International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

# Classification of Side Dish Types (Chicken, Meat, and Fish) Based of Color Using K-Nearest Neighbor (K-NN) Method

Luna Colombini Adam[1], Karmilasari[2]
[1]Faculty of Information System Management, Gunadarma University, Depok, West Java, Indonesia-16424
[2]Faculty of Information System Management, Gunadarma University, Depok, West Java, Indonesia-16424
Email address: [1]luna.c.adam@gmail.com, [2]karmila@staff.gunadarma.ac.id

***Abstract—*** *The color of the meat gives an intuitive impression of its freshness and the composition of its ingredients. The human eye can easily adapt and interpret objects, capturing information that objects can change in the real world due to the difference between day and night or the effects of light and shadow. Information technology is a technology employed for data processing, encompassing tasks like data processing, acquisition, compilation, storage, and various methods of data manipulation to generate high-quality information. One of the current uses of digital image processing techniques is to make sense of images the way a human does. Because the three types of side dishes (chicken, meat, and fish) used in this research have different textures, this study only uses color to distinguish between the chicken, meat, and fish images. In this research, the authors used the RGB Model color. The researchers use K-Nearest Neighbors (K-NN) algorithm to classify different types of side dishes based on their color, the K value used is 3. The images of chicken, meat, and salmon fish are collected through an open-source database. A total of 600 images were used for training and 30 for testing. The peak RBG color of each trained dataset image is used to identify and classify the image for testing. Next, the results achieved with the test dataset were assessed by calculating the confusion matrix to get the accuracy, precision, recall, and f1-score of the program. The accuracy result of the program is 96.7%.*

***Keywords—*** *Color Recognition: Feature Extraction: Image Classification: K-Nearest Neighbor: Machine Learning.*

## I. Introduction

Humans are visual creatures. Humans just rely on sight to understand the world around them when they see an object not only to identify and classify but also to know the difference and feel quickly. The human eye can easily adapt and interpret an object to obtain information, which in the real world an object can undergo changes either because of the difference in day and night; the influence of light and shadow. The color of a type of meat provides an intuitive impression of freshness and the composition of its ingredients. It is not only a comprehensive indicator reflecting differences in muscle biochemistry, physiology, and microbiology but also an essential factor that changes with the life span of fresh meat [1]. It is also a way to distinguish different types of meat.

Information technology is the technology used to process data, including processing, retrieving, editing, storing and manipulating data in various ways to produce high quality information. One of the current uses of digital image processing technology is to understand the meaning of an image. Pattern recognition is a branch of science that studies data patterns both numerical or symbolic so that one or more conclusions can be drawn. The input of the pattern recognition program is an image while the output is an object description. Computer vision is one of those methods. Computer vision is a non-destructive, scalable, rapid implementation method suitable for real-time evaluation.

Fresh chicken boiler meat has these characteristics: the color of the meat is generally white and bright, chewy texture, chicken skin is usually not dry and not slimy, not bruised, fine fiber, not fatty in meat fibers, and the aroma of chicken meat is not pungent, does not smell fishy and does not rot. Meanwhile, fresh salmon fillet meat color should be bright, bright red, and pink and should not be dull or too slimy with white stripes. There should be no bruising on the fish and no browning.

Therefore, because of the variety of different textures across the three different types of side dishes, the only color is used for this research in identifying the difference between chicken, meat, and fish images.

Based on the description of the background and problems above, the researcher would like to further study this thesis using the K-Nearest neighbors (K-NN) algorithm to classify different types of side dishes (chicken, meat, and fish) using feature extraction in the form of color. The objective to be achieved of this study is to implement the thesis using the K-Nearest neighbors (K-NN) algorithm to classify different types of side dishes (chicken, meat, and fish) based on color.

## II. Literature Review

In the paper written by Purwanto et al., Detection of Chicken Meat Freshness Using K-Nearest Neighbor, this study developed a fresh chicken meat detection tool using the TCS-230 RGB color sensor. The input used in the K-Nearest Neighbor is in the form of RGB color values obtained from the color sensor. In this study, meat freshness was tested using TCS-230 color sensor with an accuracy rate of 87% with a positive precision of 92% and negative precision of 67%. Based on the results of this research, it was concluded that fresh chicken meat had an R (Red) value between 68-83, G (Green) between 22-35, B (Blue) value between 19-30 [2]. The identification of the color of raw chicken will provide a clear guide on the selection of raw chicken.

