# Multiple Seed Segregation using Image Processing

Deepali Koppad
Department ofECE,*Ramaiah Inst. of Technology,* Bengaluru-560054deepali.koppad@gmail.com

Saquib Akhtar
Department of ECE
*Ramaiah Inst. of Technology*,
Bengaluru-560054
saquibakhtar0509@gmail.com

Mohammed Mehtab Alam
Department of ECE
*Ramaiah Inst. of Technology*,
Bengaluru-560054
methabalam7860@gmail.com

Soumya Kanti De
Department of ECE
*Ramaiah Inst. of Technology*,
Bengaluru-560054the.skd999@gmail.com

Suma K V,
Department of ECE,
*Ramaiah Institute of Technology*
Bengaluru – 560054
sumakv@msrit.edu

Nethra N
National Seed Project, University of Agricultural Sciences, Bengaluru-560065
nethraharsha@gmail.com

***Abstract*—Agricultural production is based on seeds and large populations are dependent on agriculture for sustenance. Seed properties are generally assessed by human inspection, which has many limitations. Previously, seed segregation hardware setup was developed, but the limitation of the setup is it can visualize a single seed at a time and no internal features were considered. This work aims at separating multiple different types of seeds as good and bad using object detection analysis. Using a Webcam, multiple images of different seeds are clicked. These seeds are kept in a tray. The images were fed into pre-trained Object detection YOLO5 model. This model classifies the seeds into two categories, good and bad and also classifies the types of seed using Image processing. It is an affordable solution. With this setup, multiple good and bad seeds are detected accurately which is also time efficient. This hardware setup can be used to segregate any number of seeds with a classification accuracy of 97%.To the best of the authors' knowledge, this is a first attempt to develop a dataset and also perform classification of different types of seeds including both good and bad types.**

**Keywords—YOLOv5, Image Processing, Seed Quality, Multiple Seeds**

## I. Introduction

Seed segregation is an important task in the agricultural industry that involves the separation of seeds based on various characteristics such as size, shape, and color. The traditional method of seed segregation involves manual sorting of seeds by farmers or laborers based on their physical characteristics such as size, shape, and color [1]. The process is usually done by visually inspecting the seeds and separating them into different categories by hand. The traditional method of seed segregation is time-consuming and labour-intensive, which can result in high labour costs and reduced efficiency. The traditional method of seed segregation has several disadvantages. First, it is prone to human error, and the quality of the seed sorting can vary depending on the experience and skill of the person doing the sorting. Second, it is time-consuming, and the process can take several hours or even days, depending on the quantity of seeds that need to be sorted. Third, the process is labour-intensive and can result in high labour costs. Despite these disadvantages, the traditional method of seed segregation is still widely used in many regions, particularly in developing countries, where automation is not readily available, and labour costs are relatively low. Seed segregation is an essential process in the agricultural industry, particularly for plant breeding, seed production, and crop improvement. It involves separating seeds based on their various characteristics such as size, shape, and color, to obtain seeds of uniform quality and characteristics. Here are some reasons why seed segregation is important:

Seed segregation ensures that the seeds are of uniform quality and characteristics, which is important for plant breeding, seed production, and crop improvement. It allows farmers and breeders to select the best quality seeds for propagation, resulting in higher crop yields and better-quality produce. Seed segregation ensures that the seeds have consistent properties, which is essential for producing uniform crops. Consistent crops are easier to harvest, process, and market, resulting in better returns for farmers. It can help prevent the spread of plant diseases.

Seeds that are infected with diseases can be segregated and removed from the production process, reducing the risk of the disease spreading to other plants. It can reduce the cost of seed production and improve the profitability of farmers. By segregating seeds, farmers can reduce the amount of seed required for planting and increase the germination rate, resulting in a more efficient use of resources. Seed segregation is often required by regulatory bodies to ensure the quality and safety of the seed. Compliance with seed segregation regulations can help farmers to avoid fines and penalties and maintain their reputation in the industry. However, the use of image processing techniques for seed segregation has the potential to revolutionize the process and make it more efficient and accurate

This work aims to develop a seed segregation system using image processing techniques. Image processing is the field of computer science that involves analyzing and manipulating digital images using various algorithms and techniques. It involves converting an image into a form that is more suitable for further analysis or processing, such as feature extraction, pattern recognition, or machine learning. The system will use digital images of seeds captured using a camera and apply various image processing algorithms to segment and classify the seeds based on their properties. The system will then automatically segregate the seeds into different categories, making the process more efficient and accurate. The use of image processing techniques for seed segregation has many advantages, including increased efficiency, accuracy, and consistency. It can also help farmers to achieve a higher quality of seed sorting, which can result in better crop yields and profits. Overall, this work has the potential to revolutionize the way seeds are sorted in the agricultural industry, making the process more automated, efficient, and accurate..

