INFORMATICS INSTITUTE OF TECHNOLOGY

UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER, UK

GIS & Remote Sensing to Find Deforestation in Sri Lanka during the Period of 2000 to 2020 and Examining How to Build an algorithm to Predict Deforestation

A dissertation by

Mr. B.M Serasinghe

Supervised by

Dr D.D Karunaratne

Submitted in partial fulfilment of the requirement for the

BEng. (Hons) in Software Engineering degree

Department of Computing

May 2020

Acknowledgement

This research is carried out in fulfilment of the requirements for BEng Honours in Software Engineering at Informatics Institute of Technology Sri Lanka. This is an individual research project and the research wouldn't have been successful without help and guidance of my university and family.

Therefore, first of all I would like to thank my supervisor Dr D.D Karunaratne who has been always providing support with his knowledge to make best out of my research project. Also I would like to thank the module leader Kaneeka Vidanage for the guidance and his words in writing the thesis.

I express my sincere gratitude to my parents for the courage given to me throughout the project which was very much needed to focus on the research.

Last but not least I would like to thank my colleagues for supporting me by filling out the survey forms and also for their continuous help during the tough times.

Abstract

The present research was conducted in order to identify the deforestation in Sri Lanka during the period of 2000-2020 and based on the collected data it will propose an algorithm to predict deforestation in coming years in Sri Lanka. Around the world deforestation is very common due to many factors and it will have unique factors from one region to another. Hence this analysis is in search of those factors in Sri Lankan forest areas. Sri Lankans are nation which loves the nature very much and we all must work in order to protect it in our best effort. From this project we will try to identify the forest areas which are in threat. By finding these data the intention is to aware the Sri Lankans and authorities about deforestation critical areas in Sri Lanka and emphasis to create necessary management plans to protect these resources.

Table of Contents

List of Figures

List of Tables

1. Introduction

Content

1.1. Chapter Overview

During this chapter it will give an approach to the problem of this research. First It will discuss about the problem considering level of threat globally and then it will show how it has been a threat in Sri Lanka. Also the other disasters which could arise from the research problem if it is couldn't avoid will be discussed during this chapter. After covering those topics objectives of the research will be explained and significance of the problem is shown. After that when implementing the concept, the requirement of hardware and software are mentioned and finally high level rich picture of the proposed solution is given.

1.2. Introduction to the Problem

Deforestation is considered as a global problem and there have been many steps been taken worldwide to control it. From the ancient times trees have been cut down for different purposes. Cutting down trees in large amount is known as deforestation (Dias & Liyanawatte, 2017). With the expansion of civilization trees have been heavily cut down for agriculture. Therefore, with the time forests and eco systems have been diminished gradually. According to Food and Agriculture organization of United Nation (FAO), Forest is known as a land area larger than 0.5 hectares where trees are taller than 5m with a crown cover exceeding 10%. According to Sri Lankan definition a forest is mentioned as a land area consists of minimum 0.05 hectares with 20% of tree canopy cover and canopy height of 3m. (Dias & Liyanawatte, 2017) Deforestation is considered as a critical problem since it can bring several environmental hazards. Deforestation could be occurred due to various reasons. Logging activities inside the forests can be identified as the major reason for global deforestation (Lentin, et al., 2013). Another major reason we can identify directly is the overpopulation cities and developing countries. Increase of the population and per capita demands for more forest products and that will cause forest lands to turn into agricultural lands or other necessary services like mining and extracting fossil fuel. From the past countries like Sri Lanka and other neighbourhoods had been prominent agricultural countries and since then people had moved to forests to create agricultural lands and use of chemicals pesticides have

brought worse life for the wild life. Also there will be an immense damage caused by creating roads to access the forest and generating hydroelectric power by using forest resources. (Tra, 2014). Sometimes it does damage even larger than the damage caused by illegal logging. Cattle grazing are also considered as one of the reason for deforestation. As the expansion of population, demand for dairy products like milk and other animal products goes higher. Therefore, we need more lands to for these purposes which directly affect forests lands to disappear. All the above factors are governed by state policies, therefore having healthy state policies is very important in protecting forests.

1.3. Problem Statement

Sri Lanka, which is considered as pearl of the Indian Ocean, takes an important place as a biological host spot with a landscape of 65,525km^2. Sri Lanka as a tropical country has a rich bio diversity. But at present Sri Lanka's flora and fauna has facing a critical problem due to deforestation. So it's necessary to manage the forests island wide to protect the all living beings.

