

# A Survey on Soybean Seed Varieties and Defects Identification Using Image Processing

Amar V. Sable<sup>1</sup>, Parminder Singh<sup>1,3</sup>, Jatinder Singh<sup>2</sup> and Mustapha Hedabou<sup>3</sup>

<sup>1</sup> School of Computer Science and Engineering, Lovely Professional University 1, Punjab, India

<sup>2</sup> School of Agriculture, Lovely Professional University 2, Punjab, India

<sup>3</sup> School of Computer Science, Mohammed VI Polytechnic University 3, Ben Guerir, Morocco

## Abstract

Agriculture has a leading part in the country's economy's growth. Specifically, in developing countries like India, agriculture has a major impact on the country's overall economy. Climate and other environmental changes in agriculture have become a serious challenge. Farmer is using various modern techniques to make the field more attractive and by using advanced techniques, farmers can plant the right crop in the ideal location to maximize crop yield production. To get the maximum yield of the crop, the beginning of the selection of seed for the crop is very important. In this paper, we reviewed various research works on seed variety identification and defect identification using various techniques. Depending upon the existing related work, we had identified the future scope to determine seed varieties/defects using computer vision and neural network technology which will help the farmers to select the right type of quality seed from available seed quality to achieve maximum crop yield in the farm.

## Keywords

Computer vision, deep learning, neural network, variety identification, seed defects.

## 1. Introduction

Soybean is an essential yield that is extensively used since it is an abundant source of minerals, a great content of protein and a substantial amount of oil [1]. The quality of soybeans marks the price and value of grain used for cultivation and feasting. Soybean infections have an important influence on crops' economic value, leading to considerable financial losses, for both the farmers and soybean production also. It is thus highly vital for soybeans and soybean growers to determine the appearance quality of their soybeans quickly and accurately [2]. Different diseases of soya affect the appearance of the seeds, including their size, shape and color. Soybeans affected by the disease may have lilac seeds, purple seed, wrinkled seeds, green seeds, tiny/split seeds and other signs. Soybean output in most cases is reliant on the grain quality, and classification quality is thus extremely essential for soybeans and soybeans producers.

To divide soybeans into various classes, a soybean screening machine is utilized. Only non-standard materials and seeds can be separated using this equipment. In contrast, it is impossible to identify by machine low-quality seeds, including dry seeds, green seeds, violet seeds and pollutants as large as ordinary soybean seeds. As a result, highly skilled employees evaluate the quality of soya grain. This technique requires many individuals to take part in this, as well as being time-consuming and sensitive to human error.

ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States

EMAIL: [amar13.sable@gmail.com](mailto:amar13.sable@gmail.com) (A. 1)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)



**Figure 1: a. Normal Soybean Seeds b. Defected Soybean Seed**



**Figure 2: Defected soybean seed type**

A further approach to assessing the quality of farm goods and meals is by using image processing technologies and machine learning approaches to evaluate their nutritional worth. Recently, these approaches have become more popular [3-8]. Mebatsion et al. is categorized by morphological or RGB color grain accuracy of 99.6%, including barley, oat, rye or wheat [9]. They used morphological and RGB color features to measure grains. Olgun et al. designed an automatic wheat grain classification approach having 88.33% accuracy, by utilizing the Dense Invariant Transforming Function (SIFT) assessed by Support vector machine (SVM) [10]. Another approach was developed with the use of picture descriptors and visual word bags, which are techniques resistant to occlusion [11]. Soya leaves were identified. Back-propagation of artificial neural networks (BP ANN's) was proposed by Tan et al. [12], for soybean disease detection and classification, for example, soybean frog-eye and mildewed soybean. In diverse soybean seeds with several diseases, 90% of the method is accurate, the scientists said. Color variations and shadow noise impeded the performance of the system with natural light, reducing its total efficiency. This idea needs to enable the light source used to uniformly stream light.

