

Tomato Maturity Classification using Naive Bayes Algorithm and Histogram Feature Extraction

Arya Kusuma*, **De Rosal Ignatius Moses Setiadi**

Department of Informatics Engineering, Dian Nuswantoro University, Semarang, Indonesia

E-mail: aryakusuma12@gmail.com, moses@dsn.dinus.ac.id

**Corresponding author*

M. Dalvin Marno Putra

Faculty of Electronic and Information Engineering, South China University of Technology, Guangzhou, China

E-mail: mrdalvin@yahoo.com

Abstract - Tomatoes have nutritional content that is very beneficial for human health and is one source of vitamins and minerals. Tomato classification plays an important role in many ways related to the distribution and sales of tomatoes. Classification can be done on images by extracting features and then classifying them with certain methods. This research proposes a classification technique using feature histogram extraction and Naïve Bayes Classifier. Histogram feature extractions are widely used and play a role in the classification results. Naïve Bayes is proposed because it has high accuracy and high computational speed when applied to a large number of databases, is robust to isolated noise points, and only requires small training data to estimate the parameters needed for classification. The proposed classification is divided into three classes, namely raw, mature and rotten. Based on the results of the experiment using 75 training data and 25 testing data obtained 76% accuracy.

Keywords – tomato, image classification, Naïve Bayes, histogram, feature extraction

1. INTRODUCTION

Tomatoes are native plants from Central America, South America, Mexico to Peru. Tomatoes come from the highlands of the west coast of South America. Tomatoes have a short life cycle and can grow as high as 1-3 meters [1]. Tomatoes can have green, yellow, and red colors that can be used as vegetables in cooking or eaten directly. Tomato fruit is a fruit that has high nutrition, in addition to being consumed as fresh fruit, but also used as flavoring ingredients and ingredients in the food and beverage industry. Delivery of tomatoes between traders can also affect the maturity of tomatoes because tomatoes have a short life cycle [2].

Along with the development of information technology, it is possible to identify fruit maturity with the computers help [3] [4]. Identification can be done by classifying the image of tomatoes with various methods such as K-Nearest Neighbor (KNN) [4] [5], Random Forest [6], Support Vector Machine(SVM) [7] [8], Naïve Bayes [7] [8] [9] [10] [11] [12] [13], etc. In this study, identification of maturity in a set of tomatoes using the Naïve Bayes algorithm. Naïve Bayes was chosen because the Naïve Bayes algorithm has high accuracy and speed when applied to a large number of databases, is robust to isolated noise points, and only requires small training data to estimate the parameters needed for classification.

But the Naïve Bayes algorithm also has a weakness, that is if the conditional probability is zero, then the probability of prediction will be zero. Therefore, to increase the

accuracy, the histogram feature extract is added. Plus the feature extraction histogram in order to know the pixel intensity values of the image, the frequency of occurrence of a relative, brightness, contrast, etc [14]. Then the Naïve Bayes algorithm will study the values that have been obtained from the histogram feature extraction and can determine the maturity level of tomatoes more accurately.

2. RESEARCH METHOD

In this research, tomato image classification is proposed based on its maturity using GLCM feature extraction and Naïve Bayes Classifier. Before the image is classified, several stages are carried out, namely tomato image data collection and acquisition, image preprocessing with the function of resizing and cropping, image color conversion, feature extraction, classification and testing described in Figure 1.

