Arabian Journal of Geosciences (2022) 15:1567
https://doi.org/10.1007/s12517-022-10844-6

ORIGINAL PAPER

Growth and yield monitoring of potato crop using Sentinel-1 data through cloud computing

Chiranjit Singha[1] · Kishore Chandra Swain[1] · Hemantha Jayasuriya[2]

Received: 30 June 2021 / Accepted: 17 September 2022 / Published online: 26 September 2022

Abstract

Agricultural crops required continuous monitoring, to protect them from climatic hazards and unusual stresses ensuring better crop yield. Potato crop is very sensitive to water stress which affects both tuber quality and yield. The cloud-free, high-resolution, freely distributed Sentinel 1 SAR data was used for near real-time monitoring of potato (*Solanum tuberosum* L.) crop in West Bengal, India. The image analysis was carried out for various phonological stages of potato crop to evaluate VH and VV backscattering values in cloud-based Google Earth Engine platform (GEE) in 50 farm plots. The VH values were found more suitable for monitoring the crop at different growth stages soundly aligned with groundtruthing-based field photographs. Sentinel 1 data was validated against the optical sensor-based Sentinel 2-NDVI values along with time series model, showing positivity for field crop applications. Receiver operating characteristic (AUROC) curves along with the potato yield maps further enhanced the suitability of backscattered images for continuous monitoring of the crop even under overcast weather in the study area. In the process, better tuber quality and crop yield can be ensured, providing higher returns to the farmers.

Keywords Sentinel 1 · Cloud computing · NDVI · VV and VH · GEE · Potato crop

Introduction

Continuous depletion of agricultural land to commercialization, urbanization, and industrialization has put immense pressure on food security in India. Low soil fertility and change of climate due to environmental pollution etc. added new threats to sustainable agriculture in the region (Gerland et al. 2014). Additionally, in developing countries including India, the farming systems, diets, culture, and economies are mostly driven by low purchasing power priority and low quality of food resulting in high malnutrition rates (Garnett et al. 2013). So, there is an immediate need to improve the current cropping pattern, practices, and intensity to meet the new challenges in Asia, especially in India with a population of 1.3 billion mostly depending on a net cropping area of 130 million hactares (Agricoop 2017). Though agriculture contributes 17% of GDP, it has been the backbone of the Indian economy, and proved its importance time and again even during the pandemic. Monsoon-driven agriculture (around 50%) is also sometimes affected in patches due to disruptive weather.

Potato, one of the major staple crops in the world, is cultivated in hundred countries including India. The annual world potato production is around 3768.27 MT in 2018–2019 (Mithiya et al. 2019). According to International Food Policy Research Institute (IFPRI) and International Potato Centre (IPC, Lima Peru) reports, India has the highest growth rate of potato production and productivity from 1993 to 2020, among the Asian countries. Potato is grown as a cash crop in India, which has wide-scale demands in the regional and international markets (Singh et al. 2016), for preparing

Responsible Editor: Biswajeet Pradhan.

✉ Kishore Chandra Swain
kishore.swain@visva-bharati.ac.in

Chiranjit Singha
singha.chiranjit@gmail.com

Hemantha Jayasuriya
hemjay@squ.edu.om

[1] Department of Agricultural Engineering, Institute of Agriculture, Visva-Bharati (A Central University), Sriniketan, West Bengal 731236, India

[2] Department of Soils, Water & Agricultural Engineering; College of Agricultural & Marine Sciences, Sultan Qaboos University, Seeb, Oman

value-added products such as chips, crisps, peeled potato, canned potato, flour, starch, animal feed, and alcohol (Malit 2020) besides direct consumption. India ranked second after China with annual production of 437.70 MT, and among the Indian states, West Bengal produced nearly 110 MT (with 0.46 million ha cropping area) ranked behind Utter Pradesh followed by Bihar and Gujarat (Agricoop 2017).

Precision agriculture (PA), a technique of applying the right amounts of crop inputs at the right place, right time, and manner, has emerged as the best alternative to meet major global challenges, such as food security and the preservation of natural resources including soil fertility and environment (Gomez et al. 2019; Hunt et al., 2018; Rusia et al. 2018; Swain et al., 2020). Remote sensing can play an important role in facilitating precision agriculture application without carrying out frequent field visits (Shafi et al. 2019; Nigon et al. 2014). Optical remote sensing has been used for vegetation analysis including long-term land cover/land-use change analysis in agriculture and allied sectors (Atzberger 2013; Swain et al. 2007). However, the importance of this type of remote sensing is severely compromised for continuous monitoring of crop due to the presence of white and dark clouds during cropping seasons (Spasova and Nedkov 2019). Synthetic Aperture Radar (SAR) sensor data facilitates an all-weather imaging capability which can be very useful in several remote sensing applications. SAR sensors are capable to provide essential, reliable and timely, observations without cloud disturbance for applications such as agricultural monitoring (Steele-Dunne et al. 2017; Jia et al. 2012) and flood scouting (Pulvirenti et al. 2016) which has been recognized for many decades. Most of the SAR data need to be purchased with typical pricing of data, varies around USD 85–150/km^2 in the year 2017 (SSTL 2017), which is predicted to fall by two thirds over the coming decade. The higher cost involved has been limiting the use of SAR data only for research purposes, leaving a few real field applications. With the lunch of Sentinel-1 (A and B) satellite by European Space Agency in 2016, SAR data is now available for all applications without any cost. The imageries collected in various weather conditions including overcast ensure the consistency of data to meet the monitoring needs of farmers, food producers, and regulatory bodies including governments (Kuenzer and Knauer 2013).

However, unlike optical data, the SAR data generally requires extensive two-dimensional, space-variant signal analysis before an image is formed through a very high speed processor (Cumming and Bennett 1979). The SAR image processing is carried out by designing parallel algorithms which mainly based on shared high-performance computing (HPC) technique in a cloud platform (Zhang et al. 2016). Google Earth Engine (GEE) type platform can be used to analyze the SAR remote sensing images and to produce relevant information from remote sensing data (Gorelick et al. 2017). The GEE platform generally hosts images in the range of petabytes and provides a high-performance computing system with significant pre-processing facility and speed. Through image sources, it can directly access Sentinel-1 satellite images for near real-time analysis enabling continuous monitoring of agricultural crop (Shelestov et al. 2017). Singha et al. (2019a) showed the scope and potential of Sentinel-1 data along with the high-performance computing GEE platform, in facilitating quick and accurate mapping of large-scale rice production areas particularly belonging to the high-precipitation regions of Bangladesh. Khabbazan et al. (2019) further reported successful application of time series Sentinel-1 backscatter data for studying reflected variation in moisture content and vertical structure associated with phenological development of potato crops and its harvesting in The Netherlands. With these benefits, SAR data from Sentinel-1 satellite would be preferred over optical sensor data, for continuous monitoring of temporally inconsistent biological entity, such as agricultural crops (Arias et al. 2018; Shelestov et al. 2016), being severely affected by the weather. The technique has already been used for crop mapping and differentiating rice cultivars from other crops in China (Tian et al. 2018) and even in France (Bazzi et al. 2019).

As we know, the potato crop is very sensitive towards variation in soil moisture level, the right amount of water should be available for proper growth of the plant and its tuber formation (FAO 2008), particularly during tuber initiation and early tuber development stages (King and Stark 1997). Though there is good potato yield in the study area (Roy et al. 2019), the profit margin is somewhat reduced due to the production of low export-quality potatoes (Ghosal 2020) associated with low selling price during crop harvesting season (Pandey 2019). Additionally, post-harvest operations are sometimes hampered due to low-quality potato and the limited availability of cold storage in the region (Sarkar 2017). So, being a cash crop, its cultivation needs to be monitored to improve tuber quality and productivity through cloud-free images of SAR satellites using cloud computing platforms.

The crop monitoring is hampered due to the unavailability of cloud-free images from optical sensor-based satellites under overcast conditions and terrestrial rainfall. The current study is focused on utilizing freely distributed highly temporal cloud-free Sentinel-1 SAR data for continuous monitoring of potato crop, particularly during its key growing seasons to improve quality of potato tuber and its yield in Tarakeswar Block, Hooghly District, West Bengal, India.

Material and methods

Study area

The study was carried out in the Alluvial Deltatic Plain formed by the Ganga River on the Southern of West Bengal

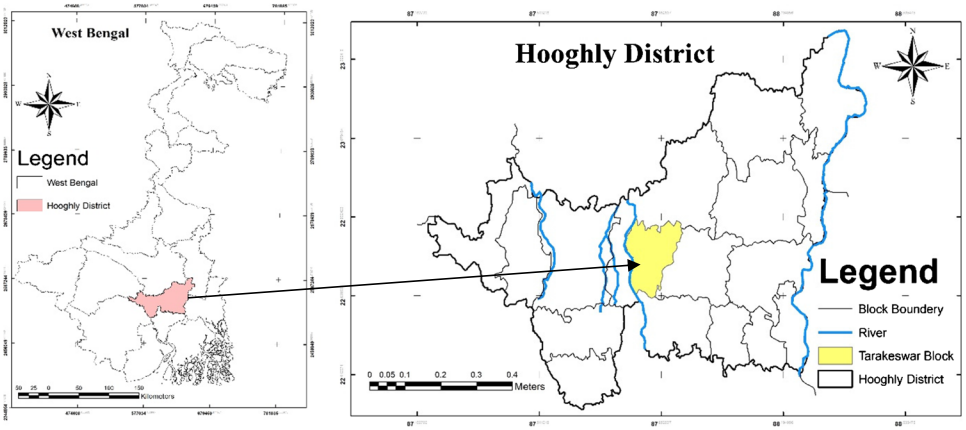


Fig. 1 Study area map of Tarakeswar Block, West Bengal

state, India. *Tarakeswar* block, Hooghly district of the state, was selected for the study, covering an area of 300 ha with alluvial type soil series. The study area has peripheral coordinates of 2,528,500 to 2,530,600°N latitude and 604,500 to 606,500°E longitude in Universal Transverse Mercator (UTM) Zone 45 of WGS 1984. The area has an elevation of the 40 m from mean sea level (Fig. 1). According to long-term weather data (IMD 2017) the climate in the region is characterized as tropical monsoon type, where monsoon contributes the majority of precipitation. The average temperature in July (summer) is 29 °C and in January (winter) is 15 °C with average annual precipitation of 1500 mm (Singha et al., 2019b). With the availability of irrigation, dual cropping system has been carried out in the study area. Major crops are rice and jute during summer (June–October) and, potato and lentil during winter season (November–April), in the study area (Singha et al. 2020). According to FAOSTAT (2016) potato is a principal winter crop in the study region, where production is doubled during the past decade (Ghosal 2020) mainly due to the horizontal growth of farm area. In West Bengal, the sowing time of potato is November–December. The harvesting of the crops is carried out around early March, but it sometimes starts in late January. Early-maturity potato varieties (*Jyoti* variety) with 75–90-day cropping type are found ideal for crop intensification without incurring major yield penalties (FAO 2008), soundly contributing to the enhancement of rural incomes.