Fish product is one of Indonesian income sources. As a maritime country, most of Indonesian that live in the coastal area are depended on fishery. Fishery have contributed to the development of domestic industries, micro industries, and export industries. When Lugatiman et al. used RGB extraction to identify the freshness of tuna fillets, he took pictures of tuna meat chunks with a Raspberry Pi and transferred them to a mobile phone for image processing. The cuts were classified using k-Nearest neighbors (k-NN) algorithm and Waikato Environment for Knowledge Analysis (WEKA) in terms of the number of hours from slaughter. The result showed 86% overall accuracy in determining tuna meat freshness [3].

In the paper titled "Colour recognition using colour histogram feature extraction and K-nearest neighbour classifier", the writer combines a histogram-based method with the K-Nearest Neighbors (KNN) algorithm to recognize colors. The KNN algorithm, trained using RGB color histograms, effectively classifies various colors. Results show accuracy depends on dataset choice, with K = 5 being optimal. Black and pink achieve the highest accuracy at 90% with K = 5 [4]. The study underscores the importance of training data, K value, and lighting conditions for accurate color classification.

The RGB color space has 3 channels, namely red channel, green channel, and blue channel. The red channel represents the intensity of the red color in the image, the green channel represents the intensity of the green color in the image, while the blue channel represents the intensity of the blue color in the image. The value ranges on each channel from 0-255 [5].

K-Nearest Neighbor is one of the most common algorithms in machine learning. A machine learning model uses a set of input values to predict an output value. K-NN is one of the simplest forms of machine learning algorithms and is mainly used for classification. Classifies data points according to how neighboring data points are classified. The flow of how the KNN algorithm works is as follows:

1. Choose the number of neighbors K.
2. Calculate the distance from the number of K neighbors.
3. Take the nearest neighbor K according to the calculated distance.
4. Among these k neighbors, calculate the number of data points in each category.
5. Assign the new data point to the category with the highest number of neighbors.

## III. RESEARCH METHOD

Image processing is a method of performing some operation on an image to obtain an enhanced image or extract useful information from it. One of the examples of image processing is content based image retrieval. Currently, most of us are looking for images through search engines using words. however, sometimes it happens that the image given by someone is not what we want. With digital image processing, it will be possible for us to search for an image by providing input examples of the image we are looking for to the search engine.

Color is one of the most important indicators of food quality. Informs about freshness, the composition of ingredients, and the presence of counterfeit products. Ratings are subjective, as colors are primarily evaluated visually. This method can be widely applied to the automation of color analysis.

From the picture below, there are a number of data points which are divided into two classes, namely A (blue) and B (yellow). For example, there is new data (black) for which we will predict the class using the KNN algorithm. From the example above, the K value used is 3. After calculating the distance between the black dot to each other data point, we get the closest 3 points consisting of 2 yellow dots and one blue dot as illustrated in the red box, then the class for new data (black dot) is B (yellow) [6].

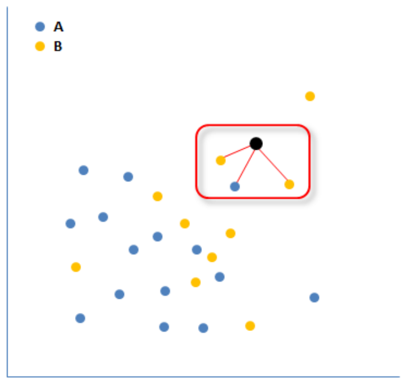



Fig. 1. Illustration of K-Nearest Neighbor [6].

This research focuses on side dish types classifying by K-Nearest Neighbors Machine Learning Classifier which is trained by R, G, and B Color. The value of k in the KNN algorithm defines how many neighbors will be checked to determine the classification of a particular query point. For example, if k=1, the instance will be assigned to the same class as its nearest neighbor. Overall, it is recommended to choose the value of k in the form of an odd number to avoid binding in the classification. Here we determine the value of k used is 3.

