**Remote vegetation analysis using machine learning**  
 *Shivani Kamerkar, Prital Ghanvat, Komal Ingale, Harshika ghadi and Sneha Mane R.A.I.T .DEPARTMENT OF EXTC ENGG , INDIA*  
  
 **Abstarct.** Our ecosystem is a geographic area where animals, plants, and organism form a life. This life mainly depend on our environment and vegetation is the most important part of ecosystem as it is mainly responsible providing oxygen and gets in carbon dioxide using photosynthesis. This information about the vegetation is very critical as global warming affect the environment. As human are mainly facing environmental catastrophic so prevent this, we can remotely sense and understand geographic of vegetation by using satellite. Our aim is to classify vegetation into three group i.e. Forest, crop and Barren Land. For this classification we use some machine learning tools and models to predict the vegetation by using satellite map and understand geographic of particular area. We build a web application that allow us to understand geographic and vegetation of particular land field. This includes crop, tree, road, barren land, river and buildings. Our main goal is to understand what kind of vegetation is used to yield in particular location and we can maintain a record and it can be used for future purpose.

**1 .Introduction**

Vegetation is one of the most crucial part of an environment and habitat. As is important for take in carbon dioxide and supply oxygen during the process of photosynthesis providing a suitable environment and atmosphere for the organism and living being to breath. Using Remote sensing is important as decrease human effort for survey and under the future effect of degradation of vegetation. Ther term vegetation is referred as the group of plant species or flora composition that cover ground or land in a particular area. Vegetation handle important functions in the biosphere, geological and spatial scales. Vegetation is distributed into five major types which include tundra, permafrost, desert, forest, ice sheet or snow and grassland. It is the play most important and resourceful factor of the ecosystem. Vegetation also plays important role in source of economy for the nation as it is responsible for providing medicines, paper, wood, food etc. The flow of natural element and bio-geochemical cycles, such as nitrogen oxygen, carbon-dioxide and water which is regulated by vegetation present in that location and due to which it is important part for the environment and ecosystem. It is mainly responsible for the characteristics of soil which includes density and volume, moisture contain, texture and chemistry of the soil. Vegetation provide shelter for many animal species and gives habitat for living organism. Due to deforestation and other human interference cause decrement in production of vegetation which cause another serious issue known as global warming. The information about natural resources and land that covers the vegetation is crucial and important. Due to advancement in the field of science and geographical research it becomes easy to get this information via satellite or other scientific means. Remote sensing is scientific technology or a tool which is used to analyze and collect information through satellite images or other scientific mean and classify data using mathematic and scientific algorithm. Remote sensing is used in many applications which include glaciers monitoring (It is used to track glaciers activity and sea level), oil spillage monitoring(Track accident of marine ship to minimize aquatic losses), land cover detection(Area cover by land of particular region, vegetation classification(Classify types of vegetation which include forest, crop etc.), deforestation monitoring(Process for monitoring forest), Earth quake detection(natural calamities and details of its aftermaths etc. Our Objectives is to maintain and get a record for vegetation and to develop an autonomous system, An automation system is an integration to sense,

controls and to perform a function with minimal or no human intervention.

**1.2 Materials and methods**

The aim concerning this study is the forecast of map feature parts using Machine Learning and Computer Vision. Our objective is a fetch a data of vegetation present in particular location of map. This mainly include classification of tree, land , crops and water bodies.

**1.3 Motivation**

Project is based on extract on data from satellite and use for research which will help to develop better environment. Supervised classification methods is used to learn patterns in data and relationship between the samples from template which measures proportionately. This measurement and others are deliberate, using specifications that anticipate our project cases which is used to classify an unknown data into MLC classifier which is widely used for satellite images also helps in statistically distribute data into pattern. Support vector machine is used to compare results and likelihood and spectral angle mapper which is obtained by different other algorithms. Using machine learning help us to improve efficiency and get results faster.

**2 Methodology**

**2.1 Data Processing**

Data processing is the technique in which data is translated, collected and manipulate large amounts of data into usable information. Data processing should be done correctly as it should not negatively affect data output or the end result. Hence, in this section firstly we add the certain co-ordinates to perform the analyzation process. After analyzing the coordinates as the input section then this data is fetched from satellite using API. API referred as application program interface which a software or piece of program that allows two program or applications to communicate with each other.. As after this image validation process is undertaken where through API the images of the particular co-ordinates are carried with respect to given longitude and latitude.

