International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 410-416, 2017, ISSN:-2319–8656

Prediction of Phosphorus Content in Different Plants: Comparison of PLSR and SVMR Methods

Sushma D Guthe
Dept. Of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, India

Dr.Ratnadeep R Deshmukh
Dept. Of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, India

Abstract: Phosphorus is one of the important biochemical components of plant organic matter and it helps to maintain the health of the plants. This study is conducted in some part of Aurangabad region, aimed to compare the PLSR and SVMR methods for predicting the phosphorus (Cp) content available in leaves in different plants using spectroscopy in Vis-NIIR reflectance spectroscopy. A total 38 leaf samples taken from jawar and maize plants were collected. In the plant 35% to 40% phosphorus content were found.

Keywords: Spectroscopy, phosphorus content (Cp), leaf, Vis-NIR, fieldspec4

1. INTRODUCTION

Plant organic matter such as nitrogen phosphorus and potassium are important biochemical components for metabolic processing. Therefore it is important to estimate the biochemical components. Earlier the analysis can be done using chemical method gives the accurate result but the chemical method is very much time consuming and complex.
The use of imaging spectroscopic analysis methods to estimate the nutrient status of maturing crops may save time, and scale back the value related to sampling and analysis. Imaging spectrographic analysis is a technology that result in getting data in narrow (<10 nm) and contiguous spectral bands. These narrow spectral bands allow the detection of some spectral features that masked within the broader bands of the multispectral scanner. [1]
The correct determination of phosphorus content in soils and plants is extremely important for agricultural science and practice. Phosphorus participates in a number of processes determining the growth, development and the productivity of the plant: formation of cell nucleus and cell multiplication, synthesis of lipids and specific proteins, transmission of hereditary properties, breathing and photosynthesis, energy transmission from richer to poorer energetic compounds, etc. It is very important to know the phosphorus status of soils to determine the necessity for phosphorus fertilizer use. Phosphorus is absorbed by the plants from soil in the form of phosphate ions. Phosphorus is a constituent of cell membranes, certain proteins, all nucleic acids and nucleotides, and is required for all phosphorylation reactions.[2] Phosphorus is known to be unique for its sensitivity and stability as a human activity indicator. Its content in soil represents a great interest to archaeologists, giving them information on type and intensity of human activity. Total phosphorus measurement gives quantitative results in contrast to mobile phosphorus and represents the best indicator for variation caused by human activity. [3]
L.K. Christensen used partial least square regression. PLSR was used on continuous spectra in the range 400-750 nm and the total N and P contents were established through chemical analyses and used as references. He was predicted P content with 74% accuracy based on the canopy spectral reflectance. He investigates the spectral reflectance in the visual range as a potential indicator for estimating nitrogen and phosphorus content in spring barley at three early growth stages. [4]

To develop calibration equations, multiple linear regression, and partial least-squares regression (PLSR) were used. This study compares between MLR and PLSR in the spectral range 1100-2500 nm. [5] Agustin Pimstein used vegetation indices and PLSR. In that study he uses specific narrow band vegetation indices instead of tradition broad band indices. It was observed that a significant improvement is obtained when the mineral total content is considered instead of the relative content. Therefore it was suggested that the biomass should also be retrieved from the spectral data. [6]
Yanfang Zhai study was based on eight different plants. He used the PLSR and SVMR. For implementation of PLSR used the parles 3.0 software and for SVMR uses the LIBSVM library. For some crop, he used the PLSR and for other SVMR. He concluded that the SVMR method combined with Vis-NIR reflectance has the potential to estimate the contents of biochemical components of different plants. [7]
Liu Yanli studied that upper side or lower side of leaves give the better spectral signature. The linear of PLS model and nonlinear of LS-SVM model fit better with spectral data of the

upper side and lower side of leaves, respectively. He used 400-1000nm wavelength sensitive to phosphorus content. [8] Stepwise linear regression was used to select wavelengths to investigate relationships between laboratory analysis results and spectral data.400-900nm wavelength sensitive to phosphorus content. [9] The results of the experiment demonstrated that radiometric measurements can be used for monitoring of N, P, S and K status in a wheat crop. Correlation analysis of nutrient status with leaf and canopy reflectance showed presence of responsive wavelengths to variable N, P, S and K status in wheat. [10]

