JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 5 No. 1 (2024)
https://doi.org/10.35877/454RI.jinav2778

Classification of Types of Dental Disease Using Principal Component Analysis (PCA) and K-Nearest Neighbor (K-NN) Methods

Nia Zanah* & Sriani

Departement of Computer Science, Faculty of Science and Tecnology, Universitas Islam Negeri Sumatera Utara, Medan, 20236, Indonesia

Abstract

Teeth are one of the organs of the human body which are quite sensitive, the many types of diseases that often occur in teeth certainly make it difficult to identify the type of disease in the teeth. This study uses 4 types of dental disease which will be classified using the Principal Component Analysis and K-Nearest Neighbor methods. For each dental disease, 10 image data were taken, with a total of 24 training data and 16 test data, a total of 40 images. The feature extraction process in this research uses RGB, HSV, and Area characteristics, for the training and testing process uses the PCA algorithm and classification uses KNN. By testing using K=1, it produces an accuracy value of 87.5% in the process of classifying types of dental disease.

Keywords: Classification, Dental Disease, Feature Extraction, Principal Component Analysis, K-Nearest Neighbor.

Received: 12 April 2024 Revised: 5 July 2024 Accepted: 20 July 2024

1. Introduction

The very rapid progress of knowledge and technology has meant that several techniques have been developed to facilitate human tasks, such as image processing, image analysis, and image utilization for various purposes and specific purposes. One of them is a variety of dental radiographic images, which can be used to identify diseases, injuries and conditions of teeth and bones that cannot be seen directly through clinical examination (Akhadi, 2020). Classification of dental disease in image processing involves the use of image analysis techniques to recognize and group types of dental disease based on the visual characteristics of dental images.

Dental disease is a situation when there is an uncomfortable sensation around the oral cavity and teeth (Nurlia, Jajuli, & Purnamasari, 2021). Indonesian people often do not pay attention to dental health, resulting in dental disease (Nas, 2019). Dental disease not only causes pain and discomfort, but can also affect various other health conditions. This dental disease is caused by dental calculus, dental caries, gingivitis, and other dental diseases.

The problem of dental disease in society today is caused by limited knowledge about the types of dental disease and how they form as well as low public awareness about the importance of caring for teeth, causing some people to ignore efforts to prevent or treat teeth. Several factors that often occur in this problem are caused by bacteria, food, saliva and others. As a result, this causes some people to experience toothache, resulting in damage to the layers of the teeth.

Adapted to this phenomenon, this research aims to develop a desktop-based application for classifying types of dental disease in order to increase public knowledge of the importance of caring for and maintaining dental health which can improve the quality of life of individuals in society at large. The classification of types of dental disease is used to help the public know the types and forms of dental disease, so that ultimately the public can understand what type of dental disease they have experienced.

Classification itself is a certain class which then carries out a process to determine whether the object belongs to a type for that class (Istiqhfarani, Cholissodin, & Bachtiar, 2020). Classification is also often done manually, but the results obtained are less accurate and inconsistent because there are human errors which can cause negligence, so a classification system is needed so that the results obtained are very accurate. Dental disease recognition is a process

* Corresponding author.
E-mail address: niazanah25@gmail.com

where technology is used to identify and classify various types of dental disease based on data and information generated through various methods.

The methods used in this research are the Principal Component Analysis (PCA) and K-Nearest Neighbor (K-NN) methods. PCA is a statistical method that has been used in several fields such as pattern recognition, remote sensing, image compression and so on (Unihehu & Suharjo, 2021). PCA is used in the feature extraction process, namely to obtain characteristics from images of dental disease. The characteristics that have been obtained will be used to see the differences between one object and another (Suryaman, Magdalena, & Sa'idah, 2021).

A method called K-NN is used to classify images of dental disease. K-NN is a case-based learning algorithm to be effective in handling several problems (Krismawan & Rachmawanto, 2022). The K-Nearest Neighbor classification process is based on the principle that objects that are similar or close in feature space will tend to have the same or similar labels.

By combining the PCA and K-NN methods, you can produce a model that is more efficient, more precise and easier to understand. The KNN method is able to classify images well, but its accuracy decreases for similar objects. Therefore, it is necessary to improve and add classification methods based on object characteristics that can reduce and maintain important information from the original characteristics in accordance with relevant criteria (Aristo Jansen Sinlae, Alamsyah, Suhery, & Fatmayati, 2022).

