OPEN ACCESS
VOL. 7, NO. 1, PP.013-021, NOVEMBER 2024

ISSN : 2622-8106 (ONLINE)
DOI: 10.20895/INISTA.V7I1.1615

Journal of Informatics, Information System, Software Engineering and Applications (INISTA)

Implementation of The Weighted K-Nearest Neighbors Algorithm in The Classification of Beef and Pork Images

Nurdi Afrianto [1], Irzon Meiditra [*2]

Information Technology, Mitra Gama Institute of Technology
Babussalam,Mandau District, Bengkalis Regency, Riau Indonesia

[1] nurdiafrianto1995@gmail.com
[2] meiditrairzon@gmail.com

Received on 25-09-2024, revised on 18-10-2024, accepted on 24-10-2024

Abstract

The demand for meat in Indonesia is still high, especially for the consumption of beef and pork, which are important commodities in the market. Although meat provides essential nutrients, pork has health risks because it contains more than 40 dangerous pathogens and various bacteria. In traditional markets in Indonesia, the fraudulent practice of mixing pork and beef to gain greater profits is a serious problem. This is very detrimental to consumers, especially Muslims who do not consume pork. The study used machine learning, the Weighted K-nearest neighbor (WKNN) algorithm, to classify meat based on color features. The stages used began with collecting a dataset of 400 images and divided into 200 images of pork and beef for each. Images were taken using a Canon EOS Kiss X50 DSLR camera at ISO 100-200 for good image quality. Feature extraction uses HSV and RGB algorithms that focus on color. Furthermore, the data is divided into 70% for training and 30% for testing. The model was evaluated with a confusion matrix, namely accuracy, precision, and F1 score, which each produced an accuracy of 85%, 86%, and 80%. The research is updated on the application of WKNN for meat classification in traditional markets.

Keywords: Machine Learning, Pork, Beef, Weighted K-Nearest Neighbors

Corresponding Author:
* Nurdi Afrianto
Faculty of Computer Science and Engineering, Mitra Gama Institute of Technology
Babussalam, Mandau District, Bengkalis Regency, Riau, Indonesia
Email: nurdiafrianto1995@gmail.com

I. INTRODUCTION

Beef provides a high-quality source of protein and the demand for meat is increasing globally [1]. Because beef is the densest source of protein and contains high levels of iron, vitamin B12, and fatty acids, eating beef raises major health concerns [2]. The demand for meat in developing and developed countries is still very high [3]. In Indonesia, meat is typically sold based on its type, such as chicken, beef, and goat. Pork serves as a source of energy, providing essential macronutrients and micronutrients for human consumption. However, most research observations assess that it affects the risk of cancer [4].

African Swine Fever (ASF), one of the pig illnesses, is extremely contagious and has reached Indonesia [5]. Pork is an important food source that can harbor several pathogens, which are all spread via food. These pathogens include *Taenia solium, Trichinella spp., Brucella spp., non-typhoidal Salmonella enterica, E. coli that produces the Shiga toxin, and Campylobacter spp***.** [6]**.** *Salmonella* bacteria are a global health burden. *Salmonella* bacteria that attack the human intestine are mainly caused by pork [7]. Pork infectious diseases have been reviewed in 57,000 publications that identified more than 40 different pathogens as priority pathogens [8]. In addition, there is the African swine fever virus (ASFV), which causes a very deadly and contagious hemorrhagic disease in domestic pigs and wild boars [9]. Foodborne infections are linked to several pathogens, such as *Salmonella enterica, Yersinia enterocolitica, Campylobacter coli and C. jejuni, Escherichia coli* O157:H7, *Arcobacter butzleri* and *A. cryaerophila, Listeria monocytogenes*, and *Salmonella enterica*. *Pseudotuberculosis* can grow in pork, which is very dangerous for the human intestine. In addition, pork can be a transmission of *Taenia solium* worms (*Cysticercosis*) and *Trichinella spiralis,* as well as the protozoan parasite *Toxoplasma gondii,* which is very dangerous for humans [10]. In pork, there are *Salmonella bacteria*, which are negative pathogens that attack the human intestine. Additionally, approximately 50% of global food poisoning incidents are attributed to bacteria linked to *Salmonella infections* [11]**.**

