



Original Research Paper

CLASSIFICATION OF RICE GRAIN VARIETIES WITH ENHANCED WILKS BASED PCA AND ENSEMBLE BASED LEARNING

Dr. S.V. Anandhi¹ and Dr. P. Chanthiya²

¹Associate Professor, Department of Artificial Intelligence and Data Science RAMCO Institute of Technology, Rajapalayam

²Associate Professor, Department of Computer Science and Engineering, V.V.College of Engineering, Tisaiyanvilai
svanandhi2020@gmail.com¹ and chanthiyapuhai@gmail.com²

ABSTRACT

One of the most extensively produced grain crops in the world, rice, has several genetic variants. These categories differ from each other due to certain characteristics. These characteristics typically include color, texture, and shape. These characteristics that set distinct varieties of rice apart allow for the classification and evaluation of seed quality. The five varieties of rice that are most frequently grown worldwide are Basmati, Jasmine, Ipsala, Doongara, Wehani, Arborio, and Karacadag. The collection has 1,75,000 grain samples, with 25,000 of each category. A second set of 121 attributes was used, comprising 17 morphological features, 9 form, and 95 color features extracted from these images. The rice dataset was classified using an ensemble learning technique and a dimensionality reduction framework. This technique lowers the misclassification rate by aggregating the predictions of multiple classifiers. To ascertain the recall, prediction, and accuracy, statistical computations were performed. The classification accuracy of this model was 100% with XGBoost, 99.63% with Adaboost, and 99.85% using gradient boosting. The findings demonstrate the applicability of the models employed in the rice variety classification study.

INTRODUCTION

Indeed, by offering innovative methods for classifying and assessing the quality of rice grains, computer vision and machine learning technologies are transforming the agricultural industry. By using these technology, the agriculture industry benefits from a more automated, precise, and effective approach than traditional manual operations. By automating the identification of crucial quality criteria including shape, size, texture, and fracture rate, these technologies help to make quality control faster, more dependable, and more cost-effective. Recent research has further demonstrated the applicability of computer vision and deep learning for automated rice analysis, with machine-learning models being evaluated using accuracy, recall, and precision for rice-related classification tasks (Tholupuluri et al., 2024). The broader adoption of artificial intelligence in agriculture has also expanded the use of data-driven methods for crop management, forecasting, and agricultural decision-making (Singh & Kaur, 2025).

A food derived from grain crops, rice is grown in many nations and consumed globally. According to Mahajan et al. [2], computer vision and image processing applications in agriculture are interesting since they are less expensive than manual approaches and have the

ability to be evaluated without causing damage. According to Barbedo et al. [7], image processing-based computer vision applications outperform conventional manual labor-based techniques. Because human error is a major factor, manually evaluating or classifying grains can be expensive and time-consuming. Because manual procedures only rely on the evaluation specialists' experience, Patrício et al. [4] suggested that they might have a different evaluation process. Rice is a grain crop-based food that is grown in many countries and consumed all over the world.

Using a two-class dataset of 1700 rice data, Sun et al. used the SVM technique to obtain 98.5% classification success in a study that was published in the literature. When 200 data points from 16 classes were analyzed in a different investigation, the SVM technique produced an accuracy of 87.16% (Liu et al., 2015). Lin et al. introduced the deep CNN technique, which achieved a 95.5% success rate in a sample consisting of three classes and 7399 data points. Ahmed et al. used CNN for classification operations following feature extraction in another study including 200 data points and three types of rice. The researchers had an 88.07% success rate.

*Corresponding Author:

Received: 25th May, 2026; Revised: 6th July, 2026; Accepted: 28th July, 2026; Available Online: 29th August, 2026
(DOI): 10.70102/AEJ.2026.18.2.29

LITERATURE SURVEY

One commodity based on grain crops that is produced in many nations and consumed globally is rice. The market uses a number of factors to influence rice prices. These components include color, texture, form, and fracture rate, according to Aukkapinyo et al. [6]. When digital product photos are obtained, a variety of machine learning techniques are employed to identify these characteristics and carry out categorization tasks.

Large data sets may be analyzed quickly and reliably thanks to machine learning algorithms. In order to comply with food safety requirements and enhance the quality of the finished product in an automated, cost-effective, efficient, and nondestructive way, such methods are crucial for producing rice (Grinberg et al.) [19]. In recent years, rice quality and classification have been assessed using a variety of digital image attributes. These include the capacity to detect cracks in rice grains, fracture rate, whiteness, and geometric measurements like length and perimeter. Systems based on image processing can be used to extract a variety of grain product properties.

Recent agricultural research has also demonstrated the use of machine-learning and ensemble frameworks for crop-related prediction, including rice, with models such as Random Forest, Gradient Boosting, and deep learning being evaluated using quantitative performance measures (Mohammad et al., 2026).

Additionally, it has been mentioned that these traits are classified using CNN (Ahmed et al. 2020) [8] and other machine learning approaches. In a study reported in the literature, Sun et al. employed the SVM technique and achieved 98.5% classification success using a two-class dataset of 1700 rice data points. 200 data points from 16 classes were examined in a different study by Liu et al.

