Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 24(3) : 249-254 (September 2022)

https://journal.agrimetassociation.org/index.php/jam

Research Paper

Village level identification of sugarcane in Sangali, Maharashtra using open source data

AMAN LONARE[1], BASANT MAHESHWARI[2] and PENNAN CHINNASAMY[1]*

[1]Center for Technology Alternatives for Rural Areas (CTARA), Indian Institute of Technology, Bombay (IITB), Mumbai, Maharashtra 400076, India

[2]Western Sydney University, Sydney 2751, Australia

**Correspondence author email : p.chinnasamy@iitb.ac.in*

ABSTRACT

Agriculture and crop monitoring are very important for an agrarian country like India. This study is done in June Khed village in the Sangli district of Maharashtra, India to assessing the efficiency of an open source cloud-based remote sensing platform Google Earth Engine (GEE), in the village-scale identification of sugarcane. The ground-truth data was collected and the efficiency of Landsat-8 and Sentinel-2 satellite data was assessed in driving GEE's inbuilt Machine Learning classifiers: Classification and Regression Tree (CART), Support Vector Machine (SVM) and Random Forest (RF). Results were validated with the ground truth data and the official data. Of the methods used, SVM outperformed RF and CART with the lowest relative deviation (+9.2%), highest F1- score (0.8) and overall accuracy (78%), using the Sentinel-2 data. Results also indicated the in-situ use of observation data with high spatio-temporal resolution data. The validated model was then up-scaled for the Walwa Taluka level, to map sugarcane area that can be used for further agriculture tasks such as crop monitoring and yield prediction, leading to better management of crop and better formulating of sugarcane farmer policy.

Keywords: Sugarcane, remote sensing, google earth engine, machine learning, crop classification

India has 60.45% of land under agriculture with majority of the population (~58%) involved in agricultural activities (Gallego *et al*. 2008). Agriculture contributes to 15.96% of Gross Domestic Product (GDP), therefore agricultural management is key in India (Gallego *et al.* 2008). In India, sugarcane is an important commercial crop. The Indian sugar industry, most of it from sugarcane, has a contribution to the national GDP at 1.1%, even though it is grown only in 2.57% of the grossed crop area, and is also supporting more than 550 sugar mills (Solomon 2016). The livelihood of more than 40 million sugarcane farmers, constituting 7.5% of the rural population is dependent on the sugarcane production (Nandhini and Padmavathy 2017). Sugarcane farming in India has been important not only for the economy of the country but also served as an alternate source of fuel. Sugarcane is projected as the crop for the future as it contributes to the production of the renewable source of green energy in the form of bioelectricity, bioethanol and many bio-based products (Solomon 2016; Malik *et al*. 2019). Thus, it is important for the farmer to monitor the sugarcane crop throughout the growing period, to have a good return on investment as expenditure in the 3-year time period is high. With such high interest and benefits from sugarcane production, sugarcane crop mapping with open source data is of utmost importance. Various remote sensing derived indices have been used for crop identification, condition and yield estimation (Prabhakar *et al.* 2019; Anil Kumar *et al.* 2022).

In this study, the primary objective is to investigate the potential of Google Earth Engine cloud platform capabilities for identifying the village level sugarcane area in the selected study area in Maharashtra. Secondary objectives include assessment of the efficiency of GEE in mapping sugarcane area, to assess the efficiency of different ML and AI models in estimating sugarcane crop area and to assess differences, if any, in using different satellite imagery data. The results of this study specifically aims to answer the following research questions:

(1) Does cloud computing based remote sensing platform, using satellite images, have the potential for mapping sugarcane at village level in India?

(2) Which satellite images (Sentinel-2 or Landsat 8) is more efficient in mapping the sugarcane at village level in India?

(3) Which machine learning algorithm is more accurate in mapping the sugarcane in the study area?

Article info - DOI: *https://doi.org/10.54386/jam.v24i3.1688*
Received: 01June 2022; Accepted: 17August 2022; Published online: 31 August 2022

MATERIALS AND METHODS

Study area

The study village, June Khed is located at 17.0700 ^{0}N (Latitude), 74.3624 ^{0}E (Longitude), in Walwa Taluka, Urun Islampur, Sangli district, Maharashtra on the banks of Krishna River. The total area of the village is 452 ha including agriculture fields as well as the Gaonthan (residence area of the village). The study area is a part of the 'sugarcane belt' of the state Maharashtra in India, which is known for high acreage of sugarcane crops.

