



Vegetation Indices-Based Rice and Potato Yield Estimation Through Sentinel 2B Satellite Imagery

8

Chiranjit Singha and Kishore C. Swain

Abstract

High-resolution optical remote sensing imagery has sound potential for future crop yield estimation. In precision agriculture adoption, these systems can provide valuable information on various factors affecting a farm's production. The current study used different crop vegetation indices, such as NDVI, NDWI, BI, DVI, SAVI, GEMI, PVI, RVI, and LAI, for estimating rice and potato crop yield on a microscale. Optical Sentinel-2B images were used for rice and potato crop estimation during 2018 and 2019 in ArcGIS 10.7 software environment for the rural neighborhood of Tarakeswar region, Hooghly (West Bengal, India). The geostatistical semi-variogram analysis with the best fitting of the exponential and spherical models determined the degree of spatial variability of rice and potato yield. In statistical Getis-Ord G_i^* analysis, the clusters of VIs indicated high yield with NDVI, RVI, and SAVI surfaces, while low vegetation indices showed low yield. Furthermore, the support vector machine, random forest, and logistic regression models were positively used in the spatial

assessment of rice and potato crop yield estimation, with AUROC values of 80–90%. However, the Naïve Bayes model was categorized as a moderate to good predictor with an AUC value between 60 and 80%. This chapter introduced a novel approach for crop yield prediction and validation with optical satellite imagery for microscale precision and agriculture adoption, which further helps using this method for other crops.

Keywords

Vegetation indices · NDVI · Sentinel-2B · Crop yield · AUROC

8.1 Introduction

In developing countries, the population explosion is the major issue for food security achievement. The future food grain yield is estimated using standard crop and environmental parameters. With the predicted crop yield, the government gets enough time to prepare itself for alternative food sources to face the food security threat (if arises). However, several optical reflectance-based vegetation indices (VIs) are very good evaluators of the local food security scenario well in advance through the earth observation-based remote sensing (RS) techniques (Lambert et al. 2018). Singha and Swain

C. Singha (✉) · K. C. Swain
Department of Agricultural Engineering, Institute of
Agriculture, Visva-Bharati, Sriniketan, West Bengal
731236, India
e-mail: singha.chiranjit@gmail.com

(2016) described the advantages of RS techniques based on multicriteria decision-making analysis, which has greater potential for sustainable agricultural planning. Integrating spatial analysis and sensor technology in the precision agriculture study can provide valuable crop information for field-specific management (ESA 2019). This also assists in monitoring variable fertilizer application, irrigation scheduling, biomass estimation, and harvesting at actual crop maturity by observing various agronomic parameters (Vallentin et al. 2021). Quick and fast crop yield estimation at the micro- to macro-level is very promising through remote sensing, further assisting in sustainable agricultural planning (Yuan et al. 2015). Assessing crop yield is one of the major issues to enhance the farmer's socio-economic development and optimize the industrial processing demand. Earth observation-based satellite systems could better monitor the changes in crop growth due to management practices, climate change impact, the emergence of pests and diseases, or water stress (ESA 2019; Yuan et al., 2015). Remote sensing resulting in several spectral VIs has helped monitor crop productivity at the farm level (Shammi and Meng 2020). Machine learning (ML)-based regression models were useful for crop yield prediction through the in-situ field information and RS-derived VIs in US Corn Belt (Ji et al. 2021). The above-said model could provide a reference for both physical and data organization. Several studies have shown that using ML techniques to analyze the data collected from the RS technique improves crop yield prediction (Shammi and Meng 2020; Guo et al. 2021). Locally field-based multi-temporal satellite remote sensing of crop VIs has good statistical significance with wheat grain yield up to 2 t ha⁻¹ (Ali et al. 2019; Panek et al. 2020). European Space Agency (ESA) provides high-resolution Sentinel 2 (S2) data through open access for the environmental monitoring and assessment of agriculture, wetland vegetation, forestry development, etc. S2 data offers a high spatial, spectral, and temporal resolution which can estimate individual farm-level crop health status, soil moisture, crop stress, and crop yield information through RGB and near-infrared-

based spectral VIs. Multiple Linear Regression (MLR), support vector machine (SVM), extreme gradient boosting (XGB), stochastic gradient descent (SGD), and random forest (RF) were employed for field scale prediction of soya yield using cloud-free S2 multispectral images derived from different VIs of specific growing periods (2018, 2019, and 2020) in Upper Austria (Pejak et al. 2022).

The simplified S2-MSI imagery can scientifically create different vegetation biophysical variables from the surface reflectance values such as Normalized Differential Vegetation Index (NDVI) (Rouse et al. 1973); Normalized Differential Water Index (NDWI) (Chandel et al. 2019); Brightness Index (BI) (Cavalaris et al. 2021), Difference Vegetation Index (DVI) (Kusul et al. 2020), Soil-Adjusted Vegetation Index (SAVI) (Nagy et al. 2021); Global Environmental Monitoring Index (GEMI) (Kobayashi et al. 2020); Perpendicular Vegetation Index (PVI) (Lu et al. 2016); Ratio Vegetation Index (RVI) (Quan et al. 2011); and Leaf Area Index (LAI) (Gasol et al. 2019). Using satellite images and VIs allows the farmers to identify different management zones on a commercial farm (Campillo et al. 2018). These indices were used to estimate vegetation characteristics as they mostly serve as indicators for crop dynamics and overall changes in biomass quantity and properties. Moreover, some indices are used to monitor changes in the water content of leaves, while others can suppress the soil's influence or eliminate the influence of the atmosphere.

Several kinds of literature reported that VI-based crop yield estimation is very popular in RS domains for wheat (Nagy et al. 2018), corn (Panda et al. 2010), maize (Fang et al. 2011), rice (Son et al. 2013), sunflower (Ali et al. 2019). In recent years, vegetation reflectance values have been further analyzed through geostatistical and machine learning techniques as an alternative approach for better crop yield prediction (van Klompenburg et al. 2020). Koutsos et al. (2021) applied the statistical hotspot autocorrelation to identify low-yield areas for special attention to the better management of the agronomic inputs as a sustainable approach. LAI and NDVI are

good predictors of crop yield through the logistic regression model (Kunapuli et al. 2015). Marino and Alvino (2021) stated that statistical hotspot analysis is a sound strategy for defining the spatial differentiation of crop yield variation on a small scale with the help of vegetation indices. The US and German organic agriculture used the clustering hotspot algorithm to better support their regional economic development (Marasteanu and Jaenicke 2016). The GIS-based interpolations techniques such as IDW, EBK, and Kriging are the most preferred technique for multi-crop yield mapping (McKinion et al. 2010). RF and SVM allowed predictive models of crop yield estimation using multi-temporal data for site-specific management in different seasons (Shah et al. 2018; Filippi et al. 2019). Even the logistic regression (LR) and Naïve Bayes (NB) models correctly predicted the corn nitrogen stress class with the help of hyperspectral sensors derived VIs (Laacouri et al. 2018; Mupangwa et al. 2020). MODIS-derived LAI estimated rice crop yield on a near real-time basis using gradient boosted regression in India (Arunugam et al. 2021). ML models are very helpful in predicting the number of crop yields for supporting food security in Africa (Chepngetich 2020; Cedric et al. 2022). Different vegetation indices, namely NDVI, OSAVI, RSI, MTCI, and BOP, incorporated with BP neural network, significantly impact the cotton yield prediction mapping in China. Ramos et al. (2020) used multispectral UAV-based 33 VIs to predict maize yield through the ranking-based RF model, where the NDVI, NDRE, and GNDVI VIs are the most precedence indices. On the contrary, Pham et al. (2022) showed that the ML approach is useful for enhancing the rice crop yield through the spatial variability of VCI/TCI data in Vietnam. Five MLS, namely elastic net (EN), linear regression (LR), support vector regression (SVR), and k-nearest neighbor (k-NN) are tested for the potato tuber quality mapping with the NDVI value in Atlantic Canada (Abbas et al. 2020). Sharifi (2020) proposed that integrating satellite remote sensing and ML techniques provides a powerful potential approach for a barley

yield prediction map based on seasonal phenological behavior in eastern Iran.

S2 MSI data with 10 m resolution has a greater potential to build a decision-making tool for crop yield estimation through numerous vegetation spectral indices. With the accessibility of open-source Python and SNAP tools along with Geostatistics, ML, and area under the curve (AUC) models, we tried to integrate several biophysical spectral indices for yield estimation of rice and potato crops in Hooghly District (West Bengal, India). Nine major crop biophysical parameters were identified, namely NDVI, NDWI, BI, DVI, SAVI, GEMI, PVI, RVI, and LAI, for accurate crop yield estimation during two crop seasons (summer, 2018, and winter, 2019).

8.2 Materials and Methods

8.2.1 Study Area

The study was carried out in the Hooghly Region (Tarakeswar ~ 2,528,500 to 2,530,600 N and 604,500 to 606,500 E in UTM-WGS84 Zone 45 N India) during the 2018 and 2019 crop seasons (Fig. 8.1). The study region experienced rich fertile alluvial soil with high productivity of rice and potato compared to other districts of West Bengal (Singha et al. 2020). The densely populated region is supported by high cropping intensity (184%). The total study area was around 300 ha with an average MSL of 18 m.

This area is categorized by tropical monsoon climate with an average annual rainfall of 1200–1700 mm and temperature of 15–35 °C (Singha et al. 2019). Rice and potato are the main crops, along with jute and vegetables. The growing season for *Kharif* rice starts with the arrival of the monsoon in July, and crops are harvested from October to December. In the *Rabi* season, the potato crop is sown in November–December and harvested in March next year. The potato crop has many advantages over other local crops in terms of high crop yield, easy market access, short crop growth, etc.

