International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 3, June 2025, pp. 3212~3225
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i3.pp3212-3225

Morphological features for multi-model rice grain classification

Suma D.[1], Narendra V. G.[1], Raviraja Holla M.[2], Darshan Holla M.[2]
[1]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India
[2]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

Article Info

Article history:

Received Sep 20, 2024
Revised Mar 6, 2025
Accepted Mar 20, 2025

Keywords:

Machine learning
Morphological features
Rice grain
Rice grain classification
Segmentation

ABSTRACT

In the realm of agriculture and food processing, the automated classification of rice grains holds significant importance. The diverse varieties of rice available demand a systematic approach to categorization. This study tackles this challenge by employing diverse machine learning models, including support vector machine (SVM), random forest (RF), logistic regression (LR), decision tree (DT), Gaussian naive Bayes (GNB), and k-nearest neighbors (K-NN). The dataset, sourced from Kaggle, features five distinct rice types: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. After the images undergo preprocessing, a set of 13 distinct morphological features is extracted. These features ensure a comprehensive representation of rice grains for accurate classification. This study aims to create an intelligent system for efficient and precise rice grain classification, contributing to optimizing agricultural and food industry processes. Among the models, K-NN demonstrated the highest classification accuracy at 97.80%, surpassing random forest (97.51%), DT (97.48%), GNB (96.99%), SVM (96.85%), and LR (96.05%). Our proposed K-NN-based classification model achieves an accuracy of 97.8%, demonstrating competitive performance and outclassing several state-of-the-art methods such as artificial neural network (ANN) and modified visual geometry group16 (VGG16) while maintaining simplicity and computational efficiency. This underscores the effectiveness of K-NN and RF in enhancing the precision of rice variety classification.

Corresponding Author:

Narendra V. G.
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE)
Manipal, Karnataka, 576104, India
Email: narendra.vg@manipal.edu

1. INTRODUCTION

Rice, a crucial dietary staple for a substantial segment of the global population, occupies a central position in ensuring worldwide food security and significantly influences economies while supporting livelihoods. With its unmatched nutritional importance, rice serves as a primary source of carbohydrates and essential nutrients, forming the dietary foundation for billions of people worldwide. The intricate relationship between the rising global production of rice and the increasing demand emphasizes the necessity for accurate classification methods. In the quest to optimize agricultural practices and ensure the production of high-quality grains, this study delves into the classification of diverse rice varieties. With rice farming serving as a cornerstone of economies, particularly in Asia, this research aims to enhance agricultural efficiency and safeguard this vital global food resource. Arora *et al.* [1] introduce an automated rice grain classification system employing machine learning and image processing techniques. Using logistic regression and support vector machine algorithms, they achieve 96% and 92% accuracy, respectively, in classifying various rice types. The system streamlines classification effectively but lacks advanced features for chalkiness detection and comprehensive grain quality evaluation, suggesting potential for further refinement.

Nagod and Ranathunga [2] present a novel method for rice quality identification using image processing and machine learning. Their approach achieves 96.0% segmentation and 88.0% classification accuracy across six rice categories, showing improved seed segmentation and reduced processing time. While efficient with lightweight algorithms, the method requires enhancements in stone identification accuracy, indicating scope for future research improvements. Kiratiratanapruk *et al.* [3] propose machine vision for automated paddy rice seed classification, offering a cost-effective alternative to manual methods. Comparing statistical (logistic regression (LR), linear discriminant analysis (LDA), k-nearest neighbors (K-NN), support vector machine (SVM)) and deep learning models (visual geometry group16 (VGG16), VGG19, Xception, InceptionV3, InceptionResNetV2), SVM achieves subgroup accuracies of 90.61%, 82.71%, and 83.9%, while InceptionResNetV2 attains 95.15%. Deep learning surpasses traditional methods by up to 11.24%, indicating potential for improved seed quality inspection in agriculture. Aukkapinyo *et al.* [4] propose an automated rice grain classification approach using a mask region-based convolutional neural network (R-CNN) based method and marker-based watershed algorithm. Their model achieves a mean average precision (mAP) of 1.0 for sticky and paddy rice grains when aligned manually, and an average mAP of approximately 0.75 for classifying five subtypes. Intriguingly, their trained classifier outperforms human experts with an average mAP of 0.80. Krishna *et al.* [5] have developed an automated system utilizing image processing techniques to categorize rice grains. Their system, utilizing MATLAB with neural networks (NN) and SVM, achieves precise and efficient assessment of rice quality, surpassing traditional manual methods. They also suggest improvements in stone identification accuracy through additional validation procedures. Ibrahim *et al.* [6] proposes an automated method for classifying rice grains using image processing methods. They apply feature extraction techniques and multi-class SVM classification to distinguish between three types of rice, achieving an impressive accuracy rate of 92.22% on a test set of 90 images. Ruslan *et al.* [7] employ image processing techniques combined with machine learning for the classification of weedy rice seeds. Their method achieves an impressive 85.3% sensitivity and 97.9% accuracy using logistic regression with RGB images. The optimized SVM model achieves a high accuracy rate of 97.3%. Singh and Chaudhury [8] present a cascade network classifier designed for the classification of rice grains. Utilizing a combination of morphological, color, texture, and wavelet features, the model achieves accuracy rates of 97.75% with morphological features and 96.75% with three selected features.

Cinar and Koklu [9] focus on classifying five rice varieties using machine learning algorithms, achieving peak accuracies of 97.99% with random forest using morphological features and 99.25% using logistic regression for color features. Farahnakian *et al.* [10] investigate novel deep-learning models, evaluating residual network (ResNet), visual geometry group (VGG) network, EfficientNet, and MobileNet. The analysis showcases EfficientNet achieving the highest accuracy (99.67%), while MobileNet excels in speed. Nga *et al.* [11] classify 17 Vietnamese rice varieties using image processing techniques, achieving accuracies of 93.94% with a novel binary particle swarm optimization (BPSO)+SVM method and 89.1% with sparse representation-based classification (SRC). Kuo *et al.* [12] achieve an 89.1% accuracy in identifying 30 rice grain varieties using image analysis and SRC. Carneiro *et al.* [13] effectively characterize rice grain physicochemical composition using near-infrared spectroscopy (NIR) with machine learning models, achieving high accuracy (93.9%) with the random tree model (RandT). Ahmed *et al.* [14] categorize image-based rice grains using geometric, deep learning, supervised, unsupervised, and statistical approaches, highlighting the efficacy of deep learning techniques. Srimulyani and Musdholifah [15] enhance rice variety identification in Indonesia using NN, achieving improved accuracy through geometry features. Singh and Chaudhury [16] classify four varieties of bulk rice grain images using back-propagation neural network (BPNN), achieving an average classification accuracy exceeding 96% across all features and datasets. Aznan *et al.* [17] employed computer vision and machine learning methods to categorize commercial rice samples based on dimensionless morphometric and color parameters extracted from smartphone photos. Their artificial neural network (ANN) model, using Bayesian regularization (BR) technique, achieved the highest classification accuracy of 93.9% among 15 rice varieties. Hamzah and Mohamed [18] discussed the significance of employing technology to classify white rice grain quality, achieving a high accuracy of 96% using BPNN. To successfully breed rice and satisfy customer desires, it is essential to determine the characteristics of rice grain quality (RGQ), which include milling, storage, cooking, nutritional value, and market qualities. Regional preferences differ; for example, Middle Eastern customers prefer fragrant, well-milled long-grain rice, whereas Europeans choose long-grain, nonaromatic rice. Global demand for high-quality rice is growing. It can be very helpful to generate new rice varieties with improved RGQ if the genetic mechanisms underlying grain quality quantitative trait loci (QTLs) and their constraints are understood [19].

