IJMECIE International Journal of Mechanical, Electrical and Civil Engineering
E-ISSN: 3047-4531
P-ISSN: 3047-4523

Research Article

Chili Pepper Variety Detection System Using the Principal Component Analysis Method

Veri Arinal [1], Frencis Matheos Sarimole [2], Sugeng [3], Rindy Julianda [4]

[1] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia; email: veriarinal@stikomcki.ac.id
[2] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia; email: frencis@stikomcki.ac.id
[3] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia; email: sugeng@stikomcki.ac.id
[4] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia; email: rindyjulianda01@gmail.com

* Corresponding Author: veriarinal@stikomcki.ac.id

Abstract: In the agricultural sector, the automatic identification of chili pepper varieties is crucial for improving production efficiency and quality. This study developed a chili pepper variety detection system based on characteristics using the Principal Component Analysis (PCA) method. The PCA method was used to reduce the dimensionality of chili pepper image data, thereby facilitating the classification process while retaining the key features necessary for chili pepper variety identification. The recognition system for chili pepper identification involves inputting chili pepper image data into a computer. The computer then interprets and identifies the chili pepper variety, and the test data utilizes a dataset of chili pepper images from various varieties. The research results indicate that the proposed system achieves a high level of accuracy in detecting and classifying chili pepper varieties. Consequently, this system can assist farmers and agricultural industry stakeholders in the chili pepper sorting and selection process, thereby improving operational efficiency and the quality of the harvest.

Keywords: Agricultural; Chili Image; Detection System; PCA Method; Type of Chili.

Received: 15 November 2023
Revised: 25 December 2023
Accepted: 20 January 2024
Published: 31 January 2024
Curr. Ver.: 31 January 2024

1. Introduction

In the trade and culinary industries, identifying chili pepper varieties is an important aspect that affects various processes, ranging from production to consumption. Each type of chili pepper has distinct colors and shapes. These variations have a direct impact on their use in different recipes and culinary applications. Therefore, the development of an automatic detection system for identifying chili pepper varieties has great potential to provide benefits to society. The process of identifying chili pepper varieties can be carried out more efficiently and accurately, supporting better decision-making at various stages of production and distribution [1].

Chili pepper data are highly complex because they involve various factors such as color, size, and shape. This variability creates unique challenges in the analysis and classification process. For example, in an image of a chili pepper, its color and shape may vary significantly depending on lighting conditions, image capture angles, and other factors. Therefore, to develop an accurate chili pepper variety detection system, careful data analysis is required to identify the most relevant and reliable features.

One of the main challenges in chili pepper data analysis is high dimensionality. With numerous features to consider, such as color and shape, the classification process becomes complex and time-consuming. In addition, processing high-dimensional data requires substantial computational resources. Dimensionality reduction is essential to address this issue because it allows important information to be retained in a lower-dimensional representation, which in turn simplifies the classification process and reduces computational burden. Through the utilization of machine learning technology, detection systems can be developed to operate automatically and efficiently [2].

Principal Component Analysis (PCA) is one of the most commonly used methods for dimensionality reduction. PCA works by identifying linear combinations of existing features that explain the greatest variation within a dataset. By applying PCA, the dimensionality of data can be reduced without losing most of the relevant information. Therefore, PCA is an attractive choice for the development of chili pepper variety detection systems, as it addresses dimensionality issues and improves classification efficiency. PCA, or Principal Component Analysis, is a statistical method used to reduce data dimensionality while preserving most of the relevant information [3].

Several previous studies have applied PCA in chili pepper analysis; however, most of them have focused on other aspects, such as image segmentation or quality analysis. Most studies have concentrated on assessing chili pepper quality, while relatively few have specifically utilized PCA for the recognition and classification of different chili pepper varieties.

This study aims to develop an efficient and accurate chili pepper variety detection system using PCA as an integral part of the analysis process. By combining image processing technology and data analysis, the proposed system is expected to make a positive contribution to the agricultural sector, culinary industry, and research on chili pepper varieties. With the implementation of this system, the identification process of chili pepper varieties is expected to become faster, more accurate, and more efficient, thereby helping to improve productivity and quality throughout the chili pepper supply chain.

2. Literature Review

Detection System

A Detection System is a method used to examine a sample or data by utilizing predefined criteria or standards in order to obtain detection results that can be used as knowledge. A detection system refers to a system that can identify and detect different types of chili peppers based on the characteristics observed in chili pepper images [23]. The detection system process is shown in Figure 1 below.

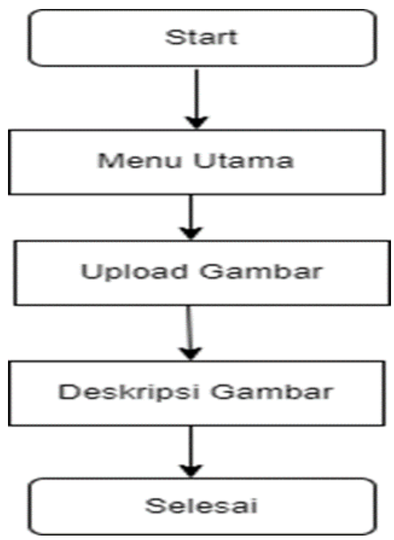


Figure 1. Detection System

Image and Image Processing

Literally, an image is a picture represented on a two-dimensional plane. From a mathematical perspective, an image is a continuous function of light intensity on a two-dimensional plane. A light source illuminates an object, and the object reflects a portion of the light rays. This reflected light is captured by optical devices such as the human eye, cameras, scanners, and other imaging devices, allowing the image of the object to be recorded.

Images can be classified into two categories: still images and moving images. A still image is a single image that does not move, whereas a moving image is a sequence of still images displayed consecutively (sequentially), creating the perception of motion to the human eye. Each image within the sequence is called a frame. The images displayed in motion pictures or on television screens essentially consist of hundreds to thousands of frames [24]. The image processing procedure can be seen in Figure 2 below.

