

# Application of the K-Nearest Neighbor (K-NN) Algorithm for Disease Classification Based on Eye Images

St. Nadira Zalzabila<sup>1</sup>, Sudarmin<sup>2</sup>, Zulkifli Rais<sup>3</sup>

<sup>1,2,3</sup>Department of Statistics, Faculty of Mathematics and Natural Sciences, Makassar State University, Indonesia

**Keywords:** Klasifikasi Citra, K-Nearest Neighbor (K-NN), Penyakit Mata, Gray Level Co-occurrence Matrix (GLCM).

## Abstract:

Eye diseases are one of the health issues that can lead to a decline in visual acuity if not detected and treated at an early stage. With advances in digital image processing and machine learning technologies, eye images can be utilised as a source of information to support the automated classification of eye diseases. This study aims to apply the K-Nearest Neighbour (K-NN) algorithm to classify eye diseases based on eye images. The dataset used in this study consists of 4,217 eye images categorised into four classes: cataract, glaucoma, diabetic retinopathy, and normal. The research stages included image pre-processing, comprising conversion to greyscale images, normalisation, segmentation, and resizing. Next, texture features were extracted using the Gray Level Co-occurrence Matrix (GLCM) method to obtain features representing the characteristics of the eye images. The data were then split into training and test sets in an 80:20 ratio before classification was carried out using the K-NN algorithm. The evaluation results show that the K-NN algorithm with  $K = 7$  achieved an accuracy rate of 81 per cent, indicating that the model is able to classify the majority of eye images accurately.

## 1. Introduction

Data mining is a crucial stage in the process of knowledge discovery in databases, also known as Knowledge Discovery in Databases (KDD) (Sudipa et al., 2023). This process not only emphasises pattern recognition but also involves various interrelated stages, ranging from data management, pre-processing, model selection, inference and pattern evaluation to the visualisation of the results obtained. Developments in data mining have been further facilitated by advances in artificial intelligence (AI), machine learning, statistics, and database systems, which enable data processing to be carried out more effectively and efficiently (Baloch et al., 2023).

One of the most widely used techniques in data mining is classification. Classification is one of the main techniques in data mining, involving the grouping of data into predetermined categories or classes based on existing attributes (Mughnyanti & Subhan, 2023). The aim is to build a model capable of predicting the category or label of unknown data. The benefit of classification in data mining is that a good classification model provides accurate predictions regarding the category or label of unknown data based on patterns learnt from the training data (Sudipa et al., 2023).

In the healthcare sector, classification techniques in data mining are widely applied to identify and diagnose various diseases. Classification enables models to predict disease categories based on patterns learnt from image data or medical records, thereby facilitating accurate and efficient diagnosis (Gunawan et al., 2025). One application that has attracted considerable attention is the classification of diseases based on eye images.

\* Corresponding author.

E-mail address: zalzabilanadira@gmail.com



The eye is one of the five senses that is vital to human survival. The eye is the organ of sight that detects light. A basic function of the eye is to distinguish between light and dark (Suwanda & Juniati, 2022). Corneal disease is one of the leading causes of blindness worldwide, ranking fourth after cataracts, glaucoma and age-related macular degeneration. Based on data from the WHO (2020) and a meta-analysis by Wang et al. (2023), it is estimated that approximately 1.5 to 2 million new cases of monocular blindness due to corneal ulceration occur each year. Incidence and prevalence rates vary significantly between developed and developing countries.

Advances in digital image processing and machine learning technologies have opened up opportunities to develop automated systems for classifying eye diseases. One algorithm frequently used in the classification process is the K-Nearest Neighbour (K-NN) algorithm. The K-NN algorithm works by classifying data based on its proximity to a number of the nearest neighbours from the training data. This method has the advantages of being simple to implement, easy to understand, and capable of delivering good performance in various classification scenarios (Eka Pratiwi et al., 2022).

Several previous studies have shown that the K-NN algorithm has good potential for medical image classification. Research conducted by Ujianto et al. (2025) revealed that the application of K-NN to medical images, with appropriate parameter settings, was able to achieve an accuracy of 90 per cent. Another study by Pamuji et al. (2025) on cataract classification recorded an accuracy of 98.33 per cent following parameter optimisation and image pre-processing. Furthermore, Ilmi et al. (2024) used the Local Binary Pattern (LBP) method for texture feature extraction in retinal fundus images and applied K-NN as the classification algorithm, resulting in an accuracy rate of 93.07%.

