International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 30 December 2022*

Remote Sensing Based Soybean Crop Health Assessment Using Sentinel-2 Vegetation Indices and Machine Learning

Rahul B. Mannade [1], Kiran Sonkamble [2]
[1] Department of Information Technology, Government College of Engineering Aurangabad, Maharashtra, 431005, India
[2] Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, 431004, India,
[1] mannade.rahul@gmail.com, [2]kiransonkamble@gmail.com

Abstract:

Remote sensing methodologies have received increased importance for the surveillance of the vitality of crops and the enablement of precision agriculture. Satellite based multispectral imagery provides critical information about the vegetation status using spectral reflectance measurements and vegetative indices. In the current investigation, imagery collected from Sentinel-2 satellite was used for assessing the health condition of soya bean crops in a 4-acre agricultural plot located in the Sillod region of Maharashtra, India. Two temporal observations, which were dated at 1st October, 2022 and 26th October, 2022 were analyzed in order to understand the temporal variations in crop vigor during the advanced growth stages of soybean cultivation. Sentinel 2 Level 2A surface reflectance data were pre-processed by cloud and spatial clipping to the delineated area of interest. The calculation of vegetation indices, i.e. Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Red Edge Index (NDRE), and Chlorophyll Index Green (CIgreen) was used to characterize the canopy vigor and the concentration of chlorophyll.Machine learning classifiers, namely, Random Forest (RF) and Support Vector Machine (SVM) were used to analyze the crop condition patterns using the extracted vegetation indices. Crop health classification using NDVI threshold values classified the plot into various zones of healthy, moderate and poor vegetation conditions. The data showed that at 1st October 2022, most of the plot had healthy crop conditions, with 42 locations having healthy conditions and two having moderate conditions. As of 26 October 2022, the number of healthy locations had fallen to 34 but moderate and poor conditions had risen to eight and two locations respectively. Mean vegetation index values also showed a decrease between the two dates with NDVI showing a decrease from 0.746 to 0.673 while CIgreen showed a decrease from 4.141 to 3.666. The average NDVI change of 0.073 shows a slow decline in canopy vigor as the soybean crop came closer to phenological maturity. The results show the value of Sentinel-2 derived vegetation indices in monitoring temporal variations in the health of crop in soybean fields.

Keywords: Remote Sensing, NDVI, Crop Analysis, Machine Learning, RF, SVM

1. Introduction

Remote sensing has become an invaluable tool in agricultural monitoring today, thanks to the ability to provide large scale, timely and non-destructive observation of crop condition. Satellite imagery allows researchers and practitioners to examine crop growth, vegetation health and environmental stress using the utilisation of spectral reflectance information. Vegetation indices obtained from multispectral imagery are widely used for monitoring crop development, detecting stressors and estimating biophysical parameters (e.g. biomass, chlorophyll content and leaf area index). With the increased availability of high resolution satellite data, such as Sentinel-2, the area of remote sensing has developed to become a powerful methodology for precision agriculture and crop management [1].

In recent years, machine learning techniques have been used more in agricultural remote sensing applications such as crop classification, yield estimation, disease detection, and crop health assessment. Algorithms like Random Forest and Support Vector Machine have shown good abilities in solving complex relations between spectral and crop attributes. These models are

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 30 December 2022*

able to process voluminous satellite derived feature sets, and extract salient patterns that lead to a better understanding of the dynamics and variability of crops in agricultural fields [2].

Soybean is one of the most important oilseed cultivars in India especially in the state of Maharashtra. The crop plays an important role in providing edible oil, animal feed and raw material for various food industries. Maharashtra is richly contributing to a major part of India's soybean production with districts like Aurangabad, Jalna and Beed becoming the key areas of cultivation. Monitoring of the health of soybean crop at different stages of growth is critical to enhance productivity and sustainable agro production. Remote sensing technologies are an effective method for monitoring the condition of crops across vast agricultural areas [3].

