



Evaluating the Impact of Data Normalization on Rice Classification Using Machine Learning Algorithms



Ahmet Çelik*[®]

Computer Technology Department, Tavsanli Vocational School, Kutahya Dumlupinar University, 43300 Kutahya, Turkey

* Correspondence: Ahmet Çelik (ahmet.celik@dpu.edu.tr)

Received: 07-18-2024

Revised: 09-05-2024 Accepted: 09-12-2024

Citation: A. Çelik, "Evaluating the impact of data normalization on rice classification using machine learning algorithms," *Acadlore Trans. Mach. Learn.*, vol. 3, no. 3, pp. 162–171, 2024. https://doi.org/10.56578/ataiml030 303.



© 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Rice is a staple food for a significant portion of the global population, particularly in countries where it constitutes the primary source of sustenance. Accurate classification of rice varieties is critical for enhancing both agricultural yield and economic outcomes. Traditional classification methods are often inefficient, leading to increased costs, higher misclassification rates, and time loss. To address these limitations, automated classification systems employing machine learning (ML) algorithms have gained attention. However, when raw data is inadequately organized or scattered, classification accuracy can decline. To improve data organization, normalization processes are often employed. Despite its widespread use, the specific contribution of normalization to classification performance requires further validation. In this study, a dataset comprising two rice varieties Osmancik and Cammeo produced in Turkey was utilized to evaluate the impact of normalization on classification outcomes. The k-Nearest Neighbor (KNN) algorithm was applied to both normalized and non-normalized datasets, and their respective performances were compared across various training and testing ratios. The normalized dataset achieved a classification accuracy of 0.950, compared to 0.921 for the non-normalized dataset. This approximately 3% improvement demonstrates the positive effect of data normalization on classification accuracy. These findings underscore the importance of incorporating normalization in ML models for rice classification to optimize performance and accuracy.

Keywords: Machine learning; Image processing; Normalization; Data processing; Rice classification; Food quality determination

1 Introduction

Due to the fact that about 67% of the world's human population is related to the agricultural sector, the production of different varieties of cereals is of great importance. Sowing different varieties of seeds together in agriculture can reduce yield and cause economic loss. Rice classification is expensive, laborious and error-prone to manual work using traditional methods. However, the use of computer vision, image processing and data evaluation methods in classification offers an up-to-date and advanced technology [1].

Rice classification becomes very important as there are many types of rice produced today. Manually classifying rice grains according to rice types is not efficient and safe because it requires a lot of time compared with automatic classification [2]. It is possible to automatically identify and classify individual rice grains using an intelligent system according to the relevant species. Computer vision techniques form the basis of such systems [3].

Rice is an important source of consumption for humans, necessitating not only the quality classification of rice products but also the identification of diseased and weed-infested rice. Some researchers have conducted studies to detect diseases on rice using ML algorithms. Jena et al. [4] classified the diseases encountered on rice species such as BrownSpot, Hispa and LeafBlast using many ML-based methods. The study was conducted using the Orange 3.26.0 interface. Ruslan et al. [5] classified weeded rice using ML and image processing methods. Weed rice is a type of weed in rice production fields. The weed rice infestation has become a general problem as it has been reported worldwide. Therefore, it is very important to classify rice at the earliest so that it is possible to take preventive measures [4].

On the basis of ML methods, many studies have also classified other agricultural products. Ozkan et al. [6] classified peanut species, including feature extraction, size reduction and size weighting stages using the advanced KNN algorithm. High achievements were achieved by using artificial neural networks (ANNs) in the classification of such products [7]. Butuner et al. [8] classified lentil species using different learning algorithms and Çelik [9] classified wheat seeds using the KNN algorithm. Ayele and Tamiru [10] classified chickpea species using many algorithms and made performance comparisons.

In this study, Osmancik and Cammeo rice classification was conducted using the KNN ML algorithm. Variable training and test data of a dataset containing 3810 records were used, each representing seven attributes of rice grains derived from an imaging system. In addition, non-normalized (Method 1) and normalized (Method 2) datasets were tested and the effect on the classification success was measured on the proposed model. In addition, the min-max normalization process was used. The most important innovative aspect of this study, which distinguishes it from other studies in the literature, is that it proves that the data normalization process on the dataset increases the classification success.