**2.2 Feature selection and Image classification**

RGB is referred as Red Green Blue, it is the primary colors in additive color synthesis model. A RGB is a image color spectrum which consists in composite layers of Red, Green and Blue, each being coded on 256 levels from 0 to 255. It corresponds to the levels or some threshold of three primary example, black corresponds color R=0, G=0, B=0, and white corresponds to the levels R=255, G=255, B=255. On the other hand, texture analysis can be defined as characterization of regions or a function of spatial variation of an image by their texture content and brightness intensity of the pixels. Texture analysis try to quantify innate qualities described by terms such as silky, bumpy, smooth, or rough. Random forest classifier can be selected for image classification and this algorithm is used in many decisions trees. This algorithm is use some feature randomness and try to create an uncorrelated forest of trees for prediction which is more accurate than that of any individual tree. Random forest is a Supervised Machine Learning Algorithm that is used for regression and classification from RGB and texture analysis which helps to covert image into HSV.

**2.3 Accuracy analysis**

Accuracy is the unit or a measurement used for analyzing model (model analyzing), identifying relationships (Decision tree) and patterns (multi-dimensional array) between variables in a dataset based on the training of raw or input data. The model can generalize data which deliver more accurate value rather than predictions and insights which can produce value based on accuracy of dataset. Scaling Analysis is a method to standardize the indiscriminate features present in the data in a fixed range. If feature scaling is not done, then a machine learning algorithm is used to determine gain of greater values, higher and consider smaller values as the lower values, regardless whatever the unit of the values.

**2.4 Evaluation of data**

Evaluation becomes highly import when it comes to validating the type of model chosen for the dataset. Evaluation helps to check whether the goals set to achieve were met or not. If the results are not satisfactory then again the above mentioned steps need to be reperformed so that the root cause behind models underperformance can be recognized and subsequently fixed

**2.4 Computer Vision**

Open CV is a library used to detect color present in the map is it is used for conversion of image from RGB to HSV and classify the image to different dataset used for analyzing location .

**3 ALGORITHMS and Library**

**3.1 Kmean**

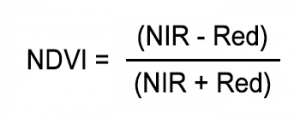
K-means unsupervised learning algorithm is used for vector quantization and processing which used to partition k number of clusters from n number of observations and each cluster centroid which serving as a prototype of the cluster or observation belongs to the cluster with the nearest mean.. *k*-means clustering shortens or minimizes within-cluster variances an mean optimizes the squared errors,. Kmean solve problem by using computationally difficult (NP-hard) and efficient heuristic algorithms which converge quickly to a local optimum .K-means clustering is a technique which is used to find clusters of comparable spatial data to some extent. k-nearest neighbor classifier has a loose relationship to the unsupervised k-means algorithm. K-means classifies data into clusters by applying 1-nearest neighbor classifier to the cluster centers.

**3.2 Numpy**

NumPy is the python library or package which mainly used for scientific computing and multidimensional array. NumPy is used in a multidimensional array object it is also used in an assortment of routines for quick operation of data this package is versatile as we can use in selection, shape manipulation sorting, mathematical function and also it can be used in masked array and matrices. This package also used for engineering mathematical problem such as Fourier transforms and other linear algebra and statistical operations. NumPy library uses ndarray which is convert from object or model into ndarray also encapsulates ndimensional arrays.. In our project we used for analysing multidimensional array and convert it into ndarray which is used for getting PCA of data of map used.

**3.3 Normalized Difference Vegetation Index**

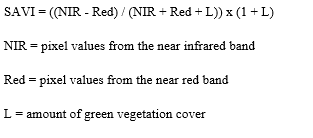
**Normalized Difference Vegetation Index (NDVI) is**  use find the NIR and red channels in image formula is given below.



Healthy vegetation (chlorophyll) show green color reflects more near-infrared (NIR) pixels and green light compared to other wavelengths. But it also absorbs more red and blue light. Due to this reason our eyes see vegetation as the **color green**. By analysing near-infrared green clor, then it could analyse for vegetation too. Satellite sensors has **Landsat** and **Sentinel-2** both have the necessary bands with NIR and red.

