

Novel approach of association rule mining for tree canopy assessment

Nilkamal More¹, V. B. Nikam², Biplab Banerjee³

^{1,2}Department of Computer Engineering, Veermata Jijabai Technology Institute, Mumbai, India

³Centre for Studies of Resource Engineering, IIT Bombay, Mumbai, India

Article Info

Article history:

Received Oct 30, 2020

Revised May 19, 2021

Accepted Jun 4, 2021

Keywords:

Google earth engine

High performance computing

PCA

Sentinel 2 dataset

SVM classifier

Tree canopy

ABSTRACT

The evolution of technology and availability of voluminous satellite images are bringing a new scenario in satellite image classification where a performance efficient method for predictive analysis of satellite images for land cover classification needs to be devised. As urban areas are growing at faster rate, special attention needs to be given to solve tree canopy assessment problem. Vegetation indices are calculated from spectral information of satellite images. Hundreds of such vegetation indices are available to detect vegetation from a satellite image. The contribution of this paper is designing an improved Apriori algorithm to select optimal number of vegetation indices for tree canopy assessment. In this research, we propose a novel computational approach that allows the improvement of results. It selects optimal combination of vegetation indices and applies principal component analysis on it. It uses a greedy approach based on Apriori algorithm. This study emphasizes on assessment of tree canopy using GPU-enabled environment for performance-efficient assessment. The results achieved, are comparable to state-of-the-art techniques, with an accuracy of 96%. The research has considered 4 years data for Mumbai city of India. This research is useful for Green India Mission of India to assess tree canopy of urban region.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nilkamal P. More

Department of Information Technology

K.J. Somaiya College of Engineering

Vidyavihar, Mumbai, India

Email: neelkamalsurve@somaiya.edu

1. INTRODUCTION

National mission for a green India is a mission of India to enhance the tree canopy of the city to reduce pollution. It is also called as green India mission (GIM). It aims at highlighting the challenge of environmental change. It is one of the eight Missions illustrated under development activity of India. Tree is an important part of the environment and human life. It is difficult to assess the tree canopy in an urban area like Mumbai because of rapid change and industrial developments. Through the green India mission, the government is targeting to solve problems like deforestation and restoring tree canopy. This research is trying to address the problem of tree canopy assessments with the help of satellite image classification in an urban area. Biological evaluation of an urban tree canopy gives significant data to the urban arranging and the board that can be utilized to ensure and upgrade environmental benefits in the urban tree canopy. The main challenge in this field is the availability of multispectral remote sensed satellite images with high resolution and computing platforms. The prime motivation of this research is to assess the tree canopy which is one of the aims of green India mission. The tree canopy assessments are done here with the help of a high

performance platform on satellite images. Tree canopy assessments is a problem which is a satellite image classification problem. So, the problem is to solve the tree canopy assessments problem using satellite images on a high performance platform to get good accuracy. This research article is planned as: Literature survey for vegetation analysis is available in section 2. It includes a summary table of research papers that have used Sentinel 2 as a dataset for vegetation analysis. Section 3 talks about the mathematical model, dataset, pre-processing of the sentinel-2 dataset, information about study area, the computing platform used and machine learning algorithm. Section 4 explains the results obtained after implementing the classification algorithm LibSVM and results are discussed in it.

2. LITERATURE SURVEY

The tree canopy assessments in an urban area are a challenge as there are networks of roads and buildings. Tree canopy assessments are a satellite image classification approach [1]. It has attracted many researchers as it affects the life of human beings. Satellite images available on the internet can not directly be used to analyse the tree canopy. So, this research involves classification on satellite images to assess tree canopy. The satellite images are explained with its spatial, spectral, and temporal resolution [2]. A typical optical satellite image consists of a number of spectral bands. The spectral vegetation indices are ratio of band information. The vegetation indices are calculated using spectral information of optical multispectral satellite images. Bannari *et al.* [3] forty such indices are available for analysis of vegetation of a geographical area. Tree canopy assessments and analysis involves phenological analysis using vegetation indices [4]. A fusion of images from Sentinel 1 and optical features of satellite images of Sentinel 2 provided a break in vegetation phenology analysis for vegetation management. Stendardi *et al.* [5] and Heckel *et al.* [6] proposed a novel idea of correlation analysis of data from Sentinel 1 (VV and VH) data to phenological vegetation analysis for the South Tyrol area. The support vector machine (SVM) approach [7] is proposed for tree canopy assessments using images of sentinel-2 [3]. It also focuses on the strength of the Sentinel-2 images for the assessments of tree canopy. The geographic area of study was the forest of Knyszyn, and forest Landscape Park in Poland. Wang *et al.* [8] did a study on the difference calculation between two images. The resultant difference image is utilized as the input for supervised classifiers. The classifiers which are used in this paper were SVM, K-nearest neighbour, ensembled methods, and random forest. The results of these classification algorithms are combined using an ensemble based method. Change detection is identified using a voting method which is a weighted method. Spectral vegetation indices calculated from spectral information of satellite images is one of the prominent tools. Table 1 shows the brief information of the techniques which used Sentinel 2 dataset.