[5], and the SVM method produced an accuracy of 87.16%. Lin et al. suggested the deep CNN technique on a sample consisting of three classes and 7399 data points. The success rate of the paper was 95.5%.

This project's objective is to use photos of several rice kinds to create a non-destructive model that improves categorization accuracy. In the suggested models, categorization was done using 106 morphological and color features that were taken from rice photos and fed into ANN and DNN. Additionally, the classification process was conducted using 75,000 rice photographs from 5 different classes as input for the CNN technique, which can categorize raw photos without the need for pre-processing. The results of CNN, DNN, and ANN categorization were then compared [1].

Proposed System

The foundation for categorizing the various varieties of rice is shown in Figure 1. There are four stages in this proposed concept. First, 170 attributes were collected from various data sources. The collected datasets need to be processed before being added to the operational model. The normalization procedure was used to organize, enhance, and simplify the dataset's scattered data. It is believed that normalization affects how well tasks involving classification and prediction perform [12–13]. Deep learning methods can also make use of this process [14].

In computer vision applications for product classification, normalization is also commonly used on photographs [18]. Machine learning algorithms employ normalization to lessen the impact of disparate attribute ranges across records. The min-max normalization approach was used to scale the dataset values in this study from 0 to 1, with 0 being the minimum and 1 representing the maximum.

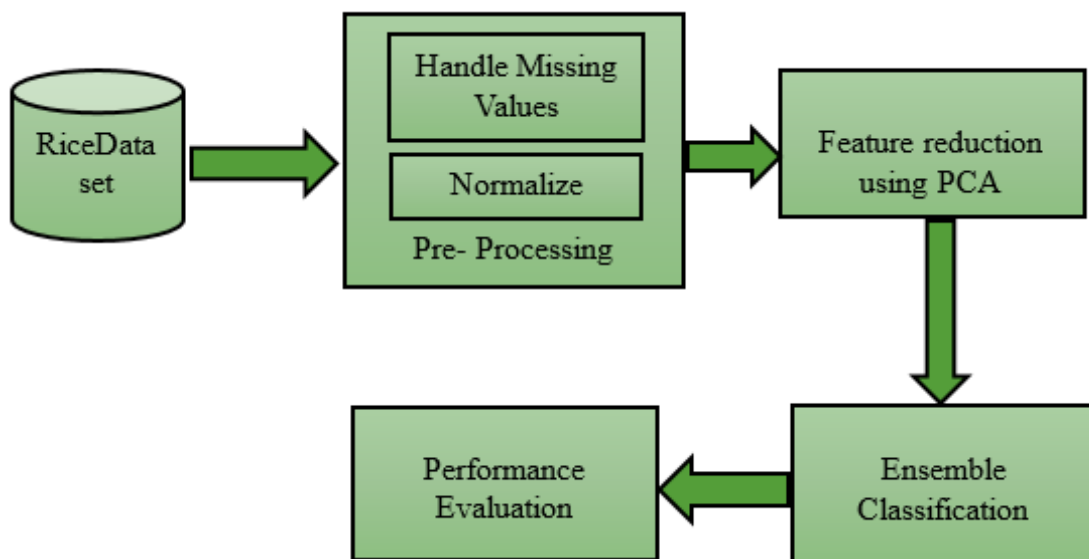


Figure 1. Proposed framework

Since there were no negative attribute values in the dataset, the Z-score approach was not used. The following describes the normalization process calculation method:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Here, x stands for the original data, y for the normalized data, x_{max} for the dataset's highest value, and x_{min} for its lowest value. In this study, parameters such as area, perimeter, major axis length, minor axis length, and convex area underwent normalization. However, as the

eccentricity and extent attributes' values were already inside the 0–1 range in the raw data, no normalization was done. Table 1 contains a selection of records from the raw dataset.

Table 1 Raw dataset of rice shown at random

Area	Perimeter	Major_Axis	Minor_Axis	Eccentricity	Eqdiasq	Solidity	Convex_Area	Extent	Compactness	Class
7805	437.915	209.8215	48.0221	0.9735	99.6877	0.9775	7985	0.3547	0.4751	Basmati
7503	340.757	138.3361	69.8417	0.8632	97.74	0.966	7767	0.6637	0.7065	Arborio
5124	314.617	141.9803	46.5784	0.9447	80.7718	0.9721	5271	0.476	0.5689	Jasmine
7990	437.085	201.4386	51.2245	0.9671	100.8622	0.9659	8272	0.6274	0.5007	Basmati
7433	342.893	140.335	68.3927	0.8732	97.283	0.9831	7561	0.6006	0.6932	Arborio
11648	445.527	178.4659	84.9327	0.8795	121.7813	0.9599	12135	0.566	0.6824	Ipsala
7621	450.325	219.0981	45.2301	0.9785	98.5056	0.9718	7842	0.3398	0.4496	Basmati
8582	367.338	146.6128	75.5406	0.857	104.532	0.974	8811	0.6423	0.713	Arborio
5450	320.362	139.9963	50.691	0.9321	83.3016	0.9626	5662	0.5502	0.595	Jasmine
6781	307.023	116.2443	74.8093	0.7654	92.9184	0.9819	6906	0.7208	0.7993	Karacdag
9656	451.287	206.7047	60.2333	0.9566	110.8801	0.9748	9906	0.5692	0.5364	Jasmine
8460	362.862	149.4804	72.9154	0.873	103.7864	0.9822	8613	0.7673	0.6943	Arborio
7090	401.778	190.9745	47.8482	0.9681	95.0119	0.9789	7243	0.4039	0.4975	Basmati
6208	300.124	113.9325	70.6069	0.7848	88.906	0.9658	6428	0.763	0.7803	Karacdag