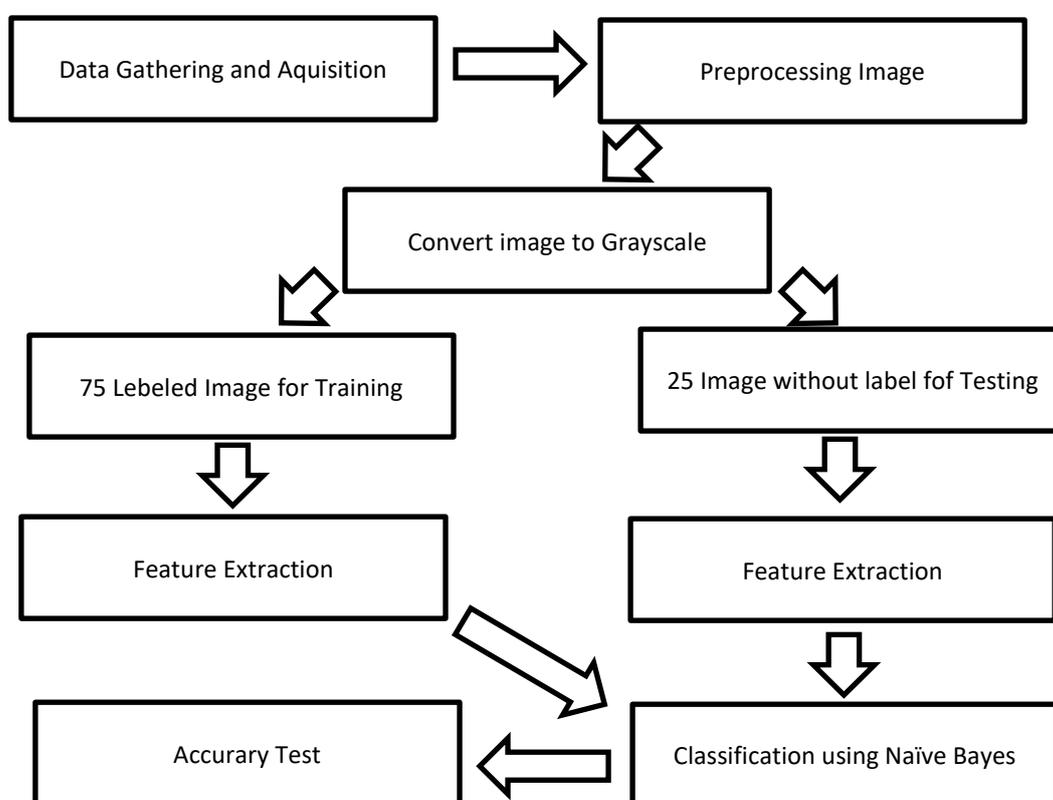


Figure 1. Proposed Method

2.1. Data Gathering and Aquisition

In this study tomato image retrieval data was taken directly using the camera of the ASUS smartphone with a Z00AD camera model. Image data used in this study were 100 tomato images with a size of 4096 * 3072, consisting of 34 raw tomatoes, 33 mature tomatoes, and 33 rotten tomatoes. Furthermore, image data is divided into 75 training data and 25 testing data. Figure 2 is the sample image of tomatoes used.

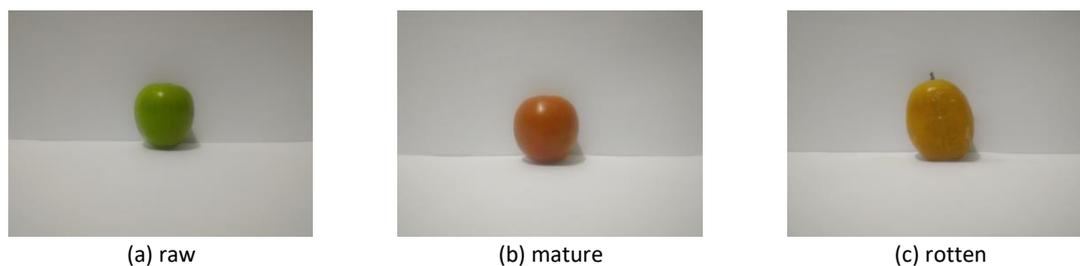


Figure 2. Sample Original Tomato Image

2.2. Pre-processing

At this stage, the image data to be classified is done by cropping and resizing to a size of 100 x 100 pixels, with the aim to normalize the image and speed up the computation process at the training and testing stages. Figure 3 is a sample of preprocessing images.



Figure 3. Sample Tomato Image after preprocessing

2.3. Convert Image to Grayscale

After the preprocessing stage is carried out, then the image color conversion from RGB to Grayscale is done using the formula (1).

$$gray = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

Where: gray is a grayscale channel, R is a red channel, G is a green channel, and B is a blue channel

2.4. Histogram Feature Extraction

A histogram is a method for obtaining a texture with a base on the histogram. An image histogram is a graph that describes the distribution of pixel intensity values of an image or a particular part of an image. In terms of histograms, it can be seen that the frequency of occurrence of a relative of an intensity in the image [15]. The histogram method is a static method of the first order to get texture features. The features contained in the histogram are mean intensity, standard deviation, energy, entropy, smoothness, and skewness [16].

