Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 3 (2022) 471 - 480 ISSN Media Electronic: 2580-0760

k-Nearest Neighbor and Feature Extraction on Detection of Pest and Diseases of Cocoa

Mohammad Yazdi Pusadan[1], Syahrullah[2], Merry[3], Ahmad Imam Abdullah[4]
[1,2,3]Information Technology, [4]Civil Engineering, Faculty of Engineering, University of Tadulako
[1]yazdi.diyanara@gmail.com, [2]syahroellah.ms@gmail.com, [3]meriinformatika@gmail.com, [4]ahmadimamabd999@gmail.com

Abstract

Knowledge and utilization of digital images are growing rapidly not only in the fields of medicine and industry but also in the field of agriculture. This knowledge can apply it to a computer-based program that is used to detect agricultural products more effectively and efficiently. this research aims to build a system to detect the types of pests and diseases of cocoa pods because in general, an inspection of pests and diseases of cocoa pods is still manual based on the visual analysis of the color of the pods visually by the human eye which has limitations, which requires more energy to sort, the level of human consistency. In terms of assessing the symptoms of pests and fruit diseases, it is not guaranteed, because humans can experience fatigue, and humans also assess symptoms of pests and fruit diseases, sometimes it is subjective. This study utilizes digital image processing techniques to extract the color features of digital images of cocoa pods, the method used to extract the color features of Hue, Saturation, Value (HSV), and the classification algorithm used by K-Nearest Neighbor. The data used as many as 150 images divided into 70% training data and 30% testing data. Based on the results of trials using k values of 5,7,11 and 13 in the holdout method, the best accuracy is 84.44% with a value of k = 5. And in the k-5 cross-validation test, the best accuracy is also found at k = 5 with a value accuracy of 99.33%.

Keywords: Cocoa pods, Detection of pests, and diseases of cocoa pods, HSV, KNN, Confusion Matrix, K-Fold Cross-Validation.

1. Introduction

Knowledge and use of digital images is growing rapidly, not only in medicine, manufacturing and sanitation, but also in agriculture for detection, quality control and classification. Based on this knowledge, it can be applied to computer-based programs that can be used to detect agricultural products more effectively. One of them is to determine the types of pests and diseases on cocoa pods.

Cocoa is one of the plantation commodities whose role is quite important for the national economy, especially a source of income and foreign exchange for the country where the beans from cocoa pods can be used as processed products known as chocolate. However, most of the cocoa plantations in Central Sulawesi Province are managed by the people. In cacao cultivation, the community often faces problems that can cause a decrease in cocoa productivity and quality. In 2015 it fell to 797 kg/ha and in 2017 it was estimated by the Directorate General of Plantations to decrease to 785 klg/ha, this was caused by pests and diseases because this plant is very susceptible to pests and diseases. Inspection of pests is generally carried out by directly observing the types of pests and diseases of cocoa pods by looking at the cocoa pods one by one based on a visual analysis of the color of the fruit skin by the human eye. This certainly affects the time needed to determine the types of pests and diseases of cocoa pods. For that reason, an automated method is needed that can improve accuracy with consistent assessments in detecting types of pests and diseases of cocoa pods.

This research will utilize digital image processing techniques, namely the *Hue, Saturation, and Value* (HSV) method for color processing. Color extraction results from image processing are used as input in the classification process. The classification method used in this study is the *k-Nearest Neighbor* (k-NN) algorithm.

The problem in this research is how to detect the types of pests and diseases of cocoa pods by utilizing digital image processing using K-Nearest Neighbor. The scope of the research is the detection of pests and diseases of cocoa in this study using the color object of cocoa pods with the Feature Extraction Hue Saturation Value (HSV) method; the types of cocoa pods studied were

Accepted: 06-05-2022 | Received in revised: 19-06-2022 | Published: 30-06-2022

farastero; and the types of diseases and cocoa pests that will be detected in this study are the cocoa pod borer (Conopomorpha Cramerella) and the fruit-sucking ladybug (Helopeltis sp.). The type of cacao disease that will be detected in this study is fruit rot (Phytophthora palmivora) in the cocoa garden of Potoya Village, Sigi Regency, Central Sulawesi Province.

This study aims to detect the types of pests and diseases of cocoa pods by utilizing digital image processing using the responsive web-based K-Nearest Neighbor algorithm. This research was conducted based on a review of several previous research results, namely as follows.

1. Identification of Diseases in Cocoa Fruits Using *Hue Saturation Value* and *Moment Invariant* with *Backprogation Algorithm* [1]. This study identifies cacao pod disease by comparing the sample of cacao pods to be studied with reference data in the database and makes cultivating cocoa farmers easier in dealing with the problem of decreasing cocoa yields due to disease attacks that are not recognized by cocoa farmers. The method used is color feature extraction using *Hue Saturation Value* (HSV) and moment invariant with backpropogation algorithm with 89.2% accuracy.

2. Identification of Disease Types in Cocoa by Digital Image Processing and *K-Nearest Neighbor* [2]. This study applies the k-Nearest Neighbor method in classifying the type of cocoa disease. The process is to classify cacao fruit diseases with the classification algorithm used is *K-Nearest Neighbor*. The method used is *Principal Component Analysis* (PCA) with the highest accuracy of 86.67%.

3. Classification of Cocoa Based on e-nose with *Neuro Fuzzy Method* [3]. The process is to apply an electronic nose for cocoa quality classification using neuro fuzzy analysis by identifying cocoa quality based on color and aroma by involving a human tester. In this study, using e-nose with the *Neuro Fuzzy* method for the quality of cocoa pods with an accuracy rate of 95.21%.

4. Classification of apples using the *K-Nearest Neighbor* method [4]. The process is to apply the *Hue Saturation Value* (HSV) and *Local Binary Pattern* (LPB) color feature extraction methods for the classification of apple types, so that the name of the apple species can be known by utilizing image processing technology, which can be identified based on the color. The method used is the k-NN classification and uses two color feature extraction methods, with the results obtained reaching an accuracy of 94%.