In light of these issues, this study applies the K-Nearest Neighbour (K-NN) algorithm to classify eye diseases based on digital images. The classification process involves image pre-processing and the extraction of texture features using the Local Binary Pattern (LBP) and Grey Level Co-occurrence Matrix (GLCM) methods, as well as parameter optimisation to improve the model's performance. This study aims to evaluate the capability of the K-NN algorithm in classifying eye diseases and to analyse the impact of parameter optimisation on classification results. The findings are expected to support the development of a more accurate and efficient early-detection system for eye diseases.

## 2. Literature Review

### 2.1. Machine Learning

Machine learning is a branch of artificial intelligence (AI) that focuses on the development of systems capable of learning from data and past experience without having to be explicitly programmed for each task they perform. Through this learning process, the system can recognise patterns, make predictions and take decisions based on information derived from the data (Kusuma, 2020).

### 2.2. Classification

Classification is a data mining technique used to build models capable of grouping data into specific classes based on patterns learnt from labelled training data (Sudipa et al., 2023). The classification process generally consists of two stages: training, to build the model, and testing, to evaluate the model's accuracy in predicting the class of new, unknown data. Models that have demonstrated good performance can subsequently be used to classify real-world data automatically (Rashad et al., 2023).

The confusion matrix comprises true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN). TP refers to positive data that are correctly detected. FN is the opposite of a true positive; that is, positive data that are incorrectly detected as negative. As for FP, this refers to negative data that is incorrectly detected as positive, whilst TP refers to negative data that is correctly detected (Rashad et al., 2023). From the confusion matrix, several evaluation metrics can be calculated to assess how well a classification model performs in predicting the data in question (Yuliarni, 2024):

- Accuracy is the proportion of correct predictions out of all observations.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision is the proportion of positive cases correctly predicted out of all cases predicted to be positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall is the proportion of positive cases correctly predicted out of all actual positive cases.

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score is the harmonic mean of precision and recall.

$$Recall = \frac{TP}{TP + FN}$$

### 2.3. *K-Nearest Neighbor (K-NN)*

K-Nearest Neighbour (K-NN) is a non-parametric machine learning algorithm widely used for classification and regression tasks. Its non-parametric nature makes this algorithm highly flexible, as it does not require any assumptions regarding the data distribution (Aziza et al., 2024). The basic principle of K-NN is highly intuitive: the algorithm assumes that objects sharing similarities in the feature space are likely to have the same label or value. In other words, K-NN predicts the class of a new data point by identifying the 'k' training data points that are closest to or most similar to that data point (Pauji et al., 2022). Proximity distances via intervals can be calculated using the Euclidean matrix and the Manhattan matrix. According to Sugriyono & Siregar (2020), the formulae for calculating these distances are as follows:

- a) Euclidean Matrix

$$d_1(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- b) Manhattan Matrix

$$d_2(x, y) = \sum_{i=1}^n (x_i - y_i)$$

The K-NN algorithm works by determining the value of k as the number of nearest neighbours to be used, then calculating the distance between the new data point and all the training data using a specific metric (I. P. Putri, 2021), such as Euclidean or Manhattan. The new data is then classified based on the most dominant class among those k nearest neighbours (Sugriyono & Siregar, 2020).

## 2.4. Digital Images

A digital image is a finite set of values representing a two-dimensional image, known as pixels. Pixel values typically represent greyscale levels, colour, height and other characteristics. Furthermore, digitisation implies that a digital image is an approximation of a real-world scene (Setiawan, 2019). Another term for a digital image is 'image', and it is a component of multimedia in the form of visual information produced through the process of digitisation (Sufie, 2023). There are several pre-processing stages for digital images that are generally carried out before proceeding to the feature extraction or classification stages. These stages include (Subandi et al., 2023).

### 2.4.1 Image Conversion

The initial stage in the pre-processing of digital images is usually the conversion of the image from colour (RGB) format to greyscale. This conversion is carried out to simplify the image data by reducing the colour dimensions from three channels (red, green, blue) to a single greyscale intensity channel.

### 2.4.2 Noise Reduction

Noise is random interference that appears in digital images due to various factors such as lighting conditions, camera sensors or data transmission. Noise can obscure important details and reduce image quality, thereby affecting the results of the analysis.

### 2.4.3 Image Normalisation

Normalisation is the process of adjusting the intensity of image pixels so that they fall within a specific range of values, for example 0 to 1 or 0 to 255. The aim of normalisation is to make the image data consistent and uniform, thereby minimising differences in intensity caused by lighting or imaging conditions.