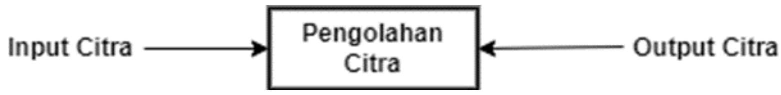


Figure 2. Image Processing

A digital image is a two-dimensional function, f(x,y), that represents light intensity, where x and y are spatial coordinates and the function value at each point (x,y) represents the gray level of the image at that point. A digital image is represented by a matrix in which the rows and columns indicate points within the image, and the matrix elements (called picture elements or pixels) represent the gray level at those points.

Digital Image Processing

According to the Kamus Besar Bahasa Indonesia (KBBI), processing refers to a method or process of transforming something into a different or more refined form. Meanwhile, an image, according to KBBI, refers to a picture or visual representation, in this context, an image obtained through a visual system. Overall, image processing refers to a method of transforming an image into another image that is more refined or desired. In other words, image processing is a process that takes an image as input and produces an image as output according to the intended objective.

According to the Webster Dictionary, an image is a representation, resemblance, or imitation of an object. For example, a photograph of an apple represents the identity of the apple in front of a camera. An image may take the form of a photograph, painting, illustration, or sketches appearing on paper, canvas, or a computer monitor screen. An image can also be described as a distribution of variations in darkness and brightness, dimness and intensity, and/or colors across a flat surface. Formally, it can be expressed numerically using values that represent variations in brightness intensity and/or color in both horizontal and vertical directions [6].

In another sense, digital image processing is the processing of two-dimensional images using computer-based media. A digital image consists of an array containing real values represented by a sequence of bits.

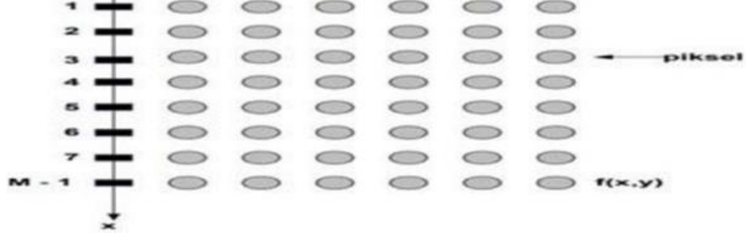


Figure 3. Coordinates in a digital image

Feature Extraction

Feature extraction is used to determine the feature values contained within a digital image that represent specific characteristics of the image. The values obtained from the feature extraction process are then processed for identification purposes. In feature extraction, various types of features can be extracted to obtain information about the characteristics of an image object so that it can be distinguished and recognized. The feature extraction process includes shape feature extraction, size feature extraction, geometric feature extraction, texture feature extraction, and color feature extraction.

In this study, the features used are color and shape features. Color feature extraction involves converting RGB (Red, Green, Blue) color values into HSV (Hue, Saturation, Value) color values to obtain specific values and characteristics of the object's color. The Red, Green, Blue (RGB) color space is a standard color space based on the acquisition of color frequencies from electronic sensors. The output of these sensors is in the form of analog signals. RGB is an additive color space, meaning that all colors begin with black and are formed by adding red, green, and blue colors. The combination of red, green, and blue creates new colors.

The Hue, Saturation, Value (HSV) color space, also known as the hexcone model, is one of the many color systems used for color selection. This color model is more commonly used than the RGB model because it more closely resembles how the human eye perceives and interprets colors. HSV (Hue, Saturation, Value) has the following main characteristics:

a. Hue represents the actual color, such as red, violet, and yellow, and is used to determine redness, greenness, and other color attributes.
b. Saturation, sometimes referred to as chroma, represents the intensity level of a color.
c. Value represents the brightness of a color. Its value ranges from 0% to 100%.
d. If the value is 0, the color becomes black. The higher the value, the brighter the color becomes, resulting in the appearance of new variations of that color.

MATLAB

MATLAB is software used for programming, analysis, and technical and mathematical computations based on matrices. MATLAB stands for Matrix Laboratory because it is capable of solving computational problems in matrix form. The first version of MATLAB was released in 1970 by Cleve Moler. Initially, MATLAB was designed to solve linear algebraic equation problems. Over time, the software has continued to evolve in terms of functionality and computational performance.

The programming language, which is now developed by MathWorks Inc., integrates programming, computation, and visualization within an easy-to-use working environment. MATLAB also offers various advantages, including data analysis and exploration, algorithm development, modeling and simulation, 2D and 3D plot visualization, and graphical user interface (GUI) application development.

In higher education, MATLAB is used as a learning tool for mathematics, engineering, and science programming at both introductory and advanced levels. In industry, MATLAB is widely used as a tool for research, development, and industrial product analysis. MATLAB can operate on Windows, Linux, and macOS operating systems. In addition, MATLAB can be integrated with other applications and external programming languages, such as C, Java, .NET, and Microsoft Excel.

MATLAB also provides toolboxes that can be used for specialized applications, including signal processing, control systems, fuzzy logic, artificial neural networks, optimization, digital image processing, bioinformatics, simulation, and various other technologies [25].

Computer Vision

Computer vision is the process of learning from and analyzing images or videos to obtain results similar to those achieved by humans. In simpler terms, computer vision attempts to imitate the way human visual perception works. Computer vision is closely related to several fields, including image processing and machine vision. There are significant similarities in the techniques and applications that encompass these three fields. This indicates that the fundamental techniques used and developed in these areas are generally similar.

More broadly, computer vision is related to and applied in various other fields such as artificial intelligence (AI), robotics, industrial automation, signal processing, physical optics, neurobiology, and many others. Today, computer vision is widely used for various purposes, such as face detection in images, facial expression recognition, and in practice, it is often combined with artificial neural networks.

Machine Learning

The term Machine Learning was first defined by Arthur Samuel in 1959. According to Arthur Samuel, Machine Learning is a field of computer science that enables computers to learn without being explicitly programmed. There are two main applications of Machine Learning, namely classification and prediction.

Basically, Machine Learning is the process through which computers learn from data. Without data, computers cannot learn anything. All Machine Learning knowledge and applications involve data. Although the data may be the same, different algorithms and approaches can be applied to obtain optimal results.

Chili Pepper

Chili pepper is one of the horticultural commodities that has significant economic and cultural value in Indonesia. In Indonesia, the two most commonly consumed chili species are Capsicum annuum L. and Capsicum frutescens L., which include several well-known varieties such as bell peppers, curly red chili peppers, and large red chili peppers.