Data pre-processing

Sentinel-1 image processing

The current study utilized the Sentinel-1A C-band Interferometric Wide (IW) mode of Level 1 processed product to backscatter coefficient ($\sigma°$) in decibels (dB). It is generally provided in dual-polarization, such as vertical transmit-vertical receive (VV) and vertical transmit-horizontal receive (VH). The spatial and temporal resolution of this imagery is 100 m^2 and 6-day time, respectively. The collection also includes the S1 Ground Range Detected (GRD) scenes, processed using the Sentinel-1 Toolbox to generate an ortho-corrected product through the cloud computing technique with Google Earth Engine (GEE) platform (https://sentinel.esa.int/web/sentinel/toolboxes/-1). The stages of Sentinel 1 image processing include thermal noise removal, radiometric calibration, and terrain correction using Shuttle Radar Topography Mission (SRTM). The final terrain-corrected values are converted to decibels via log scaling ($10 \times \log10(x)$). The image is filtered for backscatter time series using the multitemporal speckle filter, to reduce the influence of local environmental conditions and also to remove noises existing in the Sentinel-1 datasets due to speckle over the potato field during the different phenological growth stages of potato crop.

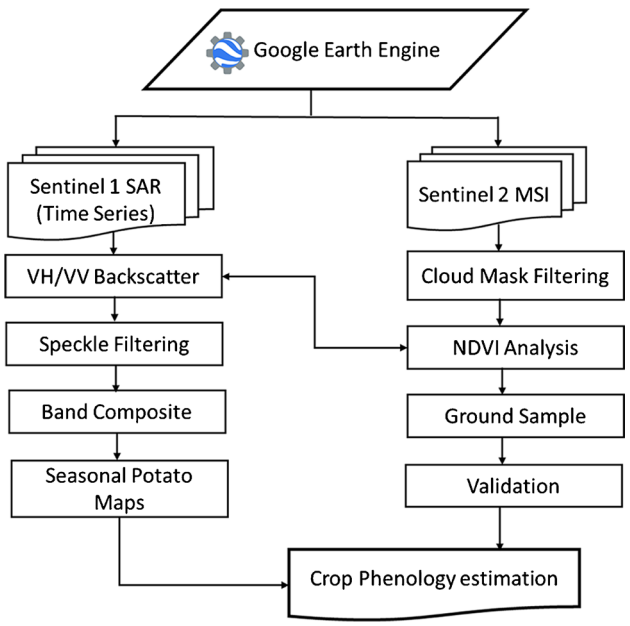


Fig. 2 Schematic diagram of workflow in this study

Table 1 Parameters tuning used by each modeling phase involves applying different machine learning techniques to validate potato yield

Model	Hyper-parameter	Optimization
SVM	Kernel	"rbf"
	C	1
	Gamma	2
	Degree	3
LR	C	1
	Intercept	24.05
	Solver	"lbfgs"
RF	Max_features	Auto
	n_estimators	10
	Criterion	"gini"
	Max depth	5
NB	var_smoothing	1e − 09

Sentinel-2 image processing

In the study, optical data Sentinel-2B MSI (Multi-Spectral Instrument), Level-1C images were used for Normalized Differential Vegetation Index (NDVI) estimation, analysis, and validation. All the images for the study were acquired from the European Union/(European Spatial Agency) ESA/Copernicus at the system through the GEE platform. Sensing orbit direction is set for descending mode with orbit number 33, programmed to cloud mask from built-in quality band filtering (Radoux et al. 2016). Data were collected during 1–20 February 2020, which is the peak growing period of the crop in the study area. From Sentinel-2 data, the surface reflectance values were obtained for the Red bands (band 4) and Infrared bands (band 8) and used to calculate the NDVI with a spatial resolution of 10 m.

Extraction of potato areas with groundtruthing

During the field visit, random 50 farm plots were selected. Field information such as soil nutrient, crop data, irrigation frequency, fertilizer, pesticide, and mechanization level for each growing season with groundtruthing in GPS point coordinate system were collected.

The whole potato plant cultivation period is divided into five growth stages, namely, sprouting, vegetative stage, tuber initiation, tuber bulking, and maturation stage (Thorton, 2020).

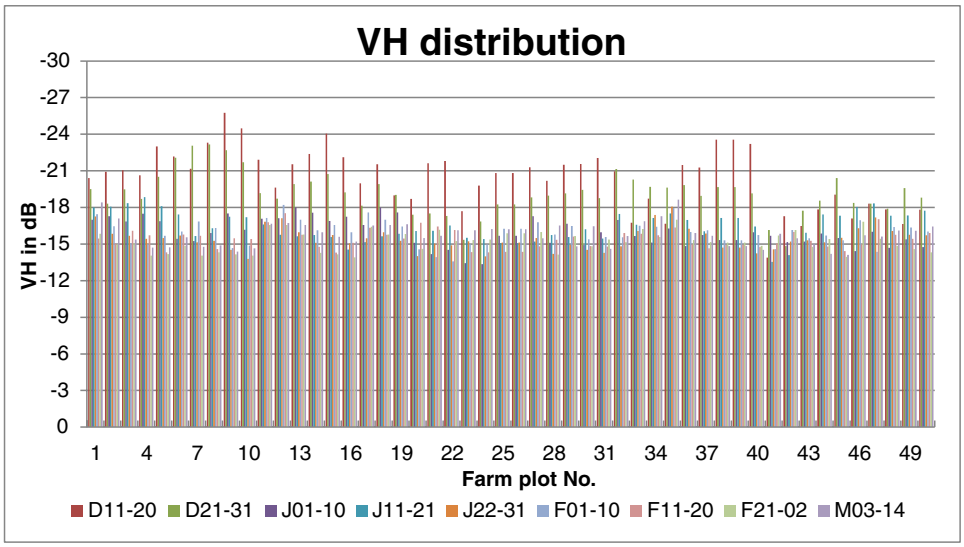
Single satellite of Sentinel-1 system has 12-day revisit time. As there is no intervention of cloud covering and bad weather, 12-day image collection system was planned for the study. In extreme conditions, two satellite constellations of 6 days of revisit data may be used to maintain uniformity in the study.

The current study facilitates quick and accurate mapping of potato areas with the existing time-series Sentinel-1 SAR and Sentinel-2B MSI images during the growing season between December 2019 and March 2020 in the ever-reliable cloud computing GEE platform. JavaScript along with the web-based interactive development environment (IDE) (https://code.earthengine.google.com) had utilized the

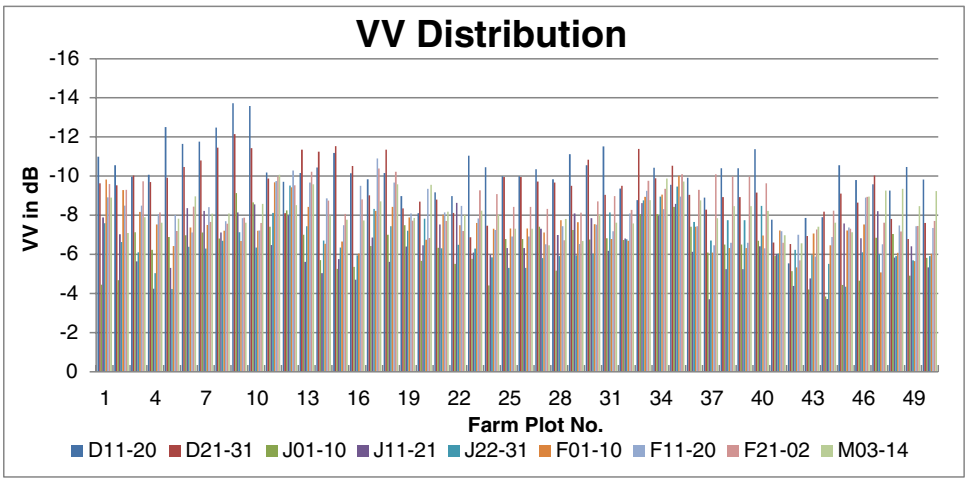
Table 2 Descriptive statistics of Sentinel-1 backscatter VH/VV (dB); data for 50 potato field location

