Differentiation Among Lettuce (L. Sativa) Seed Varieties Grown In Gourmet Farms, Silang Cavite, Philippines Using Image Processing With Fuzzy Logic And Knn As Classifiers

Vaneza May D. Manalo, Edwin R. Arboleda, Jesusimo L. Dioses Jr., Rhowel M. Dellosa

Abstract: Lettuce is one of the most widely used vegetable crops in our diet. We consume lettuce as a base for salads, on sandwiches and burgers to add texture and even as a garnish to decorate trays of food at parties. The crop comprises seven main groups of cultivars differing phenotypically. The difference between lettuce varieties is often determined by inspection of the leaves, which is time-consuming, subjective and prone to error. In this study, three lettuce cultivars: Cos or Romaine, Iceberg, and Latin lettuce are classified using its seeds by image processing. Using Fuzzy Logic and KNN as classifiers, the cultivars are differentiated according to three morphological features: area, perimeter and equivalent diameter from 216 training images and 72 testing images. The study is supposed to be beneficial to lettuce breeders and farmers as it provides a new way to classify lettuce by its seeds. Understanding the different classes of seeds has great significance as it helps in choosing the right seeds for farming.

Index Terms: Classifier, Fuzzification, Fuzzy Logic, Image Processing, KNN, Latin, Lettuce

1. INTRODUCTION

Lettuce is an important crop in the group of leafy vegetables and is frequently consumed for its pleasant and refreshing flavor, besides being easy to prepare [1]. The commonly cultivated species of lettuce, Lactuca sativa L. are produced commercially in many countries worldwide and is especially important as a commercial crop in Asia, North and Central America, and Europe[2]. The crop comprises seven main groups of cultivars including butterhead lettuce, crisphead lettuce, cos lettuce, cutting lettuce, stalk lettuce, Latin lettuce and oilseed lettuce [3]. In this study, three lettuce varieties that are commercially grown in Gourmet Farms, Silang, Cavite, Philippines were studied for seed morphological identification using image processing, and classification was done using fuzzy logic. These varieties are Romaine (Cos Lettuce), Lollo Rosso (Cutting lettuce), and Green Ice (Crisphead lettuce). Total production worldwide does not compare with the major cereal crops, especially rice, corn, and wheat, or with other commodities but among the vegetables, it ranks high [4]. Each lettuce variety is characterized by considerable morphological variation, all of which are identified on the leaf structures of the lettuce plant. On the other hand, the accurate description of the variations of lettuce seeds provides basic information useful for lettuce breeders.

- *Vaneza May D. Manalo is from the Department of Computerand Electronics Engineering, College of Engineering and Information Technology, Cavite State University.*
- *Edwin R. Arboleda is from the Department of Computer and Electronics Engineering, College of Engineering and Information Technology of Cavite State University.*
- *Jesusimo L. Dioses Jr.is from Isabela State University*
- *Rhowel M. Dellosa is from Computer Engineering Department of Asia Technological School of Science and Arts.*

2 RESEARCH METHODS

2.1 Lettuce Seeds

Lettuce seeds were from the production department of Gourmet Farms in Silang Cavite, Philippines. The sample seeds represent three (3) different varieties of lettuce organically grown in the farm for commercial purposes. The images were captured using iPhone 5s global 8 Megapixels Camera and stored for later processing using Asus X455L Laptop Computer of Intel Core i3-5005U up to 2.0 GHz 4 GB RAM and 1 TB hard disk capacity with Microsoft window 10 professional, 64-bit Operating system using MatLab platform.

2.2 Image Acquisition

The snapshots of the lettuce seed samples were taken by placing the samples on a white background. The camera set was held in a position normal to the plane of the lettuce seeds at a distance of 6 inches directly over the plane of the sample. The lettuce seed samples were well spread to avoid samples from touching each other, making the partitioning of the image into multiple segments easier and improving the accuracy of the morphological features. Images were stored in JPEG (Joint Photographic Expert Group) format with size 3264 × 2248 pixels.

