# 1. INTRODUCTION

Mangroves ecosystem are highly productive but are extremely fragile and sensitive. Mangrove forests are undergoing constant seasonal, short term, long-term changes due to their dynamic nature and to a greater extent, through various natural and biotic influences. Mangrove forest not only protect coastal organism but also protect coast from soil erosion and serve as breeding, feeding and nursery ground for estuarine and marine organisms. Therefore monitoring mangrove habitat from time to time is an essential component in a coastal marine ecological studies and coastal management.

## 1.1 Remote Sensing:

The term remote sensing can be broadly defined as obtaining information of the areas on the earth using sensors on satellites or the aircraft. These sensors collect data in the form of images and provide specialized capabilities for manipulating, analyzing, and visualizing those images. Remote sensors collect data by detecting the energy that is reflected from Earth.

Over the last few decades remote sensing technology has been used increasingly by the scientific community to describe and monitor a variety of systems on a global scale. This technology has evolved from pure visual imagery i.e. panchromatic aerial photographs to multi-spectral imagery i.e. Thematic Map. Remote sensing requires an energy source that provides electromagnetic energy. This energy travels from the source to the target where it comes in contact with the atmosphere. Once the energy reaches the target, it interacts with the target. A sensor will detect and record the energy emitted by the target. The energy from the sensor is passed to the processing station in the electronic form. The processed image is then interpreted visually or digitally. This data can be further used to reveal some more information about the target and can assist in solving a particular problem.

### 1.2 Types of Remote Sensing:

There are two types of remote sensing i.e. passive remote sensing and active remote sensing. Passive sensors detect natural energy that is reflected or emitted from the observed object. Passive sensors sense only radiation emitted by the object being viewed or reflected by the object from a source other than the sensor. Reflected sunlight is the most common external source of radiation sensed by passive sensors. Active sensors provide their own energy to illustrate the object. They send a pulse of energy from the sensor to the object and then receive the radiation that is reflected back from that object. The classification algorithm such as regression, artificial neural network and decision tree can be used to classify the image. Classification technique help to find out change in land based on land cover such as forest, development, water and mangroves etc.

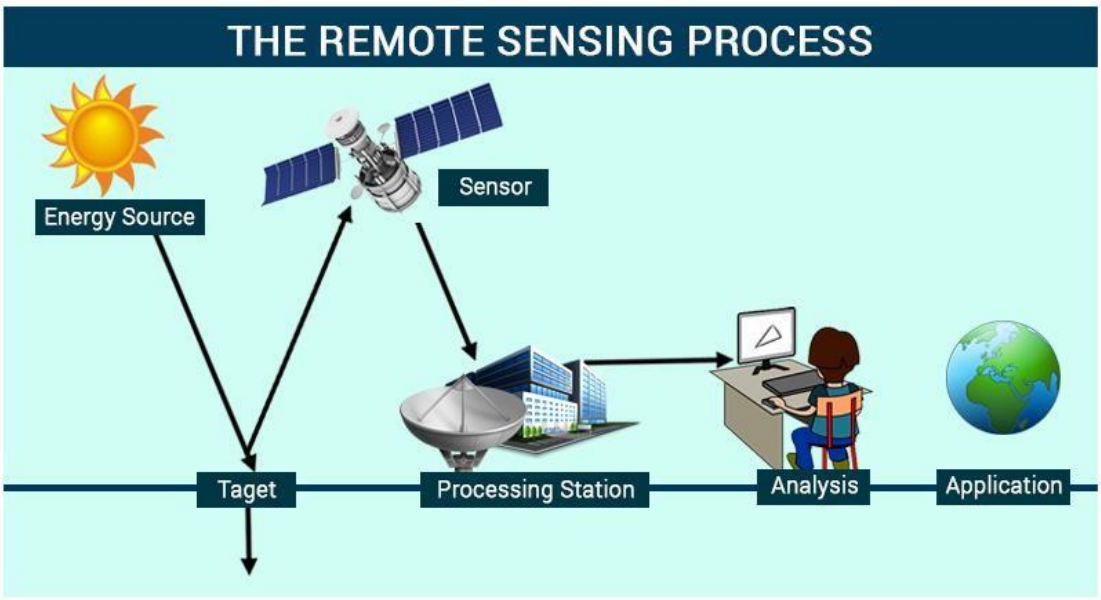


Figure 1.Remote Sensing Process (Courtesy: https://byjus.com/physics/remote-sensing)

### 1.3 Types of Sensors in IRS-P6

#### 1.3.1 High-Resolution Linear Imaging Self-Scanner (LISS-IV)

The LISS-IV is the highest-resolution sensor offering a GSD of 5.8m at nadir. The Quantization is 10 bits (selected 7 bits transmitted). Ground swath is 23 or 70 km with 5 days repeat cycle. It operates in three spectral VNIR bands. A single telescope & lens assembly is used for all bands. Band 3 (red) is placed closest to nadir, while band 2 looks ahead and band 4 looks behind the satellite velocity vector. The 12000 pixel CCD array for each band is separated into odd & even pixel, arranged in two rows with a distance of 35 microns (5 scan lines) between them.

#### 1.3.2 Medium Resolution Linear Imaging Self-Scanner (LISS-III)

The LISS-III is a medium resolution sensor offering a GSD of 23.5m. The Quantization is 7 bits (SWIR band 10 bits – selected 7 transmitted). The Ground swath is 141 km with 24-day repeat cycle. It operates in four spectral bands - three VNIR one SWIR. Each band consists of a separate lens assembly & linear array CCD. The VNIR bands use a 6000 element CCD with pixel size 10x7 microns. The SWIR band uses a 6000 element CCD with pixel size 13x13 microns. The data from the VNIR bands are digitized to 7 bits while the data from SWIR band are digitized to 10 bit.


Figure 2.LISS-III sensors (Courtesy:https://directory.eoportal.org/web/eoportal/satellite-missions/i/irs-p6)

### 1.3.3 Advanced Wide Field Sensor (AWiFS)

The AWiFS with twin cameras is a moderate-resolution sensor offering a GSD of 56m at nadir. The Quantization is 10 bits and the combined ground swath is 740km with five-day repeat cycle. It operates in four spectral bands – three VNIR one SWIR. The AWiFS camera is split into two separate electro-optic modules (AWiFS-A and AWiFS-B) tilted by 11.94 degrees with respect to nadir. The instrument used is push broom, the bands (4): 0.52-0.59, 0.62-0.68, 0.77-0.86, and 1.55-1.70 μm. The AWiFS spatial resolution is 56 m near nadir and 70 m near edge. The radiometric resolution is 10 bit and swath is 740 km.

## 1.4 Geographical Information System

GIS stands for Geographical Information System. GIS is a framework which is used to analyze, create, visualize and find trends in geographical data. It is an integrated system that capture, store, create and visualize geospatial data. GIS is used to create an interactive map from the classified image. The Classified image helps to classify the areas in Mumbai like mangroves, forest, water and development area. These characteristics are expressed by different colors like red, green, blue and yellow respectively. The interactive map will contain the information of changes in mangroves in Mumbai region.

# 2. OBJECTIVE

A Mangrove is a tree or a shrub which grows above mean sea level in the marine coastal environments and estuarine margins. Mangroves are buffers between the land and the sea. The main objective of the project is to classify the land cover in Mumbai as mangroves, forested area, water and development area using classification algorithms. Using the previous researchers, we obtained the information of mangroves in Mumbai. The land cover in Mumbai was classified into 4 classes i.e. Mangroves, forest, water and development and were represented as red, green, blue and yellow respectively.

The mangrove regions in Mumbai were represented as healthy or unhealthy based on the previous researchers. The healthy mangrove area will indicate that the mangroves in that area are healthy and are less prone to degradation. The unhealthy mangrove area will indicate that the mangroves in that area are degrading and there is a need to protect them. This will help to find out that these areas are in threat and these areas need to be taken care of else these mangroves may deplete in near future. The classification of areas in Mumbai will be done using different data mining algorithms like artificial neural network, k-nearest neighbor and decision tree algorithm. These algorithms will be individually used to classify the area in Mumbai. The classification will result in a classified map which will classify the areas based on the colors assigned to them. After the classification is done, the accuracy and the corresponding kappa value will be calculated. The accuracy and kappa value of these algorithms will be compared with each other. QGIS software will be used to display the data of those mangroves on the classified image. Pop-ups will be used to show healthy and unhealthy areas. The final result of the project will be shown on the web by using the QGIS2Web plug-in of QGIS software.

## 2.1 RESEARCH PROBLEMS

The research is generated because of different problems which are available in current system. The problems are listed below:

### 2.1.1 MANUAL SURVEY

The regulating body which are working for the coastal, wetland or mangroves have to do the manual survey of the area and based on that they have to suggest the condition and the health of mangroves which is a very difficult job.

### 2.1.2 BIASING

If the regulating body does the manual survey, there are possibilities that the data they may submit is not a legitimate one. There are possibilities that humans may make changes in the data and submit a wrong report that may lead to a wrong interpretation of the survey.

### 2.1.3 NO FREQUENT UPDATION

It is very difficult for the regulating to go and visit different places for their surveys and keep an update of the changes in the mangroves that they already visited or surveyed. To update the survey data for a region, it may take a long time.

