Available on: http://gjetr.ep-journals.org/ E-ISSN: 3051-3782, P-ISSN: 3051-3774

Global Journal of Engineering and Technology Review

Volume: 02 Issue: 02 February 2026

Soybean Seed Classification with Fusion of Convolutional Neural Networks and ViT Technique

Ghazwan Hani Hussien

Department of Computer Science, College of Computer Science and Information Technology, University of Basra, Basra, Iraq

Ahmed Khalaf Zager Alsaedi

Physic department, Collage of science, Misan University, Iraq

Abdul Hakim Moawia Dreheeb

Computer Science Department, Faculty of Humanities and Applied Sciences, Al-Zaytuna University, Tarhuna, Libya.

Published Date: February 14, 2026 **Article DOI : 10.65150/EP-gjetr/V2E2/2026-03**

ABSTRACT: Soybeans are a nutritious legume cultivated worldwide and are a staple in the human diet. They also play a vital role in animal feed production. Furthermore, soybean oil, extracted from soybeans, is widely used in the food industry. Several different convolutional neural network (CNN) models, along with image transformation (ViT) models, were used to classify soybean seeds. The seed classification comprised five categories: broken, immature, intact, shell-damaged, and defective. To enhance the generalizability of the dataset, data augmentation techniques were applied during the training process. Seventeen pre-trained CNN and ViT models with varying architectures were tested. Model performance was evaluated using performance metrics such as accuracy, fine-tuning, recall, and F1. The proposed model was constructed by combining feature extraction layers from the top-performing ViT-L/16, DenseNet201, and DenseNet169 models. The proposed model outperformed all other models, achieving an accuracy of 96.02% in the classification process. With this proposed model, we have obtained a hybrid approach to classifying soybean seeds, and it contributes to the use of artificial intelligence in the agricultural sector.

KEYWORDS: Soybean classification, Deep Learning, Convolutional Neural Networks, Vision Transformer, Feature Fusion

1. INTRODUCTION

Soybeans are a nutritious legume species grown all around the world. Soybeans occupy about 47% of the world's agricultural area[1]. In 2017, global soybean production reached 348.9 million tons, and 152.4 million tons were traded via shipments [2]. Soybeans are an essential component for human nutrition. Due to the high protein content, soybeans are frequently preferred as a protein source in place of meat, especially in vegetarian and vegan diets[3]. Additionally, it contains numerous vitamins, minerals, and antioxidants, making it a food item that aids in improving overall health. Moreover Soybean plays a critical role in animal feed production[4], particularly serving as a primary protein source for cattles[5]. Furthermore, soy oil derived from soybeans is widely used in the food industry[6]. It can be found as a main ingredient in many products such as margarine, mayonnaise, and salad sauces. Soybean oil also emerges as a significant raw material in biodiesel production[7].

Quality of the soybean seed is directly related with the yield and nutritional value of the products derived from it. Traditionally, a mechanical vibrating sieve system is used to detect damaged or low-quality seeds, or separated by human eye based on physical characteristics such as size, shape, color, and texture[8]. While the sieve method can damage seeds, visual classification is prone to error and sometimes does not produce accurate results. Therefore, there is an increasing need for more sophisticated and precise methods. The use of image processing and artificial intelligence technologies in the agricultural sector has become increasingly popular in recent years. Seed classification studies are one of these application areas. An overview of a few studies conducted using computer vision and artifical intelligence tecniques in the classification of seed. Xia et al.(2019) [9]utilized hyperspectral imaging to distinguish maize seed varieties.

They applied a wavelength selection and feature reduction / transformation technique based on MLDA algorithm and enhanced it by combining spectral and image features. Merging mean spectral feature with 5 first-order and 8 second-order textural features through the MLDA method proved to be effective for seed classification. Using the proposed method, they achieved an impressive 99.13% classification accuracy. Singh et al.[(2021) [10]developed a laser biospeckle method combined with machine learning to detect early-stage fungal infections in soybean seeds. The technique effectively identified Colletotrichum truncatum infections. Among the machine learning models, including k-NN, DT, LR, and ANN, DT was the best classifier, achieving a classification accuracy of 96.94% for infected seeds. Hamdi et al. (2022) [11]employed MobileNetV2, known its design efficiency in mobile and embedded, to classify various seeds. The authors used a dataset that contained 14 different seed types, including Sunflower, Onion, Kidney Beans, and Fenugreek. The model achieved accuracies of 98% in training and 95% in testing data. Çetin et al. (2022)[12] proposed a machine learning-based classification system for soybean seeds. They extracted features like shape, size, mass, volume, surface area, and sphericity from seed image.

