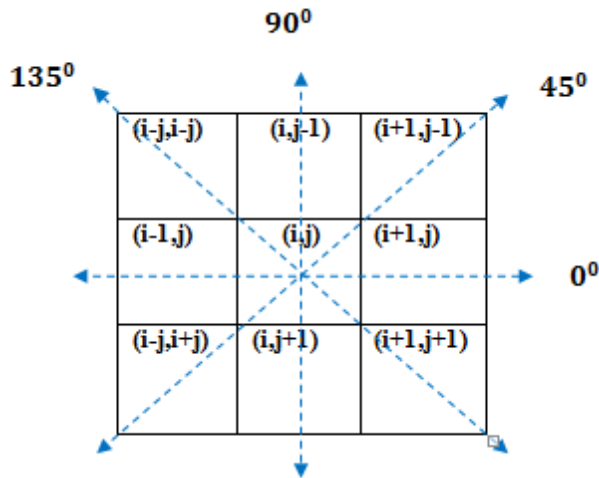
Gambar 4. Diagram Blok Preprocessing.

merupakan *Region of Interest* (RoI) pada penelitian ini. Setelah diperoleh citra pupil mata dilakukan konversi dari citra *Red Green Blue* (RGB) menjadi citra *grayscale* untuk selanjutnya memasuki tahap ekstraksi ciri menggunakan metode *Gray Level Co-occurrence Matrix* (GLCM)

2.2.3 Ekstraksi Fitur Gray Level Co-Occurrence Matrix

Ekstraksi fitur merupakan metode untuk mendapatkan nilai fitur dari sebuah citra. Terdapat tiga metode untuk memperoleh fitur tekstur dari suatu citra antara lain metode statistik, metode struktural, dan metode spektral. Pada penelitian ini dilakukan ekstraksi fitur dengan metode statistik. Analisis tekstur dengan metode statistik terdiri dari metoda statistik orde pertama dan metoda statistik orde kedua. Ekstraksi fitur orde pertama merupakan pengambilan fitur citra berdasarkan karakteristik histogram citra. Kelemahan dari statistik orde pertama ini kurang signifikan dalam mengenali perbedaan antar citra. Oleh karena itu, diperlukan pengambilan fitur statistik orde kedua dengan menggunakan metoda *Gray Level Cooccurrence Matrix* (GLCM).

Metode GLCM termasuk dalam metode statistik dengan menghitung probabilitas ketetanggaan antara dua piksel pada jarak dan orientasi sudut tertentu [11].



Gambar 5. Arah GLCM dengan Sudut 0°, 45°, 90°, dan 135°.

GLCM menghitung seberapa sering pasangan piksel dengan nilai tertentu dan relasi spasial muncul dalam citra. Sudut orientasi dari relasi spasial yaitu 0°, 45°, 90°, dan 135°.

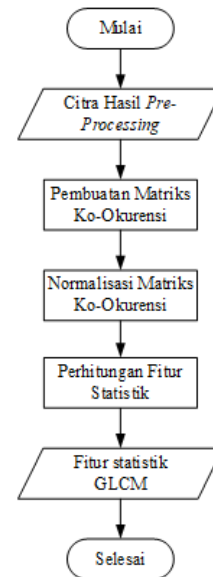
GLCM merupakan metode analisis tekstur orde dua yang berisi informasi tentang frekuensi kejadian dari kombinasi dua piksel tetangga dalam sebuah citra. Informasi tersebut direpresentasikan dalam suatu matriks ko-okurensi, selanjutnya matriks tersebut dinormalisasi. Matriks normalisasi GLCM dapat menunjukkan sifat suatu tekstur, akan tetapi tidak dapat digunakan sebagai analisis untuk membandingkan beberapa tekstur citra, sehingga setelah diperoleh matriks ko-okurensi yang telah dinormalisasi, dilakukan perhitungan fitur statistik yang diusulkan oleh Haralick *Haralick* [12]. Fitur-fitur statistik tersebut dapat menjadi nilai ciri yang dapat dijadikan analisis untuk membedakan kondisi citra mata normal, katarak imatur dan katarak matur. Gambar 6 menunjukkan proses ekstraksi ciri menggunakan metode GLCM.

Pada penelitian ini dilakukan optimasi GLCM untuk mendapatkan parameter yang optimal, sehingga pada penelitian ini parameter GLCM yang digunakan berdasarkan hasil penelitian sebelumnya yang memberikan performansi terbaik yaitu nilai jarak (d) = 1 dan 2 dan sudut orientasi (θ) sebesar 0°, 45°, 90°, 135°, sedangkan untuk fitur statistik yang digunakan pada penelitian ini adalah:

1. *Sum Entropy*

$$Sum_{ent} = - \sum_{k=2}^{2N_g} \sum_{i,j} p_d(i,j) \log(p_d(i,j)) \quad (1)$$

$i + j = k$



Gambar 6. Diagram Alir Ekstraksi Ciri.

