Vol.66: e23220781, 2023
https://doi.org/10.1590/1678-4324-2023220781
ISSN 1678-4324 Online Edition

Article - Agriculture, Agribusiness and Biotechnology

Evaluation of Remote Sensing and Meteorological parameters for Yield Prediction of Sugarcane (Saccharum officinarum L.) Crop

Preeti Saini[1]
https://orcid.org/0000-0002-4013-1617

Bharti Nagpal[2]
https://orcid.org/0000-0002-6385-1764

Puneet Garg[3]
https://orcid.org/0000-0002-4635-546X

Sachin Kumar[4*]
https://orcid.org/0000-0003-3949-0302

[1]NSUT East Campus (Formerly AIACTR), Guru Gobind Singh Indraprastha University, USICT, Delhi, India; [2]NSUT East Campus (Formerly AIACTR), Department of Computer Science and Engineering, Delhi, India; [3]Christ University, School of Sciences, Delhi, India; [4]South Ural State University, Big Data & Machine Learning Lab, Chelyabinsk, Russian Federation.

Editor-in-Chief: Bill Jorge Costa
Associate Editor: Adriel Ferreira da Fonseca

Received: 08-Oct-2022; Accepted: 02-Dec-2022

*Correspondence: kumars@susu.ru; Tel.: +7 9512471669 (S.K.).

HIGHLIGHTS

- To evaluate the applicability of remote sensing and meteorological variables for Yield prediction.
- A Novel Empirical approach was proposed based on MOD13Q1-derived products NDVI, VCI, and Historical Yield.
- The experimentation was conducted for the 75 districts using R^2, RMSE, MAE, and MSE.

Abstract: In the Agriculture sector, the farmers need a reliable estimation for pre-harvest crop yield prediction to decide their import-export policies**.** The present work aims to assess the impact of remote sensing-based derived products with Climate data on the accuracy of a prediction model for the sugarcane yield. The regression method was used to develop an empirical model based on VCI, Historical Sugarcane Yield, and Climatic Parameters of 75 districts of six major sugar-producing states of India. The MOD13Q1 product of MODIS on Board Terra Satellite at 16-day intervals was accessed during the growing season of sugarcane crop with 36 meteorological parameters for experimentation. The accuracy of the model was evaluated using R^2, Root Mean square Metric (RMSE), Mean Absolute Error (MAE), and mean square error (MSE). The preliminary results concluded that the proposed methodology achieved the highest accuracy with (R^2 =0.95, MAE=5.18, MSE=34.5, RMSE=5.87). The conclusion of the study highlighted that the coefficient of determination can be improved significantly by incorporating maximum and minimum temperature parameters with Remote sensing derived vegetation indices for the sugarcane yield.

Keywords: Farming; Yield; Prediction; Sugarcane; MLR; NDVI; VCI.

INTRODUCTION

The reliable forecasting of the Crop yield prior to harvesting is crucial in countries like India where Crops are significantly affected by weather unpredictability [1]. Sugarcane is considered one of the main cash crops in India. It contributes around 1.1% of the Country's GDP, which is notable because it is growing over 2.57% only of the cropped region [2]. All over the world, sugarcane production is rapidly increasing from 1994-2021 due to the extent of sugar consumption. In 2022, India is the most sugar-consuming country in the world (29 million tonnes) followed by European Countries (17 million tonnes). In the world, Brazil is the main sugar producer (37.04%) followed by India (18.69%) [3].

The Crop prediction models can be applied to regions where all the data availability exists in the country [4]. These models can be categorized as the sampling approach, data modeling method, and mechanism method. A sampling approach used several samples to perform yield surveying, it's a time-consuming process approach. The mechanism model includes a production and crop growth model. In the production efficiency model, a crop amount is initially computed using remote sensing data, whereas the crop growth model considers physical crop growth parameters to estimate the yield [5,40,41]. The data modeling method includes both machine learning and statistical approaches, and based on data sources, the data modeling method is categorized as meteorological-based, remote sensing-based, statistical-based other yield prediction models [5].

An adequate water supply is required for the management of sugarcane production. The productivity of this crop mainly depends on climatic factors [6]. In recent times, flood and drought conditions have increased worldwide, and adversely affecting the sugarcane crop growth. Many sugarcane mills and government agencies are working to handle these adverse conditions so that the demand-supply gap does not occur in the market.

In India, FASAL ('Forecasting Agricultural Output using space, meteorological and Land Observations') is one of the programs initiated by Govt. of India to deal with pre-harvest crop yield estimation at the state and district, and national levels across the country [7]. The application of this program includes drought assessment, and monitoring, estimation of crop area and production, analysis of cropping system, mapping of soil resources, precision farming, climatic condition influence on agriculture, etc.

LITERATURE REVIEW

Due to the advancements in technology and in the satellite-based arena, accurate crop yield forecasting can be enabled with spatial observations. In earlier studies, statistical or land observations were highly used to estimate the yield forecasting in a region. The subsequent section discussed the existing work related to the yield prediction domain.

To build the Yield estimation model, several remote sensing derived products such as NDVI (Normalized Difference Vegetation Index), Leaf Area Index (LAI), Enhanced Vegetation Index (EVI), Vegetation Condition Index (VCI), Fraction of Photosynthetically active radiation (FPAR) had been used in the literature and among them, NDVI is the most utilized vegetation index in the research studies. Dubey and Coauthors, (2018) used VCI for district-level sugar estimation under the FASAL project and found that the variability was explained in more than 60% of considered cases [1]. Dimov and coauthors (2022) employed NDVI, NREI, and phenology metrics to determine the yield in Ethiopia and investigated the sugar estimate model with R^2 value of 0.84 [8]. Rahman and coauthors (2016) used the Green Normalized Vegetation Index (GNDVI) to predict the sugarcane yield in the Bundaberg region and observed the model accuracy with R^2 value of 0.69 [9]. Bhatla and coauthors (2018) assessed the influence of Temperature, Relative Humidity, and Rainfall on sugarcane yield using 40years of data in the Gorakhpur district of Uttar Pradesh and found a relative error of 12.9% between the actual and predicted Yield in the district [10]. Geetha and coauthors (2018) used the soil samples in the Trichy region to investigate the Yield by using the modified regression method with a discretization approach [11]. Yadav and coauthors (2021) utilized the remote sensing derived NDVI with climatic factors in New Mexico and analyzed that NDVI alone results were less accurate in respect of the combination of NDVI with Climate factors for wheat, corn, and sorghum yield [12]. Khaki and coauthors (2021) proposed a five convolutional layers-based network to estimate the corn and soybean yield based on remote sensing data and estimated MAE value as 8.70 % [13]. Khaki and coauthors (2020) used the historical data of corn and soybean-based on the CNN-RNN model and achieved an RMSE of 9 percent and 8 percent [14]. Shah and coauthors (2021) Utilized the XGboost method to predict the rice yield in the Tamil Nadu region using NDVI and weather factors that evaluated 0.84 accuracy on the test variables [15]. Lopresti and coauthors (2015) examined the relation between NDVI derived via MOD13q1 and historical Yield by a simple regression method, and determined the yield thirty days prior to harvesting period with R^2 value of 0.75 [16]. Ming and coauthors (2015) investigated the relationships among the maize yield and standardized precipitation evapotranspirtation index (SPEI) in the North China region and outcomes revealed that the water conditions in June-July had a good estimation for August or for three months interval data [17]. Dubey S and coauthors (2019) highlighted the role of the VCI index as a potential estimator for wheat and rice yield prediction over the various districts of the States [18].

Pham and coauthors (2022) employed the PCA-ML method using the vegetation condition index for rice yield forecasting in Vietnam and found that the proposed model performed 18% to 45% better in comparison to the ML-based approach [19]. Cai and coauthors (2019) presented the EVI-based prediction model using a support vector machine and found a significant performance of the model with R^2 value of 0.75 [20]. Liu and coauthors (2022) performed the stepwise regression model on the 207 yield statistics of the plot with a T-test performed over the attributes to determine the best features and found the goodness of fit with 0.453 of the regression models [21]. Klompenburg and coauthors (2020) explored the various features used for experimentation in agrarian yield prediction, and focussed on the deep learning and machine learning paradigm that was highly applied to carry out the research in recent articles [22]. Shah and coauthors (2018) applied climatic parameters for yield prediction and assessed the performance of the support vector machine, random forest, and regression model over the US agriculture dataset [23]. Pandey and coauthors (2016) used the meteorological parameters for rice yield prediction in the Uttar Pradesh region and found that Maximum temperature, and minimum temperature had a strong relation in the dataset [24].

