Budy Satria

ok-3736-11630-1-CE-AAP

SINTA 2

Document Details

Submission ID

trn:oid:::3618:107955493

Submission Date

Aug 11, 2025, 10:46 PM GMT+7

Download Date

Aug 11, 2025, 10:55 PM GMT+7

File Name

ok-3736-11630-1-CE-AAP.docx

File Size

1.1 MB

12 Pages

9,010 Words

48,588 Characters

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 8 words)

Exclusions

- 1 Excluded Match

Match Groups

103 Not Cited or Quoted 15%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

10% Internet sources
10% Publications
9% Submitted works (Student Papers)

Match Groups

- **103** Not Cited or Quoted 15%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

10% Internet sources
10% Publications
9% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Publication	Viancy Vincent, Gobalakrishnan Natesan. "GAN-enhanced E-nose analysis: VTAAE ...	<1%
2	Internet	www.mdpi.com	<1%
3	Internet	jurnalsaintek.uinsby.ac.id	<1%
4	Publication	Abdull Halim Anshor, Tri Ngudi Wiyatno. "Comparative Analysis of Random Fores...	<1%
5	Publication	Lidya Ningsih, Putri Cholidhazia. "Classification Of Tomato Maturity Levels Based ...	<1%
6	Internet	dokumen.pub	<1%
7	Student papers	Stony Brook University on 2024-06-21	<1%
8	Internet	www.atlantis-press.com	<1%
9	Student papers	University of Auckland on 2024-08-06	<1%
10	Student papers	fpptijateng on 2023-09-14	<1%

11	Student papers	
Florida Polytechnic University on 2016-11-01		<1%
12	Internet	
kuwaitjournals.org		<1%
13	Internet	
repository.ipb.ac.id		<1%
14	Internet	
www.sciencegate.app		<1%
15	Student papers	
Institut Pertanian Bogor on 2025-05-26		<1%
16	Student papers	
Syntax Corporation on 2024-10-31		<1%
17	Internet	
archive.conscientiabeam.com		<1%
18	Internet	
bright-journal.org		<1%
19	Student papers	
American University of the Middle East on 2023-04-30		<1%
20	Publication	
Jiangshan Liu, Fengting Li, Chunya Yin, Ruikang Chen, Lu Han, Yingping Shi. "Pow...		<1%
21	Student papers	
University of Gibraltar on 2024-12-15		<1%
22	Internet	
repository.library.noaa.gov		<1%
23	Internet	
www.icaset.in		<1%
24	Student papers	
Curtin University of Technology on 2021-09-17		<1%

25 Publication

Fahim K Sufi, Musleh Alsulami. "Automated Multidimensional Analysis of Global E... <1%

26 Publication

Orhan Yaman. "An automated faults classification method based on binary patte... <1%

27 Student papers

Rose-Hulman Institute of Technology on 2024-07-31 <1%

28 Student papers

Universitas Muslim Indonesia on 2024-11-21 <1%

29 Student papers

Universitas Muslim Indonesia on 2025-06-08 <1%

30 Internet

proceedings.stis.ac.id <1%

31 Internet

www.fastercapital.com <1%

32 Publication

Dominik Csik, Akos Odry, Peter Sarcevic. "Comparison of RSSI-Based Fingerprinti... <1%

33 Publication

Kamda Delly Sunggara, Aviv Yuniar Rahman, Istiadi. "Fault Detection of Oil Palm ... <1%

34 Publication

Mikha Dayan Sinaga, Nita Sari Br Sembiring, Erni Darmayanti, Rizki Maulida, Edy ... <1%

35 Publication

Triyanna Widiyaningtyas, Ilham Ari Elbaith Zaeni, Putri Yula Wahyuningrum. "Self... <1%

36 Student papers

University of Liverpool on 2021-12-06 <1%

37 Internet

journal.unipdu.ac.id <1%

38 Internet

www.peeref.com <1%

39	Publication	
"Communication and Information Technologies through the Lens of Innovation", ...		<1%
40	Publication	
Bambang Sugiarto, Christy Atika Sari, M. Arief Soeleman. "KNN Algorithm Optimi...		<1%
41	Publication	
Sandhya Makkar, Gobinath Ravindran, Ripon Kumar Chakrabortty, Arindam Pal. "...		<1%
42	Student papers	
Vanderbilt University on 2023-02-07		<1%
43	Internet	
docs.qgis.org		<1%
44	Internet	
link.springer.com		<1%
45	Internet	
repositori.upf.edu		<1%
46	Internet	
scholarhub.uny.ac.id		<1%
47	Internet	
www.coursehero.com		<1%
48	Publication	
Anita Desiani, Sigit Priyanta, Indri Ramayanti, Bambang Suprihatin, Muhammat R...		<1%
49	Publication	
Assia Boukhamla, Mohamed Hatem Bouziane, Akram Laib, Nabiha Azizi, Roumais...		<1%
50	Student papers	
Clarkston Community Schools on 2023-05-05		<1%
51	Publication	
Fereidoon Shahidi, Leon J. Rubin, Lorraine A. D'Souza, Roy Teranishi, Ron G. Butte...		<1%
52	Student papers	
North South University on 2021-08-22		<1%

53 Publication

Novianti Puspitasari, Hamdani Hamdani, Heliza Rahmania Hatta, Anindita Septiar... <1%

54 Publication

Shrikaant Kulkarni, P. William, Vijaya Prakash, Jaiprakash Narain Dwivedi. "Lever... <1%

55 Student papers

The University of Manchester on 2024-04-17 <1%

56 Student papers

University of Portsmouth on 2017-04-27 <1%

57 Internet

eltikom.poliban.ac.id <1%

58 Internet

publikasi.lldikti10.id <1%

59 Internet

www.cobbsmeatprocessing.com <1%

60 Student papers

Arab Open University on 2025-05-15 <1%

61 Publication

Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelli... <1%

62 Student papers

Associatie K.U.Leuven on 2022-06-13 <1%

63 Student papers

Auckland University of Technology on 2010-06-01 <1%

64 Publication

Cyril Chinonso Ejidike, Modupe Cecilia Mewomo, Iruka Chijindu Anugwo. "Assess... <1%

65 Student papers

Erasmus University of Rotterdam on 2022-08-27 <1%

66 Student papers

Georgia Institute of Technology Main Campus on 2024-09-23 <1%

67 Student papers

Glyndwr University on 2025-08-04 <1%

68 Student papers

Karabük Üniversitesi on 2023-01-24 <1%

69 Student papers

South Bank University on 2024-03-25 <1%

70 Student papers

Universitas Putera Indonesia YPTK Padang on 2021-07-07 <1%

71 Student papers

University of Strathclyde on 2025-06-16 <1%

72 Student papers

University of Sydney on 2024-10-17 <1%

73 Student papers

University of Wales Swansea on 2024-09-30 <1%

74 Internet

ijsred.com <1%

75 Internet

iris.unive.it <1%

76 Internet

journal.uad.ac.id <1%

77 Internet

publications.scss.tcd.ie <1%

78 Internet

repository.universitasbumigora.ac.id <1%

79 Internet

www.americaspg.com <1%

80 Internet

www.nature.com <1%

81 Internet

www.ncbi.nlm.nih.gov <1%

4 Comparative Analysis of Weighted-KNN, Random Forest, and Support Vector Machine Models for Beef and Pork Image Classification Using Machine Learning

Budy Satria[a*], Nurdi Afrianto[b], Lidya Ningsih[c], Putri Sakinah[d], Acihmah Sidauruk[e], Liga Mayola[f]

[a]Department of Informatics, Universitas Andalas, Padang, 25175, Indonesia
[b]Department of Information Technology, Institut Teknologi Mitra Gama, Riau, 28784, Indonesia
[c]Department of Informatics, Universitas Telkom, Bandung, 40257, Indonesia
[d]Department of Informatics, Universitas Adzkia, Padang, 25175, Indonesia
[e]Department of Information System, Universitas Amikom, Yogyakarta, 55283, Indonesia
70 *[f]Department of Information System, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia*
*Corresponding author: *budy.satria@it.unand.ac.id*

***Abstract*—The actual problem that occurs in the sale of meat by some conventional market traders is mixing beef with pork because of**
14 **the high selling price. The difference between pork and beef lies in the color and texture of the meat. However, many people do not**
60 **understand this difference. This study aims to provide a solution to distinguish the two types of beef through a classification process by**
obtaining the best accuracy using the W-KNN, RF, and SVM models based on machine learning. This study compares the model's
73 **performance based on the number of datasets, comprising 400 original images (200 beef and 200 pork images), using a 80:20 ratio for**
43 **training and test data. The extraction process uses two algorithms: HSV (Hue, Saturation, Value) and RGB (Red, Green, Blue). The**
25 **model evaluation uses a confusion matrix that includes accuracy, Precision, Recall, and F1-score. Based on the results of the model**
1 **testing, it was found that the random forest algorithm gave the best overall results, with the highest accuracy of 98.75%, Precision of**
18 **97%, F1-score of 98%, and Recall of 99% on the number of decision trees of 400. This shows the stability and generalization of the**
superior model. The random forest algorithm is the most effective for classifying beef and pork data with minimal errors. Implications for further research include using a deep learning approach, especially for image processing, to detect differences in each meat characteristic and increase accuracy.

