Automatic change detection of forest cover change in Sri Lanka from 2008 to 2018 using satellite images

A dissertation submitted for the Degree of Master of Computer Science

M.N.T. Perera

University of Colombo School of Computing

2019

ACKNOWLEDGEMENT

This thesis could not be finished without the help and support of many people. It's an honor for me to express my gratitude to everyone.

Firstly, I would like to express my sincere gratitude and respect to my supervisor, Dr. K.L. Jayaratne for his advice, guidance and supervision throughout this project to ensure its successful accomplishment.

I wish to extend my deepest gratitude to my friends Mrs. Umeshika Panchali and Mr. E.N.R Fernando for their endless support and encouragement, even during the most demanding phases of this research.

Further, I should thank my parents, although words fail to express my indebtedness to them for their love, support and encouragement given not only during this project, but throughout my life.

ABSTRACT

According the statistics by United Nations FAO(Food and Agriculture Organization) in 2010, Sri Lanka there are 167,000 hectares as primary forest and 185,000 hectares forest as planted forest. But, unfortunately the amount of forest cover is decreasing and it stated that total forest cover reduction is 17.7% per year. Calculating the forest lost is still in the manual process in Sri Lanka and it was identified as time consuming and expensive task. Among the forest areas in Sri Lanka, Willpattu has largest coverage and it contributes to the rain and water and air purification. This study was conducted to investigate change detection of "Wilpattu" forest cover by using the satellite remote sensing data from 2000 to 2019. Here only geo-referenced RGB images of corresponding satellite images were taken which are Landsat 7 and 8. RGB values and HUE values of those satellite images were used to identify the forest cover by using statistical model and neural network. In statistical model, samples of forest, water and land area were obtained using the human perception and then the calculated minimum and maximum values were used for the classification of images. In Neural network models, Green, Red and Hue pixel values of images were used as the input to neural network. K-means clustering was used to divide the clusters as forest, water and Land. The results show 55% accuracy when compared with the data of forest department data. HUE values are used same particular images using mean and variance and results shows 65% accuracy. But it shows 352.59 km^2 forest cover reduction in statistical model and 589.61 km^2 in neural network model (RGB) and 555.45 km^2 in HUE from 2000 to 2019. Although it shows the decrease pattern of forest cover, both classification methods misclassified other vegetation types as forest. This can conclude by using only Red, Green and Hue it is not enough to identify the accurate forest cover and further development of model is still required.

ABBREVIATIONS

List of abbreviations

Abbreviation	Description
USGS	United States Geological Survey
FFNN	Feed Forward Neural Network
CNN	Convolutional Neural Network
DN	Digital Number
MATLAB	Matrix Laboratory
GIS	Geographical Information Systems

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 - Introduction

A forest is a large area covered by trees and it contains various kinds of plants and animals. Trees play an important role in the environment. It cleans air by absorbing carbon dioxide and pumping out oxygen. Furthermore, trees help to minimize soil erosion and protect watershed. According to estimation of 2010 the world's forest area measured as 4 billion hectares. And it is unevenly distributed and more than 90% of the forest area is composed of naturally generated forests[1].

1.1 The problem Domain

UN Food and Agriculture Organization reported that, there were 1,860,000 hectares of Sri Lankan forest cover in 2010, 167,000 hectares are classified as primary forest, Sri Lanka has total area of plantation 185,000 hectares. lost 24,500 ha per year from 1990 to 2010. It can be stated that, 20.9% of forest cover has been lost for Sri Lanka from 1990 to 2010 [2].

According to the Wilpattu National Park Corporate website [3]. Before 1938 willpattu is recognized as sanctuary and then declared as a National Park. It has been known as the largest and one of the oldest national park of Sri Lanka. The park is having 131, 693 hectares. It is home for many villus, or natural lakes. In the forest 605 native plants and 118 species were found, 33 out of which are unique to Sri Lanka. Willpattu is among the top national park world-recognized for its leopard population. However, in today's world deforestation in Wilpattu forest complex become a major issue. It has shown a dangerous and growing trend of deforestation since the late 2000s [4].

Forestland clearing, destroying and removing forest areas is called deforestation. Due to the deforestation there is a loss of trees. Then it can cause climate change, increased greenhouse gases and soil erosion. To mitigate these issues, country has to protect the forest.

1.2 The problem

The main problem of this research is to find fast, consistent, versatile, accurate and cost effective way to evaluate forest cover change using up to date technologies. Currently, among the existing methods single or multi date images can be used effectively with the human expertise for defining and labelling that consider as change [5]. This processing technique is based on an analyst's experience and knowledge through visual interpretation. Therefore, this process is time consuming and repetitive. Still there is no automation process in Sri Lanka to detect the change of forest with different time frame and not given the change as quantifiable amount. The proposed method addresses those limitations.

1.3 Motivation

This research idea is inspired through many previous works. But only a few studies involve change detection techniques, and most studies don't support their conclusions. Therefore, it is important to have an effective way to read satellite images of forest and display the forest cover change in different time frames. In this research, one of the main objective is to develop an algorithm for processing of satellite images in sub pixel level and use pixel based classification method for analysis. This research is combination of image processing and data mining techniques to automate the change detection of forest.

1.4 Exact Computer Science Problem

People use quantitative measurements in order to measure the change in forest cover and these are some existing methods to measure forest cover change like post and pre classification change detection. Currently among the methods, using remote sensing and GIS is also become trending methodology. Existing method doesn't have proper mechanism to read the satellite image in pixel wise, because of salt and pepper effect [6]. As well as many change detection studies consist with visual interpretation, extraction and labeling analysis to detect the change for all types of forest. The proposed methodology can be assessed forest cover change independently.

1.5 Research Contribution

Research question is generating automation process to measure the size of the change area in Wilpattu forest in years of 2000, 2013 and 2018. Satellite image contains matrices of numbers. Therefore, generate an algorithm to read matrices of satellite image which contain forest according to below categories using 3 bands.

According the green value, neural network model is developed with K-means Clustering algorithm to identify sea water and island cover. Thereafter a model is developed to separate only forest cover from other land cover using red band of pixel values. Then two data mining models are generates to analyze the size of forest cover using K-Means clustering algorithm with Artificial Neural Network. Through this research, the researcher generates a new automation process to read the satellite image and output the percentage that mentions the forest change detection.

1.6 Scope

Wilpattu forest area is taken as the study area to measure the forest cover change. Satellite images were taken from USGS website in years of 2000, 2013 and 2018. One pixel of satellite image contains 30 m x 30 m resolution in ground and pre-processed. Those images were imported into MATLAB as two dimensional matrixes. According to the imported image three bands were separated into clusters using K-means algorithm. Forest cover change detection can be classified using the pixel wise pixel based method. Because of the limited memory capacity and longer time to simulate the model, satellite images were restricted into 100*100 pixels, in order to model the forest cover change. This research is strictly based on the RGB values of the satellite images.

Chapter 2 – Related Work

2.1 Introduction about Artificial Neural Network

Artificial neural Network (ANN) process works as a biological nervous system that is similar to how brain processes the information [7]. Or simply we can say that working of neural network (NN) is similar to the function of human brain. It is made up of program systems and data structures. It is built by large number elements, named neurons to corporate each other to solve intended problem. Usually people learn by experience or examples. Similarly NN learn by examples. It trained by large amounts of data and rules about their relationships. Then a program can tell the network how to behave to a new response or can initiate activity on its own. Pattern recognition and data classification are the applications of artificial network.

NNs are largely used because its ability to learn quickly. It figures out how to achieve their functions by their own. Functions based only by sample inputs. It also has the ability to generalize to the situation. In other words it has the capacity to create reasonable outputs for given inputs which are not trained.

Applications of NNs frequently deal with supervised learning. For supervised learning, training data follows the input and target values as the user mention. When the training session was finished, inputs which don't have desired output will generate approximate desired results. But it required many training data and takes much computer time for training to be accomplishment.

2.2 Historical Background

McCulloch and Pitts(McCulloch & Pitts,1943) first designed a NN model based on neurology. For these models they consider the assumptions of the functioning of neurons. In 1958, Rosenblatt had an interest on NN and he designed the Perceptron. Again in 1960, the system ADALINE (ADAptive LInear Element) was developed by Widrow and Hoff.

Table (Table 2.1) will illustrated the development in NNs from early 1940s to recent past.

Durtion and Contributors	Contributions
Pre-1940:	
von Hemholtz, Mach, Pavlov	General theories of learning, vision, conditioning. Specific mathematical models of neuron operation.
1940s:	
Donald Hebb McCulloch and Pitts	Developed a concept fundamentally essential to the ways in which humans learn. If two nerves fire at the same time, he argued, the connection between them is enhanced. Mechanism for learning in biological neurons.Neural-like networks can compute any arithmetic function
1950s:	

Rosenblatt	First practical networks and learning rules
1960s:	
Minsky and Paper Widrow and Hoff	Demonstrated limitations of existing neural networks. Developed ADAptive LInear Element (ADALINE)
1970s:	
A. Henry Klopf Paul Werbos	Developed a basis for learning in artificial neurons based on a biological principle for neuronal learning called heterostasis Developed and used the back-propagation learning method
1980s :	
John Hopfield of Caltech	Presented a paper to create more useful machines by using bidirectional lines. (Previously, the connections between neurons were only one way).

1990s :	
Broomhead and Lowe	Developed a sub field of neural Networks; Radial bias function
Recent past :	
Amari Grossberg Hopfield, Kohonen von der Malsburg Willshaw	Contributions to the field of neural networks; such as in self-organizing maps (SOMs) and in auto associative memories.

Table 2.1 : development in Neural Networks from early 1940s to recent past

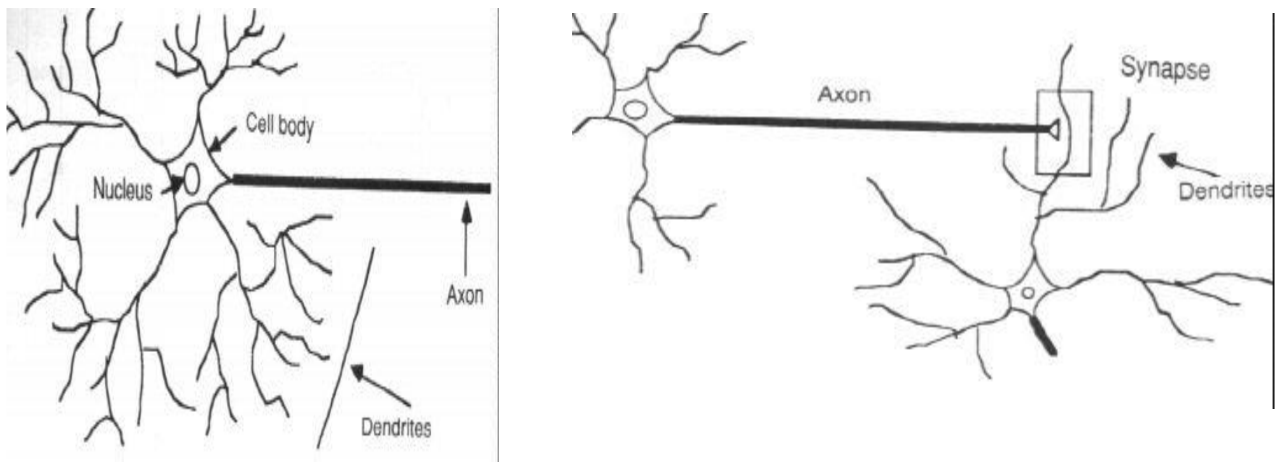
2.3 Human brain and biological Neurons

Human brain and biological neurons are based on ANNs. Neurons are considered as the fundamental unit in the nervous system.

A neuron has:

- Dendrites - Accept inputs
- Cell body or Soma - Process the inputs
- Axon - Turned the processed inputs into outputs
- Synapse (Learning) - The electrochemical contact between Neurons

Dendrite is a branched extension of a neuron that collects the signal from other neurons. Neuron sends spikes in electrical activities over an axon that separates into many branches. The axon end is almost in contact with dendrites of next neuron [8]. Neurotransmitters transfer impulse to another nerve and the link is called as a synapse. Following figure illustrates the Neuron architecture.



A NEURON THE SYNAPSE

Figure 2.1 : The Biological Structure of a Neuron

2.4 From biological Neurons to Artificial Neurons

Human brain is similar as the ANN. It is a function approximates that converts input to its output to its optimal ability. This means that we use much simpler, abstract "neurons", which capture the essence of neural computation even if they leave out much of the details of how biological neurons work. Artificial neurons simulate the tasks of the natural neurons. There are mainly four tasks which are Dendrites, Soma, Axon and Synapse.

An artificial neuron contains voluminous inputs and provides an output [7]. The neuron has two phases of operation, which are training and using phase. During the training phase, the neuron can be taught some input patterns. If the input pattern does not belong to list of input patterns, the Network will generate the output by its own.

Following figure (Figure 2.2) shows a basic artificial neuron.

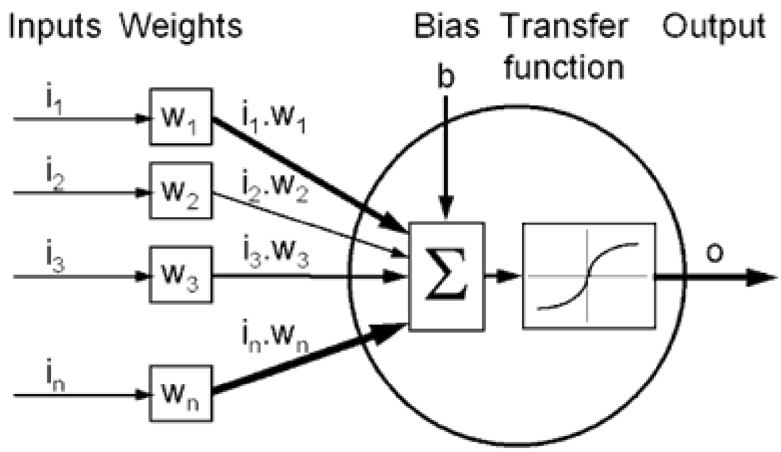


Figure 2.2 : A Basic Artificial Neuron

The above figure (Figure 2.2), many inputs are denoted by $i_{(n)}$. These weights are multiplied by its corresponding weights $w_{(n)}$. These products are summed up and then add a bias part to it. Then it fed through a transfer function which gives the desired output. Input weight plus bias is the output of a neuron. It means weighted sum plus bias of all neurons decide the manners of neural network. The weights and bias to a given problem is the main challenge while working with Neural Networks. Different architectures are used to find these weights and bias.

2.5 Components of a Neural Network

2.5.1 Network Layers

NN contains three layers of units. Those layers are Input layer, hidden layer and output layer.