### 2.1.4 TRACK THE MANGROVES

Once a survey is done and the reports are submitted then there may be a possibility that for a long period of time the region isn't inspected and if in that time any illegal practices or operations against the mangroves is done by humans then that cannot be tracked till the survey for the region is done once again. So keeping a track of the mangroves after the survey is very difficult

## 2.2 RESEARCH SOLUTION

In manual survey, the above-mentioned problems occur. In order to overcome these problems, an automated system for tracking and inspecting the mangroves is adopted. This technique makes use of Remote Sensing images and GIS to compare the change in the mangroves for a specific region for a particular period of time. Some of the advantages of this system are:

### 2.2.1 LESS TIME REQUIRED

Using Remote Sensing and GIS for tracking the mangroves takes less time as compared to a manual survey. There is no need of visiting the place of inspection again and again. We just need to use algorithms to compare different images over the time.

### 2.2.2 ACCURATE RESULTS

Since this method is automated, there is no possibility of any biasing of the data. Machines provide accurate data without any biased content.

### 2.2.3 FREQUENT UPDATION

As in a manual survey, there is no need of surveying a region again to track the change in the status of mangroves. In this automated system, we just need to apply various algorithms on the variety of images obtained from the satellite of a specific region and generate the reports. It is easier to find changes and get updated data for the mangroves.

### 2.2.4 TRACK ILLEGAL PRACTICES

Due to frequent updates of the data, the illegal practices or operations on the mangroves can be easily identified and it helps to take action on such illegal practices as soon as possible because in manual survey frequent updating is not possible.

# 3. LITERATURE REVIEW

In the year 1999, S. Agatonovic-Kustrin and R. Beresford explained the basic concept of artificial neural network (ANN) with all its applications [1]. In the year 2015, Roya Abedi and Amir Eslam Bonyad had successfully implemented K-nearest neighbor algorithm using IRS-P6 LISS-III satellite image. They got the accuracy between 80% and 93.94% [2]. In the year 2010, M K Ghose, Ratika Pradhan and Sucheta Sushan Ghose had implemented Decision Tree Classification algorithm using spectral separability matrix using the IRS LISS-III multispectral images and they got the 98% accuracy using decision tree and 95% using the maximum likelihood classifier [3]. Sahu, Suresh, Murthy and Ravindranath had done the mangrove area assessment in India in which their main focus was to show the implications of the loss of mangroves. They assessed the trends of mangroves area in India during 1987-2013 [4]. In the year 2014, Prabakaran chellamani, Chandra prakash singh and Cushma panigrahy had done health assessment of Indian mangroves using multi temporal remote sensing data. They found that around 38% and 27% of total mangroves in India belong to very-healthy and healthy categories respectively. Mr. Anand Upadhyay, Dr.Santosh kumar Singh, Mr. Aditya Shetty and Mr.Zibreel Siddiqui successfully implemented K-nearest neighbor, artificial neural network and Decision tree algorithms using the IRP- P6 LISS-III satellite image. They got the accuracy 97.04%, 95.31% and 99.82% respectively with the kappa value as 0.84, 0.95 and 0.96 respectively [6]. T G Jagtap, A G Untawale and S N Inamdar used remote sensing technique to study the areas of mangroves in Maharashtra. [7]. Kai Liu, Xia Li, Xun Shi and Shugong Wang used Decision tree with remote and GIS to monitor the changes in the mangrove forest [8].

# 4. STUDY AREA AND CHARACTERISTICS

Mumbai is the land of opportunities. People from all over India come to Mumbai in search of job opportunities. This is the main reason for the growing population of Mumbai. The population of Mumbai was around 18.41 million in 2011. Since then there was a speedy development and population pressure rapidly increased and Mumbai being the coastal area, it took the toll of mangrove land. Major mangroves are seen today in Mumbai along the Vasai Creek, Thane Creek, Manori and Malad, Mahim, Bandra, Versova, Siwari, Mumbai and few more places. Millions of citizens in Mumbai do not treat mangroves in the right way; they simply think that it's a waste material. Only little people around actually understand how important mangroves are to the quality of lifestyle. In the early 19's, over 37 sq. km. of mangroves existed in Mumbai, largely in the Thane creek, Mahim, Versova, Gorai and Ghodbunder, with regular patches in places such as Bandra, Malabar Hill and Colaba. Mumbai has probably lost 40% of all its mangroves in the past decade. Around 20 out of the 35 species of true mangroves found in India have been identified along the Maharashtra coast and 15 species of these are found in Mumbai. Our prime area of study in Mumbai region is Vikhroli, Gorai, Airoli, Mulund, Ghansoli, Charkop, Vashi and Versova.

## 4.1 LISS-III SENSOR

LISS-III camera provides multispectral data in 4 bands. LISS-III operates in three spectral bands in visible (VIS), near-infrared (NIR) and Soft Wavelength infrared (SWIR) with 24-meter spatial resolution and a space of 141 km. The fourth band has a spatial resolution of 70.5m with a ground area of 148 km. The repetitively of LISS is 24 days.

**Table 1**.LISS-III Satellite Specification

| **Specification** | LISS-III |
|---|---|
| **Spatial resolution in Nadir** | 23.5 m |
| **Swath** | 141 km |
| **Repetitivity** | 24 days |
| **Spectral Bands** | 0.52-0.59 microns(B2)<br>0.62-0.68 microns(B3)<br>0.77-0.86 microns(B4)<br>1.55-1.70 microns(B5) |
| **Quantization** | 7 Bits SWIR band has 10-bit quantization, selected 7 bits out of 10 bits will be transmitted by the data handling system |
| **No. of gains** | 4 for B2, B3, and B4. For B5 (Dynamic range obtained by sliding 7 out of 10 bits) |
| **Primary Application** | Land use, Urban planning, biodiversity characterization, Forest survey, wetland mapping, environmental impact, production estimation of major crops. |

## 4.2 LISS-III IMAGES

The LISS-III images of the Mumbai region used in the project are:

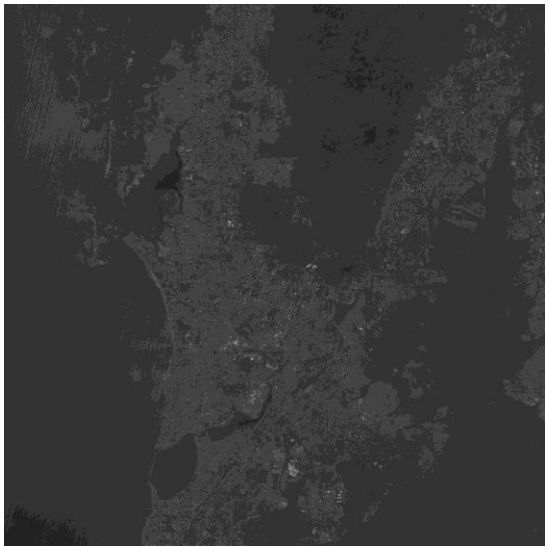

Figure 3.Band 2 Image

Figure 4.Band 3 Image

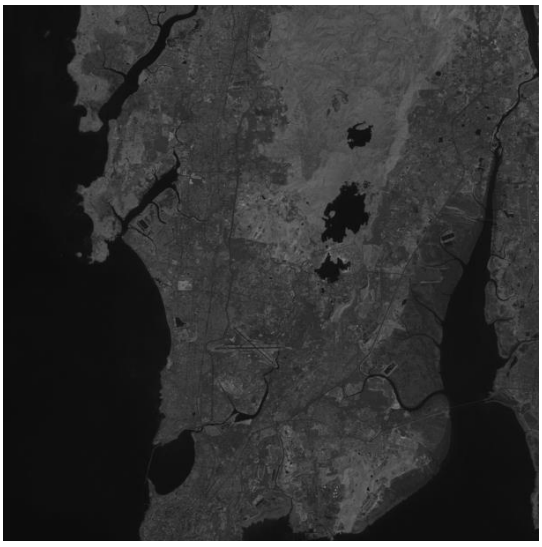

Figure 5.Band 4 Image

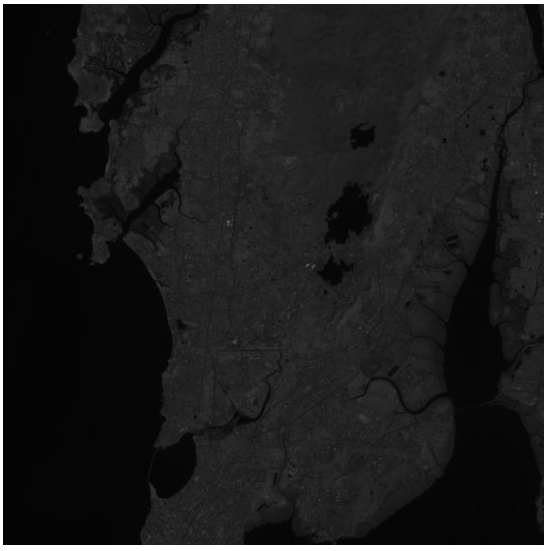

Figure 6.Band 5 Image

# 5. THEORETICAL BACKGROUND

## 5.1 Software Requirements

### 5.1.1. Python 2.7

The programming language used in this project is Python. Python is an interpreted high-level programming language for general-purpose programming. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library. Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of its features support functional programming. Many other paradigms are supported via extensions, including design by contract and logic programming. Python uses dynamic typing and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution, which binds method and variable names during program execution.