Table 1. Summary of different techniques which are using Sentinel 2 dataset

Sr. No	Techniques using Sentinel 2 as a dataset and vegetation analysis	Research papers
1	Classification of the tree canopy or forest with multi-temporal Sentinel 2 data.	[9]-[11]
2	The approach of combining Data from Sentinel 1 and Sentinel 2 is fused for the tree canopy assessments.	[5], [6]
3	Vegetation indices for vegetation analysis with Sentinel 2 dataset	[1], [12]-[14]
4	Supervised object and pixel based classification using Sentinel 2 dataset.	[4], [7]
5	Using texture or spatial information for classification using Sentinel 2 data.	[7]

Normalized difference vegetation index (NDVI) is used for vegetation analysis. Wang *et al.* [14] uses leaf area index (LAI) for vegetation analysis with remotely sensed satellite imagery. The tree canopy analysis is carried out using supervised [15] as well as unsupervised machine learning algorithms [16]. Object based and pixel based techniques are also available under supervised machine learning algorithms. Object based supervised classification is said to be better than pixel based classification method and high accuracy in less computation time is observed. Xue *et al.* discusses on hundred different types of vegetation indices. Every vegetation analysis is associated with one or more applications according to the vegetation of interest and environmental conditions with statistical implementation and precision. Vegetation indices are also applied on hyper spectral images and UAV platforms.

3. PROPOSED METHODOLOGY

The proposed method discussed in this research paper is based on an intelligently selecting optimal number of vegetation indices for the tree canopy assessments. The novelty of the algorithm involves the optimum number of principal component analysis (PCA) components on the image with selected bands and the optimum number of vegetation indices to achieve better accuracy of the classification of Sentinel 2

dataset. Figure 1 shows the steps involved in the process of classification of Sentinel 2 dataset for the tree canopy assessments. Figure 1 describes the steps followed in the process of the tree canopy assessments with the help of Sentinel 2 dataset for Mumbai and suburban regions. The proposed system works in several stages. It acquires satellite images for the Mumbai region for a particular time period. Then the pre-processing is carried out on this data to remove cloud and noise information.

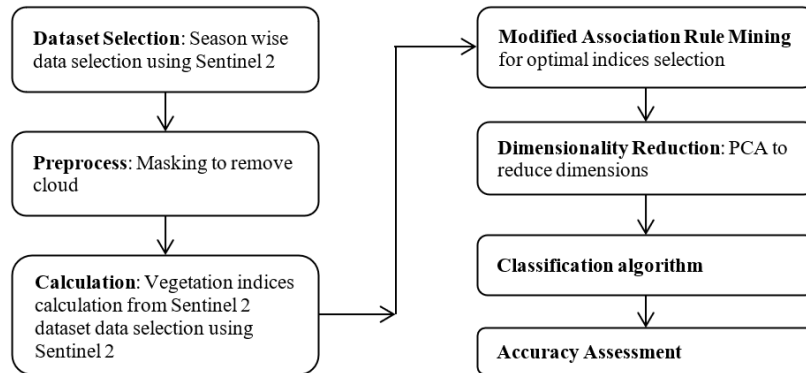


Figure 1. Optimum indices generation approach for high dimension satellite image classification

3.1. Algorithm for optimum indices selection

The intelligent module based on modified approach is used to calculate vegetation indices and select the optimum number of vegetation indices using an intelligent algorithm [17]. It is a greedy approach; it makes optimum selection at each step as it tries to find the effective combination of vegetation indices. Algorithm 1 explains the steps carried out to select optimum number of indices from the set of indices. The algorithm starts with the set of indices and their corresponding accuracy. These accuracies are calculated for individual vegetation indices or taken from literature survey. A minimum threshold is selected by reviewing the literature related to vegetation indices. The algorithm has two stages joining and pruning. Each combination generated at each stage goes through a pruning stage. In pruning stage, the accuracy is compared with minimum threshold. In pruning stage unwanted vegetation indices combination is removed by comparing it with minimum threshold. In joining stage, a new candidate set is generated by joining previous state candidate set with itself.