***Keywords*— Beef and pork image; classification; machine learning; w-KNN; random forest; support vector machine.**

Manuscript received 15 Oct. 2020; revised 29 Jan. 2021; accepted 2 Feb. 2021. Date of publication 17 Feb. 2021.

I. INTRODUCTION

Seeing the difference between beef and pork is a complicated challenge, especially for consumers unfamiliar with the characteristics of both types of meat. Although visually, beef and pork look similar, there are differences based on texture and color that are difficult to distinguish. This situation will be a serious problem when fraudulent practices, such as mixing beef with pork, disturb consumers who buy and consume the meat [1]. The consequences of fraudulent practices of mixing beef and pork have the potential to endanger consumer health and reduce buyer confidence in the meat market [2]. This problem can be solved with a machine learning-based approach, especially
23 classification algorithms such as Decision Tree, Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest, which offer practical solutions. By segmenting data based on the most relevant features, decision trees clearly understand how decisions are made in classification [3].

The Logistic regression algorithm has the advantage of classifying two classes with high probability, although it is limited to more linear data [4]. The KNN algorithm is also a machine-learning model that finds data proximity based on the distance between points, so it is suitable for classifying meat types with the same visual characteristics [5]. The SVM model can optimize the separation margin between two classes and handle subtle differences that are difficult for humans to distinguish [6]. Random forest is also an ensemble method that uses many decision trees, effectively reducing overfitting and increasing classification accuracy [7]. With the machine learning algorithm, classifying beef and pork can be done effectively and efficiently, reducing the potential for fraudulent meat sales practices and ensuring better food safety for buyers in the market [8].

Beef is a source of high-quality protein [9]. Significant health benefits are obtained when consuming beef because it is a protein source rich in iron, vitamin B12, and essential fatty acids [10]. In developed countries, purchasing power for meat reaches high levels [11]. Like in Indonesia, meat is generally sold based on categories such as chicken, beef, and goat meat. In addition, pork sales are available for those who wish to consume it. However, some studies show that consuming pork can increase the risk of cancer [12]. Pork contains Ochratoxin A (OTA), which is a dangerous mycotoxin that has toxic effects on the kidneys, liver, and nervous system, damages the immune system, and can cause congenital disabilities and cancer in humans [13].

One case of African Swine Fever (ASF) disease has been detected in Indonesia due to pork consumption [14]. Pork, as an essential food source, can be a medium for various pathogens, such as Brucella spp., non-typhoid Salmonella enterica, Shiga toxin-producing E. coli, and Campylobacter spp., which are often spread during consumption [15]. One of the major pathogens, Salmonella spp., is a significant global health threat [16]. The primary source of bacterial infection in the human body has been successfully identified in pork [17]. Based on an analysis of more than 57,000 publications, more than 40 pathogens are considered among the priority threats related to beef consumption [18]. Moreover, the lipid content and number of Enterobacteriaceae and other harmful microbes are higher in the open market [19]. Foodborne illnesses are often associated with a variety of pathogens, including Salmonella enterica, Yersinia Enterocolitica, Campylobacter coli, C. Jejuni, Escherichia coli O157:H7, Arcobacter Butzleri, A. Cryaerophila, Listeria monocytogenes, and Salmonella enterica. Pork can cause the growth of Pseudotuberculosis, which is dangerous for human intestinal health. In addition, pork can also transmit Taenia Solium worms, Trichinella spiralis, and the protozoan parasite Toxoplasma gondii [20]. Salmonella bacteria found in pork are known as pathogens that attack the human digestive system. About 50% of food poisoning cases globally are caused by bacteria associated with Salmonella infections [21].

The problem is that some irresponsible meat sellers often mix beef with pork for buyers. The texture of the two meats, which look almost the same, makes it difficult for consumers to distinguish between beef and pork when buying and consuming. The study was conducted in Pekanbaru, specifically at the Jalan Saleh Abas Market. With the advancement of technology, machine learning provides a solution to distinguish between the two types of meat through the classification process.

Research on beef and pork identification has been previously conducted using Artificial Neural Networks based on Texture Features [22]. Based on the MOP-NN Model test results, identifying digital images of beef and pork has a performance with an accuracy of 96% at 400. Image-based research on beef and pork has been conducted with the aim of predicting chemical concentrations in meat samples that can be useful in determining the quality and safety of pork and beef [23]. The study results show that pork and beef are the most commonly consumed meats worldwide. Therefore, it is crucial to rigorously evaluate the quality and safety of pork and beef, and automated detection tools are needed to ensure the quality and differentiation of meat types based on images. Research using CNN for beef and pork image classification was conducted with 450 test data [24]. According to the model, 218 pork images and 221 beef images were classified correctly. Other studies have also been conducted to classify pictures of sliced beef and pork based on color features [25]. Research on Beef and Pork Image Classification using EfficientNet B0 Feature Extraction and Visual-Based Ensemble Learning has been conducted. A dataset of 400 images, divided equally for beef and pork, obtained a 99.0% accuracy and an ROC-AUC of 0.995 [26].

One of the models with advantages in classification is Weighted K-nearest neighbors (WKNN), which has advantages over the conventional K-nearest Neighbors (KNN) algorithm. In the KNN algorithm, after calculating the distance between new data and old data, the system determines the class of new data based on the K nearest neighbors. However, the K-nearest Neighbors algorithm has weaknesses. The main weakness lies in the assumption that all features make the same contribution in determining the neighborhood without considering the relevance of each feature to the classification. WKNN has a new weighting scheme that utilizes the familiar k-nearest neighbor approach, where the weights assigned to data points used for training are determined by the squared inverse of their distance from the query point [27]. In addition, the Weighted K-nearest Neighbors algorithm gives greater weight to the nearest neighbors, which are more relevant in the classification process than the more distant neighbors [28].

This study also uses the Random Forest algorithm to compare the performance of each algorithm. The random forest algorithm is an ensemble-based machine learning algorithm that produces stable prediction models. Random forests can reduce the risk of overfitting, which is often experienced by resident trees, and improve the reliability of predictions in classification and regression [29]. In addition to overcoming overfitting, other advantages of random forests include flexibility in handling heterogeneous data, tolerance for missing values and outliers, and providing insight into the importance of features [30].

Finally, the Support Vector Machine Algorithm is also used to analyze the model's performance in this study. The support vector machine algorithm excels in handling high-dimensional datasets, both linearly and non-linearly separable, by utilizing linear polynomial kernels or radial basis functions (RBF) [31]. SVM can produce a maximum separation margin, which improves the model's generalization on new data and its resistance to overfitting on small datasets. [32]. In addition, SVM can handle outlier data using large margins and soft margins to tolerate non-conforming data [33].

A gap analysis was conducted in previous studies using the Back Propagation Neural Network (BPNN), which is very accurate. This is particularly true when combining multiple features and selecting them using the MOP method. However, the drawback lies in the lack of discussion about data sources, training duration, model testing, and the focus on GLCM feature extraction without trying other methods. Other studies on meat classification using CNN require significant computing resources and focus on high-quality images. In addition, this study only relies on the Adam Algorithm for its hyperparameter process.

Previous studies also classified pork and beef using the PNNR algorithm. Feature extraction with the HSV method has proven effective. However, to improve the accuracy and performance of the model, it is necessary to add other classification algorithms and use hyperparameter tuning to enhance the results obtained. Different studies use Spatial Fuzzy C-Means Segmentation (SFCM) for beef and pork classification, with the LVQ3 algorithm and GLCM feature extraction. Although the researchers used the Confusion Matrix for model evaluation, they did not discuss other evaluation matrices, such as precision, Recall, or F1-score. In addition, hyperparameter tuning was not performed in the study.

