

CESS

(Journal of Computer Engineering, System and Science)

Available online: <https://jurnal.unimed.ac.id/2012/index.php/cess>

ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Klasifikasi Jenis Penyakit Mata Katarak Menggunakan Metode *K-Nearest Neighbors* (KNN)

Classification of Cataract Types Using the K-Nearest Neighbors (KNN) Algorithm

Rizki Ikhwan Pamuji¹, Dian Ahkam Sani^{2*}, Rudi Hariyanto³

^{1,2,3}Teknik Informatika, Universitas Merdeka Pasuruan

Jl. Ir. H. Juanda No.68, Tapaan, Kec. Bugul Kidul, Kota Pasuruan, Jawa Timur 67129

Email: ¹ikhwanpamuji@student.unmerpas.ac.id, ²dian.ahkam@unmerpas.ac.id,

³rudihariyanto@gmail.com

*Corresponding Author

ABSTRAK

Katarak merupakan gangguan penglihatan yang terjadi akibat kekeruhan pada lensa mata, umumnya disebabkan oleh proses degeneratif, paparan radikal bebas, atau gangguan lain seperti glaukoma. Diagnosis dini sangat penting untuk mencegah komplikasi, namun metode konvensional memerlukan pemeriksaan manual yang memakan waktu. Penelitian ini bertujuan mengembangkan model klasifikasi katarak menggunakan algoritma *K-Nearest Neighbors* (KNN), sebuah pendekatan non-parametrik yang menentukan kelas berdasarkan kedekatan fitur terhadap data pelatihan. Dataset yang digunakan terdiri dari 300 data dengan 26 atribut, dan dibagi dengan rasio 80:20. Hasil evaluasi menunjukkan bahwa KNN dengan parameter $k=3$ mampu menghasilkan akurasi sebesar 98,33% dengan *precision* 99%, *recall* 99%, serta *f1-score* 99% pada *macro average* dan pada *weight average* mendapatkan *precision* 98%, *recall* 98%, serta *f-score* 98%. Temuan ini menunjukkan bahwa KNN efektif dalam mendeteksi jenis penyakit mata katarak, meskipun masih terdapat ruang untuk pengembangan lebih lanjut melalui peningkatan jumlah data, teknik preprocessing yang lebih variatif, serta eksplorasi metode klasifikasi lain guna memperoleh hasil yang lebih optimal.

Kata Kunci: *Klasifikasi; Prediksi; Katarak; K-Nearest Neighbors*

ABSTRACT

Cataract is a visual impairment caused by lens clouding, commonly resulting from degenerative processes, exposure to free radicals, or other conditions such as glaucoma. Early diagnosis is crucial to prevent complications; however, conventional methods often rely on time-consuming manual examinations. This study aims to develop a cataract classification model using the *K-Nearest Neighbors* (KNN) algorithm, a non-parametric approach that



classifies data based on the proximity of features to training data. The dataset consists of 300 records with 26 attributes, split into training and testing sets with an 80:20 ratio. Evaluation results show that KNN achieves an accuracy of 98.33%, with a macro-averaged precision, recall, and F1-score of 99%, and a weighted average precision, recall, and F1-score of 98%. These findings indicate that KNN is effective in detecting cataract types, although further improvements can be made through increased data volume, more diverse preprocessing techniques, and exploration of alternative classification methods to enhance performance.

Keywords: *Classification; Prediction; Cataract; K-Nearest Neighbors (KNN)*

1. PENDAHULUAN

Pancaindra merupakan lima alat sensorik pada tubuh manusia yang berfungsi untuk mengenali dan merespons rangsangan dari lingkungan sekitar, salah satunya yaitu mata[1]. Mata merupakan organ penglihatan yang bekerja dengan menangkap cahaya melalui lensa transparan berisi protein larut air, lalu memproyeksikannya ke retina sebagai sinyal saraf[2]. Lensa mata dilindungi oleh kapsul elastis dengan kemampuan akomodasi, namun paparan radikal bebas dapat menyebabkan penggumpalan protein tak larut air, mengurangi kejernihan lensa dan memicu gangguan mata seperti katarak.

Katarak adalah penyakit mata yang ditandai dengan kekeruhan lensa akibat denaturasi protein atau ketidakseimbangan cairan, sehingga menghambat masuknya cahaya ke retina dan menurunkan kualitas penglihatan[3]. WHO mencatat bahwa 51% kasus kebutaan global disebabkan oleh katarak. Selain lansia, katarak juga dapat dipicu oleh faktor bawaan dan penyakit mata lain seperti glaukoma dan uveitis[4]. Klasifikasi penyebab katarak menjadi enam kategori: senilis, kongenital, traumatik, komplikata, toksik, dan sekunder[5].

