6th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Comparative analysis of Image classification capabilities of Support Vector Machine (SVM) and Random Forest (RF) with Google Earth Engine (GEE) platform: A case study of Sangamner, Maharashtra

Prasad Balasaheb Wale*[1], Vinit Dhaigude [1], Satyam Mishra[1]

[1]*NIIT University, PhD Scholar, Geographic Information System (GIS) Area, Neemrana, Rajasthan, India*

Keywords
Google Earth Engine
Support Vector Machine
Random Forest
Sentinel MSI

Abstract

Support Vector Machine (SVM) and Random Forest are supervised machine learning algorithms known for their ability to precisely classify complex landscapes on earth surface. These advancements have been very productive for Geographical Information System domain to monitor natural and anthropogenic transformation using remotely sensed datasets. In the present study, Google Earth Engine (GEE) platform has been utilized to identify different land use land cover zones of Sangamner tehsil of Maharashtra. Sentinel MSI satellite images of January 2019 have been accessed and classified over GEE with both SVM and RF classifier. The classification results demonstrate that the SVM classifier performs better than RF over study area with 94.50% and 78.38% overall accuracy. The results obtained from the study illustrate that the major area is utilized for agricultural and urban practices.

1. Introduction

Machine Learning (ML) algorithms play an important role in multiple fields to solve the various complex regression and classification problems (Mariana and Dragut, 2016). In the field of Remote Sensing (RS) and Geographical Information System (GIS), ML algorithms significantly play an crucial role in different studies including Land Use Land Cover (LULC) analysis, Agricultural studies, Precise monitoring and mapping of Natural Resources (Phan Thanh Noi, 2018). Near real time face recognition, tree species identification and urban infrastructure mapping using high resolution drone images are recent applications of machine learning algorithm. Intensive transformation in anthropogenic activities to sustain increasing population with limited geographical area initiated some environmental issues including climate change. Therefore, sustainable planning and management of environment studies have got higher importance (Duraisamy, 2018).

Advancement in computer science, space technology allows us to carry out precise mapping and monitoring of natural and manmade resources at different spatial and temporal scale with greater accuracy (Claudia Maria de Almeida and Liesenberg, Veraldo, 2017). Information from RS satellites are preferable in such studies due to their global coverage and continuous illumination efficiency. The increasing number of Earth Observation (EO) satellites with higher spatial, temporal and spectral resolution revolutionizing scientific studies. (Mathias et al. , 2018).

Image classification is commonly preferred approach in GIS to study dynamics of LULC using RS datasets (Shelestov *et al.*, (n.d.)). There are two primary methods of image classification in GIS including Supervised and Unsupervised (Phan Thanh Noi, 2018). Both methods have their potentials and limitation (Donald and Weih, 2005). Supervised classification approach is further classified in pixel and object based classification approach (Blaschke *et al.*, 2000).

Recent studies of LULC analysis demonstrated that ML based classification approaches including Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Network (ANN) can detect and classify the earth surface features precisely than traditional classification techniques (Claudia Maria de Almeida and Liesenberg, Veraldo, 2017). Potential of RF in classification involves its ability to overcome the problem of overfitting very accurately whereas SVM has ability to precisely detect and separate the classes and avoid the outliers (Xiong *et al.*, (n.d.)). Recently, there have been few studies carried out to compare various supervised classification techniques. However, the accuracy may vary for each case study depending upon quality and quantity of

* **Corresponding Author**

*(prasad.wale20@st.niituniversity.in) ORCID ID 0009-0000-2701-9340
(vinit.dhaigude20@st.niituniversity.in) ORCID ID 0009-0005-4344-0561
(satyam.mishra22@st.niituniversity.in) ORCID ID 0009-0007-8349-0410

Cite this study

Wale, P. B., Dhaigude, V., & Mishra, S. (2023). Comparative analysis of Image classification capabilities of Support Vector Machine (SVM) and Random Forest (RF) with Google Earth Engine (GEE) platform: A case study of Sangamner, Maharashtra. 6th Intercontinental Geoinformation Days (IGD), 6, 113-116, Baku, Azerbaijan

training samples, dataset preferences, classifier selection and landscape complexity (Phan Thanh Noi, 2018) & (Claudia Maria de Almeida and Liesenberg, Veraldo, 2017). The major challenge of machine algorithm involves its higher computational time, high training sample requirements, more complex structure and consume more resources compare to traditional analysis techniques (Nitze, U SCHULTHESS, 2012). The satellite images, comprises of multiple spectral bands with high resolution create another challenge known is "Big Data" problem in the classification (Shelestov *et al.*, (n.d.)).