### 5.1.2. QGIS 2.14 ESSEN

QGIS is a user-friendly Open Source Geographic Information System (GIS) licensed under the GNU General Public License. QGIS is an official project of the Open Source Geospatial Foundation (OSGeo). It runs on Linux, UNIX, Mac OS X, Windows, and Android and supports numerous vector, raster, and database formats and functionalities. QGIS provides a continuously growing number of capabilities provided by core functions and plug-in. You can visualize, manage, edit, analyze data, and compose printable maps. QGIS can also be used as a graphical user interface to GRASS.

### 5.1.3. Features of QGIS

1. QGIS is a freely downloadable open source GIS software suite.
2. Since QGIS is free and open source software, there are no licensing concerns.
3. QGIS is cross-platform and can be installed and run on Windows, Mac, or Linux machines.
4. QGIS has a faster startup time than ArcGIS.
5. Geoprocessing in QGIS takes very less time than in ArcGIS.

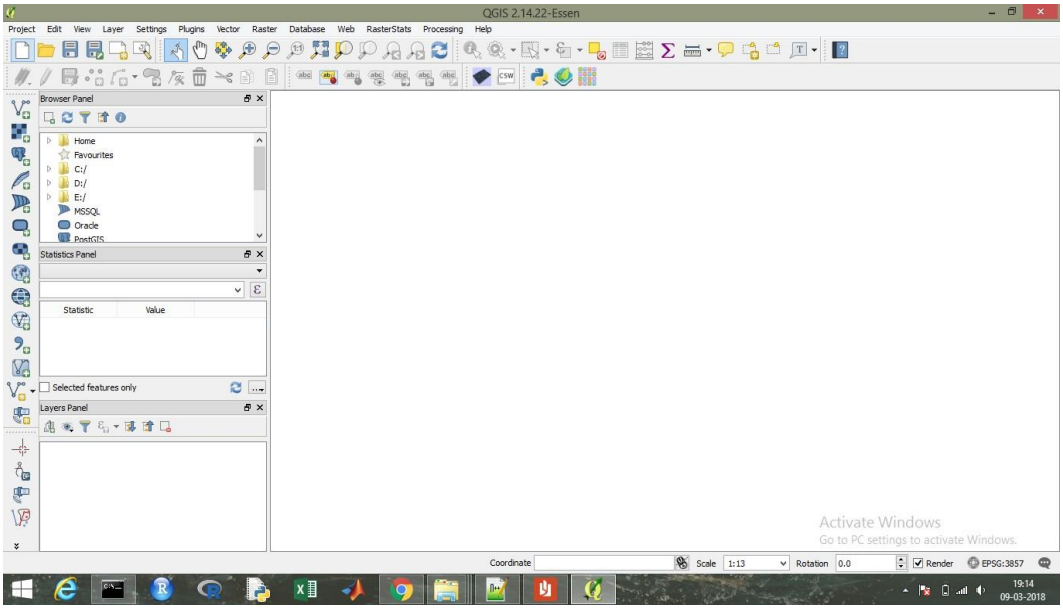


Figure 7.QGIS 2.14 Interface

## 5.2. HARDWARE REQUIREMENTS

1. Minimum 2 GB RAM (4GB recommended)
2. Windows 8.1 or above
3. Intel Dual Core Processor or above
4. Minimum 1GHz Processor

## 5.3 TOOLS AND TECHNIQUES

The libraries used in the project are:

### 5.3.1 NumPy

NumPy is the fundamental package for scientific computing with Python. It contains a powerful N-dimensional array object, many sophisticated functions, and tools for integrating C/C++ code and is useful in linear algebra, Fourier transform, and random number capabilities. NumPy is also used as an efficient multi-dimensional container of generic data.

### 5.3.2 Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hard copy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and I Python shell, the Jupiter notebook, web application servers, and four graphical user interface toolkits.

### 5.3.3 Scikit-learn

Scikit-learn is a free software machine learning library for python programming language. Scikit-learn provide a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. The library is built upon the SciPy (Scientific Python) that must be installed before you can use SciKit-learn. Scikit-learn provide:

1. Simple and efficient tools for data mining and data analysis
2. Accessible to everybody, and reusable in various contexts
3. Built on NumPy, SciPy, and Matplotlib
4. Open source, commercially usable - BSD license

### 5.3.4 OpenCV

OpenCv is a library of Python bindings designed to solve computer vision problems. OpenCv makes use of Numpy, which is a highly optimized library for numerical operations with MATLAB-style syntax. All the OpenCv array structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries such as Numpy, SciPy and Matplotlib.

# 6. ALGORITHMS USED

## 6.1 DECISION TREE

A decision tree is defined as a connected, acyclic, undirected graph, with a root node, zero or more internal nodes (all nodes except the root and the leaves), and one or more leaf nodes (terminal nodes with no children), which will be termed as an ordered tree if the children of each node are ordered (normally from left to right). Decision tree is one of the supervised classification methods. It is very powerful and fast classification and prediction method. A tree is termed as univariate, if it splits the node using a single attribute or a multivariate, if it uses several attributes. A binary tree is an ordered tree such that each child of a node is distinguished either as a left child or a right child and no node has more than one left child or more than one right child. For a binary decision tree, the root node and all internal nodes have two child nodes. All non-terminal nodes contain splits. A Decision Tree is built from a training set, which consists of objects, each of which is completely described by a set of attributes and a class label. Attributes are a collection of properties containing all the information about one object. Unlike class, each attribute may have either ordered (integer or a real value) or unordered values (Boolean value).

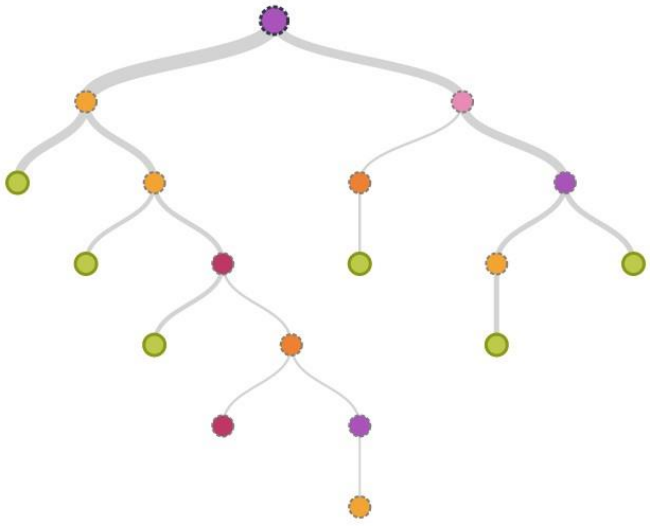

Figure 8.Decision Tree (Courtesy: https://www.iboske.com/)

### 6.1.1 DECISION TREE ALGORITHM

**Step 1**: Let T be the set of training instances.

**Step 2**: Choose an attribute that best differentiates the instances in T.

**Step 3**: Create a tree node whose value is the chosen attribute.

i. Create child links from this node where each link represents a unique value for the chosen attribute.
ii. Use the child link values to further subdivide the instances into subclasses.

**Step 4**: For each subclass created in step 3:

i. If the instances in the subclass satisfy predefined criteria or if the set of remaining attribute choices for this path is null, specify the classification for new instances following this decision path.
ii. If the subclass does not satisfy the criteria and there is at least one attribute to further subdivide the path of the tree, let T be the current set of subclass instances and return to step 2.

## 6.2. K-NEAREST NEIGHBOR

The K-nearest neighbor classification is one of the supervised classification techniques. The KNN classification algorithm is very easy to understand and implementation. It is widely used for remote sensing area mapping. The KNN classification is instance based classification techniques where the algorithm tries to make the instance of different classes based on the provided sample training data sets which is called as learning. In KNN, there is no development of internal model or there is nothing as creation of relationship between the features and classes. It simply creates instances and stores all these instances for prediction of classes. The prediction of each class is based on the voting method of nearest neighbor of each point. There are 5 distance functions used in KNN.

Given an mx-by-n data matrix X, which is treated as mx (1-by-n) row vectors x1, x2... xmx, and my- by-n data matrix Y, which is treated as my (1-by-n) row vectors y1, y2... ymy, the various distances between the vector xs and yt are defined as follows:

1. Euclidean distance: It is the distance between two points in the plane with coordinates. It is the default distance metric.

$$d_{2st} = (x_s - y_t)(x_s - y_t)'$$

2. Cityblock distance: It is sum of absolute differences.

$$d_{st} = \sum (j = 1)\text{^}n|x_{sj} - y_{tj}|$$

3. Cosine distance: One minus the cosine of the included angle between points. It is treated as vectors.

$$d_{st} = (1 - ((x_s y'_t)/\sqrt{(x_s x'_s)}\sqrt{(y_t y'_t)}\,)$$

4. Correlation distance**:** One minus the sample correlation between points.

$$d_{st} = \frac{(x_s - x'_s)(y_t - y'_t)}{(\sqrt{(x_s - x'_s)}\sqrt{(x_s - x'_s)}'\sqrt{(y_t - y'_t)}\sqrt{(y_t - y'_t)}}$$

5. Hamming distance: It is the percentage of bits that differ and is suitable only for binary data.

$$d_{st} = (x_{sj} \neq y_{tj})/n$$

## 6.2.1. APPLICATIONS OF KNN

1. Some of the applications of this method are mentioned below:
2. Classification and Interpretation- KNN method can be used in the field of medical, banking, media and legal etc.
3. Problem-solving- KNN can be used for building tools for planning and is also used as learning module in terms of pronunciation.
4. Function learning – dynamic control.
5. Teaching and aiding – help desk, user training.