#### 2.4.4 Image Segmentation

Segmentation is the process of separating objects or important areas within an image from their background. This stage is crucial for focusing on the relevant parts of the image, such as objects that are to be recognised or analysed.

#### 2.4.5 Feature Extraction

Feature extraction refers to the key characteristics of an image that can be used to distinguish one object from another or for classification. Features may include texture, shape, colour, or specific patterns extracted from a processed image. Effective feature extraction enhances the performance of pattern recognition or classification algorithms, as the extracted features represent the key characteristics of the objects in the image. The features generated by the extraction method consist of basic statistical features and texture features based on the Grey Level Co-occurrence Matrix (GLCM) method. The feature extraction used in this study comprises statistical features and texture features. The statistical features include mean, standard deviation, minimum, maximum and entropy, whilst the texture features include contrast, correlation, energy and homogeneity.

### 2.5. The Eye

The eye, one of the human sensory organs that functions as an organ of sight, is a complex structure. When we mention the word 'eye', what usually springs to mind is the eyeball. The crystalline lens, which is a transparent, colourless disc-shaped structure, plays a crucial role in our ability to see. This lens is approximately 4 mm thick and 9 mm in diameter. The crystalline lens is situated in the anterior segment of the eyeball, suspended by the zonules behind the iris. In front of the lens lies the aqueous humour, whilst behind it lies the vitreous body. The surface of the lens is divided into the anterior surface and the posterior surface. The anterior surface has an elliptical curvature with the apex of curvature at the centre, known as the anterior pole, with a radius of curvature of approximately 10 mm. The posterior surface is more convex than the anterior surface. The line connecting the anterior and posterior poles is called the lens axis (Yuliarni, 2024).

#### 2.5.1 Normal

A normal eye is one that functions optimally without any structural or pathological abnormalities. A normal eye is capable of receiving and processing light, thereby producing clear and sharp vision (G. R. Putri et al., 2023).

#### 2.5.2 Cataract

A cataract is a condition in which the eye's lens becomes cloudy and opaque. Generally, the development of a cataract is gradual and may not initially cause significant discomfort (Firnando & Suryana, 2023).

#### 2.5.3 Glaucoma

Glaucoma is a condition that affects vision. Elevated intraocular pressure can cause damage to the optic nerve, which forms part of the retina at the back of the eyeball. The optic nerve connects the light-sensitive tissue (the retina) to the part of the brain that processes visual information (Pertiwi & Utomo, 2023).

#### 2.5.4 Diabetic Retinopathy

Diabetic retinopathy is a medical condition characterised by the dilation of blood vessels in the eye; if left untreated, this can lead to the development of new blood vessels that may cover the retina, ultimately resulting in blindness (Rizal et al., 2020).

### 3. Research Methodology

#### 3.1. Type of Research

This study is a quantitative study that aims to apply the K-Nearest Neighbour (K-NN) algorithm to classify diseases based on eye images. The classification process involves image pre-processing, the extraction of statistical and textural features using the Local Binary Pattern (LBP) and Gray Level Co-occurrence Matrix (GLCM) methods, and parameter optimisation to obtain the model with the best performance. Model performance is evaluated using the metrics of accuracy, precision, recall, F1-score, and confusion matrix.

#### 3.2. Data Sources

The data used in this study is secondary data. The data consists of corneal images sourced from (<https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification>). The dataset comprises 4,217 corneal image files.

### 3.3. Research Procedure

The research procedure was carried out in several stages, namely:

1. Collecting image data from the Kaggle dataset.
2. Pre-processing the image data.
3. Building a classification model using the K-NN method.
4. Evaluating the performance of the K-NN model.
5. Drawing interpretations and conclusions based on the results of the K-NN classification model evaluation.
6. Compiling a report on the research findings.

