Emirates Journal of Food and Agriculture

Emirates Journal of Food and Agriculture 36: 1–9
doi: 10.3897/ejfa.2024.122928

RESEARCH PAPER

Enhancing soybean classification with modified inception model: A transfer learning approach

Yonis Gulzar[1]

1 *Department of Management Information Systems, College of Business Administration, King Faisal University, Al-Ahsa 31982, Saudi Arabia*

Corresponding author: Yonis Gulzar (Ygulzar@kfu.edu.sa)

Academic editor: Giovanni Buonaiuto ♦ **Received** 12 March 2024 ♦ **Accepted** 5 April 2024 ♦ **Published** 18 April 2024

Abstract

The impact of deep learning (DL) is substantial across numerous domains, particularly in agriculture. Within this context, our study focuses on the classification of problematic soybean seeds. The dataset employed encompasses five distinct classes, totaling 5513 images. Our model, based on the InceptionV3 architecture, undergoes modification with the addition of five supplementary layers to enhance efficiency and performance. Techniques such as transfer learning, adaptive learning rate adjustment (to 0.001), and model checkpointing are integrated to optimize accuracy. During initial evaluation, the InceptionV3 model achieved 88.07% accuracy in training and 86.67% in validation. Subsequent implementation of model tuning strategies significantly improves performance. Augmenting the architecture with additional layers, including Average Pooling, Flatten, Dense, Dropout, and Softmax, plays a pivotal role in enhancing accuracy. Evaluation metrics, including precision, recall, and F1-score, underscore the model's effectiveness. Precision ranges from 0.9706 to 1.0000, while recall values demonstrate a high capture rate across all classes. The F1-score, reflecting a balance between precision and recall, exhibits remarkable performance across all classes, with values ranging from 0.9851 to 1.0000. Comparative analysis with existing studies reveals competitive accuracy of 98.73% achieved by our proposed model. While variations exist in specific purposes and datasets among studies, our model showcases promising performance in soybean seed classification, contributing to advancements in agricultural technology for crop health assessment and management.

Keywords

Deep Leaning, Image Classification, Soybean seeds, Transfer Learning, Precision Agriculture

Introduction

Soybean stands out as a premier source of both oil and protein within the realm of oilseed crops (Tacarindua et al. 2013). Its cultivation spans the globe, with notable trends emerging in various regions. As of February 2024, data from statista.com underscores Brazil's dominance in soybean production for the 2023/24 period, boasting a remarkable output of 156 million metric tons. In second place, the United States contributes significantly with a production yield of 113.34 million metric tons. Argentina follows closely behind, yielding 50 million metric tons, while China and India contribute 28.84 million and 11 million metric tons, respectively, solidifying their positions within the global soybean landscape. These figures not only underscore the agricultural significance of soybean but also highlight the diverse geographical distribution of its production prowess on a global scale.

In order to continue to have high yields of soybean worldwide, it is highly important to have high quality seeds without any defects for plantation. To have high yield turnover, and successful plantation it is important to choose seeds without any issues such as intactness, spotting, immaturity, breakage, and skin damage. In order to do so an expert is needed to find out which seeds are good for sowing and which ones have defects such as intact or spotted or even immature (Zhao et al. 2021). The traditional method of assessing the seeds' quality relies on manual evaluation done by seasoned expert. However, this process is tedious and time-intensive in nature. Moreover, it is not necessary that every farmer possesses the expertise to distinguish problematic seeds among the batch (Malik et al. 2023).

The pervasive adoption of artificial intelligence (AI) for automation across various sectors, including education, retail, healthcare, automotive, manufacturing, marketing, and agriculture, underscores its pivotal role in modern society. AI's versatility and transformative potential have led to its widespread application, offering tailored solutions and innovative problem-solving capabilities. In the realm of healthcare, AI assists in diagnostics, treatment recommendations, and patient care, improving outcomes and resource utilization (Alam et al. 2022; Gulzar and Khan 2022; Anand et al. 2023; Khan et al. 2021, 2023a, 2023b; Majid et al. 2023; Mehmood et al. 2023). Retailers utilize AI for customer-centric initiatives like personalized recommendations and optimized inventory management, thereby enhancing customer satisfaction and operational efficiency (Gulzar et al. 2023a).

Moreover, AI's impact extends to agriculture, where it revolutionizes traditional farming practices through precision techniques like crop monitoring and yield prediction (Albarrak et al. 2022; Dhiman et al. 2023; Gulzar et al. 2024b). By enhancing agricultural productivity and sustainability, AI contributes to global food security and economic development. Overall, the widespread integration of AI across these sectors highlights its transformative potential and underscores the importance of further research and development in this field (Aggarwal et al. 2023; Gulzar et al. 2023b).