2. *Variance*

$$Var = \sum_i \sum_j (i - \mu)^2 p(i,j) \quad (2)$$

3. *Sum Variance*

$$Sum_{var} = \sum_{k=2}^{2N_g} \sum_{i,j} (k - Sum_{avg})^2 p_d(i,j) \quad (3)$$

$i + j = k$

$$Sum_{avg} = \sum_{k=2}^{2N_g} k \sum_{i,j} p_d(i,j) \quad (4)$$

$i + j = k$

4. *Maximal Correlation Coefficient*

$$Max_{cor-coef} = \sqrt{2^{th} \text{largest eigenvalue of } Q(i,j)} \quad (5)$$

$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(j)}, \quad (6)$$

dengan N_g adalah jumlah level keabuan yang berbeda dalam citra, $p_d(i,j)$ adalah entri dalam matriks ko-okurensi pada baris i dan kolom j , $p_x(i)$ adalah hasil penjumlahan baris dari entri matriks $p_d(i,j)$, $p_y(j)$ merupakan hasil penjumlahan kolom dari entri matriks $p_d(i,j)$, dan μ adalah nilai rata-rata.

2.2.4 *Klasifikasi K-Nearest Neighbor*

Metode *K-Nearest Neighbour* (K-NN) merupakan metode klasifikasi terhadap sekumpulan objek,

berdasarkan data pembelajaran yang sudah terklasifikasikan sebelumnya yang memiliki jarak terdekat dengan objek tersebut [13].

K-NN memiliki prinsip kerja dengan mencari jarak terdekat antara data latih dengan sejumlah k tetangga terdekatnya dalam data pelatihan. Data pelatihan diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi tersebut merepresentasikan fitur dari hasil ekstraksi fitur data latih. Pada fase pelatihan, algoritma K-NN ini melakukan penyimpanan vektor-vektor fitur hasil dari ekstraksi ciri data latih. Pada fase klasifikasi, fitur-fitur hasil ekstraksi ciri data uji akan dibandingkan dengan hasil ekstraksi ciri data latih. Jarak dari vektor ciri uji akan dihitung terhadap seluruh vektor data latih. Pengukuran jarak pada metode K-NN dilakukan dengan beberapa metode, diantaranya [14]:

1. Jarak *Euclidean*

$$j(v_1, v_2) = \sqrt{\sum_{k=1}^N (v_1(k) - v_2(k))^2} \quad (7)$$

2. Jarak *City-Block*

$$j(v_1, v_2) = \sum_{k=1}^N |v_1(k) - v_2(k)| \quad (8)$$

3. Jarak *Chebyshev*

$$j(v_1, v_2) = \max_{k=1 \rightarrow N} (|v_1(k) - v_2(k)|) \quad (9)$$

4. Jarak *Minkowski*

$$j(v_1, v_2) = \sqrt[p]{\sum_{k=1}^N |v_1(k) - v_2(k)|^p} \quad (10)$$

dengan j adalah jarak data uji ke data latih, $v_1(k)$ adalah data uji ke- k , dengan $k=1,2,N$, dan $v_2(k)$ adalah data latih ke- k , dengan $k=1,2,,N$. Tahapan algoritma K-NN yang digunakan adalah sebagai berikut:

1. Menentukan parameter k .
2. Menghitung jarak antara objek dengan data latih menggunakan persamaan jarak *Euclidean*, *Minkowski*, *Chebyshev* dan *City Block*.
3. Mengurutkan hasil pada butir 2 secara *ascending*.
4. Mengumpulkan kategori hasil klasifikasi *nearest neighbor*
5. Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat dipredisikan kategori objek.
6. Menampilkan jumlah kecenderungan hasil klasifikasi K-NN untuk setiap data berdasarkan jumlah k yang ditentukan.

Tabel 1. Model *Confusion Matrix* 3 Kelas.

	Kelas Prediksi			
	A	B	C	
Kelas Aktual	A	TP_A	E_{AB}	E_{AC}
	B	E_{BA}	TP_B	E_{BC}
	C	E_{CA}	E_{CB}	TP_C

Sebagai metoda klasifikasi K-NN memiliki beberapa kelebihan, diantaranya merupakan metoda sederhana yang dapat mengklasifikasikan data, tidak terpengaruh *noise* pada tahap klasifikasi, dan efektif untuk mengklasifikasikan data dalam jumlah besar. Ketepatan klasifikasi K-NN ini dipengaruhi oleh fitur yang digunakan serta nilai k yang digunakan [9].