Chili peppers have promising prospects in local markets. In addition to their high economic value, chili pepper production in Indonesia is substantial because they are widely needed to meet daily consumption demands. This also demonstrates the important role of chili peppers in various local cuisines.

A chili pepper type refers to a category or variety of chili pepper that possesses distinct visual characteristics, such as color, size, shape, and level of pungency. In this study, chili pepper types serve as the objects to be detected based on their characteristics [26].

Types of Chili Peppers

A chili pepper type refers to a category or variety of chili pepper that possesses distinct visual characteristics, such as color, shape, size, and texture. The following are several types of chili peppers:

Curly Red Chili Pepper

The curly red chili pepper, commonly known as cabai keriting, has a long and slender shape with a length of approximately 10–15 cm. The tip is usually curved. Its color varies from green when unripe to bright red when fully ripe. The surface texture of the curly red chili pepper is generally smooth but may also appear slightly wrinkled. It tends to be thinner than other chili pepper varieties.

Figure 4. Curly Red Chili Pepper

Bird's Eye Chili Pepper

The bird's eye chili pepper, commonly known as cabai rawit or hot chili pepper, is small in size, with a length of approximately 2–5 cm. Its shape varies from round to slightly elongated, with a pointed tip. The color changes as it ripens, starting from green when unripe and turning red when ripe. Some varieties may also appear yellow or orange. The surface texture is usually smooth and glossy. The flesh is thin but relatively firm and dense.

Figure 5. Bird's Eye Chili Pepper

Large Red Chili Pepper

The large red chili pepper, commonly known as cabai merah besar or local red chili pepper, has a long and thick shape that is generally straight with a slight curve. It measures approximately 10–20 cm in length and 2–3 cm in diameter. The chili pepper is bright red in color. Its surface texture is smooth and glossy, while the flesh is thick and slightly crunchy, with seeds located in the central part of the fruit.

Figure 6. Large Red Chili Pepper

Domba Bird's Eye Chili Pepper

The Domba bird's eye chili pepper, commonly known as white Domba chili or green Domba chili, generally has a small and slender shape with a length of approximately 2–5 cm. Its form tends to taper to a pointed tip, similar to other bird's eye chili varieties. The skin is thin and smooth, giving it a glossy appearance. The thin skin causes the chili pepper to wilt quickly after harvesting. Its flesh is also thin and not very watery, providing a sharp and immediate spicy sensation when bitten.

Figure 7. Domba Bird's Eye Chili Pepper

Large Green Chili Pepper

The large green chili pepper, commonly known as cabai hijau besar, has a long and wide shape that is generally straight with a slight curve. It measures approximately 10–20 cm in length and 2–3 cm in diameter. The chili pepper is green in color and has a smooth, glossy surface texture. Its flesh is thick and crunchy, with seeds located in the central part of the fruit.

Figure 8. Large Green Chili Pepper

Key Literature Review

Various previous studies have demonstrated that Principal Component Analysis (PCA) and artificial intelligence-based classification methods have been widely applied to solve problems related to detection, classification, and dimensionality reduction in various fields. A study conducted by Shu T., Wang S., Xu J., et al. (2021) showed that PCA can be utilized to extract positional information from wafer mark images, thereby improving wave contrast and alignment accuracy compared to conventional methods [4]. Furthermore, Duan T., Liao Z., Li T., Tang H., and Chen P. developed the State-Space Principal Component Tracking Filter (SPCTF) method for bearing fault diagnosis under varying speeds and high-noise environments. This method is capable of recovering fault information from noise-contaminated signals, thereby improving the accuracy of bearing fault diagnosis [5].

In the agricultural sector, research conducted by Siti Raysyah, Veri Arinal, and Dadang Iskandar Mulyana applied a combination of PCA and K-Nearest Neighbor (KNN) to classify coffee fruit maturity levels based on color features. The developed system achieved an accuracy of 97.77% and successfully automated a classification process that was previously performed manually [6]. Another study by He C., Li J., Liu W., et al. proposed a low-complexity Quantum Principal Component Analysis (qPCA) algorithm capable of simplifying computational circuits and reducing execution time in quantum principal component analysis processes [7].

In chili plant recognition, Ilyas Perlindungan and Risnawati utilized the Convolutional Neural Network (CNN) method to classify chili pepper types. Their research demonstrated that the application of dropout techniques could reduce the risk of overfitting in limited datasets [8]. Meanwhile, Nazila S.F., Arman Y., Wahyuni D., Nurhasanah, and Putra Y.S. developed an early detection system for pests and diseases in bird's eye chili peppers using image recognition technology and the Support Vector Machine (SVM) algorithm, achieving a classification accuracy of 82% [9].

The application of PCA was also investigated in underwater shrimp digital image processing by Setiawan A., Hadiyanto H., and Widodo C.E. Their study showed that PCA effectively reduced image dimensionality while preserving most of the important information, thereby accelerating computational processes in shrimp growth monitoring [10]. Mohamed A.F.A., Soluma F.W.Z., and Ashour M.M. utilized PCA to enhance the performance of the watershed algorithm in image segmentation. The results indicated that PCA could reduce the effects of noise and reflected light, leading to improved image segmentation quality [11].

In the field of communication networks, Franco J.D.G., Velasco J.E.P., Rizk J.L., et al. combined PCA with several supervised learning algorithms on a 5G/B5G service dataset. The results demonstrated that the Support Vector Machine (SVM) algorithm achieved a precision of 98% and an F1-score of 98.1%, highlighting the effectiveness of PCA in improving data quality before classification [12]. Additionally, Muna N., Ningsih N., and Syahroni N. implemented EfficientDet-D0 and SSD-MobileNet-V2 FPNLite models for a weed detection system. The developed system contributed to more accurate weed control and reduced excessive pesticide usage [13].

Object detection systems have also been developed by Dadang Iskandar Mulyana and M. Ainur Rofik, who created a vehicle classification system using the YOLOv5 method. Using a dataset of 1,332 images, the system achieved an accuracy rate of 90% in recognizing various vehicle types [14]. Isyada R. and Audytra H. developed a web-based early COVID-19 detection system using the SMART method to facilitate faster and more accessible preliminary disease identification [15].

