Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 5 (2023) 1026 - 1032 ISSN Media Electronic: 2580-0760

Combination of K-NN and PCA Algorithms on Image Classification of Fish Species

Rini Nuraini[1,2], Adi Wibowo[3], Budi Warsito[4], Wahyul Amien Syafei[5], Indra Jaya[6]
[1]Informatics Study Program, Faculty of Communication and Information Technology, Universitas Nasional, Jakarta, Indonesia
[2]Doctoral Study Program in Information Systems, Graduate School, Universitas Diponegoro, Semarang, Indonesia
[3]Informatics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia
[4]Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia
[5]Information Systems, Graduate School, Universitas Diponegoro, Semarang, Indonesia
[6]Faculty of Fisheries and Maritime Affairs, Institut Pertanian Bogor, Bogor, Indonesia
[1]rini.nuraini@civitas.unas.ac.id, [2]rininuraini@students.undip.ac.id, [3]bowo.adi@live.undip.ac.id, [4]budiwarsito@live.undip.com, [5]wasyafei@elektro.undip.ac.id, [6]indrajaya@apps.ipb.ac.id

Abstract

To do fish farming, you need to know the types of fish to be cultivated. This is because the type of fish will affect how it is handled and managed. So, this study aims to develop an image processing system for classifying fish species, especially fish that are cultivated, with a combination of the K-Nearest Neighbor (K-NN) algorithm and Principal Component Analysis (PCA). The feature extraction used is feature extraction based on its color and shape. The K-NN algorithm can group certain objects by considering the shortest distance from the object. According to the best criteria, the PCA method is employed in the meanwhile to decrease and keep the majority of the relevant data from the original characteristics. Based on the test results, the accuracy value obtained is 85%. The use of a combination of the K-NN and PCA algorithms in the image classification of fish species in the research that has been done has been shown to be able to increase accuracy by 7.5% compared to only using the K-NN algorithm.

Keywords: *fish species; image classification; k-nearest neighbor; principal component analysis*

1. Introduction

Fish are animals that live in water and have a myriad of benefits for human life. Among these benefits, it can be used as ornamental fish or consumed by humans because it has a high protein and nutritional content. Fish have a variety of species, both fish that live in fresh water and fish that live in sea water. Recorded more than 15,000 species of freshwater fish and more than 14,000 species of seawater fish [1]. With many benefits, fish is a superior commodity for several countries and has economic value. For this reason, many people cultivate fish, both for use as ornamental fish and for consumption. To do fish farming, you need to know the types of fish to be cultivated. This means that knowledge about the types of fish or the classification of fish species can be useful for farmers. This is because the type of fish will affect how it is handled and managed [2]. Types of fish can be identified based on their appearance because each type generally has different traits and characteristics.

Digital image processing is associated with how to perform the formation, processing, and analysis of images to obtain information from the image so that it can be more useful [3]. Image classification is one of the uses for image processing. The technique of classifying photographs into groups based on visual elements, each group having the same properties, is known as image classification [4]. To define diverse things and make them easier to recognize, image classification is helpful. [5]. Research related to the use of digital image processing in the classification and identification of fish species has been carried out by several researchers. The first researcher regarding formalin fish classification used the Support Vector Machine (SVM) approach with texture feature extraction. This study resulted in an accuracy rate of 0.784, or 78.4% [6]. The SVM technique, however, performs less well in complex class scenarios because it divides the optimum hyperplane into two classes [7]. Subsequent research regarding the identification of freshwater fish species with the Linear Discriminant

Accepted: 17-06-2023 | Received in revised: 09-08-2023 | Published: 30-08-2023

Analysis (LDA) approach using HSV color feature extraction [8]. In this study, the accuracy value reached 84.5%. But the LDA method has the drawback of being unable to handle data with samples smaller than the number of features [9]. Next, research on the classification of fish species using the K-Nearest Neighbor (K-NN) algorithm [10]. This study's accuracy score was 77.50%. By taking into account the object's closest distance, the K-NN algorithm may group certain items. The K-NN method can successfully classify with a high number of classes, despite this research's lesser accuracy than earlier studies.