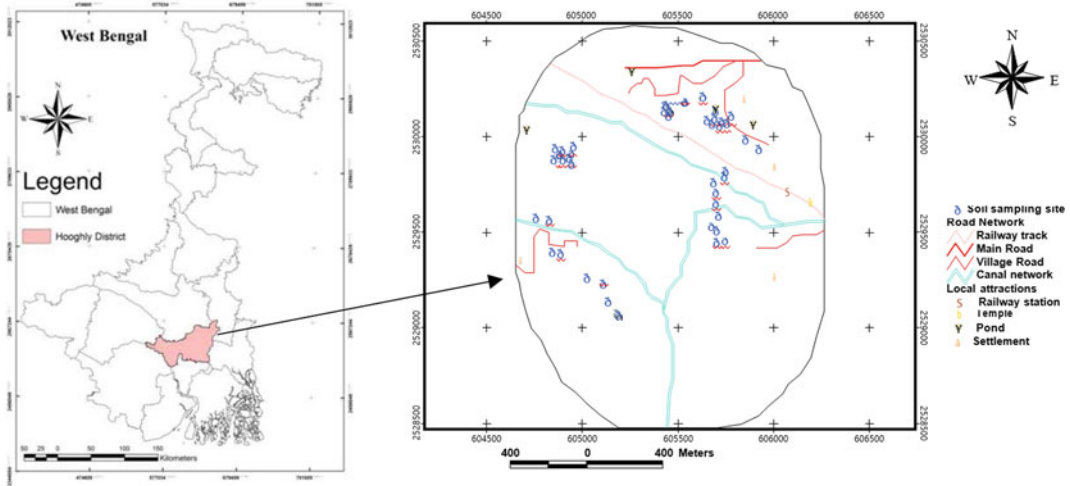


Fig. 8.1 Study area map

8.2.2 Satellite Image Processing

Cloud-free Sentinel-2B MSI imagery was selected for the *Kharif* rice (October 17, 2018) and *Rabi* potato (March 6, 2019) season, with a temporal resolution of 10 days and 290 km swath width. The images were processed for crop biophysical status mapping and yield estimation of two crops. The descending node of S2B MSI offers a 5-day temporal frequency with an orbital overpass time of approximately 10:30 a.m. (Drusch et al. 2012).

The Copernicus Open Access Hub data is easily retrieved by being atmospherically and radiometrically corrected S2B bottom of atmosphere (BOA) reflectance with 10 m pixel size imagery (<https://sci-hub.copernicus.eu/>). The Sentinel Application Platform v.8.0.0 (SNAP-ESA) was used for image pre-processing. The orthorectified images were geo-referenced in WGS84 UTM zone 45 N with Survey of India (SOI) topographical maps (No. 79B/1, 1:50,000 scale). The nine ground control points (GCPs) developed the final study area location map with a handheld GPS receiver e-Trex 20 Garmin). The detailed workflow of this research methodology is presented in Fig. 8.2.

8.2.3 Vegetation Indices (VIs)

This chapter selected the spatial resolution of 10 m for different VIs as input parameters for estimating the spatial variability of rice and potato yield, namely slope-based- (NDVI and RVI), distance-based—PVI, soil noise-based—SAVI, and NDWI, BI, GEMI, LAI (Hatfield et al. 2008). All the VIs maps were developed through ArcGIS 10.7 software for individual farm plots extracted from S2 images for two peak growing seasons. VIs are developed as the combination of numerous wavebands (red, NIR, SWIR, green, blue) and is also related to various canopy estimated parameters (Table 8.1).

8.2.4 Yield Estimation

During the post-harvest processing, a set of randomly selected 70 farms plot-wise agronomic practice details, along with crop yield information for two crops, were collected. The crop yield data were collected along with a handheld e-Trex GPS receiver (Garmin Ltd., Olathe, Kansas, USA) in October 2018 for *Kharif* rice and March 2019 for *Rabi* potato. Rice yield was estimated

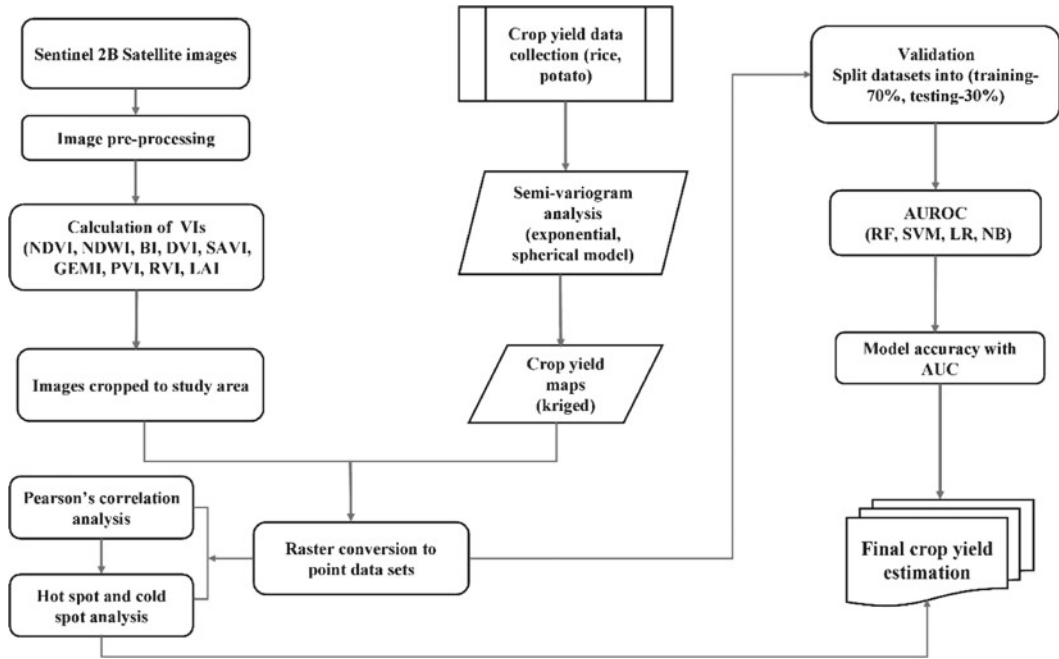


Fig. 8.2 Detailed workflow of the research work

Table 8.1 Details of selected vegetation indices (VIs)

Indices	Explanation	Formula	Source
NDVI	Normalized difference vegetation index	$\rho_{NIR} - \rho_{red} / \rho_{NIR} + \rho_{red}$	Vannoppen et al. (2018)
NDWI	Normalized difference water index	$\rho_{NIR} - \rho_{SWIR} / \rho_{NIR} + \rho_{SWIR}$	Chandel et al. (2019)
BI	Brightness index	$(\rho_{red} - \rho_{SWIR}) - (\rho_{NIR} + \rho_{blue}) / (\rho_{red} + \rho_{SWIR}) + (\rho_{NIR} - \rho_{blue})$	Cavalaris et al. (2021)
DVI	Difference vegetation index	$\rho_{NIR} - \rho_{red}$	Kussul et al. (2020)
SAVI	Soil-adjusted vegetation index	$1.5 \times (\rho_{NIR} - \rho_{red}) / (\rho_{NIR} + \rho_{red} + 0.5)$	Nagy et al. (2021)
GEMI	Global environmental monitoring index	$\eta(l - 0.25\eta) - (red - 0.125) / (l - red)$ $\eta = [2(NIR2 - red2) + 1.5NIR + 0.5red] / (NIR + red + 0.5)$ 9	Kobayashi et al. (2020)
PVI	Perpendicular vegetation index	$\sqrt{(\rho_{soil} - \rho_{veg})_{red}^2 - (\rho_{soil} - \rho_{veg})_{NIR}^2}$	Xue and Su (2017)
RVI	Ratio vegetation index	ρ_{red} / ρ_{NIR}	Quan et al. (2011)
LAI	Leaf area index	$3.618 * EVI - 0.118$	Gaso et al. (2019)

using a 10 m × 10 m square quadrant at five places for each plot. The yield data is converted into kg/ha. Similarly, potato yield was estimated for five rows, and the crop row's dimensions (length and width) were measured. The yield data in a point vector format at a spatial resolution were converted to raster GeoTiff at 10 × 10 m pixel resolution corresponding to the Sentinel-2 image pixels. Then, the yield maps were developed by the Kriging interpolation technique through ArcGIS 10.7.

8.2.5 Statistical Analysis

Statistical analysis was carried out to validate the relationship between the VIs, generic crop yield, and estimate yield data for two crops. The normality of the dataset was organized by descriptive statistics integrating minimum, maximum, range, mean, standard deviation (SD), skewness, kurtosis, and coefficient of variation. All the statistical calculations were performed in the 'R' software v.4.0.5 (University of Auckland, New Zealand) environment.

8.2.5.1 Geostatistical Analysis

ArcGIS Geostatistical Analyst tool was also utilized to study the degree of spatial dependence between Sentinel MSI-based VIs and yield data in the ArcGIS environment. The gradation of consistency for rice and potato yielding was determined using the standard equation using the geostatistical semivariogram model (Eq. 8.1) (McKinion et al. 2010).

$$\gamma(h) = \frac{1}{2N(h)} + \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (8.1)$$

where $\gamma(h)$ represents semivariance, $Z(x_i)$ is the distance between the measured sample at point X_i , and the point $Z(x_i + h)$ is the sum of the pairs separated by the lag h .

The semivariogram was fitted using exponential, spherical, Gaussian, and linear models. The residual sum of squares (RSS) was used to

determine the form and goodness of fit of the model semivariogram (Eq. 8.2).

$$RSS = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (8.2)$$

where y_i is the i th value of the variable to be predicted, x_i is the i th value of the explanatory variable, and $f(x_i)$ is the predicted value of y_i .

The statistical cross-validation method was used for the performance of the semivariogram model. The prediction accuracy of crop yield and VIs was assessed through the coefficient of determination and root mean square error value. The best model accuracy indicated the maximum R^2 and least RMSE value for a final agreement between crop yield and VIs parameters.

The sill, easily-defined range explains the exponential and spherical models for the plant and soil variability fitting (McKinion et al. 2010). Our study structured the three parameters when fitting the best semivariogram model, namely (i) Nugget C_0 represents short-scale randomness, (ii) Sill ($C_0 + C_1$) of the semivariogram is equal to the variance of the random variable when growing beyond a certain distance and becomes less or steadier around an edge value, (iii) range is defined the correlation between two properties inclined to be equivalent to zero when the distance h becomes too large. In semivariogram analysis, the trends of spatial dependence (SD) are measured by the nugget/sill ratio $C_0/(C_0 + C_1)$, %. It is described in terms of three types of spatial dependency levels; (i) strong SD < 25%; (ii) moderate SD—25–75%; (iii) weak SD > 75% (Singha et al. 2020). The geostatistical kriging analysis is incorporated with the semivariogram that estimates the known values to unknown values for spatial judgment of crop yield variation at a similar level of resolution (Li et al. 2016). Yield maps converted into raster form correspond to the same spatial resolution of S2 images for extracting multi-point values. Then Pearson correlation matrix showed the relationship between the VIs and yield for the two crops.