There has been a drastic reduction in Sri Lanka's natural forests during last century. From 80% it has reduced to 16%. During 1950 Sri Lanka was covered with 50% of forests and in the beginning1990 only quarter of the country was covered with forests. (Bogahawatte, 1999). Sri Lankan forests are categorized into five; National heritage wilderness areas, conservation forests, reserved forests, village forests and other state forests. Sri Lankan forests are reduced to half of its former extent during the rule of Portuguese, Dutch and English due to commercial agricultural works (Dias & Liyanawatte, 2017). Timber became major source of fuel after independence and that became a reason for deforestation in Sri Lanka during that period. Furthermore, with the initialization of irrigation projects like Mahaweli which was started to irrigation purposes in dry zone in north east accelerated the deforestation and damage the forest system of Sri Lanka by consuming a forest land of area of 2430km^2. One of the most recent reason for deforestation is civil war prevailed in Sri Lanka, because Sri Lanka government had to destroy the hiding places of enemies by clearing the forest areas. Also when we consider last decade wild fires and gem mining also have been a reason for deforestation in Sri Lanka (Dias & Liyanawatte, 2017).

1.4. Map of Deforestation in Sri Lanka

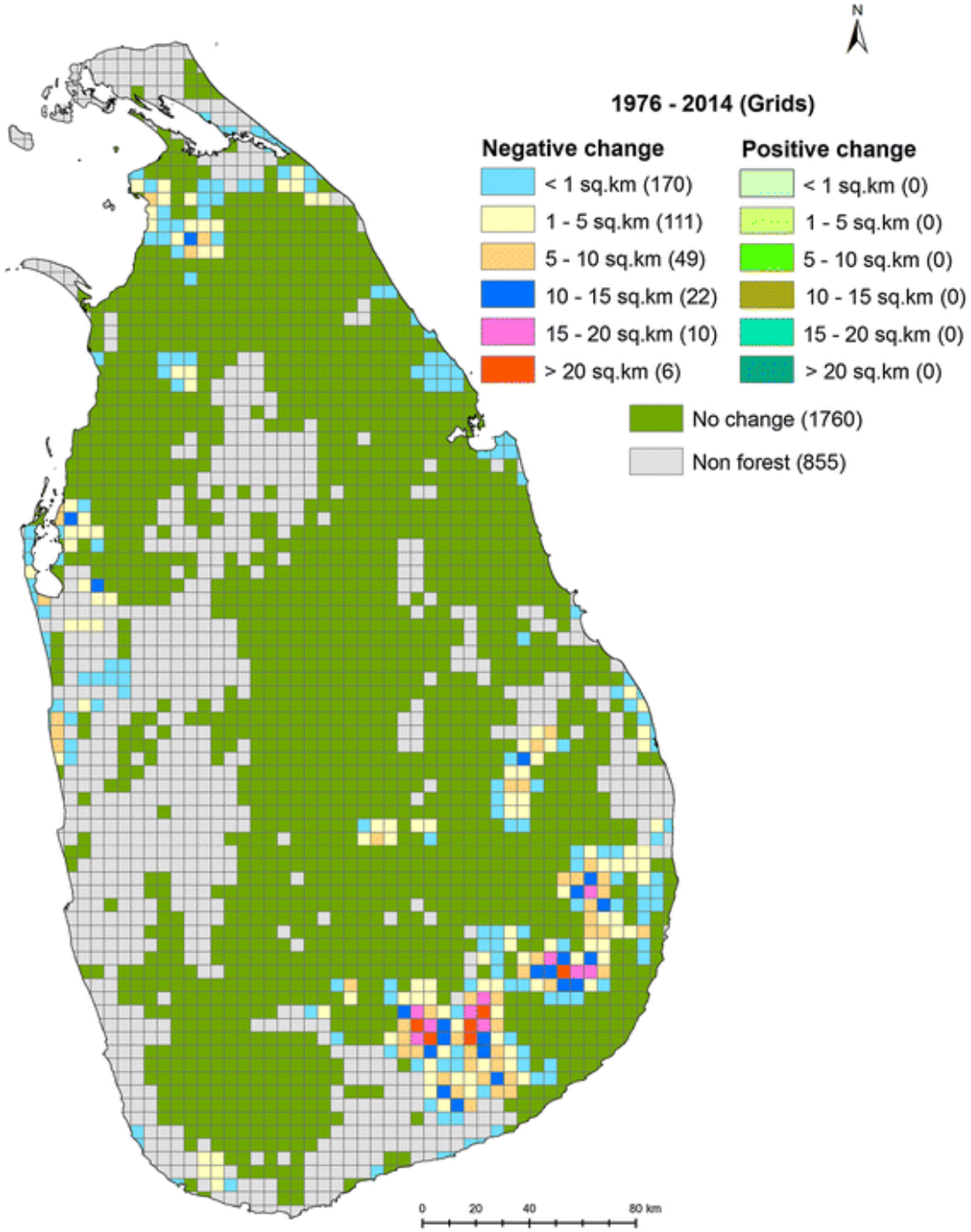

Figure 1: Deforestation of Sri Lanka in the period of 1976 to 2014 (Shekhar Jha, et al., 2016)

Figure 1 shows the deforestation that was recorded before 2016. From the different coloured grids, it shows the area of deforestation has occurred while green colour is the area there was no deforestation is recorded. From legendary given you can follow up further details about deforestation that was recorded.

1.5. Research Question

Find the deforestation happened in Sri Lanka during the period of 2000 – 2020 and how can we obtain prediction maps using a machine learning algorithm?

1.6. Motivation

In the past few years' deforestation in Sri Lanka has been a major topic in almost every media. It has created social and political awareness among the community and there were some petitions against the deforestation of forest. (Sooriyagoda, 2019) These problems have also created tension among the ethnic communities. (Jayasuriya, 2019) Therefore, it is very important to coming up with a good solution to manage forests. Not only to maintain peace, we also need to protect our eco system and maintain the bio diversity in the country.

1.7. Aim

To find the lands that have been deforested in Sri Lanka during the period of 2000 to 2020 and to produce an algorithm to find how will be the deforestation in those areas in future which will help authorities in forest management in Sri Lanka.

1.8. Research Objectives

1. Finding the deforestation in Sri Lanka during the period of 2000-2020