The first feature is the mean intensity that can be calculated by the formula (2).

$$m = \sum_{i=0}^{L-1} i \cdot p(i) \quad (2)$$

Where m is the average value of intensity, which is obtained from the results of multiplication i with p (i). Where i is the value of the grayness of an image and p (i) is the probability of the appearance of the value i.

The second feature is the standard deviation, this feature shows the size of the contrast of an image. The results of the standard deviation are obtained from the results of the reduction i with m squared and multiplied by p (i) after that the new results at the root. Where i is the gray level value of an image, p (i) is the probability of the occurrence of i and m values as a result of mean intensity, which can be calculated by the formula (3).

$$sd = \sqrt{\sum_{i=0}^{L-1} (i - m)^2 \cdot p(i)} \quad (3)$$

The third feature is energy, the value of energy which is often called the value of uniformity in an image has a maximum value of 1. Images that have a lot of gray level values

will have a little energy value compared to images that have a little gray level will have more energy values. From the above calculation, get the energy value. Energy results are obtained from the rank 2 result of the probability appearance i . Energy features can be calculated by the formula (4).

$$energi = \sum_{i=0}^{L-1} [p(i)]^2 \quad (4)$$

The fourth feature is entropy, where entropy shows the complexity of an image, so that the higher the entropy value of the image, the more complex the image is. The result of entropy is obtained from the multiplication of $p(i)$ with \log_2 and multiplied by $p(i)$. Where $p(i)$ is the prominence of the occurrence of values i . Entropy can be calculated by the formula (5).

$$entropi = \sum_{i=0}^{L-1} p(i) \log_2(p(i)) \quad (5)$$

Entropy shows the complexity of an image so that the higher the entropy value of the image the more complex the image is. The result of entropy is obtained from the multiplication of $p(i)$ with \log_2 and multiplied by $p(i)$. Where $p(i)$ is the prominence of the occurrence of values i .

The fifth feature is smoothness. The smoothness value is obtained from the result of a reduction in the value of 1 with 1 divided by 1 and added with the standard deviation value that is squared. A high smoothness value indicates that the image has a smooth intensity. Smoothness can be calculated by the formula (6).

$$smoothness = 1 - \frac{1}{1 + deviasi^2} \quad (6)$$

The sixth feature is skewness, skewness is often referred to as a third-order moment in which the negative value of the brightness distribution is left-leaning towards the mean and the positive value states that the brightness distribution is right-leaning towards the average. The skewness results are obtained from the result of the reduction in the value of i with the m raised by 3 and multiplied by $p(i)$. Where i is the gray level value of an image, m is the average intensity and $p(i)$ is the probability of the occurrence of values i . Skewness can be calculated by the formula (7).

$$skewness = \sum_{i=0}^{L-1} (i - m)^3 \cdot p(i) \quad (7)$$

2.4. Naïve Bayes Classifier

The Naïve Bayes algorithm is a classification algorithm that requires the learning process first. This algorithm predicts future opportunities based on past experience. The Naïve Bayes algorithm uses probability methods and simple statistics by assuming that one class with the other class is independent [17]. Naïve Bayes has two main processes, namely the learning process and testing [18]. The learning process uses existing data to estimate the probability distribution parameters, where it has been assumed that there is independence from each class. The data testing process uses a model that has been built into the learning process to calculate posterior opportunities and then classifies it into the largest posterior opportunity [17]. The advantage in using the Naïve Bayes algorithm is (a) it does not need to use numerical optimization, matrix, and others so that it is easier (b) the training process and more efficient use (c) can use binary or polynomial data (d) can be implemented with various kinds dataset because it has been assumed to be independent.