The K-NN algorithm can be used for classification because of its simplicity in implementation, where the algorithm groups data based on the closest data point to the data you want to predict. The K-NN algorithm is a simple instance-based algorithm, which means there are no complex training stages. So this algorithm can be applied to limited training data and can give good results because it focuses on certain information. [11]. In contrast to deep learning approaches that offer high accuracy but involve complex model architectures and require large training data as well as powerful hardware to train and implement them. However, the K-NN algorithm has weaknesses in handling deviated values, so it is vulnerable to variables that do not contain information [12]. So, it is necessary to improve and add methods based on the features formed by objects and to reduce and maintain appropriate information from their original characteristics in order to produce optimal performance. One of the algorithms that can be used to solve data reduction problems is Principal Component Analysis (PCA). This method can be used to reduce and retain most of the relevant information from the original features according to optimal criteria [13]. PCA can simplify data through linear transformation so as to form a system that has new coordinates that produce maximum variance [14].

As a result, the goal of this project is to create a fish species classification image processing system, specifically for fish that have been raised using a combination of the K-Nearest Neighbor (K-NN) algorithm and Principal Component Analysis (PCA). Feature extraction is required to identify the K-NN method since it can only function properly if it receives data or characteristics from the picture class that has to be categorized. Based on qualities of color and form, feature extraction is used. The PCA method which is used to reduce and store most of the substantial information from existing features to obtain more representative information, will be used with the K-NN algorithm to optimize and increase the accuracy of the K-NN algorithm. In this study, the type of fish used was freshwater fish that were cultivated. This is because freshwater fish that are cultivated have different handling depending on the type of fish; for this reason, a classification process is needed by the cultivator. The types of freshwater fish cultivated that are used are Goldfish, Gurami, Bangus, and belanak. These are freshwater fish that are often cultivated both for consumption and as raw materials for the production of food products.

2. Research Methods

Research must be well-organized and structured via distinct phases in order to deliver high-quality results. Due to this, the phases of this study's research were organized to be consistent with the goals of the research shown in Figure 1.

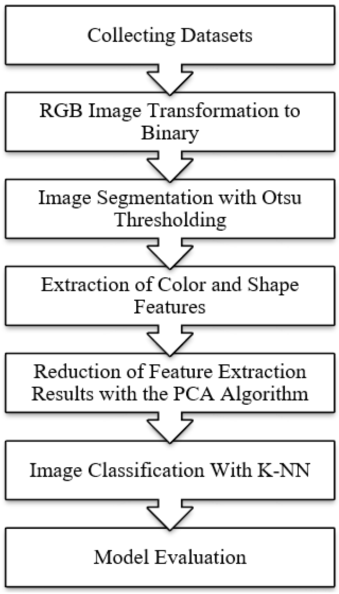


Figure 1. Phases in Conducting Research

2.1. Collecting Datasets

Image data is gathered at this point and utilized as a dataset. Datasets are a crucial component of image processing since the availability of a dataset affects a model's performance [15], [16]. The dataset used is taken from the Kaggle website with the following link https://www.kaggle.com/datasets/markdaniellampa/fish-dataset. The type of fish used is freshwater fish cultivation, including: Goldfish, Gourami, Bangus, Mullet. The total image data used is 400 images, then the image data is divided into training data and testing data. The percentage distribution used is 70:30 [17]. The percentage for training data is more than for testing data because so that the model can optimally learn from existing patterns [17]. The data used for testing is also used as data for validating the results of the training that has been carried out. So that the training data used is 280 images and the testing data is 120 images.

2.2. RGB Image Transformation to Binary

The RGB picture must then be changed into a binary image in the next phase. A binary image is one that only has the numbers 0 (black) and 1 (white) for intensity.

DOI: https://doi.org/10.29207/resti.v7i5.5178

Because the item to be segmented may be differentiated from the backdrop of the object by converting the RGB picture into a binary image, this approach tries to streamline the segmentation steps.