2.3 Image Analysis

MatLab (version R2012a) image processing toolbox and vislabels function were used to develop a computer routine algorithm to preprocess the lettuce seed sample images. The purpose of preprocessing in this paper is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. Preprocessing involved converting the sample images to black and white, filling the sample holes and removing noise to extract features and recognize patterns of the lettuce seed samples. After preprocessing, the seeds' physical attributes or morphological features were extracted for classification[4]–[6]

2.4. Data Classification

The output from Matlab image processing toolbox was gathered and tabled in Microsoft Excel according to the

morphological features of the seeds. Since lettuce seed samples may vary in geometric property, the morphological features included in this study are area, diameter, perimeter. This output was then used as the input for the Matlab Fuzzy Logic toolbox and K Nearest Neighbor (KNN) to classify the seed samples. Fuzzy logic and KNN were both used as classifiers based on the extracted features of the lettuce seed images. Fuzzy Logic involves the manipulation of fuzzy variables through a set of linguistic equations which can take the form of IF-THEN rules. The application of this method causes a class member to become a relative one and one data can belong to several classes at the same time. K nearest neighbor, on the other hand, classifies based upon the number of votes from a class' neighbors. In this algorithm a non-classified vector is classified into K similar vectors present in the training set[7][8].

2.5. Fuzzy Logic

Clustering is a process of grouping data in clusters, where data placed in one cluster are more similar to each other than those in other clusters. In 1960, Zadeh introduced fuzzy logic for handling the uncertain and imprecise knowledge in real-world applications. Fuzzy C Means centroid of a cluster is calculated as the mean of all points' value, weighted by their degree of belonging to the cluster[9][10]. The varieties of lettuce are predicted by extracting the knowledge from the results of image processing using Matlab

2.6 K Nearest Neighbor Classifier

K-Nearest Neighbor (KNN) classification is one of the most fundamental and simple classification methods. When there is little or no prior knowledge about the distribution of the data, the KNN method should be one of the first choices for classification. KNN is the most basic type of instance-based learning or lazy learning. It assumes all instances are points in n-dimensional space [11]–[13].

2.7. Classification Model

To classify three lettuce varieties, Cos, Iceberg, and Latin, three main steps are made namely, feature extraction from the lettuce seed sample images, classifying through fuzzy logic, and further verification by K-Nearest Neighbor Classifier. Figure 1 shows the process. S represents the data sets from1-96 for Iceberg, 97-192 for Cos and 193 for 288 for Latin. Classification using morphology has 3 features, namely area, perimeter, and equivalent diameter[5]. The output was represented by the three lettuce varieties, cos, iceberg, and Latin.

3 RESULT AND ANALYSIS

The lettuce seed images of figure 1 were converted into grayscale images for morphological features extraction. Thereafter, the images were enhanced by complementing image to make background black and objects white. Then the image is converted to black and white. The next step is for the holes in the image to be filled.

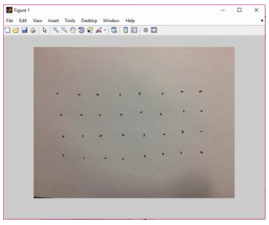

Fig 1. *Original Image*

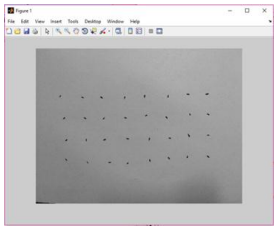

Fig 2. *Gray Image*

3.1 Morphological Features Extraction

Morphology refers to the geometric property including the shape and size of an object Three (3) morphological features: area, perimeter and equivalent diameter of each sample images were calculated from the images and their average values were taken for classification. In Figure 7, the lettuce seeds can be identified as coffee by setting the size range. Using the command, idx=find((270<=area_values)&(area_values<=1000), h=ismember(f,idx) will list all objects falling in that range. These objects are interpreted as lettuce seeds as they fall into that range. The morphological features extracted from image processing are shown in Table 1.

3.1 Feature Extraction

Table 1 shows the mean of the four features. Among the three varieties of lettuce seeds, the Latin seeds have the largest area, while Cos has the smallest diameter. Iceberg has the smallest perimeter and Latin has the largest perimeter.

TABLE 1

SAMPLE SHAPE FEATURES EXTRACTED FROM 288 LETTUCE SEEDS OF DIFFERENT VARIETIES

FEATURES	ICEBERG	COS	LATIN
AREA	381 to 850	121 to 1597	224 to 1718
DIAMETER	22.0251 to 32.8976	14.9271 to 45.0928	16.888 to 46.0705
PERIMETER	86.7696 to 120.8112	97.4975 to 301.1787	118.4648 to 291.2203

3.2 Inputs and Outputs of the Fuzzy Logic

In this paper, a fuzzy logic system is constructed with three input variables, the morphological features: area, diameter, and perimeter of each of the three lettuce varieties, and one output variable: the size of the lettuce seeds. In this order, the fuzzy logic that evaluates the seed sizes (seen as permanent system output) takes into account the possible input variables from Table 2. The variation interval for the output variable can be seen also in Table 2.