### 3.4. Data Analysis Techniques

The data analysis techniques used include:

1. Importing eye images from a dataset obtained from Kaggle
2. Image pre-processing  
Before being used in the K-NN algorithm, the images undergo a pre-processing stage to improve their quality and reduce noise. These stages include:
  - a) Grayscale conversion, i.e. converting colour images to grayscale to simplify the computational process.
  - b) Normalisation, i.e. adjusting the intensity of image pixels within a range of 0 to 1 or 0 to 255.
  - c) Segmentation, i.e. separating the relevant parts of the eye from the background.
  - d) Resizing: standardising the image size to match the algorithm's input format.
  - e) Feature extraction: extracting key features from each image.
3. Implementation of the K-NN algorithm  
The K-NN algorithm will be implemented using the Google Colab programming language. The implementation steps are as follows:
  - a) Data splitting: the feature-extracted data is divided into training and testing data sets using several splitting ratios, namely 80:20, 70:30 and 60:40. This splitting aims to evaluate the performance of the classification model across various proportions of training and test data.
  - b) Determining the value of K: Testing several different values of K (the number of nearest neighbours) to find the optimal value that yields the best accuracy.
  - c) Performing the classification process: for each test image, the algorithm calculates the distance using the Euclidean or Manhattan distance—between that image and all images in the training data. The algorithm selects the k training images with the closest distances. The class (disease label) of the test image is determined based on the majority class of these k nearest neighbours.
4. Performance Metrics  
To evaluate the performance of the K-NN model, several metrics commonly used in classification will be employed, namely:
  - a) *Accuracy*
  - b) *Precision*
  - c) *Sensitivity*
  - d) *F1-Score*

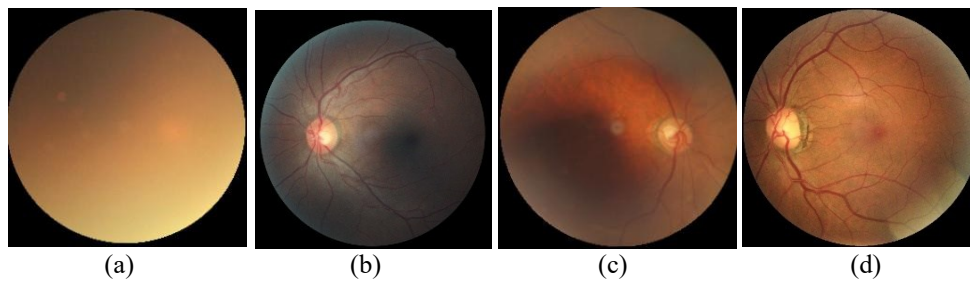
## 4. Results and Discussion

### 4.1. Data Description

This study utilised eye images as its primary subject matter, sourced from an online eye disease dataset. This dataset comprises four categories, namely cataracts, diabetic retinopathy, glaucoma and normal eyes, with a total of 4,217 eye images categorised into these groups.

**Table 4. 1** Distribution of data counts

| Disease Class        | Number of Images |
|----------------------|------------------|
| Cataracts            | 1.038            |
| Diabetic retinopathy | 1.098            |
| Glaucoma             | 1.007            |
| Normal               | 1.074            |
| <b>Total</b>         | <b>4.217</b>     |

Source : *Output Google Colab (2026)***Figure 4. 1** Examples of eye disease images

The examples of images in the dataset shown in **Figure 4.1** are intended to illustrate the differences in visual characteristics between classes prior to the pre-processing, feature extraction and classification stages using the K-NN algorithm. Figure (a) shows the original image of an eye with cataracts, Figure (b) shows the original image of an eye with diabetic retinopathy, Figure (c) shows the original image of an eye with glaucoma, and Figure (d) shows the original image of a normal eye.

## 4.2. Data Preprocessing

### 4.2.1. Grayscale Conversion

The process of converting an image to grayscale is carried out to simplify the image representation by transforming a colour (RGB) image into a grayscale image. This step aims to reduce data complexity without losing important textural information. The use of grayscale images also makes the feature extraction process more efficient as it involves only a single pixel intensity channel

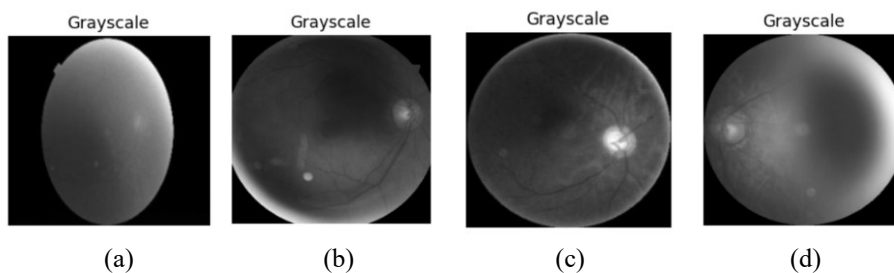
**Figure 4. 2** illustrates the stages of grayscale conversion.

