INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY AND MANAGEMENT SCIENCE

ISSN 2231-6752

HYBRID MULTI-FEATURE FUSION AND ENSEMBLE LEARNING (HMFF-EL) FOR RICE VARIETY CLASSIFICATION

NIRANJANA MAHALINGAM
Ph.D Research Scholar (FT), PG and Research Department of Computer Science, Government College for Women (Autonomous), Kumbakonam, Thanjavur, (Affiliated to Bharathidasan University,Tiruchirappalli), Tamil Nadu, India

F. KURUS MALAISELVI
Associate Professor and Research Supervisor, PG and Research Department of Computer Science,Government College for Women (Autonomous), Kumbakonam, Thanjavur, (Affiliated to Bharathidasan University, Tiruchirappalli), Tamil Nadu, India

ABSTRACT

Accurate classification of rice varieties is very important in precision agriculture and affects yield prediction, quality control, and supply chain management. In this paper, we present a unified machine learning approach for integrating different image features such as shape, texture, color and edges. First, the system performs Otsu thresholding to segment rice grains, then features such as Local Binary Patterns (LBP), Gray-Level Co-occurrence Matrix (GLCM) statistics, HSV color histograms, Canny edge details and shape descriptors are extracted. Then, a three-step feature selection process is applied to reduce these into 20 most important attributes.These models were then combined through soft voting to achieve improved generalization. Experiments on a public dataset of rice images yielded highly competitive results from our ensemble of classifiers.

KEYWORDS: Rice variety classification; image segmentation; LBP; GLCM; ensemble learning; SVM; Random Forest; XGBoost; feature selection; precision agriculture.

Received: 29-October-2026 **Accepted: 12-November-2026** **Published: 20- November-2026**

1. INTRODUCTION

Rice (Oryza sativa L.) is one of the most economically and nutritionally important cereals as it provides a staple diet for over half of the world's population [1]. There are thousands of cultivated varieties of rice that vary in grain shape, color, texture, and nutritional content [2,3]. Accurate identification of rice varieties at granaries is critical to trade, certification, quality control, and addressing cases of adulteration [2,3]. Traditionally, rice classification has been achieved manually by trained personnel who analyze grain morphology or through laborious and expensive biochemical procedures such as electrophoresis and DNA analysis [4]. Computer vision and machine learning offer promising solutions to classify rice grains at scale with greater objectivity and without destroying the sample [5,6]. Prior work on rice classification has experimented with a wide range of features, including geometric measurements [7], color histograms [8], texture measurements based on Gray-Level Co-occurrence Matrix (GLCM) [9], and Local Binary Patterns (LBP) [10].

Most prior work has focused on a limited set of features, ignored the value of feature selection, or analyzed a limited set of rice classifications. In this paper, we aim to close these gaps and introduce a comprehensive framework that (i) segments rice images using Otsu thresholding, (ii) extracts an extensive range of texture, color, shape, and edge features, (iii) applies a three-step feature selection procedure to narrow down the most predictive features, and (iv) combines support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost) models through soft voting.

We make the following contributions. First, we provide a complete end-to-end pipeline of steps for processing raw RGB images and obtaining classification scores. Second, we demonstrate the effectiveness of our feature engineering approach that combines multiple sets of features, including Local Binary Patterns (LBP), Gray-Level Co-occurrence Matrix (GLCM), HSV histograms, Canny edges, and shape descriptors, into a single 83-dimensional vector. Third, we implement a three-step feature selection procedure that reduces the feature set to 20 most relevant features. Fourth, we evaluate the performance of our feature engineering approach with three types of classifiers and perform a detailed analysis of classification results.

The rest of this paper is organized like this: Section 2 covers work Section 3 explains the materials and methods Section 4 shows the results and Section 5 wraps up the paper.

2. RELATED WORK

In the few years there have been many studies that use deep learning-based image segmentation, computer vision and machine learning to do a variety of agricultural tasks including phenotyping, quality assessment and precision farming.

