Performance Evaluation of RF and SVM for Sugarcane Classification using Sentinel 2 Time Series NDVI

Shyamal Virnodkar[1*], Dr. V. K. Pachghare [1], Dr. V. C. Patil [2], Sunil Kumar Jha [2]

[1] Department of Computer Engineering & IT, College of Engineering, Pune, SPPU, India.

[2] K. J. Somaiya Institute of Applied Agricultural Research, Karnataka, India.

[1]ssv18.comp@coep.ac.in, vkp.comp@coep.ac.in

[2]patil.vc@somaiya.com, jha.sunilkumar@somaiya.com

Abstract. Sentinel-2 optical time-series images obtained at high resolution are creditable for cropland mapping which is the key for sustainable agriculture. The presented work conducted in a heterogeneous region in Sameerwadi with an aim to classify sugarcane crops, with mainly two groups so as to provide sugarcane field map, using Sentinel-2 NDVI time series data. The potential of two better-known ML classifiers, random forest (RF) and support vector machine (SVM), was investigated to identify 7 classes including sugarcane, early sugarcane, maize, waterbody, fallow land, built up, bare land, and a sugarcane crop map is produced. Both the classifiers were able to effectively classify sugarcane areas and other land covers from the time series data. Our results show that the RF achieved higher overall accuracy (88.61 %) than the SVM having an overall accuracy of 81.86%.s This study demonstrated that utilizing Sentinel-2 NDVI time series with RF and SVM successfully classified sugarcane crop fields.

Keywords: Sentinel-2, NDVI, RF, SVM, Sugarcane classification.

1 Introduction

Agriculture plays an important role in the economy of India. To attain sustainable agriculture practice accurate crop mapping needs to be in place. Satellite imagery provides timely, accurate and detailed spatial information about an agro-ecological environment [1]. Crop mapping using satellite imagery would help in providing essential and accurate information about the crops, useful to manage many agricultural resources [2]. However, crop classification using remote sensing data is a challenging task due to crop heterogeneity and similar reflectance in fields. Various machine learning algorithms have been successfully investigated for cropland mapping from single-date to time-series remote sensing images. The cropland mapping techniques applied to time-series images have been demonstrated to perform superior to single-date mapping techniques [3-4]. For example, Muller [5] successfully differentiated cropland and pasture fields from Landsat time series and Zheng [6] applied SVM model on time series Landsat Normalised Difference Vegetation Index (NDVI) data for identification of crop type. Time series Landsat is explored with ensemble classifier and with other ML methods like SVM, Neural Network, logistic regression, extreme gradient boosting for land cover classification [7]. Senf [8] used Landsat time-series imagery, and multi-seasonal MODIS to classify crops from savannah. Jia [9] researched the adequacy of phenological features processed from MODIS NDVI time series melded with NDVI data obtained from Landsat 8 for cropland mapping. MODIS-Terra/ Enhanced Vegetation Index (EVI) time series have been effectively used to derive the phenological patterns for the classification of cotton, maize, soybean, and non-commercial crops in Brazil [10]. MODIS-Terra EVI has also been used to detect phenological stages, and MODIS NDVI to extract phenological information like the season, peak and end of the season [11] of rice crop. Double cropping, single cropping, forest, and pastures were mapped using the patterns of vegetation dynamics identified from MODIS EVI data by Maus [12].Landsat, MODIS, and Chinese HJ-1 time series have been successfully explored for sugarcane crop classification. Time series Landsat 8 [13] and time series Chinese HJ-1 CCD images [14] were used to automatically map sugarcane over large areas by applying object-based image analysis and data mining techniques. Sugarcane cropping practices, including crop type and harvest mode, were mapped using Landsat 8 NDVI time series by [15]. Time series of SPOT 5 images were integrated with crop growth model and expert knowledge to deal with the issue of missing acquisitions or uncertain radiometric values by [16] in order to detect sugarcane harvest.