5. Siam citrus disease detection based on leaf image using RGB-HSV color segmentation [5]. The process is to apply the *Hue Saturatin Value* (HSV) color feature extraction method and to detect citrus plant diseases. The object of research is citrus fruit and the method used is the fuzzy K-NN classification, so that the accuracy rate of 78% is obtained.

6. A Method for Detecting and Segmenting Infected Part of Cacao Pods [6]. This study applies a method to detect and segment the infected parts of cocoa pods. The process is to detect the severity of the disease in each object using the L*a*b* color extraction method as a feature, so that an accuracy rate of 89.2% is obtained.

7. Banana species classification system based on HSV color characteristics using the HSV method [7]. This research is to help classify the types of bananas and the level of fruit maturity properly based on HSV. The method used is HSV color feature extraction and K-NN classification method, this study obtained an accuracy rate of 82%.

8. Classification of mango fruit maturity based on HSV and K-NN images [8]. This study classified mango ripeness based on color feature extraction. The method used is HSV color feature extraction and K-NN classification method, this study obtained an accuracy rate of 80%.

9. Implementation of image processing and K-Nearest Neighbor classification to build beef and pork differentiation applications [9]. This research is to classify the image of beef and pork. The method used is HSV color feature extraction and k-NN classification method with an accuracy rate of 88.75%.

10. Identification and Classification of Fruit Diseases [10]. This study is to identify and classify fruit diseases with different fruit objects, namely apples, bananas, oranges, grapes, guavas, mangoes, papayas, peaches, pomegranates, and watermelons. The classification accuracy is relatively higher in all cases when segmented using k-Means than the C-Means clustering algorithm with the GLCM feature extraction method.

11. Detection of surface defects of mangosteen using deep learning and multilayer convolution methods [11]. This study detects defects on the surface of the mangosteen fruit based on digital image processing. The object of research is to determine the level of quality of the mangosteen fruit. The method used is deep learning CNN (Convolutional Neural Network) with an accuracy of 98%.

12. Image Processing System for Early Detection of Cocoa Fruit Pest Attack [12]. Early detection research based on image processing of pest attack symptoms on cocoa pods. The process that takes place is to detect objects on the surface of cocoa pods that are attacked by symptoms of pests and diseases by taking pictures using a white background. 100% training process and 70% data test.

13. Implementation of the backpropogation method to identify the type of defective cocoa beans based on the shape of the beans [13]. This research is to identify the

DOI: https://doi.org/10.29207/resti.v63.4064

quality of cocoa beans which is implemented into a computer-based program. The method used is edge detection and backpropogation Artificial Neural Network, the resulting accuracy rate is 76%.

14. Cocoa Care - An Android Application for Cocoa Disease Identification [14]. This study identifies cacao disease using replacing manual inspection of cacao disease by android application which identifies cacao disease from captured images and suggests possible solutions for farmers. The methods used are texture feature extraction and the k-means algorithm. The level of accuracy obtained by using this method is 100%.

15. Image Processing Approach for Grading and Identification Of Diseases On Pomegranate Fruit [15]. This study detects and assesses disease in pomegranates. The method used is texture feature extraction and k-means algorithm to detect pomegranate images on infected pixels so that pixel features do not dominate between uninfected pixels and artificial neural networks as a classification algorithm.

16. Identifying the Quality System of Cocoa Beans to Increase Productivity Using Backpropagation Neural Network Algorithm: A Case Study at Sumber Rejeki Farmers Group, Patuk Gunung Kidul [16]. This research is to produce a classification of the quality of cocoa beans to increase productivity. The process is to use a cocoa bean object to identify the quality of cocoa beans based on texture and color. The method used is the backpropagation classification algorithm with the principal component method analysis (PCA) extraction method, with an accuracy rate of 70%.

17. Classifying Physical Morphology of Cocoa Beans Digital Images using Multiclass Ensemble Least-Squares Support Vector Machine [17]. This study was to determine the quality of cocoa beans through the morphology of the image of cocoa beans. The method used is MELS-SVM, with an accuracy rate of 99.705%. Identification of Cocoa Pods with Image Processing and Artificial Neural Networks [18]. This study determines the ripe cocoa pods so that the cocoa pods can be harvested in good condition. As for identifying the maturity of cocoa pods where the image is segmented and then extracted the color features. The classification method used is a feedforward neuronal network (FNN) artificial neural network in MATLAB programming. The acquisition rate of accuracy in this study reached 91%.

18. Fermentation Level Classification of Cross Cut Cacao Beans Using k-NN Algorithm [19]. This study identified the quality of cocoa beans based on the level of fermentation, namely well fermented, under-fermented and over-fermented. The digital image processing technique in this study uses fermented cocoa beans to determine the quality of the cocoa beans with RGB feature extraction and the KNN algorithm. In this study obtained an accuracy rate of 92.50%.

2. Research Methods

Research Material

The research material used by the author comes from the acquisition of images of cocoa pods taken directly based on their color in Potoya Village, Dolo, Sigi, Central Sulawesi Province.

Data types and sources

The primary data source was obtained from the Plantation and Livestock Service Office of Central Sulawesi Province which determined the types of pests and diseases of cocoa pods, accompanied by secondary data from library materials as supporting data.

Image Acquisition

Image acquisition is done by taking pictures using a camera at a distance of approximately 20 cm with a white cocoa pod as a background. There were 10 images on each cocoa pod, 10 images of 15 cacao pods, 150 datasets obtained with 50 images of fruit rot disease, 50 images of ladybugs and 50 images of cocoa pod borer. After obtaining the image, it will be divided into training datasets and testing data, the distribution of the data using the Hold Out Method, 70% training data, 30% testing data, validation of the distribution data as training data and testing data using K-fold cross validation. In this study, k-fold cross validation was used with k = 5 to see the accuracy of the data being trained and tested. The results are calculated by calculating the average accuracy of each iteration.

