EJFA Emirates Journal of Food and Agriculture

Emirates Journal of Food and Agriculture 38: e173618
doi: 10.3897/ejfa.2026.173618

RESEARCH PAPER FEEDING THE FUTURE

Deep learning paradigms for rice seed classification: a comparative study and novel architectures

Yasir Hamid[1], Sanaa El Yassami[2], Joseph Anajemba[1], Muhammad Atif Chowhan[1], Artur Koci[1]

1 *Information Security Engineering Technology, Abu Dhabi Polytechnic, Abu Dhabi – 111499, United Arab Emirates*
2 *42 Abu Dhabi, Abu Dhabi, United Arab Emirates*

Corresponding author: Yasir Hamid (yasir.hamid@adpoly.ac.ae)

Academic editor: Zeynep Ünal ♦ **Received** 30 September 2025 ♦ **Accepted** 16 April 2026 ♦ **Published** 8 October 2026

Abstract

Automated classification of rice seeds is essential for quality assessment and sorting in modern agriculture. This paper presents a comprehensive study on rice seed classification utilizing deep learning techniques, particularly Convolutional Neural Networks (CNNs). We reviewed existing research and proposed three CNN architectures: Model1 - A basic CNN architecture; Model2 - Integrating transfer learning with MobileNetV2, and Model3 - An advanced version with additional optimization techniques. A novel dataset including 5000 rice seed images from diverse sources was utilized. Pre-processing techniques, including image cropping and resizing, ensured uniform input dimensions for CNN models. Experimental results illustrate the efficacy of the proposed models. Model3 - Incorporating various optimization techniques, achieved a peak accuracy of 95%, surpassing its predecessors. Comparative analysis revealed Model3's higher accuracy, recall, and accuracy scores, highlighting its robustness in accurately classifying rice seed varieties. In conclusion, this study emphasizes the significance of deep learning techniques in rice seed classification, offering promising solutions for automating seed quality assessment and enhancing agricultural practices.

Keywords

Agricultural practices, automated classification, deep learning, rice seed classification, transfer learning

Introduction

Automated classification of rice seeds is fundamental to modern agriculture, facilitating efficient quality assessment and sorting processes (Smith et al. 2019). Traditional methods of seed classification are often labour-intensive and time-consuming, prompting the exploration of advanced technological solutions (Jones and Patel 2018). Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have emerged as promising tools for image classification tasks due to their ability to automatically learn relevant features from datasets processed (LeCun et al. 2015).

In this paper we present a comprehensive study on the classification of rice seeds using deep learning techniques. The objective is to leverage the capabilities of CNNs and to enhance the accuracy and efficiency of seed classification, thereby contributing to the optimization of agricultural practices (Wang et al. 2020). By automating the classification process, farmers and seed producers can streamline operations, improve productivity, and ensure the quality of their seed stock (Chen and Li 2017).

This study begins with a review of existing literature on rice seed classification, highlighting various approaches and methodologies employed in previous research (Zhang and Zhang 2016). These approaches and

Copyright *Hamid, et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

techniques typically involve the extraction and analysis of morphological, colour, texture, and spectral features from rice seed images. While traditional machine learning algorithms have been utilized with moderate success, deep learning methods, especially CNNs, have shown higher performance in this domain (Sun et al. 2018).

This study provides a detailed comparison between traditional machine learning algorithms (KNN, SVM, RF) and advanced deep learning models (CNNs), demonstrating the higher performance of CNNs in classification tasks. The contribution of this research stands on providing a systematic comparison of the performance associated with traditional machine learning algorithms with state-of-the-art deep learning models. this paper proposes three distinct CNN architecture for rice seed classification to assess their effectiveness in classifying rice seeds as follows:

- Model1, a straightforward CNN architecture crafted for simplicity and effectiveness.
- Model2, which incorporates transfer learning methods by leveraging the MobileNetV2 architecture's pre-trained weights, thereby enhancing performance.
- Model3, an advanced architecture that integrates supplementary optimization techniques to augment classification accuracy even further.

The rest of the paper is organized as follows: Section 2 provides an summary of the dataset used in the study and the pre-processing techniques applied to ensure uniformity in input dimensions for the CNN models. Section 3 presents a detailed description of each proposed CNN architecture, including their design principles and implementation details. Section 4 outlines the experimental setup and the analysis of the results obtained, including comparative analysis and evaluation metrics. In section 5 we provide the conclusion and discus future research in the field of rice seed classification using deep learning techniques.

Literature review

Many authors have been proposing different machine learning-based approaches for rice seed classification. The research conducted by Zhang et al. (2017) used a combination of morphological and texture features along with machine learning algorithms such as KNN, SVM, and DT to classify rice seeds. This approach achieved an accuracy of 93.8% on a dataset of 2,000 rice seed images. In another study conducted by Mekhilef et al. (2018), a deep learning-based approach that uses CNN-based model to classify rice seeds based on their morphological and color features was proposed for rice seed classification. This new approach achieved an accuracy of 95.6% on a dataset of 5,000 rice seed images. A study by Liu et al. (2019) proposed a deep learning-based approach for rice seed classification using a stacked autoencoder (SAE) and a deep belief network (DBN). The authors used morphological and color features as inputs to the proposed model. The proposed model achieved an accuracy of 94.5% on a dataset of 1,080 rice seed images.

The research performed by Chen et al. (2019), proposed a DBN-based model for rice seed classification. The authors used morphological and texture features as inputs to the proposed model. The model achieved an accuracy of 92.2% on a dataset of 1,530 rice seed images. Another research for rise classification was conducted by Zhang et al. (2020). The author in his research proposed a deep learning-based model that utilizes a combination of CNN and long short-term memory (LSTM) networks to classify rice seeds. The proposed model achieved an accuracy of 97.50% on a dataset of 800 rice seed images. Another model offered by Nogueira et al. (2020), a CNN-based model using transfer learning was proposed for rice seed classification. The authors used pre-trained models such as VGG16 and ResNet50 as the basis for their proposed model. The proposed model achieved an accuracy of 96.55% on a dataset of 4,600 rice seed images. A study by Yang et al. (2020) proposed a deep learning-based approach for rice seed classification using a CNN and a generative adversarial network (GAN). They used morphological and texture features as inputs to the proposed approach. This new approach achieved an accuracy of 96.94% on a dataset of 1,440 rice seed images. In another study conducted by Xie et al. (2021), CNN-based model was used for rice seed classification using transfer learning. The novelty of this study was the use of pre-trained model Inception-ResNet-v2 as the basis for their proposed model. The accuracy achieved of this model was 98.15% for a dataset of 7,200 rice seed images. Huang et al. (2021) proposed a deep learning-based approach for rice seed classification using a CNN and a recurrent neural network (RNN). This model uses morphological and texture features as inputs to the proposed approach. The model achieved an accuracy of 96.4% on a dataset of 2,400 rice seed images. In 2021 Cao et al. (2021) proposed a machine learning-based approach for rice seed sorting using hyperspectral imaging. This model uses a combination of morphological and spectral features along with machine learning algorithms such as KNN, SVM, and RF to sort rice seeds. The model achieved an accuracy of 97.5% on a dataset of 1,000 rice seed images.