Algorithm 1: optimal_indices_selection (input image, Set_of_indices)

```

1.  $V_1$  = Vegetation index for Sentinel 2 dataset
2. For iterate  $\geq 2$  AND  $V_{iterate-1} \neq \Phi$ 
3. {
4.    $O_{iterate}$  = Generate_Optimum( $V_{iterate-1}$ )
5.   For all vegetation indices  $V_i \in \text{Set\_of\_indices}$ 
6.      $\{O_i = \text{subset}(O_{iterate}, V_i)\}$ 
7.     For all candidate  $c \in V_i$ 
8.     {
9.        $c.\text{counter} = c.\text{counter} + 1$ 
10.    }
11.  }
12.  $O_{iterate} = \{c \in O_{iterate} \mid c.\text{count} \geq \text{threshold}\}$ 
13.  $\text{Optimum\_no\_of\_indices} = \bigcup_{iterate} O_{iterate}$ 
14. Return( $\text{Optimum\_no\_of\_indices}$ )
}

```

The PCA is used for dimensionality reduction. It is applied to reduce the input given to the classifier. The selection of an appropriate number of coefficients is done to achieve dimensionality reduction. LibSVM is applied on it to get the desired output of classification. The research uses SVM for classification. This is a binary classifier which returns an appropriate class of the pixel of a satellite image. The algorithm used for optimal indices selection for classification of Sentinel 2 dataset for Mumbai region can be explained with the help of a pseudo code as. The intelligent module used in the proposed model works by selecting an appropriate number of vegetation indices from the given set of indices. The function selects indices based on the logic of Apriori algorithm of association rule mining algorithm which uses the threshold (accuracy) to select appropriate number of indices. The algorithm removes the drawback of association rule mining of Apriori algorithm. Apriori algorithm is used to find the frequent itemset [18]. It is easy algorithm for finding

association rules from given set of data items. But it has two major drawbacks: First is multiple scans of dataset and second is too many candidate sets are generated. This research has addressed first issue of Apriori algorithm of association rule mining [19]. Multiple scans of dataset are avoided by gathering enough information from the literature survey about the accuracy of different vegetation indices. Critical literature survey is done to calculate the threshold for this algorithm. So, it is going to avoid multiple scans of dataset. Figure 1 shows the functioning of this algorithm to find the optimum number of vegetation indices. PCA transforms images into a group of bands [20]. A dataset has many features and many of them are correlated. PCA reduces the number of bands in the feature space. Thus, it reduces computational complexity. It takes fifteen bands including vegetation indices bands of an image as an input and produces fifteen bands as an output. This research checks the number of principal components from the output of PCA by measuring the variance ratio of the principal components. This research paper uses a supervised classification [21], SVM classifiers are used for supervised classification. It maps classes to pixels. SVM maximizes separation between classes by using training dataset and annotates pixels by examining their closest class in feature space. This research involves binary classification.

The training dataset is divided into two classes i.e., a tree or non-tree region. Training dataset is created with the help of the Google earth engine (GEE) [22]. This research involves the use of LibSVM which is one of the libraries available for implementing SVM algorithms [23]. The algorithm works in two stages. In the first stage we train the model using a training dataset of satellite images and in the second stage it tests the model for a given set of training dataset of images. The training dataset of this research consists of 2200 geometrical objects. Seventy-five percent of the data is used for training and twenty-five percent of the data is used for testing. We have tried with different values for training and testing phases, like 85% for training and 15% for testing, 80% for training and 20% for testing, 70% for training and 30% for testing, 75% for training and 25% for testing. And it is found that if we use 75% for training and 25 % for testing then the better accuracies were observed. So, we have come up with 75% for the training phase, the rest of them are used for testing. The input to the algorithm is a set of thirteen bands from Sentinel 2 images and output of different vegetation indices and PCA. The optimum number of indices can be selected from these available indices in order to get the better accuracy of vegetation analysis. After experimenting with different types of kernels and gamma values, the research paper uses radial basis function (RBF) kernel for classification and gamma value of 0.5 and cost of 20. The accuracy of the algorithm accessed for accuracy with the help of Kappa coefficients.