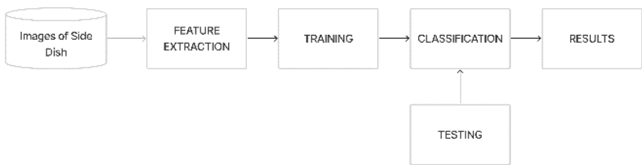



Fig. 2. Overall Design of the Program

The first phase, for dataset acquisition, the images of chicken, meat, and salmon fish are collected through an open-source database Kaggle. All side dish types have 200 images of each that will be used for training and 10 each for testing. Overall, 600 images that will be used for training and 30 for testing.

In feature extraction, the peak RBG colors of each side dish trained images are used to identify and classify the images which will then be used to test and classify images during the testing phase.

Training starts by identifying the RGB values of each training images. The model then determines to classify the input testing images and match them with the trained data. The image will have the classified outcome upon the bounding box in it.

### A. Data Acquisition

In the first phase, for dataset acquisition, the images of chicken, meat, and salmon fish are collected through an open-source database Kaggle. All side dish types have 200 images of each that will be used for training and 10 each for testing. Overall, 600 images will be used for training and 30 for testing.

### B. Feature Extraction

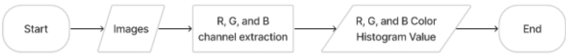



Fig. 3. Flowchart of Feature Extraction

In feature extraction, the peak RBG colors of each side dish trained images are used to identify and classify the images which will then be used to test and classify images during the testing phase.

A color histogram is an illustration of how colors are distributed in an image. In the context of digital images, it displays the count of pixels falling within specific color ranges from a predefined list, covering the entire color spectrum of the image.

### C. Training


Fig. 4. Flowchart of Training

OpenCV was used for color histogram calculations and the K-NN classifier. Numpy was used for matrix/n-dimensional array calculations. The peak of R, G, and B values of the color histogram is used for each training image and then labeled because the KNN classifier is a supervised learner and deploys these feature vectors in the CSV file. This is how you create a vector data set for training features. It is in a file named training.data. Among the 600 images trained, here are few examples of the trained images with its peak RGB values.

TABLE I. Training Images with RGB Value

| Training Images | RGV Values | Description |
|---|---|---|
|  | 190,162,126 | Chicken |
|  | 188,158,119 | |
|  | 161,109,110 | Meat |
|  | 139,91,105 | |
|  | 203,207,223 | Fish |
|  | 255,255,255 | |

### D. Testing

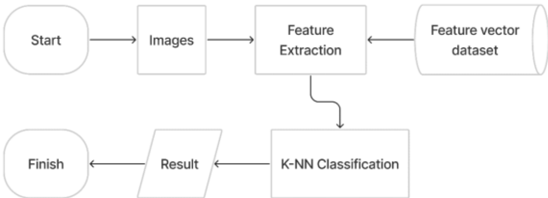



Fig. 5. Flowchart of Testing

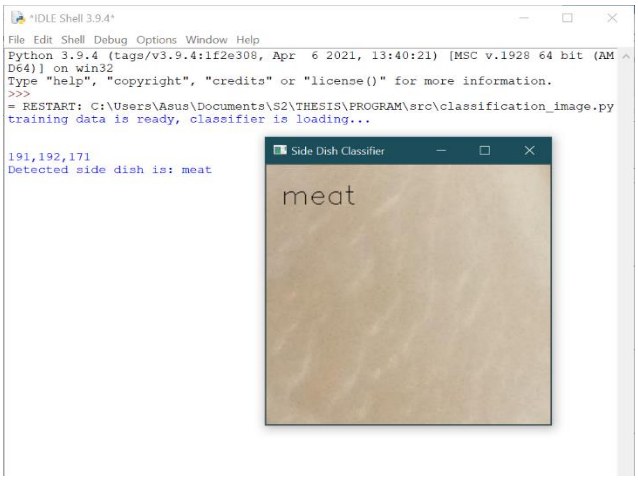



Fig. 6. One Example of Testing

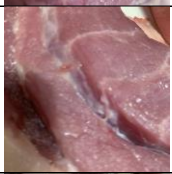
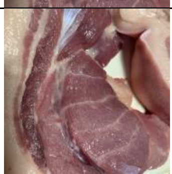
In this last phase, the program classification_image will call on two other python programs called feature_extraction which is used to create the training data by the RGB histogram feature extraction, and also knn_classifier program used for the implementation of the K-NN classifier. The program knn_classifier will fetch training.data and also fetch test image features. It will then calculate the euclidean distance and get the k value. Prediction of types of side dishes will be predicted based on their color which could be true or false compared to manual classification.