This study attempts to fill this gap by comparing more efficient machine learning algorithms such as Weighted K-Nearest Neighbors (WKNN), Random Forest (RF), and Support Vector Machines (SVM). The feature extraction used is HSV. The confusion matrix used is Accuracy, Precision, F1-Score, and Recall. Hyperparameter tuning was carried out in this study, specifically in WKNN, where the number of nearest neighbors (k value) was set to 1, 2, and 4, and the training and testing data compositions were 80-20 each. The Distance Function used is Euclidean to improve model accuracy. In a random forest, some hyperparameters include the number of trees (n_estimators). Increasing the number of trees can increase model stability. Furthermore, for SVM, the tuning process is focused on the C parameter (which controls regularization); the type of parameter used is linear and RBF. Parameters are critical in maintaining the SVM model's trade-off between bias and variance. This study provides a more practical and accurate solution for identifying beef and pork.

By utilizing the performance of the Weighted K-nearest neighbors, Random Forest, and Support Vector Machine models, this research aims to make scientific contributions to achieve optimal results and the highest accuracy in classifying beef and pork using machine learning.

This study introduces a novel approach to enhance the accuracy of beef and pork image detection by combining machine learning models, including Weighted K-Nearest Neighbors, Random Forest, and Support Vector Machine. The model's performance is analyzed using a confusion matrix, which evaluates accuracy, precision, F1-Score, and Recall values. The combination of models has not been widely carried out, so it has the opportunity to produce maximum accuracy values.

The research paper is structured as follows: introduction regarding the general description, background leading to the research topic, previous research, background of the problem, problem solution, novelty, contribution, and structure of the paper. Material and method regarding the proposed research design, machine learning models such as WKNN, RF, SVM, Feature extraction, and confusion matrix. The results and discussion regarding HSV and RGB color extraction features, testing parameters, machine learning model testing, and accuracy score results on the confusion matrix.

II. Materials and Methods

The method in this study consists of several processes. Each stage of the study aims to ensure that the survey runs according to the planned stages, including data collection, preprocessing, model development, and evaluation. The research flowchart is presented in Fig. 1.

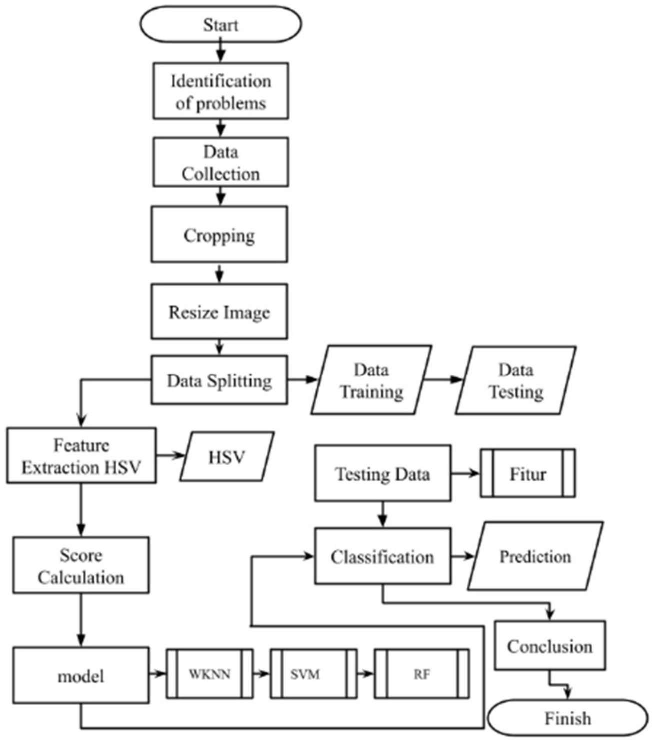

Fig. 1 Proposed research design

A. *Data Collection*

The primary data used in the study consisted of 400 images, 200 beef images, and 200 pork images. Meat was purchased in quantities of up to 1 Kg at the Pasar Bawah Pekanbaru tourist market. This market is located at Jalan Saleh Abas, Kampung Dalam Village, Senapelan District, Pekanbaru City, Riau. Samples were taken separately from the source to ensure data quality and consistency. The dataset on beef and pork can be seen in Table I as follows.

TABLE I
Result of All Datasets

No	Image	Dataset Amount of data	Total
1	Beef	200	400
2	Pork	200	

B. *Beef and Pork Samples*

The meat samples obtained were then documented in the form of image data. Image acquisition was performed using a CANON EOS Kiss X50 DSLR camera, set at ISO between 100 and 200, to produce high image quality. All beef and pork image data were taken during the day without additional lighting. Natural light was used for image data collection. This study ensures that the lighting during image collection remains consistent to avoid differences affecting the image data results. The lighting is arranged so that all image data has similar conditions to ensure the consistency of beef and pork image data.

The image format was JPG, with a shooting distance of around ± 15 cm between the camera and the object. The selection of the camera was based on its ability to capture accurate color details with optimal lighting. Adequate lighting ensures that extracting color features in the HSV color space runs smoothly. The picture of beef and pork can be seen in Figs. 2 and 3.

Fig. 2 Sample image for beef

Fig. 3 Sample image for pork

Beef and pork images were obtained by photographing using a Canon EOS KISS X50 DSLR camera with ISO settings between 100 and 200. Images were taken at a distance of 5-15 cm with a white background. The resulting data were saved in *.JPG format. The total data used in this study included 200 beef images and 200 pork images.

C. Preprocessing Data

Preprocessing data plays a crucial role in converting raw data into a more structured and interpretable format, thus enabling classification algorithms to analyze data features effectively [34]. The preprocessing method used in this study includes image cropping and image resizing. Cropping is done manually using Photoshop CS6 to ensure that only part of the image containing the research object is needed. After that, the cropped image is resized to fit a particular pixel dimension, and this step aims to speed up the data processing process.

D. Image Cropping

In this study, each image was converted into a pixel size of 400 x 400. Image cropping on one of the images can be seen in Fig. 4 as follows.

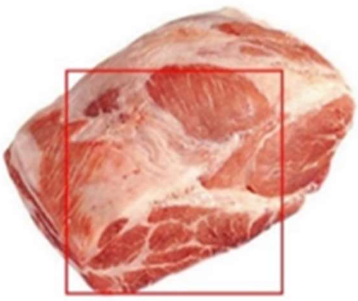

Fig. 4 Image cropping

E. Image Resize

The results of changing the image size to increase computational efficiency to 400 x 400 pixels can be seen in Fig. 5 as follows.

Fig. 5 Image resize result

F. Feature Extraction

Feature extraction process from beef and pork images to obtain relevant information about the characteristics of each meat. Feature extraction methods applied to the image class can overcome the problem of erosion and dilation with low-intensity levels [35]. The feature extraction process used in this study is RGB (Red, Green, Blue) and HSV (Hue, Saturation, Value).

1) RGB (Red, Green, Blue):

RGB is a color space that relies on the Cartesian coordinate system, where colors are represented as points defined by vectors originating from the origin [36]. RGB color space is a color model that presents the color spectrum through three main components: red, green, and blue. Each color in this model can be produced by combining the three components linearly [37]. The RGB color cube covers 256x256x256 pixels, each representing a unique combination of red, green, and blue values. For example, pure red is defined by the values (255,0,0), while cyan is represented by the values (0,255,255). This cube represents all the colors produced by the RGB color system, providing a complete visualization of the color spectrum. Various color effects can be created by manipulating the combinations of RGB values, such as variations in saturation and gradation, which allow for further exploration in digital image design and processing [38].

2) HSV (Hue, Saturation, Value):

HSV is a cylindrical coordinate system representing points (colors) in the red, green, and blue (RGB) color model. [39]. This system allows the geometric rearrangement of RGB colors into forms that better suit human visual perception and produce more intuitive and relevant colors [40]. Compared to Cartesian coordinates, the HSV approach provides a better visual in the context of color perception. Hue represents a primary color such as green, red, or magenta, which is determined by the angle of the color wheel, ranging from 0 degrees to 360 degrees [41]. In addition, saturation indicates how pure a color is or how far it is from gray. For example, as the saturation value approaches 0%, the color will appear increasingly gray, and at 100%, the color reaches its maximum purity. Another component, value (brightness), indicates the light intensity level in a color expressed as a percentage. At 0%, the color appears entirely black, while at 100%, the color becomes white [42].

G. Machine Learning

1) Weighted K-Nearest Neighbor (WKNN):

WKNN is a modification of the K-Nearest Neighbors (KNN) algorithm. The K-Nearest Neighbor algorithm represents several k values, indicating the number of closest neighbors used to assess the similarity between a new point and its points [43]. The KNN algorithm is based on instances, and classification is done by considering several nearest neighbors [44]. WKNN has good classification performance because its process is similar to KNN. However, there is a significant difference, namely in the weighting stage. Each class is given a certain weight in the scoring process, which is then used to determine the classification results of the test data [45]. Weighted K-nearest neighbors can improve classification performance by providing appropriate weights, especially on datasets with high variability or data that is not evenly distributed [46]. Calculations use the WKNN method with the following Equation.

$$W = 1/d^2 \tag{1}$$

The following are the stages in calculating the WKNN algorithm [47]:

- Determine the K parameter
- The distance between the new and all other samples is calculated individually.
- The calculated distances are sorted from smallest to largest, and the minor k is selected among these distances.
- The weight of the selected k samples is determined by calculating using Equation (1)
- 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.

