

Acta Technologica Agriculturae 3
Nitra, Slovaca Universitas Agriculturae Nitriae, 2026, pp. 179–187

SOYBEAN SEED QUALITY ASSESSMENT USING AI AND MACHINE LEARNING MODELS

Sajad SABZI^{*1}, Raziye POUARDARBANI^{*2}, Mahan MASOUMZADE³,
José Luis HERNANDEZ-HERNANDEZ⁴

¹Gorgan University of Agricultural Sciences and Natural Resources, Faculty of Water and Soil Engineering,
Department of Biosystems Engineering, Gorgan, Iran

²University of Mohaghegh Ardabili, Faculty of Agricultural and Natural Resources, Department of Biosystems
Engineering, Ardabil, Iran

³Sharif University of Technology, Faculty of Computer Engineering, Department of Computer Engineering,
Tehran, Iran, mahan.48415@gmail.com

⁴National Technological Institute of México, Technological Institute of Chilpancingo, Chilpancingo, Guerrero, México,
joseluis.hernandez@itchilpancingo.edu.mx

*correspondence: s.sabzi@gau.ac.ir (S.S.), r_pourdarbani@uma.ac.ir (R.P.)

Soybean is a key crop due to its nutritional, industrial, and agricultural benefits. Its production performance faces challenges such as improper storage, seed damage, and moisture imbalance. This paper aims to classify soybean seeds into five categories – healthy, damaged, immature, broken, and stained – using image processing and deep learning. Four classifiers DenseNet, MobileNet, ResNet, and EfficientNet were used to classify seeds on a dataset of 5513 soybean seed images. The best performance among the models used was obtained from DenseNet with an accuracy of 96% and an F1-score of 94%. The results also showed that all models achieved an AUC value above 0.99, indicating high performance of the models. This research highlights the potential of machine learning and artificial intelligence in optimising soybean classifiers.

Keywords: soybean; deep learning; convolutional neural network; image processing; quality assessment; comparative analysis

Soybean is one of the most important agricultural crops globally, valued for its high-quality protein (an excellent choice for vegetarians), diverse food applications (soy milk, tofu, tempeh, soybean oil, and plant-based proteins), health benefits (reducing the risk of heart disease, lowering LDL cholesterol, and improving bone health due to isoflavones), industrial uses (biodiesel, paints, and chemicals), and role in sustainable agriculture (improving soil quality through nitrogen fixation) (Pagano and Miransari, 2016).

Soybeans are traded in various forms of grain, oil and meal, and the largest producers are the the United States (one third of global production), Brazil, and Argentina. Other producers such as China, India, Paraguay, Canada, Mexico, and the European Union are in the second place (Dragan et al., 2018).

Soybeans face challenges in the post-harvest stage, such as physical and chemical damage to the seeds (fractures and cracks during harvest and transportation, which leads to insect and fungal attacks and accelerated chemical spoilage). Subsequently, such damage poses challenges to the grading and processing process. Failure to harvest on time and failure to observe appropriate drying conditions (for example, failure to pay attention to seed moisture at the time of harvest) can damage the quality of the product because low moisture leads to brittleness and high moisture leads to rapid spoilage. Attention to soybean storage is also very important, as pests like beetles and termites, along

with various diseases, can significantly reduce both quantity and quality (Bern et al., 2008).

Table 1 provides a detailed breakdown and explanation of soybean seeds, including damaged, broken, immature, spotted, insect-bitten, wrinkled, green, purple, and cracked seeds (Wang et al., 2002).

Sable et al. (2024) presented an efficient neural network called SSDINet for detecting defects in soybean seeds, incorporating convolutional layers, depthwise blocks, and squeeze-and-excitation mechanisms. The model achieved an impressive accuracy of 98.64% in differentiating healthy seeds from defective ones (e.g., damaged, broken, insect-bitten, wrinkled, green, purple, and cracked seeds). Zhang et al. (2023) investigated the impact of data imbalance on soybean classification CNNs. Two data-balancing methods were explored: oversampling and a loss function with class-specific weights. Experiments with four network architectures revealed that the modified loss function significantly improved accuracy, with DenseNet achieving 98.48%. However, excessive convolutional layers in augmented datasets reduced performance, while smaller networks enhanced accuracy by 1.52%. Dioses et al. (2025) applied the EfficientNet CNN to classify soybean seeds into five categories: broken, immature, intact, skin-damaged, and spotted. Over 5000 images were enhanced using noise, flipping, and contrast adjustments. The model, trained with a 75 : 25 dataset split, achieved 100% training accuracy and 93.83% validation accuracy. Intact soybeans had the highest

Table 1 Description of different kinds of damages in soybean seeds and their causes

Type	Description	Possible causes
Intact seeds	soybeans that are whole, undamaged, and in good condition (yellowish, smooth surface)	proper handling and ideal growing conditions
Damaged seeds	soybeans showing general physical harm, such as dents, abrasions, or pressure marks	mishandling during harvest, transport, or storage
Broken seeds	soybeans that are split into parts or fractured	mechanical damage from equipment or handling
Immature seeds	soybeans that are underdeveloped, smaller, or pale in colour	harvesting too early or poor growing conditions
Spotted seeds	soybeans with spots, often caused by diseases or discolouration	fungal infections, bacterial blight, or pest damage
Insect-bitten	soybeans with visible holes or damage caused by insect activity	infestation by pests like stink bugs or beetles
Wrinkled seeds	soybeans with uneven, shrivelled, or folded surfaces	poor growing conditions, drought, or disease
Green seeds	soybeans retaining a green colour, indicating immaturity or chlorophyll retention	early harvest or unfavourable weather
Purple seeds	soybeans with purple discolouration on the seed coat	genetic traits or fungal infections
Cracked seeds	soybeans with visible splits or cracks, usually superficial	rapid drying or physical/mechanical stress