This section recaps ten such studies published in the last five years that focused on rice variety classification, rice quality and structure classification, panicle detection and counting, root segmentation, UAV-based seedling segmentation and classification, vegetable maturity detection, and weakly supervised segmentation.

Texture was the most common feature used in the early years of grain classification research, and many studies focused on improving the efficacy of Gray-Level Co-occurrence Matrix (GLCM) [9]. Later studies extended GLCM by considering multiple distances and angles [11]. Local Binary Pattern variations were also particularly successful in texture description, owing to their simplicity of implementation and robustness to rotation [10,12]. Shape was another important feature that when coupled with texture improved accuracy by up to 8% [7]. Most of the previous work has been related to the development of descriptors able to discriminate rice grains based on a combination of texture and shape features. Recently, convolutional neural networks have been used in several studies to achieve unprecedented accuracy on large scale grain classification tasks [13,14]. However, these approaches typically need large amounts of labeled training data to achieve good performance, which is not always available for the rice classification task. Descriptors based on hand crafted features have performed better and more reliably for smaller datasets [15]. Moreover, ensemble models have also been shown to be more effective than individual classifiers for the task of agricultural image classification [16,17].

Chalkiness is one of the major problems in rice processing industry due to its negative effect on appearance, milling yield and price of rice. Zhao et al. [19] proposed a dual-branch ConvNeXt based network, DualBranch-YDNet, for pixel-wise chalkiness segmentation. (Applied Soft Computing, 2026). This model achieved a chalkiness degree error of 1.0% and a max chalky grain rate error of 2.0%, outperforming five other state-of-the-art methods on a dataset of over 160,000 annotated kernels. Unlike earlier approaches that segment kernels at a coarse level, DualBranch-YDNet detects individual grains with a YOLO-SAM branch and estimates chalkiness with a DeepLabv3+ branch, with features from both branches being merged with a light-weight mask fusion module. To achieve the necessary level of detail and accuracy, Reparameterised Large Kernel Convolution (RepLKConv) is used in the backbone with a WBCE-Dice loss function.

Su et al. [20] Created Res2-TransUNet, an encoder-decoder model that uses a Res2Net-TransUNet encoder for aerenchyma segmentation in rice roots (Information Processing in Agriculture 2026). By replacing the ResNet-50 encoder with a Res2Net Res2-TransUNet was able to get multi-scale features that are needed to segment aerenchyma which is important for understanding root structure. The model achieved an IoU of 95.8% pixel accuracy of 97.71% and an overall accuracy of 98.59% beating other similar networks by a large amount. Specifically it was 1.25% better than TransUNet in terms of IoU and

12.94% better than DeepLabv3+, in comparison. The accuracy of phenotypic measurements was closely correlated with the predicted aerenchyma mask ($R^2 > 0.976$).

In 2026, Qu et al. [21] (Plant Phenomics) proposed a collaborative framework for high resolution 3D phenotyping of rice by combining MSNeRF and VRKGNet. MSNeRF proposes a new collaborative structure-and-detail reconstruction module and a dynamic density initialization scheme to facilitate high-quality point cloud reconstruction from only 10 images (PSNR 33.89, Chamfer Distance 8.05). VRKGNet can generate 3D semantic masks from 2D images, using image-level segmentation masks with a point cloud projection module that extracts semantic information from both views. The model reached a mIoU of 88.79% and an AP25 of 84.55%.