2.3 Performansi Sistem

Parameter yang digunakan dalam evaluasi performansi sistem pada penelitian ini adalah akurasi sistem dan waktu komputasi.

Hasil klasifikasi kondisi mata yaitu kondisi normal, katarak imatur dan katarak matur disajikan dalam Tabel *Confusion Matrix* 3×3 . Beberapa parameter dalam *Confusion Matrix* yang akan digunakan dalam analisis keberhasilan klasifikasi sistem antara lain:

1. *True Positive*

True Positive merupakan jumlah data kelas aktual yang terklasifikasi tepat dengan kelas prediksinya.

$$\text{True Positive} = TP_A + TP_B + TP_C \quad (11)$$

2. *False Negative*

False Negative merupakan jumlah total nilai pada baris suatu kelas, selain nilai *True Positive* kelas tersebut.

$$\text{False Negative Kelas A} = E_{AB} + E_{AC} \quad (12)$$

$$\text{False Negative Kelas B} = E_{BA} + E_{BC} \quad (13)$$

$$\text{False Negative Kelas C} = E_{CA} + E_{CB} \quad (14)$$

3. *False Positive*

Merupakan jumlah total nilai pada kolom suatu kelas, selain nilai *True Positive* kelas tersebut.

$$\text{False Positive Kelas A} = E_{BA} + E_{CA} \quad (15)$$

$$\text{False Positive Kelas B} = E_{AB} + E_{CB} \quad (16)$$

$$\text{False Positive Kelas C} = E_{AC} + E_{BC} \quad (17)$$

4. *True Negative*

Merupakan jumlah total kolom dan baris, selain kolom dan baris kelas tersebut.

$$\text{True Negative Kelas A} = TP_B + E_{BC} + E_{CB} + TP_C \quad (18)$$

$$\text{True Negative Kelas B} = T_{PA} + E_{AC} + E_{CA} + T_{PC} \tag{19}$$

$$\text{True Negative Kelas C} = T_{PA} + E_{AB} + E_{BA} + T_{PB} \tag{20}$$

Akurasi sistem merupakan perbandingan antara data yang memberikan hasil klasifikasi kondisi mata yang benar terhadap jumlah data uji. Secara matematis akurasi sistem ditunjukkan pada persamaan berikut:

$$\text{Akurasi} = \frac{\sum \text{Data Benar}}{\sum \text{Data Keseluruhan}} \times 100\% \tag{21}$$

dengan $\sum \text{Data Benar} = T_{PA} + T_{PB} + T_{PC}$, $\sum \text{Data Keseluruhan} = T_{PA} + T_{PB} + T_{PC} + E_{AB} + E_{AC} + E_{BA} + E_{BC} + E_{CA} + E_{CB}$. Sedangkan persamaan untuk mendapatkan *error rate* sistem diberikan pada persamaan:

$$\text{Error} = (1 - \text{Akurasi}) \times 100\%. \tag{22}$$

Waktu komputasi adalah waktu yang dibutuhkan oleh sistem untuk melakukan suatu proses. Waktu komputasi diperoleh dengan mencari selisih dari waktu proses selesai terhadap waktu memulai proses. Secara matematis waktu komputasi dapat dituliskan sebagai berikut:

$$\text{Waktu Komputasi} = \text{Waktu Selesai} - \text{Waktu Mulai} \tag{23}$$

3. Hasil

Pada bagian ini dibahas hasil pengujian dan analisis sistem yang bertujuan untuk mengetahui performansi sistem yang telah dirancang. Adapun hal-hal yang menjadi bahasan, yaitu: hasil ekstraksi fitur untuk kombinasi parameter GLCM, yang terdiri dari jarak antar piksel $d = 1$ dan $d = 2$, sudut $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ dan empat fitur tekstur statistik, yaitu, *Sum Entropy*, *Variance*, *Sum Variance*, dan *Maximal Correlation Coefficient*. Hasil pengujian menggunakan metode K-NN dengan nilai $k = 1, 3, 5, 7, 9, 11$ serta perhitungan jarak menggunakan persamaan *Cityblock*, *Chebychev*, *Euclidean*, dan *Minkowski*. Proses pengujian tersebut untuk mengetahui performansi sistem yang meliputi akurasi sistem dan waktu komputasi.

3.1 Hasil Gray Level Co-Occurrence Matrix

Tabel 2, Tabel 3 dan Tabel 4 merupakan salah satu contoh matriks ko-okurensi skala keabuan ternormalisasi (GLCM ternormalisasi) menggunakan $d = 1$ dan $\theta = 0^\circ$ untuk kondisi mata normal, katarak imatur, dan katarak matur.