Algorithm 2: classification with optimal indices selection (sentinel2 image dataset)

```
{
    Input: Satellite images of Sentinel 2 with 13 bands.
    Output: Classified Image with the object inside it marked with green colour to
    show the trees and yellow colour to show the non-tree areas.
    1. Apply filter on the given images of Sentinel2 to get the desired image with
    cloud percentage less than 5%
    2. To apply pre-processing on the given set of images for a particular period(season
    wise)
    3. aoi= median(images)
    4. //Calculate vegetation index using bands of aoi using Algorithm 1.
    5. Opt_Vegetation_indices= optimal_indices_selection(aoi,set_of_indices)
    6. Bn = select bands from Sentinel 2 satellite image of area of interest useful for
    vegetation
    7. PCA_Coeff= PCA_CAL(Bands(aoi)  $\cup$  Opt_Vegetation_indices)
    8. out=Classify(image,PCA_Coeff(training_dataset_features),LibSVM)
    9. Accuracy_Assessment_conf_matrix(testing_dataset) //Accuracy assessment
    10. Calculate Kappa coefficient using given training and testing dataset }
```

When PCA components are uncorrelated to each other and orthogonal then we can expect to see the tree and non-tree classes as distinct classes. PCA is calculated from the features of the dataset, it does not use the information about classes, so it is an unsupervised technique. The selected components are a complex mixture of original features, so it is difficult to map it to original features. A heat-plot can be made to observe the mapping of original features with the components in the output of PCA.

3.2. Mathematical model for tree canopy assessments

In this research, tree canopy assessments are done. It is actually a problem of satellite image classification. The satellite image is represented using a matrix S , where S is $m \times n$ matrix. For satellite image classification applications, each row of R , the n -vector x_i contains the values at each frequency wavelength of the spectrum sample. Each column, λ_j contains all the observations of one value of an attribute. PCA is used to overcome problems of a large number of dimensions. This process is known as dimensionality reduction. PCA transforms the inputs $\lambda_1, \lambda_2, \dots, \lambda_N$ into another set of column vectors $\mu_1, \mu_2, \dots, \mu_N$. The vector v has

features that the input data's information content is stored in the first few coefficients called as the principal component scores. It tries to reduce the dimensions in the output. New features will be orthogonal to each other. It discards some of the components. The input matrix λ is specified by a $\beta \times \gamma$ matrix. The equation calculates the value of z which is the scaled value with (1).

$$\text{Scaled value}(\Delta) = \frac{(\forall i - \mu)}{\sigma} \quad (1)$$

In this equation of calculation of scaled value, $\forall i$ = The initial value, μ = mean, σ = standard deviation PCA generates new coefficients which are independent of each other. The covariance between any two variables $v1$ and $v2$ is calculated using the following (2) for m number of such components as:

$$\text{cov}(v1, v2) = \frac{1}{\beta - 1} \sum_j^{\beta} (V1_j - \overline{v1})(V2_j - \overline{v2}) \quad (2)$$

Then values of Eigen vector and Eigen values are derived.

In general, the eigenvector of a matrix S is the vector. It holds following relationship.

$$\lambda \bar{v} = \gamma \bar{v} \quad (3)$$

where γ is a scalar value called the eigen value. The linear transformation is defined by a formula given in (4).

$$\bar{v}(\lambda - \gamma I) = 0 \quad (4)$$

The I is the Identity matrix. The next step in PCA is to choose “ η ” number of eigenvectors with the largest value of eigenvalues. Sort the eigenvectors in descending order of eigenvalues. Then “ n ” values from them are selected. The value of “ η ” is the number of dimensions that you wish to have in the derived dataset. It is required to map our data to the generated feature space by re-organizing the data from the original space to the feature space represented by the principal components.

$$\text{Final Changed Data} = \text{Features} * Z^T \quad (5)$$

where Z^T is transpose of Z . This final transformed data is given as an input to satellite image classification algorithm. Satellite image classification can be modelled mathematically using Cellular Automata (CA). CA model considers a vast array of cells having a predefined number of positions which change at distinct time intermissions using certain transition rules. On the same line, the satellite image also has an ample number of pixels with some finite number of classes which transform at discrete intervals of time. So, the model definition of satellite image classification using SVM can be given as. Let us assume a current state of a pixel p of a satellite image, at time t as S_p^t .

$$S_p^t = \langle l_p^t, \tau_p^t, R_{\text{raster}} \rangle \quad (6)$$

In (6), l_p^t is a binary variable indicating satellite image data of particular type or not, τ_p^t is the transition indicator function, it shows whether the pixel under consideration has changed or not in current timestamp. This update rule for transition indicators can be modelled using supervised classification techniques like SVMs.