Ahad *et al.* [20] compare CNN-based deep learning architectures for detecting and localizing nine epidemic rice diseases in Bangladesh, achieving an accuracy of 98% with an ensemble model. Tran-Thi-Kim *et al.* [21] classify 17 rice grain varieties using CNN models, achieving accuracies of 92.82% with ANN, 96.41% with modified VGG16, and 97.88% with modified ResNet50. Patel and Sharaff [22] categorize ten paddy rice varieties using image processing, achieving advantages in speed and cost-effectiveness. Kurade *et al.* [23] introduce a cost-effective rice quality assessment system using machine learning algorithms, achieving a high accuracy of 77% with random forest classifier. Deepika *et al.* [24] evaluate grain quality traits in 21 rice hybrids using digital imaging, accurately classifying grain size and type, and identifying potential resources for aroma-type rice breeding programs.

This study addresses the challenges of manual rice grain classification, which is often labor-intensive and prone to subjective inconsistencies. By applying machine learning techniques, we aim to revolutionize the classification process, ensuring precision and consistency. Key contributions include:

a. Automation of rice grain classification: transitioning from traditional human-dependent methods to a fully automated system using advanced machine learning algorithms, including support vector machine (SVM), Gaussian naive Bayes (GNB), logistic regression (LR), decision tree (DT), k-nearest neighbors (K-NN), and random forest (RF).
b. Enhancing classification accuracy: leveraging data-driven algorithms to improve the accuracy and reliability of rice grain classification, meeting the demands of precision agriculture.
c. Feature extraction for grain representation: identifying and extracting critical morphological features to comprehensively represent rice grains, enabling robust classification.
d. Performance evaluation of algorithms: conducting comparative analysis of various machine learning models to identify the most effective approach for rice grain classification.
e. Rigorous validation: ensuring reliability through extensive testing and analysis to validate the system's performance across diverse rice grain varieties.

This study significantly contributes to precision agriculture by advancing the systematic categorization of rice varieties, paving the way for more consistent and efficient agricultural practices. The subsequent sections of this paper unfold seamlessly, with section 2 detailing materials and methods for rice grain classification utilizing diverse machine learning algorithms, section 3 offering a comprehensive analysis of results, and section 4 encapsulating our findings while proposing avenues for future research in the domain of rice grain classification.

2. MATERIALS AND METHODS

This section outlines the sequential process implemented to achieve the objectives of rice grain classification. The flowchart shown in Figure 1 illustrates the overall process of rice grain classification methodology. The subsections spanning from 2.1 to 2.6 will provide a brief overview of each step outlined in the flowchart.

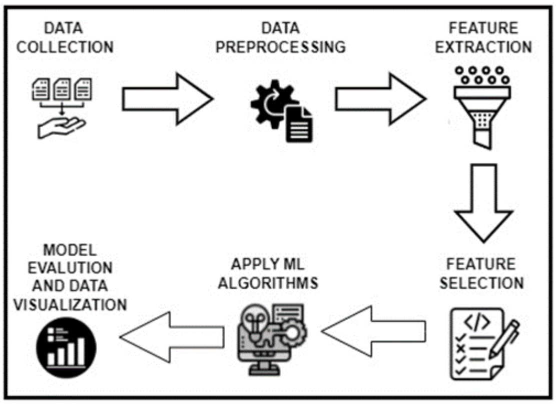


Figure 1. Infographic flowchart of overall process

2.1. Data collection

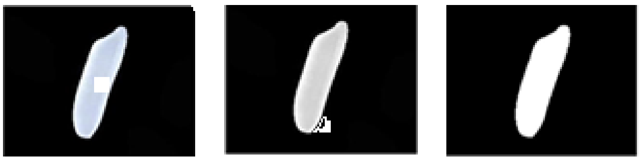
The study begins with the collection of data from a publicly available repository on Kaggle [25]. The dataset is systematically organized into a main directory named 'Rice_Dataset,' which contains five subfolders. Each subfolder represents a specific class label corresponding to different varieties of rice: Arborio, Basmati, Sala, Jasmine, and Karacadag. These subfolders collectively house a total of 75,000 images, with each class folder containing precisely 15,000 images. This well-structured dataset serves as a robust foundation for the classification task, ensuring a balanced and comprehensive representation of the five rice varieties. Such meticulous organization aids in effective preprocessing, training, and evaluation of the classification models developed in the study.

2.2. Data preparation

Upon acquiring the images from the Kaggle repository, a Pandas DataFrame, depicted in Table 1, is generated to systematically arrange the data. The DataFrame comprises two primary columns: one for storing the file paths directing to the gathered images and another for the corresponding class labels. This organized format facilitates efficient manipulation and analysis of the dataset. In this phase, images were transformed into grayscales and subjected to Otsu’s thresholding for binary segmentation. Connected component labeling identified distinct regions, facilitating essential property extraction. These preprocessing steps aim to normalize and streamline data for efficient feature extraction. The Figures 2(a), 2(b), and 2(c) depict the original, grayscale, and segmented rice grain images, respectively.

Table 1. DataFrame samples comprising image paths and corresponding labels

Sl. No.	File path	Label
1	Rice_Dataset/Arborio/Arborio (9099).jpg	Arborio
2	Rice_Dataset/Arborio/Arborio (91).jpg	Arborio
…	…	…
74999	Rice_Dataset/Karacadag/Karacadag (10897).jpg	Karacadag
75000	Rice_Dataset/Karacadag/Karacadag (10898).jpg	Karacadag



(a) (b) (c)

Figure 2. Preprocessing steps (a) original image, (b) grayscale image, and (c) segmented image

The rationale behind the pre-processing techniques employed lies in their ability to standardize and enhance the quality of the input data for subsequent analysis. Converting images to grayscale simplifies computational complexity while preserving essential visual information necessary for classification. Otsu's thresholding for binary segmentation enables the separation of rice grains from the background, facilitating more precise feature extraction. Additionally, connected component labeling aids in identifying and isolating distinct regions within the segmented images, thereby enhancing the accuracy of morphological property extraction. Overall, these pre-processing steps aim to streamline the data and optimize it for efficient feature extraction, ultimately improving the performance of the classification model.

