Spectral discrimination of vegetable crops using in situ hyperspectral data and reference to organic vegetables

Manoj Kaushik
dept. of earth and space sciences
indian institute of space science and technology, Thiruvananthapuram, India
manojkaushik93@gmail.com

Rama Rao Nidamanuri
dept. of earth and space sciences
indian institute of space science and technology, Thiruvananthapuram, India
rao@iist.ac.in

Aparna B
dept. of soil science & agricultural chemistry
college of agriculture, Vellayani, Thiruvananthapuram, India
aparna.b@kau.in

Ramiya A M
dept. of earth and space sciences
indian institute of space science and technology, Thiruvananthapuram, India
ramiya@iist.ac.in

Abstract*— Remote sensing has been evolving as a general method for multi-scale crop information extraction. Large scale multi-crop discrimination using airborne, or satellite remote sensing is required for farm level intervention and for decision making by various stake holders such as agriculture insurance companies, risk assessment agencies and local governments. Multispectral and hyperspectral remote sensing data from various platforms have been used for discriminating and mapping various crops. However, the number of crops, and the scale at which the discrimination is often limited to a few crops and is often approached as discrimination against other land cover classes. Further, multi-temporal datasets and spectral indices form the bulk of the datasets for crops discrimination. For potential operational application at field level, the ability to discriminate numerous crops – at least ten different crops at the same timeframe are vital. Furthermore, discrimination of crops grown under organic practices has promising application in the certification and quality assurance of agricultural produced sold as organic product. Theoretically, high resolution hyperspectral data has the capability to difference few tens of classes unambiguously. However, given the context of systematic spectral similarity in vegetation, especially crops, the potential of discrimination several crops are unclear. We, therefore, has assessed the spectral discrimination of as many as 23 different vegetable crops and attempted discriminating a few vegetable crops grown under organic and inorganic crop growing practices. For this, we have applied 12 different statistical and machine learning algorithms establishing the spectral discrimination and assessing its relative stand across the range of crops considered. The results indicate complex patterns of spectral discrimination wherein a few crops exhibit spectral similarity with several other crops at any scale of spectral characterization. The discrimination analysis of vegetable crops grown under organic and chemical input-based practices indicate a good discrimination. However, the quality of discrimination is substantially affected by the type of machine learning model used. We recommend coordinated multi-site and multi-phenology-based crop discrimination for establishing the stability of the discrimination observed across space and time.

Keywords*— *spectral signatures, hyperspectral imagery, machine learning, classification, organic crops, non-organic crops.

I. INTRODUCTION

Remote sensing has been one of the primary data sources for information required by governments and policy makers and large-scale insurance companies. Thanks to the evolution of aerial platforms such as drones, the relevance of remote sensing has grown by orders of magnitude to evolve as the source of data at farm level. The effective utilization of remote sensing for extracting various types of information on crops require the basic information on crop type. Multispectral and hyperspectral data have been used for discrimination and mapping various crops against other land cover classes. However, for potentially operational level application using airborne or satellite based hyperspectral sensors, it is vital to understand the theoretical limits on the spectral discrimination among crops, ignoring the co-occurrence of other land use classes. There are some multiple studies which have assessed the spectral discrimination of crops using hyperspectral data. In [1-5], the authors have studied the spectral discrimination of three crops; rice, sugarcane, and chillies, using satellite based hyperspectral imagery acquired from Hyperion. The authors report excellent spectral discriminability at the mature phenological stages of crops. In another study [6-9] have assessed the spectral discrimination of tomato plant against other land cover classes using airborne hyperspectral imagery. Most of the existing studies on crop discrimination have considered only a few crops, up to four crops simultaneously. In real, agricultural landscape in countries such India, the number of crops co-occur is more than 10 and represent a complete landscape setting. Therefore, the spectral discriminability of crops vi-a-vis is based on the probability of number of distinct crop types under the same experimental setting. Apart from that, there has been consistent awareness on the method of crop production as perceive from organic agricultural produce. Especially, vegetable crops cultivated using organic crop management practices have been gaining popularity. Commanding a premium price, vegetable crops grown under organic cultivation method is needed a major shift in the sustainable agriculture practices and for managing community health. However, despite certification process, there has been extensive marking practices that claim agriculture produce as organic and without any verifiable parameters from the consumer level. The potential of advanced spectral modelling for discriminating vegetable grown under organic cultivation practices, contrasting with produce grown under common chemical inputs-based practices helps multistage spurious produce banded as organic. The possibility of spectral discrimination of vegetable crops grown under organic practices is not yet attempted. Based on our review of literature, it is clear that spectral discrimination of crops is only attempted as a classification problem with up to four crops against other land cover classes and there are no studies which have addressed the potential of hyperspectral remote sensing for discrimination organic and chemical inputs base vegetable crops. The objective of this research is the evaluation of spectral discrimination of as many as 23 crops co-occurring in the same agricultural landscape and the potential of hyperspectral remotes sensing discriminant organic and chemical inputs-based vegetable crops.