The stages of this research are to extract the results of several types of dental disease, then collect and store them in a database. Obtaining image extraction is a PCA algorithm that uses characteristics from RGB, HSV, and Area for the training and testing process which will later be grouped using the K-NN method. The image of dental disease will show similarities in objects and classify the distance according to the distance of the closest object to the Euclidean distance. The language that will be used in this research is the MATLAB programming language which is specifically designed for technical computing, visualization and programming purposes, in the form of mathematical computing, data analysis, algorithm development, simulation, modeling and calculation graphics, which is ultimately suitable for the research that will be carried out (Andrean Nugraha, Wahyu Hidayat, & Nur Shofa, 2023). The result is a desktop-based system that can be used to identify types of dental disease using images of the dental disease.

2. Research Methods

This research uses quantitative research methodology, which is a type of research obtained through the application of mathematical or computational statistical procedures. Quantitative research has a structure that is quite organized, planned, structured from the beginning to the end of the research, and is not influenced by conditions in the field (Priadana & Sunarsih, 2021). The process carried out in this research is shown in the Figure 1.

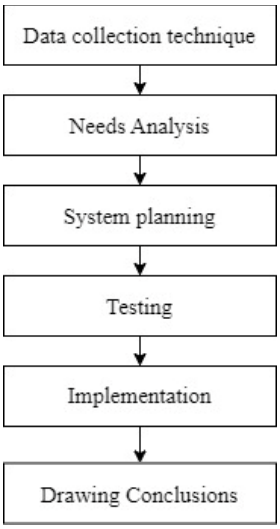

Figure 1. Research Stages

2.1. Data Collection Techniques

The data collection technique for this research is Literary Study. Literature Study involves the process of reading and first finding relevant reading sources that relate to the topic being studied. Literature studies are carried out by accessing accredited journals, library books, theses and previous research. The aim is to obtain ideas that can help in developing a conceptual framework for research methods according to the literature review. The data used in this research is a dataset obtained by accessing the website https://www.kaggle.com, namely in the form of dental disease.

2.2. Needs Analysis

Analysis of the needs for classification of types of dental disease is the process of inputting images of dental disease in Joint Photographic Expert Group (JPEG) format and carrying out stages of character extraction of dental disease images and classification of dental disease images.

2.3. System Planning

System Planning is the initial stage in the system development process. Design is the stage of implementing various techniques or principles for the purpose of determining an object.

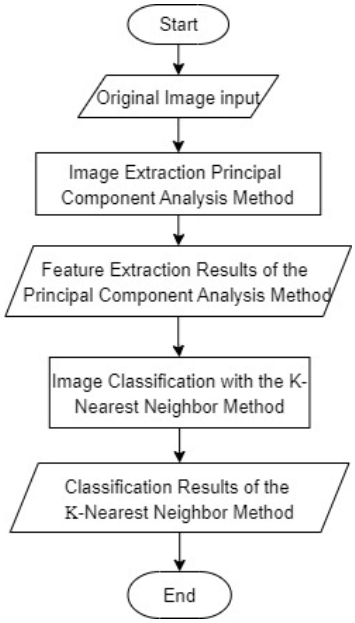

Figure 1. Flowchart System

a. Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) classification method is a complex simplification process so that the data grouping process becomes relatively easy to do. In the computing process, the data centralization stage is carried out which changes the spectral data by subtracting the value of each pixel from the average spectral value of the entire image (Rusydi & Masitoh, 2023).

b. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN), the evaluation process is carried out by finding the shortest distance between the test data and the K nearest neighbors in the training data (Solehatin & Anam, 2020). Prediction results are obtained by selecting the value that appears most frequently from the nearest neighbors. The KNN algorithm aims to categorize new objects by considering their attributes and existing training samples. In this research, the k value used is K-1, and the distance between training data and test data is calculated using Euclidean Distance. Euclidean distance can be formulated as follows:

$$d(xy) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

2.4. Testing

This research requires testing to learn how the system that has been built performs. This research aims to assess the extent of data distribution and in what direction these data groups are distributed. The data folder in image recognition is divided into 2 categories, namely training data and test data.

2.5. Implementation

By using the PCA method to reduce feature dimensions and KNN for classification, this application is an approach that can be used to classify dental diseases with accuracy that can help improve dental care.

2.6. Drawing conclusions

After extraction using the pca method and classification using KNN, the next stage is to draw conclusions. This research will produce a percentage of accuracy for the classification of dental disease images.