TABLE 2

POSSIBLE FUZZY LOGIC INPUT VARIABLES

Features	Range
Area	300-1750
Diameter	22-48
Perimeter	85-300

3.2 Fuzzification

Figure 1 shows the input and output of the fuzzy logic system. Figure 2 shows the first input variable of the system which is the area of the lettuce seed. The categories are divided into

triangular membership functions namely: small medium and large. Input 3 on the other hand, shows the perimeter of the lettuce seed as shown in figure 4. This ranges from small medium and large. Input 4 shows the last parameter which is diameter in figure 3,that is characterized by the same range.These three inputs will output a triangle membership function which will serve as the control signal and will determine the output to be used.

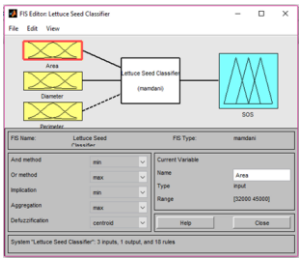

Fig 1. *Input and Output of the System*

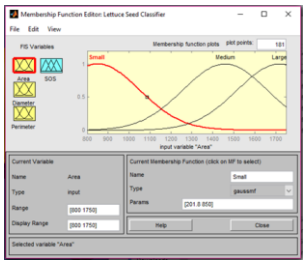

Fig 2*. Area of Lettuce Seed*

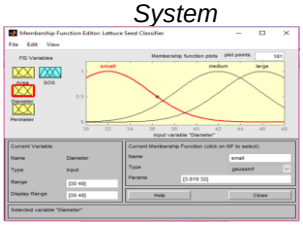

Fig 3. *The diameter of lettuce seeds*

Fig 4. *The perimeter of lettuce seeds*

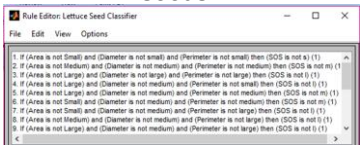

Fig 5. *Set of Rules*

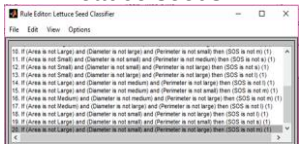

Fig 6. *Set of Rules*

Every rule was evaluated and the truth value of each input membership function used in a specific rule is examined. The minimum value is taken as the representative value for the system. That value is then assigned to each of the output membership functions used by that rule. Input functions used by a specific rule are guaranteed to be represented when that rule is evaluated by taking the minimum values of all the input functions. After evaluating each rule, all of the output membership function truth values for each particular output are examined and the maximum value for each singleton is taken as the representative value for that singleton. After this min-max process, representative values for each singleton in the outputs are for defuzzification by computing a weighted average[14].

3.3 K Nearest Neighbor Classification Using Morphology Features

In this experimental setup, three morphological features were used: area, perimeter, and equivalent diameter. Testing data is used to provide an independent measure of the proposed model's performance during and after training. In the training phase, 216 samples were used. In the testing phase, 72 samples were used.

TABLE 3
RESULTS OF CLASSIFYING COFFEE BEANS OF DIFFERENT VARIETIES USING KNN

K	Percentage of Right Classification (in %)			Mean
	Iceberg	Cos	Latin	
1	95.83	50	70.83	72.22
2	95.83	50	70.83	72.22
3	91.67	54.17	50	65.28
4	91.67	58.33	66.67	72.22
5	95.83	70.83	58.33	75.00
6	95.83	62.50	62.50	73.61
7	95.83	54.17	54.17	68.06
8	95.83	54.17	62.50	70.83
9	95.83	62.50	54.17	70.83
10	95.83	58.33	54.17	69.44

Table 3 shows the result of KNN classification of lettuce seeds of different varieties. The proposed KNN classifier has a high percentage classification for the Iceberg lettuce variety, for k=1 to 2 and k=5 to 10 it got 69 out of 72 correct classifications while at K= 3 to 4, it got 62 out of 72 correct classifications. The morphological features of Cos and Latin are overlapping thus they are very close neighbors as shown in Table 1. The results in table 3 show that there is a high probability that Cos is classified as Latin and vice versa using the KNN classifier. The k=5 yields the highest percentage of right classification aside from k=1, k=2, and k=3, which means that the features of a single seed compared to 4 of its neighbors will yield the same features of its variety.

