ISBN: 978-981-18-7950-0
2023 the 13th International Workshop on Computer Science and Engineering (WCSE 2023)
doi: 10.18178/wcse.2023.06.014

Classification of Lactuca Sativa Seed Varieties Using Spectral-Textural-Spatial Features and Decision Tree Classifier

Mark Mario Simons, Seth T. Dayao and Meo Vincent C. Caya +

School of Electrical, Electronics and Computer Engineering, Mapúa University 658 Muralla St, Intramuros, Manila, Philippines

Abstract. It is important to identify the variety of a crop before planting the seed. Identifying lettuce seed varieties manually can be a tough process. This study focuses on creating a system that can classify lettuce seed images using machine learning. Three Lettuce varieties were considered in this study, including the Romaine, Black Seeded Simpson, and Lollo Rosso. 100 seed images were captured for each lettuce seed variety in a controlled environment. The seed images were segmented using color thresholding in the HSV color space. Color, texture, and morphological features were used for variety classification including RGB color features, Solidity, Correlation, Compactness, Perimeter, Entropy. Decision Tree Classification algorithm was used to classify lettuce seeds. The system was able to successfully classify lettuce seed varieties with an average accuracy of 92%.

Keywords: machine learning, lettuce seed, variety classification, decision tree

1. Introduction

One of the popular crops in the Philippines is lettuce despite the temperatures in the country that make it impractical to plant lettuce there. Manually classifying the seeds this way can be quite laborious [1] and costly, but it is important that the agriculturists know about the seeds they are using as it helps predict the quality of the crops, and it is also important to be able to tell the variety from the seed stage for separation, purification, breeding as well as postharvest treatment and packaging. Classification of seeds can be done using computer technology through techniques such as machine learning and image processing.

Different studies have used machine learning and image processing to classify seeds for different applications including classification of healthy seeds from bad ones. Studies by [1-4] used colour thresholding to segment images to extract certain features from seed images including contrast, correlation and entropy, and studies by [3] [5-6] used morphological features such as area and perimeter. Using the data from the extracted features, they use algorithms such as K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Support Vector Machine (SVM).

One of the main issues that Lettuce producers and farmers encounter is determining the lettuce seed varieties. Previous studies have studied only a few of the lettuce seed varieties including the Chinese-loose-leaf, Grand-rapid, and Iceberg. Automating the process of identifying lettuce seed varieties will greatly decrease the amount of labour and costs that need to be put to classify the seed varieties.

The main goal of this study is to classify 3 lettuce seed varieties, namely the Romaine, Black Seeded Simpson, and Lollo Rosso. The researchers aim to do this with the use of the Decision Trees Classifier algorithm with data based on colour, texture, and morphological features of the seeds. The researchers specifically aim to (1) create a hardware setup that will be used to acquire images of the lettuce seeds. (2) To create a GUI application that predicts the variety of Lactuca Sativa seeds using DTC algorithm and (3) to evaluate the performance of the system using a confusion matrix.

With the proposed system of automating the process of classifying the lettuce seed varieties, productivity can increase in workplaces that rely on humans to manually inspect the seed samples. This gives the opportunity to let other employees work on more critical tasks.

+ Corresponding author. Tel.: + 63 2-82962652; fax: + 63 2-82962652.
E-mail address: msimons@mymail.mapua.edu.ph.

This study will train and test the classification algorithm to classify only the Romaine, Black Seeded Simpson, and Lollo Rosso Lactuca Sativa seed varieties using single kernel images. The researchers will only utilize the DTC algorithm using Python 3 to classify the Lactuca Sativa seed images, using only the morphological, colour and texture features extracted from the seed image samples. The Lactuca Sativa seed images will be captured in a controlled environment, and a sample size of 100 seed images will be taken for each seed variety.

2. Methodology

2.1. Conceptual framework

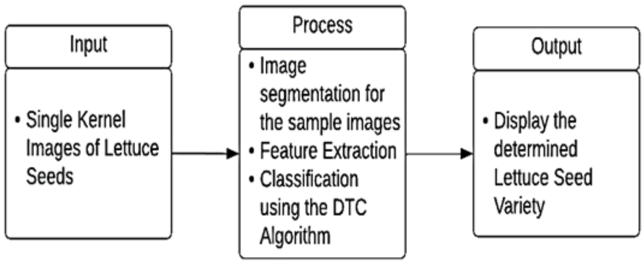

Fig. 1: Conceptual framework.