2) Random Forest (RF):

RF is an ensemble learning method that uses several decision trees to make predictions [48]. By combining many trees that tend to overfit, Random Forest can produce a more stable and accurate model. Random forest can handle data with high dimensions, sparse data, outliers, and noise [49]. The random forest algorithm is an ensemble-based regression method that combines several decision trees to predict the value of a variable [50]. This allows the trees built to be more diverse, thus making the model more accurate. Data samples not used in the bagging process are called out-of-bag (OOB). The random forest method excels in high accuracy, ability to handle noisy data, efficient performance during training, overfitting control, and ease of implementation [51]. Two important parameters in the random forest model are the number of variables selected in each partition (m) and the number of trees (B) [52]. The Random Forest algorithm can be visualized in Fig. 6 [53].

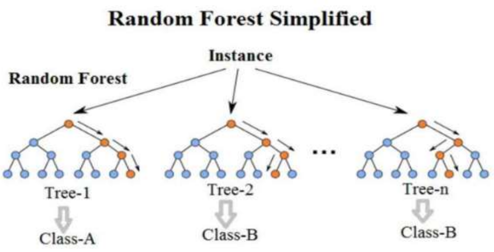

Fig. 6 Random Forest algorithm

3) Support Vector Machine (SVM)

SVM is one of the most widely used Supervised Learning methods for classification [54]. Calculations use the SVM method with the following Equation.

$$s(x) = sign[\textstyle\sum_{m=1}^{n} \delta_m s_m \, \varphi(x, x_m) + r] \tag{2}$$

In addition, SVM is also used for classification. The SVM algorithm aims to find the optimal hyperplane or dividing line that ensures the distance (margin) between two groups of data or classes [55]. The SVM algorithm works by representing observations as points in a space, whereas a large gap as possible separates points from different classes. The SVM algorithm efficiently classifies linearly inseparable patterns using the kernel trick, transforming the original input information into a high-dimensional feature space [56].

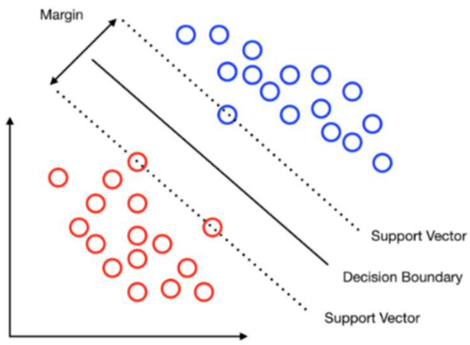

Fig. 7 Support vector machine algorithm

The SVM training process aims to find a stable hyperplane and maximize the distance, or margin, between the support vectors of the two classes shown in Figure 7. The Support Vector Machine algorithm has four main classification types: the maximum margin classifier, kernelized version, soft margin version, and soft margin kernelized version, a combination of the three [57]. Then, these types are grouped into linear SVM and non-linear SVM. In addition, the SVM classifier also has a Grid Search parameter, which is used to determine the best combination of parameters [58]. In SVM, the main element of the classifier is a separating hyperplane with a line equation in the interval [+1 -1] [59].

H. Confusion Matrix

A confusion matrix is a method for establishing a threshold probability level [60]. Classification performance can be measured and determined based on the confusion matrix. Some common ways are given as follows [61]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{5}$$

$$\text{F1- Score} = 2x\,\frac{(Precision\ x\ Recall)}{(Precision\ +Recall)} \tag{6}$$

III. Results and Discussion

A. Color Extraction Feature

The use of RGB and HSV algorithms in this study provides an overview of the color distribution of beef and pork images. The RGB model represents color as a combination of three main components: red, green, and blue. From the extraction results, the intensity of each color can be visualized to identify the dominant pattern in beef and pork samples. The results of the HSV and RGB feature extraction images can be seen in Fig. 8 as follows.

In Fig. 8, it can be explained that the color combination in the RGB model can process color data directly in a numeric format; this supports the color image classification process better. In addition, the HSV color model provides visual advantages because it approaches human perception of color. The color in the image is decomposed into three main components, namely Hue (color type), saturation (color saturation), and value (brightness). Hue identifies the primary color in the beef and pork images. Saturation describes the intensity of the color, while value indicates the brightness level. These three components work synergistically to provide a more intuitive color description for research. Furthermore, the image was resized to 400x400. This is done to help the computing process.

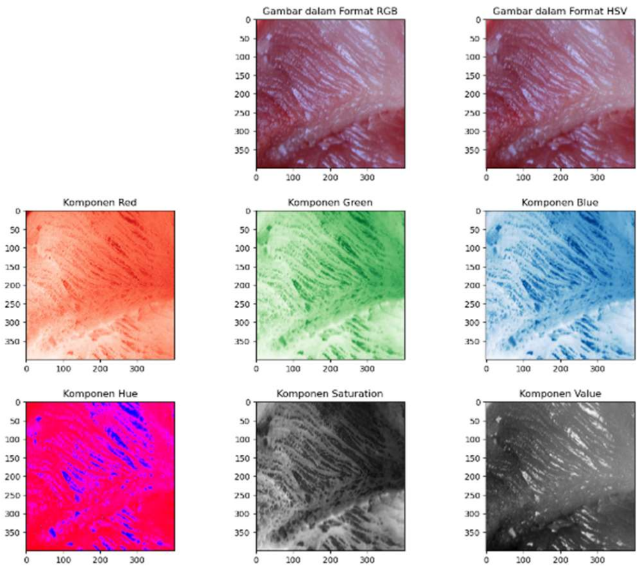

Fig. 8 HSV and RGB features extraction

B. *Testing Parameters*

The testing parameters include the training and testing data composition and the configuration of specific values applied in each algorithm. The WKNN, RF, and SVM algorithms test this parameter. The results of the parameter testing can be seen in Table II as follows.

TABLE II
TESTING PARAMETERS

No	Models	Experimental Testing	Configuration
1	WKNN	Composition of training data and testing data	80:20 (training and testing)
		K parameters	1, 2, and 4
2	RF	Composition of training data and testing data	80:20 (training and testing)
		Decision tree	100, 200, 400
3	SVM	Composition of training data and testing data	80:20 (training and testing)
		C parameters	1 and 4
		Kernel Functions	Linear, RBF Gaussian

In Table II, the researcher tested several k values in the WKNN algorithm, namely 1, 2, and 4. Smaller k values tend to cause overfitting, while larger k values can cause poor models. Therefore, the k value tested in the experiment was chosen to be moderate, as it was expected to provide a good balance between overfitting and underfitting. Models tend to be influenced by unrepresentative data, which reduces the generalization of larger and more varied data. Meanwhile, the value of k = 2 was chosen because it has a good balance between accuracy and generalization; this avoids the problem of overfitting and can capture patterns in the data. The value of k = 4 was tested to see if a larger k could reduce sensitivity to noisy data.

However, it has the risk of underfitting, which means the model is not sensitive enough to differences in data. In the experiment using the Random Forest (RF) algorithm, researchers tested three decision trees: 100, 200, and 400. The selection of the decision trees is based on the number of trees in the random forest algorithm. The better the model is at reducing variance and avoiding overfitting, the better. One hundred decision trees produce a good model but are sometimes less stable on large datasets. Testing with 200 trees yields consistent results, but there is no significant increase.

Meanwhile, 400 decision trees were tested in this study, which refers to a more significant number of trees that can improve model generalization to more complex data and reduce overfitting. In the SVM experiment, researchers tested two values of C, namely c = 1 and c = 4. The value c = 1 was chosen because research shows that a smaller parameter value of c provides a good balance between a more significant margin and the model's generalization ability on more extensive and complex datasets. This can reduce overfitting on the test data. Meanwhile, the value c = 4 is used to make the model more sensitive to misclassification so that the model can handle more complex data and increase accuracy with higher variance.

C. *Color Features*

The datasets used in this study comprise 400 images, divided equally between 200 beef images and 200 pork images. This study focuses on the differences in visual characteristics between the two types of meat. The color feature extraction method is used to determine the level of difference in meat quality.

1) *RGB Features:*

The results of RGB color feature extraction can be seen in Table III as follows.