In weather image classification, Suryaman S.A., Magdalena R., and Saidah S. combined VGG-16, PCA, and KNN methods. Their research demonstrated that the system achieved an accuracy of 87.50%, with PCA serving to reduce the dimensionality of extracted features before the classification stage [16]. Zhao W., Zhang D., Li D., Zhang Y., and Ling Q. optimized the Generalized Iterative Closest Point (GICP) algorithm using PCA for edge extraction in point cloud data. The proposed method improved registration speed by 59.04% and registration accuracy by 30.24% compared to traditional methods [17].

Computer vision-based detection systems were also developed by Mudzopar I.M. and Wiharko T. for detecting offside positions in soccer matches. The system successfully identified players, determined offside lines, and generated accurate automatic detections [18]. In the Internet of Things (IoT) domain, Pramuda A.S., Nugraha A.W.W., and Fadli A. designed a human detection system using PIR, RCWL, and infrared sensors to automatically control building lighting, thereby reducing energy waste [19].

In cybersecurity, Maulani I.E., Putra D.R.S., and Komarudin conducted a comparative study of various machine learning algorithms for intelligent intrusion detection systems. Their research provided valuable insights into the effectiveness of machine learning algorithms in enhancing network security [20]. Meanwhile, Lukmana H.H., Husaini M.A., Hoeronis I., and Puspareni L.D. developed a web-based early stunting detection information system using the User-Centered Design (UCD) approach, improving usability and supporting stunting prevention efforts [21].

Another study conducted by Li G., Yang S., Li Sai, Wang N., and Li J. applied a combination of clustering methods and PCA to identify the chemical composition of ancient glass artifacts affected by weathering. The results demonstrated that PCA could assist in classifying artifacts based on their chemical characteristics, thereby supporting historical and archaeological material studies [22].

Based on these previous studies, it can be concluded that PCA is an effective method for dimensionality reduction, feature extraction, and performance enhancement in various classification and detection systems. PCA has been successfully applied across multiple domains, including image processing, agriculture, manufacturing, communication networks, cybersecurity, and scientific data analysis. These findings provide a strong foundation for applying PCA in chili pepper type detection systems, as the method is capable of preserving important information from image data while improving the efficiency and accuracy of the classification process [4]–[22].

3. Materials and Method

Research Data

The research data used in this chili pepper type detection study consist of a collection of chili pepper images gathered from various sources. Each image is accompanied by a label indicating the corresponding chili pepper type, enabling the training and testing of classification models based on their characteristics. In addition, the dataset includes features extracted from each image, such as color values in various color spaces, including RGB, texture features obtained using methods such as Haralick or Gabor to describe the unique texture patterns of each chili pepper, chili pepper size measured in pixels to provide information about physical dimensions, and shape features represented as feature vectors describing contours and forms. With the combination of image data and extracted features, this research has a strong foundation for developing an effective detection model capable of identifying different types of chili peppers with high accuracy.

Research Object

The object of this research is a chili pepper image dataset containing images of various chili pepper types obtained from the website Kaggle. This image dataset includes different varieties of chili peppers, such as bird's eye chili peppers, curly chili peppers, and green bird's eye chili peppers, with varying resolutions and image qualities. This research focuses on the development of an automatic detection system to identify and classify these chili pepper types using the Principal Component Analysis (PCA) method.

The research process includes data preprocessing to improve image quality, image segmentation, feature extraction, and the training and evaluation of a PCA-based model to achieve optimal detection accuracy.

Methodology Implementation

The methodology applied in chili pepper type detection uses the PCA method in conjunction with Computer Vision applications. The collected dataset is processed using Computer Vision techniques capable of analyzing images through template-matching approaches.

Several stages are involved in the process of identifying chili pepper types using the Principal Component Analysis (PCA) method:

Image Collection

Collect images of various chili pepper types. Each image should be captured under consistent lighting conditions and camera angles to ensure that the acquired data are representative.

Data Preprocessing

The training and testing datasets are processed to normalize color, enhance contrast, and remove noise. This step is important to ensure good data quality before proceeding to the next stage.

Feature Extraction

Extract features from the images, such as color, texture, and shape. These features may include color histograms, texture descriptors, or shape measurements. The extracted features are then used as inputs for model development based on their characteristics.

Data Transformation

Project the original data into a new space formed by the selected principal components. This process reduces data dimensionality while preserving the most significant information.

Classification Results Analysis

The performance of various models is analyzed based on their characteristics to identify the most effective model for classifying chili pepper types. In addition, the analysis evaluates the contribution of PCA in improving the performance of the detection system.

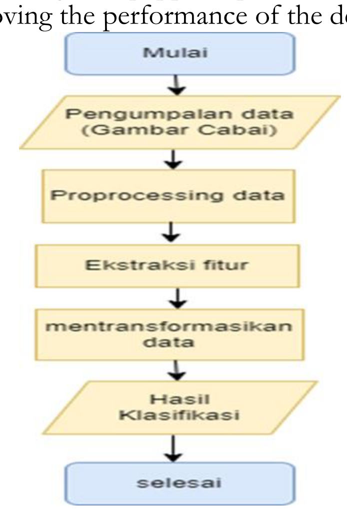


Figure 9. Application of the Methodology

Testing Design

The testing design for the chili pepper type detection system using the Principal Component Analysis (PCA) method involves a series of systematic steps to evaluate the performance of the detection system. The testing procedure is as follows:

Data Collection

The collected chili pepper image dataset is divided into two subsets: training data and testing data. This division is performed randomly to ensure a balanced representation of each chili pepper type in both datasets.

Data Preprocessing

The training and testing datasets are processed to normalize color, improve contrast, and remove noise. This step is important to ensure high-quality data before proceeding to subsequent stages.

Training and Testing

To detect chili pepper types based on images, the model is provided with a set of labeled chili pepper images. The model uses this information to learn the visual characteristics of each chili pepper type. After the model has been trained, it is tested using new chili pepper images that are not included in the training dataset to evaluate its detection capability.