Input layer	-	Raw information that served into Neural Network
Hidden Layers	-	Connect the information between input and output layers
Output layer	-	Outputs the computational results

2.5.2 Weights and bias

Many real-time inputs are received by neurons. Every input has their own weight. These weights have the same function of biological neurons [9]. Greater effort was made for some inputs than others to generate neural response.

2.5.3 Transfer function

Transfer function is the process of conversion from resulted weighted sum into a working output using an algorithmic process. Actually it is applied to the weighted sum of the inputs of a neuron to generate the output. In the transfer function the summation total can be compared with some threshold to determine the neural output. If the sum is greater than the threshold value, the processing element generates a signal [10]. If the sum of the input and weight products is less than the threshold, another signal is generated. Both types of response are significant.

2.6 How the Neural Network Learns

Output of network will generate according to their connection, weight and the given input. The network implements various functions by changing the weights while connection structure keeps as constant. However weights are changed according to objectives of the network and information available on the learning rule. There are two key learning algorithms:

1. Supervise learning
2. Unsupervised learning

2.6.1 Supervised Learning

Many neural networks trained with the supervision. It follows by, input and output are presented at the network. It means all input deliver to the network and desired output also delivered. This training is considered complete when the NN reaches a user defined performance level. This level defines that the network has achieved the desired statistical accuracy as it produces the required outputs for a given sequence of inputs.

However it needed to have large amount of data to feed the network. With bulk of data network can understand the features and relationships. Furthermore the way of representing output data is also major component. Usually numeric data is using for AN’s and because of that raw data is obtained from the environment by converting.

2.6.2 Unsupervised learning

Unsupervised learning is sometimes called as self-supervised learning. In this learning we only supply the inputs. In order to create outputs similar to the inputs network is adjusting its weights. To adjust the weights usually outer influences are not used. There are regularities in input signals and then adaptations are made affording to the function of the network. Without saying correct or incorrect network have its own way to organize information.

2.7 Neural Network Architectures

Based on their connection within layers NNs can be classified in to two groups.

- Feed-forward Networks (FFN)
- Recurrent (or Feedback) Networks

2.7.1 Feed – Forward Neural Networks

Feed forward neural networks are the most popular and most widely used models in many practical applications. They are known by many different names, such as "multi-layer perceptrons." Here, layers are organized using neurons. These layers have unidirectional connections in between. These unidirectional connection streams to one direction. The output of a layer performs as a input to another layer. FFN are static, that is, produce one output value set other than sequence of values according to given input.

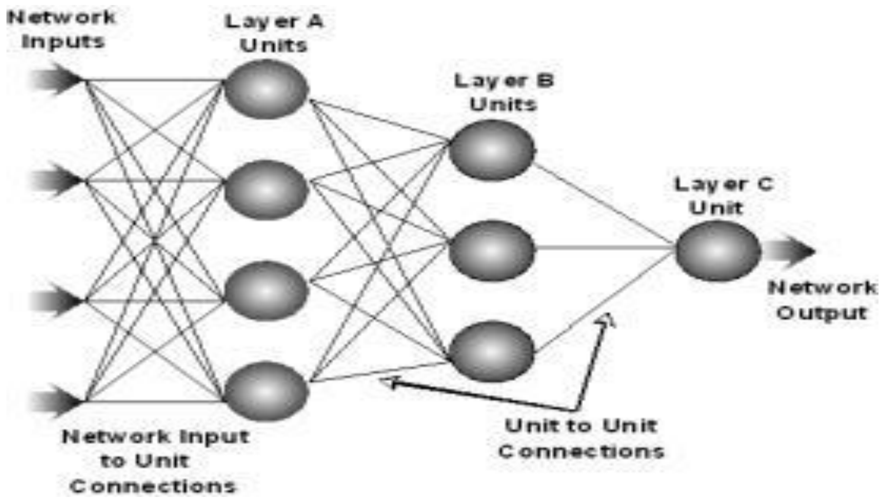


Figure 2.3 : A Feed Forward Neural Network

2.7.2 Recurrent (or Feedback) Networks

Recurrent networks can have signals travelling in both directions by introducing loops in the network. Recurrent networks are very powerful and can get extremely complicated. Recurrent networks are dynamic systems. Their state is changing continuously until they reach an equilibrium point [12]. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Following figure illustrates a Recurrent Network.

Competitive Networks, Self Organized Map (SOM) Networks, ART Networks and Hopfield Networks are some of examples for Recurrent Networks.

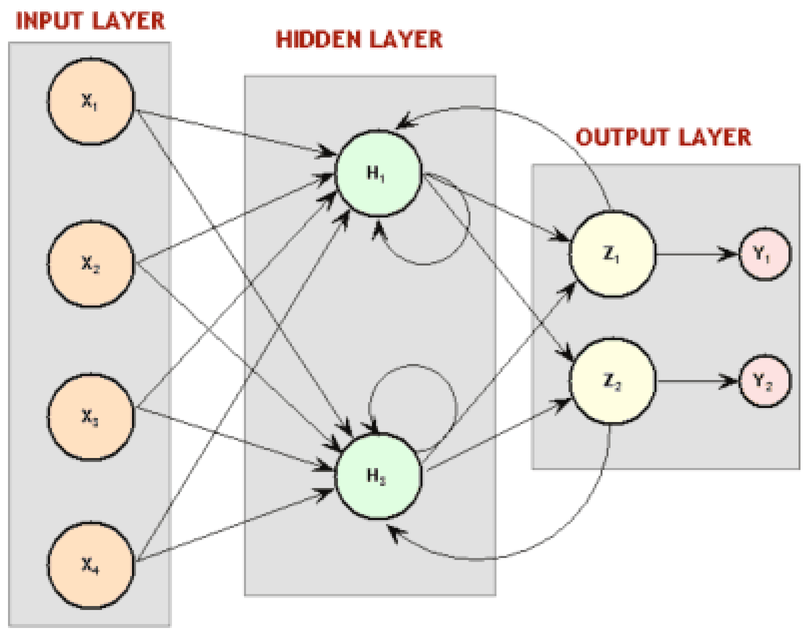


Figure 2.4 : Recurrent Neural Network

Following figure (Figure 2.5) illustrates different Neural Network architectures.

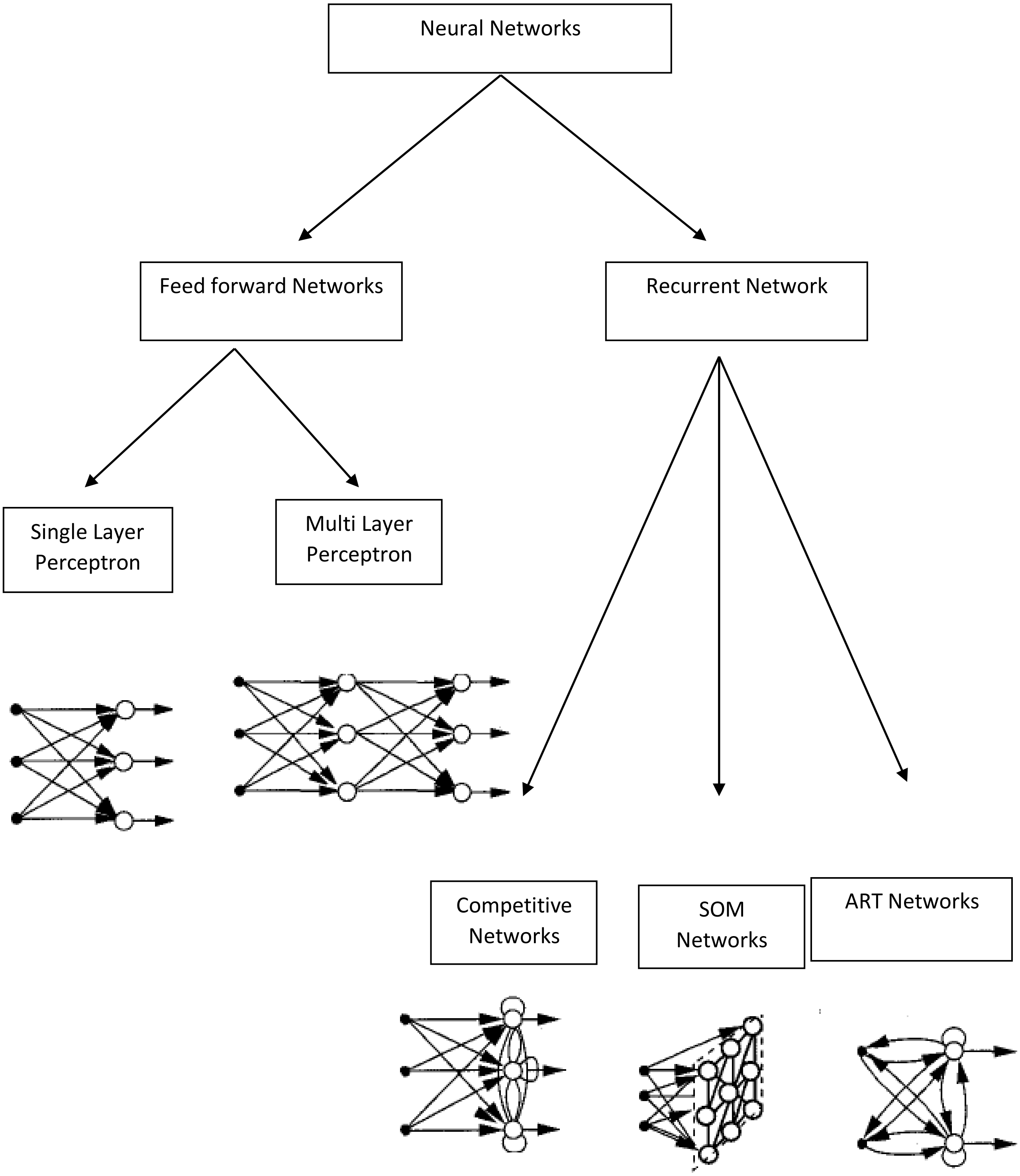


Figure 2.5 : Neural Network Architectures

2.8 Training a Neural Network – Back propagation Algorithm

Each unit has its weights and these weights have to be adjusted. The main intention is reduce the error between desired output and actual output. By performing this neural network can be trained. Back propagation is one of the methods that used to minimize the network error. Actually this back propagation algorithm is based on minimizing the error of the network using the derivatives of the error function. Calculation of derivatives flows backward through the network; hence the name back propagation appears. These derivatives point in the direction of the maximum increase of the error function. To obtain the weights in the network this is mainly used. Back propagation is a form of supervised learning.

2.9 Neural Network Applications

Neural Networks can be applied to wide verity of problems; from pattern recognition to classification of satellite imagery. Given this description of neural networks and how they work, finding what the real world applications of Neural Networks are will be interesting. Following table will briefly summarize the real world application of Neural Networks.

Areas	Applications
Science	Pattern Recognition Speech Recognition Recognizing Genes Recipes and Chemical Formulation optimization Chemical Compound Identification Physical System Modeling Ecosystem Evaluation Polymer Identification Signal Processing: Neural Filtering Biological Systems Analysis Ground Level Ozone Prognosis Odor Analysis and Identification
Finance	Stock market predictions Credit worthiness Credit ratings Fraud Detection Price Forecasts Economic Indicator Forecasts
Medicine	Medical diagnosis Detection and Evaluation of medical phenomena Patient's length of stay forecasts Treatment cost estimation
Industrial	Process control Quality control Temperature and force prediction

Data mining	Prediction Classification Change and deviation detection Time series analysis
Sales and Marketing	Sales forecasting Targeted Marketing Retail margins Forecasting
HR Management	Employee selection and hiring Employee Retention Staff scheduling Personnel Profiling
Others	Game developments Agricultural Production Estimates Sports Betting

Table 2.2 : Application of Neural Network

2.10 Forest Recognition and Neural Network

In this study, artificial neural networks have been used to predict the deforestation using satellite imagery. In the field officers from forest department faced big challenge to measure forest cover change due to environment factors such as climatic change. Therefore, they use neural network to accurately predict the forest that was destroyed.

They modeled process of deforestation with the help of various factors in defining the relationship between deforestation and environmental and socioeconomic factor. The Satellite imagery captured from 2012 to 2016 of region in Hong Kong. A simple spatial model is developed that can predict deforestation using artificial neural networks [13], [14], [15].

The Altitude, forest cover distance and proximity to residential area were selected as input to network. The data were separated as training set and test set. Levenberg-Marquadt and Back-propagation algorithm were tested in training process. Then network was designed with input layers, hidden layers and output layer

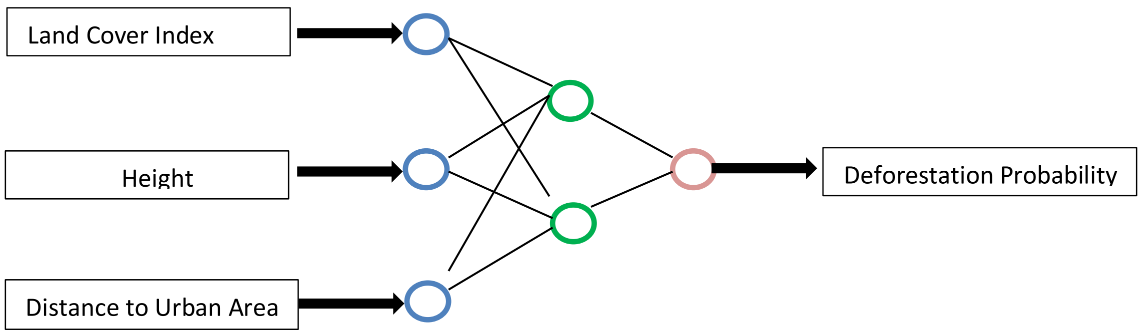


Figure 2.6 : Schematic diagram of the Network

According to above research paper [16] images were taken ArcGIS environment and Classified images into forest, urban areas and sea. Output contains one and zero. Zero identified as forest and one identified as non-forest. Then it helped to generate forest cover index. Second line of input to network is distance and proximity to the cities, according to that identifies the urban and non-urban area. Then produce the layer for the network as distance from urban areas. Last entry as the third line of the vector input is the height.

According to given input data training data must be implemented. The data which are not involved to the training process was used to test the network. The result of this stage is map separated into three classes, class one call areas without deforestation, the second class relays to areas relatively prone to deforestation and to some level of sustainable. Third class related to area which contains high risk of deforestation. Therefore, in above research, accurately predict the forest which was destroyed.

The environmentalists and managers can be used this model to develop targeted policies to control the ecological and social influences of deforestation.

Zhanhui Lin, uses a deep neural networks model to predicted forest cover type. Dataset has taken from UCI Machine learning Repository and seven types of forest cover type classes were used in this study.