Unscrupulous traders often mix beef and pork for profit. The mixing of beef and pork is difficult for buyers to distinguish and is detrimental to the surrounding community at Jalan Saleh Abas Market, Pekanbaru. Machine learning technology can distinguish beef and pork through classification. The Weighted K-nearest Neighbors (WKNN) algorithm is superior to ordinary K-nearest Neighbors (KNN). The problem with the K-nearest neighbors (KNN) algorithm is that after calculating the distance between new data and existing data, the class of the K nearest neighbors is examined to determine the class of the latest data. The problem with the KNN algorithm is that its distance metric does not consider the relevance of features to classification. All features are deemed to have the same contribution in determining the nearest neighbor. The Weighted K-nearest Neighbors (WKNN) algorithm is used to overcome this problem by giving different weights to each feature so that it can perform better classification [12]. The Weighted K-nearest Neighbors (WKNN) algorithm gives more weight to the nearest neighbors, which enhances accuracy, particularly when distant neighbors are less likely to be relevant for classification [13]. This study aims to analyze the identification of beef and pork by utilizing the Weighted K-nearest Neighbors (WKNN) algorithm to help the Muslim community, especially in the Jalan Saleh Abas market, Pekanbaru, in distinguishing the two types of meat when buying and choosing at the market.

II. RESEARCH METHOD

The research method in the study consists of several processes. Each stage of the study guides the research process to be more structured and follow what is desired. The research stages consist of data collection, data preprocessing, modeling, and testing. Figure 1 shows the research flowchart.

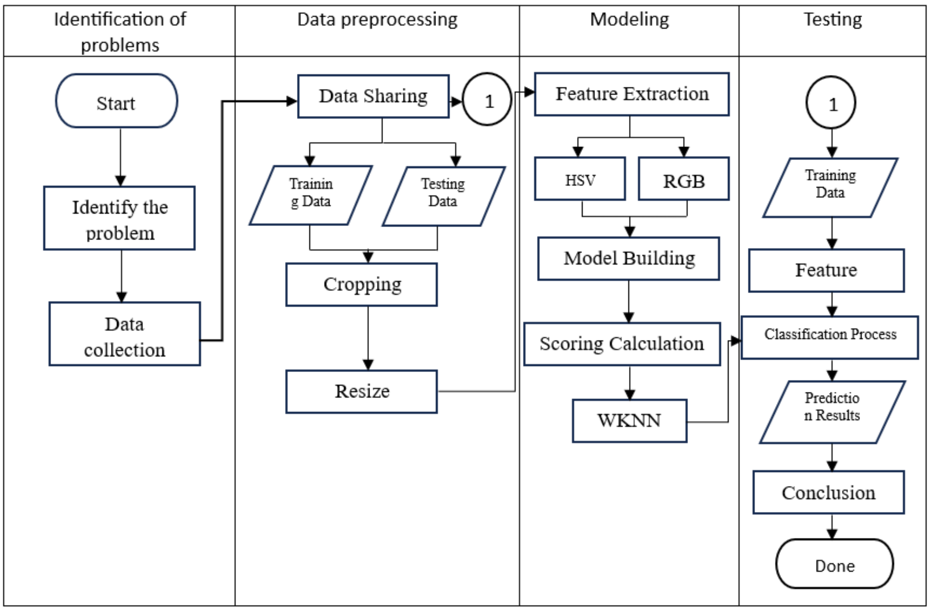

Fig. 1. Research stages

The research data were taken from research [14]. In this study, we purchased 1 kg of beef and pork at the Pasar Bawah Pekanbaru Tourist Market on Jalan Saleh Abas in Kampung Dalam Village, Senapelan District, Pekanbaru City, Riau and at the Pasar Bawah Pekanbaru Tourist Market on Jalan Saleh Abas in Kampung Dalam Village, Senapelan District, Pekanbaru City, Riau respectively.

From those meats, we take the data in the form of images taken using a CANON EOS Kiss X50 DSLR camera set to ISO 100 to 200, using the.JPG picture extension, and a shooting distance of ± 5 to 15 cm between the camera and the item. The camera was chosen because it can capture high-quality images at ISO 100 to 200, which ensures optimal lighting. Good lighting is essential for color feature extraction in HSV. Data will be gathered using an object positioned in the center of a white background. Data preprocessing is carried out after the image data is obtained.

The preprocessing method used in this study is cropping and resizing. The cropping process in this study was carried out manually using Photoshop CS6 to determine precisely the part of the image that contained the research object. Next, the cropped data undergoes a resizing process to adjust the image dimensions to the desired pixel size, which helps accelerate the computation process. The image pixels used in this study are 400 x 400.