2.3. Feature extraction

The feature extraction phase involves transforming raw data into a format suitable for modeling. In this phase, essential 13 morphological features were derived from the segmented rice grain images. This includes characteristics such as major and minor axis lengths, perimeter, eccentricity, area, convex area, extent, solidity, orientation, equivalent diameter, compactness, aspect ratio, and roundness. The major and minor axis lengths represent the longest and shortest diameters of the rice grain, respectively. Perimeter refers to the length of the rice grain's boundary, while eccentricity describes the shape of the rice grain's ellipse. Area and convex area quantify the size of the rice grain and its convex hull, respectively. Extent measures the ratio of the area of the rice grain to the area of its bounding box, providing insights into the extent of coverage. Solidity indicates the compactness of the rice grain, calculated as the ratio of the area of the rice grain to its convex area. Orientation describes the angle of the major axis of the rice grain's bounding ellipse with the horizontal axis. Equivalent diameter represents the diameter of a circle with the same area as the rice grain. Compactness quantifies how closely the rice grain's shape approaches that of a circle. Aspect ratio is the ratio of the major axis length to the minor axis length, indicating elongation or compactness. Lastly, roundness measures how close the rice grain's shape is to a perfect circle, with a value of 1 indicating a perfect circle. These morphological features were carefully selected for their relevance to rice grain characteristics and their potential to contribute to accurate classification.

These features serve as quantitative descriptors, capturing crucial information about the shape and structure of individual rice grains. The extraction process facilitates the creation of a feature matrix, forming the basis for subsequent machine learning model training and analysis. Furthermore, these extracted features were written to a CSV file for comprehensive data storage and further exploration. Table 2 visually presents the File path, Label and the 13 extracted morphological features for three rice grain images, providing a visual representation of the quantitative characteristics derived from the image analysis process.

Table 2. File path, label, and the 13 extracted morphological features of rice samples

File path	Label	Area	Perimeter	Major axis	Minor axis	Extent
Rice_Dataset/Arborio/Arborio(9099).jpg	Arborio	7835	366.132	138.316	73.319	0.625
Rice_Dataset/Arborio/Arborio(91).jpg	Arborio	7625	361.989	136.584	72.428	0.635
Rice_Dataset/Arborio/Arborio(910).jpg	Arborio	6859	342.936	135.926	64.750	0.764
…	…	…	…	…	…	…

Eccentricity	**Convex area**	**Solidity**	**Orientation**	**Equivalent diameter**	**Compactness**	**Aspect ratio**	**Roundness**
0.847	**8076**	**0.970**	**0.728**	**99.879**	**1.361**	**1.886**	**0.521**
0.847	**7872**	**0.968**	**-0.622**	**98.531**	**1.367**	**1.885**	**0.520**
0.879	**6986**	**0.981**	**-0.113**	**93.451**	**1.364**	**2.099**	**0.473**
…	**…**	**…**	**…**	**…**	**…**	**…**	**…**

2.4. Feature selection

In the feature selection stage, a meticulous process is undertaken to identify and retain the most informative subset of features among the extracted morphological descriptors. Recursive feature elimination (RFE) is employed to systematically eliminate less significant features, ensuring that the refined feature set maintains the highest relevance for subsequent machine learning model training. This strategic selection enhances model interpretability, mitigates overfitting, and optimizes the predictive capability of the chosen features.

2.5. Apply machine learning algorithms

The machine learning algorithms used in this study are summarized in Table 3, providing a comprehensive overview of the classification methods employed. Support vector machines are employed for their effectiveness in handling complex decision boundaries and high-dimensional feature spaces. Given the intricate morphological characteristics of rice grains, SVM's ability to create optimal hyperplanes for classification is advantageous. The linear kernel is chosen for its simplicity and suitability for the dataset. Random Forests are chosen for their robustness and ability to handle many features. In the context of rice grain classification, where numerous morphological features contribute to the differentiation of classes, RF provides an ensemble of decision trees for improved accuracy. Setting the tree count in the forest at 100 achieves a balance between computational efficiency and the effectiveness of the model. Logistic regression is selected for its simplicity and efficiency in binary and multiclass classification tasks. In rice grain classification, where interpretability is valuable, LR provides a straightforward probabilistic framework. The 'saga' solver is employed for optimization. The iteration limit is established at 14,000 to ensure convergence.

Table 3. Machine learning algorithms used for classification

Algorithm	Description
Support vector machine	Supervised learning model for classification tasks
Random forest	Ensemble learning method for classification tasks
Logistic regression	Regression analysis for binary classification tasks
Decision tree	Tree-like model for classification and regression tasks
Gaussian naive Bayes	Probabilistic classifier based on Bayes' theorem
K-nearest neighbors	Instance-based learning method for classification

Decision trees are selected for their interpretability and suitability for both numerical and categorical data. In rice grain classification, where understanding the decision path is crucial, decision trees provide a clear structure based on morphological features. Reduced error pruning (REP) is applied to optimize the tree structure by removing unnecessary branches and prevent overfitting. Gaussian naive Bayes is chosen for its simplicity and effectiveness in handling continuous data. In the context of rice grain classification, where morphological features can be considered as continuous variables, GNB's assumption of feature independence simplifies the modeling process. The utilization of K-NN is done due to its simplicity and capacity to capture local patterns within the data. In rice grain classification, where the similarity of morphological features is significant, K-NNs approach of assigning labels based on neighboring instances proves effective. The number of neighbors (k) is set to 5 for a balanced trade-off between bias and variance.

The rationale behind the selection of specific machine learning algorithms stems from their proven effectiveness in classification tasks, particularly in contexts like rice grain classification. Support vector machine is chosen for its ability to handle high-dimensional data and non-linear decision boundaries effectively. Random forest is selected due to its robustness to overfitting and its capability to handle large datasets with high dimensionality. Logistic regression is included for its simplicity, interpretability, and suitability for binary classification tasks. Decision tree is chosen for its intuitive representation of decision rules and ease of understanding. Gaussian naive Bayes is included for its simplicity, scalability, and efficiency, particularly in cases of small training datasets. Lastly, k-nearest neighbors are selected for their simplicity and flexibility in handling multi-class classification problems, relying on local information rather than assuming a specific data distribution. Collectively, these algorithms offer a diverse range of methodologies that complement each other, ensuring comprehensive exploration of the rice grain classification problem. Furthermore, the optimization of hyperparameters involved techniques such as grid search, which systematically explores diverse parameter combinations to identify the optimal model. Additionally, cross-validation was utilized to enhance this process by evaluating the model's generalization ability across various data subsets. Through these methodologies, our models were fine-tuned to attain optimal performance on the dataset while mitigating the risk of overfitting.