Although numerous modern imaging and machine learning techniques have been used for the improvement of soybean seeds quality, major constraints remain. There are still some key concerns that need to be fixed when seed quality through image processing is identified. Shadow noise can develop primarily when the angle of the camera and the quality of natural light are altered. This circumstance might lower the accuracy of the segmentation while also affecting the identification of the limits of soybean seed. Second, soybean seed color variations might impair the quality grading system's categorization capacity utilized for assessment. This indicates that the current colors separation research is not resilient to changes in light since the light source is customized to each specific data set. Color characteristics from the color model and components that are not subject to light fluctuations should, therefore, be incorporated to improve classification results. Third, the diversity of shape in every soybean seed class is a barrier to the soybean seed categorization. Because of these factors, combinations of color features and other features which are not form-based may increase categorization efficiency in each soybean seed class.

Deep learning (DL) is a vital artificial intellect method for machines to independently gain data knowledge [13]. Moreover, when the seed does not wet excessively (neither the seed color change nor the seed size expand), relatively moisture does not affect the seed image technique in combination with deep learning. Currently, deep networks have effectively been used to identify plant disease [14-16], monitor drought [17], classify land type [18], detect weeds, and other agricultural sectors. So far, there are few studies about deeper learning in the identification of soybean seed types and whether they also have unknown advantages.

## 2. Literature Survey

In 2020 Zhu et al. [20], designed a convolutional neural network (CNN) to detect varieties of six different soybean seeds using hyperspectral pictures, with 91% average identification rate. Here, after data augmentation, a total of 9600 pictures were collected, and the data-set were split into three parts: training set, validation set, and testing set, with ratio of 3:1:1. Transfer training was carried out using previously trained models that had been fine-tuned. It has designed and implemented the most efficient CNN model for soya seed variety detection. In addition, the standard machine learning models were developed for the identification of soya seed varieties using reflectance as input. The result shows that all of the 6 models achieved a validation set accuracy of 91% with an accuracy of 90.6%, 94.5%, 95.4%, 95.6%, 96.8% and 97.2% correspondingly were reached in the test set.

A general classification of rice identification methods may be made into two categories: chemical and physical. The prominent deep learning techniques for classification and recognition are included in the first of these models. Qiu et al. [21], in 2018 utilized a CNN to determine four different varieties of rice, and having 87% identification rate using 12,000 training samples. The performance of the model was improved by increasing the quantity of training data. The CNN architecture outperformed KNN, SVM in many instances, and showed the effectiveness of CNN in spectral data assessment. The output of this research work shows that CNN performs on par as compared to existing method for spectral data processing.

Weng et al. [22] in 2020, used a deep learning and principal component analysis (PCA) network, to train 4320 samples, resulting in the best identification rate of 98% for ten rice varieties and the best recognition rate of over 98% for ten rice types [23]. To detect rice variety, hyperspectral imaging (HSI) was used in conjunction with a deep learning network that took into account numerous characteristics such as spectroscopy, texture, and morphology. In china, HSI images of 10 popular rice varieties were determined by analyzing their contrast. In regions of interest, spectroscopy and morphology were extracted from high-resolution (HSI) photos and binary images. Monochromatic images were also used with distinguishing wavelengths which were highly related to various rice varieties used to generate the texture. The usage of a deep learning network, particularly the PCANet, were used to design identification models for rice variety, as well as machinery approaches such as KNN and random forests to compare models to the whole globe. The PCANet, in conjunction with these features, was employed. In order to remove spectral interference, Golay's First Order Savitzky, multivariate scatter correction, standard norm variation and Savitzky–Golay smoothing were used. To extract critical info from increased features and the principal component analysis, main component analysis (PCA) was used (PCA). Multi-character fusion enhanced the accuracy of identification, whereas PCANet had a major edge over other methods in classification performance. The best results were achieved by PCANet using PCA processed spectroscopic and textural features, with an accurate rating of 98.66% for training sets, and 98.57% for predictive sets, training sets and forecasts. A rice variety can be recognized properly by utilizing the proposed technique, and the system can be easily expanded into other agricultural commodities, categorization, allocation and gradation.