According to their study three mutually exclusive and distinct dataset were used to train, validate and test the predictive model. Data set size are shown in below Table 2.3

Training Dataset Size	Validation Dataset Size	Test Dataset Size
348607	116202	116203

Table 2.3 : Number of observations for each data set

The back-propagation algorithm finally discovers a weights to be estimated that minimizes Loss function(MSE). Below given the Equation of Loss function

$$E(w) = \frac{1}{2N} \sum_{n=1}^{N} \sum_{i=1}^{k} \left(d_i(n) - y_i(n)\right)^2$$ [17]

E(w) - mean square error term

w - synaptic weights

N - number of observation (input) vectors

n - single observation vector

k - number of output nodes

i - single output node

$d_i(n)$ - observed response

$y_i(n)$ - predicted response

Train the model until a minimum 100 epochs had been accomplished and Loss function was reached 0.05. For the training process two hidden layers were used and number of nodes assign to these hidden layers were changed according to 8 possible values to keep the learning rate as constant.

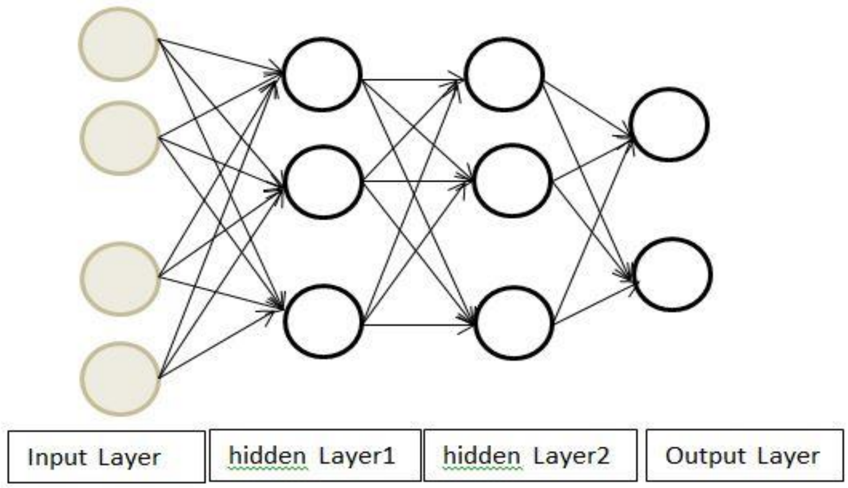


Figure 2.7 : Construction of the artificial neuron network

MLP predictions of forest covertype was generated classification accuracy of 78.5221% and XGBoost algoritm model was generated 87.0029% for classification accuracy [17].

In the study of Gail A. Carpenter, Sucharita Gopal, Byron M. Shock, and Curtis E. Woodcock , implemented a model for Land use Change Classification using Neural Network method in Nile River delta region. Images were taken between 1984 and 1993 at different time.

fuzzy ARTMAP network used in the above study and it takes less training time [18]. Large scale dataset identified with minimum training time and the process is autonomous. Using above network, change and non-change pixels were identified. But this change detection classification method contains certain types of defects. Classification method deals with high-dimensional images with different dates.

Chapter 3 – Methodology

3.1 Overview

This chapter explains design and implementation process of the forest recognition system. MATLAB is the main language used in this study to recognize the forest. The main reason is that it consisted of toolbox for matrix manipulation. Tool Box and its easiness of matrix manipulation. Forest recognition consists of following steps.

Forest Recognition process

- Data gathering
- Image Preprocessing
- Import images to MATLAB
- Construct the forest vector
- Apply K mean clustering to forest vector
- Forest Classification
- Construct the Neural Network
- Train the Neural Network
- Testing and simulation

3.2 Data Gathering

Landsat satellite images were taken from USGS Earth Explorer site. The satellite image typically consists of pixel elements having a DN located in the intersection of each row i and column j and consists of k bands. A smaller DN number indicates lower radiation in the area, while a larger number indicates high radiation characteristics in the area.

The main source of the research was Landsat data. Landsat contains visible spectrum that the human can perceive like blue, green and red light as well as near-infrared, mid-infrared and thermal-infrared that human cannot perceive.

Date	Landsat type	Cloud cover
2000	Landsat 7	18%
2002	Landsat 7	15%
2007	Landsat 5	5%
2015	Landsat 8	5%
2018	Landsat 8	5%

Table 3.1 : Details of Satellite Images

3.2 Image preprocessing

Image usually contains various types of errors. Therefore, preprocessing should be conduct to manipulate and process the image data. The USGS earth explorer provide satellite images with level 1 correction and for this analysis, images on cloud less days were selected.

Landsat Level-1 products are usually provided with after processing. For this best methods are used by scene. For each scene it has different processing levels. These levels are basically depending on various factors. Those factors are ground control points (GCP), elevation data provided by a Digital Elevation Model (DEM), and/or data collected by the spacecraft and sensor (Payload Correction Data (PCD) [19].

The processing levels Landsat Collection 1 Level-1 data are shown in Table 3.2 below.

Landsat Level-1 Processing Levels	
Processing Level	**Description**
Standard Terrain Correction L1TP	Radiometrically calibrated and orthorectified using ground control points and digital elevation model (DEM) data to correct for relief displacement. These are the highest quality Level-1 products suitable for pixel-level time series analysis.
Systematic Terrain Correction L1GT	Radiometrically calibrated and with systematic geometric corrections applied using the spacecraft ephemeris data and DEM data to correct for relief displacement.
Systematic Correction L1GS	Radiometrically calibrated and with only systematic geometric corrections applied using the spacecraft ephemeris data.

Table 3.2 : Landsat Collection 1 Level-1 Processing Levels

The Standard Terrain Correction (L1TP) and Systematic Terrain Correction (L1GT) products incorporate a Digital Elevation Model (DEM) for topographic accuracy. The Standard Terrain Correction (L1TP) products also utilize ground control points for additional geometric accuracy.

- Ground control points used for Standard Terrain Correction (L1TP) correction are derived from the GLS2000 data set. DEM sources include SRTM, NED, CDED, DTED, GTOPO 30, and GIMP.
- Systematic Terrain Correction (L1GT) is the highest level of correction possible for L7 and L8 scenes collected over Antarctica. The Ramp V2 DEM is the elevation correction source.

3.3 Soft wares used

In this study, MATLAB is used as the main software which is a full computer algebra system with a numerical computing environment. It allows for simple matrix manipulation, plotting functions and interaction with other computer languages.

MATLAB has a special feature called the 'Neural Network toolbox'. It allows the design, implementation, visualization and simulation of neural networks. Since this project use Neural Network as the forest classifier.

3.4 Construct the forest vector

Satellite image has three bands which are RED, GREEN and BLUE. Among those most important in the forest cover discovery problem is the green color. In this study GREEN and RED were selected to build the forest vector for the NN.

3.5 Clustered feature Vector using K-Means Clustering

K Means clustering is the unsupervised learning algorithm that makes suggestions from dataset without knowing outcomes. Main objective of that algorithm is find groups in the dataset which are more similar. It reveals relationships that were not previously undetected in data item.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$ [20]

J- Objective Function

k- Number of Clusters

n- Number of cases

x_i - Case i

c_j - Centroid for cluster j

$\left\| x_i^{(j)} - c_j \right\|$ - Distance Function

Steps of algorithm

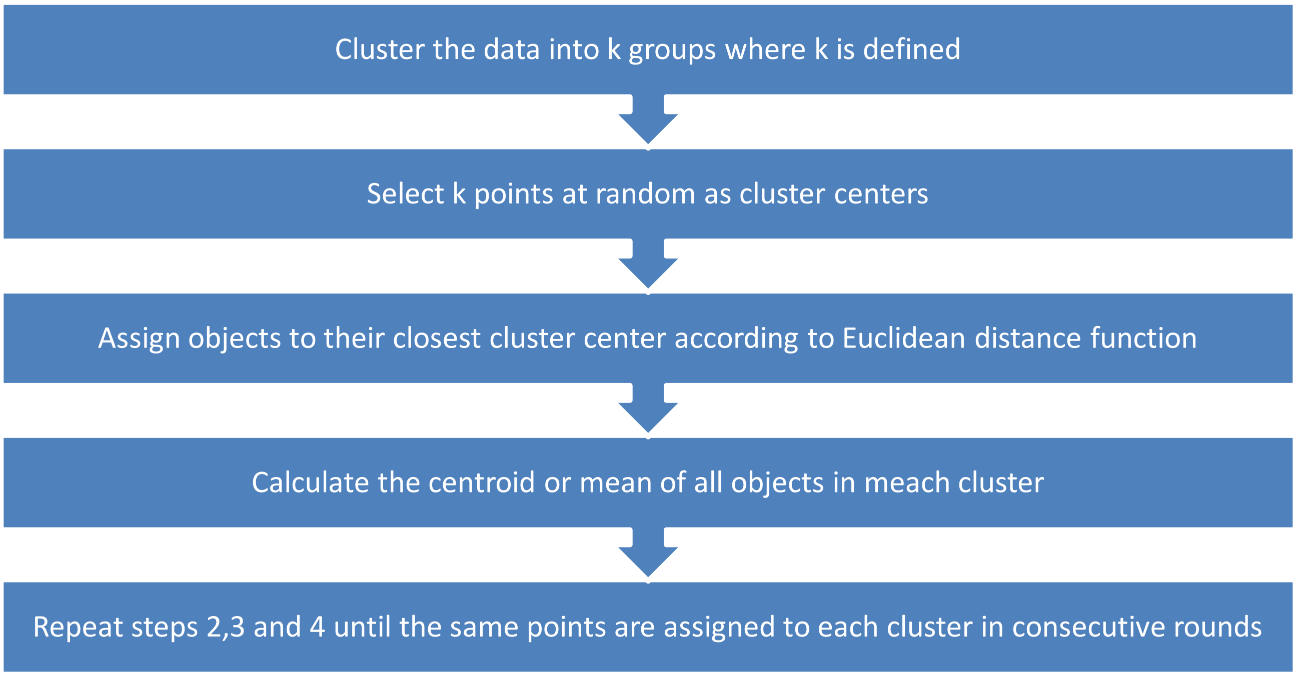


Figure 3.1 : Steps of K-means clustering algorithm

3.5.1 Clustering using RGB satellite image

Steps of applying K-means for a satellite image

Step 01: Green values of Satellite images were applied into KM algorithm and separated into two groups. Forest area is identified as group with Maximum values of green pixels. Other group of pixels was assigned black color. Sea water (non-forest area) has been removed from the original image and isolated Land and forest cover area. This was shown in Projected Image. But output image contains both land cover and forest cover as forest area.

Figure 3.2 : Input image that has applied k-means algorithm for the first time

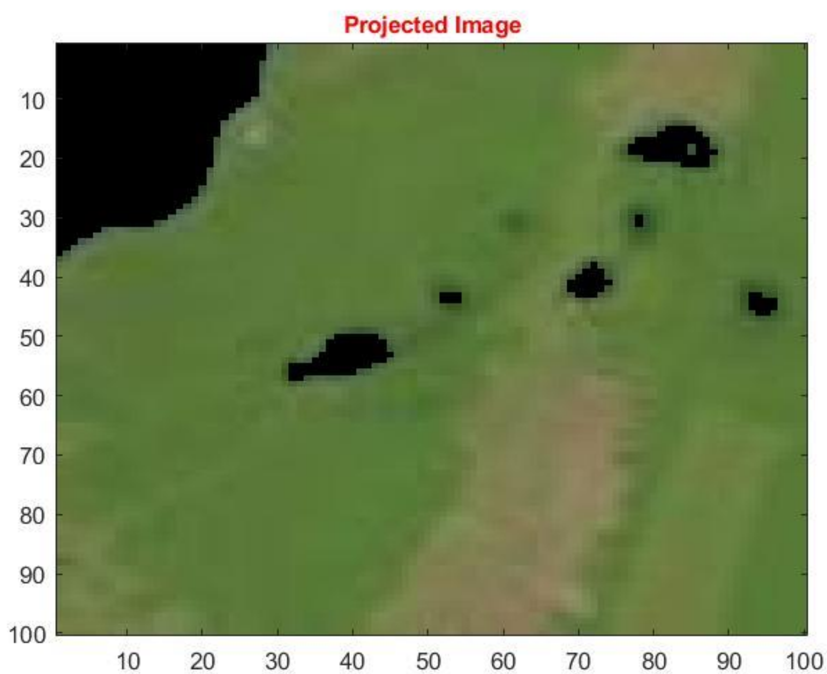


Figure 3.3 : Output image gained after applying k-means algorithm for the first time

Step 02: In order to separate forest area from previously obtained output in step 01(Figure 3.3), K-means clustering is applied and separated into three clusters using red values of the pixels. Land cover has been removed from previously obtained output (Figure 3.4) and isolated forest cover area (Figure 3.5).