Rice seeds classification has been the focus of different research, and several authors proposed different classification techniques using CNNs and other deep learning methods. For instance, a study conducted by Zhang et al. (2020) proposed a deep learning model that utilizes a combination of CNN and long short-term memory (LSTM) networks to classify rice seeds. This model achieved an accuracy of 97.50%, outperforming traditional machine learning algorithms such as support vector machines

(SVM) and random forest (RF). Identification of rice seed varieties based on Near-Infrared Hyperspectral Imaging Technology combined with Deep Learning is another study that used near-infrared hyperspectral technology (NIR) with machine learning and deep learning methods to identify different rice seed varieties. Deep learning methods have reflected higher accuracy in comparison with traditional machine learning methods. The approach also used the saliency map method to identify the most significant bands for data contribution. Overall, the study demonstrated that NIR hyperspectral imaging combined with deep learning can quickly and accurately distinguish rice seed varieties in a non-destructive manner.

Deep Convolution Neural Network with Weighted Loss to Identify Rice Seeds Vigor Based on Hyperspectral Imaging Under Sample-Imbalanced Conditions is another study that employed a deep convolutional neural network with a weighted loss function for batch and non-destructive detection of rice seed vigor using hyperspectral imaging, while effectively addressing the challenge of sample imbalance. The proposed method achieved a high accuracy compared to conventional models. To examine the feature points, a visualization analysis was conducted. The proposed method can assist in online grading based on seed vigor and other quality attributes in practical production environments.

The agricultural industry is subject of significant transformation. One of the key factors of this transformation is the use of artificial intelligence, especially deep learning methods. The research named Smart Seed Classification System based on MobileNetV2 Architecture has its main focus on seed image analysis using MobileNetV2, a DCNN, for seed classification. This study utilized 14 different seed classes, and the results demonstrated high accuracies of 98% and 95% on the training and test datasets, respectively.

The application of deep learning techniques can help overcome limitations in extracting characteristics and evidence from image data, thereby enhancing productivity and profitability in agriculture. We provide a summary of studies discussed in this section in Table 1, focusing on the approach provided, the features utilized to perform their research, dataset and the accuracy obtained.

In this section we reviewed various studies conducted on the classification and sorting of rice seeds using machine learning and deep learning techniques. The studies have used different features such as morphological, texture, color, and spectral to classify and sort rice seeds. Such proposed approaches have achieved high accuracies ranging from 92.2% to 98.15% on different datasets of rice seed images. Deep learning techniques have generally surpassed traditional machine learning algorithms in rise seeds classification. Additionally, the studies also suggest that the use of deep learning techniques can improve productivity and profitability in agriculture by overcoming limitations in extracting features and information from image data.

Material and methods

The precision of deep learning techniques for sorting and classifying rice seeds largely depends on the datasets and methods used for this classification.

The dataset utilized in this research comprises high-resolution images of rice seeds collected from multiple sources. Morphological, texture, color, and spectral features were extracted using standard image processing techniques. Morphological features include length, width, and shape descriptors, while texture features are derived using methods such as GLCM and LBP. Color features are quantified using histograms in different color spaces (RGB, HSV), and spectral features are obtained through hyperspectral imaging.

The classification is performed using several deep learning algorithms and machine learning techniques. Traditional machine learning classifiers like KNN, SVM, and RF are initially used to set a baseline. These classifiers were trained and tested using a stratified k-fold cross-validation approach to ensure the robustness and generalizability of the models.

Table 1. Literature review summary.

Study	Approach	Features	Dataset	Accuracy
Zhang et al. (2017)	Machine learning	Morphological and texture	2,000 rice seed images	93.8%
Mekhilef et al. (2018)	Deep learning	Morphological and color	5,000 rice seed images	95.6%
Liu et al. (2019)	Deep learning	Morphological and color	1,080 rice seed images	94.5%
Chen et al. (2019)	Deep learning	Morphological and texture	1,530 rice seed images	92.2%
Zhang et al. (2020)	Deep learning	Morphological and texture	800 rice seed images	97.50%
Nogueira et al. (2020)	Deep learning with transfer learning	Pre-trained models	4,600 rice seed images	96.55%
Yang et al. (2020)	Deep learning	Morphological and texture	1,440 rice seed images	96.94%
Xie et al. (2021)	Deep learning with transfer learning	Pre-trained model	7,200 rice seed images	98.15%
Huang et al. (2021)	Deep learning	Morphological and texture	2,400 rice seed images	96.4%
Cao et al. (2021)	Machine learning	Morphological and spectral	1,000 rice seed images	97.5%
Hamid et al. (2022)	Deep Learning	Pre-trained models	14 different classes of seeds	98%
Rice Seed Classification Study	Near-infrared hyperspectral imaging technology with conventional machine learning and deep learning methods	NIR hyperspectral data	5 common types of rice seeds	Most deep learning methods > 95%

Deep learning models, particularly Convolutional Neural Networks (CNNs), are then employed due to their better accuracy in image classification tasks. The CNN architecture is designed to include multiple convolutional layers followed by max-pooling and fully connected layers. Various architectures such as VGG16, ResNet50, and Inception-ResNet-v2 are explored, with transfer learning techniques applied to leverage pre-trained weights and improve classification accuracy. Data augmentation methods such as rotation, scaling, and flipping were applied to improve the robustness of the models.

The models are assessed based on precision, accuracy, recall, and F1-score metrics. The results are then compared to define the utmost efficient model for rice seed classification, highlighting the benefits of deep learning over traditional machine learning methods in this domain.

Dataset

The raw seeds used in this study were obtained from the University of Agricultural Sciences and Technology of Kashmir (SKAUST) which is one of the prominent agricultural universities serving the North-Western Himalayan hill and mountain regions, contributing to the state's efforts to achieve food and nutritional security. Raw image files were generated by scanning rice seeds using a digital scanner with an optical resolution of 2400 dpi and a color depth of 24-bit. A total of 70 rice seeds were placed on the scanner's glass plate to capture images in PNG format, which were subsequently stored on a computer disk. During the scanning process, the raw true color image of the rice seeds was maintained. Example images of the rice species used in the current study are presented in Fig. 1.

Figure 1. Rice Species used for the study.

CNN algorithm requires that all input images should have a constant dimension. Therefore, the images were cropped to a fixed resolution of 300 × 300 pixels per millimeter in daylight using python image library. The dataset is composed of 5000 images in total and is split in training set and test set in the ratio of 8:2. The trainings set contains 4000 images and the test dataset contains 1000 images. The number of images for different seed species is shown in Fig. 2.