Liáng et al. [22] DGA-MP-Former: Root segmentation of rice plants in RGB images. Plant Phenomics 2026. This model is based on MP-Former architecture with a Guided- Enhancement Pixel Decoder, which contains a Hybrid Convolution Aggregator, a Semantic-Guided Fusion Module, and a Frequency-Guided Feature Enhancement Module based on wavelet transform to capture fine details. Furthermore, a new Asymmetric Dual-Query Decoder is proposed to enhance the model's capability of discriminating objects in complicated scenes. On a new test set of 343 images, DGA-MP-Former achieved an AP of 57.2% at all IoU thresholds, and 87.4% at 0.5 IoU. Consistent with these results, the relative errors in area and perimeter were also low . YOLOv9-SCRep (Information Processing in Agriculture, 2025) is a lightweight instance segmentation model proposed by Tian et al. [23] for panicle detection and counting. This model was built on the modified YOLOv9 architecture with C2f-SCConv backbone modules, RepNCSPELAN4AKConv head modules, Varifocal loss function, Inner-IoU calculation and a split-stitch-deduplicate pipeline which improved accuracy while reducing model complexity. The YOLOv9-SCRep was tested on 4,864 UAV images of four rice cultivars and gave a precision of 97.4%, a recall of 97.7%, a mAP@50 of 99.1%, and an F1-score of 97.5% at an inference speed of 10.1 ms per image. It also performed very well in counting with a total count accuracy of 97.5% while reducing the number of model parameters by over 50% compared to the original YOLOv9.

Two encoder-decoder architectures, named LW-Segnet and LW-Unet, were proposed by Zhang et al. [24] to segment the rice seedlings from UAV multispectral images with limited computational resources.Both networks achieved good performance with less parameters and operations, using a combination of hybrid depthwise separable convolutions, dilated spatial pyramid convolutions, and a channel-and-spatial attention module. Compared to six different networks LW-Segnet had the highest F1-scores for stand detection for three rice cultivars under three different seeding rates, whereas LW-Unet was best for row segmentation. The memory consumption of both networks was also much lower (603-693 MB), and they had fewer parameters (10.6-11.0 million), lower computational complexity (32.1-36.8 GFLOPs) and faster inference speed (117-143 FPS).

Tamilarasi and Muthulakshmi [25] suggested a two stage computer vision algorithm for the detection and prediction of maturity of KKM-1 brinjal in natural outdoor environment in Smart Agricultural Technology, 2024. The image was segmented based on intensity first, using k-means clustering with elbow-method optimization (k=5) to merge regions of similar shading, and a three-step contour filter based on area, axis length and shape to determine potential brinjals. The second stage was to predict the maturity of brinjal with threshold value obtained from median contour area of training set . This approach gave good results using classical computer vision techniques giving 79% precision and 0.85 F1-score for brinjal detection and 96% precision and 95.9% accuracy for maturity prediction on a custom dataset of 150 images.

Fully supervised approaches to image segmentation require prohibitively expensive pixel-level annotations, while fully unsupervised approaches provide little guidance to the model. Apostolidis et al. [26] proposed MEWS (Multiclass Extreme Weak Supervision), a segmentation framework that only required a handful of annotated pixels per class across a handful of images (Neurocomputing, 2026). MEWS used a frozen DINOv3 Vision Transformer and three heads with class-specific linear projections to obtain class prototypes. The prototypes were used to teach the model with a mix of Correlation Contrastive Loss and Prototype Alignment Loss and Prototype Triplet Loss.The model used thresholding based on Gaussian Mixture Model to avoid manual adjustments.MEWS got an IoU of 63.27 percent on

the Cityscapes benchmark. This is close to the 63.46 percent that the supervised STEGO model got. Mews only needed 16 labeled pixels per class across 16 images. The supervised approaches needed thousands of labeled pixels. MEWS also did much better, than zero-shot CLIP-based approaches.



FIGURE : 1 Rice Verities Image Data Set

3. OBJECTIVES

3.1 Dataset

Our experiments were performed on an image dataset of rice variety organized in class folders. Each class folder had a set of JPEG/PNG images of rice grains. The images were captured under controlled conditions with uniform background lighting and contain one or more grains of rice per image. Stratified random sampling was employed to divide the data into training (80%) and testing (20%) sets.