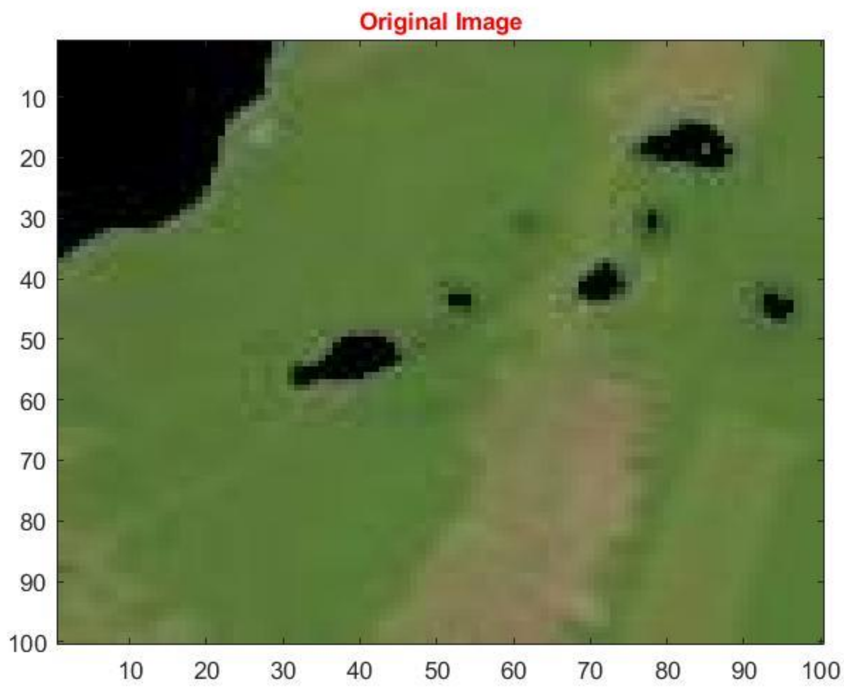


Figure 3.4 : Input image that has applied k-means algorithm for the second time

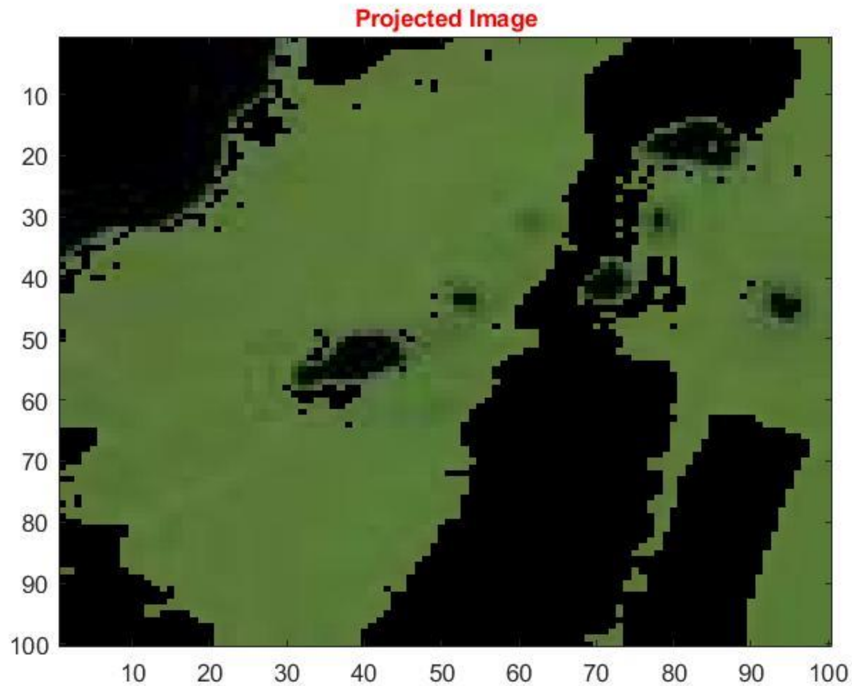


Figure 3.5 : Output image gained after applying k-means algorithm for the second time

3.5.2 Clustering using HUE value of satellite image

HUE values of Satellite image varies from 0 to 1. Because of the small variance unable to process the clustering algorithm. Therefore multiply by 255 to gain the (0-255) range of value to applying the clustering. KM algorithm was applied to resulted value and separated into six groups. Mean value obtained from K-means clustering was sorted. Then 3rd group was identified as forest cover. Reasons behind the selection of group mention in Classification section.

3.6 Forest Classification

3.6.1 Forest classification using RGB value

In this study, forest classification was done by using mean values of clusters. The highest mean values contain cluster for green vector which is named as forest (forest and land) and other group named as non-forest (water class). Figure (Figure 3.6) shows the first classification process. For an example in this study Figure 3.7 and Figure 3.8 were shows the green values of image and there 120 is selected as forest area and 34 was selected as water area.

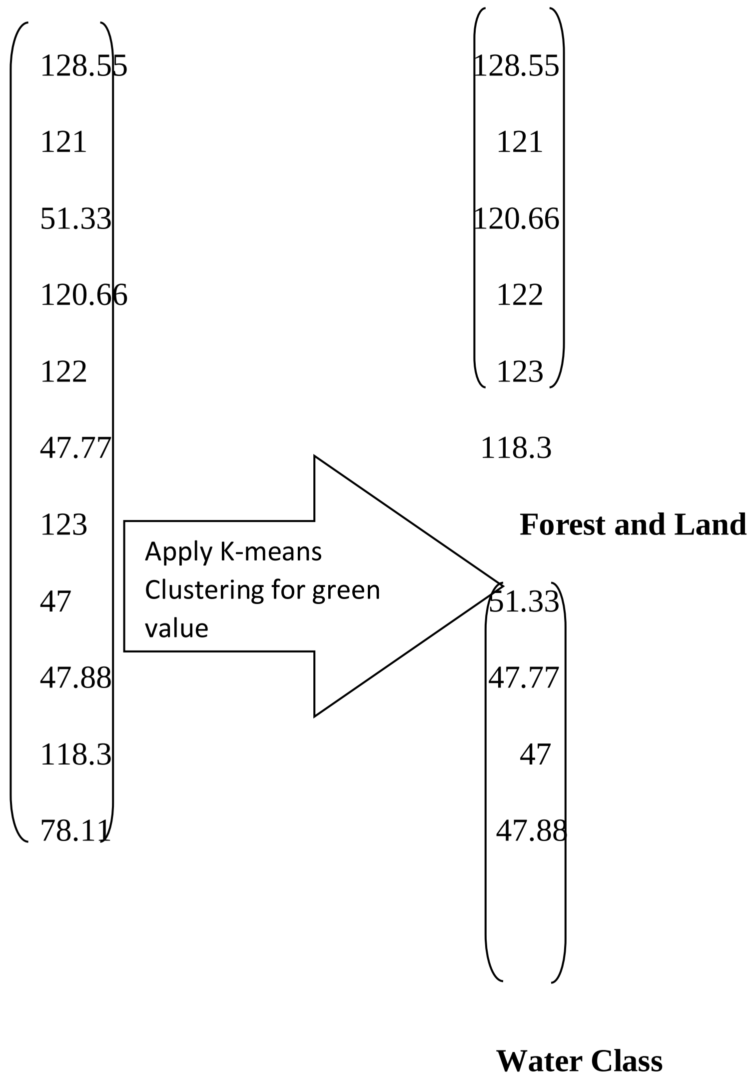


Figure 3.6 : Results when applying K-means clustering algorithm for green vector

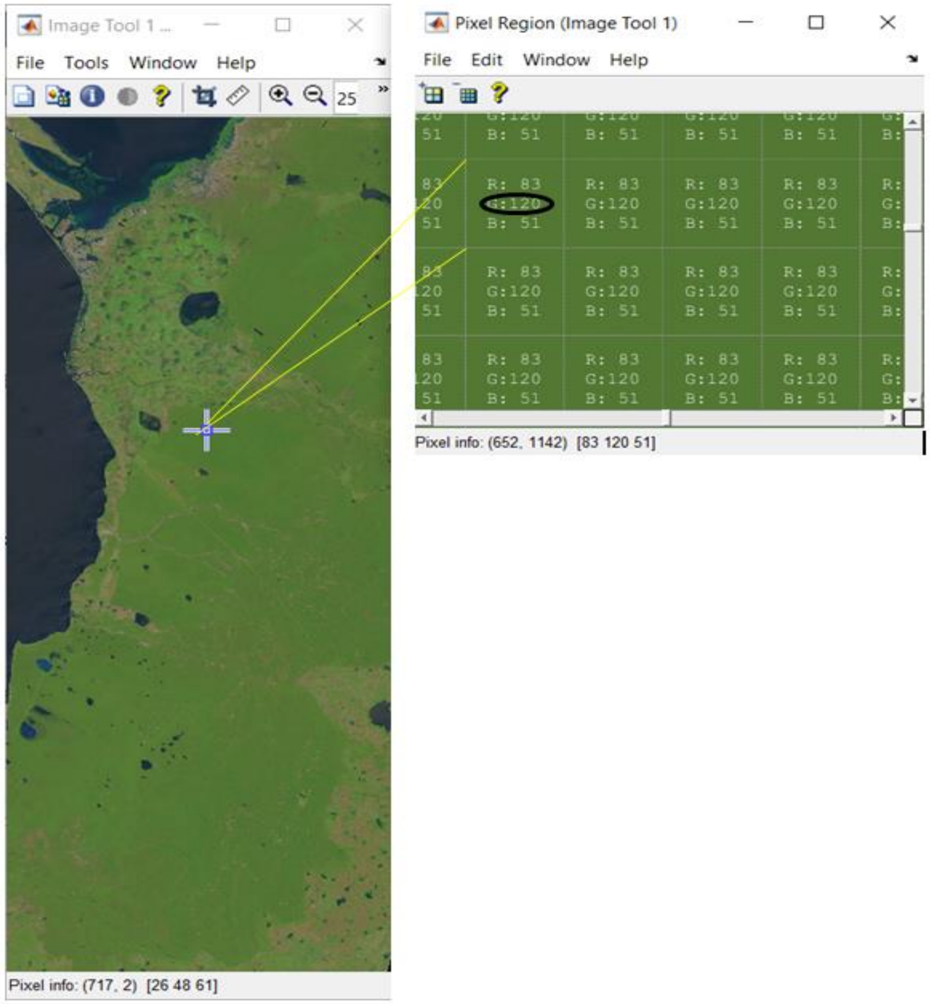


Figure 3.7 : Green value of forest class

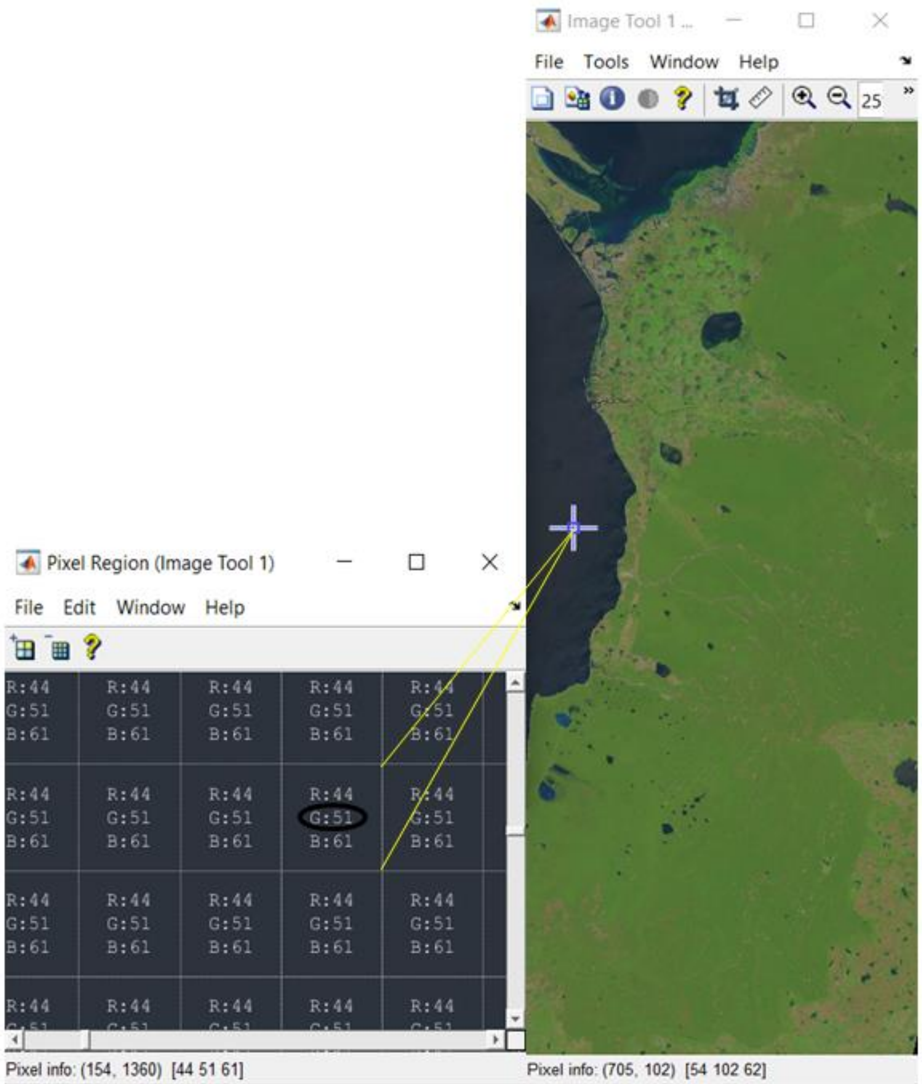


Figure 3.8 : Green value of water class

Above figures shows the green value of forest class is larger than the green value of water class. First classification technique based on that. Second classification is based on red values of the satellite images. Red values of Land and Forest classes as shown in Figure 3.9 and Figure 3.10.