A preview of the raw dataset is given in Table 1, which highlights its complexity and diversity. Principal Component Analysis (PCA) can be used in the feature reduction stage, which uses the pre-processed dataset as input. Multiple connected quantitative factors characterize each observation in multivariate data, which can be analyzed using the PCA approach. Extracting important information from the dataset and presenting it as a collection of new, uncorrelated variables known as principal components is the main objective of PCA. These elements reduce the complexity of the data. Furthermore, by charting data and variables to highlight similarities or differences, PCA [13] makes it possible to visualize patterns and correlations among them.

Finding the most important principal components in Principal Component Analysis (PCA) is made possible by a new technique that combines PCA with the Wilks' Λ -statistic. This method uses the Wilks' Λ -statistic to increase the categorization accuracy of different kinds of rough rice. The following summarizes the mathematical underpinnings of this approach, which were put forth by Yin [14] and Zhou [15–16]: Let m be the number of selected principal components, D be the sum of squares and deviations matrix inside a class, and A be the sum of squares and deviations matrix for all classes. The following are the formulas for D and A:

$$D = (d_{ij})_{m \times m} \quad (2)$$

$$A = (a_{ij})_{m \times m} \quad (3)$$

$$d_{ij} = \sum_{g=1}^c \sum_{k=1}^{N_g} (X_{igk} - u_{ig}) * (X_{jgk} - u_{jg}) \quad (4)$$

$$a_{ij} = \sum_{g=1}^c \sum_{k=1}^{N_g} (X_{igk} - u_i) * (X_{jgk} - u_j) \quad (5)$$

N_g is the sample size for the g th class (where $N_1=N_2=N_3=N_4=N_5=20$), u_{ig} is the average of the i th principal component (PC) for the g th class, and u_i is the average of the i th PC across all classes. In this case, c is the number of classes ($c=5$), which correspond to five

$$\Lambda = \frac{|D|}{|A|} \quad (6)$$

The pertinent characteristics are then fed into the ensemble-based categorization when feature reduction is complete. Using a single classifier can raise the danger of overfitting in classification tasks involving huge datasets. It has been discovered that integrating several classifiers into an ensemble performs better than depending just on one "best" model in order to solve this problem. However, choosing the best classifier models for a given dataset and job remains a substantial difficulty due to the abundance of accessible models.

The two phases of the ensemble model are classifier combination and classifier selection. KNN, MLP,

different types of rough rice. A smaller value of $|D|$ and a bigger value of $|A|$ suggest a greater differentiation between the classes, which aids in classifying them, in accordance with the Wilks' distribution principle. The definition of the Wilks' Lambda-statistic, Λ , is:

Decision Tree, and Random Forest were used in the classifier selection step, and their outputs were blended using weighted voting to improve the outcomes. The results of the experiment show how well the suggested ensemble approach for rice variety data classification works. With an accuracy of 88.52%, the ensemble classifier [20] outperformed the individual classifiers. This demonstrates that a more accurate and dependable model is produced by combining the predictions from several algorithms. Additionally, the study offers insightful information for categorizing medical data, especially for aiding medical practitioners in the diagnosis of heart conditions.