Testing Results Analysis

The analysis of testing results helps in understanding the strengths and weaknesses of the model in classifying chili pepper types. By evaluating these performance metrics, better decisions can be made regarding how the model can be optimized or applied in real-world applications.

Figure 10. Test Design

4. Results and Discussion

Testing Results

The image identification system was tested by running the main program in the system, namely main_program.m. The display result obtained from running main_program.m is shown in.

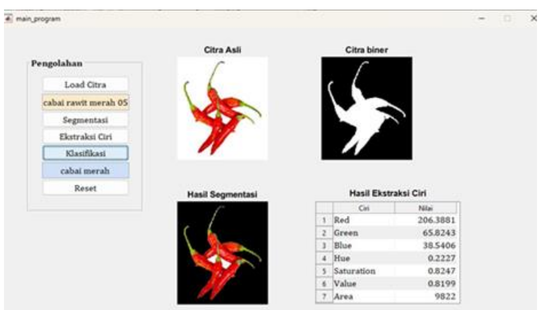
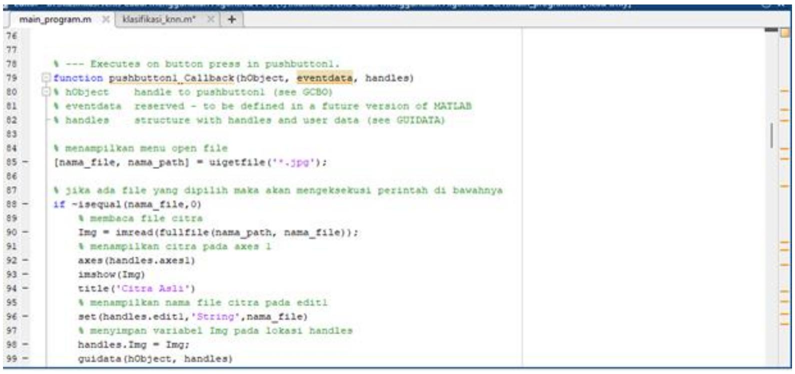


Figure 11. Testing Results

The following is the code for the Input Image button.



```
% --- Executes on button press in pushbutton1.
function pushbutton1_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton1 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% menampilkan menu open file
[nama_file, nama_path] = uigetfile('*.jpg');

% jika ada file yang dipilih maka akan mengeksekusi perintah di bawahnya
if ~isequal(nama_file,0)
    % membaca file citra
    Img = imread(fullfile(nama_path, nama_file));
    % menampilkan citra pada axes 1
    axes(handles.axes1)
    imshow(Img)
    title('Citra Asli')
    % menampilkan nama file citra pada edit1
    set(handles.edit1,'String',nama_file)
    % menyimpan variabel Img pada lokasi handles
    handles.Img = Img;
    guidata(hObject, handles)
```

Figure 12. Input Image Code

The classification process assigns the input pattern to a class based on more than one feature. In the distance measurement process, four distance metric methods are calculated, namely Sorensen, Lorentzian, Soergel, and Gower distance metrics. Distance measurement is performed by calculating the distance between the training images and the testing images. A summary of the chili pepper image classification system testing results is presented in the table below.

Table 1. Values of Segmented Chili Peppers

Characteristic	Value
RED	206.3881
BLUE	65.8243
HUE	38.5406
SATURATION	0.2227
VALUE	0.8199
AREA	9822

Data Distribution Visualization

The visualization of the training and testing data distribution using PCA components displays points that represent both the training data and the testing data. The following figure shows the distribution of data on the PCA components.

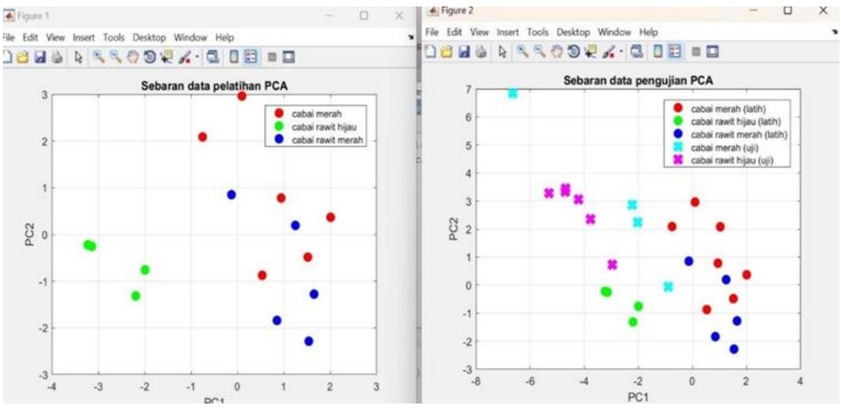


Figure 13. Distribution of Data on PCA Components for Training and Testing

Results:
Red Points: Training Red Chili Peppers
Blue Points: Training Red Bird's Eye Chili Peppers
Green Points: Training Green Bird's Eye Chili Peppers
Purple X Marks: Testing Green Bird's Eye Chili Peppers
Light Blue Points: Testing Red Chili Peppers

Feature Extraction Values

For chili peppers, the Mean, Variance, and Range values of the RGB and HSI color components obtained from the feature extraction process are used by the KNN model to determine the classification.