Dates	Polarization	Mean	Max	Min	Range	Standard deviation	Variance	Median
Dec 11–20	VH	− 20.4	− 13.9	− 25.8	11.9	2.4	6	− 20.9
	VV	− 10.1	− 5.5	− 13.7	8.2	1.4	2	− 10.2
Dec 21–31	VH	− 19.2	− 14.8	− 23.2	8.3	1.7	2.9	− 19.2
	VV	− 9.4	− 6.5	− 12.2	5.6	1.4	2	− 9.5
Jan 01–10	VH	− 16	− 13.3	− 18	4.7	1.1	1.3	− 15.8
	VV	− 6.4	− 3.8	− 9.1	5.3	1.2	1.5	− 6.7
Jan 11–21	VH	− 16.5	− 13.6	− 18.8	5.3	1.2	1.4	− 16.4
	VV	− 6.5	− 3.7	− 8.6	4.9	1.3	1.8	− 6.4
Jan 22–31	VH	− 15.5	− 13.8	− 18.1	4.3	0.9	0.9	− 15.3
	VV	− 6.7	− 4.2	− 9.5	5.3	1.2	1.4	− 6.3
Feb 01–10	VH	− 15.9	− 13.6	− 18.2	4.6	1	1	− 16
	VV	− 7.4	− 5.1	− 10	4.9	1.1	1.3	− 7.3
Feb 11–20	VH	− 15.4	− 14.3	− 17.5	3.2	0.8	0.6	− 15.4
	VV	− 7.9	− 5.9	− 10.9	5	1.1	1.3	− 7.6
Feb 21–Mar 02	VH	− 15.2	− 13.9	− 17	3.1	0.8	0.7	− 15.2
	VV	− 8.5	− 5.7	− 10.4	4.7	1.1	1.3	− 8.4
Mar 03–14	VH	− 15.8	− 14.1	− 18.6	4.5	1	1	− 15.7
	VV	− 8.2	− 6.5	− 10.1	3.6	0.9	0.8	− 8.1

Dec-December, Jan-January, Feb-February, Mar-March

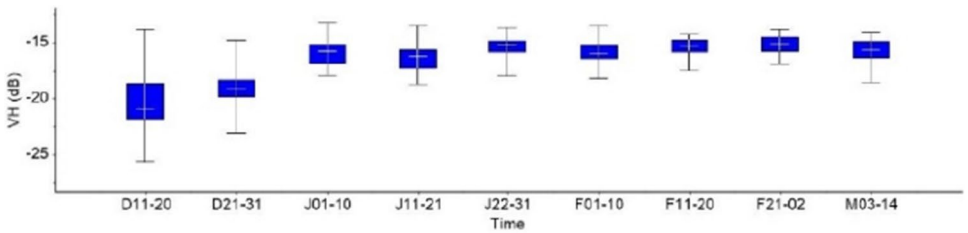


(a)

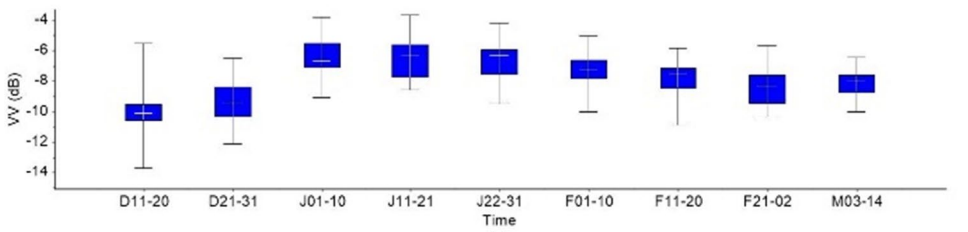


(b)

Fig. 3 Individual time-series Sentinel-1 backscatter (**a**) VH and (**b**) VV (dB) data for 50 potato fields in the study area



(a)



(b)

Fig. 4 Box plot time series backscatter (**a**) VH and (**b**) VV (dB) data. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively

parallel processing power of GEE to generate the potato crop stage maps of the study area. The segmentation was performed separately for each of the key potato growing stages along with groundtruthing points. The extracted potato maps were supplied in EPSG: 4326 (WGS-84) spatial reference system in a *Geotiff* format from the GEE. The images were supplied to Google Drive for further processing to extract backscattering (Sentinel 1A data) and NDVI values (Sentinel 2B) for the potato areas covering the 50 farm plots. Overview of adopted methodology for potato field extraction procedures comprises GEE cloud platform was drawn in flowchart form (Fig. 2).

Machine learning model

The potato yield map was generated using the kriging technique in ArcGIS environment (Dindaroglu 2014) from the 50 farm plots. The modeling phase involves applying different machine learning techniques to validate the backscatter and NDVI map. Yield data of each farmland for the study area was recorded and pre-processed. The potato yield map was generated using the kriging technique in ArcGIS environment (Dindaroglu 2014). The four most accepted classifiers such as radial basis function kernel, logistic regression, random forest, and naïve Bayes classifier were selected and applied.

Radial basis function kernel (RBF) model was used in SVM analysis for this study (Eq. (1)) (Farzin et al. 2021; Gandhi et al. 2016).

$$\mathrm{RBF}: k(x_i y_i) = \left(-y\left|\left|X_i - X_j\right|\right|\right), y > 0 \tag{1}$$

where $k(x_i, y_i)$ denotes the kernel function and γ represents the gamma term in the RBF kernel function for the potato crop yield monitoring.

Logistic regression (LR) is a classification algorithm that calculates a predicted probability for a dichotomous dependent variable based on one or more independent variables (Arabameri et al. 2021; Gandhi and Armstrong 2016). Authentication of LR model is regulated by the binary composition that indicates the probability of incidence with pair sample space, P (Eq. (2)).

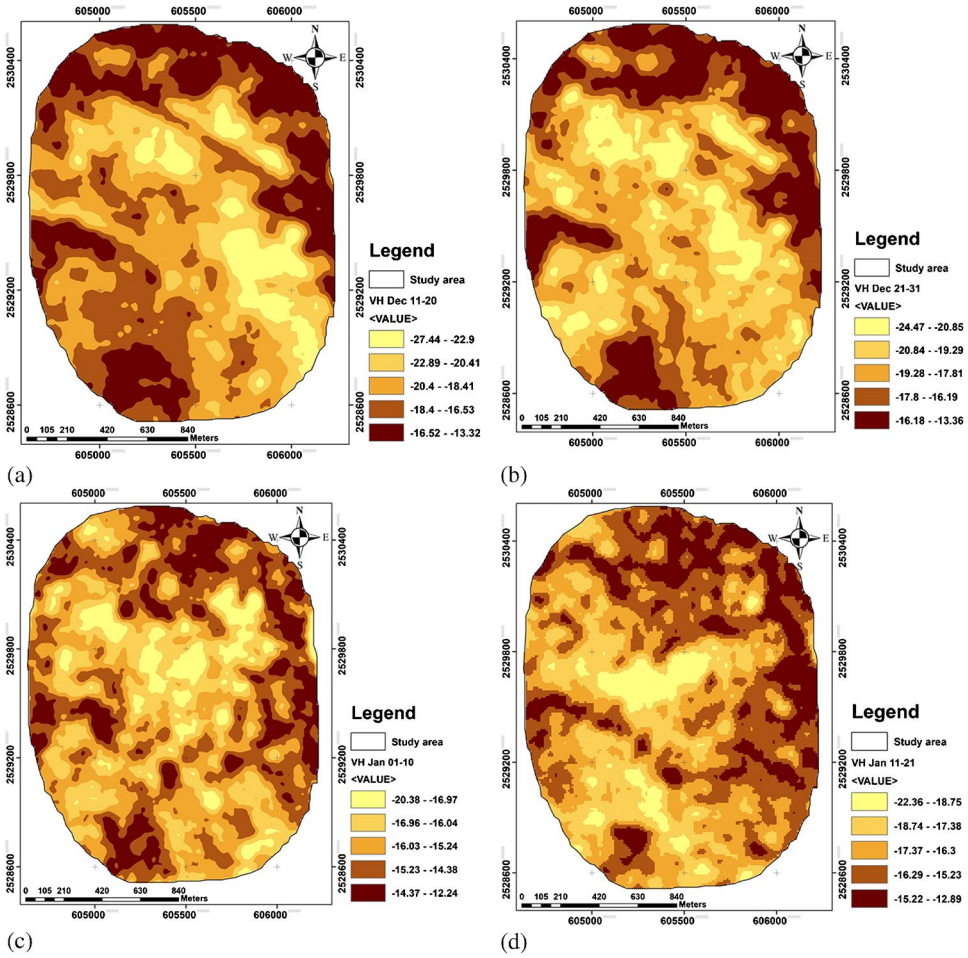
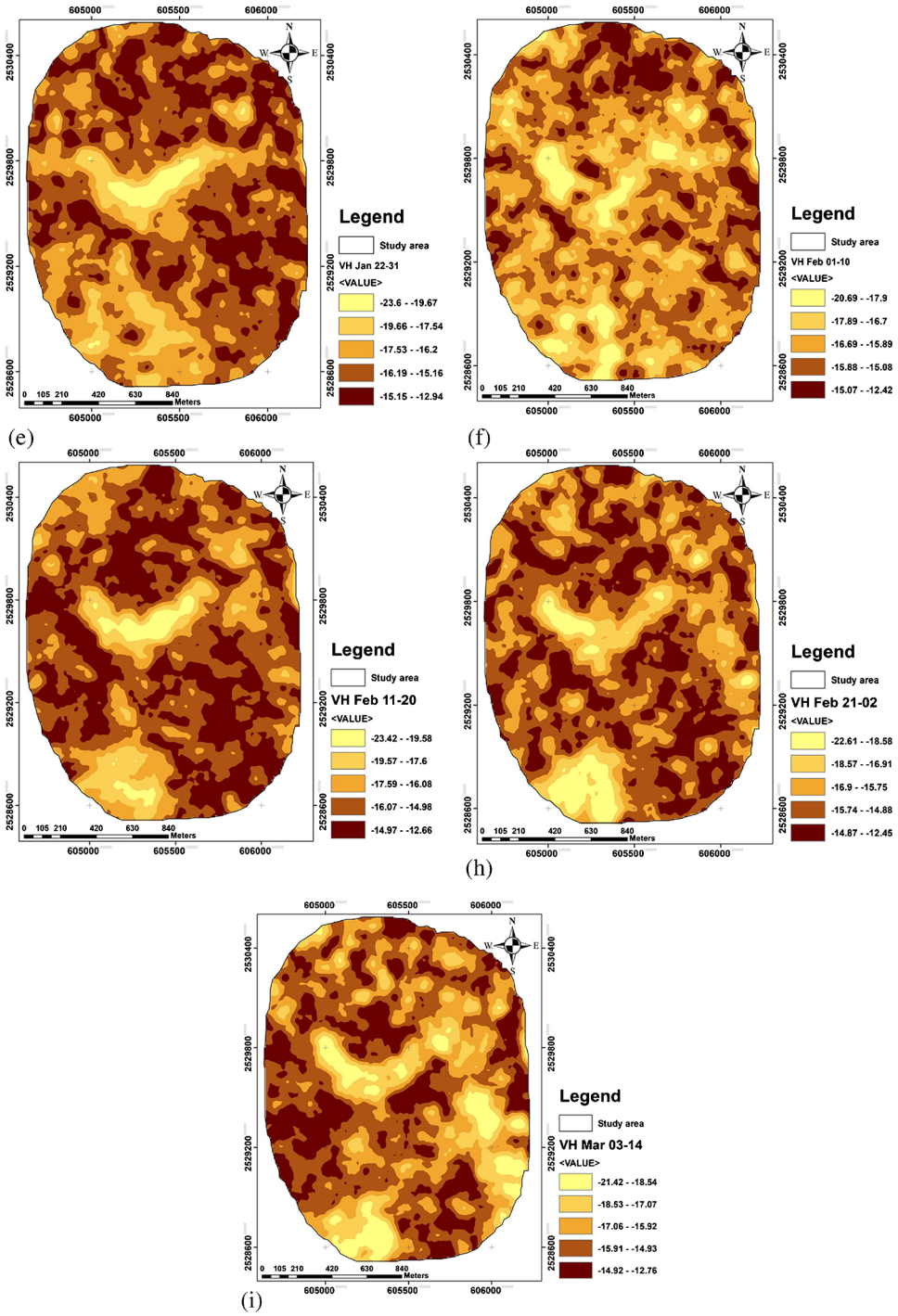


