Tuna Meat Grade Classification Using Color Histogram and Grey Level co-Occurrence Matrix

Muhammad Naimullah[1], Febryanti Sthevanie, S.T., M.T.[2], Kurniawan Nur Ramadhani, S.T., M.T.[3]

[1,2,3]Faculty of Informatics, University of Telkom, Bandung
[1]muhnaimullah@students.telkomuniversity.ac.id, [2]sthevanie@telkomuniversity.ac.id,
[3]kurniawannr @telkomuniversity.ac.id

Abstract

As one out of the biggest country that produce Tuna in the world, Indonesia supply more than 16% of total production from 7.7 million metric tons of tuna. Exported tuna must have a good quality, to know the quality of tuna it must be tested by trained panelist that required approved certification. Tuna quality can be tested by using Organoleptic that checked color, texture, smell, and appearance. In this thesis, the feature of color and texture of the tuna will be extracted using Color Histogram and Grey Level co-Occurrence Matrix. The result of the classification by using Support Vector Machine as the classifier is 81.6% with 36 data contain 13 grade A data, 11 grade B data, and 12 grade D data.

Keywords: Tuna, RGB, HSV, Grey Level co-Occurrence Matrix (GLCM), Support Vector Machine (SVM)

1. Introduction

1.1. Background

Data released by Food and Agriculture Organization in 2012 that was written in perumperindo.co.id, Indonesia is 3rd rank in the world in fisheries production below China and India. Tuna is one of the biggest products that make Indonesia become one out of the biggest country that produce tuna in the world with catches worth up to 5 Million US Dollar [2]. In 2016, Indonesia itself supply more than 16% of total production from 7.7 million metric tons of tuna. Meanwhile in 2017, Indonesia's tuna export volume reach 198.131 ton with worth 659.99 Million US Dollar. Tuna that will be exported must have a good quality meat. A good quality tuna meat must be tested by trained panelist [3].

To know the quality of Tuna meat, it can be achieved through a test called Organoleptic. Organoleptic is a quality test that use human sense to measure the quality of the meat. Organoleptic test used in tuna meat have some indicator such as color, smell, appearance, and texture [3]. The grade of tuna meat can be different in every company. Company that used to collect the data in this final year task have grade A, grade B, and grade D.

In 2007 [17], there's a research that make for checking the freshness of tuna meat is using Wavelet and Local Binary Pattern method and gain 75% accuracy. The same research was conducted in 2019 by using RGB extraction and get 86.667% accuracy [21]. To gain more evidence to know that GLCM method is good for this task is based on a research in 2018 [5], a research to classify chicken meat using GLCM and Color Feature method have an accuracy of 98%. Because of that, author suggest to use Color Histogram and Grey Level co-Occurrence Matrix (GLCM) method to be used for this research and use Support Vector Machine (SVM) as the classifier method.

1.2. Problem Identification and Limitations

Based on background described above, a problem can be formulated for this final task is:

1. How to make a tuna meat grade classification system using Color Histogram and Grey Level co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) as the classifier?

There's a limitation for this final task as follow:

1. Dataset that used was taken manually by the author.
2. Dataset that used was from PT Parlevliet Paraba Seafood
3. Grade that will be classified is grade A, grade B, and grade D only.
4. The tuna that will be used is Yellow Fin Tuna.

1.3. Goal

The aim of this final task is:

1. To make a classification system for tuna meat grade using Color Histogram and Grey Level co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) as the classifier.

2. Related Studies

2.1. Literature Review

No	Tittle	Topic	Method	Result
1	Chicken Meat Freshness Identification using Colors and Textures Feature	Image Classification	Grey Level co-Occurrence Matrix (GLCM) dan Support Vector Machine (SVM)	Generate an Accuracy of 98% with SVM using a webcam.
2	Sintesa Rrecognition Tekstur dan Warna Terhadap Daging Tuna	Image Classification	Texture Feature and Color	Have a good accuracy with R average as the dominant feature.
3	Tuna Fish Classification using Decision Tree Algorithm and Image Processing Method	Image Classification	Grey Level co-Occurrence Matrix (GLCM) dan Decision Tree	Using decision tree as the classifier, researcher resulting accuracy of 88%.
4	Fish Freshness Classification Method Based on Fish Image using k-Nearest Neighbor	Image Classification	k-Nearest Neighbor	Using 3 different kind of fish, resulting accuracy of 91.36%.