**3.4 Soil-Adjusted Vegetation Index (SAVI)**

The Soil-Adjusted Vegetation Index (SAVI) is a vegetation index is a index which analyze soil brightness by using infrared band and soil-brightness color correction factor. It is mainly used to analyze regions where vegetation is low.

 NIR and Red is referred to the bands associated with those wavelengths present in the pixel of map .

The value of L varies depending on the amount of green vegetative cover:-

L=1 (Areas in with no green vegetation cover)

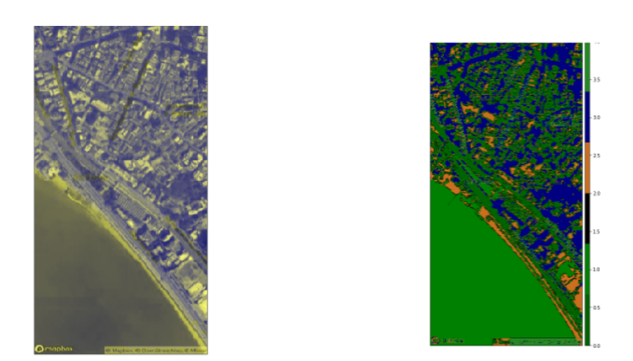
L=0.5 (Areas of moderate green vegetative cover)

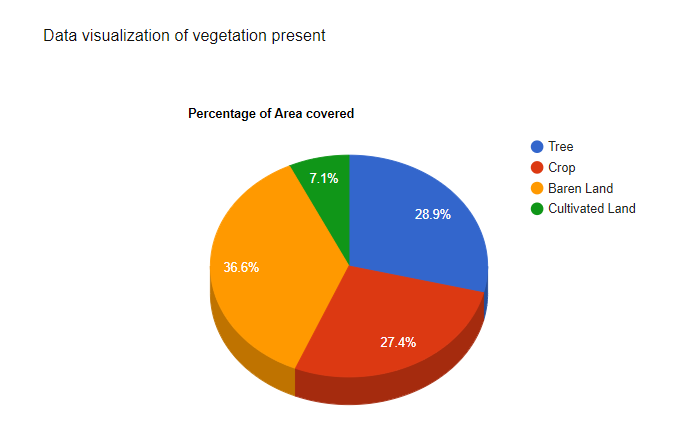
L=0 (Areas with very high vegetation cover ie equivalent )

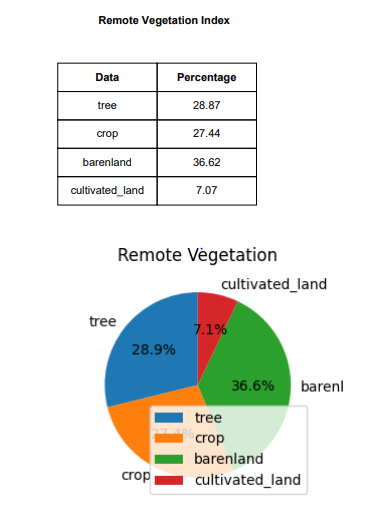
This index of L outputs values ranges between -1.0 and 1.0.

**4 Results**

Result The outcome of the project is to show the images of the map of the given input respectively. The input is considered as the particular location of the given country.In the backend the function of using api is taken into consideration for longitude and lattitude.Hence function of using api for getting status of api and further sorting of map image in specific location is performed.







**5.Conclusion**

Supervised classification and unsupervised classification is used to find the geographic of the region which include road, water, forest, crop, baren land and mountain. In this project, an image obtained from satellite through API is classified into different classes based on some features by

using machine learning algorithms and computer vision. These include Computer Vision, Support Vector machine algorithms and K-means algorithm.

Our goal is to analyze data into section as by using Machine learning .This algorithm is used to find NDVI and SAVI of the region

which help us to identity te geography of that area. Our goal is to show the existence of a particular crop, trees, water bodies or things like roads and buildings on the field in the given area.

When we analyse on the satellite image using this Computer Vision, we create a individual model of data and it categorizes the map image into sub images and give total area for each category. These category include crop, tree, water bodies and barren land. Our goal is create a model for faster analyse of survey of large are in a short time. The categorized area after applying the algorithms was compared

and create a report visualization