1. Introduction
Indonesia is one of the agrarian countries where most of its population work in the agricultural sector. Various types of fruit and vegetables have been produced to meet people's needs and export to other countries, one of which is shallots. To increase the export of shallots, the quality must be chosen so that the driving force of Indonesian shallots products is maintained[1,2]. Current technological developments can facilitate human work, in this case, to classify objects. Before being classified, objects need to be identified based on color, shape, size, and texture. Object classification can be done using imagery, wherein this study the image of shallot is used to classify its quality. The quality of the shallots can be seen based on the color and size of the shallots.

Image classification methods have been widely studied in various previous studies [3–12]. Some classifiers that are widely used are Naïve Bayes(NB)[4,9,10], K-Nearest Neighbor (KNN) [3,13] and Support Vector Machine(SVM)[5,14]. Each classifier has advantages and disadvantages of each, where SVM is a classifier that has effective performance in high-dimensional space, can use different kernel functions, can use small samples and can classify non-linear data[5,15]. But SVM kernel selection must also be appropriate because each kernel will affect the classification results. KNN is the simplest classifier and can work well with relatively few parameters, where similar samples are located around the designated point, KNN requires relatively shorter computing time in the training phase but requires more time during clarification[16,17]. The use of KNN must determine the right k value because the k value will affect the classification results. Whereas the Naïve Bayes classifier is a classifier based on simple statistics and probabilities. In this method, all variables are assumed that each variable contributes to the classification and is interdependent (conditional dependency)[16,18]. The Naïve Bayes classifier uses the Bayes theorem which is widely used in many of the more sophisticated learning methods today.

The way of learning is by referring to the development of a Bayesian probabilistic model to get a posterior class from an instant, another advantage of Naïve Bayes is that the calculations can be done in...
parallel and distributed[19]. In its application in image classification also get relatively very good results [4,9,10]. Therefore in this research, the Naïve Bayes method is implemented to classify the quality of shallots based on their image. 

Before carrying out the classification process, digital images need to be extracted in their features. Feature extraction greatly affects the classification results, therefore it is necessary to normalize the data before the features are extracted. To classify the quality of shallots visually can be done based on the color and size. One good color model used is the Hue Saturation Value (HSV), in previous studies using this color model resulted in a classification accuracy of up to 100%[3]. In fact, the HSV color model is also in harmony with the color perception of human vision and has been widely used in various computer vision sciences [5,20,21]. While the classification is based on the size of the shallot based on the wide range and circumference as parameters that represent the size characteristics. To get a wide range and perimeter carried out morphological operations and calculations based on perimeter, area, metric, and eccentricity. The results of image measurement and color detection are used to classify shallots into good, medium and poor quality.

2. Proposed Method

The proposed method for classifying the quality of shallots is carried out through the following steps: Red green blue (RGB) image input, HSV color conversion and extraction, morphology and segmentation for area and circumference calculations, object labeling, area and circumference calculation, color detection, color percentage based on the Hue component, and quality predictions based on the quality predictions of the Naïve Bayes classifier. Each step is discussed in more detail below, and to see a clearer picture of the proposed method, see Fig. 1.

![Proposed Method Diagram](image-url)

**Figure 1.** Proposed Method.
2.1. Image Acquisition
At the image acquisition stage, data is collected based on the results of the image capture. The picture is taken in the morning between the hours of 8:00 to 09:00. The image is taken from a distance of 30 cm, where the white paper is used as a background to minimize lighting differences. Taken use is the rear camera of the Samsung Galaxy J7 Prime smartphone. Furthermore, the image is cropped and resized with dimensions of 512 × 512 pixels, with a 24-bit RGB color mode. The resulting image sample is presented in Fig. 2.

![Figure 2. Sample shallot image {a) Good quality; b) Medium quality; c) Poor quality}](image)

2.2. RGB to HSV Image
HSV wheel variations can be used to select the desired color range. The Hue component is represented by an inner circle of a wheel. The horizontal axis refers to Saturation, while Value is represented by the vertical axis. The HSV color model is used because there is a color similar to the color that is normally captured by the human eye.