Many studies have been investigated the potential of a single date Sentinel-2 image to classify crops including sugarcane using RF, SVM, DT machine learning methods. Furthermore, applying RF, DTW algorithms on time series of Sentinel 2 produced the best results for cropland mapping [17], but is not yet explored for sugarcane crop classification. So, considering the affordability of high spatial-temporal resolution of Sentinel-2 data, and the potential of RF, SVM ML approaches, this study aimed to evaluate the effectiveness of time series Sentinel 2 images and potential of RF and SVM on this data to classify sugarcane crop from other land covers. Rest of the paper is organized as follows: section 2 describes the study area and the data; section 3 presents the proposed methodology; section 4 discusses the results followed by a conclusion.

2 Study Area and Data

2.1 Study Area

The study area, is located near to Sameerwadi, Karnataka, India at 16.38980 N and 75.03710 E (Fig. 1). Sameerwadi is a village situated in Mudhol taluka, Bagalkot district of Karnataka state in India. The study area covers four districts i.e. Mudhol, Jamkhandi, Raibag, and Gokak, and around 8 lack acres of land. The area has an altitude of 541 m above sea level with annual precipitation is around 545 mm. The climate is generally dry and the temperature ranges between 16.2^0 C to 38.7^0 C. Sugarcane is the main crop cultivated in this region, apart from maize, turmeric, and banana. Fig. 1 depicts the study area.

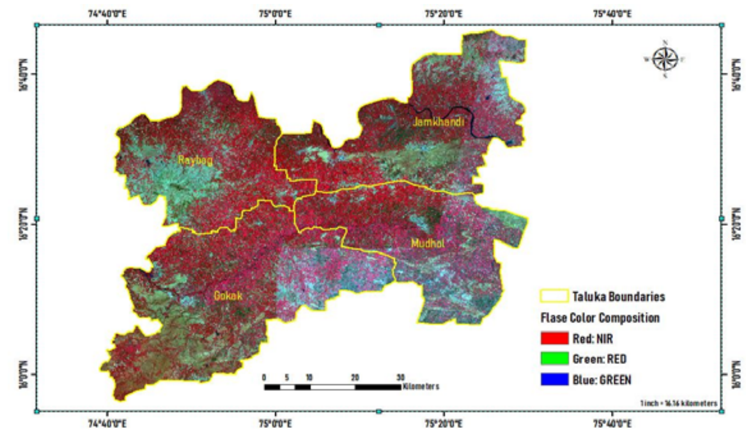

Fig. 1. FCC image of the study area

2.2 Sugarcane Crop Cycle

The phenology of sugarcane may provide valuable information for remote sensing classification in the study area. Sugarcane crop's phenological dynamics throughout its biological cycle needs to be perceived well to understand its spectral behavior, which is vital because of its great impact on classification accuracy. Depending on the planting date, sugarcane has 3 growth cycles i.e. 12 months (Early season), 14 months (Mid-late season), and 18 months (Late season) in the study region. The 12 months crop is planted in the months of January and February, 14 months crop is planted between November - December, whereas 18 months crop is planted during July - August. After harvesting for the first time, the crop is re-grown again 3-4 times and harvested after every 12 months. This practice is referred to as 'ratoon'. In addition to this, it is important to take in the growth stages and varieties of sugarcane in the classification task. There exists four stages germination, tillering, grand growth, and maturity of sugarcane with varieties of CO 86032, CO 91010, SNK 2005, 265 in the study region. Due to these properties of sugarcane, a satellite image acquired on a particular date contains variations in fields which include different growth stages of sugarcane crop, plant cane and ratoon cane, sugarcane varieties, and other crops cultivated for the crop rotation purpose. This necessitates the use of multi-temporal images to perform the classification with the best accuracy. By appropriately utilizing time-series remote sensing images, the phenology of sugarcane, which can be utilized to separate the sugarcane crop fields from the other land, may diminish the obstruction of comparative spectra from the other vegetation in the range and help in inceasing the classification accuracy.

2.3 Data

The Sentinel-2 launched on June 23, 2015, is an Earth Observation (EO) mission from the EU Copernicus program that captures optical imagery at a high resolution of 10m to 60m for the services and applications for agriculture monitoring, land cover classification, water quality, and emergencies management. It has 13 bands out of which one of the three visible bands (band 4) and the near-infrared band (band 8) were used in our study. The images were downloaded from the European Space Agency's (ESA) Sentinel Scientific Data Hub which is open source. Five number of satellite titles per month, used to obtain study area, are obtained from January 20, 2019, to May 07, 2019, as listed in Table 1. The selected temporal images were free from cloud coverage and with good quality. The images were geo-referenced to WGS 1984 UTM zone 43N projection system. EU Copernicus program provides images with geometrically and radiometrically corrections. All the images were atmospherically corrected using Semiautomatic Classification Plugin (SCP) available on QGIS 2.18 and are distributed under the GNU GPL license.