Values for Red Chili Peppers

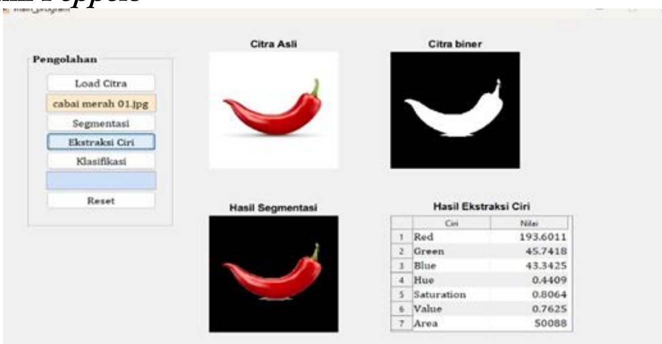


Figure 14. Feature Extraction Values of Red Chili Peppers

Values for Bird's Eye Chili Peppers

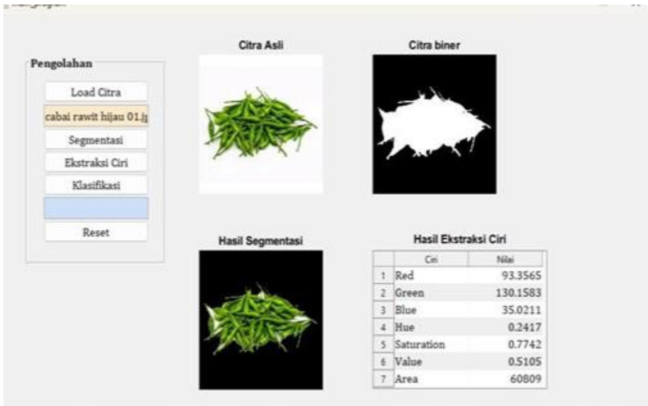


Figure 15. Feature Extraction Values of Bird's Eye Chili Peppers

Values for Red Bird's Eye Chili Peppers

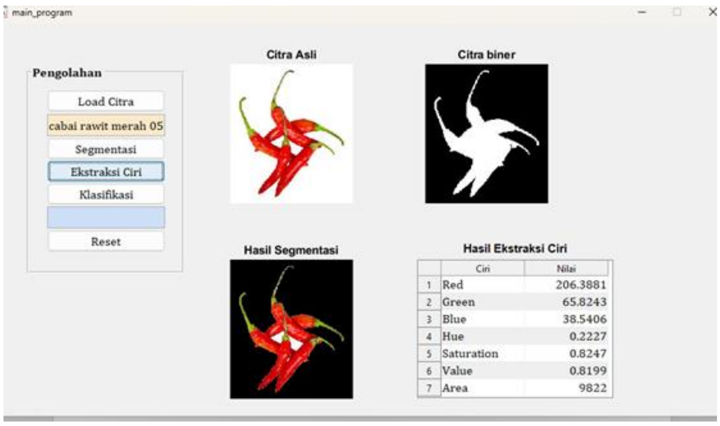


Figure 16. Feature Extraction Values of Red Bird's Eye Chili Peppers

Classification Results

The classification results indicate the chili pepper type for each tested image. Classification accuracy can be evaluated using testing data and is calculated based on the number of correctly classified chili peppers compared to the total number of testing samples. If PCA successfully reduces the data dimensionality while preserving important information, the classification accuracy will generally be high. This indicates that the extracted features are relevant and informative. The following figures present the classification results for several chili pepper testing samples.

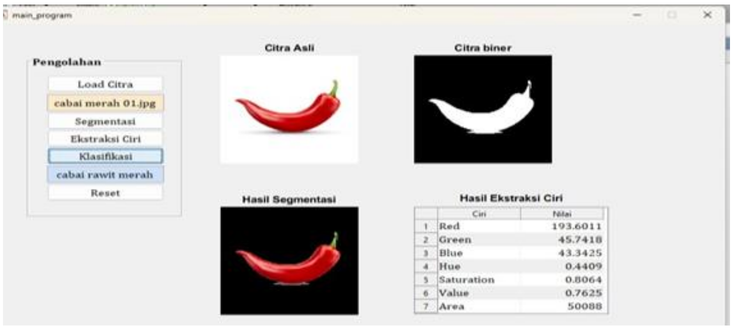


Figure 17. Classification Results of Red Chili Peppers

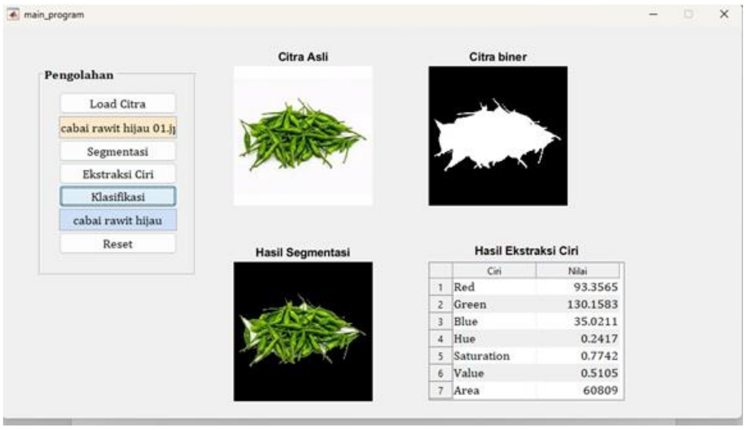


Figure 18. Classification Results of Green Bird's Eye Chili Peppers

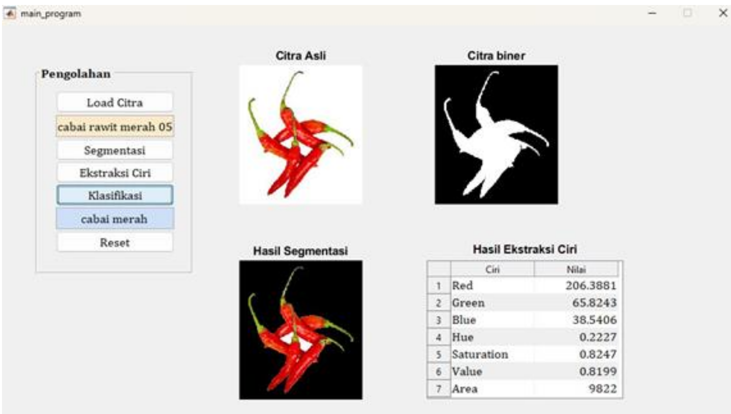


Figure 19. Classification Results of Red Bird's Eye Chili Peppers

Classification Using KNN

After extracting the mean, variance, and range features from the RGB color space, the next step is classification using the K-Nearest Neighbors (KNN) algorithm. KNN is a classification method capable of separating data into more than two classes. In this study, KNN is used to classify different types of chili peppers.

KNN works by identifying the nearest neighboring data points and determining the class of a sample based on the classes of its nearest neighbors. This study uses a linear kernel because of its simplicity and effectiveness in separating data that exhibit linear relationships, as well as its ability to provide accurate classification results without requiring many parameters to be configured.

The following figure shows a code snippet for the KNN classification process using MATLAB tools.