Berdasarkan ketiga matriks Ko-okurensi tersebut, dapat dilihat penyebaran nilai-nilai piksel matriks ko-okurensi ternormalisasi untuk kondisi normal, katarak imatur dan katarak matur terhadap diagonal utama. Pada Tabel 2 untuk kondisi mata normal, persebaran matriks

Tabel 2. Matriks Ko-Okurensi Ternormalisasi Untuk Kondisi Mata Normal.

0.1954	0.0424	0	0	0	0	0	0
0.0488	0.7135	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Tabel 3. Matriks Ko-Okurensi Ternormalisasi Untuk Kondisi Katarak Imatur.

0.3481	0.0549	0.0027	0.0002	0	0	0	0
0.0473	0.4195	0.0299	0.0004	0.0002	0	0	0
0.0069	0.0195	0.0466	0.0084	0.0004	0	0	0
0.0033	0.0015	0.0013	0.0024	0.0022	0.0009	0.0002	0
0.0004	0.0002	0.0013	0	0.0002	0.0004	0.0004	0
0	0.0002	0.0004	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

ko-okurensi berpusat pada beberapa koordinat disekitar diagonal utama. Sementara pada Tabel 3 untuk kondisi katarak imatur dan pada Tabel 4 untuk kondisi matur, matriks ko-okurensi ternormalisasi tersebar terhadap diagonal utama.

Matriks ko-okurensi ternormalisasi untuk kondisi katarak matur memiliki nilai piksel 0 paling sedikit, sementara kondisi mata normal memiliki nilai piksel 0 paling banyak, hal tersebut menunjukkan bahwa berdasarkan matriks ko-okurensi ternormalisasi yang dihasilkan diketahui citra kondisi normal memiliki tingkat keseragaman yang paling tinggi, sedangkan untuk citra katarak imatur dan citra katarak matur memiliki keseragaman yang rendah. Hal ini sesuai dengan kondisi mata normal yang tidak mengalami kekeruhan sehingga citra mata kondisi normal memiliki tingkat keseragaman tinggi, sedangkan untuk kondisi mata imatur dan matur mengalami kekeruhan yang mengakibatkan citra mata kondisi imatur dan matur memiliki tingkat keseragaman rendah.

3.2 Hasil Fitur Statistik

Gambar 7 menunjukkan hasil fitur statistik rata-rata untuk fitur *Sum Entropy*, *Variance*, *Sum Variance*, dan *Maximal Correlation Coefficient* pada citra mata normal, katarak imatur dan katarak matur. Berdasarkan Gambar 7 dapat dilihat bahwa empat fitur statistik tersebut dapat membedakan kondisi mata normal, katarak imatur dan katarak matur. Fitur *Sum Entropy* pada kondisi katarak imatur memiliki nilai yang paling tinggi, katarak imatur dipertengahan dan kondisi normal memiliki nilai *Sum Entropy* yang paling rendah. Hal ini dikarenakan fitur

Tabel 4. Matriks Ko-Okurensi Ternormalisasi Untuk Kondisi Katarak Matur.

0.1497	0.0035	0.0022	0.0026	0.0111	0.0058	0	0
0.0031	0.0004	0.0004	0.0005	0.0026	0.0013	0	0
0.0035	0.0009	0.0024	0.0007	0.0018	0.0018	0	0
0.0095	0.0018	0.0057	0.0374	0.0033	0.0013	0	0
0.0044	0.0015	0.0013	0.0175	0.1163	0.0210	0	0
0.0031	0.0002	0.0005	0.0013	0.0243	0.4978	0.0086	0
0.0011	0.0004	0.0002	0.0002	0	0.0040	0.0186	0.0058
0.0009	0.0002	0.0002	0.0002	0.0002	0.0011	0.0031	0.0128

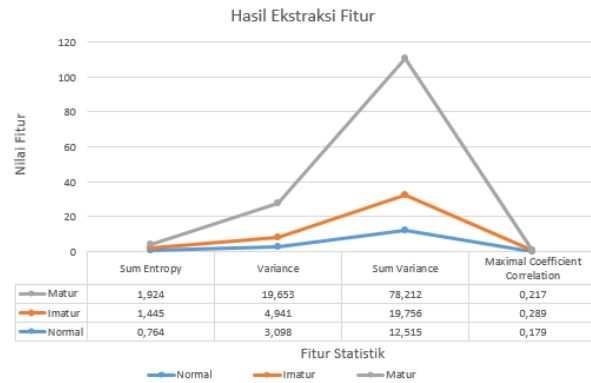
Sum Entropy menunjukkan kompleksitas suatu citra. Teksur citra yang kompleks seperti kondisi citra mata katarak yang dipengaruhi kekeruhan pada lensa mata memiliki nilai *Sum Entropy* tinggi, sedangkan untuk citra kondisi mata normal cenderung lebih seragam sehingga memiliki nilai *Sum Entropy* yang lebih rendah.