Osborne et al. used linear regression models that included reflectance at 730 and 930 nm for predicting P concentration in corn. [11]

Name	Algorithm	Elements	Spectral range	Plant
L.K. Chirstensen	PLS	Nitrogen and phosphorus	400-750nm	Spring Barley
C. petisco	MLR and PLSR	nitrogen, phosphorus and calcium	1100 to 2,500 nm	Woody plant species
Agustin Pimstein	Vegetation indices and PLSR	potassium and phosphorus	1400–1500nm and 1900–2100nm	Wheat
Yanfang Zhai	PLSR and SVMR	Nitrogen, phosphorus and potassium	350–730 nm and 1420–1800 nm	rice, corn, sesame, soybean, tea, grass, shrub, and arbour
Liu Yanli	PLS and LS-SVM	Nitrogen and phosphorus	400-1000nm	Citrus leaves

Table 1.1: different method that estimate biochemical content

2. MATERIALS AND METHODOLOGY

2.1 Study Area

The study area is Aurangabad city. Samples of leaf mainly collected from three areas Devlai, Waluj (19.850670, 75.263130) and BAMU University with jawar, maize crop. Aurangabad city is located within the Maharashtra.

2.2 Spectral Measurements and Pre-Processing of Leaf Samples

A total 38 samples were collected between 24 to 30 march 2017, from two types of plants including jawar and maize between 1pm to 5pm. The sampling sites were randomly selected based on the land areas of these plants. The leaves without leaf stalk on upper layer of canopy were cut and kept fresh in plastic bag for less than 8 hours before spectral reflectance were measured. The spectral reflectance of leaf samples measured using a Fieldspec4 spectroradiometer. [12] The spectroradiometer has a spectral range from 350 to 2500nm. It's sampling interval 3nm (from 350 to 1000nm) to 10nm (from 1000 to 2500nm). For illumination purpose a 50 W quartz halogen lamp was set over, the leaf samples at a distance of 30cm and a distance between lamp and gun is 37 cm.

The samples were put evenly on black sheet paper. The white reflectance panel was used to optimize the signal and calibrate the accuracy and detector response. Each sample was scanned 10 times within 180° rotation. Taking the mean of these 10 readings for further processing.to remove the noise values below 400nm and above 2400nm were rejected.

Before predicting the models several preprocessing transformation techniques such as first derivative of original reflectance and reflectance transformation. Data visualization and preprocessing can be done using ViewSpec pro software (ASD Inc). [13] The prediction accuracy of the model for the calibration and validation datasets was been evaluated through parameters such as R2 (Coefficient of Determination) is the correlation between predictable and observable variables, RMSE (Root Mean Square Error) provides direct assessment of modelling error expressed with original measurement unit and RPD(Residual Prediction Deviation).[14]

$$R^2 = \sum_{i=1}^{N} \frac{(\hat{y} - \bar{y})^2}{(y - \bar{y})^2}$$

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\hat{y} - y)^2}{N}}$$

$$RPD = \frac{\sigma_{val}}{RMSEv}\sqrt{n/n\text{-}1}$$

Where
$\hat{y}$is the predicted value,
y is the observed value,

$\bar{y}$ is the mean of observed values,
N is the number of sample data.

2.3 Methodology

In this study partial least square regression and support vector machine regression these two methods used. These methods are as follows:

2.3.1 Plsr

Partial least squares (PLS) is a method for constructing predictive models when the factors are many and highly collinear. Note that the emphasis is on predicting the responses and not necessarily on trying to understand the underlying relationship between the variables. [15]The PLSR method is used to find the hyperplanes of maximum variance between predictable and observable variables and develops a linear model by projecting predictable and observable variables to a new space. [16][17]

$$Y = X\beta + \varepsilon s$$

where **Y** is the vector of predictable variables (biochemical component contents in this study), **X** is the matrix of observable variables, which is a linear combination of a few latent potential factors (spectral reflectance in this study), β is the matrix of regression coefficients, and ε is the error matrix of the relationship between **X** and **Y.**

