IJE TRANSACTIONS A: Basics Vol. 38 No. 10, (October 2025) 2435-2450

International Journal of Engineering

Journal Homepage: www.ije.ir

An Optimized YOLO-ViT Hybrid Model for Enhanced Precision in Rice Classification and Quality Assessment

S. Mavaddati*[a], M. Razavi[b]

[a] Electronic Department, Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran
[b] Computer Engineering Department, Shomal University, Amol, Iran

PAPER INFO

Paper history:
Received 15 November 2024
Received in revised form 04 December 2024
Accepted 15 December 2024 2024

Keywords:
Quality Assessment
You Only Look Once
Vision Transformer
Hybrid Model
Agricultural Automation

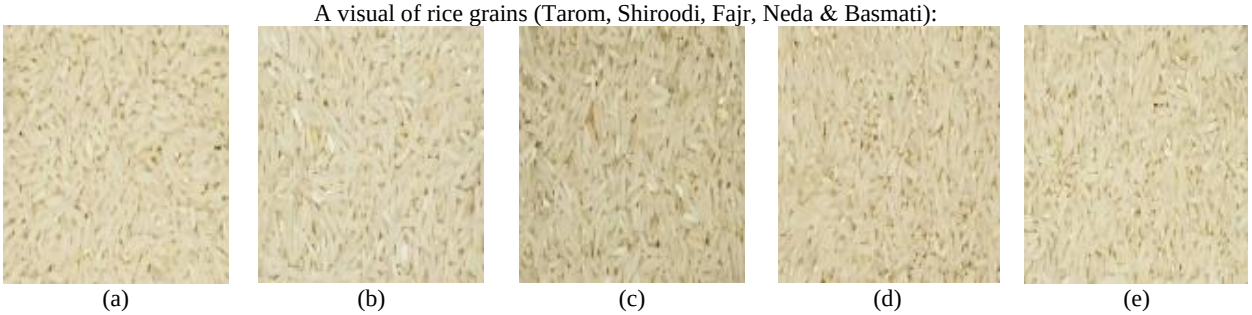
ABSTRACT

Rice holds strategic importance in economic and nutritional value, making accurate classification and quality assessment essential for agricultural production and market supply chains. This study introduces an innovative approach that combines the strengths of the You Only Look Once (YOLOv8) and Vision Transformer (ViT) models to enhance the classification of five key rice varieties like Tarom, Shiroodi, Fajr, Neda, and Basmati. That provides a comprehensive quality assessment. YOLOv8 enables rapid and precise detection of rice grains in images, while ViT captures complex spatial relationships and dependencies among image features, improving the model's ability to handle intricate patterns and contextual information. Three scenarios are explored: Scenario I employs a standalone YOLOv8 model; Scenario II implements a YOLO-ViT hybrid model for extracting spatial and relational features, and Scenario III integrates YOLO-ViT for combined detection and quality evaluation. The results demonstrate that the hybrid YOLO-ViT model significantly enhances classification accuracy and quality assessment, highlighting its effectiveness for agricultural quality control and food supply chain management. This approach innovatively leverages YOLO's fast object classification capabilities and ViT's ability to model complex relationships, providing a high-precision and efficient solution for rice quality evaluation., The proposed model has the potential to improve food safety and facilitate effective market management.The model is widely applicable in automated agricultural systems.

doi: *10.5829/ije.2025.38.10a.19*

Graphical Abstract

A visual of rice grains (Tarom, Shiroodi, Fajr, Neda & Basmati):



(a) (b) (c) (d) (e)

1. INTRODUCTION

Rice is one of the most important staple crops globally, and it is vital to ensure food security, economic stability, and nutritional value. It is a primary food source for over half of the world's population, particularly in Asia, where it holds immense cultural and economic significance (1, 2). Classifying rice varieties and assessing their quality is essential for ensuring product consistency, meeting market demands, and maintaining the integrity of

*Corresponding Author Email: s.mavaddati@umz.ac.ir
(S. Mavaddati)

Please cite this article as: Mavaddati S, Razavi M. An Optimized YOLO-ViT Hybrid Model for Enhanced Precision in Rice Classification and Quality Assessment. International Journal of Engineering, Transactions A: Basics. 2025;38(10):2435-50.

agricultural supply chains. These processes are critical for improving agricultural practices, enhancing food safety, and enabling better quality control throughout the food supply chain (3).

In recent years, various machine-learning approaches have been explored for rice classification and quality evaluation (4-6). Traditional machine learning techniques, such as support vector machines (SVM), decision trees, and random forests, have demonstrated reasonable performance. However, these methods often struggle with complex and high-dimensional data, especially when dealing with image-based inputs. Moreover, they require manual feature extraction and are prone to overfitting when working with small datasets or images with varying conditions. Deep learning models, particularly Convolutional Neural Networks (CNNs), have revolutionized the field of image classification by automating feature extraction and learning hierarchical representations directly from raw data. These models recognize complex patterns, making them ideal for tasks involving large and diverse datasets, such as rice classification. In response to the growing challenges in rice quality evaluation, hybrid models combining deep learning techniques, such as the integration of YOLO and Vision Transformer, have shown promising results by leveraging the strengths of both object detection and relational feature modeling. These models enhance both the speed and accuracy of classification, offering a comprehensive solution for rice grading and quality assessment (7, 8). This study explores these advanced deep-learning methods, demonstrating their potential to address challenges in agricultural quality control and supply chain management. These models not only outperform traditional methods in terms of accuracy and scalability but also offer significant advantages in handling the complexities of real-world data. By combining the strengths of YOLO for fast and precise object detection and Vision Transformer (ViT) for capturing intricate spatial relationships, the proposed hybrid model provides a robust framework for improving rice quality evaluation and classification. This hybrid model enhances both the speed and accuracy of classification, presenting a comprehensive solution for rice grading and quality assessment in agricultural settings. In recent years, several studies have explored the use of deep learning models for rice quality grading and classification, demonstrating the potential of automated systems in agricultural applications. For instance, a real-time, cost-effective deep learning model was developed by Jeyaraj et al. (9), which utilized high-resolution imaging combined with transfer learning to extract key features, achieving high performance in both standard datasets and real-time applications. Similarly, Pradana-López et al. (10) developed a CNN-based classifier for detecting fraudulent rice varieties, creating a rice image database to accurately differentiate between genuine and substandard rice varieties. Golpour et al. (11) employed a backpropagation neural network (BPNN) with feature selection, incorporating RGB, HSI, and HSV color spaces to classify various rice types with exceptional accuracy.

Further advancements in deep learning for rice classification include the use of residual networks (ResNet20) to categorize five rice varieties based on visual features such as size, shape, and surface texture, as demonstrated by Shrivastava et al. (12). Abbaspour-Gilandeh et al. (13) applied an artificial neural network (ANN) to classify 13 Iranian rice varieties, surpassing traditional decision-making algorithms in terms of performance. Other notable research, Deng (14) developed a CNN model trained on a large dataset of rice images under diverse conditions, enabling automatic rice sprout counting. Additionally, combining color, morphological, and shape features has been explored in studies conducted by Pazoki et al. (15), which utilized multi-layer perceptron (MLP) and neuro-fuzzy models to classify rice types, demonstrating the effectiveness of these methods in feature selection and classification.

Other efforts have focused on integrating hyperspectral imaging and CNNs for non-destructive rice classification by leveraging both spatial and spectral feature analysis (16). A pre-trained deep model, combining InceptionV3 and InceptionResNetV2, was also introduced by Ahmad Dar et al. (17) for high-accuracy rice seed type classification. Further applications of CNN-based deep learning models include classifying rice grains such as broken and fine rice (18), as well as categorizing rice products in Vietnam into whole and broken rice (19). Studies conducted by Lakshmi and Seetharaman (20) and Razavi et al. (21) have shown the effectiveness of deep models based on ResNet50, Xception, and Rice-ResNet, achieving high classification accuracy and robustness across various rice varieties and quality levels.

Razavi et al. (22) proposed a deep model named Rice-ResNet, combining ResNet and transfer learning to enhance rice classification and quality assessment. This model demonstrated improved classification accuracy and robustness across various rice varieties, highlighting the growing potential of deep learning for agricultural applications. Inspired by these advancements, the present paper introduces a novel approach to rice classification and quality assessment using a hybrid deep learning model combining YOLOv8 and ViT. YOLOv8 is employed for the precise detection of rice grains in images, while ViT excels at capturing intricate spatial relationships and dependencies among image features. This approach focuses on classifying five rice varieties commonly cultivated in northern Iran: Tarom, Shiroodi, Fajr, Neda, and Basmati. The study explores three distinct scenarios: Scenario I: Utilizes the standalone YOLO model for rice grain detection. Scenario II:

Integrates YOLO with ViT to extract both spatial and relational features. Scenario III: Combines YOLOv8 and ViT for simultaneous classification and quality assessment.