Diagnosis dini katarak penting untuk mencegah komplikasi, namun metode konvensional memerlukan pemeriksaan manual oleh dokter mata yang membutuhkan waktu dan keahlian khusus[6]. Oleh karena itu, sistem otomatis berbasis teknologi dibutuhkan untuk deteksi yang lebih cepat dan akurat[7].

Salah satu pendekatan yang menjanjikan adalah algoritma K-Nearest Neighbors (KNN)[8], metode pembelajaran non-parametrik yang mengklasifikasikan data berdasarkan kedekatan fitur terhadap data latih[9]. KNN unggul karena kesederhanaannya, fleksibel terhadap ukuran dataset, dan tidak memerlukan asumsi distribusi data[10][11].

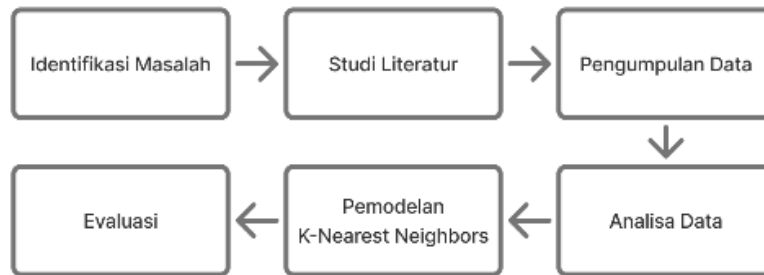
Berbagai penelitian menunjukkan performa tinggi KNN dalam klasifikasi medis. Penelitian yang mirip memperoleh akurasi 91,76% dalam klasifikasi jenis katarak[5]. Algoritma lain seperti LVQ bahkan mencapai akurasi 99%[7], dan varian fuzzy KNN mencapai 96,1% dalam diagnosis autisme[12].

Namun, KNN memiliki tantangan dalam menentukan nilai K yang optimal serta efisiensi dalam dataset besar[13]. Penelitian ini bertujuan membangun dan mengevaluasi model klasifikasi katarak berbasis KNN untuk mendukung diagnosis cepat dan akurat[14]. Diharapkan hasil penelitian ini dapat mendukung pengembangan sistem diagnosis cerdas, khususnya untuk masyarakat dengan keterbatasan akses layanan spesialis[15].

2. METODE PENELITIAN

Dalam metode penelitian ini terdapat beberapa alur yang dilakukan yaitu identifikasi masalah, studi literatur, pengumpulan data, analisa data, pemodelan menggunakan *K-Nearest*

Neighbors, dan evaluasi model. Gambar 1 dibawah ini merupakan diagram alur metode penelitian.



Gambar 1. Alur Penelitian

2.1. Identifikasi Masalah

Masalah yang diidentifikasi pada penelitian ini yaitu mengenai penyakit mata katarak yang merupakan kondisi kekeruhan pada lensa mata yang dapat mengganggu indra penglihatan, menurunkan kualitas hidup, serta dapat berpotensi akan kebutaan. Jumlah penderita katarak di dunia terus mengalami peningkatan setiap tahunnya. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasi jenis katarak antara lain seperti katarak senilis, kongenital serta komplikata berdasarkan faktor-faktor yang mempengaruhi kemunculan nya sehingga perkembangan akan penyakit ini dapat dicegah dan ditangani lebih dini.

2.2. Studi Literatur

Dalam tahap berikutnya pada penelitian ini, dilakukan eksplorasi literatur dengan melakukan pencarian referensi dari beberapa jurnal yang relevan dengan topik dan judul yang telah ditetapkan yaitu "Klasifikasi Jenis Penyakit Mata Katarak Menggunakan Metode K-Nearest Neighbors". Pada penelitian dengan judul "Implementasi Metode K-Nearest Neighbor Untuk Mengklasifikasi Jenis Penyakit Katarak" juga mengklasifikasikan katarak berdasarkan jenisnya yaitu senilis, kongenital, serta traumatik. Data yang digunakan sebanyak 170 data yang dibagi menjadi 10 bagian dengan masing-masing 16 data latih serta 1 data uji yang mendapatkan hasil akurasi sebesar 91,76%[5].