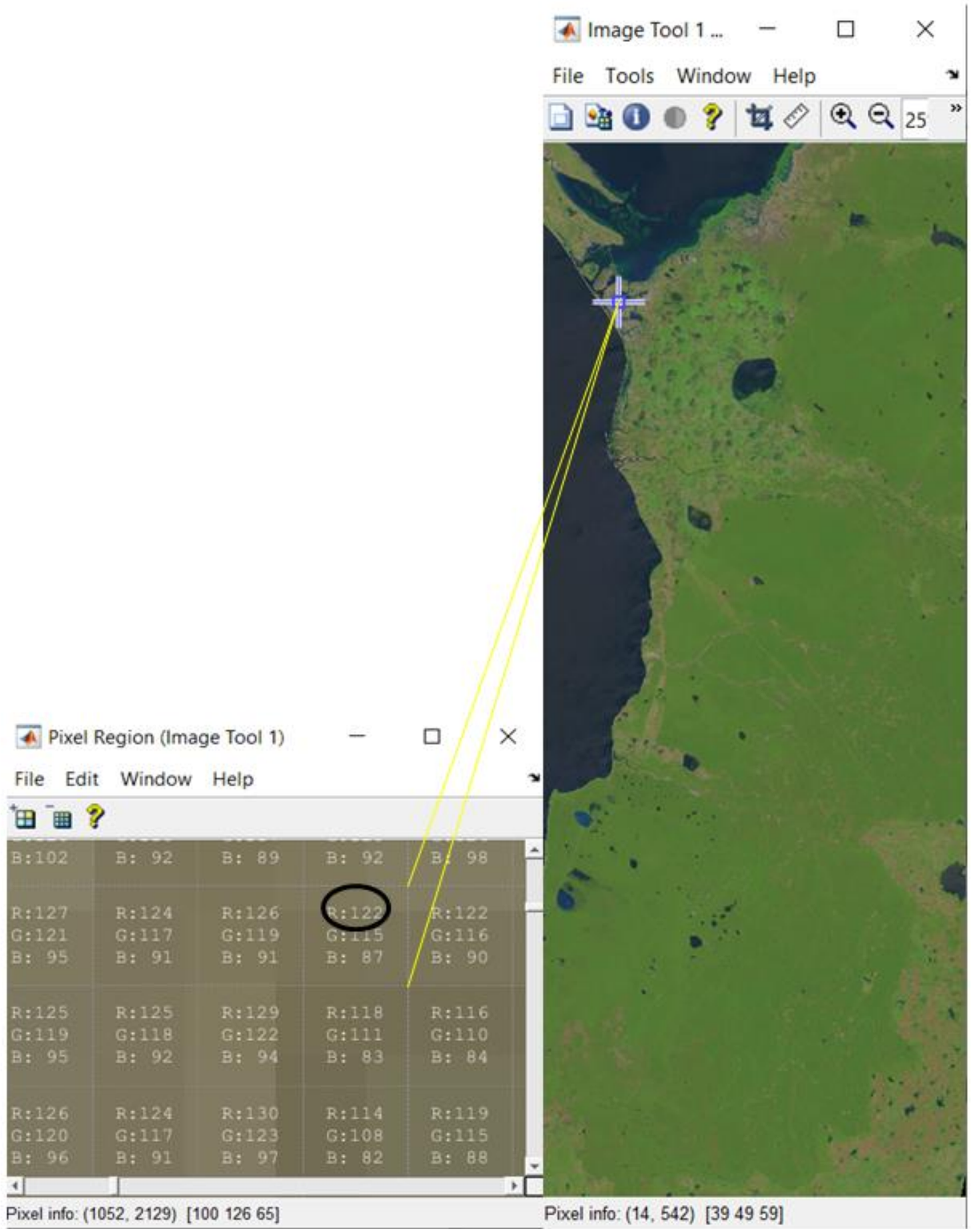


Figure 3.9 : Red value of Land class

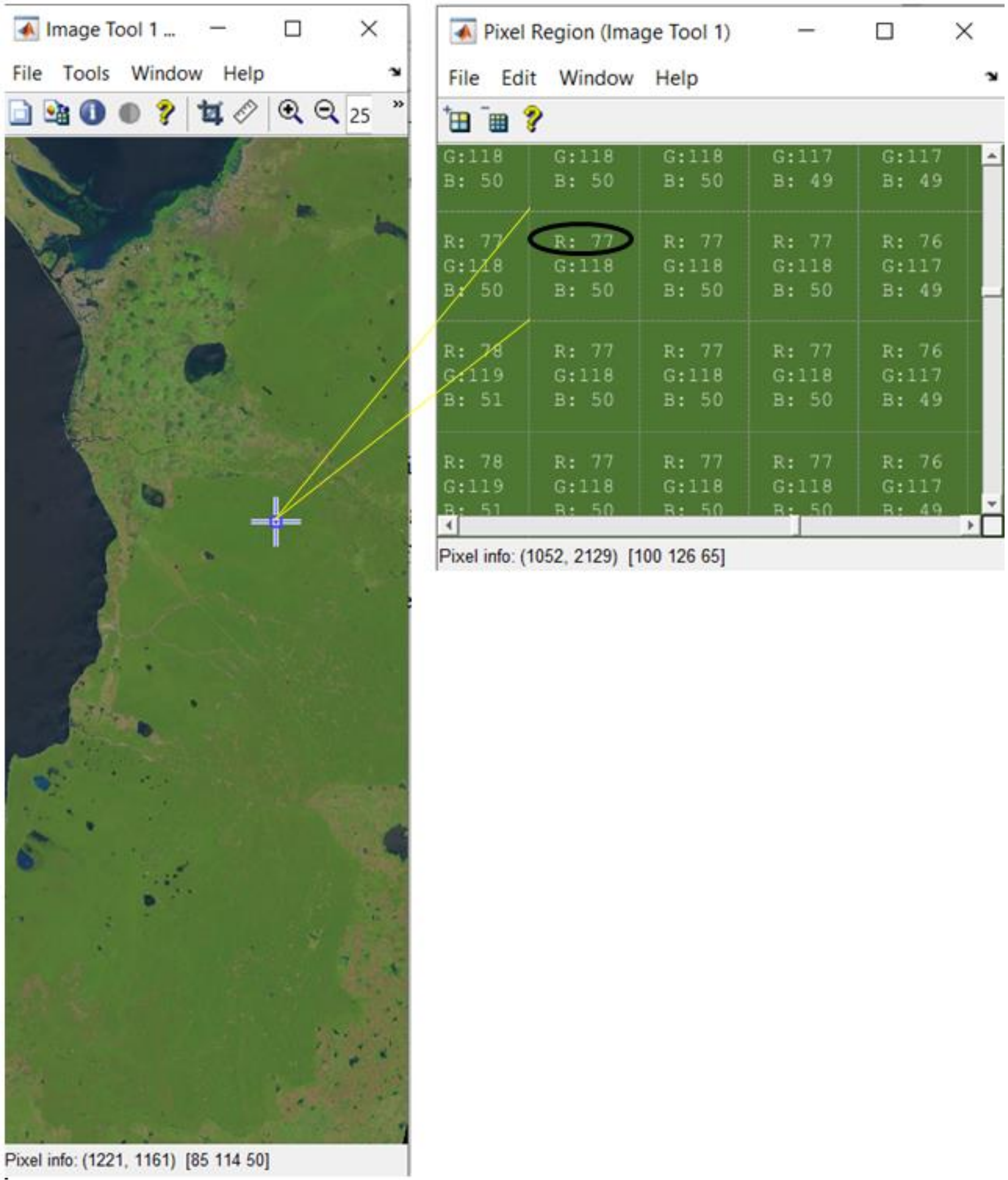


Figure 3.10 : Red value of Forest class

According to above figures red values of Land class contains high mean value when compared with forest class. Above mentioned method can be summarized as follows.

Considering Green values, separate into two clusters

Highest mean	>	Lowest mean
(Forest & Land cover)		(Water cover)

Considering Red values, separate into three classes

Highest mean	>	Middle Mean	>	Lowest Mean
(Land cover)		(Forest Cover)		(Water Cover)

3.6.2 Forest classification using HUE value

Images were cropped from satellite image where the pixel values are identified as forest. From that mean and variance of HUE value are calculated. Below table (Table 3.3) shows the results according to the selection.

Year of Satellite image	Mean (HUE value)	Variance (HUE value)
2000	0.2507 * 255 63.9285	$1.1430*10^{-5}$
2002	0.2500 * 255 63.75	$1.2430*10^{-5}$
2007	0.2358 * 255 60.129	$1.5432*10^{-5}$
2018	0.2458 * 255 62.679	$2.6817*10^{-6}$
2019	0.2551 * 255 65	$1.2450*10^{-32}$

Table 3.3 : Hue value mean and variance

According to classifying results, select the best clusters to train neural network in order to identify forest cover using HUE value.

3.7 How the Neural Network Learns

NN is used as an image classifier in this research. It can train a given satellite image as ‘forest’ or ‘non-forest. Green values and Red values of satellite images were labeled as extracted features for this study. These extracted features are given to the NN and the neural network is trained to identify forest cover. Since the input set and output set can be defined therefore training was done by supervise learning.

3.7.1 Training set

Training set consists of input matrix and output matrix. Input matrix is which columns are the feature vectors of each image. Output matrix is the column containing forest identification. Two neural networks were trained in order to identify forest cover. First figure (Figure 3.11) shows training output of identification of sea water. Second figure (Figure 3.12) shows, output of first NN taken as input to second neural network in order to identify forest.

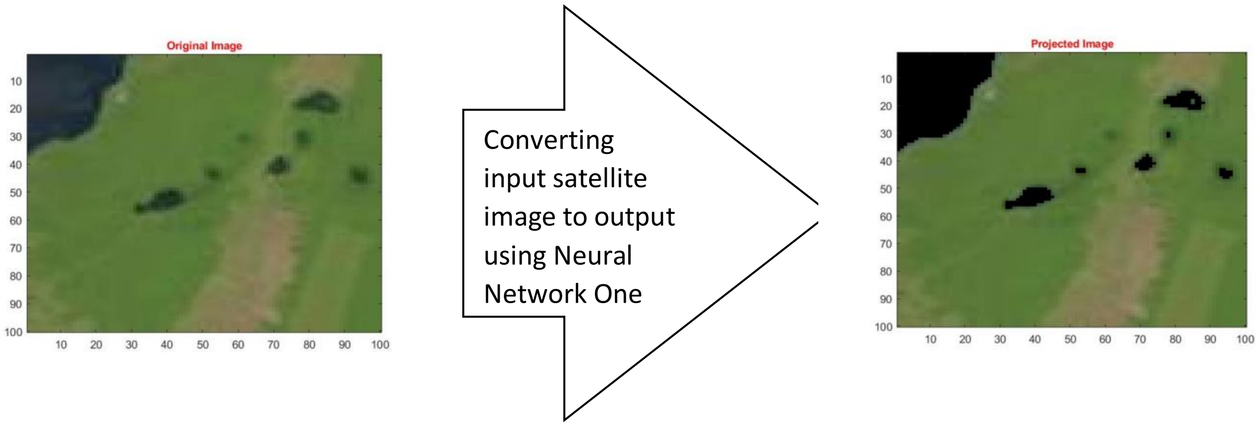


Figure 3.11 : Input and Corresponding output of first neural network

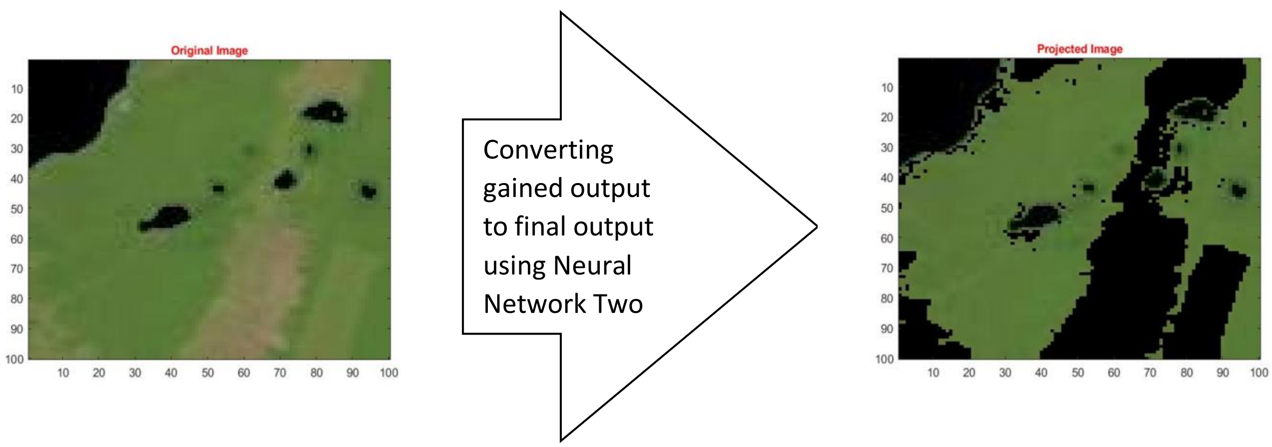


Figure 3.12 : Input and Corresponding output of second neural network

3.7.2 Training parameters

Following network parameters must be set before training the neural network.

Hidden layers = 1

numHiddenNeurons = 20;% No of neuron in the hidden layer

Division of Data for Training, Validation, Testing

net.divideParam.trainRatio = 75/100;

net.divideParam.valRatio = 10/100;

net.divideParam.testRatio = 15/100;

net.trainParam.max_fail = 1000; %Maximum validation failures

net.trainParam.epochs = 4000; %Maximum number of epochs to train

net.trainParam.goal = 0.000001; %Performance goal

net.trainParam.min_grad = 0.00000000000000000000000000000000001;

3.7.3 Train the Neural Network

After setting the above parameters, the next task is training the neural network. MATLAB command 'newpr' was used to create pattern recognition Neural Network. Input matrix, output matrix and number of hidden neurons were given as input parameters. Training will stop when Network exceeds maximum validation failures, maximum number of epochs, minimum gradient or performance goal. Following figure shows the Training state of the Neural Network.

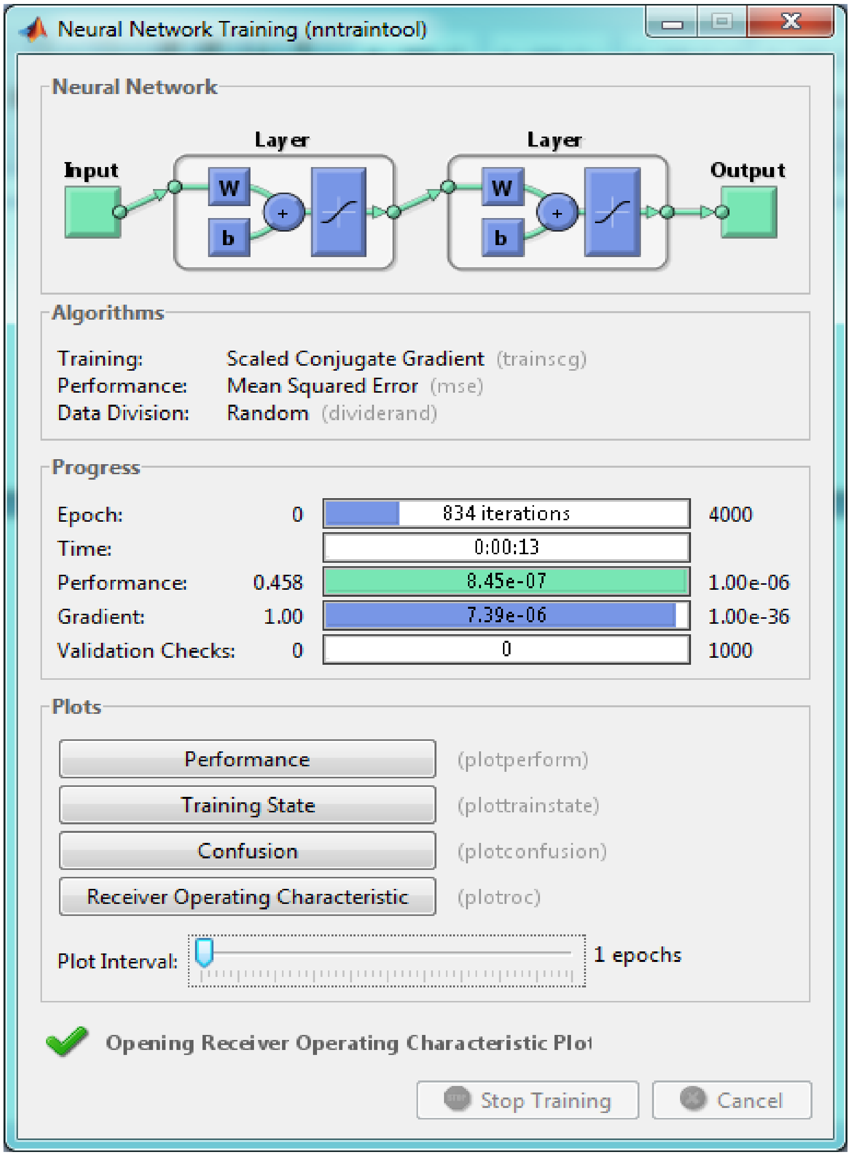


Figure 3.13 : Training state for the neural network

3.7.4 Testing and simulating the Neural Network

The next step is testing and simulating the Neural Network for new satellite images. In this step a new cropped satellite image is given to the network. Then the feature vector was extracted from that image using K-means cluster Analysis. The Neural Network is simulated using this feature vector. Output of the NN was analyzed. Following figure (Figure 3.14) illustrates this recognition procedure.

