

Detection of Formalin Cob Fish using an Android-Based Naive Bayes Classifier Method based on Fish Eye Image

Rudy Ansari ^{a,1,*}, Yuslena Sari ^{b,2}

^a Informatics Study Program, Faculty of Engineering, Universitas Muhammadiyah Banjarmasin, Banjarmasin, Indonesia

^b Information Technology Study Program, Faculty of Engineering, Universitas Lambung Mangkurat, Banjarmasin, Indonesia

¹ rudy@umb.ac.id *; ² yuzlena@ulm.ac.id

* Corresponding Author

ABSTRACT

Tuna is a popular fish among Indonesians, but formalin is frequently used as a preservative because the freshness of the fish does not last long. It is difficult for the community to distinguish between formalin and non-formalin fish due to the widespread distribution of formalin fish. The Center for Drug and Food Control (BBPOM) in Banjarmasin still requires samples to be taken to the laboratory to be tested for formalin content, which takes about a day to complete. It is believed that a technology that can identify formalin tuna would assist the community in solving this problem. The HSV technique was utilized in this study, which involved looking at the fish's eye colour and classifying them using the Naive Bayes Classifier method. Based on the testing conducted on the formalin tuna identification application based on fisheye pictures utilizing the Android-based Naive Bayes Classifier technique, it was determined that the test results had an accuracy of 80%.

KEYWORDS

Formalin
Hsv
Naive Bayes Classifier
Android
Tuna Fish



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

1. Introduction

The Indonesian people eat a lot of mackerel, which is a product of saltwater fisheries. Tuna fish has a great nutritional value, a wonderful flavour, and soft, thick flesh. Furthermore, the price of tuna is lower, making it one of the most popular fish among Indonesians [1]. Tuna is the fish with the highest production level among other marine fish species in Banjarmasin City, according to data from the Central Statistics Agency of Banjarmasin City in 2014 [2]. Other saltwater fish species in Banjarmasin City include snapper and pomfret mackerel, mackerel, and baronang fish.

Implementation of logistic regression to detect android-based formalin fish based on image and physical characteristics of fish, research by Alfian et al., acquired an accuracy of 86% of 100 data in the system testing procedure. Furthermore, the findings of data trials done directly in several traditional marketplaces in Malang revealed that the eight samples of tuna fish utilized had a high degree of accuracy become 100% [3]. In addition, Hadini et al. published an article titled android-based formalin detection system for milkfish (chanos) based on eye image using the naive Bayes classifier method. The fisheye image training data obtained the highest accuracy, namely 100%, by changing the size to 10x10. The fisheye picture testing results, on the other hand, had a high accuracy of 98.35%. The data testing is done directly, namely in the traditional marketplaces of Blimbing, Gadang, and Merjosari, yielded 100% accuracy [the 5th]. Then there's the formalin content detection model in fish with hue saturation values (HSV) image using a k-nearest neighbour study by Agustyani et al. In this investigation, the most incredible validation was 91.67% when colour feature extraction and texture were combined.

In contrast, the lowest validation was 67.67% when texture feature extraction was used. Let's pretend the confirmation is based on the fish picture portion. In such a scenario, the gill section image has the greatest validation at 86.67%, while the complete fish image receives the lowest at 81.67%. The HSV feature extraction approach is acceptable for usage based on this research since the classification procedure was carried out by looking at the colour of the tuna's eye and using the Naive Bayes classifier method.

2. Literature Review

2.1 Digital Image Processing

In the 1960s, a computer with image processing capabilities was released. It is quicker because the computer is at the forefront of technical advancement in digital picture processing [9]. As a type of visual information, the image as part of multimedia plays an essential function. Pictures have qualities that text data does not have, such as images with a lot of information [10]. Image as the data recording system's output, which might be as follows:

1. Digital data may be immediately stored on a storage medium.
2. Analog in a visual signal on a television display, such as the image above.
3. Photography is an example of optics.

2.2 Euthynnus Affinis Fish

Tuna, also known as *Euthynnus affinis* in Latin, is a pelagic fish found in the Indonesian and Indo-Pacific seas. This tuna tends to congregate in significant numbers, making separation difficult [3]. The tuna is formed like a torpedo, with tiny teeth on both jaws and a slightly slanted mouth. The second dorsal fin is separated from the first dorsal fin high on the front and then low on the rear of the second dorsal fin. The front and back of the body are grey, the sides and belly are silver, and there are up-and-coming whitish stripes on the back [12].

