IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 June 2022, accepted 19 June 2022, date of publication 13 July 2022, date of current version 25 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3190716

An Efficient Machine Learning Enabled Non-Destructive Technique for Remote Monitoring of Sugarcane Crop Health

EKTA PANWAR[1], ANJANA NAGA JYOTHI KUKUNURI[2], DHARMENDRA SINGH[2], (Senior Member, IEEE), ASHWANI KUMAR SHARMA[1], AND HARISH KUMAR[3]

[1]Department of Biosciences and Bioengineering, IIT Roorkee, Roorkee 247667, India
[2]Department of Electronics and Communication Engineering, IIT Roorkee, Roorkee 247667, India
[3]Department of Computer Science, College of Computer Science, King Khalid University, Abha 61413, Saudi Arabia

Corresponding author: Dharmendra Singh (dharmfec@gmail.com)

This work was supported in part by the Council of Scientific and Industrial Research, New Delhi; in part by the Drone Research Centre, IIT Roorkee; and in part by the Deanship of Scientific Research at King Khalid University, Abha, Saudi Arabia, under Grant R.G.P.2/198/43.

ABSTRACT Crop health can be predicted based on various biochemical variables of crops, which include chlorophyll, phenol, carbohydrate, lipid, protein, hydrogen peroxide, and proline as these variables play a critical role in maintaining the intricate phytochemistry of crop plants. In-situ monitoring of the above-mentioned variables is very cumbersome and laborious, so it is atrociously needed to identify some alternatives to monitor these variables in crop plants. Assessing these variables using satellite data may be a good choice provided it has a high spatial and temporal resolution. Sentinel-2 satellite sensor contains VNIR/SWIR spectral region, two red-edge bands, and also has good spatial as well as temporal resolution so it may be the finest option. Precise information of the field is required for the development of a retrieval algorithm for which the drone data is used, as it is highly accurate ground truth reference data. In previous studies, the researchers have primarily focused on the monitoring of biophysical and morphological parameters of crop plants like leaf area index, plant height, and stomatal conductance using these spectral features. Because monitoring biochemical variables of crop plants using satellite derivatives is still a difficult undertaking for academics, just a few studies have been published. As a result, in this study, an attempt is made to establish a methodology for monitoring sugarcane crop biochemical characteristics utilising satellite-derived variables. Satellite derivatives, i.e., vegetation indices are extracted using Sentinel-2 data while biochemical variables of the crop (as mentioned above) are analyzed using leaf samples in the laboratory using optimized protocols. Subsequently, a Machine learning-based Gaussian process regression model is developed for all the biochemical variables using the different combinations of vegetation indices. The developed model showed promising results with R^2 greater than 0.7 and normalized root mean square error (NRMSE) less than 0.2 thus holding good potential for effective monitoring the crop health condition remotely.

INDEX TERMS Crop health, non-destructive technique, remote sensing, vegetation indices, biochemical variables.

I. INTRODUCTION

The continuous monitoring of crop health is an essential aspect of agriculture to get better yield and productivity from the crops. Various physiological, biochemical, and morphological variables of crop plants are reliable indicators of crop health conditions [1]–[3]. These variables play crucial roles in maintaining the complex phytochemistry of crop plants in photosynthesis, cell division, as a building block for cell walls, signaling molecules, enzymes, energy storage, as secondary metabolites, and also in the shield of plants counter to several abiotic and biotic stresses [4]–[7].

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

The monitoring of crops with regular field visits is a very weighty process as it requires manpower and a huge amount of time. So, in the modern decade of precision agriculture, with a wide range of satellite data namely Sentinel-2, Palsar, Modis, and Landsat data readily accessible [8]–[10]; if one can monitor crops remotely, then it would be highly beneficial for farmers as well as agriculture planners. The satellite data may be one alternative that can help in monitoring crop health remotely [11]–[15]. But while considering the satellite data, a few things are significant, i.e., temporal resolution, spatial resolution, and the cost of data. However, a few satellite data are available free of cost, e.g., Sentinel-2 data, but the challenge exists in how to utilize this freely available Sentinel-2 data effectively for different agricultural applications. The spectral resolution of satellite data also has great importance while one is considering it for crop monitoring, the bands of the sensor of that satellite should have sensitivity for vegetation. Sentinel-2 turns out to be the best fit in terms of spectral as well as temporal resolution [16]–[20]. Therefore, in this study, the sensitivity of Sentinel-2 derivatives for the biochemical variables (carbohydrate, protein, lipid, phenol, proline, chlorophyll, hydrogen peroxide, and sodium dismutase) of sugarcane crops is explored to feat its role in crop health monitoring. Here we have chosen biochemical variables of the sugarcane crop as key parameters to assess the crop health because these parameters are more sensitive to environmental stresses as compared to physiological changes in crop plants and are better stress markers in crop plants [2]–[5]. In this study, we have chosen sugarcane as a model crop as it is the world's largest crop by production quality and also the native crop of India. This crop not only provides many value-added products in its raw form but its waste which is sugarcane bagasse is also a raw material for many industries like the production of papers, paperboard products, and panel boards. Sugarcane is also one of the plants with the highest bioconversion efficiency and is being widely used for biofuel production in many countries like Brazil, India, and China. So if we can develop a measure to increase its yield and productivity then it will be highly beneficial for mankind. As remote sensing has emerged as a highly advanced technique for precise agriculture monitoring [21], with a broad array of applications, including land cover assessment [22], yield and biomass estimation [23]–[25], vegetation classification [26], estimation of gross primary productivity [27], canopy water content and plant physiological status analysis [28]–[30].

Due to the extreme weather conditions, the growth, productivity, and yield of the crop got severely affected [31]. So it is mandatory to find out some suitable measures to deal with this continuous surging problem. As we know, the plant system is sophisticated machinery in which numerous biochemical reactions take place to keep the pace of growth and development and also to deal with the constraints imposed by the environment on the crop plants. If a plant experiences any difficulty from the environment in terms of abiotic (temperature, light, heat, cold, chilling, drought, flood, salt, and metal stress) or biotic (insects, weeds,pesticides, etc.) factors, the plant machinery undergoes remarkable changes to protect the crop plants from injury and irreversible damage [3], [32]–[34]. These changes in the crop plants can be observed by analyzing the disparity in the content of biochemical variables of crop plants from time to time [36]. But traditional methods of analyzing these biochemical variables in crop plants depend on plant tissue analysis, which is hugely impeded, requires a high workforce, expensive, and involves the use of a significant amount of chemical reagents. Moreover, these methods become incompetent when it comes to analyzing these variables at a larger scale like a big field of crops or grassland, etc. [37]–[38]. Also when the crop is fully grown and face sudden outburst of any kind of environmental constraints such as attacks of aphids, weed, fungi, virus, or bacteria, then it becomes impossible to collect the plant samples from the middle of the field as that can be very dangerous and impractical.

The real-time and accurate monitoring of biochemical parameters in crop plants can help farmers to make the right decision to improve the yield and productivity of crops. The remote analysis of crop fields using satellite data can be a very effective solution for the above-stated problem. It is well acknowledged that a proper understanding of the interaction between reflectance and vegetation can help extract abundant information about the crop plant using satellite data [10], [29], [30], [39]. Also, the necessity for manual monitoring of the crops can be completely sacked, if one can uncover the essential information about the crop plants remotely. The Sentinel-2 data may be used to evaluate crop health by directly employing the zero-order product, i.e. individual band reflectance [20], [28], as well as the first- order product, i.e. vegetation indices [14], [17], [40]. However vegetation indices were found to be more effective at extracting various characteristics of field crops, as these are developed by using different combinations of satellite bands that were found to be sensitive to vegetation [41]–[43]. On the other hand, employing reflectance from individual bands of distinct wavelength as a metric for estimation of the biochemical variable of the crop plants are not very productive because each wavelength has its measurement error, which increases towards both water absorption wavebands and longer wavelengths so it can only be utilized to have the initial and overall idea about the crop condition [12], [44].

Various types of active and passive sensors are available (such as optical, hyperspectral, multispectral, Radar, and Lidar), which are capable of extracting the spectral signature of vegetation in the form of reflectance (zero-order derivatives). With the use of reflectance, vegetation indices (first-order derivatives) can be calculated, which are highly effective for various applications such as monitoring biophysical, physiological, and photosynthetic variables in vegetation [15], [35]. Each vegetation index is a unique mathematical combination of at least two spectral bands, which can efficiently accentuate the spectral properties of

all types of vegetation, including herbs, shrubs, trees, forest grasslands, and crop fields [8], [10], [45]. Vegetation indices are capable of separating vegetation from other features in the satellite images. The most widely used vegetation index NDVI is reported to be in good agreement with leafy green biomass [46], green leaf cover [47], and crop yield [48]. The applications of vegetation indices are reported in various domains such as examining the climate trends [49], assessing the change in biodiversity [50], monitoring drought [51], transpiration, and evaporation in plants [52], scheduling crop irrigation [53], classify vegetation [54], and also to estimate soil water content [55], [56]. Thus, it can be considered a highly potential spectral signature for determining the biochemical variables in crop plants.

Also, observing these biochemical parameters with satellite data need perfect ground truth, which is quite challenging to obtain with simple ground measurements. Drones are capable of providing highly useful and very detailed information about the several features of land cover like population ecology, ecological monitoring, and conservation. Unmanned aerial systems have also been effectively deployed in agriculture, environmental protection, and mining [57], [58]. Therefore, we employed high- resolution drone data to obtain exact ground truth information in our investigation. As a result, an attempt has been made in this research to explore the possibility of monitoring sugarcane crop biochemical variables using the Sentinel-2 derived variables (vegetation indices) and drone data in combination.