Table 1. Sentinel 2 images used in the study

Image No.	Satellite	Date (dd/mm/yyyy)
1	Sentinel 2A	20/01/19 to 22/01/19
2	Sentinel 2A	24/02/19 to 26/02/19
3	Sentinel 2A	06/03/19 to 08/03/19
4	Sentinel 2A	10/04/19 to 12/04/19
5	Sentinel 2A	05/05/19 to 07/05/19

3 Methodology

The proposed methodology is depicted in Fig. 2 contains following steps: i) acquisition of Sentinel 2 temporal data ii) atmospheric corrections of all the images iii) NDVI computation iv) preparing an input image v) selection of training samples and generation of Region of Interest (ROI) files vi) classification using RF vii) classification using SVM viii) classification accuracy assessment.

3.1 Data Aquisition and Pre-processing

As listed in Table 1 Sentinel-2 images were obtained free of cost from the Copernicus website. All images were atmospherically corrected to reduce the effects of the atmosphere to produce the surface reflectance values. It helps in improving the use and interpretability of images.

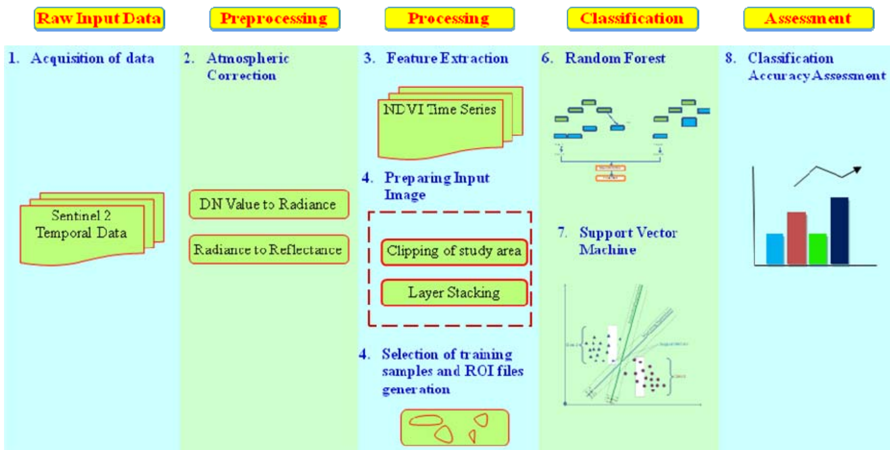

Fig. 2. The proposed methodology

3.2 Data Collection and Preparation of Training Set

The classification was performed based on the NDVI values of the crops during the period from Jan. 2019 to May 2019. We have selected the NDVI as it is proven to be the best Vegetation Index (VI) in the literature for crop mapping [6-7][9][11][15]. All pre-processed images' NDVI computation is performed to get NDVI time series. Then the study area is extracted from these images and layered stacked to generate a multispectral input image for the classification of sugarcane crop. Every pixel of the stacked image represents a vector containing NDVI values corresponding to the considered images.

Training Data Set: Training data set has been created by fields survey which was performed from January 2019 to May 2019. In this field campaign, ground truth data has been recorded by the Global Positioning System (GPS) device (Montana 680) for sugarcane and maize crops. Apart from this, samples for other classes were generated from a visual interpretation based on expert knowledge. In total, 14 sugarcane polygons, and 06 maize polygons surveyed in fields were used for training and 40 polygons were generated for all other classes. In the study area, during the sugarcane developing cycle, various phenological stages of sugarcane fields may coincide on one same date, ranging from the region of reaped sugarcane, and sugarcane in different growth phases up to the phase of grown-up sugarcane ready to harvest. In this way, we attempted to collect samples of all sugarcane phenological stages, with the goal that all the significant subclasses would be represented. The testing polygons are distinct than training polygons. The polygons were selected from the different agricultural parcels to account many other factors such as soil, water source, climate, cultivation practices, etc.