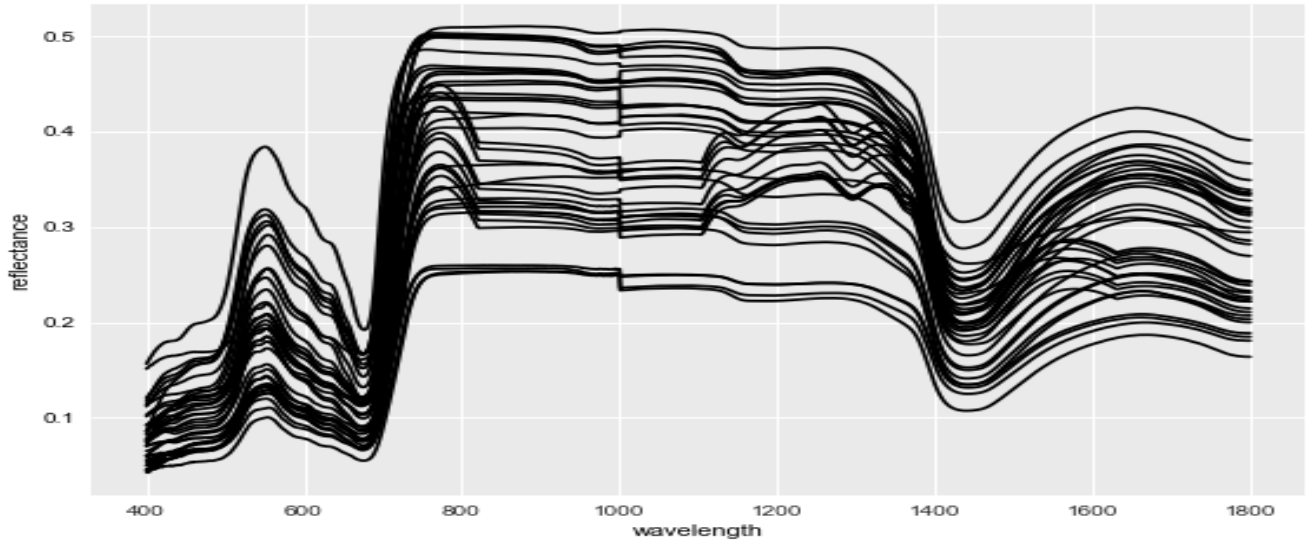
2.3.2 Svmr

The SVMR function can be expressed as [18]

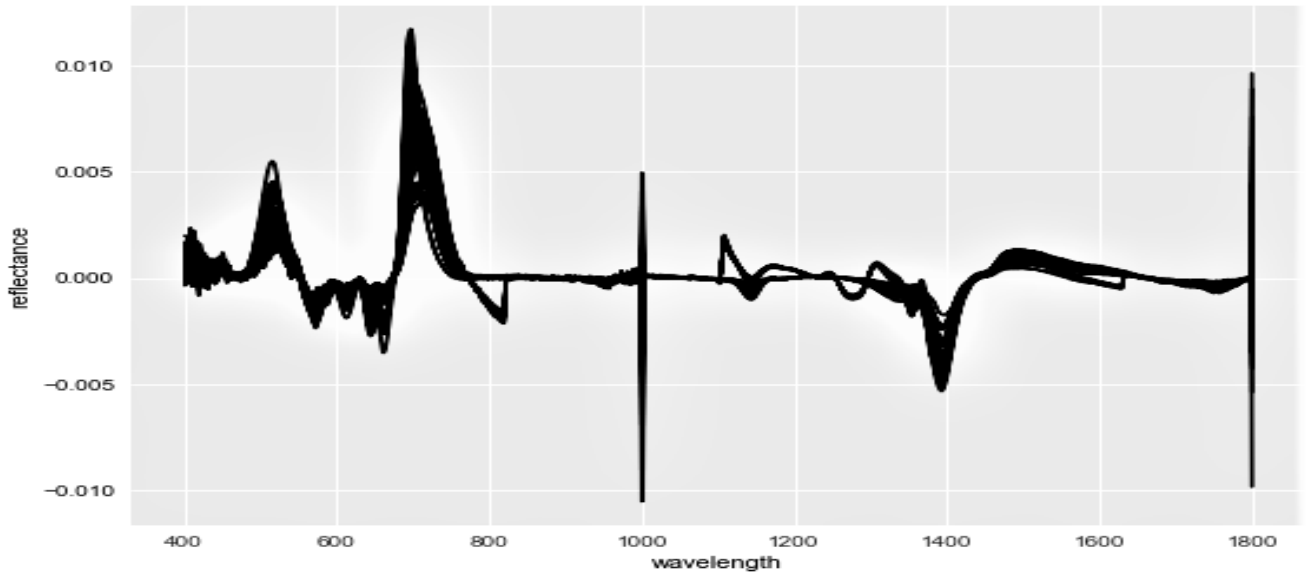
$$f(x) = \sum_{i=1}^{M} aiyiK(xi, x) - b\ 0 \leq ai \leq C$$