For the development of an effective method for retrieval of the biochemical variables of sugarcane crops using the satellite-derived vegetation indices, we have employed machine learning techniques [59]. Machine learning regression approaches could be a suitable fit for estimating biochemical characteristics of sugarcane crops using satellite-derived vegetation indicators as of their capability to fit the nonlinear data in an adaptive manner. In previous studies, researchers successfully used four cutting- edge machine learning regression techniques to examine biophysical parameters: fractional vegetation cover (FVC), crop height and density assessment, and leaf area index (LAI): Gaussian processes regression (GPR), neural networks (NN), kernel ridge regression (KRR) and support vector regression (SVR) [60]–[62]. In another study, Gaussian processes were found to be more effective than other machine learning techniques at establishing complex relationships at the canopy level due to their ability to address three key concerns for better model development: it is easier to parameterizing signal and noise ratio in the kernel function, variables were also adjusted using a proficient algorithm and GPR also provides physical insights in the problem by ranking the spectral bands according to their relevance hence solves the blank box problem of previous methods. As a result, this strategy has shown to be a useful tool for operational vegetation cover monitoring that can be readily integrated into hyper or multispectral data processing chains [63]. Researchers have also utilised a different approach, which involves training a regression model with data provided by a Radiative Transfer Model as well as real-time observations. They created a simple and reliable nonlinear nonparametric regression model for extracting the leaf area index (LAI) from Landsat data and PROSAIL-generated simulated data [64]. For the estimate of seawater chlorophyll concentration and the leaf membrane stability index of marine flora, a completely adaptive Gaussian process to signal and noise features were recently proposed in a study [65]. So in nutshell, we have attempted to create a method based on machine learning to retrieve biochemical variables of sugarcane crops using satellite-derived vegetation indices using machine learning regression analysis. For more precise and accurate ground truth information drone data is taken into consideration. The following is the format for this paper. Section 2 offers a brief synopsis of the study's contents, while Section 3 contains the methodology used. The findings are examined in Section 4, and the study's conclusion is presented in Section 5.

II. MATERIALS AND METHODS

In this study, sugarcane has been used as a model crop for exploring the efficiency of vegetation indices in predicting the different biochemical variables of the crop.

A. STUDY AREA AND COLLECTION OF SAMPLES

The study area chosen for this study, i.e., Krishi Vigyan Kendra, Dhanauri, Roorkee, Uttarakhand, India, is agricultural land with diverse vegetation comprising crop fields of wheat, rice, sugarcane, soybean, pulses, and legumes. It is located nearby Roorkee and has a centre latitude 29.930596555° N, and a longitude 77.96395277° E. The study area is used for the collection of leaf samples of sugarcane crop and also the drone data (image of 1,50,000 m^2 study area) for extracting the precise ground truth information of the sampling sites. For the remote insights into the crop field, a subset of 1,50,000 m^2 has been cropped from the complete tile of the Sentinel-2 satellite (as shown in Figure 1(c)), which corresponds to our study area, dated October 7, 2018 (as shown in Figure 1 (d)). In the cropped subset, three different sugarcane patches are marked as ss-1 (12000 m^2), ss-2 (36000 m^2), and ss-3 (12000 m^2) which corresponds to three different sampling sites (shown in Figure 1 (d)).

Leaf samples of sugarcane were plucked from three separate sugarcane crop sampling sites in the study site, namely ss-1, ss-2, and ss-3 (shown in Figure 1 (d)) for the investigation of sugarcane crop biochemical characteristics. The sugarcane crop plant's third fully-formed leaf was collected and utilised to measure major biochemical variables namely carbohydrate, protein, lipid, phenol, proline, chlorophyll, hydrogen peroxide, and sodium dismutase. The plucked leaves were brought to the laboratory in sealed polybags and were then stored at −80° C after lyophilization in liquid nitrogen. Further, the laboratory analysis was performed using these stored samples using an optimized protocol (details are given under section-C in subsection-i) for each biochemical variable, and biochemical data was logged.

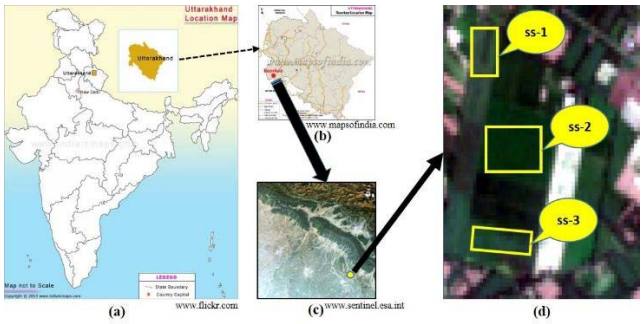


FIGURE 1. Showing the details of the study site, (a) Map of India, (b) an enlarged view of Uttarakhand state, (c) Orthorectified tile of sentinel-2 data of Roorkee Region, and (d) A processed subset of the study area showing three sampling sites(ss) named ss-1, ss-2, and ss-3.

FIGURE 2. (a) Drone image of the selected agricultural site (b) Drone image gridded into sentinel-2 equivalent pixel size.

B. DATA SET USED AND PREPROCESSING

Two different datasets have been procured for the study, i.e., drone data and sentinel-2 data. A detail of the same is given below:

1) DRONE DATA

Drone data for this study has been acquired using one drone with model type DJI Phantom quadcopter, which is hovered above the study area (Sugarcane crop field) to capture the images of the field at the height of 120 meters and spatial resolution of 6 cm from the ground level. This drone contains a single high-definition RGB camera with 4K resolution and also has an attached GPS unit, so it is capable of capturing geotagged images, which is a prerequisite to locating our sampling sites (ss-1, ss-2, and ss-3) in the Sentinel-2 data tile. The images are captured in such a manner that after the first image is captured, every next image has at least 40% overlapping with the previous image. The acquired drone images of all the sampling sites have been pre-processed by mosaicking, orthorectification, and geo-registration using pix4Dmapper software. A final output image captured using a drone and also the drone image gridded into Sentinel-2 equivalent pixel of October 5, 2018, is presented in Figures 2(a) and 2(b), separately. The dates have been picked in this manner that the time of capturing drone data remain closest to the date at which the Sentinel-2 satellite passes over the study area (as given in table-1).

TABLE 1. Particulars of drone data used for this study.

S.No.	Centre Latitude	Centre Longitude	Date of acquisition
1	29.930596555° E	77.96395277° N	January 19, 2018
2	29.930596555° E	77.96395277° N	February 09, 2018
3	29.930596555° E	77.96395277° N	March 09, 2018
4	29.930596555° E	77.96395277° N	April 13, 2018
5	29.930596555° E	77.96395277° N	May 10, 2018
6	29.930596555° E	77.96395277° N	June 12, 2018
7	29.930596555° E	77.96395277° N	October 5, 2018
8	29.930596555° E	77.96395277° N	November 22, 2018
9	29.930596555° E	77.96395277° N	December 16, 2018

2) SATELLITE DATA

Sentinel-2 is a multispectral satellite with excellent spatial and temporal resolution. Its revisiting frequency is five days, and it does so with the same view angle, making it ideal for crop field monitoring. Sentinel-2's optical sensor captures data in 13 spectral channels. Four of them, namely band-2 (blue), band-3 (green), band-4 (red), and band-8 (near-infrared), have a spatial resolution of 10 metres, while six of them, namely band-5 (vegetation red edge), band-6 (vegetation red edge), band-7 (vegetation red edge), band-8A (vegetation red edge), band-11 (short wave infrared) have a spatial resolution of 20 metres and three of them namely band 1 (coastal aerosol), band 9 (water vapour), and band 10 (short wave infrared-cirrus) have 60 meters spatial resolution. Data of Sentinel-2 can be used in a variety of ways credit to its wide range of spatial resolution. Users can download the European Space Agency's products as 100*100 Km^2 ortho-rectified tiles in the UTM/WGS84N44 zone, which additionally include a selection of user-selectable spectral bands. A total of nine tiles of the Sentinel-2 satellite have been used in this study with tiles id's; A, B, C, D, E, F, G, H, and; I as depicted in Table 2. In this study, we have used satellite data from January 2018 to June 2018 for model development and then data from October 2018 to December 2018. We have not taken data for the months July 2018 to September 2018 because during these months data is affected by the clouds in the area of our study region and we can not extract desired information from the clouded data. The Sen2cor processor, given by the European Space Agency, was used to adjust the Sentinel-2 data tiles for atmospheric influences. The Level-2A Atmosphere (BOA) reflectance product from the Level-1C Top-of-Atmosphere (TOA) product is the

TABLE 2. Particulars of drone data used for this study.

ID	Data ID	Date of acquisition
A	S2B_MSIL2A_20180120T053209_N0206_R105_T43RGP_20180120T092123	January 20, 2018
B	S2B_MSIL2A_20180209T053209_N0206_R105_T43RGP_20180209T092123	February 09, 2018
C	S2B_MSIL2A_20180311T053209_N0206_R105_T43RGP_20180311T092123	March11, 2018
D	S2B_MSIL2A_20180415T053209_N0206_R105_T43RGP_20180415T092123	April 15, 2018
E	S2B_MSIL2A_20180510T053209_N0206_R105_T43RGP_20180510T092123	May 10, 2018
F	S2B_MSIL2A_20180619T053209_N0206_R105_T43RGP_20180619T092123	June 19, 2018
G	S2B_MSIL2A_20181007T052709_N0206_R105_T43RGP_20181007T091700	October 7, 2018
H	S2B_MSIL2A_20181121T052709_N0206_R105_T43RGP_20181121T091700	November 21, 2018
I	S2B_MSIL2A_20181216T052709_N0206_R105_T43RGP_20181216T091700	December 16, 2018

final output after rectification. Atmospherically corrected, the Sentinel-2 image has been subjected to resampling with reference to the band-2 followed by the cropping of a subset from the huge tile of the Sentinel-2 as shown in section 2.1 (i.e., Figure 1(d)). The image's projection was additionally changed from UTM (Universal Transverse Mercator) to Geographic latitude and longitude format before resampling. Using drone data as ground truth reference data, the cropped image was then utilised to label the pixels of interest based on their latitude and longitude for each ROI (Region of interest).

C. PROPOSED METHODOLOGY

For the retrieval of biochemical variables of sugarcane crops utilising sentinel-2 derived vegetative indices, a machine learning-based approach is developed and implemented. Figure 3 explains the entire flowchart of the suggested methodology. The detailed processes of the suggested methodology are provided in the below-mentioned sections.