The actual input for the system is a single kernel image of the lettuce seeds which will be taken in a controlled environment. The Images are then segmented in the program for feature extraction. The features spectral, textural, spatial features and morphological features are then extracted for classification. The extracted data will be used as input to train the Decision Trees Classifier algorithm, and the train model will be used to predicted lettuce seed variety.

2.2. Hardware setup

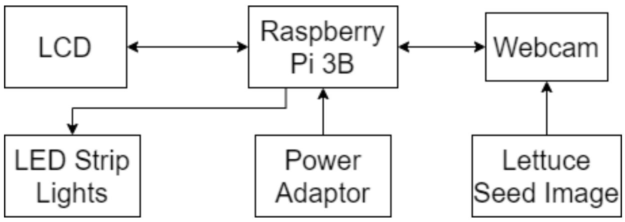

Fig. 2: System Block Diagram

The researchers have constructed a wooden box which will hold all the components as well as serve as the controlled environment where the interior of the box will be lit using LED Strip lights and the bottom will have white paper for a clearer view of the seeds as well as to differentiate the seed and the background easier. The box also holds the webcam that will be used to capture the seed images as well as the Raspberry Pi 3B which holds the program and data for classification of Lettuce Seed Images.

2.3. Image acquisition

The chosen seed variety for this study are Romaine, Black Seeded Simpson, and Lollo Rosso as they have not yet been subjects of image processing and machine learning for classification as of writing this paper. A hundred single kernel images were taken for each variety, making a total of 300 samples used to train and test the algorithm. Each sample image was taken in the controlled environment setup using a webcam placed 2 inches above the seed sample.

2.4. Software procedure

The program starts by capturing the lettuce seed image. It will then use the captured seed image as an input for the image processing step where the images are to be converted in the HSV colour space before segmenting it for feature extraction. The morphological, textural, and colour features are extracted from the processed seed image. The extracted data will then serve as the input for the Decision Trees Classifier to determine the Lactuca Sativa variety. The output of the program is the captured image as well as the determined variety of the Lactuca Sativa seed.

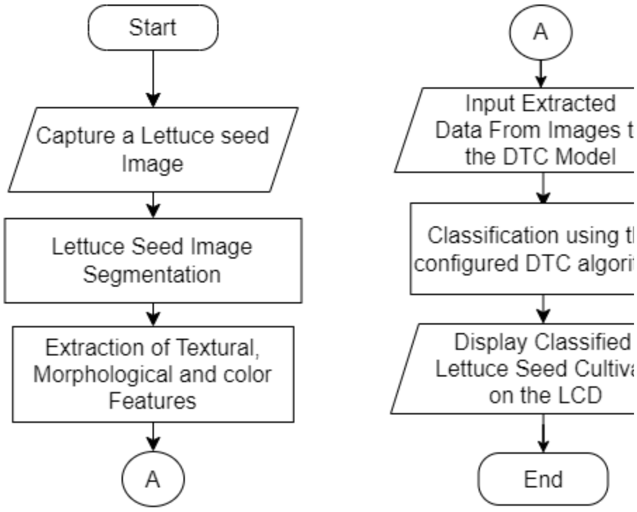

Fig. 3: System flowchart

2.5. Seed segmentation

Lactuca Sativa seed segmentation is executed to get rid of pixels that are not regions of interest and to highlight the entire seed. To properly account for the morphological and colour features, the seeds are segregated from an HSV image. The RGB images were converted to HSV colour space and by colour thresholding of the saturation channel the seeds were segmented from the images. The segmented RGB images were converted to grayscale images to be used for texture feature extraction. A binary image of the segmented image was used for extract the morphological features from the black and white seed images.

A **B** **C**

Fig. 4: Segmented Seed Images (a) Romaine, (b) Black Seeded Simpson (c) Lollo Rosso

2.6. Feature extraction

The features extracted include RGB colour features, Solidity, Correlation, Compactness, Perimeter, and Entropy. Solidity describes the measure of the curvature of the seed image. Compactness describes the curviness of the edge of the seed. Perimeter is the distance along the boundary of the seed. The gray level co-occurrence matrix (GLCM) was obtained from the segmented grayscale seed images. GLCM is a statistical texture analysis method used to obtain second order statistical texture features by evaluating the spatial relationship of pixels. The GLCM of an image is used to calculate several statistical texture features which describes the intensity variation at a target pixel. The Correlation and Entropy texture features were derived from the GLCMs. The RGB colour features were obtained from the segmented RGB seed images. For this the mean values of the red, green, and blue channels obtained from the RGB image were taken.

