

# Land Cover Classification using Machine Learning Techniques - A Survey

Vandana Singh<sup>1</sup>, Vandana Bhattacharjee<sup>2</sup>, C. Jeganathan<sup>3</sup>, Nidhi Bhattacharjee<sup>4</sup>

<sup>1,2</sup>Department of Computer Science & Engineering, BIT Mesra

<sup>3</sup>Department of Remote Sensing, BIT Mesra

<sup>4</sup>School of Computer Engineering, KIIT Bhubaneswar

**Abstract:-** Enormous amount of spatial data is being produced owing to the availability of cloud base technology such as geographic information, satellite imagery, and analysis remote sensing imagery. Land resources must be monitored, evaluated, and managed. Land cover classification based on remote sensing imagery is an important means to enable this. The aim of this research is the review of literature for classification of land cover features using machine learning techniques.

**Keywords:** Land covers classification, Machine learning

## 1. INTRODUCTION

Remote sensing is the process of monitoring and recognizing the physical characteristics of an area by measuring its considered and released radiation at a distance from the targeted area. Machine learning techniques help in remote sensing for classification and analysis of remote sensing data to classify the land cover. Image classification is of two types: supervised and unsupervised classification methods. Statistical and computational intelligence frameworks form the basis of different supervised classification algorithms. The increasing availability of modern technology such as satellite imagery, geographic information, and analysis system and global positioning system are tools that allow the production of reliable spatial data. The use of remote sensing images for land cover classification is an important means to detect, evaluate, and manage land resources.

Landcover classification explains the different land features present on the surface of the earth. Land used for different purposes such as residential area agriculture, industrial area, land cover refers to physical land type such as water bodies, forest, urban area, vegetation. A feasible solution for extracting land cover information is by classifying the image obtained from remote sensing. This review paper presents a survey of research papers applying machine learning and image analysis techniques for land cover classification.

## 2. SURVEY OF LITERATURE

Authors in [1] aim to throw light on the various societal issues of geospatial data like water management, food security, environment protection, drought, disaster, and climate monitoring. These issues are also capable of a very high impact. The authors provide all the knowledge data catalog of the Google earth engine dataset, system architecture, and index of remote sensing. Remote sensing indices especially EVI(Enhanced Vegetation Index) have been used for analyzing how land cover is changed. The

benefit of using the earth engine is that the user can work in a parallel processing environment.

In [2], the author evaluates cloud-based remote sensing platforms, and also how well Google Earth Engine (GEE) works like one. How capable it is of carrying out simultaneous temporal and spatial aggregations over a collection of satellite imagery, has also been explored. The challenges present in Singapore two subarea have been focused on by authors using EVI indices. Changes in landcover and higher computational effort within GEE while carrying out a time series analysis for small land areas, have been evaluated by the author.

In [3], Minget et al. use classification methods for accurate mapping of land cover categories are and Genetic algorithm and Random forest (RF). A genetic algorithm is a heuristic search method applied in artificial intelligence. It became popular through the work of John Holland in the early (1975). Random forest is a very powerful machine-learning classification technique. It has also been used in remote sensing. Random Forest classification has two parameters - many trees and number of variables tried each split. Classification accuracy in a random forest gets affected by these parameters. Here, selection is possible both experimentally as well as heuristically.

In [4], the authors evaluate the Vegetation Indices(EVI, NDVI, NDWI, etc). from the remote sensing-based cloud-based platform and uses effective algorithms for evaluations of vegetation cover. Remote sensing indices have been implemented within remote sensing applications with the help of different satellite datasets. No merged mathematical expression exists, that defines in google earth engine to access all Vegetation Indices. So here the author has reviewed many Vegetation Indices because of the complexity of platforms, different light, and resolutions. In this paper, the author has reviewed more than 100 Vegetation Indices.

Onesimo Mutanga and Lalit Kumar in [5] provide information about the Google Earth Engine application. Storage and processing of a huge number of data sets, for analysis and decision making purposes, is possible due to the cloud-based platform called Google Earth Engine specifically designed for it. Free availability of the Landsat series is provided by it. After storing all the data sets, Google then links them to the cloud-based computing platform for use in an open source manner. Satellite data, Geographic Information Systems (GIS) digital elevation models, demographic and climate data layers, and based vector data sets, are all currently stored in Google. User-

friendly as well as easily accessible front-end is provided by GEE application. It also ensures the presence of a convenient environment for interactive data and algorithm development. Adding and owning data and collections, as well as applying the algorithm for landcover, land change analysis, classification and weather change analysis, all are possible even at users' end.