Table 2.1 Literature Review

2.2. Tuna Fish

Tuna is a pelagic fish with a wide range of power and is in the Thunini family, which consists of several genera which have sizes from small to very large. Besides having a good taste, tuna also has many benefits for the human body. That is because tuna has various nutritional content such as A, B6, and B12 as well as high protein omega3 to rich in minerals [9].

Tuna fish that will become the dataset is originally owned by PT Parlevliet Paraba Seafood and the type of tuna fist that will become the dataset is Yellow Fin Tuna only. The 3 grade that is going to be used is:

1. Grade A
   This grade of tuna meat has bright red color all over the body until its tail.

Image 2.1 Grade A

2. Grade B
   This grade of tuna meat has red color that is not too bright like grade A, the red color has faded from body until its tail.

Image 2.2 Grade B

3. Grade D
   This grade of tuna meat has a mix color between red and brown from the body to the tail.

Image 2.3 Grade D

2.3. RGB as Color Feature

RGB Color space model have 3-main color model, there are Red (R), Green (G), Blue (B). The black color in RGB is represented in R = G = B = 0, while white color is represented in R = G = G = 255. RGB Color space have a strong relation with light [5]. That cause RGB color space is very sensitive to the change of light and noise [5].

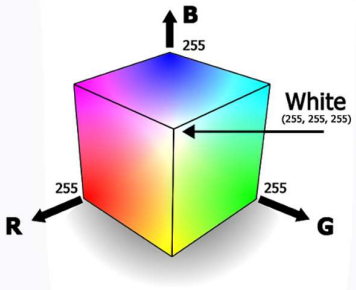


Image 2.4 RGB Color Space [5]

2.4. HSV as Color Feature

HSV consists of 3 color models, namely Hue (H), Saturation (S) and Value (V). HSV was first introduced by a.r.Smith in 1978 [5] where HSV has the advantage of the same color as the color produced by the human senses. HSV color space. The concept of the cone, seen from the side of the cone circumference,

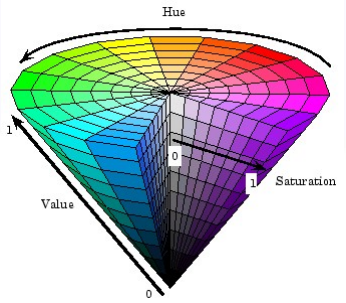


Image 2.5 HSV Color Space [Researchgate.net]

Hue is represented as the angle of each color in the cone to the 0 ° line where it represents red. Saturation is represented as the distance to the midpoint of the cone circle, the color that has a high saturation value is on the strongest side of the cone while the lowest saturation value is in the middle of the cone. Values that represent the resolution of the vertical position of the cone, at the cusp of the cone, none of them reflect all colors to black, while on the other hand all colors have control [5] [12].

2.5. Grey Level co-Occurrence Matrix (GLCM)

GLCM is a statistical method used to calculate feature textures by considering the spatial relationship of the gray pixels of an image. The GLCM matrix characterizes image textures through the calculation of several pairs of pixels that have specific spatial relationships and value relationships in an image. GLCM is made by calculating the frequency of a pixel with intensity i meeting pixels with intensity j in a specific spatial relationship [10].

The angle orientation θ is in 4 angular directions, namely 0 °, 45 °, 90 ° and 135 ° [5]. GLCM provides information about image textures including [5] [10].