GEE is a cloud based, planetary scale platform design to access, and analyze the open source geo spatial data of different satellites including Landsat and Sentinel imagery (Xiong *et al.*, (n.d.)) & (Dimosthenis Traganos, Bharat Aggarwal, Dimitris Paursanidis, 2018). The primary objective of the present study is to prepare LULC map of Sangamner tehsil of Ahmednagar district of Maharashtra. Whereas, the Second objective is to compare the SVM and RF technique with GEE platform to evaluate its efficiency in terms of computation and classification task.

2. Study Area

Sangamner is one of the fastest growing tehsils located in Ahmednagar district of Maharashtra. The spatial extend of Sangamner lies between 19.5678 N latitude to 74.2115 E longitude. Pravara river flows from the middle of the city. Pravara river originates in Western Ghats region near Ratanwadi and it merges in to the Godavari river(Veena U Joshi, 2009).

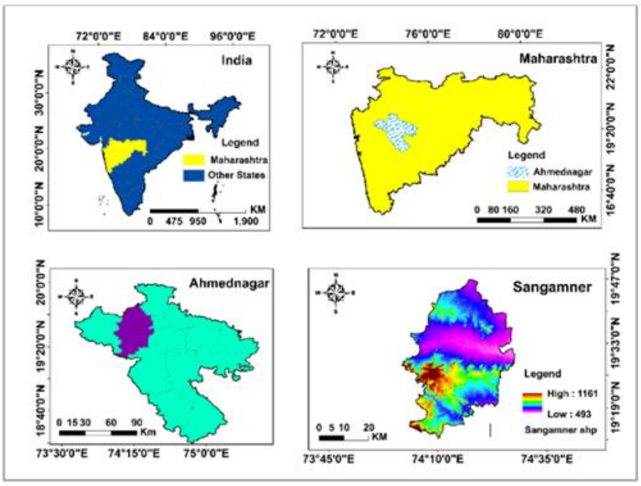

Figure 1. Study Area location with respect to country, state, and district level.

The primary activity of the study area is agriculture because of availability of fertile alluvial soil and constant water supply. Sangamner falls under rainfall shadow zone created by Western ghats, where rainfall ranges around 416mm per annual. Whereas, minimum and maximum temperature lies between 18°c to 32°c. Due to extensive agricultural practices located at the central portion of the study area, has been declared as "Overexploited" groundwater zone by Central groundwater board (Duraisamy, 2018). The construction of Nilwande damn at upper region of the study area, has brought a significant change in the socio-economic activities of the Sangamner.

3. Method

3.1. Dataset

In the present study, one Sentinel-2 MSI image captured in January 2019 has been accessed through GEE. to perform Both RF and SVM classification. Sentinel images have spatial resolution of 10 m in visible and NIR region and other 6 bands are having 20m and 3 bands are having 60 m resolution (Xiong *et al.*, (n.d.)). It Is successfully launched in 2015 and is has been providing precise information of earth surface without any cost for researchers (Son *et al.*, 2017).

3.2. Methodology

The methodology section primarily divided into three section such as accessing the sentinel MSI images of Sangamner and loading the layers on console of GEE. Supervised classification using SVM and RF classifier and analysis of their ability to separate the defined classes. Accuracy assessment and further processing of export layers and classified maps layout preparation has been prepared using ArcPro software.