Deep learning algorithms

Deep learning algorithms are designed to mimic the function of the human brain and imitate the way humans gain knowledge. Deep Learning models use multi-level structures and consist of many processing layers to predict properties for classification. Such models are widely used for image processing to complete different tasks such as image classification, image segmentation, image recognition, image colorization, and image detection. Deep learning models are highly effective in extracting complex information from images and performing image processing tasks such as noise removal and image-to-image translation.

Deep learning uses artificial neural networks to learn useful representations of features directly from data. Convolutional Neural Networks (ConvNet) are deep learning algorithms that could train datasets with a large set of parameters while significantly reducing the computation time compared to other techniques. Such algorithms are feedforward neural networks with deep structures designed to learn features' spatial hierarchies automatically and adaptively by using a backpropagation technique based on several building blocks, such as input, convolution, pooling, and fully connected and output layers which makes ConvNet extremely efficient for image processing (see Fig. 3). In image classification, the computer analyzes

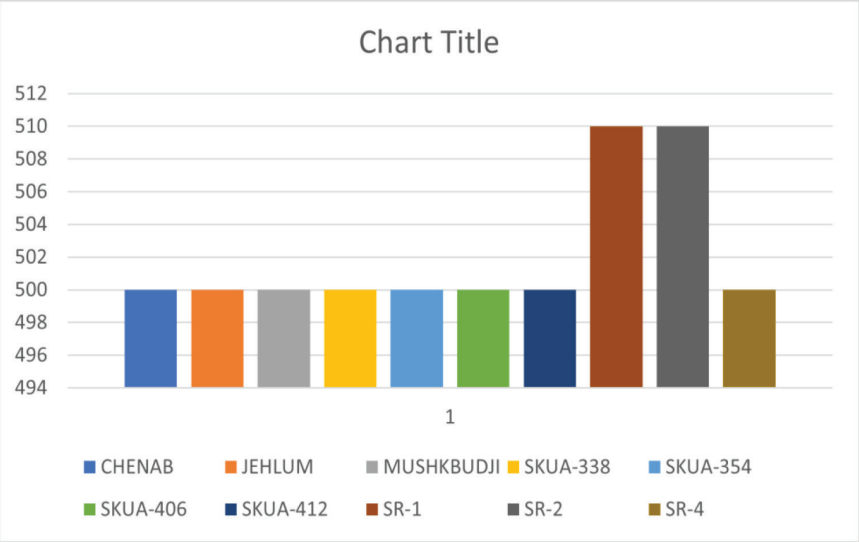

Figure 2. Number of images per rice species.

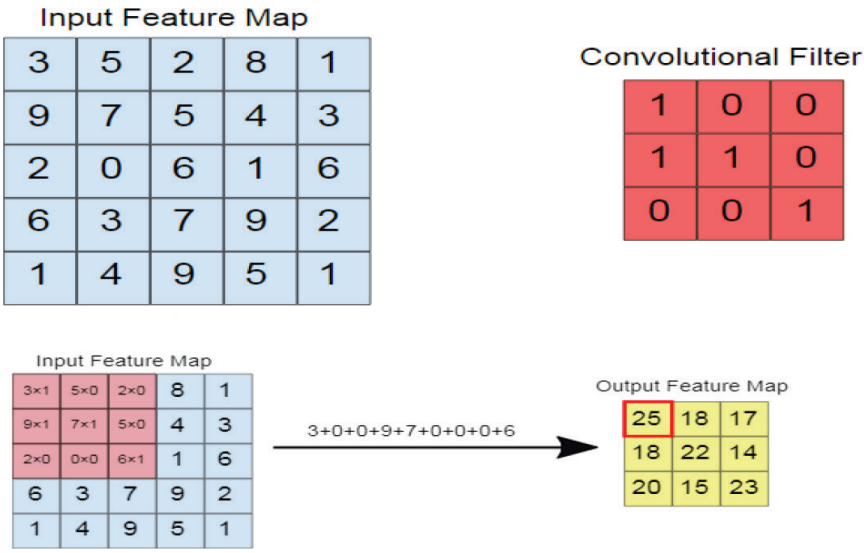

Figure 3. Convolution Layer Process.

an image and identifies the 'class' the image belongs to. A class is basically a label, the labels used in the current study are SKUA-338, SKUA-354, SKUA-406, SKUA-412, SR-1, SR-2, SR-4, CHENAB, JEHLUM, and MUSHKBEDJI.

MobileNet was proposed by Howard et al. (2017) to be used for image classification in embedded and mobile devices. MobileNet is a ConvNet variant designed to do the same convolution as ConvNet but in a different way. This model uses the idea of depth and point convolution which is different from the classical ConvNet as shown in Fig. 4. Incorporating depth convolution and point convolution increases the efficiency of ConvNet in predicting images. MobileNets are based on a streamlined architecture that uses depth-wise separable convolutions to build lightweight deep neural networks. The standard convolutions computational cost is given in Eq. 1.

$$\mathrm{D_K} \times \mathrm{D_K} \times \mathrm{M} \times \mathrm{N} \times \mathrm{D_F} \times \mathrm{D_F} \tag{1}$$

Where $D_K \times D_K$ denotes the kernel size, M represents the number of input channels, N indicates the number of output channels, and $D_F \times D_F$ corresponds to the feature map size.

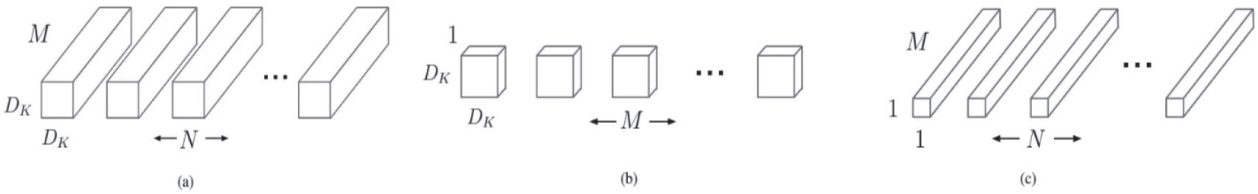

Figure 4. MobileNet Convolution Layer. **a**. Standard convolution; **b**. Depth wise convolution; **c**. Pointwise convolution.

Depthwise separable convolution processes the input channels individually and then combines them to generate new features through a linear combination using convolution. It consists of two main operations: depthwise convolution (see Eq. 2) and pointwise convolution (see Eq. 3). The total computational cost of depthwise separable convolution is calculated as the sum of the computational costs of the depthwise and pointwise convolution operations (see Eq. 4).

$$\mathrm{D_K} \times \mathrm{D_K} \times \mathrm{M} \times \mathrm{D_F} \times \mathrm{D_F} \tag{2}$$

$$\mathrm{M} \times \mathrm{N} \times \mathrm{D_F} \times \mathrm{D_F} \tag{3}$$

$$\mathrm{D_K} \times \mathrm{D_K} \times \mathrm{M} \times \mathrm{D_F} \times \mathrm{D_F} \times \mathrm{M} \times \mathrm{N} \times \mathrm{D_F} \times \mathrm{D_F} \tag{4}$$

MobileNet employs depthwise separable convolutions to decouple the relationship between the number of output channels N and the kernel size $D_K \times D_K$.