From the literature it can be seen that numerous researchers have tried to incorporate deep learning algorithms in classifying different crops (Ünal and Aktaş 2023; Ünal et al. 2024). Whether it involves identifying diseases within crops or categorizing them, the applications of deep learning have been the focal point of investigation. For instance (Farah et al. 2023) used deep learning to detect the infested leaves of soybeans. They have used VGG19 model using transfer learning to classify the infested leaves and have achieved 97.71% accuracy. In another study (Kaler et al. 2023) incorporated deep learning to identify fungal infected soybean seeds. they have proposed a model, which is the combination of CNN and ML. they claimed that their model has achieved 97.72 accuracy on test dataset. Liu et al. (2023) developed a deep learning model to improve crop mapping in Northeast China. They combined a 3D-CNN with ConvRNN, integrating multi-temporal Sentinel-1 and Sentinel-2 data. Achieving 91.7% accuracy, their model outperformed alternatives, using 10% ground truth data for weak supervision and demonstrating strong generalizability. Kansal et al. (2023) introduced an IoT-Fog computing robotic system for weed and soy plant classification in varying weather conditions. Using a 2D-CNN deep learning approach on a dataset with four classes, including soil, soybean, grass, and weeds, they achieved 97% accuracy. Their system integrates IoT and Fog computing for enhanced reliability and performance, showing promise for effective crop management in adverse conditions. Farmonov et al. (2023) utilized DESIS images to classify crops in southeastern Hungary. Employing WA-CNN, random forest, and SVM algorithms, WA-CNN achieved the highest accuracy at 97.89%, crucial for crop monitoring and yield prediction. They employed factor analysis and wavelet transform to enhance feature extraction, demonstrating DESIS data's potential for agricultural decision-making. Singh et al. (2023) address weed detection's crucial role in crop quality and productivity. They propose a low computational cost state-of-the-art architecture, C13, for detecting weeds among soybean crops. Through comparative analysis with transfer learning models, C13 achieves a high accuracy of 94.58%, outperforming others, enhancing crop cultivation efficiency. Zheng et al. (2023) address soybean kernel damage identification using a fusion of HSI and RGB images with improved ShuffleNet. Their method, HRFN, generates super-resolution HSI images, enhancing spatial fidelity. ShuffleNet_COCSP with OISEW achieves optimal recognition performance (ACCp = 98.36%). This accurate method shows promise for analyzing crop kernel quality indicators.

Gulzar et al. (2020) and Gulzar et al. (2024a) introduce a novel deep learning model for seed classification, crucial in agricultural and industrial sectors. They compare training Xception from scratch versus transfer learning with Pre-trained Xception on a dataset of 15 seeds. Pre-trained Xception exhibits superior performance, achieving perfect accuracy of 1.0000 on validation and test sets with lower loss values and quicker convergence, underscoring its efficiency and effectiveness. Whereas Aktaş et al. (2022) focuses on classifying pistachios as open or closed using deep learning. Industrial datasets were used, yielding high accuracy (96.13%–96.54%) with AlexNet and Inception V3 models.

Wu et al. (2023) introduce an off-center Bayesian deep learning method for accurate crop area estimation. Incorporating phenological features and attention mechanisms, they achieve high accuracy (90.73%) in classifying soybean, maize, and rice. Their approach offers interpretability and generalizability, vital for rapid deployment in various regions, ensuring reliable crop yield estimates for food security.

C. Zhang et al. (2023a) present a novel method for accurate soybean pod identification, crucial for obtaining phenotypic traits. Their approach considers effective and abortive seeds, addressing traditional method limitations. They introduce a comprehensive dataset and compare four object-detection models, selecting YOLOX with the best performance (mAP: 98.24%). Their algorithm enhances pod and seed counting efficiency, vital for indoor seed testing in smart agriculture. N. Zhang et al. (2023b) investigate data balancing methods for improving soybean plant damage classification accuracy. They explore oversampling and weighted loss functions in convolutional neural networks. Results indicate that the new loss function effectively enhances accuracy (98.48% with

DenseNet), while data augmentation reduces accuracy. Using fewer convolution layers with data augmentation improves accuracy by 1.52%.

Alkanan and Gulzar 2024 compare deep learning models for sunflower disease classification, crucial for yield and quality preservation. They evaluate AlexNet, VGG16, InceptionV3, MobileNetV3, and EfficientNet on a dataset. EfficientNetB3 achieves the highest accuracy of 0.979, with other models ranging from 0.865 to 0.969. Results underscore deep learning's effectiveness in disease detection, emphasizing its potential for timely management strategies. Notably, MobileNetV3 and EfficientNetB3 are recommended for their high performance and efficient training.

It can be noticed from the literature that many attempts have been made using deep learning for various aspects of soybean classification and crop management. When it comes to quality assessment of soybean seeds an automated method is needed which can classify problematic seeds from the bunch and help in acquiring good quality seeds before sowing. Therefore, in this work a deep learning model based on InceptionV3 architecture is proposed to classify five different types of problematic seeds of soybean. The modified model of InceptionV3 is designed by adding five different layers to the model before the classification layer. Furthermore, the model is fine-tuned with different techniques such as transfer learning, model checkpointing, adaptive learning rate which helps the proposed model to enhance its accuracy without having any issue of overfitting. Such modification to the model, makes it optimized for the classification and helps a common person to identify problematic soybean seeds. Following points summarize the contribution of this work.

- A detailed review has been conducted, examining the most recent works in said area using deep learning.
- Problem of classifying the problematic soybean seeds have been reintroduced and InceptionV3, a pretrained model with modification has been designed for soybean seeds.
- The modified model has been proposed by adding to the InceptionV3 architecture.
- Different model tuning techniques have been incorporated to enhance the performance of the proposed model and ensure that model does not overfit.

Material and methods

Dataset description

The dataset used in this study is a public dataset, containing the images of five types of soybean seeds: intact, spotted, immature, broken, and skin-damaged, totaling 5513 images (Lin et al. 2023). Each category comprises over 1000 images, classified according to the Standard of Soybean Classification (GB1352-2009). These images were captured using an industrial camera, focusing on soybean seeds in physical contact. Table 1 displays the image count for each class within the dataset, while Fig. 1 showcases samples from each class. The dataset images have been resized to dimensions of 299 × 299 to fit the Inception architecture. The dataset is divided into training, validation, and test subsets in the ratio of 8:1:1, as indicated in Table 1 (Ayoub et al. 2023, 2022).