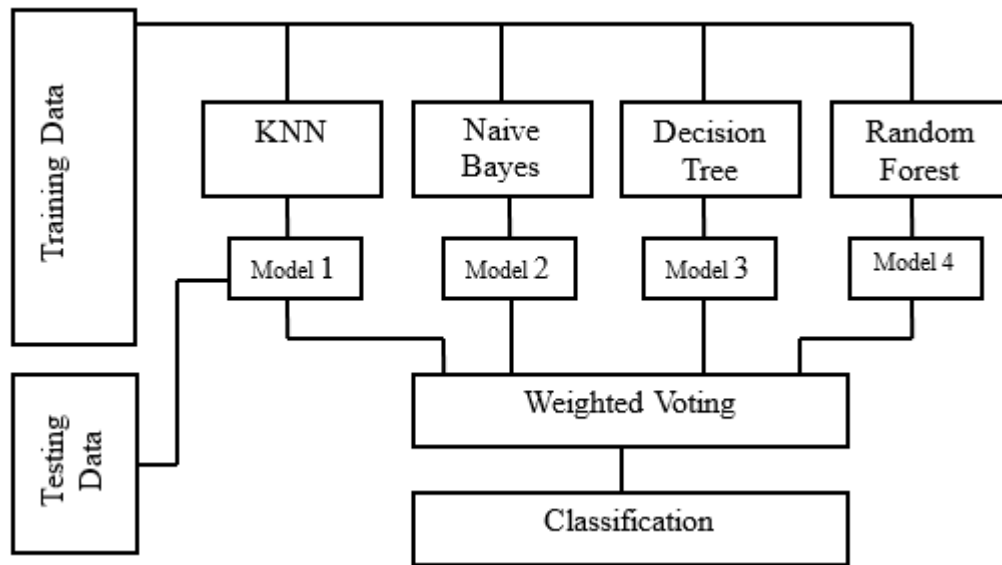


Figure2. Ensemble based classification

Early and accurate identification of at-risk patients can be facilitated by the ensemble classifier, allowing for prompt intervention and the creation of individualized treatment regimens. Our study concludes by demonstrating the potential of ensemble machine learning methods for cardiac patient data classification. Compared to individual classifiers, the ensemble

approach—which consists of KNN, MLP, Decision Tree, and Random Forest—performs better, producing a classification model that is more reliable and accurate.

A technique called weighted voting gives classifiers that do well more sway over the voting process. This method can outperform both the majority voting method and the

top individual classifier when the weights are assigned correctly. Finding the ideal weights is the main obstacle. Thus, precisely identifying these optimal weights is the main challenge [15]. Six popular methods for weighted voting are described here.

N stands for the total number of data points, K for the number of categories, M for the total number of classifiers, w_m for each classifier's weight, e_m for each classifier's errors, e_B and e_W for minimum and maximum errors, a_m for each classifier's accuracy, and a_m for each classifier's performance in a classification problem. These variables have the following relationships:

$$\text{Accuracy } (a_m) = (N - e_m) / N \quad (7)$$

Performance (a_m) is dependent on weight, accuracy, and Errors and best/worst-case situations determine weight (w_m). By minimizing errors, increasing accuracy, and optimizing weight assignment, the objective is to maximize classification performance and accuracy. Machine learning algorithms including ensemble methods, hyperparameter tuning, and classifier selection can be used to do this.

Weighted majority vote (WMV), best-worst weighted vote (BWV), quadratic best-worst weighted vote (QBWV), simple weighted vote (SWV), re-scaled weighted vote (RSWV), and Zheng and Padmanabhan's

weighted vote [17]. Weighted Majority Voting (WMV) combines optimized weights with classifier predictions to improve ensemble learning. This increases accuracy overall. Weight optimization is crucial and can be accomplished by cross-validation, gradient descent, and metaheuristic algorithms (such as EFDM and HHO).

WMV is a well-liked machine learning technique that improves performance over individual classifiers by fine-tuning weights. In this case, weighted voting is done using the weighted majority vote (WMV). Optimization of weight assignments improves integration accuracy:

$$w_m = \log \frac{a_m}{1-a_m} \quad (8)$$

Training and test sets of the dataset are separated. After testing and training, each base classifier produces a confusion matrix that evaluates its effectiveness. Each classifier's aggressive cross-efficiency ratings are determined using these matrices. The final classification results are then determined by using these scores in a weighted voting method, where predictions are weighted based on their respective cross-efficiency values.

RESULTS AND DISCUSSIONS

This study makes use of a variety of datasets from well-known data science communities and platforms such as Kaggle. The 170 features in the collected information are used to categorize various types of rice.

Table 2 Normalized values from a dataset of rice shown at random

Area	Perimeter	Major_Axis	Minor_Axis	Eccentricity	Eqdiasq	Solidity	Convex_Area	Extent	Compactness	Class
12488	458.579	191.061	84.4833	0.8969	126.0961	0.9746	12813	0.5513	0.66	Ipsala
7142	425.854	203.722	45.581	0.9746	95.3597	0.97	7363	0.5552	0.4681	Basmati
6524	366.419	168.0679	50.4087	0.954	91.1406	0.9705	6722	0.4953	0.5423	Basmati
5184	312.273	139.8793	48.0543	0.9391	81.2433	0.9752	5316	0.5906	0.5808	Jasmine
5759	285.836	110.6204	67.2578	0.7939	85.6305	0.9799	5877	0.6962	0.7741	Karacadag
12843	466.299	197.1622	83.9702	0.9048	127.8758	0.9755	13166	0.6398	0.6486	Ipsala
5789	287.704	111.4038	67.2571	0.7972	85.8533	0.9772	5924	0.6897	0.7706	Karacadag
8894	372.939	149.1019	76.9656	0.8565	106.4152	0.9627	9239	0.6285	0.7137	Arborio
6876	303.202	115.4528	76.1591	0.7516	93.5671	0.9855	6977	0.7487	0.8104	Karacadag
8055	432.913	203.8687	51.3218	0.9678	101.2716	0.974	8270	0.3728	0.4967	Basmati
10015	438.394	194.2564	66.9318	0.9388	112.9225	0.9773	10248	0.4766	0.5813	Jasmine
7248	323.896	127.5157	73.3569	0.818	96.0648	0.9852	7357	0.6639	0.7534	Karacadag
6049	338.084	149.7718	52.332	0.937	87.76	0.9688	6244	0.7143	0.586	Jasmine
7552	339.933	137.3724	70.8414	0.8568	98.0587	0.9722	7768	0.6246	0.7138	Arborio
15532	506.083	207.2012	97.0122	0.8836	140.627	0.9692	16026	0.5726	0.6787	Ipsala
6571	305.263	120.506	69.8124	0.8151	91.4683	0.9856	6667	0.6836	0.759	Karacadag
5198	325.531	145.7141	46.7827	0.9471	81.3529	0.9653	5385	0.5172	0.5583	Jasmine
7036	314.879	117.2446	77.6209	0.7495	94.6494	0.9742	7222	0.758	0.8073	Karacadag
5276	317.64	142.4069	47.9502	0.9416	81.961	0.9697	5441	0.6407	0.5755	Jasmine
8485	360.317	145.7381	74.544	0.8593	103.9396	0.977	8685	0.6559	0.7132	Arborio
6245	291.946	109.9091	72.9431	0.748	89.1705	0.9818	6361	0.7343	0.8113	Karacadag