3. Result and Discussion

3.1. Data Analysis

The data used for analysis purposes in this research are samples of tooth images measuring 3x3 pixels in the classification of types of dental disease based on tooth images extracted using the PCA method. The results of the PCA value extraction are then classified using the KNN method. The first step is to determine the dataset as training data which is extracted using the PCA method. The dataset used is 4 dental images which are classified into the types of dental disease calculus, caries, healthy and gingivitis which are used as training data variables with a size of 3x3 pixels. The dataset of the form of dental images shown on Figure 3.

Figure 2. Calculus Dental Disease, Caries, Health and Gingivitis

3.2. System Planning

Before the implementation stage of an application program, it begins with designing a classification system for types of dental disease adapted to the shape of the tooth posture so that the system can process as it should. There is a system setup starting from designing the main menu, which includes a classification menu.

3.3. System Flowchart

System flow diagrams are used to show the process flow of the system to be built. The entire flowchart of the system that will be built shown on Figure 4.

Based on the main menu flowchart image (Figure 4), it can be explained that the user starts by opening the application and uploading an image. If the image is uploaded, the application continues with segmentation to convert the image to grayscale and binary. Next, the application performs feature extraction to extract important features from the image. After the features are extracted, the application classifies the image based on the features that have been extracted. Finally, the application displays the classification results to the user, and the process is complete.

3.4. Testing

The application system testing process that has been prepared uses the Matlab program. Based on the process in testing this program, it starts with preparing training data and test data, namely dental images consisting of 24 dental images as training image data and 16 dental disease images as test images.