In earlier studies, Linear models were also employed to build the prediction model, but their main drawback is that they are not able to capture non-linear patterns among the datasets. Mishra P and coauthors (2021) presented the ARIMA modeling-based prediction for major sugar-producing states of India. The selection of the appropriate model was done based on AIC values and analyzed the accuracy with 0.04% to 0.19% [25]. Ali S and coauthors (2015) investigated the production of sugarcane and cotton crops in Pakistan using ARIMA models and found that ARIMA (0,1) & ARIMA (2,1,1) models work well for these crops [26]. Dash A and coauthors (2020) analyzed the forecasting of pulses in Odisha state using ARIMA models and presented that ARIMA (1,0,0) model performed best among all the models of forecasting the pulses [27]. Mishra P and coauthors (2022) presented the ARIMA models for predicting the five major pulse producer states of India and the results predicted that there will be a decline in pulse production in Uttar Pradesh with respect to other states [28].

The present study proposed an empirical model for predicting the sugarcane crop yield in the major sugar-producer states of India. The model was trained using the meteorological, remote sensing-derived products, and Historical Yield of the 75 districts of Uttar Pradesh, Maharashtra, Karnataka, Tamil Nadu, Bihar, and Gujarat states. In addition, no significant attempt has been done in the literature so far, to assess the impact of these predictors on the yield prediction of sugarcane crop in the study regions. The main contributions of the proposed work are as follows:

- We proposed an empirical LTP_MLR model to evaluate the suitability of remote sensing and climatic parameters on the yield.
- We evaluated the relative importance of parameters on the accuracy of yield estimation at the district level.
- We analysed the vegetation cover of the study regions to monitor the vegetation health of crops from 200-2018 years, and estimated the standard error between the observed and predicted yield.

MATERIAL AND METHODS

Study Area

In this article, we focused on one of the India’s major cash crops, i.e., the Sugarcane Crop. The investigation includes an analysis of the considered crop in six major producer states of India (Uttar Pradesh, Maharashtra, Karnataka, Tamil Nadu, Bihar, and Gujarat). As per the statistical information of 2021 year, Uttar Pradesh is the main contributor of the Sugarcane Crop (approx. 48%) in India [29]. The Study regions considered for the work are highlighted in Figure 1.

The statistics of the 2018-2019 year revealed that there was a decrement in the production of Sugarcane crops in the Uttar Pradesh region from 179.7 million tonnes (mt)to 178.4 mt. This Crop usually takes twelve to eighteen months for its growth and, known as a Kharif season Crop because it is sown in the Rainy season of India. As sugarcane is a tropical crop, it requires 21 degrees Celsius to 27 degrees Celsius temperature for its growth. The details associated with the sugarcane production requirements in the study regions are illustrated in Table 1 [30].

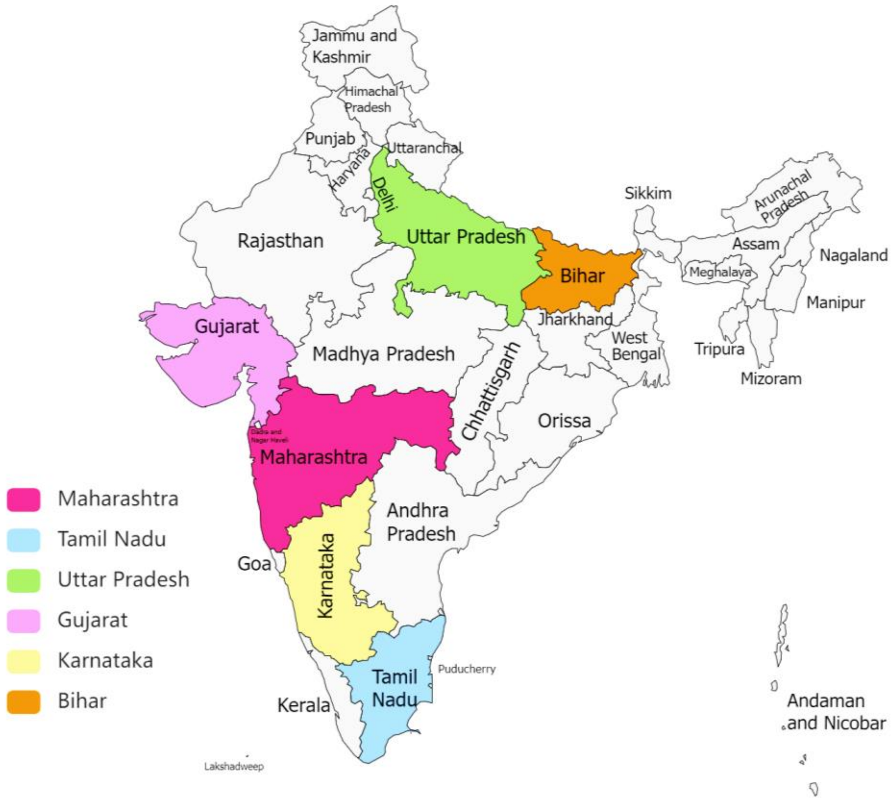

Figure 1. Study Regions of India

The statistics of the 2018-2019 year revealed that there was a decrement in the production of Sugarcane crops in the Uttar Pradesh region from 179.7 million tonnes (mt)to 178.4 mt. This Crop usually takes twelve to eighteen months for its growth and, known as a Kharif season Crop because it is sown in the Rainy season of India. As sugarcane is a tropical crop, it requires 21 degrees Celsius to 27 degrees Celsius temperature for its growth. The details associated with the sugarcane production requirements in the study regions are illustrated in Table 1 [30].

The sugarcane crop growing and harvesting seasons differ across the various states in India. It usually varies from one place to another as per the usage of water and the season of that particular region. Generally, January – March is the period of growth, and December – march is considered the harvesting duration of this crop. In Karnataka, and Maharashtra sugarcane crop is grown in three distinct seasons known as 'Adsali (June-Agust)', 'Pre-seasonal (October-November)', and 'Suru (January-February)', which takes around twelve to eighteen months for their growth.

Table 1. Details of Sugarcane Crop in study regions

States	Crop season (In months)	Water Required (ha-mm)	Yield Range (tonnes/ha)	Growing Period	Harvesting Duration	PH Range
Uttar Pradesh	9-10	1400-1600	52.3-88.8	Oct-Nov	Oct-Dec	6.0-7.0
Maharashtra	12-16	2500-3500	57.9-92.2	Oct-Nov	Jan-Mar	6.0-7.0
Tamil Nadu	10-12	1850-2150	87.1-111.5	Dec-Feb	Oct-Nov	6.0-7.0
Karnataka	12-14	2000-2200	65.8-102.9	Sep-Nov	June-Feb	6.0-7.0
Bihar	10-16	1250-2500	24.6-59.2	Jan-Mar	Dec-Mar	6.0-7.0
Gujarat	11-14	1500-2500	63.1-80.5	Dec-Feb	Nov-Feb	6.0-7.0

Dataset

The dataset used to carry out this study mainly includes Modis-based Remote sensing data, Weather parameters, and Historical sugarcane Yield. Table 2 illustrates the description of the dataset. The meteorological data includes Temperature (T2M), Maximum Temperature(T2M_Max), Minimum Temperature (T2M_Min), and Relative Humidity (RH2M) of the growing season of the Sugarcane crop. There were about 36 attributes of the climate dataset fed to build the model. The MOD13Q1 Terra products accessed by the Google Earth engine are used to derive Vegetation Condition Index based on the Normalized vegetation value of a district at a particular state [31].

Table 2. Description of Dataset

Data	Reference to Source of Data	Temporal Resolution	Spatial Resolution
Historical Sugarcane Yield	[29]	Yearly	District Level
Remote Sensing Data	[31]	16days	250m
Meteorological Data	[32]	Monthly	1°

The Statistical Sugarcane Yield data was collected from the Directorate of Economics and Statistics during 2000-2018 years of six major sugar producer states of India represented in Figure 2 [29].