Pada penelitian lainnya dengan objek yang sama yakni katarak tetapi dengan menggunakan metode yang berbeda berjudul "Klasifikasi Penyakit Mata Katarak Berdasarkan Kelainan Patologis Dengan Menggunakan Algoritma *Learning Vector Quantization*" menjelaskan bahwa penelitian ini mengklasifikasikan katarak berdasarkan data rekam medis pasien dengan data sebanyak 300 data latih dan 85 data uji yang mendapatkan hasil akurasi sebesar 99% dengan waktu komputasi hanya sekitar 0,01 detik[7].

Berbeda halnya dengan penelitian lainnya, penelitian yang berjudul "Optimasi Nilai K Pada Algoritma K-Nearest Neighbor Untuk Klasifikasi Pasien Covid-19 Yang Membutuhkan Ruang ICU" menggunakan metode K-Nearest Neighbors tetapi bukan katarak sebagai objeknya melainkan objek lain yakni pasien Covid19. Data sebanyak 32.662 data rekam medis pasien Covid19 digunakan untuk data latih 70% dan data uji 30% dengan nilai k optimal yang diperoleh dari uji k-fold: k = 16 serta akurasi terbaik yang didapatkan sebesar 86,47%[11].

Penelitian lainnya yang mirip juga dengan judul "Diagnosis Awal *Autism Spectrum Disorder* Menggunakan Algoritma *Fuzzy K-Nearest Neighbor*" menjelaskan bahwa objek penelitiannya yakni penderita autism spectrum disorder yang menggunakan salah satu varian berbeda dari K – Nearest Neighbors yaitu dengan metode Fuzzy K – Nearest Neighbors. Data

yang digunakan pada penelitian ini sebanyak 85 data responden dengan data latih 70% dan data uji 30% serta dapat menghasilkan akurasi tertinggi mencapai 96,1% dengan parameter $k = 5$ [12].

2.3. Pengumpulan Data

Pada tahapan pengumpulan data, penelitian ini menggunakan dataset yang diperoleh dari RSUD dr. Saiful Anwar Malang dengan data berupa excel. Dataset tersebut sebanyak 300 data rekam medis pasien penderita katarak dengan memuat fitur sebanyak 26 yang berkaitan dengan kondisi mata pasien dan faktor pendukung lainnya. Dataset ini akan digunakan untuk proses klasifikasi jenis katarak diantaranya, yaitu katarak senilis, katarak kongenital, dan katarak komplikata. Pemilihan data ini didasarkan pada kelengkapan atribut yang sesuai dengan klasifikasi katarak serta ketersediaan label yang telah ditentukan sebelumnya. Berikut dataset yang digunakan melalui tabel dibawah.

Tabel 1. Data Dengan Kolom Fitur No. Rekam Medis dan Nama

No	No. Rekam Medis	Nama	Usia
1	234	E	60
2	324	T	68
3	235	H	73
4	246	R	65
5	253	W	6
...
300	932	G	63

Tabel 2. Data Dengan Kolom Fitur Jenis Kelamin, Kabur, Silau, dan Trauma

Jenis Kelamin	Kabur	Silau	Trauma	...
1	1	1	0	...
1	0	1	0	...
0	1	0	0	...
1	1	1	0	...
1	1	1	0	...
...
0	0	1	0	...

Pada kolom fitur usia menjelaskan usia pasien berdasarkan tahun dan fitur jenis kelamin menjelaskan bahwa nilai 0 untuk perempuan dan 1 untuk laki-laki. Sedangkan pada fitur kabur, silau, trauma, dan yang lain menjelaskan bahwa nilai 0 untuk "tidak" mengalami dan 1 untuk "iya" mengalami.

Tabel 3. Data Dengan Kolom Fitur Tekanan Intra Okuli dan Hasil

...	Tekanan Intra Okuli	...	Hasil
...	21	...	Senilis
...	34	...	Senilis
...	27	...	Senilis

...	25	...	Senilis
...	34	...	Kongenital
...
...	38	...	Senilis

Pada kolom tekanan intra okuli berisikan informasi angka tekanan pada pemeriksaan sedangkan kolom hasil berisikan informasi diagnosa jenis penyakit katarak yang didapat.