Remote sensing data

Open source and cloud free satellite images of Landsat 8 and Sentinel – 2, for the study area, were downloaded from the GEE platform for the identification of the sugarcane crop. Different composites derived from Landsat – 8 imagery are available in GEE. In this study, Landsat 8, 8 – Day Top-of-Atmosphere (TOA) reflectance composites with 13 bands, 30 m spatial resolution and 15 days temporal resolution were used. An 8-day composite was selected over the 32 day composite to have better temporal resolution. Similarly, Sentinel – 2, multispectral instrument Level – 1C and Level – 2A images with 13 bands, with 10 meters spatial resolution and 5 days temporal resolution images were used for the analysis.

In-situ sugarcane crop data collection

In this study, Krishi Vibhag of the Walwa Taluka was visited where records of the agriculture related information of the villages constituting the Taluka is stored. For verifying/ground-truthing satellite data, and to understand the cropping patterns and issues, a detailed social survey and field survey were conducted. The field-based ground-truths for the classification were collected in the second week of March, 2020 (*Rabi* season). Ground truths, for validating satellite data, were collected through an android application called SW (SoftWel) Maps, which is used for labelling the field/crop type as a respective class by surveying the field (Softwel 2020).

Crop classification

The classification of the multi temporal satellite images was done as per pixel basis. A pixel is a fundamental unit of a satellite image and in pixel-based classification individual image pixels are analysed for the spectral information and each group of unique pixels is labelled as one class (Tsiligirides 1998). The methodology flowchart used in the study is shown in Fig.1.

Vegetation indices

Six spectral bands, including blue, red, green, near infrared (NIR), shortwave infrared 1 (SWIR1) and shortwave infrared 2 (SWIR2) were selected for the classification task. Additionally, three vegetation indices (VI) have been used namely, Normalised Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Land Surface Water Index (LSWI) were estimated as per (Tucker *et al*. 1985; Sellers *et al*. 1992; Arvor *et al*. 2011; Xia *et al*. 2002). In addition to these, percentile and mean values were calculated using GEE inbuilt methods, i.e. percentile composites and interval mean, to take advantage of the time series of images so that phenology aspect of the classes can be considered which will further improve the accuracy of classification (Hansen *et al.* 2013, Potapov *et al*. 2012). Classification and Regression tree (CART), Random Forest (RF) and Support Vector Machine (SVM) classifiers are used to process the result, further ground truth data and census data are used for assessing the accuracy of the results from classifiers.

Classification using machine learning models

CART, RF and SVM classifiers, inbuilt in the GEE, were employed to identify sugarcane, grapes, wheat, water, urban, forest, barren and groundnut. These classification algorithms are first analysed and then used for the classification of the multi-temporal Landsat 8 TOA composites and Sentinel-2 images acquired from the GEE.

SVM is an approximate implementation of the method of structural risk minimisation (Fletcher 2000) and the quadratic programming (QP) problem is globally optimised. Classification and Regression Tree (CART) is a supervised machine learning algorithm is widely used for crop classification. Random Forest