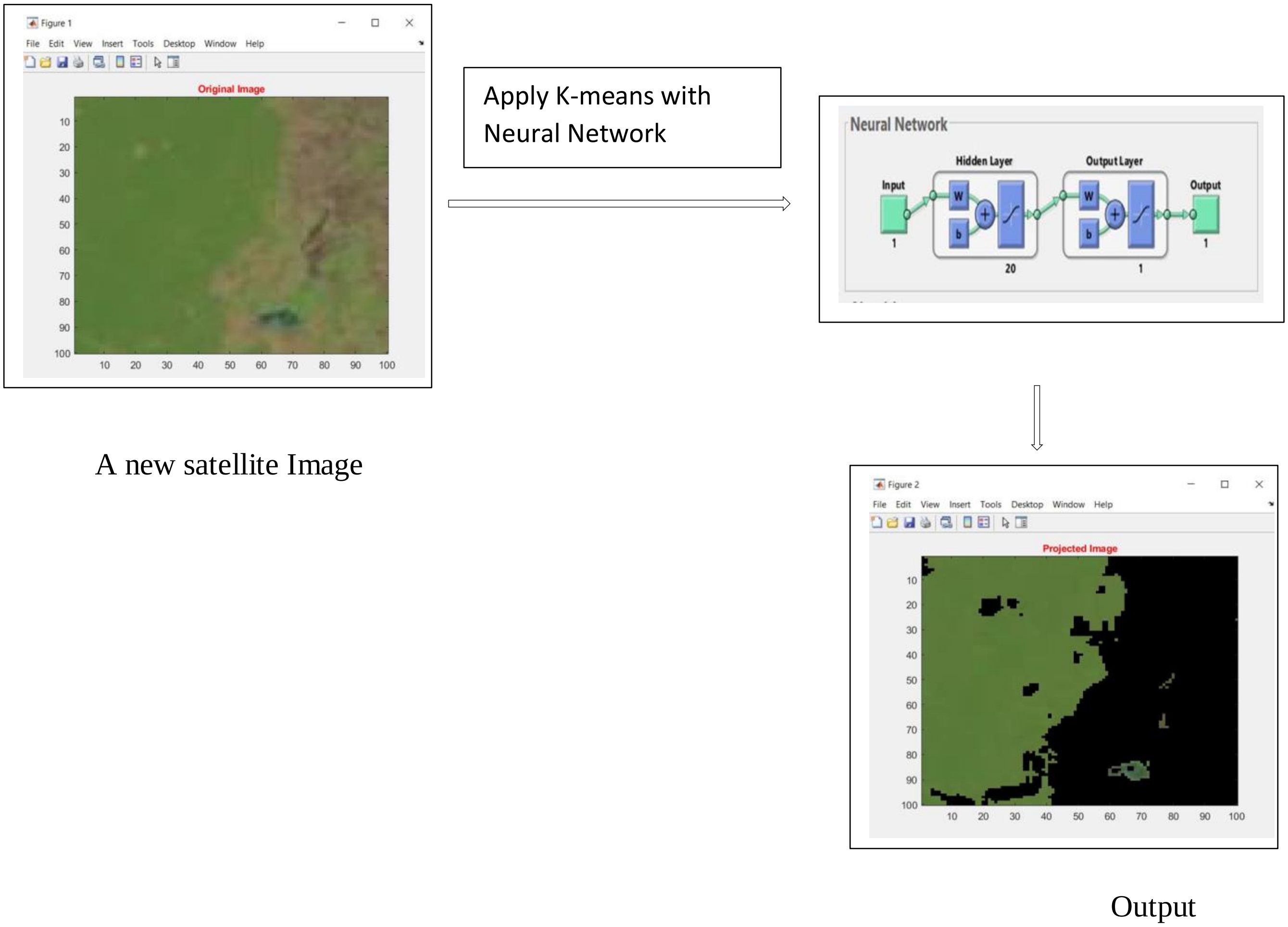


Figure 3.14 : Recognition Procedure

Model was identified the forest cover which shows in green and applied black color for other pixels.

Proposed other Method using statistical model

By using human perspective, forest, land and water are identified from samples of satellite image in different years. The sample size could be 50 * 50. Minimum value and maximum value of the forest, Land and water was analyzed statistically and defined a range value for forest, land and water classes.

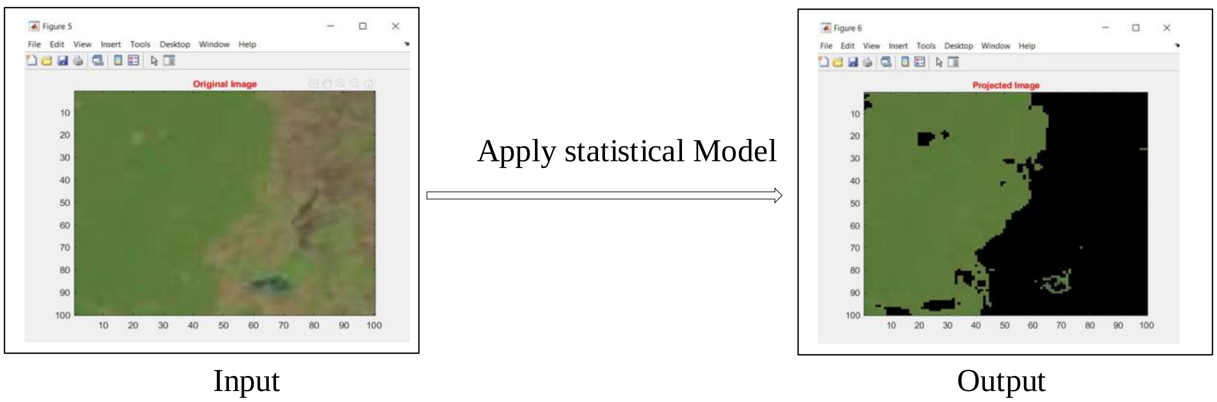


Figure 3.15 : Recognition procedure for Statistical model

Chapter 4 – Results and Discussion

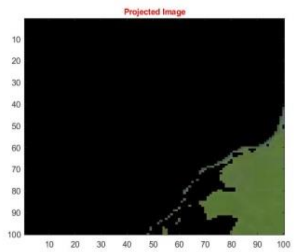
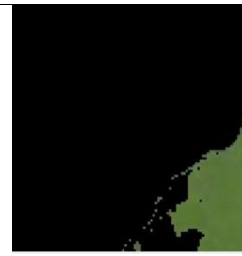
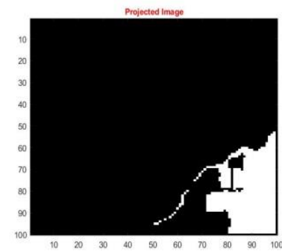
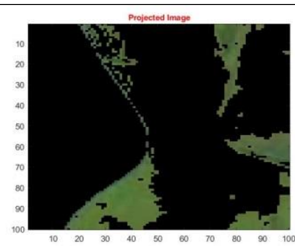
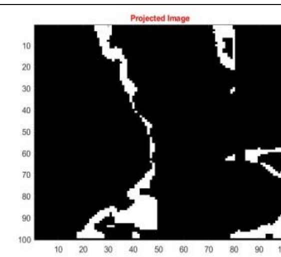
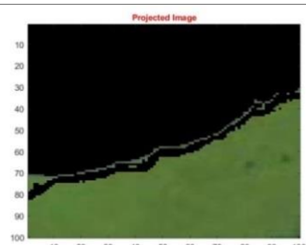
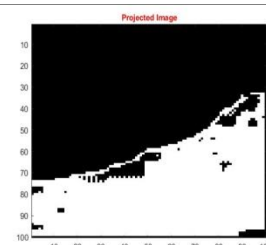
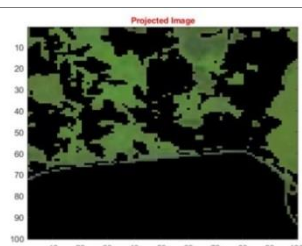
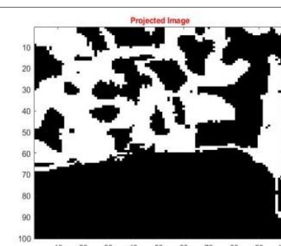
4.1 Overview

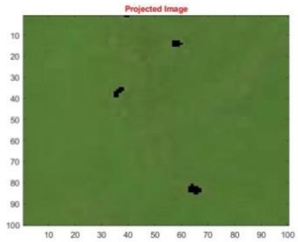
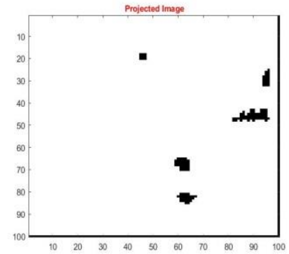
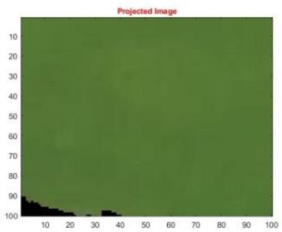
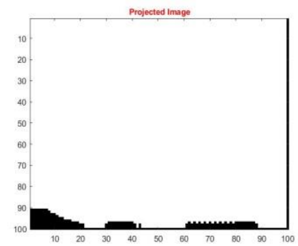
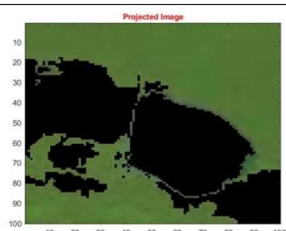
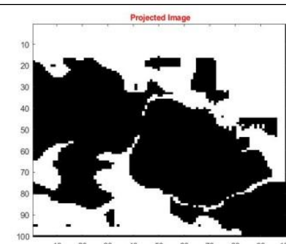
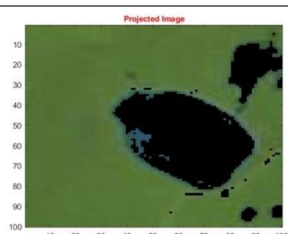
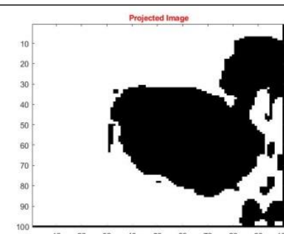
This chapter describes the results of applying Statistical Model and NN models using RGB and HUE values. It summarized the study and compares the different models according to given input.

4.2 Results

The most important information in the image of the forest cover discovery problem is the green color, but in this research it is difficult to separate forest cover using only green value. Therefore consider red values along with green values used to isolate forest cover. NN and statistical model were design in this study.

12 different cropped satellite images were tested in three models. The test results shown in below table (Table 4.1).

Input	Forest Cover Result by Neural Network- RGB	Forest Cover Result by Statistical Model - RGB	Forest Cover Result by Neural Network- HUE values
	 10.84%	 10.96%	 8.76%
	 16.36%	 15.68%	 8.53%
	 40.57%	 40.25%	 38.63%
			

	29.23%	34.06%	36.41%
	 99.59%	 99.68%	 98.6%
	 98.71%	 98.90%	 97.60%
	 57.41%	 57.68%	 48.93%
	 79.24%	 76.14%	 66.15%

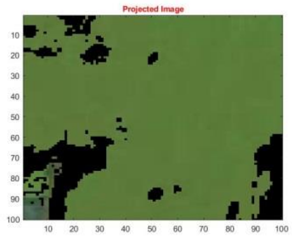
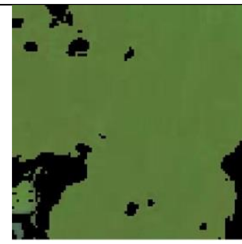
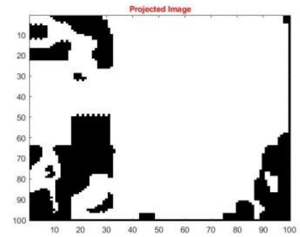
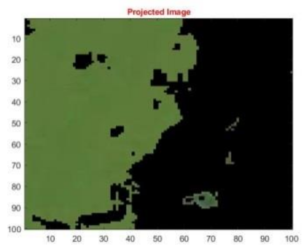
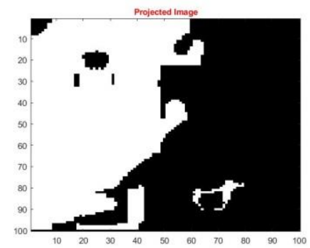
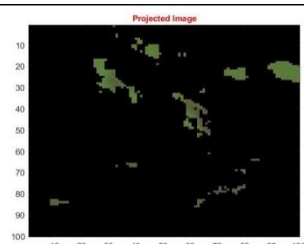
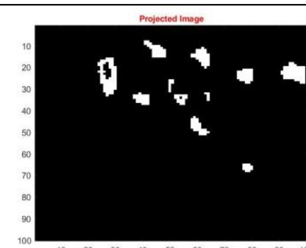
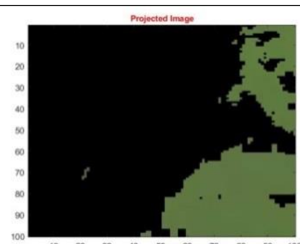
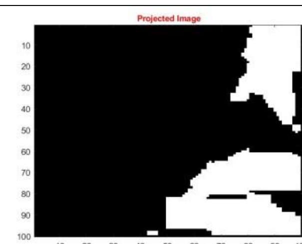
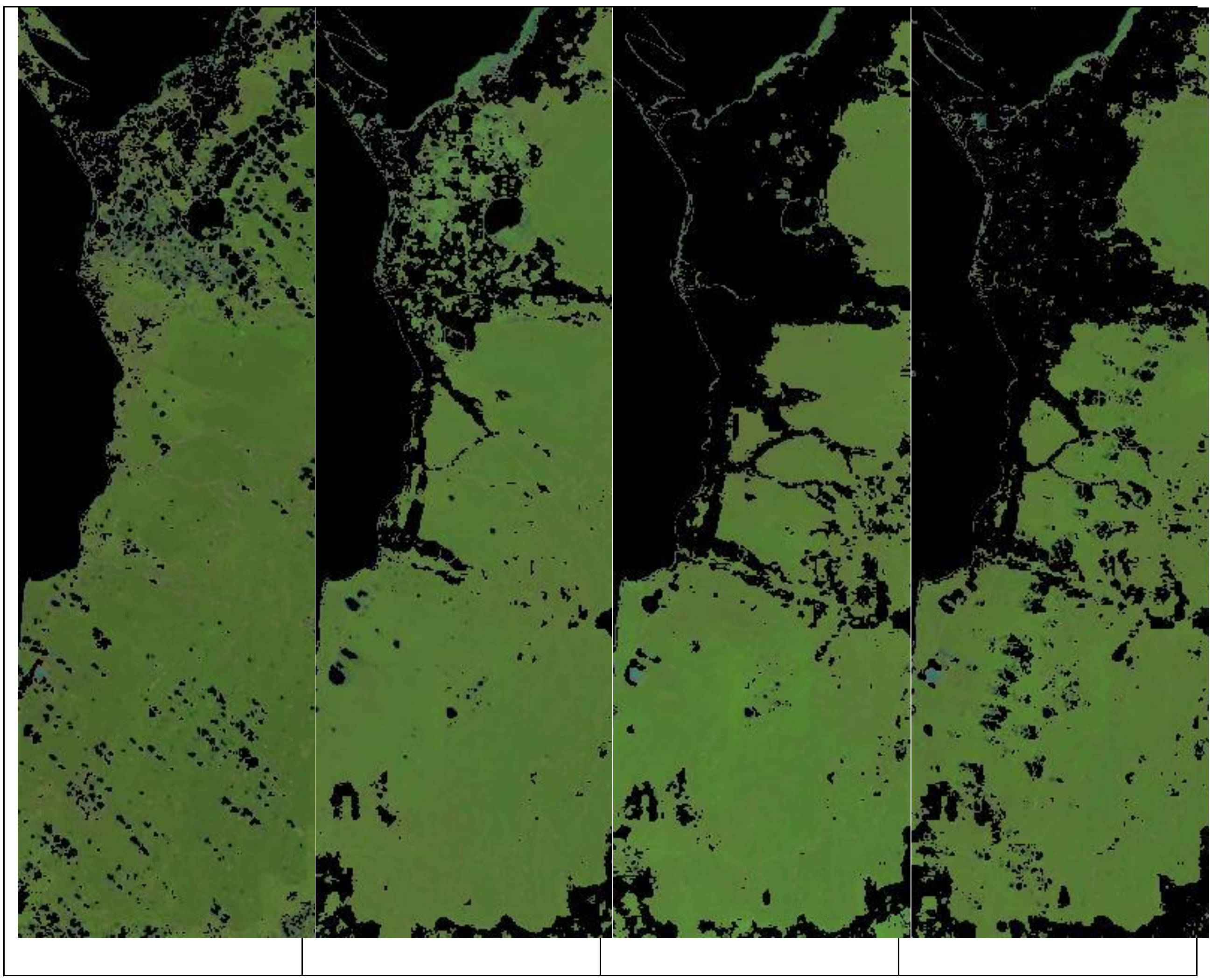
	 86.61%	 88.32%	 81.07%
	 48.73%	 51.57%	 46.41%
	 4.35%	 5.73%	 3.53%
	 24.59%	 29.32%	 23.26%