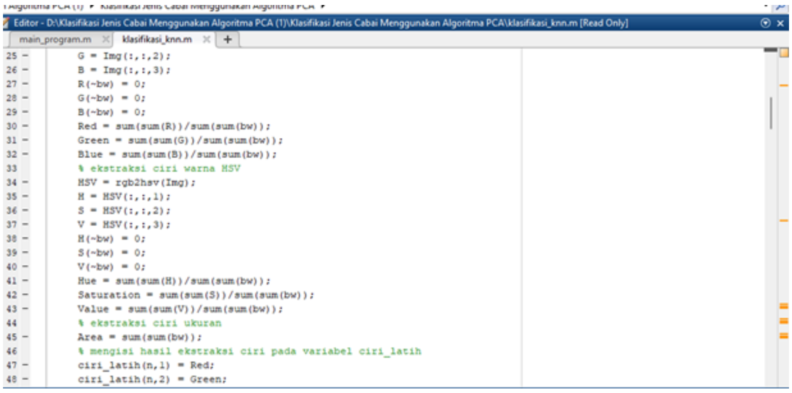


Figure 20. Classification Code Using the KNN Method

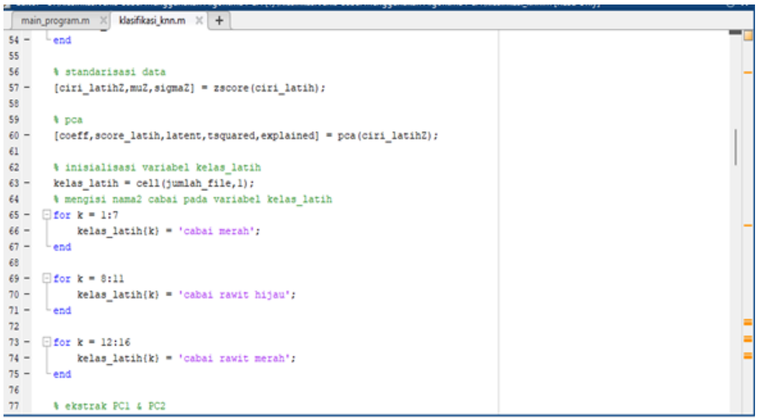


Figure 21. Data Standardization Process Code

5. Conclusion

The identification of chili pepper types is an important aspect that influences various processes, ranging from production to consumption. Each type of chili pepper has distinct colors and shapes, and these variations directly affect their use in various recipes and culinary applications. The development of an automatic detection system for identifying chili pepper types has great potential to benefit society, as the process of identifying chili pepper varieties can be carried out efficiently and accurately, supporting better decision-making at various stages of production and distribution.

Chili pepper data are complex because they involve various factors such as color, size, and shape. Accurate analysis of chili pepper data involves handling high-dimensional data, and processing such data requires substantial computational resources. Principal Component Analysis (PCA) is a commonly used method for reducing data dimensionality while preserving most of the relevant information.

This study successfully implemented a deep learning approach using the Principal Component Analysis (PCA) algorithm to detect chili pepper types from images. The PCA algorithm was applied to process chili pepper images uploaded by users, enabling the system to automatically analyze and identify various chili pepper types. This implementation demonstrates that PCA can be adapted for image classification tasks involving chili pepper types, providing a faster and more accurate solution compared to traditional manual methods.

The PCA model generated through the training process in this study demonstrated significant capability in identifying chili pepper types. Using the available dataset, the PCA model was trained to recognize specific patterns in chili pepper images associated with their respective types.

Suggestions

Based on the results of this study on a chili pepper type detection system using Principal Component Analysis (PCA), two main recommendations for improving the effectiveness and implementation of the developed system are as follows:

To address variability in chili pepper data, it is recommended to use a dataset that includes various lighting conditions, image capture angles, and variations in chili pepper shapes and colors. This will help the model learn to recognize chili pepper types more consistently and accurately.

Conducting comprehensive testing and validation of the developed system by involving various chili pepper types in real-world scenarios will ensure that the system can be relied upon under actual production and distribution conditions.

References

[1] E. Siaga, M. Meihana, S. M. Lumbantoruan, J.-I. Sakagami, and B. Lakitan, "Karakter Morfo-agronomi Tanaman Cabai Merah (Capsicum annum L.) Fase Awal Vegetatif pada Kondisi Stres Jenuh Air," J. Ilmu Pertan. Indones., vol. 29, no. 2, pp. 236–243, 2024, doi: 10.18343/jipi.29.2.236.

[2] Z. H. Batubara, Y. Hamonangan, M. Arfan, and A. Hidayatno, "Perancangan Sistem Deteksi Pelanggaran Penggunaan Helm Dengan Metode Deep Learning Menggunakan Yolov5 Ultralytic," Transient J. Ilm. Tek. Elektro, vol. 13, no. 1, pp. 11–20, 2024, doi: 10.14710/transient.v13i1.11-20.

[3] T. L. Tyasi and W. P. B. Putra, "Principal Component Analysis (PCA) in the Body Measurements of Nguni Cows," Pak. J. Zool., vol. 55, no. 3, pp. 1469– 1472, 2023, doi: 10.17582/journal.pjz/20210605170634.

[4] S. Pan, F. Mechanics, F. Dai, and F. Mechanics, "Measurement algorithm for wafer alignment based on principal component analysis," no. June, 2021, doi: 10.1364/AO.425767.

[5] T. Duan, Z. Liao, T. Li, H. Tang, and P. Chen, "Bearing Fault Diagnosis Based on State-Space Principal Component Tracking Filter Algorithm," IEEE Access, vol. 9, pp. 158784–158795, 2021, doi: 10.1109/ACCESS.2021.3131494.

[6] S. R. Raysyah, A. Veri, and I. M. Dadang, "Classification of Coffee Fruit Maturity Level Based on Color Detection Using Knn and Pca Method," JSiI (Journal Inf. Syst., vol. 8, no. 2, pp. 88–95, 2021.

[7] C. He, J. Li, W. Liu, J. Peng, and Z. J. Wang, "A Low Complexity Quantum Principal Component Analysis Algorithm," IEEE Trans. Quantum Eng., vol. PP, p. 1, 2021, doi: 10.1109/TQE.2021.3140152.