The present study focuses its efforts on assessing the health of soybean crop using the utilisation of Sentinel-2 satellite imagery and vegetation indices. A four-acre sockets soybean field located in Sillod region of Maharashtra was analysed by the use of multispectral imagery captured on two dates, i.e., 1 October 2022 and 26 October 2022. Vegetation indices, such as NDVI, GNDVI, NDRE and CIgreen were extracted to evaluate crop condition. Moreover, machine learning classifiers, namely Random Forest and Support Vector Machine were used to analyse crop health patterns. The objective of this investigation was to study the temporal changes in the crop vigor and assess the effectiveness of vegetation indexes in monitoring the health of soybean crop.

2. Literature Survey

Several studies have been carried out on the use of remote sensing and vegetation indices in monitoring crop's health and agricultural productivity.

A study published in MDPI looked at the use of Sentinel-2 vegetation indices in crop monitoring and in chlorophyll estimation. The research proved that red-edge based indexes like NDRE are better sensitive to the chlorophyll content than traditional indexes like NDVI. The study showed the benefit of the application of Sentinel-2 red-edge bands in capturing small changes in crop physiological conditions as well as crop growth dynamics during the growing season [4].

Another research work was done on using machine learning algorithms for classification of crops based on multispectral satellite imagery. The article's authors tested various machine learning techniques, such as Random Forest and Support Vector Machine, for the purposes of agricultural mapping and crop health assessment. Their results showed that Random Forest is excellent in processing non-linearity between spectral features and crop parameters while SVM provides good generalization ability for classification tasks [5].

A further MDPI study discussed the application of vegetation indices based on Sentinel-2 imagery for monitoring crop growth and detecting crop stress. The results of this research demonstrated that several indices such as GNDVI and CIgreen are sensitive to chlorophyll variations and can be successfully used to track the crop nutrition status and plant vigor. The results confirmed that combining several vegetation indices is beneficial for improving accuracy of crop condition assessment and delivering useful information for applications in precision agriculture [6].

These studies collectively show the importance of vegetation indices and machine learning approaches in agricultural monitoring based on remote sensing approaches.

3. Methodology

In the methodology of this work, the crop health of soybean crop is the focus of attention that is assessed using satellite remote sensing data and vegetation indices acquired by Sentinel-2 images. The overall workflow consists of a number of sequential stages, such as the acquisition of the data sets, pre-processing of the satellite imagery, feature extraction using the calculation of the vegetation index and the analysis of the crop health using machine learning classifiers. Sentinel 2 Multispectral imagery was processed to derive spectral information relevant to vegetation monitoring. Preprocessing steps, including cloud masking and cropping to the region of interest were performed to ensure proper analysis in the region of interest, which in this case was a soybean field. Subsequently, vegetation indices such as NDVI, GNDVI, NDRE and CIgreen were calculated to represent crop vigor and chlorophyll contents. These indices were then used as input features in machine-learning models, i.e., Random Forest and Support Vector Machine, to study crop health patterns. The last

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 30 December 2022*

stage was the evaluation of vegetation index trends, along with the spatial distribution of crop condition between the two observation dates in order to understand changes in soybean crop condition across time.In this research we have used following methodology figure 1,

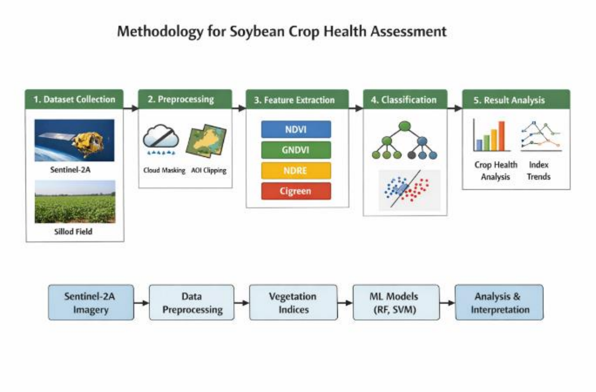

Figure 1: Proposed Methodology

3.1 Dataset

The data set which was used for this investigation is Sentinel-2 satellite data along with ground observations taken from a soybean field located in the Sillod region of Aurangabad district, Maharashtra, India. The Sentinel-2 mission, a multispectral Earth-observation satellite operated by the European Space Agency, provides high-resolution imagery from spatial resolutions of 10 m to 60 m. The instrument captures a number of spectral bands (visible, near-infrared and red-edge wavelengths), which are particularly friendly to vegetation monitoring [7].