3.3 Classification:

Random Forest: Random forest is a nonparametric, ensemble method [18] based on the Classification and Regression Trees (CART). A classification tree iteratively splits the bootstrap data into pure subsets. Many such independent classification trees are generated by setting Ntree and mtry hyper parameters. The ensemble's final decision is taken from the majority vote of the predictions of all the trees. RF has shown magnificent performance in remote sensing applications [19-22] due to the capability of handling large input variables, run on large dataset, to handle outliers and to provide the importance of predictive variables on final model performance [23-24]. RF also achieved significant accuracy in sugarcane classification [1-2].

Support Vector Machine: SVM is a statistical learning method used for solving classification as well as regression problems. It does not assume the distribution of data and finds an optimal hyperplane between the two classes to be classified. It is basically a two-class classification method but can be extended for multiclass problems [25-26]. The main capability of SVM of achieving high accuracy even with less number of training samples made them very useful in remote sensing applications [27-28]. SVM is proven to be one of the best ML methods in various remote sensing applications mainly include crop classification [26], biotic stress detection [29], yield estimation, and LULC [25][30-31.

Sugarcane crop has varying crop cycle and diverse planting and harvesting dates make classification complex. We first, classified Sentinel-2 NDVI time series using ground truth data and supervised classification into 7 classes using RF and SVM classifiers. The classes are sugarcane (sugarcane crops having age more than 6 months), early sugarcane (sugarcane with age less than 6 months), maize, water body, fallow land, built-up, and bare land. Both the models were trained using training dataset. Both models are widely used models in the crop classification and are tuned with the hyper parameters to achieve maximum accuracy. Open-source R software is used to implement RF and SVM classifiers. Then, recoding of the assigned classes was performed in post-classification through ENVI software. This resulted in one early sugarcane class and other grown-up sugarcane class. This formed the sugarcane map for 4 districts region.

3.4 Accuracy Measures

In remote sensing, accuracy is the measure to validate the correctness and quality of the generated classification maps. The evaluation is performed through the overall accuracy and kappa coefficient measures, and accuracy of an individual class is measured through producer's and user's accuracy. Sometimes F1-score has been used to determine class-wise accuracy [2]. In this work, the accuracy was determined with overall accuracy and kappa coefficient measures.

4 Results and Discussion

Sentinel-2 five tiles, covering the study area, in every month for five months was obtained, layered stacking of NDVI was performed and the resultant image was used for sugarcane and other land cover classification. Two well-known ML classifiers RF and SVM discriminated the sugarcane and other classes very well. The RF's overall accuracy is obtained as 88.61% and the kappa coefficient is 0. 8387 (Table 2).

Table 2. Accuracy Assessment of RF and SVM

	Overall Accuracy	Kappa Coefficient
RF	88.61 %	0.8387
SVM	81.86 %	0.7623

The classified image using RF is shown below in Fig. 3. The optimum accuracy was achieved by tuning the parameter mtry with value 2. The SVM's achieved overall accuracy as 81.86% and kappa coefficient is 0.7623 (Table 2) and the classified image is given in Fig. 4. From Table 3 and 4 it is observed that the work resulted in classifying sugarcane, early sugarcane, builtup and bareland classes more accurately by RF than SVM. Fallow land class achieved the lowest producer's accuracy with RF and Maize is less accurately classified by SVM.

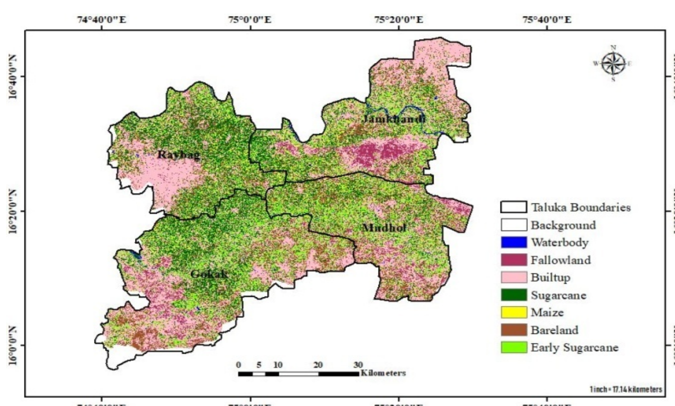

Fig. 3. The classified image by RF

Table 3. Producer's and user's accuracy for RF

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy in %	User's Accuracy in %
Waterbody	10	11	9	90.00	81.00
Fallowland	13	8	5	38.46	62.50
Builtup	87	92	85	97.70	92.39
Sugarcane	29	29	26	89.66	89.66
Maize	10	8	7	70.00	87.50
Bareland	22	22	20	90.91	90.91
Early Sugarcane	30	31	27	90.00	87.10