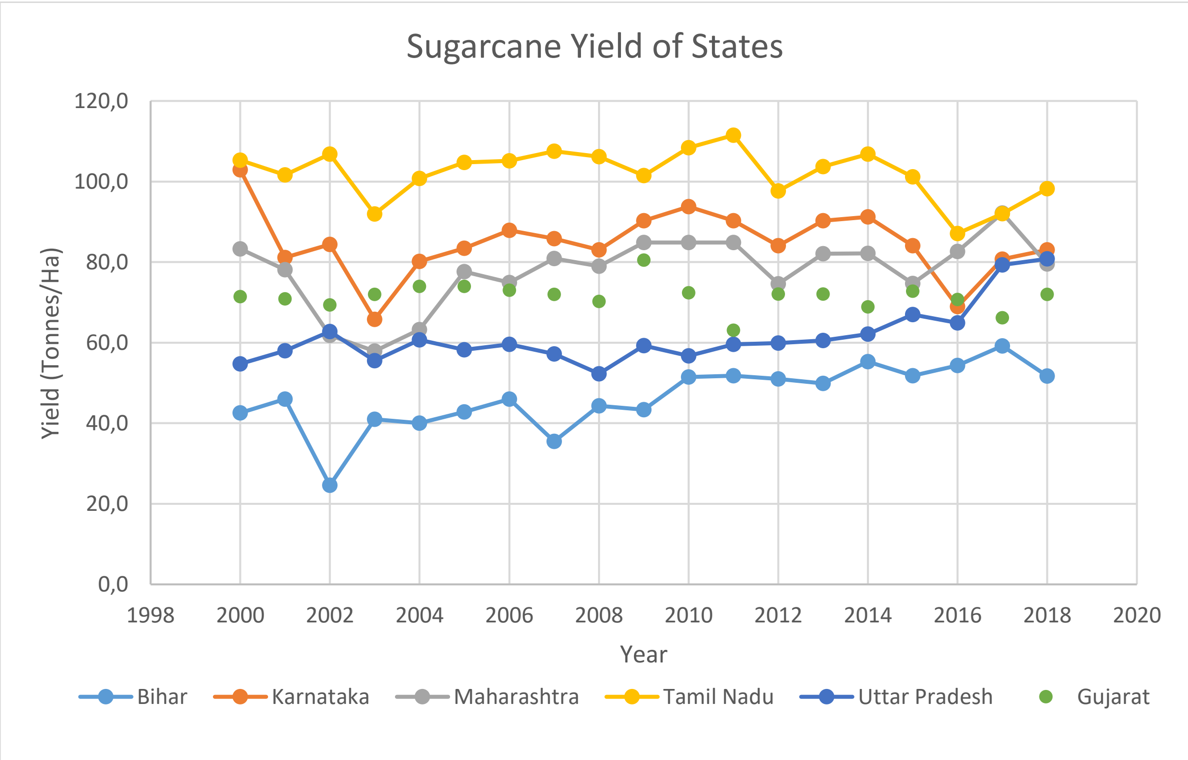

Figure 2. Historical Yield of selected states from the 2000-2018 Year

Experimentation Tool

The proposed approach was implemented using Python Programming in Jupyter Notebook and GEE. The hardware requirements include an i5-11300H processor, an X-64-based PC, and 16 GB RAM.

Proposed Methodology

The present study focused on the analysis of Sugarcane Yield in India using a remote sensing-based approach collected from MODIS satellite imagery data and predicting the Yield from 2000-to 2018 year respectively. This section highlights the steps to be followed to implement the empirical model in major sugarcane producing states. Finally, a solution is built to overcome these issues to predict the crop yield in study regions. The proposed approach used in this study is shown in Figure 3.

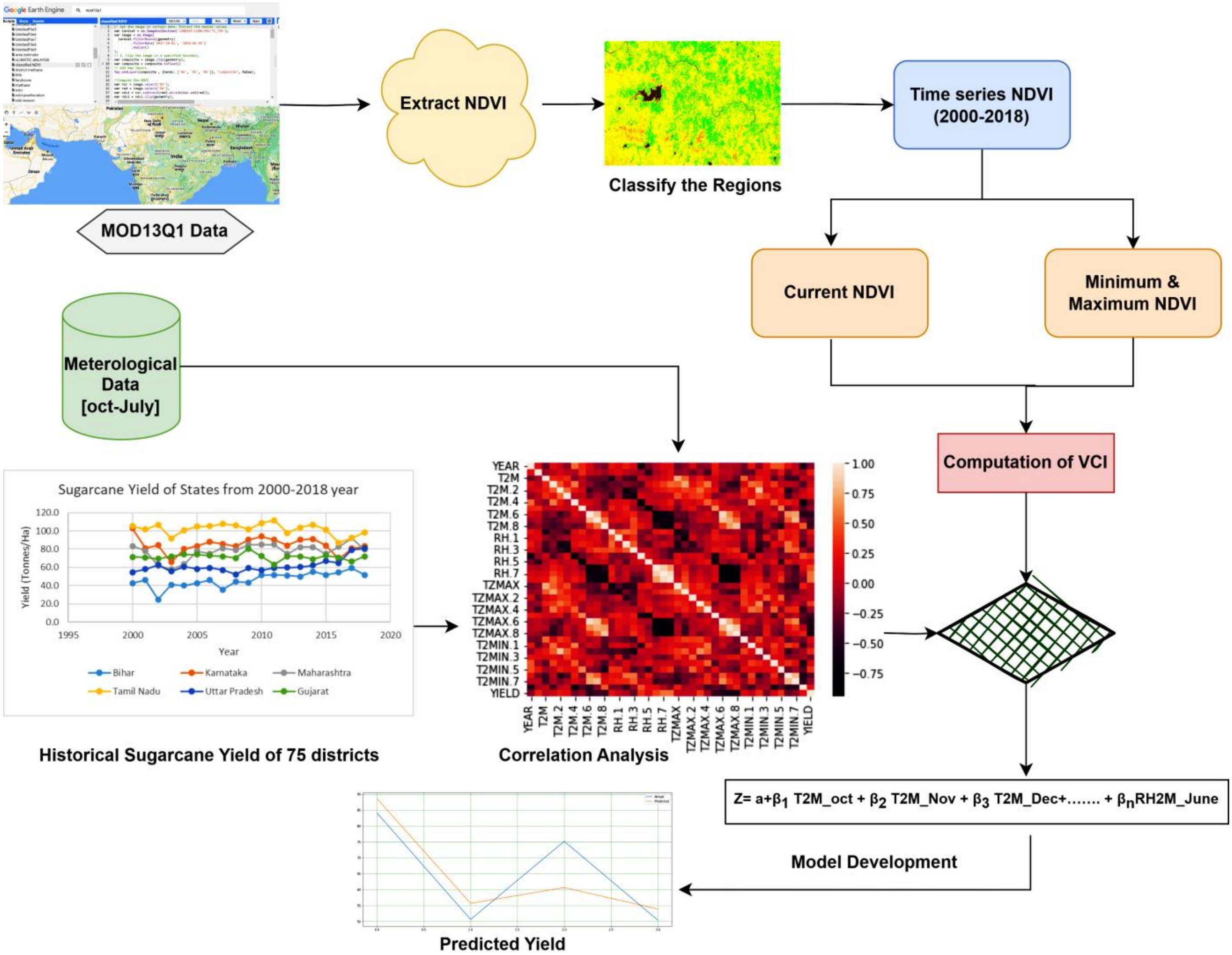

Figure 3. Proposed Methodology using Remote Sensing based data (MOD13Q1) & Ground Truth Data with 36 Climatic attributes at the district level

Computation of Vegetation Indices

Normalized Difference Vegetation Index ((NDVI)) depicts the vegetation region by estimating the differences between near-infrared and red light [33,34]. The range of NDVI varies from -1 to +1. The formula used to compute the NDVI is represented by Equation (1),

$$\text{NDVI}=\frac{NIR-RED}{NIR+RED} \tag{1}$$

VCI is an NDVI-derived based index that basically isolates the short-term ecological signals from the longer ones [35,36,42]. In this present study, 18 years (2000-2018) of NDVI were derived from the VCI of a particular region. This can be computed using Equation 2 [35,36].

$$\text{VCI} =\left(\frac{N_i-N_m}{N_x-N_m}\right)\times 100 \tag{2}$$

Where N_i is the NDVI at current i time; N_m is the minimum historical NDVI; N_x is the maximum historical NDVI.

Correlational Analysis

The correlational analysis was conducted to estimate the most viable predictors in the yield prediction model. To perform the correlation a person's coefficient method was used to estimate the linear connectivity among the parameters. The formula used to compute the Pearson coefficient (c) is denoted by Equation (3).

$$c=\frac{\sum(y_i-\bar{y})(z_i-\bar{z})}{\sqrt{\sum(y_i-\bar{y})^2\ \sum(z_i-\bar{z})}} \tag{3}$$

The principal component analysis and correlational analysis of 37 parameters were evaluated to predict the correlation among the operational predictors [37,38,39,42]. The principal component analysis was used to determine the highly influential factors in building the prediction model. The Jupyter notebook was used to evaluate the Principal Components and Correlational analysis. The most relevant variables were identified and utilized to build the prediction model and this approach of selection is also known as the 'Stepwise method'.