2.3. Image Segmentation with Otsu Thresholding

It is the aim of the segmentation procedure to distinguish certain items from other things in the image [18]. The limits of the region whose organization and form are comparable are used to divide the object. The Otsu thresholding segmentation technique is used. This technique uses an automated threshold search to identify the foreground, to be determined, and background of an image [19]. The result of this operation will be the separation of the necessary items from the backdrop, leaving just the necessary elements visible.

2.4. Extraction of Color and Shape Features

The process of identifying an item in an image and differentiating it from other objects is called feature extraction. At the classification step, the retrieved features are subsequently employed as input parameters [20]. Applying feature extraction based on color and form. The RGB and HSV values will be used by the color feature. Regarding the properties of the form dependent on the size of the object area.

2.5. Reduction of Feature Extraction Results with the PCA Algorithm

In order to achieve optimum performance, the results of feature extraction will now be reduced to two principal components, which will subsequently be reduced to one principal component. A technique called PCA uses several lines and planes to reduce dimensionality. Without significantly reducing the amount of information used to describe the full dataset, the PCA technique may condense the dimensions of the observed data into fewer dimensions. In accordance with ideal criteria, PCA will minimize and keep the majority of the relevant data from the original characteristics. [21]. Data may be simplified using linear transformations via the PCA technique to create a system with new coordinates that yield the most variance [22]. To do this, it is necessary to convert the data or pictures that need to be dimensionally reduced into a collection of column matrices, $\Gamma_1, \Gamma_2, \Gamma_3, \ldots, \Gamma_M$, where M is the number of available samples. The following procedures may be used to determine the main element of each data set:

This phase begins with the data being prepared by generating a set S using Formula 1.

$$(\Gamma_1, \Gamma_2, \Gamma_3, \ldots \Gamma_M) S = \Gamma_1, \Gamma_2, \Gamma_3, \ldots, \Gamma_M \tag{1}$$

The second step involves utilizing Formula 2 to get the middle value (Ψ).

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n \tag{2}$$

By using Formula 3, the third step determines the average value and the difference value (Φ) from the training data (Γ_i).

$$(\Psi) \Phi_i = \Gamma_i - \Psi \tag{3}$$

The covariance matrix (C) value is determined at the fourth step using Formula 4

$$C = \frac{1}{\mathrm{M}} \sum_{n=1}^{M} \Phi_n \Phi_n^T = AA^T \tag{4}$$

The next step involves solving Formula 5 to determine the eigenvalues (λ) and eigenvectors of the covariance matrix (C).

$$C \, x \, v_i = \lambda_i \, x \, v_i \tag{5}$$

The eigenface (μ) employs Formula 6 once the eigenvector (v) is produced.

$$\mu_i = \sum_{k=1}^{M} v_{lk} \Phi_k \quad l = 1, \ldots, M \tag{6}$$

2.6. Image Classification with K-NN

Using the K-Nearest Neighbor (KNN) technique, classification and regression issues may be resolved. This method falls under the category of supervised learning, where the training set must consist of labeled data. K-NN is a classification technique that bases its learning on data that is closest to the item being processed [23]. Non-parametric algorithms like K-NN employ a variable number of parameters, and when more data becomes available, the number of parameters often rises. Non-parametric algorithms take fewer data-related assumptions but are computationally slower. Figure 2 serves as an example of how the K-NN algorithm classifies items.

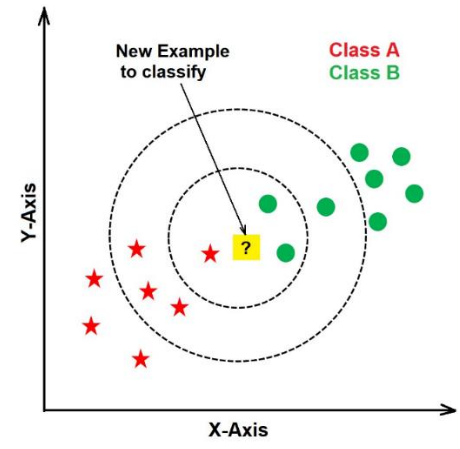


Figure 2. Illustration of Classification Using the K-NN

Figure 2 shows how K-NN classes objects using a learning pattern that takes into account the closest distance to other objects; if an item is near to an object that is being classified, it is regarded as a member of the closest object. In the K-NN algorithm, data is entered into a multidimensional space, and each dimension is associated with different data attributes. The accuracy value of the KNN algorithm is highly dependent on