They classified five varieties of soybean with various machine learning algorithms. For the Ceyhan and Traksoy varieties, they achieved classification accuracies of 90.00% using the RF algorithm and 89.00% with the MLP algorithm, respectively. Huang et al. (2022) [13] developed a two-step method for soybean seed classification, utilizing Mask R-CNN for segmentation and a novel model, Soybean Network (SNet), for classification. SNet incorporates a Mixed Feature Recalibration (MFR) module, enhancing its focus on key damage areas. The model achieved 96.2% classification accuracy. Li et al. (2023) [14] proposed a DL technique for online soybean seeds classification. Initially, they employed the multi-scale Retinex with color restoration (MSRCR) technique for segmentation. Subsequently, they used a CNN to classify the seeds into four categories normal, abnormal, damaged, also non-classifiable soybeans. They achieved F-scores of 95.97%, 97.41%, 97.25%, and 96.14% for each category, respectively, and achieved an average accuracy of 95.63%

Table 1. Summary of literature

Study	Year	Method Used	Details	Classification Accuracy / Results
Xia et al. [9]	2019	Hyperspectral imaging with MLDA	Combined spectral and image features for maize seed variety classification.	99.13% accuracy
Singh et al. [10]	2021	Laser biospeckle and machine learning	Detected early-stage fungal infections in soybean seeds.	96.94% with DT
Hamdi et al. [11]	2022	MobileNetV2	Classified various seeds including Sunflower, Onion, etc.	98% training, 95% testing
Çetin et al. [12]	2022	Machine learning-based system	Extracted physical features to classify soybean varieties.	90% with RF, 89% with MLP
Huang et al. [13]	2022	Mask R-CNN and SNet	Used for soybean seed classification focusing on damage areas.	96.2% accuracy
Li et al. [14]	2023	Deep learning with MSRCR and CNN	Online classification of soybean seeds into categories.	F-scores 95.97% - 97.41% - 97.25% - 96.14%, Avg. 95.63%

Existing methods for classifying seeds have achieved high accuracy rates. However, the approaches used in these studies typically rely on a single model, and outcomes can vary under different situations and conditions.

Our main contributions as follows;

- Seventeen pre-trained CNN models with various blocks like dense, inception, residual, separable convolution, also stacked were used for experimentation.
- Additionally, ViT models with varying patch and layer sizes were also tested.
- Vision Transformers have been introduced to this field for the first time in the literature.
- Among the CNN and ViT models, three top-performing models were selected as feature extractors, also features from these models were concatenated. The proposed model demonstrated robustness and high capacity for generalization in soybean seed classification.

Remainder of paper is structured as follows Section two outlines materials also methods employed in this study, including dataset also data augmentation techniques. Section 3 presents experimental setup and various CNN and ViT models tested. Section 4 discusses results by comparing different models performance. Finally, Section 5 concludes the paper by summarising the main findings and suggesting future research directions.

2. MATERIALS AND METHODS

This section describes in detail Materials and Methods used in this study, which was started with a publicly available dataset. Data augmentation techniques including rotations, shifts, shear transformations and zoom adjustments were applied to improve the generalisability of the model. Seventeen pre-trained CNN models such as DenseNet, Inception, MobileNet, NASNet, ResNet, VGG and Xception, as well as various ViT models with different patch and layer sizes were evaluated in the experiments. Features from three best performing models were combined to form a robust feature vector, resulting in a model with high accuracy and generalisation capacity for soybean seed classification. Figure 1. shows schematic diagram of the proposed study.