Figure (a) shows a cataract-affected eye that has undergone the grayscale conversion process; Figure (b) shows an eye with diabetic retinopathy that has undergone the grayscale conversion process; Figure (c) shows a glaucoma-affected eye that has undergone the grayscale conversion process; Figure (d) shows a normal eye that has undergone the grayscale conversion process.

#### 4.2.2. Normalisation

The normalisation stage is carried out with the aim of scaling the pixel intensity values of an image to a specific range, namely between 0 and 1. This process aims to reduce the impact of differences in lighting between images and to standardise the data scale so that the distribution of pixel values becomes more consistent. Normalisation is carried out by dividing each pixel intensity value—which originally fell within the range 0–255—by the maximum value of 255. Through this process, distance calculations in the K-NN algorithm become more stable and efficient, thereby improving performance during the feature extraction and classification stages.

#### 4.2.3. Segmentation

Following the normalisation stage comes the image segmentation stage. This stage is carried out with the aim of separating the main object from the image background. In this study, the segmentation process is used to highlight the eye area, which contains important information relating to disease characteristics. Through segmentation, irrelevant areas can be minimised so that the analysis focuses more on the parts of the image that make a significant contribution to the feature extraction process. It is hoped that this will improve the accuracy of the features produced and enhance performance and accuracy during the classification stage using the K-NN algorithm.

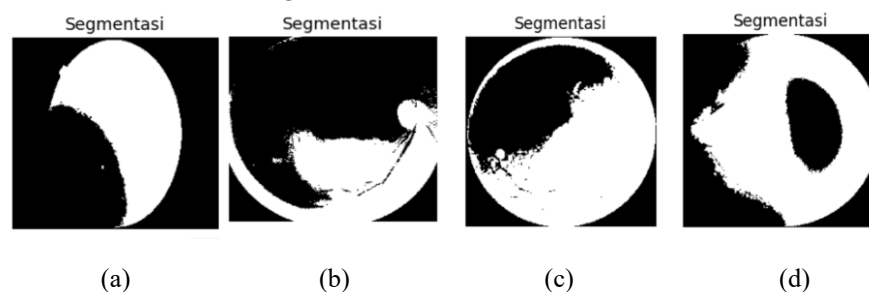


**Figure 4.3** Segmentation stages

**Figure 4.3** illustrates the stages of segmentation. Figure (a) shows an eye with cataracts that has undergone the segmentation process; Figure (b) shows an eye with diabetic retinopathy that has undergone the segmentation process; Figure (c) shows an eye with glaucoma that has undergone the segmentation process; and Figure (d) shows a normal eye that has undergone the segmentation process.

#### 4.2.4. Resizing

The resizing stage involves adjusting the image size to  $224 \times 224$  pixels. This step aims to standardise the dimensions of all images used in the study so that each image has a consistent size. Standardisation of image size is necessary to ensure that feature extraction and distance calculations in the K-NN algorithm can be performed optimally. With images of the same dimensions, differences in the number of pixels between images can be minimised, resulting in more stable and accurate analysis results.

#### 4.2.5. Feature Extraction

The feature extraction stage is carried out to convert the pre-processed images into a set of numerical values representing the key characteristics of the corneal image. In this study, feature extraction aims to capture statistical and textural information relevant to distinguishing between normal eyes and those with abnormalities. Based on the results of the feature extraction, a total of 4,217 image data points were obtained, each represented

by 21 numerical features. These features consist of basic statistical features and texture features extracted using the Grey Level Co-occurrence Matrix (GLCM) method.

### 4.3. Implementation of the K-NN Algorithm

#### 4.3.1. Data splitting

The training and test data were split using several ratios, namely 80:20, 70:30 and 60:40. This data split was carried out to enable the model to learn patterns optimally from the training data, whilst also providing representative test data to evaluate the model's accuracy and reliability against data that had not been used in the training process. The results of the training and test data split are shown in Table 4.2.

**Table 4. 2** Distribution of Training and Testing Data

| <i>Training</i>  | <i>Testing</i>   | <i>Total</i>      |
|------------------|------------------|-------------------|
| 80% (3.373 data) | 20% (844 data)   | 100% (4.217 data) |
| 70% (2.951 data) | 30% (1.266 data) | 100% (4.217 data) |
| 60% (2.530 data) | 40% (1.687 data) | 100% (4.217 data) |

#### 4.3.2. Determining the Value of $k$

The value of  $k$  represents the number of nearest neighbours used as a reference in determining the class of a test data point. In this study, the value of  $k$  was obtained by testing several variations of  $k$  using the training data. To obtain the most optimal value of  $k$ , the model's performance was evaluated using cross-validation, whereby each value of  $k$  was tested and compared based on the resulting accuracy. The value of  $k$  that yielded the highest accuracy was then selected as the optimal value and used in the classification of the test data. This approach was adopted to minimise the likelihood of overfitting and underfitting, ensuring that the resulting K-NN model exhibits more stable and optimal performance. Table 4.8 below shows the value of  $k$  that yielded the highest accuracy.