2.4. Analisis Data

Pada tahap analisis data dalam penelitian ini, dilakukan pemeriksaan terhadap data untuk memperoleh informasi yang sesuai serta ekstraksi data menjadi numerik. Selain itu, analisis ini bertujuan untuk memahami pola-pola yang terdapat dalam data. Langkah-langkah yang dilakukan meliputi identifikasi tipe data pada setiap kolom, pengecekan data yang hilang (*missing values*), serta analisis korelasi antar data menggunakan visualisasi *heatmap*.

2.5. Pemodelan K-Nearest Neighbors

Setelah tahap analisis data, dilakukan proses pemodelan menggunakan algoritma *K-Nearest Neighbors* (KNN) untuk melakukan klasifikasi jenis katarak. *K-Nearest Neighbors* (KNN) adalah algoritma klasifikasi berbasis jarak yang membandingkan data baru dengan data latih menggunakan *Euclidean Distance*. Algoritma ini menerapkan prinsip dasar *geometri Euclidean* yang diadaptasi dari teorema *Pythagoras* untuk menentukan kedekatan antar data.

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

Keterangan:

d = Jarak Euclidean

x = Data baru

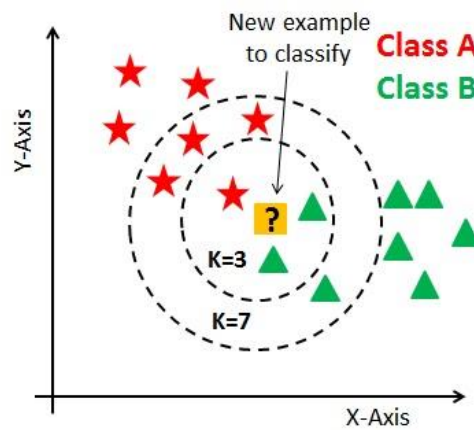
y = Data referensi

\sum = Penjumlahan semua fitur

x_i = Data fitur x

y_i = Data fitur y

Metode KNN bekerja dengan cara menghitung jarak antara data uji dan data latih, kemudian menentukan kelas berdasarkan mayoritas dari sejumlah tetangga terdekat (k) yang memiliki jarak paling dekat. Semakin dekat jaraknya, semakin besar pengaruhnya terhadap hasil klasifikasi.



Gambar 2. Metode KNN

Algoritma ini tidak memerlukan proses pelatihan model secara eksplisit, karena termasuk dalam kategori *lazy learner*. Pemilihan nilai k dan metode pengukuran jarak seperti *Euclidean Distance* menjadi komponen penting dalam membangun model klasifikasi yang akurat dan efektif dalam mengelompokkan jenis katarak.

2.6. Evaluasi

Setelah proses pemodelan menggunakan metode *K-Nearest Neighbors* dilakukan, maka akan masuk ke tahapan evaluasi model. Dalam evaluasi model pada penelitian ini metode yang digunakan adalah *confusion matrix*. *Confusion Matrix* adalah metode yang digunakan untuk menilai kemampuan model klasifikasi dalam mengenali data dari berbagai kategori. Evaluasi kinerja model ini biasanya dilakukan dengan mengukur nilai *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Adapun rumus-rumus perhitungan tersebut didasarkan pada *Confusion Matrix*.

Tabel 4. *Confusion Matrix*

<i>Confusion Matrix</i>	<i>Predictive Positive</i>	<i>Predictive Negative</i>
<i>Actual Positive</i>	TP	FP
<i>Actual Negative</i>	TN	FN

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{F1-Score} = \frac{2TP}{2TP+FP+FN} \quad (5)$$

Keterangan:

TP = Nilai prediksi positif dengan nilai sebenarnya positif.

TN = Nilai prediksi positif dengan nilai sebenarnya negatif.

FP = Nilai prediksi negatif dengan nilai sebenarnya positif.
 FN = Nilai prediksi negatif dengan nilai sebenarnya negatif

3. HASIL DAN PEMBAHASAN

3.1. Import Dataset

Import dataset merupakan proses memperoleh serta menyusun data secara teratur untuk keperluan analisis atau pemrosesan lebih lanjut. Adapun tahapan dalam proses import dataset sebagai berikut:

```
# koneksi ke google drive
from google.colab import drive
drive.mount('/content/drive')

# masukan file excel
import pandas as pd
df = pd.read_excel('/content/drive/My Drive/Rekam Medis Katarak.xlsx')
print(df.head())
```

Gambar 3. Proses Import Library dan Input Dataset

Langkah awal yaitu koneksikan *google drive* ke *google collab* untuk memulai pemrograman. Kemudian masukkan dataset dari *google drive* ke *google collab*. Jika dataset berupa file excel maka pada kode tambahkan *xlsx* agar excel dapat terbaca. Berikut tampilan *output* dataset pada Gambar 4.