Two acquisitions of Sentinel-2 imagery were used corresponding to 1 October 2022 and 26 October 2022. Ground-truth data, which consist of geographic coordinates and chlorophyll concentration measurements, were collected from the soybean field in order to validate the satellite-determined estimates.

3.2 Preprocessing

Preprocessing of satellite imagery is an important prerequisite for the accurate calculation of the vegetation indices. In the present study, Sentinel 2 Level 2A surface reflectance products were used. Cloud masking was implemented to remove the pixels that are affected by atmospheric interference and cloud cover. The study area was defined by a shapefile representing a 4 acres soybean field boundary [8].

The imagery was clipped to the area of interest (AOI) in order to isolate the soybean field from neighboring land cover. This step ensured that only relevant pixels in the field were used in the vegetation index calculations and further analysis [9].

3.3 Feature Extraction

Vegetation indices were calculated from Sentinel-2 spectral bands in order to measure crop condition and canopy vigour. The indices calculated were NDVI, GNDVI, NDRE and CIgreen.

Normalized Difference Vegetation Index (NDVI) is a commonly used index for assessing vegetation health and biomass in the form of a quantification of the difference between near infrared and red reflectance [10].Green Normalized Difference Vegetation Index (GNDVI) uses the green band instead of the red band and has an increased sensitivity to the concentration of chlorophyll.Normalized Difference Red Edge Index (NDRE) exploits the red-edge wavelengths and provides improved sensitivity to crop chlorophyll and canopy structure.Chlorophyll Index Green (CIgreen) is used to estimate the chlorophyll content using the combination of near-infrared and green spectral bands [11][12].

These vegetation indices provide useful information on plant vigour, chlorophyll content and crop physiological status.

3.4 Classification

Machine learning techniques were used to classify the health of soybean crop depending on the obtained vegetation indices. Two popular classifiers were used for this purpose, Random Forest and Support Vector Machine.

Random Forest is an ensemble learning framework which builds several decision trees and combines the output of decision trees, thereby enhancing the classification performance. The method is good at dealing with nonlinear relationships and large amounts of data [13].

Support Vector Machine A supervised algorithm for finding optimal decision boundaries for segregating classes in a high dimensional feature space. SVM is well suited for the small sample scenarios and provides a good generalization [14].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 30 December 2022*

Both classifiers were used with the vegetation indices as input features to classify crop condition into different health levels.

3.5 Result Analysis and Discussion

The crop health classification results were analyzed for two observation dates: 1 October 2022 and 26 October 2022.

Crop health classification based on NDVI values produced the following results as depicted in table 1:

Table 1: Crop health classification based on NDVI

Date	Healthy	Moderate	Poor
1 Oct 2022	42	2	0
26 Oct 2022	34	8	2

The results indicate that the majority of the soybean field exhibited healthy vegetation conditions on 1 October 2022. However, by 26 October 2022, the number of healthy locations decreased while moderate and poor crop conditions increased.

The mean vegetation index values for the two dates are presented below in table 2:

Table 2: Mean Vegetation Index Values

Date	NDVI	GNDVI	NDRE	CIgreen
1 Oct 2022	0.7458	0.6691	0.4849	4.1411
26 Oct 2022	0.6729	0.6335	0.4773	3.6655

All vegetation indices show a slight decline between the two observation dates. The average NDVI change between the two dates was −0.0728, indicating a reduction in vegetation vigor.

This decline in vegetation indices may be associated with the soybean crop entering later growth stages such as pod filling and physiological maturity. During these stages, chlorophyll concentration decreases and leaf senescence begins, resulting in lower spectral reflectance in the near-infrared region.The spatial distribution maps generated from the vegetation index data further confirm the transition of some areas from healthy to moderate crop condition during the study period.