II. STUDY SITE AND DATA ACQUISITION

The study site selected for this experiment is the experimental farming station of College of Agriculture Vellayani, Thiruvananthapuram, Kerala, India, located at (8°25'53.35"N, 76°59'10.86"E). Geographically, the agriculture college is surrounded by fresh water from three sides. There are dedicated 2.52 sq. km. cropland areas for research within the college campus. Thiruvananthapuram has a tropical savanna climate and a tropical monsoon climate. The average maximum temperature is 34.5°C, and the average minimum temperature is 21.3°C.

We collected spectral measurements of various vegetable crops and other co-occurring plants reflectance spectra in December 2021. The typical crops grown in the study site are rice, red spinach, brinjal, cluster beans, cucumber, paddy, ginger, ridge gourd fruit, winged beans, etc. Some of these crops are available both in organic and non-organic varieties. Two types of spectroradiometers, e.g., ASD FieldSpec3 and SVC HR1024i, are used to collect the spectral reflectance signatures. The wavelength range lies between 350 to 2500 nm for both instruments. The samples were taken on a clear sunny day between 11 am to 1 pm and at a height of 50 cm above the plant. Fig. 1 shows the study area site in an agriculture university and the example visual of spectral samples collection.

Fig. 1. Study site and data collection by the team at the university of agriculture, Vellayani

The number of reflectance measurements and the types of crops considered are listed in Table 1. There are 536 samples of all categories included in this research work.

TABLE I: CROP TYPES AND NUMBER OF SAMPLES CONSIDERED FOR SPECTRAL MEASUREMENTS

SVC-ASD samples combination		
	Classes	**Number of samples**
1	Bitter gourd fruit	20
2	Bitter gourd leaf	30
3	Bottle gourd	20
4	Brinjal	29
5	Cauliflower vegetable	19
6	Cauliflower leaf	31
7	Cluster bean	19
8	Cucumber	10
9	Ladyfinger	20
10	Organic Brinjal	69
11	Organic Banana	5
12	Organic Cabbage	25
13	Organic Ginger	5
14	Organic Green chili	57
15	Organic Red spinach	5
16	Paddy medium	13
17	Paddy grown	10
18	Ridge gourd fruit	20
19	Ridge gourd leaf	21
20	Red spinach	20
21	Snake gourd	58
22	Tomato	20
23	Winged beans	10
		536

III. METHODOLOGY

The adopted framework to conduct this experiment is shown in Fig. 2. The measured spectral measurements are converted from radiance to reflectance using the reference radiance measurements acquired over the reference panel. As the spectral measurements were acquired using two different instruments, to generate a common baseline spectral data, we applied spectral normalization and wavelength resampling to a uniform spectral interval of 5nm. The spectral reflectance spectra processed for analysis contain 369 spectral channels. Data cleaning and outlier removal were also performed based on visual examination in the two-dimensional feature space. As the number of spectral samples for the different crops considered is less than the dimensionality of the spectra, we reduced the dimensionality of the reflectance spectra using principal component analysis (PCA) algorithm. Amongst the 23 crop types considered, two crops, brinjal and red spinach were cultivated under organic and chemical inputs-based crop growing practices. We, therefore, assessed the spectral discrimination of organic vegetable crops for these two crops only. In contrast with the approach of statistical separability, most of the past studies employed, for quantifying the spectral separability, we evaluated the crops discrimination from a functional perspective as a classification task wherein the quality of spectral discrimination is directly ascribed to the classification accuracy. Further, the method used for the classification substantially influences the quality of discrimination. Therefore, we have applied twelve different machine learning (ML) algorithms: Logistic regression (LR), Ridge classifier, SGD classifier, K-nearest neighbours (KNN) classifier, Decision Tree classifier, Linear SVM, SVM with RBF kernel, Gaussian Naive Bayes (NB), AdaBoost classifier, Random Forest (RF) classifier, Gradient boosting (GB) classifier, and Quadratic discriminant analysis (QDA) for classification. A description of these classifiers is out of scope for this paper. We suggest the readers refer to [10-11] for description of these machine learning algorithms.

