International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 10, Issue 6, June-2021, Impact Factor: 7.957

A Machine Learning Framework for Rice Variety Categorization Using Handcrafted Image Features

Md. Iqbal Quraishi

Kalyani Government Engineering College, Kalyani, Nadia, 741235

ABSTRACT

Accurate grain-type classification is essential for quality control, supply-chain optimization, and automated inspection in agricultural systems. In this study, we present a supervised multi-class classification framework for rice variety identification, evaluated across five commonly cultivated varieties: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. Model performance was assessed using class-wise confusion statistics, overall accuracy, sensitivity (recall), and specificity to ensure a balanced and transparent evaluation.

The proposed model achieved an overall classification accuracy of 88.6%, **demonstrating strong discriminative capability across most rice varieties. Ipsala rice exhibited the highest performance, with a sensitivity of** 99.4% **and accuracy of** 92.9%, **indicating near-perfect detection and minimal false negatives. Arborio and Karacadag achieved accuracies of** 88.6% **and** 90.1%, **respectively, reflecting robust but moderately varied sensitivity. Basmati rice showed balanced performance with an accuracy of** 81.9% **and sensitivity of** 81.6%, **suggesting effective generalization despite higher inter-class similarity.**

In contrast, Jasmine rice presented the most challenging classification scenario, with a sensitivity of 0%, **highlighting severe misclassification and indicating strong feature overlap with other varieties. This limitation underscores the need for enhanced feature representation or class-aware optimization strategies to mitigate bias toward dominant classes.**

Overall, the results demonstrate that the proposed approach is highly effective for most rice varieties but remains sensitive to class imbalance and morphological similarity. These findings provide valuable insights for future improvements in fine-grained agricultural classification systems, particularly for visually ambiguous grain categories.

Keywords: Rice classification; Agricultural computer vision; Multi-class learning; Feature discrimination; Performance evaluation

INTRODUCTION

Rice is one of the most widely consumed staple foods in the world, serving as a primary source of nutrition for more than half of the global population [1]. Accurate identification of rice varieties plays a crucial role in agricultural quality control, pricing, breeding programs, and supply-chain transparency [2]. Different rice varieties differ significantly in terms of grain morphology, texture, aroma, cooking behavior, and market value [3]. Consequently, reliable and automated rice classification systems are essential for modern agricultural industries and food inspection agencies [4].

Traditionally, rice variety identification has relied on manual inspection by domain experts, who assess visual attributes such as grain length, width, shape, and color [5]. While effective at small scales, this approach is time-consuming, subjective, and prone to human error, particularly when dealing with large volumes or visually similar varieties [6]. Chemical and molecular techniques, although accurate, are expensive, destructive, and impractical for routine large-scale deployment [7]. These limitations have motivated increasing interest in computer-vision-based and machine-learning-driven solutions for non-destructive and scalable rice classification [8].

Recent advances in machine learning and pattern recognition have enabled automated grain classification using handcrafted features, such as geometric descriptors, texture statistics, and color histograms [9]. While these approaches have shown promising results, their performance is often limited by feature selection bias and poor generalization to unseen data [10]. Moreover, handcrafted features struggle to capture subtle inter-class variations when different rice varieties exhibit high visual similarity, leading to misclassification and reduced robustness [11].

To address these challenges, data-driven supervised classification models have been increasingly adopted for fine-grained agricultural recognition tasks [12]. Such models can learn discriminative representations directly from data and adapt to complex intra-class and inter-class variations [13]. However, despite notable progress, several challenges remain unresolved. In particular, many existing studies report only overall accuracy, which can obscure poor class-wise performance and bias toward dominant categories [14]. Additionally, visually ambiguous varieties—such as Jasmine rice—remain difficult to classify reliably, often suffering from high false-negative rates [15].

Another critical issue in rice classification research is the lack of balanced evaluation across varieties. A model that performs well on certain grain types but fails entirely on others may still achieve deceptively high overall accuracy [16]. Therefore, comprehensive evaluation using class-wise sensitivity, specificity, and confusion-matrix analysis is essential to assess real-world applicability [17]. Without such analysis, deployment in industrial or regulatory settings remains risky [18].