The decline in the vegetation indices between 1st October 2022 to 26th October 2022 shows a perceptible change in the physiological condition of the soybean crop in the late growth stage. NDVI dropped from 0.7458 to 0.6729 and GNDVI and CIgreen also showed reductions which suggested a decrease in chlorophyll concentration and canopy vigor. Similarly, NDRE showed a slight decrease, which is further evidence of decrease in plant chlorophyll content. These changes are typical of the usual soybean growth phase when crops start to transition from an active vegetative growth phase to a mature and senescent phase. The shift toward the higher moderate and poor crop condition classes at the second observation date suggests that areas of decreased vigor of the vegetation had occurred on some parts of the field perhaps from natural aging of the crop, nutrient limitations, or localized environmental stress. Overall, the results prove that vegetation indices generated from Sentinel-2 imagery are able to capture slight temporal changes in the health of crops in a relatively small agricultural field.

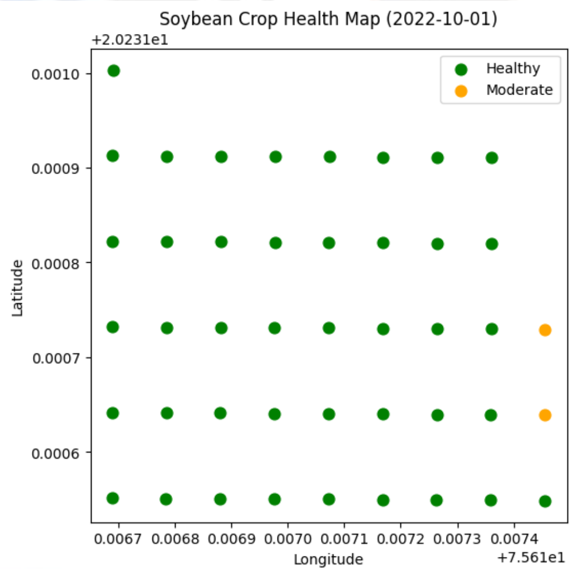

Figure 2: Soybean Crop Health First Date

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 30 December 2022*

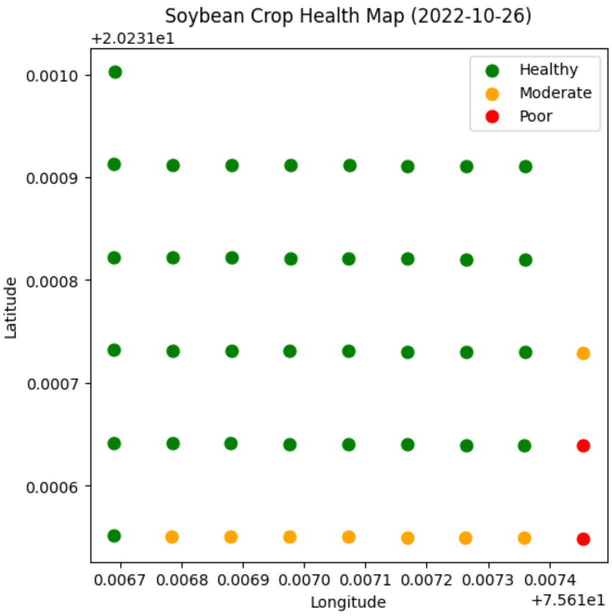

Figure 3: Soybean Crop Health SecondDate

In order to explain the spatial and temporal change in the health of the soybean crop, it is a good idea to use several visual representations in the analysis as shown in figure 2 and figure 3 for both dates. Spatial crop health maps based on the NDVI values provide a good representation of the distribution of healthy, moderate, and poor vegetation distribution over the study area for both the observation dates, and hence, show the progression of the crop condition in early-to-late-October. Additionally, bar charts can be used to compare the number of sampling points that fall into the healthy, moderate and poor categories on each sampling occasion, helping to uncover changes in the categories of crop health from sampling period to sampling period. Line graphs or column charts showing mean values of NDVI, GNDVI, NDRE and CIgreen for both dates also show further evidence of the decline in vegetation vigor during the time of the study. Scatter plots comparing vegetation indices with ground based chlorophyll measurements may also be prepared to visualize the correlation of the spectral indices with physiological status of the crop. Collectively, these graphical outputs enable an intuitive grasp of the dynamics of crop health and increase the interpretability of satellite-based analyses.