2 Methodology



Figure 1. Osmancik and Cammeo rice structures



Figure 2. Flow chart of the designed system

ML methods have a major role for the classification, identification, and analysis of different data for various applications [4]. The dataset used in this study was downloaded from open source storage. The dataset [3] was created and recorded by capturing different rice images according to the relevant species in the first stage. In the second stage, the captured images were processed using image processing methods and the morphological features of the rice samples were extracted. In the third stage, the attributes of the samples belonging to the rice classes were recorded in the dataset [2, 3].

In this study, two different types of rice grown in Turkey were classified. Osmancik rice type has had a large cultivation area since 1997 and the weight of a thousand grains is 23-25 grams. The Cammeo rice type, on the other hand, first grown in 2014, has a thousand-grain weight of 29-32 grams [3]. These rice species structures are shown in Figure 1.

In this study, the flow chart of the designed model is shown in Figure 2. In the study, in the first stage, normalization was performed on the attributes of rice classes on the dataset. Then, classification was carried out with the KNN algorithm according to 70% and 50% training rates. In the study, classification was performed on non-normalized data using the same algorithm. In the last stage, the effect of the normalization process on the classification process was measured.

2.1 Dataset

In the open source dataset used, there are a total of 3810 records belonging to the Osmancik and Cammeo rice classes. The data have seven attributes for each record, i.e., area, perimeter, major axis length, minor axis length, eccentricity, convex area and extent. The examples of raw dataset records [3] are shown in Table 1. The Osmancik and Cammeo rice varieties used in this study were produced in Turkey. Shape-based morphological features were used in feature selection. Therefore, it is thought that it can be used in many rice classifications. The attributes were created with the data obtained through the image processing steps.

Table 1. Attribute (non-normalized) values of randomly selected rice from the dataset

Area	Perimeter	Major Axis Length	Minor Axis Length	Eccentricity	Convex Area	Extent	Class
15231.00	525.57897949	229.7498779	85.09378815	0.928882003	15617.00	0.572895527	Cammeo
14656.00	494.31100464	206.0200653	91.73097229	0.895404994	15072.00	0.615436316	Cammeo
14634.00	501.12200928	214.1067810	87.76828766	0.912118077	14954.00	0.693258822	Cammeo
13447.00	455.64801025	183.9575806	94.45813751	0.858102858	13867.00	0.625907660	Osmancik
13233.00	459.85900879	192.5907135	88.34671783	0.888576806	13436.00	0.588735163	Osmancik
12538.00	452.66000366	188.8052826	86.10971832	0.889940381	12846.00	0.684164584	Osmancik

In addition, the descriptions of the rice attributes within the dataset [3] are shown in Table 2. Rice attributes and images of each rice grain were calculated after image processing methods were applied and recorded in the dataset.

Table 2.	Explana	ations c	of rice	attributes	[3]

Explanation	Attribute
The total number of pixels within the boundaries of a rice grain image	Area
Circumference of the image of a rice grain	Perimeter
The largest radius of the image of a rice grain	Major axis length
The smallest radius of the image of a rice grain	Minor axis length
The roundness ratio of the rice grain image relative to an ellipse having the same moments	Eccentricity
On the region formed by the image of a rice grain, the total number of pixels of the smallest convex shell	Convex area
The ratio of the region formed by a rice grain image to the bounding box pixels	Extent

2.2 Min-Max Normalization

The normalization process was used to organize, improve and simplify scattered data in the dataset. Thus, it is thought that it may affect classification and prediction successes [11]. The normalization process can also be used in deep learning methods [12]. In computer vision applications used for product classification, normalization operations are also performed on images [1].

In ML methods, normalization is used to reduce the impact of the attribute data range of each record. In this study, the min-max normalization process was used, with 0 selected as the minimum and 1 as the maximum. Thus, it is intended to arrange the values in the dataset between 0 and 1. In the study, the Z-score method was not chosen because there was no negative attribute value.

The calculation of the normalization process is shown as follows [13]:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where, x is the base data, y is the normalized data, x_{max} is the greatest value of the underlying data, and x_{min} is the smallest data value of the basic data. In the study, the attribute values concerning area, perimeter, major axis length, minor axis length and convex area were normalized. No normalization was performed for the attributes of eccentricity and extent, as their values already ranged between 0 and 1 in the raw data. Examples of records belonging to the normalized dataset are shown in Table 3.