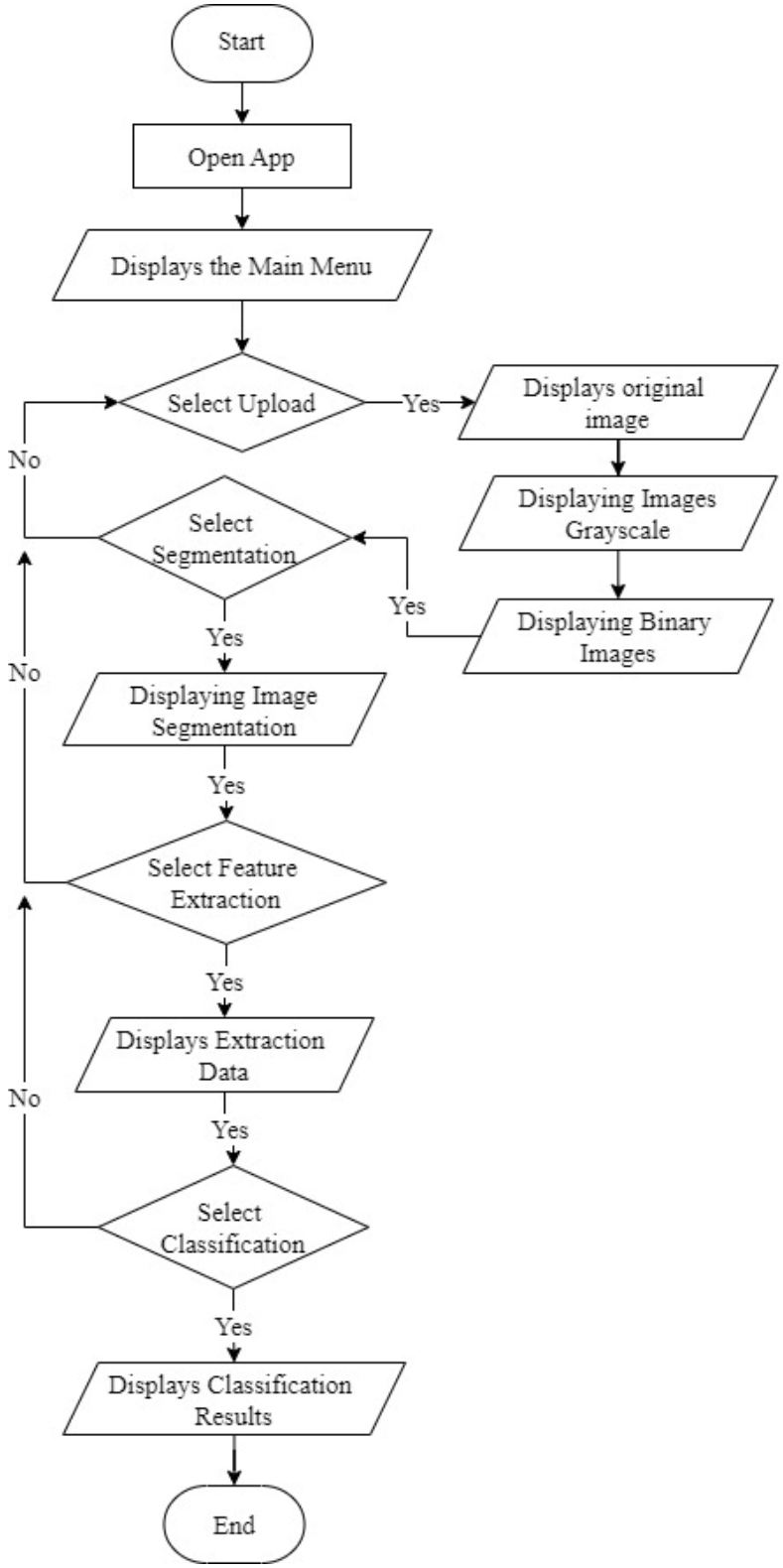

Figure 3. Main Menu Flowchart

In accordance with the tooth image information table (Table 1), the next step is to prepare training tooth image data and test tooth images which are placed in the "Read Data" folder and the "Test Image" folder as shown on Figure 5 and 6.

Table 1. Test Data and Training Data Information

Image Data	Amount	Type of Dental Diseases	Amount
Training Image	24 Image	Dental Calculus	6
		Dental Caries	6
		Healthy Teeth	6
		Gingivitis	6
Test Image	16 Image	Dental Calculus	4
		Dental Caries	4
		Healthy Teeth	4
		Gingivitis	4

Figure 4. Training Dental Image

Figure 5. Testing Dental Image

Next, classify the types of dental disease according to the image of the test tooth that has been prepared in Figure 6. The stages in the system shown on Figure 7.

Figure 6. Classification Process Menu

Based on this Figure 7, in the initial step of classifying types of dental disease according to the shape of the tooth image, the user can press the "Upload" button so that the Figure 8 search menu is displayed.

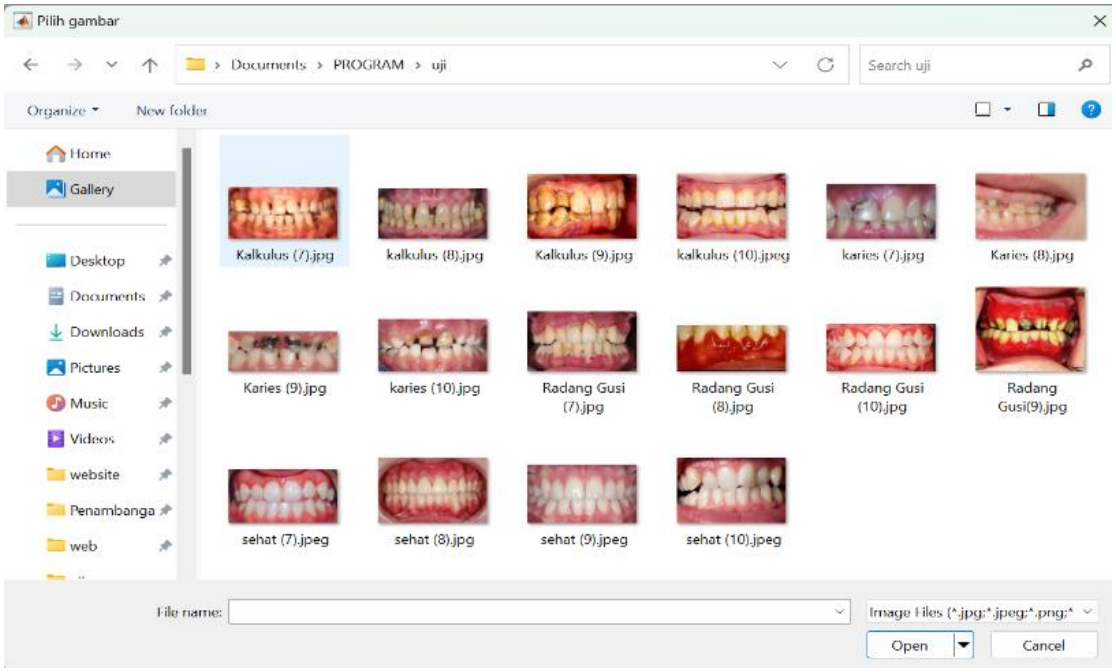

Figure 7. Image Search Menu

According to the image, the user determines the test image to be classified, then in this stage the user points to the image "Kalkulus (7).jpg" or the image of dental calculus, then selects "open". The results were obtained on Figure 9.