Fig. 5 Sentinel-1 backscatter VH (dB) mapping: (**a**) Dec 11–20, (**b**) Dec 21–31, (**c**) Jan 01–10, (**d**) Jan 11–21, (**e**) Jan 22–31, (**f**) Feb 01–10, (**g**) Feb 11–20, (**h**) Feb 21–02, and (**i**) Mar 03–14

Fig. 5 (continued)



$$P(y = \pm 1 \mid x, w) = \frac{1}{1 + \exp\left(-y\left(w^T x + b\right)\right)} \tag{2}$$

where, b signifies the intercept, T signifies the transfer matrix, and the k-dimensional coefficient, $w = (w_1, w_2, \ldots w_k)^T$, comprises the parameters that are essential to be assessed.

Random forest (RF) is a technique that uses multiple decision trees designed to perform the estimation using bootstrapping, resampling, and then applying majority voting or averaging with the generalization error (GE) (Eq. (3)) (Son et al. 2020; Sahu et al. 2017).

$$\mathrm{GE} = P_{x,y}(mg(x, y) < 0) \tag{3}$$

where P is the predictor, x and y are the potato crop monitoring input parameters that denote the likelihood space (x, y), and mg represents the function (Eq. (4)).

$$mg(x,y) = av_k I\left(h_k(x) = y\right) - \max_{j \neq y} av_k I\left(h_k(x) = j\right) \quad (4)$$

where $I(*)$ described the pointer function, j is a unification of hyper-rectangles, and h_k characterizes the union of hyper-rectangles.

Naïve Bayes classifier (NB) is a simple probabilistic classifier with strict feature independence assumptions based on Bayes' theorem (Eq. (5)) (Wang et al. 2021; Pandith et al. 2020).

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)} \quad (5)$$

where $P(Y|X)$ denotes the conditional probability of the parameter input trajectory x, in relation to the output parameter y; $P(Y)$ represents the prior probability; and $P(X)$ is the attribute aspect.

Hyper-parameters were tuned using the grid search technique, which is widely used for finding out the best settings of parameters in Python programming language. Training and testing data sets are used during the modeling hyper-parameter optimization for the model training, evaluation, validation, and selection (Table 1).

Python offers rich and efficient libraries for applying machine learning methods like Pandas and Numpy libraries for data manipulation. Subsequently, the Scikit-learn library containing various sets of supervised or unsupervised machine learning methods could be used (Youssef et al. 2019). For model development and validation purpose, the data was split into 70% (35 plots) and 30% (15 plots) groups, for training and testing, respectively. The Area Under the Receiver Operating Characteristics (AUROC) curve was generated using Anaconad-3 inbuilt Jupyter notebook 6.0.1.

Statistical measures

Descriptive statistics enable us to present the data in a more meaningful way, which allows a simpler interpretation of the data (Lund and Lund 2018). Descriptive statistics (such as mean, max, min, standard deviation, variance) of the VH, VV values for the images were carried out to observe the quality and pattern of the data. The outliers are removed from the data and supplied for further analysis. Graphs and box plots were also developed in "R" (Opensource software developed in University of Auckland, New Zealand) software environment for displaying and making a better understanding of the trends.

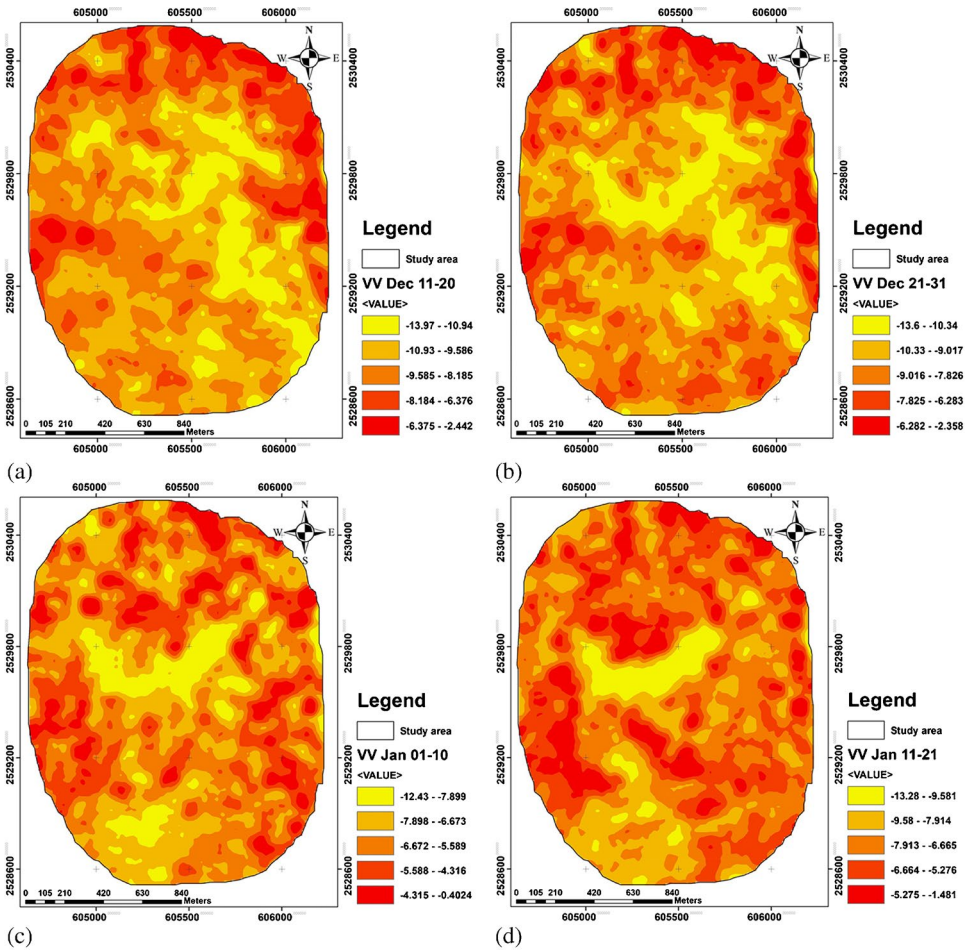
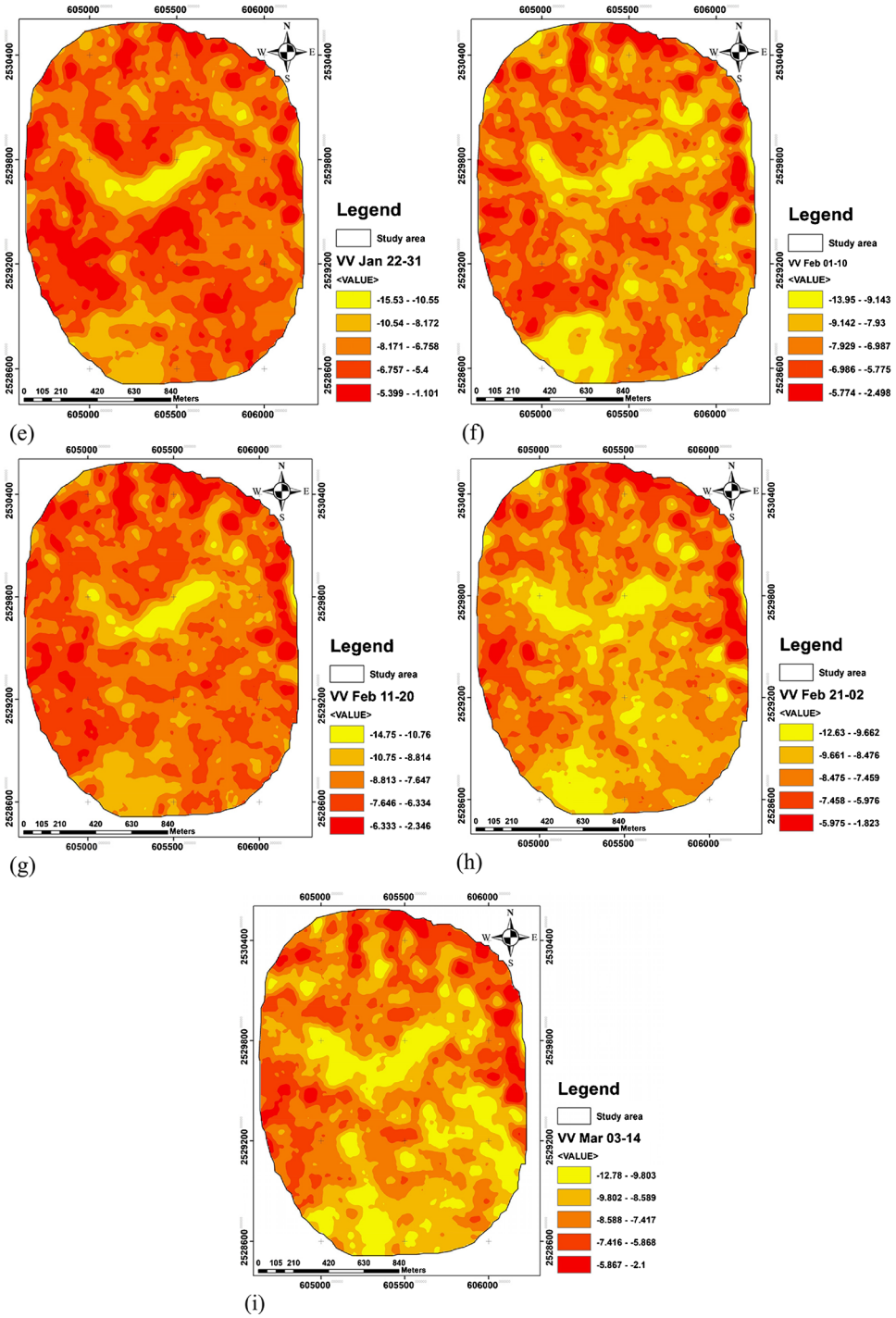