F1-score (97.35%), though minor misclassifications occurred for similar categories. The results highlight the model's effectiveness for agricultural quality assessment. Chen et al. (2025) used hyperspectral imaging to classify soybeans harvested with the 4LZ-1.5 combine. Preprocessing methods like baseline correction and normalisation refined the data, while feature wavelengths were selected using successive projection algorithm (SPA) and competitive adaptive reweighted sampling (CARS). A random forest model, optimised with Scholarly articles for optimised with particle swarm optimisation (PSO) and differential evolution (DE), achieved 100% accuracy during training. Yang et al. (2024) introduced a method for classifying soybean seeds using hyperspectral RGB image reconstruction. Hyperspectral images of seven soybean varieties (H1–H7) were captured, and the red, green, and blue bands with key features were selected to reconstruct images with distinct textures and colours. A new dataset for seed segmentation was created, and seven models were compared for classification performance. Experiments with ResNet34 showed an accuracy improvement from 88.87% to 91.75% after RGB reconstruction, confirming that RGB reconstruction enhances image features. Classification accuracies for ResNet18, ResNet34, ResNet50, ResNet101, CBAM-ResNet34, SENet-ResNet34, and SENet-ResNet34-DCN were 72.25%, 91.75%, 89%, 88.48%, 92.28%, 92.80%, and 94.24%, respectively. The SENet-ResNet34-DCN model achieved the best accuracy (94.24%) with a loss of 0.3, proving its effectiveness in soybean seed classification. Jitanan and Chimlek (2019) extracted robust colour features from the H component of the hyperspectral imaging (HIS) model. By combining the H component's colour histogram with Gray level co-occurrence matrix (GLCM) texture features, a support vector machine (SVM) classifier was developed to grade soybean seeds, addressing shape variations. Trained on a dataset of 6600 images, the model achieved an overall accuracy of 99.2%. Specifically, it reached 100% accuracy for purple, green, and other seeds, 98.1% for wrinkled seeds, and 97.9% for normal seeds. De Oliveira et

al. (2024) used AI and image analysis to classify greenish and yellow seeds of soybeans by colour, improving seed evaluation. A decision tree trained on 12,000 images from the GroundEye® S800D achieved 89.7% accuracy. Yellow seeds had better germination and growth than greenish ones. Hue angle (h) and luminosity (L) were key features, with hue angles below 97.5° linked to higher germination rates ($R^2 = 73\%$). Wang et al. (2004) used near-infrared (NIR) spectroscopy to classify healthy and fungal-damaged soybean seeds, distinguishing the types of fungal damage. A diode-array NIR spectrometer (400–1700 nm) collected seed spectra. Partial least square (PLS) and neural network models were developed. The two-class PLS model achieved over 99% accuracy in the 490–1690 nm range. Neural networks performed better in five-class classification, with 93.5% accuracy for calibration and 94.6% for validation. Validation accuracies were 100% (healthy), 99% (*Phomopsis*), 84% (*Cercospora kikuchii*), 94% (SMV), and 96% (downy mildew). In recent studies, such as Xia et al. (2025), hybrid models have been investigated. They proposed a combination of CNN and Vision Transformer (ViT) models. The accuracy of classifier on a similar five-class soybean dataset reached 96.02%, indicating the effectiveness of combining the strengths of different architectures. As observed, various studies have been reported on the potential of machine learning and deep learning techniques in classification (Adnan et al., 2026; Shirzad et al., 2025), but so far, no comprehensive and complete comparison of several advanced CNN architectures has been conducted on the soybean. Different models provide different results in terms of accuracy, computational complexity, and training stability. On the other hand, the use of intact and high-quality seeds is essential for achieving optimal growth and higher yield. Sowing damaged seeds for planting will result in poor germination and reduced yield.

This paper aims to address this gap by developing four classifiers – namely, DenseNet, MobileNet, ResNet, and EfficientNet – to classify soybean seeds into five

categories based on image processing and artificial intelligence: healthy, immature, broken, stained, and skin-damaged.

This research has significant practical implications. Automating seed classification can improve the agricultural supply chain, increasing productivity and quality control. This, in turn, will increase crop yield and profitability, and ensure a continuous supply of high-quality soybeans, maintaining crop stability.

Material and Methods

The Soybeans Seeds dataset was obtained from Kaggle (Lin et al., 2023; Pereira et al., 2024) and includes a publicly available collection of 5513 RGB images of individual soybean seeds. This dataset was chosen because of being publicly available. Moreover, it provides well-defined class categories and has a sufficient number of images to evaluate deep learning models. However, it has certain limitations, such as a fixed background and controlled lighting, which may not fully represent the complexity of real-world situations. Class labels are assigned based on expert visual assessment. Each image belongs to one of five quality classes: intact, spotted, immature, broken, and skin-damaged (Fig. 1). The images are resized to 227×227 pixels from their original resolution of 3072×2048 pixels.