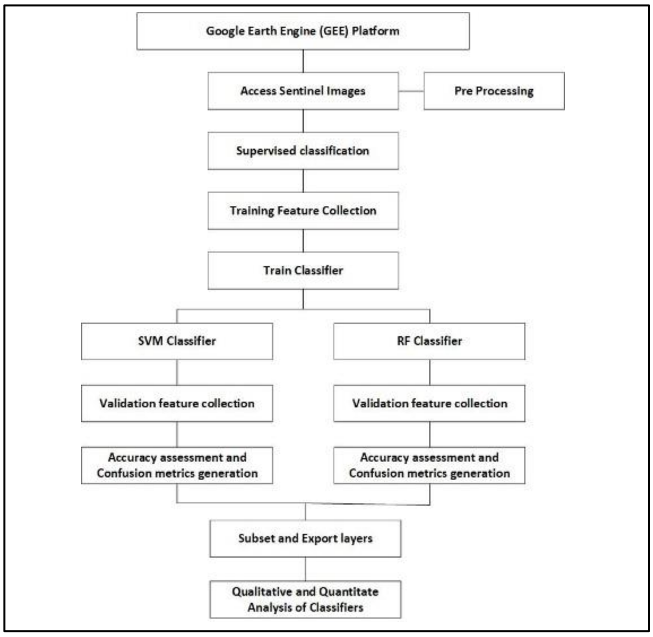

Figure 2. Methodological flow chart showing the major steps which have been followed in present study.

Fundamentally, satellite images are very complex in nature, and it is very difficult to classify with parametric classifiers. Therefore, Non- parametric classifiers e.g., RF, SVM, ANN have been booming and getting more attention from RS discipline.

In the present study, we have utilized SVM and RF classifiers to analyze LULC analysis of the study area. In order to compare both classification techniques, the same training samples and validation sample feature collection dataset have been used in this study.

3.2.1 Random Forest (RF) and Support Vector Machine (SVM) classifiers:

RF classification technique has efficiency to handle the large and complex dataset and overcome the very common issue of overfitting accurately (Raczko and Zagajewski, 2017). RF classification has performed using RandomForest classifier command in GEE. The decision trees are created by making divisions of training samples and decision trees train randomly to avoid overfitting (Kremic and Subasi, 2016). Each decision tree independently classifies the probability of feature to classify in one class among the classes defined by user. The final classification result takes all decision trees probability into account to split features into various classes(Mariana and Dragut, 2016).

$$\{h(x,\theta_k), k = 1,2,3..\ i..\} \quad (1)$$

Where, h is RF classifier, x is input variable or vector, θ_k is independent and randomly distributed sample vectors use to create decision tree(Son *et al.*, 2017). More detailed illustration of random forest classifier mentioned in (Mariana and Dragut, 2016).

SVM is another commonly use classifier introduced by vapnik (Chang-an *et al.*, 2019). SVM classifier can precisely classify the linearly and non- linearly distributed data by hyperplane. Hyperplane is line that can separate the two classes in n-dimensional space distinctly (Kremic and Subasi, 2015). The data points or Vectors which are located near to hyperplane are called as support vectors (Kremic and Subasi, 2016). Support vectors plays an extraordinary role to separate two classes with high margin. Margin is distance between two hyperplane that split the classes from each other. We must choose the hyperplane, which has the highest and equal distance from support vectors. Following figure no 3 gives basic idea of SVM and referred from (Shujun Huang, Nianguang Cai, Pedro Penzuti Pacheco, 2018).

In the linear classification, SVM can accurately classified two classes but in the real world we deal with more complex classes where linear classification may give poor results. To avoid this problem, we use non-linear SVM classification techniques. Nonlinear function uses kernel trick to avoid problem of mix classification and dimensionality (Kremic and Subasi, 2016). Kernels generally transform the low dimensional input space into higher and accurately separable dimensional space. There are four important kernels in non-linear SVM classification these are: Linear, Polynomial, Radial Basis Function (RBF) and sigmoid (Claudia Maria de Almeida and Liesenberg, Veraldo, 2017).

