Feature Extraction and Classification of Green Mung Bean Using Machine Learning

Teklehaymanot Girma, Sudhir Kumar Mohapatra, Befkadu Belete

Abstract— Development of a machine learning vision system aiming in the establishment of technological and innovative approaches towards sample green mung bean raw quality value classification by extracting the relevant green mung bean features is the focal issue of this exploratory research. The major objective of the research spans extraction and selection of the important green mung bean morphological and color features that are useful for the purpose of classification of the raw quality grade level of sample green mung beans by designing, analyzing and testing a digital image processing model. The automated raw quality value classification experimentation comprised the analysis of images of green mung bean using major attributes of morphological structures (shape and size), and color features. The sample of green mung bean providing a total of 65 samples, which yielded 386 sample images after a series of re-sampling measures of same into 3 sub-samples. The overall image processing work to develop models and depict trends for an efficient raw quality value classification involved sequential phases of image acquisition, image enhancement and segmentation, feature extraction, attribute selection, classification and performance evaluation. Discretization of the raw quality value in to three interval classes was done to improve the performance of the model. 80% split evaluation technique was implemented for the Logistic Regression, kNN, SVM, Naive Bayes and RF classifiers. In kNN classifier yielded higher model performance (96.89% correctly classified), followed by RF (93.58%).

Index Terms— Green mung, kNN, RF, SVM, logistic regression, naïve bayes, feature extraction, classification.

---

1 Introduction

The Mung bean, known in Amharic as MASHO, is an annual crop. It looks like a garden bean than like a soya bean plant. Mung bean is a warm season crop requiring 90–120 days and it is an erect plant which is branched and is about 60 to 76 cm tall. 150,000 to 200,000 quintals of Mung bean are produced per year in Ethiopia. It is mostly found in North Shewa and Southern Wollo but it also found in areas like Assosa, Hararge, Ilubabor, Gamogofa, Tigray, Gonder, Konso, South Omo zone and Konta [1, 2].

Green mung tra[1]ding and circulation in Ethiopia is subjected to a standard set by the. Quality and Ethiopian Standards Agency (ESA). This standard sets criteria by which green mung quality is evaluated. The standard is based on morphological and color characteristics of green mung. When consignment of green mung arrives at the authority, small amounts of green mung are taken from different sacks and get mixed. Based on this, the green mung sample are manually classified into impurities, defect and healthy classes

Green mung bean is a very important crop in the export of Ethiopia as large amount is exported around the world market. ESA controls the quality grading of green mung bean in traditional way.

Currently, green mung bean quality prediction is assessed manually. However, manual evaluation takes significant amount of time and it requires trained experts. This is especially evident during large scale inspection in the process of exporting the crop. The farmer who need to take his product to the place where it can be graded. If there is any such proven technology available, then the farmer can judge their crop quality at home. The process of grading is time taking and boring process for human experts [3].

Quality prediction system using machine learning for green mung bean is relevant to explore the possibilities of adopting faster system which saves more precise in grading of green mung bean by reducing experts effect of bias related to the quality standard that improves the commercial needs. Therefore, replacing manual inspection of green mung bean with a superior speed, precise, cost effective, consistent, non-destructive automated system is necessary for green mung bean in which it generates huge amountof income and forging currency to the country.

The proposed system has a great advantage for minimizing the needs of experts and experience. Because, the developed system is an automated system and it minimize the needs of expert in that area. It also used for avoiding the subjectivity, bias, inconsistency and tiredness associated with human nature. The system facilitating the exchange of green mung bean between the dealer and customer**.**

2 Related Work

Youwen Tian et al. proposed the prediction mechanism of crop dieses by using computer vision. In this research the proposed method uses image processing, segmentation and statistical calculation. [4].

Narendra V G et al have discussed about the usage of image analysis and computer vision in the prediction and sorting of agricultural and food products. Both image analysis and computer vision are cost effective and non-destructive techniques that we can use them for prediction and sorting of agricultural and food products [5].