Table 3. Attribute (normalized) values of randomly selected rice from the dataset

Area	Perimeter	Major Axis Length	Minor Axis Length	Eccentricity	Convex Area	Extent	Class
0.6759373	0.87923163	0.9012159	0.5324174	0.928882003	0.693917018	0.572895527	Cammeo
0.6253300	0.71409491	0.6480872	0.6706631	0.895404994	0.646009142	0.615436316	Cammeo
0.6233938	0.75006612	0.7343491	0.5881245	0.912118077	0.635636428	0.693258822	Cammeo
0.5189227	0.50990259	0.4127440	0.7274672	0.858102858	0.540084388	0.625907660	Osmancik
0.5000880	0.53214229	0.5048347	0.6001726	0.888576806	0.502197609	0.588735163	Osmancik
0.4389192	0.49412192	0.4644550	0.5535782	0.889940381	0.450334037	0.684164584	Osmancik

2.3 KNN ML Algorithm

The KNN is a widely used supervised ML algorithm. In this algorithm, analysis of records with well-defined classes and attributes is performed. The class of the new sample record is calculated by measuring the distances to the existing classes with distance metrics and is determined according to the majority of the class to which the nearest k sample belongs [13, 14]. It can be expected that tests can be performed using different k neighbor values and different success results can be obtained. In previous studies, the value of k neighbors has been mostly chosen as 3 by default. Therefore, in the developed model, a k-value of 3 was chosen. The primary purpose of this study is to prove its contribution to the classification performance on normalized datasets.

The KNN algorithm is known as a widely used and easily interpretable model. In addition, the algorithm is used in multiple classification applications. In the KNN algorithm, the probabilities of multiple classes are calculated with an approach called majority voting labeling.

The KNN algorithm has been successfully used in the classification of food products [9] and estimation processes [15]. Success rates may change depending on the change in the k neighbor value. Selecting the most appropriate k neighborhood value can increase the classification success [14]. Euclid, Chebyshev, Manhattan and Mahalanobis distance metric methods have been used with the KNN algorithm [9]. The most common Euclidean distance metric has been used and its calculation is shown as follows [15]:

$$d_{Euclid} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$
(2)

where, x_i is new sample value, y_i is a previously stored sample value in the database, n is the number of attributes, and d_{Euclid} is the distance metric value of x_i and y_i .

3 Experimental Results

In this study, classification successes were measured with the KNN algorithm using non-normalized (Method 1) and normalized (Method 2) datasets, taking into account the variable training and testing data. Accuracy, F1-score, precision and recall were chosen, which are widely used as success measurement metrics. Two configurations were tested: one with 70% training and 30% testing data, and another with a 50% split for both training and testing. The classification success rates for each configuration are presented in Tables 4 and 5.

Table 4. Success of the KNN algorithm on the non-normalized dataset (Method 1)

Training & Testing Rates	AUC	F1	Precision	Recall
70% training & 30% testing	0.921	0.875	0.875	0.875
50% training & 50% testing	0.917	0.869	0.869	0.869

 Table 5. Success of the KNN algorithm on the normalized dataset (Method 2)

Training & Testing Rates	AUC	F1	Precision	Recall
70% training & 30% testing	0.950	0.92	0.921	0.92
50% training & 50% testing	0.949	0.906	0.907	0.906

Table 4 illustrates the classification performance using the raw, non-normalized attribute data. The highest accuracy, achieved with 70% training and 30% testing, was 0.921. According to the F1-score, precision and recall success metrics, classification success rates ranging from 0.869 to 0.875 were obtained.

The classification performance measurement performed on normalized attribute data is shown in Table 5. In this case, the highest accuracy of 0.950 was obtained with 70% training and 30% testing. According to the F1-score, precision and recall success metrics, classification achievements ranging from 0.906 to 0.921 were obtained.



Figure 3. Classification performance using the model on the normalized and non-normalized datasets: (a) Success values based on 70% training and 30% testing data rates; (b) Success values based on 50% training and 50% test data rates



Figure 4. Classification success graphs (ROC curves) of two types of rice: (a) Cammeo rice; (b) Osmancik rice

The results demonstrate that increasing the proportion of training data in the model led to improved classification success. Additionally, the classification accuracy of the KNN algorithm was significantly enhanced when the normalized dataset was employed. Figure 3 compares the classification performance between the normalized and non-normalized datasets. The performance graph of the model designed by selecting 70% training and 30% testing rate on the normalized and non-normalized datasets is shown in subgraph (a) of Figure 3.