Proposed models for classification of soybean seeds

In this paper, four classification models were used: MobileNetV2, ResNet-18, DenseNet-121, and EfficientNet-B0. The goal was to perform a comprehensive comparison between the models selected due to their distinct features: DenseNet-121 due to high accuracy, MobileNetV2 due to computational efficiency, and ResNet-18 and EfficientNet-B0 as balanced performance baselines.

MobileNetV2

This model is a lightweight model that is proposed for use in mobile phones and embedded devices. The most important innovation of that is the inverse residual structure with linear bottlenecks. This reduces the computational cost and model size, while minimising the loss of accuracy. For this reason, this model is proposed as a candidate for real-time tasks on devices with limited resources (Sandler et al., 2018).

DenseNet-121

This model was proposed by introducing a new connection. In fact, each layer is connected to the other layer in a feedforward manner. The three main advantages of this type of connection are feature reuse, enhanced feature propagation, and reduced vanishing gradient problem (Huang et al., 2017). In this paper, DenseNet-121 was used because it has a deep architecture and is computationally manageable.

EfficientNet

This model is a systematic model that scales the resolution, depth, and width of the network using a composite factor. The main advantage of that is the trade-off between accuracy and computational resources (FLOPS). This paper employs EfficientNet-B0, a base model known for its strong performance-to-efficiency ratio (Tan and Le, 2019).

ResNet-18

This model allows gradients to flow easily through deep networks by enabling layers to learn residual functions, which is derived from the concept of identity shortcut connection. The advantage was that it enables the training of very deep and powerful models. This paper utilises ResNet-18, a shallower variant that provides a strong baseline without demanding high computational resources (He et al., 2016).

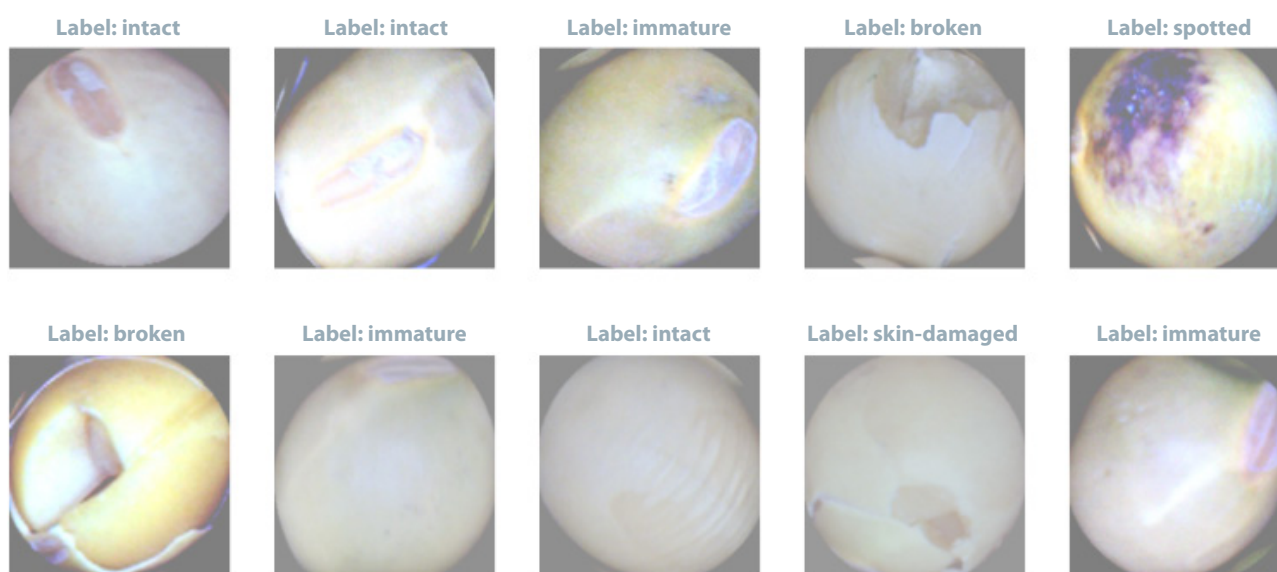


Fig. 1 Examples of soybean seeds labelled into five classes: intact, immature, broken, spotted, and skin-damaged

Evaluation criteria of classifiers

To ensure the reliability of our machine-learning models, a diverse set of evaluation metrics was applied to analyse their performance and behaviour at various stages of development.

Confusion matrix and accuracy, F1-score, and micro-averaged precision (AP)

It compares predicted and true labels to assess classification accuracy across classes by criteria of precision, recall, and the precision–recall curve. Precision and recall are the key metrics in classification tasks. Precision measures the proportion of correct positive predictions out of all positive predictions, calculated as Eq. 1, where TP and FP represent true positives and false positives, respectively.

$$\text{precision} = \frac{TP}{TP + FP} \quad (1)$$

Recall, or sensitivity, represents the proportion of actual positives correctly identified, calculated as Eq. 2, where FN is the count of false negatives.

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

Accuracy measures the proportion of correctly classified samples out of the total dataset, serving as a straightforward indicator of model performance (Eq. 3).

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

F1-score, a harmonic mean of precision and recall, is calculated as Eq. 4. This metric is particularly valuable when both precision and recall are equally crucial.

$$F1\text{-score} = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \quad (4)$$

Using the analysis of these criteria, it can be stated that if recall and F1-score are high and accuracy is low, it

means that the model performs well in identifying most of the true positives but has problems in the accuracy of predictions.

Experimental setup and training

All models were implemented and trained using the Python programming language and PyTorch framework. Experiments were conducted on a system with the following hardware specifications: Intel Core i7-11800H processor, 32 Gb RAM, and an NVIDIA RTX 3060 GPU with 6 Gb of VRAM.