2.6. Model evaluation and data visualization

In this phase, a comprehensive set of metrics and techniques has been employed to assess and portray the performance of the classification models. The confusion matrix presents a detailed breakdown of predicted and actual class labels, shedding light on the model's classification performance. Figure 3 illustrates the confusion matrix for multiclass classification, aiding in the interpretation of classification results. From the confusion matrix, several metrics are calculated to holistically assess the performance of the rice grain classification model. Equations (1)-(4) provide formulas for averaged precision, averaged recall, F1-score, and averaged accuracy calculations.

		Prediction Class				
		C1	C2	C3		Cn
Actual class	C1	T1	F12	F13	...	F1n
	C2	F21	T2	F23	...	F2n
	C3	F31	F32	T3	...	F3n
	...	...	...	...	...	...
	Cn	Fn1	Fn2	Fn3	...	Tn

Note: C: Class, T:True, and F: False

Figure 3. Confusion matrix for multiclass classification

$$\text{Average precision } = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l} \times 100 \tag{1}$$

$$\text{Average recall } = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l} \times 100 \tag{2}$$

$$\text{F1} - \text{score } = \frac{2 \times \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l} \times \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}}{\frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l} + \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}} \times 100 \tag{3}$$

$$\text{Average accuracy } = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l} \times 100 \tag{4}$$

3. RESULTS AND DISCUSSION

The performance evaluation of our rice grain classification models was carried out on a system equipped with a dual-core Intel® Core™ i3-2370M processor and 4GB of RAM. To assess the effectiveness of the models in distinguishing between different varieties of rice, we employed various metrics as outlined in section 3, including 3.1 confusion matrix, 3.2 performance metrics, 3.3 receiver operating characteristic (ROC) curve, and 3.4 performance comparison and justification of the proposed model. These metrics, along with the subsequent discussion, provide a comprehensive evaluation of the classification models.

3.1. Confusion matrix

In Tables 4 to 9, the confusion matrices depict the performance of various models-SVM, RF, LR, DT, GNB, and K-NN respectively. Each row signifies the true class, and each column denotes the predicted class. Diagonal elements (top-left to bottom-right) signify accurate predictions, while off-diagonal elements indicate misclassifications.

Table 4. Confusion matrix for SVM

		Predicted Labels				
		Arborio	Basmati	Ipsala	Jasmine	Karacadag
True Labels	Arborio	3618	0	1	10	134
	Basmati	0	3539	0	190	0
	Ipsala	10	0	3862	10	0
	Jasmine	8	131	6	3548	1
	Karacadag	90	0	0	0	3592

Table 5. Confusion matrix for RF

		Predicted Labels				
		Arborio	Basmati	Ipsala	Jasmine	Karacadag
True Labels	Arborio	3601	0	7	8	147
	Basmati	0	3564	0	75	0
	Ipsala	7	0	3862	13	0
	Jasmine	7	97	7	3582	1
	Karacadag	98	0	0	0	3584

Table 6. Confusion matrix for LR

		Predicted Labels				
		Arborio	Basmati	Ipsala	Jasmine	Karacadag
True Labels	Arborio	3528	0	4	25	206
	Basmati	0	3531	0	198	0
	Ipsala	43	0	3834	5	0
	Jasmine	19	135	11	3525	4
	Karacadag	92	0	0	0	3590

Table 7. Confusion matrix for DT

		Predicted Labels				
		Arborio	Basmati	Ipsala	Jasmine	Karacadag
True Labels	Arborio	3580	0	3	11	169
	Basmati	1	3661	0	67	0
	Ipsala	9	0	3858	14	1
	Jasmine	4	105	11	3573	1
	Karacadag	75	0	0	0	3607

Table 8. Confusion matrix for GNB

		Predicted Labels				
		Arborio	Basmati	Ipsala	Jasmine	Karacadag
True Labels	Arborio	3603	0	0	18	142
	Basmati	0	3560	0	169	0
	Ipsala	`17	0	3480	25	0
	Jasmine	0	79	1	3613	1
	Karacadag	112	0	0	0	3570

3.2. Performance metrics

Figures 4 to 7 compare the precision, recall, F1-score, and accuracy of six machine learning models in classifying five different types of rice grains. In terms of precision in Figure 4, K-NN achieves the highest average precision (97.60%), followed by RF (97.40%), DT (97.40%), GNB (97.00%), SVM (96.60%), and LR (96.00%). For recall in Figure 5, K-NN again leads with the highest average recall (97.80%), with RF and DT (both 97.40%), GNB (97.00%), SVM (96.80%), and LR (96.20%) following. Similarly, for the F1-score in Figure 6, K-NN maintains the highest average F1-score (97.80%), followed by RF and DT (both 97.60%), SVM and GNB (both 97.00%), and LR (96.00%). Finally, in terms of accuracy in Figure 7, K-NN emerges as the highest-performing classifier with 97.80% accuracy, followed by RF (97.51%), DT (97.48%), GNB (96.99%), and both SVM and LR (96.85%).

Table 9. Confusion matrix for K-NN

		Predicted Labels				
		Arborio	Basmati	Ipsala	Jasmine	Karacadag
True Labels	Arborio	3610	0	1	8	144
	Basmati	2	3564	0	73	0
	Ipsala	`11	0	3856	15	0
	Jasmine	6	176	6	3605	1
	Karacadag	78	0	0	0	3604

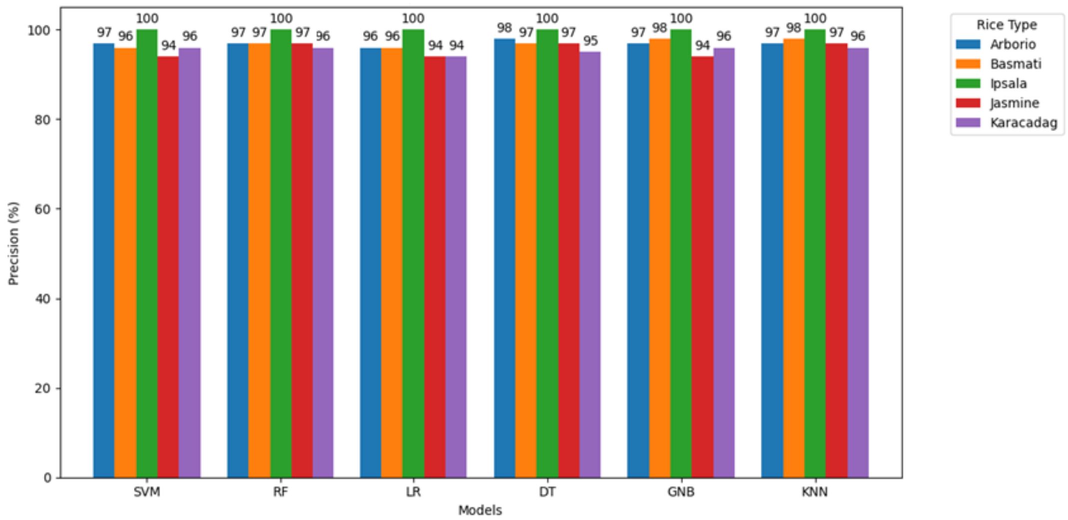


Figure 4. Comparison of precision of various machine learning models for rice grain classification