1. Contrast
   Contrast is the difference of a pixel from the value next to it which can be calculated using equation (1). The contrast value is between 0 and (N1) 2 where N is the number of lines in GLCM
   $$\sum_{i,j}|i-j|^2 p(i,j) \quad (1)$$

2. Correlation
   Correlation is a feature that calculates the correlation of a pixel with other pixels which can be calculated using equation (2). The range of correlation values between -1 and 1.
   $$\sum_{i,j}\frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j} \quad (2)$$

3. Homogeneity
   Homogeneity is a feature that reverses the value of the proximity distribution between GLCM elements and diagonal GLCM which can be calculated using equation (3). The range of homogeneity values is between 0 to 1.
   $$\sum_{i,j}\frac{p(i,j)}{1+|i-j|} \quad (3)$$

4. Energy
   Energy is a feature that reverses the value of the number of boxes in each element in GLCM which can be calculated using equation (4). The range of energy values is between 0 to 1.
   $$\sum_{i,j} p(i,j)^2 \quad (4)$$

2.6. Support Vector Machine (SVM)

Support Vector Machine (SVM) has been used to analyze and identify a pattern that can solve classification and analysis problems [14] [15]. SVM was introduced by Vapnik and Cortes which aimed at the classification process to maximize margins between class classification boundaries [15] [16]. Basically, SVM functions to find the best hyperplane that has the highest margin, the margin itself has the meaning as the sum of the hyperplane distance to the nearest point of each classification class.

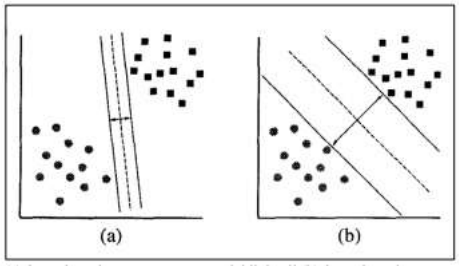


Image 2.6 (a) hyperplane with small margin (b) hyperplane with big margin

3. Built Systems

3.1. System Overview

In this final project research, there are several stages that will be carried out during the research. These stages are image capture, after capturing the image that will be used as dataset, the image next is cropped manually. Next step is preprocessing the data which convert the data into greyscale for GLCM feature extraction and image enhancement for color feature extraction. After preprocessing, next step feature extraction using GLCM and Color Histogram then combining both feature into one and last is classification using SVM classifier.

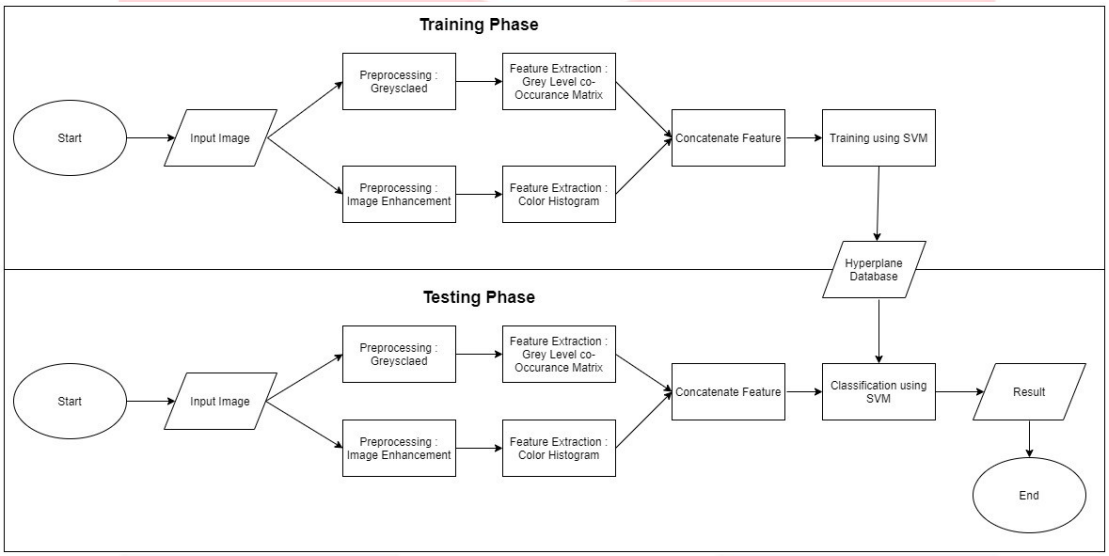


Image 3.1 System Overview Flowchart

3.2. Data Retrieval

The dataset that will be used is a picture of tuna meat from PT Parlevliet Paraba Seafood that will be taken using a camera. The shooting technique uses a mini studio that will be purchased which will maximize the results of the images.