4. RESULTS AND DISCUSSION

Result analysis section gives information about the Sentinel 2 data set of Mumbai and Navi Mumbai region. Data cleaning using cloud removal is carried out on the dataset. Then information about the geographic location used for study is explained in this section. Google earth engine (GEE) and Google co-laboratory is used in this research as computing platforms. The results obtained in empirical study of this research are discussed in this section.

4.1. Dataset

Sentinel-2 images are collected from two European satellites. It gives access to wide-swath (up to 290 km). There are thirteen spectral bands. The high-resolution images with spatial resolution of 10 m, 20 m and 60 m are available for Sentinel-2. Data is available at 5-day revisit frequency. The Sentinel-2 dataset

with multispectral bands has 13 spectral bands: visible, red, green, blue and near-infrared (NIR) at 10 meters, red edge, and SWIR at 20 meters, and atmospheric bands at 60 meters of spatial resolution. Sentinel2 is useful in many applications like vegetation change detection, water bodies' detection, soil texture analysis in coastal as well as urban areas. The Sentinel-2 images are downloaded from the Google earth engine (GEE) or SciHub. The capability of an instrument to differentiate differences in light intensity and reflectance is called a radiometric resolution of a satellite image. The accurately sensed satellite image can be obtained with the greater radiometric resolution. Bits are used to express the radiometric resolution. Eight to sixteen bits is the typical range of radiometric resolution.

4.2. Data cleaning using cloud removal

A remote sensed satellite image has to be cleaned before applying classification algorithms as noisy images often lead to ambiguous results. This data cleaning approach used here reduces annotation unpredictability and salt-and-pepper noise. In this research paper, cloud masking technique is used prior to the classification process. It uses the metadata available with the satellite image dataset on Google earth engine. Data cleaning involves removal of noise from satellite images of the Sentinel-2 dataset. The algorithm for pre-processing Sentinel 2 imagery explained in algorithm 3.

Algorithm 3: Cloud_Masking (satellite_image_AOI)

```
{
    Input: Image of area of interest of Sentinel 2.
    Output: Cloud masked satellite image.
    1. Band_qa = image.select('QA60');
    2. Cloud RemovalBitMask = 1 << 10;
    3. Cirrus RemovalBitMask = 1 << 11;
    4. m_k=Band_qa.bitwiseAnd(cloudRemovalBitMask).eq(0).and(qa.bitwiseAnd
        (cirrusRemovalBitMask).eq(0));
    5. return (Updated_Data.updateMask(m_k).divide(10000));
}
```

4.3. Compute platform

Google earth engine (GEE) (<https://earthengine.google.com/>) is a GPU/TPU enabled high performance computing based cloud computing platform for geospatial analysis of satellite image datasets. It assigns resources dynamically to cater with the computation intensive tasks. It is available for research oriented, academic, and non-commercial application purposes. It provides an efficient way to handle computationally intensive tasks of advanced image processing. It can be used for development purposes with an uncomplicated online application interface of GEE code editor. It enables users to train, test and develop algorithms interactively. It provides better visualization of results of analysis. Various distributed technologies are also available to process this geospatial data [24].

4.4. Geographic area for study

In this work, Mumbai and Navi Mumbai which are India's mega cities, and the business capital of Maharashtra are considered for study. Mumbai and Navi Mumbai area which is used for study consists of 964 sq. km. Mumbai is located at a latitude and longitude of 19.076090 N and 72.877426 E, respectively. It is on the west coast of India. It is a densely populated city with a population of approximately 12.5 million. As Mumbai is a densely populated and there is lots of industrial area, it is difficult to assess the tree canopy. On this foundation, this research paper chooses Mumbai for research purposes.

4.5. Experimental results

The study area measures approximately 963.78 km², which includes Mumbai and Navi Mumbai area of the city. Figure 2 shows the typical input and output of the system. Government can use this tool to assess the tree canopy of a particular area. So, measures can be carried out to select an area for plantation to reduce pollution. In Figure 2, the left block shows an input image with boundaries marked for the area of interest that is Mumbai and the right block shows a classified image with green colour showing the trees in the area of interest and yellow colour shows deforestation area. The plot in Figure 3 shows fifteen PCA coefficients on the X-axis and percentage of explained variances components on Y axis. After finding eigenvectors, order the eigenvalues in descending order. The eigenvectors give the components in order of their significance. This plot is useful to decide the number of components useful for classification. So, we can see, the first three components are useful for classification. The average thematic accuracy for the given dataset is 91.10% to 96.49%. This is compared with the accuracy of a research paper which uses image fusion approach which also uses GPU-enable environment of GEE [25]. Figure 4 shows the accuracies obtained using confusion matrix method.