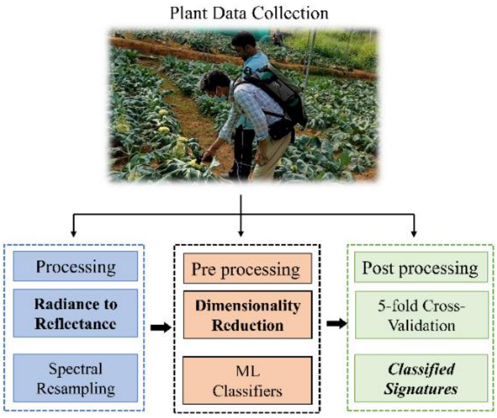

Fig. 2. Methodology adopted for spectral discrimination of vegetable crops

B. Validation of the Results

Various performance matrices - accuracy, precision, recall, f-1 score, and mean squared error (MSE), expressed in the following equations (1-6), are used to validate the performance of classification results.

$$Accuracy\ (Acc) = \frac{\text{Total correct predictions}}{\text{Total number of subjects}}$$

$$= \frac{(\text{TP} + \text{TN})}{\text{TP+TN+FN+FP}} \quad (1)$$

$$Average\ Accuracy = \frac{\sum_{n=1}^{10} Acc_n}{10} \quad (2)$$

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})} \quad (3)$$

$$\text{Recall} = \frac{(\text{TP})}{(\text{TP} + \text{FN})} \quad (4)$$

$$\text{F1 Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (5)$$

$$MSE(y, \hat{y}) = \frac{1}{s} \sum_{i=0}^{s-1} (y - \hat{y})^2 \quad (6)$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

IV. RESULT AND ANALYSIS

A. Analysis of all plant spectral reflectances

The results of the five-fold cross-validation of various ML algorithms are shown in Table 2. Other parameters like precision, F1 score, recall, and MSE are also relatively high, as seen in Table 2. The classification results indicate complex patterns of spectral discriminations. The over performance of crop discrimination, as quantified by different accuracy metrics, varies from 34% to 90% indicating the delicate nature of spectral differences amongst different crop types and the limitations of several classification algorithms to detect the finer spectral differences. Compared to the overall accuracy, the per-crop accuracy estimates as presented in the Table 4 for the best case of classification algorithm present the widest range of the variation of spectral discrimination as indicated by the accuracy metrics. While many crops are classified with accuracy of 90% and above, there are a few crops, e.g., bitter gourd exhibiting lowest spectral discriminability (20%). However, the reflectance measures for bitter gourd correspond to the fruiting level measurements while most of the other crops are green leaf measurements. Results show that Random Forest (RF) classifier performed best on the given dataset with the highest accuracy of 89.5%.

TABLE II: SUMMARY OF THE ACCURACY ASSESSMENT OF THE CROP DISCRIMINATION USING DIFFERENT ML MODELS

Classifiers	Accuracy (%)	Precision (%)	F1 Score (%)	Recall (%)	MSE
LR	0.51	0.37	0.36	0.38	51.40
Ridge Classifier	0.60	0.41	0.40	0.44	46.19
SGD Classifier	0.51	0.43	0.41	0.44	56.54
KNN	0.63	0.54	0.52	0.53	37.10
Decision Tree	0.65	0.62	0.60	0.62	37.20
Linear SVC	0.69	0.56	0.54	0.57	34.32
SVC_RBF	0.76	0.71	0.68	0.68	27.96
Gaussian NB	0.80	0.77	0.75	0.75	22.23
Ada Boost	0.24	0.11	0.12	0.15	93.91
RF Classifier	0.90	0.85	0.83	0.83	9.98
GB Classifier	0.79	0.74	0.70	0.70	21.58
QDA	0.34	0.05	0.07	0.13	48.57

In Table 3, we present per-class accuracies of the best-performed classifier, RF classifier. We can see that organic ginger, ridge gourd fruit, cluster beans, cauliflower leaf, and Lady's finger have the highest 100% accuracy, highlighted in green colour. However, some plants, like bitter gourd fruit and cauliflower vegetables, are poorly distinguished.