MobileNet-v2 is very similar to the original MobileNet that uses the depthwise separable convolution, except it introduces two new features. The first feature is the linear bottlenecks between the layers that encode the model's intermediate inputs and outputs, and the second feature is the shortcut connections between the bottlenecks that enable faster training and better accuracy compared to the traditional residual connection. MobileNet-v2 was proposed in 2018 and has fewer parameters than the original MobileNet. This new version uses a lightweight depthwise convolution that allows better propagation of the gradients.

Evaluation metrics

Evaluation metrics are essential to assess the performance of proposed approaches. Several evaluation metrics have been used over the years and the most significant evaluation for machine learning-based models metric is accuracy. However, to compare the performance of classification models comprehensively, various evaluation metrics can be used including Precision, Recall, and F1-Score. These metrics are calculated based on four sub-metrics that are true positive (t_p), false positive (f_p), true negative (t_n), and false negative (f_n). Our classification model predicts the class of each instance of the dataset by assigning a predicted label to each sample. In our classification model, each sample falls in one of the 10 possible labels such as: SKUA-338, SKUA-354, SKUA-406, SKUA-412, SR-1, SR-2, SR-4, CHENAB, JEHLUM, MUSHKBEDJI.

In this research, to evaluate the effectiveness and performance of the proposed models, we employed three statistical measures based on confusion matrices. The used metrics are F1-Score (Eq. 5), Recall (Eq. 6), and Precision (Eq. 7).

$$\mathrm{F1\ \ Score} = 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{5}$$

$$\mathrm{Recall} = \frac{\mathrm{t_p}}{\mathrm{t_p} + \mathrm{f_n}} \tag{6}$$

$$\mathrm{Precision} = \frac{\mathrm{t_p}}{\mathrm{t_p} + \mathrm{f_p}} \tag{7}$$

Where the Precision represents the positive predictive values and can be interpreted as the ability of a model to not label a sample as positive when it is actually negative. Whereas the Recall represents the true negative rate and can be interpreted as the ability of a model to correctly label all positive samples. The F1-Score represents the harmonic mean of the precision and recall; thus, precision and recall are contributing relatively equally to the F1-Score.

Computational details and architecture

The architecture of our model comprises 10 convolution layers with a filter size of (5 × 5 × 16) for the first two layers, a filter size of (3 × 3 × 32) for the third and fourth layer, a filter size of (3 × 3 × 64) for layers five and six, filter size of (3 × 3 × 128) for layers seven, eight and nine and a filter size of (3 × 3 × 192) with a stride of 1 for layer ten. Each convolution layer is followed by batch normalization, ReLU activation is followed by max pooling layers of size (2 × 2). The convolution layers are followed by a fully connected layer of size (192). A dropout layer with a probability of 0.5 was used to avoid overfitting. Lastly, output layer uses SoftMax activation. The optimizer used is ADAM with a learning rate of 0.00002. Table 2 provides an overview of the proposed model architecture.

The three CNN architectures are modified and improved as follows: Model1: MobileNetV2 is a pre-trained deep learning model that has been fine-tuned by removing the last layer and adding a new layer for classification.

Model2: The MobileNetV2 model has been further fine-tuned using transfer learning by adding 5 different heads. The head model consists of an AveragePooling2D layer, a Flatten layer, a Dense layer with ReLU activation, and a final Dense layer with softmax activation for classification.

Model3: It is an advanced version of MobileNetV2 that includes several additional techniques for improving performance. The techniques used include Model Checkpointing to save the best model during training, Decayed Learning Rate to reduce the learning rate as the training progresses, Data Augmentation to increase the diversity of the training data, Training the model separately to prevent overfitting, and Dropout to reduce overfitting.

Research flow

The operational flow of the research starts with loading MobileNetV2 model with pre-trained weights, setting a solid foundation for further modifications. Ath this stage, three distinct models are defined: Model1, Model2, and Model3. For Model1, fine-tuning involves removing the last layer and adding a new classification layer tailored to the specific dataset. Model2 is enhanced through transfer learning by integrating five different heads into the MobileNetV2 architecture, thereby increasing its adaptability and learning capacity. Model3 employs a series of advanced techniques to boost performance. These techniques include model Checkpointing to save the best-performing model, Decayed Learning Rate to optimize the learning process, Data Augmentation to increase the diversity of the training data, separate training of the model components, and Dropout to prevent overfitting. Subsequently, all three models are trained and evaluated using a rice seed image dataset comprising 10 classes. Their performances are compared using various metrics such as accuracy, loss, validation accuracy, and validation loss to determine which model yields the best results. Research flow is exhibited in the Fig. 5.

Table 2. Architecture of the proposed model.

Layer (type)	Output Shape	Param #
Conv2d_2 (Conv2D)	(None, 128, 128, 32)	896
Activation_1 (Activation)	(None, 128, 128, 32)	0
Max_pooling2d_1 (Maxpooling2)	(None, 64, 64, 32)	0
Conv2d_3 (Conv2D)	(None, 64, 64, 64)	8256
Activation_2 (Activation)	(None, 64, 64, 64)	0
Max_pooling2d_2 (Maxpooling2)	(None, 64, 32, 32)	0
Flatten_1 (Flatten)	(None, 65536)	0
Dense_1 (Dense)	(None, 256)	16,777,472
Activation_3 (Activation)	(None, 256)	0
Dropout_1 (Dropout)	(None, 256)	0
Dense_2 (Dense)	(None, 10)	2,570
Total params: 16,789,194		
Trainable params: 16,789,194		
Non-trainable params: 0		

Results and discussion

Figs 6–9 describes the performance of model-1 on the dataset. It's evident from the graphs in Fig. 1 that the model's performance is notably lacking in maintaining high accuracy. Initially, the model achieved an accuracy of merely 20%, which peaked at a modest 0.6 after 70 iterations, remaining constant thereafter. This represents a firm contrast to acceptable performance standards. This trend is mirrored in the validation accuracy and loss displayed in Figs 4, 5. Analyzing these visual representations, it's evident that model-1 has failed to converge effectively on this dataset, with a performance ratio far below acceptable thresholds, signaling a significant gap between observed and desirable outcomes.

As described in Fig. 6A–D present the performance visualizations for Model2, showing significant improvements compared to Model1. In contrast to Model1, Model2 presents enhanced training and validation accuracies, both peaking at a result of 88%. This substantial increase in accuracy is indicative of the improvements implemented with Model2. Additionally, the validation loss for this model has seen a significant decrease, reaching a minimum of 0.8, a significant improvement over the previous model's performance. Notably, a validation loss rate of 0.1 is considered acceptable for model performance. Moreover, by iteration 70, Model2 achieved an accuracy rate of 70%, maintaining consistency thereafter. Comparison of this performance with the performance of the Model1, it underlines the substantial improvements achieved in Model2, emphasizing its superior performance and efficacy in handling the dataset.