In the HSV color model, each component has a different meaning. The Hue component represents the actual color values, such as green, yellow, red, purple. The saturation component represents the strength or purity of color. While the value represents the color brightness with a range from 0 - 100%. Convert RGB images to HSV images or vice versa using the theory contained in the research [22]. The results of image sample conversion are presented in Table 1.

| Color Channels | High-Quality Shallot | Medium Quality Shallot | Poor Quality Shallot |
|----------------|----------------------|------------------------|----------------------|
| Hue            | ![Hue](image)        | ![Hue](image)          | ![Hue](image)        |
| Saturation     | ![Saturation](image) | ![Saturation](image)   | ![Saturation](image) |
| Value          | ![Value](image)      | ![Value](image)        | ![Value](image)      |
2.3. Thresholding
Thresholding is a technique used to separate objects from their background. Thresholding is also one of the methods of image segmentation, the process of which is based on a gray image [13]. The thresholding process is also sometimes called binarization because thresholding can produce binary images. Thresholding can be done by Eq. 1.

\[
r(x, y) = \begin{cases} 
1 & \text{if } p(x,y) > T \\
0 & \text{if } p(x,y) \leq T 
\end{cases}
\] (1)

Where \( p \) is the pixel value designated \( x \) and \( y \) coordinates, while \( T \) is the threshold value. In this research, the threshold channel is the Saturation (S) channel. The results of the Thresholding process are presented in Fig. 3.

![Figure 3. Thresholding results \{(a) Good quality; (b) Medium quality; (c) Poor quality\}](image1)

2.4. Filling Holes and Opening
Filling Holes are used to fill all parts of the image with a value of 1 using guidelines based on the value of neighboring pixels. Filling Holes in the image are used to get solid object segments. While the opening is one of the commonly used morphological operations. The opening is an erosion process that is followed by a widening process, generally producing the effect of removing small and thin objects, can break objects at an oblique point, and remove thin bulges. The filling holes and opening process in the image are used to manage the segmentation results so that a better calculation is obtained. The results of the filling holes and opening stages in the image sample are presented in Fig. 4.

![Figure 4. Filling holes and opening results \{(a) Good quality; (b) Medium quality; (c) Poor quality\}](image2)

2.5. Object Labelling
After the whole process is done on the whole image, the labeling process is done on the whole image to mark the image class in the training process.
2.6. Size identification
Size identification is used to find out how big the size of the image object. Because in the classification of shallot quality, the size of the shallot is one of the factors in determining its quality. In this study, size identification using three parameters in determining the size of images of shallots is perimeter, area, and metric. Perimeter is the number of pixels located on the boundary object. The area is the number of pixels that make up a particular object. While the metric is the ratio of values between the area and the perimeter of the object. Metric values have vulnerable 0 to 1. Objects that have a spherical shape are metric values close to 1[13].

2.7. Color Identification
Color is one of the visual characteristics possessed by the object. Color is formed from a perfect spectrum of light, the level of brightness of light in colors can determine the color value. In this study, identification of the color of shallot images was carried out on the Hue component. The Hue component was chosen because it represents the true color value seen by human visuals. Therefore, in color identification, the average value of Hue is calculated for each shallot image and after that classifies the value into predetermined color classes.

2.8. Naïve Bayes Classifier
Naïve Bayes Classifier is a classification method that adopts Bayes theorem, which uses probability and statistics. Wherein said future predictions can be made based on past experience. Naïve Bayes Classifier also has special features that are strong assumptions of independence in any event or condition. According to Xhemali et al. [23], Naïve Bayes Classifier works very well compared to other classifier models because it has better accuracy. Some of the advantages of using the Naïve Bayes method are:

- Uses polynomial or binary data.
- This method allows for the implementation of various data sets.
- Do not use matrix calculations, numerical optimization, etc. so it is relatively easy to apply.

In this study, the classification of shallots quality is classified into three qualities, i.e. good, medium, and bad by looking at the image size which includes area, perimeter, metric and color image based on the Hue component value and the probability value of each shallot image.