8.2.5.2 Hotspot Analysis

Hotspot analysis is a statistical procedure that identifies statistically significant cold and hot spots through the Getis-Ord G_i^* statistic. It utilizes a set of features that are weighted to identify these cluster areas. These statistics have three components: G_i Z Score, G_i p-value, and G_i -Bin values of the selected criteria. The resultant G_i Z Score and G_i p-Value define the high or low values of the cluster in a spatial process. A high Z score and a small p-value indicate a significant hot spot positively correlated with a low negative Z score. The more intense the hotspot or coldspot clustering, the more it is distributed using kriging interpolation within the region. This approach identifies high and low clusters of yield data correspondence to VIs with similar areas. The ArcGIS spatial statistics tools did spatial autocorrelation of the VIs in hotspots analysis in the mapping cluster approach. The Getis-Ord G_i^* statistic was explained in (Eqs. 8.3, 8.4, and 8.5) (Abdulhafedh 2017).

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}x_j - \bar{X} \sum_{j=1}^n w_{ij}}{\sqrt{\frac{s}{n-1} \left[n \sum_{j=1}^n w_{ij}^2 - \left(\sum_{j=1}^n w_{ij} \right)^2 \right]}} \quad (8.3)$$

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (8.4)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (8.5)$$

where G^* indicates the G statistics of i , x_j is the VIs and yield of j , w_{ij} describes the spatial weight between VIs and yield of i and j , \bar{X} , and S represent mean, variance, and n indicate the sum of the VIs and yield parameters.

When evaluating crop yield and VIs agricultural data with their surrounding neighbors, the hotspots represent positive autocorrelation with high production, and coldspots represent negative autocorrelation with low production in the study area. The attempts were made to carry out hotspot analysis for point data on a microscale.

8.3 Validation

Validation is carried out using the survey training dataset of rice and potato yield information collected from seventy farm plots in the study area. For validation purposes, we use the area under the receiver operating characteristic (AUROC) curve incorporated into the four ML techniques, namely (i) logistic regression (LR), (ii) support vector machine (SVM), (iii) random forest (RF), and (iv) Naïve Bayes (NB) which measured the consistent relationship between yield data and VIs.

LR is a classification algorithm that calculates a predicted probability for a dichotomous dependent variable based on one or more independent variables (Ahmed and Sajjad 2018). Authentication of the LR model is regulated by the binary composition that indicates the probability of incidence with pair sample space, P (Eq. 8.6)

$$P(y = \pm 1|x, w) = \frac{1}{1 + \exp(-y(w^T x + b))} \quad (8.6)$$

where b signifies the intercept, T signifies the transfer matrix, and the k -dimensional coefficient, $w = (w_1, w_2, \dots, w_k)^T$, comprises the essential parameters to be assessed. Similarly, SVM, RF, and NB models can be estimated using their respective standard equations.

The model's prediction accuracy was estimated at 70% for training and 30% for testing data through the area under the AUROC curve (Eq. 8.7) (Allen 2015).

$$AUC = \frac{\sum TP + \sum TN}{P + N} \quad (8.7)$$

where TP—true positive and TN—true negative represents the number of farm plots correctly predicted crop yield, P is the total number of plots with heavy specs, and N is the total number of plots without substantial specs.

All-model performances were optimized for choosing the best tuning hyperparameter to

determine the high accuracy (AUC) results (Table 8.2). The AUROC curve was produced by the Anaconda python Jupyter notebook v. 6.0.1.

8.4 Results and Discussion

8.4.1 Descriptive Statistical Analysis

Descriptive statistical analysis was carried out for all the selected VIs and crop yields (Table 8.3). The mean rice and potato yield rate of the study area was 7.11 and 28.04 (t/ha), respectively. The spatial variability of crop yield and VIs data were good estimators of the ecological vegetation process that resulted in a standard deviation (SD), verified a CV value of 23.25% for rice yield and 29.88% for potato crop yield. The VIs of NDVI, NDWI, BI, DVI, GEMI, SAVI, PVI, RVI, and LAI had low CVs in rice crop; similarly, the NDVI, DVI, SAVI, PVI, and RVI had a moderate level of CVs for the potato crop. The NDVI and SAVI ranged between 0.45 to 0.58 and 0.27 to 0.34 for rice crops, and similarly for potato crops between 0.16 to 0.55 and 0.11 to 0.35. The high mean values for RVI were around 3.33, followed by LAI (2.06), GEMI (0.64), and NDWI (0.62) in the rice crop, whereas the potato crop is high mean values of RVI (2.25), followed by LAI (0.80), GEMI (0.56), and NDVI (0.38), respectively. The

positive (1.29) and negative (− 2.08) skewness were found with BI and NDWI for the rice crop. However, the negative (− 0.32) skewness found in NDVI and positive (0.44) in the LAI was found in potato crop, creating the symmetrical type distribution. The kurtosis interrelated to DVI, SAVI, and PVI showed platykurtic behavior associated with normal distribution for rice crops. In the case of NDWI and BI, high kurtosis (− 0.89 and 0.27) was observed, which may represent the leptokurtic behavior of distribution for the potato crop.

8.4.2 Relationship Between VIs and Crop Yields

Pearson correlations matrix described the interrelation between the selected VIs and crop yield for two specific crop seasons (Tables 8.4 and 8.5). Higher r^2 values reported the most significance in crop growth and yield patterns around the experimental farm plots. Rice yield was significant/positively correlated with NDVI, and RVI, while a moderate correlation with SAVI, DVI, and PVI. VIs of RVI, NDVI, SAVI, DVI, and PVI were best positively correlated with the actual rice yield with high r values that ranged from 0.80 and 0.92 (Table 8.4).

Similarly, potato yield was strongly correlated with NDVI and moderately correlated with RVI,

Table 8.2 Optimization of parameters for AUROC validation

Model	Tuning parameter	Feature selection
RF	n_estimators	10
	max_features	Auto
	Max depth	5
	Criterion	'gini'
SVM	Kernel	'rbf'
	Gamma	2
	C	1
	Degree	3
NB	var_smoothing	1e−09
LR	C	1
	Solver	'lbfgs'
	Intercept	24.05

Table 8.3 Descriptive statistical analysis of selected VIs and crop yields (tons/hector)

Crops	Parameters	NDVI	NDWI	BI	DVI	GEMI	SAVI	PVI	RVI	LAI	Yield
Rice	Min	0.45	0.42	0.08	0.14	0.55	0.27	0.10	2.64	1.26	3.40
	Max	0.58	0.68	0.11	0.19	0.68	0.34	0.13	3.74	2.68	12.00
	Mean	0.54	0.62	0.09	0.17	0.64	0.31	0.12	3.33	2.06	7.12
	SD	0.03	0.04	0.00	0.01	0.03	0.02	0.01	0.23	0.26	2.01
	Skewness	- 1.19	- 2.08	1.29	- 0.33	- 1.62	- 0.55	- 0.34	- 0.77	- 0.40	0.53
	Kurtosis	2.34	8.21	3.21	- 0.33	3.96	- 0.03	- 0.34	1.18	0.86	0.25
	CV	4.66	7.07	4.44	6.47	3.90	5.75	6.67	6.76	12.83	28.25
Potato	Min	0.16	- 0.02	0.10	0.06	0.43	0.11	0.05	1.46	0.36	9.90
	Max	0.55	0.55	0.14	0.21	0.69	0.35	0.15	3.37	1.43	39.00
	Mean	0.38	0.27	0.12	0.13	0.56	0.22	0.09	2.25	0.80	28.04
	SD	0.09	0.16	0.01	0.04	0.06	0.06	0.03	0.49	0.29	8.38
	Skewness	- 0.32	- 0.25	1.14	0.19	0.08	0.03	0.20	0.31	0.44	- 0.47
	Kurtosis	- 0.37	- 0.90	0.26	- 0.74	- 0.34	- 0.81	- 0.73	- 0.65	- 0.39	- 0.54
	CV	24.23	59.26	8.42	28.50	10.26	27.18	28.50	21.73	36.17	29.88

N.B. *SD* standard deviation; *CV* coefficient of variation; *n* No. of sample

Table 8.4 Pearson correlations matrix between VIs and rice yield (*n* = 70)

	NDVI	NDWI	BI	DVI	GEMI	SAVI	PVI	RVI	LAI	Rice yield
NDVI	1.000									
NDWI	0.704**	1.000								
BI	- 0.701**	- 0.770**	1.000							
DVI	0.821**	0.353*	- 0.178	1.000						
GEMI	0.542**	0.248	- 0.116	0.661**	1.000					
SAVI	0.908**	0.470**	- 0.332*	0.984**	0.645**	1.000				
PVI	0.819**	0.354*	- 0.175	0.999**	0.665**	0.983**	1.000			
RVI	0.996**	0.688**	- 0.681**	0.836**	0.565**	0.918**	0.833**	1.000		
LAI	0.804**	0.850**	- 0.696**	0.555**	0.615**	0.648**	0.555**	0.813**	1.000	
Rice yield	0.895**	0.551**	- 0.563**	0.806**	0.582**	0.862**	0.800**	0.920**	0.746**	1.000

** Correlation is significant, 0.01 level; * Correlation is significant, 0.05 level

SAVI, DVI, PVI, and LAI of *r* values from 0.761 to 0.908 (Table 8.5). BI was insignificant/negatively correlated with both the crop yield (rice and potato) because of the variability of canopy reflectance and light use efficiency (LUE) with site-specific field management.