In this study we have used Radial Bias Function (RBF) kernel for SVM classification, because it is the most popular kernel and shows better performance than other kernels (Phan Thanh Noi, 2018). More detailed illustration about SVM classifier and RBF kernel mentioned in (Phan Thanh Noi, 2018) and (Raczko and Zagajewski, 2017).

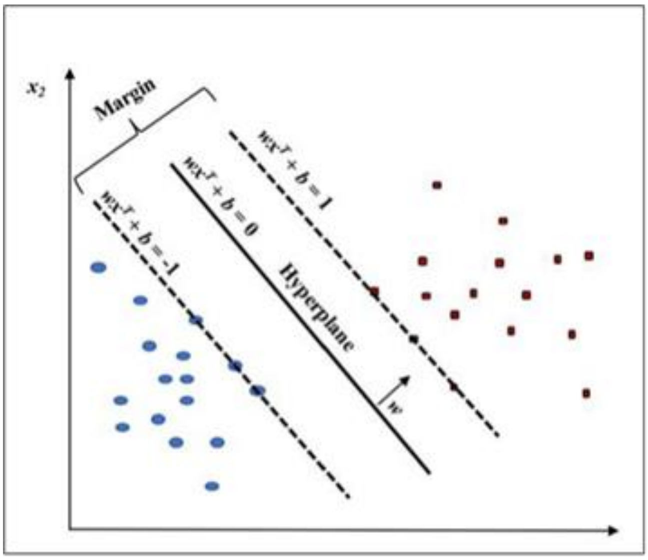

Figure 3. Detailed illustration of SVM algorithm its important components including Hyperplane, Margin and support vectors

4. Results and Discussion

In the present study LULC map of study area is prepared based on sentinel MSI data of January 2019. SVM and RF classification techniques have been utilized with GEE platform to use to classify image of Sangamner tehsil into different classes. The image has been classified into 4 different LULC classes including Forest, Water, Agriculture and Barren Land.

Table 1. Indicates the area (sq km) of Water, Urban, Agriculture and Barren land as per the result of SVM and RF classification analysis

Class	Area sq km (SVM)	Area Sq Km (RF)
Water	14.35	45.22
Urban	55.64	558.85
Agriculture	196.20	147.00
Barren Land	1394.03	909.15

The barren land indicates the most dominant class over study area due to the presence of plateau whereas agriculture class ranks second in terms of land coverage as per SVM classification results. Whereas, RF algorithm indicates urban as second largest land cover with an area of ~550 sq km. Some pixels of barren land got mix-up with urban area, therefore it's an overestimation of an urban area. Some barren land pixels in RF classification got assigned in water class as well. The visual interpretation of classified maps obtained through SVM and RF indicates that SVM classifier performed better than RF classifier to make LULC map of study area.

The central portion of the study area indicates the dominance of the agriculture area due to closeness of Pravara River. Whereas the Sangamner city could be found in the NW side of map as a urban area dominated region with red color. As per the results of RF classification, we can easily identify the issue of mix pixels between urban and barren land class.

We have calculated the overall accuracy for both SVM and RF classifiers which is are 94.50% and 78.38% respectively.

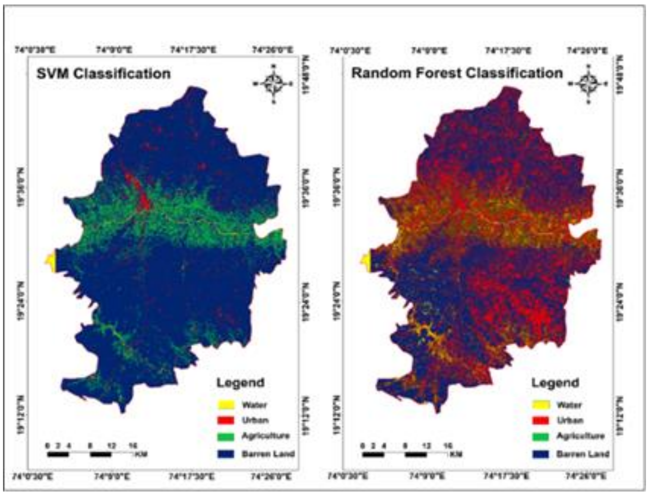

Figure 4. - Land Use Land Cover (LULC) classification of Sangamner using SVM and RF classifiers