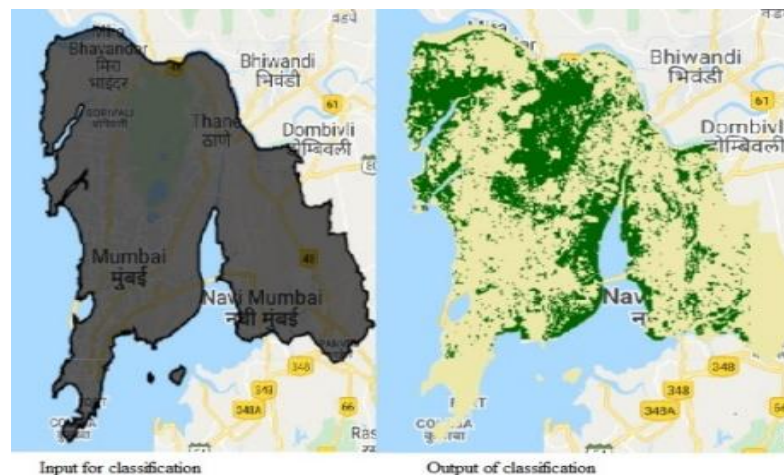


Figure 2. Input and output to the classification

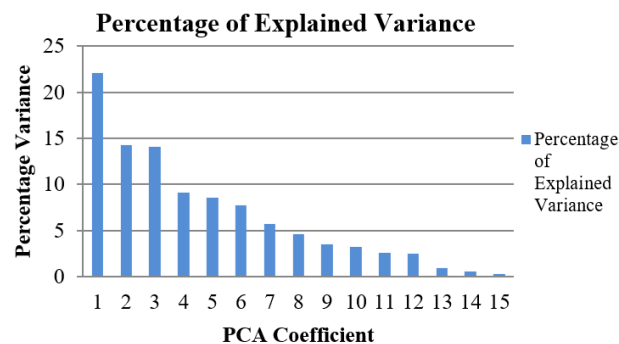


Figure 3. PCA Coefficients vs Percentage of variances

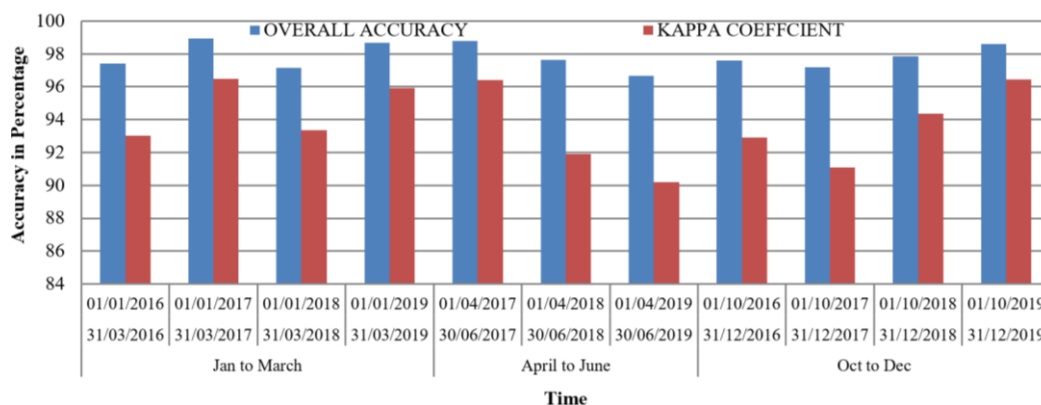


Figure 4. Kappa coefficient and overall accuracy for classification

The season is selected in such a way that it gives efficient vegetation analysis and as per guidelines given on website of GIM [26]. So, from Figure 4, we will come to know that the overall accuracy is ranging from 96.67% to 98.94%. This is observed because of the optimal number of indices selected and PCA. Figure 4 explains the values for Kappa coefficient and overall accuracy for the method of optimal indices selection for tree canopy assessments. The data is selected for three seasons i.e., January to March, April to June and October to December. The analysis is done for 4 years 2016 to 2019. Accuracy plot is obtained by taking seasons on x axis and accuracy on Y axis. Figure 5 shows the results of classification process in assessing tree canopy. So, it is found that combination of all the vegetation indices along with PCA transformation gave more accuracy for classification.