TABLE III
RGB COLOR FEATURES

No	R	G	B	Label
1	119.047506	113.531787	162.834775	Pork
2	131.469681	95.355738	159.209487	Beef
3	116.728906	111.117050	147.536837	Pork
4	90.083975	90.083975	172.860456	Beef
5	108.102706	106.223156	158.358675	Pork
…	…	…	…	…
396	82.019006	85.044188	156.449069	Pork
397	123.373100	107.029506	150.627500	Beef
398	82.959137	84.896144	160.012200	Beef
399	162.096950	159.219750	204.895563	Pork
400	129.518781	93.944569	157.994050	Beef

2) HSV Features:

This is done because the color of the meat is an essential factor in distinguishing its type and quality. The color feature extraction used in this study is the RGB and HSV color methods. The HSV color feature used in this study is calculated through a multiplication process that incorporates a weighting technique. Each channel is multiplied by alpha, beta, and gamma so that the hue channel is multiplied by alpha, the beta channel is multiplied by saturation, and the value channel is multiplied by gamma. This study uses an alpha value of 0.3, a beta value of 0.5, and a gamma value of 0.2. The HSV color feature extraction results can be seen in Table IV as follows.

TABLE IV
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
…	…	…	…	…
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

D. Model Testing

After the feature extraction process, the model was tested using beef and pork images. In this test, the data was divided into 80% for model training and 20% for testing. The parameters used differed, but were in the same data composition, namely 80% for training data and 20% for testing data. Evaluation of each algorithm using a convolution matrix: Accuracy, Precision, F1-Score, and recall metrics.

1) Weighted K-Nearest Neighbor (WKNN):

The experiment was conducted with k values of 1, 2, and 4. The Weighted K-Nearest Neighbor Algorithm results can be seen in Table V as follows.

TABLE V
WEIGHTED K-NEAREST NEIGHBOR

K	Ratio	Accuracy	Precision	F1-Score	Recall
1	80:20	85%	96%	96%	96%
2	80:20	96.25%	96%	96%	96%
4	80:20	91.25%	96%	96%	96%

For k=1, the accuracy reaches 85%, with high Precision, F1-score, and recall values of 96%. However, this accuracy is lower than other k, indicating the possibility of overfitting. When the parameter is set to k=2, the accuracy increases to 96.25%, with the confusion matrix remaining at 96%. This shows that the model has good performance at k=2. For parameter k=4, the accuracy decreases slightly to 91.25%, although the precision, F1-Score, and recall values remain stable at 96%. This justifies that the model becomes less specific with a larger k value. The confusion matrix in the WKNN model test can be seen in Figs. 9, 10, and 11 as follows.

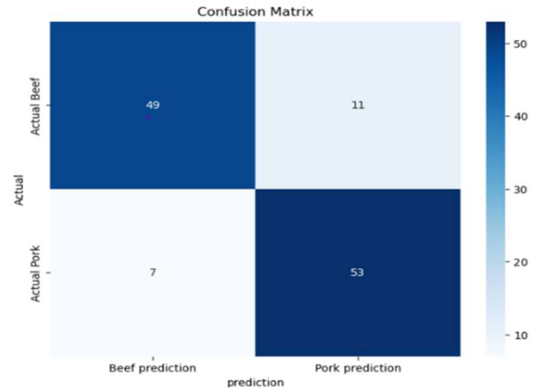

Fig. 9 WKNN algorithm with k=1

In Fig. 9, the WKNN algorithm with k=1 successfully classifies 49 beef and 53 pork data. However, 11 beef samples are incorrectly classified as pork, and seven are incorrectly classified as beef. This error indicates that a value of k=1 tends to cause overfitting, resulting in less stable model performance.

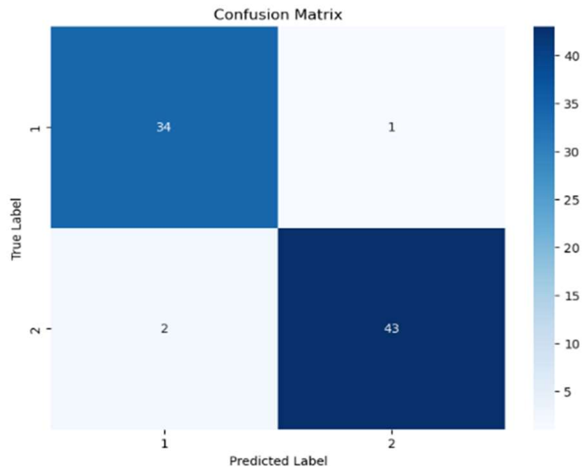

Fig. 10 WKNN algorithm with k=2

In Fig. 10, the experiment is continued by changing the value of k = 2. As a result, the model shows significant improvement. The model only misclassifies one beef sample as pork and two as beef. This indicates that k = 2 provides a good balance between generalization and specification on the dataset.

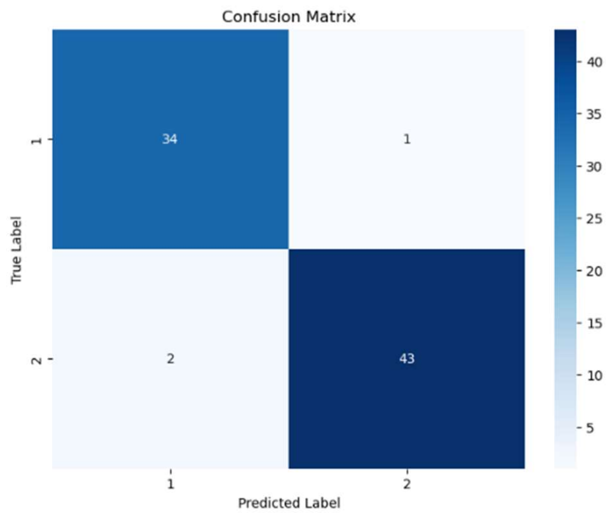

Fig. 11 WKNN algorithm with k=4

In Fig. 11, the value of k=4 produces the same results as k=2; namely, only one error occurred in beef data and 2 in pork data. This consistency shows that increasing the value of k within certain limits does not significantly affect the model's performance on the dataset.

2) Random Forest (RF):

Experiments on the random forest algorithm were carried out with different numbers of decision trees, namely 100, 200, and 400, with a ratio of 80:20. The results obtained on the Random Forest Algorithm can be seen in Table VI as follows.

TABLE VI
RANDOM FOREST

Estimators	Accuracy	Precision	F1-Score	Recall
100	98%	97%	98%	98.50%
200	97.50%	98%	97%	98%
400	98.75%	97%	98%	99%

At 100 decision trees, the accuracy reached 98% with a precision value of 97%, an F1-score of 98%, and a recall value of 98.5%. This shows an excellent performance. Then, when the number of trees in the experiment was increased to 200, the accuracy decreased slightly to 97.50%, although the precision and recall values remained high at 98%. This shows the stability of the model. The experiment was continued with 400 decision trees. The accuracy was increased to 98.5%, with a precision value of 97%, an F1-score of 98%, and a recall value of 99%. This justifies the number of trees that provide better stability and generalization.

The confusion matrix in the RF model testing can be seen in Figs. 12, 13, and 14 as follows.

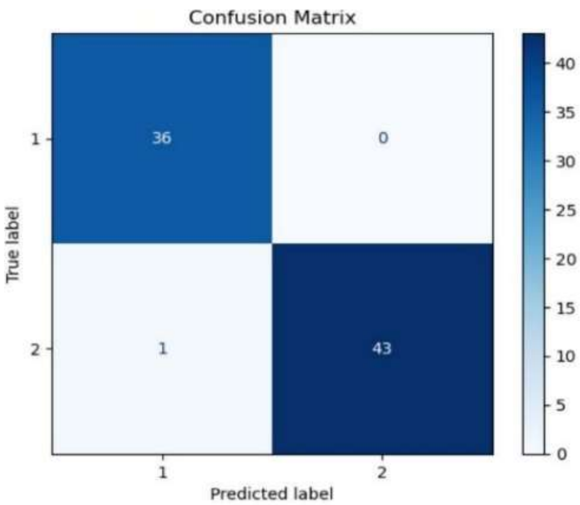

Fig. 12 RF algorithm with n estimators=100

In Fig. 12, the random forest algorithm with 100 decision trees (n_estimators=100), the model correctly classified all 36 beef data points and only made a mistake on one pork data point classified as beef.