Fig. 6 Sentinel-1 backscatter VV (dB) mapping: (**a**) Dec 11–20, (**b**) Dec 21–31, (**c**) Jan 01–10, (**d**) Jan 11–21, (**e**) Jan 22–31, (**f**) Feb 01–10, (**g**) Feb 11–20, (**h**) Feb 21–02, and (**i**) Mar 03–14

Fig. 6 (continued)



In order to test the performances of the applied models, the following statistical metrics will be used: precision, recall, *F*-measure (F_1), and accuracy (Suganya et al. 2020).

Precision and recall are used to express the quality and quantity of the results. In other words, precision indicates the ratio of correct results to irrelevant ones, and recall indicates the overall relevant results retrieved expressed in percentage. Precision and recall are measured through Eqs. (6) and (7), respectively.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

The harmonic mean of the precision and recall of the model developed is expressed as *F*-measure. *F*-measure = 1 shows the best precision and recall. However, *F*-measure = 0 means one

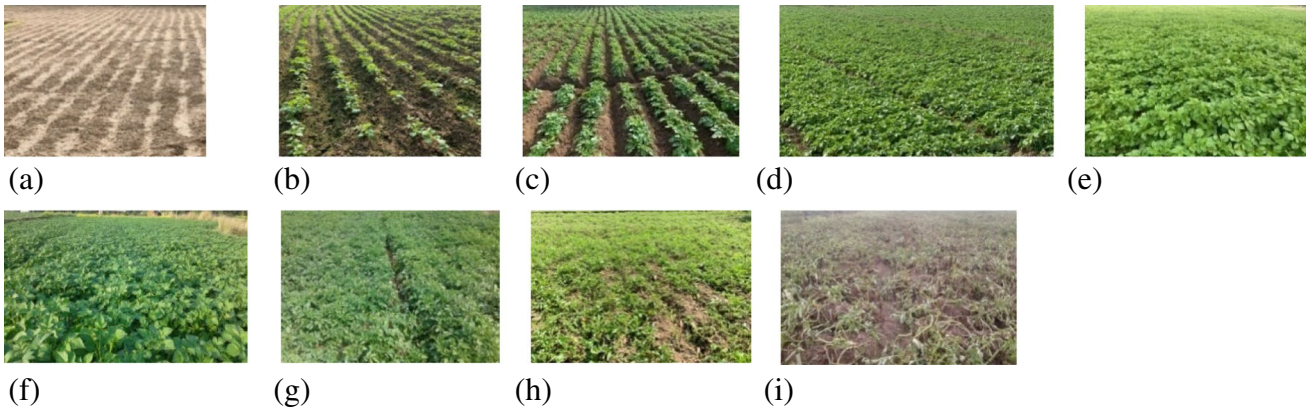


Fig. 7 Field photographs for key date interest in the study region: (**a**) Dec 11–20, (**b**) Dec 21–31, (**c**) Jan 01–10, (**d**) Jan 11–21, (**e**) Jan 22–31, (**f**) Feb 01–10, (**g**) Feb 11–20, (**h**) Feb 21–02, and (**i**) Mar 03–14

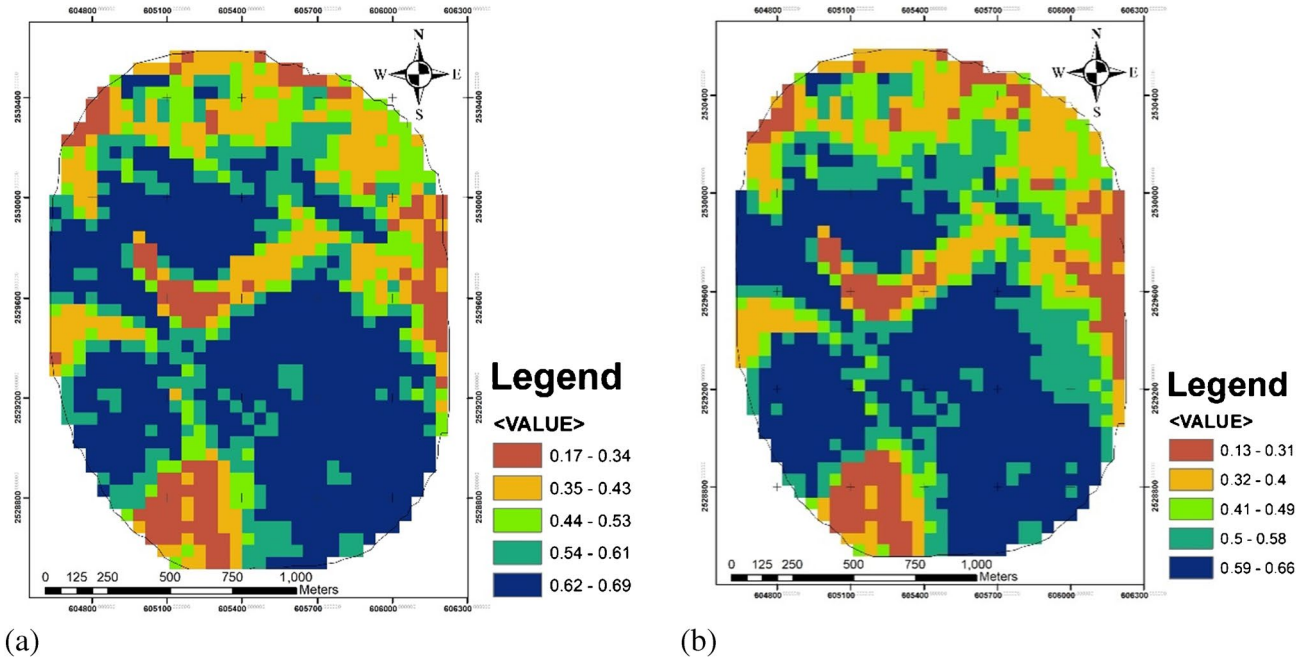


Fig. 8 Peak NDVI growing periods in the study region: (**a**) F01–10 and (**b**) F11–20

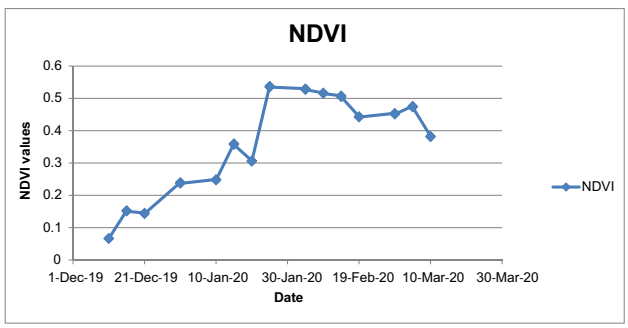


Fig. 9 NDVI time series analysis for potato crop (emergence to closure periods)

of the precision and recall is zero. *F*-measure (F_1) works as an accuracy indicator of a test and is calculated based on precision and recall parameters as follows (Eq. (8)):

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{8}$$

Accuracy is one of the most important classification evaluation criteria that could be calculated as given in Eq. (9). In order to assess the ability of the models to predict potato crop yield, the Area Under Curve (AUC) values were used. The AUC can be calculated as follows (Eq. (10)):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (9)$$

$$\text{AUC} = \frac{\sum \text{TP} + \sum \text{TN}}{P + N} \quad (10)$$

where TP (true positive) and TN (true negative) are the numbers of respondents that are correctly classified, P is the total number of potato farm plots with heavy spectacles, and N is the total number of potato farm plots without heavy spectacles.

This study considered several metrics to estimate the performance of different methods to correctly classify the cases in an independent dataset and avoid overfitting. All classification algorithms were executed using a tenfold cross-validation procedure (Palanivel and Surianarayanan 2019). The tenfold cross-validation process divides the collection of data into ten approximately equal sections. The nine remaining parts are used to train the model, and the test error is computed by classifying the given part. Lastly, the results often tests are averaged.

Results and discussion

Backscatter time series

The time series of Sentinel-1 backscatter (VH, VV) was measured along with the phenological development period of the potato crop. The sowing of potato seeds takes place during 11–20 December 2019 in winter season. The plant attends the vegetative stage around 1–10 January, 2020 which continues with stem elongation up to 11–21 January. The reproductive stage starts with the initiation of tubers during 22–31 January, around the blooming stage during 1–10 February. The tuber bulking stage is witnessed after 11–20 February followed by harvesting of the crop during 3–14 March 2020.