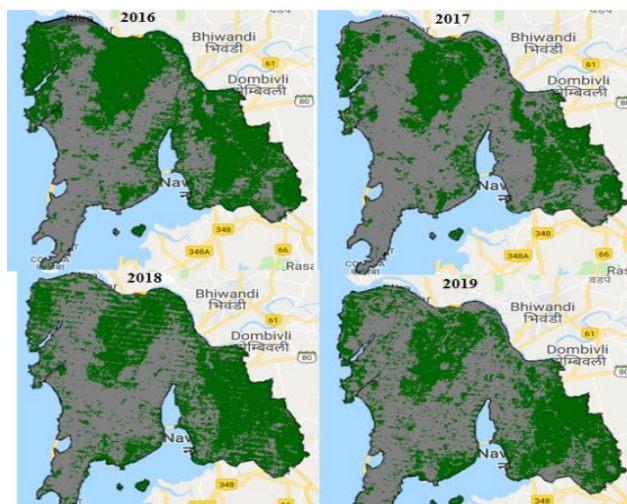


Figure 5. Output of classification for last four years

5. CONCLUSION

This research experimentally proves that tree canopy detection in the Mumbai and Navi Mumbai area based on optimal number of combinations of vegetation indices of high spectral resolution images of sentinel 2 dataset gives more accurate thematic accuracy. The study area of Mumbai is observed from 2016 to 2019. Season wise critical analysis of tree canopy plays an important role as there will be the same environmental conditions available throughout the season. The Season wise analysis gives more accurate overall accuracy. The thematic accuracy observed with kappa coefficient in this technique is 96%. It is more than what is observed in recent literature. They got 90% thematic accuracy. This study can be further extended to produce time efficient tree canopy assessments in the city. More advanced machine learning and advanced deep learning techniques can be used in to further improve the performance of the system. GIM can use this technology to assess tree canopy in the city and decide area for plantation.

ACKNOWLEDGEMENTS

The authors acknowledge the Principal Dr. Shubha Pandit of K.J. Somaiya College of Engineering. We would also like to express our gratitude to the Faculty Development Centre (VJTI and Dr. Babasaheb Ambedkar Technological University), which was founded under the aegis of PMMMNMTT, MHRD, Govt of India, New Delhi, to help us advance this study.

REFERENCES

- [1] D. Dai and W. Yang, "Satellite Image Classification via Two-Layer Sparse Coding with Biased Image Representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 173-176, Jan. 2011, doi: 10.1109/LGRS.2010.2055033.
- [2] M. Gallagher, J. Enterkine, and N. Glenn, *Leveraging Google Earth Engine to couple Landsat and MODIS for detecting phenological changes in semi-arid ecosystems*. Boise State University, ScholarWorks, 2018. [Online]. Available: https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=1051&context=gss_2018
- [3] A. Bannari, D. Morin, F. Bonn, and A. R. Huete, "A review of vegetation indices," *Remote sensing reviews*, vol. 13, no. 1-2, pp. 95-120, 1995, doi: 10.1080/02757259509532298.
- [4] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote sensing of environment*, vol. 204, pp. 509-523, 2018, doi: 10.1016/j.rse.2017.10.005.
- [5] L. Stendardi *et al.*, "Exploiting Time Series of Sentinel-1 and Sentinel-2 Imagery to Detect Meadow Phenology in Mountain Regions," *Remote Sensing*, vol. 11, no. 5, p. 542, 2019, doi: 10.3390/rs11050542.
- [6] K. Heckel, M. Urban, P. Schratz, M. Mahecha, and C. Schmullius, "Predicting Forest Cover in Distinct Ecosystems: The Potential of Multi-Source Sentinel-1 and -2 Data Fusion," *Remote Sensing*, vol. 12, no. 2, pp. 302, 2020, doi: 10.3390/rs12020302.
- [7] X. Wang, S. Liu, P. Du, H. Liang, J. Xia, and Y. Li, "Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning," *Remote Sensing*, vol. 10, no. 2, pp. 276, 2018, doi: 10.3390/rs12020302.