NO	NOREK	NAMA	JENIS KELAMIN	UMUR	KABUR	SILAU	TRAUMA	\
0	1	234	E	1	60	1	1	0
1	2	324	T	1	68	0	1	0
2	3	235	H	0	73	1	0	0
3	4	246	R	1	65	1	1	0
4	5	253	W	1	6	1	0	0

	RADANG BILIK MATA DEPAN	LENSA KERUH TIDAK RATA	...	RW. KEHAMILAN	\
0	1	1	1	...	0
1	0	0	0	...	0
2	1	1	1	...	0
3	1	1	1	...	0
4	0	1	1	...	1

	Rw. Skt MATA UVETIS	Rw. Skt MATA GLAUCOMA	\
0	0	0	
1	0	0	
2	0	0	
3	0	0	
4	0	0	

	RW. KELUARGA TIDAK BISA MELIHAT	KELAINAN JANTUNG BAWAAN	\
0	0	0	
1	0	0	
2	0	0	
3	0	0	
4	1	1	

	GANGGUAN PENDENGARAN	PARASIT TOXO PLASMA	ROBELA/CAMPAK	\
0	0	0	0	
1	0	0	0	

Gambar 4. Dataset

Pada dataset tersebut berjumlah sebanyak 300 data dengan 26 fitur memuat no rekam medis, nama pasien, jenis kelamin, umur, kabur, riwayat keluarga tidak bisa melihat, tekanan intra okuli, dll. Pada fitur jenis kelamin nilai 0 artinya “perempuan” sedangkan nilai 1 artinya “laki-laki”. Pada fitur lainnya nilai 0 adalah “tidak” yang artinya pasien tidak memiliki gejala pada fitur tersebut sedangkan nilai 1 adalah “iya” yang artinya pasien memiliki gejala tersebut. Untuk fitur nama pasien sudah disamarkan dengan mengambil huruf depannya saja.

3.2. Analisa Data

Analisis data merupakan tahapan untuk memahami karakteristik dari suatu dataset dengan meninjau informasi seperti jenis data apa saja pada setiap atribut, distribusi nilai, serta

identifikasi nilai yang hilang (*missing values*). Berikut merupakan tahapan dalam proses eksplorasi data:

```
Tipe Data Setiap Kolom:  
NO int64  
NOREK int64  
NAMA object  
JENIS KELAMIN int64  
UMUR int64  
KABUR int64  
SILAU int64  
TRAUMA int64  
RADANG BILIK MATA DEPAN int64  
LENSA KERUH TIDAK RATA int64  
LENSA KERUH RATA int64  
LENSA KERUH PENCAIRAN CORTEX int64  
MICROFTALMIA int64  
MICROCEPALI int64  
TEKANAN INTRA OKULI / TIO int64  
ANTARIAR/COKLAT int64  
RW. KEHAMILAN int64  
RW. Skt MATA UVETIS int64  
RW. Skt MATA GLAUCOMA int64  
RW. KELUARGA TIDAK BISA MELIHAT int64  
KELAINAN JANTUNG BAWAAN int64  
GANGGUAN PENDENGARAN int64  
PARASIT TOXO PLASMA int64  
ROBELA/CAMPAK int64  
CITO MEGALO VIRUS int64  
HASIL object  
dtype: object
```

Gambar 5. Tipe Data

```
Jumlah Missing Value per Kolom:  
NO 0  
NOREK 0  
NAMA 0  
JENIS KELAMIN 0  
UMUR 0  
KABUR 0  
SILAU 0  
TRAUMA 0  
RADANG BILIK MATA DEPAN 0  
LENSA KERUH TIDAK RATA 0  
LENSA KERUH RATA 0  
LENSA KERUH PENCAIRAN CORTEX 0  
MICROFTALMIA 0  
MICROCEPALI 0  
TEKANAN INTRA OKULI / TIO 0  
ANTARIAR/COKLAT 0  
RW. KEHAMILAN 0  
RW. Skt MATA UVETIS 0  
RW. Skt MATA GLAUCOMA 0  
RW. KELUARGA TIDAK BISA MELIHAT 0  
KELAINAN JANTUNG BAWAAN 0  
GANGGUAN PENDENGARAN 0  
PARASIT TOXO PLASMA 0  
ROBELA/CAMPAK 0  
CITO MEGALO VIRUS 0  
HASIL 0  
dtype: int64
```

Gambar 6. Cek Missing Values

Pada Gambar 5. dan Gambar 6. Menjelaskan bahwa mayoritas tipe data bersifat integer dan tidak ada data kosong atau tidak diisi pada dataset tersebut sehingga aman untuk tahapan selanjutnya.