The next step is extraction to obtain relevant information from the beef and pork data to collect the characteristics from the images. This study uses the Hue, Saturation, Value (HSV) and Gray Level Co-occurrence Matrix (GLCM) algorithms. The HSV algorithm is used to obtain the best color features based on the advantages of the HSV color space in modeling human perception of color. At the same time, the GLCM algorithm is used to obtain texture feature extraction from the meat. After obtaining the color and texture features of the meat, the modeling process is carried out. Modeling uses the Weighted K-nearest neighbors (WKNN) classification algorithm. The Weighted K-nearest Neighbors (WKNN) algorithm was chosen because of its ability to handle unstructured data in the classification process. Thus, the best classification results are obtained from beef and pork images.

A. Image Data

This study used beef and pork as data sources. Meat data were obtained from Cipuan Market and the lower market in Pekanbaru City. Table I shows the image of beef and pork.

TABLE I. IMAGE OF PORK AND BEEF

No.	Image Pork	Image Beef
1.		
2.		
3.		
4.		
5.		

The images of beef and pork were captured with a Canon EOS KISS X50 DSLR camera set at an ISO range of 100-200. The distance used was between 5-15 cm. The background of the image used is white. Also, the formatted data is presented in *.JPG format. The dataset for this study consists of 200 images of beef and pork each.

B. Data Preprocessing

Preprocessing is an important stage in image processing. This stage aims to prepare raw images into images that are ready to be processed by the algorithm. The image cropping phase is illustrated in Figure 2. Figure 3 shows the results of resizing data preprocessing to improve computing efficiency into 400 x 400 pixels.

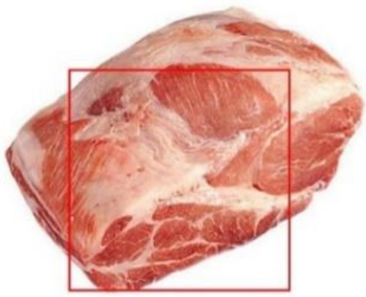

Fig. 2. Cropping Image

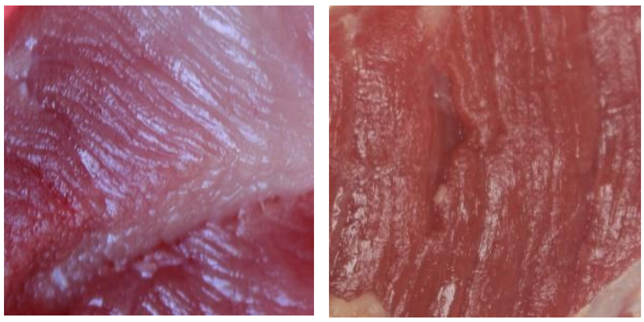

Fig. 3. Resized Image

C. RGB Color

RGB is a color space that relies on the Cartesian coordinate system, with colors represented as points determined by vectors originating from the origin [15]. RGB color space is a color model that uses three components (Red, Green, and Blue) to represent a wide spectrum of colors. Each color in RGB space can be defined by a linear combination of these three components. The mapping for all pixels in RGB space, namely the RGB cube, can be seen in Figure 4.

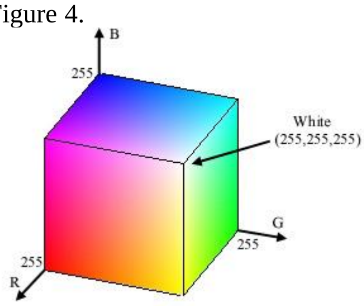

Fig. 4. RGB colors (Red, Green, Blue)

RGB Cube has 256x256x256 pixels with RGB values pixel (255, 0, 0) indicating red and pixel value (0, 255, 255) indicating cyan [16]. RGB color cube represents the visual display of all colors that the RGB color system can produce. By combining RGB values, it can create various color effects such as tone, saturation, and gradient.

D. Hue Saturation Value (HSV)

HSV is a cylindrical coordinate system that represents points in the Red, Green, and Blue (RBE) color model [17]. HSV can rearrange the geometry of Red, Green, and Blue (RGB) for more relevant perceptual results compared to the Cartesian coordinate representation. Figure 5 shows the HSV color model.