As mentioned in recent years, deep learning models have shown significant promise in the field of agricultural quality assessment, particularly for rice classification. Various approaches have been explored, with many studies focusing on the application of CNNs, ResNet, and other models for rice grading and quality evaluation. Despite these advancements, a few critical research gaps remain that hinder the further application of these techniques in real-world agricultural settings. Although existing methods, such as the use of CNNs for detecting rice varieties (9), and ResNet models for rice classification (21), have shown promising results, these models often struggle with handling the complex spatial dependencies and relational features inherent in rice grain images. Moreover, while Ahmad Dar et al. (17) have demonstrated the power of transfer learning for rice seed classification, they typically fail to integrate both spatial and relational features effectively, which are critical for accurate rice classification and quality evaluation. Additionally, previous research has largely relied on a single model or simple combinations of models for rice classification. For example, the Rice-ResNet model (22) combined ResNet with transfer learning to improve accuracy. However, this model still faces challenges in dealing with the fine-grained features of rice varieties, particularly in scenarios where the data is diverse and includes different rice types under varying environmental conditions. To address these limitations, the current work introduces a novel hybrid model combining YOLOv8 and Vision Transformer (ViT). The novelty of this approach lies in its ability to leverage YOLOv8 for precise object detection, which ensures accurate localization of rice grains in images, and ViT for modeling complex spatial relationships among features. This dual-focus approach not only improves classification accuracy but also enables a more robust quality assessment, which has been an overlooked aspect in prior research. Furthermore, unlike many existing studies that focus on either classification or quality evaluation separately, our method simultaneously addresses both aspects by integrating YOLOv8 and ViT. This holistic approach marks a significant step forward in rice quality monitoring and agricultural applications, offering a scalable, efficient solution that can be adapted to broader agricultural settings. Thus, the research gap identified in the literature lies in the need for a comprehensive and integrated deep learning model that combines the strengths of both object detection and relational feature extraction. By filling this gap, the current study advances the field and provides a novel solution for improving rice classification and quality assessment. Key contributions of the proposed classification method are:

- Innovative Hybrid YOLO-ViT Model for Rice Classification**:** This approach presents a unique combination of YOLO and ViT, harnessing their complementary strengths to improve rice classification accuracy.
- Improved Classification Performance**:** By integrating YOLO's spatial feature extraction with ViT's ability to capture complex relational dependencies, this method significantly enhances the classification of rice varieties, especially when handling intricate and diverse image data.
- Precise Quality Assessment**:** The hybrid YOLO-ViT model goes beyond mere classification, offering precise rice quality evaluation, thus addressing key challenges in agricultural quality control.
- Balanced Performance Across Rice Varieties**:** The method ensures reliable classification and quality assessment for five rice varieties—Tarom, Shiroodi, Fajr, Neda, and Basmati—minimizing biases from data imbalance, which often complicate agricultural datasets.
- Efficient and Scalable Solution**:** This approach optimizes rice classification systems' efficiency, overcoming challenges such as feature representation and model complexity. It is adaptable for broader agricultural applications, making it suitable for large-scale implementation.
- Contribution to Agricultural Monitoring**:** The method provides a scalable solution for rice quality monitoring, enhancing the field of automated rice classification and improving food supply chain management.

This innovative approach plays a crucial role in advancing the field of rice quality assessment and classification, enhancing both the efficiency and accuracy of agricultural production and market supply management.

This paper is structured as follows: Section 2 outlines the step-by-step process of rice classification. Section 3 provides a detailed description of the deep learning models applied in this research. Section 4 presents the experimental findings, confirming the model's effectiveness in real-world agricultural applications. Finally, Section 5 summarizes the key outcomes and emphasizes the method's potential to revolutionize rice grading and quality evaluation.

2. MATERIALS AND METHODS

Rice plays a pivotal role in the Iranian economy, ranking the country among the world's top producers and consumers. As a staple crop, accurate and efficient classification of rice is essential for ensuring its quality

and safety throughout the production and distribution processes. Proper classification supports various stages of the rice value chain—ranging from cultivation to consumption—by enabling precise identification of rice varieties, which in turn improves storage, processing, and handling procedures. Additionally, this process contributes to reducing waste, enhancing farmers' productivity, and providing benefits to all stakeholders in the rice industry. The proposed method involves several key phases, including the creation of an extensive rice image dataset, the application of preprocessing strategies to optimize the accuracy of classification, and the development of a deep learning model. The specifics of each step are discussed in detail in the following sections.

2. 1. Image Database Creation for Rice Classification

Creating an image database for rice classification follows a series of organized steps. Initially, rice samples from various varieties are collected to ensure they represent the key rice types consumed in north of Iran. These varieties include Tarom, Shiroodi, Fajr, Neda, and Basmati. High-resolution images of the rice grains are captured using cameras with optimal lighting conditions to ensure clarity and detail. The images are taken from a consistent top-down perspective, making the rice's distinctive textures and features visible. A specially designed LED light tent, measuring 23cm×23cm×23cm, was constructed to facilitate the accurate capture of the rice grain characteristics. The tent features two LED strips, each 23cm long, positioned strategically along the interior and top to provide uniform lighting and reduce the occurrence of shadows. The upper panel of the tent is fitted with a 5 cm diameter circular aperture, through which the camera, such as a smartphone, is positioned to capture images of the rice samples. The rice samples are evenly distributed on the bottom panel of the tent, and measured using a 50-gram cup to ensure consistency across all images. For imaging, the Sony Imaging Camera with a 300-imx sensor and 19-megapixel resolution was employed to capture the images of the rice grains (23). In the subsequent step, each input image is labeled according to its corresponding rice variety, thereby creating a labeled dataset. A comprehensive collection of more than 10,000 images representing five distinct rice varieties—Tarom, Shiroodi, Fajr, Neda, and Basmati—has been assembled, forming a robust dataset for in-depth analysis. Examples of each rice variety are shown in Figure 1.

2. 2. Data Preparation and Preprocessing for Rice Classification

The preprocessing phase is a pivotal step in the rice classification pipeline, designed to optimize image quality and prepare the dataset for training deep learning models. This stage involves a series of critical operations that ensure the consistency, clarity, and relevance of the data used by the model. First, image resizing is performed to standardize the dimensions of all images, which helps reduce variability and ensures that the model receives input images of consistent size. This process is particularly important for neural networks, as it ensures uniformity and enhances training efficiency. Next, color space transformation is applied, where each image is converted to a specific color space, such as RGB to HSV, to standardize the color representation. This transformation helps mitigate the effects of lighting variations and other environmental factors, ensuring that the model can focus on the inherent characteristics of the rice grains rather than external influences. Noise reduction is another essential step, aimed at removing unwanted disturbances in the images that could interfere with the feature extraction process. The images undergo a median filtering process to remove noise and shadow artifacts, thereby improving the visual quality of the images and ensuring that the deep learning model focuses on relevant features. In this study, images are resized to a uniform resolution of 256×256 pixels, followed by the application of a median filter to further enhance image quality. These preprocessing techniques significantly improve the model's ability to accurately identify and classify the rice varieties by ensuring that the data fed into the model is as clean, uniform, and representative as possible. The rice dataset initially consists of a total of 10,287 images across five different rice cultivars, with 8,228 images designated for training and 2,059 images allocated for testing.