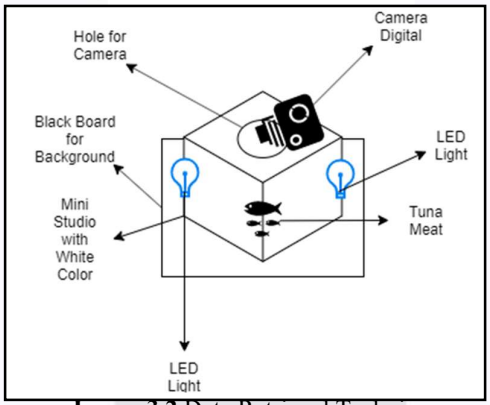


Image 3.2 Data Retrieval Technique

The distance between the digital camera and the tuna is 30cm. The LED lights and black background are useful for maximizing the results of shooting meat which will be used as data for this research.

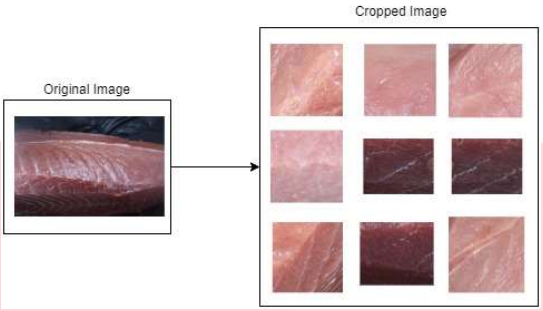


Image 3.3 Cropping Image

Each picture taken will be labeled in advance by experts from the relevant company so that after the classification process can be immediately determined its accuracy. After getting the image of tuna meat, next thing is to crop the image into several image so the dataset that will be used is increasing for better result.

3.3. Feature Extraction

To get the required values, the image that was taken previously will be convert the image to grayscale. In the next step, the converted image will be entered into the image feature extraction process using the GLCM method which will produce the values needed for the next process, namely classification.

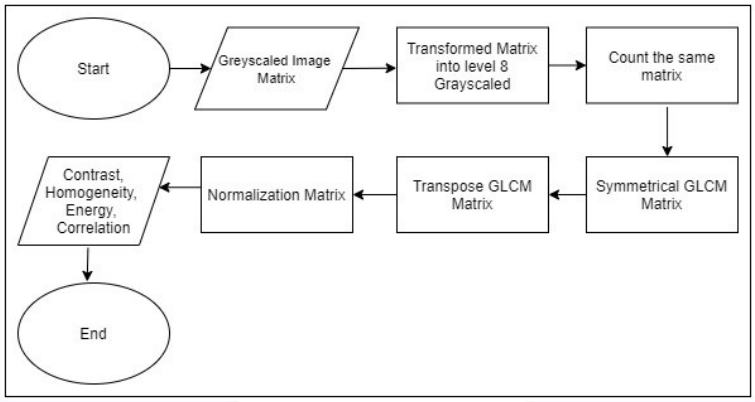


Image 3.4 GLCM Flowchart [22]