## IV. Results and Discussions

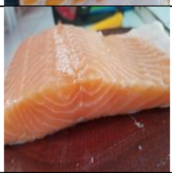
The testing of the program is used in which 30 different images of side dish types are run through the classifier. There are 10 for each different type of side dish types. The results are as follows:

TABLE II. Testing Results

| Images | Program Classification | Manual Classification | Results |
|---|---|---|---|
|  | Meat (191,192,171) | Chicken | Not Matched |
|  | Chicken (191,175,140) | Chicken | Matched |
|  | Chicken (169,135,94) | Chicken | Matched |
|  | Chicken (222,207,176) | Chicken | Matched |
|  | Chicken (225,211,181) | Chicken | Matched |
|  | Chicken (183,140,104) | Chicken | Matched |
|  | Chicken (205,184,149) | Chicken | Matched |
|  | Chicken (204,179,135) | Chicken | Matched |
|  | Chicken (185,155,122) | Chicken | Matched |

| | | | |
|---|---|---|---|
|  | Chicken (184,153,121) | Chicken | Matched |
|  | Meat (159,117,122) | Meat | Matched |
|  | Meat (158,113,129) | Meat | Matched |
|  | Meat (170,114,131) | Meat | Matched |
|  | Meat (98,69,170) | Meat | Matched |
|  | Meat (129,101,104) | Meat | Matched |
|  | Meat (138,90,99) | Meat | Matched |
|  | Meat (124,77,76) | Meat | Matched |
|  | Meat (144,96,104) | Meat | Matched |
|  | Meat (152,89,88) | Meat | Matched |

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

| | | | |
|---|---|---|---|
|  | Meat (141,101,107) | Meat | Matched |
|  | Fish (208,127,32) | Fish | Matched |
|  | Fish (207,144,73) | Fish | Matched |
|  | Fish (209,36,41) | Fish | Matched |
|  | Fish (255,138,0) | Fish | Matched |
|  | Fish (255,86,0) | Fish | Matched |
|  | Fish (255,255,255) | Fish | Matched |
|  | Fish (208,127,32) | Fish | Matched |
|  | Fish (255,255,255) | Fish | Matched |
|  | Fish (255,255,255) | Fish | Matched |

| | | | |
|---|---|---|---|
|  | Fish (169,193,202) | Fish | Matched |

A confusion matrix is a commonly employed table to characterize how well a classification model, or "classifier," performs on a test dataset where the actual values are known. Some of the performance metrics that are common and often used in the form of classification reports are Accuracy, Precision, Recall, and F1 score [7]. This study uses the confusion matrix parameter to measure the accuracy of the model, so that it can be known by making a confusion matrix table to determine the values for True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

- True Positive (TP): When an observation is positive and predicted to be positive.
- False Negative (FN): When an observation is positive but predicted to be negative.
- True Negative (TN): When an observation is negative and predicted to be negative.
- False Positive (FP): When an observation is negative but predicted to be positive.

Accuracy gives you the overall accuracy of the model, meaning the fraction of the total samples that were correctly

classified by the classifier. To calculate accuracy, apply the subsequent formula:

$$Accuracy = \frac{TN+TP}{TN+FP+TP+FN} \quad (1)$$

Precision tells you what fraction of predictions as a positive class were actually positive. To calculate precision, apply the subsequent formula:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall tells you what fraction of all positive samples were correctly predicted as positive by the classifier. To calculate Recall, apply the subsequent formula:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score combines precision and recall into a single measure. Mathematically it's the harmonic mean of precision and recall. To calculate, apply the subsequent formula:

$$F1 - Score = 2\,x\,\frac{Recall\ x\ Precision}{Recall+Precision} \quad (4)$$

TABLE III: Confusion Matrix for Multi-Class Classification

| | | True Class | | |
|---|---|---|---|---|
| | | **Chicken** | **Meat** | **Fish** |
| **Predicted Class** | **Chicken** | **9** | **0** | **0** |
| | **Meat** | **1** | **10** | **0** |
| | **Fish** | **0** | **0** | **10** |

All fish and meat images classified using program classification results are matched perfectly with its manual classification while nine out of image of chicken are accurate with one of them classified as meat.