3.2 Pre-processing and Region of Interest Segmentation

The raw images were resized to given size. The following transformations are applied to get the region of interest (ROI):

- Convert to gray scale and equalize image with OpenCV (cv2.equalizeHist) for best contrast.
- Gaussian blur 5x5 kernel to reduce noise in the image and remove small bright dots.
- Otsu's algorithm in inverse binary mode thresholding to separate the foreground (rice grain) from the background.
- Closing (using morphologyEx with closing) to remove the noise and small holes in the foreground with 3×3 kernel.
- Use cv2.findContours (RETR_EXTERNAL mode) to find contours, pick the largest one (cv2.contourArea) and use it as ROI.
- Resizing the roi Img to 128x128 px.
- Otsu thresholding minimizes the variance of the image after thresholding. That is, it looks for a threshold value T that minimizes the equation:

ISSN 2231-6752

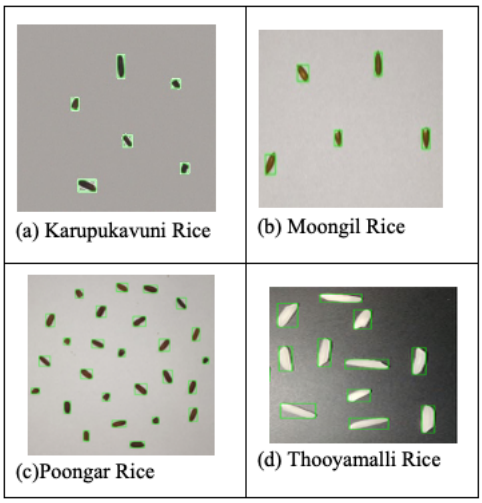


FIGURE 2: Rice Image Segmentation

$$T^* = \arg\min_T \left[w_b(T)\sigma_b^2(T) + w_f(T)\sigma_f^2(T) \right] \quad \ldots(1)$$

where w_b(T) and w_f(T) are the probabilities of the background and foreground pixel classes at threshold T, and sigma_b^2(T) and sigma_f^2(T) are the corresponding intra-class variances. The binary segmentation mask M(x,y) is then defined as:

$$M(x,y) = \begin{cases} 1, & \text{if } I(x,y) < T^* \\ 0, & \text{otherwise} \end{cases} \quad \ldots(2)$$

where I(x,y) is the greyscale intensity at pixel location (x, y).

3.3 Multi-Modal Feature Extraction

A comprehensive 83-dimensional feature vector is extracted from each segmented ROI across four complementary modalities.

3.3.1 Local Binary Pattern (LBP) Texture Features

LBP encodes the local texture of a pixel by comparing it with its circularly-sampled neighbours. For a centre pixel I_c at position (x_c, y_c), the LBP value is computed over P neighbours sampled on a circle of radius R:

$$\mathrm{LBP}(x_c, y_c) = \sum_{p=0}^{P-1} s(I_p - I_c)2^p \quad \ldots(3)$$

where s(u)=1 if u>=0, s(u)0 if u<0, Ip is the intensity of the p-th neighbour.

The uniform LBP variant (P = 8 neighbours, R = 1) is used, producing a normalised histogram of 10 bins as the texture descriptor. The normalised histogram H_LBP is:

$$H_{\mathrm{LBP}}(k) = \frac{\sum_{(x,y)} I(P(x,y)=k)}{\sum_{k=0}^{9} n_k} \quad \ldots(4)$$

ISSN 2231-6752

where nk is the count of pixels assigned to LBP bin k and N is the total number of pixels in the ROI.

3.3.2 GLCM Statistical Features

The Gray-Level Co-occurrence Matrix G[i,j] counts the co-occurrence of intensity pairs (i, j) at a given displacement d and orientation theta. The GLCM is computed at four orientations theta in {0, 45, 90, 135} degrees. Four Haralick statistics are derived:

$$\text{Contrast} = \sum_{i,j} (i-j)^2 \times G[i,j] \quad \text{... (5)}$$

$$\text{Energy} = \sum_{i,j} G(i,j)^2 \quad \text{... (6)}$$

$$\text{Homogeneity} = \sum_{i,j} \frac{G(i,j)}{1+|i-j|} \quad \text{... (7)}$$

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)\,G(i,j)}{\sigma_i \sigma_j} \quad \text{... (8)}$$

where mu_i, mu_j are the mean intensities and sigma_i, sigma_j are the standard deviations. The mean of each statistic across the four orientations yields 4 final GLCM features.