The general form of the naïve Bayes theorem for categorical data can be calculated by the formula (8) [10].

$$P(X|Y) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (8)$$

Where:

$P(X|Y)$: the probability of data with X vector in class Y

- $P(Y)$: the initial probability of class X
 $P(X_i|Y)$: the independent probability of Y class from features in X vector
 $P(X)$: the probability of X

3. RESULTS AND DISCUSSION

Before performing the testing process on the test image data, the input image is first entered into the pre-processing and feature extraction stages. From the results of the test image processing, the classification procedure is then performed using the Naive Bayes algorithm calculation. The stages of the classification process can be explained as follows:

1. Calculate Prio probability $P(C_i)$ training data for each class. $P(C_i)$ is the probability value of the amount of training data per class with the entire data. In this study three classes of image types are used, namely raw, mature, rotten which each contains 25 data so that the value of $P(C_i)$ of each class can be calculated as follows:

$$P(C_i = \text{"raw"}) = 25 / 75 = 0,33$$

$$P(C_i = \text{"mature"}) = 25 / 75 = 0,33$$

$$P(C_i = \text{"rotten"}) = 25 / 75 = 0,33$$

2. Calculate the mean (μ) and standard deviation (σ) of each attribute in the training data based on the class. To calculate the mean (μ) equation (9) and standard deviation (σ) are used equation (10).

$$\mu = \frac{\sum_{i=1}^n X^i}{n} \quad (9)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n-1}} \quad (10)$$

Where

X = attribute value in the training data used,

n = the value of the total training data for each class

Based on the data in the training database and equations (9) and (10) the values of μ and σ for each attribute of the training data based on class can be calculated as follows:

$$\mu_{\text{raw}} = (143.7128 + 147.8919 + \dots + 145.2599) / 25 = 148.8326$$

$$\sigma_{\text{raw}} = \sqrt{(1/24(143.7128 - 148.8326)^2 + (147.8919 - 148.8326)^2 + \dots + (145.2599 - 148.8326)^2)} = 2.62248180919509$$

$$\mu_{\text{mature}} = (151.4861 + 150.0413 + \dots + 149.4776) / 25 = 147.623464$$

$$\sigma_{\text{mature}} = \sqrt{(1/24(151.4861 - 147.623464)^2 + (150.0413 - 147.623464)^2 + \dots + (149.4776 - 147.623464)^2)} = 5.13543829025452$$

$$\mu_{\text{rotten}} = (153.0672 + 151.6064 + \dots + 144.7321) / 25 = 150.393912$$

$$\sigma_{\text{rotten}} = \sqrt{(1/24(151.4861 - 150.393912)^2 + (150.0413 - 150.393912)^2 + \dots + (149.4776 - 150.393912)^2)} = 3.82641955860801$$

Furthermore, the calculation is performed on all the attributes of the training data based on each class. As for the overall results of the calculation of μ and σ , each attribute of the training data based on rotten, mature and raw classes can be seen in tables 1, 2, and 3.

Table 1. The values of μ and σ for each attribute of training data are based on rotten class

No	Mean intensity	energy	entropy	contrast	smoothness	skewness
1	175.8689	0.01770976	6.1207290545	25.227412	0.998431181	-1.70986044
2	176.5476	0.01827368	6.1150787328	25.325306	0.998443267	-1.78535384
3	167.2656	0.02038912	5.9428218563	21.033465	0.997744734	-1.53933269
-	-	-	-	-	-	-

-	-	-	-	-	-	-
-	-	-	-	-	-	-
25	167.9758	0.0171193	6.1329027119	24.22888	0.998299432	-1.45661311
μ	170.01818	0.0188395	6.0554905302	23.260243	0.998116285	-1.641503297
σ	3.9701639	0.0011124	0.0785830297	1.3820489	0.000225584	0.107169019

Table 2. The values of μ and σ for each attribute of training data are based on mature class

No	Mean intensity	energy	entropy	contrast	smoothness	skewness
1	175.8689	0.01770976	6.1207290545	25.227412	0.998431181	-1.70986044
2	176.5476	0.01827368	6.1150787328	25.325306	0.998443267	-1.78535384
3	167.2656	0.02038912	5.9428218563	21.033465	0.997744734	-1.53933269
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
25	167.9758	0.0171193	6.1329027119	24.22888	0.998299432	-1.45661311
μ	170.01818	0.0188395	6.0554905302	23.260243	0.998116285	-1.641503297
σ	3.9701639	0.0011124	0.0785830297	1.3820489	0.000225584	0.107169019