The dataset, consisting of 5513 images, was randomly split into three subsets: a training set containing 3302 images (about 60%), a validation set with 1106 images (about 20%), and a test set also comprising 1106 images (about 20%). The specifications of the hyperparameters are very important for training the models. In this study, the best hyperparameters were selected for all models using trial and error as follows. The models were trained using the Adam optimiser with a learning rate of 0.0001, a batch size of 32, and 100 epochs. Cross-Entropy Loss was employed as the objective function for all models.

Results and Discussion

Training and validation performance

Figure 2a–b shows the training and validation losses of all the proposed models. A closer look at the loss curves reveals that the DenseNet model shows signs of overfitting, as indicated by the divergence between its training and validation losses. In contrast, the MobileNet model shows less overfitting, with the validation loss closely following the training loss. The fluctuations observed in the validation loss curves for EfficientNet and ResNet indicate some instability during training.

Classification performance

Figure 3 shows the confusion matrix for the test dataset. It can be concluded from this figure that most of the samples are located on the main diagonal of the matrix, which means that the number of correct classifications for each

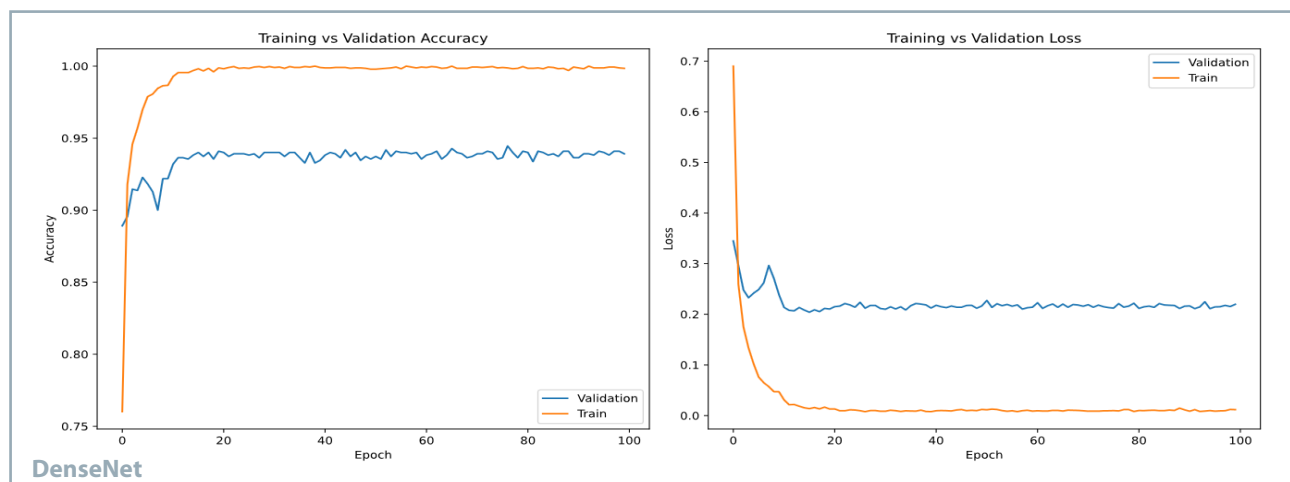


Fig. 2a Training and validation loss of proposed models

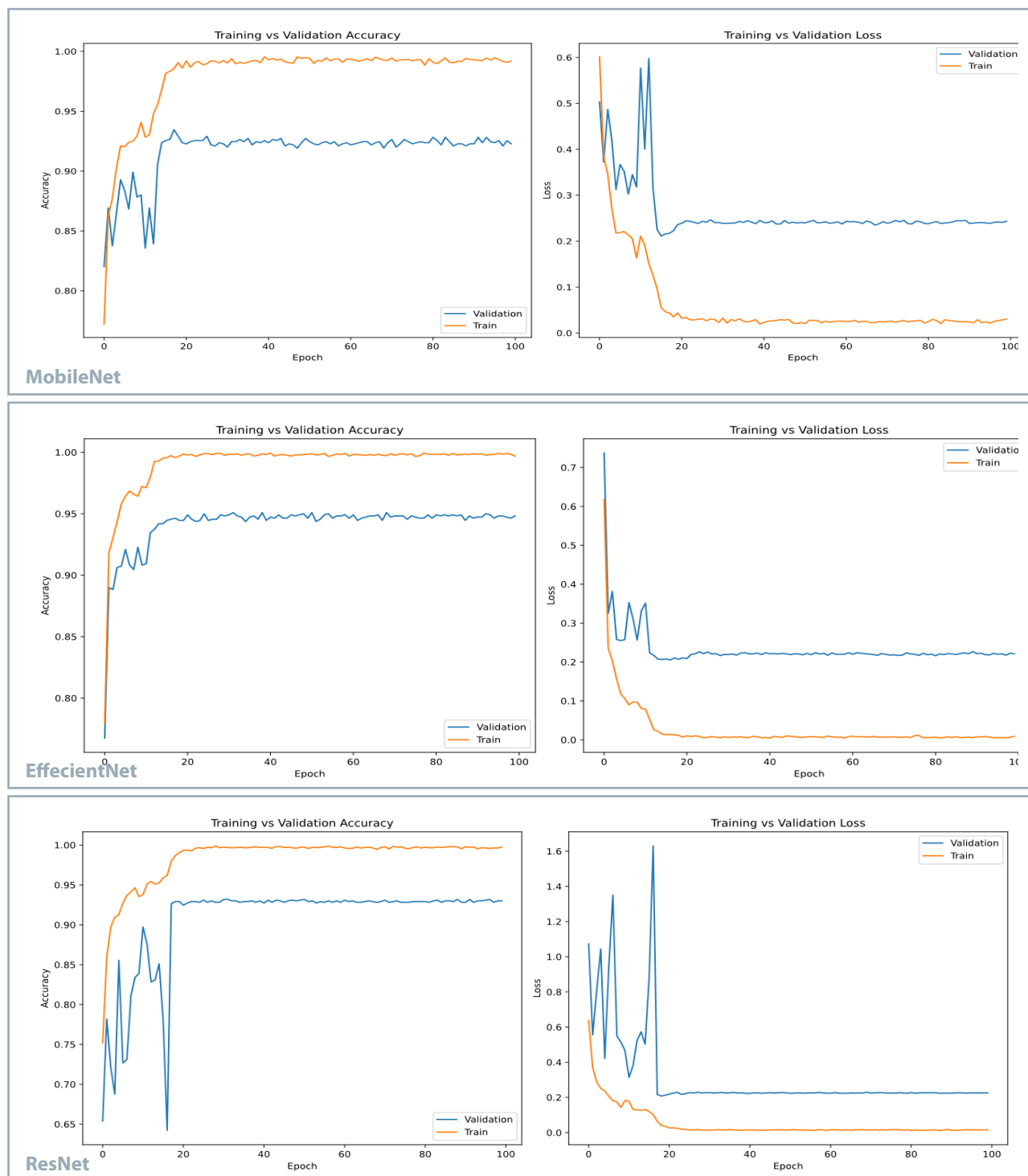


Fig. 2b Training and validation loss of proposed models

class is high. Table 2 gives the values of the various criteria for evaluating the different models used in this study. As can be seen, all models have high performance, and the values of the criteria are above 0.91. A closer look at the table shows that the Intact and Spotted classes have the highest recall

rates. On the other hand, the Skin-damaged and Immature classes are predicted with high accuracy. In conclusion, the DenseNet method achieves high and balanced scores across all criteria.

