*Machine Learning Based Peanut Maturity Classification from Hyperspectral Image*

Sree Nirmillo Biswash Tushar

Department of Electrical and Computer Engineering, Texas State University,

San Marcos, TX-78666, United States of America

***Abstract-* The maturity of the seed is key information to ensure quality of the crops and better economic returns. But the assessment of the maturity of the peanut requires exocarp removal which suffers from observer’s color assessment skill and experience. Moreover, it requires a great amount of time for a large number of peanut and often ends up in a blasting peanut pod. In order to find an optimal solution, researchers tried to apply digital image processing method. Although this method does not cause pod blasting and does not require personal inspection, the method demands exocarp removal. Recently, a research group has shown why traditional RGB image fails to classify peanut maturity and they have introduced a hyper-spectral unmixing based classifier to solve the problem. In this project, I have worked the same problem with logistic regression, random forest and support vector machine to compare their performance with the previous classifier.**

Keywords ***– Machine Learning, Hyperspectral image, Peanut Maturity***

1. INTRODUCTION

Peanut ([*Arachis*](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/arachis) hypogaea)  is consumed all over the work due to its nutrition value and deliciousness. To be specific, peanut is source of protein, fat, carbohydrates, vitamin, and minerals and it prevents cardiovascular disease, cancer, diabetics, and obesity [1]. Therefore, peanut is a vastly cultivated crop in the United States. In 2021, the total production of peanut was around 6.63 billion pounds from 1.63 million acres of land in US [2]. To maximize the peanut production, maturity level of the peanut pod is one of the most important things to know. If the pod is immature while harvesting, it will not provide optimum seed quality, grade, and flavor. On the contrary, over-mature peanut pod can break the pegs where pods are attached [AZ,3].

Visual inspection has been a popular method among agricultural community for maturity classification for a long time. The most recognized method for peanut maturity level classification is MPB (Maturity Profile Board) which maps the color of the mesocarp to the five different maturity categories and the shades within the color into different subcategories [2]. But this method needs exocarp removal and it often causes pod blasting. Since, immature pods are

fragile compared to mature pods, most immature pod blows apart. Focusing on not destroying too many immature pods consumes a considerable amount of time. Moreover, color categorization using MPB requires human visual inspection, which is very subjective to human visual condition, lighting condition, observer skill in discriminating color that leads to the possibility of large error. This is very time consuming when one has to classify large number of pods. Authors in [6] also introduced a mesocarp color and pod size based nearest neighbor classifier for pod maturity classification which requires pod blasting but free from visual inspection. Thereby, recent advancement of vision technology has inspired researchers to automate the peanut maturity assessment which can speed up the process with optimum accuracy.

The idea of non-destructive peanut maturity classification is initiated by authors of [5], who have shown a positive correlation between tannin and maturity level. Interestingly, hyperspectral reflectance also varies with different chemical composition. Therefore, HSI image can be used as a peanut pod tannin distinguisher and hence their maturity level. Using all the information, authors in [4] established a linear unmixing model and fully constrained least square algorithm to classify mature and immature pods. They have made the classification based on the multispectral reflectance of the pericarp which does not require exocarp removal. They have also shown that visible part of the spectrum is indistinguishable for mature and immature pod.

But linear hyperspectral unmixing is a computationally costly process where spectrum of every pixel needs to participate. However, when all pixels of mature and immature peanut are averaged, the resulting average spectrum shows clear distinction between mature and immature peanut [AZ]. Moreover, average spectral feature based peanut maturity classification with a few key wavelength features can leverage a lot of computational time. Earlier, standard machine learning algorithm (SVM and random forest) has shown promising result for pod maturity classification of snap bean based on spectral and biophysical features [Am].

Therefore, the scope of this paper is to explore traditional machine learning (logistic regression, support vector machine and random forest) classifier to classify peanut pod maturity by selecting few key features and demonstrate the trade off between the number features and performance of the model. In this work, multiple peanuts have been identified and classified simultaneously which speed up the process for identifying optimum matured peanut pod. It is shown in this paper that it is feasible to automate the optimum matured peanut identification task by intelligent feature selection and classification.