The descriptive statistics study was carried out for VH and VV values, to visualize the trend of collected information. The results showed a low mean backscattering signature during the seedling periods of VH and VV polarization − 20.42 dB and − 10.13 dB, respectively, whereas higher mean backscattering was experienced at the peak vegetation stage of VH and VV polarization (− 15.21 dB and − 6.39 dB) and decreased in the values after attaining the maturity stage (Table 2). The graph and box plot showed and validate the statistics trend of backscattering signatures of potato fields (Figs. 3 and 4). In the box plot potato parcel showed low median backscattering values during the seedling periods with VH and VV polarization at − 20.94 dB and − 10.15 dB, respectively, whereas higher median backscattering was experienced at peak vegetation stage of VH and VV polarization with − 15.20 dB and − 6.34 dB, respectively (Fig. 4).

Lower backscattering of VV polarization was witnessed compared to the VH polarization just after the seedling. The lower values during the period 11–20 December 2019 are due to the presence of a bare soil surface. So, the dynamics in VH and VV backscatter were due to land management practice (tillage) variations and surface soil moisture in the potato fields during seedling (Figs. 5 and 6).

By the start of the phenological monitoring, the potatoes had already reached vegetation stage 1–10 January 2020. The rapid increase in both VH and VV backscatter from 11 to 21 January onward corresponds to the period of leaf formation and stem elongation in the later stage of vegetation

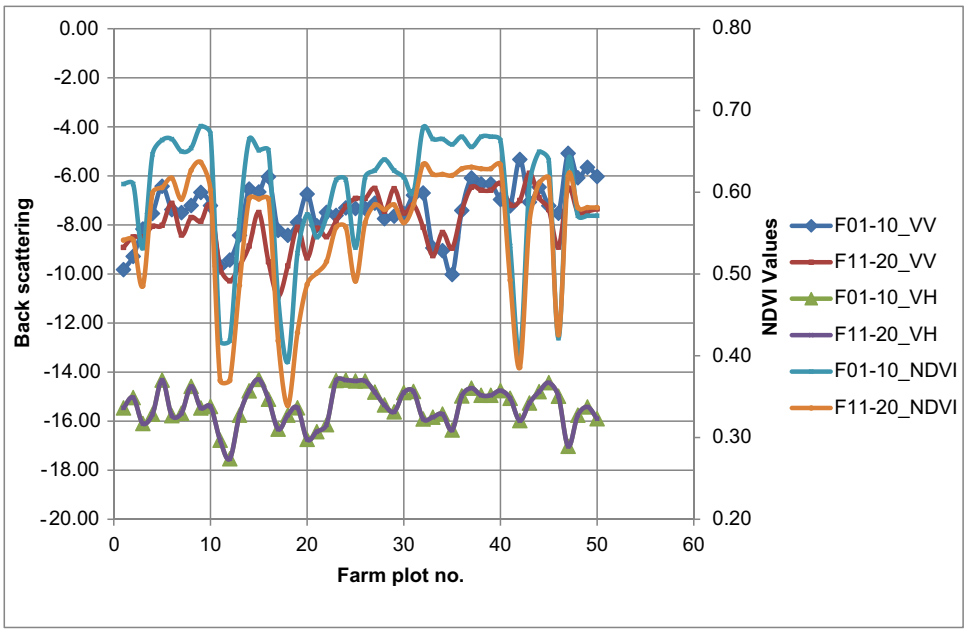
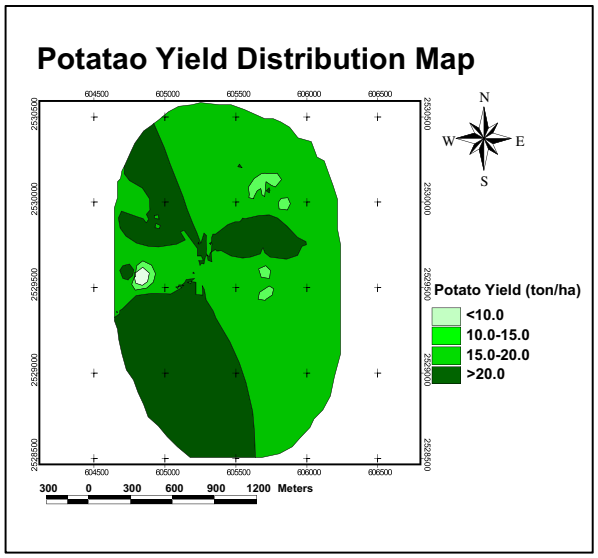


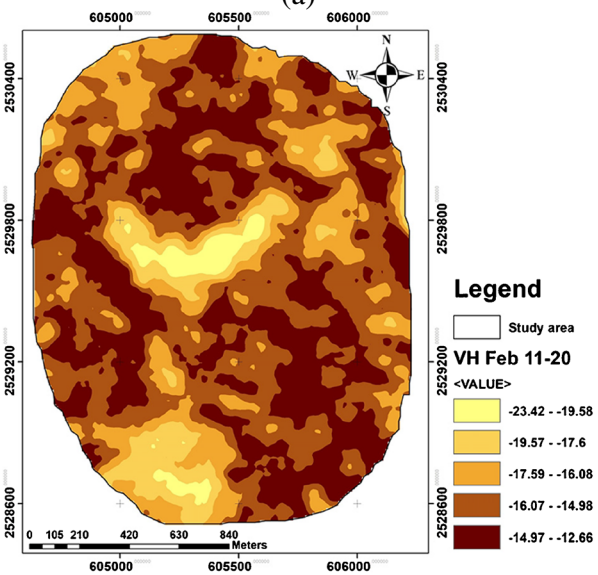
Fig. 10 Temporal analysis within VV, VH, and NDVI values in 50 farm plots during the periods 01–20 February 2020



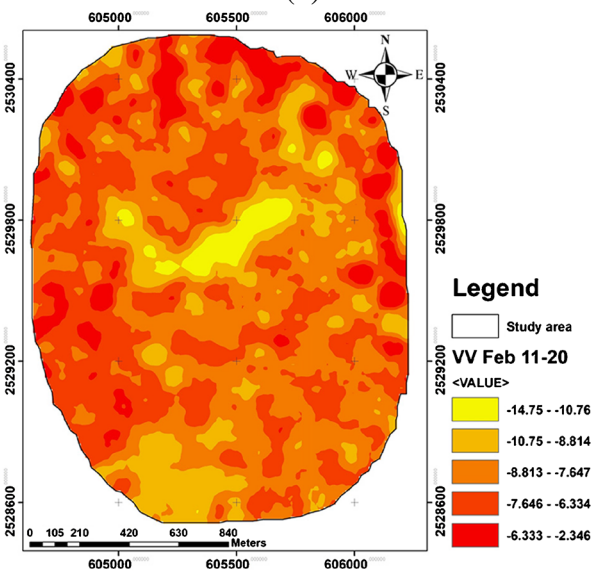
(a)

Crop field during 11-20 February

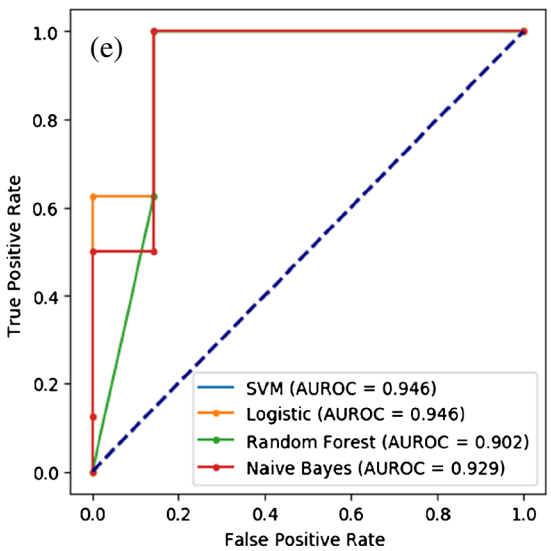
(b)



(c)



(d)



◂**Fig. 11** Validation with potato yield. (a) Potato yield distribution map. (b) Field map during 11–20 February. (c) VH map during 11–20 February. (d) VV map during 11–20 February. (e) Receiver operating characteristic (AUROC) curves

growth. During this time, the potato plants enter a stage of exponential growth until the ground cover reaches to maximum in the middle of tuber initiation stage at rounds 22–31 January (Figs. 5e and 6e). Above ground biomass stabilizes during this stage (22–31 January) and vertical crop growth ceased. During the end of tuber bulking stage (1–10 February 2020) and the beginning of tuber maturity (11–20 February) stages, the average height of the plants decreases. Variations in planting date and field management practices (e.g., ridge structure and nutrient application) intended that the amount of above ground biomass could vary between the parcels. This could effectively explain the decrease in backscatter around this time. Additionally, the rate at which the plants reached the yellowing and brownish stages (fully maturity) vary. However, variation in backscattering values could be used in predicting crop maturity and harvesting schedule. The abrupt decrease in VH backscatter from 21 February to 14 March 2020 is related to haulming, the process of destroying the haulms (i.e., stalks or stems) before harvesting. The trend and variation in VH and VV can be easily co-related to field images taken at different crop growth stages during field visits (Fig. 7).

The VV, VH backscatter values from Sentinel-1 found to show clearly the status of potato crop growth. As the Sentinel-1 can provide cloud-free images throughout the year, these data can be used for near real-time monitoring of potato crop. Even using these data, diseased crop areas can be identified for special attention. Water stress in potato causes considerable losses in yield (Romero et al., 2017), sometimes causing less pulp in the tubers. Sentinel-1 images may be used to identify the water-stressed area in the potato crop, which will ensure better return to the farmers.

NDVI estimation from Sentinel-2

NDVI values estimated from optical satellite-based images have established itself as the leading tool for vegetation index in studying disturbances in vegetation covers. The free-of-cost, high-resolution (up to 10 m) optical Sentinel-2 data generally have superior reliability in modeling the NDVI for agricultural crop monitoring throughout their phonological growth periods and are also well documented (Saini and Ghosh 2018). However, Sentinel-2 optical sensor is handicapped by bad weather including rain and cloud covers. SAR data may be used as an alternative or at least a counterpart approach to the optical index data, such as EVI, NDVI, and SAVI (Torbick et al. 2017; Navarro et al. 2016).