## II. LITERATURE SURVEY

Deep learning and machine learning techniques have the potential to detect, classify, and recognize objects, thus enabling researchers to automate the separation of objects. In recent years, a variety of research has been conducted to assess the quality of seeds, as well as to determine how they should be classified according to their grade and quality.

Using seeds of pearl millet and maize classification was done using YOLOv5 for healthy and diseased seeds [2]. The model reported precision and recall accuracy of 99%. Convolution Neural Network have been used to classify soya seeds into good and bad based on its external features with the accuracy of 93% using hardware setup [3].Color and texture features of soyabean leaf were taken into consideration using Image processing in the pre-processing and Artificial Neural network with the aid of MATLAB. An Accuracy of about 95.7% is obtained in this classification. [4].

The paper [5] discusses the use of YOLOv5 model for classification of garlic. The model is trained on images of healthy and unhealthy garlic dataset. The precision of the YOLOv5 model achieved was 99% though the dataset had only 570 images in total. One-class classifier was used by [6] to classify tomato seed. For this, tomato seed images were captured using a camera and LED as a light source.

The paper [5] discusses the use of YOLOv5 model for classification of garlic. The model is trained on images of healthy and unhealthy garlic dataset. The precision of the YOLOv5 model achieved was 99% though the dataset had only 570 images in total. One-class classifier was used by [6] to classify tomato seed. For this, tomato seed images were captured using a camera and LED as a light source.

The authors in [7], use YOLOv5 model for detection and classification of maritime objects. The original Singapore Maritime Dataset was modified for better annotation and then used in the YOLOv5 classification. For the original YOLOv5 model, a transformed encoder block was added [8] to improve the sensitivity of the model. This modified model TIA-YOLOv5 was used for real-time detection of crop and weeds in the field.

The paper [9] discusses the use of YOLOv5 model for detection of the black variety bean kernel. The paper reported an accuracy of 99.8% for detecting the black variety of the beans. The authors in [10], use YOLOv5 for detecting tomato plants phenotyping traits with an aim to identify nodes, fruits and flowers on the plant.

Comparing and classifying the germinated and non-germinated seed samples was done by [11]. They achieved an accuracy of 75% using neural networks. Using hyperspectral images, variety of rice seeds were studied in [12]. Three different models namely K-nearest neighbors, support vector machine (SVM) and CNN models were used for classification. For the different models with different training dataset, the accuracy varied between 70-80%.

The papers mentioned in the literature either use an existing dataset or have developed a new dataset, but none of those are for the seeds that the current work is exploring. Hence in this work, a dataset has been generated for different types of seeds. For this, a hardware setup has also been developed.

## III. PROPOSED METHODOLOGY

In this work, five varieties of seeds such as Soyabean White Matar, Green Matar, Black Chana and Kabuli Chana are considered. The images of the seeds are taken and then classified and graded as good and bad depending upon its external features like shape, size, edges etc.

### *A. Generalised Block Diagram*

The general block diagram is shown in Fig: 1. The images of seeds are captured in the data collection step. After pre-processing thedataset is prepared. This dataset is split into three categories for training, validation and testing the AI model. The YOLOv5 model is used in this work. This model is trained and validated using the generated dataset. The YOLOv5 model is optimized to classify the different types of seeds into good or bad based on their physical characteristics and the features extracted using Image pre-processing.

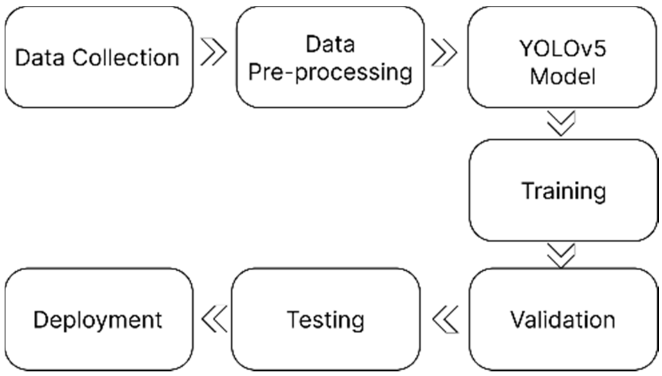


Fig 1. Block Diagram

### *B. Dataset*

For the five types of seeds, both good and bad seed images were captured using the developed hardware. For each category (good and bad) of each seed type number of images were taken as shown in Table 1. A mixture of all these types of seed images were also captured since in this work multiple seed segregation is being carried out.