2.7. Classification

The data gathered from extracting features from the seed images are split into training and testing data, where the researchers used 80% of the data for training and 20% for testing. This method of splitting data is also known as the hold out method. The Decision Trees Classifier (DTC) will be used to train and classify the data as it has been used by previous researchers and had success with it as it proved accurate when it comes to classifying seeds. The DTC in this study is implemented using Python 3 through scikit learn.

2.8. Experimental setup

Fig. 5 shows the experimental setup which contains the controlled environment, where the seed images will be taken, and the hardware components including the Raspberry Pi 3B, webcam, LCD, and power supply. A controlled environment is used to collect seed images to systematize the process of acquiring the seed images and make it reproducible. The box is approximately 5 inches in length and width, and approximately 6 inches tall. To avoid shadows the LED strip lights are placed inside the box on the top. The camera will be placed on top of the box to capture the image of the seed.

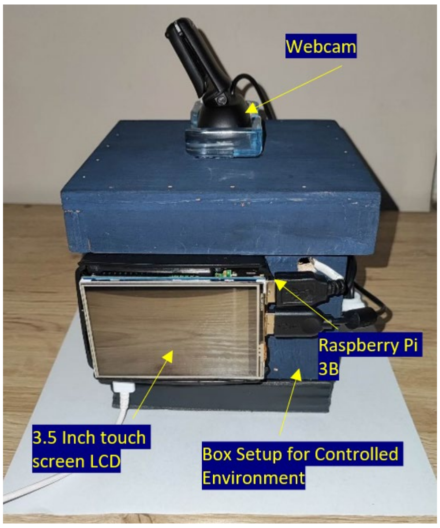

Fig. 5: Experimental setup

3. Results and Discussion

A sample image of each lettuce variety was taken in the controlled environment. The captured seed images are of good quality and clearly display the distinctive characteristics and traits of each seed type. The images of the lettuce seeds are brightly lighted to bring out even the tiniest characteristics of the seeds against a clean, white backdrop. The lettuce seed images still contain some faint shadows. Fig. 4 shows an example picture of the segmented seed for each type. The images were segmented using only the saturation channel, with a lower threshold value of 30 and an upper value of 255. The segmented image has a black backdrop and the whole seed is highlighted with little to no error. The segmented seed picture preserves the colour and characteristics of the seed.

3.1. Statistical treatment

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

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

Equation 1 shows the formula used in the study to calculate for the Accuracy. In the formula, 4 variables can be found which are the TP, TN, FP and FN. TP is the true positives while the TN is the true negatives. FP is the number of false positives and FN is the number of false negatives. Equation 2 shows the formula for precision which will also be used to judge the performance of the lettuce seed classification system.

3.2. Classification results

The confusion matrix shown in Table 1 shows the results of the testing of the decision tree model which is used to determine the accuracy. The Lollo Rosso seed variety had the best classification accuracy (98.33%), followed by Black Seeded Simpson (90%) and Romaine (88.33%). The Lollo Rosso class had a precision of 0.95, the Romaine class got a precision of 0.85, and the Black Seeded Simpson had the lowest precision of 0.8. The algorithm built a model that could accurately categorize lettuce seed varieties using all three types of features, including colour, texture, and morphology. To make sure that all pertinent features were included, the algorithm included features such as correlation, entropy, perimeter, compactness, and the r and g channels. This shows that the algorithm successfully used the most crucial traits and patterns from the seed data, enabling it to produce accurate predictions based on a small number of features. Overall, the findings show that the decision tree classification model, with an average accuracy of 92%, was very good at predicting the lettuce seed varieties.

Table 1: Confusion matrix

		Prediction		
		Romaine	Black Seeded	Lollo Rosso
Actual	Romaine	18	2	0
	Black Seeded	4	16	0
	Lollo Rosso	1	0	19

4. Conclusion and Recommendations

The hardware setup created was able to provide a controlled environment to obtain lettuce seed images. From the features extracted the decision tree model used correlation, entropy, perimeter, compactness, and the r and g channels. The GUI application was able to use the hardware setup to create a setup that can be used to classify lettuce seed varieties with great accuracy. Using the Decision Trees Classifier, the system was able to classify seeds of the selected varieties with an average accuracy of 92%. The researchers recommend testing deep learning algorithms for lettuce seed classification. Further studies can also test other combinations of features to increase the model's accuracy.

5. References

[1] S. D. Fabiyi *et al*., "Varietal Classification of Rice Seeds Using RGB and Hyperspectral Images," in *IEEE Access*, vol. 8, pp. 22493-22505, 2020, doi: 10.1109/ACCESS.2020.2969847.