1) LABORATORY ANALYSIS OF PLANT SAMPLES FOR OBTAINING BIOCHEMICAL DATA

The biochemical variables of the sugarcane crop were investigated using freeze-dried leaf samples and optimised procedures, as detailed below. The Metzner method is the traditional method for estimating chlorophyll [66], while the Ninhydrin approach has been used to measure proline [67]. Similarly, the Bradford method was used to measure protein [68], and the Folin-ciocalteau method was used to determine phenol [69]. The Trichloro-acetic acid method is used to calculate hydrogen peroxide [2], while the Anthrone method is used to calculate carbohydrates [70]. The following are the specifics of the optimised protocols:

a: CARBOHYDRATE ESTIMATION

Plant samples were hydrolyzed in 2.5 N hydrochloric acid before being neutralized with a Na_2CO_3 solution (sodium carbonate). After centrifuging the lysate for 10 minutes at

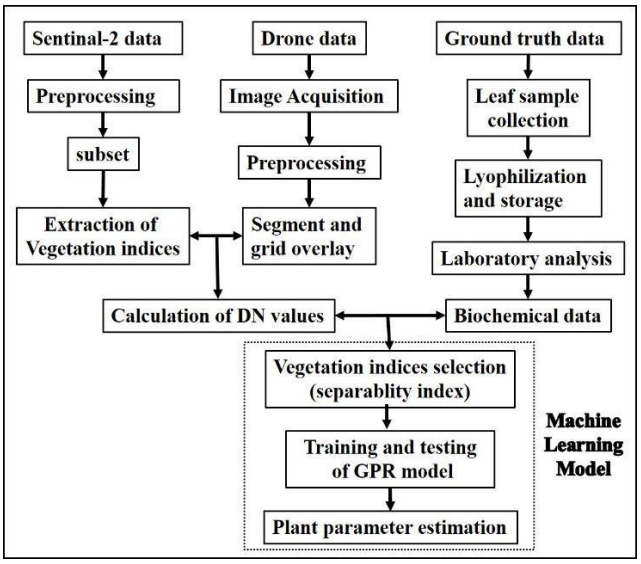


FIGURE 3. Flowchart of the proposed methodology.

10000 rpm, the supernatant was collected in a separate container. The optical density of each sample was measured at 660 nm after freshly produced anthrone reagent was added to the supernatant. Glucose was used to create the standard.

b: PROTEIN ESTIMATION

To measure protein content, the leaf samples were homogenised in phosphate buffer (pH 7.0) with 0.05 percent Triton X-100, 1 mM Ascorbic acid, 2% PVP (polyvinyl pyrrolidine). and 1 mM EDTA (ethylene diamine tetraacetic acid) at 4°C, the aforementioned solution was subjected to centrifugation at 12000 rpm for 4 minutes, and the collected supernatant was stored for later testing. To determine the amount of protein in the sample extract, the original supernatant was dissolved in phosphate buffer (pH-7), and then the Bradford reagent was mixed with the aforementioned solution. The mixture was set alone for a few minutes to allow the red dye to attach to the protein and turn blue. The optical density of the sample was measured at 595 nm. The assay's standard was made with bovine serum albumin (BSA).

c: PHENOL ESTIMATION

The leaf samples were homogenised in 80 % ethanol, and the phenol concentration was calculated using the supernatant. The supernatant was then mixed with folin-Ciocalteau reagent and allowed to sit for five minutes. The aforementioned solution was given a 20 % Na_2CO_3 boost after five minutes. The sample's optical density was measured at 650 nm. The assay's standard was made with gallic acid.

d: LIPID ESTIMATION

The leaf samples were homogenised in a solution of 0.25 percent thiobarbituric acid in 10 % Trichloroacetic acid, which was gestated at 95°C for 15 minutes before being chilled on ice. At 600 nm and 532 nm, the absorbance of the supernatant

was measured. after centrifuging the homogenate for 10 minutes at 10000 rpm. Using 155 mM^{-1} cm^{-1} as the extinction value, the degree of lipid peroxidation was measured in micromole MDA per gram fresh weight of the sample. The malondialdehyde (MDA) was utilised as the assay's standard.

e: PROLINE ESTIMATION

The proline amino acid was utilized as the standard for this experiment. The leaves were homogenised with 3% sulpho salicylic acid and subjected to centrifugation at 10000 rpm for 20 minutes at room temperature. The supernatant was removed and added with orthophosphoric acid and glacial acetic acid to make a solution. The aforementioned solution was then incubated for one hour in a hot water bath (100°C) before being cooled on ice. In addition, 4 ml toluene reagent was added to the aforementioned solution and vortexed for a minute. Finally, the sample's optical density was measured at 520 nm.

f: CHLOROPHYLL ESTIMATION

The leaf samples were crushed at room temperature in 80% acetone, then homogenised at 5000 rpm in a centrifuge for 5 minutes at 4°C. Then the supernatant was separated into its own test tube, and the remaining leaves were vortexed once again with 80 percent acetone. The aforementioned solution was then centrifuged for 5 minutes at 4°C at 5000 rpm. This method was carried out several times until the residue became colourless. Finally, using 80% acetone as a blank, the optical density of the collected test tube solution was measured at 645, 663, and 652 nm.

g: HYDROGEN PEROXIDE ESTIMATION

After homogenising the leaf samples in 0.1% Trichloroacetic acid (w/v), they were centrifuged at 15000 rpm for 20 minutes. The extracted supernatant was then combined with 0.1M phosphate buffer and 1M potassium iodide reagent. The solution was then incubated for one hour at room temperature in the dark. Finally, at 390 nm, the optical density of the aforementioned solution was calculated.

h: SODIUM DISMUTASE ESTIMATION

The leaves were crushed in phosphate buffer (pH 7.0) having 0.05 % Triton-x-100, 1 mM Ascorbic acid, 1 mM EDTA (ethylene diamine tetra acetic acid), and 2 % PVP (Polyvinyl Pyrollidine), and centrifuged at 12000 rpm and 4°C. The collected supernatant was then used in the subsequent testing. To 2.9 ml of reaction mixture containing 75 M nitro blue tetrazolium, 13 M methionine, 0.1 MmM EDTA, and 16.7 M riboflavin, 0.1 ml of enzyme extract was added. After 20 minutes of incubation under 15-watt fluorescent lamps, for another 15 minutes, the test tubes were left in the dark. When the incubation period was finished, the absorbance was measured at 560 nm.

The biochemical data of sugarcane crop calculated using these optimized protocols as explained above was logged and further used in the development of machine learning-based method for retrieval of biochemical data of sugarcane crop using the sentinel-2 derived vegetation indices as testing and training data set and also for the validation of the developed model.

2) EXTRACTION OF VEGETATION INDICES USING SENTINEL-2 DATA

This research employed a total of 14 vegetation indices including NDVI, SAVI, PVI, IPVI, MSAVI, MSAVI2, DVI, GNDVI, SLAVI, NDI45, IRECI, MTCI, PSSRa, and S2REP have been extracted using Sentinel-2 data [71]. Vegetation indices are derived using spectral reflectance of vegetation at two or more bands acquired using optical sensors and highlighting a particular canopy feature/property [72], [73]. The satellite-derived vegetation indices are reported to be in good agreement with several physiological and photosynthetic variables of the crops [74]. Scientists have created a variety of vegetation indexes for diverse applications. Table 3 provides the list of various vegetation indices taken from the literature along with their respective formulae and significance. After successful extraction of vegetation indices using the sentinel-2 data. This information is converted into the form of digital number values and logged for further analysis.

3) MODEL DEVELOPMENT

Satellite data-derived vegetation indices offer an effective solution for monitoring crop biochemical parameters at a large scale in real-time. However, as it is can be seen from Table 3, over the past four decades, several researchers have proposed numerous vegetation indices that have been utilised in a variety of applications and have been discovered to be sensitive to various plant factors such as vegetation water content, leaf area index, chlorophyll content, nitrogen, and phosphorous content and so on. [46], [47], [52], [55], [56]. The rigorous sensitivity analysis between the derived vegetation indices and biochemical parameters showed that the relationship between the two is quite complex and we need to toil out such approaches which are capable of predicting the complex relationship between these variables [40]. Using all of the obtained vegetation indices to model the link between biochemical plant factors and vegetation indices would also be complicated and time-consuming. Hence, there is a need to obtain the optimal indices that have the best discriminatory power against each of the considered biochemical parameters of the crop using separability index analysis.

Furthermore, machine learning (ML) algorithms provide flexible parameter estimate tools and can understand the underlying intricate nonlinear relationship among vegetation indices and crop biochemical parameters by fitting a flexible yet robust model to the training data [63], [82]. The machine learning model may be used to assess the parameters of any unknown dataset once it has been properly trained. In the recent past, the use of non-parametric kernel- based methods is on the rise as they do not make prior assumptions about the data set. Gaussian process regression (GPR) is a

TABLE 3. List of the vegetation indices that were employed in this study. And derived using sentinel-2 data. The term R stands for the reflectance of different sentinel-2 bands, which refers to the wavelength mentioned with it.

Vegetation Index	Algorithm	Significance	Reference
Normalized Difference Vegetation Index (NDVI)	$(R_{842}-R_{665})/(R_{842}+R_{665})$	Used in research such as canopy structure, leaf area index, and canopy photosynthesis.	[76]
Difference Vegetation Index (DVI)	$(R_{842}-R_{665})$	Used to track the vegetation's biological environment because it is sensitive to changes in soil backdrop. The Environmental Vegetation Index is	[77]
		another name for it (EVI).	
Perpendicular Vegetation Index (PVI)	$(-\sin 45(R_{842})\cos 45(R_{665}))$	Used to invert surface vegetation characteristics (chlorophyll content, grass yield) as well as to calculate LAI, identification, and classification.	[77]
Infrared Percentage Vegetation Index (IPVI)	$(R_{842}/R_{842}+R_{665})$	The normalised difference vegetation index is functionally and linearly equal to this index. It has the advantage of being both computationally faster and never being negative.	[71]
Soil Adjusted Vegetation Index (SAVI)	$[(R_{842}-R_{665})/(R_{842}+R_{665}+0.5)]*(1+0.5)$	Designed by combining the soil conditioning index (L), which runs from 0 to 1, to improve the NDVI's sensitivity to soil background.	[78]
Modified Soil Adjusted Vegetation Index (MSAVI)	$(R_{842}+0.5-(0.5*\text{ sqrt}((2*R_{842}+1)\text{^}2-8*(R_{842}(2*R_{665})))))$	MSAVI minimizes the impact of bare soil on SAVI	[79]
Modified Soil Adjusted Vegetation Index-2 (MSAVI2)	$(0.5*((2*(R_{842}+1))-(((2*R_{842})+1)\text{^}2-8(R_{842}-R_{665}))\text{^}0.5))$	Employed in desertification research, drought monitoring, grassland yield estimation, leaf area index assessment, plant growth analysis, soil organic matter analysis, and soil erosion analysis.	[79]
Green Normalized Difference Vegetation Index (GNDVI)	$(R_{842}-R_{560})/(R_{842}+R_{560})$	This measure is a tweaked version of NDVI that is more sensitive to crop chlorophyll concentration variations.	[48]
Specific Leaf Area Vegetation Index (SLAVI)	$(R_{842}/(R_{665}+R_{1610}))$	Useful for calculating the specific leaf area of forest vegetation cover.	[81]
Inverted Red-Edge Vegetation Index (IRECI)	$(R_{842}-R_{665})/(R_{705}/R_{740})$	The biophysical factors of trees in the forest can be estimated using this index.	[17]
MERIS Terrestrial Chlorophyll Index (MTCI)	$(R_{842}-R_{705})/(R_{705}-R_{665})$	Created with Merris datasets in mind to quantify chlorophyll content.	[75]
Normalized Vegetation Index-45 (NDI-45)	$(R_{705}-R_{665})/(R_{705}+R_{665})$	Enables comparisons of terrestrial photosynthetic activity and canopy structure alterations over time and space.	[16]
Pigment Specific Simple Ratio (PSSRa)	R_{842}/R_{665}	Strong association with the concentration of the specific pigment.	[42]
Sentinel-2 Red Edge Position (S2REP)	$R_{705}+35*((((R_{842}+R_{665})/2)-R_{665})/(R_{740}-R_{705}))$	Association with leaf chlorophyll content retrieval.	[17]