DOI: https://doi.org/10.29207/resti.v7i5.5178

whether there are features that are not meaningful, or if the feature weights do not correspond to the object's significance [24]. The KNN algorithm makes the assumption that comparable things will be present nearby or in nearby neighborhoods. This implies that comparable data will be located next to one another. The KNN algorithm classifies data or new cases using all available data and similarity or distance functions. The class that houses the majority of the adjacent data is then given the new data. The test sample, for instance, X, is modified to resemble a feature vector shape (X_1, X_2, ..., X_m), in order to apply the preceding K-NN. The whole training sample is this sample. Then, using the sample values at (d_{i1}, d_{i2}, ..., d_{im}), compute the similarity between the training sample and the X test sample using Formula 7.

$$Similarity\ (X, d_i) = \frac{\sum_{j=1}^{m} X_j . d_{ij}}{\sqrt{\left(\sum_{j=1}^{m} X_j\right)^2} \sqrt{\left(\sum_{j=1}^{m} d_{ij}\right)^2}} \tag{7}$$

Next, choose the k samples with the highest degree of similarity, and use Formula 8 to compute the probability.

$$P(X, C_j) = \sum_d Similarity\ (X, d_i). y(d_i, C_j) \tag{8}$$

2.7. Model Evaluation

This evaluation stage is carried out to assess the effectiveness of the model created [25]. To provide an evaluation of the model built using the confusion matrix by looking for precision, recall and accuracy values. These values can be obtained using Formula 9,10, and 11.

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

where, TP (True Positive) indicates positive data that is correctly classified, TN (True Negative) indicates negative data indicates positive data that is correctly classified, FP (False Positive) indicates negative data that is classified with positive data results, FN (False Negative) shows positive data which is classified with negative data results.

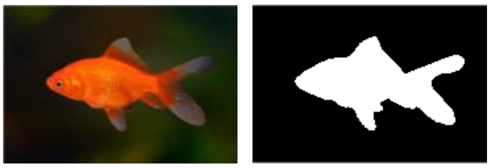
3. Results and Discussions

To classify images of fish species using a combination of the K-Nearest Neighbor (K-NN) algorithm and Principal Component Analysis (PCA) the first step is to prepare a dataset in advance, where the data will later be used as training and testing of the model being built. Datasets are needed in pattern recognition and learning in order to find information in the image so that the classification process can be carried out. The total image data used is 400 images, then the image data is divided into training data and testing data. The percentage distribution used is 70:30. So that the training data used is 280 images and the testing data is 120 images. The types of fish used are cultivated fish including: Goldfish, Gourami, Bangus, Mullet. Samples from the image dataset of fish species used can be seen in Table 1.

Table 1. Dataset Image Samples Used

Type of Fish Used	Image Sample
Goldfish	
Gourami	
Bangus	
Mullet	

The created model is then put into use in the MATLAB program for both training and testing. Beginning with the conversion of RGB pictures to binary images, the constructed model is trained. By transforming the RGB picture into a binary image, the item that has to be segmented may be separated from the background object, simplifying the segmentation step of the process. In this procedure, an RGB picture is converted into an image with a value of 0 or 1, or into an image that only contains two colors, black and white, depending on the item. Figure 3 displays picture examples of several fish species that emerged after the conversion into binary images.