Fig. 7A–D shows the visual representation of Model3 performance. As can be seen from the figures, Model3 demonstrates significant performance improvement, boasting a peak training accuracy of 95% and validation

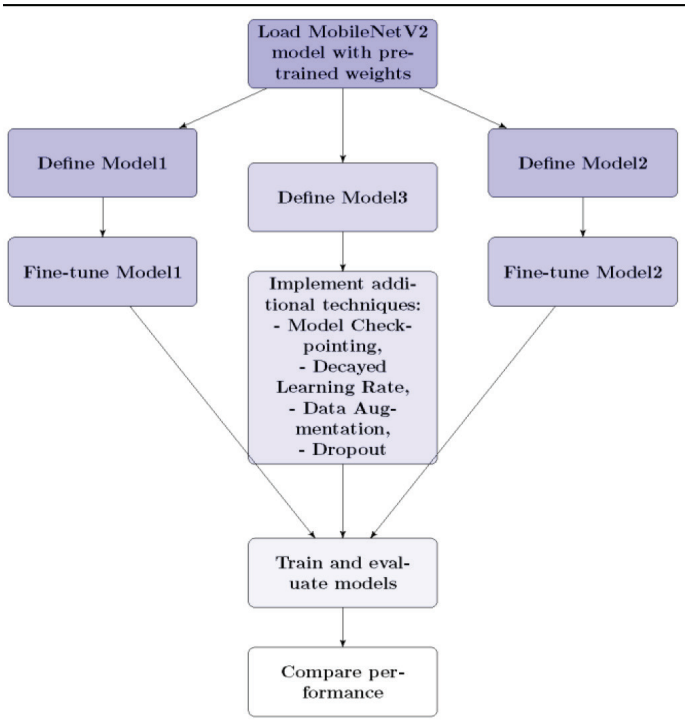

Figure 5. Research flow.

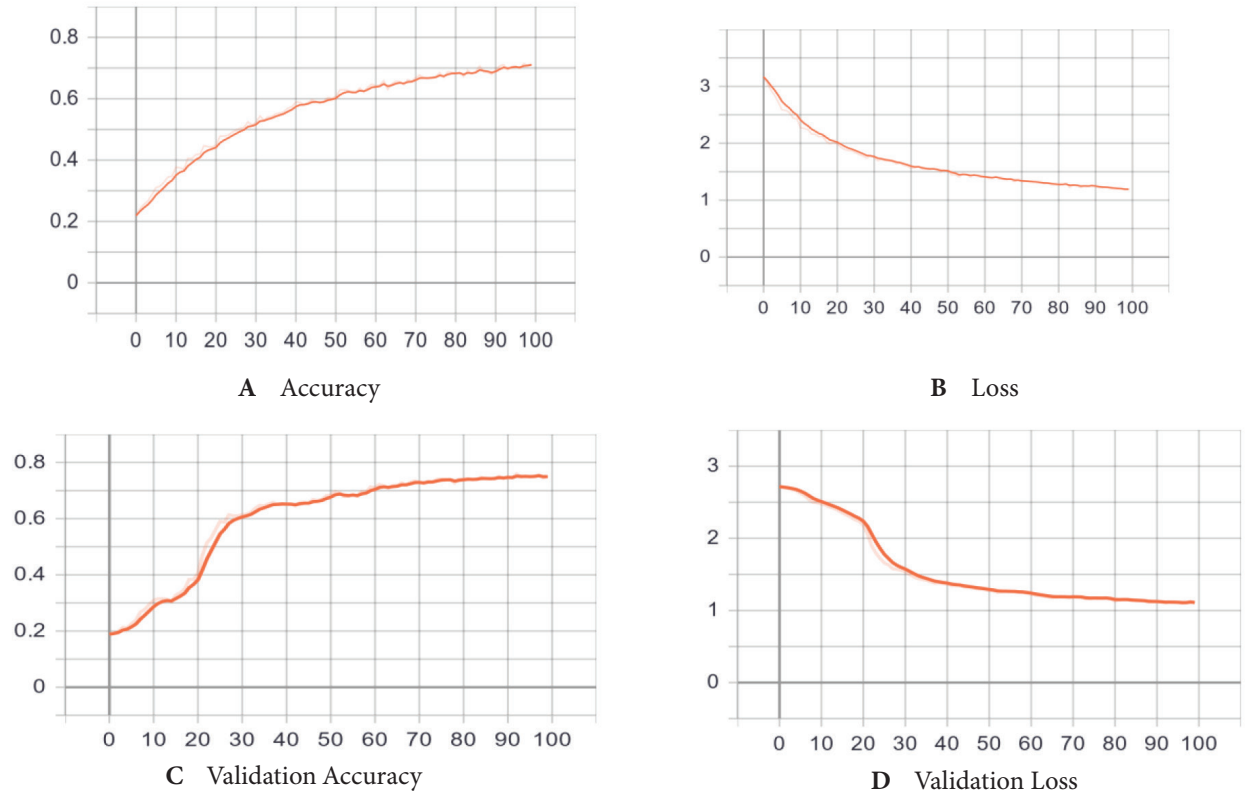

Figure 6. Model1.

accuracy of 92%. The training and validation losses are also significantly low, standing at 0.6 and 0.5 respectively. The performance trajectory of Model3 presents a stable progression, beginning with an initial accuracy of 56% which steadily increases until reaching its peak around iteration 80, after which it experiences a slight decline. Notably, a striking deviation is observed in the validation set's performance, with accuracy steadily increasing from iteration 10 onwards, reaching a commendable 90% and maintaining consistency thereafter.