Nilai fitur *Variance* dan *Sum Variance* pada kondisi katarak matur lebih tinggi dibandingkan dengan kondisi imatur dan normal, fitur *Variance* dan *Sum Variance* menunjukkan tingkat keabuan dari suatu citra semakin bervariasi tingkat keabuan suatu citra, semakin tinggi nilai *Variance* dan *Sum Variance*. Sedangkan, untuk fitur *Maximal Correlation Coefficient*, citra katarak pada kondisi imatur memiliki nilai yang lebih tinggi dibanding kondisi lainnya. Berdasarkan hasil yang diperoleh, empat fitur statistik tersebut cukup signifikan dalam membedakan setiap kondisi citra mata, sehingga dapat dijadikan fitur latih yang akan digunakan pada tahap klasifikasi. Selain cukup signifikan dalam membedakan kondisi mata normal, katarak imatur dan katarak matur secara umum, fitur statistik yang digunakan dapat lebih detail membedakan perubahan kondisi mata dari mulai kondisi normal memasuki kondisi imatur tahap awal, kondisi matur, kondisi imatur mendekati matur dan kondisi matur.

Gambar 8 menunjukkan perbandingan hasil fitur statistik untuk setiap kondisi mata, nilai fitur *Variance*, *Sum Variance*, *Sum Entropy* semakin tinggi sesuai dengan tingkat kekeruhan pada lensa, hal ini dikarenakan semakin tinggi kompleksitas suatu citra dan semakin beragam tingkat keabuan suatu citra ketiga nilai fitur tersebut akan semakin tinggi. Sedangkan untuk fitur *Maximal Correlation Coefficient* secara umum lebih tinggi pada kondisi imatur ketika lensa mata mulai terjadi kekeruhan akan tetapi masih terdapat bagian pupil yang hitam, sedangkan untuk kondisi mata normal dan kondisi mata matur nilai fitur *Maximal Correlation Coefficient* lebih rendah dibandingkan dengan kondisi imatur.

3.3 Hasil Pengujian Sistem

Pada tahap klasifikasi sistem menggunakan K-NN, performansi pengujian sistem dilakukan dengan menganalisis pengaruh metode perhitungan jarak yang digunakan, yaitu *City Block*, *Chebychev*, *Euclidean* dan



Gambar 7. Hasil Fitur Statistik.

Fitur Statistik	Kondisi mata normal	Kondisi imatur tahap awal	Kondisi imatur	Kondisi imatur mendekati matur	Kondisi matur
Variance	1,042725	2,336016	4,225722	8,109579	19,91734
Sum Variance	4,310811	9,248356	17,0053	32,38605	79,26479
Sum Entropy	0,181319	0,97549	1,234837	1,770998	2,10923
Maximal Correlation Coefficient	0,070667	0,360052	0,20496	0,236438	0,215427

Gambar 8. Perbandingan Hasil Fitur Statistik untuk Setiap Kondisi Mata.

Minkowski, serta nilai k untuk nilai $k = 1, 3, 5, 7, 9$ dan 11 terhadap performansi akurasi sistem. Hasil yang diperoleh pada pengujian empat fitur tekstur statistik hasil ekstraksi matriks ko-okurensi pada berbagai d , θ dan k dapat dilihat pada Tabel 5, Tabel 6, dan Tabel 7.

Tabel 5 menunjukkan akurasi terbaik yang dihasilkan pada pengujian empat fitur statistik menggunakan jarak *Chebyshev* yaitu sebesar $91,11\%$. Hasil tersebut diperoleh pada penggunaan $k = 9$ dimana empat fitur tekstur statistik tersebut merupakan hasil ekstraksi matriks ko-okurensi dengan nilai $d = 1$ dan $\theta = 135^\circ$. Pada pengujian menggunakan jarak *Chebyshev*, semakin besar nilai k hasil akurasi sistem cenderung lebih baik dan mencapai optimal pada saat nilai $k = 9$.

