RESEARCH ARTICLE OPEN ACCESS

Correlation Of NDVI and CRI2 for Crop Health Assessment

Sulochana Shejul [1], Vijay Dhangar [2], Pravin Dhole [3], Bharti Gawali [4]

Department of CS and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad - India.

ABSTRACT

The objective of the current study was to assess the feasibility of using hyperspectral data to measure chlorophyll and carotenoid levels in crops and assessment of health, using Vegetation Indices in addition to machine learning methods. We believe that combining various existing vegetation indices could lead to better results for the health assessment of crops than relying just on a single index. In examining this concept, two different cash crops namely Cotton and Maize was selected from Aurangabad region. Healthy and unhealthy leaves of each was selected for data collection and then created spectral signature of each using the ASDFieldSpec4 Spectroradiometer. Then pre-processing has been done by applying 2nd derivative smoothening of signature. Important feature bands has been extracted using bad band removal process. NDVI and CRI2 vegetation indices was calculated and using the correlation of these two indices, we benchmark the boundary for the health of the selected crops. And finally the machine learning algorithms has been applied to the vegetation indices. Logistic Regression gives the accuracy of 96.4% and SVM gives the 93.3% accuracy for Maize leaves and for Cotton, Logistic Regression and SVM gives 96.7%and 93.7% respectively.

Keywords: Hyperspectral data, Spectral Signature, NDVI, CRI2, Logistic Regression, SVM.

I. INTRODUCTION

The substances known as pigments give the things around us colour. Natural pigments known as photosynthetic pigments aid in the process of photosynthesis. These chemicals, known as pigments, effectively absorb visible light. The wavelengths that are either reflected or transmitted by the plant tissue are changed as a result of their interactions with sunlight [6]. Most widespread pigments in photosynthetic plants are chlorophyll and Carotenoids. Carotenoid absorb light energy and convert it to chemical energy for use in photosynthesis process [7]. If

photosynthesis happens properly, then the plant will be healthy. And healthy plant shows the good amount of chlorophyll. We can say for the healthy plant these two pigments should be appropriate range. To calculate these pigments, there are various vegetation indices are available in remote sensing like NDVI and CRI2. But monitoring vegetation still faces a lot of difficulties when trying to pinpoint a specific illness or stress using remote sensing methods.

There are various research papers has been observed for the health analysis of vegetation by using NDVI index [23] [24] . Very less amount of the papers has been observed which has used NDVI, CRI2, or more combinations of the spectral vegetation indices for health analysis. In this research study, we have built The relation between the NDVI and CRI2 vegetation indices for the health analysis of Cotton and Maize crops by using MATLAB simulation tool.

It is necessary to examine qualitative and quantitative variations between healthy and diseased plants' spectral reflectance. Consequently, a spectral measurement of a separate waveband or a mix of wavebands can be applied to optically differentiating between healthy and diseased plants. Pigments absorb the majority of visible light. The leaf morphology causes numerous scattering inside the leaf and is correlated with the percentage of air gaps, determines the reflectance of wavelengths in the NIR region [13]. The basis for spectral research and remote sensing is being able to quantify electromagnetic energy at different wavelengths when it works with a substance. Each substance has its own unique manner that its physical properties lead the potential for electromagnetic radiation to be reflected, distorted, or absorbed. The spectral signature of a material is a distinctive shape that results from the measurement of these interactions across discrete portions of the spectrum.

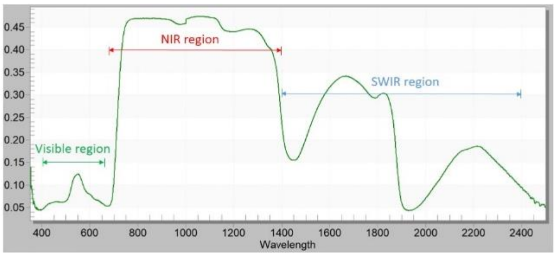

Figure 1. Vegetation's spectral signature.

Compared to other natural materials, vegetation has a different interaction with solar radiation. The plant spectrum (figure 1) often reflects in the green and near infrared (NIR) wavelengths, absorbs in the red and blue, and has prominent absorption patterns at the range of wavelengths where there is atmospheric water. Further variations in the spectrum result from many plant components, including their water content, pigments, carbon, nitrogen, and other properties. It is possible to learn useful information on the environment's stressors, water quality, and other vital properties of plants by measuring these fluctuations and examining how they interact with one another. These connections are frequently referred to as vegetation indices. [11].