B. Ni et al. proposed a new system for quality prediction the size of corn kernels by the help machine vision. For the grading length, width, thickness and projection area are used. These all parameters are used for classifying features [6].

N.S. Visen1et al. proposed a new algorithm used to extract different features of bulk grain samples by the using of neural network. They have used five different grain types. These are

Teklehaymanot Girma, Department of Software Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia.
Sudhir Kumar Mohapatra, Department of Software Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia.
Befkadu Belete, Department of Software Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia.

oats, wheat, barley, rye and durum. On the training of the algorithm more than 150 color and textural features are extracted. The proposed algorithm is based on back propagation neural network-based classifier. If there is unknown grain type is found, the proposed algorithm can easily identify it. Finally, 98% of accuracy is achieved for the classification [7].

S. Gunasekaran et al. have proposed a system for soybeans in which to evaluate mold contamination, external physical damage and floury-to-vitreous endosperm ratio by using computer vision. Thresholding algorithms were developed and it has the threshold value of 23,27, and 25 respectively which are used to evaluate mold contamination in corn, mold contamination in soybeans, and external damage evaluation [8].

Tom Patten et al. proposed a system for the automatic analysis of Heliothiszea insect by using of image analysis. This insect eats the corn crop and there are different insecticidal bio-toxins are developed to kill it or to stop the growth of it. In the proposed system by following three steps; insect segmentation, regional processing and instar and life classification, the insect was analyzed. The proposed algorithm follows a back-propagation neural network and it has a result for insect count, instar and life with average value of 84%, 62% and 69% respectively [9].

Geng Ying et al. have studies about the recognition of crop dieses based on image processing methods. First by using image clipping technology, the leaf with spots is separated from the complex background. Then two filters, simple filter and median filter, are compared. The median filter is selected and it is used to wiping noises for the image. At last by using threshold the leaf is separated from its background and by using Laplacian operator technique of edge detection and Snake model the spots are separated from the normal leaf. This leads to good segmentation effect [10].

Cheng-Jin et at. have discussed about different algorithms in which we can use for evaluating food quality using computer vision. They briefly discussed about different computer vision techniques like artificial neural network (ANN), statistical learning (SL), fuzzy logic, genetic algorithm and decision tree [11].

Based on no free lunch (NFL) theorem, there is no single nature-inspired optimization technique, which can optimally solve all optimization problems [69]. This means that an optimization algorithm is competent for solving a certain set of problems but ineffective on other class of problems [12].

3 Proposed Methodology

In order to accomplish the objectives of the research a total of 40 literatures are reviewed then sampling. Sampling is one of the main procedures in green mung bean classification and quality assessment. In the current practice of the manual system, the sample drawer draws a representative sample of 3 kg per 10 tons of a truck, which is an average carrying capacity of a truck, on its arrival.

In this regard, we have taken 65 images in which the images contain 386 green mung beans. From these samples, 80% were used for training and 20% were used for testing purposes. The samples of green mung beans were obtained from ECX in which under the Ministry Agriculture and Rural Development.

A digital camera model DSC- S650, SONY 7.2 Mega Pixel, was used to record green mung bean images. When images were taken, the camera was mounted on a stand which provides easy vertical movement and stable support for the camera. The camera was fixed at a distance of 130 mm from the sample table in-order to get clear images of green mung beans.

Extraction of features from an image for further measurement or interpretation is termed as feature extraction. To implement the proposed model for the classification of green mung kNN, Logistic Regression, RF, SVM and Naive Bayes classification approaches were compared. The classification systems were supervised because classes were predefined that correspond to the crop quality type.

Our system is developed and tested on a PC of Intel Core i7-3632QM CPU with 2.20GHZ speed, 8192 MB of RAM, 500GB of hard Disk capacity, with Microsoft Windows 10 Enterprise operating system.

In this research our interest is to address the prediction problem of green mung bean by using machine learning. The schematic representation of green mung bean feature extraction procedure is described in Figure 1.