5726	273.63	102.7642	71.2113	0.721	85.3848	0.9912	5777	0.735	0.8309	Karacadag
14446	488.048	203.3706	91.7769	0.8924	135.6216	0.9761	14800	0.6465	0.6669	Ipsala

The pre-processed dataset is randomly displayed in Table 2. At first Gather and prepare the rice dataset, which may contain a variety of characteristics such as grain width, length, texture, color, and others. There may be a strong correlation between these qualities. Standardizing the data is crucial because PCA is sensitive to the feature scales. In this stage, the data is centered by scaling it to unit variance and removing the mean. Determine the

dataset's covariance matrix. The links between the various traits and their collective variation are captured in this matrix. This study utilizes various datasets sourced from platforms like Kaggle, a renowned data science community. The compiled dataset comprises 170 features, which are employed to classify different rice varieties.

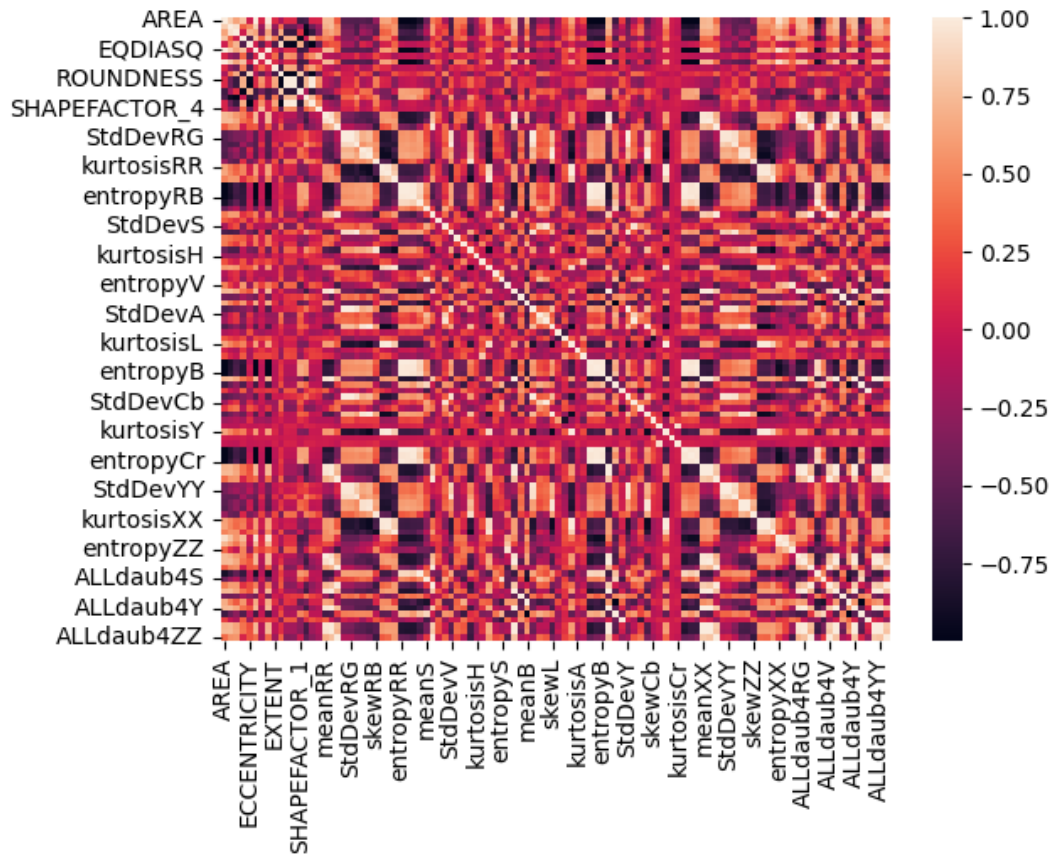


Figure.2 Covariance Matrix

The Covariance Matrix, which shows the covariance between every pair of variables in the multivariate data, is shown in Figure 2. This matrix provides information on how data points are dispersed throughout two dimensions by displaying the size and direction of the data distribution in a multidimensional space. Determine the covariance matrix's eigenvalues and eigenvectors.