## 6.2.2. KNN ALGORITHM

1. For each training sample, add the sample to the list of training samples.
2. Given a query instance xq ¨ given a query instance x to be classified, q to be classified.
    i. Let x1, x2….xk denote the k instances from training samples that are nearest to xq.
    ii. Return the class that represents the maximum of the k instances

## 6.3. ARTIFICIAL NEURAL NETWORK

An artificial neural network is a biologically inspired computational model formed from hundreds of single units, artificial neurons, connected with coefficients (weights) which constitute the neural structure. They are also known as processing elements as they process information. Each processing element has weighted inputs, transfer function and one output. Processing element is essentially an equation which balances inputs and outputs. ANNs are also called connectionist models as the connection weights represent the memory of the system. ANNs are capable of processing extensive amounts of data, however, and making predictions that are sometimes surprisingly accurate.

### 6.3.1 NEURONS

The artificial neuron is the building component of the ANN designed to simulate the function of the biological neuron. The arriving signals, called inputs, multiplied by the connection weights (adjusted) are first summed (combined) and then passed through a transfer function to produce the output for that neuron. The activation function is the weighted sum of the neuron's inputs and the most commonly used transfer function is the sigmoid function.

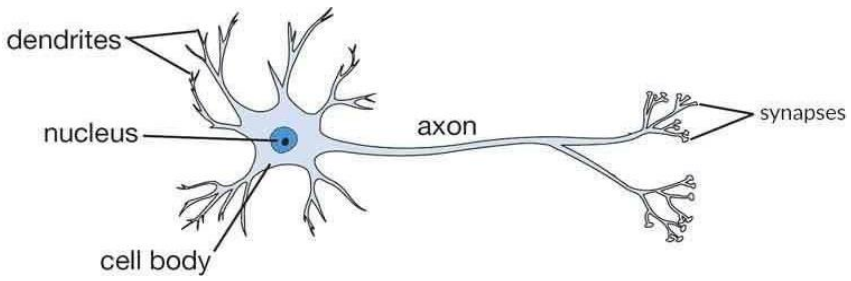


Figure 9.Biological Neuron (Courtesy: https://laptrinhx.com/topic/32279/overview-of-artificial-neural-networks-and-its-applications)

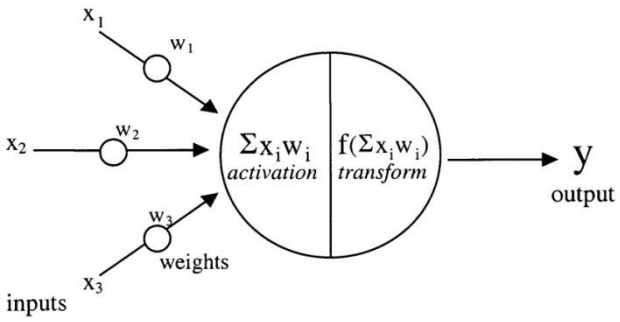


Figure 10.Model of an artificial neuron (Courtesy: https://www.researchgate.net/)

### 6.3.2 PATTERN RECOGNITION

Pattern recognition is the assignment of a label to a given input value. Well known application include other typical applications of pattern recognition techniques that are automatic recognition, classification, the automatic recognition of handwritten postal codes on postal envelopes, automatic recognition of images of human faces, or handwriting image extraction from medical forms.

### 6.3.3 CLUSTERING/CATEGORIZATION

Categorization is the process in which ideas and objects are recognized, differentiated, and understood. Categorization implies that objects are grouped into categories, usually for some specific purpose. Applications of clustering include data mining, data compression and exploratory data analysis.

### 6.3.4 PREDICTION/FORECASTING

It has a significant impact on decision making in business, science and engineering. Stock market prediction and weather forecasting are typical application of prediction/forecasting techniques.

### 6.3.5 CHARACTERISTICS OF ANN

1. The Network Structure of ANN is simple and easy.
2. Parallel Processing Ability.
3. Distributed Memory.
4. Fault Tolerance Ability.
5. ANN is an interconnected system the output of a system is a collective output of various input so the result is summation of all the outputs which comes after processing various inputs.
6. Learning Ability.

### 6.3.6 ALGORITHM

**Step 1**: Applying artificial neural network on dataset.

a. Feed forward neural network with back propagation of error

b. Feed forward neural network using back propagation of error algorithm

c. Training of artificial neural networks

**Step 2**: Reading the image and converting to array.

**Step 3**: Reconstructing the image and applying different classes.

**Step 4**: Calculating accuracy and kappa-value.

# 7. METHODOLOGY

## 7.1 PREPROCESSING

The LISS-III images obtained are present in different bands i.e. Band 2, Band 3, Band 4 and Band 5. These bands represent different spectral values. There bands need to be combined to produce the RGB image of the area. This combining of the LISS-III images is done by the merge () in python. The merge () combines the band images (2, 3, 4) and produces an RGB image.

## 7.2 FEATURE EXTRACTION

Digital image classification is the process of assessing pixels to classes. Each pixel is treated as individual unit, and by comparing pixels of unknown and pixels of known identity it is possible to assemble groups of pixels into classes. For assigning the pixel values, Matlab software is used. The RGB image obtained by stacking process is used as an input. The pixels can be assigned using impixel(). The impixel() returns the value of pixels in the specified image, where the image can be a grayscale, binary, or RGB image. The impixel() displays the image specified and waits for the user to select the pixels in the image using the mouse. We need to assign the pixel values on the image according to the desired need. In this project, we have created 4 classes in the image i.e. Mangroves, Forest, Water and Development. These classes are represented using different color i.e. red, green, blue and yellow respectively. A Classified image is created using this classification method. Classified image helps to classify the image based on classes.

## 7.3 CLASSIFICATION

After the design and training with the training dataset, the algorithm is ready to be tested. These training datasets are used to test the algorithm. Based on the testing dataset, the accuracy of the algorithm is calculated. The confusion matrix and kappa coefficient are the two important techniques which are used to calculate the accuracy. These training datasets are used to classify the satellite images under the land cover as mangrove, forest, water and development.

## 7.4 PROPOSED METHOD

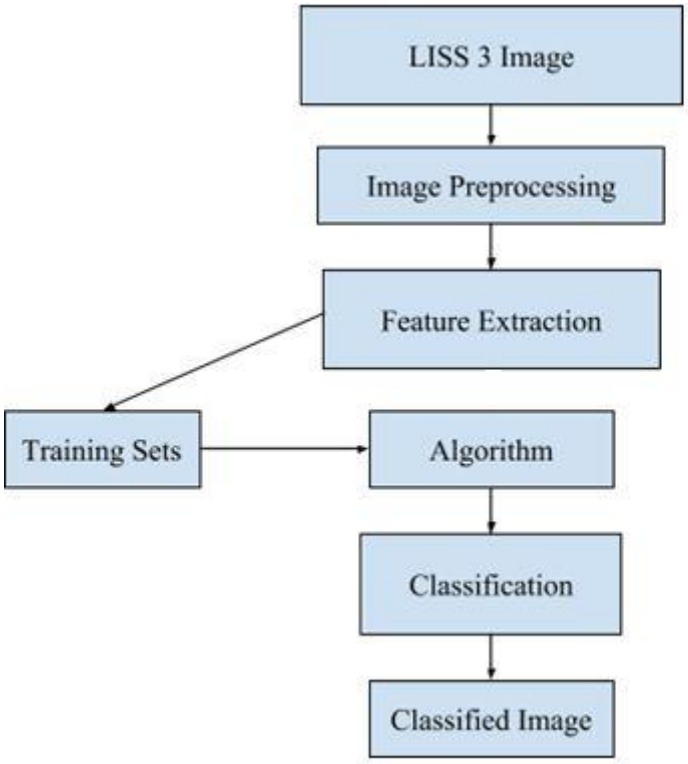


Figure 11.Proposed Method

# 8. GRAPHICAL USER INTERFACE

The GUI for the project is created in python using Tkinter. Tkinter is Python's standard GUI package. It is the most commonly used GUI Package in python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit. Creating a GUI application using Tkinter is an easy task. All you need to do is perform the following steps −

i. Import the Tkinter module.
ii. Create the GUI application main window.
iii. Add one or more widgets to the GUI application.
iv. Enter the main event loop to take action against each event triggered by the user.

The Tkinter widgets include Button, Canvas, Checkbutton, Entry, Frame, Label, ListBox, MenuButton, Menu, Message, RadioButton, Scale, Scrollbar, Text, TopLevel, Spinbox, PanelWindow, LabelFrame and tkMessageBox.