After applying data augmentation, the number of training images for each cultivar is significantly increased, resulting in a total of 33,022 images for training. The following summarizes the number of images before and after augmentation, Tarom: 2,026 total images, with 1,620 for training and 406 for testing. After augmentation, the number of training images increases to 8,100. Shiroodi: 2,114 total images, with 1,691 for training and 423 for testing. After augmentation, the number of training images increases to 8,455. Fajr: 2,053 total images, with 1,642 for training and 411 for testing. After augmentation, the number of training images increases to 8,210. Neda: 2,050 total images, with 1,640 for training and 410 for testing. After augmentation, the number of training images increases to 8,200. Basmati: 2,044 total images, with 1,635 for training and 409 for testing. After augmentation, the number of training images increases to 8,175. After augmentation, the total number of training images across all rice cultivars



(a) (b) (c) (d) (e)

Figure 1. Sample images of rice varieties used in the proposed classification model: a) Tarom rice, b) Shiroodi rice, c) Fajr rice, d) Neda rice, and e) Basmati rice

reaches 33,022, with 8,228 images in the original training set and 24,794 images added through augmentation. The total number of images in the dataset (including training and testing) increases from 10,287 to 35,281. Figure 2 illustrates the distribution of different rice varieties and their respective categories. The bar chart on the left shows the distribution of sample quantities for each rice variety, including Tarom, Shiroodi, Fajr, Neda, and Basmati, clearly displaying the proportion of each variety in the dataset. This chart aids in comparing the frequency of each rice variety. On the right, the pie chart depicts the distribution of rice categories based on quality features and classifications. It represents the percentage of each category relative to the total dataset, providing a detailed view of the quality assessment and different characteristics of the rice samples. To enhance the effectiveness of deep learning models, large datasets are essential; however, collecting and labeling such datasets comes with challenges, including high costs and time demands. Advanced computational infrastructures are also needed to train complex, large-scale models. Increasing data volume can improve model performance and reduce the risk of overfitting. To diversify data, data augmentation techniques can be employed to artificially generate new data, which is especially helpful for classes with limited samples. In this study, data augmentation methods—such as rotation, scaling, flipping, and adding noise—were applied. These techniques increased data diversity and improved the model's ability to generalize to new images and make more accurate classifications.

2. 3. Cross-Validation for Model Assessment K-fold cross-validation is an essential evaluation method for rice classification models, providing a robust measure of the model's generalization to unseen data and mitigating potential training challenges. In this approach, the dataset is divided into multiple subsets (folds), with each fold serving as the validation set once, while the remaining folds are used for training. This rotation through various data partitions offers a more accurate and reliable model performance evaluation. K-fold cross-validation is particularly effective in identifying issues like overfitting and underfitting, offering insights into the model's adaptability across different data conditions. In this study, a 5-fold cross-validation method was applied to assess the performance of the deep learning model. The dataset was split such that 80% was used for training and 20% for validation in each fold. This process was iterated across all five folds, ensuring that each subset served as the validation set once, while the model was trained on the remaining data. After each iteration, the model's performance was evaluated on the validation set, and the final accuracy was calculated by averaging the accuracy scores across all five folds. This comprehensive approach not only minimizes the limitations inherent in model training but also provides deeper insights into the model's generalizability across new datasets.

3. DEEP MODEL ARCHITECTURE FOR RICE CLASSIFICATION AND QUALITY EVALUATION

The architecture of the deep learning models used in this study integrates two powerful techniques—YOLOv8 and ViT—to improve both the classification of rice varieties and the evaluation of their quality. These models are designed to capture a wide range of features and dependencies within the rice images, ensuring high precision and efficiency.

3. 1. YOLOV8 Deep Model for Rapid Classification YOLOv8 is a real-time object detection model known for its speed and accuracy (24, 25). In the context of rice classification, YOLOv8 is used to detect and localize individual rice grains within an image. The architecture of YOLOv8 involves dividing the image into a grid and predicting bounding boxes and class probabilities for each grid cell, all in one forward pass. This enables YOLOv8 to perform simultaneous classification, making it particularly suitable for real-time applications. In this study, YOLOv8 is employed to identify the presence of rice grains and isolate them from the background. The model has been optimized to recognize five specific rice varieties: Tarom, Shiroodi, Fajr, Neda, and Basmati. The speed and accuracy of YOLO's object classification capabilities make it an ideal choice for large-scale rice classification tasks.

3. 2. Vision Transformer for Capturing Spatial Relationships The Vision Transformer model has gained attention for its ability to model complex spatial relationships in images by leveraging transformer architectures, which were originally designed for sequential data (26). Unlike CNNs, which use filters to extract features locally, ViT treats an image as a sequence of patches and applies self-attention mechanisms to capture global dependencies across the entire image (27). In the rice classification task, ViT is integrated into the model to process the intricate patterns and spatial dependencies present in rice grains. By using ViT, the model gains a deeper understanding of the contextual relationships between different parts of the image, improving its ability to classify rice varieties with greater accuracy. ViT is particularly useful for distinguishing subtle differences between similar rice types and assessing quality features such as size, shape, and color variations.

3. 3. Hybrid YOLO-VIT Model for Enhanced Feature Extraction The hybrid YOLO-ViT model combines the strengths of both YOLOv8 and ViT to

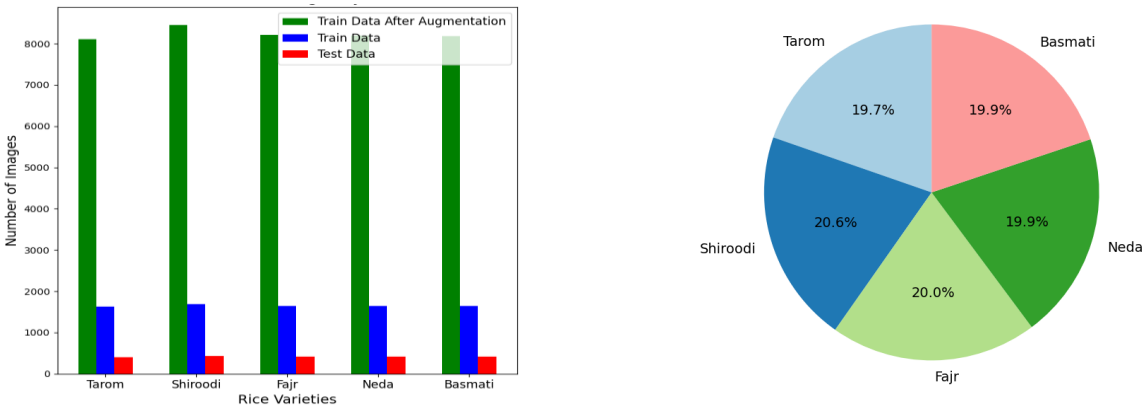


Figure 2. Bar chart of rice varietal distribution and pie chart of rice category distribution

achieve superior performance in rice classification. The hybrid approach utilizes YOLOv8 for rapid and precise classification of rice grains, followed by ViT to capture complex relationships between the grains and their surrounding environment. This integration allows the model to extract both spatial and relational features, leading to a more robust and accurate classification process. In Scenario II, the YOLO-ViT hybrid model is employed to not only detect the rice grains but also analyze their contextual information. The ability to extract both types of features enable the model to recognize rice varieties more effectively and assess their quality with higher precision. For the quality evaluation of rice, the YOLO-ViT hybrid model goes beyond classification to assess the quality attributes of the grains. The quality assessment is based on factors such as the uniformity of the grains, the presence of defects, and the level of adulteration. In Scenario III, the YOLO-ViT model is integrated to handle both classification and quality assessment tasks. The model is trained to identify defects and variations that impact rice quality, such as cracks, discoloration, and impurities, in addition to classifying the rice varieties. The hybrid YOLO-ViT model offers several advantages over traditional deep learning models such as YOLO's fast classification capabilities ensure real-time performance, while ViT enhances the model's ability to handle complex patterns and spatial relationships. Enhanced Classification: The combination of YOLO's classification and ViT's feature extraction improves the classification accuracy of rice varieties, even in challenging conditions with subtle differences. Also, integrating quality assessment capabilities makes the model suitable for practical applications in agricultural quality control and food safety. Overall, the hybrid architecture of YOLOv8 and ViT presents a novel solution for rice classification and quality evaluation, offering both speed and precision in automated agricultural systems. Figure 3 presents the block diagrams of the proposed methodologies based on the fine-tuned architectures of the deep learning models used in each scenario, including YOLOv8, ViT, and the hybrid YOLO-ViT model. Detailed explanations of each model, the network structures, and the additional techniques applied are provided in the subsequent sections. In this study, three distinct algorithmic approaches were designed to address the challenges of rice variety classification and quality assessment. Each scenario employs a unique configuration of deep learning techniques to enhance classification accuracy and robustness: Scenario I uses the Standalone YOLOv8 model, which serves as the foundation of the study. YOLOv8 is employed for rapid and precise classification of rice grains within images. This baseline model offers a straightforward approach to rice classification, focusing on object detection with minimal computational cost. The pseudocode for the Standalone YOLOv8 model is shown in Table 1. Scenario II introduces the YOLO-ViT Hybrid Model. This model combines the fast classification capabilities of YOLOv8 with the ability of ViT to capture complex spatial relationships and dependencies in images. Integrating both models improve the model's capacity to handle intricate patterns and enhances the overall accuracy of rice variety classification. The pseudocode for the YOLO-ViT Hybrid model is shown in Table 2. Scenario III employs the YOLO-ViT Hybrid Model for Classification and Quality Evaluation. In addition to classifying rice varieties, this integrated model assesses the quality of the grains. By combining YOLO's object detection with ViT's feature extraction capabilities, this approach provides a comprehensive solution for both classification and quality control. The pseudocode for this hybrid model is stated in Table 3.