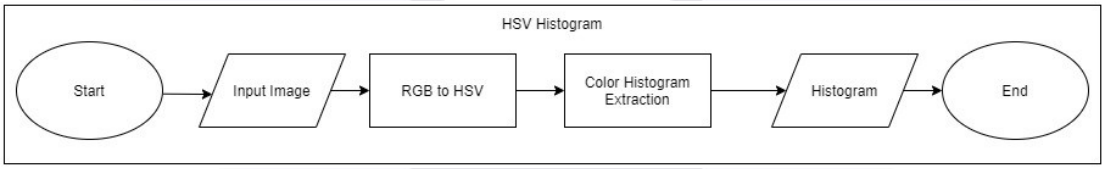


Image 3.5 HSV Color Histogram Flowchart

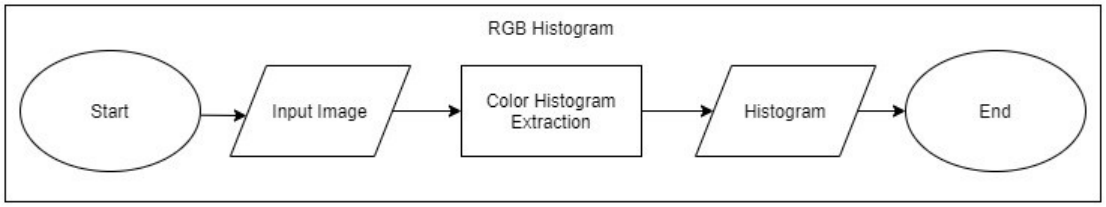


Image 3.6 RGB Color Histogram Flowchart

3.4. Classification

In the classification process, Support Vector Machine is used as a classification method where the data that has been obtained from the feature extraction will then be input in this method. After the data is entered, the data will be normalized and determine the parameters. In the next step, the data will be trained and tested based on the evaluation of the model. After the evaluation process is completed, the classification process will be carried out by SVM.

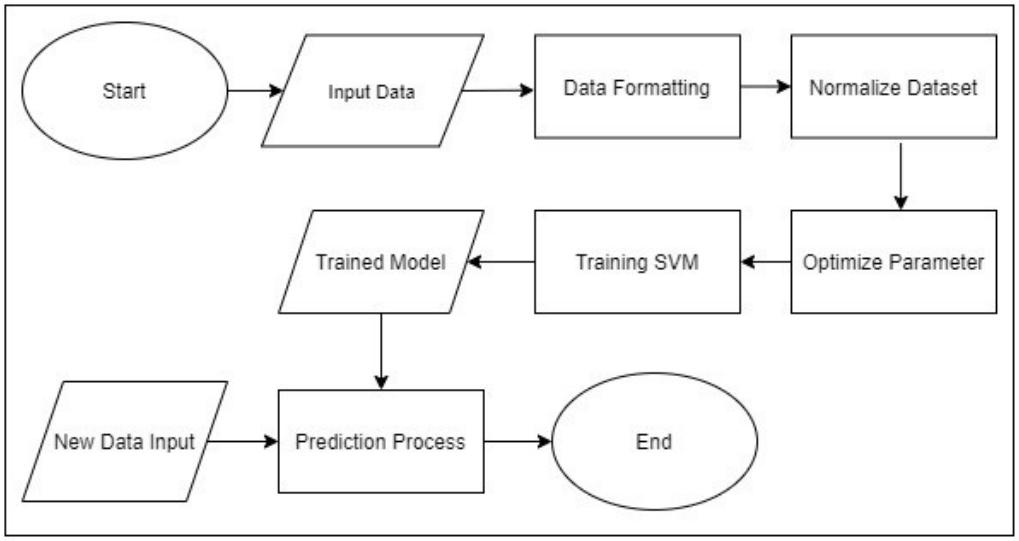


Image 3.7 SVM Classifier Flowchart [Researchgate.net]

3.5 Evaluation Design

This is needed to evaluate the model and the results of the evaluation refer to the accuracy value that can be calculated with the formula:

$$accuration = \frac{correct\ number\ of\ guesses}{amount\ of\ test\ data} x100\% \quad (5)$$

4. Evaluation

The program will be executing with 4 different scenarios to get the comparison between each scenario. Scenario 1 will use GLCM and RGB Histogram, Scenario 2 will use GLCM and HSV Histogram, Scenario 3 will use both RGB and HSV Histogram and GLCM with default SVM parameter, and Scenario 4 will use both Color histogram and GLCM with tuned SVM parameter. Each scenario will be validated using Stratified 10-Fold Cross Validation. Therefore, the actual feature for each grade is shown in Table 4.1 .