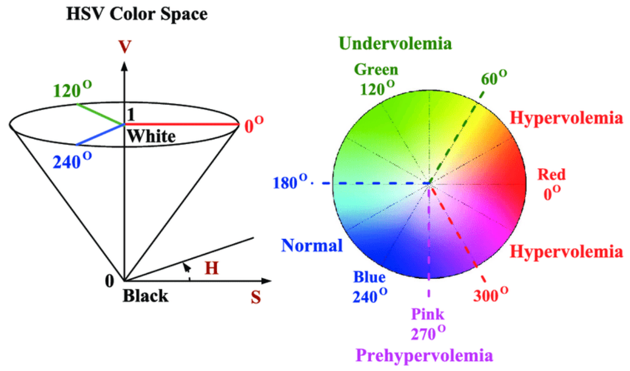

Fig. 5. Saturation Value (HSV) Color

Hue represents pure colors such as Green, Red, or Magenta and spans a range from 0 degrees to 360 degrees. Saturation refers to the intensity of color purity. When the value ranges from 0% (gray) to 100% (pure color), it indicates the level of color purity. Value can be expressed in percentages (05 TO 100%). A value of 0% indicates black, while 100% indicates white [18]. Image transformation from RGB to HSV color space using Equations (1), (2), (3).

$$H = arctan\left\{\frac{\sqrt{3}\,(G-B)}{(R-G)+(R-B)}\right\} \tag{1}$$

$$S = 1 - \frac{\min(R,G,B)}{V} \tag{2}$$

$$V = \frac{R+G+B}{3} \tag{3}$$

Equations (4), (5), and (6) are used to determine the RGB value.

$$r = \frac{R}{R+G+B} \tag{4}$$

$$g = \frac{G}{R+G+B} \tag{5}$$

$$b = \frac{B}{R+G+B} \tag{6}$$

Equations (7), (8), (9), and (10) are used to transform the normalized RGB values to HSV after the RGB normalization procedure is complete.

$$v = \max(r,g,b) \tag{7}$$

$$s = \begin{cases} 0 \\ v \end{cases} \tag{8}$$

$$H = \begin{cases} 0 & if\ s = 0 \\ 60^0\ x\left[0 + \frac{(G-B)}{S\ x\ V}\right] & if\ v = r \\ 60^0\ x\left[2 + \frac{b-r}{s\ x\ v}\right] & if\ v = g \\ 60^0\ x\left[4 + \frac{r-g}{s\ x\ v}\right] & if\ v = b \end{cases} \tag{9}$$

$$H = H + 360 \qquad if\ H < 0 \tag{10}$$

The Equation uses the following notations: V stands for the highest brightness level; S for color saturation; and H for Hue, which describes the color inside the domain. Equations 9 and 10 explain how to calculate the hue value H, which depends on the Red, Green, and Blue (RGB) components. If the value v is equal to r (red is the largest), then the color tends to be closer to red. The difference between Green G and Blue B is calculated to determine how far the red color is closer to green or blue. If green G is greater than blue B, then Hue H moves towards green-yellow. If the value v is equal to g (green is the largest), then the difference between blue B and red R is used in the calculation and then multiplied by 60 degrees. Because green g is the largest component, the first term in Equation 2 is to ensure the Hue position is in the correct part of the color wheel. If the value v is equal to B (blue is the largest), then the calculation is done using the difference between red R and green G. The first term in Equation 4 is because the Hue in blue is in the fourth quadrant of the color wheel. After calculating the Hue, adjustments are made so that the obtained value is 0 to 360 degrees. If the computed H value is negative, then 360 is added to ensure that the color wheel positively charges the value. This Equation ensures that H is always in the range between 0-360 degrees because negative hue values are not in the HSV color model.