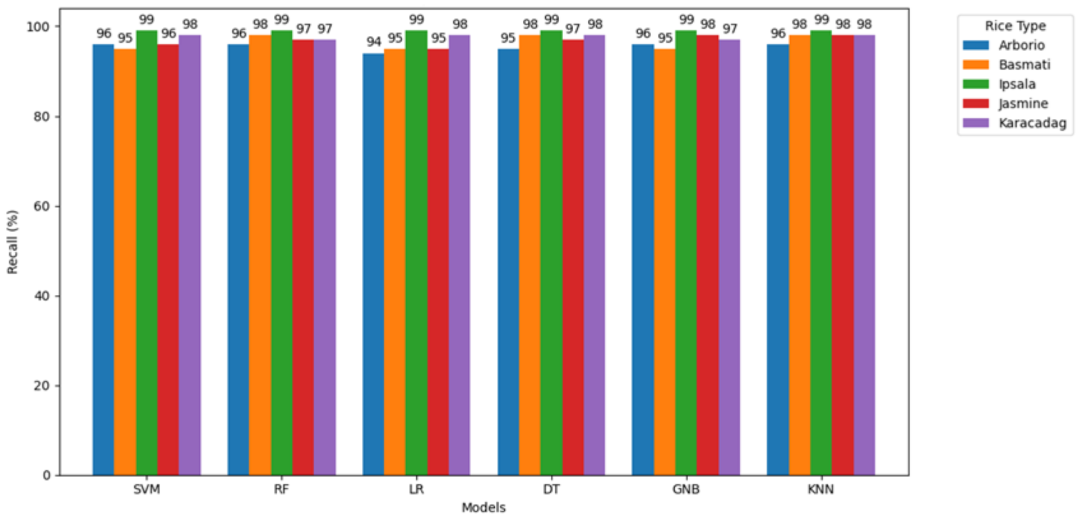


Figure 5. Comparison of recall of various machine learning models for rice grain classification

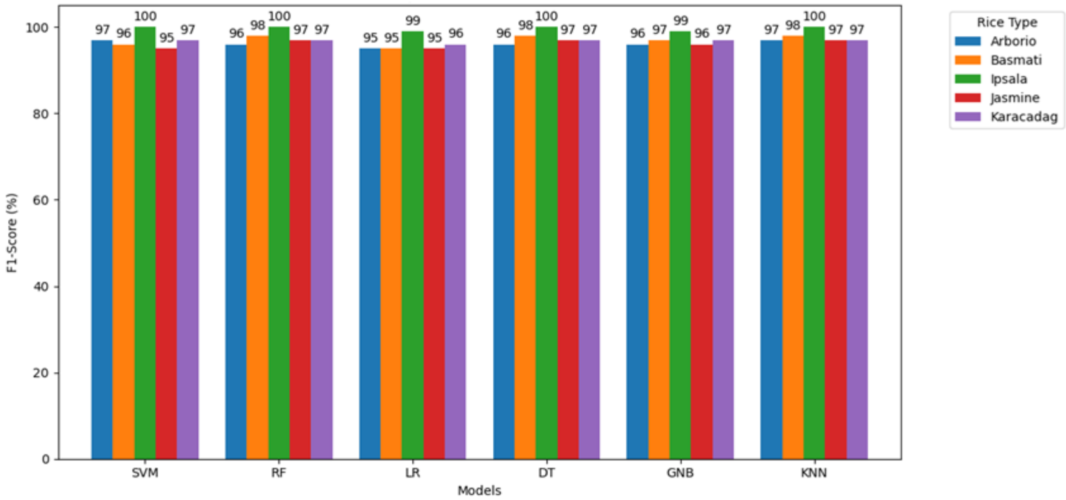


Figure 6. Comparison of F1-score of various machine learning models for rice grain classification

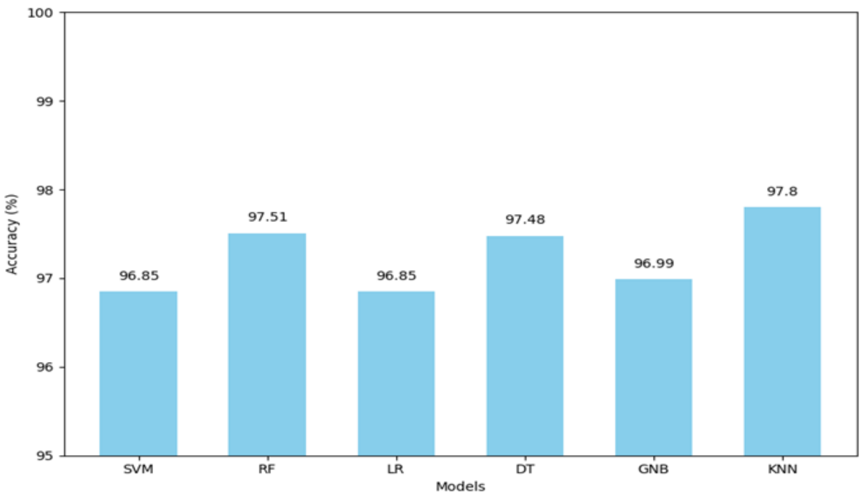


Figure 7. Comparison of accuracy of various machine learning models for rice grain classification

3.3. Receiver operating characteristic (ROC) curve

In multi-class classification, the ROC curve signifies the overall discriminatory power of the model across all classes. The area under the ROC curve (AUC) provides a measure of the model's ability to distinguish between different classes, with higher AUC indicating better performance. Overall, the ROC curve for multi-class classification provides valuable insights into the model's classification accuracy across multiple classes and the effectiveness of the chosen discrimination threshold. To enhance the interpretability of the results, ROC curves have been utilized. Additionally, the AUC is calculated, providing a summarized metric for the model's discrimination ability. The robust elucidation of the models' effectiveness in the intricate task of rice grain classification is achieved through the concerted application and integration of a diverse array of model evaluation techniques and sophisticated data visualization methods. We performed a one-way analysis of variance (ANOVA) statistical test to assess if there were any statistically significant variations in accuracy across the means of the six different models utilized in the analysis. The obtained high p-value indicates that our models reveal no notable disparities in accuracy among the machine learning models.