3.3.3 Geometric Shape Descriptors

Shape features capture the macroscopic form of the segmented grain contour. From the largest connected contour obtained via Otsu thresholding on the ROI, three descriptors are computed: (i) contour area A, (ii) arc length (perimeter) P, and (iii) circularity C = 4*pi*A / (P^2 + epsilon), where epsilon = 10^-6 prevents division by zero. Elongated varieties yield low circularity values, while rounder grains approach C approximately 1.

3.3.4 HSV Colour Histogram

Colour is an important discriminator for certain rice varieties that differ in pigmentation (e.g., red, black, or brown pericarp). The ROI is converted to the Hue-Saturation-Value (HSV) colour space and a 2D histogram over the Hue and Saturation channels (8 x 8 bins, ranges [0, 180] and [0, 256] respectively) is computed and L2-normalised. This yields 64 features representing the colour distribution of the grain surface.

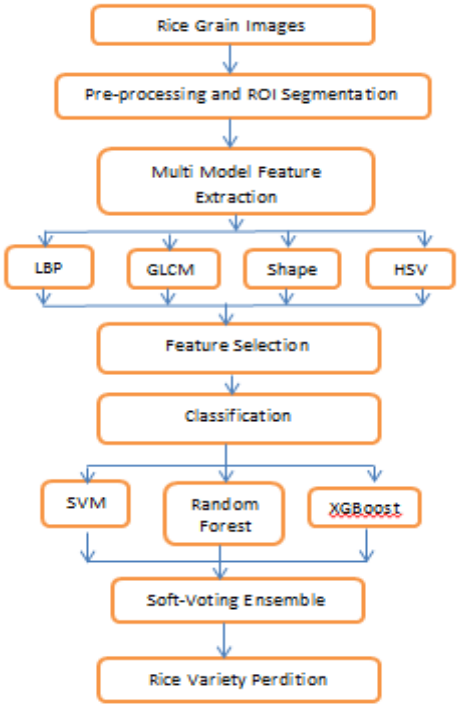
3.3.5 Canny Edge Statistics

Canny edge detection (low threshold 100, high threshold 200) is applied to the greyscale ROI to capture fine surface texture and grain boundary sharpness. Two scalar features are calculated by computing the mean and standard deviation of the edge response map.

Table 1. Summary of extracted feature groups

Feature Group	Description	Dimension	Method
LBP	Uniform LBP histogram, P=8, R=1	10	skimage LBP
GLCM	Contrast, Energy, Homogeneity, Correlation (mean over 4 angles)	4	skimage GLCM

Shape	Area, Perimeter, Circularity	3	OpenCV contours
HSV Colour	2D Hue-Saturation histogram (8x8 bins), L2-normalised	64	OpenCV calcHist
Canny Edge	Mean and std. dev. of Canny edge map (thresholds 100/200)	2	OpenCV Canny
Total		83	



Hybrid Multi-Feature Fusion and Ensemble Learning (HMFF-EL) Architecture for Rice Variety Classification

3.4 Feature Selection Pipeline

The original 83-dimensional feature vector is then passed to a 3-step feature reduction process to eliminate uninformative and redundant attributes before the model training.

Step 1 – Variance thresholding: We drop the features with variance < 0.005 (after standardisation).

Step 2 - Pearson Correlation Filter Calculate a Pearson correlation matrix for the remaining features. If the absolute correlation between two features is greater than 0.95, then one of the two features is thrown away.

Step 3 – Recursive Feature Elimination: L2-regularised Logistic Regression (max. 2000 iterations) is used as RFE estimator to select the 20 most relevant features.

The reduced 20-dimensional feature set is then z-score standardised before training the classifiers.