E. Weighted K-Nearest Neighbors (WKNN)

The Weighted K-nearest neighbors (WKNN) algorithm is a variant of the K-nearest neighbors (KNN) algorithm. Weighted K-nearest Neighbors (WKNN) is an instance-based algorithm where classification decisions are based on several neighbors nearest to unknown data. The Weighted K-Nearest Neighbors (WKNN) method can perform good classification because the method is not much different from the KNN algorithm. The difference is the weighting for each type/class in the score calculation process to determine the classification of the test data [19]. Steps for calculating the Weighted K-Nearest Neighbors (WKNN) algorithm [12]:

- Determine the K parameter
- The distance between the new sample and all other samples is calculated one by one
- The calculated distances are sorted from smallest to largest, and the smallest k is selected among these distances
- The weight of the selected k samples is determined by calculating using Equation (12)
- The weights of the same classes are summed and the class of the new sample is determined by looking at the total of the nearest neighboring classes.

$$w = 1 / d^2 \tag{12}$$

III. RESULTS AND DISCUSSION

This study aims to classify images of beef and pork using color feature extraction. Hue, Saturation, Value (HSV), and Red, Green, and Blue (RGB) algorithms are used for color feature extraction. The dataset is divided into 30% for data testing, while the remaining 70% is used for training. Thirty percent of the testing data consists of 60 images taken by a camera, while seventy percent consists of 140 images to train the model. The HSV color feature is chosen as a visual representation of meat for the classification process. In this study, image data used as many as 400 images obtained from beef and pork. Color features are extracted to identify the classification of pork and beef. Part of what determines the maturity of beef and pork is color. RGB and HSV feature extraction are the methods used to extract color features. Table 1 shows the results of RGB feature extraction, and Table II shows the results of HSV feature extraction.

TABLE II. HSV COLOR FEATURES

No.	H	S	V	Label
1.	123.856100	78.882456	162.835475	pork
2.	138.569206	107.725962	161.722231	beef
3.	125.048519	65.558594	147.626900	pork
4.	118.392538	124.607331	172.860456	beef
5.	121.219706	85.550275	158.358675	pork
….	……	…..	……	Beef
396.	118.915406	124.247669	156.449069	pork
397.	134.658206	77.275394	151.954612	beef
398.	119.405294	125.430988	160.012200	pork
399.	122.245719	58.704475	204.895788	beef
400.	138.299725	109.204475	160.600550	pork

TABLE III. RGB COLOR FEATURES

No.	R	G	B	Label
1.	119.047506	113.531787	162.834775	pork

No.	R	G	B	Label
2.	131.469681	95.355738	159.209487	beef
3.	116.728906	111.117050	147.536837	pork
4.	90.083975	94.680750	172.860456	beef
5.	108.102706	106.223156	158.358675	pork
….	……	…..	……	Beef
396.	82.019006	85.044188	156.449069	pork
397.	123.373100	107.029506	150.627500	beef
398.	82.959137	84.896144	160.012200	pork
399.	162.096950	159.219750	204.895563	beef
400.	129.518781	93.944569	157.994050	pork

We carry out model testing using images of beef and pork after the feature extraction process is complete. In this test, 70% of the data is used to train the model, while 30% is used to test the model. The test was carried out using the Weighted K-nearest Neighbor (WKNN) algorithm with a predetermined K value. A confusion matrix is used to calculate the model performance evaluation and its accuracy. The test results show that the Weighted K-nearest Neighbor (WKNN) algorithm using color feature extraction managed to obtain an accurate value of 85% with K = 1 and 60 test data. The confusion matrix produces a precision value analysis of 86%, indicating the proportion of correct positive predictions from all positive predictions. The recall value is 81%, which indicates the model's ability to identify positive cases correctly. The F1-Score value is 80%. Table III shows the complete matrix. However, the accuracy results obtained are still low compared to previous studies that used K-nearest neighbor (KNN) for tomato classification [20]. The accuracy obtained was 91.25%. The decrease in accuracy can be caused by differences in the training and testing process of the data, as well as the number of trials conducted.

TABLE IV. CONFUSION MATRIX

	Beef prediction	Pork meat prediction
actual beef	49	11
actual pork	7	53

Based on the confusion matrix shown in Table IV, the level of accuracy can be formulated using Equation (13).

$$Accuracy = \frac{49+53}{49+53+7+11} = \frac{102}{120} = 0.85\ (85\%) \tag{13}$$