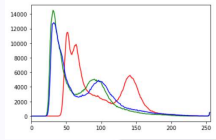
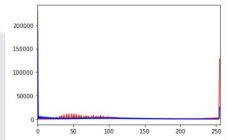
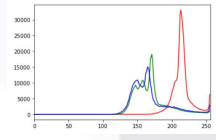
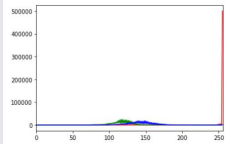
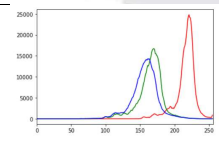
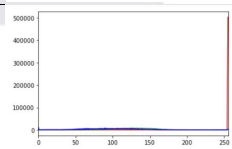
No	Grade	Histogram before Enhance	Histogram After Enhance	Grey Level co-Occurrence Feature
1	Grade A			**Contrast = 0.28211748** **Homogeneity = 0.92449763** **Energy = 0.38446501** **Correlation = 0.9734567**
2	Grade B			**Contrast = 0.02665029** **Homogeneity = 0.98667485** **Energy = 0.82721782** **Correlation = 0.94768606**
3	Grade D			**Contrast = 0.07856289** **Homogeneity = 0.96187542** **Energy = 0.53122642** **Correlation = 0.96814319**

Table 4.1 Actual Feature for each grade

4.1.Testing Results

4.1.1. Scenario 1

In this scenario, only use GLCM Feature and RGB Histogram that will be used to run the program. It produces an accuracy of 61.66%.

FOLD 1		FOLD 2		FOLD 3		FOLD 4		FOLD 5	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	32	4	32	4	32	4	32	4
Result = 75		Result = 50		Result = 75		Result = 50		Result = 25	
FOLD 6		**FOLD 7**		**FOLD 8**		**FOLD 9**		**FOLD 10**	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	33	3	33	3	33	3	33	3
Result = 75		Result = 66		Result = 66		Result = 66		Result = 66	

Table 4.2 Result of GLCM and RGB Histogram

4.1.2. Scenario 2

In this scenario, only use GLCM Feature and HSV Histogram that will be used to run the program. It produces an accuracy of 79.1%.

FOLD 1		FOLD 2		FOLD 3		FOLD 4		FOLD 5	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	32	4	32	4	32	4	32	4
Result = 50		Result = 50		Result = 75		Result = 75		Result = 75	
FOLD 6		**FOLD 7**		**FOLD 8**		**FOLD 9**		**FOLD 10**	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	33	3	33	3	33	3	33	3
Result = 100		Result = 100		Result = 100		Result = 100		Result = 66.6	

Table 4.3 Result of GLCM and HSV Histogram

4.1.3. Scenario 3

In this scenario, use both Color Histogram and GLCM Feature will be used to run the program with default SVM parameter. It produces an accuracy of 76.6%.

FOLD 1		FOLD 2		FOLD 3		FOLD 4		FOLD 5	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	32	4	32	4	32	4	32	4
Result = 50		Result = 50		Result = 75		Result = 75		Result = 50	
FOLD 6		**FOLD 7**		**FOLD 8**		**FOLD 9**		**FOLD 10**	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	33	3	33	3	33	3	33	3
Result = 100		Result = 100		Result = 100		Result = 100		Result = 66.6	

Table 4.4 Result of GLCM and both Color Histogram with default SVM Parameter

4.1.4. Scenario 4

In this scenario, use both Color Histogram and GLCM Feature will be used to run the program with tuned SVM parameter. Which is penalty = l1, dual = False. It produces an accuracy of 81.6%.