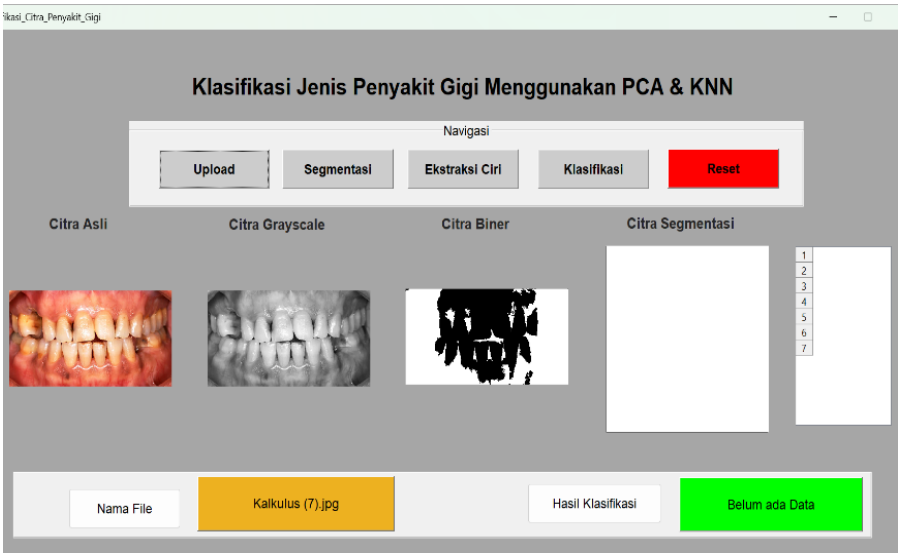

Figure 8. Image of Selected Teeth

According to the image, after selecting the image of dental disease to be classified, the system can read the image name information. Next, when the image classification process begins, it can be carried out by pressing the "Segmentasi" and "Ektraksi" feature buttons first in order to obtain a "Klasifikasi" of the type of dental disease as shown on Figure 10.

Figure 9. Image of Calculus 7 Classified as Dental Calculus

Based on this image, the classification stage was successfully implemented in the application system, when the calculus test image was successfully classified as "Gigi Kalkulus" type, which ultimately resulted in the system classification being correct. For all results, you can see the system test results. By obtaining a classification test with 16 test images consisting of 4 types of dental disease, the results were obtained presented on Table 2.

Table 2. Test Result

No	Nama Image	Type Dental Diseases	Classification Diseases	Information
1.	kalkulus (7).jpg	Dental Calculus	Dental Calculus	Correct
2.	kalkulus (8).jpg	Dental Calculus	Dental Calculus	Correct
3.	kalkulus (9).jpg	Dental Calculus	Dental Calculus	Correct
4.	kalkulus (10).jpg	Dental Calculus	Dental Calculus	Correct
5.	karies (7).jpg	Dental Caries	Dental Caries	Correct
6.	karies (8).jpg	Dental Caries	Dental Caries	Correct
7.	karies (9).jpg	Dental Caries	Dental Caries	Wrong
8.	karies (10).jpg	Dental Caries	Dental Caries	Correct
9.	sehat (7).jpeg	Healthy Teeth	Healthy Teeth	Correct
10.	sehat (8).jpg	Healthy Teeth	Healthy Teeth	Correct
11.	sehat (9).jpeg	Healthy Teeth	Healthy Teeth	Correct
12.	sehat (10).jpeg	Healthy Teeth	Healthy Teeth	Correct
13.	Radang Gusi (7).jpg	Gingivitis	Gingivitis	Correct
14.	Radang Gusi (8).jpg	Gingivitis	Gingivitis	Correct
15.	Radang Gusi (9).jpg	Gingivitis	Gingivitis	Correct
16.	Radang Gusi (10).jpg	Gingivitis	Gingivitis	Wrong

Based on the test results of the Table 2, 14 test images were classified correctly and 4 test images were classified incorrectly. So the next step is to calculate the accuracy level according to the test image used. The following formula is:

$$Accuracy = \frac{Number\ of\ Correct\ Classifications}{Amount\ of\ Data}\ x\ 100\%$$

$$Accuracy = \frac{14}{16}\ x\ 100\% = 87{,}5\%.$$

According to the results of the accuracy test, an accuracy value of 87.5% was obtained in the classification stage of types of dental disease based on images of 16 types of tooth data.