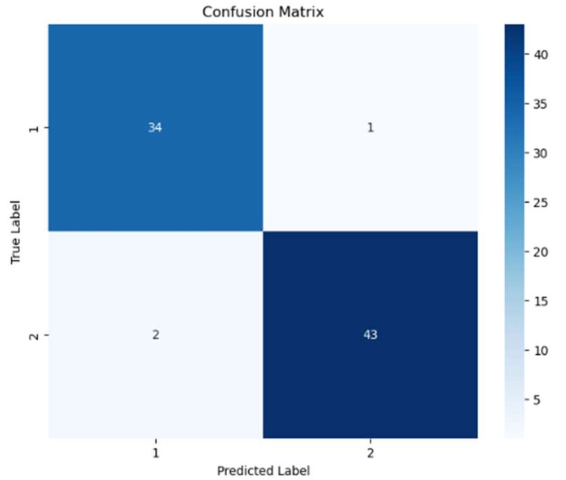

Fig. 13 RF algorithm with n_estimators=200

With a more significant number of decision trees, namely n_estimators=200, the model slightly decreases performance. 2 Pork data is misclassified as beef, although all beef data is still correctly classified.

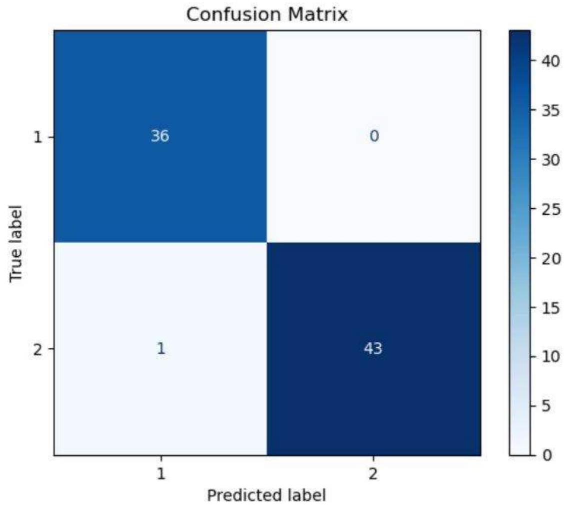

Fig. 14 RF algorithm with n_estimators=400

The experiment was continued with the number of decision trees (n_estimators=400), and the model achieved the best performance by only making a mistake on one pork data sample. In contrast, all 36 beef data samples were successfully classified correctly. This justifies that increasing the number of decision trees improves the model's generalization and produces a stable performance.

3) Support Vector Machine (SVM):

Furthermore, the experimental results were carried out using the Support Vector Machine (SVM) algorithm. In SVM with a regularization parameter = 1 and a linear kernel, the evaluation results showed an accuracy of 96.10%, with Precision, F1-Score, and Recall values consistently at 95%. This indicates that the model performs well by separating the data linearly. However, in the SVM algorithm with a regularization parameter = 4, RBF kernel, and a gamma value of 0.01, the accuracy decreased to 68.83%, with a precision value of 80%, F1-Score of 69%, and a Recall of 84%. These results indicate that the gamma parameter is too small, causing the model to fail to separate data well in high-dimensional space, so that model performance decreases. Overall, the random forest algorithm with 400 decision trees gave the best results compared to the Weighted K-Nearest Neighbor algorithm in terms of accuracy (98.75%) and Recall (99%). This justifies the random forest algorithm being better for the dataset. The results obtained in the Support Vector Machine Algorithm can be seen in Table VII as follows.

TABLE VII
SUPPORT VECTOR MACHINE

C	Ratio	Accuracy	Precision	F1-Score	Recall
1	80:20	96.10%	96%	96%	96%
4	80:20	68.83%	80%	69%	84%

The confusion matrix in the SVM model testing can be seen in Figs. 15 and 16 as follows.

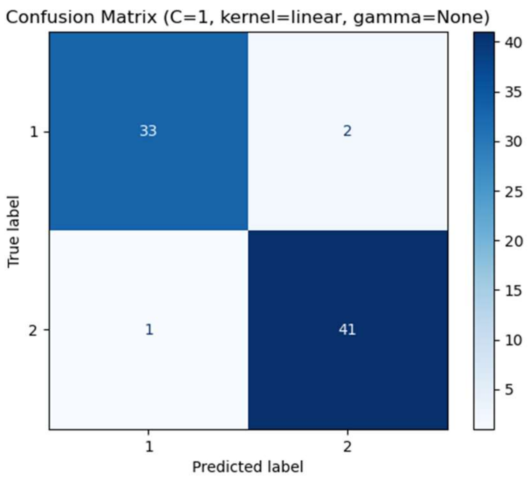

Fig. 15 SVM algorithm with C=1

In Fig. 15, the SVM algorithm with the regulation parameter (c=1) correctly predicts 74 meat data points, including 33 beef data points, but incorrectly predicts two meat data points as pork data. For pork data, 41 are correctly predicted, and only one pork data point is incorrectly predicted as beef data.

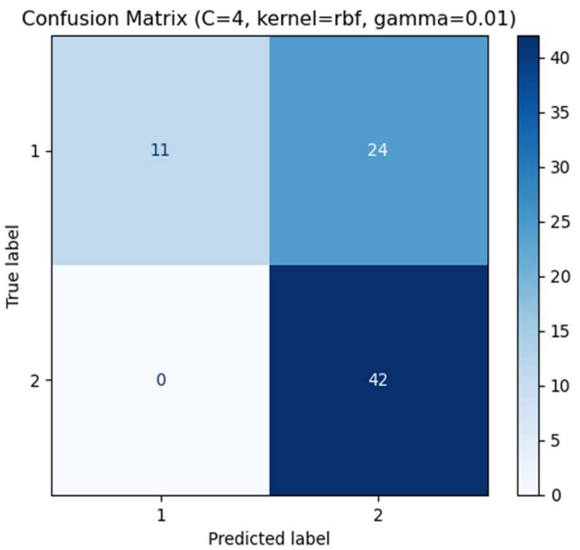

Fig. 16 SVM algorithm with C=4

The experiment was continued with the regulation parameters (c=4); the kernel is RBF, and the gamma is 0.01. The results showed good performance in classifying pork data. 43% of the pork data was predicted correctly. However, the model struggled to recognize beef data, correctly predicting only 11 out of 35 data points, while incorrectly predicting 24 as pork data. This indicates that the model exhibits a bias towards pork data, resulting in suboptimal performance for the dataset studied. Overall, the random forest algorithm with n_estimators=400 gave the best results compared to the WKNN algorithm and the support vector machine algorithm, with minimal errors in the classification of beef and pork. The accuracy comparison results of all tested models can be seen in Fig. 17, as follows.

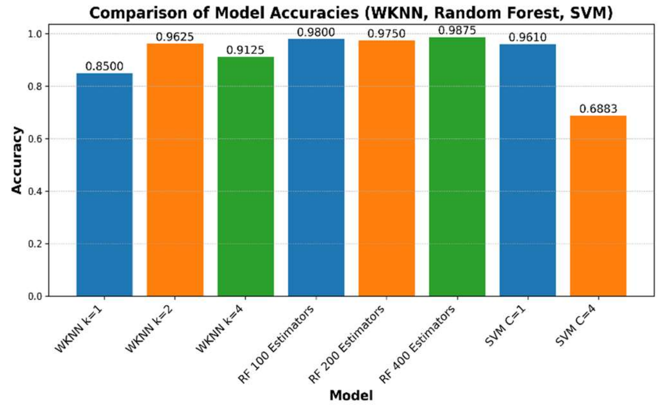

Fig. 177 Accuracy results of all models

From the results of the accuracy test in Fig. 17, it can be seen that the superior model overall is Random Forest with 400 estimators (accuracy level 98.75%). The WKNN and SVM models performed well with specific optimal parameters but were more susceptible to parameter variations than Random Forest. The least successful model was SVM with C=4 (accuracy only 66.83%).

E. Score Calculation

Based on the performance of the test results of the three algorithms, the comparison values obtained can be seen in Table VIII as follows.

TABLE VIII
ACCURACY COMPARISON RESULTS

No	Algorithm	Accuracy
1	Weighted K-Nearest Neighbor	96.25%
2	Random Forest	98.75%
3	Support Vector Machine	96.10%

Table VIII presents the model performance comparison results for three algorithms: Weighted K-Nearest Neighbor (accuracy = 96.25%), Random Forest (accuracy = 98.75%), and Support Vector Machine (accuracy = 96.10%). The random forest algorithm is the most effective for classifying beef and pork data with minimal error.