3. Implementation and Testing
In this research, a total of 120 images of shallot were used, which consisted of 60 training images and 60 testing images. From 60 training images, there are 20 images of high quality, medium, and poor quality shallots. As for the testing images, each consists of 20 high quality, medium, and poor shallot images. Furthermore, the extraction of color and size features according to the proposed method, where the sample results are presented in Table 2.

| No | Image | Area | Perimeter | Metric | Hue | Class |
|----|-------|------|-----------|--------|-----|-------|
| 1  | ![Image](image1.png) | 3814 | 299.22    | 0.77   | 0.64| Good  |
| 2  | ![Image](image2.png) | 4294 | 289.03    | 0.80   | 0.57| Good  |
| 3  | ![Image](image3.png) | 4520 | 271.32    | 0.77   | 0.65| Good  |

|       |       |       |           |        |     |       |
|-------|-------|-------|-----------|--------|-----|-------|
|       |       |       |           |        |     |       |
|       |       |       |           |        |     |       |

Table 2. Area, Perimeter, Metric, Hue Feature Extraction Results from Training Images
Then the training is done using the data set above, then tested with the Naïve Bayes method with several steps such as:

1. Perform probability density testing probability data with the equation (2).
   \[ p_{\mu \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2) \]

2. Calculate the probability of each class that has obtained the value of the probability density based on the probability equation (3).
   \[ P = P(X|C_i) \times P(C_i) \quad (3) \]

3. After the testing process, searching for prediction then compare the value \( P(1|X_i) \), \( P(2|X_i) \), and \( P(3|X_i) \) such as the following rules:
   a. When \( P(X1_i|Y1_i) > P(X2_i|Y1_i) \), \( P(X1_i|Y1_i) > P(X3_i|Y1_i) \) or \( P(X1_i|Y4_i) > P(X2_i|Y4_i) \), \( P(X1_i|Y4_i) > P(X3_i|Y4_i) \) so it can be classified into the good quality of shallots.
   b. When \( P(X2_i|Y1_i) > P(X3_i|Y1_i) \), \( P(X2_i|Y1_i) < P(X1_i|Y1_i) \) or \( P(X2_i|Y2_i) > P(X3_i|Y2_i) \), \( P(X2_i|Y2_i) < P(X1_i|Y2_i) \) so it can be classified into the medium quality of shallots.
   c. When \( P(X3_i|Y1_i) < P(X1_i|Y1_i) \), \( P(X3_i|Y1_i) < P(X2_i|Y1_i) \) or \( P(X3_i|Y2_i) > P(X1_i|Y2_i) \), \( P(X3_i|Y2_i) < P(X2_i|Y2_i) \) so it can be classified into poor quality of shallots.

After training on the training image dataset, the testing process is then carried out on image testing based on the steps above. The results of calculating the testing image dataset are presented in Table 3.
The proposed method is 91.67%.

Furthermore, to calculate the accuracy used confusion matrix table in table 4. The results of classified correctly, 3 images are classified with medium quality and 2 images are classified with poor quality. Image of shallots were classified correctly and 5 other images were incorrect. Misclassification results are found in the image of shallots with good quality, from 20 images of good quality shallots. Only 15 images can be classified correctly, 3 images are classified with medium quality and 2 images are classified with poor quality. Furthermore, to calculate the accuracy used confusion matrix table in table 4. The results of the calculation of the confusion matrix presented in table 4. shows that the accuracy generated from the proposed method is 91.67%.

| Image Name | Target Class | Probability | Classification Results |
|------------|--------------|-------------|------------------------|
| G1         | M1           | 0.02423444  | 7.24202235 4.87135266 | Good       |
| G2         | M2           | 0.07385153  | 7.65309948 5.14427975 | Good       |
| G3         | M3           | 0.00021118  | 2.26728232 0.46856443 | Poor       |
| G4         | M4           | 0.00382527  | 3.87852053 0.78049415 | Medium     |
| G5         | M5           | 0.00049394  | 2.33740821 1.28994532 | Poor       |
| G6         | M6           | 0.00408053  | 3.96814760 2.44972946 | Medium     |
| G7         | M7           | 0.00400434  | 4.79117500 3.19371690 | Medium     |
| G8         | M8           | 0.02442344  | 7.24202235 4.87135266 | Good       |
| G9         | M9           | 0.00491327  | 5.78490525 4.50636253 | Good       |
| G10        | M10          | 0.00170086  | 7.18708233 5.14427975 | Good       |