The pathogen's physiological modifications to plant metabolism frequently cause disease symptoms. However, the effect of plant diseases on a plant's physiology and phenology differs according to the relationship in between a host and a pathogen and may result in changes to the plant's pigments, contained water, and tissue's ability to operate or the presentation of bacterial-specific structures. In reality, each of these influences could change the plant's spectral pattern. The Hyperspectral differentiation of diseased and healthy leaves and elements of canopy benefits from knowledge of the physiological effects of illnesses on the metabolism and structures of plants. [12].

1.1 Study area and Database:

The targeted study area is located in the district of Aurangabad, Maharashtra (India) and is a significant corn producer, and Maize cash crops [3] [4]. The central coordinate of the study area is 19056 '12.76" N, latitude and 75022'.8.48" E longitude. The Hyperspectral non–imaging data is utilized in this investigation. The database is gathered through ASD Field Spec4 Spectroradiometer. Which is having the spectral range of 350-2500nm. In this wavelength range, it collects visible, near infrared, and short wave infrared spectrum data.

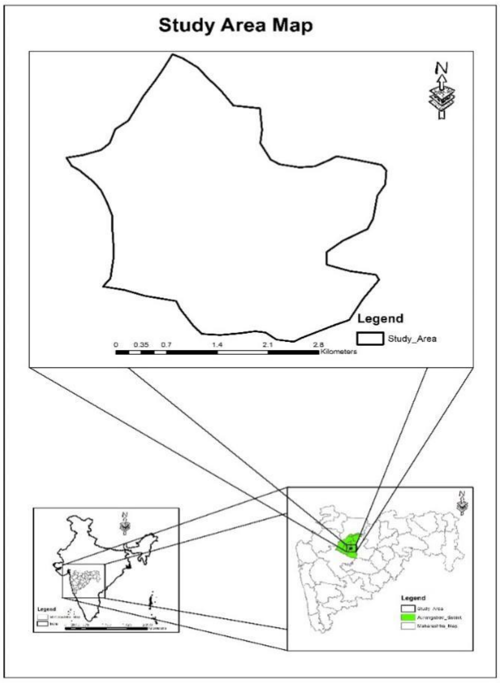

Fig. 2. Selected Study area

II. LITERATURE

Spectral imaging is developing as a tool as an independent platform for plant assessment. This is because it permits the determination of biochemical factors such as proteins, various minerals, and water as well as information on plant pigments like chlorophylls, anthocyanin, and carotenoids. When compared to a healthy plant, a sick plant will display different physicochemical properties, which the Spectral Imaging may measure as a function of light reflected or absorbed [1]. Hyperspectral technology is widely being used in precision agriculture for identifying plant diseases and protecting plants. The electromagnetic spectrum contains up to several hundred bands which can be measured by a hyperspectral sensor within its wavelength range, as opposed to RGB cameras, which can only detect three visible-band wavelengths (red, green, and blue colours). The Hyperspectral camera (sensor) has a high spectral resolution since Each of these bands in the spectrum only measures a small portion of the electromagnetic spectrum. By doing this, In a Hyperspectral photograph, each pixel is able to gather its own unique set of data regarding the reflection (or transmittance) for every spectrum band. Without additional spatial informationnon-imaging sensors using hyperspectral technology assess the aggregate of these data, known as a spectral signatures [2].

Izzuddin, M. A., et al. has studied that the CRI2 index is sensitive to the leaf carotenoid pigment which helps plant to absorb light energy for use in the photosynthesis process. Unhealthy vegetation usually contains lower carotenoid concentration. They also suggested that the cucumber leaves having lower chlorophyll and Carotenoid content after infection [21].

Tuominen et al. has assessed the ENVI The capacity of the Forest Health tool to identify seepage and dust-contaminated forest areas. For that they have used the combination of the NDVI, CRI2, and water band vegetation indices [22].