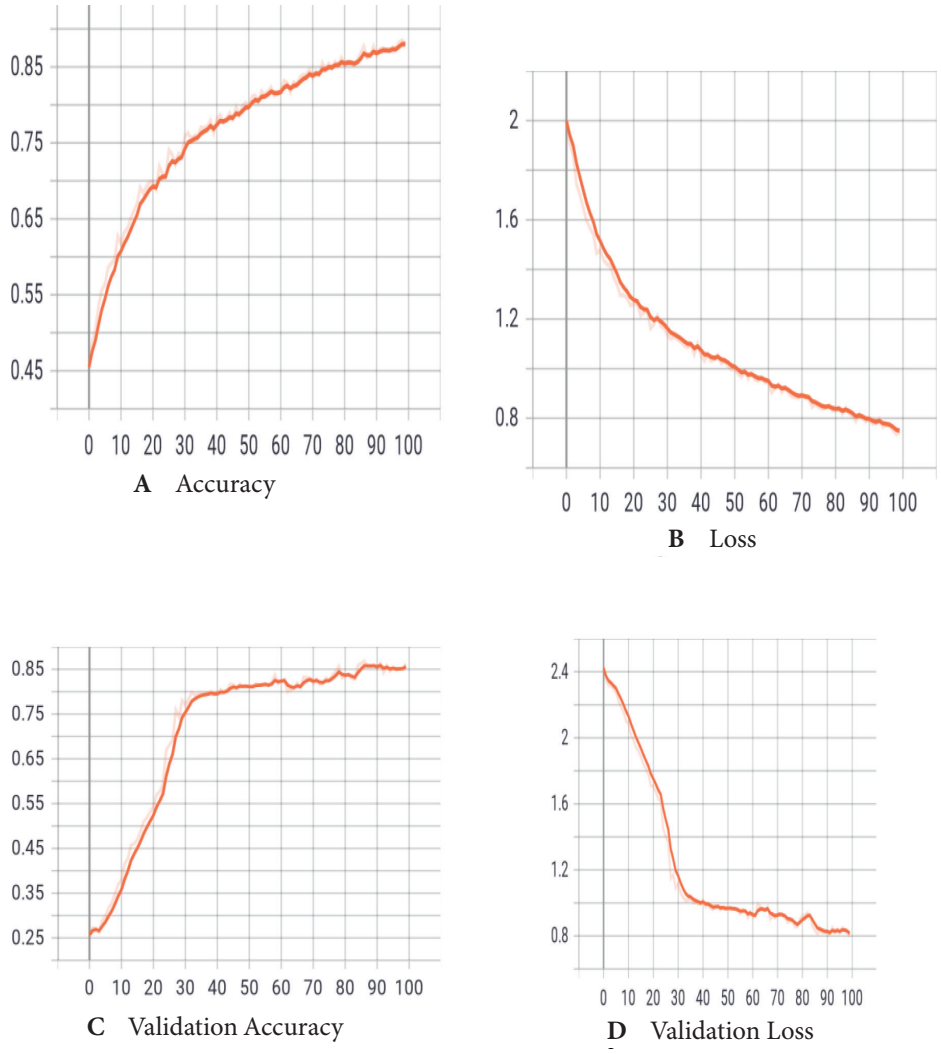

Figure 7. Model2.

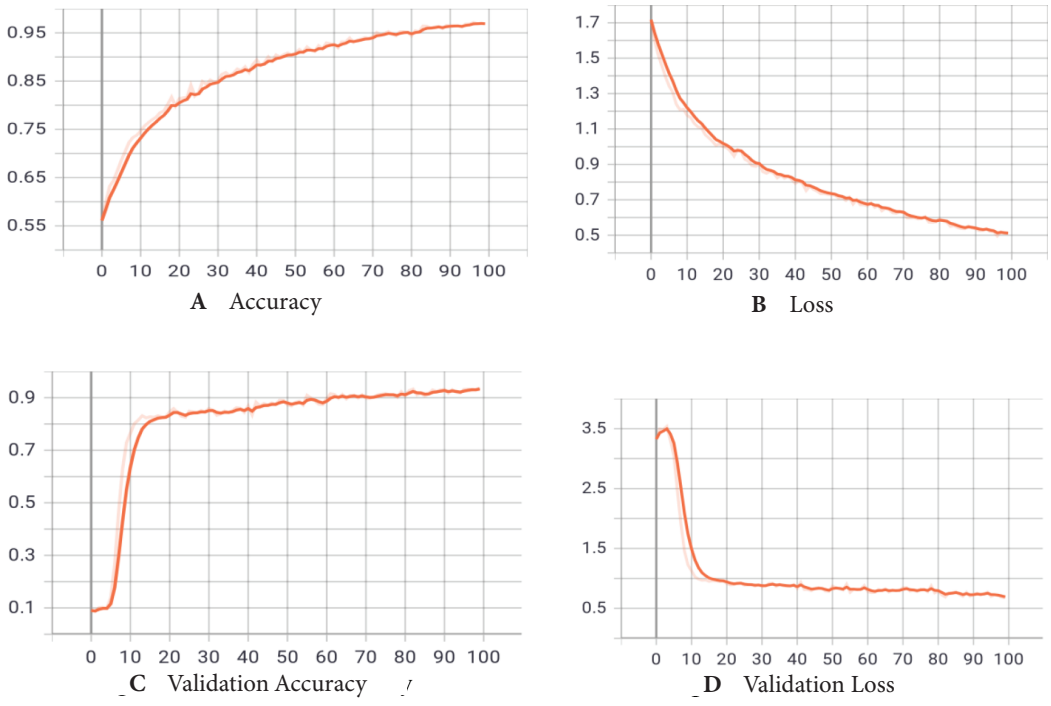

Figure 8. Model3.

Analyzing the performance for the three models, the Model3 demonstrates a significant improvement compared to Model1 and Model2. Model3 outperforms both predecessors by a considerable margin, demonstrating higher accuracy rates and lower loss values. Specifically, compared to Model1, which achieved an initial accuracy of 20% and peaked at 0.6 accuracy by iteration 70, Model3's performance is notably superior, with initial accuracy at 56% and peaking at 95% accuracy in training. Similarly, compared to Model2's peak training accuracy of 88%, Model3's performance exceeds expectations, achieving a remarkable 95% accuracy.

Fig. 9 given below shows the performance results of three machine learning models, Model1, Model2, and Model3, in terms of precision, recall, and accuracy. Model1 has a precision of 0.75, recall of 0.75, and accuracy of 0.75. This means that when the model makes a positive prediction, 75% of the time it is correct, and when the model makes a prediction, 75% of the actual positive cases are correctly identified. The accuracy of the model is also 75% which means that in 75% of the cases, the model makes the correct prediction.

Model2 has a precision of 0.87, recall of 0.86, and accuracy of 0.86. This means that when the model makes a positive prediction, 87% of the time it is correct, and when the model makes a prediction, 86% of the actual positive cases are correctly identified. The accuracy of the model is also 86% which means that in 86% of the cases, the model makes the correct prediction. Model3 has a precision of 0.94, recall of 0.94, and accuracy of 0.94, indicating that when the model makes a positive prediction, 94% of the time it is correct, and when the model makes a prediction, 94% of the actual positive cases are correctly identified. The accuracy of the model is also 94% which means that in 94% of the cases, the model makes the correct prediction. Based on the results, Model3 has the highest precision, recall, and accuracy scores compared to Model1 and Model2. The precision, recall and accuracy are all 0.94 for Model3, while they are 0.75, 0.75 and 0.75 for Model1 and 0.87, 0.86, and 0.86 for Model2 respectively. This means that Model3 has the highest percentage of correctly predicted positive cases and the lowest percentage of incorrectly predicted positive cases, while also having the highest overall percentage of correctly predicted cases.