where **K**(**xi**, **x**) is the kernel function, **xi** is the input vector, **x** is an item used to create higher-dimensional feature space, **yi** is the corresponding output vector, **ai** is the Lagrange multiplier (called the support value), M is the number of samples, and **b** is the bias term. [19] The radial basis function (RBF) [20] was employed as the kernel function in this study:

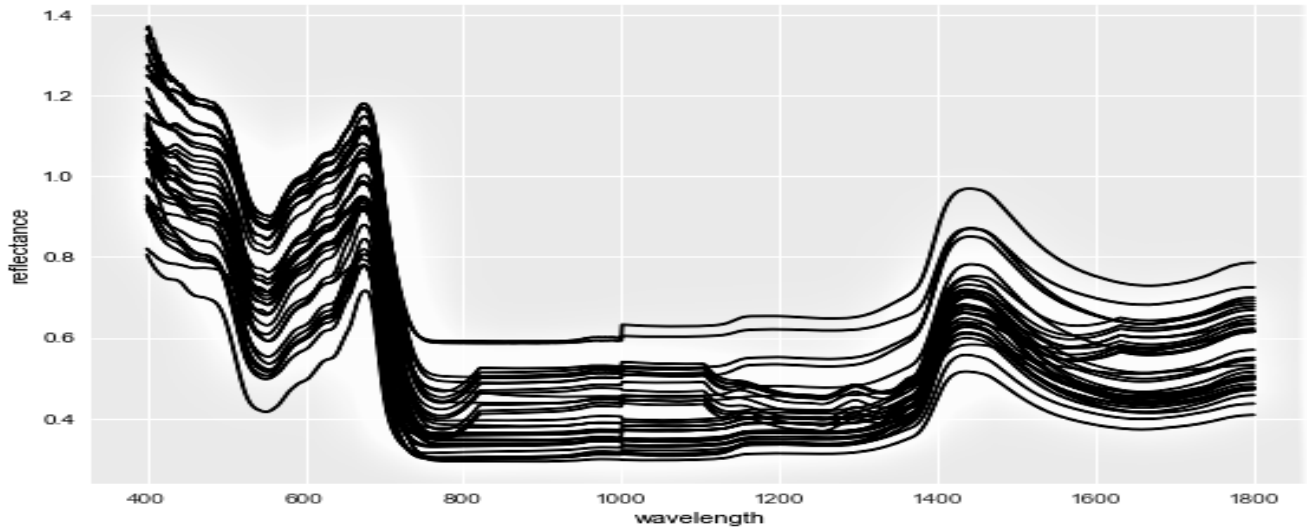
$$\boldsymbol{K(x_i, x) = e^{-y(x_i - x)^2}}$$



(a)



(b)



(c)

Figure 2.1: Average Spectral Reflectance curves (a) original spectral reflectance, (b) the first derivative of reflectance and (c) Reflectance transformation of the reflectance