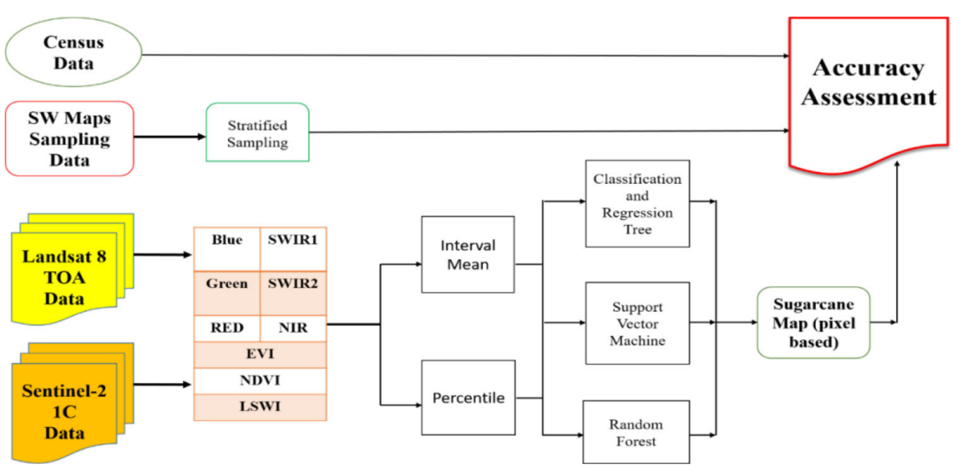

Fig. 1: Methodology for crop classification using cloud computing based methods

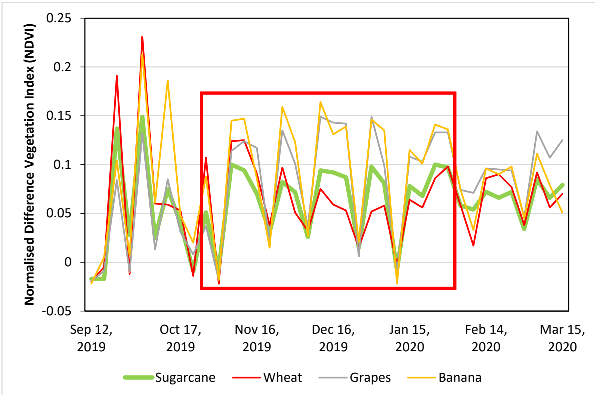

Fig. 2: Temporal variation in the NDVI values of the 4 crops in June

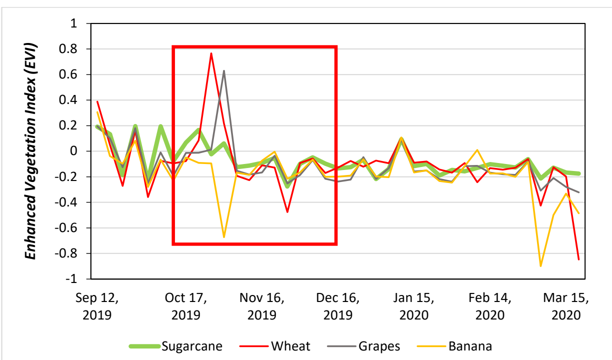

Fig. 3: Temporal variation in the EVI values of the 4 crops in June Khed, India

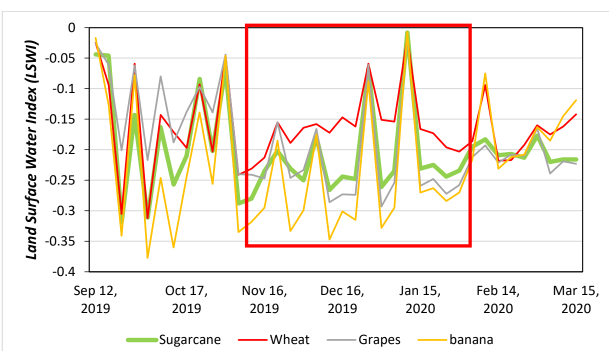

Fig. 4: Temporal variation in the LSWI values of the 4 crops in June Khed study area in Maharashtra, India

(RF) is an ensemble of Decision Trees (DT) which overcomes one of the major shortcoming of the Decision Trees i.e. overfitting (Mathur and Foody 2008).

Accuracy assessment

Of the total sites, where ground-truth data were collected, 20% were used as validation sites while the remaining were used for training the classifier. The GEE has inbuilt function 'errorMatrix' which calculates the 'ConfusionMatrix' to assess the accuracy of the classifier. Five accuracy metrics were calculated for comparing the results of classifiers.

i. Relative Deviation (RD): It is the percentage by which the model overestimated or underestimated the area of the sugarcane by the model.

ii. Overall Accuracy (OA): It is the proportion of validation sites that were mapped correctly by the model.

iii. Producer's Accuracy (UA): It is defined as the ratio of the number of validation sites of a class classified accurately to the total number of the validation sites for the class.

iv. User's Accuracy (UA): It is defined as the ratio of the number of validation sites of a class correctly classified to the number of sites classified as that class.

v. F1-score: It is the harmonic mean of UA and PA.