III. METHODOLOGY

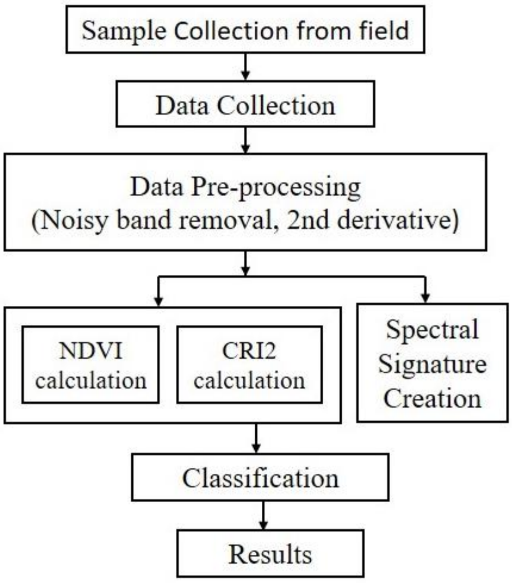

Fig 3. Proposed Methodology for research work

3.1 Data Collection

From the selected study area, 20 samples of each crop for both healthy and diseased has been selected. The selected crops were Cotton and Maize. Well-developed leaves were selected for the research and the crop age was 90 days. In 90 days the crops are so matured and able to create its fruit. So this phase is so important to pay attention to it. After cutting the leaves of crop from plants, they were stored into air tight plastic bag and kept in cool environment to maintain moisture of leaf. Within 2 hours we came to a dark room where the ASD Field Spec4 device is set up. This device is warmed up for 20 minutes. And then first off all white reference panel is taken for calibration. After showing the full reflectance i.e. 1, we started to collect spectral signatures of all the collected samples. 10 iterations were taken for each sample. To show and store the signatures, RS3 software is used after collecting spectral signatures, pre-processing has been done by using ViewSpecPro software.

iseased Cotton leaf ealthy Cotton Leaf

iseased Maize Leaf ealthy Maize Leaves

Fig. 4. Healthy and Diseased Sample leaves

3.2 Pre-processing techniques:

After creating the spectral signatures of samples, first off all we have calculated mean of every ten utterances. The bands from 350-400 and 2400-2500 have been dropped due to the noisy bands or having no useful information. And then we have calculated its 2nd derivative to smoothening of the curve. Important bands that are having the pigment information has been extracted. Using that extracted bands vegetation indices have been calculated.

3.3 Spectral Signature creation

Signatures for healthy and unhealthy cotton maize has been created as shown in the figure (Fig.4). In the visible region (400-700nm), the signature for the healthy vegetation always reflects more in green band (550) as it is having good amount of chlorophyll content and absorbs in red and blue bands (450 and 650). Whereas the diseased vegetation reflects less in the green band or reflects more in the red and blue bands as compare to healthy vegetation. There for, in the NIR region (700-1300nm) healthy vegetation reflects more at the range 700-860nm because of the leaf pigments

(Chlorophyll) [19][20]. And at the 1450nm and 1900nm there is absorption because of the moisture available in the leaf.