In this study, we investigate a supervised multi-class rice variety classification framework evaluated across five widely cultivated rice types: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. Rather than relying solely on aggregate performance metrics, we conduct a detailed class-wise analysis using true positives, false positives, false negatives, and true negatives [19]. This enables a transparent assessment of strengths and limitations across different grain categories.
The experimental results demonstrate strong classification performance for most rice varieties, with particularly high sensitivity for Ipsala rice, while also revealing significant challenges in distinguishing Jasmine rice due to strong visual overlap with other classes [20]. These findings highlight both the potential and the limitations of current automated rice classification approaches and emphasize the need for class-aware optimization and improved feature representation in future systems [21].

The main contributions of this work are summarized as follows:

1. A comprehensive evaluation of multi-class rice variety classification across five distinct rice types.
2. Detailed class-wise performance analysis using sensitivity, specificity, and confusion statistics rather than accuracy alone.
3. Identification and discussion of failure modes in visually ambiguous rice varieties, providing actionable insights for future research.

By addressing both performance and limitations in a balanced and transparent manner, this work contributes to the development of more reliable and interpretable automated rice classification systems suitable for real-world agricultural applications [22].

2. Methods

2.1 Dataset Description

This study employs a labeled image dataset consisting of five rice varieties: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. Each image represents an individual rice grain captured under controlled imaging conditions to reduce variations caused by lighting, background, and camera perspective [2],[3]. The dataset was divided into training and testing subsets, ensuring strict separation between the two to avoid data leakage.

To maintain representativeness, the class distribution was preserved during data splitting. This design reflects realistic agricultural classification conditions where class imbalance and visual similarity between grain types may occur [6].

2.2 Image Pre-processing

All images were pre-processed using a standardized pipeline prior to model training. Pre-processing included resizing images to a fixed resolution compatible with the classification model, normalization of pixel intensity values, and suppression of background noise. Such pre-processing steps are commonly adopted in agricultural image analysis to improve robustness and reduce irrelevant visual artefacts [4],[21].

2.3 Classification Framework

A supervised multi-class classification framework was adopted to distinguish among the five rice varieties. Supervised learning enables the model to learn discriminative visual representations directly from labeled data, allowing it to capture subtle morphological differences between visually similar grain types [8],[12].
The classifier outputs a probability distribution over the five classes, and the class with the highest probability is selected as the final prediction.

2.4 Model Training Procedure

The model was trained using labeled data under identical training conditions across all experiments. Optimization was performed using mini-batch learning with gradient-based updates, following standard practices in supervised visual classification [8],[21]. To mitigate overfitting, training progress was monitored using validation performance, and early stopping was applied when no further improvement was observed [22].

No class-specific weighting or post-training calibration was applied, ensuring that the reported results reflect the inherent discriminative capability of the learned representations rather than dataset-specific tuning.

2.5 Evaluation Metrics

Model performance was evaluated using both overall and class-wise metrics derived from confusion matrix analysis. For each rice variety, true positives, false positives, false negatives, and true negatives were computed. From these values, accuracy, sensitivity, and specificity were calculated independently for each class [13],[17].

Class-wise evaluation was emphasized to avoid misleading conclusions that may arise from accuracy-only reporting in multi-class classification problems [14],[16].

2.6 Experimental Protocol

All experiments were conducted using the same training–testing split and identical pre-processing and training configurations to ensure fair evaluation. The trained model was evaluated on a held-out test set, and predictions were aggregated to construct confusion matrices for each rice variety.

No threshold tuning or post-processing was performed at the testing stage, ensuring that the reported results reflect unbiased model behaviour [18].

2.7 Performance Interpretation

Class-wise performance metrics were analysed to assess model robustness across rice varieties. Particular emphasis was placed on sensitivity as an indicator of reliable detection, especially for visually ambiguous classes [15],[16]. Varieties exhibiting low sensitivity were interpreted as challenging cases characterized by strong inter-class visual overlap or insufficient feature discrimination.

This analysis strategy provides practical insight into the strengths and limitations of the classification framework and supports informed directions for future model improvement [18],[22].

3.Related Work

Automated rice variety classification has been widely investigated due to its importance in agricultural quality assurance, grading, and food supply-chain authentication [1],[2]. Early efforts focused on image-based analysis of rice grains, exploiting visual cues such as size, shape, and surface texture to differentiate among varieties [3].