A technique for automated grading of tomatoes based on computer vision approaches was suggested by Arekeri and Lakshmana et al. [23] in their paper. To create a feed-forward neural system between defective and healthy tomato slices, they employ color, shape and texture information. When it comes to tomato categorization, high prediction accuracy has been recorded.

Based on their findings, concluded that binary logistic regression network classifiers outperformed single feature models when it came to distinguishing between high-quality seeds and low-quality seeds [24]. According to Tohidul Islam et al. [25], the feasibility of employing a small-scale picture database for attaining excellent classification accuracy with DCNNs has been studied. According to the results

of the UEC-FOODI00/256 datasets, classification accuracy ranged between 67.57% and 78.77% when AlexNet, a novel DCNN architecture, was used.

After a quick outline of deep learning, Lei Zhou et al. [26] gained a deeper understanding of the nature and strategies for the model formation of different major architectural deep neural networks. A large number of papers were reviewed, including those used as a data analysis tool for solving food problems and challenges, which include food recognition, vegetable quality detection, caloric estimating, fruit quality detection, meat and aquatic products, the chain of foodstuffs and the contamination of foodstuffs. Each study looked at the specific questions, data sets, pre-processing procedures, networks and frameworks, the results gained and contrasts with other common research solutions. They have also explored the idea of employing profound learning in food sensory and consumer research as an improved data mining approach which they found to be encouraging. During the survey, deep learning was shown to overcome other techniques, such as manual feature extractors and standard machine learning techniques. DL is also found to be a viable food quality and safety inspection tool.

Using the covariances of features for a trained DCNN as the representation of pictures, Tatsuma and Aono et al. [27] established a new technique for food recognition that is both efficient and accurate. On the Food-101 data set [28], they were able to achieve an accuracy rate of 58.65%. Deep learning models were also used to target more precise targets in this study. Ili Zu and colleague's et al. [29] researched to determine the presence of foreign objects in walnuts to assure food and safety and quality. To achieve a consistent imaging environment, artificial lightroom was used. The authors said deep learning models had the highest accuracy of classification, with a maximum accuracy of 99.5%. Two recent studies on the subject of seed identification have been undertaken. For example, DCNN models were employed in the study published by Ni et al. [30] to check faulty maize kernels, and DCNN [31] for hybrid okra seed with near-infrared hyperspectral images were utilized in the studies reported by Nie et al. [31]. Regarding the issue of binary classification, both studies revealed more than 90% classification accuracy.

The seed classification technique presented by Yonis Gulzar et al. uses CNN and transfer learning. The proposed system includes a model that uses sophisticated deep learning techniques to classify 14 well-known seeds. In this study, decaying learning rate, model checkpointing, and hybrid weight modification were all used. During data collection, this study employs symmetry in the sampling of seed photos. When photos are resized and labelled to extract their characteristics, symmetry makes them more homogeneous. Because of this, the training set's classification accuracy was 99 %. For the 234 images in the test set, the suggested model had a precision of 99 % [32].

A computer vision (CV) method to canola classification was the focus of this research done by S. Qadri et al. There were images of eight canola kinds to choose from for input. In the artificial neural network, binary features, first-order histogram features, spectral feature and second-order statistical texture features of three bands, blue (B), green (G), and red (R) were used. The classification was accomplished through the use of a ten-fold stratified cross-validation approach. The classifier performed best with data from regions of interest (512 512), with accuracy rates ranging from 95 % to 98 % [34].