With multi-class classification, we could calculate its recall, precision, and f1-score for each class but we can also combine the f1-score of each class to have a single measure for the whole model. One of the ways to do that is called Micro F1. It is calculated by considering the total TP, total FP, and total FN of the model, so it calculates the metrics globally instead of individual classes [8].

- Total TP = (9 + 10 + 10) = 29
- Total FP = (0+0) + (1+0) + (0+0) = 1
- Total FN = (1+0) + (0+0) + (0+0) = 1

From the data above, we could get the value of precision, recall, and micro f1-score as follows:

$$Precision = \frac{29}{(29+1)} = 0.966 \quad (5)$$

$$Recall = \frac{29}{(29+1)} = 0.966 \quad (6)$$

$$Micro\ F1 = 2\,x\,\frac{(0.966\ x\ 0.966)}{(0.966+0.966)} = 0.966 \quad (7)$$

$$Accuracy = \frac{29}{30}\ x\ 100\% = 96.66\% \quad (8)$$

The accuracy of this application is 96.7%. This application can be said to have a fairly good accuracy because it has an accuracy rate of greater than 50%.

## V. Conclusions and Suggestions

Based on the research done above, the accuracy of classification of different side dish types (chicken, meat, and fish) based on color using the K-Nearest Neighbor method is high at 96.7%.

There are more than one testing image of fish that RGB value comes out as (255, 255, 255) which is the color white because that is the maximum peak value of the image. That is why there the data used for both testing and training should be clean and went through pre-processing first to ensure the data represent its image accurately.

Even though the accuracy is very high, we could tell that the training data has a hugely important role in classification accuracy. The training data used in this research have different sizes, pixels, and focus and that can largely impact the accuracy of the program. The training data should have the same dimensions and focus directly on the object. Pre-processing here could be used to enhance the images first before feature extraction. The texture of meat and different colored fish meat should be used to further increase the accuracy of this classification as well.

## VI. Acknowledgment

We would like to express my deep gratitude to my advisor for her patient guidance, advice and assistance in keeping my progress on schedule. We would also like to extend our thanks to all the previous person and team who published their research online. Without their proceeding and journal, this study wouldn't be done. Finally, we wish to thank our parents and peers for their support and encouragement throughout the study.

## REFERENCES

[2] M. You, J. Liu, J. Zhang, M. Xv, and D. He, "A novel chicken meat quality evaluation method based on color card localization and color correction," IEEE Access, vol. 8, pp. 170093–170100, 2020. doi:10.1109/access.2020.2989439

[3] I. Purwanto, M. Afriansyah, and K. Kusrini, "Deteksi Tingkat kesegaran daging ayam menggunakan k-nearest neighbor," CCIT Journal, vol. 12, no. 2, pp. 177–185, 2019. doi:10.33050/ccit.v12i2.688

[4] 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), 2019. doi:10.1109/hnicem48295.2019.9073468

[5] R. Bayraktar, B. A. Akgul, and K. S. Bayram, "Colour recognition using colour histogram feature extraction and K-nearest neighbour classifier," New Trends and Issues Proceedings on Advances in Pure and Applied Sciences, no. 12, pp. 08–14, 2020. doi:10.18844/gjpaas.v0i12.4981

[6] P. Hidayatullah, Pengolahan Citra Digital: Teori Dan Aplikasi Nyata. Bandung, West Java: Penerbit Informatika Bandung, 2017.

[7] L. Afifah, "Algoritma K-nearest neighbor (KNN) Untuk Klasifikasi," IlmudataPy, https://ilmudatapy.com/algoritma-k-nearest-neighbor-knn-untuk-klasifikasi/?amp=1 (accessed Sep. 2, 2022).

[8] GeeksforGeeks, "Confusion matrix in machine learning," GeeksforGeeks, https://www.geeksforgeeks.org/confusion-matrix-machine-learning/ (accessed Aug. 29, 2022).

[9] J. Mohajon, "Confusion matrix for your multi-class machine learning model," Medium, https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826 (accessed Sep. 2, 2022).