2. To identify the areas which are in threat of deforestation
3. Trying to reduce the problems in society arisen due to deforestation
4. To produce a suitable algorithm that will predict the areas in threat

1.9. Project Scope and Delimitation

This research is conducted to find the deforestation and to produce an algorithm to predict the forest areas which are in threat of deforestation by using the technologies remote sensing and GIS. The study area is limited to areas where there are forests in Sri Lanka. The satellite images which are going to use are going to obtain from USGS Earth Explorer and Copernicus Open Access Hub. This study takes satellites images from 2000 to 2020 for the consideration to analyse the data on deforestation in Sri Lanka.

1.10. Significance of the Problem

As we have discussed above deforestation is an environmental issue which has been a serious threat to the eco systems in Sri Lanka and it has taken the attention of everyone during last few years. Regardless of the important of forests many countries do not show statistics about deforestation extent. (Ahmadi, 2018)Therefore, through this research paper we will try to investigate and come with a solution by using technologies available in the world. Through the literature review of this document you will have clear idea about what technologies we intend to use, what are the existing projects available and how our research is different from them. So first we will know the technologies have used so far. It's especially machine learning methods are taken in to consideration when making prediction algorithm with data. Geographic Information Systems (GIS) can create a good environment for making predictions with the use of machine learning algorithms and satellite images. Spatial analysis can have used for various aspects in spatial problems in various scales. Deforestation can be modelled considering various factors which depends the relationship between deforestation and environmental and socioeconomic factors. (Ahmadi, 2018)

1.11. Software and Hardware requirements

Software Requirements
Operating System of Windows 7 or above
It will be useful in performing functions
QGIS Software
To process satellite images and mapping the deforestation
Preferred Web Browsers – Chrome, Safari
To download necessary satellite images from web sites
Documentation – Microsoft word, Google Sheets
In completion of the reports and rest of the documents

Table 1: Software Requirements Table

Hardware Requirements

For Web Application
CPU: i3 or similar (i5 or i7 desirable), 2 cores or more
To process the functions smoothly within a shorter period of time
Minimum of RAM: 2GB (8 Gb desirable)
allows high speed access to the data
Disk Space of 10GB+
To provide required storage
Graphics: dedicated, 128 MB (256 MB desirable)
It gives higher resolution