| Image Name | Target Class | Probability | Classification Results |
|------------|--------------|-------------|------------------------|
| M1         | M1           | 0.00174516  | 4.21386315 2.29815012 | Medium     |
| M2         | M2           | 0.00181062  | 4.51161150 2.44972946 | Medium     |
| M3         | M3           | 0.00286412  | 4.80723020 2.69267465 | Medium     |
| M4         | M4           | 0.00302028  | 5.29193799 3.10163671 | Medium     |
| M5         | M5           | 0.00321579  | 5.68896657 3.19371690 | Medium     |
| M6         | M6           | 0.00436580  | 5.72280146 3.34875909 | Medium     |
| M7         | M7           | 0.00471943  | 5.97242665 3.51236415 | Medium     |
| M8         | M8           | 0.00115697  | 5.29193799 3.94672604 | Medium     |
| M9         | M9           | 0.00143121  | 5.69746434 4.23201363 | Medium     |
| M10        | M10          | 0.00174516  | 4.21386315 2.29815012 | Medium     |

| Image Name | Target Class | Probability | Classification Results |
|------------|--------------|-------------|------------------------|
| P1         | P1           | 0.00007446  | 0.76776636 0.46856443 | Poor       |
| P2         | P2           | 0.00014790  | 1.55462448 0.72775255 | Poor       |
| P3         | P3           | 0.00021551  | 1.85097675 0.84634347 | Poor       |
| P4         | P4           | 0.00025318  | 2.21126445 0.98416170 | Poor       |
| P5         | P5           | 0.00028874  | 2.24885774 1.25729462 | Poor       |
| P6         | P6           | 0.00031304  | 3.04097789 1.28994532 | Poor       |
| P7         | P7           | 0.00036129  | 3.40747000 2.14676565 | Poor       |
| P8         | P8           | 0.00021551  | 1.85097675 0.84634347 | Poor       |
| P9         | P9           | 0.00049394  | 2.26728232 0.84634347 | Poor       |
| P10        | P10          | 0.00021118  | 0.51649703 0.84634347 | Poor       |

Based on the data presented in table 3, it appears that from the classification of 60 test images, 55 images were classified correctly and 5 other images were incorrect. Misclassification results are found in the image of shallots with good quality, from 20 images of good quality shallots. Only 15 images can be classified correctly, 3 images are classified with medium quality and 2 images are classified with poor quality. Furthermore, to calculate the accuracy used confusion matrix table in table 4. The results of the calculation of the confusion matrix presented in table 4. shows that the accuracy generated from the proposed method is 91.67%.
Table 4. Confusion Matrix Results

| Actual Class | Result Class | Good | Medium | Poor |
|--------------|--------------|------|--------|------|
|              | Good         | 15   | 3      | 2    |
|              |              | 25.0%| 5.0%   | 3.3% |
|              | Medium       | 0    | 20     | 0    |
|              |              | 0%   | 33.3%  | 0%   |
|              | Poor         | 0    | 0      | 20   |
|              |              | 0%   | 0%     | 33.3%|
|              |              | 100% | 86.96% | 90.91%|
|              |              | 0.0% | 13.04% | 9.09%|
|              |              |      |        | 91.67%|
|              |              |      |        | 8.33%|

4. Conclusions
This research proposes a method of classifying the quality of shallots using the Naïve Bayes classifier based on HSV color extraction and calculation of object size based on metric, perimeter, and area. Channel hue is used as a benchmark for calculating the quality of shallots based on their color, while for calculating the size of shallots a saturation channel is used. Before the shallot size is calculated, several morphological operations such as thresholding, filling holes and opening are performed to improve the accuracy of the calculation of the object's distribution. The choice of HSV color models and color channels is adjusted to the theory of human vision. Furthermore, the results of all calculations are calculated and then the quality is determined based on the rules that have been explained in the implementation section. Based on the results of trials conducted in this research, the proposed method produces an accuracy of 91.67%.