3.3. Preprocessing Data

Tahapan *preprocessing* adalah proses perubahan data mentah menjadi data yang siap digunakan untuk tahapan pemodelan. Pada penelitian ini, proses *preprocessing* mencakup proses pembuangan data yang tidak dibutuhkan, penerapan label encoding pada data kategorikal yang masih digunakan serta korelasi semua fitur dengan visualisasi *heatmap*.


```
# Hilangkan kolom no, norek, dan nama
df_cleaned = df.drop(columns=['no', 'NOREK', 'NAMA'])

# Tampilkan semua kolom saat print
pd.set_option('display.max_columns', None)

# Buat drop aman biar gak error kalau kolom gak ketemu
df_cleaned = df.drop(columns=['no', 'NOREK', 'NAMA'], errors='ignore')

# Lihat data awal
df_cleaned.head()
```

JENIS KELAMIN	UMUR	KABUR	SILAU	TRAUMA	RADANG BILIK MATA DEPAN	LENSA KERUH TIDAK RATA	LENSA KERUH RATA	LENSA KERUH PENCAHRAIN CORTEX	MICROFTALMIA	MICROCEPALI	TEKANAN INTRA OKULI / TIO
0	1	60	1	1	0	1	1	0	0	0	21
1	1	68	0	1	0	0	0	1	1	0	34
2	0	73	1	0	0	1	1	0	0	0	27
3	1	65	1	1	0	1	1	0	0	0	25
4	1	6	1	0	0	0	1	0	0	1	34

Gambar 7. Penghapusan Fitur Tidak Terpakai

```
from sklearn.preprocessing import LabelEncoder

# Initialize LabelEncoder
le = LabelEncoder()

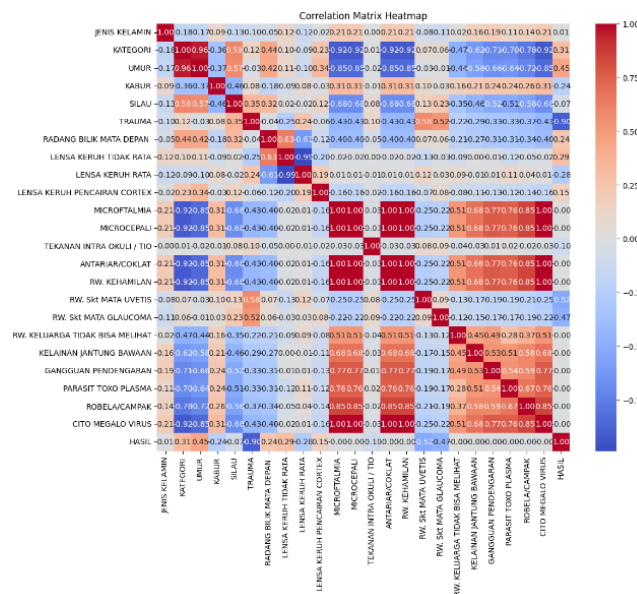
# Apply label encoding to 'HASIL'
df_cleaned['HASIL'] = le.fit_transform(df_cleaned['HASIL'])

# Display the updated DataFrame
df_cleaned.head()

# Tampilkan mapping label-nya
print("Mapping hasil LabelEncoder:")
for i, label in enumerate(le.classes_):
    print(f"{label} → {i}")
```

Mapping hasil LabelEncoder:
 KOMPLIKATA → 0
 KONGENITAL → 1
 SINILIS → 2

Gambar 8. Label Encoding



Gambar 9. Korelasi Data

Pada Gambar 7. fitur yang tidak diperlukan untuk tahapan selanjutnya antara lain: fitur no, no rekam medis, dan nama pasien. Sedangkan pada Gambar 8. menunjukkan pada fitur hasil akan dilakukan label encoding dimana komplikata sebagai 0, kongenital sebagai 1 dan senilis sebagai 2. Setelah itu pada Gambar 9. yaitu mengkorelasikan semua fitur yang dimana jika warna semakin merah maka terindikasi antar fitur saling berkorelasi, sedangkan jika warna semakin biru maka antar fitur tidak saling berkorelasi.