sophisticated kernel-based machine learning (ML) regression algorithm that has been made use of to successfully estimate plant biophysical characteristics using satellite data [61], [62]. As a result, we investigated the use of machine learning approaches to spectrally assess crop biochemical variables using satellite-derived vegetation indicators in this research. A two-step procedure is carried out in implementing the proposed machine learning model given in figure 3. The detailed methodology carried out in implementing each step is given below.

Step 1 Separability Index Analysis:

The sentinel-2 data yielded a total of 14 vegetation indices, which are listed in Table 3. A statistical scale-based separability index (SI) study is cast-off to find the optimum vegetation indices that best characterize the biochemical Variables. SI analysis is a distance metric that uses the statistical mean and standard deviation to order vegetation indices by distance. The main idea behind adopting SI measure is that it can provide the vegetation indices that are having the best discriminative power to the corresponding plant parameter. SI of all the vegetation indices (i) and the plant

parameters (j) is obtained as given in (1) [82].

$$SI_{ij} = \left| \frac{|\mu_i - \mu_j}{\sigma_i + \sigma_j} \right| \tag{1}$$

where $^{\mu}$ and $^{\sigma}$ are the mean and standard deviation, respectively. Any SI value larger than 0.8 implies that the vegetation index in question has a good chance of identifying the plant parameter in question.

Step 2 Gaussian Process Regression:

Satellite data-derived vegetation indices offer an effective solution for monitoring Gaussian process regression (GPR) is a sophisticated machine learning model that performs well on small datasets and is based on the Bayesian method. As with all Bayesian methods, it begins with a prior distribution (Gaussian) on the latent target function and predicts the target function values by combining the prior and Gaussian likelihood distribution (noise) in the form of the posterior distribution [58]. Due to the inherently probabilistic nature of the method, it provides predictions as well as confidence levels of those predictions making the model more reliable. The mathematical implementation of the GPR model is discussed below:

a: DATA SET

We have a data set consisting of predictor- response pairs $D = \{(x^i, y^i)_{i=1}^N\} = \{X, y\}$; where the predictor variables $X \in ¡^N$ contain N measurements of the vegetation indices obtained from satellite data and the response variable $y \in ¡$ contains the corresponding N values of the plant parameters. The data set is split into training (X) and testing (X_*) subsets. The goal is to identify functional relationship $f : ¡^N \rightarrow ¡$ between predictor-response pairs such that

$$y = f(x) + \varepsilon \tag{2}$$

where $^{\varepsilon}$ is the inherent noise existing in the observations which are supposed to be independently identically distributed Gaussian additive noise with zero mean and distribution over all available functions that match the data to discover the fitted function.

b: GAUSSIAN PROCESS PRIOR

According to GP prior, the training and testing data follow joint multivariate Gaussian distribution well-defined by a mean vector m(X) and also a covariance matrix K.

$$\begin{bmatrix} f(X) \\ f(X_*) \end{bmatrix} \sim N' \left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right) \tag{3}$$

In GPR, assume that the function f(x) is Gaussian distributed and defined as:

$$f(x) \sim GP\left(m(x), k\left(x, x'\right) + \sigma_n^2 I\right) \tag{4}$$

where m(x)=E[f(x)] is the mean function that reflects the average of all functions in the distribution evaluated at 'x'. The prior mean is always set to zero for simplicity. k(x, x') = E[(f(x)-m(x))(f(x')-m(x'))] is the covariance matrix, often accepted as the kernel function, that describes the influence of one point on another. The kernel describes the smoothness of a function in the distribution. I is the identity matrix.

c: MODEL SELECTION

Fitting a GPR model involves estimating the mean function and kernel parameters. The mean is usually considered either zero or a constant equal to the average of all the training data points. Though several kernels are available in the literature, a squared exponential kernel is the most popular kernel used in GP modelling and is accepted as [84]:

$$k\left(x, x'\right) = \sigma^2 \exp\left(-\frac{1}{2l^2}\left(x - x'\right)^T \left(x - x'\right)\right) \tag{5}$$

where, two main parameters of the kernel function are the signal variance ($^{\sigma}$) and the length scale (l), respectively.

d: HYPER PARAMETER TUNING

The hyper parameters of the kernel, which influence the form of the covariance function and vary the confidence of the prediction, are a key stage in GPR. hyper parameters $\theta : (l, \sigma^2)$ in the covariance matrix are estimated by getting the most out of the log- likelihood function:

$$P(y/X, \theta) = -\frac{1}{2}\log|k| - \frac{1}{2}y^T k^{-1} y - \frac{1}{2}\log 2\pi \tag{6}$$

where k is the covariance matrix. To determine the hyper parameters that optimize the aforementioned function, the gradient-based technique is used. variance N(0, σ^2). For a given set of predictor-response values, there are potentially infinitely many functions that fit the data. The purpose of GPR is to calculate the probability

e: JOINT PROBABILITY DISTRIBUTION

The joint probability distribution spans the space of possible function values for the function that we need to forecast. The joint distribution of the training f(X) and testing f(X*) outputs, according to the previous, is

$$\begin{bmatrix} f(X) \\ f(X_*) \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \tag{7}$$

where $K(X, X)$ and $K(X_*, X_*)$ are the covariance matrices for all training and testing data, respectively; $K(X, X_*)$ is the covariance matrix for training and testing data, $K(X_*, X)$ is the covariance matrix for testing and training data, σ_n^2 is the noise variance and I is the identity matrix.

f: PREDICTIVE POSTERIOR DISTRIBUTION

To perform regression (predictive distribution), condition the joint prior distribution to get the functions that agree with the training data. After conditioning, the obtained analytical equations are:

$$f_* \mid X, y, X_* \sim N\left(\bar{f}_*, \Sigma(f_*)\right) \tag{8}$$

TABLE 4. Separability index of considered vegetation indices with each of the biochemical variables of sugarcane crop where; carbs = carbohydrate, H_2O_2 = Hydrogen Peroxide, chloro = chlorophyll, and SoD = sodiumdismutase.

SI stats	NDVI	SAVI	DVI	IPVI	MSAVI2	GNDVI	SLAVI	NDI45	MSAVI	IRECI	MTCI	PSSRA	S2REP	PVI
Carbs	0.09	0.51	0.81	0.83	0.66	0.1	0.62	0.52	0.56	0.82	0.55	0.66	0.61	0.9
Phenol	0.08	0.79	1.17	0.74	0.55	0.08	0.52	0.81	0.85	1.18	0.8	1	0.6	1.29
Lipid	0.31	0.25	0.52	1.06	0.89	0.34	0.82	0.25	0.29	0.52	0.32	0.38	0.64	0.6
Proline	0.35	0.41	0.8	1.4	1.13	0.39	1.01	0.41	0.46	0.81	0.47	0.6	0.66	0.93
Protein	0.58	0.01	0.27	1.45	1.24	0.63	1.12	0.01	0.04	0.26	0.1	0.11	0.68	0.34
H_2O_2	0.94	0.33	0.04	2.01	1.73	1.03	1.54	0.34	0.28	0.06	0.16	0.23	0.73	0.02
Chloro	0.38	0.16	0.42	1.13	0.96	0.42	0.88	0.16	0.2	0.41	0.24	0.28	0.65	0.49
SOD	0.19	0.48	0.83	1.07	0.86	0.22	0.79	0.49	0.53	0.84	0.53	0.66	0.63	0.94

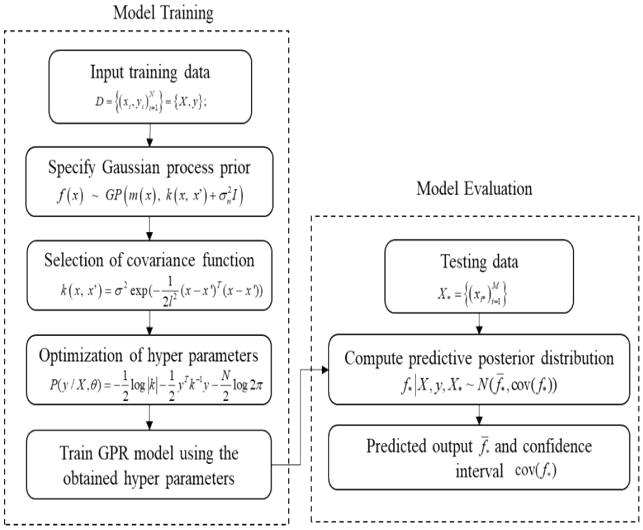


FIGURE 4. Flow chart showing the development of the GPR model.

$$f_* \triangleq E\left[f_* | X, y, X_*\right] = K(X_*, X)\left[K(X, X) + \sigma_n^2 I\right]^{-1} y \tag{9}$$

$$\Sigma\left(f_*\right) = K\left(X_*, X_*\right) - K\left(X_*, X\right)\left[K(X, X) + \sigma_n^2 I\right]^{-1} K(X, X_*) \tag{10}$$

Function values at any new test data f_* can be predicted by evaluating the mean function $\bar{f}_*$. The prediction's confidence interval is then calculated using $\Sigma(f_*)$ as given in equations 9 and 10. A flow chart depicting the mathematical implementation involved in GPR is given in Figure 4.