Subgraph (b) of Figure 3 shows the performance graph of the model designed by selecting 50% training and 50% testing rate on the normalized and non-normalized datasets. In the figure, it can be observed that the normalization process had a positive effect on the classification performance. Specifically, the use of normalization resulted in a 3.2% increase in classification success, as measured by the accuracy metric, and a 4.6% improvement in the F1-score metric when 50% training data was used.

According to the model in which the highest classification achievements were obtained in the study (normalized dataset +70% training and 30% test data rates), the classification success of each class was evaluated separately. According to the results obtained, it can be observed that Cammeo rice was classified with a higher success rate than Osmancik rice.

Figure 4 shows the Receiver Operating Characteristic (ROC) curve graph, indicating the classification success rates of the two types of rice. On the designed model, subgraph (a) of Figure 4 shows the classification success graph of the Cammeo rice, and subgraph (b) of Figure 4 shows that of the Osmancik rice.



Figure 5. Convolution matrix of the model

The correct and incorrect classification results of the designed model can be analyzed by using convolution matrices. In this study, the highest success rate was obtained from the convolution matrix of the model using Method 2, as shown in Figure 5. On the model, 2667 pieces of rice were used for training (70% rate) and 1143 pieces of rice for testing (30% rate).

In the figure, 493 Cammeo and 650 Osmancik rice classes were tested on the model. In the test process, 440 of the 493 rices belonging to the Cammeo class were classified correctly, but 53 were classified incorrectly (as Osmancik rice). In addition, in the test process, 610 of the 650 rices belonging to the Osmancik class were classified correctly, but 40 were classified incorrectly (as Cammeo rice).

3.1 Correlation Analysis of Attributes

Correlation analysis was performed on the developed model to show the relationship and direction of each attribute between classes. The correlation value levels of the attributes show the effects of the classification process. Positive correlation values are represented by +1, while negative values are represented by -1. When the correlation values are close to the limit values (-1 and +1), it is determined that there is a high-level correlation. When the values are close to 0, there is a low-level correlation. The correlation value of the attribute used has a great impact on the classification process when it is at a high level. If the correlation value is at a low level, it has little effect on the classification process. Table 6 shows the correlation values and levels of the attributes of each class used in the dataset. It can be seen that the highest correlation value is in the major axis length attribute and the lowest one is in the extent attribute. In addition, the medium correlation value is in the eccentricity attribute.

Attribute	Osmancik	Cammeo	Correlation Level
Major axis length	-0.992	+0.892	High
Perimeter	-0.879	+0.879	High
Convex area	-0.837	+0.837	High
Area	-0.835	+0.835	High
Eccentricity	-0.676	+0.676	High
Minor axis length	-0.439	+0.439	Medium
Extent	+0.170	-0.170	Low

Table 6. Correlation values of dataset attributes

4 Discussion

Some research on rice classifications has been conducted using different methods in classification processes. The success rates vary depending on the used methods. The comparison between the model developed in this study and other studies is shown in Table 7.

In the study by Cinar and Koklu [3], a total of 3810 rice grains belonging to Osmancik and Cammeo classes were imaged. Then they were processed by image processing methods and the attributes of each rice grain were created. Seven morphological attributes were used for each grain of rice. In the study, models were created using Logistic Regression (LR), Multilayer Perceptron (MLP), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB) and KNN ML algorithms. Classification performance measurement values were obtained for each algorithm. According to the results, the highest classification success rate was measured at 93.02% with the LR algorithm.

Hong et al. [16] used six datasets, with one of them containing the highest number of 4152 rice data records. The RF model was used to classify the rice in the datasets. According to the results, they measured a classification success rate of 90.54% for the RF algorithm. Nga et al. [17] used an optimized SVM model to classify the rice within the dataset containing 3400 rice data records and achieved a classification success rate of 93.94%. Using the same dataset,

Nga et al. [18] performed the classification process using improved convolutional neural network (CNN) models. According to the results, they measured 96.41% classification success rate with modified VGG16 and 97.88% with modified ResNet50 algorithm. Nusrat et al. [1] used RiceNet, InceptionV3, and ResNetInceptionV2 models of CNN to classify the rice in the dataset containing 4748 rice data records. According to the results, the classification success rates with RiceNet, InceptionV3 and ResNetInceptionV2 were 94%, 84% and 81%, respectively. In addition, Koklu et al. [19] conducted a research comparing the classification success of rice varieties using different deep learning methods.