Table 3. The values of μ and σ for each attribute of training data are based on the raw class

No	Mean Intensity	energy	entropy	contrast	smoothness	skewness
1	159.7868	0.01691892	6.1775189564	23.08702	0.998127377	-1.37964629
2	159.9474	0.01694832	6.1400599437	23.13551	0.998135205	-1.28712746
3	173.6774	0.01741918	6.1475969455	27.20166	0.998650346	-1.71182209
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
25	160.2225	0.01718214	6.1383776899	22.90605	0.998097728	-1.31483781
μ	176.55406	0.01644402	6.2515778595	28.23254	0.998733787	-1.79026506
σ	4.1045704	0.00072188	0.0516634745	1.226313	0.000103886	0.07994722

The calculation results of the μ and σ values will then be used in the third stage to calculate $P(X_k.uji | C_i)$ for each class.

- The third stage calculates $P(X_k.uji | C_i)$ for each attribute of test data based on class. Based on the calculation of the mean (μ) and standard deviation (σ) each attribute of the training data is based on the class and the results of the extraction of mature image features that have been done before, then the value of $P(X_k.uji | C_i)$ of each attribute of test data based on class can be calculated using equation (11):

$$P(X_k.uji | C_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2(\sigma^2)}} \quad (11)$$

Where,

σ = the result of calculation σ each attribute of the training data is based on the class that has been done before.

μ = the result of the calculation of μ each attribute of the training data is based on the class that has been done before.

$$\Pi = 3.14 \quad e = 2.718282$$

The calculation results of $P(X_k.uji | C_i)$ above can be seen in table 4.

Table 4. Calculation results of P (Xk.uji | Ci)

	Raw Tomato Image	Mature Tomato Image	Rotten Tomato Image
Mean Intensity	0.0189851750740846	0.0252164463524086	0.0167631530520624
Energy	1.18195007020561	8.37367179010836e-13	1.16173005020341
Entropy	0.0928787047429147	6.60710359649696e-05	0.0726565025227125
Contrast	0.0197430373740571	0.0521668538149243	0.0175210151520351
Smoothness	215.614886690774	1100.98466842882	113.412664470552
Skewness	1.01559870429582	0.860310235334722	1.01337650227362

4. The fourth stage of the probability P (Ci | Xk.uji) posterior count for each class P (Ci | Xk.uji) is the probability of the Ci hypothesis based on the condition X test data. From the stage that has been done before, the value of P (Ci | Xk.uji) for each class can be calculated using equation (12).

$$P(Ci|Xk.uji) = \frac{P(Xk.uji|Ci)P(Ci)}{P(Xk.uji)} \quad (12)$$

Where:

P (Ci|Xk.uji) = probability Ci hypothesis based on the condition X test data (posterior probability)

P (Xk.uji | Ci) = probability X test data based on conditions in the Ci hypothesis

P (Ci) = probability hypothesis Ci (prior probability)

P (Xk.uji) = probability from X test data

Calculations with equation (12) are described as follows:

$$P(Ci = \text{"raw"}|Xk.uji) = 0.0189851750740846 \times 1.18195007020561 \times 0.0928787047429147 \times 0.0197430373740571 \times 215.614886690774 \times 1.01559870429582 = 0.0090104$$

$$P(Ci = \text{"mature"}|Xk.uji) = 0.0252164463524086 \times 8.37367179010836e-13 \times 6.60710359649696e05 \times 0.0521668538149243 \times 1100.98466842882 \times 0.860310235334722 = 6.8935e-17$$

$$P(Ci = \text{"rotten"}|Xk.uji) = 0.0167631530520624 \times 1.16173005020341 \times 0.0726565025227125 \times 0.0175210151520351 \times 113.412664470552 \times 1.01337650227362 = 0.002849227124$$

5. The last step is to look for the probability P (Ci | Xk.uji) posterior highest value of each class. This final step will look for the highest value from the calculation of P (Ci | Xk.uji) for each class. The results of this stage will then be used as a determinant of class predictions for a data entry.

$$P(Ci = \text{"raw"}|Xk.uji) = 0.0090104$$

$$P(Ci = \text{"mature"}|Xk.uji) = 6.8935e-17$$

$$P(Ci = \text{"rotten"}|Xk.uji) = 0.002849227124$$

From the results above it can be seen that the highest P (Chi | Xk.uji) value of each class is 0.0090104 which is owned by the posterior probability P (Ci = "raw" | X). These results indicate that the test image is classified into the raw type class.