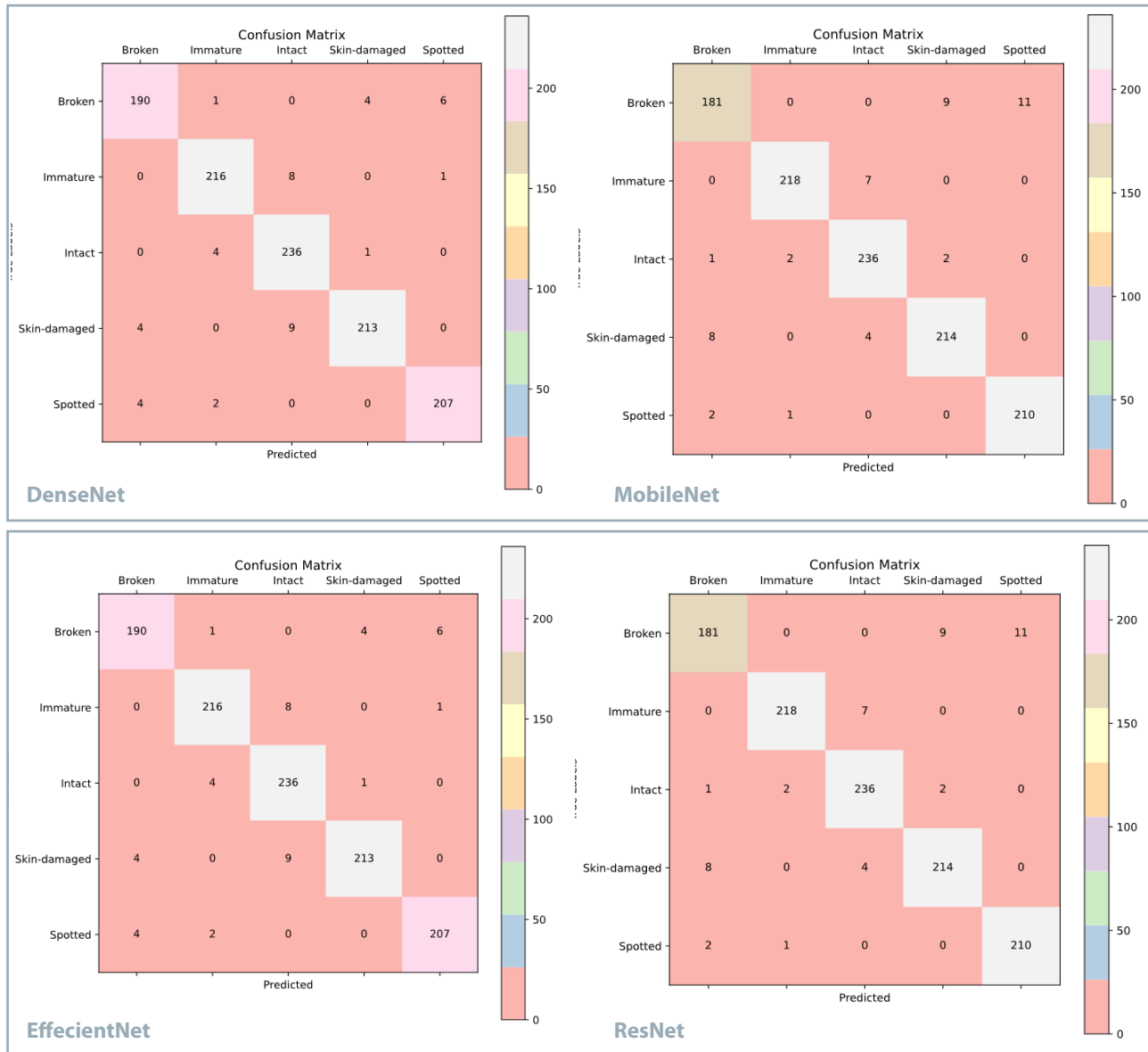


Fig. 3 Confusion matrix of proposed classifiers

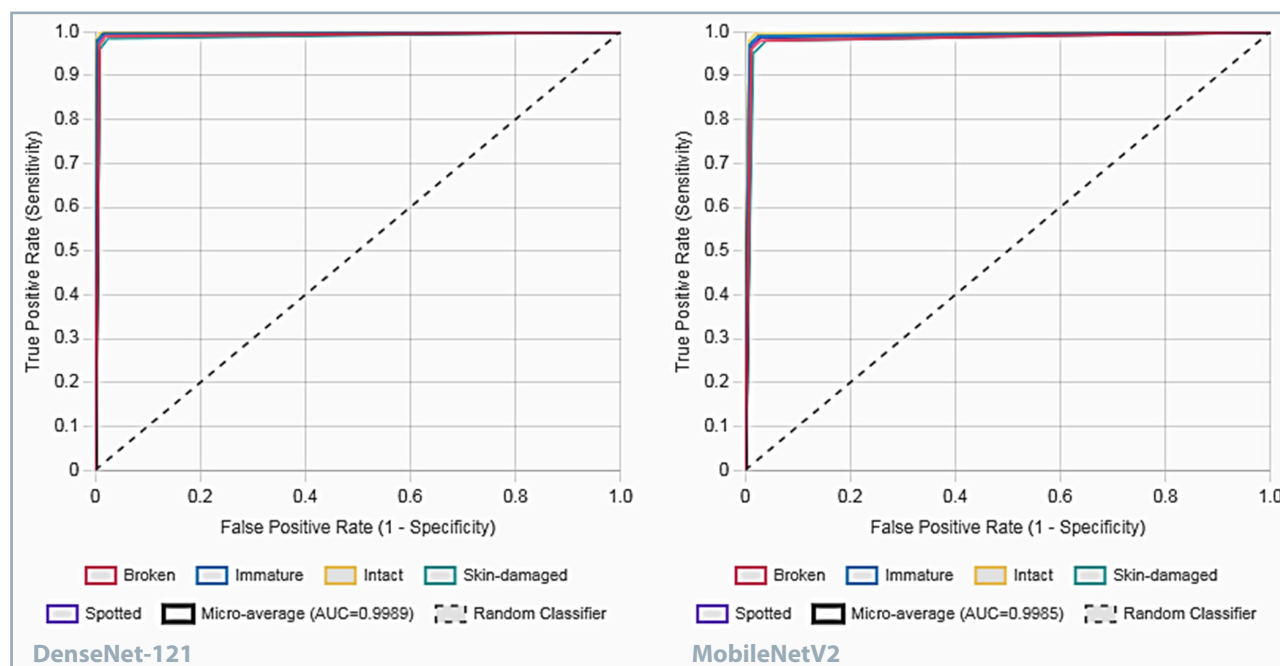
Table 2 Evaluation criteria extracted from confusion matrix for all proposed classifiers

Model name	Class	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
DenseNet	broken	93.4	91.0	92.2	95.7
	immature	97.8	96.9	97.3	
	intact	95.2	98.3	96.7	
	skin-damaged	97.7	93.8	95.7	
	spotted	94.1	97.7	95.9	
ReseNet	broken	96.0	94.5	95.2	96.0
	immature	96.9	96.0	96.4	
	intact	93.3	97.9	95.5	
	skin-damaged	97.7	94.2	95.9	
	spotted	96.7	97.2	97.0	
EfficiencyNet	broken	94.3	90.0	92.1	95.8
	immature	98.6	96.9	97.8	
	intact	95.5	97.9	96.7	
	skin-damaged	95.1	94.7	94.9	
	spotted	95.0	98.6	96.8	
MobileNet	broken	93.1	94.0	93.6	95.5
	immature	96.4	96.0	96.2	
	intact	94.4	97.1	95.7	
	skin-damaged	98.1	92.9	95.5	
	spotted	95.4	97.2	96.3	

ROC curve of all proposed classifiers

Figure 4 shows the receiver operating characteristic (ROC) curves used to evaluate the performance of a model at all classification thresholds. The area under the curve (AUC) is a key metric derived from these plots that quantifies the overall discriminatory power of the model. All models

showed excellent discriminatory ability, achieving average AUC values above 0.99. Specifically, the AUC values were 0.9989 for DenseNet, 0.9985 for MobileNetV2, 0.9992 for ResNet, and 0.9988 for EfficientNet. This indicates an outstanding ability to discriminate between different initial classes.