The greater applicability of SAR data can be supported by the strong relationship between the NDVI and backscatter coefficients (Filgueiras et al. 2019). It is probable to witness the vegetation variability at the central pivots by detecting exclusively the dynamics of the values of the backscatter coefficients VH and VV. The set consisting of an NDVI image of Sentinel 2 and the backscatter coefficients in the different polarizations of Sentinel-1 were compared.

Sentinel 2 optical data shows maximum NDVI values recorded in the first and second week of February (Figs. 8 and 9). The current study observed that peak growing in terms of NDVI value (>0.5) is a good estimator to VH/VV backscatter value within key dates of interest for the 50 farm plot location (Fig. 10).

By critically observing, the Sentinel-1-based backscattering data has a similar pattern to that of Sentinel-2-based NDVI values, showing the suitability of Sentinel-1 data for potato crop growth monitoring. Additionally, Sentinel-1 will enable continuous monitoring of the crop in the peak growing seasons.

Potato yield mapping and model comparison

Crop yield is one of the major indicators of crop growth and sustainability (Kancheva and Georgiev 2011). Potato yield (t/ha) distribution map was developed in ArcGIS 10.5 software environment (Fig. 11a). The average potato yield was around 18.8 t/ha which is sound compared to other areas in the region; however, the quality of potato needs to be monitored. Fifty farm plot ground backscatter and NDVI value validate with the crop yield data through the potato crop yield map. In order to demonstrate the accuracy, area under the receiver operating characteristic (AUROC) curve was used in this analysis with different machine learning techniques, such as support vector machine, logistic regression, random forest, and naïve Bayes, which compared the potato yield data with the backscatter and NDVI map of the study area.

The cross-validation test classification results were obtained by support vector machines (SVM), logistic regression (LR), random forest (RF), and naïve Bayes (NB) algorithms for the potato yield prediction (Table 3), reporting some of the most popular measures/indicators of success: train accuracy, test accuracy, precision, recall, and the F_1 score.

The random forest model classifier obtained the best training and testing accuracy, precision, recall, and F_1 scores (Prasad et al. 2021). The resulting hyperparameters from the grid search pointed out the construction of maximum depth 5 and number of estimators 10 within the forest model; with these settings, the precision of the model reached 99%, the recall was 83%, while the F_1 score was 90%. The classification error was zero, with an overall accuracy of 90%.

SVM and naïve Bayes classifier obtained second-best training and testing accuracy, precision, recall, and F_1 score

value. The support vector machine classification algorithm is very common and has several hyperparameters to tune up the outcome with gamma 2, C 1. The choice of kernel, which will control how the input variables are projected, is perhaps the most important parameter (Gonzalez-Sanchez et al. 2014). The C parameter is also important to choose as it represents the penalty and can contain a variety of values. It has a significant impact on the shape of the resulting regions for each class. In this case, the resulting hyperparameters obtained by grid search pointed out to RBF kernel with 1.0 penalty value. The SVM model train and test accuracy is 97% and 80%, respectively. The algorithm achieved an F_1 score of 88%, while recall and precision scores reached 91% and 84%, respectively.

In the naïve Bayes classifier, the precision of the model reached 90%, the recall was 83%, while the F_1 score was 87% with 88% train accuracy and 80% test accuracy.

There are no critical hyperparameters to tune within the logistic regression classifier. No useful differences in performance were achieved with various solvers, regularization, and C parameter, which controls the regularization strength. In this study "lbfgs" solver was used in LR model. The final results achieved by LR were obtained under the default settings where the scores for precision and recall were 90% and 75%, respectively, with the train and test accuracy that is > 72%.

Discussion

West Bengal is a leading state in India, in terms of the total area under potato farming. Still, the low price during harvesting and low potato quality has drastically reduced the profit margin of the farmers. Potato in the surrounding regions of the study area mostly maintained a similar time frame and similar potato variety (*Pukhraj* or *Chandramukhi*) for cropping. According to Wegmüller et al. (2011) range and direction of Sentinel-1 data may be highly influenced by ridge orientation of the potato fields, so in this work, during the periods 21–31 December VV backscatter in parcel dB were higher than the others due to the formation of ridges during potato cultivation in bare soil.

With the variation in VH and VV values, the growth stage of potato can be evaluated and monitored. Particularly, VH values showed a sound inclination to the growth of potato plant in the field. The application of a cloud computing system with Google Earth Engine (GEE) enables quick acquisition and analysis of Radarsat data for potato crop. GEE platform removed the final constraint of fast analysis of Radarsat data (Swain et al. 2020). The uninterrupted data will enable to identification crop stress and disturbance areas for special attention.

The variation of VV, VH values from Sentinel-1 imagery with NDVI values estimated from optical Sentinel-2 imagery at the peak crop growth stages showed identical trends (Fig. 10), validating the suitability of backscattering data. The time series curve for NDVI throughout the cropping season shows a symmetric pattern to that of VH and VV backscatter values, which further validated the suitability of Sentinel-1-based SAR data.

Free of cost, multitemporal Sentinel-1 data may be used for quick and near real-time crop monitoring in the future (Khabbazan et al. 2019). Sentinel-1 data with 5–12-day revisit period is very suitable for crop analysis giving enough time to take positive crop production decisions (Macedo and Kawakubo 2018). This shows the suitability of SAR data providing consistent judgment for monitoring agricultural crops. Additionally, agriculture being a biological entity can easily be affected by a slight change in weather and the emergence of pests, insects, etc., which need continuous monitoring.

Yield estimation of potato shows that the backscattering values can easily predict potato yield well in advance, through the models. The VV and VH backscatted values during 11–20 February has better correlation with the potato yield distribution map (Fig. 11). As the 11–20 February has the optimum vegetation growth with minimum exposed soil. The AUROC result showed that SVM and logistic regression were well performed at 0.946, naïve Bayes at 0.929, and random forest at 0.902. So, the AUROC results further validated the suitability of backscatter-based crop monitoring. Foughalia et al. (2019) used blight forecast model in Cloud-IoT system to prevent potato late blight disease. Here, the able to reduce the overall cost of the system to control the disease.

Potato being a cash crop in the region needs special attention to maintain crop yield as well as tuber quality, which may be achieved through Sentinel-1 imagery in GEE platforms. The study may be extended to support in updating irrigation scheduling for efficient water use in dry seasons, climate change pattern, cropland suitability evaluation, agriculture practice and crop

Table 3 Comparison of the performance measures of applied classification algorithms

Classifiers	Train accuracy	Test accuracy	Precision	Recall	F_1 score
SVM	0.971	0.800	0.846	0.917	0.880
Logistic regression	0.829	0.733	0.900	0.750	0.818
Random forest	0.99	0.867	0.99	0.833	0.909
Naïve Bayes	0.886	0.800	0.909	0.833	0.870

rotation selection, and crop yield prediction ensuring regional food security for sustainable agricultural planning.

Conclusion

The current study demonstrated a simple but efficient approach to monitor potato cropping areas using the VH- and VV-polarized images of Sentinel-1 C band during different phenological periods, through the rapid temporal unified framework of Google Earth Engine cloud computing environment. The system helped in quick pre-processing and fast-paced data acquisition and processing. The VH and VV backscatter data were used to identify crop growth stages and successfully validated with NDVI optical sensor data and its time series curves. Receiver operating characteristic (AUROC) curves along with potato yield map further enhanced the suitability of backscatter images for continuous monitoring of the potato crop in the study area. Key growth stage curves of backscatter were also validated using field visit photographs. The study enhanced the potential use of freely accessible Sentinel-1 SAR data for crop monitoring in the future.

Declarations

Conflict of interest The authors declare no competing interests.

References

Arabameri A, Pal SC, Rezaie F, Nalivan OA, Chowdhuri I, Saha A, Lee S, Moayedi H (2021) Modeling groundwater potential using novel GIS-based machine-learning ensemble techniques. J Hydrol Regio Studi 36:100848. https://doi.org/10.1016/j.ejrh.2021.100848

Agricoop (2017) Pocket book of agricultural statistics. 128, www.agricoop.nic.in. Accessed 26 May 2020

Atzberger C (2013) Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens 5:949–981

Arias M, Campo-Bescós MA, Álvarez Mozos J (2018) Crop type mapping based on Sentinel-1 backscatter time series. In Proceedings of the IGARSS -2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018: 6623–6626.

Bazzi H, Baghdadi N, El Hajj M, Zribi M, Minh DHT, Ndikumana E, Courault D, Belhouchette H (2019) Mapping paddy rice using Sentinel-1 SAR time series in Camargue. France Remote Sens 11:887

Cumming I, Bennett J (1979) Digital processing of Seasat SAR data. Proceedings of ICASSP '79, IEEE International Conference on Acoustics, Speech, and Signal Processing, Washington, DC, pp 710–718

Dindaroglu T (2014) The use of the GIS Kriging technique to determine the spatial changes of natural radionuclide concentrations in soil and forest cover. J Environ Health Sci Eng 12(1):130

FAO (2008) Potato and water resources. http://www.fao.org/potato-2008/en/potato/water.html. Accessed 20 May 2020

Farzin M, Avand M, Ahmadzadeh H, Zelenakova M, Tiefenbacher JP (2021) Assessment of ensemble models for groundwater potential modeling and prediction in a karst watershed. Water 13:2540. https://doi.org/10.3390/w13182540

Filgueiras R, Mantovani EC, Altho D, Fernandes Filho EI, França da Cunha F (2019) Crop NDVI monitoring based on Sentinel 1. Remote Sens 11:1441

Foughalia K, Fathallahb K, Frihidab A (2019) A cloud-IOT based decision support system for potato pest prevention. Procedia Comput Sci 160:616–623

Gandhi N, Armstrong L (2016) Applying data mining techniques to predict yield of rice in humid subtropical climatic zone of India. In: Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom. 1901–1906. https://ieeexplore.ieee.org/abstract/document/7724597/ Accessed 25 May 2020

Gandhi N, Armstrong LJ, Petkar O, Tripathy A K (2016) Rice crop yield prediction in India using support vector machines. 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE): 1–5 https://doi.org/10.1109/JCSSE.2016.7748856

Garnett T, Appleby MC, Balmford A, Bateman IJ, Benton TG, Bloomer P, Burlingame B, Dawkins M, Dolan L, Fraser D, Herrero M, Hoffmann I, Smith P, Thornton PK, Toulmin C, Vermeulen SJ, Godfray HCJ (2013) Sustainable intensification in agriculture: premises and policies. Sci 341(6141):33–34

Gerland P, Raftery AE, Ševčíková H, Li N, Gu D, Spoorenberg T, Alkema L, Fosdick BK, Chunn J, Lalic N, Bay G (2014) World population stabilization unlikely this century. Sci 346(6206):234–237

Ghosal S (2020) West Bengal to produce 10% higher potatoes this year. Economic Times, 10 January 2020 Accessed 12 February 2020.