2.4 Data Analysis

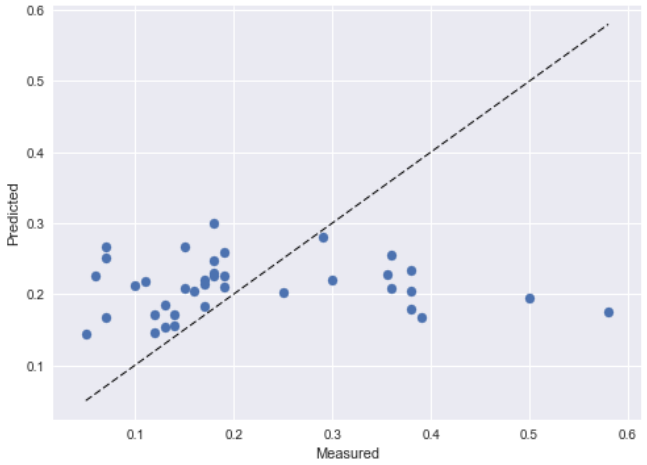
In this study, the PLSR and SVMR models for leaf sample was implemented within the Anaconda2 (32 bit) software used. [21] This software provides all packages that are required to implement model. In Anaconda2 Spyder 3.1.2 environment was used for python programming .this environment used the python 2.7.13 version. Wavelength range between 400-1000nm we sensitive to phosphorus content

3. RESULTS AND DISCUSSION

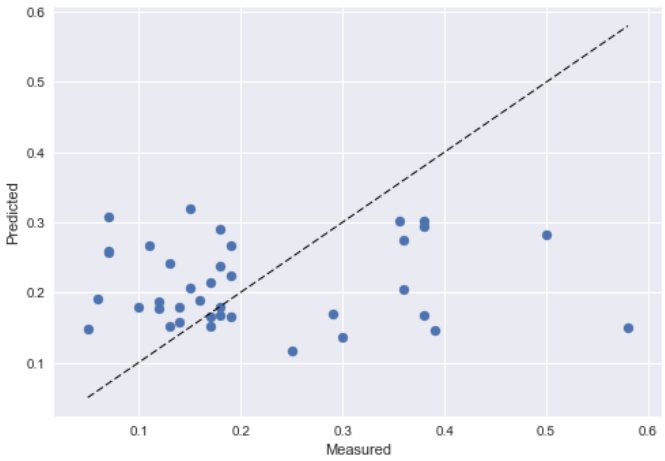
The average raw reflectance spectra of leaves of various plants had similar patterns as shown in figure 2.1. The descriptive statistics of Cp of the 38 leaf samples are shown in table 3.1. The maximum value is almost ten times the minimum value of Cp.

	Mean	SD	Median	Max	Min
Cp(%)	0.25	0.14	0.24	0.50	0.04

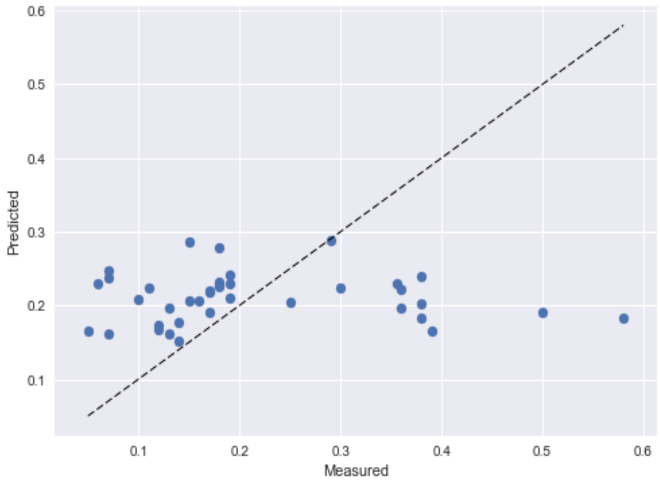
Table 3.1: Descriptive statistics of phosphorus (Cp) content in 38 leaf samples



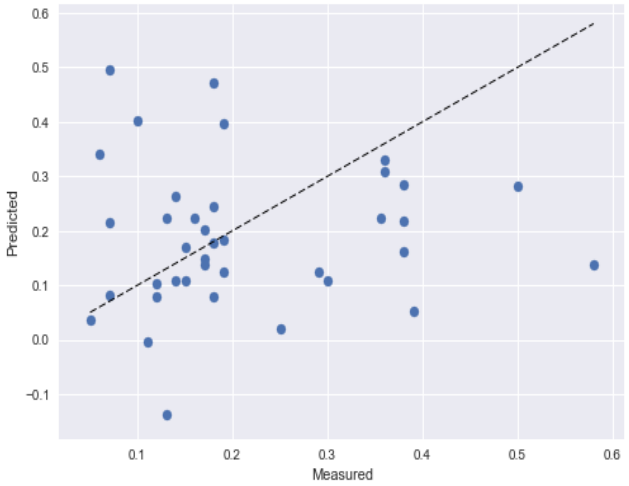
(a)