RESULTS AND DISCUSSIONS

Normalised difference vegetation index (NDVI)

Results indicated spatial variations in NDVI results across the crop types and temporally depending on the growth period. Fig. 2 shows the seasonal temporal variation of NDVI, used as an input band for preparing the classification map for sugarcane. It can be seen that during the mid-growth period (mid-November to late January) of the sugarcane (highlighted in the red box), the NDVI values of sugarcane (in light green) is different from that of the other crop types. Thus, with this spatiotemporal variation in NDVI, the sugarcane crop can be distinguished from the other crops (banana, grapes and wheat) and was used as a feature for classification in the further steps.

Khed, India

Enhanced vegetation index (EVI)

Since EVI is responsive to the canopy structural variations, while NDVI is chlorophyll sensitive, EVI results showed signatures of sugarcane that was used for classification. Fig.3 shows the temporal variation of EVI, used as an input band for preparing the crop classification. Results indicate that in the initial growth stage (early October to mid-December) of the crops (highlighted in the red box), the EVI value of sugarcane crops (in light green) can be easily distinguished from other crops. Thus EVI, as one of the features, was used for generating the classification map of the study area.

Land surface water index (LSWI)

Results of the LSWI estimation, shown in Fig. 4, indicate temporal variation and hence were also used as an input band for preparing the classification map. It can be seen that during the mid-growth period (mid-November to late January) of the sugarcane (highlighted in the red box), the LSWI values of sugarcane (in light green) is different from that of the other crop types. Thus, with this temporal variation in LSWI, the sugarcane crop can be distinguished from the other crops (banana, grapes and wheat) and used as a feature for classification.

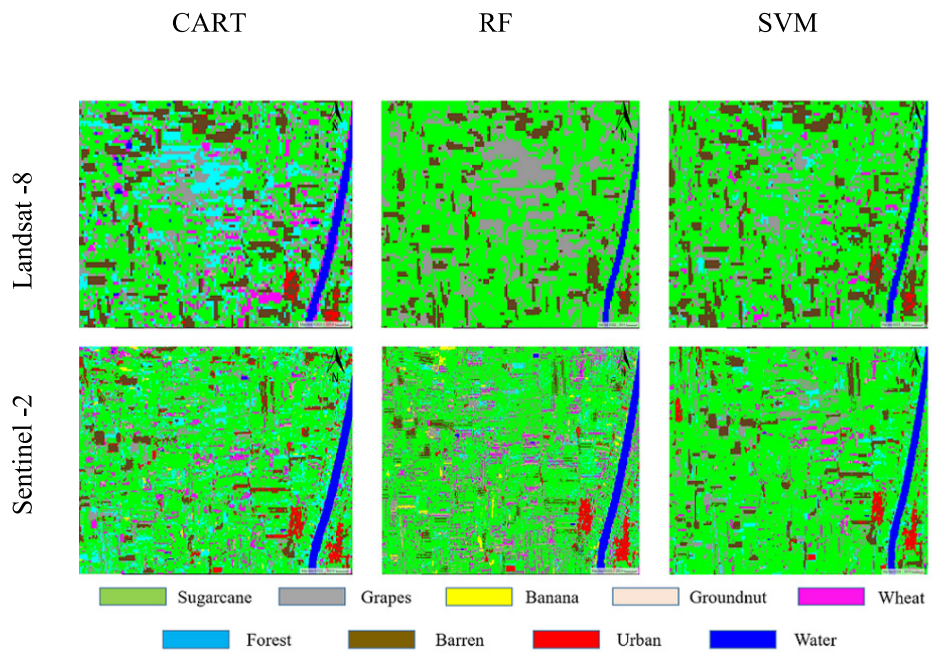

Fig. 5: Crop classification map produced using Classification and Regression Tree (CART), Random Forest (RF) and Support Vector Machine (SVM) on Landsat 8 (top row) and Sentinel-2 (bottom row) satellite imagery for the study areas in Maharashtra

Table 1: Classification algorithm's estimation accuracy for sugarcane crop area using Landsat- 8 and Sentinel-2 data

Classification Algorithm	Landsat- 8		Sentinel-2	
	Estimated area (ha)	Relative deviation (%)	Estimated area (ha)	Relative deviation (%)
Classification and regression tree (CART)	196	-19%	271.73	+11%
Support vector machine (SVM)	276	+12%	267.59	+9.2%
Random forest (RF)	296	+21%	298.41	+21.8%

Sugarcane crop area classification

Using the VIs estimated from above (NDVI, EVI and LSWI) and the methodology discussed in Fig. 1, the crop classification was mapped using the three different algorithms (CART, RF and SVM) for two different datasets: Landsat-8 and Sentinel-2.