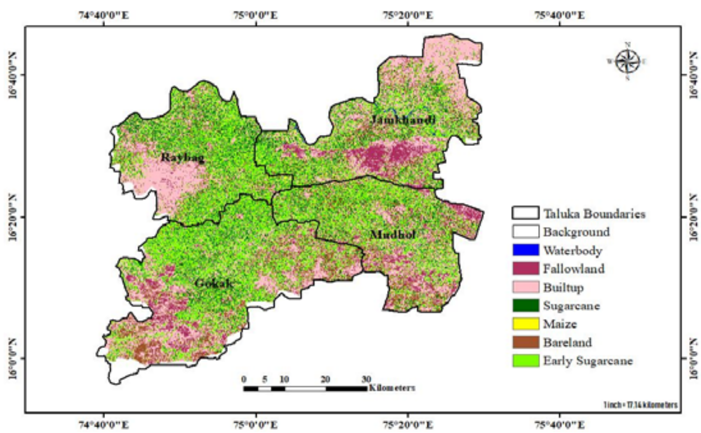

Fig. 4. The classified image by SVM

Table 4. Producer's and user's accuracy for SVM

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy in %	User's Accuracy in %
Waterbody	6	7	5	83.33	71.43
Fallowland	34	25	23	67.65	92.00
Builtup	39	36	33	84.62	91.67
Sugarcane	43	47	38	88.37	80.85
Maize	15	11	9	60.00	81.81
Bareland	23	30	22	95.65	73.33
Early Sugarcane	33	37	28	84.85	75.68

The total area classified into each of the classes by RF and SVM is presented in Fig. 5.

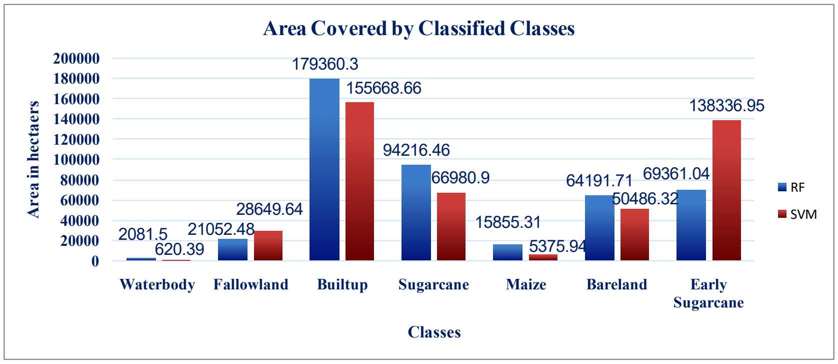

Fig. 5. Class wise coverage of total area in hectors

After classifying the time series image into seven classes, reclassification was performed resulted in two sugarcane classes (early sugarcane and groun up sugarcane) and sugarcane map is generated which are shown in Fig.6 and Fig. 7.

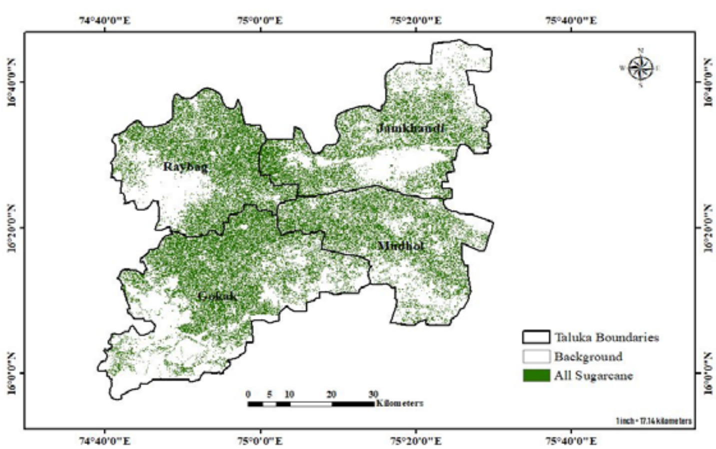

Fig. 6. Sugarcane map on RF classified image

5 Conclusion

In this study, we evaluated the potential of RF and SVM to discriminate the sugarcane crop from other land covers using Sentinel-2 NDVI time series images and a limited number of training polygons. We utilized Sentinel-2 images of five