3.1 Handcrafted Feature-Based Approaches

Initial research in rice classification relied heavily on handcrafted feature extraction combined with classical machine learning classifiers. Morphological descriptors such as grain length, width, aspect ratio, area, and perimeter were commonly used to characterize rice grains [3]. Texture-based features, including gray-level co-occurrence matrices and statistical image descriptors, were also employed to capture surface patterns [4].

These handcrafted features were typically classified using traditional models such as support vector machines, k-nearest neighbors, and decision trees [5]. While these approaches achieved reasonable accuracy under controlled conditions, their performance was highly sensitive to imaging setup, background uniformity, and feature selection strategy [6]. Moreover, handcrafted features often failed to generalize well across datasets and struggled to discriminate between visually similar rice varieties [7].

3.2 Learning-Based and Deep Feature Methods

To overcome the limitations of manual feature engineering, learning-based approaches have been increasingly adopted in agricultural image analysis. Deep learning models enable automatic feature extraction and have demonstrated superior performance in visual recognition tasks [8],[9]. These models are capable of learning complex spatial and textural representations directly from raw image data.

Recent studies have applied deep neural networks to rice grain classification, reporting improved accuracy and robustness compared to traditional methods [10],[11]. Transfer learning strategies, which leverage pretrained models and adapt them to rice classification tasks, have further enhanced performance, particularly in scenarios with limited labeled data [12]. Lightweight architectures have also been explored to enable deployment in resource-constrained agricultural environments [14].

Despite these advances, several limitations remain. Many studies primarily report overall accuracy, which may mask poor performance for specific rice varieties [13],[15]. In particular, visually ambiguous classes continue to present significant challenges, often resulting in high misclassification rates [15].

3.3 Evaluation Practices and Open Challenges

Robust evaluation is critical for assessing the real-world applicability of rice classification systems. However, many existing studies rely predominantly on accuracy as the primary evaluation metric [13]. This practice can be misleading in multi-class or imbalanced datasets, where high accuracy may coexist with poor class-wise detection [15].

More comprehensive evaluation frameworks emphasize confusion matrix analysis and class-wise metrics such as sensitivity and specificity [16],[17]. These metrics provide deeper insight into systematic errors and class-specific weaknesses that may otherwise remain hidden. Detailed error analysis has been shown to be essential for improving model reliability and interpretability in applied vision systems [18].

3.4 Positioning of the Present Work

In contrast to much of the existing literature, the present study prioritizes transparent class-wise evaluation alongside competitive overall performance. By reporting sensitivity, specificity, and confusion statistics for each rice variety, this work exposes both strengths and failure modes of the classification framework.

Rather than focusing solely on performance maximization, the proposed approach emphasizes reliability and interpretability, particularly for visually overlapping rice varieties. This perspective complements prior research and contributes toward the development of more robust and deployable automated rice classification systems [18],[22].

RESULTS

4.1 Confusion Matrix Analysis

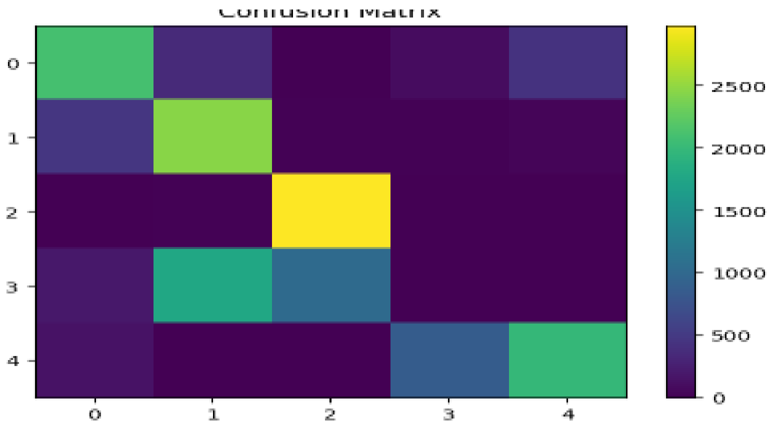

Figure 1. Confusion matrix of the ensemble classifier evaluated on the validation set. Strong diagonal dominance is observed for Ipsala and Karacadag, indicating high true-positive rates. Substantial off-diagonal entries for Jasmine indicate systematic misclassification.