[8] I. Perlindungan and Risnawati, "Pengenalan Tanaman Cabai Dengan Teknik Klasifikasi Menggunakan Metode CNN," Semin. Nas. Mhs. ilmu Komput. dan Apl., pp. 15–22, 2020.

[9] S. F. Nazila, Y. Arman, D. Wahyuni, N. Nurhasanah, and Y. S. Putra, "Deteksi Dini Serangan Hama Penyakit pada Cabai Rawit Menggunakan Metode Image Recognition," J. Tek. Inform. dan Sist. Inf., vol. 9, no. 2, pp. 232–241, 2023, doi: 10.28932/jutisi.v9i2.6342.

[10] A. Setiawan, H. Hadiyanto, and C. E. Widodo, "Dimensional Reduction of Underwater Shrimp Digital Image Using the Principal Component Analysis Algorithm," E3S Web Conf., vol. 448, pp. 1–12, 2023, doi: 10.1051/e3sconf/202344802061.

[11] A. F. Ahmed Mohamed, F. W. Z. Solima, and M. M. Ashour, "Enhance watershed algorithms using principal component analysis capabilities," Al- Qadisiyah J. Eng. Sci., vol. 17, no. 1, pp. 5–15, 2024, doi: 10.30772/qjes.2024.145116.1055.

[12] J. D. Gonzalez-Franco, J. E. Preciado-Velasco, J. E. Lozano-Rizk, R. Rivera- Rodriguez, J. Torres-Rodriguez, and M. A. Alonso-Arevalo, "Comparison of Supervised Learning Algorithms on a 5G Dataset Reduced via Principal Component Analysis (PCA)," Futur. Internet, vol. 15, no. 10, 2023, doi: 10.3390/fi15100335.

[13] Nailul Muna, Norma Ningsih, Nanang Syahroni, Abd. Malik Syamlan, Vina Larasati, and Karimatun Nisa', "Implementasi Algoritma EfficientDet-D0 dan SSD-MobileNet-V2 FPNLite untuk Sistem Deteksi Gulma," Indones. J. Comput. Sci., vol. 13, no. 1, pp. 1324–1333, 2024, doi: 10.33022/ijcs.v13i1.3723.

[14] D. Iskandar Mulyana and M. A. Rofik, "Implementasi Deteksi Real Time Klasifikasi Jenis Kendaraan Di Indonesia Menggunakan Metode YOLOV5,"J. Pendidik. Tambusai, vol. 6, no. 3, pp. 13971–13982, 2022, doi: 10.31004/jptam.v6i3.4825.

[15] R. Irsyada and H. Audytra, "Inovasi Metode Smart (Simple Multi Attribute Rating Technique) Sebagai Sistem Deteksi Covid-19," J. Simantec, vol. 11, no. 2, pp. 157–166, 2023, doi: 10.21107/simantec.v11i2.17258.

[16] S. A. Suryaman, R. Magdalena, and S. Sa'idah, "Klasifikasi Cuaca Menggunakan Metode VGG-16, Principal Component Analysis Dan K- Nearest Neighbor," J. Ilmu Komput. dan Inform., vol. 1, no. 1, pp. 1–8, 2021, doi: 10.54082/jiki.1.

[17] W. Zhao and D. Zhang, "Optimized GICP registration algorithm based on principal component analysis for point cloud edge extraction," vol. 57, no. 1, pp. 77–89, 2024, doi: 10.1177/00202940231193022.

[18] I. M. Mudzopar and T. Wiharko, "Pengembangan Sistem Deteksi Offside Berbasis Metode Yolo dalam Video Pertandingan Sepak Bola," Digit. Transform. Technol., vol. 3, no. 2, pp. 524–530, 2023, doi: 10.47709/digitech.v3i2.2908.

[19] A. Setia Pramuda, A. W. Widhi Nugraha, and A. Fadli, "Perancangan Sistem Deteksi Manusia Menggunakan Sensor PIR, RCWL, dan Infrared Pada Sistem Manajemen Lampu Gedung Berbasis Internet of Things," J. Pendidik. dan Teknol. Indones., vol. 3, no. 1, pp. 1–11, 2023, doi: 10.52436/1.jpti.224.

[20] I. Elan Maulani, D. Rayhan Sunandar Putra, and K. Komarudin, "Sistem Deteksi Intrusi Cerdas: Studi Perbandingan Algoritma Pembelajaran Mesin Untuk Keamanan Siber," J. Sos. Teknol., vol. 3, no. 11, pp. 918–923, 2023, doi: 10.59188/jurnalsostech.v3i11.987.

[21] T. Vol, "PERANCANGAN SISTEM INFORMASI DETEKSI DINI STUNTING BERBASIS," vol. 14, no. 3, 2023.

[22] G. Li, S. Yang, S. Li, N. Wang, and J. Li, "Research and Application of System-based Clustering and Principal Component Analysis Algorithms," Acad. J. Sci. Technol., vol. 5, no. 3, pp. 9–13, 2023, doi: 10.54097/ajst.v5i3.7349.

[23] Zaini Miftach, "済無No Title No Title No Title," pp. 53–54, 2018.

[24] U. Hasdiana, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," Anal. Biochem., vol. 11, no. 1,pp. 1–5, 2018, [Online]. Available: http://link.springer.com/10.1007/978-3-319-59379-1%0Ahttp://dx.doi.org/10.1016/B978-0-12-420070-8.00002-7%0Ahttp://dx.doi.org/10.1016/j.ab.2015.03.024%0Ahttps://doi.org/10.1080/07352689.2018.1441103%0Ahttp://www.chile.bmw- motorrad.cl/sync/showroom/lam/es/

[25] A. Tjolleng, "Buku Pengantar pemrograman MATLAB: Panduan praktis belajar MATLAB," ReasearchGate, no. August, pp. 1–6, 2018.

[26] A. B. Kaswar, F. Adiba, and D. D. Andayani, "Sistem Klasifikasi Tingkat Kematangan Buah Cabai Katokkon Berdasarkan Fitur Warna LAB Menggunakan Artificial Neural Network Backpropagation," J. Embed. Syst. Secur. Intell. Syst., vol. 4, no. November, pp. 149–157, 2023.