Table 1. Number of images per class of soybean dataset.

#	Class	Training	Validation	Test	Number of Images
1	Broken	800	101	101	1002
2	Immature	901	112	112	1125
3	Intact	961	120	120	1201
4	Skin-damaged	901	113	113	1127
5	Spotted	846	106	106	1058

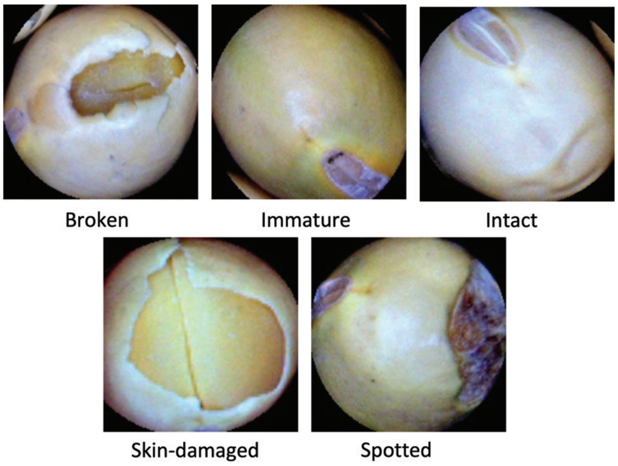

Figure 1. Samples of different classes of soybean dataset.

Proposed model

In this study to classify problematic seeds of soybean an InceptionV3 architecture is used as base model due to its well-known performance of handling complexities in the data and its ability to capture intricate patterns within images. Its deep and efficient convolutional neural network design makes it well-suited for tasks requiring detailed feature extraction. In this study, the original InceptionV3 architecture has been enhanced by incorporating five additional layers before the classification layer, aiming to boost efficiency and performance. Among these added layers are the Average Pooling layer (7×7), which aids in reducing spatial dimensions and controlling overfitting by summarizing information from previous layers. The Flatten layer serves to transform multidimensional data into a one-dimensional array, ensuring compatibility with subsequent fully connected layers essential for classification tasks. Additionally, the inclusion of a Dense layer (0.5) enhances the model's capacity to learn intricate patterns and relationships within the data. The Dropout layer further combats overfitting by randomly deactivating neurons during training, promoting the acquisition of more generalized representations. Finally, the softmax layer at the output enables the conversion of raw predictions into probabilities, facilitating intuitive interpretation and aiding in the selection of the most probable class for classification. This comprehensive modification aims to improve the efficiency and effectiveness of the InceptionV3 architecture in accurately identifying problematic soybean seeds. The architecture of proposed model is shown in Fig. 2.

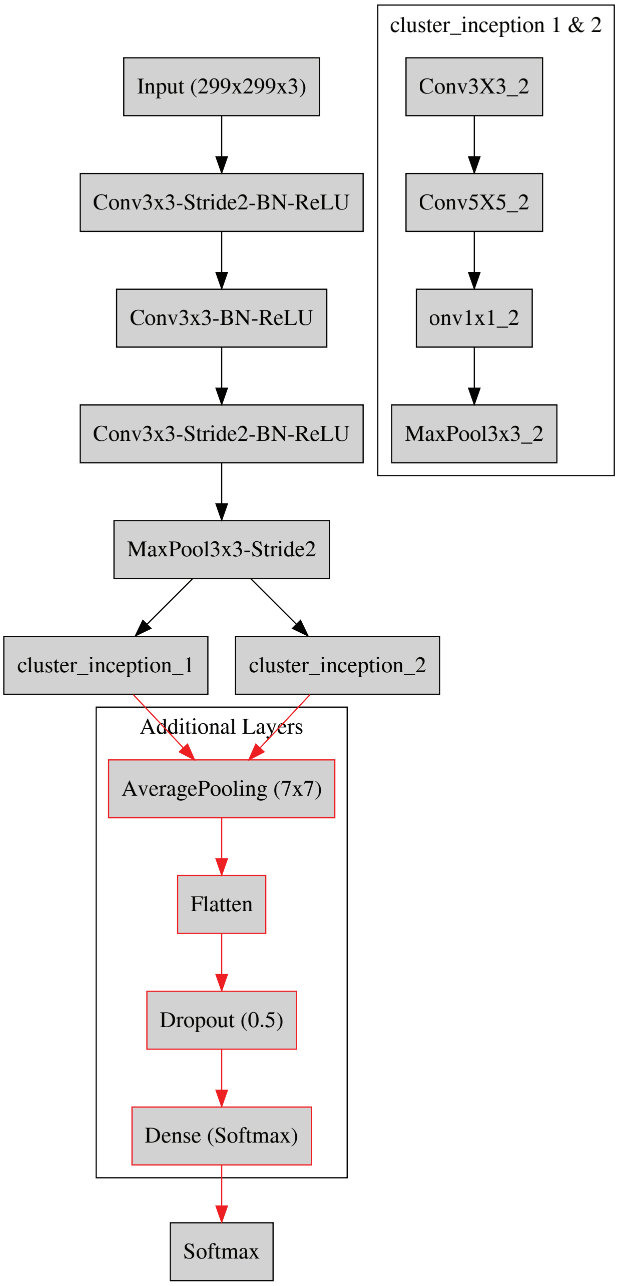

Figure 2. Proposed Model Architecture.