Figures 8(a) to 8(f) illustrates the ROC curves for six classification models-SVM, RF, LR, DT, GNB, and K-NN respectively. These curves visually represent each model's ability to distinguish between true positive and false positive rates across different classification thresholds. The ROC analysis facilitates a comprehensive assessment of each model's efficacy in binary classification tasks, aiding in the selection of optimal models based on their discriminatory power and overall accuracy. The corresponding AUC values, summarized in Table 10, provide a concise measure of discriminative performance, with higher AUC values indicating superior discriminatory abilities. Different models perform differently for each rice grain class. For example, In Table 10, in the Arborio class, the GNB model achieves the highest AUC value of 0.9976, indicating strong performance in distinguishing Arborio rice grains. Similarly, in the Jasmine class, the RF model achieves the highest AUC value of 0.9981. While RF generally performs well across all classes, certain models may excel in specific classes. For example, the LR model achieves a particularly high AUC value of 0.9997 for the Ipsala class, indicating its effectiveness in distinguishing Ipsala rice grains.

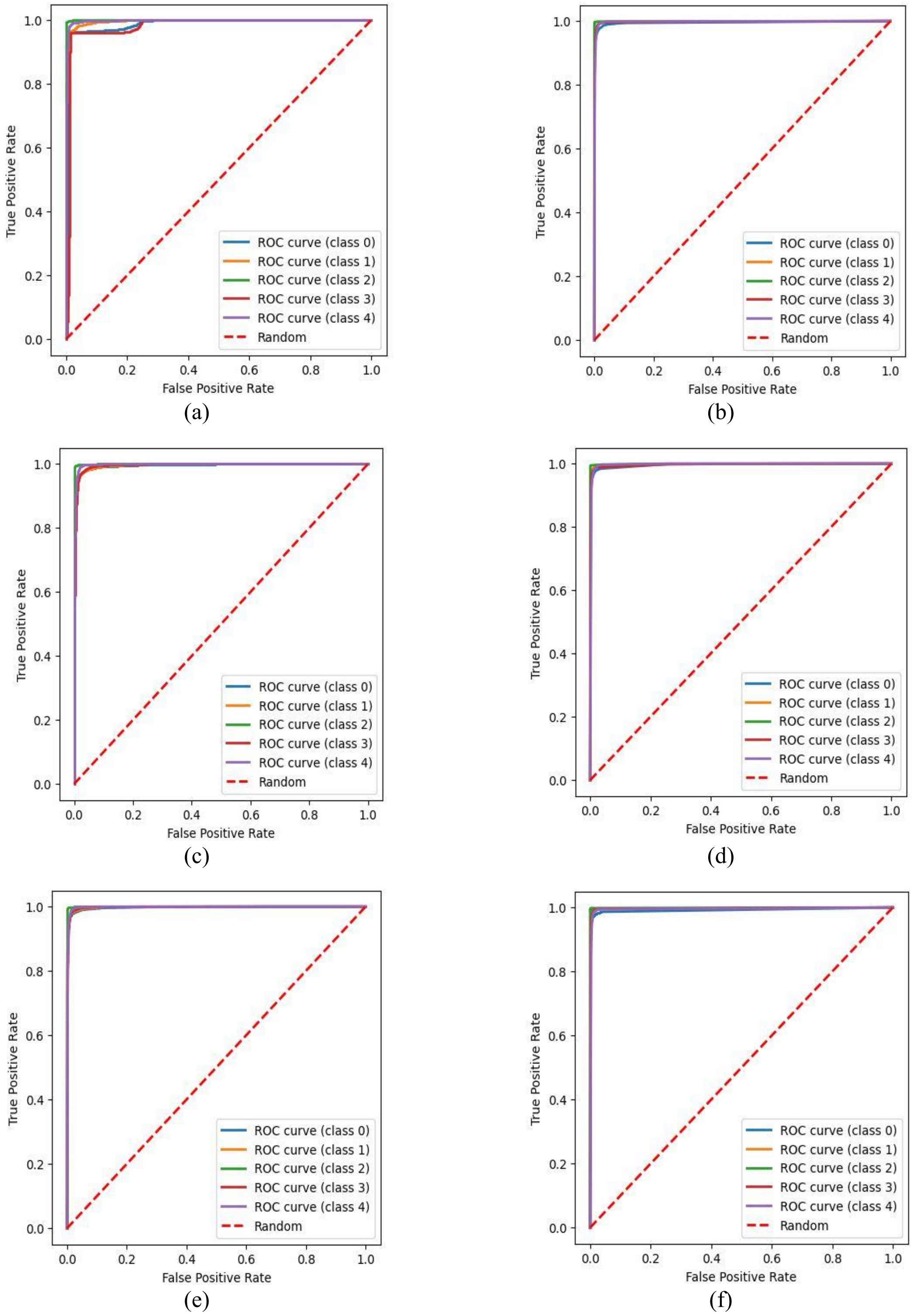


Figure 8. ROC for different classification models (a) SVM, (b) RF, (c) LR, (d) DT, (e) GNB, and (f) K-NN

Table 10. AUC for each class

Models	Class				
	Arborio	Basmati	Ipsala	Jasmine	Karacadag
SVM	0.9863	0.9964	0.9998	0.9851	0.9979
RF	0.9970	0.9980	0.9990	0.9981	0.9978
LR	0.9957	0.9967	0.9997	0.9951	0.9976
DT	0.9954	0.9976	0.9994	0.9967	0.9969
GNB	0.9976	0.9983	0.9999	0.9974	0.9983
K-NN	0.9943	0.9957	0.9985	0.9962	0.9966

3.4. Performance comparison and justification of the proposed model

Table 11 highlights the performance of various methods for rice grain classification and related tasks, demonstrating the strengths and trade-offs of each approach. Our proposed K-NN classifier achieved an impressive accuracy of 97.80%, closely matching advanced models like the modified ResNet50 (97.88%) and CNN-based ensemble models (98%). Despite being a simpler and computationally less intensive algorithm compared to deep learning or ensemble approaches, our K-NN model offers a competitive edge by delivering high accuracy with reduced complexity. This simplicity makes it more accessible and efficient for practical implementation, particularly in scenarios where computational resources are limited. In contrast to methods focusing on specific traits or emphasizing cost-effectiveness, our model strikes an excellent balance between simplicity and performance, proving its effectiveness in rice grain classification tasks.