Table 2: Hardware Requirements Table

1.12. A high Level rich picture of the proposed solution

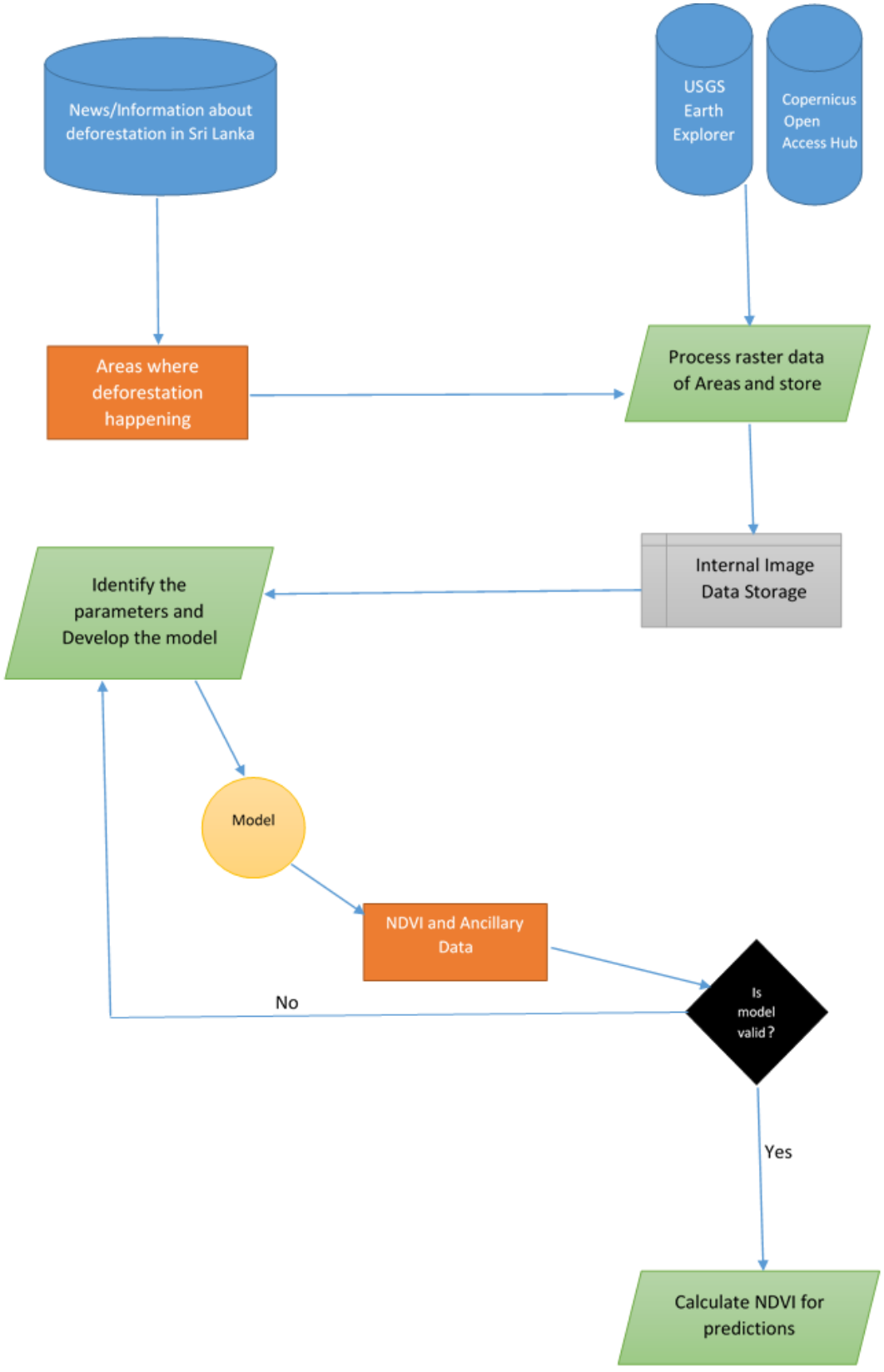

Figure 2- High-level Architecture of the research Design

1.13. Chapter Summary

In this chapter we have discussed in detail about the problem domain that have been chosen and it is justified by the recent set of events. Also this chapter has given you aim and objectives of the research. Further it has given you the requirements in order to make this research successful. Therefore, by referring to this chapter you can have overall idea what is this research about and what it will deliver to the society and thereby significance of the research is justified.