1. METHODOLOGY
2. Data Preprocessing

All the data preprocessing steps are shown in figure-1 as a part of the methodology. Spectral Python library (SPy) [7] has been used to read hyperspectral image data. Since each image file contains 15 images total, firstly, the peanut pixels have been segmented from the background. After segmentation, calibration panel and non-uniform background have been removed by cropping the segmentation mask beyond calibration panels’ location. Then, all the pixels in the image have been identified and labeled according to which peanut they belong to. For spectral based peanut maturity classification, peanut spectrum has been calculated by averaging the spectrum of all pixels of a peanut. For Spatial-spectral based classification, spectrum at different regions of the peanut has been estimated which has decided the maturity of the peanut.

Diagram

Description automatically generated

Fig. 1 Training and Testing Procedures

1. *Image segmentation*

First of all, reflectance values of Red, Green and Blue wavelength from HSI image were merged together to form an RGB image for segmentation purpose. For red, green and blue channel, 450 nm, 550 nm and 650 nm wavelength were considered respectively. Ostu thresholding has shown good peanut segmentation performance for all of the images. An example of ostu thresholded image is shown in figure-2 where peanut pixels are labeled as white and backgrounds are labeled as black. In the segmentation mask, some peanut’s region has been labeled as background (figure-2). Implementation of the morphological dilation operation with kernel size (5,5) has shown improvement in filtering those false background pixels in peanut region but it enlarges the peanut mask at the edges (figure-3). Segmentation mask found after all of those mentioned operation has been used for the peanut indentification and feature matrix creation.



A picture containing text

Description automatically generated

1. *Indentify each peanut from the image:*

The identification of each peanut in the segmentation mask is a tricky task. The pixel position, row and column of that pixel, of one peanut is far away from other peanuts. Therefore, k-means clustering was algorithm used to group peanut pixel based on the distance between the pixel position. Figure-4 is showing the masks of all identified peanut. The convention of reading peanut number is given in the caption of figure-3. In figure-4, it is clear that kmeans algorithm does not label all the peanut squentially. In figure-4, first peanut (top-left) has been labeled as 8. This issue has been handled by re-ordering the label of all the peanut pixel.

Shape

Description automatically generated

1. *Create Feature Matrix for classification*

Two types of feature matrixes have been created for classification – spectral feature based, spectral-spatial feature based. Section 2 only returns position of each peanut in the image. For spectral feature creation, average reflectange at all the pixel position of a peanut at certain wavelength is calculated first and the average reflectance of all wavelengths gives the spectrum of that peanut. For spectral-spatial feature matrix, whole peanut image has been divided into (2,8) matrix. Then average spectrum is calculated for all the regions of a peanut. Lastly, the information of all the spectrum has been organised as an image of shape (16, Number of wavelength) with all the regions as a row and reflectance at all wavelength in the column.

*B. Create a Dataframe:*

Finally a dataframe has been created where Features are the spectrum of each peanut witheach peanut’s label. The final dataframe is of shape (number peanuts, number of wavelength).

*C. Classifier selection , Classification and Analysis:*

Lastly, a suitable classifier is chosen for maturity classification and hyperparameters are tuned to find out best model. Detailed analysis of classifier selection and performance of the classifier will be discussed in the Section IV.

1. DATASET

The dataset used by [4] were collected from the field experiment of the North Florida Research and Education Centre. In 2016 and 2017, total five and seven cultivators, with 3 replicated plots for each cultivator, were used respectively for experiment. In my project, I have just used 2016 dataset due to the fact of large size of data file. The name of the cultivators used in 2016 are TUFRunner 511, FloRun 157, Georgia-06G and TUFRunner 257, FloRun 331. Total 200 samples were randomly selected from 15 plots of 2016 and again 15 samples from each plot were used to capture HSI image. Thereby, the final dataset contained 225 (5×3×15) peanut pod HSI images from 2016 sample collection. In the provided dataset, some data were found corrupted and therefore, I could only read 205 peanut image. The HSI images were taken using pushbroom line-scan hyperspectral imaging system. Then, all images were denoised using 4th order Savitzky eGolay filter and the resultant images had the shape of 1376×467 for the spectral range of 400nm-1000nm with 1.4 nm spectral resolution. The provided dataset has two files for each 15 peanuts (one header file with an extension .hdr and one data file with an extension .hyp). The quantization of each image was 32 bit and therefore every reflectance value returns 4 bytes.