Gomez D, Salvador P, Sanz J, Casanova JL (2019) Potato yield prediction using machine learning techniques and Sentinel 2 data. Remote Sens 11(15):1745

Gonzalez-Sanchez A, Frausto-Solis J, Ojeda-Bustamante W (2014) Predictive ability of machine learning methods for massive crop yield prediction. Span J Agr Res 12(2):313–328

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27

Hunt ER, Horneck DA, Spinelli CB, Turner RW, Bruce AE, Gadler DJ, Brungardt JJ, Hamm PB (2018) Monitoring nitrogen status of potatoes using small unmanned aerial vehicles. Preci Agr 19:314–333

IMD (2017) Regional Meteorological Centre Kolkata India Meteorological Department, Ministry of Earth Sciences Government Of India. http://imdkolkata.gov.in/ Accessed 10 March 2020

Jia K, Li Q, Tian YC, Wu B, Zhang F, Meng J (2012) Crop classification using multi configuration SAR data in the North China Plain. Int J Remote Sens 35:170–183

Kancheva R, Georgiev G (2011) Spectral and agronomical indicators of crop yield. Proceedings of SPIE - Int Soc Opt Eng 1–12.

Kuenzer C, Knauer K (2013) Remote sensing of rice crop areas. Int J Remote Sens 34:2101–2139

Khabbazan S, Vermunt P, Steele-Dunne S, RateringArntz L, Marinetti C, van der Valk D, van der Sande C (2019) Crop monitoring using Sentinel-1 data: a case study from the Netherlands. Remote Sens 11(16):1887. https://doi.org/10.3390/rs11161887

King BA, Stark JC (1997) Potato irrigation and management. Bulletin No 789, Published by U.S. Department of Agriculture. Assessed on 10 April 2021.

Lund A, Lund M (2018) Laerd statistics: descriptive and inferential statistics. 4https://statistics.laerd.com/ Accessed 15 May 2020

Macedo L, Kawakubo FS (2018) Using Sentinel-1 SAR data for crop phonological development monitoring associated with Sentinel 2 data in Rio Verde-GO. Proc Remote Sens for Agr, Ecosyst and Hydrol XX; 10783:1078304.

Malit R (2020) Value addition: 8 products you can make from potatoes, 2017,http://www.herbusiness.co.ke/ Accessed 25 May 2020.

Mithiya D, Mandal K, Datta L (2019) Forecasting of potato prices of hooghly in West Bengal: time series analysis using SARIMA model. Int J Agric Econ 4(3):101–108. https://doi.org/10.11648/j.ijae.20190403.13

Navarro A, Rolim J, Miguel I, Catalão J, Silva J, Painho M, Vekerdy Z (2016) Crop monitoring based on SPOT-5 Take-5 and Sentinel-1A data for the estimation of crop water requirements. Remote Sens 8:525

Nigon TJ, Mulla DJ, Rosen CJ et al (2014) Evaluation of the nitrogen sufficiency index for use with high resolution, broadband aerial imagery in a commercial potato field. Precis Agric 15:202–226

Pandith V, Kour H, Singh S, Manhas J, Sharma V (2020) Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis. J Sci Res 64(2):394–398

Pandey K (2019) Slumping potato prices hit farmers, but worse is yet to come.https://www.downtoearth.org.in/ Accessed 12 January 2020.

Palanivel K, Surianarayanan C (2019) An approach for prediction of crop yield using machine learning and big data techniques. Int J Comput Eng Tech 10(3):110–118

Prasad NR, Patel NR, Danodia A (2021) crop yield prediction in cotton for regional level using random forest approach. Spat Inf Res 29:195–206

Pulvirenti L, Chini M, Pierdicca N, Boni G (2016) Use of SAR data for detecting floodwater in urban and agricultural areas: the role of the interferometric coherence. IEEE Trans Geosci Remote Sens 54(3):1532–1544

Radoux J, Chome G, Jacques DC, Matton N, Lamarche C, D'Andrimont R, Defourny P (2016) Sentinel-2's potential for sub-pixel landscape feature detection. Remote Sens 8(6):488

Romero AP, Alarcon A, Valbuena RI, Galeano CH (2017) Physiological assessment of water stress in potato using spectral information. Fron Plant Sci 8:1608

Roy A, Siddique G, Mandal M, (2019) A study on the productivity of Potato in Hugli, West Bengal, India. J Geo Environ Earth Sci Int 1–14.

Rusia DK, Swain KC, Singha C (2018) Integrated geospatial technique for potential groundwater zone (PGZ) identification. J Agro Nat Reso Manag 5(3):142–150

Sahu S, Chawla M, Khare N (2017) An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach, 2017 International Conference on Computing, Communication and Automation (ICCCA): 53–57 https://doi.org/10.1109/CCAA.2017.8229770.

Saini R, Ghosh S (2018) Crop classification on single date Sentinel-2 imagery using random forest and support vector machine. ISPRS - Int Arc Photo Remote Sens Spat Inform Sci 17(5):683–688

Sarkar D (2017) Shortage of cold-storage space hurts West Bengal's potato farmers. 09 March 2017, https://economictimes.indiatimes.com Accessed 12 January 2020.

Shafi U, Rafia Mumtaz R, García-Nieto J, Hassan SA, Zaidi SAR, Iqbal N (2019) Precision agriculture techniques and practices: from considerations to applications. Sensors 19:3796

Shelestov A, Lavreniuk M, Kussul N, Novikov A, Skakun S (2017) Exploring Google Earth Engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping. Front Earth Sci 5(17):1–10

Shelestov A, Lavreniuk MS, Kussul N (2016) Large scale crop mapping in ukraine using Google Earth Engine. American Geophysical Union, Fall Meeting 2016, abstract #IN51B-1851.

Singh U, Praharaj CS, Singh SS, Singh NP (2016) Biofortification of food crops, Kindly Edition; Springer Publications, p 925

Singha M, Dong J, Zhang G, Xiao X (2019a) High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. Sci Data 6(26) https://doi.org/10.1038/s41597-019-0036-3

Singha C, Swain KC, Saren BK (2019b) Land suitability assessment for potato crop using analytic hierarchy process technique and geographic information system. J Agr Eng 56(3):78–87

Singha C, Swain KC, Swain SK (2020) Best crop rotation selection with GIS-AHP technique using soil nutrient variability. Agr-Basel 10:213

Son NT, Chen CF, ChenGuo CRHY, ChengChen YSSL, Lin HS, Chen SH (2020) Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan. Int J Remote Sens 41(20):7868–7888. https://doi.org/10.1080/01431161.2020.1766148

Spasova T, Nedkov R (2019) On the use of SAR and optical data in assessment of flooded areas. Proc. SPIE Vol11174 in, Seventh Int Conference on Remote Sensing and Geoinformation of the Environ

SSTL (2017) SSTL announces NovaSAR-1 data deal with Australia's CSIRO, SSTL Press Release, 26 Sept. http://www.sstl.co.uk/Press/SSTL-announces-NovaSAR, 2017,(Accessed on 16 May 2020).

Steele-Dunne SC, McNairn H, Monsivais-Huertero A, Judge J, Liu PW, Papathanassiou K (2017) Radar remote sensing of agricultural canopies: a review. IEEE J Sel Top Appl Earth Obs Remote Sens 10:2249–2273

Suganya M, Dayana R, Revathi R (2020) Crop yield prediction using supervised learning techniques (June 30. Int J Compu Eng Tech 11(2):9–20

Swain KC, Singha C, Nayak L (2020) Flood susceptibility mapping through the GIS-AHP technique using the cloud. ISPRS Int J Geo-Inform 9(12):720

Swain KC, Jayasuriya HPW, Salokhe VM (2007) Low altitude remote sensing (LARS): a potential substitution to satellite based remotes sensing for precision agriculture adoption in fragmented and diversified farming conditions. Agr Eng Int CIGR Ej 9(12):1–16

Thorton M (2020) Potato growth and development, College of Agriculture and Life Science, University of Idaho, USA report, browsed on 25th April, 2020.

Tian H, Wu M, Wang L, Niu Z (2018) Mapping early, middle and-late rice extent using sentinel-1A and Landsat-8 data in the poyanglakeplain. China Sensors 18(1):185

Torbick N, Chowdhury D, Salas W, Qi J (2017) Monitoring rice agriculture across Myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2. Remote Sens 9:119

Youssef M, Mohammed S, Hamada EK, Wafaa BF (2019) A predictive approach based on efficient feature selection and learning algorithms' competition: case of learners' dropout in MOOCs. Educ Inf Technol 24:3591–3618

Wang H, Wang H, Wu Z, Zhou Y (2021) Using multi-factor analysis to predict urban flood depth based on naive Bayes. Water 13:432. https://doi.org/10.3390/w13040432

Wegmüller U, Santoro M, Mattia F, Balenzano A, Satalino G, Marzahn P, Fischer G, Ludwig R, Floury N (2011) Progress in the understanding of narrow directional microwave scattering of agricultural fields. Remote Sens Environ 115:2423–2433

Zhang F, Li G, Li W, Hu W, Hu Y (2016) Accelerating spaceborne SAR imaging using multiple CPU/GPU deep collaborative computing. Sensors 16(4):494–499