In [6], the authors explore the uptake as well as usage of GEE platform. This exploration has been done mainly in terms of the datasets used, the broad fields of study, and the geographic location of users. The major question of this manuscript ends up being invaluable, as to provide a platform for planetary-scale geospatial analysis, which is also accessible for everyone, is one of the main goals of GEE. Authorship patterns, geographic scope of analysis, as well as the major area of application, have been assessed using peer-reviewed literature. Even though the applications of GEE with respect to subject matter are immensely diverse, analysis shows that GEE's use is mostly done by developed countries, considering both user nationality (as provided by institutional affiliation), and geographic application. Quite a large number of opportunities for earth observation and geospatial applications are provided by GEE. It may also remove some barriers, specially in the developing world. But all of this needs to be worked on there are some opportunities to improve on this too. Desktop processing machines cannot provide a scale suitable for storage and analysis of remotely sensed data. It can be said that overall this problem has been resolved by the new big data model opened up by GEE. A lot of research is being done to find ways to increase applications in developing countries. Some suggestions are towards a more aggressive intervention approach.

In [7], the a vegetation monitoring framework is developed. It is applicable at a planetary scale and is based on the BACI (Before-After, Control-Impact) design. Google Earth Engine is used in this approach. A web based app named EcoDash is developed. It maps vegetation EVI from Moderate Resolution Imaging Spectroradiometer (MODIS) products during 2000 and 2002. The authors show that the approach is a cost-effective and useful tool for developers. Image classification is an important issue in remote sensing and other applications [8]. To achieve the remote sensing image classification novel supervised algorithm based on the immune network theory is used in this paper. Artificial immune network and antibody network is designed and implemented for classifications of multi hyperspectral remote sensing images. As mentioned by the model, every AB(Antibody) has Trained samples carried to analyze the performance of ABNet using different types of a satellite image. The supervised trained sample provides high classification accuracy

In [9] the author aims to discriminating weed crops using artificial intelligence methods. Radial basis function Neural network approach is used. The limitation of a neural network is that output may sometimes become unstable when applied to a large problem. With a strong mathematical bases one can guarantee better performance of a neural network. In terms of accuracy, this method

gives 75% accuracy which is good as the image analysis technique.

In [10] the author uses mean shift segmentation and backpropagation neural network for enhancing the green vegetation color from the plant. With shadow and presence of blue color, separating green vegetation becomes a very complex task. This paper aims to improve the segmentation rate of the vegetation images containing green vegetation using mean shift segmentation. Operations like opening, closing filling, thickening, etc are used to extract the components present in the image. The backpropagation neural network is then used to classify the vegetation image into two parts: green and non-green. 100 vegetation images were used in the study.

Finally in [11], landcover features have been classified using segmentation technique. MLC(Maximum likelihood), neural network, support vector machine have also been used for landcover classification.

[12] shows that Machine Learning algorithms have been developing surprisingly successfully in the last few years. Fields like speech recognition, search engines and robotics, or generally, data intensive scientific and technical fields, have had a huge benefit due to these developments. One such data intensive application is remote sensing. Providing data over a wide range of the electromagnetic spectrum like VIS, IR, UV, NIR and Radar, happens because of remote sensing today. The sensors can handle single band images as well as multi- and even hyperspectral data, among others. Many times, applications in the remote sensing field end up being monitoring tasks, because of that in the focus of image exploitation, remain long time series data. Many machine learning algorithms, ranging from very basic algorithms like K-Means and PCA to little complicated classification and regression frameworks like decision trees, artificial neural networks, SVMs and Random Forests, have been used for remote sensing purposes since quite a few years. Deep learning methods and convolutional networks (ConvNets) have been in the focus for image exploitation purposes since last few years and are standing on the verge between tremendous success and unbelievable hype. This has been possible through a mix of specialized hardware, algorithmic progress and data availability. How these relatively new approaches can be utilised in remote sensing applications, which questions have been answered and which are still open, all have been explored in this overview.

In [13], publications related to Deep Learning in almost all sub-areas of the remote sensing field are studied. Techniques like image fusion, image registration, scene classification, object detection, LULC classification, segmentation that use deep learning in remote sensing were summarized. This study describes the various fields of remote sensing, along with opportunities and challenges where deep learning is used.