4. Conclusion

This investigation supports the usefulness of Sentinel 2 satellite images and indices of vegetation, for monitoring the physiological condition of a soybean crop grown in a 4 acre of Sillod, Maharashtra. By taking advantage of multispectral information, the study gives a framework for satellite-based evaluation of plant health in a isolated agricultural setting. Vegetation indices including the Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Red Edge (NDRE) and the CIgreen index, were spectrally calculated from the Sentinel-2 imagery for the quantification of the crop vigor and the concentration of chlorophyll. Concurrently, machine learning classifiers such as Random Forest and Support Vector Machine were used to identify spatial patterns of health of the crops and thus combining predictive analytics and remote sensing observations. Empirical findings show that most of soybean field had a good vigor, at early October, 2022, and a progressive attenuation of vegetation indices was documented at late October. The observed decrement in index values along with the shift from healthy to moderate crop condition indicated that the soybean crop was moving towards physiological maturity during the study period. The present study highlights the potential of satellite based remote sensing methodology to real time monitoring of crop health and facilitating precision agriculture practices. Prospective research should include longer temporal monitoring, more field data sets, and more advanced machine learning applications to further improve the precision of crop status monitoring and yield forecasting systems.

References

1) Khanal, S., Kc, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote sensing*, *12*(22), 3783.
2) Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote sensing*, *12*(19), 3136.
3) Rathore, V., Sharma, H., &Narvariya, R. (2021). Growth rate of cost of cultivation of soybean in Maharashtra States of India. *Pharm. Innov. J*, *10*(3), 84-89.

4) Segarra, J., Buchaillot, M. L., Araus, J. L., & Kefauver, S. C. (2020). *Remote sensing for precision agriculture: Sentinel-2 improved features and applications.*
5) Holtgrave, A. K., & Förster, M. (2020). Comparing Sentinel-1 and Sentinel-2 data and indices for agricultural vegetation monitoring. Remote Sensing, 12(18), 2919. https://doi.org/10.3390/rs121829
6) Kayad, A., et al. (2019). Monitoring within-field variability of corn yield using Sentinel-2 vegetation indices and machine learning techniques. Remote Sensing, 11(23), 2873. https://doi.org/10.3390/rs11232873
7) Kobayashi, N., Tani, H., Wang, X., &Sonobe, R. (2020). Crop classification using spectral indices derived from Sentinel-2A imagery. *Journal of Information and Telecommunication*, *4*(1), 67-90.
8) Candra, D. S., Phinn, S., & Scarth, P. (2020). Cloud and cloud shadow masking for Sentinel-2 using multitemporal images in global area. *International Journal of Remote Sensing*, *41*(8), 2877-2904.
9) Řezník, T., Chytrý, J., & Trojanová, K. (2021). Machine learning-based processing proof-of-concept pipeline for semi-automatic sentinel-2 imagery download, cloudiness filtering, classifications, and updates of open land use/land cover datasets. *ISPRS International Journal of Geo-Information*, *10*(2), 102.
10) Ramos, A. P. M., Osco, L. P., Furuya, D. E. G., Gonçalves, W. N., Santana, D. C., Teodoro, L. P. R., ... &Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*, 105791.
11) Chea, C., Saengprachatanarug, K., Posom, J., Wongphati, M., & Taira, E. (2020). Sugar yield parameters and fiber prediction in sugarcane fields using a multispectral camera mounted on a small unmanned aerial system (UAS). *Sugar Tech*, *22*(4), 605-621.
12) Zhou, X., Kono, Y., Win, A., Matsui, T., & Tanaka, T. S. (2021). Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. *Plant Production Science*, *24*(2), 137-151.
13) Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., &Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308-6325.
14) Demidova, L. A., Klyueva, I. A., & Pylkin, A. N. (2019). Hybrid approach to improving the results of the SVM classification using the random forest algorithm. *Procedia Computer Science*, *150*, 455-461.