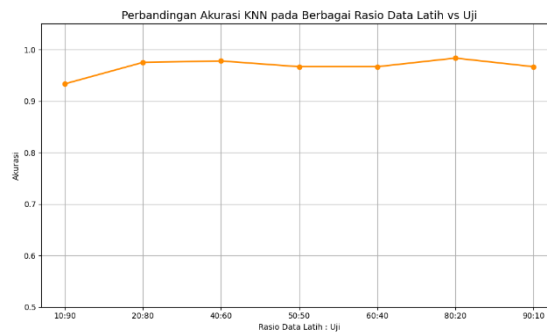
3.4. Pemodelan K-Nearest Neighbors

Setelah tahapan ekstraksi data sudah siap, maka tahapan selanjutnya yakni pembagian data latih dan data uji. Rasio pembagian bisa dilihat pada Gambar 10.

Rasio 10:90 → Train: (30, 23), Test: (270, 23)
 Rasio 20:80 → Train: (60, 23), Test: (240, 23)
 Rasio 40:60 → Train: (120, 23), Test: (180, 23)
 Rasio 50:50 → Train: (150, 23), Test: (150, 23)
 Rasio 60:40 → Train: (180, 23), Test: (120, 23)
 Rasio 80:20 → Train: (240, 23), Test: (60, 23)
 Rasio 90:10 → Train: (270, 23), Test: (30, 23)

Gambar 10. Rasio Data

Setelah pembagian rasio data, maka akan dilanjut ke tahapan penghitungan metode *K-Nearest Neighbors* dengan menggunakan parameter $k=3$ sebagai acuan.



Gambar 11. Diagram Rasio Perbandingan

Akurasi: 98.33%

Classification Report:				
	precision	recall	f1-score	support
0	0.96	1.00	0.98	24
1	1.00	1.00	1.00	12
2	1.00	0.96	0.98	24
accuracy			0.98	60
macro avg	0.99	0.99	0.99	60
weighted avg	0.98	0.98	0.98	60

Gambar 12. Hasil Klasifikasi K-Nearest Neighbors

Pada Gambar 11. bisa dilihat bahwa rasio tertinggi berada pada rasio 80:20 dengan parameter $k=3$ yang memberikan hasil akurasi sebesar 98,33% dengan *precision* 99%, *recall* 99%, serta *f1-score* 99% pada *macro average* dan pada *weight average* mendapatkan *precision* 98%, *recall* 98%, serta *f-score* 98% bisa dilihat pada Gambar 12.

3.5. Evaluasi

Tahapan akhir adalah evaluasi dimana akan ditampilkan menggunakan *confusion matrix* yang ditampilkan pada Gambar 13.

Confusion Matrix:

$$\begin{bmatrix} 24 & 0 & 0 \\ 0 & 12 & 0 \\ 1 & 0 & 23 \end{bmatrix}$$

Gambar 13. Confusion Matrix

Hasil prediksi menunjukkan bahwa kelas 0 sebagai komplikata memberikan hasil sempurna serta kelas 1 sebagai kongenital juga memperoleh hasil sempurna. Sedangkan kelas 2 sebagai senilis memberikan prediksi benar 23 dan 1 salah yang memprediksi ke kelas 0 sebagai komplikata.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode *K-Nearest Neighbors* mampu menghasilkan klasifikasi yang efektif dalam mendiagnosis jenis penyakit mata katarak, berdasarkan data yang diperoleh dari layanan kesehatan. Dataset yang digunakan terdiri dari 300 data berisikan 26 fitur. Hasil prediksi menggunakan metode ini menghasilkan akurasi sebesar 98,33% dengan *precision* 99%, *recall* 99%, serta *f1-score* 99% pada *macro average* dan pada *weight average* mendapatkan *precision* 98%, *recall* 98%, serta *f-score* 98% pada rasio 80:20 dengan parameter $k=3$. Meskipun performa tersebut tergolong tinggi, namun hal ini membuka peluang untuk pengembangan lebih lanjut, seperti peningkatan variasi dan jumlah data, serta penerapan teknik *preprocessing* yang lebih beragam guna mengoptimalkan akurasi model. Selain itu, eksplorasi metode lain juga disarankan untuk melakukan perbandingan kinerja dalam klasifikasi penyakit mata katarak. Dengan mempertimbangkan aspek tersebut, diharapkan penelitian lanjutan dapat menghasilkan model klasifikasi yang lebih optimal.