III. RESULTS AND DISCUSSION

A. SEPARABILITY INDEX ANALYSIS

To obtain the vegetation indices that are having the best discriminative power to the corresponding plant parameter, SI analysis has been carried out as described in section C(3). The mean and standard deviation of all vegetation indicators, as well as the associated plant biochemical Variables, were calculated at first. Then, using the mean and standard deviation numbers as a starting point, the separability index of each vegetation index with each of the plant parameters has been calculated using equation 1. Table 4 shows the obtained SI of all the biochemical variables with all the considered vegetation indices. The results show that different vegetation indices are having different separability indexes with each biochemical variable. Since the biochemical plant Variables are organic compounds in nature, the absorption, reflections, and transmittance characteristics are different for each vegetation index, resulting in variable sensitivity of each parameter. Vegetation indices having a separability index greater than 0.8 are considered to be having good potential to model the corresponding biochemical Variable and are used for further analysis. The SI values of any biochemical variable-vegetation index pair greater than 0.8 have been marked in bold. As clearly seen in the table, the vegetation index IPVI is having SI $>$ 0.8 for all the considered biochemical Variables except phenol.

Similarly, MSLAVI2 is also having SI $>$ 0.8 for a majority of the considered biochemical Variables except for phenol and carbs. It is evident from the tabulated results (Table 4) that, the vegetation indices IPVI, MSAVI2, and SLAVI are the most influential candidates for the considered biochemical Variables followed by DVI, PVI, and IRECI. All the vegetation indices that are having either SI $>$ 1.0 or the top three vegetation indices that are having SI $>$ 0.8 are considered to model each of the biochemical Variables. A total of seven vegetation indices have been identified that best model the considered biochemical plant Variables. Table 5 provides the list of vegetation indices considered for each biochemical Variable for further analysis. Further, these vegetation indices are cast-off to train a machine learning model to spectrally assess the biochemical plant Variables.

B. GAUSSIAN PROCESS REGRESSION

The link between biochemical variables and satellite- derived vegetation indices was modelled using laboratory- tested biochemical variables of sugarcane crop samples obtained at different growth stages (on the dates listed in Table 1). In this study, we divided the available data into two sets: training and testing. Of the available samples, 80% of samples have been used for training the GPR model while the rest 20% have been used for validating the trained model. The standard GPR

TABLE 5. List of the combination of satellite-derived vegetation indices considered for each biochemical Variable as obtained from SI analysis.

Biochemical variable of plant	Considered vegetation indices (SI > 0.8)
Carbohydrate	PVI, IRECI, IPVI
Phenol	DVI, PVI, IRCEI
Lipid	IPVI, MSAVI2, SLAVI
Proline	IPVI, MSAVI2, SLAVI
Protein	IPVI, MSAVI2, SLAVI
Hydrogen Peroxide	IPVI, MSAVI2, SLAVI, GNDVI
Chlorophyll	IPVI, MSAVI2, SLAVI
Sodium Dismutase	IPVI, PVI, MSAVI2

formulation described in section C(3) is used for spectrally estimating the plant biochemical Variables.

A fitted GP regression model that takes satellite- derived vegetation indices as input and returns the plant biochemical Variables as the output is developed with the optimized kernel parameters using the chosen GP prior and likelihood function as stated in the preceding section. The hyper parameters of the chosen kernel are optimized by getting the most out of the log-likelihood function given in equation (6). To prevent having to deal with the issue of overfitting, the GPR model is trained to minimize five-fold cross-validation loss and obtained the model that best fits the given vegetation indices and the biochemical plant Variables. The generated model's performance is then validated using a new data set after it has been trained (testing data).

The obtained regression plots that show the correlation of the predicted plant biochemical Variables with their laboratory-tested values for both training and testing data as obtained from the GPR model have been shown in Figures 5 and 6. Figures 5(a)-(h) represents the regression fit plots of the GPR model obtained from the training data set for the biochemical plant Variables Carbohydrate, Hydrogen Peroxide, Chlorophyll, Lipid, Phenol, Proline, Protein, and Sodium Dismutase respectively.

A well-fitted regression model provides the predicted values close to the true observed values. It can be seen from the goodness of fit plots as obtained from the training data (Figure 5(a)-(h)), that the trained GPR model is well capable to comprehend the underlying complex relationships amongst the considered vegetation indices and the corresponding biochemical Variables. Similarly, Figures 5(a)-(h) represent the regression fit plots of the trained GPR model obtained from the testing data set for the biochemical Variables Carbohydrate, Hydrogen Peroxide, Chlorophyll, Lipid, Phenol, Proline, Protein, and Sodium Dismutase, respectively. The visual analysis of the regression plots of the testing data set (Figures 6(a)-(h)) indicates that the predicted plant biochemical Variables from the developed GPR model are pretty much close to the observed values even on the new data set.

To evaluate the developed regression model, widely used accuracy metrics regression loss (L), mean absolute error (MAE), root mean square error (*RMSE*), normalized root mean square error (N*RMSE*), and coefficient of determination (R^2) have been calculated as given by the equations 11 to 15 [84]:

$$L = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{13}$$

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_{\cdot} - y\right)^2} \tag{15}$$

where y_i, $\hat{y}_i$ are respectively the real and forecast values of ith sample, $\bar{y}$ is the mean value of y_i, $y_{\min}$ and $y_{\max}$ are the minimum and maximum values y_i and n are the number of testing samples.

Regression loss estimates the loss in the predictions of GPR model based on the predicted and laboratory-tested values. Since the goal of any regression model is to reduce the loss, any value close to zero indicates a better model. It can be seen from Table 6 that the regression loss of all the developed GPR models except for SoD is close to zero indicating that the models are quite good. The reason for higher regression loss for SoD is due to the fact that the data ranges of SoD are relatively higher compared to other biochemical Variables. Also squaring the error between the actual and predicted values in regression loss inflates the average error when there are any large errors or outliers. In such cases, MAE performs better than L. The average absolute error between actual and expected biochemical plant Variables is termed MAE. The calculated MAE values of all the GPR predicted biochemical Variables (Table 6) are well below the average values of the biochemical Variables given in Table 7. The computed MAE of SoD (1.65) can also be seen which is quite less than the regression loss observed for SoD (4.0146). Another widely used performance metric to evaluate the goodness of regression fit is RMSE, which is an extension of regression loss. The standard deviation of the error in the model projected values is measured by the RMSE. Regression loss L is used to train a regression model while RMSE is used to evaluate the model. The ideal value of RMSE is zero indicating that all the predicted values matched well with the actual values. The RMSE observed for all the biochemical Variables from both the training and testing datasets is shown in Table 6. The observed RMSE of all the models on training data ($RMSE_{train}$) is pretty

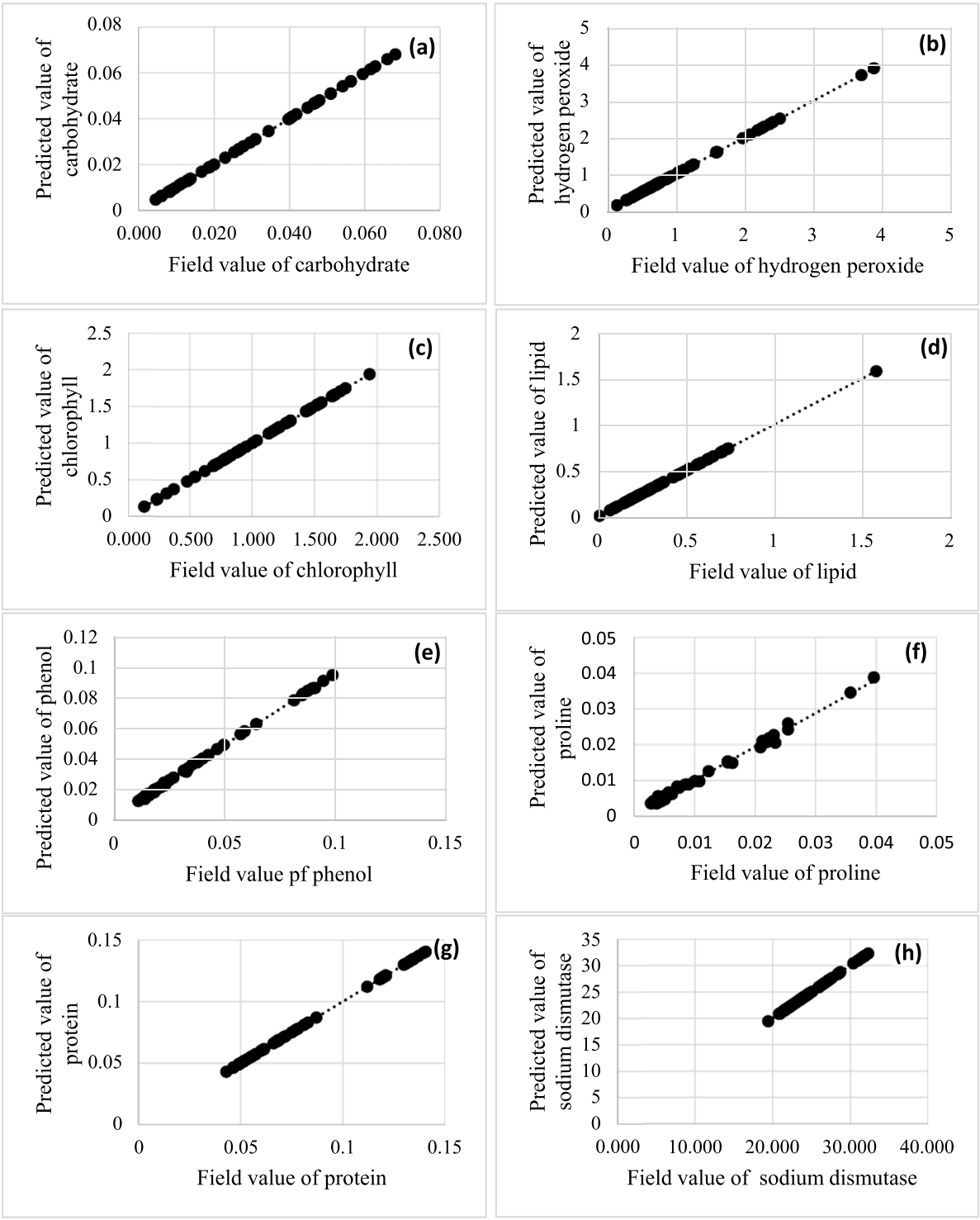


FIGURE 5. Regression fit plots obtained from the training data set (a) Carbohydrate (b) Hydrogen Peroxide (c) Chlorophyll (d) Lipid (e) Phenol (f) Proline (g) Protein, and (h) Sodium Dismutase.