Deep convolutional neural networks (CNNs) are used as general feature extractors in the study by Shima Javanmardi et al., which is a novel method. To classify the collected features, ANN, cubic SVM, quadratic SVM, boosted trees, bagged trees and linear discriminant analysis (LDA) were used. The classification accuracy of corn seed types was higher for models trained using CNN-extracted features than for models using simply basic features. Corn seed variants can be intelligently classified using a CNN-ANN classifier, according to this study [35].

J. Zhang et al. investigated the use of hyperspectral imaging and DCNN to analyze damaged-freeze corn seeds. The average spectra were taken from the area of embryo hyperspectral images spanning the wavelength range of 450–979 nm. In the next step, four models were developed for five-category classifications (5 frozen conditions) and four-category classifications (severe freezing, moderate freezing, slight freezing and no freezing), and the values of the evaluation indexes were evaluated for comparison (specificity, accuracy, precision and sensitivity). The visual categorization map was created using DCNN findings. It shows that DCNN and hyperspectral imaging can quickly identify freezing damage in corn seeds [36].

Q. Zhou et al. used NIR visible hyperspectral imaging, subregional voting, and a DL model. It was quite accurate in detecting maize seed varieties. Raw spectra were processed with Savitzky–Golay smoothing and an FD algorithm to emphasize spectral differences across types. An improved LeNet-5 network model based on subregional pixels was developed because the sample picture size was too small for the standard CNN model. The suggested model performed well in identifying the variety of normal maize seeds as well as sweet maize seeds [37].

**Table 1**  
A review on seeds identification technique

Ref.	Technique	Experiment Setup	Remark
[38]	CNN architecture (AlexNet, DenseNet, VGGNet, P-ResNet, GoogLeNet, MobileNet, ShuffleNet, and EfficientNet) for classification of maize seed.	Used 8080 maize seed image of five different variety in china.	P-ResNet gives the highest accuracy of 97.8 %.
[39]	DCNN (MobileNetv2) architecture to determine different variety of seed.	3:1 training to testing ratio.	Achieved accuracy of 98%.
[33]	Extract 17 different semantic feature (colour, seed-shape, binary feature etc.) from multispectral image to analyse seed using canonical discriminant analysis (CDA).	20 different sugar beet seed from 3 maturity class used.	Multispectral imaging technique is used for the assessment of sugar beet seed quality.
[34]	first order histogram, RST Invariant feature, textural/spectral/binary feature etc., were used to deploy ANN architecture to identify canonical seed	1600 images of 8 different canola seed variety were used.	Achieved accuracy in between 95 % to 98%.
[35]	ANN, SVM, KNN, LDA, boosted tree and bagged tree are used to classify extracted feature.	3:1 ratio for training and validation of neural network.	CNN-ANN approach is efficient for the identification of corn seed.
[36]	Identification of frozen corn seed by DCNN model using hyper spectral images.	1920 corn images were used for training, testing and validation in the ratio 4:1:1.	DCNN model gives better result than KNN, SVM and ELM model.
[37]	CNN and sub regional voting model were used to identify maize seed variety using hyper spectral image.	2430 samples used for training and testing.	For 6 different variety of normal maize seed accuracy is between 93% to 95 %.
[22]	PCANet, KNN and Random forest approaches were used to determine variety of rice seed.	4320 samples of rice variety were collected.	PCANet achieves the best result of 98%.
[32]	CNN (VGG16) is used to identify 14 different variety of seed.	234 images of different seed, randomly categorized into training and testing.	Achieved accuracy of 99%.