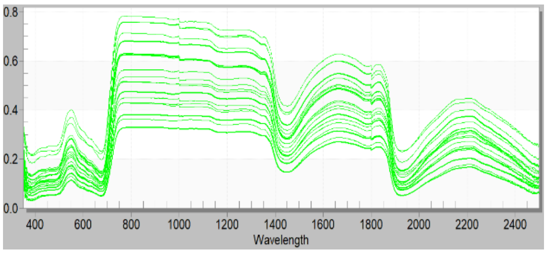

Fig.4a. Maize Healthy Signature

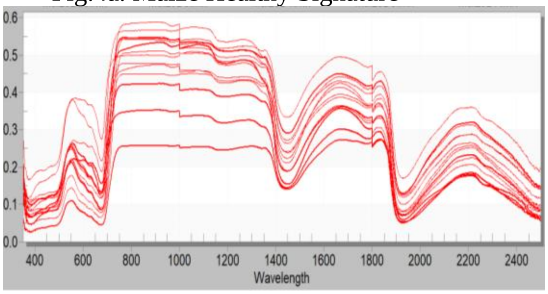

Fig.4b. Maize Unhealthy Signature

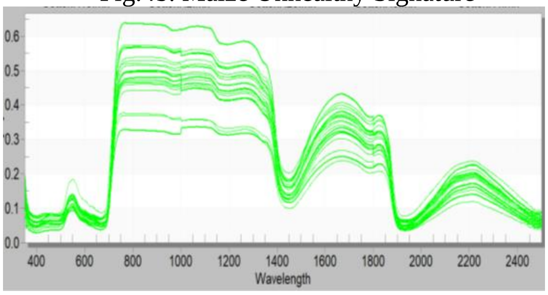

Fig.4c. Cotton Healthy Signature

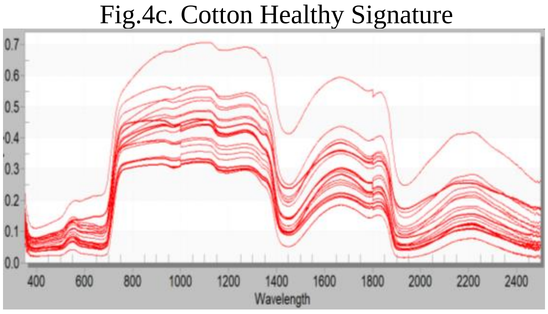

Fig.4d. Cotton Unhealthy Signature

3.4 Spectral Vegetation indices

according to the absorption and transmittance of each plant's vegetative spectrum's bands and every sample of data, it is feasible to apply indices for the identification of hyperspectral characteristics impacting vegetation health. Spectral vegetation indices helps to get more precise information. It is the combination of two or more bands. In this research, we have applied two pigment content indices i.e. NDVI and CRI2. These two indices are relative to the chlorophyll and carotenoid pigments respectively. Chlorophylls and Carotenoids are the main pigments of the green leaves. A leaf's morphology Plant physiological states throughout development, senescence, tolerance, and adaptation to various environments and pressures are frequently diagnosed using car content and its proportion to chlorophyll [18].

3.4.1 NDVI

The most used vegetation index for crop or any study on the health of plants is NDVI. It is a great predictor of the absorption of chlorophyll pigment of greenery. NDVI shows the difference between the reflectance at 750 nm, which is related to internal structure scattering in leaves, and the reflectance at 705 nm, which is related to chlorophyll absorption in leaves, normalised by dividing both reflectance's by their sum. This measure is responsive to minute variations in leaf cover and plant senescence [5]. The range of this index is from -1 to 1 but it varies for all the type of vegetation within range [10].

$$NDVI = (R800-R670)/(R800+R670) \quad (1)$$

3.4.2 CRI2

Cri2 is the index which is responsible for the carotenoid pigment in the leaves. Like chlorophyll, Carotenoid pigment in the leaf is important for photosynthesis process, in which it absorbs the energy. Common pigments called carotenoids are essential for photosynthesis. By absorbing in the blue-green region of the sun spectrum and transmitting the absorbed energy to chlorophylls, they expand the spectrum of light wavelengths that can fuel photosynthesis. Chlorophyll's influence on the 510 nm reciprocal reflectance is eliminated using the reciprocal reflectance at 700 nm. This index has values between 0 and greater than 15. Green vegetation often falls within the range of 1 to 11. CRI2 is calculated by the formula given below:

$$CRI2 = (1/R510)-(1/R700) \quad (2)$$

3.4.3 Vegetation indices results

After pre-processing has been done, we have calculated the NDVI and CRI2 vegetation indices for further processes. Whose results are stated in the table below (table 1). The NDVI shows the values for healthy cotton and maize with the mean of 0.80 and 0.75 respectively. And for Unhealthy Cotton and Maize, it shows 0.61 and 0.63 respectively. The CRI2 shows the values for healthy Cotton and Maize as 5.5 and 3.2 Unhealthy Cotton and Maize as 3.75 and 1.45 respectively.

Table 1: Results of vegetation indices

Cotton Healthy		Cotton Unhealthy		Maize Healthy		Maize Unhealthy	
NDVI	CRI2	NDVI	CRI2	NDVI	CRI2	NDVI	CRI2
0.82	5.30	0.58	3.07	0.77	4.41	0.65	1.49
0.83	6.04	0.63	3.89	0.76	4.14	0.65	1.81
0.81	6.38	0.62	3.75	0.75	3.10	0.60	1.12
0.80	6.64	0.65	3.79	0.73	3.03	0.60	1.31
0.77	8.06	0.58	3.97	0.74	4.19	0.66	1.61

The Cotton leaf which is having the NDVI and CRI2 values greater than 0.7 and 4.25 respectively is considered as the healthy. Whereas the Maize leaf having the NDVI and CRI2 Values greater than 0.7 and 2 respectively is also considered as healthy leaf [22].

IV. CLASSIFICATION TECHNIQUES

After calculating vegetation indices, we have applied some machine learning techniques on the results of vegetation indices to classify. For that we have used supervised machine learning algorithms Logistic Regression and Support Vector Machine (SVM). For that we have given training and testing with an 80:20 ratio. Two classes were created by the morphological study for classification of leaf as Healthy and unhealthy.