2. Literature Review

Content

2.1. Chapter Introduction

During the literature review, research problem will be discussed in detail. First it will discuss about the causes of deforestation and then effects which has been occurred due to deforestation will be taken into consideration. Once the effects are identified, the control strategies for the deforestation will be discussed. Finally, this chapter discuss about the technologies that has been used in previous researches to map deforestation and will show the research gap.

2.2. Importance of Managing Deforestation

Forests can be considered as an important part of the Earth and it occupies third of the earth's land surface. It provides many benefits including control hydrologic cycle, preventing climatic change, protects bio diversity and soil conservation. (Sheram & Katherine, n.d.)Forest resources can deliver long-term economic benefits at the national level. It takes prominent place in wood production and at least 145 countries are currently involved in it. There is ample evidence to prove that, due to heavy deforestation, the entire world is facing an environmental crisis. No one knows how much of the rain forest in the world has been destroyed and continues to be destroyed every year. Data is always unreliable and interpretable in a particular way. Nevertheless, amid increased environmental advocacy and recognition, it is obvious that tropical rains are declining and the rate of tropical rainforest destruction is growing worldwide. Deforestation can be considered as conversion of forest to a permanent non-forested land for different uses like agriculture, grazing and for different kinds of infrastructure developments (G. Cornelius van & Bulte, 2000). Deforestation has become a major problem for developing countries as it decreases areas of tropical forests (Barraclough, 2000) causing loss of bio diversity a producing greenhouse effects. FAO considers the planting of trees mainly for the production of timber as forests and therefore does not recognize the transformation of natural forests into planting (But still it is recorded as a natural forest loss). But FAO does not consider tree plantation that supplies non-wood products as forests, even though rubber plantations are listed as forests. Generally, 30% of the earth's land or 3.9 billion hectares are owed by forests

and it had estimated that original forest cover was nearly 6 billion hectares (Bryant, et al., 1997). With 53% of the world's total land area, the Russian Federation, Brazil, Canada, America, and China have become the richest countries in the world. In less than ten percent of the total area of the forest, another 64 countries with a combined population of 2 billion have registered forests and sadly, ten of these countries have no forests. Among the 16 countries, the forest areas of Chad, Islamic Republic of Iran and Mongolian countries were comparatively high, each with its more than 1 million hectares and three countries with more than 10 million forest hectares. The forest region in North and Central America remained relatively stable despite its growth in Europe in the past decade. In particular, in India and China, Asia has reported a net gain in forest areas in the last decade because of its large-scale afforestation program. The annual net loss of forest areas was documented by South America, Africa and Oceania.