Model tuning

In this study, several model tuning techniques have been employed as mentioned follows. These model tuning techniques are crucial for enhancing the performance and stability of classification models, particularly in identifying problematic soybean seeds. They prevent overfitting, optimize learning rates, utilize pre-trained knowledge effectively, and ensure model stability and performance monitoring, collectively leading to more accurate and reliable results.

- Transfer Learning: Transfer learning involves leveraging pre-trained models, such as InceptionV3, and fine-tuning them on a specific task. By utilizing the knowledge learned from a large dataset, transfer learning enables quicker convergence and improved performance, especially when the target dataset is limited (Gulzar 2023; Mamat et al. 2023).
- Adaptive Learning Rate: Adaptive learning rate methods dynamically adjust the learning rate during training based on the performance of the model. Setting the learning rate to 0.001 indicates a relatively small and gradual adjustment, which can help prevent overshooting or oscillation during optimization, leading to smoother convergence and potentially better generalization.
- Model Checkpointing: Model checkpointing involves saving the model's weights and architecture during training at specific intervals (e.g., after each epoch or when a performance metric improves). This technique ensures that the best-performing model is retained, even if training is interrupted, and allows for easy resumption from the last checkpoint in case of unexpected termination.
- Dropout: Dropout is a regularization technique commonly used to prevent overfitting in neural networks. A dropout rate of 0.5 indicates that during training, each neuron has a 50% probability of being temporarily "dropped out" or deactivated. This forces the network to learn more robust and generalized representations by preventing co-adaptation of neurons, ultimately improving the model's ability to generalize to unseen data.

Experimental environment settings and performance evaluation metrics

The proposed model was developed using Python version 3.8, employing OpenCV version 4.7 and the Keras Library version 2.8 on the Windows 10 Pro operating system. The hardware configuration consisted of an Intel *i*5 processor clocked at 2.9 GHz, an Nvidia RTX 2060 GPU, and 16 GB of RAM.

Various evaluation metrics were applied to gauge the efficacy of classifying corn seed diseases. These metrics included standard indicators like accuracy, precision, recall, and the F1-score. Accuracy denotes the proportion of correctly identified samples across all classes, while recall measures the ratio of accurately classified positive instances relative to all actual positives. Precision quantifies the proportion of accurately identified positive instances among all expected positives. These metrics were computed using Equations (1) to (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

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

$$F1 - Score = 2x \frac{RecallxPrecision}{Recall + Precision} \quad (4)$$

Results and discussion

In this section, we present and discuss the results of our proposed model for the classification of soybean seeds. We begin by showcasing the performance metrics of our model, including loss and accuracy. Following that, we delve into the analysis of the confusion matrix obtained during testing. Additionally, we provide insights into the precision, recall, and F1-score of our model for comprehensive evaluation and discussion.

During the initial evaluation, we employed the InceptionV3 model to train on the soybean dataset without any modifications or tuning techniques. The training phase yielded an accuracy of 88.07%, while the validation accuracy stood at 86.67%. However, the relatively low accuracy observed during both training and validation urged us to implement various model tuning strategies to enhance performance.

To address this, we applied transfer learning, adjusting the adaptive learning rate to 0.001, and implemented model checkpointing. These techniques were instrumental in improving the model's efficacy. Additionally, we augmented the InceptionV3 architecture by incorporating additional layers before the classification layers. These added layers include Average Pooling, Flatten, Dense, Dropout, and Softmax. Adding these layers have been very significant when it comes to performance of the model. Average Pooling is used to reduce the spatial dimensions of the previous convolutional layer. Flatten is utilized to convert the pooled feature maps into a single column, facilitating input into the subsequent Dense layer. The Dense layer enhances the model's capability to learn complex patterns by increasing the number of trainable parameters. Dropout is applied to prevent overfitting by randomly deactivating a fraction of neurons during training. Finally, Softmax activation is utilized in the output layer to generate probability distributions across multiple classes, enabling effective classification.

In Fig. 3, we present the training and validation curves illustrating the accuracy and loss of our proposed model. These curves offer valuable insights into the learning dynamics of our model throughout the training process. As it can be observed from the training accuracy curve, the model's initial learning rate of approximately 28% at epoch 1, gradually improved with each iteration. By the 27th iteration, the model achieved a remarkable accuracy of around 99%, maintained this high accuracy level until the conclusion of training. Conversely, the validation accuracy curve demonstrated a similar upward trend, albeit starting at a lower initial accuracy of 42%. However, with consistent training, the validation accuracy steadily risen, reaching approximately 98–99% by the 31st iteration. Examining the training and validation loss curves, we observed an inversely proportional relationship with accuracy. Initially, both training and validation losses are high, but they steadily decreased with each iteration, indicating successful learning. Notably, some oscillations are observed in the validation loss curve until around the 125th iteration, after which a noticeable convergence occurs.

The impact of transfer learning on training accuracy and loss is evident from these curves, showcasing its significant influence on model performance. The proposed model was trained on ImageNet dataset and its weights were retained while training on soybean dataset. by adopting a hybrid approach we have frozen the model's actual layers and let the additional layers get trained on soybean dataset until 20th iteration. After 20th iteration we unfroze all the layers to make the slight adjustment of the original layers with the said dataset. Adding new layers to the model played a crucial role in enhancing the model accuracy. For instance, the Average Pooling layer helped in dimensionality reduction, whereas the Flatten layer prepared the data for input into subsequent layers. The Dense layer further deepened the model's ability to discern intricate patterns by augmenting trainable parameters. Dropout regularization mitigated overfitting by randomly deactivating neurons during training, promoting generalization. Finally, the Softmax activation function confirmed the generation of probability distributions across multiple classes, enabling effective classification.