While the eigen values display the level of variation along each eigen vector, the eigenvectors indicate the direction of maximal variance. Choose the top eigenvectors that correspond to the biggest eigenvalues after sorting the eigenvalues in descending order. The major components that explain the greatest amount of the data's variance are formed by these eigenvectors.

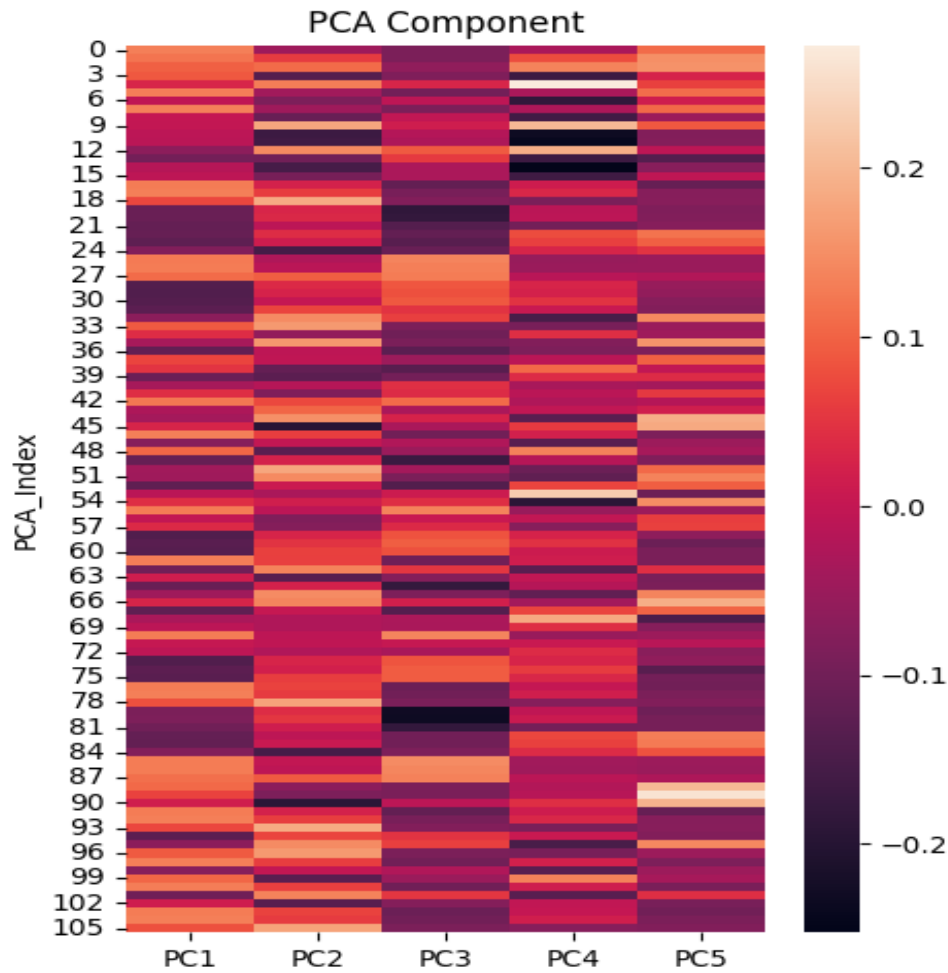


Figure.3 PCA based feature reduction

The PCA-based feature reduction is displayed in Figure 3. In a rice dataset, feature reduction based on Principal Component Analysis (PCA) entails breaking down the initial set of features (variables) into a more manageable set of uncorrelated characteristics known as principal components. These elements reduce dimensionality while capturing the majority of the data's variation. When the objective is to distinguish between various rice types or attributes, this method works especially well for datasets with a lot of features, as those pertaining to rice categorization.

For classification jobs, use the smaller dataset (which has fewer dimensions). By eliminating noise and unnecessary information, the PCA-reduced data can increase the effectiveness and precision of machine learning models. By streamlining the data without sacrificing important information, PCA-based feature reduction in rice datasets might enhance classifier performance by facilitating faster model training and better generalization.

Three important metrics—recall, precision, and accuracy—are defined by equations (9)–(11) and are used to assess how well the suggested human classification system works.