Research	Algorithms and Methods Used	Dataset Used	Success Rates	
Cinar and Koklu [3]	LR, MLP, SVM, DT, RF, NB and KNN	Dataset containing 3810 rice sample data	According to the accuracy success metric: LR=93.02% MLP=092.86% SVM=92.83%	
			DT=92.49% RF=92.39% NB=91.71% KNN=88.58%	
Hong et al. [16]	RF	Six Vietnam rice seed datasets	According to the accuracy success metric: RF=90.54%	
Nga et al. [17]	SVM combined with binary particle swarm optimization	Dataset containing 3400 rice sample data	According to the accuracy success metric: SVM=93.94%	
Nga et al. [18]	Modified VGG16 and modified ResNet50	Dataset containing 3400 rice sample data	According to the accuracy success metric: Modified VGG16=96.41% Modified ResNet50=97.88%	
Nusrat et al. [1]	RiceNet, InceptionV3 and ResNetInceptionV2	Sher-eKashmir University of Agriculture Sciences and Technology (SKUAST) Srinagar. The dataset used in this study consisted of 4748 rice image data.	According to the accuracy success metric: RiceNet=94% InceptionV3=84% ResNetInceptionV2=81%	
Proposed model in this study	Min-max normalization and KNN	Dataset containing 3810 rice sample data	According to the accuracy success metric: KNN=95.0%	

Table 7. Comparison between the proposed model in this study and other studies

A dataset containing 3810 records was used in this study, which belong to the Osmancik and Cammeo rice classes and were shared as open source at the University of California, Irvine (UCI) [20]. In order to classify the rice in the dataset, min-max normalization was performed in the first stage. Then the classification success rates were measured on the proposed model using the KNN algorithm according to the variable training dimensions (50% and 70%). A classification success rate of 95% was obtained with the KNN algorithm on the normalized dataset. In this developed model, unlike other studies, the normalization preprocessing step was performed in the dataset. In addition, this study reveals that the min-max normalization of the dataset increases the classification success rate, which is its innovative strength.

5 Conclusions

In this study, different types of rice, an important food type in the world and widely used by humans, were classified. Osmancik and Cammeo rice species cultivated and consumed in Turkey were selected for classification. The dataset was downloaded from the UCI repository, which is available as open source. A model was designed using the KNN ML algorithm with variable training and test data rates. Before the classification process, the min-max normalization was performed on the existing dataset records, thereby arranging the attribute data of the records. The normalized dataset and non-normalized datasets were tested on the proposed model. In the testing processing, it was observed that the classification success rate increased in the normalized data. As a result, this study proves that the min-max data normalization process performed on the datasets can increase the classification success rate of intelligent systems. In future studies, classification performance could be measured by applying normalization

processes on different datasets and various learning algorithms, thereby proving the success of normalization processes on different models.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflict of interest.