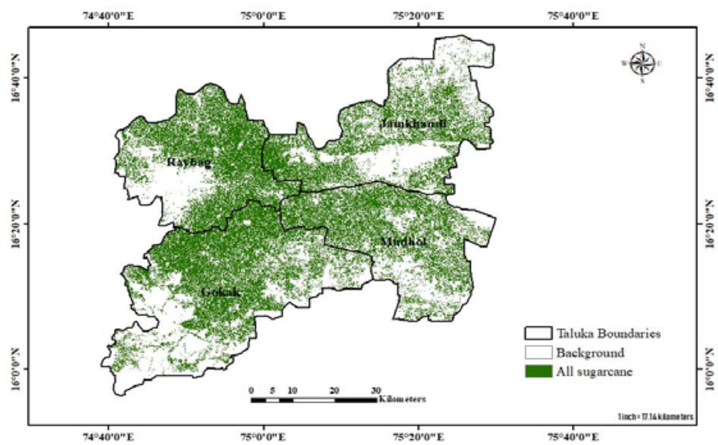

Fig. 7. Sugarcane map on SVM classified image

months from January to May 2019 which covers two main phenology of sugarcane i.e. tillering and grand growth, January to December temporal coverage is required for precise crop classification. The achieved producer's and user's accuracies are reaching to 97.70 and 92.39 resp. Total accuracy for RF is reached about 88.61%, whereas SVM reached at 81.86% states that RF worked well for classification. Thus from results, we conclude that our spectral temporal approach for classification gave reliable discrimination between sugarcane and other land covers. Future investigation will be to evaluate different vegetation indices like GNDVI, EVI, etc from time series data to discriminate all four phenology of sugarcane crop in the study area.

Acknowledgments

The authors would like to thank the staff of KIAAR and GBL, Sameerwadi, Karnataka, India for their support and efforts in collecting ground truth data of crop plots used as the training set in this study.

References

1. Everingham, Y. L., Lowe, K. H., Donald, D. A., Coomans, D. H., & Markley, J.: Advanced satellite imagery to classify sugarcane crop characteristics. Agronomy for Sustainable Development, 27(2), 111–117 (2007).