[20]	CNN used to detect hyper spectral pictures (AlexNet, ResNet18, Xception, InceptionV3, DenseNet201, and NASNetLarge).	Used 9600 hyperspectral image of 10 different soybean seed variety.	An average identification rate is 91%.
[30]	DCNN architecture (ResNet, VGG, AlexNet) were used to identify maize kernel.	To train neural network 1632 images of maize were used.	Achieved Accuracy of 98.2 %.
[31]	DCNN architecture is used to identify hybrid loofah and Okra seed using hyper spectral image.	6136 images of hybrid okra and 4128 images of loofah seed were used for training and testing.	Classification accuracy is 95%.
[21]	KNN, SVM and CNN used to identify rice variety using hyper spectral images.	1500 samples of rice is randomly categorized into training and validation.	CNN gives better result as compare to SVM and KNN.
[24]	Binary logistic regression and neural network is used to select high quality peeper seed.	50% data used for training, 25 % for testing and remaining 25% for final assessment of neural network.	Selection rate of peeper seed is 90%.
[25]	CNN is used to identify food image.	Food-11 dataset is used which consist of 16643 images.	Inceptionv3 gives highest accuracy of 92%.

### 3. Findings

Based on the work performed by which is surveyed in section 2, researchers have scope to perform the work in new direction.

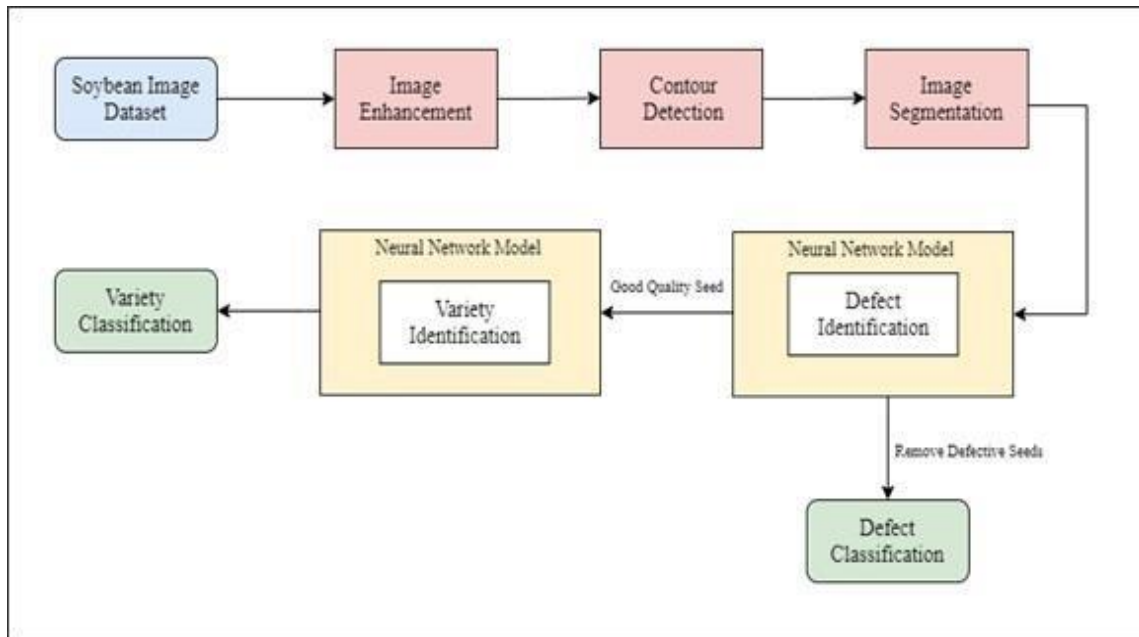
Earlier the images are taken from real-world entities, so it will contain some noise. The shadow noise affects the accuracy of segmentation for seed images. So, there will be a scope to improve the accuracy by eliminating noise.

Seed identification is hampered by shape variation in each seed class. Various researchers had tried to find out the variety or defects of seed. However, according to the literature survey, no researchers has worked on both concepts together. Some efforts are also done using hyperspectral techniques to detect the seed varieties but then the setup cost of hyperspectral is on the higher side. Also, according to the review of papers, none of the researchers had tried to identify varieties of seed on Indian species till date.