The same classification calculation was carried out on 25 training images to test the correctness of the system. The following results of the overall classification of 25 test images using the Naïve Bayes algorithm are shown in table 5.

Table 5 Classification results of data testing

Image	Posterior Probability Raw	Posterior Probability Mature	Posterior probability Rotten	Results	Target	True / False
Tomato1.jpg	35823.0221	0.0024995	10487.9428	Mature	Rotten	False
Tomato2.jpg	6688.3109	540.7488	2225.1225	Rotten	Rotten	True
Tomato3.jpg	208793.8158	0.010599	54397.7039	Rotten	Rotten	True
Tomato4.jpg	634.3044	5.6211e-20	9698.8758	Rotten	Rotten	True

Tomato5.jpg	0.9156	5.0507e-37	81.1582	Rotten	Rotten	True
Tomato6.jpg	2375.6269	1.0446e-17	23207.9649	Rotten	Rotten	True
Tomato7.jpg	4181.8946	3.2781e-09	29315.7475	Mature	Rotten	False
Tomato8.jpg	9.3408	4.8781e-19	362.2356	Mature	Rotten	False
Tomato9.jpg	137989.0808	0.54743	14015.1328	Rotten	Mature	False
Tomato10.jpg	179350.1484	0.0093558	42685.4361	Rotten	Mature	False
Tomato11.jpg	0.00066246	317108.1508	0.12965	Mature	Mature	True
Tomato12.jpg	2.5076	302560.6264	4.2767	Mature	Mature	True
Tomato13.jpg	0.00046026	535648.7303	0.26635	Mature	Mature	True
Tomato14.jpg	0.00042504	418328.7556	0.87299	Mature	Mature	True
Tomato15.jpg	1.6446e-06	344605.8458	0.0010426	Mature	Mature	True
Tomato16.jpg	1.1147e-10	35355.8793	2.0806e-06	Mature	Mature	True
Tomato17.jpg	72286.4936	1.1561	12426.0246	Mature	Raw	False
Tomato18.jpg	30.1574	6.6712e-26	0.63273	Raw	Raw	True
Tomato19.jpg	1.8677	4.193e-12	0.032477	Raw	Raw	True
Tomato20.jpg	0.089203	3.5512e-37	7.1966e-05	Raw	Raw	True
Tomato21.jpg	0.0018123	1.1686e-13	0.00010342	Raw	Raw	True
Tomato22.jpg	7511.6832	1.6654e-13	31538.1921	Raw	Raw	True
Tomato23.jpg	45800.192	3.3802e-10	65809.3251	Raw	Raw	True
Tomato24.jpg	212.0307	65890.5947	75.9584	Raw	Raw	True
Tomato25.jpg	3568.3469	15515.9899	526.258	Raw	Raw	True

6. From table 5, it can be concluded that the testing of the testing data above as many as 25 tomatoes were tested consisting of 9 raw tomatoes, 8 cooked tomatoes, and 8 rotten tomatoes. Then the accuracy can be calculated as follows:

$$\frac{19}{25} \times 100\% = 76\%$$

Based on the results of accuracy calculations, obtained an accuracy value of 76% in the classification of the maturity of tomatoes using the Naïve Bayes method. Based on experimental results in this study, accuracy is still not better than similar studies as in research [18] [17]. So in the next research, it is necessary to do some experiments using different features or better preprocessing stages to improve the accuracy of tomato image classification.

4. CONCLUSION

Based on the results of research on the Classification of Tomato Fruit Maturity Level Using the Naïve Bayes Algorithm and Histogram Feature Extraction, it can be concluded that the naïve Bayes algorithm can be implemented to classify the image of tomatoes as raw, ripe, and rotten. From the results of the experiments carried out and calculating the accuracy that the results of the naïve Bayes algorithm for tomato images produce an accuracy rate of 76%.