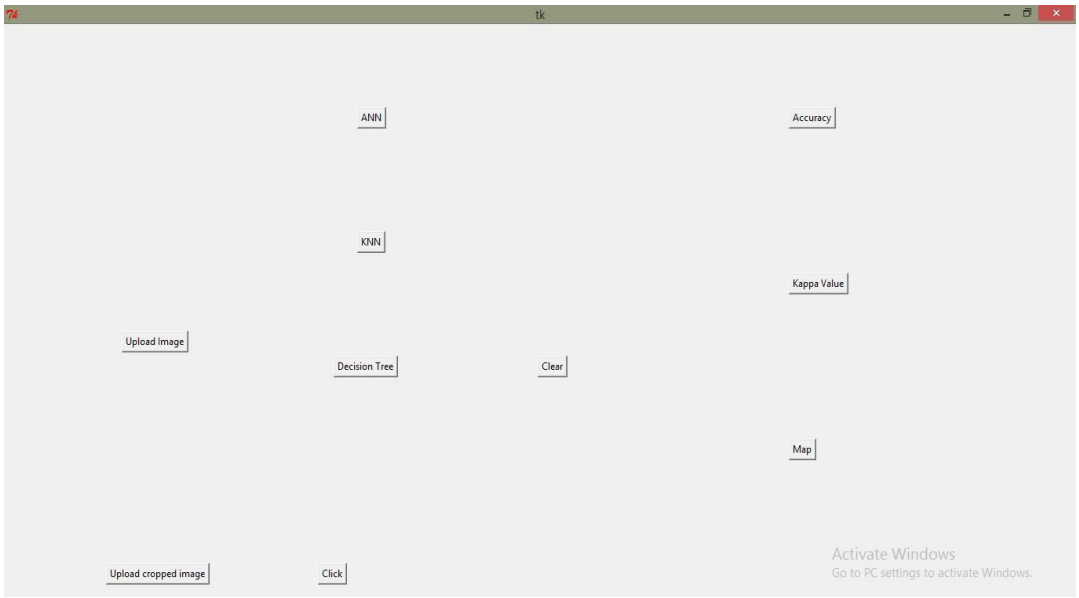


Figure 12.GUI -1

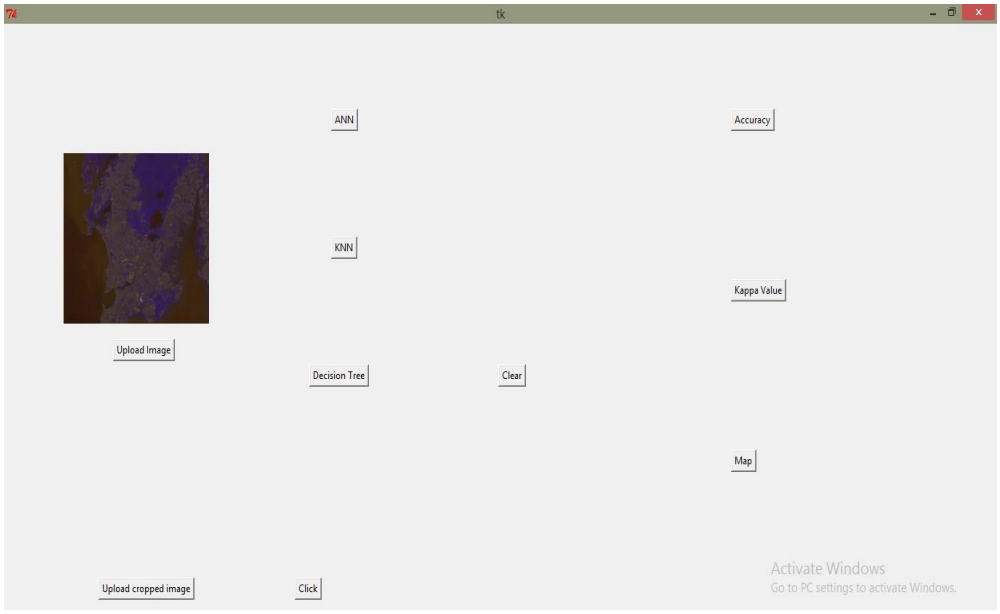


Figure 13.GUI-2


Figure 14.GUI-3

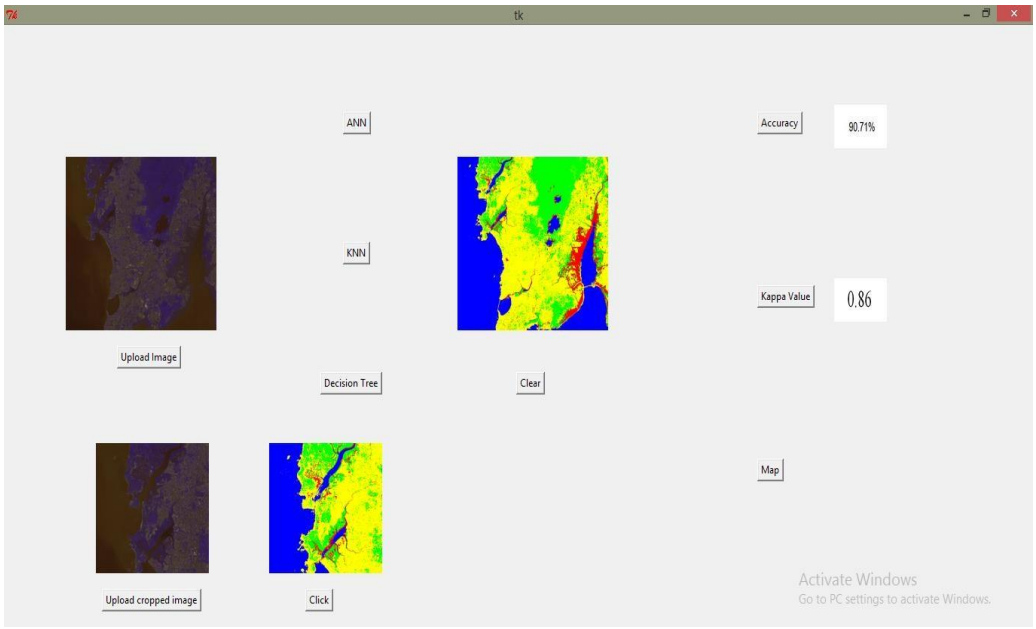


Figure 15.GUI-4

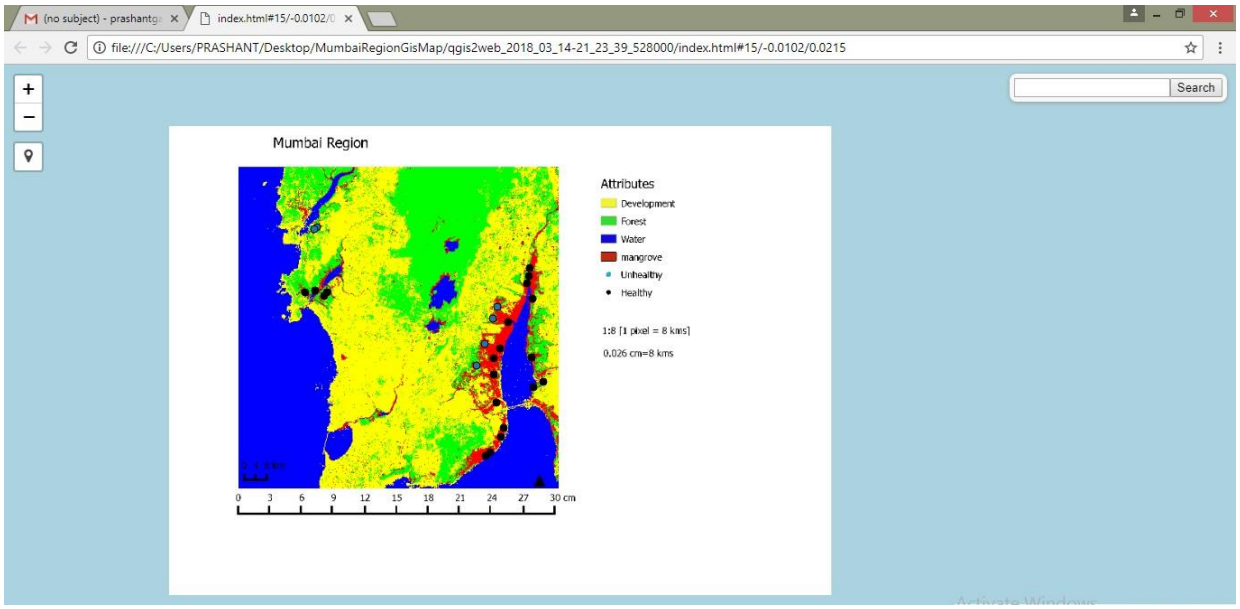


Figure 16.GUI-5

# 9. CREATION OF INTERACTIVE MAP

The following are the steps involved in creating an interactive map in QGIS:

**STEP 1:** Loading the raster image

- The first step in creating an interactive map is to load the raster image in the QGIS application. The classified image of the Mumbai region is used as the raster image.

**STEP 2:** Create a new print composer

- We need to create a print composer from the project menu. Print composer is used to add the map components on the classified image. Map components like legend, north arrow, scale bar, and label can be added. Legend, scale bar, label and the north arrow are added on the classified image.

**STEP 3:** Saving the raster image

- The image is saved with the entire necessary component added in it.

**STEP 4:** Open the new raster image

- The newly created raster image is loaded in the QGIS application for further processing.

**STEP 5:** Creating layers

- The next step is creating point layers. There are two layers to be created i.e. Healthy and Unhealthy. These two layers are point layers which represent a point on the image.

**STEP 6:** Plotting the points

- Two layers are created namely healthy and unhealthy. The two layers specify the conditions of the mangroves in Mumbai. The points are plotted on the image according to the information collected.

**STEP 7:** Creating a print composer

- A new print composer is created with the changes in the classified image. Two new layers are added in the image i.e. healthy and unhealthy point layers.

**STEP 8:** Deploying the map on the web:

- After the complete interactive map is created, we need to deploy it on the web. This can be done by using the QGIS plug-in QGIS2Web. QGIS2Web provides web mapping of the GIS data. QGIS2Web uses Open Layers and Leaflet libraries of the project. After the mapping is complete, the map can be viewed in the web browser which provides the abilities of zooming the map, searching for an particular area and it becomes easier to traverse to the data source that is provided in the point layer.