For each seed variety, the captured images are split into three types, namely: training, validating and testing datasets.

TABLE 1. Dataset

| Seeds Name | Training | Validation | Testing |
|---|---|---|---|
| **Good Soyabean** | *1233* | *82* | *55* |
| **Good White Matar** | *1233* | *82* | *55* |
| **Good Green Matar** | *1233* | *82* | *55* |
| **Good Black Chana** | *1233* | *82* | *55* |
| **Good Kabuli Chana** | *1233* | *82* | *55* |
| **Bad Soyabean** | *1233* | *82* | *55* |
| **Bad White Matar** | *1233* | *82* | *55* |
| **Bad Green Matar** | *1233* | *82* | *55* |
| **Bad Black Chana** | *1233* | *82* | *55* |
| **Bad Kabuli Chana** | *1233* | *82* | *55* |

The seeds used in this work include Good Soya seed, Good Black Chana, Good White Matar, Good Green Matar, Good Kabuli Chana, Bad Soya seed, Bad Black Chana, Bad WhiteMatar, Bad Green Matar, Bad Kabuli Chana. A sample of the seed images is shown in Fig. 2.

Fig 2. Dataset

A total of 13700 seed images were taken. Out of these, 12330 were used for training, 822 images for validation and remaining 548 images were used as testing set.All the three categories (training, validation and testing) have no overlap of images.

TABLE 2. Total Dataset

| **Total Seeds** | **Training** | **Validation** | **Testing** |
|---|---|---|---|
| *13700* | *12330* | *822* | *548* |

After the images were captured, the data pre-processing was done for all the images. This involved, resizing so that all pictures are of the same scale which is 640*640. Auto orientation was done in order to discard the EXIF rotation and standardize pixel orientation. The contrast, hue, shadow adjustment was not done in order to avoid complexity and keep the seeds format to its original RGB format

The next step after pre-processing is data augmentation. In machine learning, data augmentation involves transforming or modifying an existing dataset to generate additional training data to improve model performance. In order to accomplish this, a number of operations or transformations are applied to the existing data to produce new samples that are similar, but not identical, to the original samples. By forming new and different examples to train datasets, data augmentation is useful for improving machine learning models' performance and outcomes. When the dataset is rich and sufficient, machine learning models perform better and are more accurate. The process of collecting and labelling data for machine learning models can be tedious and expensive. Data augmentation techniques can be used to transform datasets to reduce operational costs.For this step the following types of augmentation were performed: Flipping, 90° rotate clock and anti-clock wise, shear of plus /minus 10°, and rotation to plus/minus 10°.

### C. *YOLOv5 model*

The YOLOv5 algorithm was developed for the detection of objects. It is an upgrade to the previous versions of YOLO[You Only Look Once], which is consider to be a real-time object detection system that can detect objects in videos and images. YOLOv5 has some improvements over the previous versions, including a new architecture, faster training times, a smaller model size, and improved accuracy. It uses a single-stage detection pipeline that predicts bounding boxes and class probabilities directly from the image.

Computer vision tasks such as object detection involve locating and identifying objects in images. A variety of applications rely on it, such as robots, automatic vehicles, surveillance, or robotVacuum. There are two main categories of object detection algorithms: single-shot detectors and two-stage detectors. The R-CNN (Regions with CNN features) model was among the earliest successful attempts to solve the object detection problem using deep learning. To detect and locate objects in images, this model combined region proposal algorithms with convolutional neural networks (CNNs). Based on how many times an input image is passed through a network, object detection algorithms can be broadly classified into two categories. Single-shot object detection predicts the presence and location of objects in an image using only one pass of the input image. A single pass is used to process an entire image, which makes them computationally efficient. A single-shot object detection method, however, is generally less accurate and less effective at detecting small objects than other methods. Real-time detection of objects in resource-constrained environments can be achieved using such algorithms. To process an image, YOLO uses a fully convolutional neural network (CNN). An object can be detected in two passes of an input image using two-shot object detection.