Table 4.1 : Comparison of Neural Network Model and Statistical Model

The classification is started by clipping the image into several parts due to the longer time taken by the neural network. With the medium capacity computers, its time consuming task. To obtain the results of whole study area resolution of pixels were merge into one pixel using 5*5 neighborhood pixels. As a result 30*30 m^2 pixel become 150*150 m^2 in final output.

4.2.1 Results of Neural Network Model using RGB values

The resulted forest cover area in Neural Network model using RGB values was shown in the figure (Figure 4.1) from 2000 to 2019.



a b c d

Figure 4.1 : Results of neural network model using RGB values for Wilpattu forest cover during the considered time period (a)2000 (b)2015 (c) 2018 (d) 2019

4.2.2 Statistical Model using RGB values

The statistical model using RGB value is applied in to the same data set and the outputs are shown in the figure (Figure 4.2).

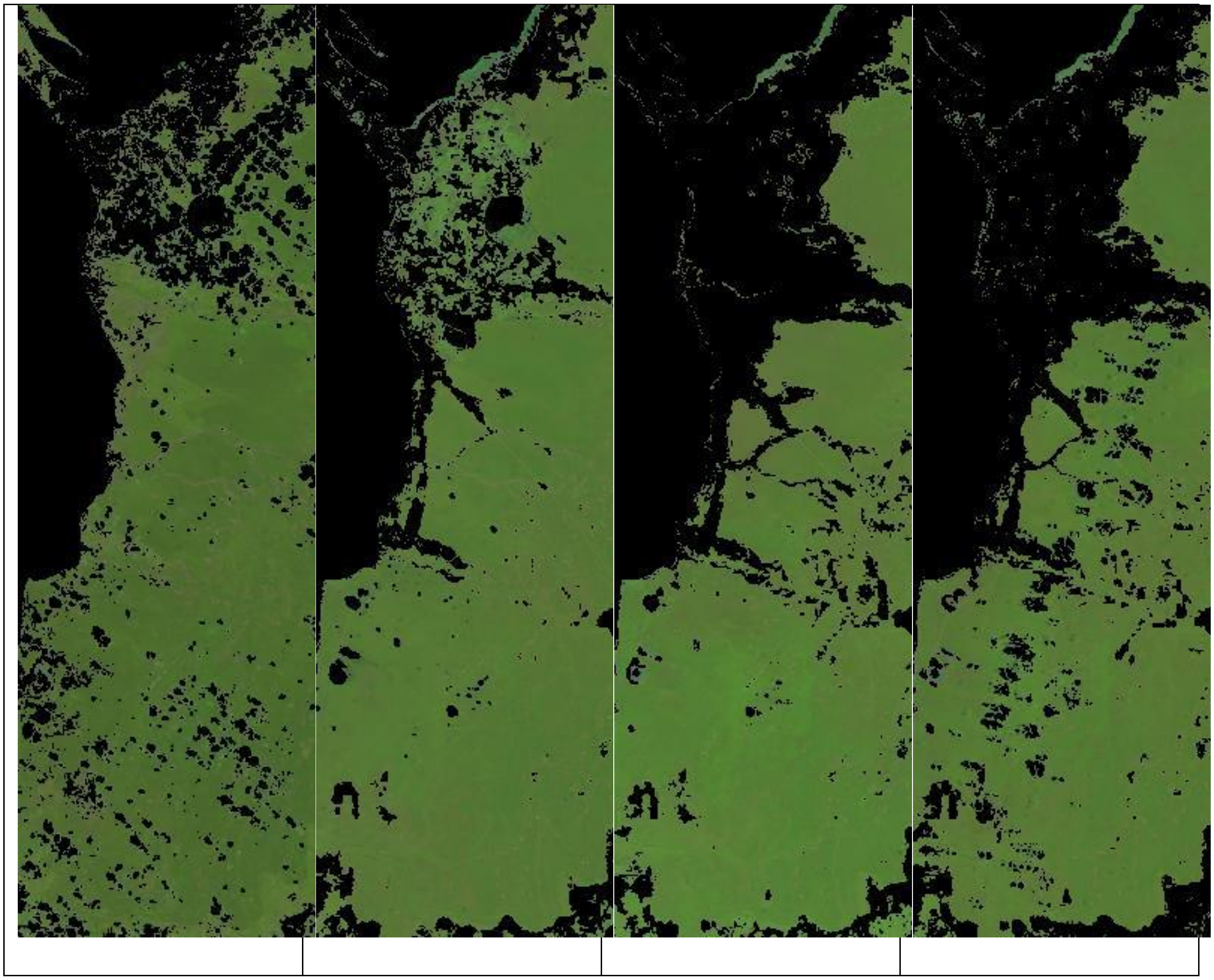


Figure 4.2 : Results of statistical model for Willpattu forest cover during the considered time period (a)2000 (b)2015 (c) 2018 (d) 2019

4.2.3 Neural Network Model using HUE value

The neural network model using HUE value is applied in to the same data set and the outputs are shown in the figure (Figure 4.3).

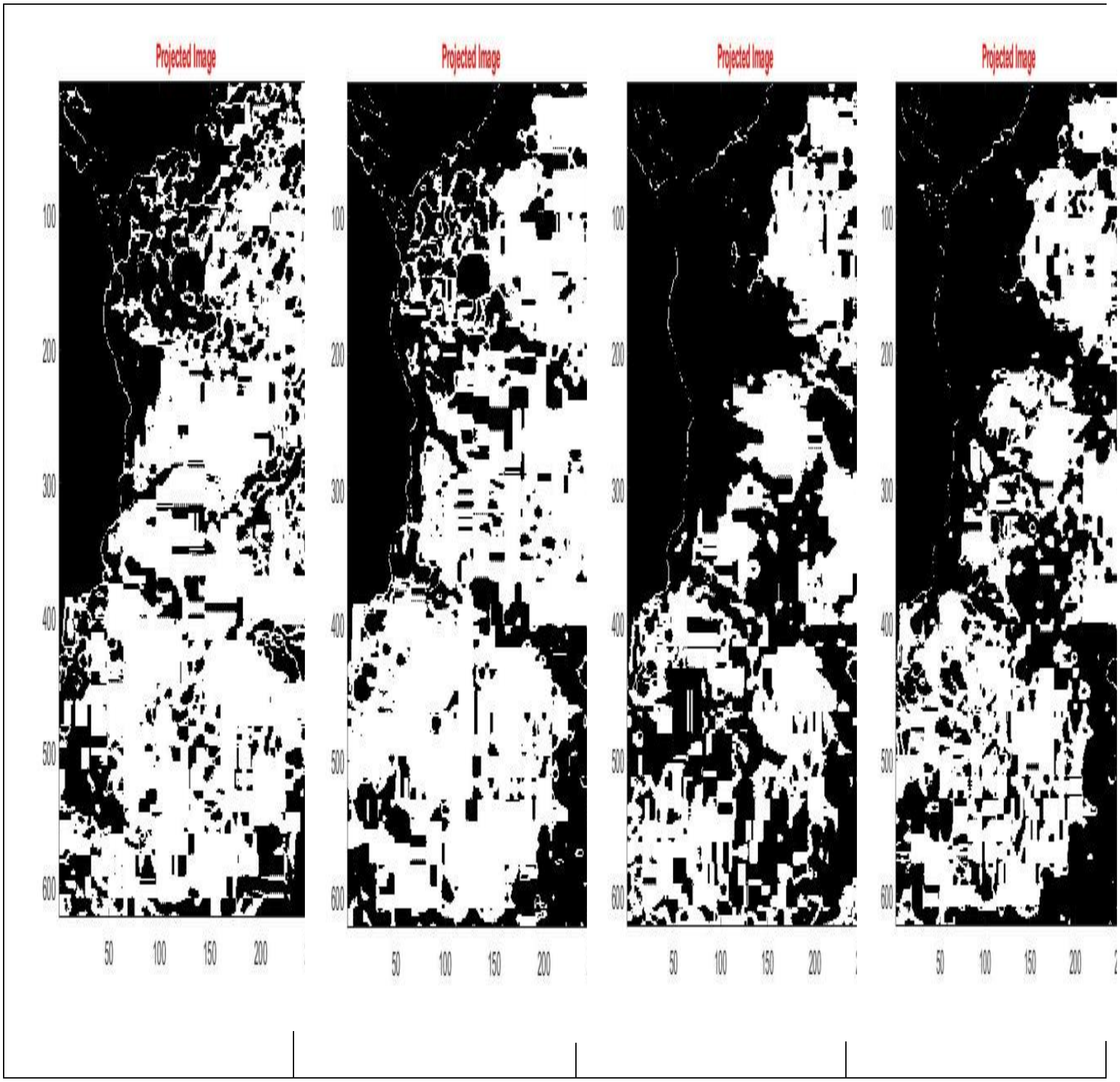


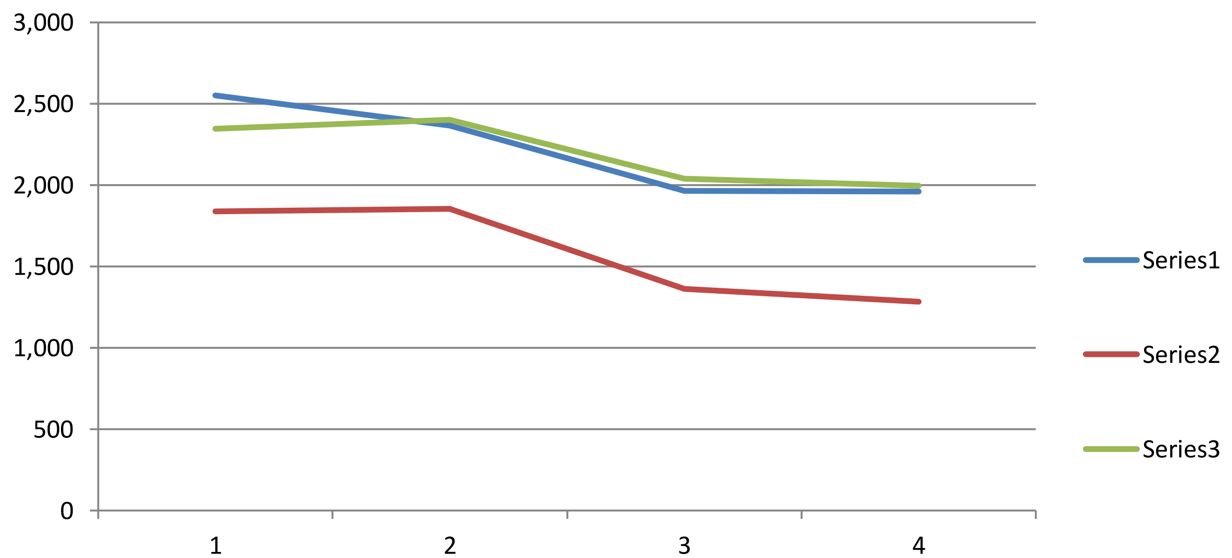
Figure 4.3: Results of neural network model(Hue) for Willpattu forest cover during the considered time period (a)2000 (b)2015 (c) 2018 (d) 2019

4.3 Evaluation

To conduct the evaluation procedure, forest cover of the selected Wilpattu area was calculated using the forest map of Survey department (2000) and values shown as 1,347,570,803.29 m^2. Evaluation is done by comparing the total forest cover compared to that.