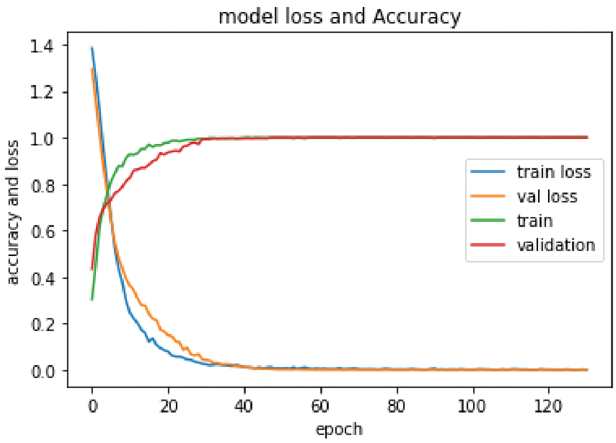

Figure 3. Proposed Model Loss and Training During Training and Validation.

Fig. 4 illustrates the confusion matrix generated from the testing phase of our proposed model. Upon examination, it becomes evident that the majority of instances across all classes were accurately predicted, with only a few misclassifications observed. Specifically, within the broken seeds class, out of 101 instances, two instances were erroneously classified as skin-damaged, while one instance out of 112 from the immature class was incorrectly identified as intact. Similarly, within the intact class, two instances out of 120 were misclassified as immature. Regarding the skin-damaged class, out of 112 instances, two were mistakenly predicted as broken. Interestingly, all 106 instances within the spotted class were correctly identified as such, demonstrating the model's robustness in distinguishing this particular category. Overall, while the proposed model demonstrates high accuracy across most classes, the observed misclassifications highlight areas for potential refinement and optimization. These insights gleaned from the confusion matrix are invaluable for further enhancing the model's performance and reliability in soybean seed classification.

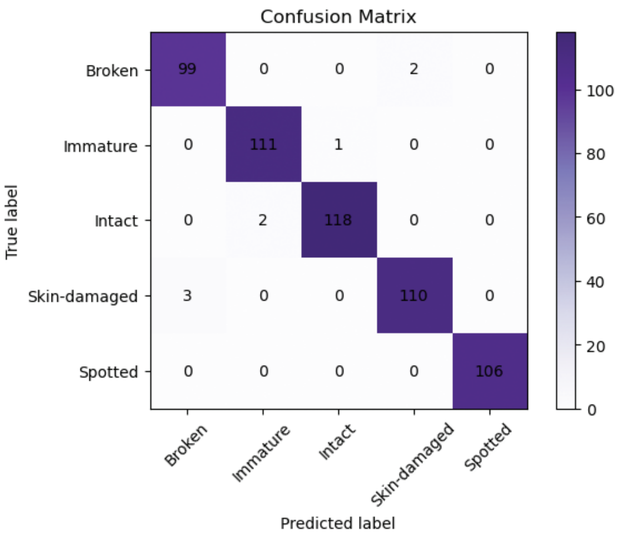

Figure 4. Confusion Matrix of Proposed Model during Testing.

The precision, recall, and F1-score metrics provided in Table 2 offer valuable insights into the performance of our proposed model across different classes during testing. These metrics, alongside the overall accuracy of 98.73%, serve as crucial indicators of the model's effectiveness in soybean seed classification. Starting with precision, which measures the proportion of correctly predicted instances among all instances classified as a particular class, we observe consistently high values across all classes. For instance, the precision for the Broken, Immature, Intact, and Skin-damaged classes ranges from 0.9706 to 0.9916, indicating a high level of accuracy in classifying instances within these categories. The highest precision score of 1.0000 is achieved for the Spotted class, signifying that all instances classified as Spotted were indeed correctly predicted.

Moving on to recall, which measures the proportion of correctly predicted instances among all instances that truly belong to a particular class, we again observe commendable performance across all classes. The recall values of 1.0000 for Broken, Immature, Skin-damaged, and Spotted classes indicate that the model effectively captures all instances belonging to these categories. While the recall for the Intact class is slightly lower at 0.9833, it still demonstrates the model's ability to correctly identify the majority of instances within this class.

The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of a model's performance. Here, we see impressive F1-scores ranging from 0.9851 to 1.0000 across all classes, highlighting the model's capability to achieve both high precision and recall simultaneously. Particularly noteworthy is the perfect F1-score of 1.0000 for the Spotted class, reflecting the model's exceptional performance in accurately classifying instances within this category.

These metrics collectively underscore the significance of employing model tuning techniques and incorporating additional layers in our proposed model. The consistently high precision, recall, and F1-scores, coupled with the impressive overall accuracy, affirm the effectiveness of these strategies in enhancing the model's performance in soybean seed classification. By fine-tuning parameters and refining the model architecture, we have achieved remarkable accuracy and reliability, thus demonstrating the profound impact of meticulous model development and optimization in machine learning tasks. Fig. 5 illustrates the visualization showcasing the accuracy of all classes in the proposed model during the testing phase.

Table 2. Precision, Recall and F1-Score, Accuracy of all classes of proposed model during Testing.