4. RESULTS AND DISCUSSION

In this study, the rice images undergo extensive pre-processing and augmentation to enhance their quality and

ensure the robustness of the classification model. After these steps, the models are trained to classify the rice images into five varieties: Tarom, Shiroodi, Fajr, Neda, and Basmati. Three distinct scenarios are explored to evaluate the performance of different model architectures. Scenario I uses a standalone YOLOv8 model for rapid classification of rice grains in images, focusing on efficient classification without considering complex spatial dependencies. Scenario II combines YOLOv8 with ViT, a hybrid approach designed to capture both spatial and relational features from the images. This combination enables the model to better handle intricate patterns and dependencies, improving classification accuracy. Scenario III integrates the YOLO-ViT hybrid model not only for classification but also for quality evaluation, leveraging ViT's ability to model complex relationships for a more comprehensive assessment of rice quality. To assess the performance of these models, various performance metrics are used, and the models are evaluated under different conditions.

The incorporation of ViT in Scenarios II and III significantly enhances classification accuracy and quality assessment by leveraging ViT's ability to capture long-range dependencies and complex patterns. The results are rigorously validated through a series of experiments, which include hyperparameter optimization and comparisons between the standalone YOLOv8 model and the hybrid models. These evaluations provide a comprehensive understanding of the strengths and limitations of each scenario. Detailed experimental setups and results will be presented in the subsequent sections, demonstrating the effectiveness of the proposed YOLO-ViT approach in improving both precision and efficiency for rice quality evaluation.

4. 1. Experimental Setup And Model Configuration

To evaluate the performance of the proposed rice cultivar classification model, a dataset consisting of over 10,000 color images for each rice variety (Tarom, Shiroodi, Fajr, Neda, and Basmati) was collected, as detailed in Section 2.1. Each image was resized to 256×256 pixels, followed by processing with a median filter to normalize pixel values and reduce noise. This step helps mitigate shadow artifacts and the impact of irrelevant pixels, leading to improved classifier accuracy and overall algorithm performance.

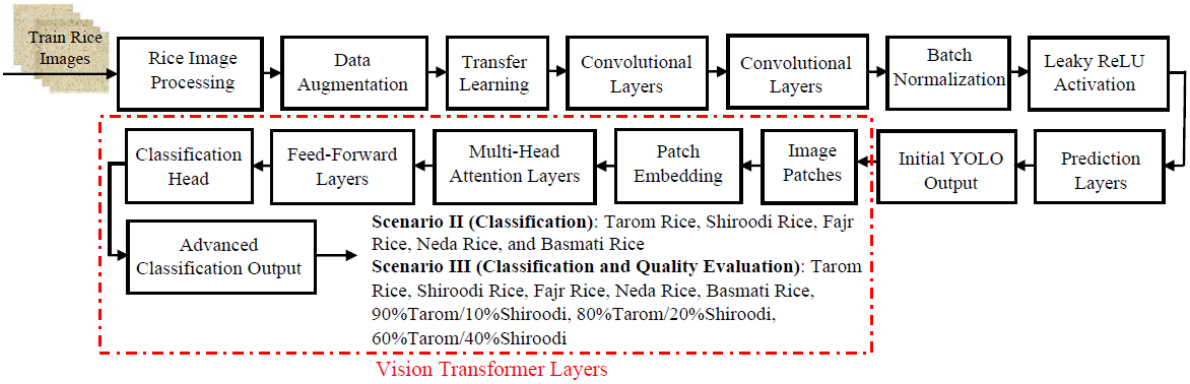


Figure 3. a) Block diagram of Scenario I, illustrating the standalone YOLOv8 model for rice grain classification. b) Scenarios II and III block diagrams show the combined YOLO-ViT model with preprocessing, data augmentation, and transfer learning. Scenario II focuses on feature extraction, while Scenario III includes both classification and quality assessment of rice grains.

TABLE 1. Pseudocode for Scenario I. Standalone YOLOv8: Basic feature extraction for efficient rice variety classification

Scenario I: Baseline YOLOv8 Model for Rice Variety Classification

Input: Rice grain images
Output: Classification label for each rice variety

1. Load rice grain images dataset.
2. Preprocess images:
- Resize images to the input size for YOLOv8.
- Normalize pixel values.

3. Load pre-trained YOLO model.
4. Run the YOLOv8 **model for object detection:**
- Detect rice grains in the image.
- Extract bounding boxes around detected grains.

5. Classify rice varieties:
- Use extracted features from detected grains for classification.

6. Output the classification labels for each rice grain in the image.

TABLE 2. Pseudocode for Scenario II. YOLO-ViT Hybrid Model: improved classification accuracy by combining fast detection and advanced feature extraction

Scenario II: YOLO-ViT Hybrid Model

Input: Rice grain images
Output: Classification label for each rice variety

1. Load rice grain images dataset.
2. Preprocess images:
- Resize images to the required size for both YOLOv8 and ViT.
- Normalize pixel values.

3. Load pre-trained YOLO model.
4. Run the YOLOv8 model for object detection:
- Detect rice grains in the image.
- Extract bounding boxes for detected grains.

5. Load pre-trained ViT model.
6. Extract spatial features using ViT:
- Apply ViT to the detected regions from YOLOv8.
- Capture complex spatial dependencies from the rice grain images.

7. Classify rice varieties:
- Use features extracted by ViT to classify rice varieties.

8. Output the classification labels for each rice grain in the image.

TABLE 3. Pseudocode for Scenario III. YOLO-ViT Hybrid Model for Classification and Quality Evaluation: integrated rice classification and quality assessment through feature extraction and quality metrics

Scenario III: YOLO-ViT Hybrid Model for Classification and Quality Evaluation

Input: Rice grain images
Output: Classification label and quality assessment for each rice variety

1. Load rice grain images dataset.
2. Preprocess images:
- Resize images to the required size for both YOLOv8 and ViT.
- Normalize pixel values.

3. Load pre-trained YOLO model.
4. Run the YOLOv8 model for object detection:
- Detect rice grains in the image.
- Extract bounding boxes around detected grains.

5. Load pre-trained ViT model.
6. Extract spatial features and quality indicators using ViT:
- Apply ViT to detected regions and extract quality-related features (e.g., texture, defects).

7. Classify rice varieties:
- Use features extracted by ViT for classification.

8. Evaluate rice quality:
- Assess rice quality based on extracted quality metrics (e.g., defects, texture).

9. Output the classification labels and quality assessment for each rice grain in the image.

Three distinct deep-learning model scenarios were developed for rice cultivar classification:

Scenario I: In this scenario, the standalone YOLOv8 model is used for detecting and localizing rice grains in images. YOLOv8 is known for its fast and precise object detection capabilities, making it ideal for accurately identifying the position of rice grains in the images without relying on more complex feature extraction.

Scenario II: This scenario introduces a YOLO-ViT hybrid model, combining the strengths of YOLO for spatial feature extraction and ViT for capturing complex spatial relationships and dependencies among image features. YOLOv8 is used for the precise detection of rice grains, while ViT enhances the model's ability to understand intricate patterns and contextual information, leading to better classification performance.