much close to zero indicating that the underlying intricate relationship between biochemical Variables and vegetative indices is well modelled by trained models. Similarly, the observed RMSE of all the models on testing data ($RMSE_{test}$) is also close to zero except for SoD. Since RMSE measures the absolute errors on the same scale as the original data and the range of SoD are in the order of above 20, the value of the observed RMSE is also in the same order as the data range. However, the observed RMSE of SoD is well below the average value of SoD given in Table 7. Relatively higher RMSE is observed in the case of the test data set compared to the training data set. This could be due to, the total available data

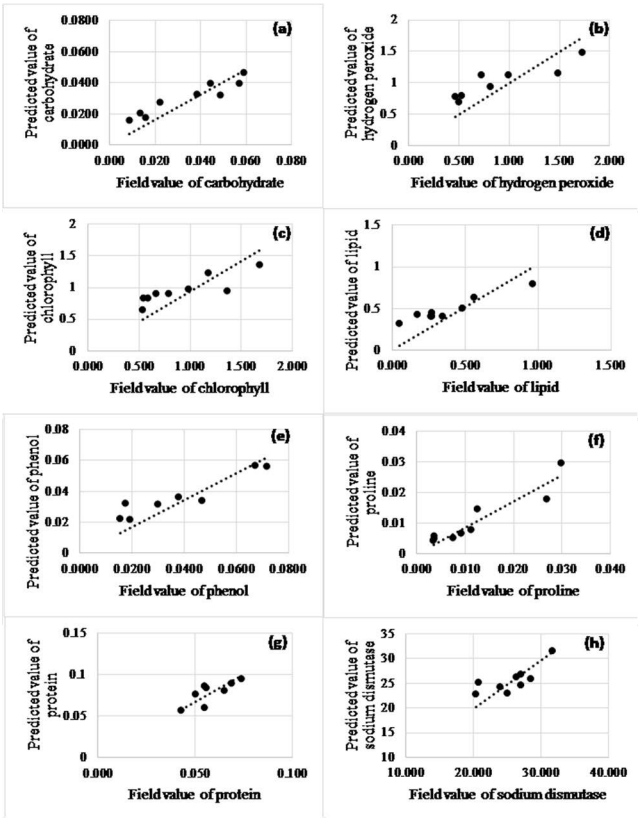


FIGURE 6. Regression fit plots obtained from the validation data set (a) Carbohydrate (b) Hydrogen Peroxide (c) Chlorophyll (d) Lipid (e) Phenol (f) Proline (g) Protein, and (h) Sodium Dismutase.

TABLE 6. Performance metrics of the trained GPR model.

Plant biochemical variable	Regression loss (L)	MAE_{test}	RMSE		$NRMSE_{test}$	R^2_{test}
			$RMSE_{train}$	$RMSE_{test}$		
Phenol	0.0001	0.0083	0.00168	0.01001	0.178	0.79
Lipid	0.0258	0.1332	0.00004	0.16053	0.192	0.67
Protein	0.0003	0.0144	0.00004	0.01607	0.172	0.73
Chlorophyll	0.0402	0.1541	0.00006	0.20048	0.174	0.77
Sodium Dismutase	4.0146	1.6500	0.23850	2.00364	0.188	0.61
Carbohydrate	0.0001	0.0087	0.00940	0.00340	0.067	0.74
Hydrogen Peroxide	0.0707	0.2484	0.00032	0.26597	0.210	0.65
Proline	0.0000	0.0027	0.00019	0.00364	0.137	0.86

TABLE 7. Comparison of the laboratory-measured and GPR estimated plant biochemical Variables statistics.

Plant biochemical variable	Mean value		Standard deviation	
	Laboratory measured	GPR estimated	Laboratory measured	GPR estimated
Phenol	0.040	0.037	0.02	0.01
Lipid	0.416	0.494	0.24	0.13
Protein	0.076	0.084	0.03	0.03
Chlorophyll	0.926	0.911	0.38	0.26
Sodium Dismutase	25.688	25.333	3.41	2.68
Carbohydrate	0.034	0.030	0.02	0.01
Hydrogen Peroxide	0.905	1.005	0.44	0.25
Proline	0.012	0.011	0.01	0.01

that consists of a wide range of plant biochemical Variables observed over the entire crop growth season of a sugarcane field being split into training and testing data sets. Because organic Compounds make up these biochemical variables, the extent of reflectance and absorption characteristics for each variable at different growth stages would be varied. making the trained GPR model sensitive to these changes. The accuracy metrics regression loss, MAE, and RMSE measure the loss on the same scale as the given data. However, the data ranges of the considered biochemical Variables are not the same. To compare the model fit of different models having different data ranges, normalized RMSE is used. As it can be seen from Table 6, the obtained NRMSE values for all the biochemical Variables are below 0.2 indicating that all the developed models can perform well on the validation data set. The lowest NRMSE of 0.067 is observed in the case of carbohydrates while the highest observed NRMSE of 0.2 is observed in the case of hydrogen peroxide. Another relative measure of model fit is the coefficient of determination, R^2 which is a measure of how well the predicted biochemical Variables obtained from the developed GPR model fit with the laboratory- tested values. R^2 value spans from 0 to 1. A value near 1 suggests an excellent match. All of the biochemical Variables have R^2 values ranging from 0.61 to 0.86 (Table 6), demonstrating that the proposed GPR model is capable of accurately predicting plant biochemical Variables from satellite-derived vegetative indices. proline with a R^2=*0.86,* had the highest coefficient of determination. R^2 for most of the considered biochemical plant Variables is above 0.7.

Further, the model's performance is also evaluated in terms of the laboratory tested in addition to the GPR anticipated mean and standard deviation values. Table 7 shows the results of a comparison of the mean and standard deviation of both the laboratory tested and GPR predicted biochemical Variables. It can be perceived from the tabulated values that the average values of all the biochemical Variables in the validation set are closely matching with the average values of the GPR predicted biochemical Variables. In the case of standard deviation, a similar pattern can be seen. The standard deviation observed with GPR projected values is slightly smaller than the standard deviation observed with laboratory testes data. Overall, it can be inferred from these results that, the developed GPR model is well capable of capturing the underlying complex relation between the plant biochemical Variables and the satellite-derived vegetation indices, and the biochemical Variables estimated from the GPR model are well within the ranges of the laboratory tested values, making it possible and paving the way for further research to spectrally assess biochemical plant Variables using satellite data.

IV. CONCLUSION

A Gaussian Process Regression model is created and implemented in this study for remote monitoring of sugarcane crops utilising Sentinel-2 and drone data. This model cannot only nullify the need of making regular field visits to keep track of crop health conditions but it is also environmental friendly as it dismissed the need for those laborious laboratory analyses which involve the use of many toxic and hazardous chemicals. It also provides insight into those areas of the crop fields which are otherwise inaccessible to the farmers. It offers the advantage of regular monitoring of the crop field without being to the field every day. Also if the crop is facing any kind of stress due to environmental conditions or due to the sudden exposure to harmful biotic factors like any weed, aphid can be detected before the onset of symptoms thus allowing for timely shielding of the crop. The machine learning technique that is Gaussian process regression is implemented for the development of the model for crop monitoring employing sentinel-2 derived vegetation indices as an input variable and biochemical variables of sugarcane crop as an output variable. The developed model has shown promising results with the R^2 value close to 1 for all the training models and R^2 value ranging from 0.61 to 0.86 for all the validation models thus holding high potential for monitoring the sugarcane crop remotely. The proposed methodology also has the potential to be extended for several other different crops.

REFERENCES

[1] A. Arora, R. K. Sairam, and G. C. Srivastava, "Oxidative stress and antioxidative system in plants," *Current Sci.*, vol. 82, no. 10, pp. 1227–1238, 2002.

[2] P. M. Christy, R. D. Preetha, S. Vasantha, and D. Divya, "Biochemical and molecular analysis of sugarcane genotypes response to salinity and drought," *Int. J. Appl. Biol. Pharmaceutical Technol.*, vol. 1, pp. 210–218, 2013.

[3] T. Jaiphong, J. Tominaga, K. Watanabe, M. Nakabaru, H. Takaragawa, R. Suwa, M. Ueno, and Y. Kawamitsu, "Effects of duration and combination of drought and flood conditions on leaf photosynthesis, growth and sugar content in sugarcane," *Plant Prod. Sci.*, vol. 19, no. 3, pp. 427–437, Jul. 2016.

[4] C. M. D. Santos and M. de Almeida Silva, "Physiological and biochemical responses of sugarcane to oxidative stress induced by water deficit and paraquat," *Acta Physiologiae Plantarum*, vol. 37, no. 8, pp. 1–14, Aug. 2015.

[5] L. Galon, G. Concenço, E. A. Ferreira, I. A. F. A. da Silva, C. L. Giacobbo, and A. Andres, "Influence of biotic and abiotic stress factors on physiological traits of sugarcane varieties," in *Photosynthesis*. London, U.K.: IntechOpen, 2013.

[6] K. K. Verma, X.-H. Liu, K.-C. Wu, R. K. Singh, Q.-Q. Song, M. K. Malviya, X.-P. Song, P. Singh, C. L. Verma, and Y.-R. Li, "The impact of silicon on photosynthetic and biochemical responses of sugarcane under different soil moisture levels," *Silicon*, vol. 12, no. 6, pp. 1355–1367, Jun. 2020.

[7] D. Kumar, N. Malik, and R. S. Sengar, "Physio-biochemical insights into sugarcane genotypes under water stress," *Biol. Rhythm Res.*, vol. 52, no. 1, pp. 92–115, Jan. 2021.

[8] E. Adam, O. Mutanga, and D. Rugege, "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review," *Wetlands Ecol. Manage.*, vol. 18, no. 3, pp. 281–296, Jun. 2010.