Class	Precision	Recall	F1 Score	Accuracy
Broken	0.9706	1.0000	0.9851	0.9802
Immature	0.9802	1.0000	0.9900	0.9911
Intact	0.9916	0.9833	0.9874	0.9833
Skin-damaged	0.9821	1.0000	0.9909	0.9735
Spotted	1.0000	1.0000	1.0000	1.0000

Furthermore, the proposed model is compared with the existing works in the area of soybean seed and plant analysis as presented in Table 3. From the table it can be noticed that (Kaler et al. 2023) utilized a CNN-based approach with two classes of soybean seeds, achieving an accuracy of 97.72% in classifying problematic seeds. (Zheng et al. 2023) employed ShuffleNet and DenseNet (N. Zhang et al. 2023b) architectures, both achieving higher accuracies of 98.36% and 98.48%, respectively, for classifying four classes of soybean seeds. (Mathew et al. 2023) implemented YOLOv7 for soybean pod counting, however, specific accuracy values are not provided. (Tirkey et al. 2023) utilized YOLOv5 for detecting insects on soybean crops, achieving a high accuracy of 98.75%. (Goshika et al. 2024) also utilized YOLOv5, focusing on the identification of diseases on soybean leaves, although specific accuracy values are not reported.

In comparison, the proposed model employs the InceptionV3 architecture for classifying five classes of problematic soybean seeds, achieving an accuracy of 98.73%. This accuracy is competitive with previous studies, demonstrating the effectiveness of the proposed model in accurately classifying problematic soybean seeds. The consistent high accuracies across various architectures

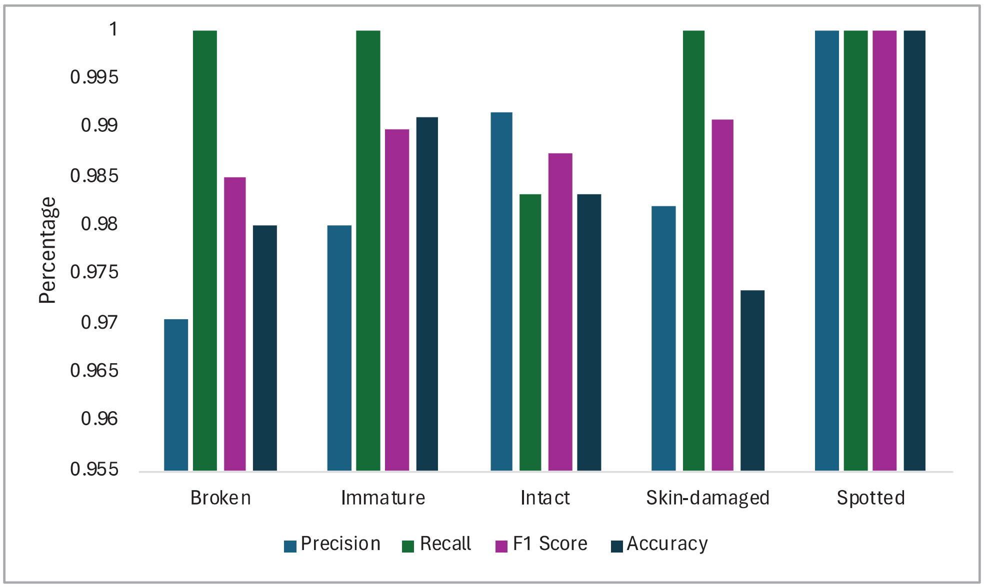

Figure 5. Precision, Recall, F1-Score and Accuracy of Proposed Model During Testing.

Table 3. Contrast between the Proposed Model and previous research investigations.

Work	Base Model	Classes	Dataset type	Accuracy	Purpose
(Kaler et al. 2023)	CNN	Two	Soybean seeds	97.72%	Classification of problematic soybean seeds
(Zheng et al. 2023)	ShuffleNet	Four	Soybean seeds	98.36%	Classification of problematic soybean seeds
(N. Zhang et al. 2023b)	DenseNet	Four	Soybean seeds	98.48%	Classification of problematic soybean seeds
(Mathew et al. 2023)	YOLOV7	Four	Soybean plant	-	Counting of soybean pods
(Tirkey et al. 2023)	YOLOv5	Five	Soybean leaf disease	98.75%	Detection of insects on soybean crop
(Goshika et al. 2024)	YOLOv5	Five	Soybean leaf disease	-	Identification of diseases on soybean leaves
Proposed Model	InceptionV3	Five	Soybean seeds	98.73%	Classification of problematic soybean seeds

indicate the robustness of deep learning approaches in soybean analysis tasks. However, it is noteworthy that while accuracies are high in most cases, the specific purposes and datasets vary among the studies, indicating the importance of selecting appropriate models tailored to the specific task and dataset characteristics. Overall, the proposed model showcases promising performance in soybean seed classification, contributing to the advancement of agricultural technology for crop health assessment and management.

Conclusion

In this study, an effective model tailored for the classification of problematic soybean seeds has been introduced. Leveraging the InceptionV3 architecture as our foundation, we have enhanced its efficiency and performance by adding five additional layers before the classification layer. Through rigorous training on a dataset featuring five distinct problematic soybean seed classes, our model achieved an impressive accuracy of 98.73%. Notably, the integration of model tuning techniques such as transfer learning and dropout regularization played a pivotal role in refining the model's accuracy, fostering robust and generalized representations. Furthermore, our comparative analysis with existing studies revealed that our model outperformed others in terms of accuracy, showcasing its efficacy in soybean seed classification tasks.