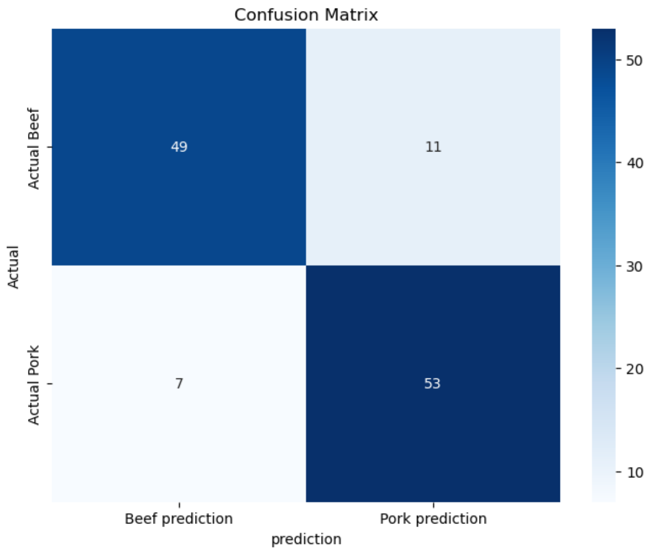

Fig. 6. Confusion Matrix

Figure 6 shows the plot of the results of the confusion matrix. Each box in the plot shows the number of predictions made by the model. Darker boxes indicate more correct predictions, while lighter boxes indicate fewer. Out of 60 beef data points, the model successfully predicted 49 correctly. While out of 60 pork data, the model was able to predict 53 correctly.

IV. CONCLUSION

This study shows that the Weighted K-nearest Neighbors (WKNN) algorithm can classify beef and pork well. The accuracy rate is 85%, precision 86%, recall 81%, and F1-score 80%. These results indicate that the model has stable performance in detecting both types of meat with a low error rate. The precision score illustrates that most of the positive predictions from the model are correct, so the risk of error in identifying pork and beef is relatively low. The recall value of 81% indicates that the model can detect beef and pork effectively. The F1-Score obtained is 80%, which means a balance between precision and recall; this illustrates the solid performance of the model. For further research, it is recommended to add additional variables from the K-nearest neighbors (KNN) algorithm, such as using Grid Search or Random Search to find the best combination of weights on different features. Combining the Weighted K-nearest neighbors (WKNN) algorithms with other algorithms, such as SVM, can help improve the shortcomings of each algorithm. Conducting experiments with varying compositions of training data and test data, as well as varying the K value, is also expected to increase the diversity of results.

ACKNOWLEDGMENT

Nurdi Afrianto and Irzon Meiditra wrote this journal from the Faculty of Computer Science and Engineering, Institut Teknologi Mitra Gama (ITMG), based on research entitled Implementation of the Weighted K-nearest neighbors (WKNN) Algorithm in the Classification of Beef and Pork Images. This research was funded by the Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRPTM) *Kemenristekdikti* through the Research and Community Service program for the 2024 fiscal year. The entire contents of the article are the author's full responsibility.

REFERENCES

[1] P. L. Greenwood, " Review: An Overview of Beef Production From Pasture and Feedlot Globally, As Demand For Beef And The Need For Sustainable Practices Increase," *Animal*, vol. 15, no. 1, p. 100295, 2021, doi: 10.1016/j.animal.2021.100295.

[2] N. Johnson, K. Truatman, S. Stastny, R. McGrath, and K. Hackney, "Beef Consumption and Functional Performance in Middle-Aged and Older Adults: A Systematic Review," vol. 3, no. 1, pp. 18–31, 2019.

[3] H. Dagevos, " Finding Flexitarians: Current Studies on Meat Eaters and Meat Reducers," *Trends Food Sci. Technol.*, vol. 114, no. June, pp. 530–539, 2021, doi: 10.1016/j.tifs.2021.06.021.

[4] L. P. Penkert, R. Li, J. Huang, A. Gurcan, M. Chung, and T. C. Wallace, "Pork Consumption and Its Relationship to Human Nutrition and Health: A Scoping Review," *Meat Muscle Biol.*, vol. 5, no. 1, 2021, doi: 10.22175/mmb.12953.

[5] E. Mighell and M. P. Ward, " African Swine Fever Spread Across Asia, 2018–2019," *Transbound. Emerg. Dis.*, vol. 68, no. 5, pp. 2722–2732, 2021, doi: 10.1111/tbed.14039.

[6] T. Kayano, J. Pulford, and L. Thomas, "Identifying Pig- and Pork-Associated Zoonotic and Foodborne Hazards in Eastern and Southern Africa: A Systematised Review," *Zoonotic Dis.*, vol. 3, no. 2, pp. 120–133, 2023, doi: 10.3390/zoonoticdis3020011.

[7] L. Soliani, G. Rugna, A. Prosperi, C. Chiapponi, and A. Luppi, "Salmonella Infection in Pigs: Disease, Prevalence, and a Link between Swine and Human Health," *Pathogens*, vol. 12, no. 10, 2023, doi: 10.3390/pathogens12101267.