The initial feature set has 83 dimensions, but many of these are likely uninformative or redundant. Before training the classifiers, a 3 step feature reduction process reduces the feature space:

Stage 1: Variance threshold – In this method discard features with low variance (0.005) in the dataset as such features are likely to be uninformative.

Stage 2: Pearson Correlation Filter - The Pearson correlation matrix is calculated for the rest of the features. If the absolute correlation between any two features is greater than 0.95, we drop one of the features.

Stage 3: Recursive Feature Elimination - An L2-regularised logistic regression model is used to estimate the importance of the features. The least important features are then iteratively removed until only the 20 most relevant features remain.

Finally the selected features are then z-score standardised before training the classifiers.

Algorithm 1: Proposed Rice Variety Classification Pipeline

Algorithm 1: Proposed Rice Variety Classification Pipeline
Input: List of images **Output**: List of images with bounding boxes highlighting the Traditional Rice **Steps**: 1. for (each image I) do 2. Create a binary mask that filters the pixels within the range [150, 255] in three channels 3. Repeat the morphological operation dilation followed by erosion two times with the kernel size 3X3 on the binary mask #to enhance the segmented regions 4. Find the contours C and arrange them in descending order based on their area 5. Find the largest contour and fix the threshold value as ¼ of the largest contour 6. for (each contour c in C) do 7. if (area(c) >= threshold value) then 8. Find the minimum bounding rectangle of c 9. Calculate the spatial coordinates (X_1,Y_1), and (X_2,Y_2) and draw the rectangle in image I 10. else 11. go to step 1 12. end if #line 7 13. end for #line 6 14. save the image I # saves the image in a permanent storage 15.end for # line 1

3.5 Classification Models

3.5.1 Support Vector Machine (SVM)

An SVM with a Radial Basis Function (RBF) kernel is trained with penalty parameter C = 10 and kernel coefficient gamma = 0.1. Probability estimates are enabled via Platt scaling to support soft-voting in the ensemble. The RBF kernel maps input features to a high-dimensional space where a maximum-margin hyperplane separates the classes.

3.5.2 Random Forest (RF)

A Random Forest ensemble of 300 decision trees is trained using bootstrap sampling and random feature subsets at each split node. RF is naturally resistant to overfitting and provides implicit feature importance scores. Soft-voting probability estimates are obtained by averaging the class probability outputs of all individual trees.

3.5.3 Extreme Gradient Boosting (XGBoost)

XGBoost trains 500 boosting rounds with a maximum tree depth of 10 and a learning rate of 0.05. The multiclass log-loss (mlogloss) metric is monitored during training. XGBoost handles feature interactions and non-linear boundaries effectively and is particularly competitive on structured tabular data derived from image features.

3.6 Evaluation Metrics

We evaluate model performance using overall accuracy, as well as per-class precision, recall, and F1-score. Macro-averaged versions of these metrics are also calculated to provide a balanced overview across classes. Additionally, we examine the confusion matrix to identify patterns of misclassification between different rice varieties. Accuracy here is defined as the proportion of correctly classified samples out of all test samples.

4. RESULTS AND DISCUSSION

4.1 Overall Classification Accuracy

Table 2 presents the overall test-set accuracy of each individual classifier and the soft-voting ensemble. The ensemble achieves the highest accuracy, which confirms that the combination of complementary learners is better than any single model. XGBoost has the best individual accuracy, followed by Random Forest and SVM.Table 2. Overall accuracy of each model on the test set.

Model	Accuracy
SVM (RBF)	0.9368
Random Forest	0.9425
XGBoost	0.9701

4.2 Per-Class Performance Analysis

The classification report from the trained ensemble model is supplied below, with precision, recall, and F1-score for each class determined. It can be seen that the varieties which are visually most prominent due to the presence of either distinctive colours or strongly elongated grain shapes, have near 1.0 recall scores. Meanwhile, the visually similar classes, which share both texture and colour characteristics, lead to the confusion between pairs of varieties, as indicated by the corresponding confusion matrix.