**Fig. 4** Comparison of ROC curves for the best model (DenseNet-121) and the most computationally efficient model (MobileNetV2)

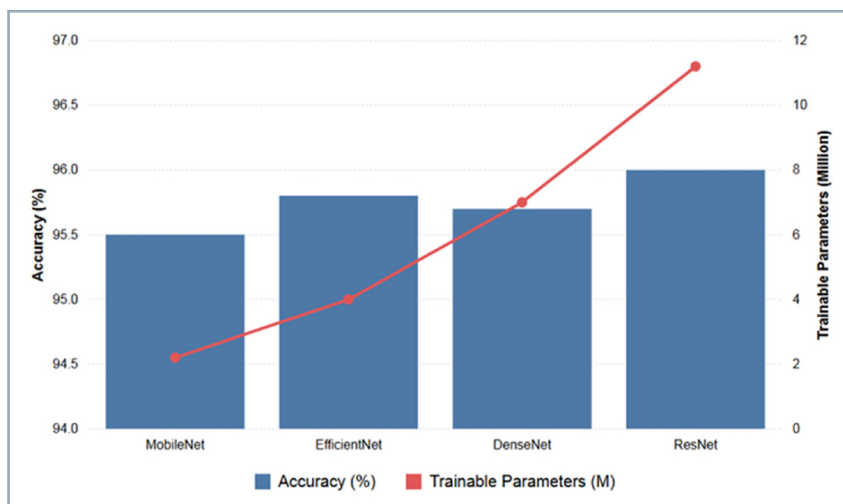


Fig. 5 Accuracy (%) vs. number of trainable parameters (million) in all proposed classifiers

The high values of AUC (averaged at the micro scale $AUC = 0.9989$ for DenseNet and 0.9985 for MobileNetV2) indicate exceptional discrimination ability for both models. The dotted diagonal line represents a random classifier ($AUC = 0.5$).

Figure 5 shows an interesting comparison between the number of trainable parameters of different models and the accuracy obtained by each model. As can be seen, DenseNet-121 and ResNet-18 have the highest number of trainable parameters, while MobileNet with a lower number of trainable parameters (and therefore lower complexity) achieved an accuracy of over 94%. This makes MobileNet a candidate for real-time implementation of soybean sorting systems, where processing power and speed are critical constraints.

The findings indicated the success of deep learning methods in classifying soybean classes. Among the models, DenseNet-121 achieved the highest performance, consistent with Zhang et al. (2023), who reported an accuracy of over 98% for DenseNet after applying data balancing techniques. Also, this study demonstrated significant success in scaling the EfficientNet hybrid model, aligning with the results of Diones et al. (2025), who achieved a validation accuracy exceeding 93% on a similar dataset. Furthermore, the DenseNet model in our study achieved an accuracy of 96%, which is comparable to the 96.02% reported by Xia et al. (2025) using a CNN-ViT

hybrid architecture. This indicates that standard CNNs optimised for the size of more complex hybrid models are effective. In terms of performance, MobileNet is an excellent choice for real-time soybean sorting systems with limited computational resources due to its much lower trainable parameters, achieving accuracies exceeding 95%. This is especially significant when compared to more advanced models, such as the lightweight SSDINet by Sable et al. (2024), which achieves a higher accuracy of 98.64%. Nevertheless, it establishes a benchmark for the further optimisation of lightweight architectures.

According to the confusion matrix analysis, most of the classification errors occurred between similar visual categories, such as “broken” and “damaged skin”. This challenge is common in fine-grained visual classification tasks and highlights areas for further improvement in model accuracy.

Conclusion

This study highlights the critical role of soybean seeds in global agriculture and the challenges associated with ensuring their quality. Advanced AI-driven methods have significantly improved soybean seed classification by accurately identifying defects such as damage, immaturity, and breakage. Among the evaluated models, DenseNet exhibited the best performance due to its efficient feature reuse and gradient flow, achieving

the highest accuracy and smooth training curves. MobileNet, ResNet, and EfficientNet also demonstrated strong performance, with precision, recall, and F1-scores exceeding 91% across all classes.

From a practical perspective, these results are very promising. The trained models, especially the lightweight MobileNetV2, can be integrated into a real-time sorting system. Such a system, consisting of a conveyor belt, a high-resolution camera, and an embedded processing unit, can automatically classify and sort thousands of grains per minute. This approach has the potential to significantly increase productivity, minimise the need for manual labour, and ensure higher quality outputs. Although our study is primarily computational, an important next step involves developing a physical prototype in both laboratory and field settings to validate the practical application of these findings. Despite the very hopeful results, there are three major limitations in this study. The first limitation is related to the dataset, in fact, this study was conducted on a dataset with 5 soybean classes, while there are different soybean quality classes worldwide that were not considered in this study. The second limitation is that the images were taken under controlled light conditions and it is better to take images in ambient conditions. Finally, the third limitation is that classification was performed on single seeds, which would be interesting if the study were conducted on images with several soybean seeds. Of course, it should be noted that for each type of limitation mentioned, a new problem definition must be made, and the study must be handled from scratch.

Future work should focus on the following to overcome these limitations:

1. Expanding the dataset by collecting more diverse samples from multiple sources, covering a broader spectrum of defect classes and varying lighting conditions.

2. Applying advanced augmentation techniques, such as adding noise (e.g., Gaussian, salt-and-pepper) and altering brightness and contrast, to enhance the model's robustness and generalisation capability.
3. Designing a comprehensive end-to-end system that incorporates object detection or segmentation to isolate individual seeds from bulk images before performing classification.