System Development Method

The system development method is Waterfall model and using the k-nearest neighbor algorithm are as follows: i) the stages of adding data are carried out as a process for recording data acquisition of cocoa pods as training data. ii) the detection stage is chosen if the data addition stage has been carried out. iii) the classification results are the final stage of the system, namely displaying the results of the analysis of the detection of pests and diseases of cocoa pods; and iv) HSV feature extraction process. The stages of HSV feature can be seen in Figure 1.

The explanation of the stages is as follows:
i) Image acquisition is a data retrieval process where the data will be stored in the database. The image is processed in the form of a matrix, so that a series of matrices will be obtained that are ready to be trained.

ii) Preprocessing, at this stage image processing is carried out to produce a better image to be processed to the next stage which consists of several processes including cropping, resizing, and normalizing RGB images. Cropping is the process of cutting an image, to

DOI: https://doi.org/10.29207/resti.v63.4064

remove parts of the image that are not needed. Resize is used to normalize the size of the image so that it has the same size. While the normalization of RGB image aims to make the values for each color component can be compared with each other because the image is taken under different intensity conditions.

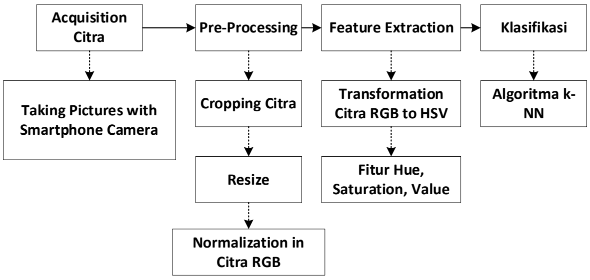

Figure 1. The stages of the HSV feature extraction pattern recognition process

iii) Feature extraction, at this stage of taking object features that can be used as a distinguishing material from other objects. In this study, using HSV color feature extraction, the resulting value becomes the input value in the image classification process. To get the value of the previous HSV feature, an RGB to HSV image transformation is carried out by calculating the equation (2.4 to 2.7); and iv) classification, at this stage the algorithm is able to classify objects in the existing class based on the characteristics of the object using the K-Nearest Neighbor (K-NN) algorithm. The class categories are fruit rot disease, pbk pests, fruit-sucking ladybugs, The process stages of the K-NN algorithm can be seen in Figure 2 with the following series of steps.

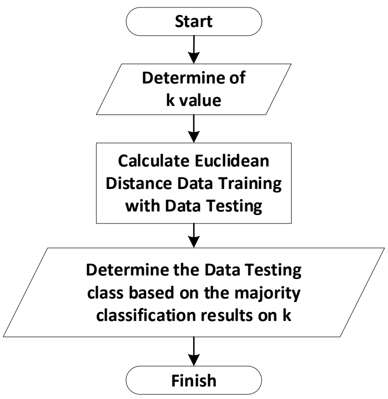

Figure 2. Stages of the K-Nearest Neighbor (K-NN) Algorithm

The explanations from Figure 2 are: i) determining the value of k as the nearest neighbor parameter to be selected by testing which k has the best accuracy; ii) calculating the distance between data, at this stage the distance calculation is carried out using the Euclidean distance formula in equation (2.8); iii) sorting the data from the smallest distance, at this stage sorting the objects into groups or classes that have the smallest eclidean distance; and iv) determine the testing data class based on the majority on k, at this stage the most majority class can be predicted as an object category.

1. Data Input, after collecting cocoa fruit image data, the data is inputted into the detection system for types of pests and diseases of cocoa pods based on needs.

2. Testing, testing the K-NN algorithm is carried out using testing data and direct testing on the image of cocoa pods found in the owner of the Potoya Village cocoa garden. The test is carried out by comparing the results of the system implementation and the test results from the Plantation and Livestock Service Office of Central Sulawesi Province, Palu City. The confusion matrix can be seen in Table 1..

Table 1. Confusion matrix multi class

TC = True Class		Classification		
FC = False Class		C1	C2	C3
Class Target	C1	TC1	FC1	FC1
	C2	FC2	TC2	FC2
	C3	FC3	FC3	TC3

Description:

C1 = Desease BB; C2 = Pest PBK; C3 = Pest KPB.

1. Conclusion, at this stage the conclusion is the final result that is expected to be able to answer the purpose of proposing this technology product, namely the detection of pests and diseases of cocoa pods using the k-nearest neighbor algorithm. Based on the stages of the proposal above, the following is a flow chart of the proposal in Figure 3 below.

In its application, the user is divided into two parts, namely admin and user. The admin in this case is the head of the farmer group, while the users are a group of farmers or the general public who want to detect pests and diseases of cocoa plants.

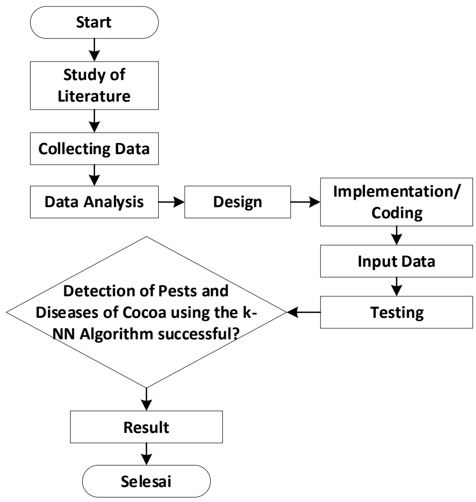

Figure 3. Flow Diagram

DOI: https://doi.org/10.29207/resti.v63.4064

Admin is in charge of updating data and information about pests and plant diseases, as well as conducting data training.