8.4.3 Geostatistical Analysis

Plot-based yield information and extracted VIs datasets were analyzed using the geostatistical semivariogram method (Table 8.6). The lag size of both crops was between 14.90 and 64.46,

Table 8.5 Pearson correlations matrix between VIs and potato yield ($n = 70$)

	NDVI	NDWI	BI	DVI	GEMI	SAVI	PVI	RVI	LAI	Potato yield t/ha
NDVI	1									
NDWI	0.899**	1								
BI	- 0.854**	- 0.866**	1							
DVI	0.938**	0.917**	- 0.830**	1						
GEMI	0.880**	0.934**	- 0.738**	0.946**	1					
SAVI	0.944**	0.928**	- 0.862**	0.998**	0.940**	1				
PVI	0.938**	0.917**	- 0.830**	1.000**	0.945**	0.998**	1			
RVI	0.934**	0.928**	- 0.857**	0.993**	0.936**	0.994**	0.994**	1		
LAI	0.909**	0.971**	- 0.835**	0.943**	0.950**	0.947**	0.943**	0.961**	1	
Potato yield	0.908**	0.783**	- 0.781**	0.808**	0.741**	0.821**	0.807**	0.801**	0.761**	1

** Correlation is significant, 0.01 level; * Correlation is significant, 0.05 level

that's were depending on the spatial variability of the crop, VI, and soil nutrients. The high lag size was 57.76 for rice (BI) and 64.46 for potato (NDWI) with a spherical semivariogram. Associating range values of rice VIs, the lowest value was measured for GEMI (95.59 m), and the highest value was measured at NDWI (255.01 m). Similarly, the VIs of potato crops varied amply between 246.46 and 349.83 m. Nuggets were highest for PVI (5.29) and lowest for BI, GEMI (0.0001) of rice crops, where the potato crop has the lowest nugget for GEMI variables. The nine VIs of the rice crop revealed a sill size between 0.002 (BI, GEMI) and 6.42 (PVI), which stated a relatively parallel total variance. The VIs of potato crop sill varied from 0.008 (GEMI) to 0.093 (NDVI). The ratio of percentage between the two parts, a nugget to sill variance, ranged noticeably between RVI (0.039%) and PVI (82.37%), where C_0 represented 0.039% and 82.37% of $C_0 + C$ with strong spatial dependency (0.41%) for rice yield, and the other potato crops where C_0 denoted 9.87% to 60.36% of $C_0 + C$ with moderate spatial dependency (34.27%) for potato yield. The exponential and spherical models are very good of a fit in semivariogram analysis with the lowest RMSE value, which means they are highly

significant. Exponential models were suitable alternatives to the experimental semivariograms for NDVI, NDWI, DVI, SAVI, PVI, RVI, and LAI, whereas BI and GEMI values were best fitted with a spherical model for rice crops (Fig. 8.3). NDVI and BI are very well-explained by the exponential models; whereas NDWI, DVI, GEMI, SAVI, PVI, RVI, and LAI, are the best suited to the spherical model for potato crops (Fig. 8.3). R^2 was always > 0.75 estimated for the crop yield, with an error of RMSE very low, particularly concerning average VI. The best R^2 values of 0.846, 0.801, and 0.743 were found for RVI, NDVI, and SAVI, respectively, while a weak R^2 was found with NDWI and BI for rice crops (Figs. 8.3 and 8.4). On the other hand, the NDVI, SAVI, DVI, PVI, and RVI were associated with the highest corresponding R^2 values of 0.824, 0.674, 0.653, and 0.652 for the potato crop. The RMSE of VIs ranged between 0.004 and 0.41, concerning rice yield of 1.77 t/ha. Similarly, the VIs of potatoes range between 0.006 and 0.29, with a potato yield of 8.4 t/ha (Table 8.6). The maps indicated that the north-eastern field had a lower crop yield than the southwestern part of the field. The spatial variability of the crop was also exhibited in the generated rice and potato yields.

Table 8.6 Semivariograms model performance with the exponential and spherical models for both the crop VIs and yields analysis

		Model	C_0	C_1	$(C_0 + C_1)$	$C_0/(C_0 + C_1)$ (%)	SD	a (m)	Lag size	R^2	RMSE
Rice	NDVI	Exponential	0.001	0.002	0.003	39.604	M	250.126	20.843	0.80	0.024
	NDWI	Exponential	0.001	0.01	0.011	4.757	S	255.01	21.25	0.30	0.035
	BI	Spherical	0.0001	0.001	0.002	18.072	S	171.72	57.76	-0.32	0.004
	DVI	Exponential	0.003	0.001	0.005	75.764	W	111.36	17.17	0.65	0.012
	GEMI	Spherical	0.0001	0.002	0.002	13.634	S	95.592	14.9	0.74	0.021
	SAVI	Exponential	0.002	0.001	0.004	64.12	M	266.8	22.23	0.34	0.018
	PVI	Exponential	5.286	1.131	6.417	82.375	W	126.35	17.175	0.64	0.008
	RVI	Exponential	0.0022	0.0034	0.0057	0.0392	S	240.629	20.052	0.85	0.0207
	LAI	Exponential	0.001	0.019	0.021	6.978	S	242.186	20.18	0.56	0.206
	Rice	Exponential	0.037	0.053	0.09	0.414	S	203.468	16.995		1.777
Potato	NDVI	Exponential	0.0092	0.084	0.0932	9.871	S	349.834	29.152	0.82	0.074
	NDWI	Spherical	0.0027	0.019	0.0217	12.442	S	279.85	64.464	0.61	0.0815
	BI	Exponential	0.005	0.008	0.013	38.462	M	361.3	30.108	-0.61	0.0067
	DVI	Spherical	0.01	0.07	0.08	12.5	S	247.15	31.703	0.65	0.0232
	GEMI	Spherical	0.0017	0.007	0.0087	19.54	S	260.74	27.686	0.67	0.0347
	SAVI	Spherical	0.0104	0.0749	0.0853	12.192	S	276.062	28.779	0.55	0.0395
	PVI	Spherical	0.0097	0.0711	0.0808	12.005	S	246.46	31.908	0.65	0.0164
	RVI	Spherical	0.0067	0.033	0.0397	16.877	S	265.344	34.255	0.64	0.2968
	LAI	Spherical	0.0166	0.0109	0.0275	60.364	M	289.486	30.697	0.58	0.1529
	Potato	Exponential	0.0629	0.1206	0.1835	34.278	M	368.364	30.697		8.4

N.B. C_0 Nugget; C_1 Partial sill; $(C_0 + C_1)$ Sill; $C_0/(C_0 + C_1)$ (%) Nugget/Sill; a Range (m), $RMSE$ root mean square error

Remote sensing-based S2-derived VIs maps and the responses crop yield maps were established for similar spatial resolution (Figs. 8.5 and 8.6). A total of nine VIs are selected through their associations with crop productivity. The study region also observed a spatial pattern consistent with crop yield distribution. High values found in the south–southwest part of the farm plot were associated with high crop yield compared to the northern part of the region.

8.4.4 Hot Spots Analysis

The concept of the Z score indicates a significant hotspot and a negative Z score indicates a significant coldspot in Getis-Ord G_i^* statistic. The analysis data revealed that the high-yielding

areas were mainly concentrated in the RVI, NDVI, and SAVI surfaces. This observation explained that high-yield clusters exist in the RVI, NDVI, and SAVI surfaces with significant cluster distribution.

This analysis suggested that RVI, NDVI, and SAVI indices were superior in guiding maps for ideal demonstration of the spatial variability for rice and potato yield patterns at small farm levels (Figs. 8.7 and 8.8). BI values are highly significant (hotspot) and range between 90 and 95% confidence level in the northeastern region due to soil humidity and pH variability with low crop yield for both crops (Fig. 8.7). Similarly, the other VIs (DVI, GEMI, PVI, and LAI) showed a highly significant (90–95%) cluster with high yield due to better crop practice management in the southwest part of the farm areas for both the

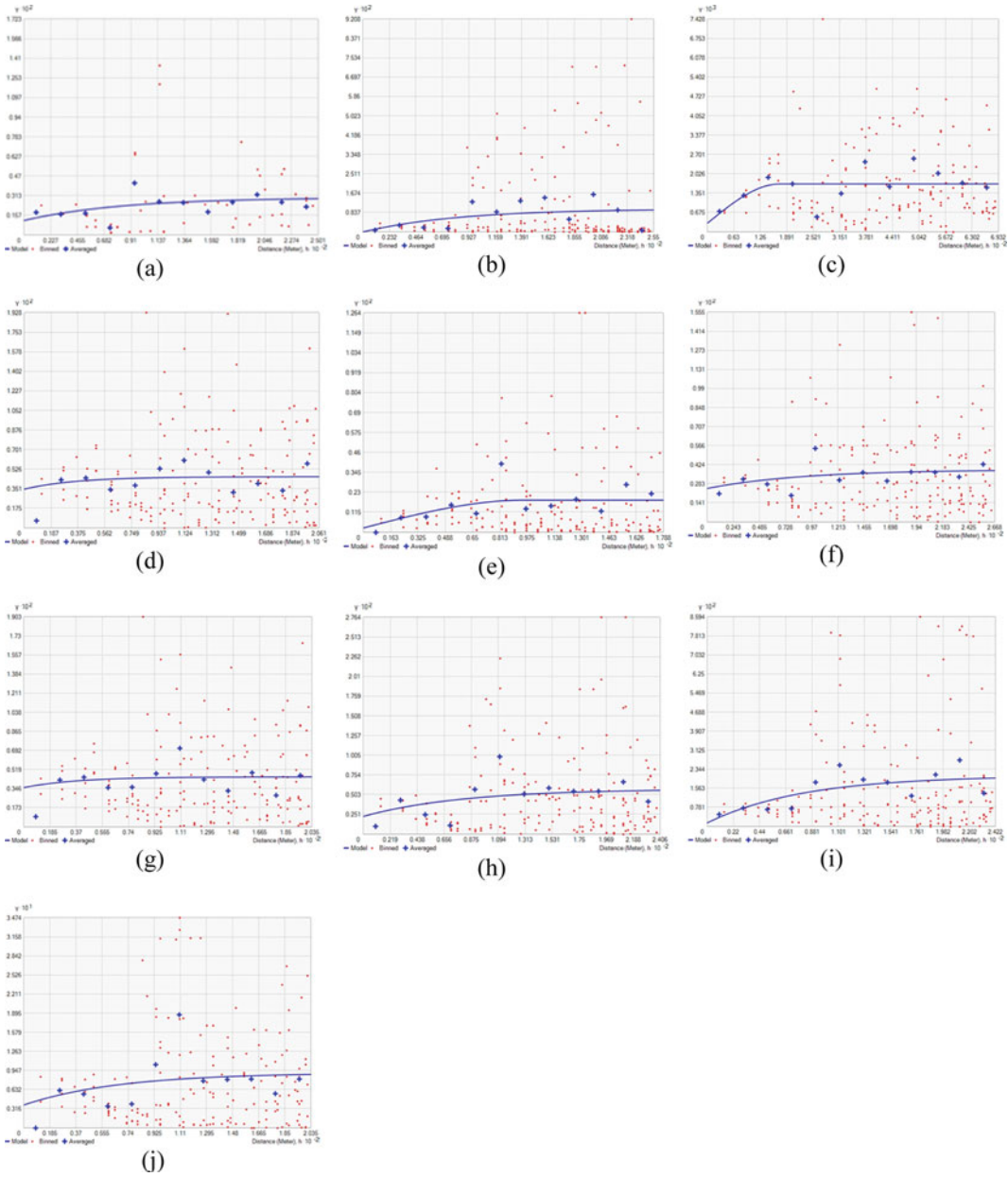


Fig. 8.3 Semivariograms graph for rice crop; **a** NDVI, **b** NDWI, **c** BI, **d** DVI, **e** GEMI, **f** SAVI, **g** PVI, **h** RVI, **i** LAI, and **j** rice yield

crop (Figs. 8.7 and 8.8). In this chapter, the hotspot analysis described the spatial pattern of both crop yield distribution with VIs: coldspot significance level found in the northern part of the farm

plot, and in contrary hotspot was in significance level in the south–southwestern side of the farming area. The significance levels of hotspot maps presented a high number of low-distant VIs.