Long-time Period Regression Model

The regression model used to predict the yield based on single and multiple predictors represented by the following equation (4),

$$z_i = \sum_{i=1}^{m} x_i y_i \tag{4}$$

Where, z_i denotes the crop yield, x_i is the predictor coefficient of y_i, and m is the total number of parameters used to build the model.

The simplest method of regression includes a single predictor and as per correlation result analysis, other variables are added to the regression model to predict the crop yield. In our case, the T2M, T2M_Max, T2_Min, NDVI, VCI, R2H, etc. parameters were added for the longer duration of months as sugarcane crop growth period varies over the several states. In consideration of this fact, mostly all the month's data were utilized for computed the NDVI and VCI.

If a regression model contains a single dependent variable and several independent variables, then the technique is known as 'Multiple Regression'. In this approach, every variable has a weight assigned with it that can be represented by Equation 5.

$$Z= a+ \beta_1 \text{ T2M\_oct} + \beta_2 \text{ T2M\_Nov} + \beta_3 \text{ T2M\_Dec}+\ldots\ldots. + \beta_n \text{ RH2M\_June} \tag{5}$$

Here, T2M_Oct, T2M_Nov, T2M_Dec……RH2M_June are the independent variables in our study, $\beta_{1..n}$ are the weights to ensure the prediction of the dependent variable (z_i) from the independent parameters.

RESULTS AND DISCUSSION

The objective of the proposed work is to build a model to explore the remote sensing and meteorological data for predicting the yield, in earlier studies either one of these datasets used to explore the efficiency of the prediction model. In this study, we utilized both sets of data along with historical ground truth to estimate the yield at the district level in major sugarcane producer states of India. This study provides directions to the farmers for their decision-making policies. This section highlights the experimental outcomes.

Evaluation Criteria

The model was developed to establish a relation between the 37 independent variables (VCI & Meteorological attributes) and dependent variables (Ground truth Yield data) in the study sites. The performance of the proposed work was assessed using the Coefficient of determination (R^2), Root Mean square error (RMSE), Mean square error (MSE), and Mean Absolute Error (MAE). The model is considered as a good model, if R^2 value is predicted high whereas a lower value for other metrics. These metrics were evaluated using Equations 6-9.

$$\text{MSE} = \frac{1}{M}\sum_{j=1}^{M}(w_j - \widehat{w})^2 \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(w_j - \widehat{w})2} \tag{7}$$

$$MAE = \sum_{j=1}^{M}\left|\frac{o_j - p_j}{M}\right| \tag{8}$$

$$R^2=1-\frac{Rs}{TS} \tag{9}$$

Where, M= no. of dataset values; $\hat{w}$ = Observed Yield; w_j=yield's mean value; RS= Residual's sum; TS= Sum of Squares.

Analysis of Sugarcane Vegetation Index & Climatic Variables

Figure 4 exhibits the trend analysis of the NDVI values derived from MOD13Q1 products over a growing season in major sugar producer districts of the studied states. The peak in the NDVI range was observed in the growing season January-March whereas from July-August in some regions of the Country. The highest VCI range was observed at 72.27 and the lowest at 18.87. Most of the district's NDVI shows a rise in NDVI value from October to June, therefore these seasons' data were considered for modeling the prediction model. In Uttar Pradesh, the Manipuri district observed the highest NDVI with 0.92 in 2018 and the lowest with -5.79 in Buduan district.

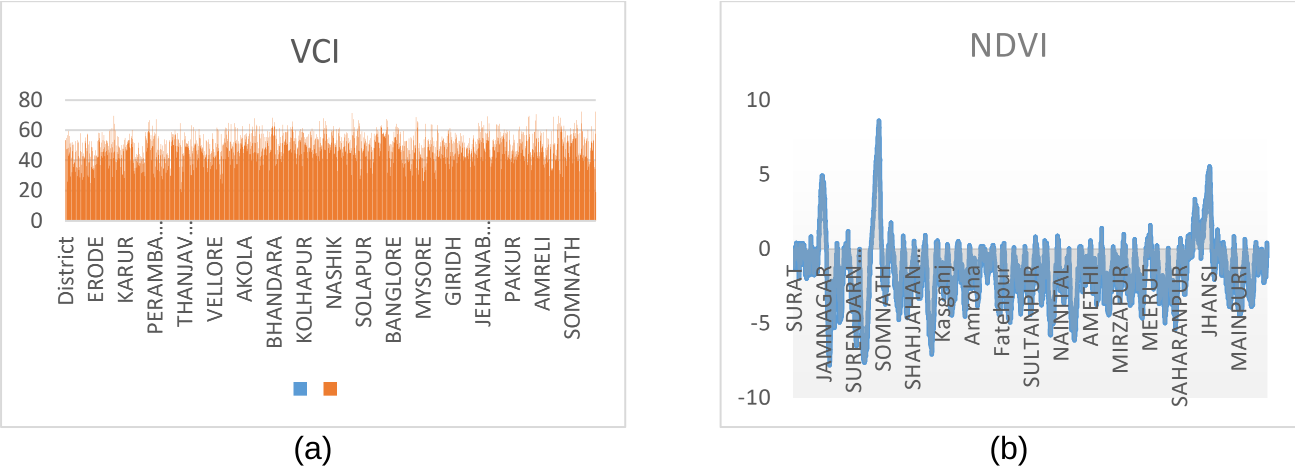

Figure 4. Trend Analysis of (a) VCI, (b) NDVI

The classification results of NDVI over these states from the 2017-2018 year highlight the value of NDVI over the region in Figure 5. The vegetation data can be analyzed based on Legends specified in each classified image to depict the range of NDVI variation in the region over a period. In general, the range of NDVI lies between -1 to +1. The negative value depicts the likelihood of water in the region whereas the high positive value denotes the dense vegetation area. The region that has a value less than zero, is represented by a brown color for the barren land & rocks. The dark green color signifies a value range greater than 0.6 for the tropical and temperate rainforests. The value lies in the range from 0.2-0.6 specifying the Crop area in the State, highlighted with Yellow and light green colors in the classified images.

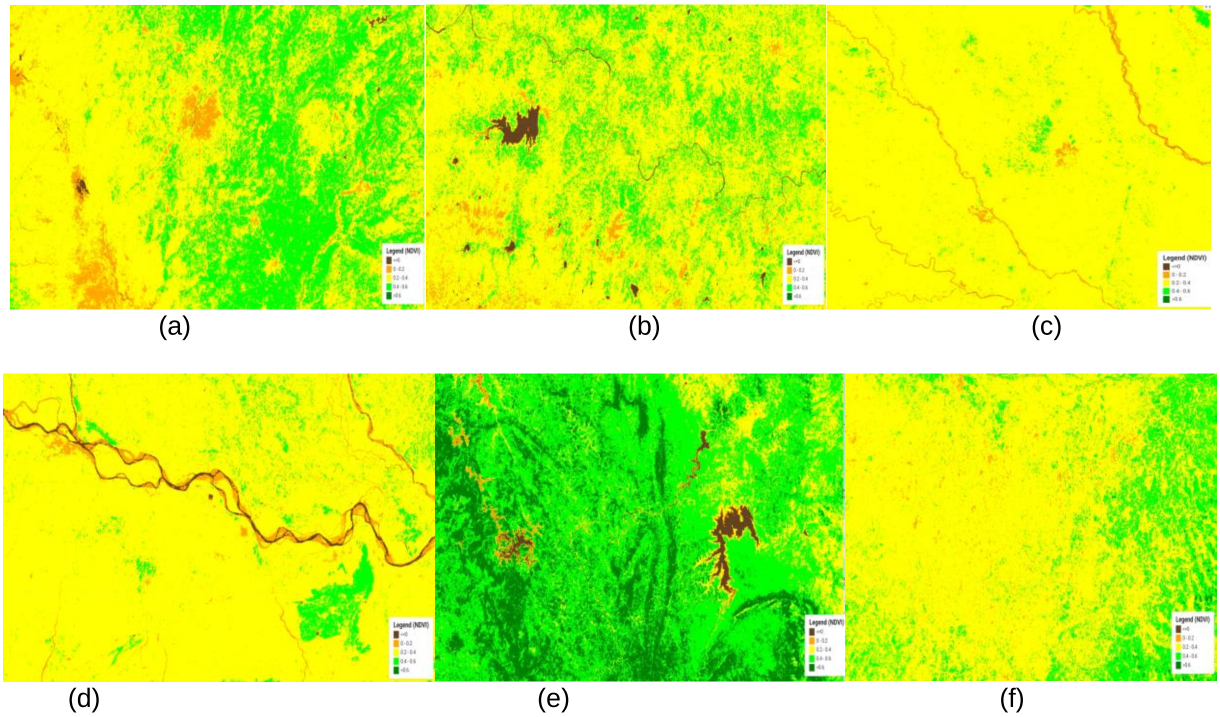

Figure 5. NDVI Classification of Regions in the States (a) Gujarat, (b) Maharashtra, (c) Uttar Pradesh, (d) Bihar, (e) Karnataka, (f) Tamil Nadu.