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \quad (9)$$

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \quad (10)$$

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{FalsePositive} + \text{FalseNegative} + \text{TrueNegative}} \quad (11)$$

Table 3 Performance measures of various rice classification algorithms

Dataset	Proposed Work			Existing Work 1			Existing Work 2		
	Recall	Precision	Accuracy	Recall	Precision	Accuracy	Recall	Precision	Accuracy
T1	99.8	99.25	99.02	98.05	98.30	98.85	96.26	96.37	96.05
T2	99.22	99.13	99.90	98.13	98.05	98.47	96.77	96.62	96.53
T3	99.75	99.17	99.97	98.72	98.77	98.04	96.11	96.45	96.47
T4	99.82	99.02	99.91	98.51	98.84	98.42	96.16	96.53	96.77
T5	99.93	99.67	99.99	98.81	98.33	98.59	96.45	96.71	96.16
T6	99.01	99.20	99.98	98.89	98.15	98.51	96.57	96.65	96.15
T7	99.84	99.39	99.94	99.01	98.29	98.42	96.47	96.27	96.87
T8	99.87	99.66	99.92	98.24	98.44	98.05	96	96.44	96.37
T9	99.22	99.53	99.95	98.15	98.26	98.06	96.74	96.56	96.45
T10	99.33	99.79	99.97	98.04	98.63	98.65	96.66	96.13	96.11
T11	99.66	99.17	99.95	98.41	98.06	98.37	96	96.47	96.15
T12	99.8	99.03	99.93	98.07	98.89	98.2	96.26	96.23	96.24
T13	99.78	99.19	99.95	98.94	98.05	98.57	96.42	96.51	96.06
T14	99.89	99.11	99.92	98.8	98.99	98.15	96.26	96.69	96.45
T15	99.6	99.04	99.96	98.44	98.9	98.51	96.67	96.52	96.51
T16	99.14	99.71	99.93	98.05	98.62	98.16	96.56	96.14	96.37
T17	99.52	99.2	99.94	98.47	98.14	98.31	96.22	96.81	96.64
T18	99.76	99.41	99.97	98.03	98.35	98.99	96.15	96.66	96.37
T19	99.06	99.65	99.97	98.06	98.47	98.4	96.24	96.87	96.67
T20	99.91	99.72	99.92	98.87	98.47	98.77	96.17	96.71	96.91
Overall	99.60	99.35	99.90	98.43	98.45	98.42	96.36	96.52	96.42

With average recall, precision, and accuracy rates of 99.60%, 99.35%, and 99.90%, respectively, across 20 different datasets, the suggested rice categorization system performs exceptionally well (Table 3).

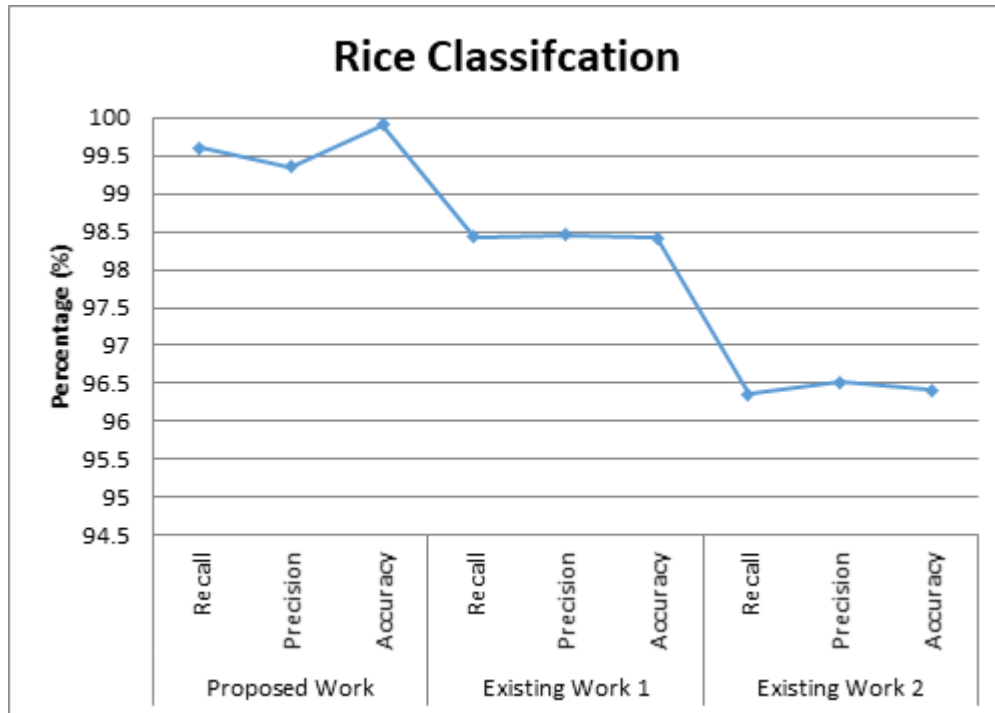
**Figure. 4** Evaluation metrics for rice classification algorithms

Table 3 compares various rice categorization algorithms, and Figure 4 shows the techniques' performance characteristics.



Figure. 5 shows the misclassification error

The misclassification error for the suggested rice classification methods is displayed in Figure 5. This approach is more effective because it makes use of feature-based classification. Important performance indicators consist of Robust class identification across five different classes is ensured by the F1 score, which is a balanced assessment of false positives and false negatives, recall, precision, and the capacity to detect all positive cases.