Results on sugarcane classification using Landsat-8 indicate variations between classification algorithms (Table 1) with RF giving the highest area (296 ha), followed by SVM and CART with 276 and 196 ha, respectively. On the other hand, the sugarcane area prediction using Sentinel-2 data also indicated variations between algorithms (Table 2), with RF giving the highest area (298 ha) followed by CART and SVM, with 271 and 267 ha, respectively. The sugarcane area estimation comparison between models and between the different satellites data are shown in Fig. 5, which clearly indicates spatial resolution differences, given that the Setinel-2 has a higher spatial resolution of 10m, compared to 30m of Landsat-8 data. However, due to the presence of variations in results obtained between algorithms and between the satellites data used, it was necessary to conduct an accuracy assessment.

Accuracy assessments

The total sugarcane cultivation area in the village June Khed (from Government sources and ground survey) is 245 ha, i.e.54% of the total village area. The difference between the estimated and the actual value of the sugarcane area from the Landsat-8 and Sentinel-2 images is given in the Table 1 and Table 2, respectively.

From the above Tables 1, it is evident that the SVM model had the least relative deviation % while driven by both satellite datasets, Sentinel-2 and Landsat-8, thus indicating that SVM performed well for this area. Also, of the two datasets, it was also evident that the Sentinel-2 driven SVM model was the best with least relative deviation of 9.2%. In addition to the relative deviation assessment, Table 2 shows the Overall Accuracy (OA), User's Accuracy (UA), Producer's Accuracy (PA), and F1-score calculated from the confusion matrix method for the three algorithms on Landsat-8 and Sentinel-2 data.

Similar to the relative deviation results, Table 2 indicates that SVM model driven with Sentinel-2 data had the

Table 2: Validation accuracy obtained for sugarcane classification between different datasets and different algorithms

Classifier	Satellite	Overall accuracy	User's accuracy	Producer's accuracy	F1-score
Classification and Regression Tree	Landsat-8	57%	48%	98%	0.64
	Sentinel-2	70%	57%	97%	0.72
Support Vector Machine	Landsat-8	63%	63%	97%	0.76
	Sentinel-2	78%	68%	98%	0.80
Random Forest	Landsat-8	62%	55%	97%	0.70
	Sentinel-2	78%	65%	99%	0.79

best performance when compared against other models and data. The GEE based identification of the sugarcane crop is found to be reliable with an overall accuracy of 78% using Sentinel-2 satellite images with SVM classifier. The F1-score value is also found to be high (0.80) as compared to the previous studies (Roy *et al.* 2014) in which the coarser resolution image (250 m) were used for large scale classification. The accuracy of Sentinel-2 satellite images based classification, from the current study at a 10m spatial resolution across a village, is found to be higher than Landsat-8 based maps in majority of the cases. The major reason being higher spatial and temporal resolution of Sentinel-2 images (10 m, 5 days) as compared to Landsat 8 images (30 m, 15 days) and due to the presence of Multi Spectral Imager with higher number of bands in the red region of the spectrum.

CONCLUSIONS

This study focused on assessing the potential of open source cloud-based online remote sensing platform and Google Earth Engine in the identification of the sugarcane crop at the June Khed village in India. Various vegetation indices viz. NDVI, EVI and LSWI, different ML models (including SVM, RF and CART) driven by different satellite data (Sentinel-2 and Landsat-8) were used for sugarcane crop classification and comparisons were made between models, satellite and observation data. The overall accuracy and F1 - score of SVM driven by Sentinel-2 data, was found to be the highest as 78% and 0.80, respectively. The highest accuracy for classification algorithm are achieved on Sentinel-2 data as the spatial resolution (10 m) and temporal resolution (5 days) are better than that of Landsat 8 (30 m, 15 days). The sugarcane maps of the areas (village and taluka level) prepared can aid in further development of Decision Support Tools (DSTs) which can estimate acreage, predict the yield of the sugarcane, aid in growth monitoring tasks and can eventually help farmers in better farm management and thus, maximizing the profit.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the open source satellite repositories from NASA (USA) and European Space Agency (ESA). The authors acknowledge the local sugarcane mills for providing data on yield.