REFERENCES

- [1] A. F. Smith, *The Tomato in America*, University of Illinois Press, 1994.
- [2] S. R. Rupanagudi, B. Ranjani, P. Nagaraj and V. G. Bhat, "A cost effective tomato maturity grading system using image processing for farmers," in *International Conference on Contemporary Computing and Informatics (IC3I)*, Mysore, 2014.
- [3] V. Pavithra, R. Pounroja and B. S. Bama, "Machine vision based automatic sorting of

- cherry tomatoes," in *International Conference on Electronics and Communication Systems (ICECS)*, Coimbatore, 2015.
- [4] O. R. Indriani, E. J. Kusuma, C. A. Sari, E. H. Rachmawanto and D. R. I. M. Setiadi, "Tomatoes classification using K-NN based on GLCM and HSV color space," in *International Conference on Innovative and Creative Information Technology (ICITech)*, Salatiga, 2017.
 - [5] T. Sutojo, P. S. Tirajani, D. R. I. M. Setiadi, C. A. Sari and E. H. Rachmawanto, "CBIR for classification of cow types using GLCM and color features extraction," in *International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, 2017.
 - [6] W. Man, Y. Ji and Z. Zhang, "Image classification based on improved random forest algorithm," in *International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, 2018.
 - [7] A. Krishna, D. Edwin, and S. Hariharan, "Classification of liver tumor using SFTA based Naïve Bayes classifier and support vector machine," in *International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*, Kannur, 2017.
 - [8] G. Karthick and R. Harikumar, "Comparative performance analysis of Naive Bayes and SVM classifier for oral X-ray images," in *International Conference on Electronics and Communication Systems (ICECS)*, Coimbatore, 2017.
 - [9] X. Deng, J. Guo, Y. Chen and X. Liu, "A Method for Detecting Document Orientation by Using Naïve Bayes Classifier," in *International Conference on Industrial Control and Electronics Engineering*, Xi'an, 2012.
 - [10] M. A. Jabbar and S. Samreen, "Heart disease prediction system based on hidden naïve Bayes classifier," in *International Conference on Circuits, Controls, Communications and Computing (IAC)*, Bangalore, 2016.
 - [11] K. Baati, T. M. Hamdani, and A. M. Alimi, "Diagnosis of Lymphatic Diseases Using a Naive Bayes Style Possibilistic Classifier," in *IEEE International Conference on Systems, Man, and Cybernetics*, Manchester, 2013.
 - [12] J. Polpinij and C. Sibunruang, "Thai heritage images classification by Naïve Bayes image classifier," in *International Conference on Digital Content, Multimedia Technology and its Applications*, Seoul, 2010.
 - [13] S.-C. Hsu, I.-C. Chen and C.-L. Huang, "Image classification using pairwise local observations based Naive Bayes classifier," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Hong Kong, 2015.
 - [14] A. Z. Ichsan, A. and D. Yendri, "Perancangan dan Pembuatan Sistem Visual Inspection Sebagai Seleksi Buah Tomat Berdasarkan Kematangan Berbasis Web Camera," p. 9, 2013.
 - [15] F. Yan, M. Hamit, A. Kutluk, C. Yan, L. Li, W. Yuan and D. Kong, "Feature extraction and analysis on X-ray image of Xinjiang Kazak Esophageal cancer by using gray-level histograms," in *IEEE International Conference on Medical Imaging Physics and Engineering*, Shenyang, 2013.
 - [16] A. Ciputra, D. R. I. M. Setiadi, E. H. Rachmawanto and A. Susanto, "Klasifikasi Tingkat Kematangan Buah Apel Manalagi dengan Algoritma Naive Bayes dan Ekstraksi Fitur Citra Digital," *Simetris*, vol. 9, no. 1, pp. 465-471, 2018.
 - [17] F. Y. Manik and K. S. Saragih, "Klasifikasi Belimbing Menggunakan Naive Bayes Berdasarkan Fitur Warna RGB," *Indonesian Journal of Computing and Cybernetics Systems*, vol. 11, no. 1, p. 10, 2017.

- [18] A. Ghofur, "Implementasi Metode Klasifikasi Naive Bayes Untuk Memprediksi Kualitas Cabai," *Jurnal Ilmiah Informatika*, vol. 1, no. 1, p. 7, 2016.