2. Saini, R., & Ghosh, S. K.: Crop Classıfıcatıon on Sıngle Date Sentınel-2 Imagery usıng Random Forest and Suppor Vector Machıne. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences (2018).
3. Gomez, C., White, J. C., & Wulder, M. A.: Optical remotely sensed time series data for land cover classification: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 116, 55–72 (2016).
4. Long, J. A., Lawrence, R. L., Greenwood, M. C., Marshall, L., & Miller, P. R.: Object-oriented crop classification using multitemporal ETM+ SLC-off imagery and random forest. GIScience & Remote Sensing, 50(4), 418–436 (2013).
5. Muller, H., Rufin, P., Griffiths, P., Siqueira, A. J. B., & Hostert, P. :Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. Remote Sensing of Environment, 156, 490–499 (2015).
6. Zheng, B., Myint, S. W., Thenkabail, P. S., & Aggarwal, R. M.: A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. International Journal of Applied Earth Observation and Geoinformation, 34, 103–112 (2015).
7. Man, C. D., Nguyen, T. T., Bui, H. Q., Lasko, K., & Nguyen, T. N. T. :Improvement of land-cover classification over frequently cloud-covered areas using Landsat 8 time-series composites and an ensemble of supervised classifiers. International Journal of Remote Sensing, 39(4), 1243–1255 (2018).
8. Senf, C., Leitao, P. J., Pflugmacher, D., van der Linden, S., & Hostert, P. Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sensing of Environment, 156, 527–536 (2015).
9. Jia, K., Liang, S., Zhang, N., Wei, X., Gu, X., Zhao, X., … Xie, X. :Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data. ISPRS Journal of Photogrammetry and Remote Sensing, 93, 49–55 (2014).
10. Boschetti, M., Stroppiana, D., Brivio, P. A., & Bocchi, S. :Multi-year monitoring of rice crop phenology through time series analysis of MODIS images. International Journal of Remote Sensing, 30(18), 4643–4662 (2009).
11. Arvor, D., Jonathan, M., Meirelles, M. S. P., Dubreuil, V., & Durieux, L. : Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. International Journal of Remote Sensing, 32(22), 7847–7871 (2011).
12. Maus, V., Câmara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., & de Queiroz, G. R. :A time-weighted dynamic time warping method for land-use and land-cover mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(8), 3729–3739 (2016).
13. Vieira, M. A., Formaggio, A. R., Renno, C. D., Atzberger, C., Aguiar, D. A., & Mello, M. P. (2012). Object based image analysis and data mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. Remote Sensing of Environment, 123, 553–562.
14. Zhou, Z., Huang, J., Wang, J., Zhang, K., Kuang, Z., Zhong, S., & Song, X. (2015). Object-oriented classification of sugarcane using time-series middle-resolution Remote Sensing data based on adaboost. PloS One, 10(11), e0142069.
15. Mulianga, B., Begue, A., Clouvel, P., & Todoroff, P. (2015). Mapping cropping practices of a sugarcane-based cropping system in Kenya using remote sensing. Remote Sensing, 7(11), 14428–14444.
16. El Hajj, M., Begue, A., Guillaume, S., & Martine, J.-F. :Integrating SPOT-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices—The case of sugarcane harvest on Reunion Island. Remote Sensing of

Environment, 113(10), 2052–2061 (2009).
17. Belgiu, M., & Csillik, O. :Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sensing of Environment, 204, 509–523 (2018).
18. Breiman, L. :Random forests. Machine Learning, 45(1), 5–32 (2001).
19. Mohite, J., Karale, Y., Pappula, S., TP, A. S., Sawant, S. D., & Hingmire, S. :Detection of pesticide (Cyantraniliprole) residue on grapes using hyperspectral sensing. In Sensing for Agriculture and Food Quality and Safety IX (Vol. 10217, p. 102170P) (2017).
20. Poona, N., Van Niekerk, A., & Ismail, R. :Investigating the utility of oblique tree-based ensembles for the classification of hyperspectral data. Sensors, 16(11), 1918 (2016).
21. Yin, H., Pflugmacher, D., Li, A., Li, Z., & Hostert, P. :Land use and land cover change in Inner Mongolia-understanding the effects of China's re-vegetation programs. Remote Sensing of Environment, 204, 918–930 (2018).
22. Loggenberg, K., Strever, A., Greyling, B., & Poona, N. :Modelling water stress in a Shiraz Vineyard using hyperspectral imaging and machine learning. Remote Sensing, 10(2), 202 (2018).
23. Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93–104 (2012).
24. Truong, Y., Lin, X., & Beecher, C. :Learning a complex metabolomic dataset using random forests and support vector machines. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 835–840) (2004).
25. Mountrakis, G., Im, J., & Ogole, C. :Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), 247–259 (2011).
26. Khobragade, A., Athawale, P., & Raguwanshi, M. :Optimization of statistical learning algorithm for crop discrimination using remote sensing data. In Advance Computing Conference (IACC), 2015 IEEE International (pp. 570–574) (2015).
27. Zheng, B., Myint, S. W., Thenkabail, P. S., & Aggarwal, R. M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. International Journal of Applied Earth Observation and Geoinformation, 34, 103–112 (2015).
28. Foody, G. M., & Mathur, A. :A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on Geoscience and Remote Sensing, 42(6), 1335–1343 (2004).
29. Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., & Plümer, L. :A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precision Agriculture, 16(3), 239–260 (2015).
30. Hawrylo, P., Bednarz, B., Wkezyk, P., & Szostak, M. :Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2. European Journal of Remote Sensing, 51(1), 194–204 (2018).
31. Warner, T. A., & Nerry, F. :Does single broadband or multispectral thermal data add information for classification of visible, near-and shortwave infrared imagery of urban areas? International Journal of Remote Sensing, 30(9), 2155–2171 (2009).