### 4. Proposed Methodologies

The soybean image data-set is taken as an input to image enhancement technique where the quality of image will be increase by removing the shadow noise & background noise from the image. After enhancement of the image, we carry out the contour detection, which identifies the object boundary. It is often the first step in image processing after an enhancement, which helps identify and recognize visual objects and at last segmentation will be perform using IP techniques. The output of segmentation

will be provided as input to the neural network to detect defective and good quality soybean seed. Good quality soybean seed are given as an input to our next neural network model which identify the variety of seed. Depending upon the network architecture that we utilize during development, the training and testing ratio may finalize.



**Figure 3:** Proposed system architecture

## 5. Conclusion

This paper, gives brief review of latest articles related to soybean seed defects and variety identification based on various methodologies. Majorly the papers had analyzed the use of DL and ML classification for soybean seed defects and variety. From performance point of view, DL approaches gives better result as compared to the existing method reviewed in this paper. Most of the authors had performed the classification of soybean seed varieties and achieved good results but then soybean seed defects identification is still an untouched area and use of deep learning methods with the help of enhanced use of computer vision technology for better and more accurate soybean seed variety and defect identification will be the area of research in future direction.

## 6. References

- [1] V. Kumar, A comparative assessment of total phenolic content, ferric reducing-anti-oxidative power, free radical-scavenging activity, vitamin C and isoflavones content in soybean with varying seed coat color, in: *Food Research International* 43(1), 323-328, 2010.
- [2] M. Carmona, Development and validation of a fungicide scoring system for management of late-season soybean diseases in Argentina, in: *Crop Protection* 70, 83-91, 2015.
- [3] T. Brosnan, Improving quality inspection of food products by computer vision-a review, in: *Journal of Food Engineering* 61(1),3-16, 2004.
- [4] H. Sabrol, Recognition of Tomato Late Blight by using DWT and Component Analysis, in: *International journal of electrical and computer engineering* 7, 194-199, 2017.
- [5] R. R. Parmar, Unified approach in food quality evaluation using machine vision, in: *Advances in computing and communications*, 239-248, 2011.
- [6] S. Lilik, Digital Image-Based Identification of Rice Variety Using Image Processing and Neural Network, in: *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)* 16, 182-190, 2015.