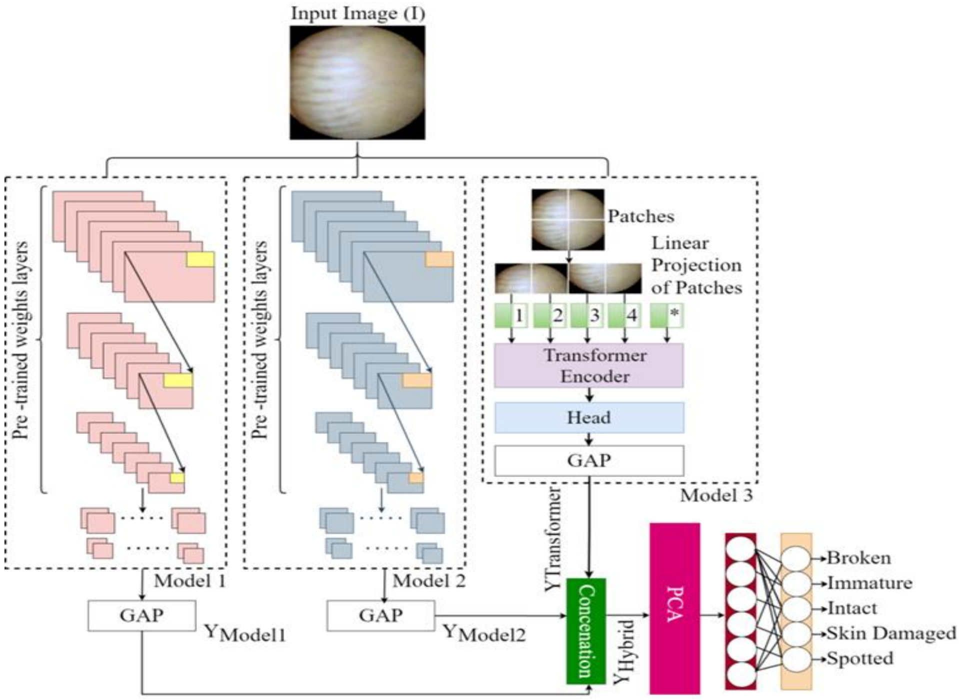

Fig. 1. Schematic Diagram of the proposed study

2.1. DATASET

In this study, we used a publicly available dataset [15] [16] contains 5513 soybean images in 5 categories: broken, immature, intact, skin damaged, also spotted. Table 2 shows distribution of data.

Table 2. Details of dataset

Broken	Immature	Intact	Skin Damaged	Spotted	Total
1002	1125	1201	1127	1058	5513

Moreover, data augmentation techniques were applied to increase generalization capability of model. Parameter details are shown in Table 3.

Table 3. Data Augmentation Parameters

Parameter	Rotation range	Width shift range	Height shift range	Shear range	Zoom range
Value	0.2	0.3	0.4	0.5	0.6

Sample images for each category are presented in Fig. 2.