### D. *YOLOv5 Architecture*

The 1st twenty convolution layers of YOLOv5 model are pre-trained by plugging in an average pooling and fully connected layer, using ImageNet as shown in Fig. 3. Convolutional and connected layers are then added to the pre-trained network to improve performance, as demonstrated by [13].

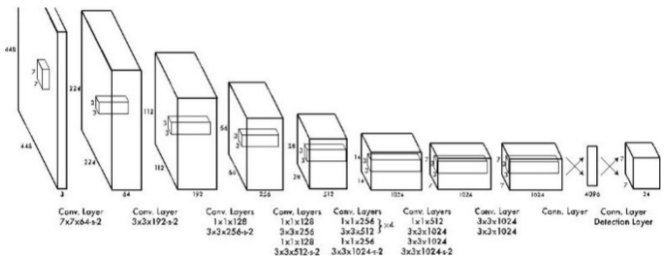


Fig 3. YOLOv5 Architecture

As the part of YOLO final fully connected layer, bounding box coordinates and class probabilities are predicted. An input image is divided into a grid size of S×S by YOLO algorithm. The grid cell that can detect an object if the center of the object falls into it, model detects that as object. There are confidence scores for each cell in the grid that predicts a bounding box as well as B bounding boxes. Using the confidence scores, we can assess the reliability of the prediction and the confidence that the box contains the object. Each grid cell is predicted to have multiple bounding boxes by YOLO.

### *E. Hardware Setup*

In this work, a hardware system is specifically designed for seed analysis and identification. The block diagram of the same has been shown in Fig. 4.

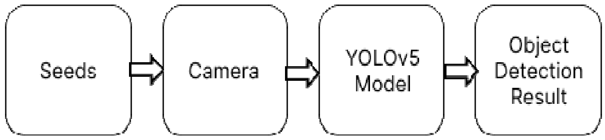


Fig 4. Hardware Block Diagram

The Dual-Camera Seed Imaging Device. This innovative device is equipped with two high-resolution cameras strategically positioned on the top and bottom to capture comprehensive images of seeds placed between them on a glass. Seamlessly integrated with the YOLOv5 model, this hardware solution offers high accuracy and efficiency in seed analysis. Fig. 5shows the prototype of this proposed model.

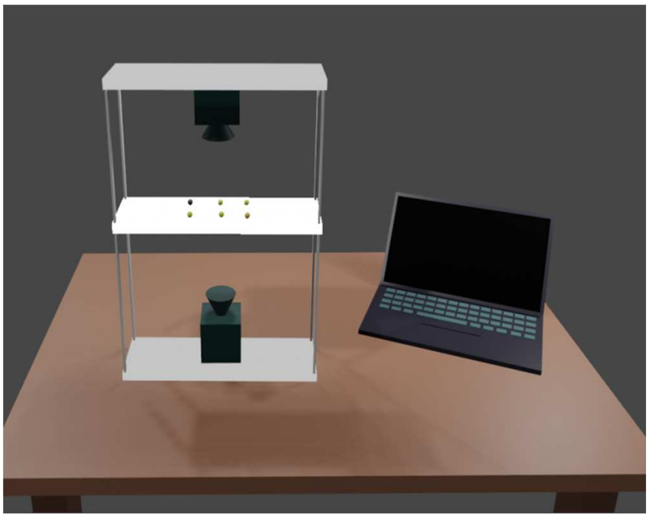

Fig 5. Hardware Prototype

The top and bottom cameras are meticulously calibrated to ensure optimal imaging conditions, enabling precise capture of the seeds from different angles. The top camera provides a downward view of the seed, while the bottom camera captures an upward view. This dual-perspective approach eliminates blind spots and offers a comprehensive visual representation of the seed, capturing critical details such as shape, color, texture, and surface characteristics. The Dual-Camera Seed Imaging Device utilizes a highly efficient data transfer system to seamlessly feed the captured images into the YOLOv5 model. YOLOv5 is a state-of-the-art object detection model known for its outstanding performance in real-time image analysis. The actual hardware implementation is shown in Fig. 6.