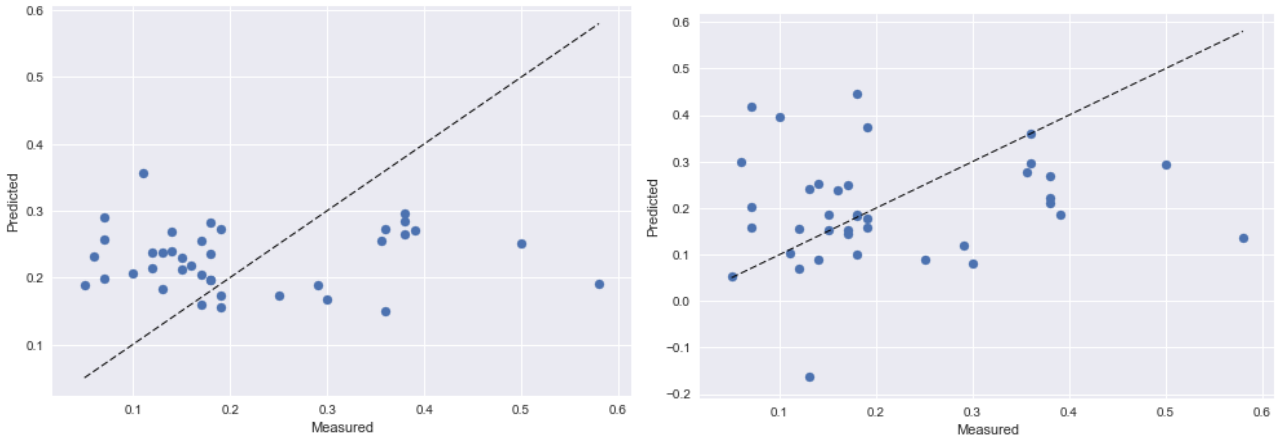
(b)



(c)



(d)



(e) (f)

Figure 3.1: correlations between measured and estimated values of phosphorus contents in Partial Least Square Regression (PLSR) (a) original reflectance (b)1st derivative (c)reflectance transformation and Support Vector Machine Regression (SVMR) (d) original reflectance (e) 1st derivative (f) reflectance transformation

Pre processing tech	RMSE		R^2	
	Cal	Val	Cal	Val
Original reflectance	0.0793	0.0920	0.0573	0.1072
First derivative	0.1331	0.0926	0.2050	0.0947
Log(1/R)	0.1456	0.0935	0.0477	0.0780

(a)

Pre processing Tech	RMSE		R^2	
	Cal	Val	Cal	Val
Original reflectance	0.1449	0.0776	0.7172	0.3649
first derivative	0.0889	0.08095	0.6450	0.3085
Log(1/R)	0.0818	0.07865	0.6995	0.3473

(b)

Table 3.2: the results of model calibration and validation (a) PLSR (b) SVMR

In the PLSR model, considering raw spectral reflectance it gives poor performance for this model. When we done the preprocessing it gives better results as shown in the table 3.2(a). Log(1/R) preprocessing method gives the poor performance (R^2_v =0.0780) while first derivative give better performance than Log(1/R). In the SVMR model, considering raw spectral reflectance it gives poor or moderate results when estimating Cp(R^2_v=0.365). Preprocessing method results as shown in table 3.2(b).

The SVMR model gives the better results when estimating Cp as compared to PLSR model.

4. CONCLUSION

In this study, we compared PLSR and SVMR models for estimating the phosphorus content of leaves of two different crops with laboratory based Vis-NIR reflectance data. We concluded that the SVMR method after applying preprocessing method estimates biochemical content of different plants.

ACKNOWLEDGEMENT

This work is supported by Dept. of Computer Science and Information Technology under the funds for Infrastructure under science and Technology (DST-FIST) with sanction no. SR/FST/ETI- 340/2013 to Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India. The authors would like to thank Department and University Authorities for providing the infrastructure and necessary support for carrying out the research.