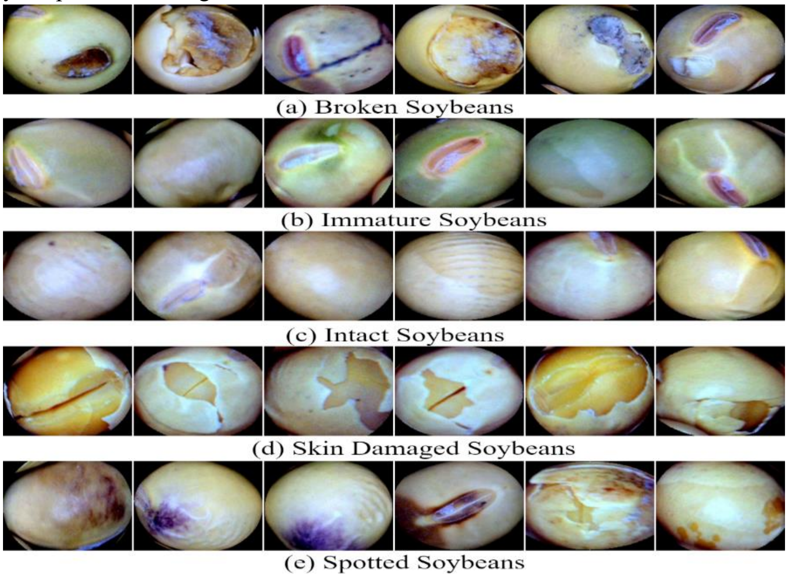

Fig. 2. Sample soybean images from dataset

2.2 CONVOLUTIONAL NEURAL NETWORKs (CNNs)

Convolutional Neural Networks (CNNs) are deep learning algorithms inspired by the mammalian visual cortex system[17]. CNNs are widely used in image analysis and have achieved successful results in various fields such as medical[18], education[19], and agriculture[20]. A CNN contains three major components the convolution, pooling, also fully connected layers. Convolution layer extracts features from an image using a kernel and creates a feature map. Pooling layer reduces size of feature map which in turn reduces computational complexity. Final fully connected layer is classification layer that uses neural networks and a classifier function, such as softmax. Using these three core components, various architectures have been proposed. In experimental studies, seventeen well-known pre-trained CNN models, namely DenseNet121, DenseNet169, DenseNet201, InceptionResNetV2, InceptionV3, MobileNet, MobileNetV2, NASNetMobile, ResNet50, ResNet50V2, ResNet101, ResNet152, ResNet152V2, VGG16, VGG19, and Xception, which contain different blocks such as dense, inception, residual, separable convolution, and stacked, were tested.

2.3. VISION TRANSFORMER

ViT is a deep learning model used in image processing and, unlike classical CNNs, it processes images by dividing them into small patches. ViT learns the features of each patch in sequence, thus obtaining information about the whole image. ViT uses the transformer mechanism to analyse image data; it allows it to learn the relationship between each part and the others. ViT, which performs especially well on large datasets, can effectively learn complex image features and achieve high accuracy rates in tasks such as classification and object recognition. A typical ViT architecture is shown in Fig. 3.

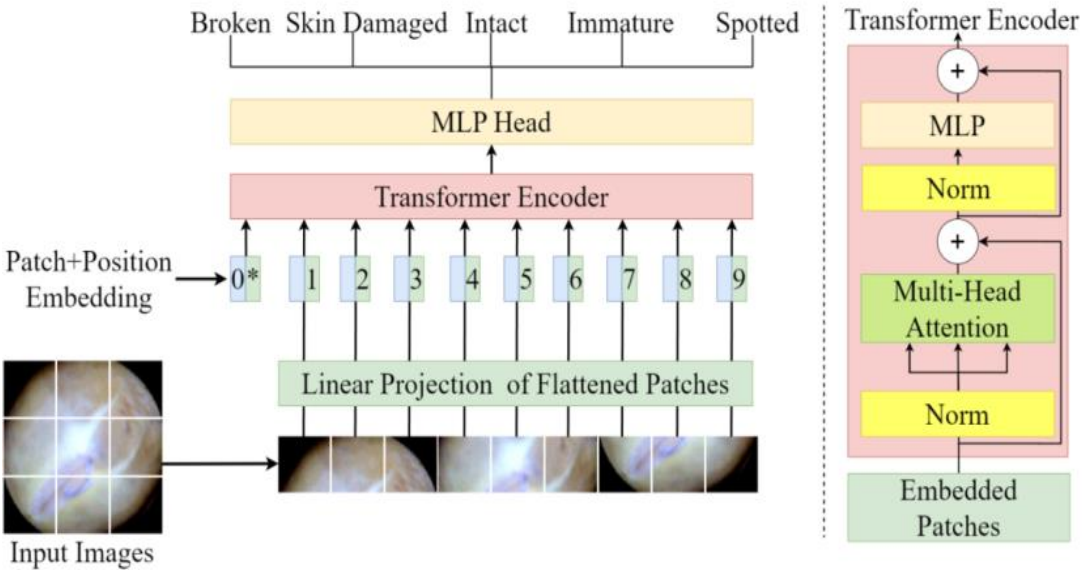

Fig. 3. Classical ViT architecture

2.4. TRANSFER LEARNING

Transfer learning (TF) is the process of transferring knowledge from one task to another. In CNNs, TF means that a model is first trained on one dataset and then, using its weights, is retrained on another dataset. The model learns basic information from the first dataset and better adapts to the new dataset. In this study, all the CNNs used for experiments were previously trained on ImageNet dataset. ImageNet is a vast visual database designed for use in visual object recognition software research and contains 1000 classes, each represented by thousands of images. The final classification layer of models of used was removed, and a Global Average Pooling (GAP) layer and softmax function were added for classification. Main reason for using GAP is to effectively reduce the size of the feature vector. Fig. X shows the transfer learning process from a model trained on ImageNet dataset to the Soybean dataset.