Logistic Regression (LR):

Logistic regression which is a supervised learning which is utilised to figure out the chances that a binary event will occur. LRA with a categorical outcome variable. Situations involving categorical outcomes are rather common in practise. Predictions for the binary outcomes of success/failure or improved/not-improved may be made, for instance, in the context of reviewing an educational programme. Similar to this, a medical result could be the existence or absence of a disease [8][14][15][16].

It is a statistical model that is frequently uses the logistic function to model a binary dependent variable. Another name Sigmoid function describes the logistic function, and it is presented as:

F(x) = 1/(1+e-x)=ex/(ex+1) (3)

With the help of this function, the logistic regression model able to condense the information from (-k, k) to (0, 1). Despite the fact that it supports multiclass categorization, the majority of the time, logistic regression is employed for binary classification tasks.

A sigmoid function, which in this equation is composed of log-odds, is used to compress the output of the linear equation to a probability between 0 and 1. Also, we can establish a decision boundary and carry out a classification task using this probability.

SVM:

SVM's accuracy is one of the primary justifications for use in health assessment. Additional benefits include its simplicity, directness, and memory effectiveness [].In a multidimensional setting, SVM essentially functions as a linear separator between two data points to identify two distinct classes. SVM fundamentally explains the relationship between features and repeated features. Using an n-dimensional space vector and the SVM, two vector sets are created from the dataset. The SVM algorithm essentially builds a hyper plane environment in to evaluate each piece against a distinct linear line. It is proposed to use the hyper-plane concept to carry out data separation based on greatest distance analysis to find the classes. Lowering the error ratio requires defining the classifier with the largest margin. [17].

A supervised machine learning technique, the Support Vector Machine is typically applied to classification tasks. SVM can be applied to multivariate classification as well as binary classification. Multimodal SVM is used in this study since there are more than two classes that need to be predicted. Several categorical and continuous data points can be handled by SVM. The objective of the SVM algorithm is to locate the largest marginal hyperplane in an N-dimensional space that clearly categorises data points in classification. Finally, in order to reduce error, Support Vectors are iteratively created on the ends of comparable featured data points. Marginal distance is the distance between two parallel support vectors. Support vectors are positioned so that the maximum marginal distance exists and the vectors are perpendicular to one another. After that, a hyperplane is formed between the support vectors to categorise the data points into several groups. SVM is a highly favoured classification technique since it offers notable accuracy while using minimal processing power.

Classification Results

The healthy and unhealthy leaves are classified based on calculated vegetation indices NDVI and CRI2. The obtained results are stated in the table 2. The logistic regression gives the accuracy of 96.4% and 96.7% for Maize and Cotton respectively. And the SVM gives the accuracy of 93.3% and 93.7% for Maize and Cotton respectively. The logistic regression is good for the health analysis.

Table 2: Classification Results.

Crop	Logistic regression	SVM
Maize	96.4%	93.3%
Cotton	96.7%	93.7%

The following figure shows the classification results plotted. In that NDVI index is plotted on the x axis and the CRI2 index is plotted on the y axis. The blue colour shows the healthy sample points and red colour shows the Unhealthy sample points.