TABLE III: CROP-WISE CLASSIFICATION ACCURACIES FOR THE BEST CASE OF CLASSIFIER (RF ALGORITHM)

Plants Name	Abbreviation	Average accuracy crop-wise
O_Green_chilli	O_GC	0.90
O_Red_spinach	O_RS	0.95
O_Brinjal	O_B	0.73
O_Cabbage	O_Cab	0.73
O_Ginger	O_G	1.00
O_Banana	O_Ban	0.86
Ridge_gourd_leaf	RGL	0.87
Snake_gourd	SG	0.97
Ridge_gourd_fruit	RGF	1.00
Red_spinach	RS	0.90
Paddy_meduim	PM	0.85
Bitter_gourd_fruit	BGF	0.20
Clusterbean	Cbean	1.00
Brinjal	Brinjal	0.84
Bitter_gourd_leaf	BGL	0.40
Cucumber	Cucum	0.98
Winged_beans	WB	0.80
Cauliflower_vegetable	CV	0.50
Tomato	Tomato	0.93
Cauliflower_leaf	CL	1.00
Ladyfinger	LadyF	1.00
Paddy_grown	PG	0.90
Bottle_gourd	BG	0.70
Overall average accuracy		**0.84**

B. Discrimination of Organic vs. Non-organic vegetable crops

Two crops, brinjal and red spinach, have both organic and non-organic varieties in the plant species spectral dataset. The discrimination analysis and comparison of the spectral variability of these two types of vegetable crops are separately performed. Results indicate best spectral discrimination as quantified by the classification accuracy from using the Support vector machine with RBF kernel as the classifier.

The average five-fold confusion matrix by SVM RBF classifier of these four crops, e.g., brinjal, organic brinjal (O_B), red spinach (RS), and organic red spinach (O_RS), is shown in Fig. 3. In Table 4, there are individual accuracies of each plant shown. From Table 4, we can observe that organic brinjal and red spinach have higher accuracies than non-organic. It is noticeable that non-organic plants also performed well, and there is relatively less difference in performance compared to organic plants. The results suggest the possibility of spectrally discriminating organic vs. non-organic vegetable crops using hyperspectral data. However, the conformity analysis of the distinctness of this discrimination requires experimentation across sites and multiple crops.

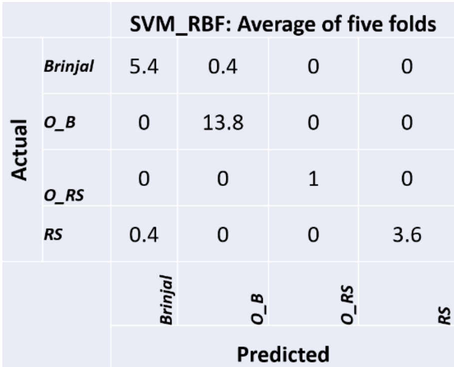

Fig. 3. Average of five-fold confusion matrices of organic vs non-organic crops

TABLE IV: CROP-WISE ACCURACY OF DISCRIMINATING ORGANIC VS. NON-ORGANIC VEGETABLE CROPS

Plants Name	**Abbreviation**	**Average accuracy plant-wise**
Brinjal	Brinjal	0.93
O_Brinjal	O_B	1.00
O_Red_spinach	O_RS	1.00
Red_spinach	RS	0.90
	Overall average accuracy	**0.96**

V. CONCLUSION

In this research work, we have assessed the potential of spectral discrimination of as many as 23 vegetable crops using reflectance spectra acquired from in situ measurements. Further, the experimental evaluation of the possibility of discriminating two different vegetable crops grown under organic and chemical inputs-based crop growing practices has been attempted for the first time. When there are several crops, simultaneous discrimination seems a marred with spectral confusion even with using hyperspectral data. The discrimination is largely determined the type of classification algorithm used, with the Random Forest classifier performing best on the plant spectral reflectance dataset when all the plant samples were considered. We also compared organic vs. non-organic plant discrimination capabilities and find that organic plant spectra are more discriminative than non-organic plant spectra suggesting the promise of using hyperspectral data for mapping the spatial distribution of organic crops cultivation.