# 10. ACCURACY ASSESSMENT

## 10.1 CONFUSION MATRIX

A confusion matrix is a table that contains information about actual and predicted classifications done by a classification system. It is used to describe the performance of a classifier on set of test data for which the true values are known.

**Table 2**.Confusion Matrix

| | | **Predicted** | |
|---|---|---|---|
| | | **Class1** | **Class2** |
| **Actual** | **Class1** | **A** | **B** |
| | **Class2** | **C** | **D** |

The entries in the confusion marix have the following meaning in the context to our studies:

A is the number of correct predictions that an instance is class 1.

The entries in the confusion matrix have the following meaning in the context of our study: A is the number of **correct** predictions that an instance is **class1**.

B is the number of **incorrect** predictions that an instance is **class2**. C is the number of **incorrect** of predictions that an instance **class1.** D is the number of **correct** predictions that an instance is **class2**.

Several standard terms have been defined for the 2 class matrix:

1. The **accuracy** is the proportion of the total number of predictions that were correct. It is determined using the equation

$$AC = \frac{(a+d)}{(a+b+c+d)}$$

2. The **true positive rate** is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$TP = \frac{d}{c+d}$$

3. The **false positive rate** is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the equation:

$$FP = \frac{b}{a+b}$$

4. The **true negative rate** is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

$$TN = \frac{a}{a+b}$$

5. The **false negative rate** is the proportion of positives cases that were incorrectly classified as negative, as calculated using the equation:

$$FN = \frac{c}{c+d}$$

6. Finally, *precision* (*P*) is the proportion of the predicted positive cases that were correct, as calculated using the equation:

$$P = \frac{d}{b+d}$$

The drawback of this measure is that it does not tell you anything about how well individual classes were classified. The user and producer accuracy are two widely used measures of class accuracy.

The producer's accuracy refers to the probability that a certain land-cover of an area on the ground is classified as such, while the user's accuracy refers to the probability that a pixel labeled as a certain land-cover class in the map is really this class.

## 10.2. KAPPA STATISTIC

The Kappa statistic was derived to include measures of class accuracy within an overall measurement of classifier accuracy. It provides a better measure of the accuracy of a classifier than the overall accuracy, since it considers inter-class agreement. The kappa value is calculated using the below formula:

$$k = \frac{N \sum_{i=1}^{r} xii - \sum_{i=1}^{r} (Xi + * X + i)}{N2 - \sum_{i=1}^{r} Xi + * X + i}$$

# 11. RESULTS AND OBSERVATION

A confusion matrix is an also a tool to summarize accuracy of the classifiers. It basically used for the supervised classification of the data. It consists of the information about classified verses misclassified data in the supervised learning. The diagonal elements in the confusion matrix show the correctly classified data whereas the non-diagonal elements are misclassified data of the classification. So with the help of that the accuracy of the classifier is easily calculated and it will help the algorithm designer to understand the importance and applicability of particular classifier for the specific data.

## 11.1. ACCURACY AND KAPPA-VALUE OF ARTIFICIAL NEURAL NETWORK

**Table 3**.Accuracy and kappa-value of artificial neural network

| **Class** | **Mangroves** | **Forest** | **Water** | **Development** | **Total** | **User's Accuracy** |
|---|---|---|---|---|---|---|
| **Mangroves** | 25 | 2 | 0 | 3 | 30 | 83.33 |
| **Forest** | 6 | 62 | 0 | 13 | 81 | 76.54 |
| **Water** | 4 | 0 | 119 | 0 | 123 | 96.75 |
| **Development** | 0 | 2 | 0 | 87 | 89 | 97.75 |
| **Total** | 35 | 66 | 119 | 103 | | |
| **Producer's Accuracy** | 71.43 | 93.94 | 100 | 84.47 | | 90.71% |

Accuracy = ((25+62+119+87)/323)*100

Accuracy= 90.71%

Kappa value= 0.86

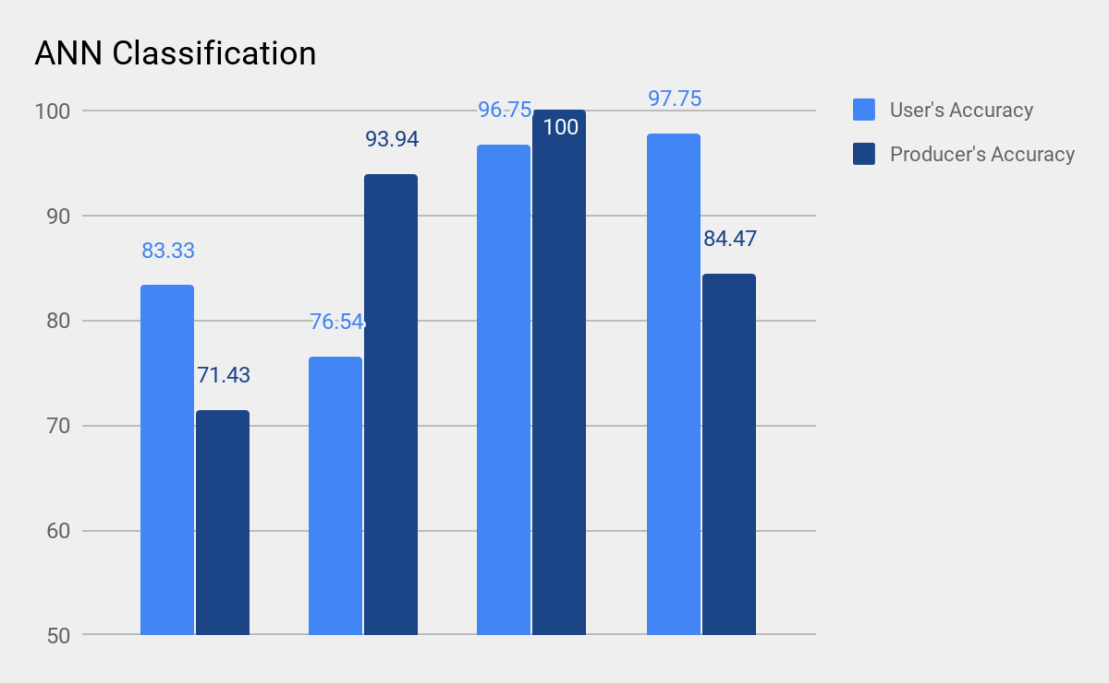


Figure 17.Classification Accuracy Assessment of Artificial neural network

The confusion matrix based accuracy assessment of LISS-III satellite image shows that the accuracy of classification using Artificial Neural Network is 90.71%. Similarly kappa coefficient of ANN is 0.86 which is good. The accuracy of the classifiers can be increased by increasing the number of features and sample data.

## 11.2. ACCURACY AND KAPPA-VALUE OF K-NEAREST NEIGHBOR

**Table 4**.Accuracy and kappa-value of K-nearest neighbor

| **Class** | **Mangroves** | **Forest** | **Water** | **Development** | **Total** | **User's Accuracy** |
|---|---|---|---|---|---|---|
| **Mangroves** | 27 | 2 | 0 | 0 | 29 | 93.10 |
| **Forest** | 3 | 61 | 0 | 14 | 78 | 78.20 |
| **Water** | 4 | 0 | 119 | 0 | 123 | 96.75 |
| **Development** | 1 | 3 | 0 | 89 | 93 | 95.70 |
| **Total** | 35 | 66 | 119 | 103 | | |
| **Producer's Accuracy** | 77.14 | 92.42 | 100 | 86.40 | | 91.64% |

Accuracy = ((27+61+119+89)/323)*100

Accuracy= 91.64%

Kappa value= 0.88

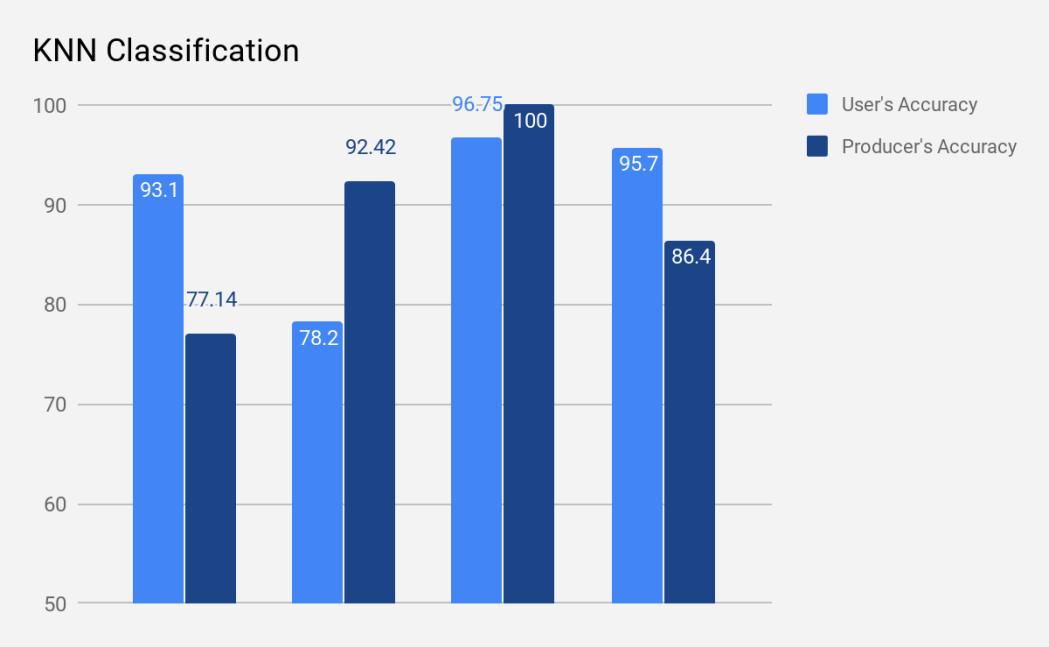


Figure 18.Classification Accuracy Assessment of K-nearest neighbor classification

The confusion matrix based accuracy assessment of LISS-III satellite image shows that the accuracy of classification using K-Nearest Neighbor is 91.64%. Similarly kappa coefficient of KNN is 0.88 which is good. The accuracy of the classifiers can be increased by increasing the number of features and sample data.