Tabel 6 menunjukkan akurasi terbaik yang diperoleh pada pengujian empat fitur tekstur statistik menggunakan jarak *City block* yaitu sebesar $91,11\%$. Hasil tersebut diperoleh pertama kali pada penggunaan $k = 7$ terhadap matriks ko-okurensi dengan nilai $d = 1$ dan $\theta = 45^\circ$. Pada pengujian menggunakan jarak *City block* semakin besar nilai k hasil akurasi sistem cenderung lebih baik dan mencapai optimal pada saat nilai $k = 9$.

Tabel 7 menunjukkan hasil pengujian empat fitur statistik menggunakan jarak *Minkowski* dan *Euclidean* yang memberikan akurasi terbaik yang sama yaitu sebesar $93,33\%$, untuk mengetahui performansi sistem

Tabel 5. Hasil Pengujian Menggunakan Jarak Chebyshev.

Nilai k	d=1 θ=0°	d=2 θ=45°	d=1 θ=90°	d=1 θ=135°	d=2 θ=0°	d=2 θ=45°	d=2 θ=90°	d=2 θ=135°
1	77.78 %	80.00 %	77.78 %	73.33 %	75.56 %	84.44 %	75.56 %	77.78 %
3	66.67 %	77.78 %	68.89 %	75.56 %	73.33 %	80.00 %	73.33 %	77.78 %
5	84.44 %	84.44 %	84.44 %	84.44 %	86.67 %	86.67 %	84.44 %	88.89 %
7	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %
9	88.89 %	88.89 %	86.67 %	91.11 %	88.89 %	88.89 %	88.89 %	88.89 %
11	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %

Tabel 6. Hasil Pengujian Menggunakan Jarak City Block.

Nilai k	d=1 θ=0°	d=2 θ=45°	d=1 θ=90°	d=1 θ=135°	d=2 θ=0°	d=2 θ=45°	d=2 θ=90°	d=2 θ=135°
1	75.56 %	71.11 %	71.11 %	71.11 %	73.33 %	77.78 %	73.33 %	68.89 %
3	66.67 %	66.67 %	66.67 %	66.89 %	66.89 %	71.11 %	66.67 %	71.11 %
5	86.67 %	84.44 %	86.67 %	86.67 %	86.67 %	84.44 %	84.44 %	86.67 %
7	86.67 %	91.11 %	91.11 %	91.11 %	84.44 %	91.11 %	84.44 %	91.11 %
9	88.89 %	91.11 %	88.89 %	91.11 %	91.11 %	91.11 %	91.11 %	91.11 %
11	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	91.11 %	88.89 %	91.11 %

secara lebih terperinci dapat dilihat pada *confusion matrix* pada Tabel 8. *Confusion matrix* pada Tabel 8 tersebut merupakan hasil pengujian terhadap data uji dengan jumlah 15 data kelas normal, 15 data kelas imatur dan 15 data kelas imatur.

Berdasarkan Tabel 8 dapat diperoleh parameter - parameter performansi sistem sebagai berikut:

1. *True Positive* (TP)

True Positive menunjukkan jumlah data kondisi normal, imatur, dan matur di kelas aktual terdeteksi tepat sesuai dengan kondisi di kelas prediksinya. Berdasarkan Tabel 8 dapat diperoleh hasil TP sebagai berikut.

$$TP = 14 + 15 + 13 = 42$$

Dari hasil tersebut dapat diketahui bahwa 42 data dari 45 data uji terdeteksi tepat sesuai dengan kondisi kelasnya.

2. *False Negative* (FN)

False Negative menunjukkan jumlah kondisi kelas aktual yang terdeteksi salah menjadi kelas lain di kelas prediksinya.

FN kelas normal = 1, artinya terdapat 1 data

Tabel 7. Hasil Pengujian Menggunakan Jarak Euclidean dan Minkowski.

Nilai k	d=1 θ=0°	d=2 θ=45°	d=1 θ=90°	d=1 θ=135°	d=2 θ=0°	d=2 θ=45°	d=2 θ=90°	d=2 θ=135°
1	77.78 %	75.56 %	77.78 %	75.56 %	77.78 %	80.00 %	77.78 %	75.56 %
3	75.56 %	71.11 %	66.67 %	68.89 %	64.44 %	73.33 %	71.11 %	77.78 %
5	84.44 %	84.44 %	84.44 %	86.67 %	86.67 %	91.11 %	86.67 %	93.33 %
7	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	91.11 %	84.44 %	91.11 %
9	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	91.11 %	88.89 %	91.11 %
11	88.89 %	88.89 %	88.89 %	88.89 %	88.89 %	91.11 %	88.89 %	91.11 %

Tabel 8. Confusion Matrix Pengujian Sistem.

		Kelas Prediksi		
		Normal	Imatur	Matur
Kelas Aktual	Normal	14	1	0
	Imatur	0	15	0
	Matur	0	2	13

kondisi normal pada data uji yang terdeteksi salah oleh sistem menjadi kondisi kelas lain.