Scenario III: In this scenario, the YOLO-ViT hybrid model is used for both rice classification and quality assessment. This model leverages YOLOv8 for detecting rice grains, ViT for extracting both spatial and relational features, and the integration of both helps in achieving high accuracy in classifying rice varieties and assessing their quality. In summary, these three scenarios progressively increase the model's complexity, from a standalone YOLOv8 model in Scenario I to the more advanced YOLO-ViT hybrid models in Scenarios II and III, enhancing both classification and quality assessment accuracy. The performance of the proposed models is assessed through various parameters, including model architecture, hyperparameters, and preprocessing techniques. A detailed summary of the hyperparameters for each scenario, along with comparisons to traditional classifiers, is provided in Table 4. For example, the batch size was set to 128, meaning the optimizer processes 128 training examples per iteration. Experimental results indicated that larger batch sizes led to higher specific loss and reduced test accuracy, prompting adjustments to the learning rate and batch size during training to optimize training accuracy and convergence rate. In this Table, various hyperparameters are configured for different deep learning models and their respective scenarios. The number of epochs varies depending on the complexity of the model, with Adam and SGD being used as the optimizers. The batch size is consistent at 64 for most models, except for MLP, and ResNet34, 32. The kernel size is set to 3×3 for convolutional layers across all models, while the activation function is mostly ReLU, with exceptions like Leaky ReLU for YOLO (Scenario I) and Sigmoid for MLP. The learning rate ranges from 10^{-3} to 10^{-4}, depending on the model's sensitivity. The weight decay is set to 10^{-4} for most models to prevent overfitting, except for MLP which does not apply weight decay. The number of parameters increases significantly with more complex models, ranging from 1 Million for MLP to 70 million for YOLO-ViT models, indicating the model's size and learning capacity. Finally, the number of layers varies, with the simplest model, MLP, having 4 layers and the more complex YOLO-ViT models containing 24 layers. Table 5 presents the configuration details for each scenario of the YOLO-based deep learning model used in the proposed rice classification and quality evaluation algorithm.

TABLE 4. Hyperparameters for various deep model scenarios in rice classification and quality assessment

	Epochs	Optimizer	Batch Size	Kernel size	Activation function	Learning rate	Weight decay	#of parameters	# of layers
ResNet50 (21, 22)	200	Adam	128	3×3	Sigmoid	10^{-4}	10^{-4}	23M	50
MLP (15)	100	SGD	32	-	Sigmoid	10^{-4}	-	1M	4
Resnet34 (12)	100	Adam	32	3×3	ReLU	10^{-4}	10^{-4}	21M	34
RNN	20	Adam	64	3×3	ReLU	10^{-3}	10^{-4}	2M	10

YOLOv8 (Scenario I)	100	Adam	64	3×3	Leaky ReLU	10^{-3}	10^{-4}	64M	20
YOLO-ViT (Scenario II)	200	Adam	64	3×3	ReLU	10^{-3}	10^{-4}	70M	24
YOLO-ViT (Scenario III)	200	Adam	64	3×3	ReLU	10^{-4}	10^{-4}	70M	24

TABLE 5. Configuration details for each scenario of the YOLO-based deep model in the proposed algorithm (Scenario I: YOLO, Scenario II: YOLO-ViT, Scenario III: YOLO-ViT for classification and quality assessment)

Layers	Name of layer	YOLOv8 (Scenario I)	YOLO-ViT (Scenario II, III)
1	Input Layer	Input Image (256x256)	Input Image (256x256)
2	Convolutional Layer	Conv2D (3x3, 64 filters)	Conv2D (3x3, 64 filters)
3	Max Pooling	MaxPool (2x2)	MaxPool (2x2)
4	Convolutional Layer	Conv2D (3x3, 128 filters)	Conv2D (3x3,128 filters)
5	Max Pooling	MaxPool (2x2)	MaxPool (2x2)
6	YOLO Detection Layer	YOLO Detection (Bounding Box)	YOLO Detection (Bounding Box)
7	Flatten	Flatten	Flatten
8	Transformer Encoder Layer	N/A	ViT Encoder (Multi-Head Attention)
9	Fully Connected Layer	Dense (256 units)	Dense (256 units)
10	Dropout	Dropout (0.5)	Dropout (0.5)
11	Fully Connected Layer	Dense (128 units)	Dense (128 units)
12	Dropout	Dropout (0.5)	Dropout (0.5)
13	Output Layer	Dense (3 units-Softmax)	Dense (3 units-Softmax)
14	Activation	ReLU	ReLU
15	Output Layer	Softmax (Classification)	Softmax (Classification and Quality Detection)

The algorithm presented in this study was developed and trained on advanced computing hardware, featuring an Intel Xeon E5 2600 CPU with 16 cores running at 3.20 GHz. The system is equipped with 3.20 GB of RAM, providing sufficient resources for executing deep learning models efficiently. The training was carried out on a Linux-based machine with 192 GB of DDR4 RAM, which supports handling larger models and datasets while enabling the simultaneous execution of multiple instances of the algorithm. To assess its performance, the proposed algorithm was benchmarked against several deep learning models, including ResNet50 (21, 22), MLP (24), ResNet34 (12), and RNN. Comparisons were also made with algorithms presented by Mavaddati (23) and Mavaddati (28), which use sparse representation techniques and GMM classifiers for rice classification. Mavaddati (23) describes a model that incorporates various features—color, statistical, and texture-based—to accurately represent different rice varieties. This model uses sparse representation and dictionary learning to build over-complete models, alongside sparse principal component analysis (SPCA) and sparse structured principal component analysis (SSPCA) to reduce dimensionality for classifying four rice varieties. The combination of these methods aims to improve both the accuracy and efficiency of the classification process. Additionally, Mavaddati (28) introduces a model-based learning approach for classifying four rice cultivars, which integrates fractal-based and texture features for model construction using a GMM classifier. The use of sparse structured principal component analysis further reduces the dimensionality of the feature vectors, resulting in improved classification accuracy and faster computational time. This approach effectively combines the strengths of fractal and texture features enhancing classification performance.

4. 2. Performance Evaluation of the Proposed YOLO-based Algorithm for Rice Classification and Quality Assessment In complex systems such as deep learning models, evaluating the impact of system benefits and performance is essential to determine their effectiveness and reliability (29). The evaluation of different models and criteria is crucial when

implementing complex systems such as deep learning models for rice classification (30). To evaluate the performance of the proposed CNN-based rice classification algorithm, we utilize several key evaluation metrics: accuracy, specificity, positive prediction rate, sensitivity, and F-measure. Accuracy measures the overall ability of the model to correctly classify rice varieties, while specificity assesses its ability to correctly identify negative instances. The positive prediction rate reflects the proportion of true positive instances among all predicted positives, and sensitivity evaluates the model's capacity to detect relevant rice varieties. The F-measure combines precision and recall into a single metric, offering a balanced evaluation by considering both false positives and false negatives. These metrics are employed to compare the performance of the CNN models across three scenarios: Scenario I (YOLO), Scenario II (YOLO-ViT), and Scenario III (YOLO-ViT for classification and quality assessment), enabling a clear understanding of each model's ability to classify rice varieties and assess their quality. These evaluation measures define as follows:

$$\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Spe} = \frac{TN}{TN+FP} \quad (2)$$

$$\text{Ppr} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Sen} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{F-Measure} = 2 \times \frac{\text{Ppr} \times \text{Sen}}{\text{Ppr} + \text{Sen}} \quad (5)$$

In this context, TP, TN, FP, and FN represent the rates of True Positives, True Negatives, False Positives, and False Negatives, respectively. The True Positive rate (TP) indicates the proportion of positive instances correctly identified as such by the classifier, while the True Negative rate (TN) reflects the proportion of negative instances correctly identified. The False Positive rate (FP) corresponds to the negative instances incorrectly classified as positive, and the False Negative rate (FN) refers to positive instances wrongly labeled as negative. Better classifier performance is indicated by higher TP and TN rates, coupled with lower FP and FN rates. A higher TP shows the classifier's effectiveness in identifying relevant classes, while a higher TN demonstrates its accuracy in rejecting irrelevant ones. A lower FP means fewer negative instances are mistakenly classified as positive, and a reduced FN rate signifies fewer positive instances are missed. These metrics are crucial for comparing classifiers, and achieving high TP and TN rates while minimizing FP and FN is key to ensuring both accuracy and reliability in classification performance. The performance of the proposed algorithm is thoroughly evaluated using key metrics including accuracy, specificity, positive predictive value, sensitivity, and F1-Score, which are presented in Table 6. These metrics offer a robust framework for assessing the model's effectiveness in detecting and classifying various rice attributes. Table 7 displays the confusion matrix for the fine-tuned YOLO-ViT algorithm in Scenario III, applied to the classification of rice cultivars. This matrix provides a detailed overview of the model's performance, showing the true positive, false positive, true negative, and false negative counts for each of the five rice varieties. It highlights the algorithm's ability to correctly classify the rice cultivars and reveals any misclassifications that occurred during the evaluation process.