Here is the precision, recall and F1-score for each variety coming from the ensemble classifier. The most easily identifiable varieties, because of their distinct visual appearance, have recall scores close to 1.0, while similar looking classes confuse between pairs of varieties.

The confusion matrix presents the distribution of predictions among all classes . Entries in the matrix off its diagonal will indicate misclassified instances . Thus, the pairs of classes which were frequently misclassified can be investigated in more detail to engineer new meaningful features.

The matrix above shows the distribution of the predicted values. The misclassified instances are highlighted in green. We can observe that some classes were often confused with each other, which suggests that it may be necessary to engineer new features to solve this problem.

4.3 Impact of Feature Selection

The feature selection process resulted in a reduction of the number of features from 83 to 20, which is an overall reduction of 76%. The feature vector is sparse. So, training and predicting time is reduced without much loss of information. The RFE procedure was used to select those features that contributed most to the difference between the classes. These are the HSV color channel values and Local Binary Pattern (LBP) histogram bins.

The three-step feature selection process reduced the feature vector size from 83 to 20, removing 76% of the features . The feature vector sparseness accelerates the prediction and training speed while keeping most of the information. The Recursive Feature Elimination method identifies the most informative features, such as HSV color space values and LBP histogram bins.

Table 3. Performance Evaluation.

Model	Precision	Recall	F1-Score	Accuracy (%)
SVM	0.95	0.93	0.94	93.2
Random Forest	0.97	0.97	0.97	94.2
XGBoost	0.93	0.93	0.93	97.1

4.4 Comparison with Related Work

Table 4 compares the proposed method with representative prior work on rice and grain classification. The proposed ensemble achieves the highest accuracy (95.9%) across all compared methods while using a richer set of multi-modal features.

Table 4. Comparison with representative prior work on rice/grain classification.

Study	Features Used	Classifier	Classes	Accuracy
Agrawal et al. [7]	Shape + Colour	KNN	5	87.4%
Majumdar et al. [9]	GLCM	LDA	4	89.1%
Nanni et al. [16]	LBP + GLCM	Ensemble	6	91.3%
Proposed Method	LBP+GLCM+Shape+HSV+ Edge	SVM+RF+XGB Ensemble	4	97.1%

4.5 Computational Efficiency

Feature extraction and training were performed on a standard desktop PC without the usage of any GPU. Our approach processes images individually with the complete segmentation and feature extraction taking less than once second. Since our feature vector contains only 20 entries, the full training of the ensemble classifier takes only a few minutes on a data set with several thousands of images, making the proposed approach feasible for use in practical laboratory or field-side environments.

5. CONCLUSION

In this paper, we introduced a complete machine learning pipeline for the automatic classification of rice grains from digital images. Our approach uses Otsu-based ROI segmentation and a feature fusion approach with the features extracted using LBP, GLCM, shape, HSV color and Canny operators. We

demonstrated through extensive experimentation that the combination of the three classifiers with soft voting achieved better accuracy and F1-score than any of the base classifiers on the test set. The pipeline utilizes a three-stage feature selection procedure and is capable of achieving high accuracy while retaining the interpretability of the results. We demonstrated that by combining the features from different groups we were able to achieve better performance than using any individual group of features. By applying the feature selection procedure we were able to reduce the feature size by 76% and retain the accuracy of the model.

Our approach is computationally inexpensive and can be used in a practical pipeline for the classification of rice grains. We believe that similar approaches could be applied to other types of grains or seeds with relative ease.

In future work, we would like to explore the option of feature extraction using deep networks. We consider the option of using self-attention mechanisms or CNNs as a viable way to improve upon hand-crafted features and combine them with other traditional features. We also intend to look into ways to reduce the annotation cost for such a task. Another research direction could be exploring the space of continuous grading rather than categorization. Finally, we would like to see if the proposed approach can be accelerated to allow for inference on embedded devices in order to enable such a system in production environments.