**Table 4. 3** Value of  $k$

| <i>Training</i> | <i>Nilai <math>k</math></i> | <i>Akurasi</i> |
|-----------------|-----------------------------|----------------|
| 80%             | 7                           | 0,808          |
| 70%             | 8                           | 0,796          |
| 60%             | 3                           | 0,800          |

**Table 4.3** above presents the results of the tests carried out; a split of 80% training data and 20% test data yielded the highest accuracy value, namely 0.808 with  $k = 7$ . With a split of 70% training data and 30% test data, an accuracy value of 0.796 was obtained with  $k = 8$ . Furthermore, with a split of 60% training data and 40% test data, the model achieved an accuracy of 0.800 with  $k = 3$ . Overall, the test results indicate that using a larger proportion of training data tends to yield better classification performance. Based on this, the configuration comprising 80% training data and 20% test data, with  $k = 7$ , was selected as the optimal configuration in this study. This finding confirms that the balance between the amount of training data and the selection of the optimal value of  $k$  plays a crucial role in determining the performance of the K-NN algorithm.

#### 4.3.3. Calculating distances

The classification process in the K-NN algorithm is carried out by calculating the distances between the test data and all the training data. The purpose of these distance calculations is to determine the degree of proximity between data points based on their feature values. The training data with the smallest distance value relative to the test data is then considered the nearest neighbour and used as the basis for determining the class. The distance calculation methods applied are the Manhattan distance and the Euclidean distance. These methods calculate the distance by summing the absolute

differences in the values of each feature between the test data and the training data. Table 4.4 shows a comparison of accuracy for the Manhattan and Euclidean distances.

**Table 4. 4** Comparison of Distance Measures

| Distance  | Value of $k$ | Accuracy |
|-----------|--------------|----------|
| Manhattan | 7            | 0,808    |
| Euclidean | 7            | 0,807    |

The table above shows that the Manhattan distance method yields a higher accuracy than the Euclidean method for the same value of  $k$ .

#### 4.4. Performance Metrics

The following table shows the model's performance metrics by class.

**Table 4. 5** Model performance metrics by class

| Class                | Precision | Recall | F1-Score |
|----------------------|-----------|--------|----------|
| Cataract             | 0,85      | 0,85   | 0,85     |
| Diabetic Retinopathy | 0,96      | 0,89   | 0,92     |
| Glaucoma             | 0,72      | 0,67   | 0,69     |
| Normal               | 0,71      | 0,81   | 0,76     |

Based on the evaluation results, the diabetic retinopathy class demonstrated the best performance with the highest precision and F1-score values, This indicates that the texture features applied are capable of effectively representing the characteristics of this condition. Conversely, the glaucoma class achieved the lowest recall value, indicating that errors still occur in identifying data within this class. This is thought to be due to the similarity in image texture between glaucoma and other classes, particularly the normal class.

Overall, the application of the K-NN model combined with GLCM and LBP texture features produced fairly good classification performance, with an accuracy rate exceeding 80%.

#### 4.5. Discussion

The classification model achieved an accuracy of 81%, indicating that the majority of the test data was successfully classified correctly. This result demonstrates that the features used were able to represent the characteristics of eye images. Similarly, in the study by Prabowo et al. (2021), which utilised the K-NN algorithm and the GLCM feature extraction method, an accuracy of 82% was achieved. This reinforces the finding that the application of the K-NN algorithm using GLCM feature extraction is effective in improving the accuracy of the classification model. In this study, several data splitting schemes were applied, namely 80% training data with 20% test data, 70% training data with 30% test data, and 60% training data with 40% test data. The evaluation results showed that the 80% training data and 20% test data splitting scheme yielded the highest accuracy compared to the other schemes, This because a larger volume of training data enables the K-NN algorithm to obtain more comprehensive information when learning the patterns and feature distributions within each class. Conversely, in schemes with a smaller proportion of training data, model performance tends to decline due to the limited information available during the learning process.