2.3 Formalin

Formalin is an odourless, colourless solution with a strong odour. When formaldehyde is mixed with water, it contains around 37% formaldehyde, which becomes 15% preservative when added methanol. Formalin is often used to eliminate pests (disinfectants) and is also widely utilized in industrial processes. Formol, Methylene aldehyde, Paraformin, Morbucid, Oxomethane, Polyoxymethylene glycols, Methanal, Formoform, Superlysoform, Formaldehyde, and Formalith, are some of the various names for formalin [14].

2.4 HSV Color Space

HSV (Hue, Saturation, Value) is one of several colour systems people use to select colours (such as paint and ink) from the colour wheel or palette. This colour model is viewed more than the RGB model as a way humans try and describe colour sensations. The HSV model is formulated by finding the RGB colour cube along the grey axis (the combined axis of black and white dots), which results in a hexagonal shape from the colour palette [8]. The process of converting the RGB colour model to HSV uses the equation 1 to 4.

$$R = \frac{R}{255}; G = \frac{G}{255}; B = \frac{B}{255} \quad (1)$$

$$v = \max \quad (2)$$

$$s = \begin{cases} 0, & \text{if } \Delta = 0 \\ \frac{\Delta}{\max}, & \text{if } \Delta \neq 0 \end{cases} \quad (3)$$

$$h = \begin{cases} 60 \frac{G - B}{\Delta}, & \text{jika } \max = R \\ 120 + 60 \frac{B - R}{\Delta}, & \text{jika } \max = G \\ 240 + 60 \frac{R - G}{\Delta}, & \text{jika } \max = B \end{cases} \quad (4)$$
$$h = h + 360, \text{ if } H < 0$$

The HSV value must then be normalized to a simpler number. In other words, the 3-dimensional vector picture must be converted into a 1-dimensional vector image to reduce classification time [15]. The equation 5 to 7 is used to quantify.

$$H = \begin{cases} 0 & \text{if } h \in [316, 20] \\ 1 & \text{if } h \in [21, 40] \\ 3 & \text{if } h \in [76, 155] \\ 4 & \text{if } h \in [156, 190] \\ 5 & \text{if } h \in [191, 270] \\ 6 & \text{if } h \in [271, 295] \\ 7 & \text{if } h \in [296, 315] \end{cases} \quad (5)$$

$$S = \begin{cases} 0 & \text{if } s \in [0, 0.2] \\ 1 & \text{if } s \in [0.2, 0.7] \\ 2 & \text{if } s \in [0.7, 1] \end{cases} \quad (6)$$

$$s = \begin{cases} 0, & \text{if } \Delta = 0 \\ \frac{\Delta}{\max}, & \text{if } \Delta <> 0 \end{cases} \quad (7)$$

Equation 8 is the quantization result is utilized as the input value for a 1-dimensional HSV vector (G) [16]:

$$G = (9 \times H) + (3 \times S) + V \quad (8)$$

2.5 Naive Bayes Classifier

The Naive Bayes Classifier is a basic probabilistic classification algorithm that adds the frequencies and combinations of values from a dataset to produce a sequence of probabilities. The Bayes theorem is used in this algorithm, which assumes that the values determine all independent or non-interdependent characteristics in the class variables [7]. According to another interpretation, the Naive Bayes Classifier technique is also known as Bayesian Classification, which is a statistical classification approach for estimating the likelihood of membership in a class. The Bayes theorem underpins the Naive Bayes Classifier technique, which offers classification benefits similar to decision trees and neural networks. Furthermore, the Naive Bayes Classifier technique has demonstrated excellent accuracy and speed [5]. The following equation function is used to calculate using the Naive Bayes Classifier method:

$$g_j X = \log(p(w_j)) - \sum_{i=1}^d \log(\sigma_{ij}) - \frac{1}{2} \sum_{i=1}^d \frac{(f_i - \mu_{ij})^2}{\sigma^2_{ij}} \quad (9)$$

Where ω_j is a feature vector in class j with the vector mean parameter μ_{ij} , and *covariance* Σ is the training data estimate result. In addition, equation 10 demonstrates its use in the data classification process.