For future direction we plan to explore the scalability and applicability of our model to broader agricultural contexts beyond soybean seed classification. Additionally, further investigations will be conducted to delve into fine-tuning the model parameters and explore ensemble methods to potentially enhance classification performance even further.

Data availability statement

A public dataset has been used in this study (Lin et al. 2023).

Conflicts of interest

The author declare no conflict of interest.

Acknowledgments

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under the Project GRANT6,170.

References

Aggarwal S, Gupta S, Gupta D, Gulzar Y, Juneja S, Alwan AA, Nauman A (2023) An artificial intelligence-based stacked ensemble approach for prediction of protein subcellular localization in confocal microscopy images. Sustainability 15(2): 1695. https://doi.org/10.3390/su15021695

Aktaş H, Kızıldeniz T, Ünal Z (2022) Classification of pistachios with deep learning and assessing the effect of various datasets on accuracy. Journal of Food Measurement and Characterization 16(3): 1983–1996. https://doi.org/10.1007/s11694-022-01313-5

Alam S, Raja P, Gulzar Y (2022) Investigation of machine learning methods for early prediction of neurodevelopmental disorders in children. Wireless Communications and Mobile Computing 2022: 5766386. https://doi.org/10.1155/2022/5766386

Albarrak K, Gulzar Y, Hamid Y, Mehmood A, Soomro AB (2022) A Deep Learning-Based Model for date fruit classification. Sustainability 14(10): 6339. https://doi.org/10.3390/su14106339

Alkanan M, Gulzar Y (2024) Enhanced corn seed disease classification: leveraging MobileNetV2 with feature augmentation and transfer learning. Frontiers in Applied Mathematics and Statistics 9: 1320177. https://doi.org/10.3389/fams.2023.1320177

Anand V, Gupta S, Gupta D, Gulzar Y, Xin Q, Juneja S, Shaikh A (2023) Weighted average ensemble Deep Learning Model for stratification of brain tumor in MRI Images. Diagnostics 13(7): 1320. https://doi.org/10.3390/diagnostics13071320

Ayoub S, Gulzar Y, Reegu FA, Turaev S (2022) Generating image captions using bahdanau attention mechanism and transfer learning. Symmetry 14(12): 2681. https://doi.org/10.3390/sym14122681

Ayoub S, Gulzar Y, Rustamov J, Jabbari A, Reegu FA, Turaev S (2023) Adversarial approaches to tackle imbalanced data in machine learning. Sustainability 15(9): 7097. https://doi.org/10.3390/su15097097

Dhiman P, Kaur A, Balasaraswathi VR, Gulzar Y, Alwan AA, Hamid Y (2023) Image acquisition, preprocessing and classification of citrus fruit diseases: A systematic literature review. Sustainability 15(12): 9643. https://doi.org/10.3390/su15129643

Farah N, Drack N, Dawel H, Buettner R (2023) A Deep Learning-Based approach for the detection of infested soybean leaves. IEEE Access 11: 99670–99679. https://doi.org/10.1109/ACCESS.2023.3313978

Farmonov N, Amankulova K, Szatmari J, Sharifi A, Abbasi-Moghadam D, Mirhoseini Nejad SM, Mucsi L (2023) Crop type classification by DESIS hyperspectral imagery and machine learning algorithms. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16: 1576–1588. https://doi.org/10.1109/JSTARS.2023.3239756

Goshika S, Meksem K, Ahmed KR, Lakhssassi N (2024) Deep Learning Model for classifying and evaluating soybean leaf disease damage. International Journal of Molecular Sciences 25(1): 106. https://doi.org/10.3390/ijms25010106

Gulzar Y (2023) Fruit image classification model based on MobileNetV2 with deep transfer learning technique. Sustainability 15(3): 1906. https://doi.org/10.3390/su15031906

Gulzar Y, Khan SA (2022) Skin lesion segmentation based on vision transformers and convolutional neural networks—A comparative study. Applied Sciences 12(12): 5990. https://doi.org/10.3390/app12125990

Gulzar Y, Hamid Y, Soomro AB, Alwan AA, Journaux L (2020) A Convolution Neural Network-Based Seed Classification System. Symmetry 12(12): 2018. https://doi.org/10.3390/sym12122018

Gulzar Y, Alwan AA, Abdullah RM, Abualkishik AZ, Oumrani M (2023a) OCA: Ordered Clustering-Based Algorithm for E-Commerce Recommendation System. Sustainability 15(4): 2947. https://doi.org/10.3390/su15042947

Gulzar Y, Ünal Z, Aktas HA, Mir MS (2023b) Harnessing the power of transfer learning in sunflower disease detection: A comparative study. Agriculture 13(8): 1479. https://doi.org/10.3390/agriculture13081479

Gulzar Y, Ünal Z, Ayoub S, Reegu FA (2024a) Exploring Transfer Learning for Enhanced Seed Classification: Pre-trained Xception Model. In Lecture Notes in Civil Engineering. https://doi.org/10.1007/978-3-031-51579-8_14

Gulzar Y, Ünal Z, Ayoub S, Reegu FA, Altulihan A (2024b) Adaptability of deep learning: datasets and strategies in fruit classification. BIO Web of Conferences 85: 01020. https://doi.org/10.1051/bioconf/20248501020

Kaler N, Bhatia V, Mishra AK (2023) Deep Learning-Based Robust Analysis of Laser Bio-Speckle Data for Detection of Fungal-Infected Soybean Seeds. IEEE Access 11: 89331–89348. https://doi.org/10.1109/ACCESS.2023.3305273