Image analysis starts with image acquisition. The selection of radiation (light) sources and sensors (such as cameras) has to be considered very carefully. The geometry of the viewing situation, i.e., the relative positioning of sources and camera with respect to the objects of interest, usually also has a major impact on the contrast between these objects and their background. Figure 2 (a) shows image accusation of green mung.

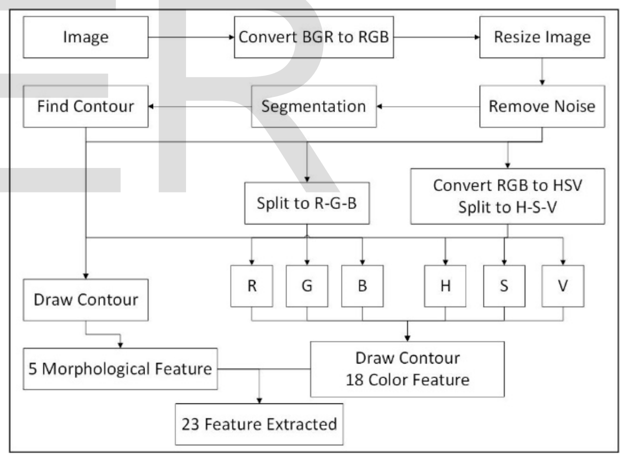


Fig. 1. Schematic representation of green mung feature extraction procedure.

Once the image captured the next step is started from rendering of images on computer screen and storing images on hard disks or other media for further processing. It includes all tasks in image pre-processing, image segmentation and removal of noises.

From the original green mung bean images, the first step image pre-processing starts by converting the image color format from BGR to RGB, hence Python OpenCV reads image in BGR format.

After the color format changed then the image is resized to

354 x 234, because we need to get a standard sized image. The final step in image pre-processing is removing the noise by using Gaussian Blur function.

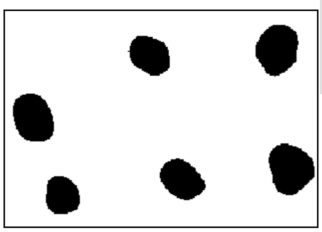
Hence, as shown in Figure 2 (b) the captured image was blurred by using Gaussian Blur. We use (5, 5) as the input for the blurring function. Then the blurred image is converted to the gray scale image.

In Figure 2 (c) the gray scale image was segmented with threshold value of 127 by using histogram analysis
The background is subtracted in-order to avoid blurs, light distortions and other noises that are formed due to illumination effects. The enhancement of original image by subtracting the background was shown in Figure 2 (c).

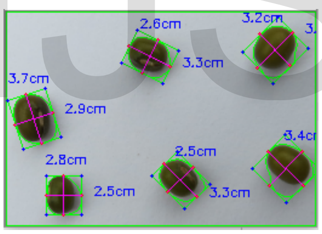
As described above, image analysis is the process of extracting meaningful information from images that are used for classification of images to different categories. For image analysis of green mung bean morphological and color features are used as a classification parameter that was shown in Figure 2 (d).

(a) A green mung bean

(b) Gaussian blur



(c) Image segmentation



(d) Measuring size of objects in an image with OpenCV

Fig. 2. Experimental result

4 Experiment

4.1. Feature extraction

Morphology and color feature are extracted on this research. Morphological feature means the geometric property of images. In our case it is the size and shape characteristics of green mung beans. It can be obtained from the analysis of binarized images. From morphology of green mung beans, the following geometric features were extracted from the binarized images as described in the previous section.