- [7] C. J. Du, Learning techniques used in computer vision for food quality evaluation: A review, in: *Journal of food engineering* 72, 39-55, 2006.
- [8] A. Tannouche, A fast and efficient shape descriptor for an advanced weed type classification approach, in: *International Journal of Electrical and Computer Engineering* 6, 1168-1175, 2016.
- [9] H. K. Mebatsion, Automatic classification of non-touching cereal grains in digital images using limited morphological and color features, in: *Computers and Electronics in Agriculture* 90, 99-105, 2013.
- [10] M. Olgun, Wheat grain classification by using dense SIFT features with SVM classifier, in: *Computers and Electronics in Agriculture* 122, 185-190, 2016.
- [11] R. D. L. Pires, Local descriptors for soybean disease recognition, in: *Computers and Electronics in Agriculture* 125, 48-55, 2016.
- [12] K. Z. Tan, Identification of diseases for soybean seeds by computer vision applying BP neural network, in: *International Journal of Agricultural and Biological Engineering* 7, 43-50, 2014.
- [13] Jordan, M.I, Machine learning: Trends, Perspectives, and Prospects, in: *Science* 349, 255–260, 2015.
- [14] Barbedo, J.G.A, Plant Disease Identification from Individual Lesions and Spots Using Deep Learning, in: *Biosys. Eng.* 180, 96–107, 2019.
- [15] Barbedo, J.G.A, Impact of Dataset Size and Variety on the Electiveness of Deep Learning and Transfer Learning for Plant Disease Classification, in: *Compute. Electron. Agric.* 153, 46–53, 2018.
- [16] DeChant, C, Automated Identification of Northern Leaf Blight-Infected Maize Plants from Field Imagery Using Deep Learning, in: *Phytopathology* 107, 1426–1432, 2017.
- [17] Shen, R.P, Construction of a Drought Monitoring Model Using Deep Learning Based on Multi-Source Remote Sensing Data, in: *Int. J. Appl. Earth Obs.* 79, 48–57, 2019.
- [18] Jin, B.X, Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use-Type Classification of Remote Sensing Images, in: *J. Indian Soc. Remote Sens* 47, 951–965, 2019.
- [19] Rasti P, Supervised Image Classification by Scattering Transform with Application to Weed Detection in Culture Crops of High Density, in: *Remote. Sens.* 11, 249, 2019.
- [20] Zhu, S, A Rapid and Highly Efficient Method for the Identification of Soybean Seed Varieties: Hyperspectral Images Combined with Transfer Learning, in: *Molecules* 25, 152(2020) Zhu, S: A Rapid and Highly Efficient Method for the Identification of Soybean Seed Varieties: Hyperspectral Images Combined with Transfer Learning. *Molecules* 25, 152, 2020.
- [21] QiuZ, Variety Identification of Single Rice Seed Using Hyperspectral Imaging Combined with Convolutional Neural Network, in: *Appl. Sci.* 8, 212, 2018.
- [22] Weng, S, Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion, in: *Spectrochim Acta Mol Biomol Spectrosc* 234, 2020.
- [23] Arakeri M, Computer Vision-Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry, in: *Procedia Computer Science* 79, 2016.
- [24] K.L. Tu, Selection for high-quality pepper seeds by machine vision and classifiers, in: *J. Integr. Agric.* 17(9), 1999–2006, 2018.
- [25] Md Tohidul Islam, Food Image Classification with Convolutional Neural Network, in: *Intelligent Informatics and Biomedical Sciences (ICIIBMS) International Conference* 3, 257-262, 2018.
- [26] Zhou, Application of Deep Learning in Food: A Review, in: *Comprehensive Reviews in Food Science and Food Safety*, 2019.
- [27] TATSUMA, Food Image Recognition Using Covariance of Convolutional Layer Feature Maps, in: *IEICE Transactions on Information and Systems*, 2016.
- [28] Food-101 – Mining Discriminative Components with Random Forests, in: *Lecture Notes in Computer Science*, 1611-3349, 2018.
- [29] Ili Zhu, Deep learning and machine vision for food processing: A survey, in: *Current Research in Food Science* 4, 233-249, 2021.
- [30] C. Ni & D. Wang, Automatic inspection machine for maize kernels based on deep convolutional neural networks, in: *Biosyst. Eng.* 178, 131–144, 2019.
- [31] Nie, Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning, in: *Sensors and Actuators B: Chemical* 296, 2019.



- [32] Gulzar, Yonis, A Convolution Neural Network-Based Seed Classification System, in: *Symmetry*, 2020.
- [33] Jiang, A deep learning approach for fast detection and classification of concrete damage, in: *Automation in Construction* 128, 2021.
- [34] Qadri S, Classification of canola seed varieties based on multi-feature analysis using computer vision approach, in: *International Journal of Food Properties*, 2021.
- [35] Shima Javanmardi, Computer-vision classification of corn seed varieties using deep convolutional neural network, in: *Journal of Stored Products Research* 92, 2021.
- [36] Zhang, Identification of Corn Seeds with Different Freezing Damage Degree Based on Hyperspectral Reflectance Imaging and Deep Learning Method, in: *Food Analytical Methods*, 2021.
- [37] Zhou Q, Identification of the variety of maize seeds based on hyperspectral images coupled with convolutional neural networks and subregional voting, in: *J Sci Food Agric* 101(11), 2021.
- [38] Peng Xu, Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning, in: *MDPI agriculture*, 2022.
- [39] Yasir Hamid, Smart Seed Classification System based on MobileNetV2 Architecture, in: *International Conference on Computing and Information Technology (ICIT)* 217-222, 2022.