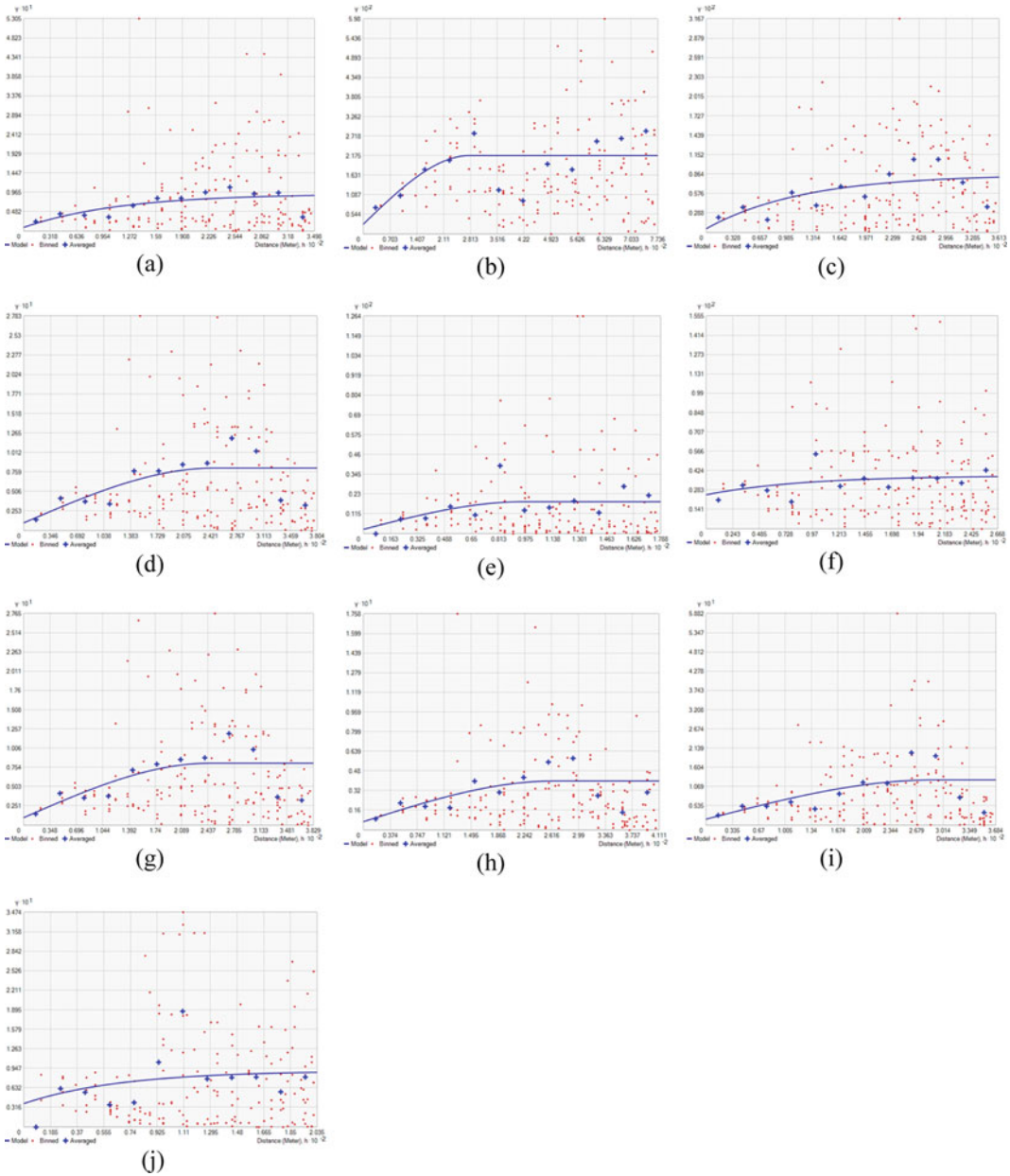


Fig. 8.4 Semivariograms graph for potato crop; **a** NDVI, **b** NDWI, **c** BI, **d** DVI, **e** GEMI, **f** SAVI, **g** PVI, **h** RVI, **i** LAI, and **j** potato yield

8.4.5 AUROC Validation

The outcomes of the validation model of AUROC with four MLAs techniques between the yield and VIs datasets were assessed for both crops (Fig. 8.9). SVM, LR, and RF algorithms led to

the spatial assessment of rice and potato crop yield estimation with AUC accuracy values between 80 and 90%, categorizing them as very good predictors, while NB had an accuracy value between 60 and 80%, categorized as a moderate to the good predictor for both the crops. More

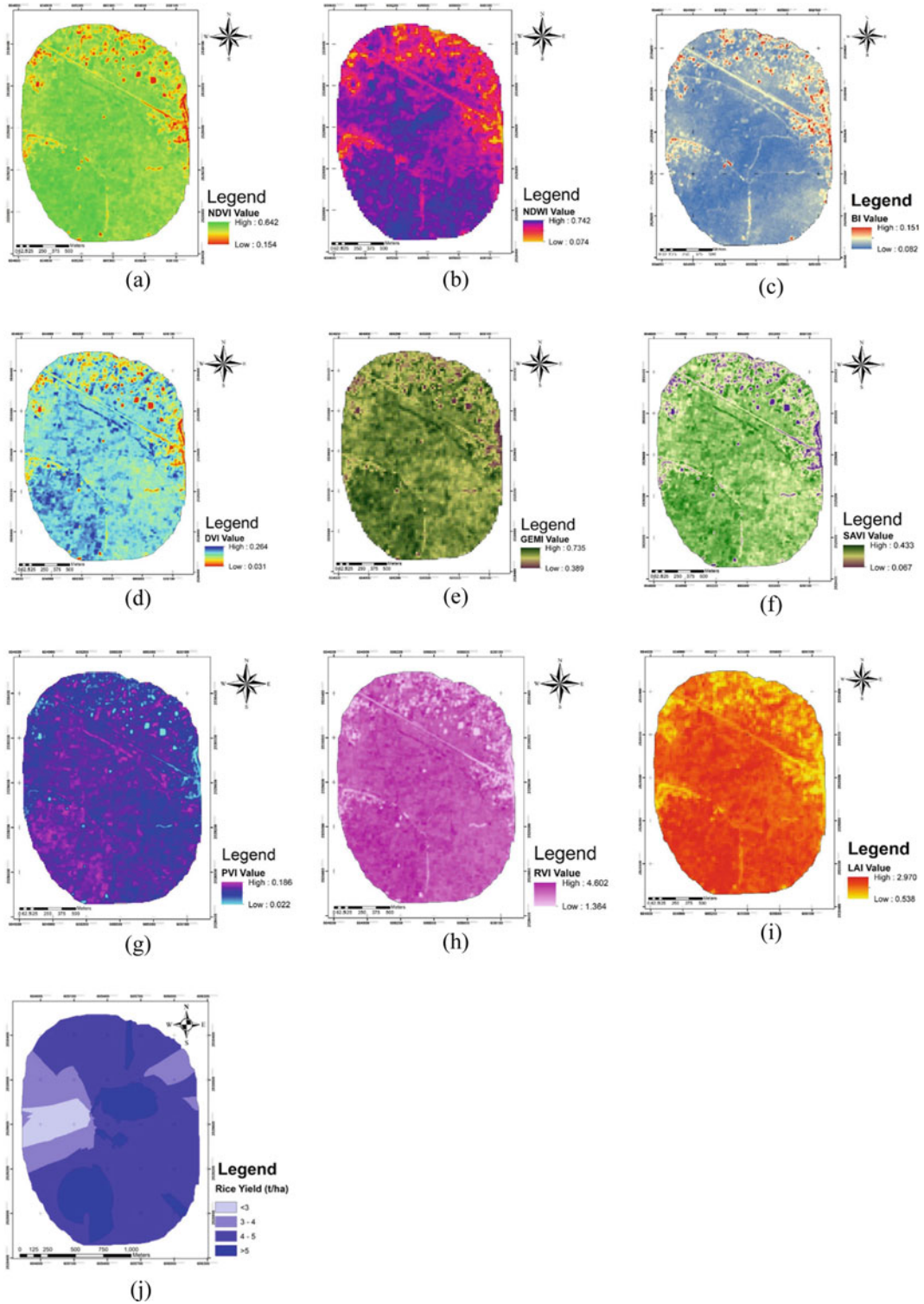


Fig. 8.5 Vegetation indices and yield distribution map for rice crop; **a** NDVI, **b** NDWI, **c** BI, **d** DVI, **e** GEMI, **f** SAVI, **g** PVI, **h** RVI, **i** LAI, and **j** rice yield