1. Area (A): The number of pixels inside the region covered by a green mung bean, including the boundary region. It is measured by square pixels.
2. Perimeter (P): The length of the outside boundary of the region covered by a green mung bean.
3. Major Axis Length (Major): It is the distance between the end points of the longest line that could be drawn through the green mung bean. The major axis end points are found by computing the pixel distance between every combination of border pixel in the green mung bean boundary and finding the pair with the maximum length.
4. Minor Axis Length (Minor): It is the distance between the end points of the longest line that could be drawn through the green mung bean while maintaining perpendicularity with the major axis.
5. Aspect Ratio (Elongation): The ratio of the length of the major axis to the length of the minor axis (Elongation =Major/Minor)

Natural color images incorporate three sensors that are spectrally sensitive to the red, green and blue portions of the light spectrum which are collectively described as RGB images. Graphics file formats store RGB images as 24-bit images, where the red, green, and blue components are 8 bits each [13]

In relation with RGB colors, there are three common perceptual descriptors of a light sensation. They are intensity (I), Saturation (S) and hue (H). Intensity is a measure of the brightness of a given image while saturation describes the amount of whiteness of a light source in a given image. The hue is also an attribute of light that distinguishes one color from the other, for example a red color from green or yellow color. The mathematical formula that converts RGB color space to HSI is given as follows [13].

	A	B	C	D	E
1	area	perimeter	major axis	minor axis	aspect ratio
2	7.624618182	119.3969686	3.001818182	2.54	1.181818182
3	8.49253315	130.3675311	3.319973475	2.55801235	1.29787234
4	12.44110193	153.1959581	3.848484848	3.232727273	1.19047619
5	10.58086097	145.3969686	3.617575758	2.924848485	1.236842105
6	11.42805037	149.1959581	3.838079874	2.977543653	1.289008767
7	8.510902071	127.8822498	3.14965041	2.702173563	1.165598855

Fig. 3. Morphological features computation

Hence, the color features are extracted by computing the mean values of RGBs and HSVs of green mung bean images. That is, the mean value of red, mean value of green, mean value of blue, mean value of hue, mean value of saturation and mean value of intensity are computed from each component. The results were shown in Figure 3. The measured values of each selections.

	F	G	H
1	mean red	median red	std red
2	97.48051948	96	43.58946924
3	137.2654321	137	10.2714955
4	174.765625	177	39.08873806
5	61.35638298	59	23.54704688
6	131.8846154	134	80.8195853
7	85.51851852	86	40.88982822

Fig. 4. Color feature extraction for Red, Green and Blue

As described above we have identified 18 color features. They were the mean value of RGBs (Red, Green and Blue) components and the mean value of HSVs (Hue, Saturation and Value) components. Therefore, to compute the mean val-

ISSN 2229-5518

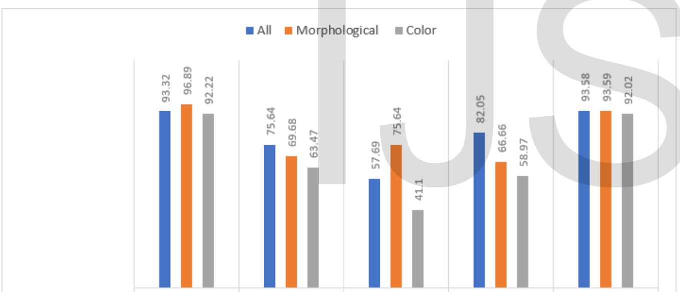
ue of each component of these color spaces we need to split each component to separate image stack. To do these, we have a built-in function called split. It reads RGB image and returns three values which are the red, green and blue.

After each color space was split, the mean, median and standard deviation values of each component were computed by OpenCV built-in function particle analyzer method. As an example, the results of red color feature component were shown in Figure 4.

In the training phase, data is repeatedly presented to the classifier, while weights are updated to obtain a desired response. In testing phase, the trained system is applied to data that it has never seen to check the performance of the classification. Hence, we need to design the classifier by partitioning the total data set into training and testing data set. From the total data set of each region, 80% was used to build training and the remaining 20% of the total was used for testing data. Hence, from the total of 386 data sets, 308 were used for training and 77 were used for testing.