$$\hat{C} = \arg \max_c (g_j(x)), j = 1, \dots, C \dots \dots \dots \quad (10)$$

2.6 Android

Google acquired a startup named Android in 2000 before smartphones were available on the market. Android is a relatively new firm that specializes in producing embedded operating systems and mobile applications. The creators of Android, including Andi Rubin, Rich Miner, Nick Sears, and Chris White, went to Google when Android Inc. became a Google group. The Android OS was created exclusively for Google's internal purposes at the time, and it was not yet released open source. Even though Android was the first open-source mobile embedded OS, many people assumed it was just a piece of software for phones. Android is developing within the boundaries of the operating system and building apps utilizing the Android SDK, despite its rising commercial success. The increasing popularity of Android OS-based smartphones has impacted the high optimism of several manufacturers who are members of the OHA group, changing the dominance of the mobile device market, which Nokia and Apple [17] formerly controlled.

3. Method

3.1 Research Materials and Tools

The research materials and tools that will be used in the research are as follows:

1. Computer / Laptop with Core i5 specifications, 4 GB RAM.
2. Smartphone camera to take photos of tuna fish eye.
3. Android Studio.

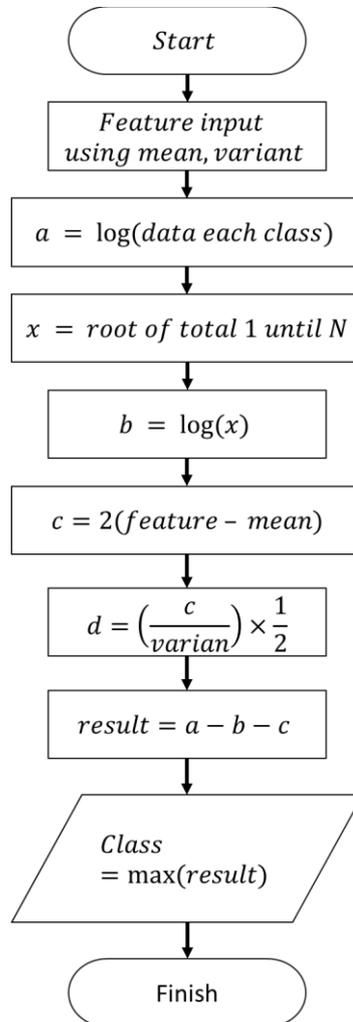


Fig. 1. Workflow method

3.2 Data Pre-Processing

On the system that will be built, data will be retrieved. The primary data was obtained by direct observation in the field, namely by visiting marketplaces in Banjarmasin City such as Cemara Market, Old Market, and Belitung Market to get tuna fish data. The data is acquired from the market, and then the picture data is created in the shape of a tuna's eye. Then there's the training data, which comes in tuna eye pictures, with up to 60 of them. Thirty images of formalin tuna eyes and 30 photographs of unformulated tuna eyes are divided into two groups. The data is processed and stored in the system as training data to get parameter estimates for the mean and variance values. The tuna categorization method will utilize this number.

3.3 Research Workflow

Fig. 1 shows a flowchart that will be used to calculate the Naive Bayes Classifier technique. Where the mean and variant feature values are supplied, the algorithm calculates the logarithmic value of the data for each class, then takes the root of the total data 1 to N for the complete data. The findings are logarithmically computed and then averaged over all characteristics and the prior average value. 2 times

the value of c is computed from the results and divided by the variant's results. By subtracting the values of a , b , and c , the final result is achieved, and the greatest value is the class obtained.

4. Results And Discussion

4.1 Result

The Naive Bayes Classifier must be calculated using previously provided training data to obtain the classification results from the testing data. After the data is ready, the calculating procedure may begin. The following example demonstrates how the Nave Bayes Classifier is calculated: The data prepared is training data; in this case, three formalin tuna eyes and three formalized tuna eyes with a size of 5×4 pixels were employed. The picture is first transformed from RGB to HSV, and then the HSV values are merged, as demonstrated in all of the training data photos.

Table 1. All 5×4 Pixel Image Training Data Features

	No Formalin			Formalin		
1	46	37	0	46	46	40
x2	46	37	1	46	37	28
x3	37	37	28	45	10	19
x4	28	37	28	45	13	22
x5	46	37	46	46	46	40
x6	46	39	45	45	46	37
x7	36	39	46	46	19	19
x8	28	37	37	45	10	19
x9	49	37	46	46	46	49
x10	46	39	51	45	48	46
x11	45	39	48	46	46	19
x12	19	37	46	28	10	10
x13	48	37	46	45	49	46
x14	46	39	48	46	49	46
x15	27	37	51	28	1	19
x16	19	37	46	28	10	10
x17	48	46	40	45	46	37
x18	46	46	37	46	37	28
x19	28	28	37	28	10	19
x20	19	28	37	27	13	13

Furthermore, the values for each feature in each image class are not formalin and formalin, and the results are in Table 2.