TABLE 6. Detailed performance evaluation of various rice classifiers based on Accuracy, Specificity, Positive Prediction Rate, Sensitivity, and F-Measure metrics

		Acc (%)	Spe (%)	Ppr (%)	Sen (%)	F-Measure(%)
ResNet50 (21, 22)	Tarom	99.09	98.35	98.59	99.10	98.84
	Shiroodi	98.26	98.66	97.89	98.24	98.06
	Fajr	98.11	98.45	98.44	98.32	98.38
	Neda	98.55	98.70	98.37	98.51	98.44
	Basmati	99.01	99.05	98.91	98.62	98.79
MLP (15)	Tarom	92.28	93.82	93.53	92.29	92.91
	Shiroodi	93.34	94.27	93.64	93.36	93.50
	Fajr	93.95	93.35	93.49	94.08	93.78
	Neda	94.67	94.51	94.63	93.96	94.29
	Basmati	94.91	93.98	94.18	93.87	94.02
ResNet34 (12)	Tarom	96.25	95.98	96.53	96.61	96.60
	Shiroodi	96.16	95.71	95.89	96.68	96.28

	Fajr	97.11	95.69	96.57	96.32	96.44
	Neda	96.81	96.73	96.91	97.10	97.00
	Basmati	96.49	97.22	96.59	95.89	96.24
	Tarom	95.83	96.02	96.75	95.91	96.31
	Shiroodi	95.58	94.13	95.98	96.59	96.36
RNN	Fajr	96.80	96.17	95.85	96.23	95.97
	Neda	95.90	95.94	95.87	96.46	96.10
	Basmati	95.82	96.76	96.71	96.57	96.60
	Tarom	93.76	93.65	94.30	93.19	93.72
	Shiroodi	93.52	93.28	93.27	93.09	93.12
GMM-based (28)	Fajr	93.29	93.58	93.33	94.40	93.86
	Neda	94.12	93.37	93.12	93.85	93.54
	Basmati	94.00	94.06	93.82	93.33	93.47
	Tarom	93.32	93.49	94.27	94.29	94.27
	Shiroodi	93.47	94.69	93.51	94.39	94.06
Dictionary learning-based (23)	Fajr	93.92	93.78	93.49	94.27	93.92
	Neda	94.06	93.46	94.98	94.04	94.53
	Basmati	93.17	93.23	93.38	93.62	93.43
	Tarom	98.27	98.21	98.73	98.54	98.69
	Shiroodi	98.76	98.73	98.03	98.17	98.04
YOLOv8 (Scenario I)	Fajr	99.09	98.59	98.69	98.59	98.68
	Neda	98.69	98.43	97.99	99.06	98.46
	Basmati	98.28	98.61	98.85	98.59	98.67
	Tarom	99.05	99.25	99.19	99.08	99.23
	Shiroodi	99.20	99.11	99.13	99.27	99.39
YOLO-ViT (Scenario II)	Fajr	98.96	98.96	98.93	99.21	99.14
	Neda	99.17	98.86	98.71	99.06	99.17
	Basmati	99.02	99.14	99.09	98.95	99.38
	Tarom	99.74	99.43	99.94	99.61	99.77
	Shiroodi	99.77	99.75	99.74	99.95	99.85
YOLO-ViT (Scenario III)	Fajr	99.94	99.58	99.63	99.68	99.66
	Neda	99.95	99.50	99.94	99.77	99.87
	Basmati	99.71	99.62	99.75	99.94	99.85

TABLE 7. Confusion matrix for the fine-tuned YOLO-ViT (Scenario III) algorithm in classifying rice cultivars

	Tarom (%)	**Shiroodi (%)**	**Fajr (%)**	**Neda (%)**	**Basmati (%)**
Tarom (%)	**99.74**	0.01	0.03	0.14	0.08
Shiroodi (%)	0.06	**99.77**	0.09	0.04	0.04
Fajr (%)	0.03	0.00	**99.94**	0.01	0.02
Neda (%)	0.03	0.01	0.01	**99.95**	0.00
Basmati(%)	0.07	0.09	0.07	0.06	**99.71**

Figure 4 presents the training progress of the proposed YOLO-ViT model in Scenario III, applied to the rice classification task. As mentioned, the proposed YOLO-ViT model combines the unique strengths of both YOLOv8 and ViT architectures, offering a high-precision approach for rice classification and quality assessment. This model stands out due to several key factors. Firstly, YOLOv8 is known for its speed and accuracy in classification, which enables rapid and precise categorization of rice grain varieties. This

efficiency is particularly beneficial in agricultural settings, where large datasets need to be processed quickly and consistently across various conditions. YOLOv8 processes images in real-time, making it well-suited for applications requiring high throughput and reliability.

The addition of ViT in this model enhances its ability to capture intricate spatial dependencies and contextual relationships within the images. For rice quality assessment, this is critical, as rice grains exhibit subtle variations in texture, color, and size that conventional convolutional models may overlook. ViT's attention mechanisms allow the model to focus on significant parts of the image, enhancing the learning of complex, non-linear patterns and dependencies that are vital for accurately distinguishing between different rice types. The study evaluates three different configurations of this model, each demonstrating its capabilities in various classification scenarios. In the first configuration, YOLOv8 is applied independently for basic classification, showcasing its strengths in high-speed and accurate grain categorization. The second configuration introduces a YOLO-ViT hybrid model, where both spatial and relational features are extracted, resulting in improved classification accuracy and enabling more precise quality assessments. The third configuration leverages YOLO-ViT for a comprehensive classification and quality evaluation, achieving the highest accuracy levels by fully utilizing YOLOv8's efficiency and ViT's attention-based feature mapping. This approach proves especially effective in complex classification scenarios, where YOLO's speed is combined with ViT's capability to identify complex interdependencies within the data. The adaptability of the YOLO-ViT model to various conditions makes it broadly applicable in automated agricultural systems. Its high precision and efficiency support enhanced food safety, quality control, and market management, promoting a standardized approach to large-scale rice quality evaluation. The experimental results underscore the model's remarkable performance in classifying and assessing the quality of five distinct rice varieties, Tarom, Shiroodi, Fajr, Neda, and Basmati. Specifically, the third configuration of YOLO-ViT consistently demonstrated superior accuracy, as reflected in precision, specificity, sensitivity, and F-measure metrics that outperformed traditional models such as MLP, CNN, RNN, GMM-based, and dictionary learning-based methods. This robust performance indicates the model's effectiveness not only in distinguishing between rice varieties but also in evaluating each variety's quality with high reliability.

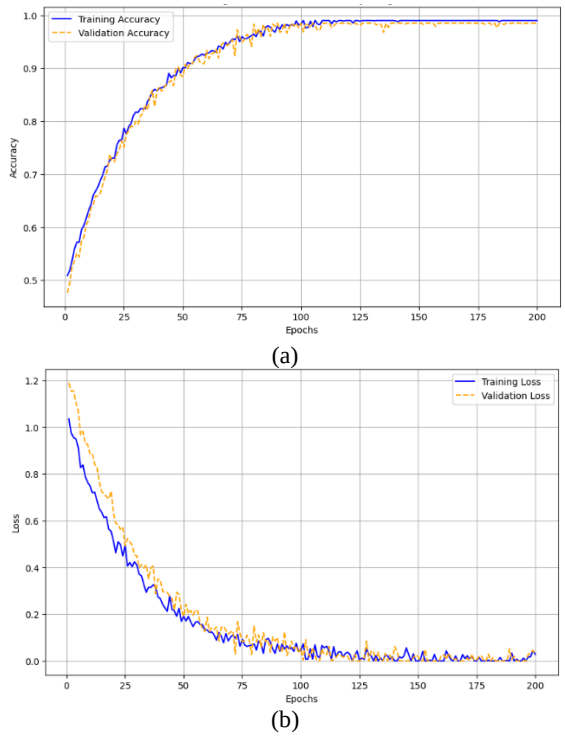


Figure 4. Training progress plots for the proposed YOLO-ViT model in Scenario III applied to the rice classification problem: a) Accuracy plot showing model convergence towards high precision, and b) Loss function plot illustrating the optimization process for minimizing classification errors across the five rice varieties

For each rice variety, high F-measure values, reaching up to 99% in some cases, validate the model's balanced performance across precision and recall metrics. This high level of accuracy across multiple rice varieties, including complex varieties like Basmati, highlights the model's robustness and reliability in rice classification and quality assessment. Additionally, the specificity and sensitivity scores show the YOLO-ViT model's ability to identify fine-grained differences among rice grains accurately, reducing misclassifications and ensuring consistent quality evaluations. Table 8 presents the statistical analysis of different scenarios for rice quality detection, utilizing various detection methods. Table 8 compares the performance of each method across multiple evaluation metrics, providing insights into their effectiveness in accurately identifying rice quality. It offers a comprehensive view of the strengths and limitations of each detection approach, highlighting the impact of different scenarios on model performance. Table 9 presents the confusion matrix for the YOLO-ViT model applied in Scenario III, which evaluates the model's performance in classifying different rice cultivars. The matrix displays TP, FP, TN, and FN values for each rice variety, providing a clear representation of how accurately the model distinguishes between the different cultivars and where misclassifications occur.