TABLE IX
RESULTS OF ANOVA AND T-TEST STATISTICAL TESTS

No	Comparison Model	Anova test statistics (F) and T test (t)	P-Value	Conclusion

1	ANOVA (WKNN, RF, SVM)	F = 1.6886	0.2752	Not significant (P > 0.05)
2	WKNN vs RF	t = -2.2141	0.0912	Not significant (P > 0.05)
3	SVM vs RF	t = -1.5352	0.2223	Not significant (P > 0.05)
4	WKNN vs SVM	t = 0.7610	0.5020	Not significant (P > 0.05)

In Table IX, the analysis conducted shows no significant difference in accuracy between the models tested. Through the ANOVA test comparing the WKNN, Random Forest (RF), and SVM algorithms, an F value of 1.6886 was obtained with a p-value of 0.2752, higher than 0.05. Hence, the difference between the models is insignificant. In addition, the t-test conducted to compare the accuracy of the WKNN and Random Forest algorithm models showed a t value of -2.2141 with a p-value of 0.0912, indicating that the difference in accuracy between the two models is insignificant ($p > 0.05$). The same thing also applies to the comparison between SVM and Random Forest, with a t value of -1.5352 and a p-value of 0.2223, which means this difference is insignificant. Finally, the test between the WKNN and SVM models produced a t-value of 0.7610 with a p-value of 0.5020, indicating no significant difference in the accuracy of the two models. All comparisons suggest no significant difference between the existing models at the significance level of $\alpha = 0.05$.

Test results show that the Rando forest algorithm provides the best performance results compared to other models in terms of accuracy and stability. One of the main reasons is that random forest combines many decision trees, which makes the model more stable and better at dealing with data variations. Each decision tree in the random forest algorithm is trained using a different subset of data so that errors that occur in one decision tree do not significantly affect the overall results of the model. Experiments with 400 decision trees justify that the more trees used, the more accurate and effective the model is at handling patterns in more complex data. With its ability to handle noise and variation in data, random forest proves to be a reliable model for distinguishing beef and pork images.

IV. Conclusion

Based on the results of the tests that have been carried out, a conclusion was obtained from the performance of the WKNN model, showing optimal performance at k = 2 with an accuracy of 96.25%. However, its performance decreased at k = 1 due to overfitting and at k = 4 due to reduced model specifications. In RF mode, the highest accuracy is 98.75%, Precision 97%, F1-Score 98%, and recall 99% at the number of decision trees of 400. This shows that the stability and generalization of the model are excellent. The SVM model performs well on the linear kernel with regularization C=1; the accuracy obtained is 96.1%. However, its performance decreased when using the RBF linear kernel with a gamma value of 0.01. The random forest algorithm is the best algorithm for classification problems with pork and beef data.

Although the tested classification model shows good results, the research object must consider several future challenges. The first is the lighting at the Bawah Pekanbaru market, which often changes; this can affect the image quality, which will later affect the model's accuracy. Second, meat freshness is different because of the uncertainty of market sales. This can affect the texture and color of the meat, making the classification model more difficult. Differences in image processing, such as cropping and resizing beef and pork images, can affect the quality of the data and information in the image. Therefore, in the future, model performance can be improved by converting images to a more stable color space, such as Lab or YCbCr, as a novelty for further research. In addition, more diverse data collection, including images of beef and pork freshness levels, is needed. Finally, in the future, using a deep learning algorithm for classification problems and helping to detect textures and colors related to the level of meat freshness will increase the variance of experimental results and accuracy.

In further research, it is expected to be able to separate images into three classifications, namely beef images, pork images, and mixed meat images. Implications for further research include using a deep learning approach, especially for image processing, to detect differences in each meat characteristic and increase accuracy.

References

[1] R. Liu, Z. Gao, H. A. Snell, and H. Ma, “Food safety concerns and consumer preferences for food safety attributes: Evidence from China,” *Food Control*, vol. 112, no. September 2019, p. 107157, 2020, doi: 10.1016/j.foodcont.2020.107157.

[2] H. Zhang *et al.*, “Effects of initial temperature on microbial community succession rate and volatile flavors during Baijiu fermentation process,” *Food Res. Int.*, vol. 141, no. May, p. 109887, 2021, doi: 10.1016/j.foodres.2020.109887.

[3] X. Gong, J. D. Morton, Z. F. Bhat, S. L. Mason, and A. E. D. A. Bekhit, “Comparative efficacy of actinidin from green and gold kiwi fruit extract on in vitro simulated protein digestion of beef Semitendinosus and its myofibrillar protein fraction,” *Int. J. Food Sci. Technol.*, vol. 55, no. 2, pp. 742–750, 2020, doi: 10.1111/ijfs.14345.

[4] V. D. Paup, S. M. Barnett, C. Diako, and C. F. Ross, “Detection of Spicy Compounds Using the Electronic Tongue,” *J. Food Sci.*, vol. 84, no. 9, pp. 2619–2627, 2019, doi: 10.1111/1750-3841.14709.

[5] P. N. Darma and M. Takei, “High-Speed and Accurate Meat Composition Imaging by Mechanically-Flexible Electrical Impedance Tomography with k-Nearest Neighbor and Fuzzy k-Means Machine Learning Approaches,” *IEEE Access*, vol. 9, pp. 38792–38801, 2021, doi: 10.1109/ACCESS.2021.3064315.

[6] L. Tang, M. Zhang, and L. Wen, “Support Vector Machine Classification of Seismic Events in the Tianshan Orogenic Belt,” *J. Geophys. Res. Solid Earth*, vol. 125, no. 1, pp. 1–19, 2020, doi: 10.1029/2019JB018132.

[7] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, “Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization,” *Geosci. Front.*, vol. 12, no. 1, pp. 469–477, 2021, doi: 10.1016/j.gsf.2020.03.007.

[8] S. K. Punia, M. Kumar, T. Stephan, G. G. Deverajan, and R. Patan, “Performance analysis of machine learning algorithms for big data classification: Ml and ai-based algorithms for big data analysis,” *Int. J. E-Health Med. Commun.*, vol. 12, no. 4, pp. 60–75, 2021, doi: 10.4018/IJEHMC.20210701.oa4.

[9] 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. xxxx, pp. 1–16, 2021, doi: 10.1016/j.animal.2021.100295.

[10] N. Johnson, K. Stone, S. N. Stastny, R. McGrath, and K. J. Hackney, “Beef Consumption and Functional Performance in Middle-Aged and Older Adults: A Narrative Review,” *J. Food Nutr. Sci.*, vol. 3, no. 1, pp. 18–31, 2021, doi: 10.1057/jfns000027.

[11] H. Dagevos, "Finding flexitarians: Current studies on meat eaters and meat reducers," *Trends Food Sci. Technol.*, vol. 114, no. December 2020, pp. 530–539, 2021, doi: 10.1016/j.tifs.2021.06.021.

[12] 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.

[13] M. Vlachou, A. Pexara, N. Solomakos, and A. Govaris, "Ochratoxin A in Slaughtered Pigs and Pork Products," *Toxins (Basel).*, vol. 14, no. 2, 2022, doi: 10.3390/toxins14020067.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] S. Magqupu, C. L. F. Katiyatiya, O. C. Chikwanha, P. E. Strydom, and C. Mapiye, "Quality and safety of pork sold in the informal urban street markets of the Cape Metropole, South Africa," *Meat Sci.*, vol. 204, no. April, p. 109270, 2023, doi: 10.1016/j.meatsci.2023.109270.

[20] 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.

[21] 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.

[22] K. Anwar and S. Setyowibowo, "The Identification of Beef and Pork Using Neural Network Based on Texture Features," *J. Eng. Res.*, pp. 1–19, 2022, doi: 10.36909/jer.17983.

[23] P. D. C. Sanchez, H. B. T. Arogancia, K. M. Boyles, A. J. B. Pontillo, and M. M. Ali, "Emerging nondestructive techniques for the quality and safety evaluation of pork and beef: Recent advances, challenges, and future perspectives," *Appl. Food Res.*, vol. 2, no. 2, 2022, doi: 10.1016/j.afres.2022.100147.

[24] Salsabila, A. Fitrianto, and B. Sartono, "Image Classification of Beef and Pork Using Convolutional Neural Network in Keras Framework," *Int. J. Sci. Eng. Inf. Technol.*, vol. 5, no. 02, pp. 245–248, 2021, doi: 10.21107/ijseit.v5i02.9864.

[25] A. A. Baiti, M. Fachrie, and S. Diwandari, "Classification of Beef and Pork Images Based on Color Features and Pseudo Nearest Neighbor Rule," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 8, no. 2, pp. 156–163, 2023, doi: 10.21831/elinvo.v8i2.64810.

[26] A. T. Akbar, S. Saifullah, H. Prapcoyo, B. Yuwono, and H. C. Rustamaji, "Robust Classification of Beef and Pork Images Using EfficientNet B0 Feature Extraction and Ensemble Learning with Visual Interpretation," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 11, no. 1, pp. 41–53, 2025, doi: 10.26594/register.v11i1.4045.