Table 2. All 5×4 Pixel image training data features

	No Formalin	Formalin
x1	83	132
x2	84	111
x3	102	74
x4	93	80
x5	129	132
x6	130	128
x7	121	84
x8	102	74
x9	132	141
x10	136	139
x11	132	111
x12	102	48
x13	131	140
x14	133	141
x15	115	48
x16	102	48
x17	134	128
x18	129	111
x19	93	57
x20	84	53

The next step is to compute the mean and variance in each class or each in the training data after getting the value from the training data. The results of the mean and variance in the informal class are shown in Table 3. The results of the mean and variance in the formalin class are shown in Table 4.

Table 3. Results of mean and variance values in non-formalin class

Features	Total Image	Mean	Varian
X1	83	27.67	1530.89
X2	84	28	1568
X3	102	34	2312
X4	93	31	1922
X5	129	43	3698
X6	130	43.33	3755.56
X7	121	40.33	3253.56
X8	102	34	2312
X9	132	44	3872
X10	136	45.33	4110.22
X11	132	44	3872
X12	102	34	2312
X13	131	43.67	3813.56
X14	133	44.33	3930.89
X15	115	38.33	2938.89
X16	102	34	2312
X17	134	44.67	3990.22
X18	129	43	3698
X19	93	31	1922
X20	84	28	1568

Table 4. Results of Values and Variants in Formalin Class

Features	Total Image	Mean	Varian
X1	132	44	3872
X2	111	37	2738
X3	74	24.67	1216.89
X4	80	26.67	1422.22
X5	132	44	3872
X6	128	42.67	3640.89
X7	84	28	1568
X8	74	24.67	1216.89
X9	141	47	4418
X10	139	46.33	4293.56
X11	111	37	2738
X12	48	16	512
X13	140	46.67	4355.56
X14	141	47	4418
X15	48	16	512
X16	48	16	512
X17	128	42.67	3640.89
X18	111	37	2738
X19	57	19	722
X20	53	17.67	624.22

After collecting the mean and variance of the training data, the following step in the data testing procedure may begin. The size is 5 x 4 pixels of tuna eye imaging data are used in this demonstration example. The picture is transformed from RGB to HSV, then the HSV value is quantized, then merged, and the image data testing values are obtained, as given in Table 5. The Naive Bayes Classifier computation may then be carried out in each class, namely the informal and formal classes when the results of the picture data testing are obtained. Calculate the class probability by dividing each class's data by the real data and then logging the result as follows.

$$a = \log(\text{data each class} / \text{number of data})$$

$$a = \log(3/6)$$

$$a = -0,30102999$$

Table 5. Image Testing Data Conversion RGB to HSV

H = 191	H = 191	H = 190	H = 190	H = 195
S = 15,3	S = 16,1	S = 10,3	S = 19,1	S = 12,9
V = 56,5	V = 53,7	V = 68,6	V = 36,9	V = 60,8
H = 190	H = 192	H = 191	H = 191	H = 196
S = 23,7	S = 100	S = 10,6	S = 11,6	S = 26,3
V = 29,8	V = 5,9	V = 62,7	V = 57,3	V = 22,4
H = 196	H = 189	H = 190	H = 190	H = 190
S = 75	S = 8,8	S = 12	S = 13,5	S = 17,4
V = 7,8	V = 58	V = 39,2	V = 34,9	V = 27,1
H = 190	H = 196	H = 192	H = 192	H = 203
S = 9,4	S = 22,4	S = 8,8	S = 8,1	S = 17
V = 49,8	V = 19,2	V = 44,7	V = 48,6	V = 18,4

Table 6. Feature of Image Testing Data

x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
46	40	51	46	46	51	37	48	37	46
x11	x12	x13	x14	x15	x16	x17	x18	x19	x20
37	46	37	46	37	46	46	49	37	45

Then, as previously stated in each class, namely formal and informal classes, compute the second and third stages (Table 7 and Table 8).