Additionally, Model3 has the highest recall score, which means it correctly identifies the highest percentage of actual positive cases when compared to the other two models. High recall is important when the goal is to identify as many actual positive cases as possible, such as in medical diagnosis or fraud detection. Furthermore, Model3 has the highest precision score, which means it has the lowest percentage of incorrectly predicted positive cases when compared to the other two models. Overall, Model3's results show a strong performance in both precision and recall, and a high accuracy. This indicates that Model3 can correctly identify a high percentage of actual positive cases while also minimizing the number of false positives. This combined performance makes Model3 the best model among the three, in terms of its ability to generalize well and make accurate predictions on unseen data.

Table 3. given below provides the summary of the results of the three models.

Table 3. Result of the three models.

Model	Precision	Recall	F1-Score	Accuracy
Model1	0.75	0.75	0.75	0.75
Model2	0.87	0.86	0.86	0.86
Model3	0.94	0.94	0.94	0.94

Fig. 10 shows the precision results of three machine learning models, Model1, Model2, and Model3. The precision values range from 0.6 to 1, with Model3 having the highest overall precision values. Model1's precision values range from 0.6 to 0.99, Model2's range from 0.58 to 0.97, and Model3's range from 0.9 to 1. From the fig. 13, Model3 has the highest precision values overall, with all its values being above 0.9 and several being 1. Model2 also has mostly high precision values, but with a few slightly lower values. Model1 has the lowest overall precision values, with several falling below 0.8. It appears that Model3 consistently performs better than the other two models, while Model1's performance is more inconsistent.

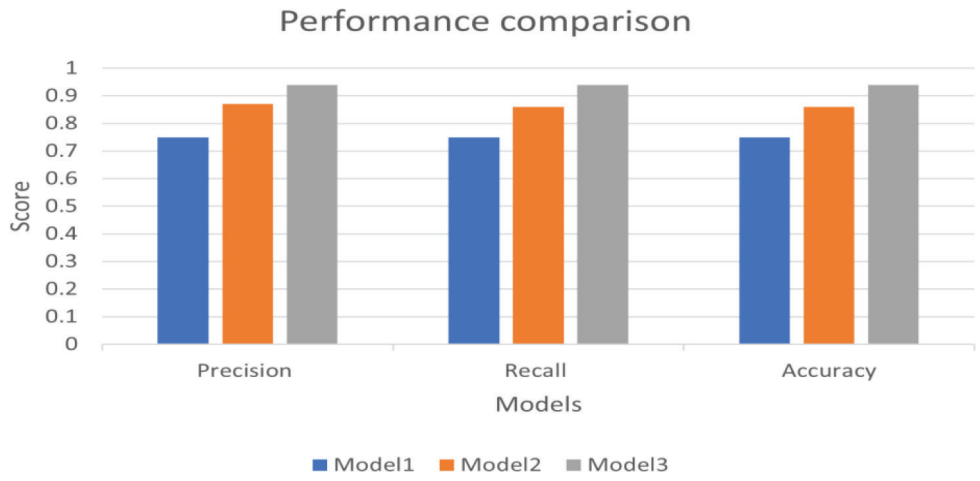

Figure 9. Performance comparison of three proposed models.

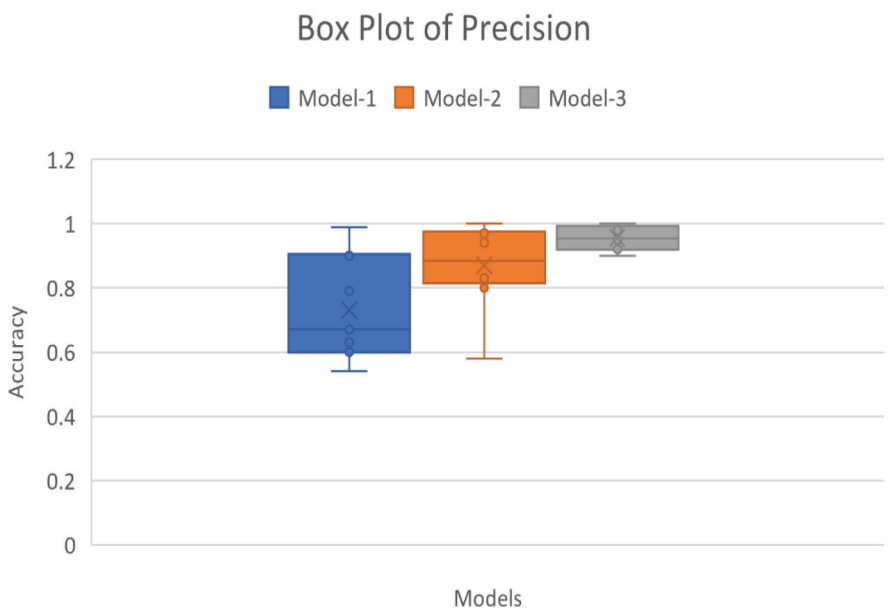

Figure 10. Precision.

Fig. 11 shows the recall results of three machine learning models, Model1, Model2 and Model3, on a dataset with 10 classes. The recall values range from 0.44 to 1, with Model1 and Model3 having the highest overall recall values. Model1's recall value ranges from 0.44 to 0.99, Model2's range from 0.55 to 0.87, and Model3's range from 0.8 to 1.

From the data, it appears that Model1 and Model3 have the highest recall values overall, with all of their values being above 0.8 and several being 1. Model2 has mostly moderate recall values, with a few slightly lower values. Model3 consistently performs better than Model2 in terms of recall, while Model1 has a more consistent performance across the dataset.

It's also worth emphasizing that, even though Model3 has the highest recall scores, it might have more false positives. On the other hand, Model1 has a bit lower recall score, but it might have less false positives.

Quantitative analysis of the results

Model1 demonstrated limited convergence behavior, with training accuracy initially at 20% and peaking at approximately 60% after 70 epochs. Validation performance followed a similar trend, indicating underfitting and insufficient feature representation capability.

Model2, which incorporated transfer learning through MobileNetV2, showed substantial improvement. Training and validation accuracy both increased to approximately 88%, demonstrating the benefit of leveraging pretrained feature extractors. Validation loss decreased significantly compared to Model1, indicating better generalization.

Model3, the optimized architecture, achieved the highest performance. Training accuracy reached 95%, with validation accuracy stabilizing at 92%. Training and validation losses were reduced to 0.6 and 0.5 respectively.