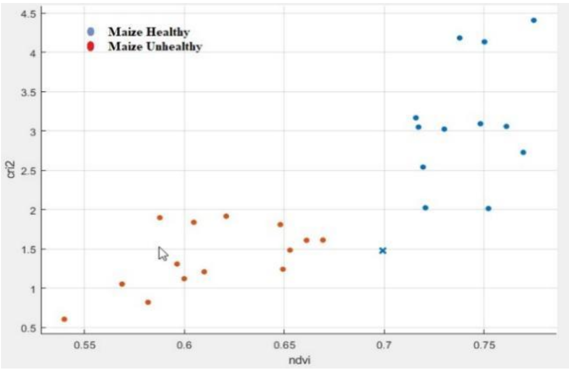

Fig.4a. Logistic Regression for Maize

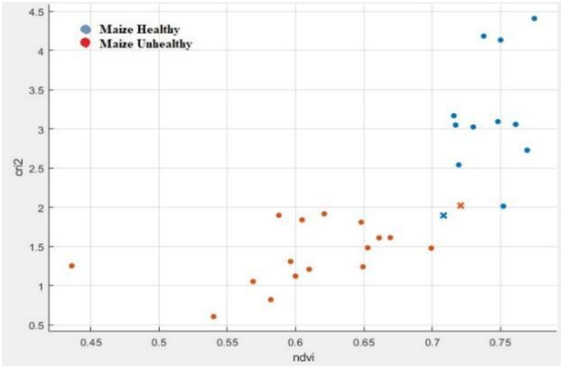

Fig.4b. SVM for Maize

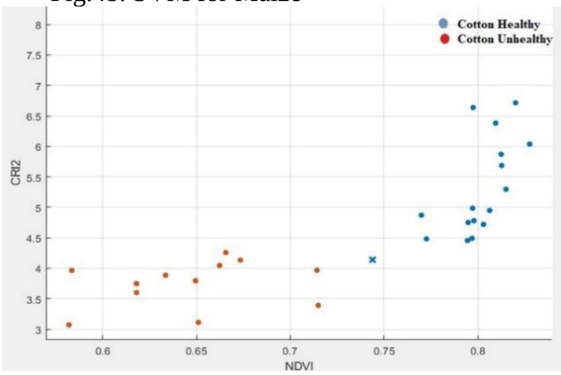

Fig.4c. Logistic Regression for Cotton

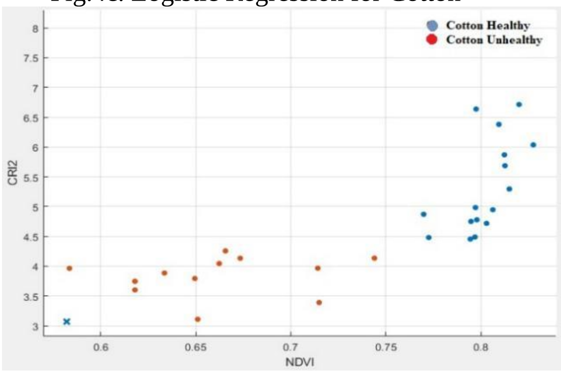

Fig.4d. SVM for Cotton

Fig.4. Classification results plotting for Cotton and Maize health assessment.

In the above figure fig.4, a, b, c, and d shows the scatter plots of data set when Logistic regression and SVM applied. In that figure, fig. 4. a & b shows that if the NDVI and CRI2 is greater than 0.7 and 2 respectively, then the crop leaf is healthy otherwise unhealthy. And the fig. 4. c & d shows that if the NDVI and CRI2 is greater than 0.7 and 4.25 respectively, then it is healthy otherwise not healthy. And when the logistic regression and SVM is applied on these plots, it is giving the confusion matrix as shown in the table 3.

11	1

11	1

0	18

a. Logistic Regression for Maize

1	17

b. SVM for Maize

17	0
1	12

c. Logistic Regression for Cotton

16	1
0	13

d. SVM for Cotton

Fig. 5. Confusion Matrices for applied confusion matrices

V. CONCLUSION

Both the vegetation indices can be useful for crop health assessment. The correlation of both index is positive. When NDVI increased, the CRI2 also increases. Normally NDVI and The CRI2 typically ranges from 0 – 1 and 0 – 15 respectively, but in the selected study area NDVI index varied between 0.54 - 0.78 for maize and 0.58 – 0.83 for Cotton. The CRI2 index varied between 0.6 and 4.4 for the Maize and 3.0-8.0 for cotton. From the study it is found that, logistic regression gives good results for NDVI and CRI2 indices.

The Cotton leaf which is having the NDVI and CRI2 values greater than 0.7 and 4.25 respectively is considered as the healthy. Whereas the Maize leaf having the NDVI and CRI2 Values greater than 0.7 and 2 respectively is also considered as healthy leaf.

ACKNOWLEDGEMENT

VI. The authors wants to acknowledge and thanks to the Chhatrapati Shahu Maharaj Research, Training and Human Development Institute (SARTHI), Pune for awarded the fellowship and also thanks to the Mutispectral research laboratory and Geospatial research laboratory facilities, Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India. For their support in research work.