CONCLUSION

To categorize the various types of rice grains, this study recommends using ensemble-based learning and PCA with Wilks' Λ -statistic technique. The created ensemble based learning approach yields a lower false rate and better accuracy compared to other current algorithms. According to the results, the accuracy average across 20 datasets is 99.90%.

REFERENCES

1. A. R. Singh and A. Kaur, "AI adoption in agriculture: A managerial framework for sustainable agribusiness development," *Journal of Artificial Intelligence in Agriculture and Applied Research*, vol. 1, no. 1, pp. 1–12, 2025. <https://agrpjournals.com/journals/journal/index.php/jaiaar/article/view/2>
2. Abdi, H.; Williams, L.J., "Principal Component Analysis", *WIREs. Computational Statistics*, 2010, pp.433–459.
3. Ahmed, T. Rahman, C. R. Abid, M., Mahmud, F., "Rice grain disease identification using dual phase convolutional neural network-based system aimed at small dataset", 2020.
4. Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K., Taha, K., "Efficient machine learning for big data: a review", *Big Data Research*, 2015, Vol.2, Issue 3, pp.87–93.
5. Aukkapinyo, K., Sawangwong, S., Pooyoi, P., Kusakunniran, W., "Localization and classification of rice-grain images using region proposals-based convolutional neural network", *International Journal of Automation and Computing*, 2019, pp.1–14.
6. Barbedo, J.G.A., "A review on the main challenges in automatic plant disease identification based on visible range images", *Biosystems Engineering*, 2016, Vol. 144, pp. 52–60.
7. Barbedo, J.G.A., "A review on the main challenges in automatic plant disease identification based on visible range images". *Biosystems Engineering*, 2016, Vol. 144, pp.52–60.
8. G. B. Tholupuluri, B. Charith, and U. M. Gopal Krishna, "Deep learning algorithms for rice leaf disease detection," *Bulletin of Pure and Applied Sciences-Zoology*, vol. 43B, no. 2S, pp. 331–343, 2024. <https://bpasjournals.com/zoology/index.php/journal/article/view/243>
9. Grinberg, N.F., Orhobor, O.I., King, R.D., "An evaluation of machine-learning for predicting

- phenotype: studies in yeast, rice, and wheat”, *Machine Learning*, 2020, Vol. 109, Issue 2, pp.251–277.
10. 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,” *Journal of Thermal Analysis and Calorimetry*, 2023, Vol. 148, pp. 3155–3177.
11. Liu,T., Chen, W., Wu, W., Sun, C., Guo, W., Zhu, X., “Detection of aphids in wheat fields using a computer vision technique”, *Biosystems Engineering*, 2016, Vol. 141, pp.82–93.
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, 2024, pp. 245–259.
13. 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 Systems with Applications*, 2023, Vol. 235, p. 121-134.
14. Mahajan, S., Das, A., Sardana, H.K.,”Image acquisition techniques for assessment of legume quality”, *Trends Food Science and Technology*, 2015, Vol. 42, Issue 2, pp.116–133.
15. Murat Koklu, Ilkay Cinar and Yavuz Selim Taspinar, “Classification of rice varieties with deep learning methods”, *Computers and Electronics in Agriculture*, 2021, 187, 106285.
16. Patricio, D. I., & Rieder, R., “Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review”, *Computers and Electronics in Agriculture*, 2018, Vol.153, pp.69–81.
17. Qingxian An, Siwei Huang, Yuxuan Han, You Zhu, “Ensemble learning method for classification: Integrating data envelopment analysis with machine learning”, *Computers and Operations Research*, 2024, Vol. 169, pp. 1-19.
18. S. I. Mohammad, A. Vasudevan, and T. M. Al-Fawwaz, “Machine learning approaches for yield forecasting in major crop systems,” *Journal of Experimental Biology and Agricultural Sciences*, vol. 14, no. 2, pp. 58–62, 2026, doi: 10.18006/2026.14(2).58.62.
19. Sai Xu , Zhiyan Zhou, Huazhong Lu, Xiwen Luo and Yubin Lan, “Improved Algorithms for the Classification of Rough Rice Using a Bionic Electronic Nose Based on PCA and the Wilks Distribution”,*Sensors*, 2014, Vol.14, pp.5486-5501.
20. Yin, Y.; Tian, X.L, “Classification of Chinese drinks by a gas sensors array and combination of the PCA with Wilks distribution” *Sensors and Actuators*, 2007, Vol.124, pp.393–397.
21. Zheng, H., Zhang, Y., Yang, L., Liang, P., Zhao, Z., Wang, C., & Chen, D. Z, “A new ensemble learning framework for 3D biomedical image segmentation”. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, Vol. 33, Issue 1, pp. 5909-5916.
22. Zhou, H.T.; Yin, Y, “The defect and improved method of PCA used in electronic nose detection and analysis”, *Sensors World*, 2008, Vol.10, pp.39–42.
23. Zhou, Z.H., “Ensemble methods: foundations and algorithms”, 2012, CRC Press.