4.2. Classification

For the classification we have used five supervised machine learning classification algorithms are taken for comparison. This are k-Nearest Neighbor (kNN), Logistic Regression, Naïve Bayes, Support Vector Machine (SVM) and Random Forest (RF).



	kNN	Logistic Regression	Naïve Bayes	SVM	Random Forest
All	93.32	75.64	57.69	82.05	93.58
Morphological	96.89	69.68	75.64	66.66	93.59
Color	92.22	63.47	41.1	58.97	92.02

Fig. 5. Performace of the modle in different classifiers

Statistical computations dealing with green mung bean raw quality value classification yielded almost similar model outputs for the various classification tools employed (Table 5.13), The higher model performance value in kNN is attributed to suitability of the classifier under smaller number of datasets. This together with the resulting higher model statistic values and lower mean and relative error rates for performance evaluation depict the suitability and possibility of implementing the automated classification system using similar dataset orientation and tools. Better performances could even be achieved through further in-depth implementations involving larger samples [14].

It becomes obvious that the two different green mung bean features when used in combination and separately in the classification tools yielded various model outputs. The combined feature classification models provide the higher model performance values, succeeded with those of the morphological features only Figure 5.

5 Conclusion

Automated feature extraction and classification systems for agricultural products are proven to be less costly, efficient and non-destructive. Application of this technology makes effective quality control and inspection aspects for such economically important commodities. With this regard, this research has focused on using image processing techniques and approaches to classify raw quality value of sample green mung beans by employing logistic regression and k-nearest neighbor classification approaches. The achievements obtained in this research work indicate the possibility of applying classification of raw quality value using computer vision system. Morphological and color features were the attributes extracted from the sample green mung bean images of various grade levels and used for the classification purpose. Image pre-processing techniques on the original images enhanced the harvesting of such important features in a reliable and effective manner. The classification models built with all the classifier tools were evaluated for almost a homogeneous performance, with the k-Nearest Neighbor yielding the highest model performance (96.89% correctly classified), Random Forest (93.58%), Support Vector Machine (88.05%), Naive Bayes (75.64%) and Logistic regression (75.64%). Combined aggregate feature values of both the morphological and color features were used to build and evaluate the base quality value classification model with all classifiers used in this research.

Sensitivity of the classifier model was however attempted by running the classification model for the separate aggregate feature values, i.e., morphological and color features. The separate trials produced higher performances in all classifiers for the morphological features than the color features, reflecting the suitability of the morphological features for grading and sorting the green mung bean samples than the color features. It should however be underlined that the highest overall performances and hence suitability for classification purposes in this research is concluded for the combined morphological and color features.

An important recommendation to forward from this research could be the launching of developing a working base classification model for raw quality value classification purposes by utilizing larger number of datasets from each grade level of green mung bean sample. This could also be supported by another future research that aims at prediction of actual raw quality values by utilizing a meaningfully larger dataset. Important is due consideration to the number and type of samples and to the data acquisition environment to effectively achieve the intended end-goals of developing an applicable automated computer vision system for this purpose. This automated technique might also be a potential approach in Ethiopia assisting quality control and grading/sorting activities of other important agricultural products like fruits and cereals.

To conclude the future work, the following recommendations are made for further research and improvements.

1. Test the classification with other classification algo-

rithms

2. Include the moisture content analysis and mass determination
3. Use mobile phone as image acquisition tool

ACKNOWLEDGMENT

First of all, I would like to express my deepest gratitude to my advisor Dr. Sudhir for his motivative and constructive guidance right from the moments of problem formulation to the completion of the work. Many thanks and appreciations go to him for the discussions with him always made me think that things are possible. His enthusiasm and encouragement have always inspired me to accelerate to the completion of the work. I am also very thankful to my instructors and all staff members of the department of Software Engineering for their contribution in one way or another for the success of my study. I would like to forward my special thanks to the Ethiopian Commodity Exchange for supplying samples of green mung bean. I would like to extend my appreciation and thanks to laboratory experts and all staffs members of the organization for their kind cooperation. Particularly, many thanks go to Ato Daniel for his constant assistance in explaining the work in the problem domain and giving me reading materials. Finally, I am very grateful to my families, friends, and peers who, in one or the other way brought me up to a success in my academic endeavor.