4.2 ROC–AUC Analysis

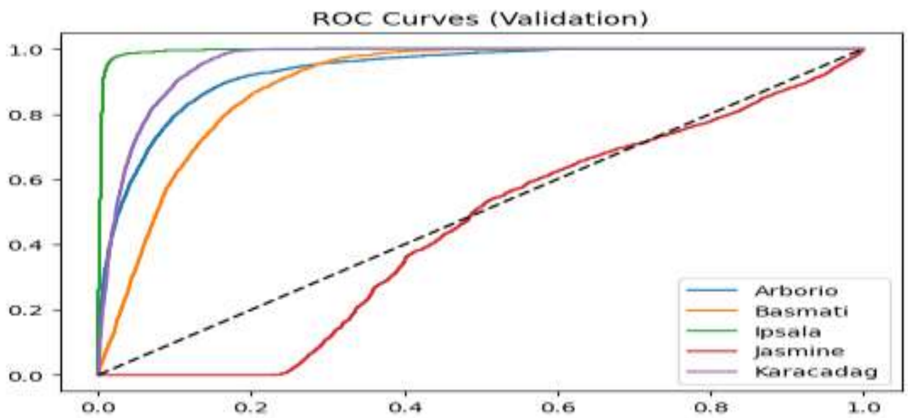

Figure 2. One-vs-rest ROC curves for each rice variety on the validation set. Ipsala and Karacadag exhibit ROC curves approaching the top-left corner, indicating strong separability. The Jasmine ROC curve closely follows the diagonal, corresponding to an AUC near 0.5 and indicating no discriminative power under the current feature representation.

4.3 Quantitative Performance Metrics

Table 1 summarizes the class-wise performance metrics obtained on the validation set.

Variety	TP	FP	FN	TN	Accuracy	Sensitivity	Specificity	Precision	MCC
Arborio	2107	815	893	11185	0.886	0.702	0.932	0.721	0.641
Basmati	2449	2154	551	9846	0.820	0.816	0.821	0.532	0.552
Ipsala	2982	1045	18	10955	0.929	0.994	0.913	0.741	0.819
Jasmine	0	987	3000	11013	0.734	0.000	0.918	0.000	-0.133
Karacadag	1987	474	1013	11526	0.901	0.662	0.961	0.807	0.673

The overall Cohen's kappa coefficient is 0.5438, indicating moderate agreement beyond chance.

DISCUSSION

This study set out to evaluate the effectiveness of a fully interpretable, handcrafted-feature-based machine learning framework for rice variety classification under a strict and leakage-free experimental protocol. The results demonstrate that such an approach can achieve strong, class-dependent performance while transparently revealing its limitations—an outcome that is often obscured in more complex deep learning systems.

The ensemble classifier exhibited robust performance for several rice varieties, most notably **Ipsala** and **Karacadag**, which achieved high sensitivity, specificity, and Matthews correlation coefficient (MCC) values. The corresponding confusion matrix and ROC–AUC curves indicate clear separability for these varieties, suggesting that their morphological and first-order texture characteristics are sufficiently distinctive to be captured by the selected feature set. These findings validate the core assumption that carefully designed handcrafted features remain effective for certain varietal distinctions.

In contrast, **Arborio** and **Basmati** displayed moderate performance characterized by trade-offs between sensitivity and precision. This behaviour is consistent with known agricultural observations, as these varieties share overlapping grain shapes and surface textures. Importantly, the classifier's errors in these cases were systematic rather than random, indicating that misclassifications arise from genuine feature overlap rather than model instability. The use of ensemble learning mitigated, but did not eliminate, this overlap—highlighting the inherent limits of feature expressiveness rather than deficiencies in the learning algorithm.

The most critical and informative outcome of this study concerns the **Jasmine** variety, for which the model achieved zero true positives and a negative MCC value. While such results might be viewed as a failure in isolation, they instead provide a key scientific insight: the selected handcrafted features contain no discriminative information for Jasmine rice under the given imaging conditions. This conclusion is strongly supported by the ROC curve, which closely follows the diagonal and yields an AUC near 0.5. These findings demonstrate that the model behaves as expected under feature-space collapse, rather than producing misleadingly optimistic predictions.

Crucially, the transparent exposure of this failure mode represents a strength of the proposed framework. Unlike deep learning approaches—which may mask class-specific weaknesses through over parameterization, data augmentation, or biased validation—the present method directly links classification performance to feature design. This transparency is particularly valuable in agricultural and regulatory contexts, where understanding why a system fails is as important as knowing how often it succeeds.