Table 7. No Formalin

Feature	Data feature	Image Total	Mean	Varian	$\sum_{i=1}^d \log(\sigma_{ic})$	$\frac{1}{2} \sum_{j=1}^n \frac{(x_i - \mu_{ic})^2}{\sigma_{ic}}$
X1	46	83	27.67	1530.89	1.59	0.22
X2	40	84	28	1568	1.60	0.10
X3	51	102	34	2312	1.68	0.13
X4	46	93	31	1922	1.64	0.12
X5	46	129	43	3698	1.78	0.01
X6	51	130	43.33	3755.56	1.79	0.02
X7	37	121	40.33	3253.56	1.76	0.01
X8	48	102	34	2312	1.68	0.08
X9	37	132	44	3872	1.79	0.01
X10	46	136	45.33	4110.22	1.81	0.00
X11	37	132	44	3872	1.79	0.01
X12	46	102	34	2312	1.68	0.06
X13	37	131	43.67	3813.56	1.79	0.01
X14	46	133	44.33	3930.89	1.80	0.01
X15	37	115	38.33	2938.89	1.73	0.01
X16	46	102	34	2312	1.68	0.06
X17	46	134	44.67	3990.22	1.80	0.00
X18	49	129	43	3698	1.78	0.01
X19	37	93	31	1922	1.64	0.02
X20	45	84	28	1568	1.60	0.18
Total					34.43	1,036

After you've gotten your results, move on to the last stage, which is to remove each result from each class and determine the maximum value.

Non-formalin

$$\text{result} = -0,30102999 - 34,4283986 - (1/2 * 1,035903654)$$

result = - 0,30102999 - 34,4283986 - 0,517951827
 result = - 35,24738042

Formalin

result = - 0,30102999 - 32,73669981 - (1/2 * 7,478714133)
 result = - 0,30102999 - 32,73669981 - 3,739357067
 result = - 36,77708687

The non-formalin class's results are higher than the formalin class, as can be seen from the two classes' calculations so that the non-formalin category contains the outcomes of the testing data.

Table 8. Formalin

Feature	Feature testing	Number of images	Mean	Varian	$\sum_{i=1}^d \log(\sigma_{ic})$	$\frac{1}{2} \sum_{j=1}^n \frac{(x_i - \mu_{ic})^2}{\sigma_{ic}}$
X1	46	132	44.00	3872.00	1.79	0.00
X2	40	111	37.00	2738.00	1.72	0.00
X3	51	74	24.67	1216.89	1.54	0.57
X4	46	80	26.67	1422.22	1.58	0.26
X5	46	132	44.00	3872.00	1.79	0.00
X6	51	128	42.67	3640.89	1.78	0.02
X7	37	84	28.00	1568.00	1.60	0.05
X8	48	74	24.67	1216.89	1.54	0.45
X9	37	141	47.00	4418.00	1.82	0.02
X10	46	139	46.33	4293.56	1.82	0.00
X11	37	111	37.00	2738.00	1.72	0.00
X12	46	48	16.00	512.00	1.35	1.76
X13	37	140	46.67	4355.56	1.82	0.02
X14	46	141	47.00	4418.00	1.82	0.00
X15	37	48	16.00	512.00	1.35	0.86
X16	46	48	16.00	512.00	1.35	1.76
X17	46	128	42.67	3640.89	1.78	0.00
X18	49	111	37.00	2738.00	1.72	0.05
X19	37	57	19.00	722.00	1.43	0.45
X20	45	53	17.67	624.22	1.40	1.20
		Total			34.43	7.48

4.2 Accuracy Testing

From the data processing to the classification of the tuna eye picture, a total of 30 non-formalin data and 30 formalin data were used for training, and five informal data and five formalin data were used for testing. After that, the following are the outcomes.

4.2.1 Non-formalin

Five data from the system findings are non-formalin or formalin, based on 30 training data and five non-formalin testing data. Table 9 shows the results of the ground truth obtained system comparison.