FN kelas imatur = 0, artinya semua kondisi kelas imatur pada data uji berhasil diklasifikasikan secara tepat oleh sistem, tidak terjadi kesalahan klasifikasi untuk kondisi kelas imatur.

FN kelas matur = 2, artinya terdapat 2 data kondisi matur yang terdeteksi salah oleh sistem menjadi kondisi kelas lain.

3. *False Positive* (FP)

FP kelas normal = 0, artinya tidak ada kelas imatur dan kelas matur data uji yang terdeteksi sebagai kelas normal oleh sistem.

FP kelas imatur = 3, artinya terdapat 3 data dari data uji kelas normal atau kelas matur yang terdeteksi sebagai kelas imatur oleh sistem.

FP kelas matur = 0, artinya tidak terdapat kelas normal dan kelas imatur pada data uji yang terdeteksi sebagai kelas matur oleh sistem.

4. *True Negative* (TN)

TN kelas normal = 30, artinya 15 data kelas imatur dan 15 data kelas matur pada data uji tidak ada yang terdeteksi sebagai kelas normal.

TN kelas imatur = 27, artinya 14 data pada kelas normal dan 15 data kelas pada data uji tidak ada yang terdeteksi sebagai kelas imatur, akan tetapi terdapat 3 data (1 kondisi normal dan 2 kondisi matur) yang salah terdeteksi sebagai kelas imatur oleh sistem.

TN kelas matur = 30, artinya 15 data kelas normal dan 15 data kelas imatur pada data uji tidak ada yang terdeteksi sebagai kelas matur.

Tabel 9. Waktu Komputasi.

No.	GLCM		K-NN		Akurasi	Komputasi(s)
	<i>d</i>	θ	Jarak	K		
1.	2	135°	<i>Minkowski</i>	5	93.33 %	1.27
2.	2	135°	<i>Euclidean</i>	5	93.33 %	1.28
3.	1	45°	<i>Chebyshev</i>	7	91.11 %	1.27
4.	1	135°	<i>Cityblock</i>	9	91.11 %	1.29

Berdasarkan hasil tersebut dapat dilihat bahwa kesalahan klasifikasi terjadi pada kelas normal yang terdeteksi sebagai kelas imatur oleh sistem, dan kondisi kelas matur yang terdeteksi sebagai kelas imatur oleh sistem. Hal tersebut dapat terjadi dikarenakan kondisi imatur merupakan kondisi transisi dari kondisi normal menjadi kondisi katarak matur sehingga memiliki potensi untuk terjadinya kesalahan klasifikasi oleh sistem. Akan tetapi, tidak ada data normal yang terdeteksi sebagai kondisi matur ataupun sebaliknya kondisi matur yang terdeteksi sebagai kondisi normal, sehingga dapat disimpulkan bahwa sistem yang dikembangkan cukup berhasil dalam mengenali kondisi mata sesuai dengan kelas seharusnya dengan akurasi sebesar 93,33% dan *error rate* sebesar 6,67 %.

$$Akurasi = \frac{42}{45} \times 100\% = 93,33\%$$

$$Error\ Rate = \frac{3}{45} \times 100\% = 6,67\%$$

Hasil tersebut menunjukan bahwa metode ekstraksi ciri yang digunakan yaitu GLCM dengan fitur statistik *Sum Entropy, Variance, Sum Variance, dan Maximal Correlation Coefficient* merupakan fitur yang cukup signifikan dalam membedakan kondisi mata normal, kondisi imatur dan kondisi matur sehingga dapat disimpulkan metoda ekstraksi ciri yang diusulkan memberikan fitur ciri yang dapat meminimalisir terjadinya *overfitting* pada tahap klasifikasi.

Pada tahap klasifikasi pengujian menggunakan jarak *Minkowski* dan *Euclidean* akurasi sistem mencapai optimal pada saat nilai $k = 5$, untuk nilai k yang lebih besar dari $k = 5$ akurasi sistem kembali menurun. Hal ini sesuai dengan teori bahwa nilai k yang semakin besar dapat mengurangi *noise* pada tahap klasifikasi akan tetapi dapat membuat batasan antar setiap klasifikasi menjadi tidak jelas sehingga dapat menurunkan akurasi. Oleh karena itu diperlukan pengujian untuk mendapatkan kombinasi paling optimal yang memberikan akurasi terbaik.