In [14] authors have presented a review of several research works. It is claimed that Support Vector Machines, Random Forests, and boosted Decision Treess have been shown to be very powerful methods for classification of remotely sensed data, and in general, these methods appear to produce overall accuracies that are high compared to

alternative machine classifiers such as single decision tree and k-Nearest neighbour. They further state that the best algorithm for a specific task may be case-specific and may depend on the classes being mapped, the nature of the training data, and the predictor variables provided. Further, the parameters play an important role. Some algorithms have been reported to be robust to parameters settings, such as random forest, and machine learning may still outperform parametric classifiers. Nevertheless, if possible, parameter optimization should be performed to obtain the best classification performance. The sample size and quality of training data also have an impact on classification accuracy. Classification accuracy may also be affected by training data imbalance hence it is important to consider training data imbalances, especially if there is a need to map rare classes with accuracy.

In [15] the author has applied machine learning techniques for land cover classification. The techniques used are Support vector machine, Random Forest, CART, and decision tree-based classification. Image acquisition and Preprocessing of the satellite image have been done before using them for classification. And then classification algorithms have been applied to identify the Land Cover Feature and the difference of accuracy in these three classification algorithms is studied. Comparative study has been conducted by applying the vegetation indices along with machine learning techniques. Upon application of these techniques on a specific geographical area, it is seen that the combined approach gives more accuracy. Google earth engine platform is used for the classification of the Landcover feature.

This section presented a literature survey of papers in the domain of analysing remote sensing images for various classification purposes.

### 3. CONCLUSION

Application of Machine learning techniques in analyzing remote sensing data that are summarised in this paper have a big scope in future. In addition, for more accuracy, we need to re-frame the satellite images in a bid to secure uniqueness in colors as found in images and to manage different colors as are shown in objects of the same class structure. One may conclude that generally, machine-learning methods, and especially SVM, RF, and boosted DTs, have been shown to be more robust to large or complex feature spaces in comparison to parametric methods. As part of our ongoing work, we aim to build models for land cover classification using various machine learning techniques.

### 4. ACKNOWLEDGEMENTS

The authors are grateful for the financial support provided by the SKYMAP GLOBAL (SMG) Pvt. Ltd. Singapore, under the "SMG-BIT Agri Research program" in carrying out this research.

### 5. REFERENCES

- [1] Noel Gore lick (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone
- [2] Nanki Sidhu(2018)Using Google Earth Engine to detect land cover change: Singapore as a use case
- [3] Minget al. (2016): Land covers classification using random forest with genetic algorithm-based parameter optimization
- [4] Jinru\_Xue and Baofeng Su, Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications, Journal of Sensors Volume 2017, Article ID 1353691, 17 pages
- [5] Onesimo Mutanga and Lalit Kumar(March 2019), Google earth engine application, MDPI Publications, 2019
- [6] Lalit Kumar and Onisiono Mutanga, Google Earth Engine Application since Inception usage, Trends and potential
- [7] Ate Poortinga, Nicholas and David Saah, An Operation Before – After control Impact(BACI) Designed platform for vegetation monitoring at planetary scale
- [8] Yanfei Zhong, An adaptive artificial immune network for Supervised classification of multi hyperspectral remote sensing imagery
- [9] M. J Aitkenhead, Weed and crop discrimination using image analysis and artificial intelligence method, Computers and Electronics in Agriculture 39 (2003) 157- 171
- [10] Liying Zheng, Mean Shift based color segmentation of image containing green vegetation, Computers and Electronics in Agriculture 65 (2009) 93–98
- [11] Xiangrong Zhang, Joint Segmentation and classification of Hyperspectral Image using mean shift and Sparse Representation classifier
- [12] Karsten Schulz, Ronny Hänsch, and Uwe Sörgel "Machine learning methods for remote sensing applications: an overview", Proc. SPIE 10790, Earth Resources and Environmental Remote Sensing/GIS Applications IX, 1079002 (9 October 2018); <https://doi.org/10.1117/12.2503653>
- [13] LeiMa, YuLiu, XuLiangZhang, YuanxinYe, GaofeiYin, Brian AlanJohnso, Deep learning in remote sensing applications: A meta-analysis and review, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 152, June 2019, Pages 166-177
- [14] Aaron E. Maxwell, Timothy A. Warner & Fang Fang (2018) Implementation of machine-learning classification in remote sensing: an applied review, International Journal of Remote Sensing, 39:9, 2784-2817, DOI: 10.1080/01431161.2018.1433343
- [15] Vandana Singh, Land Cover Classification Using Machine Learning Technology, M.Tech Thesis, Birla Institute of Technology, Mesra, 2020