The correlation analysis was performed to evaluate the parameter which is strongly correlated to estimate the Yield in the mentioned states. These matrices represent the Pearson's correlation coefficient value. The range of value lies between -1 to +1 where -1 denotes the negative linear relationship between the parameters and, +1 denotes the positive linear relation. The strength and correlation among the climatic and remote sensing-based variables are represented as dark colors signifying positive relation (above 0.75) whereas the negative relation (less than -0.50) is highlighted by light shades of colors in Figure 6 for Uttar Pradesh, Maharashtra, Tamil Nadu, Karnataka, Gujarat, and Bihar respectively.

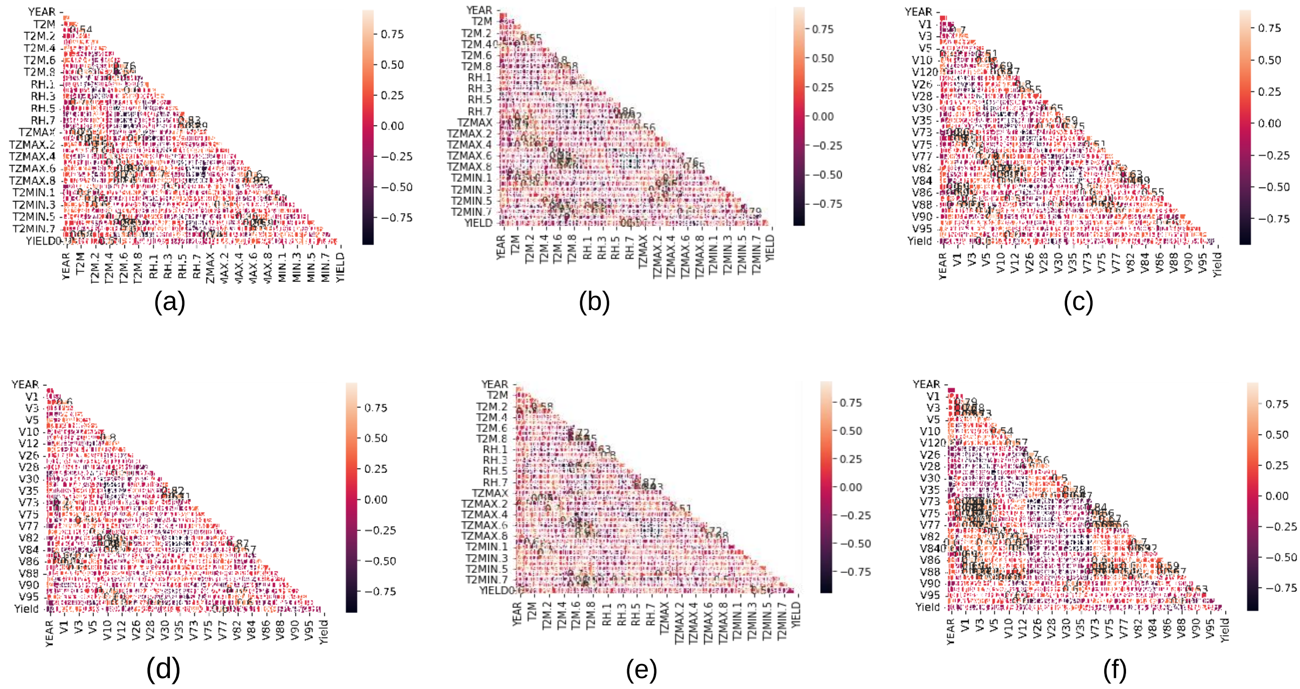

Figure 6. Correlation Analysis among experimental attributes of regions (a) Uttar Pradesh, (b) Maharashtra, (c) Tamil Nadu, (d) Gujarat, (e) Bihar, (f) Karnataka.

As per correlation analysis results, T2M, and T2M_Max climatic factors exhibit a good correlation with VCI in the range of 0.7 to 0.8 in the Maharashtra and Uttar Pradesh state districts. The correlation coefficient provided significant results at a 0.5 level among RS_Y, T2M, and S_Y. The correlation exhibit detrended linear relation among T2M_Min and RS_Y at values lesser than 0 and in the negative range.

Results based on Proposed methodology

The LTP_MLR technique was used to build the empirical model for predicting the yield using Weather variables and remote sensing derived VCI. The range of different measures to assess the performance of the proposed model was illustrated in the district map of the studied regions in Figure 7 and Table 3. Among the 75 districts, 45 districts had a R^2 value greater than 0.60 (with the highest prediction in Aurangabad and Nashik with 0.95 value), 13 districts had a R^2 value ranging from 0.40 to 0.60, and the rest of the districts had a value range from lower than 0.40. Table 3 represents that the proposed district yield model performed well for the Maharashtra State followed by Uttar Pradesh (with a maximum R^2 value of 0.93), Tamil Nadu, Karnataka, Bihar, and Gujarat. The lower prediction results in various districts may be because their longer crop season patterns exist in a particular region that's why NDVI in these regions shows an increasing trend in the study duration as shown in Figure 4.