In the SVM classification we have not observed the mix pixel problem but in case of RF classification Barren land has mixed with urban areas. The results obtained in the present study validates the same. Moreover, in some cases, RF may perform better than SVM depending on quality and quantity of training samples, dataset variations, and spatial extend of the study area

5. Conclusion

In the present study, we have compared the efficiency of SVM and RF supervised classification techniques to classify freely available sentinel-2 images of ROI. Accuracy assessment carried out for quantitative comparison of both classification techniques. We have also implemented the GEE platform as the most popular and computationally very efficient framework for spatial data analysis. It is very important to study various classification techniques, their workflow, back-end process to precise selection of classifier for various studies to improve the accuracy and quality of studies.

References

Blaschke, T., Lang, S., Lorup, E., Strobl, J., & Zeil, P. (2000). Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. Environmental information for planning, politics and the public, 2(1995), 555-570.

Liu, C. A., Chen, Z. X., Yun, S. H. A. O., Chen, J. S., Hasi, T., & PAN, H. Z. (2019). Research advances of SAR remote sensing for agriculture applications: A review. Journal of integrative agriculture, 18(3), 506-525. https://doi.org/10.1016/S2095-3119(18)62016-7.

Sothe, C., Almeida, C. M. D., Liesenberg, V., & Schimalski, M. B. (2017). Evaluating Sentinel-2 and Landsat-8 data to map sucessional forest stages in a subtropical forest in Southern Brazil. Remote Sensing, 9(8), 838. https://doi.org/10.3390/rs9080838.

Traganos, D., Aggarwal, B., Poursanidis, D., Topouzelis, K., Chrysoulakis, N., & Reinartz, P. (2018). Towards global-scale seagrass mapping and monitoring using Sentinel-2 on Google Earth Engine: The case study of the aegean and ionian seas. Remote Sensing, 10(8), 1227.

Enderle, D. I., & Weih Jr, R. C. (2005). Integrating supervised and unsupervised classification methods to develop a more accurate land cover classification. Journal of the Arkansas Academy of Science, 59(1), 65-73.

Duraisamy, V., Bendapudi, R., & Jadhav, A. (2018). Identifying hotspots in land use land cover change and the drivers in a semi-arid region of India. Environmental monitoring and assessment, 190(9), 535.

Kremic, E., & Subasi, A. (2016). Performance of random forest and SVM in face recognition. The International Arab Journal of Information Technolog, 13(2), 287-293.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Wessel, M., Brandmeier, M., & Tiede, D. (2018). Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. Remote Sensing, 10(9), 1419. https://doi.org/10.3390/rs10091419.

Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil, 79, 3540.

Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors, 18(1), 18. https://doi.org/10.3390/s18010018.

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing, 50(1), 144-154. https://doi.org/10.1080/22797254.2017.1299557.

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017, July). Large scale crop classification using Google earth engine platform. In 2017 IEEE international geoscience and remote sensing symposium (IGARSS) (pp. 3696-3699). IEEE.

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. Cancer genomics & proteomics, 15(1), 41-51. https://doi.org/10.21873/cgp.20063.

Son, N. T., Chen, C. F., Chen, C. R., & Minh, V. Q. (2018). Assessment of Sentinel-1A data for rice crop classification using random forests and support vector machines. Geocarto international, 33(6), 587-601. https://doi.org/10.1080/10106049.2017.1289555.

Joshi, V., & Nagare, V. (2009). Land use change detection along the Pravara river basin in Maharashtra, using remote sensing and GIS techniques. Acta geographica Debrecina Landscape & Environment series, 3(2), 71-86.

Xiong, J., Thenkabail, P. S., Tilton, J. C., Gumma, M. K., Teluguntla, P., Oliphant, A., ... & Gorelick, N. (2017). Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine. Remote Sensing, 9(10), 1065. https://doi.org/10.3390/rs9101065.