3.5. Implementation

The application/use of the system is to classify types of dental disease based on dental images. Teeth are classified according to their pattern and shape. By using digital image processing, it is possible to identify images similar to human vision. The application of this system can be used by pupils and students who want to learn the classification of types of dental disease according to dental images. This system can facilitate future researchers who want to conduct research on classifying dental diseases, finally the differences and processes can be known more clearly.

4. Conclusion

This research uses the Principal Component Analysis (PCA) method to reduce feature dimensions and K-Nearest Neighbor (KNN) to classify types of dental disease based on dental images. The KNN method can carry out classification by finding the closest value of the transformed test image to the stored training image. From the test results, the system built was able to classify dental disease with K=1, producing an accuracy level of 87.5%, with the accuracy value categorized as Good. Obtaining test results, classification accuracy can definitely change with more data tested.

References

Akhadi, M. (2020). *Sinar-X Menjawab Masalah Kesehatan*. Yogyakarta: Deepublish.

Andrean Nugraha, R., Wahyu Hidayat, E., & Nur Shofa, R. (2023). Klasifikasi Jenis Buah Jambu Biji Menggunakan Algoritma Principal Component Analysis dan K-Nearest Neighbor. *Generation Journal*, *7*(1), 1–7.

Aristo Jansen Sinlae, A., Alamsyah, D., Suhery, L., & Fatmayati, F. (2022). Classification of Broadleaf Weeds Using a Combination of K-Nearest Neighbor (KNN) and Principal Component Analysis (PCA). *Sinkron*, *7*(1), 93–100.

Istiqhfarani, W. A., Cholissodin, I., & Bachtiar, F. A. (2020). Klasifikasi Penyakit Dental caries menggunakan Algoritme Modified K- Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *4*(5), 1499–1506. Retrieved from https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7265/3498

Krismawan, A. D., & Rachmawanto, E. H. (2022). Principal Component Analysis (PCA) dan K-Nearest Neighbor (KNN) dalam Deteksi Masker pada Wajah. *Prosiding Sains Nasional dan Teknologi*, *12*(1), 382.

Nas, C. (2019). Sistem Pakar Diagnosa Penyakit Gigi Dan Mulut Menggunakan Metode Case-Based Reasoning. *Jurnal Digit*, *9*(2), 202.

Nurlia, E., Jajuli, M., & Purnamasari, I. (2021). Penerapan Naïve Bayes Untuk Klasifikasi Tingkat Risiko Diagnosis Gigi Di Uptd Puskesmas Cingambul. *JIKO (Jurnal Informatika dan Komputer)*, *4*(2), 127–132.

Priadana, S., & Sunarsih, D. (2021). *Metode Penelitian Kuantitatif*. Tangerang Selatan: Pascal Books. Retrieved from https://www.google.co.id/books/edition/METODE_PENELITIAN_KUANTITATIF/9dZWEAAAQBAJ?

Rusydi, A. N., & Masitoh, F. (2023). *Teknologi Pengindraan Jauh Untuk Pengelolaan Lingkungan Perairan*. Malang: UB Press. Retrieved from https://play.google.com/books/reader?id=0P7tEAAAQBAJ&pg=GBS.PR17&hl=id

Solehatin, & Anam, C. (2020). *E-Deteksi Kematangan Buah Jeruk Banyuwangi Menggunakan Metode KNN Berbasis Android*. Yogyakarta: Deepublish. Retrieved from https://www.google.co.id/books/edition/E_Deteksi_Kematangan_Buah_Jeruk_Banyuwan/yuYREAAAQBAJ?

Suryaman, S. A., Magdalena, R., & Sa'idah, S. (2021). Klasifikasi Cuaca Menggunakan Metode VGG-16, Principal Component Analysis Dan K-Nearest Neighbor. *Jurnal Ilmu Komputer dan Informatika*, *1*(1), 1–8.

Unihehu, A. L., & Suharjo, I. (2021). The Klasifikasi Jenis Ikan Berbasis Jaringan Saraf Tiruan Menggunakan Algoritma Principal Component Analysis (PCA). *Jurnal Ilmiah Ilmu Komputer*, *7*(2), 27–32.