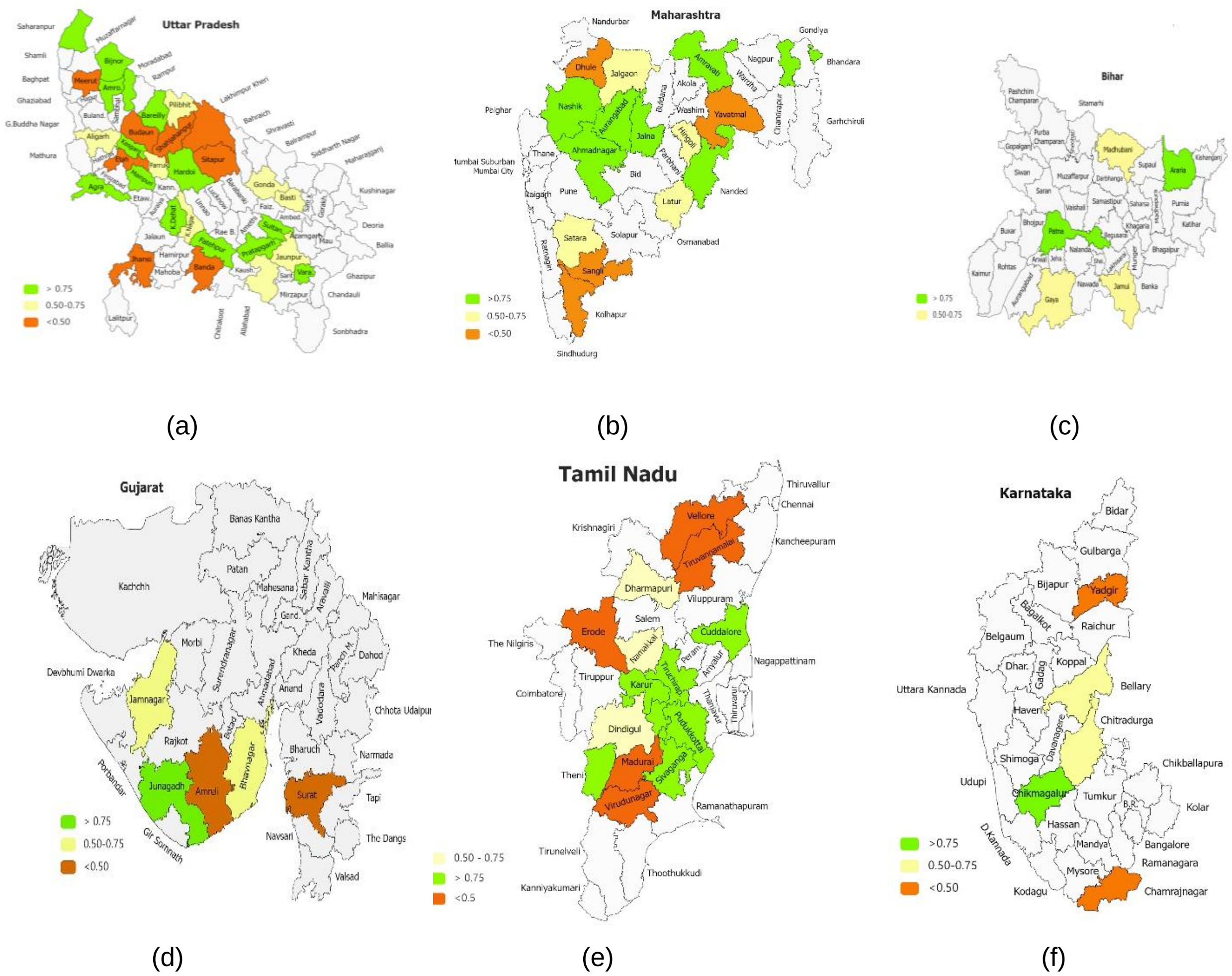

Figure 7. Representation of R^2 Analysis of considered districts (a) Uttar Pradesh, (b) Maharashtra, (c) Bihar, (d) Gujarat, (e) Tamil Nadu, (f) Karnataka.

The result analysis had been presented on the map of these districts to highlight the range of coefficient of determination in a particular district during the model development.

Table 3. Results for State-wise range of various evaluation metrics for Yield Prediction

State	No. of Districts	R^2 (Range)	MAE	MSE	RMSE
Uttar Pradesh	30	0.30 - 0.93	2.1-15.8	6.9 – 351.9	2.63 - 18.7
Maharashtra	15	0.24 - 0.95	6.88 – 28.9	67.35 – 1060.4	8.2 – 32.56
Tamil Nadu	15	0.55 - 0.93	5.18 – 20.89	34.5 – 446.4	5.87 – 21.49
Karnataka	5	0.40 - 0.93	6.22 – 15.8	45.86 – 446.4	6.77 – 21.13
Bihar	5	0.52 - 0.75	5.8 – 26.09	146.12 – 790.4	6.79 – 25.53
Gujarat	5	0.29 - 0.80	12.5 – 22.7	179.8 – 731.7	13.4 – 27.05

Validation Results of Model

The Empirical model was validated based on actual ground yield data versus the predicted yield as shown in Figure 8. The final Sugarcane Yield was the sum of the statistical yield (S_Y), Remote sensing-based Yield (RS_Y), and meteorological variable (M_Y). The similarity between the observed and predicted values was close to the diagonal and the coefficient of determination of most of the districts lies in the range of 0.80 to 0.95. The results in Figure 5 highlight the reliability of estimated measures having good accuracy.

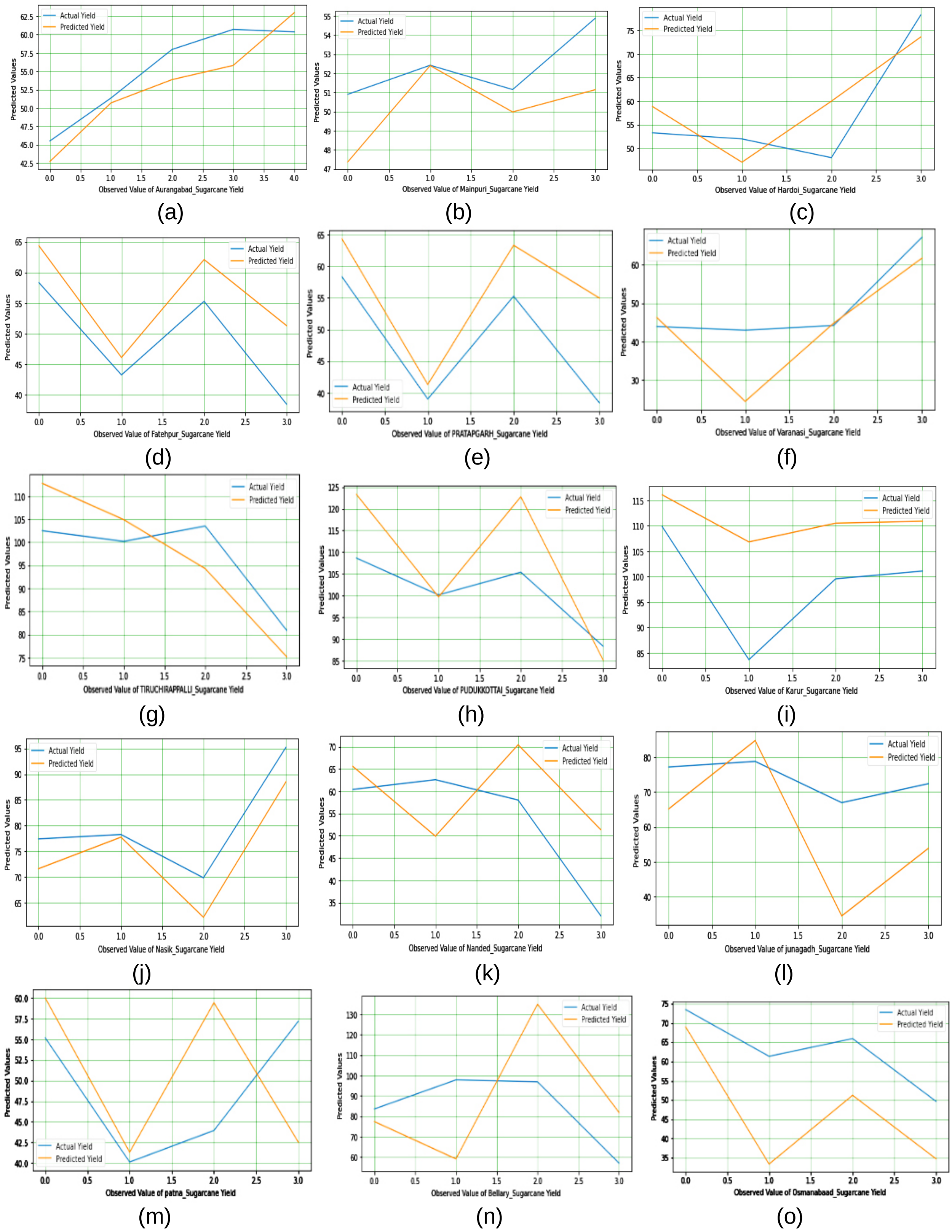

Figure 8. Comparison of Actual Yield with Predicted Results of various districts (a) Aurangabad: R^2 =0.94, (b) Mainpuri: R^2 =0.90, (c) Hardoi: R^2=0.86, (d) Fatehpur: R^2=0.89, (e) Pratapgarh: R^2=0.85, (f) Varanasi: R^2 =0.83, (g) Tiruchirappalli: R^2=0.93, (h) Pudukkottai: R^2=0.92, (i) Karur: R^2=0.85, (j) Nashik: R^2=0.95, (k) Nanded: R^2=0.81,(l) Junagadh: R^2=0.80, (m) Patna: R^2=0.74, (n) Bellary: R^2=0.68, (o) Osmanabad: R^2=0.76.

The standard error was found to be low in most of the districts of Maharashtra state with 1.04%, in the Bihar region the error was highest at 21.9% which can be improved further by incorporating July to September VCI and NDVI values in the model. The results of Figure 8 and Table 4 estimated the potential of incorporating remote sensing-based data with meteorological parameters in the prediction of sugarcane at the district level.

However, the predicted errors may be due to the NDVI poor signals that can be influenced by the cloud and contamination of atmospheric moisture. The Nashik, Aurangabad, and Tiruchirappalli district of Maharashtra, & Karnataka performed well with R^2 of 0.94-0.95.