[9] A. Baez-Gonzalez, J. Kiniry, M. Meki, J. Williams, M. Alvarez-Cilva, J. Ramos-Gonzalez, A. Magallanes-Estala, and G. Zapata-Buenfil, "Crop parameters for modeling sugarcane under rainfed conditions in Mexico," *Sustainability*, vol. 9, no. 8, p. 1337, Jul. 2017.

[10] J. Baluja, M. P. Diago, P. Balda, R. Zorer, F. Meggio, F. Morales, and J. Tardaguila, "Assessment of vineyard water status variability by thermal and multispectral imagery using an unmanned aerial vehicle (UAV)," *Irrigation Sci.*, vol. 30, no. 6, pp. 511–522, 2012.

[11] G. S. Birth and G. R. McVey, "Measuring the color of growing turf with a reflectance spectrophotometer," *Agronomy J.*, vol. 60, no. 6, pp. 640–643, Nov. 1968.

[12] P. J. Curran, "Remote sensing of foliar chemistry," *Remote Sens. Environ.*, vol. 30, no. 3, pp. 271–278, 1989.

[13] P. J. Curran, J. L. Dungan, and D. L. Peterson, "Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: Testing the Kokaly and Clark methodologies," *Remote Sens. Environ.*, vol. 76, no. 3, pp. 349–359, 2001.

[14] C. S. T. Daughtry, C. L. Walthall, M. S. Kim, E. B. de Colstoun, and J. E. McMurtrey, III, "Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance," *Remote Sens. Environ.*, vol. 74, no. 2, pp. 229–239, 2000.

[15] G. Guyot and F. Baret, "Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux," in *Proc. Int. Colloq. Spectral Signatures Objects Remote Sens.*, vol. 287, 1988, p. 279.

[16] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors*, vol. 11, no. 7, pp. 7063–7081, 2011.

[17] W. J. Frampton, J. Dash, G. Watmough, and E. J. Milton, "Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation," *ISPRS J. Photogramm. Remote Sens.*, vol. 82, pp. 83–92, Aug. 2013.

[18] E. Panwar, A. Agarwal, D. Singh, and V. Pruthi, "An efficient application of satellite image for biochemical parameters study in sugarcane crop," in *Proc. Conf. Inf. Commun. Technol. (CICT)*, Oct. 2018, pp. 1–5.

[19] O. Rozenstein, N. Haymann, G. Kaplan, and J. Tanny, "Estimating cotton water consumption using a time series of Sentinel-2 imagery," *Agricult. Water Manage.*, vol. 207, pp. 44–52, Aug. 2018.

[20] A. Veloso, S. Mermoz, A. Bouvet, T. L. Toan, M. Planells, J.-F. Dejoux, and E. Ceschia, "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," *Remote Sens. Environ.*, vol. 199, pp. 415–426, 2017.

[21] W. A. Dorigo, R. Zurita-Milla, A. J. W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman, "A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 2, pp. 165–193, May 2007.

[22] T. Purevdorj, R. Tateishi, T. Ishiyama, and Y. Honda, "Relationships between percent vegetation cover and vegetation indices," *Int. J. Remote Sens.*, vol. 19, no. 18, pp. 3519–3535, Dec. 1998.

[23] A. A. Gitelson, A. Viña, T. J. Arkebauer, D. C. Rundquist, G. Keydan, and B. Leavitt, "Remote estimation of leaf area index and green leaf biomass in maize canopies," *Geophys. Res. Lett.*, vol. 30, no. 5, Mar. 2003, Art. no. 1248.

[24] A. A. Gitelson, Y. Gritz, and M. N. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," *J. Plant Physiol.*, vol. 160, no. 3, pp. 271–282, 2003.

[25] W. R. Raun, J. B. Solie, G. V. Johnson, M. L. Stone, E. V. Lukina, W. E. Thomason, and J. S. Schepers, "In-season prediction of potential grain yield in winter wheat using canopy reflectance," *Agronomy J.*, vol. 93, no. 1, pp. 131–138, Jan. 2001.

[26] Y. Xie, Z. Sha, and M. Yu, "Remote sensing imagery in vegetation mapping: A review," *J. Plant Ecol.*, vol. 1, no. 1, pp. 9–23, Mar. 2008.

[27] A. J. Norton, P. J. Rayner, E. N. Koffi, M. Scholze, J. D. Silver, and Y.-P. Wang, "Estimating global gross primary productivity using chlorophyll fluorescence and a data assimilation system with the BETHY-SCOPE model," *Biogeosciences*, vol. 16, no. 15, pp. 3069–3093, Aug. 2019.

[28] A. Lamsal, S. M. Welch, J. W. Jones, K. J. Boote, A. Asebedo, J. Crain, X. Wang, W. Boyer, A. Giri, E. Frink, X. Xu, G. Gundy, J. Ou, and P. G. Arachchige, "Efficient crop model parameter estimation and site characterization using large breeding trial data sets," *Agricult. Syst.*, vol. 157, pp. 170–184, Oct. 2017.

[29] F. Zhang and G. Zhou, "Estimation of canopy water content by means of hyperspectral indices based on drought stress gradient experiments of maize in the north plain China," *Remote Sens.*, vol. 7, no. 11, pp. 15203–15223, Nov. 2015.

[30] Y. Zhang, L. Guo, Y. Chen, T. Shi, M. Luo, Q. Ju, H. Zhang, and S. Wang, "Prediction of soil organic carbon based on Landsat 8 monthly NDVI data for the Jianghan plain in Hubei province, China," *Remote Sens.*, vol. 11, no. 14, p. 1683, Jul. 2019.

[31] R. A. Azevedo, R. F. Carvalho, M. C. Cia, and P. L. Gratão, "Sugarcane under pressure: An overview of biochemical and physiological studies of abiotic stress," *Tropical Plant Biol.*, vol. 4, no. 1, pp. 42–51, Mar. 2011.

[32] J.-K. Zhu, "Plant salt tolerance," *Trends Plant Sci.*, vol. 6, no. 2, pp. 66–71, Feb. 2001.

[33] A. K. Parida and A. B. Das, "Salt tolerance and salinity effects on plants: A review," *Ecotoxicol. Environ. Saf.*, vol. 60, no. 3, pp. 324–349, Mar. 2005.

[34] M. A. Sliva, J. A. G. Silva, J. L. Da Jifon, and V. Sharma, "Use of physiological parameters to detect differences in drought tolerance among sugarcane genotypes," in *Proc. Int. Soc. Sugar Cane Technol.*, vol. 26, 2007, pp. 541–547.

[35] P. Fu, K. Meacham-Hensold, K. Guan, J. Wu, and C. Bernacchi, "Estimating photosynthetic traits from reflectance spectra: A synthesis of spectral indices, numerical inversion, and partial least square regression," *Plant, Cell Environ.*, vol. 43, no. 5, pp. 1241–1258, May 2020.

[36] C. Li, Q. Nong, M. K. Solanki, Q. Liang, J. Xie, X. Liu, Y. Li, W. Wang, L. Yang, and Y. Li, "Differential expression profiles and pathways of genes in sugarcane leaf at elongation stage in response to drought stress," *Sci. Rep.*, vol. 6, no. 1, Sep. 2016, Art. no. 25698.

[37] B. Reddersen, T. Fricke, and M. Wachendorf, "Effects of sample preparation and measurement standardization on the NIRS calibration quality of nitrogen, ash and NDFom content in extensive experimental grassland biomass," *Animal Feed Sci. Technol.*, vol. 183, nos. 3–4, pp. 77–85, Jul. 2013.

[38] J. A. Lyn, M. H. Ramsey, R. J. Fussell, and R. Wood, "Measurement uncertainty from physical sample preparation: Estimation including systematic error," *Analyst*, vol. 128, no. 11, pp. 1391–1398, 2003.

[39] N. Dangwal, N. R. Patel, M. Kumari, and S. K. Saha, "Monitoring of water stress in wheat using multispectral indices derived from Landsat-TM," *Geocarto Int.*, vol. 31, no. 6, pp. 682–693, Jul. 2016.

[40] E. Panwar, D. Singh, and A. K. Sharma, "Exploring the possibility of assessing biochemical variables in sugarcane crop with Sentinel-2 data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 3712–3715.

[41] A. Bannari, D. Morin, F. Bonn, and A. R. Huete, "A review of vegetation indices," *Remote Sens. Rev.*, vol. 13, nos. 1–2, pp. 95–120, 1995.

[42] G. A. Blackburn, "Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 657–675, Jan. 1998.

[43] A. Damm, S. Cogliati, R. Colombo, L. Fritsche, A. Genangeli, L. Genesio, J. Hanus, A. Peressotti, P. Rademske, U. Rascher, D. Schuettemeyer, B. Siegmann, J. Sturm, and F. Miglietta, "Response times of remote sensing measured sun-induced chlorophyll fluorescence, surface temperature and vegetation indices to evolving soil water limitation in a crop canopy," *Remote Sens. Environ.*, vol. 273, May 2022, Art. no. 112957.

[44] P. Yang, W. Verhoef, and C. van der Tol, "The mSCOPE model: A simple adaptation to the SCOPE model to describe reflectance, fluorescence and photosynthesis of vertically heterogeneous canopies," *Remote Sens. Environ.*, vol. 201, pp. 1–11, Nov. 2017.

[45] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *J. Sensors*, vol. 2017, pp. 1–17, May 2017.

[46] I. Sandholt, K. Rasmussen, and J. Andersen, "A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status," *Remote Sens. Environ.*, vol. 79, nos. 2–3, pp. 213–224, Feb. 2002.

[47] R. R. Gillies, W. P. Kustas, and K. S. Humes, "A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the normalized difference vegetation index (NDVI) and surface e," *Int. J. Remote Sens.*, vol. 18, no. 15, pp. 3145–3166, Oct. 1997.

[48] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 289–298, 1996.

[49] L. Eklundh and L. Olsson, "Vegetation index trends for the African Sahel 1982–1999," *Geophys. Res. Lett.*, vol. 30, no. 8, Apr. 2003, Art. no. 1430.

[50] N. Pettorelli, S. Ryan, T. Mueller, N. Bunnefeld, B. Jedrzejewska, M. Lima, and K. Kausrud, "The normalized difference vegetation index (NDVI): Unforeseen successes in animal ecology," *Climate Res.*, vol. 46, no. 1, pp. 15–27, Jan. 2011.