References

- M. U. D. Nusrat, A. Assif, D. Rayees, B. Muzafar, U. S. Saqib, M. Tabasum, I. Zahir, G. Wahid, and Y. Aamir, "RiceNet: A deep convolutional neural network approach for classification of rice varieties," *Expert Syst. Appl.*, vol. 235, p. 121214, 2023. https://doi.org/10.1016/j.eswa.2023.121214
- [2] B. Arora, N. Bhagat, L. R. Saritha, and S. Arcot, "Rice grain classification using image processing & machine learning techniques," in 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 205–208. https://doi.org/10.1109/ICICT48043.2020.9112418
- [3] I. Cinar and M. Koklu, "Classification of rice varieties using artificial intelligence methods," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 3, pp. 188–194, 2019. https://doi.org/10.18201/ijisae.2019355381
- [4] K. K. Jena, S. K. Bhoi, D. Mohapatra, C. Mallick, and P. Swain, "Rice disease classification using supervised machine learning approach," in 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2021, pp. 328–333. https://doi.org/10.1109/I-SMAC52330.2021.964 1054
- [5] R. Ruslan, S. Khairunniza-Bejo, M. Jahari, and M. F. Ibrahim, "Weedy rice classification using image processing and a machine learning approach," *Agriculture*, vol. 12, no. 5, p. 645, 2022. https://doi.org/10.3390/agricultur e12050645
- [6] I. A. Ozkan, M. Koklu, and R. Saracoglu, "Classification of pistachio species using improved KNN classifier," *Prog. Nutr.*, vol. 23, no. 2, p. e2021044, 2021. http://doi.org/10.23751/pn.v23i2.9686
- [7] D. Singh, Y. S. Taspinar, R. Kursun, I. Cinar, M. Koklu, I. A. Ozkan, and H. N. Lee, "Classification and analysis of pistachio species with pre-trained deep learning models," *Electronics*, vol. 11, no. 981, pp. 1–14, 2022. https://doi.org/10.3390/electronics11070981
- [8] R. Butuner, I. Cinar, Y. S. Taspinar, R. Kursun, M. H. Calp, and M. Koklu, "Classification of deep image features of lentil varieties with machine learning techniques," *Eur Food Res. Technol.*, vol. 249, pp. 1303–1316, 2023. https://doi.org/10.1007/s00217-023-04214-z
- [9] A. Çelik, "Determination of the classification success of KNN algorithm distance metric methods on wheat seeds dataset," *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, vol. 23, no. 5, pp. 1142–1149, 2023. https://doi.org/10.35414/akufemubid.1263900
- [10] N. A. Ayele and H. K. Tamiru, "Developing classification model for chickpea types using machine learning algorithms," *Int. J. Innov. Technol. Explor. Eng.*, vol. 10, no. 1, pp. 5–11, 2020. https://doi.org/10.35940/ijitee. A8057.1110120
- [11] K. Karunamurthy, A. A. Janvekar, P. Palaniappan, V. Adhitya, T. T. K. Lokeswar, and J. Harish, "Prediction of IC engine performance and emission parameters using machine learning: A review," J. Therm. Anal. Calorim., vol. 148, pp. 3155–3177, 2023. https://doi.org/10.1007/s10973-022-11896-2
- [12] M. Shah, K. Banker, J. Patel, and D. Rao, "Comparative analysis of deep learning architectures for rice crop image classification," in *Proceedings of 4th International Conference on Artificial Intelligence and Smart Energy*, Coimbatore, India, 2024, pp. 245–259. https://doi.org/10.1007/978-3-031-61471-2_18
- [13] Y. Zhang, Q. F. Wang, X. F. Chen, Y. C. Yan, R. M. Yang, Z. T. Liu, and J. H. Fu, "The prediction of spark-ignition engine performance and emissions based on the SVR algorithm," *Processes*, vol. 10, no. 2, p. 312, 2022. https://doi.org/10.3390/pr10020312
- [14] A. Çelik, "Improving iris dataset classification prediction achievement by using optimum k value of kNN algorithm," J. ESTUDAM Inf., vol. 3, no. 2, pp. 23–30, 2022. https://doi.org/10.53608/estudambilisim.1071335
- [15] P. Guru, J. Sathyapriya, K. V. R. Rajandran, J. Bhuvaneswari, and C. Parimala, "Product sales forecasting and prediction using machine learning algorithm," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 4, pp. 355–366, 2023.
- [16] T. T. Hong, T. T. Thanh Hai, L. T. Lan, V. T. Hoang, V. Hai, and T. T. Nguyen, "Comparative study on vision based rice seed varieties identification," in *IEEE Seventh International Conference on Knowledge and Systems Engineering (KSE)*, Ho Chi Minh City, Vietnam, 2015, pp. 377–382. https://doi.org/10.1109/KSE.2015.46

- [17] T. T. K. Nga, P. V. Tuan, D. M. Tam, I. Koo, V. Y. Mariano, and D. H. Tuan, "Combining binary particle swarm optimization with support vector machine for enhancing rice varieties classification accuracy," *IEEE Access*, vol. 9, pp. 66062–66078, 2021. https://doi.org/10.1109/ACCESS.2021.3076130
- [18] T. T. K. Nga, P. V. Tuan, D. M. Tam, I. Koo, V. Y. Mariano, and D. H. Tuan, "Enhancing the classification accuracy of rice varieties by using convolutional neural networks," *Int. J. Electr. Electron. Eng. Telecommun.*, vol. 12, no. 2, pp. 150–160, 2023. https://doi.org/10.18178/ijeetc.12.2.150-160
- [19] M. Koklu, I. Cinar, and Y. S. Taspinar, "Classification of rice varieties with deep learning methods," *Comput. Electron. Agric.*, vol. 187, p. 106285, 2021. https://doi.org/10.1016/j.compag.2021.106285
- [20] D. Dua and C. Graff, "UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA," UCI Machine Learning Repository, 2023. https://archive.ics.uci.edu