2.3. Classification of the Topic Deforestation

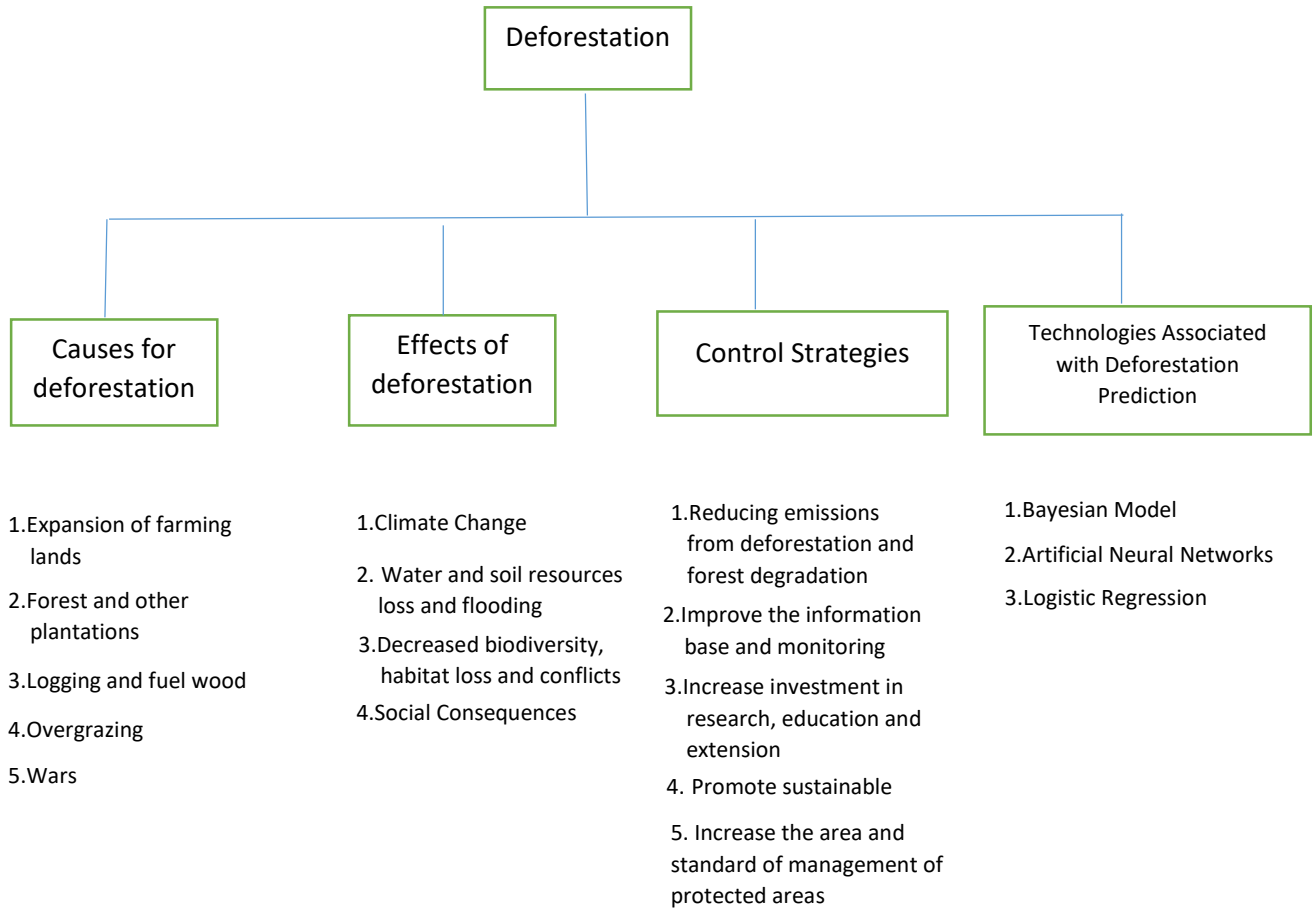

Figure 3: Classification of the topics

In this diagram It has shown the topics that we are going to discuss below in detail during the literature review. It will help you to have a clear understanding of the topics overall. Each topic has arranged in a manner to give good approach to the technologies associated with the research.

2.3.1. The causes of deforestation

To understand the major determinants of deforestation, distinguishing between the deforestation agents and their causes is necessary. Deforestation agents include slash and burn farmers, commercial farmers, loggers, firewood gathering firms, developers of infrastructures and others who cut down the forests. Forces that motivate agents to clear the forests are causes of deforestation. However, most of the literary works currently available varies generally among two rates of factors that cause deforestation directly and indirectly.

2.3.1.1. Expansion of farming land

Usually 60 percent of tropical moist forests are cut down for agricultural purposes (Myers, et al., n.d.) along with logging and making paths to access and urbanization can be considered as the rest. However, when the lands for settlements won't enough people are forced to migrate and occupy forest lands (Christensen & Pfaff, 2008). The growth of agricultural land is responsible for deforestation. This is because agriculture growth is commonly seen to contribute approximate 60 percent for overall tropical deforestation as a mainly source of deforestation. Shifting agriculture, which is often referred to as slash and burn agriculture, is the clearing of the forested land for the planting or planting of crops before resources are depleted and or the area is overtaken by crops

2.3.1.2. Forest and other plantations

Plantation is a good benefit which will help reduce the deforestation rate. But it won't necessarily mean to reduce the burden of timber on natural forests. Agricultural expansion as the major cause for deforestation in the tropics might replace forestry by plantation in the remaining natural forests. Plantations may also facilitate deforestation by building roads to increase access to the forest frontier for shifting farmers and others.

2.3.1.3. Logging and fuel wood

Logging does not necessarily cause deforestation. However, logging can do a bigger damage by the forest degradation. Nonetheless, logging provides roads for settlers to follow and log scales can help fund the expense of clearing existing trees and preparing land for crop or pasture planting. Therefore, logging can be considered as a factor that will increase deforestation (Chomitz, et al., 2007).

2.3.1.4. Overgrazing

Overgrazing in more dry areas of the tropics is more popular where soil erosion occurs on pastures overgrazed. Stripping trees for the food of grassing animals may be a concern, but possibly does not lead to a substantial deforestation in some dry regions of the tropics.