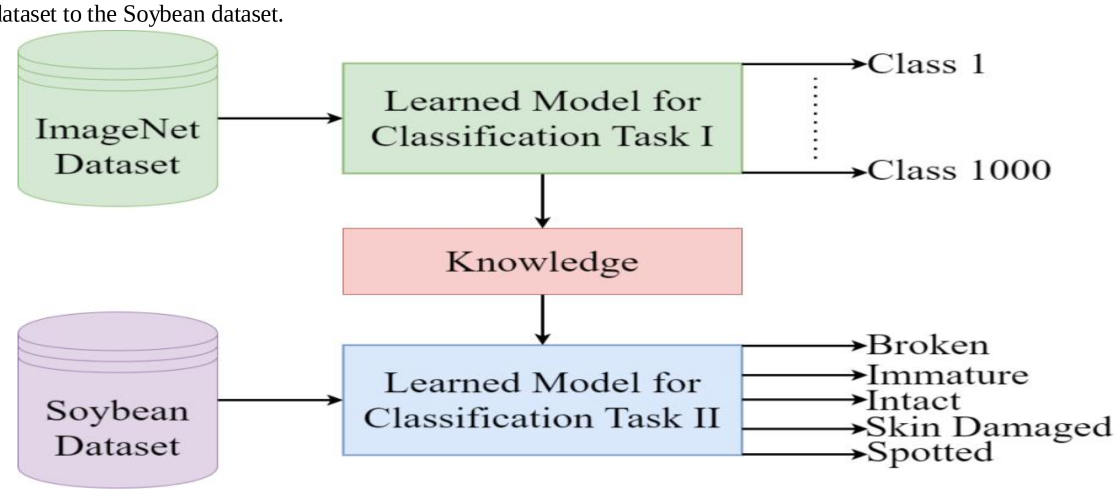

Figure 4. Transfer Learning Framework from ImageNet to Soybean Dataset.

2.5 FEATURE CONCATENTION

Feature concatenation is a common technique in which features from multiple sources, using different techniques, are combined into a single feature vector. Combining information from multiple feature sources results in a robust set of complementary features that enhance the model's performance. In experimental studies on CNNs and ViT models, three most successful models were selected as feature extractors, and their final layers were concatenated to create a feature vector.

3. EXPERIMENTAL SETUP

Classification models are evaluated using various metrics. Accuracy measures the ratio of correctly predicted samples. Precision assesses how many of samples predicted as positive are actually positive. Recall indicates how many of all actual positive samples were correctly predicted. F1 Score represents harmonic mean of precision also recall. In this context, True Positive (TP) refers to samples correctly identified as positive, True Negative (TN) to samples correctly identified as negative, False Positive (FP) to negative samples mistakenly labeled as positive, and False Negative (FN) to positive samples incorrectly labeled as negative. The mathematical representation of these metrics spans from Eq. 1 to Eq. 4.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Precision} = TP/(TP+FP) \quad (2)$$

$$\text{Recall} = TP/(TP+FN) \quad (3)$$

$$\text{F1 Score} = 2 \times (\text{Precision} \times \text{Recall})/(\text{Precision}+\text{Recall}) \quad (4)$$