FOLD 1		FOLD 2		FOLD 3		FOLD 4		FOLD 5	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	32	4	32	4	32	4	32	4
Result = 50		Result = 100		Result = 75		Result = 75		Result = 50	
FOLD 6		**FOLD 7**		**FOLD 8**		**FOLD 9**		**FOLD 10**	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
32	4	33	3	33	3	33	3	33	3
Result = 100		Result = 100		Result = 100		Result = 100		Result = 66.6	

Table 4.5 Result of GLCM and both Color Histogram with tuned SVM Parameter

4.2.Analysis of Results of Testing

Based on test result where 4 scenarios where each scenario have different result. From scenario 1 to scenario 4, the accuracy is increasing while adding more feature extraction method and tuned the SVM parameter. The best accuracy is obtained from Scenario 4 where both color histogram and GLCM is used with tuned parameter in SVM classifier.

Total data used for this test is 36 data where it contains 13 data for Grade A, 11 data for grade B, and 12 data for grade D. Confusion matrix will provide more information for the model. Table 4.6 provide information from the best scenario which is scenario 4.

		Predicted Class		
		A	B	D
Actual class	A	1	0	0
	B	0	1	1
	D	0	0	1

Table 4.6 Confusion Matrix

The wrong predicted data from table above show data from grade B is predicted as grade D, it applies the other way round. It is because both grades almost have similar feature from color histogram and GLCM as show in table 4.7.

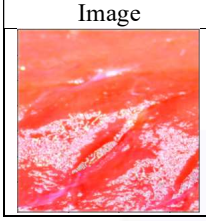
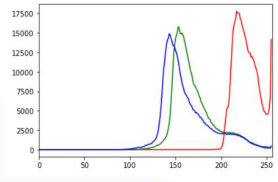
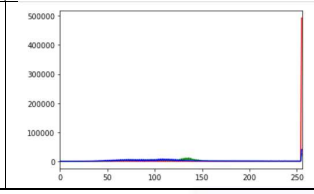
Image	Histogram Before Enhance	Histogram After Enhance	Actual	Predicted
			Grade B	Grade D

Table 4.7 Wrong Prediction

The data is predicted wrong because after enhancing the image by adding contrast and color, it has similar value in red color with grade D.

5. Conclusion

The use of Color Histogram combines with Grey Level co-Occurrence Matrix method for this system resulting good accuracy 81.6% with some tuning in classifier which is Support Vector Machine. The result of cross validation can be improved by adding more data. The model that is created is not yet by any means perfect.

More feature extraction method can be added, tuning more parameter from both feature extraction method and classifier, adding more data, and the process of taking pictures is made as well as possible, it aims to make the feature extraction results better and make better accuracy for future research.

Reference

[1]Humas Perum Perindo,2015, Potensi Indonesia sebagai Negara Maritim,Perum Perindo, http://www.perumperindo.co.id/publikasi/artikel/21-potensi-indonesia-sebagai-negara-maritim (diakses 16 September 2019)

[2]Ambaranie Nadia Kemala Movanita,2019, Diakui AS, Indonesia Kuasai Pasar IkanTunaDunia,Kompas.com,https://money.kompas.com/read/2019/04/03/111413126/diakui-asindonesia-kuasai-pasar-ikan-tuna-dunia?page=all (diakses 16 September 2019)

[3] Bawinto, A.S., Mongi, E.L. and Kaseger, B.E., 2015. Analisa kadar air, pH, Organoleptik, Dan Kapang Pada Produk Ikan Tuna (Thunnus Sp) Asap, Di Kelurahan Girian Bawah, Kota Bitung, Sulawesi Utara. MEDIA TEKNOLOGI HASIL PERIKANAN, 3(2).

[4] Sidik, F., Nurani, T.W. and Wisudo, S.H., 2019. Mutu dan Perdagangan Ikan Tuna Hasil Tangkapan Longline yang Didaratkan di Pps Nizam Zachman Jakarta. Buletin PSP, 21(2).

[5] R. A. Asmara et al., "Chicken Meat Freshness Identification using Colors and Textures Feature," 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, 2018, pp. 93-98.