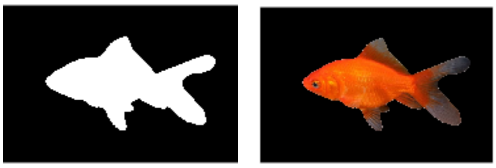
(a) (b)

Figure 3. (a) RGB Image and (b) Binary Image

Figure 3 (b) shows the results of the image that has been transformed into a binary image, which is then followed

DOI: https://doi.org/10.29207/resti.v7i5.5178

by the image segmentation step. The purpose of image segmentation is to be able to distinguish between certain things in a picture and other objects. The limits of the region whose organization and form are comparable are used to divide the object. The Otsu thresholding segmentation technique is used. As a consequence of this procedure, the necessary item has been distinguished from the backdrop and is now the only thing that is visible. Figure 4 shows image samples of several fish species that have been segmented using the Otsu thresholding technique.



(a) (b)

Figure 4. (a) Binary Image and (b) Image Segmentation

Following image segmentation, a feature extraction procedure is used to determine the qualities of the item in the picture that has to be identified. The feature extraction procedure uses color and shape-based feature extraction. The RGB and HSV values will be used by the color feature. Regarding the attributes of the shape dependent on the area or value of the item. It will thus compute the RGB, HSV, and area values of the item under test at this point. Figure 5 displays image examples of several fish species from which features have been retrieved.

	Features	Value
1	Red	181.0261
2	Green	76.9212
3	Blue	26.5591
4	Hue	0.1032
5	Saturation	0.7753
6	Value	0.7107
7	Area	50331

(a) (b)

Figure 5. (a) Image Segmentation and (b) Feature Extraction Value

The next step transforms the feature extracted using the PCA approach into a principal component, which is then split into two principal components to achieve optimum performance. According to the best criteria, the PCA algorithm will perform data reduction and retain most of the relevant data from its original characteristics. To achieve optimal performance, feature extraction results will now be reduced to two main components, which will then be reduced to one main component. The PCA algorithm uses multiple lines and planes to reduce dimensions. Without significantly reducing the amount of information used to describe the complete data set, PCA techniques can condense the observed data dimensions into fewer dimensions. Furthermore, the K-NN algorithm is used in the classification process. The K-NN algorithm uses a variable number of parameters, and as more data is available, the number of parameters often increases. The K-NN algorithm performs classification by basing its learning on data that is close to the item being processed. From the closest distance, K-NN performs grouping so that objects can be classified. Figure 6 displays a plot of the data distribution in each class.

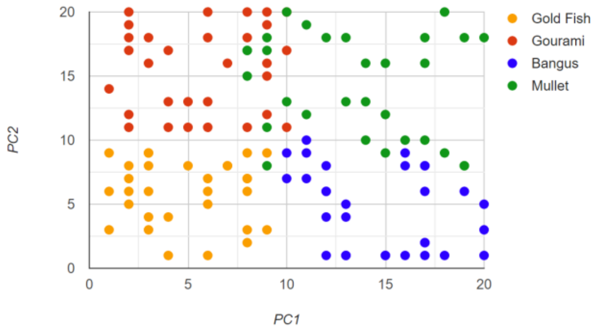


Figure 6. K-NN Training Data Distribution Plotting

The next step after creating the training model is to test it by turning it into an image categorization system. MATLAB was used to construct this system's GUI-based user interface, which makes it simpler to operate. Users may input photographs, execute image segmentation, feature extraction, and classification in the system that was established. Each process' results will be shown by the system, including those from segmentation, feature extraction, and classification as well as the outcomes of the RGB to binary transformation. Figure 7 displays the GUI application of the fish species categorization system created using MATLAB.

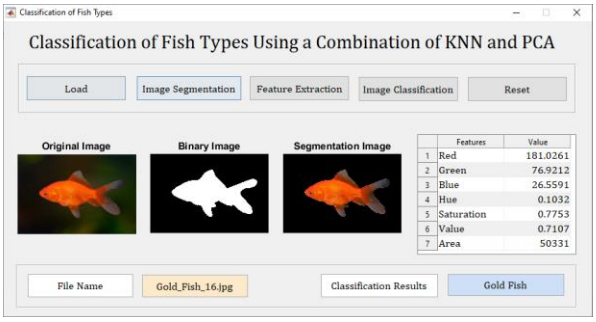


Figure 7. Fish Species Image Classification System Interface

The constructed model will be assessed to see how well it performs after being applied to the MATLAB program. In this process measurements are carried out so that the model can be known whether it can perform the task properly. There are 120 images of the fish species utilized as test data. The test is conducted by contrasting the model's categorization output with the data actually in existence. So, to get a value on the confusion matrix model will be tested by entering a test image on the system that was built then the values of TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) will be obtained.