In this study, feature extraction was carried out using the Grey Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP). GLCM features were used to capture global texture information in images based on spatial relationships between pixels, such as contrast, correlation, energy and homogeneity. These features play a key role in

characterising differences in texture patterns in images of eyes affected by health conditions, Meanwhile, LBP is used to extract local texture information by encoding the intensity patterns of pixels surrounding each central pixel. The combination of these two methods enables a more comprehensive representation of image texture, at both the local and global levels.

At the classification stage, the K-NN algorithm is applied with a parameter value of  $k = 7$ , meaning that the class assignment of a test data point is based on its seven nearest neighbours in feature space. The choice of  $k = 7$  is based on test results showing that this value provides the best performance compared to other values of  $k$ , The use of a relatively larger value of  $k$  reduces the influence of noise and outliers, thereby making classification decisions more stable and less reliant on just one or two nearest neighbours.

## 5. Conclusion

This study successfully applied the K-NN algorithm to the classification of eye diseases based on eye images, achieving an accuracy rate of 81% in classifying images into several disease categories. This demonstrates that this method is effective in recognising image patterns, particularly following the pre-processing and feature extraction stages.

The evaluation results show that the K-NN model with  $k = 7$  achieved an accuracy rate of 81%, indicating that the majority of the test data could be classified correctly. The selection of this value for  $k$  was based on test results showing the most optimal performance compared to other values of  $k$ .

## References

- Aziza, L. N., Astuti, R. Y., Maulana, B. A., & Hidayati, N. (2024). Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Ketahanan Pangan di Provinsi Jawa Tengah. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 404–412. <https://doi.org/10.57152/malcom.v4i2.1201>
- Baloch, L., Bazai, S. U., Marjan, S., Aftab, F., Aslam, S., Neo, T. K., & Amphawan, A. (2023). A Review of Big Data Trends and Challenges in Healthcare. In *International Journal of Technology* (Vol. 14, Nomor 6, hal. 1320–1333). Faculty of Engineering, Universitas Indonesia. <https://doi.org/10.14716/ijtech.v14i6.6643>
- Eka Pratiwi, N., Suryadi, L., Ardhy, F., & Riswanto, P. (2022). Penerapan Data Mining Prediksi Penjual Mebel Terlaris Menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus: Toko Zerita Meubel). In *Jurnal Sistem Informasi Musirawas* Lusi Suryadi, Ngajiyano (Vol. 7, Nomor 2).
- Firmando, D., & Suryana, E. (2023). Cataract Disease Diagnostic Expert System Using Forward Chaining Method Sistem Pakar Diagnosa Penyakit Katarak Menggunakan Metode Forward Chaining. In *Jurnal Media Computer Science* (Vol. 2, Nomor 1).
- Gunawan, R., Fathurrahman, R., Widyaningrum, A. I. S., Issandra, F., Abdurachman, M. A., Putra, Y. E., & Naufal. (2025). Pendekatan Transfer Learning untuk Klasifikasi Penyakit Mata Menggunakan Citra dengan CNN InceptionV3. *Jurnal CoSciTech (Computer Science and Information Technology)*, 6(1), 60–67. <https://doi.org/10.37859/coscitech.v6i1.8509>
- Ilmi, N., Novitri, A., & Pratama, H. R. (2024). Detection of Diabetic Retinopathy using Local Binary Pattern Variant and K-Nearest Neighbor. *Journal of Informatics and Communication Technology (JICT)*, 6(1), 168–177. [https://doi.org/10.52661/j\\_ict.v6i1.279](https://doi.org/10.52661/j_ict.v6i1.279)
- Kusuma, D. P. D. (2020). *MACHINE LEARNING*. DEEPUBLISH CV. BUDI UTAMA. <https://repository.deepublish.com/media/publications/593576-machine-learning-teori-program-dan-studi-5c99fd07.pdf>
- Mughnyanti, M., & Subhan, H. N. G. (2023). Data Mining Manhattan Distance dan Euclidean Distance Pada Algoritma X-Means Dalam Klasifikasi Minat dan Bakat Siswa. *remik*, 7(1), 835–842. <https://doi.org/10.33395/remik.v7i1.12162>