The results presented in Tables 5 and 6 collectively demonstrate the high effectiveness of various scenarios in detecting rice adulteration using different

classification methods. Table 6, which shows the confusion matrix for the YOLO-ViT model in Scenario III, illustrates the model's performance in classifying different rice cultivars. Given the model's high accuracy in distinguishing between rice varieties, it can be concluded that the model is highly effective at detecting subtle differences, which is crucial for identifying rice adulteration.

TABLE 8. Statistical comparison of various scenarios for rice quality detection using different detection methods, highlighting the performance metrics and effectiveness of each method across different scenarios

		Acc (%)	ρ-value	Mean ρ-value	Mean Acc (%)
YOLOv8 (Scenario I)	Tarom	98.67	0.005	0.0060	98.39
	Shiroodi	99.04	0.004		
	Fajr	98.61	0.007		
	Neda	98.47	0.004		
	Basmati	98.51	0.006		
	90% Tarom/10%Shiroodi	98.22	0.007		
	80% Tarom/20%Shiroodi	98.41	0.008		
	60% Tarom/40%Shiroodi	98.16	0.007		
YOLO-ViT (Scenario II)	Tarom	99.35	0.005	0.0045	99.32
	Shiroodi	99.49	0.003		
	Fajr	99.35	0.005		
	Neda	99.36	0.003		
	Basmati	99.29	0.004		
	90% Tarom/10%Shiroodi	98.88	0.005		
	80% Tarom/20%Shiroodi	98.89	0.005		
	60% Tarom/40%Shiroodi	98.96	0.006		
YOLO-ViT (Scenario III)	Tarom	**99.73**	**0.002**	**0.0025**	**99.22**
	Shiroodi	**99.81**	**0.002**		
	Fajr	**99.67**	**0.002**		
	Neda	**99.70**	**0.002**		
	Basmati	**99.59**	**0.001**		
	90% Tarom/10%Shiroodi	**98.22**	**0.004**		
	80% Tarom/20%Shiroodi	**98.41**	**0.004**		
	60% Tarom/40%Shiroodi	**98.33**	**0.003**		

TABLE 9. Confusion matrix for the proposed YOLO-ViT model in Scenario III, showcasing the results for different rice cultivars

	Tarom	**Shiroodi**	**Fajr**	**Neda**	**Basmati**	**90%Tarom/ 10%Shiroodi**	**80%Tarom/ 20%Shiroodi**	**60%Tarom/ 40%Shiroodi**
Tarom	**99.73**	0.01	0.03	0.09	0.03	0.04	0.04	0.03
Shiroodi	0.01	**99.80**	0.01	0.01	0.08	0.00	0.03	0.06
Fajr	0.06	0.03	**99.67**	0.05	0.06	0.04	0.04	0.05
Neda	0.05	0.06	0.07	**99.70**	0.04	0.03	0.03	0.02
Basmati	0.03	0.05	0.01	0.06	**99.59**	0.12	0.03	0.11
90%Tarom/10%Shiroodi	0.27	0.26	0.29	0.28	0.26	**98.22**	0.27	0.15
80%Tarom/20%Shiroodi	0.25	0.19	0.30	0.16	0.15	0.28	**98.41**	0.26
60%Tarom/40%Shiroodi	0.26	0.25	0.16	0.17	0.19	0.30	0.34	**98.33**

To further elaborate on the comparisons made in the study, the proposed YOLO-ViT model showcases a significant advancement over traditional and alternative deep learning approaches. Specifically, while standalone models like ResNet50, RNN, and GMM-based methods demonstrate solid performance, they fall short in capturing the nuanced spatial and contextual relationships necessary for high-precision classification. The integration of YOLOv8 and Vision Transformer (ViT) in the proposed model combines the speed and efficiency of YOLO with the sophisticated attention mechanisms of ViT, which focus on intricate patterns within the data. This hybrid configuration, particularly in Scenario III, achieves superior results across all evaluation metrics, as seen in the higher accuracy, sensitivity, specificity, and F-measure values compared to other models. By leveraging ViT's ability to extract complex interdependencies and YOLO's real-time processing capabilities, the model adapts exceptionally well to the diverse characteristics of rice grains, including texture, size, and color variations. These advantages not only enhance classification precision but also improve the robustness and scalability of the approach for large-scale agricultural applications.

5. CONCLUSION

This study introduced a novel approach for rice classification and quality assessment by combining YOLO and ViT models. The aim was to enhance accuracy and efficiency in evaluating rice quality, which is critical for agricultural production and food supply chain management. The hybrid YOLO-ViT model significantly improved the classification of five rice varieties—Tarom, Shiroodi, Fajr, Neda, and Basmati—and quality evaluation. YOLO, known for fast object detection, was combined with ViT's ability to model complex spatial relationships and contextual features, improving the overall analysis of rice quality. The study explored three scenarios: Scenario I used the standalone YOLO model, Scenario II introduced the YOLO-ViT hybrid for feature extraction, and Scenario III combined the hybrid model for classification and quality evaluation. The YOLO-ViT hybrid model achieved the highest accuracy, demonstrating its effectiveness in rice variety classification and quality assessment. This approach provides a comprehensive solution for automated rice quality control, improving precision, efficiency, and food safety in agricultural systems. The proposed model not only classifies rice varieties with high accuracy but also performs quality assessment with remarkable precision and efficiency. The combination of YOLO for fast object detection and ViT for modeling complex spatial relationships enables the model to better understand the texture and structural features of the image. This combination, especially for assessing rice quality and identifying its visual characteristics, can significantly improve the processes of rice production and distribution in the market. Moreover, the application of this model in agriculture and food quality monitoring can enhance food safety and improve supply chain management.

6. FUTURE WORK

In the future, further refinement of the YOLO-ViT hybrid model can be explored by incorporating additional rice varieties and environmental factors that may influence rice quality. Moreover, the model's performance in real-world agricultural settings should be evaluated to ensure its robustness and scalability. Future research can also focus on enhancing the computational efficiency of the model, ensuring faster processing times for large-scale agricultural applications. Additionally, integrating this system with IoT devices and remote sensing technologies could offer real-time monitoring and automation in rice quality control, contributing to the advancement of precision agriculture.

7. ACKNOWLEDGMENTS

This research was financed by a research grant from the University of Mazandaran. The authors appreciate for the financial support provided by the University of Mazandaran.