## 11.3. ACCURACY AND KAPPA-VALUE OF DECISION TREE

**Table 5**.Accuracy and kappa-value of Decision tree

| **Class** | **Mangroves** | **Forest** | **Water** | **Development** | **Total** | **User's Accuracy** |
|---|---|---|---|---|---|---|
| **Mangroves** | 26 | 6 | 0 | 5 | 37 | 70.27 |
| **Forest** | 6 | 54 | 0 | 24 | 84 | 64.28 |
| **Water** | 3 | 1 | 119 | 0 | 123 | 96.75 |
| **Development** | 0 | 5 | 0 | 74 | 79 | 93.67 |
| **Total** | 35 | 66 | 119 | 103 | | |
| **Producer's Accuracy** | 74.29 | 81.81 | 100 | 71.84 | | 84.52% |

Accuracy = ((26+54+119+74)/323)*100

Accuracy= 84.52%

Kappa value= 0.78

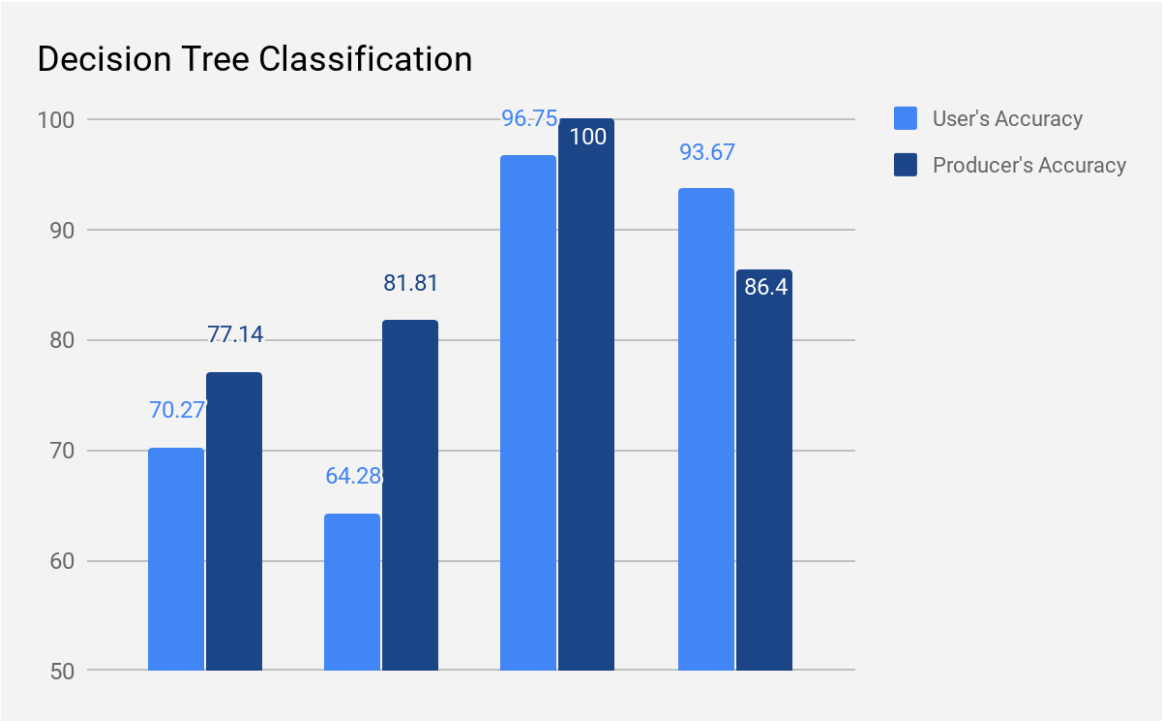


Figure 19.Classification Accuracy Assessment of Decision tree classification

The confusion matrix based accuracy assessment of LISS-III satellite image shows that the accuracy of classification using Decision tree is 84.52%.Similarly kappa coefficient of Decision Tree is 0.78 which is bad. The accuracy of the classifiers can be increased by increasing the number of features and sample data.

The overall Accuracy and Kappa value are:

**Table 6**.Overall Accuracy and kappa value

| Sr no | Classifier used | Accuracy | Kappa Value |
|---|---|---|---|
| 1 | K-Nearest Neighbor | 91.64% | 0.88 |
| 2 | Decision Tree | 84.52% | 0.78 |
| 3 | Artificial Neural Network | 90.71% | 0.86 |

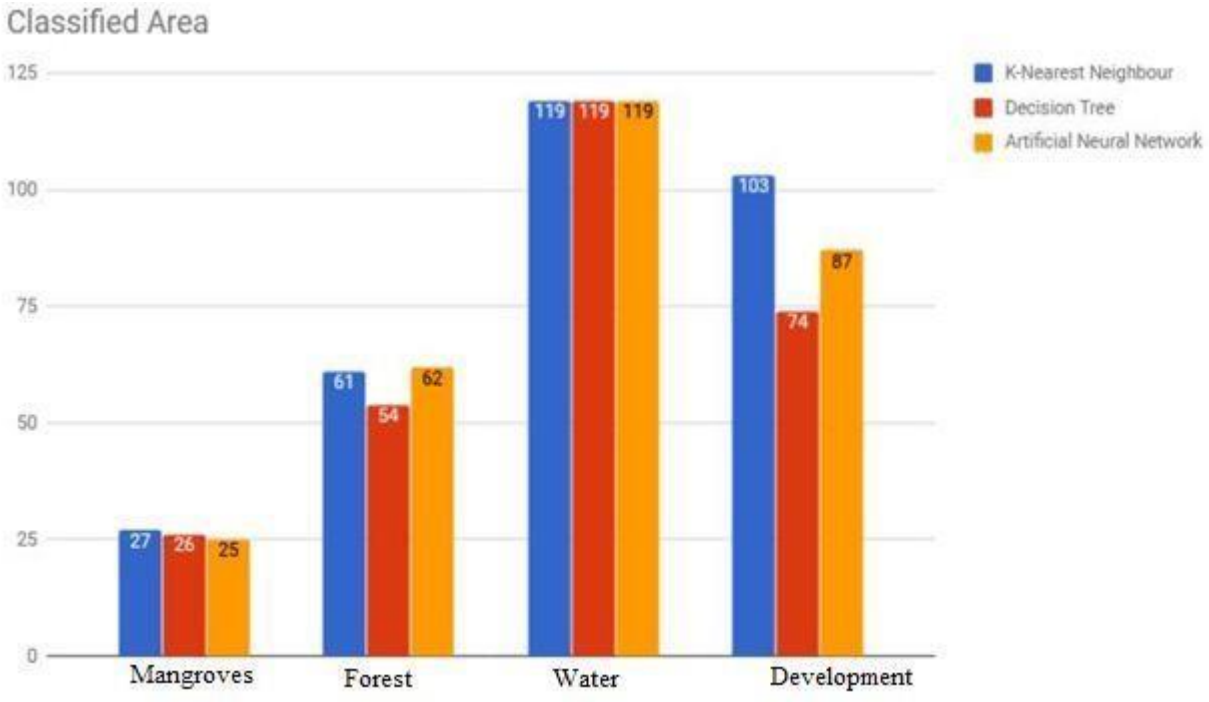


Figure 20.Overall Classified Area

The above graph shows the area that is correctly classified by the algorithms. For example, 27 pixels of mangroves were correctly classified by K-nearest Neighbor algorithm while Decision tree had 26 and artificial neural network had 25.