[27] J. Chauvin *et al.*, "Simulated annealing-based hyperspectral data optimization for fish species classification: Can the number of measured wavelengths be reduced?," *Appl. Sci.*, vol. 11, no. 22, pp. 1–20, 2021, doi: 10.3390/app112210628.

[28] X. Peng, R. Chen, K. Yu, F. Ye, and W. Xue, "An improved weighted k-nearest neighbor algorithm for indoor localization," *Electron.*, vol. 9, no. 12, pp. 1–14, 2020, doi: 10.3390/electronics9122117.

[29] S. Mustafa, A. Jaffar, M. Rashid, S. Akram, and S. M. Bhatti, "Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet-Random Forest for enhanced segmentation and classification," *PLoS One*, vol. 20, no. 1, p. e0315120, 2025, doi: 10.1371/journal.pone.0315120.

[30] R. P. Ram Kumar and S. Polepaka, "Performance comparison of random forest classifier and convolution neural network in predicting heart diseases," *Adv. Intell. Syst. Comput.*, vol. 1090, pp. 683–691, 2020, doi: 10.1007/978-981-15-1480-7_59.

[31] M. Premalatha and C. Vijayalakshmi, "SVM approach for non-parametric method in classification and regression learning process on feature selection with ε-insensitive region," *Malaya J. Mat.*, vol. S, no. 1, pp. 276–279, 2019, doi: 10.26637/MJM0S01/0051.

[32] J. Liang, Z. Qin, J. Ni, X. Lin, and X. Shen, "Practical and Secure SVM Classification for Cloud-Based Remote Clinical Decision Services," *IEEE Trans. Comput.*, vol. 70, no. 10, pp. 1612–1625, 2021, doi: 10.1109/TC.2020.3020545.

[33] R. Rodríguez-Pérez and J. Bajorath, "Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery," *J. Comput. Aided. Mol. Des.*, vol. 36, no. 5, pp. 355–362, 2022, doi: 10.1007/s10822-022-00442-9.

[34] S. H. Apandi, J. Sallim, R. Mohamed, and N. Ahmad, "Data Pre-processing of Website Browsing Records: To Prepare Quality Dataset for Web Page Classification," *Int. J. Informatics Vis.*, vol. 8, no. 1, pp. 239–246, 2024, doi: 10.62527/joiv.8.1.1618.

[35] T. O. Qadir, N. S. A. M. Taujuddin, and N. Fuad, "A New Feature Extraction Approach in Classification for Improving the Accuracy in Iris Recognition," *Int. J. Informatics Vis.*, vol. 7, no. 4, pp. 2161–2166, 2023, doi: 10.62527/joiv.7.4.1373.

[36] 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.

[37] D. T. Joy, G. Kaur, A. Chugh, and S. B. Bajaj, "Computer Vision for Color Detection," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 9, no. 3, pp. 53–59, 2021, doi: 10.21276/ijircst.2021.9.3.9.

[38] V. Skala, "Multispectral Image Generation from RGB Based on WSL Color Representation: Wavelength, Saturation, and Lightness," *Computers*, vol. 12, no. 9, 2023, doi: 10.3390/computers12090182.

[39] A. K. Al-Musawi, F. Anayi, and M. Packianather, "Three-phase induction motor fault detection based on thermal image segmentation," *Infrared Phys. Technol.*, vol. 104, 2020, doi: 10.1016/j.infrared.2019.103140.

[40] H. C. Kang, H. N. Han, H. C. Bae, M. G. Kim, J. Y. Son, and Y. K. Kim, "Hsv color-space-based automated object localization for robot grasping without prior knowledge," *Appl. Sci.*, vol. 11, no. 16, 2021, doi: 10.3390/app11167593.

[41] Y. Zhou, Z. Wang, and J. Wang, "Real-Time Adaptive Threshold Adjustment for Lane Detection Application under Different Lighting Conditions using Model-Free Control," *IFAC-PapersOnLine*, vol. 54, no. 20, pp. 147–152, 2021, doi: 10.1016/j.ifacol.2021.11.167.

[42] M. K. Othman and A. A. Abdulla, "Enhanced Single Image Dehazing Technique based on HSV Color Space," *UHD J. Sci. Technol.*, vol. 6, no. 2, pp. 135–146, 2022, doi: 10.21928/uhdjst.v6n2y2022.pp135-146.

[43] M. Kurniawan, G. E. Yuliastuti, A. Rachman, A. P. Budi, and H. N. Zaqiyah, "Implementing K-Nearest Neighbors (k-NN) Algorithm and Backward Elimination on Cardiotocography Datasets," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 3, pp. 1239–1245, 2024, doi: 10.62527/joiv.8.3.1996.

[44] S. Xu, C. C. Chen, Y. Wu, X. Wang, and F. Wei, "Adaptive residual weighted k-nearest neighbor fingerprint positioning algorithm based on visible light communication," *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–24, 2020, doi: 10.3390/s20164432.

[45] R. Zhou, Y. Yang, and P. Chen, "An rss transform—based wknn for indoor positioning," *Sensors*, vol. 21, no. 17, 2021, doi: 10.3390/s21175685.

[46] S. A. Bahanshal, R. S. Baraka, B. Kim, and V. Verdhan, "An Optimized Hybrid Fuzzy Weighted k-Nearest Neighbor with the Presence of Data Imbalance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 4, pp. 660–665, 2022, doi: 10.14569/IJACSA.2022.0130476.

[47] N. Afrianto and I. Meiditra, "Implementation of The Weighted K-Nearest Neighbors Algorithm in The Classification of Beef and Pork Images," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 7, no. 1, pp. 13–21, 2024, doi: 10.20895/inista.v7i1.

[48] X. Wang *et al.*, "Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–14, 2021, doi: 10.1186/s12911-021-01471-4.

[49] H. Chen, L. Wu, J. Chen, W. Lu, and J. Ding, "A comparative study of automated legal text classification using random forests and deep learning," *Inf. Process. Manag.*, vol. 59, no. 2, 2022, doi: 10.1016/j.ipm.2021.102798.

[50] G. Dudek, "A Comprehensive Study of Random Forest for Short-Term Load Forecasting," *Energies*, vol. 15, no. 20, 2022, doi:

10.3390/en15207547.

[51] H. Hairani, A. Anggrawan, and D. Priyanto, "Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 258–264, 2023, doi: 10.30630/joiv.7.1.1069.

[52] H. A. Zeini, D. Al-Jeznawi, H. Imran, L. F. A. Bernardo, Z. Al-Khafaji, and K. A. Ostrowski, "Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil," *Sustain.*, vol. 15, no. 2, 2023, doi: 10.3390/su15021408.

[53] C. Avci, M. Budak, N. Yagmur, and F. B. Balcik, "Comparison between random forest and support vector machine algorithms for LULC classification," *Int. J. Eng. Geosci.*, vol. 8, no. 1, pp. 1–10, 2023, doi: 10.26833/ijeg.987605.

[54] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting," *Ann. Data Sci.*, vol. 10, no. 1, pp. 183–208, 2023, doi: 10.1007/s40745-021-00344-x.

[55] D. A. Pisner and D. M. Schnyer, *Support vector machine*. Elsevier Inc., 2019. doi: 10.1016/B978-0-12-815739-8.00006-7.

[56] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Support Vector Machines and Support Vector Regression*. 2022. doi: 10.1007/978-3-030-89010-0_9.

[57] A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," *Reliab. Eng. Syst. Saf.*, vol. 233, no. April, 2023, doi: 10.1016/j.ress.2023.109126.

[58] A. E. Minarno, M. Fadhlan, Y. Munarko, and D. R. Chandranegara, "Classification of Dermoscopic Images Using CNN-SVM," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 2, pp. 606–612, 2024, doi: 10.62527/joiv.8.2.2153.

[59] S. N. Khan, S. U. Khan, H. Aznaoui, C. B. Şahin, and Ö. B. Dinler, "Generalization of linear and non-linear support vector machine in multiple fields: a review," *Comput. Sci. Inf. Technol.*, vol. 4, no. 3, pp. 226–239, 2023, doi: 10.11591/csit.v4i3.pp226-239.

[60] G. Phillips *et al.*, "Setting nutrient boundaries to protect aquatic communities: The importance of comparing observed and predicted classifications using measures derived from a confusion matrix," *Sci. Total Environ.*, vol. 912, 2024, doi: 10.1016/j.scitotenv.2023.168872.

[61] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf. Sci. (Ny).*, vol. 340–341, pp. 250–261, 2016, doi: 10.1016/j.ins.2016.01.033.