4. RESULTS

In this study, performance of various CNN also ViT models in classifying soybean images is compared. Performance of models is evaluated on key classification metrics such as accuracy, precision, recall also F1 score. Results show how different model structures perform in the classification of soybeans. Table 3. shows classification performance of models. The three models with the highest accuracy are ViT-L/16, DenseNet201 and DenseNet169. ViT-L/16 is highest performing model with an accuracy of 95.74%; it also showed very high results in other metrics like precision, recall also F1-score. DenseNet201 followed ViT-L/16 with 94.57% accuracy, followed by DenseNet169 with 94.38% accuracy. These models have achieved very successful results in terms of classification performance and are remarkable for their high accuracy and consistency in other metrics. The models with the lowest accuracy are ResNet50 (88.41%), VGG19 (89.23%) and VGG16 (90.95%). In terms of classification performance, these models presented relatively lower results and could not approach the performance of more advanced models such as ViT and DenseNet. In particular, the low performance of models such as VGG19 and VGG16 in metrics like accuracy also F1 score reveals the advantage of more complex model structures in classifying soybean images. In this study, a proposed model was proposed by combining the feature extraction layers of the three highest performing models (ViT-L/16, DenseNet201 and DenseNet169). This model, called Proposed Model, achieved highest classification success in the study with an accuracy rate 96.02%.

Table 3. Comparisons of the results between CNN, ViT Models and proposed method

Model	Accuracy	Precision	Recall	F1-Score
DenseNet121	0.90950	0.92178	0.90950	0.91105
DenseNet169	0.94389	0.94531	0.94389	0.94397
DenseNet201	0.94570	0.94679	0.94570	0.94554
InceptionResNetV2	0.92217	0.92386	0.92217	0.92213
InceptionV3	0.92670	0.92843	0.92670	0.92667
MobileNet	0.92850	0.92997	0.92851	0.92838
MobileNetV2	0.92941	0.93166	0.92941	0.92949
NASNetMobile	0.92488	0.92521	0.92489	0.92452
ResNet50	0.88416	0.89434	0.88416	0.88439
ResNet50V2	0.91583	0.91888	0.91584	0.91563
ResNet101	0.91855	0.92037	0.91855	0.91797
ResNet101V2	0.93575	0.93701	0.93575	0.93559
ResNet152	0.91855	0.92207	0.91855	0.91816
ResNet152V2	0.93574	0.93641	0.93575	0.93557
VGG16	0.90950	0.91275	0.90950	0.90872
VGG19	0.89231	0.89850	0.89231	0.89181
Xception	0.92488	0.92487	0.92489	0.92474
ViT-B/16	0.93122	0.93204	0.93122	0.93125
ViT-B/32	0.93484	0.93559	0.93484	0.93482
ViT-L/16	0.95747	0.95763	0.95747	0.95737
ViT-L/32	0.94210	0.94124	0.94443	0.94123
Proposed	0.96018	0.96042	0.96018	0.96018

The fact that this combined model achieved higher accuracy and consistency in other metrics than the individual models indicates that combining different model structures improves classification performance. To sum up, the ViT and DenseNet models used in this study provided good results in soybean classification, and the proposed model improved this performance even further. A more robust and highly accurate classification model was produced by integrating different model structures, thus achieving great success in the categorisation of soybean images. These results demonstrate the benefits of model fusion in visual classification and provide an important reference point for future work.

5. CONCLUSION

This study examines the utilization of CNN also ViT models for the image-based classification of soy bean seeds. The results demonstrate that advanced models, such as ViT and DenseNet, exhibit superior classification performance. Furthermore, the classification success can be enhanced by integrating diverse model features. Proposed model exhibits highest performance with an accuracy rate 96.02%, indicating that it is both feasible and effective to automatically classify agricultural products through image processing and artificial intelligence techniques. This study provides a crucial reference for the advancement of artificial intelligence-based solutions in the agricultural sector in the future.

DATA AVAILABILITY STATEMENT

All data, models and code generated or used during the study appear in the submitted article.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declare(s) that there is no conflict of interest.