REFERENCES

1. Mishra, P., Polder, G., & Vilfan, N. (2020). Close range spectral imaging for disease detection in plants using autonomous platforms: a review on recent studies. Current Robotics Reports, 1(2), 43-48.(j)
2. Thomas, S., Kuska, M. T., Bohnenkamp, D., Brugger, A., Alisaac, E., Wahabzada, M. & Mahlein, A. K. (2018). Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. Journal of Plant Diseases and Protection, 125(1), 5-20.(J)
3. Available on (https://agricoop.nic.in/sites/default/files/Aurangabad.pdf) access date 13/2/2023.(OL)
4. Available on (https://www.mgmkvk.com/about_district.php#:~:text=Agriculture%20in%20the%20Aurangabad%20district,Cotton%20is%20major%20cash%20crops.) access date: 13/02/2023. (OL)
5. Dutta, D., et al. "Assessment of vegetation health quality parameters using hyperspectral indices and decision tree classification." Proceedings of the ISRS Symposium, Nagpur, Maharashtra. 2009.(J)
6. Available on (https://byjus.com/biology/photosynthetic-pigments/#:~:text=Photosynthetic%20pigments%20are%20the%20most,transmitted%20by%20the%20plant%20tissue.) access date: 16/02/2023.
7. Izzuddin, M. A., et al. "Analysis of airborne hyperspectral image using vegetation indices, red edge position and continuum removal for detection of Ganoderma disease in oil palm." J. Oil Palm Res 30 (2018): 416-428. (O)
8. Yang, Ce, Won Suk Lee, and Jeffrey G. Williamson. "Classification of blueberry fruit and leaves based on spectral signatures." Biosystems engineering 113.4 (2012): 351-362.(O)
9. Online available at [https://www.researchgate.net/figure/Leaf-colour-chart-LCC-SPAD-and-NDVI-values-of-sweet-corn-recorded-at-different-stages_tbl3_305032515] accessed on date: 20/02/2023.(OL)
10. Nandibewoor, Archana, Saleem B. Hebbal, and Ravindra Hegadi. "Remote monitoring of maize crop through satellite multispectral imagery." Procedia Computer Science 45 (2015): 344-353.(P)
11. Solutions, Harris Geospatial, et al. "Vegetation analysis: using vegetation indices in ENVI." Harris Geospational Solutions [Online]. Available: https://www. harrisgeospatial. com/Learn/Whitepapers/TabId/2359/ArtMID/10212/ArticleID/16162/Vegetation-Analysis-Using-Vegetation-Indices-in-ENVI. aspx (2013). (OL)

12. Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N.: Assessing carotenoid content in plant leaves with reflectance spectroscopy. Photochemistry and photobiology, 75(3), 272-281 (2002).(O)
13. Ustin, S. L., & Jacquemoud, S. (2020). How the optical properties of leaves modify the absorption and scattering of energy and enhance leaf functionality. In Remote sensing of plant biodiversity (pp. 349-384). Springer, Cham.(O)
14. Online available [https://www.healthknowledge.org.uk/e-learning/statistical-methods/specialists/logistic-regression] accessed on date: 20/02/2023. (OL)
15. Sperandei, Sandro. "Understanding logistic regression analysis." Biochemia medica 24.1 (2014): 12-18.(O)
16. Dayton, C. Mitchell. "Logistic regression analysis." Stat 474 (1992): 574.(O)
17. Janardhanan, Padmavathi, and Fathima Sabika. "Effectiveness of support vector machines in medical data mining." Journal of communications software and systems 11.1 (2015): 25-30.(J)
18. Gitelson, Anatoly A., et al. "Assessing carotenoid content in plant leaves with reflectance spectroscopy¶." Photochemistry and photobiology 75.3 (2002): 272-281.
19. Zhang, Yao, et al. "Transfer-learning-based approach for leaf chlorophyll content estimation of winter wheat from hyperspectral data." Remote Sensing of Environment 267 (2021): 112724. (J)
20. Qi, Haixia, et al. "Hyperspectral inversion model of chlorophyll content in peanut leaves." Applied Sciences 10.7 (2020): 2259.
21. Izzuddin, M. A., et al. "Analysis of airborne hyperspectral image using vegetation indices, red edge position and continuum removal for detection of Ganoderma disease in oil palm." J. Oil Palm Res 30 (2018): 416-428.
22. feature extraction based on Narrowband vegetation indices of cotton and maize crops by EO-1 Hyperion Data
23. Kyratzis, Angelos C., et al. "Assessment of vegetation indices derived by UAV imagery for durum wheat phenotyping under a water limited and heat stressed mediterranean environment." *Frontiers in Plant Science* 8 (2017): 1114.**(J)**
24. Golhani, K., Balasundram, S. K., Vadamalai, G., & Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture*, *5*(3), 354-371.**(P)**