[2] C. H. Mendigoria *et al*., "Seed Architectural Phenes Prediction and Variety Classification of Dry Beans (Phaseolus vulgaris) Using Machine Learning Algorithms," *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)*, Bangalore, India, 2021, pp. 01-06, doi: 10.1109/R10-HTC53172.2021.9641554.

[3] R. Concepcion II, S. Lauguico, K. Siphengphet, J. Alejandrino, E. Dadios and A. Bandala, "Variety Classification of Lactuca Sativa Seeds Using Single-Kernel RGB Images and Spectro-Textural-Morphological Feature-Based Machine Learning," *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Manila, Philippines, 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9400015.

[4] X. Yi, M. Eramian, R. Wang and E. Neufeld, "Identification of Morphologically Similar Seeds Using Multi-kernel Learning," *2014 Canadian Conference on Computer and Robot Vision*, Montreal, QC, Canada, 2014, pp. 143-150, doi: 10.1109/CRV.2014.27.

[5] E. O. Güneş, S. Aygün, M. Kırcı, A. Kalateh and Y. Çakır, "Determination of the varieties and characteristics of wheat seeds grown in Turkey using image processing techniques," *2014 The Third International Conference on Agro-Geoinformatics*, Beijing, China, 2014, pp. 1-4, doi: 10.1109/Agro-Geoinformatics.2014.6910610.

[6] K. Kiratiratanapruk and W. Sinthupinyo, "Color and texture for corn seed classification by machine vision," *2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, Chiang Mai, Thailand, 2011, pp. 1-5, doi: 10.1109/ISPACS.2011.6146100.

[7] C. Hortinela et al., "Classification of Cane Sugar Based on Physical Characteristics Using SVM," 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Laoag, Philippines, 2019, pp. 1-5, doi: 10.1109/HNICEM48295.2019.9072699.

[8] A. N. Yumang, M. Chloe M. Sta. Juana and R. L. C. Diloy, "Detection and Classification of Defective Fresh Excelsa Beans Using Mask R-CNN Algorithm," *2022 14th International Conference on Computer and Automation Engineering (ICCAE)*, Brisbane, Australia, 2022, pp. 97-102, doi: 10.1109/ICCAE55086.2022.9762416.

[9] C. C. Hortinela and K. J. R. Tupas, "Classification of Cacao Beans Based on their External Physical Features Using Convolutional Neural Network," *2022 IEEE Region 10 Symposium (TENSYMP)*, Mumbai, India, 2022, pp. 1-5, doi: 10.1109/TENSYMP54529.2022.9864337.

[10] M. V. C. Caya, R. G. Maramba, J. S. D. Mendoza and P. S. Suman, "Characterization and Classification of Coffee Bean Types using Support Vector Machine," *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Manila, Philippines, 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9400144.

[11] V. V. Quiñones, M. J. Macawile, A. Ballado, J. D. Cruz and M. V. Caya, "Leukocyte segmentation and counting based on microscopic blood images using HSV saturation component with blob analysis," *2018 3rd International Conference on Control and Robotics Engineering (ICCRE)*, Nagoya, Japan, 2018, pp. 254-258, doi: 10.1109/ICCRE.2018.8376475.

[12] K. M. A. Bejerano, C. C. Hortinela IV and J. J. R. Balbin, "Rice (Oryza Sativa) Grading classification using Hybrid Model Deep Convolutional Neural Networks - Support Vector Machine Classifier," *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia, 2022, pp. 1-6, doi: 10.1109/IICAIET55139.2022.9936869.

[13] V. A. M. Luis, M. V. T. Quiñones and A. N. Yumang, "Classification of Defects in Robusta Green Coffee Beans Using YOLO," *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia, 2022, pp. 1-6, doi: 10.1109/IICAIET55139.2022.9936831.

[14] C. C. Hortinela *et al.*, "Milled Rice Grain Grading using Raspberry Pi with Image Processing and Support Vector Machines with Adaptive Boosting," *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Manila, Philippines, 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9400102.

[15] Z. Mabunga, J. Dela Cruz, G. Magwili and A. Samortin, "Development of Sanitary Landfill's Groundwater Contamination Detection Model Based on Machine Learning Algorithms," *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Manila, Philippines, 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9400140.

[16] D. Cruz *et al.*, "Determination of blood components (WBCs, RBCs, and Platelets) count in microscopic images using image processing and analysis," *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Manila, Philippines, 2017, pp. 1-7, doi: 10.1109/HNICEM.2017.8269515.