Furthermore, on the user side, which is only the detection process on cocoa pods whether there are pests or plant diseases. After the system is optimally validated and has good accuracy in the detection process, the resulting information can be used by policy makers to determine regulations regarding cocoa plants. These regulations include: i) purchase of superior seeds accompanied by anticipation of pests and plant diseases, ii) grouping of diseased and pest attacked cocoa pods, whether they can be used again by farmer groups; and iii) assistance of superior seeds from the government to farmer groups.

3. Results and Discussions

System Analysis

Data collection, the data to be used is the image of cocoa pods based on the types of diseases and pests, namely BB Disease, PBK Pests and KPB Pests which will be training data and test data. Cocoa fruit image data retrieval is done using a smartphone camera, cacao fruit image capture is carried out 10 times using a rear camera with a camera screen position, namely portrait, the process of taking cocoa fruit images is assisted using white paper so that the fruit object is clearer and a ruler so that the distance between objects which will be captured from the camera position remains the same, namely 15 cm to 20 cm.

Data Training, in the training process, researchers used images of cocoa pods taken using a smartphone camera. Next, cropping is done which only takes the focus of the cocoa fruit image object. The resizing process is carried out to uniform the size of the image. After that, normalize the RGB image so that the values for each color component can be compared with each other because the image is taken in different intensity conditions, after that the RGB to HSV image transformation process is carried out after the feature extraction of Hue, Saturation, Value features will proceed to the next step. K-NN algorithm calculation.

a) Data input, the first process is inputting training data, where the data is obtained using a smartphone camera, then the image is transferred to a computer for upload via file browse. The data used in the training data process is 105 data consisting of 35 images of BB disease category data, 35 images of CPB pest category data and 35 data images of KPB category data. The data input process can be seen in Figure 4.

Procedure for uploading images that will be used as input in the training data process.

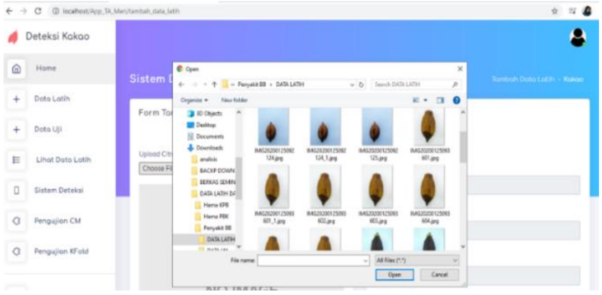

Figure 4. Upload Cocoa Image

b) Cropping is the process of cutting the image at certain coordinates in the image area. To cut part of the image, two coordinates are used, namely the initial coordinate which is the initial coordinate for the cutting image, the final coordinate which is the final coordinate point of the cropped image. So that it will form a rectangular shape where each pixel in a certain coordinate area will be stored in a new image. The results of cropping before and after the image of cocoa pods can be seen in Figure 5.

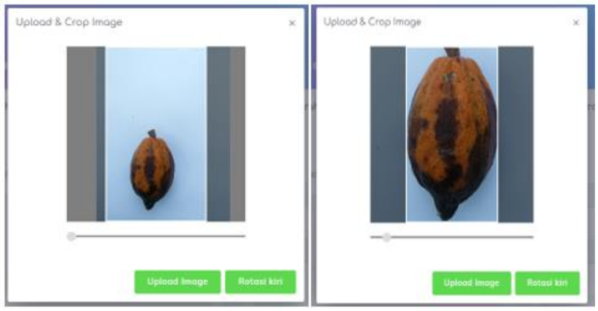

Figure 5. Cropping Image Process

c) Feature extraction plays an important role in distinguishing the types of objects. The processed characteristics are in the form of values that can be used to distinguish between an object and another object. Characteristics are expressed by an array of numbers that can be used to identify feature objects to be processed using color feature extraction obtained from resizing results using the Hue, Saturation, Value feature extraction method. Procedure for extracting HSV features that will be used as identical parameters for each cocoa pod image.

d) The k-NN algorithm process, k-NN uses the value of k parameter in determining the decision results. Then calculate the distance between the new data and all existing data in the training data and then determine the nearest neighbor based on the kth minimum distance which uses the category that has the most frequency from the nearest neighbor, as the predicted value or the classification result of the new data.

System Test

The test is divided into several parts, namely testing the system function using blackbox testing, then the second

DOI: https://doi.org/10.29207/resti.v63.4064

test is carried out to measure the accuracy of the detection system for pests and diseases of cocoa pods using the HSV color feature extraction and the K-NN algorithm. Furthermore, testing is carried out on the training data to calculate the confusion matrix so that the accuracy and error of this system can be known and the K-fold cross validation test on the cocoa pod image dataset used for training data. The following are several stages of system testing.

a) Blackbox testing Testing system function testing is done by using blackbox testing. For testing system functions can be seen in Table 2.

Table 2. Testing with blackbox testing method

Num.	Function of testing	System expectation	Result
1.	Verification Login (Form Login)	Can perform login verification, that is, if the username and password are entered correctly, the admin can enter the main menu, if incorrect, a message will appear to log back in.	Succeed
2.	Input Data (Form Data Training)	input training data	Succeed
3.	Input Data Testing (Form Data Testing)	Can input data test	Succeed
4.	Detail, Update, Delete Data Training (Form View Data)	Can display details, can change and delete training data	Succeed
5	Process Detection Cacao (Form System Detection)	Can know the results of detection of cocoa pods	Succeed
6	Function Capture Citra Cacao	Can capture Cacao image	Succeed
7	Function Upload Citra Cacao	Can upload Cacao image	Succeed
8	Proses Cropping	Can cropping Cacao image	Succeed

a) Cocoa fruit detection test, in testing the Cocoa fruit detection system the total data is 150 images, then the data will be divided into 70% training data and 30% as test data. Where 105 as training data which is divided from 35 images of fruit rot disease, 35 images of cocoa pod borer pests and 35 images of fruit-sucking ladybugs. For the test data, 45 images were divided into 15 images of fruit rot disease, 15 images of the cocoa pod borer and 15 images of fruit-sucking ladybugs. From the whole training data, it will be tested using various k values, namely 5,7,11 and 13. The following diagram shows the percentage of accuracy results from testing the detection system can be seen in Figure 6.