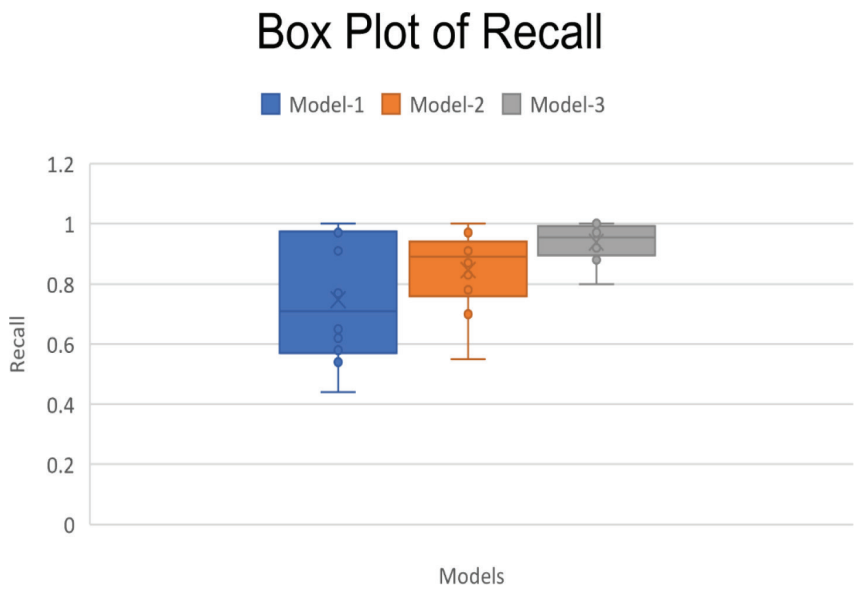

Figure 11. Recall.

The learning curves demonstrate stable convergence and reduced overfitting due to the integration of data augmentation, learning rate decay, dropout regularization, and checkpointing.

Table 3 summarizes the overall performance comparison among the three models.

Discussion

The results show that transfer learning and architectural optimization improve rice seed classification performance. Model1 produced the lowest accuracy, indicating that a basic CNN struggles to learn sufficiently discriminative features across similar rice varieties. Model2 achieved better performance by leveraging MobileNetV2 pre-trained weights, which improved feature extraction and training stability. Model3 delivered the best results, showing that adding optimization techniques (e.g., augmentation, dropout, and learning rate scheduling) can further enhance generalization. Overall, the achieved accuracy is competitive with prior studies while relying only on standard RGB images, making the proposed approach practical for real-world agricultural applications.

Conclusion

This research shows a significant improvement in rice seed classification accuracy when using deep learning techniques, particularly Convolutional Neural Networks (CNNs), over traditional machine learning methods. Traditional algorithms like KNN, SVM, and RF provided a useful baseline, however, CNNs, enhanced by transfer learning and data augmentation, consistently outperformed them in terms of accuracy, precision, recall, and F1-score. This research emphasizes the potential of advanced deep learning models in agricultural applications, suggesting that future work could further optimize these models and apply them in real-time systems for automated seed sorting. Further optimization of such models promises to enhance efficiency and accuracy in agricultural processes, contributing to better crop management and quality control.

Future research could focus on integrating more sophisticated deep learning models and exploring potentials methods that combine multiple techniques to further improve classification performance. Additionally, using larger and more diverse datasets can help enhance the generalizability of these models. Real-time implementation of these models in automated systems for seed sorting could be investigated to provide practical solutions for the agricultural industry.

References

Cao M, Zhou Y, Wang Y (2021) A machine learning-based rice seed sorting system using hyperspectral imaging. Journal of Spectroscopy 2021: 5582648. https://doi.org/10.1155/2021/5582648

Chen J, Li H (2017) Convolutional neural networks for rice seed classification. IEEE Transactions on Agricultural Imaging 3(2): 89–97.

Chen X, Li Q, Chen H (2019) Rice seed classification based on deep belief network. International Journal of Agricultural and Biological Engineering 12(3): 128–135.

Hamid Y, Wani S, Soomro AB, Alwan AA, Gulzar Y (2022) Smart seed classification system based on MobileNetV2 architecture. In: 2022 2nd international conference on computing and information technology (iccit). IEEE Xplore, 217-222. https://doi.org/10.1109/ICCIT52419.2022.9711662

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv, 1704.04861.

Huang L, Liu J, Zhang L (2021) Rice seed classification based on convolutional neural network and recurrent neural network. Journal of Physics: Conference Series 1865(1): 012102. https://doi.org/10.1088/1742-6596/1865/1/012102

Jones R, Patel S (2018) Automation in agriculture: challenges and opportunities. Agricultural Engineering Today 21(3): 45–56.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553): 436–444. https://doi.org/10.1038/nature14539

Liu X, Wang J, He Y, Chen G, Zhang Y (2019) Rice seed classification based on stacked autoencoder and deep belief network. Journal of Food Engineering 260: 1–8. https://doi.org/10.1016/j.jfoodeng.2019.05.008

Mekhilef R, Abdullah-Al-Wadud M, Shanto M (2018) A deep learning-based approach for rice seed classification. Computers and Electronics in Agriculture 152: 312–319. https://doi.org/10.1016/j.compag.2018.07.023

Nogueira AB, Andrade EL, Lopes FA, Silva FM (2020) Deep learning for rice seed classification: A comparison between VGG16 and ResNet50. Computers and Electronics in Agriculture 176: 105608. https://doi.org/10.1016/j.compag.2020.105608

Smith A, Johnson B, Williams C (2019) Advances in rice seed classification: a review. Journal of Agricultural Technology 14(2): 87–102.

Sun Q, Wang Z, Li J (2018) Comparative analysis of machine learning and deep learning methods for rice seed classification. International Journal of Computer Applications 167(7): 15–20.

Wang L, Liu K, Zhang M (2020) Deep learning applications in agriculture: a comprehensive review. Computers and Electronics in Agriculture 176: 105650.

Xie X, Yang J, Liao Q, Zhou W (2021) Rice seed classification based on transfer learning with Inception-ResNet-v2. IEEE Access 9: 13966–13975. https://doi.org/10.1109/ACCESS.2021.3055421

Yang Y, Li X, Li Y, Li W (2020) Rice seed classification based on convolutional neural network and generative adversarial network. Journal of Food Engineering 288: 110151. https://doi.org/10.1016/j.jfoodeng.2020.110151

Zhang X, Zhang Y (2016) Feature extraction techniques for rice seed classification: a comparative study. Journal of Agricultural Science and Technology 14(4): 287–301.

Zhang Y, Hu H, Shi Z, Liu Y, Zhang C, Yu J (2017) A machine learning-based approach for rice seed classification. Computers and Electronics in Agriculture 140: 21–28. https://doi.org/10.1016/j.compag.2017.05.010

Zhang Y, Ma Y, Wu Q, Chen Y, Zhang C, Yu J (2020) Rice seed classification based on convolutional neural network and long short-term memory network. Computers and Electronics in Agriculture 171: 105315.

Additional information

Conflict of interest
The authors have declared that no competing interests exist.

Ethical statement
No ethical statement was reported.

Artificial Intelligence (AI) use
Regarding the use of AI in the preparation of this manuscript, the authors declare the following:
Generating the text

Funding
No funding was reported.

Author contributions
Conceptualization: SEY. Data curation: JA. Formal analysis: MAC. Writing - original draft: YH. Writing - review and editing: AK.

Author ORCIDs
Y. Hamid https://orcid.org/0000-0003-1334-2651

Data availability
All of the data that support the findings of this study are available in the main text.