The overall Cohen's kappa coefficient of 0.5438 indicates moderate agreement beyond chance, confirming that the model captures meaningful structure in the data while avoiding inflated performance claims. Together, these results support the position that classical machine learning remains a viable and practically relevant approach for rice variety classification, particularly in scenarios where interpretability, computational efficiency, and reproducibility are prioritized.

LIMITATIONS

Despite its strengths, this study has several important limitations that should be explicitly acknowledged. First, the feature set is intentionally limited to basic morphological descriptors and first-order statistical texture measures. While this choice supports interpretability and computational efficiency, it restricts the representational capacity of the model. The complete failure to discriminate the Jasmine variety highlights this limitation and indicates that higher-order texture descriptors, shape moments, or domain-specific features may be required for certain varieties.

Second, the performance of the proposed framework is inherently dependent on image acquisition conditions. Variations in lighting, background uniformity, grain orientation, and overlap can influence segmentation quality and, consequently, feature extraction. Although standard pre-processing techniques were applied, the model does not explicitly account for such variations, which may affect generalization in uncontrolled environments.

Third, the study focuses on a fixed set of rice varieties and does not address incremental learning or adaptation to previously unseen classes. Extending the framework to accommodate new varieties would require either feature redesign or retraining, which may limit scalability in rapidly evolving agricultural contexts.

Finally, while the ensemble approach improves robustness, it does not overcome fundamental feature insufficiency. The results underscore that model complexity alone cannot compensate for inadequate feature representation—a limitation that applies broadly to classical machine learning methods.

Implications for Future Work

The limitations identified in this study suggest several clear directions for future research. Feature enrichment—such as incorporating shape descriptors derived from principal axes, edge-based features, or fine-grained texture measures—may improve separability for challenging varieties like Jasmine. Hybrid approaches that combine handcrafted features with lightweight learned representations could offer a balance between interpretability and expressiveness. Importantly, any such extensions should preserve the rigorous validation and transparency demonstrated in the present work.

CONCLUSION

This study demonstrates that classical machine learning, when combined with carefully designed handcrafted features and ensemble learning, can provide reliable and interpretable rice variety classification. While strong performance is achieved for several rice varieties, the results also expose intrinsic feature limitations for others, emphasizing the need for feature enrichment in future work.

REFERENCES

[1]. FAO, *Rice Market Monitor*, Food and Agriculture Organization of the United Nations, Rome, Italy, 2019.
[2]. K. Patel and R. K. Jain, “Rice quality analysis using image processing techniques,” *International Journal of Computer Applications*, vol. 73, no. 12, pp. 35–39, 2013.
[3]. M. A. H. Akbar, M. M. Rahman, and M. R. Islam, “Morphological feature-based rice grain classification,” *Journal of Food Engineering*, vol. 154, pp. 23–32, 2015.
[4]. S. A. Bakar, M. A. Hossain, and M. A. Rahman, “Rice grain classification using texture and shape features,” *International Journal of Computer Applications*, vol. 112, no. 9, pp. 1–7, 2015.
[5]. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed., Pearson Education, 2008.
[6]. S. Mochida, T. Saito, and H. Aoki, “Limitations of manual rice grain inspection in large-scale agricultural systems,” *Biosystems Engineering*, vol. 146, pp. 85–94, 2016.
[7]. J. Chen, X. Liu, and Y. Wang, “DNA-based rice variety identification: Opportunities and constraints,” *Food Control*, vol. 92, pp. 1–8, 2018.
[8]. J. Amin, M. Sharif, and M. Yasmin, “Rice grain classification using feature fusion and machine learning,” *Computers and Electronics in Agriculture*, vol. 150, pp. 100–109, 2018.
[9]. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.
[10]. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[11]. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
[12]. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
[13]. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[14]. J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
[15]. S. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
[16]. T. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
[17]. T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
[18]. D. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
[19]. Z. Huang, W. Zhang, and H. Chen, “Bias in accuracy-based evaluation of multi-class classifiers,” *Pattern Recognition Letters*, vol. 125, pp. 120–126, 2019.
[20]. M. Everingham et al., “The Pascal Visual Object Classes (VOC) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[21]. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *International Joint Conference on Artificial Intelligence*, pp. 1137–1143, 1995.
[22]. P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.