- [8] A. Mironczuk and A. Hościło, "Mapping tree cover with Sentinel-2 data using the Support Vector Machine (SVM)," *Geoinformation Issues*, vol. 9, no. 1(9), pp. 27-38, 2017.
- [9] T. B. Ottosen, G. Petch, M. Hanson, and C. A. Skjøth, "Tree cover mapping based on Sentinel-2 images demonstrate high thematic accuracy in Europe," *International Journal of Applied Earth Observation and Geoinformation*, vol. 84, pp. 101947, 2020, doi: 10.1016/j.jag.2019.101947.
- [10] M. Persson, E. Lindberg, and H. Reese, "Tree Species Classification with Multi-Temporal Sentinel-2 Data," *Remote Sensing*, vol. 10, no. 11, pp. 1794, 2018, doi: 10.3390/rs10111794.
- [11] N. Puletti, F. Chianucci, and C. Castaldi, "Use of Sentinel-2 for forest classification in Mediterranean environments," *Ann. Silv. Res.*, vol. 42, pp. 32-38, 2018, doi: 10.3390/land8040058.
- [12] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *Journal of sensors*, vol. 2017, 2017, doi: 10.1155/2017/1353691.
- [13] Z. Wu and Q. Qin, "Retrieving LAI and LCC Simultaneously from Sentinel-2 Data Using Prosail and PSO-Coupled BI-Lut," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 2895-2897, doi: 10.1109/IGARSS.2018.8517839.
- [14] Z. Wang, Y. Sun, T. Zhang, H. Ren, and Q. Qin, "Optimization of Spectral Indices for the Estimation of Leaf Area Index Based on Sentinel-2 Multispectral Imagery," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 5441-5444, doi: 10.1109/IGARSS.2018.8517747.
- [15] Y. D. Zhang *et al.*, "Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation," *Information Fusion*, vol. 64, pp. 149-187, 2020, doi: 10.1016/j.inffus.2020.07.006.
- [16] R. K. Gupta and S. Bhargava, "A Novel Approach for Cluster Outlier Detection in High Dimensional Data," *International Journal of Science and Research (IJSR)*, vol. 2, no. 7, pp. 131-134, 2013.
- [17] L. Wang, J. Yan, L. Mu, and L. Huang, "Knowledge discovery from remote sensing images: A review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 5, pp. e1371, 2020, doi: 10.1002/widm.1371.
- [18] K. I. Lakhtaria and S. S. Iyer, "Data Analytics Using Satellite Remote Sensing in Healthcare Applications," in *Machine Learning and Data Mining in Aerospace Technology*, Springer, Cham, 2020, pp. 125-146, doi: 10.1007/978-3-030-20212-5_7.
- [19] K. Li *et al.*, "Impact factors analysis on the probability characterized effects of time of use demand response tariffs using association rule mining method," *Energy Conversion and Management*, vol. 197, pp. 111891, 2019, doi: 10.1016/j.enconman.2019.111891.
- [20] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "PCA-based feature reduction for hyperspectral remote sensing image classification," *IETE Technical Review*, pp. 1-21, 2020, doi: 10.1080/02564602.2020.1740615.
- [21] J. M. Górriz, GEE: Google Earth Engine link. Available: https://developers.google.com/earthengine/datasets/catalog/COPERNICUS_S2.
- [22] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines [CP]," Software, [Online]. Available: <http://www.csie.ntu.edu.tw>.
- [23] P. Harrington, *Machine learning in action*. Shelter Island: Manning, 2012. [Online]. Available: http://pzs.dstu.dp.ua/DataMining/bibl/Machine_Learning_in_Action.pdf.
- [24] N. P. More, V. B. Nikam, and S. S. Sen, "Experimental Survey of Geospatial Big Data Platforms," *2018 IEEE 25th International Conference on High Performance Computing Workshops (HiPCW)*, 2018, pp. 137-143, doi: 10.1109/HiPCW.2018.8634070.
- [25] N. More, V. B. Nikam, and B. Banerjee, "Machine learning on high performance computing for urban greenspace change detection: satellite image data fusion approach," *International Journal of Image and Data Fusion*, vol. 11, no. 3, pp. 218-232, 2020, doi: 10.1080/19479832.2020.1749142.
- [26] R. R. Rashmi *et al.*, Green India Mission (GIM) Overview and National Perspective [PowerPoint slides]. [Online]. Available: <http://moef.gov.in/wp-content/uploads/2018/03/Green-India-Mission.pdf>.