REFERENCES

[1] J. Keatinge, W. Easdown, R. Yang, M. Chadha, and S. Shanmugasundaram, "Overcoming chronic malnutrition in a future warming world: the key importance of mungbean and vegetable soybean," *Euphytica,* vol. 180, no. 1, pp. 129–141, 2011.

[2] G. Wedajo, "Adaptation study of improved mung bean (vigna radiata) varieties at alduba, south omo, ethiopia," *Research Journal of Agriculture and Environmental Management,* vol. 4, no. 8, pp. 339–342, 2015.

[3] M. Ahmed, M. Siraj, and S. Mohammed, "Marketing practices and challenges of mung bean in ethiopia amhara regional state: North shewa zone in focus," *African Journal of Marketing Management*, vol. 9, no. 4, pp. 35–45, 2017.

[4] Q. Su, N. Kondo, M. Li, H. Sun, D. F. Al Riza, and H. Habaragamuwa, "Potato quality grading based on machine vision and 3d shape analysis," *Computers and Electronics in Agriculture*, vol. 152, pp. 261–268, 2018.

[5] E. S. Pariente, J. C. Cancilla, K. Wierzchos, and J. S. Torrecilla, "Onsite images taken and processed to classify olives according to quality–the foundation of a high-grade olive oil," *Postharvest Biology and Technology*, vol. 140, pp. 60–66, 2018.

[6] T. Brosnan and D.-W. Sun, "Inspection and grading of agricultural and food products by computer vision systems—a review," *Computers and electronics in agriculture*, vol. 36, no. 2-3, pp. 193–213, 2002.

[7] L. W. Steenhoek, M. K. Misra, C. R. Hurburgh Jr, and C. J. Bern, "Implementing a computer vision system for corn kernel damage evaluation," *Applied Engineering in Agriculture*, vol. 17, no. 2, p. 235, 2001.

[8] A. Mohapatra, S. Shanmugasundaram, and R. Malmathanraj, "Grading of ripening stages of red banana using dielectric properties changes and image processing approach," *Computers and Electronics in Agriculture,* vol. 143, pp. 100–110, 2017.

[9] Y. Tian, L. Wang, and Q. Zhou, "Grading method of crop disease based on image processing," in *International Conference on Computer and Computing Technologies in Agriculture*. Springer, 2011, pp. 427–433.

[10] V. Narendra and K. Hareesha, "Prospects of computer vision automated grading and sorting systems in agricultural and food products for quality evaluation," *International Journal of Computer Applications*, vol. 1, no. 4, pp. 1–9, 2010.

[11] C.-J. Du and D.-W. Sun, "Learning techniques used in computer vision for food quality evaluation: a review," *Journal of food engineering*, vol. 72, no. 1, pp. 39–55, 2006.

[12] M. Jain, V. Singh, and A. Rani, "A novel nature-inspired algorithm for optimization: Squirrel search algorithm," *Swarm and evolutionary computation*, vol. 44, pp. 148–175, 2019.

[13] T. Acharya and A. K. Ray, *Image processing: principles and applications*. John Wiley & Sons, 2005.

[14] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

AUTHORS PROFILE

Teklehaymanot Girma is a final year MSc. in Software Engineering student at Addis Ababa Science and Technology University (AASTU) under college of Electrical and Mechanical Engineering. He has BSc. degree in Computer Science with distinction. Currently he has completed course study and working on his final thesis

Sudhir Kumar Mohapatra is working as associate professor in Software Engineering department at Addis Ababa Science and Technology University. He has involved in several researches related with image processing, machine learning and software testing field of studies.

Befekadu Belete is working as lecturer in Software Engineering department at Addis Ababa Science and Technology University. He has also a department head of Software engineering department.