[8] S. Pfuderer, R. M. Bennett, A. Brown, and L. M. Collins, " A Flexible Tool for the Assessment of the Economic Cost of Pig Disease in Growers and Finishers at Farm Level," *Prev. Vet. Med.*, vol. 208, no. September, p. 105757, 2022, doi: 10.1016/j.prevetmed.2022.105757.

[9] E. Radulovic *et al.*, " The Baseline Immunological and Hygienic Status of Pigs Impact Disease Severity of African Swine Fever," *PLoS Pathog.*, vol. 18, no. 8, pp. 1–22, 2022, doi: 10.1371/journal.ppat.1010522.

[10] N. Bor *et al.*, "Prevalence of Antibiotic Residues in Pork in Kenya and the Potential of Using Gross Pathological Lesions as a Risk-Based Approach to Predict Residues in Meat," *Antibiotics*, vol. 12, no. 3, 2023, doi: 10.3390/antibiotics12030492.

[11] R. G. Ferrari, D. K. A. Rosario, A. Cunha-Neto, S. B. Mano, E. E. S. Figueiredo, and C. A. Conte-Juniora, "Worldwide epidemiology of Salmonella serovars in animal-based foods: A meta-analysis," *Appl. Environ. Microbiol.*, vol. 85, no. 14, 2019, doi: 10.1128/AEM.00591-19.

[12] F. Tarakci and A. Ozkan, "Comparison of classification performance of kNN and WKNN algorithms," *Selcuk Univ. J. Eng. Sci.*, vol. 20, no. 02, pp. 32–37, 2021, [Online]. Available:

http://sujes.selcuk.edu.tr/sujes.

[13] C. Dewi and Y. K. Arbawa, "Performance Evaluation of Distance Function in KNN and WKNN for Classification of Soil Organic Matter," *Proc. 2019 4th Int. Conf. Sustain. Inf. Eng. Technol. SIET 2019*, pp. 196–199, 2019, doi: 10.1109/SIET48054.2019.8986030.

[14] A. M. Priyatno, F. M. Putra, P. Cholidhazia, and L. Ningsih, "Combination of Extraction Features Based on Texture and Colour Feature for Beef and Pork Classification," *J. Phys. Conf. Ser.*, vol. 1563, no. 1, 2020, doi: 10.1088/1742-6596/1563/1/012007.

[15] F. García-Lamont, J. Cervantes, A. López-Chau, and S. Ruiz-Castilla, "Color Image Segmentation Using Saturated RGB Colors and Decoupling the Intensity from the Hue," *Multimed. Tools Appl.*, vol. 79, no. 1–2, pp. 1555–1584, 2020, doi: 10.1007/s11042-019-08278-6.

[16] Z. Yan, L. Xu, A. Suzuki, J. Wang, J. Cao, and J. Huang, "RGB Color Model Aware Computational Color Naming and Its Application to Data Augmentation," *Proc. - 2022 IEEE Int. Conf. Big Data, Big Data 2022*, pp. 1172–1181, 2022, doi: 10.1109/BigData55660.2022.10020750.

[17] Y. A. Anwar, B. E. W. Asrul, and S. Zuhriyah, "Implementasi Algoritma Hue Saturation Value (HSV) Pada Penentuan Kualitas Beras Berbasis Android," *Pros. Semin. Nas. Sist. Inf. dan Teknol.*, pp. 139–143, 2023.

[18] S. Simbolon *et al.*, "Image Segmentation using Color Value of the Hue in CT Scan Result," *J. Phys. Conf. Ser.*, vol. 2394, no. 1, 2022, doi: 10.1088/1742-6596/2394/1/012017.

[19] S. Dewi Purba and P. Sirait, "Credit Card Risk Classification Using K-Nearest Neighbor Weighted Algorithm Based on Forward Selection," *J. Mantik*, vol. 4, no. 3, pp. 1551–1559, 2020, [Online]. Available: https://iocscience.org/ejournal/index.php/mantik.

[20] L. Ningsih and P. Cholidhazia, "Classification Of Tomato Maturity Levels Based on RGB And HSV Colors Using KNN Algorithm," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 1, no. 1, pp. 25–30, 2022, doi: 10.31004/riggs.v1i1.10.