Kansal I, Khullar V, Verma J, Popli R, Kumar R (2023) IoT-Fog-enabled robotics-based robust classification of hazy and normal season agricultural images for weed detection. Paladyn 14(1): 1–11. https://doi.org/10.1515/pjbr-2022-0105

Khan F, Ayoub S, Gulzar Y, Majid M, Reegu FA, Mir MS, Elwasila O (2023a) MRI-based effective ensemble frameworks for predicting human brain tumor. Journal of Imaging 9(8): 163. https://doi.org/10.3390/jimaging9080163

Khan F, Gulzar Y, Ayoub S, Majid M, Mir MS, Soomro AB (2023b) Least square-support vector machine based brain tumor classification system with multi model texture features. Frontiers in Applied Mathematics and Statistics 9: 1324054. https://doi.org/10.3389/fams.2023.1324054

Khan SA, Gulzar Y, Turaev S, Peng YS (2021) A modified HSIFT descriptor for medical image classification of anatomy objects. Symmetry 13(11): 1987. https://doi.org/10.3390/sym13111987

Lin W, Fu Y, Xu P, Liu S, Ma D, Jiang Z, Su Q (2023) Soybean Seeds 6. https://doi.org/10.17632/V6VZVFSZJ6.6

Liu N, Zhao Q, Williams R, Barrett B (2023) Enhanced crop classification through integrated optical and SAR data: a deep learning approach for multi-source image fusion. International Journal of Remote Sensing. https://doi.org/10.1080/01431161.2023.2232552

Majid M, Gulzar, Y, Ayoub S, Khan F, Reegu FA, Mir MS, Soomro AB (2023) Enhanced transfer learning strategies for effective kidney tumor classification with CT imaging. International Journal of Advanced Computer Science and Applications 14(8): 2023. https://doi.org/10.14569/IJACSA.2023.0140847

Malik I, Ahmed M, Gulzar Y, Baba SH, Mir M S, Soomro AB, Elwasila O (2023) Estimation of the extent of the vulnerability of agriculture to climate change using analytical and Deep-Learning methods: A case study in Jammu, Kashmir, and Ladakh. Sustainability 15(14): 11465. https://doi.org/10.3390/su151411465

Mamat N, Othman MF, Abdulghafor R, Alwan AA, Gulzar Y (2023) Enhancing image annotation technique of fruit classification using a Deep Learning approach. Sustainability 15(2): 901. https://doi.org/10.3390/su15020901

Mathew J, Delavarpour N, Miranda C, Stenger J, Zhang Z, Aduteye J, Flores P (2023) A novel approach to pod count estimation using a depth camera in support of soybean breeding applications. Sensors 23(14): 6506. https://doi.org/10.3390/s23146506

Mehmood A, Gulzar Y, Ilyas QM, Jabbari A, Ahmad M, Iqbal S (2023) SBXception: A Shallower and Broader Xception architecture for efficient classification of skin lesions. Cancers 15(14): 3604. https://doi.org/10.3390/cancers15143604

Singh V, Gourisaria MK, Harshvardhan GM, Choudhury T (2023) Weed detection in soybean crop using deep neural network. Pertanika Journal of Science and Technology 31: 401–423. https://doi.org/10.47836/pjst.31.1.24

Tacarindua CRP, Shiraiwa T, Homma K, Kumagai E, Sameshima R (2013) The effects of increased temperature on crop growth and yield of soybean grown in a temperature gradient chamber. Field Crops Research 154: 74–81. https://doi.org/10.1016/j.fcr.2013.07.021

Tirkey D, Singh KK, Tripathi S (2023) Performance analysis of AI-based solutions for crop disease identification, detection, and classification. Smart Agricultural Technology 5: 100238. https://doi.org/10.1016/j.atech.2023.100238

Ünal Z, Aktaş H (2023) Classification of hazelnut kernels with deep learning. Postharvest Biology and Technology 197: 1122 https://doi.org/10.1016/j.postharvbio.2022.112225

Ünal Z, Kızıldeniz T, Özden M, Aktaş H, Karagöz Ö (2024) Detection of bruises on red apples using deep learning models. Scientia Horticulturae 329: 113021. https://doi.org/10.1016/j.scienta.2024.113021

Wu Y, Wu P, Wu Y, Yang H, Wang B (2023) Remote sensing crop recognition by coupling phenological features and off-center bayesian deep learning. Remote Sensing 15(3): 674. https://doi.org/10.3390/rs15030674

Zhang C, Lu X, Ma H, Hu Y, Zhang S, Ning X, Jiao J (2023a) High-Throughput Classification and Counting of Vegetable Soybean Pods Based on Deep Learning. Agronomy 13(4): 674. https://doi.org/10.3390/agronomy13041154

Zhang N, Zhang E, Li F (2023b) A soybean classification method based on data balance and deep learning. Applied Sciences (Switzerland) 13(11): 6425. https://doi.org/10.3390/app13116425

Zhao G, Quan L, Li H, Feng H, Li S, Zhang S, Liu R (2021) Real-time recognition system of soybean seed full-surface defects based on deep learning. Computers and Electronics in Agriculture 187: 106230. https://doi.org/10.1016/j.compag.2021.106230

Zheng L, ZhaoM, Zhu J, Huang L, Zhao J, Liang, D, Zhang D (2023) Fusion of hyperspectral imaging (HSI) and RGB for identification of soybean kernel damages using ShuffleNet with convolutional optimization and cross stage partial architecture. Frontiers in Plant Science 13: 1098864. https://doi.org/10.3389/fpls.2022.1098864