Year	**By using neural network – RGB values (m^2)**	**By using neural network – HUE values (m^2)**	**By using statistical method - RGB values (m^2)**
2000	2,549,520,000	1,836,720,000	2,346,772,500
2015	2,364,457,500	1,854,180,000	2,401,335,000
2018	1,964,227,511	1,362,555,000	2,040,075,000
2019	1,959,907,500	1,281,262,500	1,994,175,000

Table 4.2: Results on forest cover when applying Neural Network Model and Statistical Model



Series1 – By using neural network (RGB values)

Series2 – By using neural network (HUE values)

Series3 – By using statistical model

Figure 4.4 : Graph for displaying forest cover when applying Neural Network Model and Statistical Model

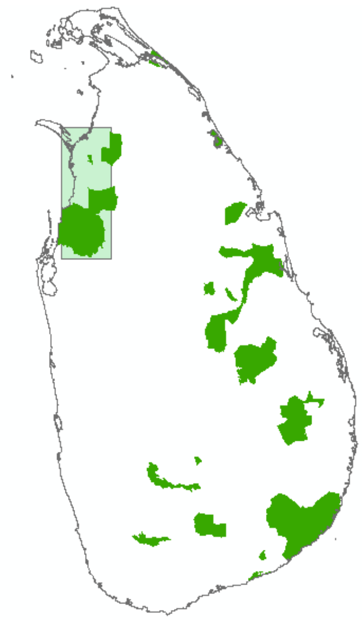


Figure 4.5 : Total Forest covers in Sri Lanka (2000)

Above table (4.2 Table) shows neural network models and statistical model output of the 2000, 2015, 2018 and 2019 satellite images. Three models show that forest cover is decreased according to year increases. Then we conclude that deforestation has been occurred in the study area from 2000 to 2019. But when we consider real forest cover in 2000 which was obtained from forest Survey department map (Figure 4.5), total area in 2000 obtained from both neural network models and statistical model is more than the real value. Because training the image using GREEN with RED and HUE values cannot accurately differentiate the forest cover from other vegetation types which are paddy, garden, shrubs and etc. Both green and red pixel values show similar characteristics in all vegetation types. By using more comprehensive method with blue pixels too vegetation types can be differentiate.

4.4 Discussion

In this study satellite images obtained from USGS web site. It contains 40 years of historical data when compared with other sources like NOAA Class [21]. Availability of the data was main reason for select the USGS web site as the main source for the research input. The Site provides users to select best satellite image from the result set.

In the methodology, Feed Forward Neural Network was used. Because FFNN is known as the one of the best supervised machine learning tasks which input and output already known rather than other supervised classification methods (Penatti et al., 2015). FFNN is a high-precision good model that provides predictions that are very close to actual values. In many applications FFNN is used X-ray images of the superimposed objects to train inspectors [22] and Deblocking of Low Bit Rate Coded Images [23].

CNN is another classifying machine learning mode which can derive from FFNN. CNN is deep artificial neural network that are mainly used to classify images. Many researchers explained about CNN is the best methods for classifying and in 2015, Penatti et al. shows that the accuracy as 99.5%.

Main objective of the research is find forest cover from Landsat images. RGB images of corresponded satellite images were taken as input and giving the output is determined whether the given pixel value can be identified as forest or not. It means the output values contain the Boolean results. After taking into the FFNN is used as best method which satisfy the main objective if the study out of other methods.

MATLAB was used here to read the RGB values of satellite images and assign it to a 2D matrix. The matrix values are applied to the NN model and the image is regenerated again based on simulation results. MATLAB provides a Neural Network Toolkit that hopes to use neural networks in conjunction with MATLAB.

This study specially based on the Land cover classification techniques to identify forest cover. K-means clustering is a vector quantization method to divide n objects into k clusters, each of which belongs to a cluster with the most recent mean. It is one of the best method for classifying and in both prediction of students' Academic performance [24] and detect an infected part of fruit [25] .

Chapter 5 – Conclusion

Because of its good forcefulness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance, neural network is very suitable for solving the problems of data mining.

Compared with other statistical methods where they is no prior knowledge about the data, NN are useful. They offer a powerful and distributed computing architecture, with significant learning abilities and they are able to represent highly nonlinear and multivariable relationships.

Artificial Neural Networks offer qualitative methods for business and economic systems. In traditional quantitative tools which are available in statistics and econometrics, it cannot quantify due to the complexity in translating the systems into precise mathematical functions.

NN and data mining have been become promising fields in the research areas where there are large mass of dataset are available. The ability of neural networks to detect and assimilate the relationship between large numbers of variables is become trending reason for use of the NN.

5.1 Future Work

The aim of this research is to examine the forest area from a given RGB image. The study used Artificial Neural Network Method and statistical method depends on RGB values of images from USGS website. Though the study is primarily based on RGB values of satellite images, more accurate results can be obtained by using more quality images with high spatial resolution.

Satellite images were taken when the amount of clouds were very less. Therefore removing clouds from the image was not done in this research. Also there is no major preprocessing was done, even though the images has gone through radiometric and geometric correction in the initial preprocessing part. As a future work, cloud removal and other preprocessing steps can be used to get more accurate results.

The image was separated into two clusters as forest and non-forest based on the green value of the RGB unit vector. When considering the red value, the land area of the forest was identified.

However more accurate results can be obtained in future studies by using both red and green values along with blue values.

The real satellite image contains several bands relative to the wave lengths of the electromagnetic spectrum. In order to reduce the complexity of the research, RGB images were extracted from satellites images. By considering the band values of multispectral images to train the neural network, more accurate and efficient results can be obtained.

References

[1] A. B. F. D. D. S. B. D. Marian Vittek, "Land Cover Change Monitoring Using Landsat MSS/TM Satellite Image Data over West Africa between 1975 and 1990," *Remote Sensing,* pp. 658-676, 01 2014.

[2] 08 2018. [Online]. Available: https://rainforests.mongabay.com/deforestration/2000/Sri_Lanka.htm.

[3] 08 2018. [Online]. Available: http://www.wilpattunationalpark.com/visiting_wilpattu_national_park.php?lId=02.

[4] D. K.U.J.Sandamali, "Deforestation or Reforestation, A Time Series Remote Sensing Perspective of Wilpattu National Park, Sri Lanka," *Journal of Applied Mathematics and Computation,* 10 2018.

[5] 08 2018. [Online]. Available: http://www.fao.org/3/i1757e/i1757e02.pdf.

[6] P. B. P. Baudouin Desclée, "Forest change detection by Stastical object-based method," *Remote Sensing of Environment,* vol. 102, no. 1-2, pp. 1-11, 05 2006.

[7] 08 2018. [Online]. Available: https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html.

[8] 08 2018. [Online]. Available: https://www.scribd.com/document/250449108/Definition-of-Soft-Computing.

[9] A. Tözer, "NEURAL NETWORK SYSTEM FOR CHARACTER RECOGNITION".

[10] 08 2018. [Online]. Available: https://www.termpaperwarehouse.com/essay-on/Neural-Network/166432.

[11] 08 2018. [Online]. Available: http://irrigation.rid.go.th/rid15/ppn/Knowledge/Artificial Neuron Networks Technology/4.0 Neural Networks Components.htm.

[12] 09 2018. [Online]. Available: http://myneuralnetwork.blogspot.com/.

[13] P. G. P. H. H. Mu¨ller, "Long-term deforestation dynamics in the Brazilian Amazon—Uncovering historic frontier development along the Cuiabá–Santarém highway," *International Journal of Applied Earth Observation and Geoinformation ,* p. 61–69, 02 2016.

[14] V. L. C. P. C. B.-L. O. W. C. M. S. B. V. A.-R. J. B. A. C. e. a. M. Pfeifer, "Creation of forest edges has a global impact on forest vertebrates," *Nature,* 11 2017.

[15] P. V. P. R. M. M. H. S. T. A. T. D. T. S. S. G. T. L. e. a. M. C. Hansen, "High-resolution global maps of 21st-century forest cover change," *F1000 - Post-publication peer review of the biomedical literature ,* p. 850–853, 11 2013.

[16] V. Ahmadi, "Using GIS and Artificial Neural Network for Deforestation Prediction," 03 2018.

[17] Z. Lin, "Artificial neural networks in predicting forest cover types from cartographic variables".

[18] X. X. ,. Z. Y. ,. L. X. ,. Z. F. ,. Z. X. Yuan Yanbin, "The Application of Improved Fuzzy ARTMAP Neural Network in Remote Sensing Classification of Land-use," *2010 Second International Conference on Multimedia and Information,* 2010.

[19] [Online]. Available: https://www.usgs.gov/land-resources/nli/landsat/data-tools.

[20] 03 2019. [Online]. Available: https://www.saedsayad.com/clustering_kmeans.htm.

[21] 03 2019. [Online]. Available: https://gisgeography.com/free-satellite-imagery-data-list/.

[22] W. ZhangJian, "The application of feed-forward neural network for the X-rayimage fusion," *Journal of Physics,* 2011.

[23] J.-S. M. S.-H. I. a. D.-S. L. K.-K. K. M.-S. Yang, "Application of Feedforward Neural Network for the Deblocking of Low Bit Rate Coded Images," *Lecture Notes in Computer Science Artificial Intelligence and Simulation,* p. 547–555, 2005.

[24] O. O. O. I. C. O. . J, "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance," *International Journal of Computer Science and Information Security,* vol. 7, 2010.

[25] P. D. N. S. J. P. G. D. S. R., "Infected Fruit Part Detection using K-Means Clustering Segmentation Technique," *International Journal of Interactive Multimedia and Artificial Intelligence,* pp. 65-72, 06 2013.

Appendices

Train HUE values to network

```
clc
clear all;

[FileName,PathName] = uigetfile('*.jpg','Select the Training jpg file');
full_file_name = fullfile(PathName, FileName);
I = imread(full_file_name);

hsv = rgb2hsv(I);

[m n k]=size(I);
k=1;

arr = zeros(((m-2)*(n-2)),1);

for i=2:m-1
    clear j;
    for j=2:n-1

        arr(k)= round(mean(mean(double(hsv(i-1:i+1,j-1:j+1,1))*255)));
        k=k+1;

    end
end

IDX=kmeans(arr,6);

P1=arr(IDX==1,:);
P2=arr(IDX==2,:);
P3=arr(IDX==3,:);
P4=arr(IDX==4,:);
P5=arr(IDX==5,:);
P6=arr(IDX==6,:);

number=min(size(P1),size(P2));
number=min(number,size(P3));
number=min(number,size(P4));
number=min(number,size(P5));
number=min(number,size(P6));

number=number(1);

arrSort = zeros(6,1);
out1=round(mean(P1));
arrSort(1)= out1;
arrSort(2)= round(mean(P2));
arrSort(3)= round(mean(P3));
arrSort(4)= round(mean(P4));
arrSort(5)= round(mean(P5));
arrSort(6)= round(mean(P6));
% arrSort(7)= round(mean(P7));
```

```
B=sort(arrSort);


for i=1:6
   if arrSort(i)==B(3)
         thirdPos=i;
   else
   end
end


 if thirdPos ==1
                    P=[P1(1:number,:);P2(1:number,:); P3(1:number,:);
P4(1:number,:); P5(1:number,:); P6(1:number,:)]';
                    T=[ones(1,number) zeros(1,number) zeros(1,number)
zeros(1,number) zeros(1,number) zeros(1,number)];

 elseif thirdPos==2
                    P=[P1(1:number,:);P2(1:number,:); P3(1:number,:);
P4(1:number,:); P5(1:number,:); P6(1:number,:)]';
                    T=[zeros(1,number) ones(1,number) zeros(1,number)
zeros(1,number) zeros(1,number) zeros(1,number)];

 elseif thirdPos==3
       P=[P1(1:number,:);P2(1:number,:); P3(1:number,:); P4(1:number,:);
P5(1:number,:) P6(1:number,:)]';
       T=[zeros(1,number) zeros(1,number) ones(1,number) zeros(1,number)
zeros(1,number) zeros(1,number)];

 elseif thirdPos==4
       P=[P1(1:number,:);P2(1:number,:); P3(1:number,:); P4(1:number,:);
P5(1:number,:) P6(1:number,:)]';
       T=[zeros(1,number) zeros(1,number) zeros(1,number) ones(1,number)
zeros(1,number) zeros(1,number)];

 elseif thirdPos==5
       P=[P1(1:number,:);P2(1:number,:); P3(1:number,:); P4(1:number,:);
P5(1:number,:) P6(1:number,:)]';
       T=[zeros(1,number) zeros(1,number) zeros(1,number) zeros(1,number)
ones(1,number) zeros(1,number)];

 else
       P=[P1(1:number,:);P2(1:number,:); P3(1:number,:); P4(1:number,:);
P5(1:number,:) P6(1:number,:)]';
       T=[zeros(1,number) zeros(1,number) zeros(1,number) zeros(1,number)
zeros(1,number); ones(1,number)];


 end

net2_hsv_1=newpr(P,T,20);
 %Set up Division of Data for Training, Validation, Testing
net2_hsv_1.divideParam.trainRatio = 75/100;  % Adjust as desired
net2_hsv_1.divideParam.valRatio = 10/100;  % Adjust as desired
net2_hsv_1.divideParam.testRatio = 15/100;  % Adjust as desired
```

```
net2_hsv_1.trainParam.max_fail = 10000; %Maximum validation failures
net2_hsv_1.trainParam.epochs = 4000;   %Maximum number of epochs to train
net2_hsv_1.trainParam.goal = 0.000001;    %Performance goal
net2_hsv_1.trainParam.min_grad = 0.000000000000000000000000000000000001;

% Train and Apply Network
[net2_hsv_1,tr] = train(net2_hsv_1,P,T);
save('final2_hsv_1.mat','net2_hsv_1');
```

Browse the cropped image and test the network

```
load final2_hsv_1.mat;

[FileName,PathName] = uigetfile('*.jpg','Select the Training jpg file');
full_file_name = fullfile(PathName, FileName);
INew = imread(full_file_name);

figure;

title('Original Image','FontWeight','bold','Fontsize',10,'color','red');


hsvResult  = rgb2hsv(INew);
 [x z v]=size(INew);
INew2=zeros(x,z,3,'uint8');
arrVal = zeros(x*z,1);

 g=1;


for s=1:x
    for f=1:z
%          output=0;
        output=sim(net2_hsv_1,(double(hsvResult(s,f,1))*255));
        arrVal(g)=output;
        g=g+1;

        if sim(net2_hsv_1,round(double(hsvResult(s,f,1))*255)) >.9999
            rr=I(s,f,1);
            INew2(s,f,1)=255;
            valR=INew2(s,f,1);
            gg=I(s,f,2);
            INew2(s,f,2)=255;
            bb=I(s,f,3);
            INew2(s,f,3)=255;
        else
            INew2(s,f,1)=0;
            INew2(s,f,2)=0;
            INew2(s,f,3)=0;


        end
    end
end

figure;
image(INew2);
imwrite(INew2,'E:\Final_Demo\2000_Crop1_result_2.jpg');

title('Projected Image','FontWeight','bold','Fontsize',10,'color','red');
```