[6] R. A. Asmara, F. Rahutomo, Q. Hasanah and C. Rahmad, "Chicken meat freshness identification using the histogram color feature," 2017 International Conference on Sustainable Information Engineering and Technology (SIET), Malang, 2017, pp. 57-61
[7] Widiastuti, I. and Putro, S., 2010. Analisis mutu Ikan Tuna selama lepas tangkap. Maspari Journal, 1(1), pp.22-29.
[8] Mulyawan, H., 2011. Identifikasi dan Tracking Objek Berbasis Image Processing secara Real Time. EEPIS Final Project.
[9] IKAN TUNA dan PRODUK OLAHAN TUNA (Market Brief) oleh Atase Perdagangan London Kedutaan Besar Republik Indonesia http://djpen.kemendag.go.id/membership/data/files/dcabc-market-brief-tuna.pdf 20/9/2019
[10] W. N. Khotimah, A. Z. Arifin, A. Yuniarti, A. Y. Wijaya, D. A. Navastara and M. A. Kalbuadi, "Tuna fish classification using decision tree algorithm and image processing method," 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Bandung, 2015, pp. 126-131.
[11] Cardani, D., 2001. Adventures in hsv space. Laboratorio de Robótica, Instituto Tecnológico Autónomo de México.
[12] D. S. Y. Kartika and D. Herumurti, "Koi fish classification based on HSV color space," 2016 International Conference on Information & Communication Technology and Systems (ICTS), Surabaya, 2016, pp. 96100.
[13] N. M. S. Iswari, Wella and Ranny, "Fish freshness classification method based on fish image using kNearest Neighbor," 2017 4th International Conference on New Media Studies (CONMEDIA), Yogyakarta, 2017, pp. 87-91.
[14] M. Al-Sarayreh, M. M. Reis, W. Q. Yan and R. Klette, "Deep Spectral-spatial Features of Snapshot Hyperspectral Images for Red-meat Classification," 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ), Auckland, New Zealand, 2018, pp. 1-6.
[15] M. H. Selamat and H. Md Rais, "Image face recognition using Hybrid Multiclass SVM (HM-SVM)," 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Bandung, 2015, pp. 159-164
[16] V. N. Vapnik, "An overview of statistical learning theory," in IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 988-999, Sept. 1999.
[17] A. Mateo, F. Soto, J. A. Villarejo and C. Fernandez, "Automatic Inspection System for Quality Evaluation of Fresh Tuna Meat," 2007 IEEE International Symposium on Industrial Electronics, Vigo, 2007, pp. 16691674.
[18] W. A. Saputra and D. Herumurti, "Integration GLCM and geometric feature extraction of region of interest for classifying tuna," 2016 International Conference on Information & Communication Technology and Systems (ICTS), Surabaya, 2016, pp. 75-79
[19] S. Widiyanto, Y. Sukra, S. Madenda, D. T. Wardani and E. P. Wibowo, "Texture Feature Extraction Based On GLCM and DWT for Beef Tenderness Classification," 2018 Third International Conference on Informatics and Computing (ICIC), Palembang, Indonesia, 2018, pp. 1-4.
[20] 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," 2017 International Conference on Innovative and Creative Information Technology (ICITech), Salatiga, 2017, pp. 1-6.
[21] K. Lugatiman, C. Fabiana, J. Echavia and J. J. Adtoon, "Tuna Meat Freshness Classification through Computer Vision," 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Laoag, Philippines, 2019, pp. 1-6, doi: 10.1109/HNICEM48295.2019.9073468.
[22] Hidayat, A.S., 2017. Implementasi Algoritma GLCM Dan MED pada Aplikasi Pendeteksi Kolesterol Melalui Iris Mata. MIND Journal, 2(2), pp.23-42.
[23] Pakaya, R. and Suleman, S., 2018, November. SINTESA RECOGNITION TEKSTUR DAN WARNA TERHADAP DAGING TUNA. In Prosiding SEHATI (Seminar Nasional Humaniora dan Aplikasi Teknologi Informasi) (Vol. 4, No. 1, pp. 21-27).