References

- Adnan, A., Ramdhani, T.I., Suryana, Y., Aziz, A., Rochmadi, T., Kamaruddin, A., Faiza, N.N., 2026. Comparative study of NIR-based machine learning for predicting soil nutrients in Indonesian farmlands. *Acta Technologica Agriculturae* 29(1), 27–34. <https://doi.org/10.2478/ata-2026-0004>
- Bern, C.J., Hanna, H.M., Wilcke, W.F., 2008. 3 – Harvesting, storing, and post-harvest management of soybeans. In Johnson, L.A., White, P.J., Galloway, R. (eds.), *Soybeans, Chemistry, Processing, and Utilization*. AOCs Press, Champaign, IL, USA, pp. 67–92. <https://doi.org/10.1016/B978-1-893997-64-6.50006-8>
- Chen, M., Chang, Z., Jin, C., Cheng, G., Wang, S., Ni, Y., 2025. Classification and recognition of soybean quality based on hyperspectral imaging and random forest methods. *Sensors* 25(5), 1539. <https://doi.org/10.3390/s25051539>
- De Oliveira, V.R., Souza, N.O.S., da Silva Arantes, M., Carmona, R., 2024. Artificial intelligence applied to the classification of greenish seeds and prediction of physiological quality in soybean. *Ciência e Agrotecnologia (Science and Agrotechnology)* 48, e003624. <https://doi.org/10.1590/1413-7054202448003624>
- Dioses, I.A.M., Dioses, J.L., Aquino, R.N., Feliciano, C.G.M., Tababa, J.B., Hermosura, L.F., 2025. Soybean seed quality assessment classification using EfficientNet B0 algorithm. In 2025 21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA), pp. 144–149. Pulau Pinang, Malaysia. <https://doi.org/10.1109/CSPA64953.2025.10933286>
- Dragan, M., Mijić, K., Jakšić, D., 2018. Opportunistic management behavior in reporting earnings of agricultural companies. *Custos e Agronegócio on line* 14(1), 125–142. <http://www.custoseagronegocioonline.com.br/numero1v14/OK%206%20opportunistic%20english.pdf>
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. Las Vegas, NV, USA. <https://doi.org/10.1109/CVPR.2016.90>
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. Honolulu, HI, USA. <https://doi.org/10.1109/CVPR.2017.243>
- Jitanan, S., Chimlek, P., 2019. Quality grading of soybean seeds using image analysis. *International Journal of Electrical and Computer Engineering (IJECE)* 9(5), 3495–3503. <https://doi.org/10.11591/ijece.v9i5.pp3495-3503>
- Lin, W., Fu, Y., Xu, P., Liu, S., Ma, D., Jiang, Z., Zang, S., Yao, H., Su, Q., 2023. Soybean Seeds. Mendeley Data. Version 6. <https://doi.org/10.17632/v6vzvfszj6.6>
- Pagano, M.C., Miransari, M., 2016. 1 – The importance of soybean production worldwide. In Miransari, M. (ed.), *Abiotic and Biotic Stresses in Soybean Production; Soybean Production: Volume 1*, pp. 1–26. Academic Press, Cambridge, MA, USA. <https://doi.org/10.1016/B978-0-12-801536-0.00001-3>
- Pereira, G.M.L., Foleis, J.H., de Souza Brito, A., Bertolini, D., 2024. A database for soybean seed classification. In 2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 1–6. IEEE, Manaus, Brazil. <https://doi.org/10.1109/SIBGRAPI62404.2024.10716268>
- Sable, A., Singh, P., Kaur, A., Driss, M., Boulila, W., 2024. Quantifying soybean defects: A computational approach to seed classification using deep learning techniques. *Agronomy* 14(6), 1098. <https://doi.org/10.3390/agronomy14061098>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. MobileNetV2: Inverted residuals and linear bottlenecks. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4510–4520. Salt Lake City, UT, USA. <https://doi.org/10.1109/CVPR.2018.00474>
- Shirzad, N., Shahgholi, G., Ardabili, S., Szymanek, M., 2025. Investigating the effects of spectroscopic method in estimating soluble solid content values and firmness of cherries from an environmental point of view: Prediction of environmental parameters with machine learning method. *Acta Technologica Agriculturae* 28(1), 17–25. <https://doi.org/10.2478/ata-2025-0003>
- Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (PMLR) 97, pp. 6105–6114. Long Beach, CA, USA. <https://proceedings.mlr.press/v97/tan19a.html>
- Wang, D., Dowell, F.E., Ram, M.S., Schapaugh, W.T., 2004. Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. *International Journal of Food Properties* 7(1), 75–82. <https://doi.org/10.1081/JFP-120022981>
- Wang, D., Ram, M.S., Dowell, F.E., 2002. Classification of damaged soybean seeds using near-infrared spectroscopy. *Transactions of the ASAE* 45(6), 1943–1948. <https://doi.org/10.13031/2013.11410>
- Xia, Y., Zhu, R., Ji, F., Zhang, J., Chen, K., Huang, J., 2025. Soybean seed classification and identification based on corner point multi-feature segmentation and improved MobileViT. *AgriEngineering* 7(10), 354. <https://doi.org/10.3390/agriengineering7100354>
- Yang, X., Ma, K., Zhang, D., Song, S.H., An, X., 2024. Classification of soybean seeds based on RGB reconstruction of hyperspectral images. *PLOS ONE* 19(9), e0307329. <https://doi.org/10.1371/journal.pone.0307329>
- Zhang, N., Zhang, E., Li, F., 2023. A soybean classification method based on data balance and deep learning. *Applied Sciences* 13(11), 6425. <https://doi.org/10.3390/app13116425>