Table 9. Testing Results

Image Data Tuna Eyes	Ground Truth	Result
Image 1	No formalin	No formalin
Image 2	No formalin	No formalin
Image 3	No formalin	No formalin
Image 4	No formalin	No formalin
Image 5	No formalin	formalin

4.2.2 Formalin

Five data from the system are not formalized, out of 30 training data, and 5 formalized testing data. Table 10 shows the comparison of ground truth obtained system outcomes. Tables 9 and 10 compare the

results of all system tests, which may be used to continue the accuracy testing process. The system's accuracy level may then be determined as follows.

$$\begin{aligned} \text{Accuracy} &= \frac{\text{Amount of correct test result data}}{\text{total testing data}} \times 100 \% \\ &= \frac{8}{10} \times 100 \% \\ &= 80 \% \end{aligned}$$

Table 10. Testing Results of Formalin

Image Data Tuna Eyes	Ground Truth	Result
Image 1	formalin	formalin
Image 2	formalin	No formalin
Image 3	formalin	formalin
Image 4	formalin	formalin
Image 5	formalin	formalin

According to the test findings in the formalin tuna detection application based on fisheye pictures utilizing the android-based naive Bayes classifier technique, the results acquired an accuracy of 80%.

5. Conclusions

The application for identifying formalin or unformalized tuna has been effectively created utilizing the naive Bayes classifier approach, according to the findings of this study. The app is well-designed and performs nicely on Android. The system's accuracy is 80% after successfully implementing the naive Bayes classifier technique as a classification method.

References

- [1] C. Novia, "Kajian Kelayakan Teknis dan Finansial Produksi Nugget Jamur Tiram Putih (*Pleurotus ostreatus*) Rasa Ikan Tongkol (*Euthynnus aletrates*) Skala Industri Kecil," p. 1, 2011.
- [2] M. S. Milo, "Mutu Ikan Tongkol (*Euthynnus affinis* C.) Di Kabupaten Gunungkidul dan Sleman Daerah Istimewa Yogyakarta," p. 1, 2013.
- [3] A. N. Alfian, "Implementasi Regresi Logistik Untuk Mendeteksi Ikan Berformalin Berbasis Android Berdasarkan Citra dan Sifat Fisik Ikan," 2016.
- [4] B. I. Siahaan, "Pendeteksi Ikan Berformalin Melalui Citra Mata Menggunakan Metode *Probabilistic Neural Network* Berbasis Android," p. 1, 2018.
- [5] F. M. Hadini, "Sistem Pendeteksi Ikan Bandeng (*Chanos chanos*) Berformalin Berbasis Android Berdasarkan *Image* Mata Menggunakan Metode *Naive Bayes Classifier*," 2016.
- [6] Y. Agustyani, S. Setyaningsih dan A. Qur'ania, "Model Deteksi Kandungan Formalin Pada Ikan Dengan Citra *Hue Saturation Values* (HSV) Menggunakan *K-Nearest Neighbor*".
- [7] A. Saleh, "Implementasi Metode Klasifikasi *Naive Bayes* Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec Journal*, pp. 209-212, 2015.
- [8] E. Prasetyo, *Pengolahan Citra Digital dan Aplikasinya Menggunakan Matlab*, Andi, 2011.
- [9] D. Putra, *Pengolahan Citra Digital*, Yogyakarta: ANDI, 2010.
- [10] R. Munir, *Pengolahan Citra Digital*, Bandung: Informatika Bandung, 2002.
- [11] S. Madenda, *Pengolahan Citra dan Video Digital Teori, Aplikasi dan Pemrograman Menggunakan MATLAB*, Erlangga, 2015.
- [12] A. K. Atmaja, "Aplikasi Asap Cair Redestilasi Pada Karakterisasi Kamabako Ikan Tongkol (*Euthynnus affinis*) Ditinjau Dari Tingkat Keawetan dan Kesukaan Konsumen," pp. 14-17, 2009.
- [13] D. Destari, "Kemunduran Mutu Ikan Tongkol (*Euthynnus affinis*) Segar Di Pasar Modern dan Tradisional Kota Pekanbaru," 2017.

- [14] N. Hasnah, "Identifikasi Kandungan Formalin Pada ikan Asin Yang Dijual Di Kota Kendari Sulawesi Tenggara," pp. 5-8, 2018.
- [15] G. F. Rohmi, W. B. Zulfikar dan Y. A. Gerhana, "Implementasi Citra Digital Berdasarkan Nilai HSV Untuk Mengidentifikasi Jenis Tanaman Mangga Menggunakan Algoritma *K-Nearest Neighbor*," vol. 1, pp. 142-147, 2018.
- [16] D. Syahid, J. dan D. Nursantika, "Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode *K-Nearest Neighbor* (KNN) Berdasarkan Nilai *Hue, Saturation, Value* (HSV)," vol. 1, 2016.
- [17] J. E. Istiyanto, Pemrograman Smart Phone Menggunakan SDK Android dan *Hacking* Android, Graha Ilmu, 2013.