DOI: https://doi.org/10.29207/resti.v7i5.5178

Figure 8 displays the outcomes of the confusion matrix from the tests that were conducted.

		Truth data				
		Gold Fish	Gourami	Bangus	Mullet	Classification overall
Classifier results	Gold Fish	32	1	0	1	34
	Gourami	1	24	3	3	31
	Bangus	0	2	23	2	27
	Mullet	0	2	3	23	28
	Truth overall	33	29	29	29	120

Figure 8. Confusion Matrix Results

Figure 8 displays the confusion matrix's findings, which were then used to determine values for precision, recall, and accuracy. The results are shown in Table 2.

Table 2. Test Result

Class Name	Precision	Recall	Accuracy
Goldfish	94.12	96.97	85.00
Gourami	77.42	82.76	
Bangus	85.19	79.31	
Mullet	82.14	79.31	

The accuracy results for the K-NN and PCA methods are then compared using just the K-NN algorithm and not the PCA approach that was used to reduce the data. Figure 9 shows a comparison graph of the accuracy of fish species picture categorization using the K-NN method alone vs a combination of the K-NN and PCA techniques.

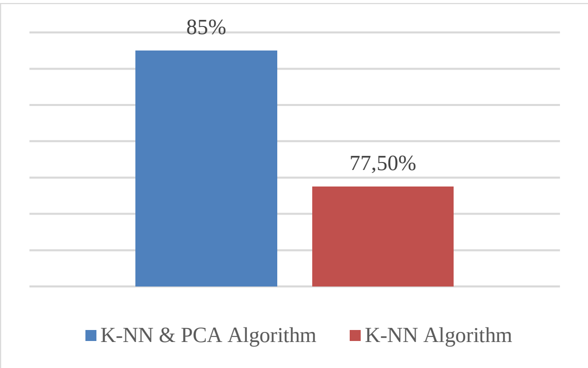


Figure 9. Comparison Graph Between the Combination of K-NN and PCA Algorithms with the K-NN Algorithm Only

In Table 2 it can be seen that the accuracy value of all test cases shows a value of 85.00 or 85%. Furthermore, these results are then converted into assessment categories with the following references: Good, if you get a score of 76% to 100%; Enough, with a value of 56% to 5%; Less Good, with a value of 40% to 55%, and Less Good, with a value of <40% [26]. The generated model falls within the "good" category based on these criteria. However, the test results reveal that the Bangus and Mullet class values result in a Recall value of 79.31%, which is below ideal. This is so because the class shares traits in terms of color and form with other classes. However, when compared to the K-NN algorithm alone, the K-NN plus PCA method combination delivers superior accuracy overall. This is shown by the graph in Figure 6, which shows a 7.5% accuracy difference between the K-NN method alone and the K-NN algorithm when PCA and NN are used together.

The experimental findings, however, indicate that the error rate may exceed 15%. This can be attributed to a number of reasons, including: 1) The K-NN method only uses the nearest neighbor when classifying; 2) Since the classes being utilized are almost identical, additional traits that are more representational may be employed instead of only color and form characteristics; 3) Pre-processing is necessary to create the best dataset since the model has trouble classifying images with different views and backgrounds; 4) The number of datasets used is considered not large, so experiments are needed using more datasets for better learning patterns.