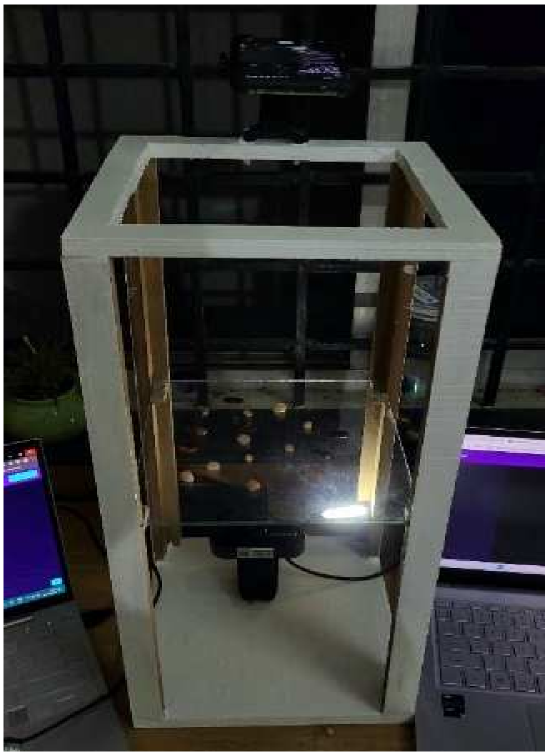

Fig 6. Hardware Setup

## IV. Results and Discussion

Using the designed hardware, images of different types of seeds images are captured as mentioned in Table 1. After capturing the images, data augmentation is performed. This augmented data is split into three categories that can be fed to the YOLOv5 model for training. The model is trained on HP laptop with configuration: Processor-Intel Core i5-10210U CPU @ 1.60GHz, 2112 MHz, 4 Cores, GPU-NVIDIA GeForce MX350, 2MP-720P Webcam.

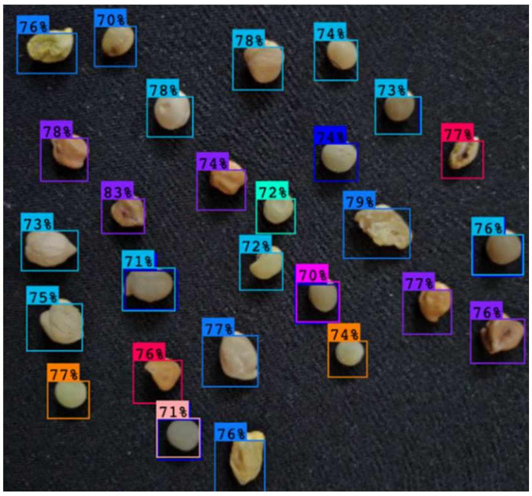

Fig 7. Output Result of First batch

After training and validating the YOLOv5 model, the model is tested using images of seeds that were not used in the training of the model. The YOLOv5 model is able to detect different types of seeds as good or bad with accuracy of about 98%.

In order to test the model with respect to the confidence score, the following two batches of images was used. In first batch as shown in Fig.7 seven types of seeds were used: Good and bad seeds of white matar, green matar and kabulichanna. The highest confidence score was obtained for each type of seed:Good White Matar: 74%, Good Green Matar: 77%, Good Kabuli Chana: 78%, Bad Black Chana: 77%, Bad White Matar: 70%, Bad Green Matar: 72%, Bad Kabuli Chana: 74%.

In Second batch as shown in Fig. 8, eight types of seeds were used that includes their maximum confidence score:Good White Matar: 75%, Good Green Matar: 69%, Good Kabuli Chana: 76%, Bad Black Chana: 69%, Bad White Matar: 71%, Bad Green Matar: 71%, Bad Kabuli Chana: 75%, Bad Soya Seed: 72%.

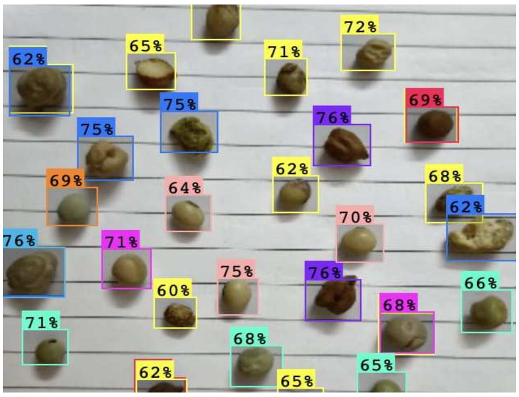

Fig 8. Output Result of Second batch

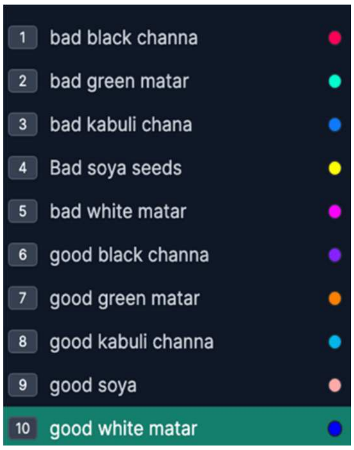


Fig 9. Color coding of boxes

In Fig.9 color coding is shown, red box is bad black chana, cyan is bad green matar, similarly othersseeds colors are also coded.