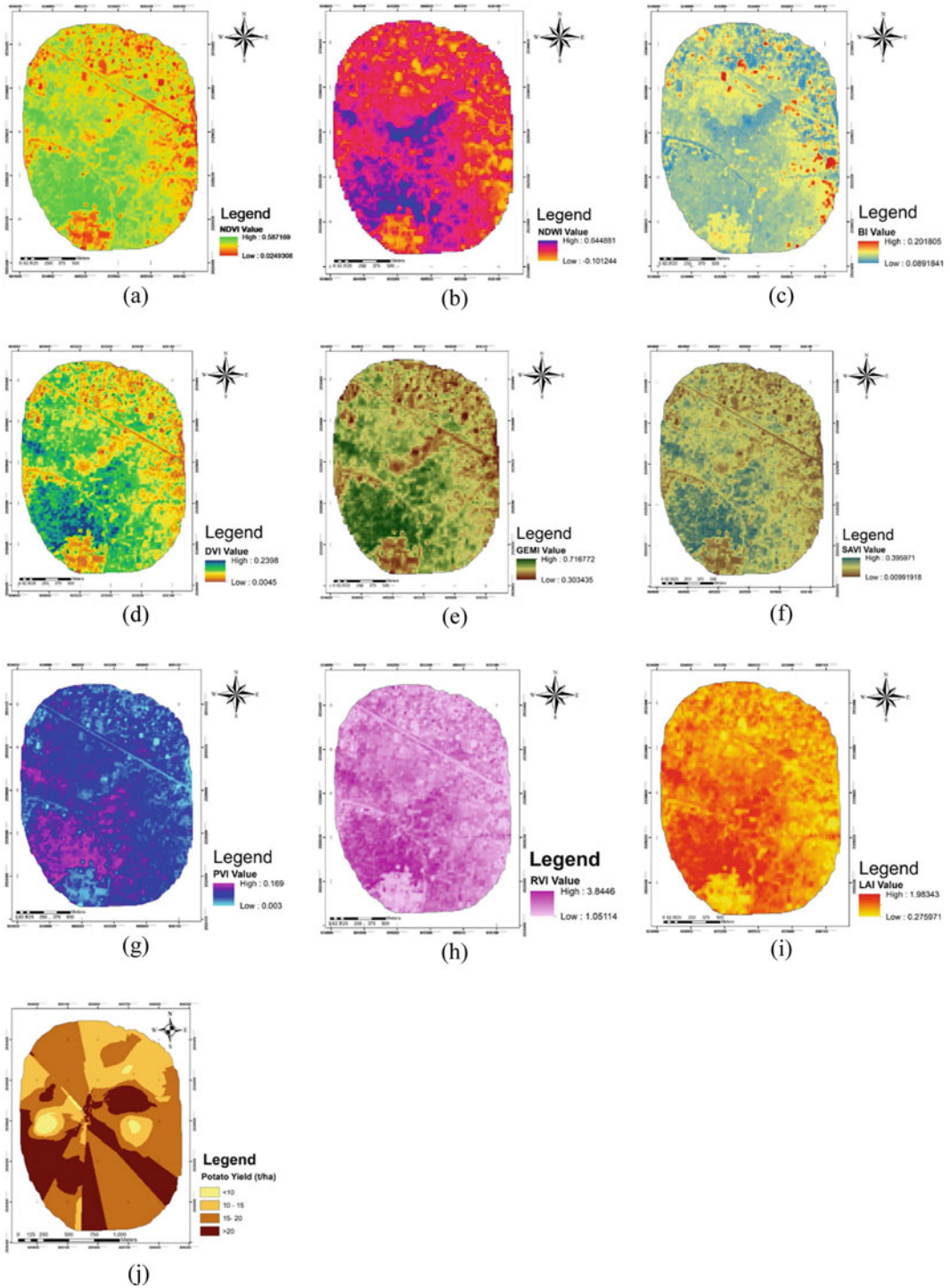


Fig. 8.6 Vegetation indices and yield distribution map for potato crop; a NDVI, b NDWI, c BI, d DVI, e GEMI, f SAVI, g PVI, h RVI, i LAI, and j potato yield

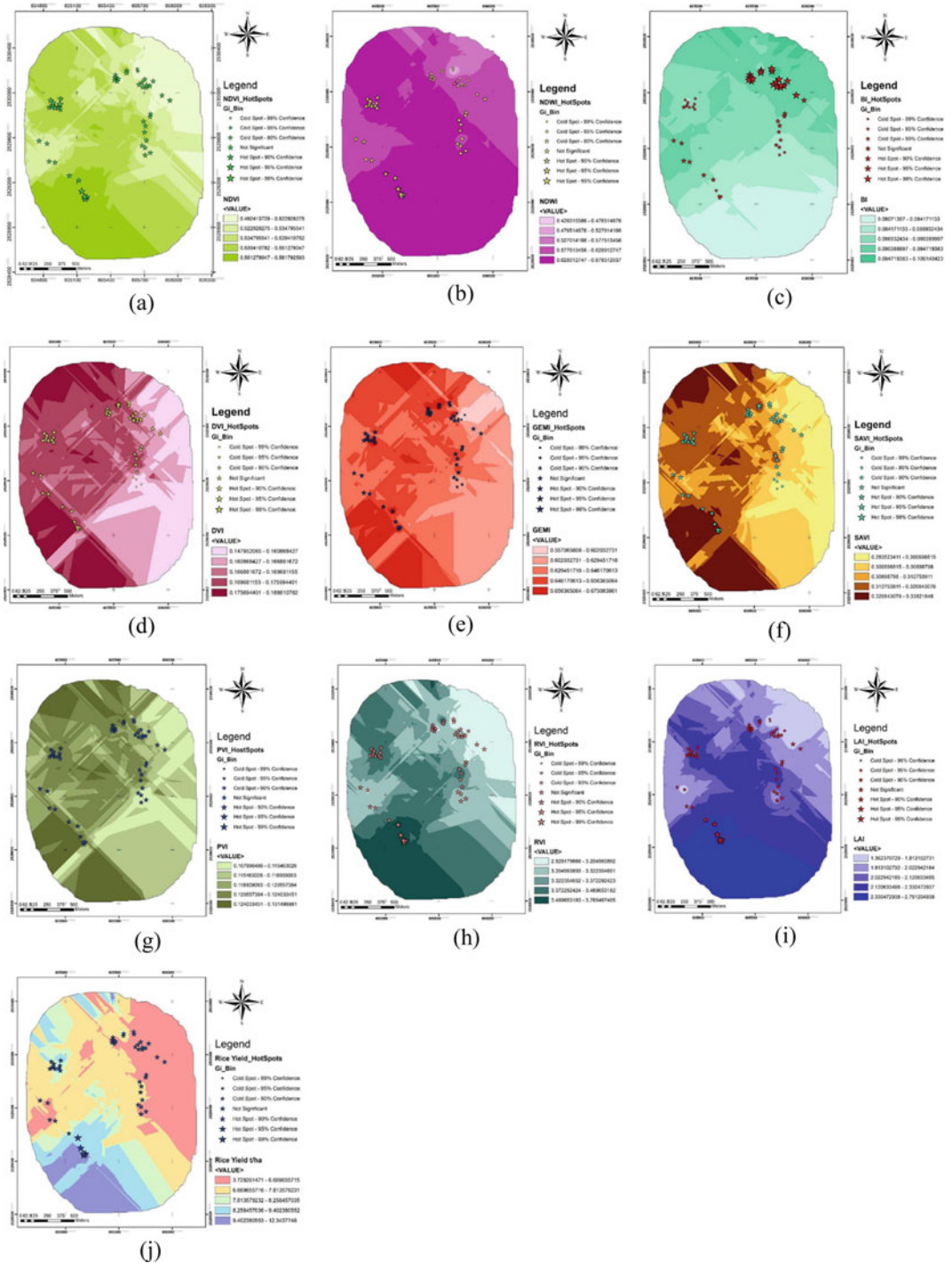


Fig. 8.7 Hotspot analysis for rice crop; a NDVI, b NDWI, c BI, d DVI, e GEMI, f SAVI, g PVI, h RVI, i LAI, and j rice yield

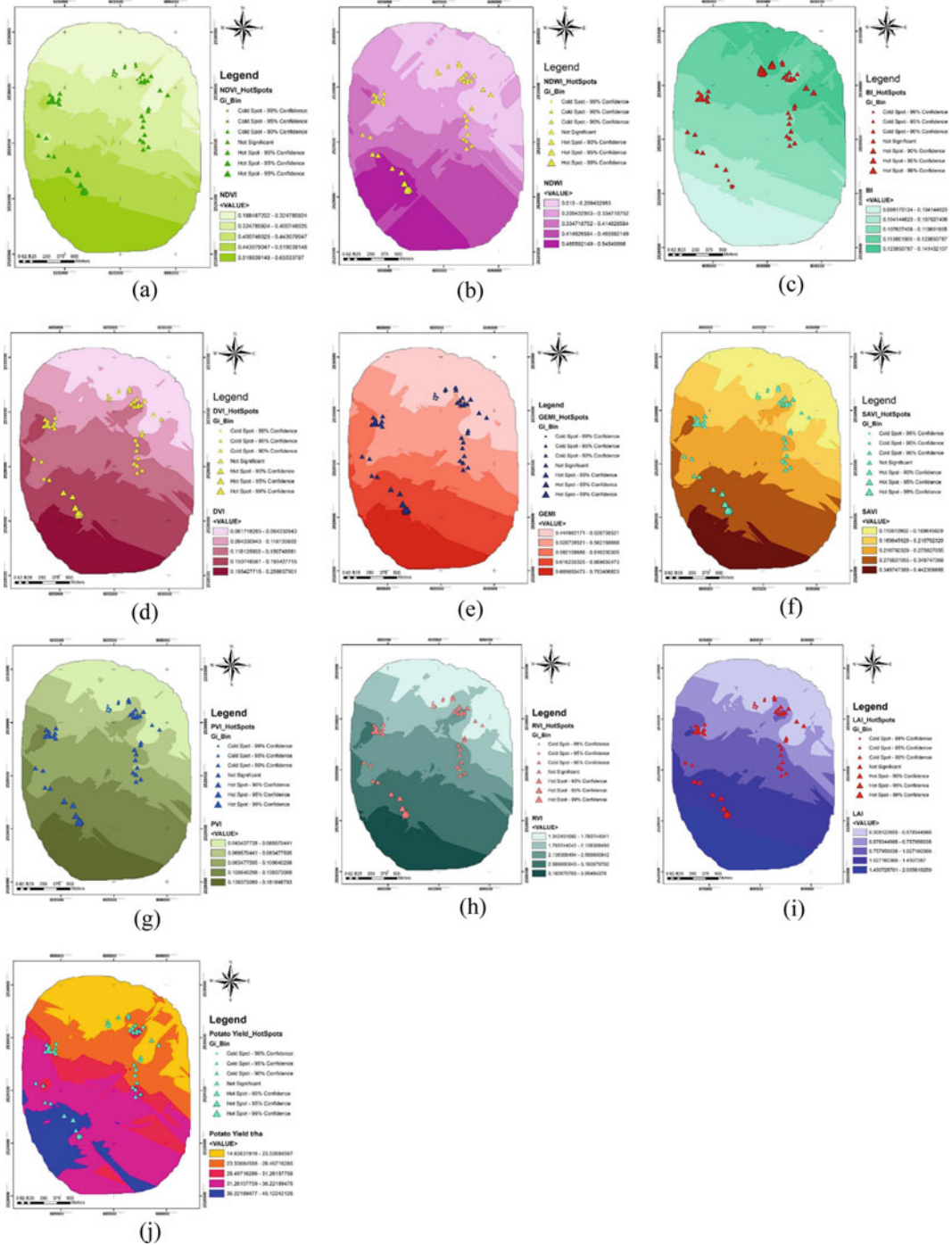


Fig. 8.8 Hotspot analysis for potato crop **a** NDVI, **b** NDWI, **c** BI, **d** DVI, **e** GEMI, **f** SAVI, **g** PVI, **h** RVI, **i** LAI, and **j** potato yield