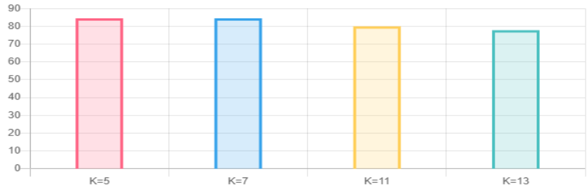

Figure 6. Percentage of K-NN Algorithm Testing Accuracy

b) Confusion Matrix testing, this test is carried out to get accuracy and error rate. Accuracy aims to add up the correct prediction results and the error rate is used to add up the wrong prediction results. The results of the confusion matrix test can be seen in Table 3, Table 4, Table 5, and Table 6.

The calculation of accuracy and error rate at the value of **k = 5** is as follows.

a. *Accuracy*
Accuracy = 13 + 13 +12 / 45 x 100% = **84,44 %**

b. *Error rate*
Error rate = 100 – 84,44% = **15,56%**

Table 3. Confusion Matrix Multi Class with K = 5

		Classification Result		
		Diseases BB	Pest PBK	Pest KPB
Class Target	Disease BB	13	2	0
	Pest PBK	1	13	1
	Pest KPB	2	1	12

The calculation of accuracy and error rate at the value of **k = 7** is as follows.

a. *Accuracy*
Accuracy = 13 + 13 +12 / 45 x 100% = **84,44 %**

b. *Error rate*
Error rate = 100 – 84,44% = **15,56%**

Table 4. Confusion Matrix Multi Class with K = 7

		Classification Result		
		Disease BB	Disease BB	Disease BB
Class Target	Disease BB	13	1	1
	Pest PBK	1	13	1
	Pest KPB	2	1	12

The calculation of accuracy and error rate at the value of **k = 11** is as follows.

a. *Accuracy*
Accuracy = 12 + 13 +11 / 45 x 100% = **80,00 %**

b. *Error rate*
Error rate = 100 – 80,00% = **20,00%**

Table 5. Confusion Matrix Multi Class with K = 11

		Classification Result		
		Disease BB	Disease BB	Disease BB
Class Target	Disease BB	12	3	0
	Pest PBK	1	13	1
	Pest KPB	2	2	11

The calculation of accuracy and error rate at the value of **k = 13** is as follows.

a. *Accuracy*
Accuracy = 11 + 13 +11 / 45 x 100% = **77,78 %**

b. *Error rate*
Error rate = 100 – 77,78% = **22,22%**

Table 6. Confusion Matrix Multi Class with K = 13

		Classification Result		
		Desease BB	Desease BB	Desease BB
Class Target	Desease BB	11	3	1
	Pest PBK	1	13	1
	Pest KPB	2	2	11

c) Testing k-fold cross validation. The testing phase to validate the training of cocoa fruit image data is carried out using the 5-fold cross validation method. The dataset is divided, which initially amounted to 150 data, into 5 subsets (sections) with each subset totaling 30 data. In the first fold, a combination of 4 different subsets is combined and used as training data, while 1 subset is used as test data. The results of the test accuracy can be seen in Table 7, Table 8, Table 9, and Table 10.

Table 7. Result of K-fold cross validation with K = 5

K-fold cross validation with K = 5				
Feature	fold	Amount Data Comparison		Accuracy
		Data Test	Data Training	
Colour Feature	fold-1	30	120	100%
	fold-2	30	120	100 %
	fold-3	30	120	96.67%
	fold-4	30	120	100%
	fold-5	30	120	100%
	Average Accuracy		99,33%	

Table 8. Result of K-fold cross validation with K = 7

K-fold cross validation with K = 7				
Feature	fold	Amount Data Comparison		Accuracy
		Data Test	Data Training	
Colour Feature	fold-1	30	120	36.67%
	fold-2	30	120	100%
	fold-3	30	120	96,67%
	fold-4	30	120	100%
	fold-5	30	120	100%
	Average Accuracy		86,67%	

In the process of training and testing carried out until the last fold. The results of data validation that have been tested based on the table above can be seen that in fold-2 every k value in KNN does not experience significant changes. However, based on the table above, it can also be seen that the lowest data validation is found in fold-1 at the value of k on the k-NN experiencing a significant change with the lowest accuracy at 6.67%. The highest average number of accuracy in this test is found at the value of K = 5 with an accuracy value of 99.33%. And the lowest average accuracy is found at the value of K = 13 with an accuracy value of 78.67%. From the data validation results, it is known that the best K value used is K = 5, with fold-2 as test data and the rest of the partitions as training data. The results of k-fold cross validation on the system are displayed in the form of a diagram which can be seen in Figure 7.

Table 9. Result of K-fold cross validation with K = 11

K-fold cross validation with K = 11				
Feature	fold	Amount Data Comparison		Feature
		Data Test	Data Training	
Feature Colour	fold-1	30	120	6.67%
	fold-2	30	120	100%
	fold-3	30	120	100%
	fold-4	30	120	100%
	fold-5	30	120	90%
	Average Accuracy		79,33%	

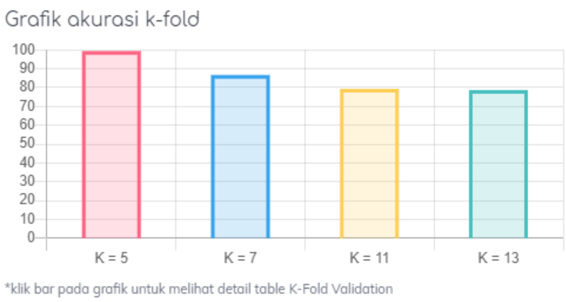

Figure 7. K-Cross Validation Accuracy

The average accuracy of the K-fold cross validation test on the k value of the KNN experiment can be seen in Table 10.