Berdasarkan Tabel 9, waktu komputasi paling kecil adalah sebesar 1,27 detik dan akurasi sistem tertinggi sebesar 93,33 % menggunakan fitur *Sum Entropy, Variance, Sum Variance, dan Maximal Correlation Coefficient* yang merupakan hasil ekstraksi matriks ko-okurensi dengan $d = 2$ dan $\theta = 135^\circ$, dan

menggunakan perhitungan jarak *Minkowski* ($p = 5$) dengan $k = 5$ sebagai klasifikator. Nilai k yang lebih kecil dapat memberikan waktu komputasi lebih cepat walaupun perbedaannya tidak signifikan.

4. Kesimpulan

Hasil dari penelitian ini menunjukkan bahwa sistem deteksi dini katarak dan klasifikasi kondisi mata berhasil dalam mengenali kondisi mata ke dalam tiga kondisi, yaitu, normal, katarak matur dan katarak imatur. Hasil ini lebih berkembang dari penelitian sebelumnya, dimana sistem hanya dapat mengenali dua kondisi mata yaitu kondisi normal dan katarak saja. Berdasarkan hasil pengujian yang telah diperoleh, ketepatan klasifikasi menggunakan K-NN dipengaruhi oleh parameter GLCM yang digunakan berupa jarak antar piksel (d), sudut (θ), dan fitur statistik yang digunakan, serta nilai k dan persamaan jarak dalam K-NN. Kombinasi yang tepat dari parameter yang digunakan akan memberikan performansi sistem yang optimal. Performansi terbaik dalam penelitian ini mencapai akurasi sistem 93,33% dan *error rate* sebesar 6.67 % dengan waktu komputasi 1,27 detik. Performansi sistem terbaik tersebut diperoleh dengan menggunakan parameter terbaik dalam GLCM yaitu jarak $d = 2$, $\theta = 135^\circ$, serta fitur statistik *Sum Entropy, Variance, Sum Variance, Maximal Correlation Coefficient*, dan nilai parameter terbaik dalam k-NN yaitu, nilai $k = 5$ dengan menggunakan perhitungan jarak *Euclidean* dan *Minkowski*.

Daftar Pustaka

- [1] K. K. R. Indonesia, "Gangguan penglihatan masih menjadi masalah kesehatan. pusat komunikasi publik sekretariat jenderal kementerian kesehatan ri; 2010 [cited 2013 february 9]."
- [2] B. Gary and H. Taylor, "Cataract blindness—challenges for 21st century," *Bulletin of the World Health Organization*, 2001.
- [3] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [4] I. Wijayanto, A. Rizal, and S. Hadiyoso, "Multilevel Wavelet Packet Entropy and Support Vector Machine for Epileptic EEG Classification," in *2018 4th International Conference on Science and Technology (ICST)*. IEEE, aug 2018, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/8528634/>
- [5] R. Supriyanti, H. Habe, M. Kidode, and S. Nagata, "Extracting appearance information inside the pupil for cataract screening." in *MVA*, 2009, pp. 342–345.

- [6] J. Nayak, "Automated classification of normal, cataract and post cataract optical eye images using svm classifier," in *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1, 2013, pp. 23–25.
- [7] M. Caixinha, E. Velte, M. Santos, and J. B. Santos, "New approach for objective cataract classification based on ultrasound techniques using multiclass svm classifiers," in *Ultrasonics Symposium (IUS), 2014 IEEE International*. IEEE, 2014, pp. 2402–2405.
- [8] Y. N. Fuadah, A. W. Setiawan, T. L. Mengko *et al.*, "Mobile cataract detection using optimal combination of statistical texture analysis," in *Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), 2015 4th International Conference on*. IEEE, 2015, pp. 232–236.
- [9] G. Batista and D. F. Silva, "How k-nearest neighbor parameters affect its performance," in *Argentine symposium on artificial intelligence*. sn, 2009, pp. 1–12.
- [10] M. Palondongan and Fuadah, "Performansi optimasi metode gray level co-occurrence matrix (glcm) untuk klasifikasi katarak," in *Prosiding Nasional SNETO 2017*. SNETO, 2017.
- [11] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using glcm approach," *International Journal of Scientific and Research Publications*, vol. 3, no. 5, p. 1, 2013.
- [12] R. M. Haralick, K. Shanmugam *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.
- [13] K. Ganis, I. Santoso, R. R. Isnanto *et al.*, "Klasifikasi citra dengan matriks ko-okurensi aras keabuan (gray level co-occurrence matrix-glcm) pada lima kelas biji-bijian," Ph.D. dissertation, Jurusan Teknik Elektro Fakultas Teknik Undip, 2011.
- [14] A. Kadir and A. Susanto, "Pengolahan citra teori dan aplikasi," *Andi, Yogyakarta*, 2012.