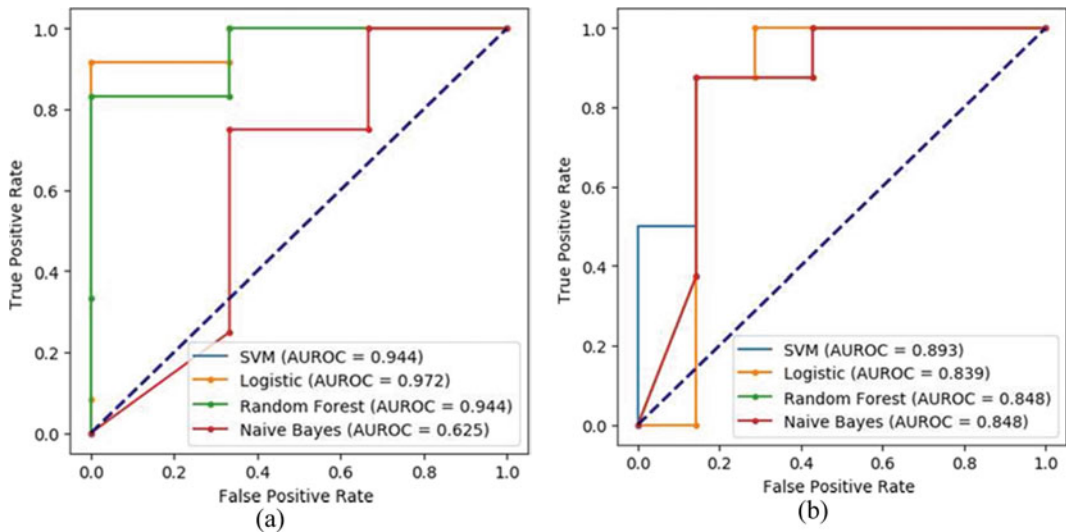


Fig. 8.9 AUROC validation between VIs and crop yield (rice and potato); **a** rice crop, **b** potato crop

precisely, the outputs revealed that the three MLAs techniques, SVM, LR, and RF algorithms, produced competitively better performances than the NB ML algorithm for both the crop yield data.

8.5 Discussion

We have used different spectral vegetation indices, estimated from the Sentinel-2B satellite data, to carry out the geostatistical-based crop yield estimation of rice and potato crops in two different seasons in Tarakeswar, Hooghly region (West Bengal). The spatial variability of crop yielding is delineated by the climate, topography, soil factor, water availability, and farm mechanization parameters (Singha and Swain 2022). A non-destructive yield prediction map was developed using a multiyear dataset. The map was developed using a geostatistical tool and an artificial intelligence model (Panek et al. 2020). A correlation coefficient analysis shows the spatial pattern of the VIs and yield. The VIs of RVI, NDVI, SAVI, DVI, and PVI were found to be significantly positive and correlated to the rice yield in the *Kharif* season.

The most commonly used VIs utilizes the information in the red and near-infrared

(NIR) canopy reflectance or radiances. They are combined in ratios: ratio vegetation index (RVI). RVI is the most frequently used VI with a significant correlation with grain yield (Pinter et al. 2003; Ali et al. 2019). The RVI, which measures the relative importance of rice yield to the farm plot, showed the best correlation with the actual rice yield, followed by the NDVI > SAVI > DVI > PVI > LAI > GEMI > NDWI > BI (Fig. 8.4). The plant signals attained from SAVI and NDVI had high correlation with the Sentinel-2-measured canopy reflectance during the *Kharif* rice and *Rabi* potato growing seasons. Similarly, during the *Rabi* seasons, the R^2 values of NDVI, SAVI, DVI, PVI, RVI, NDWI, and LAI decreased toward potato yield distribution. These relationships showed a sound correlation coefficient ranging between, 0.761 and 0.908, with potato crop yield.

A weak correlation was observed between the various parameters, namely NDWI, BI of crop concentration, and actual yield. It was attributed to the fact that the crop's maturity caused an increase in visible reflectance (Kumhálová and Matějková 2017). RVI, NDVI, and SAVI indices were the better-controlling factors to variate crop yield. On the other hand, for potato crops, the NDVI, SAVI, DVI, PVI, and RVI indices

were the highest R^2 values of 0.824, 0.674, 0.653, and 0.652, respectively (Table 8.6). Seasonal NDVI values indicate the best yield estimation factor for rice and potato crop (yield) predictions. The maximum NDVI was retained as the most significant variable for predicting field-level yield. Borowik et al. (2013) reported that the relationship between vegetation above-ground biomass and NDVI reflects each habitat type in seasonal variation. The selected NDVI and RVI indices have comparatively higher correlations with grain yield than SAVI (Ali et al. 2019).

Geostatistical semivariogram analysis was performed with best-fitting exponential and spherical models to develop the RS-based VIs and crop yield maps (Singha et al. 2020). VIs maps combined with RGB and NIR bands have a greater potentiality for plant sensitivity analysis in different crop growing seasons (Pinter et al. 2003). The ML methods such as support vector regression (SVR) and ensemble multilayer perceptron neural networks are successfully predicting the rice yield in the Cauvery Delta Zone (CDZ) and Gujarat India (Bhojani and Bhatt 2020; Yu et al. 2021). The chapter's contribution is simple and efficient in providing effective structures within appropriate site-specific agronomic management. The sill variance depends on canopy cover reflectance with the plant growth condition assumed by the VIs semivariograms analysis (Pradhan et al. 2014). Data of all VIs and crop yield were signified according to the Getis-Ord G_i^* statistic for identifying the degree of significance in the study area. In statistical hotspot analysis, high clusters of yield were detected in RVI, NDVI, and SAVI surfaces and low yield while the low significant clusters observed in RVI, NDVI, and SAVI surfaces were superior to guide map and best alternatives for rice and potato yield prediction in a small scale of village level. The hotspot analysis was used to define a cluster's higher or lower values in a spatial process, specifying the number of clusters to be detected (Marino and Alvino 2021). VIs map correspondence to yield map have similar trends in hotspot analysis. Generally, the northern part of the study area has some limitations

with coldspots due to high soil pH (acidic soil), least organic carbon, high electrical conductivity, and low farm mechanization level. Similarly, high yield and VIs values were found in the northwestern region with good correlations because there has been remarkable crop growth represented by higher canopy biomass in agreement with the final yield.

The validation process was made through the AUROC machine learning model between VIs and crop yield (rice and potato) for two different seasons. SVM, LR, and RF models have good validation results with > 83% accuracy for both crops. Crop yield estimation can be improved using modern advanced techniques such as hybrid ML and deep learning, satellite data fusion, LiDAR, UAV, IoT, SAR, and GEE cloud with a higher number of influential factors (soil factor, climate, topography, hydrology, AGB, farm mechanization level, and socioeconomic background).

8.6 Conclusion

The Sentinel-2 mission has opened up new scenarios for monitoring the performance of smallholder farming systems. The various temporal and spatial resolutions of satellite images were possible to estimate a farm's crop yield using crop observation VIs. Evaluation of farm decisions can then be made based on the data collected by observing different crop practice parameters. High-resolution images can now accurately classify crop biophysical data for individual farmlands. The data collected by the NDVI and SAVI revealed a high correlation between the observed canopy reflectance and the plant signals. The hotspot analysis further identified the areas with high crop yields and problem areas with low crop yields. These signals indicated that the various interventions designed to improve smallholder farms' productivity may not address these communities' diverse socioeconomic conditions. The results of the crop models may be provided to the farmer to take necessary steps to improve crop productivity, particularly in the northern part of the concerned study area.

Acknowledgements The authors acknowledge the contribution of Visva-Bharati (A Central University), West Bengal, India, for facilitating this research work.

References

- Abbas F, Afzaal H, Farooque A, Tang S (2020) Crop yield prediction through proximal sensing and machine learning algorithms. *Agron* 10:1046
- Abdulhafedh A (2017) A novel hybrid method for measuring the spatial autocorrelation of vehicular crashes: combining Moran's Index and Getis-Ord G_i^* statistic. *Open J Civil Eng* 7:208–221. <https://doi.org/10.4236/ojce.2017.72013>
- Ahmed R, Sajjad H (2018) Analyzing factors of ground-water potential and its relation with population in the Lower Barpani Watershed, Assam, India. *Nat Resour Res* 27:503–515
- Allen LM (2015) Influence of corn seeding rate, soil attributes, and topographic characteristics on grain yield, yield components, and grain composition. Graduate Theses and Dissertations. 14949. <https://lib.dr.iastate.edu/etd/14949>
- Ali A, Martelli R, Lupia F, Barbanti L (2019) Assessing multiple years' spatial variability of crop yields using satellite vegetation indices. *Remote Sens* 11:2384
- Arumugam P, Chemura A, Schauburger B, Gornott C (2021) Remote sensing based yield estimation of rice (*Oryza Sativa* L.) using gradient boosted regression in India. *Remote Sens* 13:2379. <https://doi.org/10.3390/rs13122379>
- Bhojani SH, Bhatt N (2020) Wheat crop yield prediction using new activation functions in neural network. *Neu Comput Appl* 32:13941–13951
- Borowik T, Pettorelli N, Sonnichsen L, Jedrzejewska B (2013) Normalized difference vegetation index (NDVI) as a predictor of forage availability for ungulates in forest and field habitats. *Eur J Wild Res* 59:675–682
- Campillo C, Carrasco J, Gordillo J, Cordoba A, Macua J (2018) Use of satellite images to differentiate productivity zones in commercial processing tomato farms. In: *Proceedings of the XV International Symposium on Processing Tomato 1233*, Athens, Greece, pp 97–104
- Cavalaris C, Megoudi S, Maxouri M, Anatolitis K, Sifakis M, Levizou E, Kyriarissis A (2021) Modeling of durum wheat yield based on sentinel-2 imagery. *Agron* 11:1486. <https://doi.org/10.3390/agronomy11081486>
- Cedric LS, Adoni WYH, Aworka R, Zoueu JT, Mutombo FK, Kimpolo CLM (2022) Crops yield prediction based on machine learning models: case of West African countries. *Smart Agric Technol* 2:100049. <https://doi.org/10.2139/ssrn.4003105>
- Chandel NS, Tiwari PS, Singh KP, Jat D, Gaikwad BB, Tripathi H, Golhani K (2019) Yield prediction in wheat (*Triticum aestivum* L.) using spectral reflectance indices. *Current Sci* 116(2):272–278. <https://doi.org/10.18520/cs/v116/i2/272-278>
- Chepngetich J (2020) Vegetation index based crop yield prediction model using convolution neural network: a case study of Kenya. Dissertation paper Strathmore University, pp 1–57
- Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, Bargellini P (2012) Sentinel-2: ESA's optical high-resolution mission for GMES. *Oper Serv* 120:25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- ESA (2019) European Space Agency, EO4SD-Earth Observation for Sustainable Development. Agriculture and Rural Development/Service Portfolio (2019). https://www.eo4idi.eu/sites/default/files/eo4sd_agri_portfolio_170529_singlepag.pdf. Accessed 01 July 2020
- Fang H, Liang S, Hoogenboom G (2011) Integration of MODIS LAI and vegetation index products with the CSM-CERES-Maize model for corn yield estimation. *Int J Remote Sens* 32:1039–1065
- Filippi P, Jones EJ, Wimalathunge NS, Somarathna PDSN, Pozza LE, Ugbaje SU, Bishop TFA (2019) An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Preci Agric*. <https://doi.org/10.1007/s11119-018-09628-4>
- Gaso DV, Berger AG, Ciganda VS (2019) Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. *Comput Electron Agric* 159:75–83. <https://doi.org/10.1016/j.compag.2019.02.026>
- Guo Y, Fu Y, Hao F, Zhang X, Wu W, Jin X, Bryant CR, Senthilnath J (2021) Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecol Ind* 120:106935
- Hatfield JL, Gitelson AA, Schepers JS, Walthall C (2008) Application of spectral remote sensing for agronomic decisions. *Agron J* 100, S-117–S-131
- Ji Z, Pan Y, Zhu X, Wang J, Li Q (2021) Prediction of crop yield using phenological information extracted from remote sensing vegetation index. *Sensor* 21:1406. <https://doi.org/10.3390/s21041406>
- Kobayashi N, Tani H, Wang X, Sonobe R (2020) Crop classification using spectral indices derived from Sentinel-2A imagery. *Info Telecom* 4(1):67–90. <https://doi.org/10.1080/24751839.2019.1694765>
- Koutsos TM, Menexes GC, Mamolos AP (2021) The use of crop yield autocorrelation data as a sustainable approach to adjust agronomic inputs. *Sustain* 13:2362. <https://doi.org/10.3390/su13042362>
- Kumhálová J, Matějková Š (2017) Yield variability prediction by remote sensing sensors with different spatial resolution. *Int Agroph* 31:195–202
- Kunapuli SS, Rueda-Ayala V, Benavidez-Gutierrez G, Cordova-Cruzatty A, Cabrera A, Fernandez C, Maignushca J (2015) Yield prediction for precision