Table 11. Comparison of proposed K-NN classifier with benchmark methods for rice grain classification

Author(s)	Method	Task	Accuracy	Key remarks
Our method	K-NN classifier	Classify rice grains	97.8%	High accuracy achieved using K-NN
Ahad *et al.* [20]	CNN-based deep learning ensemble model	Detect and localize nine epidemic rice diseases	98%	Ensemble model used for rice disease detection
Tran-Thi-Kim *et al.* [21]	CNN, ANN, Modified VGG16, Modified ResNet50	Classify 17 rice grain varieties	ANN: 92.82%, VGG16: 96.41%, ResNet50: 97.88%	Modified ResNet50 achieved highest accuracy
Patel *et al.* [22]	Image processing	Categorize ten paddy rice varieties	87%	Speed and cost advantages highlighted
Kurade *et al.* [23]	Random forest classifier	Rice quality assessment system	77%	Focus on cost-effective rice quality assessment
Deepika *et al.* [24]	Digital imaging	Evaluate grain quality traits in 21 rice hybrids	91%	Focus on aroma-type rice breeding potential

K-NN emerged as the best-performing algorithm among the six algorithms used in our study. The specific advantages of the K-NN approach, such as its simplicity, effectiveness in handling non-linear data, and potential for easy implementation and interpretation, make it a compelling choice for rice grain classification tasks, particularly in scenarios where complex decision boundaries and interpretability are paramount. This may include investigating alternative feature selection techniques, exploring ensemble methods, or incorporating domain-specific knowledge to improve the overall performance and robustness of the K-NN approach in future research endeavors.

Acknowledging potential biases or limitations associated with the chosen machine learning algorithms is crucial for ensuring a robust analysis. These algorithms may exhibit biases due to their inherent assumptions, parameter settings, or sensitivity to outliers and noise in the dataset. To mitigate these challenges, strategies such as sensitivity analysis, robustness testing, and ensemble methods can be employed. Sensitivity analysis and robust testing enable evaluation under various scenarios and parameter settings, identifying and addressing vulnerabilities. Ensemble methods, including combining multiple classifiers or using model averaging techniques, can help mitigate individual algorithm biases and enhance classification performance. Additionally, techniques like cross-validation and regularization aid in reducing overfitting and improving generalization ability. By proactively addressing these biases and implementing appropriate mitigation strategies, we can ensure a more reliable analysis of the rice grain classification problem.

Our study's findings hold practical implications for precision agriculture, aiding farmers in accurately classifying rice varieties for optimized crop management and resource allocation. The developed intelligent system offers real-time classification support, potentially integrating into existing agricultural technology. Beyond agriculture, these methodologies extend to medical imaging and industrial quality control, showcasing the versatility and broader impact of our research. The adoption of machine learning solutions in smart farming can have positive environmental and economic impacts. These technologies promote sustainable practices by optimizing resource usage and reducing waste through precise application of inputs. Economically, farmers benefit from increased productivity and reduced operational costs, leading to improved profitability. This integration into smart farming practices not only boosts efficiency but also opens avenues for new revenue streams and business opportunities.

4. CONCLUSION AND FUTURE WORK

This study has explored the classification of rice grains utilizing a diverse set of machine learning algorithms, including support vector machine (SVM), random forest (RF), logistic regression (LR), decision tree (DT), Gaussian naive Bayes (GNB), and K-nearest neighbors (K-NN). Following a thorough evaluation K-NN emerged as the top-performing model, demonstrating exceptional precision, recall, F1-score, and accuracy. Leveraging morphological features extracted from segmented rice grain images and effective preprocessing techniques such as grayscale conversion and Otsu's thresholding during segmentation, K-NN achieved an impressive accuracy of 97.80%. The experimental results validate that our K-NN model not only offers a simpler alternative but also outperforms several state-of-the-art approaches, such as ANN and Modified VGG16, in terms of accuracy and efficiency for rice grain classification. This outcome highlights the robustness and potential of K-NN as a reliable tool for rice grain classification, laying a strong foundation for future advancements in quality assessment within the agricultural and food processing industries.

Addressing limitations, our study could benefit from a more diverse dataset to better represent real-world rice varieties. To mitigate biases, future research could explore automated hyperparameter tuning methods for more objective parameter selection. In future research, there is potential to explore the integration of advanced deep learning methods, such as convolutional neural networks (CNNs), to improve model accuracy. Expanding the dataset to encompass a broader range of rice varieties would enhance the depth of the evaluation process. Further refinement could be achieved by incorporating additional morphological and textural features and exploring ensemble methods. Additionally, investigating the real-time implementation of these models for practical applications in rice quality assessment would be a valuable avenue for future exploration.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration. Suma D: Conceptualization (lead); Methodology (lead); Formal analysis (lead); Writing – original draft (lead). Narendra V G: Supervision (lead); Writing – review and editing (equal). Raviraja Holla M: Methodology (supporting); Software development (equal); Writing – original draft (supporting). Darshan Holla M: Software development (lead); Writing – review and editing (equal). All authors have read and approved the final version of the manuscript.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Suma D.	✓	✓			✓				✓					
Narendra V. G.										✓		✓		
Raviraja Holla M.		✓	✓						✓					
Darshan Holla M.			✓							✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, NVG, upon reasonable request.

REFERENCES

[1] B. Arora, N. Bhagat, L. Saritha, and S. Arcot, "Rice grain classification using image processing machine learning techniques," in *Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020*, Feb. 2020, pp. 205–208. doi: 10.1109/ICICT48043.2020.9112418.

[2] N. Nagoda and L. Ranathunga, "Rice sample segmentation and classification using image processing and support vector machine," in *2018 13th International Conference on Industrial and Information Systems, ICIIS 2018 - Proceedings*, Dec. 2018, pp. 179–184. doi: 10.1109/ICIINFS.2018.8721312.

[3] K. Kiratiratanapruk *et al.*, "Development of paddy rice seed classification process using machine learning techniques for automatic grading machine," *Journal of Sensors*, vol. 2020, pp. 1–14, Jul. 2020, doi: 10.1155/2020/7041310.

[4] K. Aukkapinyo, S. Sawangwong, P. Pooyoi, and W. Kusakunniran, "Localization and classification of rice-grain images using region proposals-based convolutional neural network," *International Journal of Automation and Computing*, vol. 17, no. 2, pp. 233–246, Dec. 2020, doi: 10.1007/s11633-019-1207-6.

[5] C. V. Krishna, B. Suchitra, L. Sujihelen, M. S. Roobini, S. Cherukullapurath, and A. Jesudoss, "Quality analysis of rice grains using morphological techniques," in *2022 International Conference on Communication, Computing and Internet of Things, IC3IoT 2022 - Proceedings*, Mar. 2022, pp. 1–5. doi: 10.1109/IC3IOT53935.2022.9767925.

[6] S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, and M. R. M. Noordin, "Rice grain classification using multi-class support vector machine (SVM)," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 3, pp. 215–220, Dec. 2019, doi: 10.11591/ijai.v8.i3.pp215-220.

[7] R. Ruslan, S. Khairunniza-Bejo, M. Jahari, and M. F. Ibrahim, "Weedy rice classification using image processing and a machine learning approach," *Agriculture (Switzerland)*, vol. 12, no. 5, p. 645, Apr. 2022, doi: 10.3390/agriculture12050645.