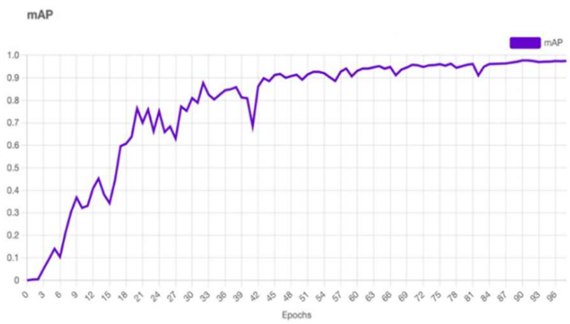


Fig 10. mAP Graph

A. Mean average precision (mAP): It is used in the YOLOv5 Model. The mAP compares the ground-truth bounding box to the detected box and returns a score. The higher the score, the more accurate the model is in its detections. mAP is a way to measure how well the YOLOv5 model can detect and locate objects in an image, and how confident it is in its predictions. The mAP of the current model is as shown in Fig.

B. Box loss: This is a way to measure how much the predicted location of an object's bounding box differs from the actual location. The result of the same is shown in Fig.11.

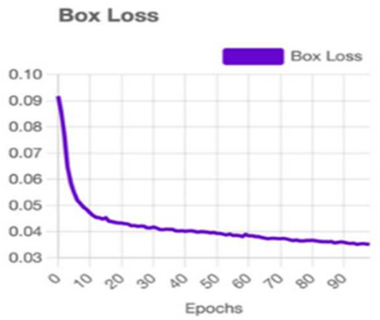


Fig 11. Box Loss Graph

C. Class loss: this is a way to measure how accurate the predicted object class is compared to the actual class as shown in Fig.12.

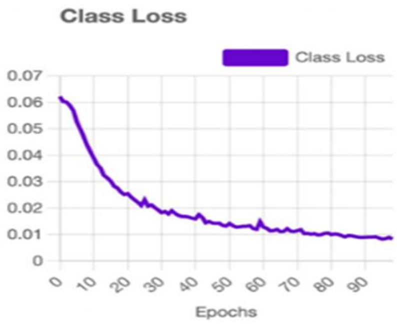


Fig 12. Class Loss Graph

D. Object loss: is a way to measure how confident the model is in predicting that an object exists in a certain location. The result of which is shown in Fig.13.


Fig 13. Object Loss Graph

In the current work, five variety of seeds are considered and two types of classification is performed. The first classification is identifying the type of seed and second classification involves identifying if the seed is of good or bad quality with a good accuracy.

To the best of the author's knowledge this is the first attempt for such experimentations. Hence, the only comparison with the papers in literature is to compare the performance of the YOLOv5 model.
The results of YOLOv5 model used for testing barley seeds was 98.1% accuracy as shown in [14]. Rice granules were classified with an accuracy of 98% in [15]. The results in [16] discuss the use of YOLOv5 model in the study and analysis of wheat grain detection. The paper has achieved a precision of around 96%.

The results of the current work, where five variety of seeds are taken into account, are in sync with those obtained in the literature survey as far as the YOLOv5 model is concerned.

## CONCLUSION AND FUTURE WORK

This paper discusses the application of AI in the field of agriculture. A hardware prototype has been developed which is used to capture images of multiple seeds. Two types of images for each seed are captured and then pre-processed for further experimentation. These captured seed images were fed to an AI-based model YOLOv5 to classify into good and bad seeds.

A total of 13700 seed images were taken. Out of these, 12,330 were used for training, 822 images for validation and remaining 548 images were used as testing set. The model was able to classify the seeds with a 97% testing accuracy.

For the current work, only external features are considered for classification. In future, the internal structures / defects can also be taken in account for better segregation of seeds. Either a light source can be used while capture images or X-ray images of the seeds can be capture in order to view the internal defects.