## 11.4. CLASSIFIED AND UNCLASSIFIED IMAGES

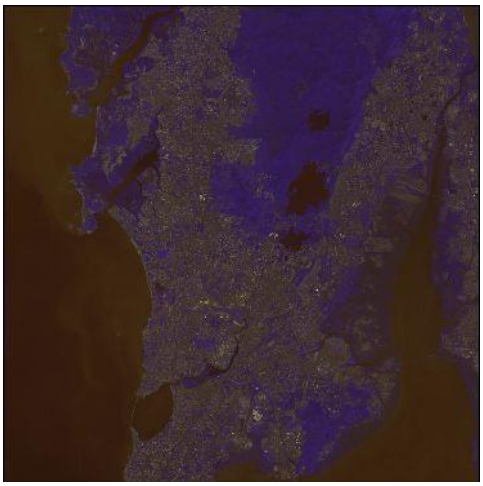

Figure 21.False color image before Classification

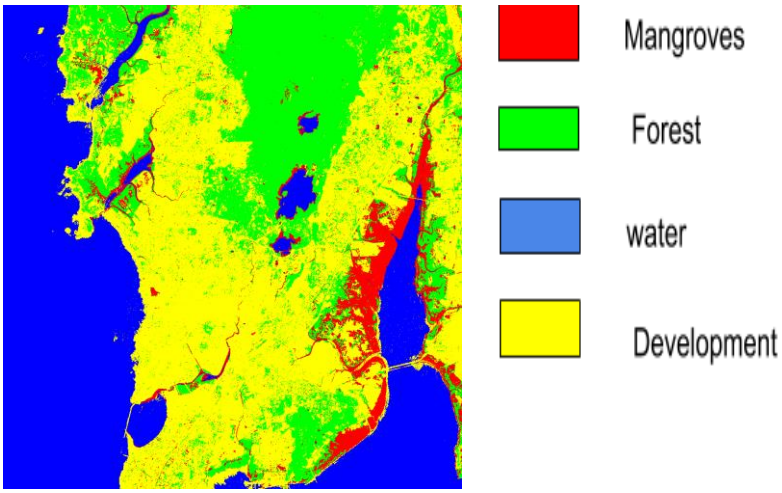


Figure 22.False color image after Classification

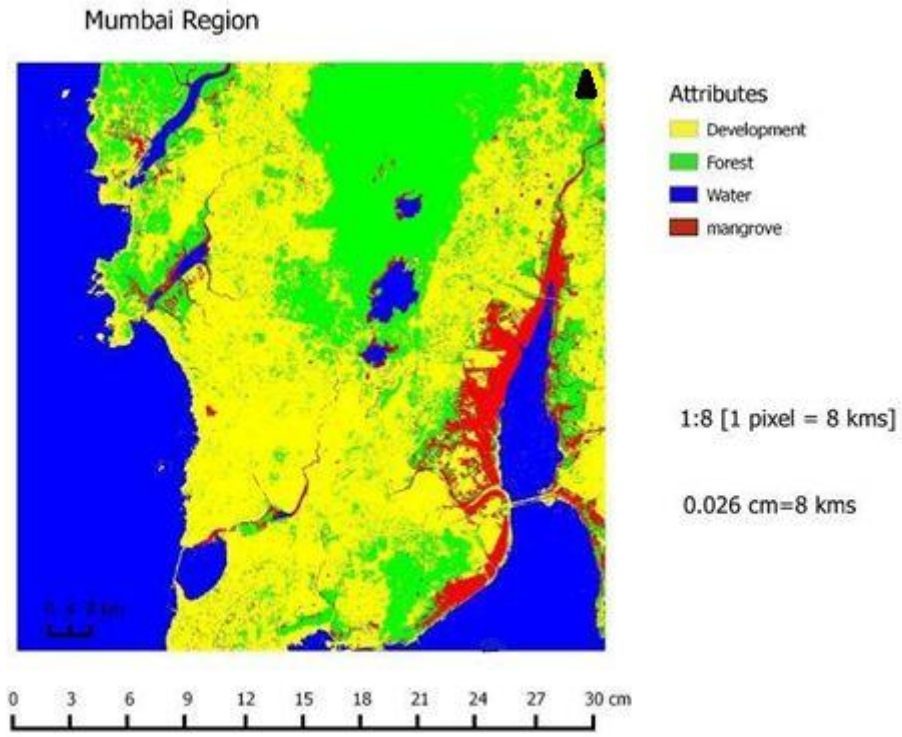


Figure 24.Map

# 12. DISCUSSIONS

The accuracy assessment of different pixel based approaches, have proved that all the classification techniques have produced results with satisfactory accuracy, which proves that all this techniques are suitable for classifying and mapping the Mangrove area in Mumbai region. By conducting statistical analysis of the accuracies, we find that the KNN based approach has better performance in terms of accuracy and kappa-value as compared to different classifiers implemented.

First of all, we implemented Artificial Neural Network classifier from which the accuracy of 90.71% and its corresponding kappa-value was 0.86 achieved. Then after going through various research papers we got an idea for implementing decision tree. After implementing decision tree classifier we got that the accuracy as 84.52% and its kappa-value as 0.86. After classifying using decision tree we got that the accuracy and kappa-value is more as compared to Artificial Neural Network classifier. With the aim of gaining more accuracy we then implemented KNN. KNN gave us the accuracy of 91.64% with corresponding accuracy as 0.88. The accuracy and kappa-value gained from KNN is more as compared to other two classifiers.

# 13. CONCLUSION

We have successfully created the thematic map using classified image that can be served as thematic map or as input to QGIS for creating interactive and informative map. The satellite image based classification of Mangroves is very wide area of study and research and so many people's are doing research in terms of making efficient algorithm, performance, data handling, training or time. So, here in the same respect the Decision tree, KNN and ANN has been used for classifying LISS-III satellite image and accuracy is calculated using confusion matrix. The results show that the accuracy of the KNN is highest that is 91.64% with the kappa value 0.88. This study also shows that the classification accuracy increases with the increase in the size of the training set but up to a certain limit. Better accuracy can be achieved by increasing the number training samples during learning process. Number of training samples depends on the complexity of the study area. If study area is simple and it consists of well- defined crisp classes then less number of pixels can also give better accuracy.

# 14. FUTURE ENHANCEMENT

If the LISS-III dataset of further year will available then we can use the forecasting algorithm and some GPS tool on these image to predict future of mangroves and provide some metrics like which area will have mangroves growth and in which area mangroves will be depleted and provide accurate mapping region wise, so that we can take necessary steps to protect them.

The accuracy of the entire process can be increased by classifying the data set with different classifiers. Classifiers like Radial basis network function, Principal component analysis, SVM and LVQ techniques can be used to improve the classification accuracy. In future other statistical parameters are also used to improve classification accuracy. The accuracy can further be improved by using various techniques for classification, such as fuzzy logic and genetic algorithms with all these different feature values.

# 15. REFERENCES


[1] Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research by S. Agatonovic-Kustrin *, R. Beresford School of Pharmacy, University of Otago, PO Box 913, Dunedin, New Zealand Received 25 August 1999; received in revised form 17 September 1999; accepted 23 October 1999.

[2] Roya Abedi, Amir Eslam Bonyad," Estimation and Mapping Forest Attributes Using "k Nearest Neighbor" Method on IRS-P6 LISS III Satellite Image Data", ECOLOGIA BALKANICA, Vol. 7, Issue 1 June 2015, pp. 93-102.

[3] Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matrix

[4] M. K. Ghose, Ratika Pradhan, Sucheta Sushan Ghose Department of Computer Science and Engineering Sikkim Manipal Institute of Technology,Sikkim, INDIA.

[5] Mangrove Area Assessment in India: Implications of Loss of Mangroves, Sahu SC1*, Suresh HS2, Murthy IK1 and Ravindranath NH1.

[6] Assessment of the health status of Indian mangrove ecosystems using multi temporal remote sensing data, prabakaran chellamani, Chandra prakash singh & sushma panigrahy.

[7] Mr. Anand Upadhyay, Dr.Santosh kumar Singh, Mr. Aditya Shetty and Mr.Zibreel ,Siddiqui," Land use and Land cover classification of LISS-III satellite image using KNN and Decision tree", Proceeding of the 10th INDIACom;3rd International conference on computing for sustainable global development, Organized by BVICAM,16-18 March,2016, IEEEXplore, ISBN:978-93-80544-19-9.

[8] Study of mangrove environment of Maharashtra using remote sensing data, T G Jagtap, A G Untawale and S N Inamdar.

[9] Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning Kai Liu, Xia Li, Xun Shi, and Shugong Wang, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences.

## 15.1 LINKS


[1] https: //www.analyticsvidhya.com/blog/2014/10/ann-work-simplified [12-JAN 2018]

[2] https://www.techopedia.com/definition/5967/artificial-neural-network-ann [28-DEC 2017]

[3] https://machinelearningmastery.com/k-nearest-neighbors-for-machine-learning/ [4-DEC 2017]

[4] http://scikit-learn.org/stable/modules/tree.html [21-JAN 2018]

[5] http://scikit-learn.org/stable/modules/neural_networks_supervised.html [13-JAN 2018]

[6] http://scikit-learn.org/stable/modules/neighbors.html [10-DEC 2017 ]

[7] http://www.sunday-guardian.com/investigation/godrej-helps-preserve-mumbai-mangroves.html [6-SEP 2017]

[8] https://www.hindustantimes.com/mumbai-news/two-mangroves-from-mumbai- region-on-list-the-12-unique-wetlands-in-india/story- EiVR3zcmlQBL2y19wYLMiJ.html [10-AUG 2017]

[9] http://indianexpress.com/article/cities/mumbai/bird-man-may-soon-meet-up-along-mulund-mangrove/ [15-AUG 2017]

[10] http://www.wildgraphy.com/albums/mulund-mangroves/ [22-AUG 2017]

[11] https://timesofindia.indiatimes.com/city/navi-mumbai/nmmc-plans-a-cable-bridge- over-ghansoli-mangrove-area/articleshow/57965215.cms [13-AUG 2017]

[12] https://timesofindia.indiatimes.com/city/navi-mumbai/nmmc-plans-a-cable-bridge- over-ghansoli-mangrove-area/articleshow/57965215.cms [11-SEP 2017]

[13] https://www.mumbailive.com/en/environment/high-court-declines-the-decision-of-cutting-mangroves-in-mumbai-for-recreation-9982 [25-OCT 2017]

[14] https://oceanservice.noaa.gov/facts/remotesensing.html [23-DEC 2017]

[15] http://eo.belspo.be/Directory/SensorDetail.aspx?senID=72[27-DEC 2017]

[16] http://eoedu.belspo.be/en/satellites/irs.htm [24-NOV 2017]