Table 10. Result of K-fold cross validation

K-fold cross validation	Nilai K			
	K =5	K =7	K = 11	K=13
Average Accuracy Fold 1 -5	99,33%	86,67%	79.33%	78.67%
	Rata – Rata Accuracy			86%

The result of color feature extraction of cocoa pods

Extraction of Hue Saturation Value color features, on the image of cocoa pods based on equations (2.1) to (2.7) produces feature values of the color characteristics of cocoa pods that contain types of pests and diseases. The range of H feature values in cocoa fruit images ranges from a minimum value of 12555.9 to a maximum value of 28882.6. The range of S feature values is in the range of 14700.5 to 39801.9 and the range of V feature values is in the range of 33964.9 to 47470.3. The range of values from the feature extraction of cocoa pods can be seen in Table 11 and the total range of values for

DOI: https://doi.org/10.29207/resti.v63.4064

pests and diseases of Cocoa pods can be seen in Table 12.

Table 11. Range Value each Color Feature

Color Feature	Min	Max
H	12555.9	28882.6
S	4700.5	39801.9
V	33964.9	47470.3

Table 12. Overall Value Range of Pests and Diseases of Cocoa Fruit

Category	H		S		V	
	(-)	(+)	(-)	(+)	(-)	(+)
Disease	1427	2888	1470	3825	3396	4738
BB	3.6	2.6	0.5	8.7	4.9	1.4
Pest PBK	1390	2498	2301	3837	3898	4747
	5.1	7.4	9.5	6.3	.1	0.3
Pest KPB	1255	1636	3178	3980	3564	4685
	5.9	6.7	3.7	1.9	8.6	9.4
(-)	12555.9		14700.5		3898.1	
(+)	28882.6		39801.9		47381.4	

Information:
(-): Min, (+): Max

Cocoa fruit image detection results

Detection of pests and diseases of cocoa pods using the KNN algorithm with various k values of 5,7,11, and 13 based on equation (2.8) produces a different Euclidean distance limit value for each k value. For k = 5 it ranged from 176,235 to 5013, for k = 7 it ranged from 176,235 to 5224.15, for k = 11 it ranged from 176,235 to 4824.79, and for k = 13 it ranged from 176,235 9016.63.

The value of the Euclidean distance limit for each k value can be seen in Table 13.

Table 13. Distance limit with Euclidean

Category	K = 5		K = 7		K = 11		K = 13	
	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
Desease	17	95	17	51	17	43	17	4844
BB	2.0	4.7	2.0	83.	2.0	69.	2.0	.24
	12	69	12	63	12	84	12	
Pest PBK	42	46	42	56	42	61	42	6686
	6.9	59.	6.9	02.	6.9	50.	6.9	.18
	89	73	89	23	89	1	89	
Pest KPB	23	92	23	57	23	61	23	6792
	9.0	8.0	9.0	61.	9.0	15.	9.0	.09
	96	32	96	7	96	69	96	
(-)	172.012		172.012		172.012		172.012	
(+)	954.769		5761.7		6150.1		6792.09	

Discussion

The results of testing the k-NN algorithm using the value of K

This test uses 45 test data, where the test data is divided into 15 images of BB disease, 15 images of CPB pests, and 15 images of CPB pests. Tests conducted at the Department of Plantation and Animal Husbandry of Central Sulawesi Province, UPT. Plantation Plant Protection where the original test data image is validated by an expert on the type of cocoa plant then the results are compared with the results of the implementation of the K-NN algorithm. The result is that there are 2 data on BB disease category detected in the CPB Pest category, in the CPB pest category there is 1 data detected as BB disease and 1 detected as KPB pest, as for the KPB pest category 2 data was detected as BB disease and 1 data was detected as CPB pest. These results can be seen in Table 14.

Table 14. The Results of the k-NN Algorithm with a value of K = 5

DATA	Detection Result		
	Data Test	UPT Plant Protection Kakao Result	System Result
Disease BB	15	15	13
Pest PBK	15	15	13
Pest KPB	15	15	12

Test results Multiclass confusion matrices

Multiclass confusion matrices testing was carried out to determine the accuracy and error rate used as parameters for the success and accuracy of the K-NN algorithm in detecting images of pests and diseases of cocoa pods. The results of the Multiclass confusion matrices test can be seen in Table 18.

Table 15. Test Results of Multiclass Confusion Matrices

Multiclass confusion matrices	K Value			
	K =5	K =7	K = 11	K=13
Accuracy	84.44%	84.44%	80.00%	77.78%
Error Rate	15.56%	15.56%	20.00%	22.22%
		Average Accuracy		81,67%
		Average Error Rate		18,33%

K-fold cross validation test results

Validation of test results that have been tested for validity using the K-fold cross validation method. Where the dataset will be divided into training data and test data, where the training data will be divided into several partitions. Then a number of K-experiments were carried out, where each experiment used the K-th partition data as test data (validation) and used the remaining partitions as training data. k-fold can reduce computation time while maintaining the accuracy of the estimate. In other words, there will be no data bias. The partition K value used is 5, while the cocoa fruit image data set contains 150 image data consisting of 3 categories, namely BB disease, CPB pests, KPB pests, 5 iterations of experimental K times by calculating the accuracy value. To get the amount of test data for each iteration as follows.

$$k = \frac{150}{5} = 30$$

Which is:

K: 5 partition; *N*: 150 data set; dan *k*: 30 the amount of test data for each iteration.

The division of partitions is done randomly but the categories are entered in order so that the distribution of

DOI: https://doi.org/10.29207/resti.v63.4064

data used in each partition is balanced. The results of the partition data division can be seen in Table 19.