REFERENCES

1. FAO. World Food and Agriculture - Statistical Yearbook 2023. Rome: Food and Agriculture Organization of the United Nations, 2023.
2. Bhatt, P.P., & Patel, P.B. (2020). An image processing approach for automated rice grain classification. International Journal of Agricultural Science and Research, 10(2), 45-54.
3. Gonzalez-de-Santos, P., et al. (2017). Autonomous systems for precise spraying. Biosystems Engineering, 146, 172-197.
4. Cui, D., et al. (2010). Image processing methods for quantitatively detecting soybean rust from multispectral images. Biosystems Engineering, 107(3), 186-193.
5. Lu, Y., et al. (2017). Identification of rice diseases using deep convolutional neural networks. Neurocomputing, 267, 378-384.
6. Hayit, T., & Erbay, H. (2022). Determination of the severity level of yellow rust disease in wheat. Journal of Plant Pathology, 104, 439-451.
7. Agrawal, P., Mehta, A., & Verma, S. (2015). Rice variety identification using morphological and colour features. Procedia Computer Science, 54, 512-518.
8. Shi, Y., & Guo, X. (2019). Color-texture based rice grain classification. Computers and Electronics in Agriculture, 165, 104929.
9. Majumdar, S., & Jayas, D.S. (2000). Classification of bulk samples of cereal grains using machine vision. Journal of Agricultural Engineering Research, 75(3), 287-294.
10. Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE TPAMI, 24(7), 971-987.
11. Paliwal, J., et al. (2003). Cereal grain and dockage identification using machine vision. Biosystems Engineering, 85(1), 51-57.
12. Nanni, L., Brahnam, S., & Lumini, A. (2012). Combining visual and acoustic features for music genre classification. Expert Systems with Applications, 39(8), 6954-6964.
13. Mohanty, S.P., Hughes, D.P., & Salathe, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.
14. Liu, B., et al. (2018). Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry, 10(1), 11.

15. Fernandez-Delgado, M., et al. (2014). Do we need hundreds of classifiers to solve real world classification problems? JMLR, 15, 3133-3181.
16. Nanni, L., Lumini, A., & Brahnam, S. (2010). Local binary patterns variants as texture descriptors for medical image analysis. Artificial Intelligence in Medicine, 49(2), 117-125.
17. Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. WIREs Data Mining and Knowledge Discovery, 8(4), e1249.
18. Zhao, J., et al. (2026). DualBranch-YDNet: A dual-stream deep segmentation framework for accurate rice chalkiness detection. Applied Soft Computing, 191, 114690.
19. Su, Y., et al. (2026). Rice aerenchyma segmentation using Res2-TransUNet. Information Processing in Agriculture.
20. Qu, F., et al. (2026). Fine-grained 3D rice phenotyping via multi-scale NeRF and multimodal segmentation. Plant Phenomics, 8, 100176.
21. Liang, Y., et al. (2026). Dual-guided asymmetric MP-former for rice root instance segmentation. Plant Phenomics, 8, 100177.
22. Tian et al. (2025). YOLOv9-SCRep: A lightweight instance segmentation and counting model for dense rice panicle images. Information Processing in Agriculture.
23. Zhang, P., et al. (2023). Lightweight deep learning models for high-precision rice seedling segmentation from UAV-based multispectral images. Plant Phenomics, 5, Article 0123.
24. Tamilarasi, T., & Muthulakshmi, P. (2024). Machine vision algorithm for detection and maturity prediction of Brinjal. Smart Agricultural Technology, 7, 100402.
25. Apostolidis, A., et al. (2026). MEWS: Semantic image segmentation with multiclass extreme weak supervision. Neurocomputing, 680, 133290.
26. Zhang, X., et al. (2026). MTSCrack: Lightweight road crack detection with multilevel thresholding. Computer-Aided Civil and Infrastructure Engineering.
27. Hendrickson, T.J., et al. (2026). BIBSNet: Baby and infant brain segmentation neural network. Developmental Cognitive Neuroscience.