DAFTAR PUSTAKA

- [1] A. Sandy Wardhani, F. Tri Anggraeny, and A. Mustika Rizki, "Penerapan Model Hibrida CNN-KNN Untuk Klasifikasi Penyakit Mata," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, Jun. 2024.
- [2] D. Hananta Firdaus, B. Imran, L. Darmawan Bakti, and E. Suryadi, "Klasifikasi Penyakit Katarak Pada Mata Menggunakan Metode Convolutional Neural Network (CNN) Berbasis Web," *Jurnal Kecerdasan Buatan dan Teknologi Informasi (JKBTI)*, vol. 1, no. 3, pp. 18–26, Dec. 2022.
- [3] P. Astari, "Katarak: Klasifikasi, Tatalaksana, dan Komplikasi Operasi," *CDK*, vol. 45, pp. 748–753, 2018.
- [4] R. B. J. Simanjutak, Y. Fu'adah, R. Magdalena, S. Saidah, A. B. Wiratama, and I. D. S. Ubaidah, "Cataract Classification Based on Fundus Images Using Convolutional Neural Network," *International Journal on Informatics Visualization*, pp. 33–38, Mar. 2022, [Online]. Available: www.joiv.org/index.php/joiv
- [5] M. Safaat, A. Sahari, and D. Lusiyanti, "Implementasi Metode K-Nearest Neighbor Untuk Mengklasifikasi Jenis Penyakit Katarak," *Jurnal Ilmiah Matematika Dan Terapan*, vol. 17, no. 1, pp. 92–99, Jun. 2020, doi: 10.22487/2540766x.2020.v17.i1.15184.
- [6] J. A. P. Taba, "Katarak Kongenital: Skrining dan Diagnosis," *CDK*, pp. 399–405, 2021.
- [7] R. Hariyanto, A. Basuki, and R. N. Hasanah, "Klasifikasi Penyakit Mata Katarak berdasarkan Kelainan Patologis dengan menggunakan Algoritma Learning Vector Quantization," *Journal of Mechanical Engineering and Mechatronics Submitted*, vol. 1, no. 2, pp. 86–92, Aug. 2016.
- [8] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Nature Research*, Dec. 2022. doi: 10.1038/s41598-022-10358-x.

- [9] P. K. S, R. N. Shreya, Y. B. C, S. M. Lakshmi, and R. B. N, "Color Detection using K-Nearest Neighbors Classification Algorithm," *International Research Journal of Engineering and Technology*, no. 8, Aug. 2021, [Online]. Available: www.irjet.net
- [10] V. Immanuel Sunarko, D. Lizard Sambawo Dimara, P. Sandya Etniko Siagian, D. Manalu, and F. Tri Anggraeny, "Implementasi K-Fold Dalam Prediksi Hasil Produksi Agrikultur Pada Algoritma K-Nearest Neighbor (KNN)," *INTEGER*, vol. 10, pp. 10–16, Mar. 2025, [Online]. Available: <https://satudata.pertanian.go.id/>
- [11] R. A. Pangestu, Taslim, Y. Yunefri, and E. Sabna, "Optimasi Nilai k Pada Algoritma K-Nearest Neighbor Untuk Klasifikasi Pasien Covid-19 Yang Membutuhkan Ruangan ICU," *JURNAL INOVTEK POLBENG*, vol. 7, no. 1, 2022.
- [12] M. A. Putri, C. A. Putra, and W. S. J. Saputra, "Diagnosis Awal Autism Spectrum Disorder Menggunakan Algoritma Fuzzy K-Nearest Neighbor," *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 6, no. 2, pp. 182–187, May 2024.
- [13] J. Uni, U. Nysa, A. Mahmudi, and K. Auliasari, "Perbandingan Jarak Euclidean, Manhattan, Chebyshev Pada Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbors (KNN)," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 4, 2023.
- [14] D. Anitasari, F. Yulianto, T. Azhima, and Y. Siswa, "Klasifikasi Kecelakaan Lalu Lintas di Kota Samarinda Menggunakan Algoritme K-Nearest Neighbor," *MNEMONIC*, vol. 8, no. 1, Feb. 2025.
- [15] A. Khikam, N. Martyan Anggadimas, and M. Udin, "Implementasi *Decision Tree* Untuk Klasifikasi Obesitas," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 3, 2025.