4 CONCLUSION

This paper introduces Fuzzy Logic and K nearest neighbor for lettuce seed classification. The computer can classify the three varieties of lettuce seeds grown for commercial use in Gourmet Farms, Silang Cavite, via the seed images loaded from digital cameras. The fuzzy system defined in this paper makes possible the correlation between three lettuce seed varieties. Three features are extracted and processed to form the input vector of the KNN. The experimental result indicates that our algorithm is workable with accuracy of 75% (k=5) on three lettuce seed varieties: Cos, Iceberg, and Latin.

REFERENCES

[1] E. Kristkova, A. Lebeda, A. Novotna, I. Dolezalova, and T. Berka, "Morphological variation of Lactuca serriola L . achenes as a function of their geographic origin," Acta Bot. Croat., vol. 73, no. 1, pp. 1–19, 2014.

[2] É. P. Botrel and L. A. A. Gomes, "Evaluation of crisphead lettuce cultivars in different cover types by manova and discriminant analysis," pp. 439–444, 2013.

[3] I. C. Valenzuela et al., "Quality Assessment of Lettuce using Artificial Neural Network," in 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2017, pp. 1–5.

[4] E. R. Arboleda, "Comparing Performances of Data

Mining Algorithms for Classification of Green Coffee Beans," Int. J. Eng. Adv. Technol., vol. 8, no. 5, pp. 1563–1567, 2019.

[5] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "Classification of Coffee Bean Species Using Image Processing, Artificial Neural Network, and K Nearest Neighbors," 2018 IEEE Int. Conf. Innov. Res. Dev., no. May, pp. 1–5, 2018.

[6] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "An Image Processing Technique for Coffee Black Beans Identification," in 2018 IEEE International Conference on Innovative Research and Development (ICIRD), 2018, no. May, pp. 1–5.

[7] R. E. T. Bae, E. R. Arboleda, A. Andilab, and R. M. Dellosa, "Implementation Of Template Matching, Fuzzy Logic And K Nearest Neighbor Classifier On Philippine Banknote Recognition System," Int. J. Sci. Technol. Res., vol. 8, no. 08, pp. 1451–1453, 2019.

[8] C. L. Macalalad, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa, "Morphological Based Grain Comparison of Three Rice Grain Variety," Int. Journal Sci. Technol. Res., vol. 8, no. 08, pp. 1446–1450, 2019.

[9] K. E. N. T. Rabe, E. R. Arboleda, and R. M. Dellosa, "Fuzzy Logic Based Vehicular Congestion Estimation Monitoring System Using Image Processing And KNN Classifier," Int. J. Sci. Tehnol. Res., vol. 8, no. 08, pp. 1377–1380, 2019.

[10] A. E. Andaya, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa, "Meat Marbling Scoring Using Image Processing with Fuzzy Logic Based Classifier," Int. J. Sci. Technol. Res., vol. 8, no. 08, pp. 1442–1445, 2019.

[11] I. M. I. Subroto and A. Selamat, "Plagiarism Detection through Internet using Hybrid Artificial Neural Network and Support Vectors Machine.," Telkomnika, vol. 12, no. 1, pp. 209–218, 2014.

[12] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, and D. Amorim Fernández-Delgado, "Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?," J. Mach. Learn. Res., vol. 15, pp. 3133–3181, 2014.

[13] N. F. Ghazali, N. Shahar, N. A. Rahmad, N. A. J. Sufri, M. A. As'ari, and H. F. M. Latif, "Common sports activity recognition using inertial sensor," in Proceedings - 2018 IEEE 14th International Colloquium on Signal Processing and its Application, CSPA 2018, 2018, no. March, pp. 67–71.

[14] G. M. Reddy and T. G. Manohar, "Fuzzy Logic Controller for Power Balance Theory-Based STATCOM for Grid Connected Wind Turbine System," Indones. J. Electr. Eng. Comput. Sci., vol. 11, no. 1, pp. 409–412, 2018.