Conflict of Interest Statement: The author (s) declares (s) that there is no conflict of interest.

Disclaimer: The contents, opinions and views expressed in the research article published in Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Publisher's Note: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Anil Kumar, D., T. L. Neelima, P. Srikanth, M. Uma Devi, K. Suresh, & C. S. Murthy. (2022). Maize yield prediction using NDVI derived from Sentinal 2 data in Siddipet district of Telangana state. *J. Agrometeorol.*, *24*(2): 165–168. https://doi.org/10.54386/jam.v24i2.1635

Arvor, D., Jonathan, M., Meirelles, M.S.P., Dubreui, V., D urieux, L. (2011). Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *Int. J. Remote Sens.* 32(22): 7847-7871. doi.org/10.1080/01431161.2010.531783

Fletcher R. (2000). Practical methods of optimization. John Wiley & Sons. New York. 80, 4: ISBN: 978-0-471-49463-8.

Gallego, J., Carfagna, E., Baruth, B. (2008). Accuracy, objectivity and efficiency of remote sensing for agricultural statistics. *Agric. Surv. Methods*, 10 (4): 193-211. doi.org/10.1002/9780470665480.ch12

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Kommareddy, A. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160): 850-853. doi: 10.1126/science.1244693

Malik, K., Verma, D.K., Srivastava, S., Mehta, S., Kumari, N., Verma, A., Singh, K.P. (2019). Sugarcane Production and Its Utilization as a Biofuel in India: Status, Perspectives, and Current Policy. *Sugarcane Biofuels*, 123-138. doi.org/10.1007/978-3-030-18597-8_6

Mathur, A., Foody, G.M. (2008). Crop classification by support vector machine with intelligently selected training data for an operational application. *Int. J. Remote Sens.,* 29(8): 2227-2240. doi.org/10.1080/01431160701395203

Nandhini, K.D. and Padmavathy, V. (2017). A Study on Sugarcane Production in India. *Int. J. Adv. Res. Botan.,* 3(2): 13-17.

doi.org/10.1080/01431160701395203

Potapov, P.V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., Justice, C.O. (2012). Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. *Remote Sens. Environ.* 122: 106-116. doi.org/10.1016/j.rse.2011.08.027

Prabhakar, M., Merugu Thirupathi, G. Srasvan Kumar, U. Sai Sravan, M. Kalpana, K.A. Gopinath, & N. Ravi Kumar. (2019). Damage assessment of chilli thrips using high resolution multispectral satellite data. *J. Agrometeorol.*, *21*(4): 480–487. https://doi.org/10.54386/jam.v21i4.284

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Scambos, T.A. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.,* 145: 154-172. doi.org/10.1016/j.rse.2014.02.001

Sellers, P.J., Berry, J.A., Collatz, G.J., Field, C.B., Hall, F.G. (1992). Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sens. Environ.,* 42(3): 187-216. doi.org/10.1016/0034-4257(92)90102-P

Softwel (2020). Softwel Productivity, SW Maps - GIS & Data Collector. Available Online: https://play.google.com/store/apps/details?id=np.com.softwel.swmaps (Accessed on 30th June 2020)

Solomon, S. (2016). Sugarcane production and development of sugar industry in India. *Sugar Tech.* 18(6): 588-602. doi.org/10.1007/s12355-016-0494-2

Tsiligirides, T.A. (1998). Remote sensing as a tool for agricultural statistics: a case study of area frame sampling methodology in Hellas. *Comput. Electron. Agric.* 20(1): 45-77. doi.org/10.1016/S0168-1699(98)00011-8

Tucker, C.J., Vanpraet, C.L., Sharman, M.J., Van Ittersum, G. (1985). Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel: 1980–1984. *Remote Sens. Environ.,* 17(3): 233-249. doi.org/10.1016/0034-4257(85)90097-5