5. References

[1] Rahmadona L, Fariyanti A and, B 2017 Dayasaing Komoditas Bawang Merah di Kabupaten Majalengka, Jawa Barat J. Hortik. Indones. 8 128
[2] Kustiari R 2018 Perilaku Harga dan Integrasi Pasar Bawang Merah di Indonesia J. Agro Ekon. 35 77
[3] Indriani O R, Kusuma E J, Sari C A, Rachmawanto E H and Setiadi D R I M 2018 Tomatoes classification using K-NN based on GLCM and HSV color space Proceedings - 2017 International Conference on Innovative and Creative Information Technology: Computational Intelligence and IoT, ICITech 2017 vol 2018-Janua (Salatiga: IEEE) pp 1–6
[4] Kusuma A, Setiadi D R I M and Putra M D M 2018 Tomato Maturity Classification using Naive Bayes Algorithm and Histogram Feature Extraction J. Appl. Intell. Syst. 3 39–48
[5] Fu L, Duan J, Zou X, Lin G, Song S, Ji B and Yang Z 2019 Banana detection based on color and texture features in the natural environment Comput. Electron. Agric. 167
[6] Yang X, Zhang R, Zhai Z, Pang Y and Jin Z 2019 Machine learning for cultivar classification of apricots (Prunus armeniaca L.) based on shape features Sci. Hortic. (Amsterdam). 256
[7] Hameed K, Chai D and Rassau A 2018 A comprehensive review of fruit and vegetable classification techniques Image Vis. Comput. 80 24–44
[8] Mohanty N, John A L S, Mamatha R and Rath T M 2013 Shape-based image classification and retrieval Handbook of Statistics vol 31 (Elsevier B.V.) pp 249–67
[9] Luoyong N, Jiaming H, Hongbo H, Bishi H and Zhe X 2015 The automatic classification research to medical imaging disease based on Naïve nayesion algorithm Proceedings - 2014 10th International Conference on Computational Intelligence and Security, CIS 2014 (Kunming: Institute of Electrical and Electronics Engineers Inc.) pp 308–11
[10] Zaw H T, Maneerat N and Win K Y 2019 Brain tumor detection based on Naïve Bayes classification Proceeding - 5th International Conference on Engineering, Applied Sciences and Technology, ICEAST 2019 (Luang Prabang: Institute of Electrical and Electronics Engineers Inc.)

[11] Saleem G, Akhtar M, Ahmed N and Qureshi W S 2019 Automated analysis of visual leaf shape features for plant classification Comput. Electron. Agric. 157 270–80

[12] Siedliska A, Baranowski P, Zubik M, Mazurek W and Sosnowska B 2018 Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging Postharvest Biol. Technol. 139 115–26

[13] Sari C A, Kuncoro M W, Setiadi D R I M and Rachmawanto E H 2018 Roundness and eccentricity feature extraction for Javanese handwritten character recognition based on K-nearest neighbor 2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018 (Institute of Electrical and Electronics Engineers Inc.) pp 5–10

[14] Rahmayuna N, Rahardwika D S, Sari C A, Setiadi D R I M and Rachmawanto E H 2018 Pathogenic Bacteria Genus Classification using Support Vector Machine 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (Yogyakarta: IEEE) pp 23–7

[15] Kavitha J C and Suruliandi A 2016 Texture and color feature extraction for classification of melanoma using SVM International Conference on Computing Technologies and Intelligent Data Engineering (Kovilpatti: IEEE) pp 1–6

[16] Jadhav S D and Channe H P 2016 Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques Int. J. Sci. Res. 5 1842–1842

[17] Enriko I K A, Suryanegara M and Gunawan D 2016 Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient’s Health Parameters J. Telecommunication, Electron. Comput. Eng. 8 59–65

[18] Jiang L, Zhang L, Yu L and Wang D 2019 Class-specific attribute weighted naive Bayes Pattern Recognit. 88 321–30

[19] Berrar D 2019 Bayes’ Theorem and Naive Bayes Classifier Encyclopedia of Bioinformatics and Computational Biology (Elsevier) pp 403–12

[20] Hamuda E, Mc Ginley B, Glavin M and Jones E 2017 Automatic crop detection under field conditions using the HSV colour space and morphological operations Comput. Electron. Agric. 133 97–107

[21] Miao R H, Tang J L and Chen X Q 2015 Classification of farmland images based on color features J. Vis. Commun. Image Represent. 29 138–46

[22] Smith A R 1978 Color Gamut Transform Pairs Comput Graph (ACM) vol 12 (New York: SIGGRAPH-ACM) pp 12–9

[23] Xhemali D, Hinde C J and Stone R 2009 Naïve Bayes vs. Decision Trees vs. Neural Networks in the classification of training web pages