Table 16. Distribution of Data Partition

id	Feature Color			Category	K Partition
	H	S	V		
1	24846.8	28794.4	34632.7	Disease BB	1
:	:	:	:	:	:
10	28431.2	14846.8	40735.4	Disease BB	1
11	28413.6	14792.2	40766	Disease BB	2
:	:	:	:	:	:
20	14612.1	29176.5	41076.6	Disease BB	2
21	14612.1	29176.5	41076.6	Disease BB	3
:	:	:	:	:	:
30	17575	37279.2	40529	Disease BB	3
31	16929.8	38258.7	39844.8	Disease BB	4
:	:	:	:	:	:
40	27775.6	14868.9	40662.1	Disease BB	4
41	27858.8	15017.6	39033	Disease BB	5
:	:	:	:	:	:
50	17243.6	37082.8	41290.3	Disease BB	5
51	19560.3	25093.7	46150.1	Pest PBK	1
:	:	:	:	:	:
60	18081.2	32220.2	43495.3	Pest PBK	1
61	24987.4	25801	44082.4	Pest PBK	2
:	:	:	:	:	:
70	24095.5	26321.4	43496.7	Pest PBK	2
71	24960.4	25895.9	44209.2	Pest PBK	3
:	:	:	:	:	:
80	23208.6	25925.6	45068.7	Pest PBK	3
81	23443.4	25536.9	45421.3	Pest PBK	4
:	:	:	:	:	:
90	16492.9	36118.1	41946.9	Pest PBK	4
91	16951.1	35401.6	42276.8	Pest PBK	5
:	:	:	:	:	:
100	23693.5	25363.6	45500.2	Pest PBK	5
101	15290.3	34418.1	39295.3	Pest KPB	1
:	:	:	:	:	:
110	13771	32157	36674.4	Pest KPB	1
111	13114.4	33288.2	35989.3	Pest KPB	2
:	:	:	:	:	:
120	16314.6	37226.5	46657.9	Pest KPB	2
121	16330.7	36965.6	46605	Pest KPB	3
:	:	:	:	:	:
130	13388.2	37471.7	45112.3	Pest KPB	3
131	14606.2	36937.8	46859.4	Pest KPB	4
:	:	:	:	:	:
140	17723.9	25749.7	40584.5	Pest KPB	4
141	16141.5	28317	39110.7	Pest KPB	5
:	:	:	:	:	:
150	12240.1	40285.7	45060.8	Pest KPB	5

The results of the test accuracy with Confusion matrices and the results of the validation with K-fold cross validation there are no significant differences. Test accuracy with Confusion matrices with validation results with K-fold cross validation is shown in Table 20.

Table 17. The results of the Hold Out Method and K-fold Cross Validation Test Accuracy

Training Method	Value K on k-NN Algorithm	Accuracy with HoldOut Method (%)	Accuracy of Validation with K-fold cross validation (%)
k-Nearest Neighbor	K = 5	84.44%	99,33%
	K = 7	84.44%	86,67,33%
	K = 11	80.00%	79,33%
	K = 13	77.78%	78,67%

In cross validation with the fold = 5 test there is no significant difference between the accuracy of the HoldOut Method, this is due to the distribution of training data and test data having balanced classes and k-partitions. The cause of the low or far significant accuracy in the fold is data bias. The value of the resulting color features have similarities with the image characteristics of other categories of cocoa pods so that there is an error in detection. This is also caused by the unbalanced range of feature values in each parameter that can affect the quality of the image data results. Therefore, a data segmentation process is needed to improve the accuracy of the results of classifying datasets.

Discussion of System Detection Results

Testing the K-Nearest Neighbor algorithm which was carried out using 150 data divided into 105 training data and 45 test data of cocoa fruit image data, it was found that the detection process for pests and diseases of cocoa pods has a fairly good level of accuracy based on the Multiclass confusion matrices test. that by using k values of 5 and 7 produces a fairly good accuracy of 84.44% so that the detection results are quite good. However, in the tests carried out on several k values in the K-NN algorithm, it was found that there were inappropriate detection results so that there was a decrease in the accuracy value for every increase in the value of k. This is because the value of the neighboring limit that is determined exceeds the amount of training data for each cocoa pod image used during the training. Based on the working principle of the K-NN algorithm, it is known that the value of k is very dependent on the data, a high value of k will reduce the effect of noise on the classification, but make the boundaries between each classification more blurred.

Several other factors that cause errors in the detection of pests and diseases of cocoa pods, among others, are caused by the quality of the dataset which is influenced by the level of light entering the camera which makes the intensity and color contrast of the image not match so that the resulting color feature values have similarities with the features. image of pests and diseases of cocoa pods in other classes so that there is an error in the classification and also the unbalanced feature value range for each attribute that can affect the quality of the image data. In addition, in this study, researchers conducted process experiments on the detection system where the system detection results were strongly influenced by the object input, in the results of this experiment the system was able to detect correctly when the input object was in a vertical position either in reverse rotation of the object, but the system detection results on the input the object in the horizontal position of the object cannot be detected correctly or incorrectly in the system classification output result.

DOI: https://doi.org/10.29207/resti.v63.4064

4. Conclusion

Based on the results of testing and analysis of the detection system for pests and diseases of cocoa pods, it can be concluded that the detection of pests and diseases of cocoa pods can be done using the k-Nearest Neighbor (KNN) algorithm based on the extraction of Hue, Saturation, Value (HSV) color features on image of cocoa pods. This system can detect cocoa pods based on existing categories, but not all data is detected correctly because the accuracy is not too good. Based on the trials that have been carried out using the data distribution holdout method with a data sharing ratio of 70% for training data and 30% for testing data with a value of k on the K-NN which is varied, namely 5,7,11 and 13 for the best accuracy is the value of k = 5, which is 84.44%, calculated using the confusion matrix test method and in the K-fold cross validation test method, a trial is also carried out by dividing the data by 5 partitions with the k value of K-NN which varies the best accuracy is at the value of k = 5, which is 99, 33%.