REFERENCES

1) J. Pereira, M. Lopes, J. Parish, A. Silva, and M. PicanÇO, "Impact of RR soybeans and glyphosate on the community of soil surface arthropods," *Planta Daninha,* vol. 36, p. e018171324, 2018.
2) K.-C. Lee *et al.*, "Rapid Identification Method for Gamma-Irradiated Soybeans Using Gas Chromatography–Mass Spectrometry Coupled with a Headspace Solid-Phase Microextraction Technique," *Journal of agricultural and food chemistry,* vol. 68, no. 9, pp. 2803-2815, 2020.
3) B. Behailu and M. Abebe, "Effect of soybean and finger millet flours on the physicochemical and sensory quality of beef meat sausage," *Asian Journal of Chemical Sciences,* vol. 7, no. 1, pp. 6-14, 2020.
4) X. Zhu, W. L. Leiser, V. Hahn, and T. Würschum, "Identification of QTL for seed yield and agronomic traits in 944 soybean (Glycine max) RILs from a diallel cross of early-maturing varieties," *Plant Breeding,* vol. 140, no. 2, pp. 254-266, 2021.
5) M. A. McNiven, P. Robinson, and J. MacLeod, "Evaluation of a new high protein variety of soybeans as a source of protein and energy for dairy cows," *Journal of dairy science,* vol. 77, no. 9, pp. 2605-2613, 1994.
6) M. Friedman and D. L. Brandon, "Nutritional and health benefits of soy proteins," *Journal of agricultural and food chemistry,* vol. 49, no. 3, pp. 1069-1086, 2001.
7) A. E. Atabani, A. S. Silitonga, I. A. Badruddin, T. Mahlia, H. Masjuki, and S. Mekhilef, "A comprehensive review on biodiesel as an alternative energy resource and its characteristics," *Renewable and sustainable energy reviews,* vol. 16, no. 4, pp. 2070-2093, 2012.
8) P. Lin *et al.*, "Rapidly and exactly determining postharvest dry soybean seed quality based on machine vision technology," *Scientific Reports,* vol. 9, no. 1, p. 17143, 2019.
9) C. Xia, S. Yang, M. Huang, Q. Zhu, Y. Guo, and J. Qin, "Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis," *Infrared Physics & Technology,* vol. 103, p. 103077, 2019.
10) P. Singh *et al.*, "Development of an intelligent laser biospeckle system for early detection and classification of soybean seeds infected with seed-borne fungal pathogen (Colletotrichum truncatum)," *Biosystems Engineering,* vol. 212, pp. 442-457, 2021.
11) Y. Hamid, S. Wani, A. B. Soomro, A. A. Alwan, and Y. Gulzar, "Smart seed classification system based on MobileNetV2 architecture," in *2022 2nd international conference on computing and information technology (iccit)*, 2022: IEEE, pp. 217-222.
12) N. Çetin, "Machine learning for varietal binary classification of soybean (Glycine max (L.) Merrill) seeds based on shape and size attributes," *Food Analytical Methods,* vol. 15, no. 8, pp. 2260-2273, 2022.
13) Z. Huang *et al.*, "Deep learning based soybean seed classification," *Computers and Electronics in Agriculture,* vol. 202, p. 107393, 2022.
14) W. Lin, L. Shu, W. Zhong, W. Lu, D. Ma, and Y. Meng, "Online classification of soybean seeds based on deep learning," *Engineering Applications of Artificial Intelligence,* vol. 123, p. 106434, 2023.
15) "Kaggle Soybean Seeds Dataset." [Online]. Available: https://www.kaggle.com/datasets/warcoder/soyabean-seeds.
16) W. Lin *et al.*, "Soybean image dataset for classification," *Data in Brief,* p. 109300, 2023.
17) C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
18) Ünlükal, N., Ülker, E., Solmaz, M., Uyar, K., & Taşdemir, Ş. (2024). Histological tissue classification with a novel statistical filter-based convolutional neural network. Anatomia, Histologia, Embryologia, 53(4), e13073.
19) J. Wang, S. C. Satapathy, S. Wang, and Y. Zhang, "LCCNN: a lightweight customized CNN-based distance education app for COVID-19 recognition," *Mobile Networks and Applications,* pp. 1-16, 2023.
20) Cengel, T. A., Gencturk, B., Yasin, E. T., Yildiz, M. B., Cinar, I., & Koklu, M. (2024). Apple (Malus domestica) Quality Evaluation Based on Analysis of Features Using Machine Learning Techniques. Applied Fruit Science, 1-11.

Corresponding Author: Ghazwan Hani Hussien