ACKNOWLEDGMENT

The spectral measurements were collected as a group activity by the students of Master of Technology in Geoinformatics of the Indian Institute of Space Science and Technology. We thank Manohar Kumar, Prankur Sharma, Pyare Lal, Chetan Bajiya, Palaparthi Anindya, Jayati Vijaywargiya, Gopika Das K, Aswathy S S, and Aseem Garg, for participating in the field data acquisitions.

REFERENCES

[1] Rao, N. Rama, Pradeep Kumar Garg, and Sanjay Kumar Ghosh. “Development of an agricultural crops spectral library and classification of crops at cultivar level using hyperspectral data.” Precision Agriculture 8.4 (2007): 173-185.

[2] Som-ard, Jaturong, Clement Atzberger, Emma Izquierdo-Verdiguier, Francesco Vuolo, and Markus Immitzer. "Remote sensing applications in sugarcane cultivation: A review." Remote sensing 13, no. 20 (2021): 4040.

[3] Nihar, Ashmitha, N. R. Patel, and Abhishek Danodia. “Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India.” Journal of the Indian Society of Remote Sensing (2022): 1-12.

[4] MMbengue, F., Faye, G., Talla, K., Adama Sarr, M., Ferrari, A., Mbaye, M., Semina Dramé, M. and Sagne, P. 2022. Evaluation Of Machine Learning Classification Methods For Rice Detection Using Earth Observation Data: Case Of Senegal. European Scientific Journal, ESJ. 18, 17 (May 2022), 214. DOI:https://doi.org/10.19044/esj.2022.v18n17p214.

[5] Kumar, Mukesh, Ayan Das, K. N. Chaudhari, Sujay Dutta, K. K. Dakhore, and B. K. Bhattacharya. "Field-scale assessment of sugarcane for mill-level production forecasting using Indian satellite data." Journal of the Indian Society of Remote Sensing 50, no. 2 (2022): 313-329.

[6] Rinaldi, M., A. Castrignanò, D. De Benedetto, D. Sollitto, S. Ruggieri, P. Garofalo, F. Santoro, B. Figorito, S. Gualano, and R. Tamborrino. "Discrimination of tomato plants under different irrigation regimes: analysis of hyperspectral sensor data." Environmetrics 26, no. 2 (2015): 77-88.

[7] Mothapo, Mologadi Clodean, Timothy Dube, Elfatih Abdel-Rahman, and Mbulisi Sibanda. "Progress in the use of geospatial and remote sensing technologies in the assessment and monitoring of tomato crop diseases." Geocarto International 37, no. 16 (2022): 4784-4804.

[8] F. M. Lacar, M. M. Lewis and I. T. Grierson, “Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia.” IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), 2001, pp. 2875-2877 vol.6, doi: 10.1109/IGARSS.2001.978191.

[9] Eddy, P. R., A. M. Smith, B. D. Hill, D. R. Peddle, C. A. Coburn, and R. E. Blackshaw. "Weed and crop discrimination using hyperspectral image data and reduced bandsets." Canadian Journal of Remote Sensing 39, no. 6 (2014): 481-490.

[10] Cravero, Ania, Sebastian Pardo, Samuel Sepúlveda, and Lilia Muñoz. "Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review." Agronomy 12, no. 3 (2022): 748.

[11] Abioye, Emmanuel Abiodun, Oliver Hensel, Travis J. Esau, Olakunle Elijah, Mohamad Shukri Zainal Abidin, Ajibade Sylvester Ayobami, Omosun Yerima, and Abozar Nasirahmadi. 2022. "Precision Irrigation Management Using Machine Learning and Digital Farming Solutions" AgriEngineering 4, no. 1: 70-103. https://doi.org/10.3390/agriengineering4010006