- territorial management in maize using spectral data. In: Precision Agriculture 2015—Papers presented at the 10th European conference on precision agriculture, ECPA 2015, pp 199–206
- Kussul N, Lavreniuk M, Kolotii A, Skakun S, Rakoid O, Shumilo L (2020) A workflow for sustainable development goals indicators assessment based on high-resolution satellite data. *Int J Digi Earth* 13(2):309–321. <https://doi.org/10.1080/17538947.2019.1610807>
- Laacouri A, Nigon T, Mulla D, Yang C (2018) A case study comparing machine learning and vegetation indices for assessing corn nitrogen status in an agricultural field in Minnesota. In: Proceedings of the 14th international conference on precision agriculture. International Society of Precision Agriculture, Monticello, IL
- Lambert MJ, Pierre C, Traoréb S, Blaesa X, Baretta P, Defourny P (2018) Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sens Environ* 216:647–657
- Li Z, Liu S, Zhang X, West TO, Ogle SM, Zhou N (2016) Evaluating land cover influences on model uncertainties—A case study of cropland carbon dynamics in the Mid-Continent Intensive Campaign region. *Ecol Model* 337:176–187. <https://doi.org/10.1016/j.ecolmodel.2016.07.002>
- Lu G, Li C, Yang G, Yu H, Zhao X, Zhang X (2016) Retrieving soybean leaf area index based on high imaging spectrometer. *Soyb Sci* 35:599–608
- Marasteanu I, Jaenicke E (2016) Hot spots and spatial autocorrelation in certified organic operations in the United States. *Agric Res Econ Rev* 45(3):485–521. <https://doi.org/10.1017/age.2016.5>
- Marino S, Alvino A (2021) Vegetation indices data clustering for dynamic monitoring and classification of wheat yield crop traits. *Remote Sens* 13:541. <https://doi.org/10.3390/rs13040541>
- McKinion JM, Willers JL, Jenkins JN (2010) Spatial analyses to evaluate multi-crop yield stability for a field. *Comput Electron Agric* 70:187–198
- Mupangwa W, Chipindu L, Nyagumbo I, Mkuhlani S, Sisito G (2020) Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. *SN Appl Sci* 2:952. <https://doi.org/10.1007/s42452-020-2711-6>
- Nagy A, Fehér J, Tamás J (2018) Wheat and maize yield forecasting for the Tisza river catchment using MODIS NDVI time series and reported crop statistics. *Compu Electron Agric* 151:41–49
- Nagy A, Szabó A, Adeniyi OD, Tamás J (2021) Wheat yield forecasting for the Tisza River catchment using Landsat 8 NDVI and SAVI time series and reported crop statistics. *Agron* 11:652. <https://doi.org/10.3390/agronomy11040652>
- Panda SS, Ames DP, Panigrahi S (2010) Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sens* 2:673–696
- Panek E, Gozdowski D, Stepien M, Samborski S, Rucinski D, Buszke B (2020) Within-field relationships between satellite-derived vegetation indices, grain yield and spike number of winter wheat and triticale. *Agron* 10:1842
- Pejak B, Lugonja P, Antić A, Panić M, Pandžić M, Alexakis E, Mavrepis P, Zhou N, Marko O, Crnojević V (2022) Soya yield prediction on a within-field scale using machine learning models trained on Sentinel-2 and soil data. *Remote Sens* 14:2256. <https://doi.org/10.3390/rs14092256>
- Pinter PJ, Hatfield JL, Schepers JS, Barnes EM, Moran MS, Daughtry CST, Upchurch DR (2003) Remote sensing for crop management. *Photogram Eng Remote Sens* 69(6):647–664
- Pham HT, Awange J, Kuhn M, Nguyen BV, Bui LK (2022) Enhancing crop yield prediction utilizing machine learning on satellite-based vegetation health indices. *Sensors* 22(3):719. <https://doi.org/10.3390/s22030719>
- Pradhan S, Bandyopadhyay KK, Sahoo RN, Sehgal VK, Singh R, Gupta VK, Joshi DK (2014) Predicting wheat grain and biomass yield using canopy reflectance of booting stage. *J Indian Soc Remote Sens* 42(4):711–718. <https://doi.org/10.1007/s12524-014-0372-x>
- Quan Z, Xianfeng Z, Miao J (2011) Eco-environment variable estimation from remote sensed data and eco-environment assessment: models and system. *Acta Scien Nat Univ Pekine* 47:1073–1080
- Ramos APM, Oscob LP, Furuyaa DEG, Gonçalves WN et al (2020) A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices. *Compu Electron Agric* 178:10579. <https://doi.org/10.1016/j.compag.2020.105791>
- Rouse JW, Haas RH, Schell JA, Deering DW (1973) Monitoring vegetation systems in the Great Plains with ERTS (Earth Resources Technology Satellite). In: Proceedings of 3rd earth resources technology satellite symposium, Greenbelt, 10–14 December, SP-351, pp 309–317
- Shah A, Dubey A, Hemnani V, Gala D, Kalbande DR (2018) Smart farming system: crop yield prediction using regression techniques. Springer, Singapore, pp 49–56. https://doi.org/10.1007/978-981-10-8339-6_6
- Shammi SA, Meng Q (2020) Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. *Ecol Ind* 121:107124. <https://doi.org/10.1016/j.ecolind.2020.107124>
- Sharifi A (2020) Yield prediction with machine learning algorithms and satellite images. *J Sci Food Agric* 10696. <https://doi.org/10.1002/jsfa.10696>
- Singha C, Swain KC, Swain SK (2020) Best crop rotation selection with GIS-AHP technique using soil nutrient variability. *Agric* 10:213
- Singha C, Swain KC, Saren B (2019) Land suitability assessment for potato crop using analytic hierarchy process technique and geographic information system. *J Agric Eng* 56(3):78–87

- Singha C, Swain KC (2016) Land suitability evaluation criteria for agricultural crop selection: a review. *Agric Rev* 37:125–132
- Singha C, Swain KC (2022) Rice and potato yield prediction using artificial intelligence techniques. In: Pattnaik PK, Kumar R, Pal S (eds) *Internet of things and analytics for agriculture*, Volume 3. *Studies in Big Data*, vol 99. Springer, Singapore. https://doi.org/10.1007/978-981-16-6210-2_9
- Son NT, Chen CF, Chen CR, Chang LY, Duc HN, Nguyen LD (2013) Prediction of rice crop yield using MODIS EVI–LAI data in the Mekong delta, Vietnam. *Int J Remote Sens* 34:7275–7292
- Vallentin C, Harfenmeister K, Itzerott S, Kleinschmit B, Conrad C, Spengler D (2021) Suitability of satellite remote sensing data for yield estimation in northeast Germany. *Preci Agric* 23:52–82. <https://doi.org/10.1007/s11119-021-09827-6>
- Vannoppen Sui J, Qin Q, Ren H, Sun Y, Zhang T, Wang J, Gong S (2018) Winter wheat production estimation based on environmental stress factors from satellite observations. *Remote Sens* 10(6):962–972. <https://doi.org/10.3390/rs10060962>
- van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: a systematic literature review. *Comp Electron Agric* 177:105709. <https://doi.org/10.1016/j.compag.2020.105709>
- Xue J, Su B (2017) Significant remote sensing vegetation indices: a review of developments and applications. *J Sens* 1–17. <https://doi.org/10.1155/2017/1353691>
- Yuan W, Cai W, Nguy-Robertson AL, Fang H, Suyker AE, Chen Y, Dong W, Liu S, Zhang H (2015) Uncertainty in simulating gross primary production of cropland ecosystem from satellite-based models. *Agric Forest Meteorol* 207:48–57
- Yu J, Wang J, Leblon B (2021) Evaluation of soil properties, topographic metrics, plant height, and unmanned aerial vehicle multispectral imagery using machine learning methods to estimate canopy nitrogen weight in corn. *Remote Sens* 13:3105. <https://doi.org/10.3390/rs13163105>