8. REFERENCES

1. Mohidem NA, Hashim N, Shamsudin R, Che Man H. Rice for food security: Revisiting its production, diversity, rice milling process and nutrient content. Agriculture. 2022;12(6):741. 10.3390/agriculture12010028
2. Tang L, Risalat H, Cao R, Hu Q, Pan X, Hu Y, et al. Food security in China: a brief view of rice production in recent 20 years. Foods. 2022;11(21):3324. 10.3390/foods11213324
3. Bandumula N. Rice production in Asia: Key to global food security. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences. 2018;88:1323-8. 10.1007/s40011-017-0884-0
4. Hamzah AS, Mohamed A. Classification of white rice grain quality using ANN: a review. IAES International Journal of Artificial Intelligence. 2020;9(4):600. 10.11591/ijai.v9.i4.10756
5. Xu J, Wang X, Gao H, Zhai Z, Li R, editors. Rice Identification and Classification System Based on SVM Algorithm. 2022 International Conference on Microwave and Millimeter Wave Technology (ICMMT); 2022: IEEE. 10.1109/ICMMT55580.2022.10023270
6. Filipović V, Panić M, Brdar S, Brkljač B, editors. Significance of morphological features in rice variety classification using

hyperspectral imaging. 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA); 2021: IEEE. 10.1109/ISPA52656.2021.9552086

7. Tarimo SA, Jang M-A, Ngasa EE, Shin HB, Shin H, Woo J. WBC YOLO-ViT: 2 Way-2 stage white blood cell detection and classification with a combination of YOLOv5 and vision transformer. Computers in Biology and Medicine. 2024;169:107875. 10.1016/j.compbiomed.2024.107875
8. Hassan NM, Hamad S, Mahar K. YOLO-based CAD framework with ViT transformer for breast mass detection and classification in CESM and FFDM images. Neural Computing and Applications. 2024;36(12):6467-96. 10.1007/s00521-023-09364-5
9. Jeyaraj PR, Asokan SP, Nadar ERS. Computer-assisted real-time rice variety learning using deep learning network. Rice Science. 2022;29(5):489-98. 10.1016/j.rsci.2022.02.003
10. Pradana-López S, Pérez-Calabuig AM, Rodrigo C, Lozano MA, Cancilla JC, Torrecilla JS. Low requirement imaging enables sensitive and robust rice adulteration quantification via transfer learning. Food Control. 2021;127:108122. 10.1016/j.foodcont.2021.108122
11. Golpour I, Parian JA, Chayjan RA. Identification and Classification of Bulk Paddy, Brown, and White Rice Cultivars with Colour Features Extraction using Image Analysis and Neural Network. Czech Journal of Food Sciences. 2014;32(3). 10.17221/238/2013-CJFS
12. Shrivastava P, Singh K, Pancham A. Classification of grains and quality analysis using deep learning. International Journal of Engineering and Advanced Technology. 2021;11(1):244-50. 10.35940/ijeat. A3213.1011121
13. Abbaspour-Gilandeh Y, Molaee A, Sabzi S, Nabipur N, Shamshirband S, Mosavi A. A combined method of image processing and artificial neural network for the identification of 13 Iranian rice cultivars. Agronomy. 2020;10(1):117. 10.3390/agronomy10010117
14. Deng R, Jiang Y, Tao M, Huang X, Bangura K, Liu C, et al. Deep learning-based automatic detection of productive tillers in rice. Computers and electronics in agriculture. 2020;177:105703. 10.1016/j.compag.2020.105703
15. Pazoki A, Farokhi F, Pazoki Z. Classification of rice grain varieties using two Artificial Neural Networks (MLP and Neuro-Fuzzy). 2014.
16. Chatnuntawech I, Tantisantisom K, Khanchaitit P, Boonkoom T, Bilgic B, Chuangsuwanich E. Rice classification using spatio-spectral deep convolutional neural network. arXiv preprint arXiv:180511491. 2018.
17. Dar RA, Bhat DM, Assad A, Islam ZU, Gulzar W, Yaseen A. Classification of rice grain varieties using deep convolutional neural network architectures. Muzafar and Bhat, Dr Muzafar and Assad, Assif and Islam, Zahir Ul and Gulzar, Wahid and Yaseen, Aamir, Classification of Rice Grain Varieties Using Deep Convolutional Neural Network Architectures. 2022. 10.2139/ssrn.4028123
18. Amit PV, Manjunath V, editors. Convolutional neural network with transfer learning for rice type classification. Proc SPIE; 2018. 10.1117/12.2312113
19. Son NH, Thai-Nghe N, editors. Deep learning for rice quality classification. 2019 international conference on advanced computing and applications (ACOMP); 2019: IEEE. 10.1109/ACOMP.2019.00021
20. Lakshmi V, Seetharaman K, editors. Rice classification and quality analysis using deep neural network. 2022 International Conference on Intelligent Innovations in Engineering and Technology (ICIIET); 2022: IEEE. 10.1109/ICIIET55458.2022.9967584
21. Razavi M, Mavaddati S, Koohi H. ResNet deep models and transfer learning technique for classification and quality detection of rice cultivars. Expert Systems with Applications. 2024;247:123276. 10.1016/j.eswa.2023.123276
22. Razavi M, Mavaddati S, Kobti Z, Koohi H. Rice-ResNet: Rice classification and quality detection by transferred ResNet deep model. Software Impacts. 2024;20:100654. 10.1016/j.simpa.2024.100654
23. Mavaddati S. Rice classification and quality detection based on sparse coding technique. International Journal of Engineering Transactions B: Applications. 2018;31(11):1910-7. 10.5829/ije.2018.31.11b.15
24. Al Mudawi N, Qureshi AM, Abdelhaq M, Alshahrani A, Alazeb A, Alonazi M, et al. Vehicle detection and classification via YOLOv8 and deep belief network over aerial image sequences. Sustainability. 2023;15(19):14597. 10.3390/su151914597
25. Samaniego LA, Peruda SR, Brucal SGE, Yong ED, De Jesus LCM, editors. Image Processing Model for Classification of Stages of Freshness of Bangus using YOLOv8 Algorithm. 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE); 2023: IEEE. 10.1109/GCCE59613.2023.10315381
26. Dey A BS. Shot-ViT: Cricket batting shots classification with Vision Transformer network. International Journal of Engineering Transactions C: Aspects. 2024;37(12):2463-72. 10.5829/ije.2024.37.12c.04
27. Huo Y, Jin K, Cai J, Xiong H, Pang J, editors. Vision transformer (Vit)-based applications in image classification. 2023 IEEE 9th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS); 2023: IEEE. 10.1109/BigDataSecurity-HPSC-IDS58521.2023.00033
28. Mavaddati S. Rice classification with fractal-based features based on sparse structured principal component analysis and Gaussian mixture model. Journal of AI and Data Mining. 2021;9(2):235-44. 10.22044/jadm.2021.9583.2090
29. Hamidi H. A model for impact of organizational project benefits management and its impact on end user. Journal of Organizational and End User Computing (JOEUC). 2017;29(1):51-65. 10.4018/JOEUC.2017010104
30. Hamidi H. A combined fuzzy method for evaluating criteria in enterprise resource planning implementation. Intelligent Systems: Concepts, Methodologies, Tools, and Applications: IGI Global; 2018. p. 639-70.

COPYRIGHTS
©2025 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.

Persian Abstract

چکیده

برنج از نظر ارزش اقتصادی و تغذیه‌ای اهمیت استراتژیکی دارد، به‌طوری که طبقه‌بندی دقیق و ارزیابی کیفیت آن برای تولید کشاورزی و زنجیره‌های تأمین بازار ضروری است. این مطالعه یک روش نوآورانه را معرفی می‌کند که از ترکیب مدل‌های YOLOv8 و Vision Transformer (ویژن ترنسفورمر) استفاده می‌کند تا دقت طبقه‌بندی پنج نوع اصلی برنج (طارم، شیرودی، فجر، ندا و باسماتی) و ارزیابی جامع کیفیت آنها را بهبود بخشد. مدل YOLOv8 امکان شناسایی سریع و دقیق دانه‌های برنج را در تصاویر فراهم می‌کند، درحالی‌که ViT با استخراج روابط فضایی پیچیده و وابستگی‌های میان ویژگی‌های تصویری، توانایی مدل را در مدیریت الگوها و اطلاعات متنی پیچیده ارتقاء می‌دهد. سه سناریو در این پژوهش مورد بررسی قرار گرفته‌اند: سناریوی اول استفاده از مدل YOLOv8 به‌تنهایی، سناریوی دوم به‌کارگیری مدل ترکیبی YOLO-ViT به منظور استخراج ویژگی‌های فضایی و رابطه‌ای، و سناریوی سوم ترکیب مدل YOLO-ViT برای شناسایی و ارزیابی کیفیت هم‌زمان. نتایج نشان می‌دهند که مدل ترکیبی YOLO-ViT دقت طبقه‌بندی و ارزیابی کیفیت را به طور قابل توجهی بهبود می‌بخشد و اثربخشی آن را برای کنترل کیفیت کشاورزی و مدیریت زنجیره تأمین غذایی اثبات می‌کند. این روش از توانایی‌های YOLO در طبقه‌بندی سریع و قابلیت ViT در مدل‌سازی روابط پیچیده بهره می‌برد و راه‌حلی دقیق و کارآمد برای ارزیابی کیفیت برنج ارائه می‌دهد. مدل پیشنهادی که به‌طور گسترده در سیستم‌های کشاورزی خودکار قابل استفاده است، پتانسیل بهبود ایمنی غذایی و مدیریت مؤثر بازار را داراست.