[51] A. J. Peters, E. A. Walter-Shea, L. Ji, A. Vina, M. Hayes, and M. D. Svoboda, "Drought monitoring with NDVI-based standardized vegetation index," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 1, pp. 71–75, Jan. 2002.

[52] F. N. Kogan, "Application of vegetation index and brightness temperature for drought detection," *Adv. Space Res.*, vol. 15, no. 11, pp. 91–100, Jan. 1995.

[53] Z. Wan, P. Wang, and X. Li, "Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA," *Int. J. Remote Sens.*, vol. 25, no. 1, pp. 61–72, 2004.

[54] W. P. Kustas and J. M. Norman, "Use of remote sensing for evapotranspiration monitoring over land surfaces," *Hydrol. Sci. J.*, vol. 41, no. 4, pp. 495–516, 1996.

[55] D. Lloyd, "A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery," *Int. J. Remote Sens.*, vol. 11, no. 12, pp. 2269–2279, Dec. 1990.

[56] H. G. Jones, "Irrigation scheduling: Advantages and pitfalls of plant-based methods," *J. Exp. Botany*, vol. 55, no. 407, pp. 2427–2436, Sep. 2004.

[57] J. Shahmoradi, E. Talebi, P. Roghanchi, and M. Hassanalian, "A comprehensive review of applications of drone technology in the mining industry," *Drones*, vol. 4, no. 3, p. 34, Jul. 2020.

[58] H. Tao, H. Feng, L. Xu, M. Miao, H. Long, J. Yue, Z. Li, G. Yang, X. Yang, and L. Fan, "Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data," *Sensors*, vol. 20, no. 5, p. 1296, Feb. 2020.

[59] J. Verrelst, Z. Malenovský, C. Van der Tol, G. Camps-Valls, J.-P. Gastellu-Etchegorry, P. Lewis, P. North, and J. Moreno, "Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods," *Surv. Geophys.*, vol. 40, no. 3, pp. 589–629, May 2019.

[60] G. Camps-Vails, L. Gómez-Chova, J. Muñoz-Mari, J. Vila-Francés, J. Amoros, S. D. Valle-Tascon, and J. Calpe-Maravilla, "Biophysical parameter estimation with adaptive Gaussian processes," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 4, Jul. 2009, p. 69.

[61] V. Van Vaerenberg, M. Lázaro-Gredilla, and I. Santamaría, "Kernel recursive least-squares tracker for time-varying regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1313–1326, Aug. 2012.

[62] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.

[63] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, "Retrieval of vegetation biophysical parameters using Gaussian process techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1832–1843, May 2012.

[64] D. H. Svendsen, L. Martino, M. Campos-Taberner, F. J. Garcia-Haro, and G. Camps-Valls, "Joint Gaussian processes for biophysical parameter retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1718–1727, Mar. 2018.

[65] M. Najafzadeh and S. Niazmardi, "A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters," *Natural Resour. Res.*, vol. 30, no. 5, pp. 3761–3775, Oct. 2021.

[66] H. Metzner, H. Rau, and H. Senger, "Untersuchungen zur synchronisierbarkeit einzelner pigmentmangel-mutanten von chlorella," *Planta*, vol. 65, no. 2, pp. 186–194, 1965.

[67] L. S. Bates, R. P. Waldren, and I. D. Teare, "Rapid determination of free proline for water-stress studies," *Plant Soil*, vol. 39, no. 1, pp. 205–207, Aug. 1973.

[68] M. M. Bradford, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding," *Anal. Biochem.*, vol. 72, nos. 1–2, pp. 248–254, May 1976.

[69] V. L. Singleton, R. Orthofer, and R. M. Lamuela-Raventos, "Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent," *Methods Enzymol.*, vol. 299, pp. 152–178, Jan. 1999.

[70] W. E. Trevelyan, R. S. Forrest, and J. S. Harrison, "Determination of yeast carbohydrates with the anthrone reagent," *Nature*, vol. 170, no. 4328, pp. 626–627, Oct. 1952.

[71] R. E. Crippen, "Calculating the vegetation index faster," *Remote Sens. Environ.*, vol. 34, no. 1, pp. 71–73, 1990.

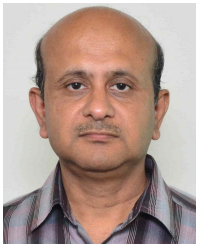
[72] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, May 1979.

[73] J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd ed. London, U.K.: Pearson, 2009.
[74] L. Leroux, C. Baron, B. Zoungrana, S. B. Traore, D. Lo Seen, and A. Begue, "Crop monitoring using vegetation and thermal indices for yield estimates: Case study of a rainfed cereal in semi-arid West Africa," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 347–362, Jan. 2016.
[75] J. Dash and P. J. Curran, "The MERIS terrestrial chlorophyll index," *Int. J. Remote Sens.*, vol. 25, no. 23, pp. 5403–5413, Dec. 2004.
[76] J. Rouse, Jr., R. Haas, J. Schell, and D. Deering, "Monitoring vegetation systems in the great plains with ERTS," 2007.
[77] A. J. Richardson and C. L. Wiegand, "Distinguishing vegetation from soil background information," *Photogramm. Eng. Remote Sens.*, vol. 43, no. 12, pp. 1541–1552, Dec. 1977.
[78] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, 1988.
[79] J. Qi, A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian, "A modified soil adjusted vegetation index," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 119–126, May 1994.
[80] B. Pinty and M. M. Verstraete, "GEMI: A non-linear index to monitor global vegetation from satellites," *Vegetatio*, vol. 101, no. 1, pp. 15–20, Jul. 1992.
[81] L. Lymburner, P. J. Beggs, and C. R. Jacobson, "Estimation of canopy-average surface-specific leaf area using Landsat TM data," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 2, pp. 183–192, 2000.
[82] A. N. J. Kukunuri, D. Murugan, and D. Singh, "A neural network approach to classify mixed classes using multi frequency SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 1747–1750.
[83] S. Krishnan, K. Samudravijaya, and P. V. S. Rao, "Feature selection for pattern classification with Gaussian mixture models: A new objective criterion," *Pattern Recognit. Lett.*, vol. 17, no. 8, pp. 803–809, Jul. 1996.
[84] C. E. Rasmussen, "Gaussian processes in machine learning," in *Proc. Summer School Mach. Learn.*, Berlin, Germany: Springer, 2003, pp. 63–71.
[85] J. Gareth, W. Daniela, H. Trevor, and T. Robert, *An Introduction to Statistical Learning: With Applications in R*. New York, NY, USA: Springer, 2013.

EKTA PANWAR received the bachelor's and master's degrees (Hons.) in biotechnology from Banasthali University, Rajasthan, India, in 2011 and 2013, respectively. She is currently pursuing the Ph.D. degree in environmental biotechnology with the Department of Biosciences and Bioengineering and the Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, India. Her research interests include precise agriculture monitoring, retrieval of biochemical parameters of crops using satellite data, monitoring biotic and abiotic stresses in crops, and assessing crop health conditions remotely. She has received fellowships from the Council of Scientific and Industrial Research, New Delhi, India, for her Ph.D. research work.

ANJANA NAGA JYOTHI KUKUNURI received the bachelor's degree in electronics and communication engineering from Andhra University, Visakhapatnam, India, in 2004, and the master's degree in communication systems from Graphic Era University, Dehradun, India, in 2014. She is currently pursuing the Ph.D. degree in RF and microwave engineering with the Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, India. Her research interests include multi-sensor satellite data analysis, developing machine-learning models for crop classification, crop parameter retrieval, and agriculture drought assessment. She has received fellowships from the Indian Space Research organization, India, and the Ministry of Human Resource and Development, Government of India, for her Ph.D. research work.

DHARMENDRA SINGH (Senior Member, IEEE) received the Ph.D. degree in electronics engineering from IIT (BHU) Varanasi, India. He was a Visiting Scientist Postdoctoral Fellow with the Department of Information Engineering, Niigata University, Niigata, Japan; the German Aerospace Center, Cologne, Germany; the Institute for National Research in Informatics and Automobile, France; the Institute of Remote Sensing Applications, Beijing, China; the Karlsruhe University, Karlsruhe, Germany; and the Polytechnic University of Catalonia, Barcelona, Spain. He is currently a Senior Professor with the Department of Electronics and Communication Engineering and the Coordinator of the Drone Research Centre, Indian Institute of Technology Roorkee, Roorkee, India. He has more than 25 years of experience in teaching and research. He has published more than 350 papers in various national/international journals and conferences. His research interests include the development of radar, microwave/millimeter/THz imaging, DRONE-based applications, antidrone, microwave, optical remote sensing, artificial intelligence, machine learning, the IoT, cloud-based solution, image processing/image analysis, image fusion, image enhancement, radar polarimetry and interferometry, stealth technology, and electromagnetic modeling. He has received various fellowships and awards from national and international bodies.



ASHWANI KUMAR SHARMA received the Ph.D. degree in biochemistry and biophysics from the All India Institute of Medical Sciences, New Delhi, India. He was a Postdoctoral Research Associate at the University of North Carolina at Chapel Hill, USA. He is currently a Professor with the Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee, Roorkee, India. He has more than 22 years of experience in teaching and research. He has published more than 100 papers in various reputed national/international journals and conferences. His research interests include biochemistry, biophysics, plant biotechnology, macromolecular crystallography, protein biochemistry, and structural biology.

HARISH KUMAR received the Bachelor of Engineering degree in computer science and engineering from Visveswaraiah Technological University, India, the Master of Technology degree in computer cognition technology from Mysore University, India, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology Roorkee. He did his postdoctoral research at the INRIA Bordeaux—Sud-Ouest, France. He has worked as the Chief Engineer at SEL, Samsung India Electronics Private Ltd., India. He is currently working as the University Professor with the Department of Computer Science, College of Computer Science, King Khalid University, Abha. He did his internship in the Samsung Global Internship Program 2008 at Samsung Electronics Company Ltd., Suwon, South Korea (he was one among two students from India, selected for the year 2008). He has several quality research papers in well-reputed journals. His research interests include augmented reality, virtual reality, security and privacy, HCI, steganography, machine learning, pattern recognition, natural language processing, computer vision, insider threats, fuzzy logic, deep learning, and soft computing. He has received the Young Scientist Award 2010 from the Second National Young Scientist Symposium—2010, Uttarakhand, India.