Table 4. Actual Yield versus Predicted Yield using District level yield model

State	District	Observed Yield (Tonnes/Ha)	Predicted Yield (Tonnes/Ha)	Standard Error (%)
UP	Kanpur	65.51	62.0	5.36
Tamil Nadu	Tiruchirappalli	80.88	76.5	5.41
Maharashtra	Latur	56.59	56.0	1.04
Gujarat	Junagadh	71.8	63.3	11.8
Bihar	Patna	61.13	47.7	21.9

*****Standard error={{observed-predicted}/{observed}} *100

Comparison of Importance of various parameters

There are various stages that occur during the growth of sugarcane crops which are as follows: germination, Tillering, Early growth, Active growth, and Elongation. The germination stage of the sugarcane crop requires 32-degree Celsius to 38-degree Celsius temperature. Approx. 12-degree Celsius to 14-degree Celsius temperatures are required for mature of the sugarcane crop. The productivity of sugarcane is highly dependent upon the water and optimal weather conditions in the region. If the temperature range lies above 38 degrees Celsius, it results in the respiration process. The fluctuation in temperature at High and low levels can deteriorate the quality of the sugarcane. The tropical regions are considered suitable regions for the growth of the sugarcane crop. These facts were taken into consideration to incorporate T2M, T2M_Max, and T2M_Min parameters to assess the influence of these variables on the yield.

Table 5. Accuracy Assessment of Empirical Model based on Parameters

	Inclusion of Weather Parameters				Exclusion of Weather Parameters			
Districts	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
AURANGABAD	0.94	6.89	67.35	8.2	0.23	14.55	266.9	16.33
KANPUR	0.93	3.2	16.4	4.05	0.89	1.22	2.55	5.07
TIRUCHIRAPPALLI	0.93	6.2	43.5	6.6	0.11	7.28	62.89	7.93
CHIKMAGALUR	0.93	6.22	45.86	6.77	0.24	9.34	168.4	12.9
ARARIA	0.75	5.8	46.1	6.79	0.30	10.3	100.4	14.3
JUNAGADH	0.80	16.9	383.8	19.5	0.13	14.04	225.3	15.01

Table 5 shows the investigation analysis results about the impact of meteorological factors on the accuracy of the prediction model. The result was analyzed in the well-performing districts of each state and found that the coefficient of determination dropped significantly with the increment in the RMSE value by 30 to 50%. In the Tiruchirappalli district, the R^2 dropped to 0.11 if only remote sensing data were included in the experimentation, whereas in the case of Junagadh and Araria the mean absolute error raised by 60%. We can infer from these outcomes to necessarily ponder the meteorological parameters to build the Yield prediction models.

Several researchers working in this domain utilized various techniques as mentioned in Table 6 to explore the applicability of remote sensing-based data to carry out yield prediction of a particular crop. We performed a comparison with existing solutions to prove the originality and contribution of our research findings in the domain of Agriculture. The proposed methodology provided the outcomes that best support the inclusion of meteorological and remote sensing derived attributes with statistical data to enhance the Model's prediction accuracy.

Table 6. Description of the existing studies and the Proposed Model

Study	*Approach*	*Statistical Parameters*	*Overview*
Dubey et al. (2018)	Stepwise Regression	VCI	VCI-based approach for Yield prediction
Cai et al. (2019)	SVM	EVI+SIF	EVI-based approach for wheat yield prediction
Zhu et al. (2021)	Random Forest	NDVI+SPI	NDVI-based approach for maize yield prediction
Yadav et al. (2021)	Regression Model	NDVI+Temp+Prec	NDVI-based approach for Wheat Yield Prediction
Dimov et al. (2022)	OLS Regression	NDVI + NDREI	Phenological metric-based Sugarcane Yield prediction
Proposed Methodology	LTP_MLR	NDVI_Oct…+NDVI_July + VCI_Oct…+VCI_July + T2M +T2M_Max_Oct…+ T2M_Max_July + T2M_Min_Oct…+ T2M_Min_July + R2H_Oct…. +R2H_July + Yield	NDVI, VCI & meteorological-based approach for sugarcane Yield prediction

Figure 9 represented the comparative accuracy assessment of various proposed methods for the yield prediction domain. The given analysis manifests the performance of the proposed method in comparison to recent studies. It is the crucial domain of discussion as per 2030 sustainable goals therefore many researchers are attracted to this research domain, we had also utilized the MOD13Q1 dataset to contribute to this research arena. Moreover, we had developed an approach based on artificial intelligence that can aid support farmers to handle their decisions regarding management issues. The comparative analysis highlighted the achievement of the baseline model with existing state-of-the-art techniques.

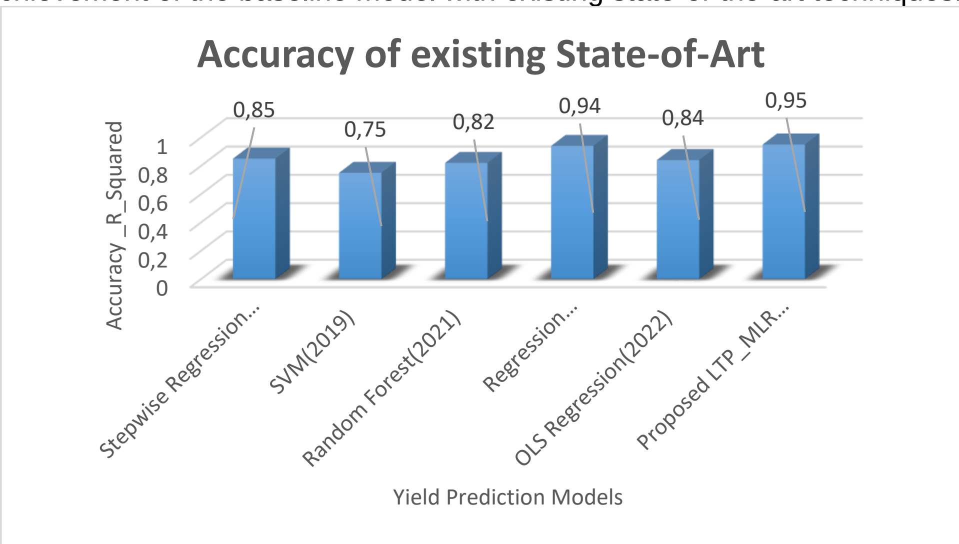

Figure 9. Comparative Analysis with existing state-of-the-art methods

CONCLUSION AND FUTURE DIRECTIONS

This study examined the impact of remote sensing and meteorological parameters on the sugarcane crop yield prediction model. The experimental analysis was conducted on the 75 districts of the major sugarcane producer states of India. The investigation outcomes were assessed using several performance metrics and concluded that the accuracy of the proposed methodology achieved R^2 value in a range of 0.95 and 0.94 in most of the districts of Maharashtra state with a standard error of 1.04% followed by the Uttar Pradesh state with 5.36%. The lowest accuracy was observed in some of the districts of Bihar state due to noisy seasonal NDVI data for analyzing these regions' predictions. Although the overall result analysis of the considered states was found satisfactory, and among them, Maharashtra districts provided the best predictions in reference to the ground truth. The results also highlight that the coefficient of determination value dropped significantly in the case of remote sensing-based parameters that had been utilized alone for predicting the yield. The assessment of the prediction model accuracy based on the 37 parameters presented that the yield prediction of a crop depends on the timely capture of the weather and NDVI data of a crop during the growing season. The study also evaluated the vegetation cover of the crops in the study regions to quantify vegetation health. As per the NDVI analysis, the Karnataka region had a denser forest area covered rather than the Crop regions in the 2018 year. The results also exhibit a strong correlation of VCI with T2M, T2M_Max, and T2M_Min meteorological parameters in the prediction model.

However, in some districts, the coefficient of determination was found to be low, so in these cases, the incorporation of other vegetation indices (NDWI, EVI, SAVI) along with the meteorological factors of a longer period may improve the accuracy of the model. This study work can be extended in the future by adding additional information of soil fertility level, disease persistence level, irrigation frequency, etc. in the prediction model.

Funding: This research received no external funding.
Acknowledgments: The authors are grateful to the Directorate of Economics and Statistics, Department of Agriculture, Cooperation, and Farmers Welfare to make availability of data.
Conflicts of Interest: The authors declare that they have no conflict of interest.