- Pamuji, R. I., Sani, D. A., & Hariyanto, R. (2025). Klasifikasi Jenis Penyakit Mata Katarak Menggunakan Metode K-Nearest Neighbors (KNN). *CESS (Journal of Computer Engineering, System and Science)*, 10(2), 690–701. <https://doi.org/10.24114/cess.v10i2.67786>
- Pauji, A., Aisyah, S., Surip, A., Saputra, R., & Ali, I. (2022). Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika dan Komputer*, 4(1), 21–27. <https://doi.org/10.32485/kopertip.v4i1.114>
- Prabowo, A., Erwanto, D., & Rahayu, P. N. (2021). Klasifikasi Kesegaran Daging Sapi Menggunakan Metode Ekstraksi Tekstur GLCM dan KNN Freshness Classification of Beef Using GLCM Texture Extraction Method and KNN. *Jurnal Elektro Lucea*, 7(1).
- Putri, G. R., Maulana, M. A., & Lestari, T. (2023). *Identifikasi Mata Juling dan Mata Normal Pada Anak Dengan Metode Convolutional Neural Network (CNN)* (Vol. 13, Nomor 2). <https://jurnal.umj.ac.id/index.php/just-it/index>
- Putri, I. P. (2021). Analisis Performa Metode K-Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular. *Indonesian Journal of Data and Science (IJODAS)*, 2(1), 21–28.
- Rashad, I., Isnanto, R. R., & Widodo, C. E. (2023). Klasifikasi Penyakit Jantung Menggunakan Algoritma Analisis Diskriminan Linier. *J. Sistem Info. Bisnis*, 13(1), 29–36. <https://doi.org/10.21456/vol13iss1pp29-36>
- Setiawan, M. R. (2019). *Klasifikasi citra makanan menggunakan K-Nearest Neighbor dengan fitur bentuk simple morphological shape descriptors dan fitur warna grayscale histogram* [Universitas Brawijaya]. [https://repository.ub.ac.id/id/eprint/168843/1/Muhammad Rizky Setiawan.pdf](https://repository.ub.ac.id/id/eprint/168843/1/Muhammad%20Rizky%20Setiawan.pdf)
- Subandi, R., Herman, & Yudhana, A. (2023). Pre-Processing Pada Klasifikasi Citra Medis Pneumonia. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(1), 86–93. <https://doi.org/10.51454/decode.v4i1.198>
- Sudipa, I. G. I., Andi Putra, T. E., Jurnaidi Wahidin, A., Alfa Syukrilla, W., Khrisna Wardhani, A., Heryana, N., Indriyani, T., Willyanto Santoso Tutuk Indriyani, L., & Willyanto Santoso, L. (2023). *DATA MINING*. [www.globaleksekutifteknologi.co.id](http://www.globaleksekutifteknologi.co.id)
- Sugriyono, S., & Siregar, M. U. (2020). Preprocessing kNN algorithm classification using K-means and distance matrix with students' academic performance dataset. *Jurnal Teknologi dan Sistem Komputer*, 8(4). <https://doi.org/10.14710/jtsiskom.2020.13874>
- Suwanda, A. E., & Juniati, D. (2022). Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Retina Menggunakan Dimensi Fraktal Box Counting Dan Fuzzy K-Means. *Jurnal Penelitian Matematika dan Pendidikan Matematika*, 10.
- Wang, E. Y., Kong, X., Wolle, M., Gasquet, N., Ssekasanvu, J., Mariotti, S. P., Bourne, R., Taylor, H., Resnikoff, S., & West, S. (2023). Global Trends in Blindness and Vision Impairment Resulting from Corneal Opacity 1984–2020. *Ophthalmology*, 130(8), 863–871. <https://doi.org/10.1016/j.ophtha.2023.03.012>
- Yuliarni, T. (2024). *Deteksi penyakit mata menggunakan Algoritma Reinforcement Learning* [Universitas Negeri makassar]. [https://lib.unm.ac.id/storage/file\\_thesis/obF4ug7avU32mB5CmpZBZ1Ouit6r2qx9gHwvshVs.pdf](https://lib.unm.ac.id/storage/file_thesis/obF4ug7avU32mB5CmpZBZ1Ouit6r2qx9gHwvshVs.pdf)
- Ujjianto, N. T., Gunawan, Fadillah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. (2025). Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis, *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 4(1), 33–43. <https://doi.org/10.24246/itexplore.v4i1.2025,pp33-43>
- Pertiwi, A. A., & Utomo, A. N. (2023). Implementation of Support Vector Machine (Svm) Method on Fundus Image of Retina of the Eye for the Detection of Glaucoma, *Jurnal Rekayasa Informasi*, 12(1), 19–27,