4. Conclusion

A model for categorizing photos of fish species using a combination of the K-NN and PCA algorithms has been built as a result of the study. The feature extraction process is based on the object's color and form. By taking into account the object's closest distance, the K-NN algorithm may group certain items. The PCA method, which is used to minimize and retain the majority of the relevant information from the original features in accordance with optimum criteria, will be integrated with the K-NN algorithm in order to optimize and increase the accuracy of the K-NN algorithm. According to the test findings, the accuracy value is 85%, which is excellent. According to the study, employing both the K-NN and PCA algorithms together to classify images of fish species may enhance accuracy by 7.5% compared to using only the K-NN method alone.

There are various recommendations for improvement for future study, including the need to take into account more representative characteristics in order to get a better learning plot. In addition, deep learning may be used to address the complexity of visual characters. A vast number of datasets must be used in order to get the best learning results.

References

[1] S. Manel *et al.*, "Global determinants of freshwater and marine fish genetic diversity," *Nat. Commun.*, vol. 11, no. 1, p. 692, 2020, doi: 10.1038/s41467-020-14409-7.

[2] A. Ssekyanzi, N. Nevejan, R. Kabbiri, J. Wesana, and G. Van Stappen, "Knowledge, Attitudes, and Practices of Fish Farmers Regarding Water Quality and Its Management in the Rwenzori Region of Uganda," *Water (Switzerland)*, vol. 15, no. 1, pp. 1–22, 2023, doi: 10.3390/w15010042.

[3] L. Zhang, L. Zhang, and L. Zhang, "Application research of digital media image processing technology based on wavelet transform," *Eurasip J. Image Video Process.*, vol. 2018, no. 138, pp. 1–10, 2018, doi: 10.1186/s13640-018-0383-6.

[4] H. Mayatopani, R. I. Borman, W. T. Atmojo, and A.

DOI: https://doi.org/10.29207/resti.v7i5.5178

Arisantoso, “Classification of Vehicle Types Using Backpropagation Neural Networks with Metric and Ecentricity Parameters,” *J. Ris. Inform.*, vol. 4, no. 1, pp. 65–70, 2021, doi: 10.34288/jri.v4i1.293.

[5] R. I. Borman, F. Rossi, D. Alamsyah, R. Nuraini, and Y. Jusman, “Classification of Medicinal Wild Plants Using Radial Basis Function Neural Network with Least Mean Square,” in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022.

[6] E. P. Wanti, A. Pariyandani, S. Zulkarnain, and S. Idrus, “Utilization of SVM Method and GLCM Feature Extraction in Classifying Fish Images with Formalin,” *Sci. J. Informatics*, vol. 8, no. 1, pp. 168–175, 2021, doi: 10.15294/sji.v8i1.26806.

[7] R. I. Borman, F. Rossi, Y. Jusman, A. A. A. Rahni, S. D. Putra, and A. Herdiansah, “Identification of Herbal Leaf Types Based on Their Image Using First Order Feature Extraction and Multiclass SVM Algorithm,” in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2021, pp. 12–17. doi: 10.1109/ICE3IS54102.2021.9649677.

[8] R. Nuraini, “Identification of Freshwater Fish Types Using Linear Discriminant Analysis (LDA) Algorithm,” *IJICS (International J. Informatics Comput. Sci.*, vol. 6, no. 3, pp. 147–154, 2022, doi: 10.30865/ijics.v6i3.5565.

[9] M. J. Alrawashdeh, T. Radwan, and K. Abunawas, “Performance of linear discriminant analysis using different robust methods,” *Eur. J. Pure Appl. Math.*, vol. 11, no. 1, pp. 284–298, 2018.

[10] S. Kaharuddin and E. W. Sholeha, “Classification of Fish Species with Image Data Using K-Nearest Neighbor,” *Int. J. Comput. Inf. Syst.*, vol. 02, no. 02, pp. 54–58, 2021.

[11] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, “Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction,” *Sci. Rep.*, vol. 12, no. 6256, pp. 1–11, 2022, doi: 10.1038/s41598-022-10358-x.

[12] M. A. Imron and B. Prasetiyo, “Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn,” *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 56–62, 2020, doi: 10.52465/joscex.v1i1.7.