Further research for development is to focus on:
(i) It is necessary to add training data for the category of pests and diseases of cocoa pods so that there are more types of cocoa pods so that it is hoped that the system built will be more accurate in detecting the image of the test data.

(ii) It is necessary to develop a detection system that is not only based on fruit, but a combination of fruit, leaves, and stems in order to provide information with a high level of confidence.

(iii) It is necessary to add a method of separating the object image of cocoa pods and the background (segmentation) so that when taking images of cocoa pods it is not necessary to use a white background, this can also increase the accuracy of the results of dataset detection; (iv) for the feature extraction method, the accuracy of the detection results can be improved by comparing the use of other feature extraction methods such as texture or shape characteristics; (v) need to add appropriate handling solutions to the system detection results; and (vi) for the classification algorithm, other classification algorithms can be added besides k-nearest neighbor.

Acknowledgment

The authors would like to thank the Faculty of Engineering, University of Tadulako, which has provided financial assistance in this research. This research was funded through the DIPA Research Fund, Faculty of Engineering, Fiscal Year 2022. The authors also thank several parties who supported the availability of data during the research. The parties involved are the Plantation and Livestock Service Office of Central Sulawesi Province, the Plantation Plant Protection Unit and the Sinar Tani Cocoa Farmer Group, Hamlet III RT 9, Potoya Village, Dolo District, Sigi Regency, Central Sulawesi Province.

Reference

[1] U. S. U. Syafrizal Lubis, "Identifikasi Desease pada Buah Kakao Menggunakan Hue Saturation Value dan Moment Invariant dengan Algoritma Backproga," 2018.
[2] G. H. Yogiswara, R. Magdalena, and H. F. T. S. Putra, "Identifikasi Jenis Desease Pada Kakao Dengan Pengolahan Citra Digital Dan K-nearest Neighbor," in *e-Proceeding of Engineering*, 2016, vol. 5, no. 1, pp. 371–377.
[3] K. T. Ikhsan Nur Rahmani, Danang Lelono, "Klasifikasi Kakao Berbasis e-nose dengan Metode Neuro Fuzzy," vol. 8, no. 1, 2018, doi: 10.22146/ijeis.25512.
[4] N. Wijaya and A. Ridwan, "Klasifikasi Jenis Buah Apel Dengan," *Sisfokom*, vol. 08, no. 1, pp. 74–78, 2019.
[5] F. R. Lestari *et al.*, "Deteksi Desease tanaman jeruk siam berdasarkan citra daun menggunakan segmentasi warna rgb-hsv," pp. 276–283, 2018.
[6] D. S. Tan, R. N. Leong, A. F. Laguna, and C. A. Ngo, "A Method for Detecting and Segmenting Infected Part of Cacao Pods," no. March, 2016.
[7] P. H. S. Lestari, Zeni Dwi, Nur Nafi'iyah, "Sistem Klasifikasi Jenis Pisang Berdasarkan Ciri Warna HSV Menggunakan Metode K-NN," pp. 11–15, 2019.
[8] H. Khotimah and M. Nafi, Nur, "Klasifikasi Kematangan Buah Mangga Berdasarkan Citra HSV dengan KNN," vol. 1, no. 2, pp. 4–7, 2019.
[9] E. Budianita, Jasril, and L. Handayani, "Implementasi Pengolahan Citra dan Klasifikasi K- Nearest Neighbour Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi," vol. 12, no. 2, pp. 242–247, 2015.
[10] R. K. N. C. H. K. Naveena, "Identification and Classification of Fruit Diseases," vol. 6, no. 7, pp. 11–14, 2016.
[11] F. F. Azizah, Laila Marifatul, Sitti Fadillah Umayah, "Deteksi Kecacatan Permukaan Buah Manggis Menggunakan Metode Deep Learning dengan Konvolusi Multilayer," vol. 21, no. 2, pp. 230–236, 2018, doi: 10.18196/st.212229.
[12] R. T. Basri, Harli, Indrabayu, Areni, Intan Sari, "Image Processing System for Early Detection of Cocoa Fruit Pest Attack Image Processing System for Early Detection of Cocoa Fruit Pest Attack," 2019, doi: 10.1088/1742-6596/1244/1/012003.
[13] S. Nurmuslimah, "Implementasi metode," vol. 2, no. 2, pp. 91–98, 2016.
[14] N. Harivinod, P. Pooja, N. K. H, B. S. Ashritha, and G. G. Hegde, "Cocoa Care - An Android Application for Cocoa Disease Identification," pp. 440–444.
[15] D. S. Gaikwad and K. J. Karande, "Image Processing Approach for Grading And Identification Of Diseases On Pomegranate Fruit : An Overview," vol. 7, no. 2, pp. 519–522, 2016.
[16] and E. S. W. U Lestari, R A Kumalasanti, "Identifying the Quality System of Cocoa Beans to Increase Productivity Using Backpropagation Neural Network Algorithm : A Case Study at Identifying the Quality System of Cocoa Beans to Increase Productivity Using Backpropagation Neural Network Algorithm :," 2019, doi: 10.1088/1742-6596/1413/1/012033.
[17] A. L. dan Y. Adhitya, "Classifying Physical Morphology of Cocoa Beans Digital Images using Multiclass Ensemble Least-Squares Support Vector Machine Classifying Physical Morphology of Cocoa Beans Digital Images using Multiclass Ensemble Least-Squares Support Vector Machine," 2018.
[18] S. A. Veites-campos and R. Ramírez-betancour, "Identification of Cocoa Pods with Image Processing and Artificial Neural Networks," no. 7, 2018.
[19] R. E. Angelia and N. B. Linsangan, "Fermentation Level Classification of Cross Cut Cacao Beans Using k-NN Algorithm."