[8] K. R. Singh and S. Chaudhury, "A cascade network for the classification of rice grain based on single rice kernel," *Complex and Intelligent Systems*, vol. 6, no. 2, pp. 321–334, Feb. 2020, doi: 10.1007/s40747-020-00132-9.

[9] I. Cinar and M. Koklu, "Identification of rice varieties using machine learning algorithms," *Tarim Bilimleri Dergisi*, vol. 28, no. 2, pp. 307–325, May 2022, doi: 10.15832/ankutbd.862482.

[10] F. Farahnakian, J. Sheikh, F. Farahnakian, and J. Heikkonen, "A comparative study of state-of-the-art deep learning architectures for rice grain classification," *Journal of Agriculture and Food Research*, vol. 15, p. 100890, Mar. 2024, doi: 10.1016/j.jafr.2023.100890.

[11] T. T. K. Nga, T. V. Pham, D. M. Tam, I. Koo, V. Y. Mariano, and T. Do-Hong, "Combining binary particle swarm optimization with support vector machine for enhancing rice varieties classification accuracy," *IEEE Access*, vol. 9, pp. 66062–66078, 2021, doi: 10.1109/ACCESS.2021.3076130.

[12] T. Y. Kuo, C. L. Chung, S. Y. Chen, H. A. Lin, and Y. F. Kuo, "Identifying rice grains using image analysis and sparse-representation-based classification," *Computers and Electronics in Agriculture*, vol. 127, pp. 716–725, Sep. 2016, doi: 10.1016/j.compag.2016.07.020.

[13] L. de Oliveira Carneiro *et al.*, "Characterizing and predicting the quality of milled rice grains using machine learning models," *AgriEngineering*, vol. 5, no. 3, pp. 1196–1215, Jul. 2023, doi: 10.3390/agriengineering5030076.

[14] S. B. Ahmed, S. F. Ali, and A. Z. Khan, "On the frontiers of rice grain analysis, classification and quality grading: A review," *IEEE Access*, vol. 9, pp. 160779–160796, 2021, doi: 10.1109/ACCESS.2021.3130472.

[15] W. Srimulyani and A. Musdholifah, "Identification of rice variety using geometric features and neural network," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 3, p. 301, Jul. 2019, doi: 10.22146/ijccs.48203.

[16] K. R. Singh and S. Chaudhury, "Efficient technique for rice grain classification using back-propagation neural network and wavelet decomposition," *IET Computer Vision*, vol. 10, no. 8, pp. 780–787, Jul. 2016, doi: 10.1049/iet-cvi.2015.0486.

[17] A. Aznan, C. G. Viejo, A. Pang, and S. Fuentes, "Computer vision and machine learning analysis of commercial rice grains: A potential digital approach for consumer perception studies," *Sensors*, vol. 21, no. 19, p. 6354, Sep. 2021, doi: 10.3390/s21196354.

[18] A. S. Hamzah and A. Mohamed, "Classification of white rice grain quality using ann: A review," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4, pp. 600–608, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp600-608.

[19] S. Sultana, M. Faruque, and M. R. Islam, "Rice grain quality parameters and determination tools: a review on the current developments and future prospects," *International Journal of Food Properties*, vol. 25, no. 1, pp. 1063–1078, May 2022, doi: 10.1080/10942912.2022.2071295.

[20] M. T. Ahad, Y. Li, B. Song, and T. Bhuiyan, "Comparison of CNN-based deep learning architectures for rice diseases classification," *Artificial Intelligence in Agriculture*, vol. 9, pp. 22–35, 2023, doi: 10.1016/j.aiia.2023.07.001.

[21] N. Tran-Thi-Kim, T. Pham-Viet, I. Koo, V. Mariano, and T. Do-Hong, "Enhancing the classification accuracy of rice varieties by using convolutional neural networks," *International Journal of Electrical and Electronic Engineering and Telecommunications*, vol. 12, no. 2, pp. 150–160, 2023, doi: 10.18178/ijeetc.12.2.150-160.

[22] B. Patel and A. Sharaff, "Rice variety classification & yield prediction using semantic segmentation of agro-morphological characteristics," *Multimedia Tools and Applications*, vol. 82, no. 29, pp. 45567–45584, May 2023, doi: 10.1007/s11042-023-15549-w.

[23] C. Kurade *et al.*, "An automated image processing module for quality evaluation of milled rice," *Foods*, vol. 12, no. 6, p. 1273, Mar. 2023, doi: 10.3390/foods12061273.

[24] D. C, G. R P, T. K, R. N, and K. A, "Linear discriminant analysis of grain quality traits in rice (Oryza sativa L.) using the digital imaging technique," *Journal of Cereal Science*, vol. 109, p. 103609, Jan. 2023, doi: 10.1016/j.jcs.2022.103609.

[25] M. Koklu, "Rice image dataset," *Kaggle*, 2022. https://www.kaggle.com/datasets/muratkokludataset/rice-image-dataset (accessed: Apr. 15, 2024).

BIOGRAPHIES OF AUTHORS

Suma D. is working as an assistant professor in the Department of Computer Science and Engineering, Manipal Institute of Technology (a constituent institution of Manipal Academy of Higher Education), Manipal. She completed B.E.(ECE) from Kuvempu University, India and M. Tech. (CSE) from Visvesvaraya Technological University, India. Her areas of interest include image processing, high-performance computing, and machine learning. She can be contacted at email: suma.d@manipal.edu.

Narendra V. G. received the B.E. degree in CSE from Karnataka University Dharwad, in 1997, the M. Tech. degree from Visvesvaraya Technological University (VTU), Belgaum, in 2001, and the Ph.D. degree from Manipal University, Manipal, India, in 2017. He is currently a Professor with the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. His research interests include image processing, agricultural image processing, machine learning, soft computing, and computer vision-based applications for agriculture. He is on the editorial board of some journals. He can be contacted at email: narendra.vg@manipal.edu.

Raviraja Holla M. is working as an assistant professor in the Department of Information and Communication Technology Department, Manipal Institute of Technology (a constituent institution of Manipal Academy of Higher Education, Manipal. He completed B.E. (CSE) from Bangalore University, India and M. Tech. (CSE) from KSOU Mysore, India. He pursued his Ph.D. from the Department of Computer Engineering, National Institute of Technology Karnataka (NITK), Surathkal. His areas of interest include information security, high-performance computing, and semantic web. He can be contacted at email: raviraj.holla@manipal.edu.

Darshan Holla M. is currently pursuing the bachelor's degree in computer and communication engineering with the Information and Communication Technology Department, Manipal Institute of Technology. He is deeply interested in the field of data mining and machine learning and their applications in the healthcare domain. He aspires to contribute to impactful research in this domain and use technology to make significant improvements to the existing domain. He can be contacted at email: darshan.m1@learner.manipal.edu.