[13] R. I. Borman and B. Priyopradono, “Implementasi Penerjemah Bahasa Isyarat Pada Bahasa Isyarat Indonesia (BISINDO) Dengan Metode Principal Component Analysis (PCA),” *J. Inform. J. Pengemb. IT*, vol. 03, no. 1, pp. 103–108, 2018.

[14] R. I. Borman, R. Napianto, N. Nugroho, D. Pasha, Y. Rahmanto, and Y. E. P. Yudoutomo, “Implementation of PCA and KNN Algorithms in the Classification of Indonesian Medicinal Plants,” in *ICOMITEE 2021*, 2021, pp. 46–50.

[15] A. Mulyanto, W. Jatmiko, P. Mursanto, P. Prasetyawan, and R. I. Borman, “A New Indonesian Traffic Obstacle Dataset and Performance Evaluation of YOLOv4 for ADAS,” *J. ICT Res. Appl.*, vol. 14, no. 3, pp. 286–298, 2021.

[16] Z. Abidin, Permata, R. I. Borman, U. Ardiyatno, F. Rossi, and Y. Jusman, “Computer-aided Translation Based on Lampung Language as Low Resource Language,” in *International Conference on Computer Science, Information Technology, and Electrical Engineering (COMITEE)*, 2021, pp. 7–11. doi: 10.1109/ICOMITEE53461.2021.9650324.

[17] A. Gholamy, V. Kreinovich, and O. Kosheleva, “Why 70/30 or 80/20 Relation Between Training and Testing Sets : A Pedagogical Explanation,” *Sch. UTEP*, vol. 2, pp. 1–6, 2018.

[18] I. Ahmad, Y. Rahmanto, R. I. Borman, F. Rossi, Y. Jusman, and A. D. Alexander, “Identification of Pineapple Disease Based on Image Using Neural Network Self-Organizing Map (SOM) Model,” in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022.

[19] S. Bhahri and R. Rachmat, “Transformasi Citra Biner Menggunakan Metode Thresholding Dan Otsu Thresholding,” *J. Sist. Inf. dan Teknol. Inf.*, vol. 7, no. 2, pp. 195–203, 2018.

[20] W. K. Mutlag, S. K. Ali, Z. M. Aydam, and B. H. Taher, “Feature Extraction Methods: A Review,” *J. Phys. Conf. Ser.*, vol. 1591, no. 1, pp. 1–10, 2020, doi: 10.1088/1742-6596/1591/1/012028.

[21] P. Haque, B. Das, and N. N. Kaspy, “Two-Handed Bangla Sign Language Recognition Using Principal Component Analysis (PCA) And KNN Algorithm,” in *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019.

[22] N. Nikita, H. Sadawarti, and B. Kaur, “Classification of Renal Cancer using Principal Component Analysis (PCA) and K-Nearest Neighbour (KNN),” *Int. J. Eng. Res. Technol.*, vol. 8, no. 16, pp. 156–159, 2020.

[23] S. Sanjaya, M. L. Pura, S. K. Gusti, F. Yanto, and F. Syafria, “K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors,” *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 101, 2019, doi: 10.24014/ijaidm.v2i2.7975.

[24] N. K. A. Wirdiani, P. Hridayami, N. P. A. Widiari, K. D. Rismawan, P. B. Candradinata, and I. P. D. Jayantha, “Face Identification Based on K-Nearest Neighbor,” *Sci. J. Informatics*, vol. 6, no. 2, pp. 150–159, 2019, doi: 10.15294/sji.v6i2.19503.

[25] Z. Abidin, R. I. Borman, F. B. Ananda, P. Prasetyawan, F. Rossi, and Y. Jusman, “Classification of Indonesian Traditional Snacks Based on Image Using Convolutional Neural Network (CNN) Algorithm,” in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022, pp. 18–23.

[26] R. I. Borman, Y. Fernando, and Y. Egi Pratama Yudoutomo, “Identification of Vehicle Types Using Learning Vector Quantization Algorithm with Morphological Features,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 2, pp. 339–345, 2022, doi: 10.29207/resti.v6i2.3954.

DOI: https://doi.org/10.29207/resti.v7i5.5178