REFERENCES

1. Sushma D Guthe, Ratnadeep R Deshmukh, May-June 2017, "Estimation of Phosphorus Content in Leaves of Plants using PLSR and SVMR: A Review", IJARCS, Volume 8, No. 4.
2. NCERT Solutions - Biology for Class 11th, MINERAL NUTRITION
3. Krasimir Ivanov, Penka Zaprjanova, Violina Angelova, Georgi Bekjarov and Lilko Dospatliev, August 2010, "ICP determination of phosphorous in soils and plants", © 2010 19th World Congress of Soil Science, Soil Solutions for a Changing World.
4. L.K. Christensen, B.S. Bennedsen, R.N. Jorgensen, H. Nielsen, May 2004, "Modelling Nitrogen and Phosphorus Content at Early Growth Stages in Spring Barley using Hyperspectral Line Scanning", Biosystems Engineering, Volume 88, Issue 1, Pages 19–24.
5. C. Petisco , B. Garcı´a-Criado , B. R. Vxa´ zquez de Aldana, 2005, "Use of near-infrared reflectance spectroscopy in predicting nitrogen, phosphorus and calcium contents in heterogeneous woody plant species", Springer-Verlag , Anal Bioanal Chem (2005) 382: 458–465.
6. Agustin Pimstein, ArnonKarnieli, Surinder K. Bansal, David J. Bonfil, February 2011, "Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy", Field Crops Research Volume 121, Issue 1, Pages 125–135
7. Yanfang Zhai , Lijuan Cui , Xin Zhou , Yin Gao , TengFei and WenxiuGao, April 2013, "Estimation of nitrogen, phosphorus, and potassium in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods", International Journal of Remote SensingVol. 34, Issue 7, 2502–2518.
8. Liu Yanli, LyuQiang, He Shaolan, Yi Shilai, Liu Xuefeng, XieRangjin, ZhengYongqiang, Deng Lie , April 2015, "Prediction of nitrogen and phosphorus contents in citrus leaves based on hyperspectral imaging", Int J Agric&BiolEng, Vol. 8,Issue 2 80－88.
9. Y. Özyigit, and M. Bilgen, *2013*, "Use of Spectral Reflectance Values for Determining Nitrogen, Phosphorus, and Potassium Contents of Rangeland Plants", J. Agr. Sci. Tech. Vol. 15: 1537-1545.
10. G. R. Mahajan, R. N. Sahoo, R. N. Pandey, V. K. Gupta, Dinesh Kumar, 2014, "Using hyperspectral remote sensing techniquesto monitor nitrogen, phosphorus, sulphur and potassium in wheat (Triticum aestivum L.)", Precision Agric , 15:499–522.
11. Osborne, Schepers, Francis, & Schlemmer, 2002, "Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements", Agronomy Journal, 94, 1215–1221.
12. Fieldspec4 user manual, ASD Document 600979, Rev. A, December 2011.
13. ViewSpec Pro™ User Manual, ASD Document 600555 Rev. A © 2008 by ASD Inc.
14. Vasques, G. M., S. Grunwald, and W. G. Harris, 2010, "Spectroscopic Models of Soil Organic Carbon in Florida, USA", *Journal of Environmental Quality* 39: 923–34.
15. Randall D. Tobias, 1995, "An Introduction to Partial Least Squares Regression", SAS Institute Inc., Cary, NC.
16. Efron, B., and G. Gong, 1983, "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation", The American Statistician 37: 36–48.
17. Lecture, Modeling Input/output Data: Partial Least Squares (PLS), cited on www.sbcny.org/pdfs.
18. Zhang, X. J., and M. Z. Li, 2008, "Analysis and Estimation of the Phosphorus Content in Cucumber Leaf in Greenhouse by Spectroscopy." *Spectroscopy and Spectral Analysis* 28: 2404–8.
19. Zhang, X. G (2000), "Introduction to Statistical Learning Theory and Support Vector Machines", Acta Automatica Sinica 26: 32–42.
20. Burges, C. J. C (1998), "A Tutorial on Support Vector Machines for Pattern Recognition." *Data Mining and Knowledge Discovery* 2: 121–67.
21. "Anaconda software Download 32-bit" cited on https://www.continuum.io/download