REFERENCES

1. Dubey SK, Gavli AS, Yadav SK, Sehgal S, Ray SS. Remote Sensing-Based Yield Forecasting for Sugarcane (Saccharum officinarum L.) Crop in India. J. Indian Soc. Remote. Sens. 2018; 46:1823–33.DOI: https://doi.org/10.1007/s12524-018-0839-2
2. Solomon S. Sugarcane Production and Development of Sugar Industry in India. Sugar Tech. 2016; 18(6): 1-15. DOI: https://doi.org/10.1007/s12355-016-0494-2
3. Som-ard J, Atzberger C, Izquierdo-Verdiguier E, Vuolo F, Immitzer M. Remote Sensing Applications in Sugarcane Cultivation: A Review. Remote Sens. 2021; 13(20):4040. DOI: https://doi.org/10.3390/rs13204040
4. Verma AK, Garg PK, Hari Prasad KS, Dadhwal VK, Dubey SK, Kumar A. Sugarcane Yield Forecasting Model Based on Weather Parameters. Sugar Tech.2021; 23:158–66. DOI: https://doi.org/10.1007/s12355-020-00900-4
5. Zhu X, Guo R, Liu T, Xu K. Crop Yield Prediction Based on Agrometeorological Indexes and Remote Sensing Data. Remote Sens.2021;13, 2016. DOI: https://doi.org/10.3390/rs13102016
6. Zhao D, Li YR. Climate change and sugarcane production: Potential impact and mitigation strategies. Int. J. Agron.2015; 547386.
7. Ray SS, Neetu, Sanapala M, Gupta S. Use of Remote Sensing in Crop Forecasting and Assessment of Impact of Natural Disasters: Operational Approaches in India. In conference FAO Expert Meeting on Crop Monitoring for Improved Food Security, Vientiane, Lao PDR;2014.
8. Dimov D, Uhl JH, Löw F, Seboka GN. Sugarcane yield estimation through remote sensing time series and phenology metrics. Smart Agricultural Technology.2022; 2:2772-3755. DOI: https://doi.org/10.1016/j.atech.2022.100046
9. Rahman MM, Robson AJ. A Novel Approach for Sugarcane Yield Prediction Using Landsat Time Series Imagery: A Case Study on Bundaberg Region. Advances in Remote Sensing.2016;5(2):93-102. DOI: https://doi.org/10.4236/ars.2016.52008
10. Bhatla R, Dani B, Tripathi A. Impact of climate on sugarcane yield over Gorakhpur District, UP using a statistical model. Vayu Mandal.2018; 44(1): 11–22.
11. Geetha MCS, Shanthi IE. Predicting the soil profile through modified regression by discretisation algorithm for the crop yield in Trichy district, India. Int. J. Grid Util. Comput. 2018; 9:235–42.
12. Yadav K, Geli HME. Prediction of Crop Yield for New Mexico Based on Climate and Remote Sensing Data for the 1920–2019 Period. Land.2021;10:1389. DOI: https://doi.org/10.3390/land10121389
13. Khaki S, Pham H, Wang L. Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. Sci Rep.2021;11: 11132. DOI: https://doi.org/10.1038/s41598-021-89779-z
14. Khaki S, Wang L, Archontoulis SV. A CNN-RNN framework for crop yield prediction. Front. Plant Sci. 2020; 10:1750.
15. Shah A, Agarwal R, Baranidharan B. Crop Yield Prediction Using Remote Sensing and Meteorological Data. International Conference on Artificial Intelligence and Smart Systems (ICAIS). 2021;952-960. DOI: https://doi.org/10.1109/ICAIS50930.2021.9395849
16. Lopresti MF, Di Bella CM, Degioanni AJ. Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. Information Processing in Agriculture.2015; 2:73–84.
17. Ming B, Guo Y, Tao H, Liu G, Li S, Wang P. SPEIPM-based research on drought impact on maize yield in North China Plain. J. Integr. Agric.2015; 14: 660–9
18. Dubey S, Gavli A, Ray S. Vegetation Condition Index: A Potential Yield Estimator. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch., XLII-3/W6; 2019.
19. Pham HT, Awange J, Kuhn M, Nguyen BV, Bui LK. Enhancing Crop Yield Prediction Utilizing Machine Learning on Satellite-Based Vegetation Health Indices. Sensors (Basel).2022; 22(3):719.

20. Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J, et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol .2019;274:144–59. DOI: https://doi.org/10.1016/j.agrformet.2019.03.010
21. Liu F, He X, Ling F. Research on Rice Yield Prediction Model Based on Deep Learning. Computational Intelligence and Neuroscience.2022; 1922561. DOI: https://doi.org/10.1155/2022/1922561
22. Klompenburg TV, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture. 2020; 177: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2020.105709
23. Shah A, Dubey A, Hemnani V, Gala D, Kalbande DR. Smart Farming System: Crop Yield Prediction Using Regression Techniques. In Proceedings of International Conference on Wireless communication Springer, Singapore.2018; 49-56. DOI: https://doi.org/10.1007/978-981-10-8339-6_6
24. Pandey KK, Maurya D, Gupta G, Mishra SV. Yield Forecasting Models based on Weather Parameters for Eastern U.P. Int. J. Plant Res..2016; 29(1):22. DOI: https://doi.org/110.5958/2229-4473.2016.00006.9
25. Mishra P, Al Khatib MG, Sardar I, Mohammad J, Karakaya K, Dash A et al. Modeling and Forecasting of Sugarcane Production in India. Sugar Tech. 2021;23: 1317–24. DOI: https://doi.org/10.1007/s12355-021-01004-3
26. Ali S, Badar N, Fatima H. Forecasting production and yield of sugarcane and cotton crops of Pakistan for 2013–2030. Sarhad J. Agric.2015; 31 (1): 1–10.
27. Dash A, Mahapatra SK. Using ARIMA model for yield forecasting of important pulse crops of Odisha, India. Amazon. J. Plant Res.2020; 4:646-59.
28. Mishra P, Yonar A, Yonar H, Kumari B, Abotaleb M, Das SS, Patil SG. State of the art in total pulse production in major states of India using ARIMA techniques. Curr Res Food Sci. 2021:28 (4):800-6. DOI: https://doi.org/10.1016/j.crfs.2021.10.009
29. Crop Production Statistics [Internet]. Directorate Of Economics & Statistics, Ministry of Agriculture and Farmers Welfare; 1998 [cited 2022 Mar 2]. Available from: https://aps.dac.gov.in/APY/Public_Report1.aspx
30. Bhardwaj A[Internet]. Benefits of Micro Irrigation System Sugar Recovery & Productivity. Indian Sugar Mills Association; 2013 [cited 2022 Mar 20]. Available from: http://www.indiansugar.com/uploads/Workshop-Micro-irrigation.pdf
31. MODIS Terra MOD13Q1based Product [Internet]. NASA; [cited 2022 Jan 20]. Available from: https://lpdaac.usgs.gov/products/mod13q1v006/ .
32. Meteorological Data. [Internet]. NASA; [cited 2022 Jan 20]. Available from: https://power.larc.nasa.gov/data-access-viewer/
33. Panek E, Gozdowski D. Analysis of relationship between cereal yield and NDVI for selected regions of Central Europe based on MODIS satellite data. Remote Sens. Appl.: Soc. Environ.2020; 17: 100286, 2352-9385. DOI: https://doi.org/10.1016/j.rsase.2019.100286
34. Hunag J, Wang X, Li X, Tian H. Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR. PLoS ONE.2013; 8(8): e70816. DOI: https://doi.org/10.1371/journal.pone.0070816
35. Dubey S, Gavli A, Ray S. Vegetation Condition Index: A Potential Yield Estimator. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.2019; XLII-3/W6. DOI: https://doi.org/10.5194/isprs-archives-XLII-3-W6-211-2019
36. Drought Monitoring System [Internet]. NDRMS. [cited 2022 Mar 10]. Available from: https://dms.sig.ac.in/facts/indices.html
37. Friendly M. Corrgrams: Exploratory displays for correlation matrices. Am. Stat. 2002; 56:316–324.
38. Hahsler M, Hornik K, Buchta C. Getting things in order: An introduction to the R package seriation. J. Stat. Softw. 2008; 25:1–34.
39. Murdoch DJ, Chow ED. A graphical display of large correlation matrices. Am. Stat.1996: 50:178–80.
40. Hammer RG, Sentelhas PC, Mariano JCQ. Sugarcane Yield Prediction Through Data Mining and Crop Simulation Models. Sugar Tech.2020; 22:216–25. DOI: https://doi.org/10.1007/s12355-019-00776-z
41. Dammalage TL, Sirisena T, Susaki J. MODIS satellite data-based rice yield-forecasting model for Sri Lanka: a pilot study on kurunegala district. AJG. 2018;17(3):24-33.
42. Priya SRK, Balambiga RK, Mishra P, Das SS. Sugarcane yield forecast using weather-based discriminant analysis. Smart Agricultural Technology.2022;3:2772-3755. DOI: https://doi.org/10.1016/j.atech.2022.100076