## REFERENCES


[1]. Se Micaela Colley, Alex Stone, Linda Brewer, "Organic Seed Processing: Threshing, Cleaning and Storage" Organic Seed Resource Guide, January 2009

[2]. Dulal R, Zheng L, Kabir M, McGrath S, Medway J, Swain D and Swain W. (2022). Automatic Cattle Identification using YOLOv5 and Mosaic Augmentation: A Comparative Analysis 2022 International Conference on Digital Image Computing: Techniques and ApplicationsSeed Quality Analyzer and Segregator, Dr. Suma K V, Department of ECE, Centre for Imaging Technologies Ramaiah Inst. of Technology (of VTU Aff.), Bangalore, India. IEEE Region 10 Conference (TENCON 2019)

[3]. Ranjith Bose M, Ranjith K, Suraj Prakash, Subham Kumar Singh, Dr Vishwanath Y, Intelligent Approach for Classification of Grain Crop Seeds Using Machine Learning, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 05, May-2018, e-ISSN: 2395-0056

[4]. Harika, Sai & Reddy, Vishnuvardhan & Reddy, Divyamsh & Soppari, Kavitha. (2022). Seed Quality Testing Using Deep Learning And YOLO V5. 10.24327/ijrsr.2022.1305.0250.

[5]. Yasmin, J.; Lohumi, S.; Ahmed, M.R.; Kandpal, L.M.; Faqeerzada, M.A.; Kim, M.S.; Cho, B.-K. Improvement in Purity of Healthy Tomato Seeds Using an Image-Based One-Class Classification Method. *Sensors* 2020, *20*, 2690.

[6]. Kim, J.-H.; Kim, N.; Park, Y.W.; Won, C.S. Object Detection and Classification Based on YOLO-V5 with Improved Maritime Dataset. *J. Mar. Sci. Eng.* 2022, *10*, 377

[7]. Wang A, Peng T, Cao H, Xu Y, Wei X, Cui B. TIA-YOLOv5: An improved YOLOv5 network for real-time detection of crop and weed in the field. Front Plant Sci. 2022 Dec 22;13:1091655.

[8]. Rodríguez-Álvarez, José Luis and García Alcaraz, Jorge L. and Puig i Vidal, Rita and Soto-Cabral, Alondra, Contrasting Bean Analysis System Based on Yolov5 and a Neural Network Model Using the Interval Type-2 Fuzzy Set Approach.

[9]. Angelo Cardellicchio, Firozeh Solimani, Giovanni Dimauro, Angelo Petrozza, Stephan Summerer, Francesco Cellini, Vito Renò, "Detection of tomato plant phenotyping traits using YOLOv5-based single stage detectors", Computers and Electronics in Agriculture, Volume 207, 2023, 107757, ISSN 0168-1699,

[10]. D.Sivakumar, K.SuriyaKrishnaan, P.Akshaya, G.V.Anuja, G.T.Devadharshini, Computerized Growth Analysis of Seeds Using Deep Learning Method, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S5, April 2019.

[11]. Qiu, Z, Chen, J, Zhao, Y, Zhu, S, He, Y, Zhang C, "Variety Identification of Single Rice Seed Using hyperspectral Imaging Combined with Convolutional Neural Network",. Appl. Sci.2018,8, 212.

[12]. Ziwen Chen, Lijie Cao, and Qihua Wang, "YOLOv5-Based Vehicle Detection Method for High-Resolution UAV Images", College of Mechanical and Power Engineering, Dalian Ocean University, School of Information Engineering, Dalian Ocean University, School of Medical Information Engineering, Jining Medical University, China, Published 2 May 2022.

[13]. Shi, Y.; Li, J.; Yu, Z.; Li, Y.; Hu, Y.; Wu, L. Multi-Barley Seed Detection Using iPhone Images and YOLOv5 Model. *Foods* 2022, *11*, 3531.

[14]. Miss Shivpriya Desai, Dr. A. P. Rao, Seed Quality Analysis using Image Processing and ANN, International Journal of Trend in Scientific Research and Development, Volume 1(4), ISSN: 2456-6470.

[15]. Wenyi Zhao, Shiyuan Liu, Xinyi Li, Xi Han, Huihua Yang, "Fast and accurate wheat grain quality detection based on improved YOLOv5", Computers and Electronics in Agriculture, Volume 202, 2022, 107426, ISSN 0168-1699,