2.3.1.5. Wars and role of the military

Military operations have been known for the cause of deforestation. In addition to military operations, Southeast Asia and South America have recorded the role of the military in deforestation. The writers have noted that Brazilian diplomacy plays a significant role in the degradation of the Amazon forest.

2.3.2. Effects of Deforestation

2.3.2.1. Climate change

Deforestations will alter the global shift in energy by increasing the atmospheric carbon dioxide concentration and by not just micrometeorological processes (Pinker, et al., n.d.) because carbon dioxide absorbs ambient thermal radiation in infrared. Deforestation in lowland areas turns into higher rates of cloud formation and rainfall (Hooper, et al., 2005). Deforestation disrupts ordinary weather patterns that produce hotter and drier weather, thereby rising drought and desertification, crop failures, polar ice caps melting, coastal floods and the displacement. Land degradation has become a growing issue in dry forest zones and has resulted in severe desertification cases. Global warming or environmental change involves climate and ecological anthropogenic problems as the recent obvious climate and precipitation regimes in some regions, increase in sea levels, stratospheric ozone depletion, soil and forest pollution. Due to global deforestation, carbon dioxide release is equal to an estimated 25% of fossil fuel combustion emissions (Okia, 1990).

2.3.2.2. Water and soil resources loss and flooding

The global water cycle is also disrupted by deforestation. The region cannot retain as much water to create a drier atmosphere as a result of the loss of part of the trees. Deforestation-affected waters include drinking water, fishing and marine ecosystems, flood protection, waterway management and dams, less desirable recreation for water, and crop damage and erosion and turbidity irrigation systems. The protection of urban water is theoretically one of the key facilities for forests (Chomitz, et al., 2007). It's expensive to filter and treat water. Deforestation can also contribute to rivers and streams which can no longer sustain and control water flows. After they have gone, too much water will lead to downstream water, many of which in many parts of the world have caused disasters. The downstream flow contributes to a depletion of land, so that streams, reservoirs and dams are also silted. In two factors, deforestation causes floods. Firstly, the soil is more likely to be completely saturated with water with a smaller' tree fountain' effect. Second, deforestation also leads to a compaction of soil which cannot absorb rain.

2.3.2.3. Decreased biodiversity, habitat loss and conflicts

Forests in particular in the tropics act as biodiversity storage areas, and deforestation, fragmentation and destruction thus threatens the entire ecosystem and habitat of migratory animals, including endangered species, some of them yet to be categorized. Two thirds of all species are found in tropical forests and 65% of the world's 10,000 endeavour species are endangered (Myers, et al., n.d.). Another detrimental consequence of deforestation is that human-animal disputes are gradually impacting conservation performance in order to alienate the interest of the public in conservation.

2.3.2.4. Social consequences

The introduction of civilization typically means for indigenous peoples the destruction / change of their traditional way of living and the dismantling of social structures primarily with their removal from their ancestors. The interference of externals is disruptive to traditional life patterns, traditions and religious beliefs, which increase infrastructure growth such as road construction, which also leads to border expansion with social and land conflicts. (Schmink, 1992).

2.3.3. Strategies to reduce Deforestation

2.3.3.1. Reducing emissions from deforestation and forest degradation

Many international organizations, including the United Nations and the World Bank, have been implementing deforestation mitigation programs primarily through REDD, which use direct monetary and other incentives to enable developing countries to restrict / reduce deforestation.

2.3.3.2. Improve the information base and monitoring

Information is inadequate on the global biodiversity distribution and forest poverty. It seems easy to know how much wood, where and what it consists of, but, unexpectedly, most basic information is not always available. By first realizing, it is not possible to maintain a forest environment correctly. New remote sensing technology makes deforestation hotspots viable and affordable.

2.3.3.3. Increase investment in research, education and extension

Stakeholder education and awareness helps people understand how negative environmental impacts of deforestation and forestry activities can be avoided and minimized and, if possible, effective steps taken. Research confirms and contributes to understanding the issue, its cause and its mitigation. This arena is behind the backdrop because of a lack of funds and investment. In the general population, forestry and forestry are not recognized and knowledge is missing.

2.3.3.4. Promote sustainable management

This must be sustainable ecologically, economically and socially in order to encourage sustainable forest management. Encouraging sustainable sustainability does not harm the forest's natural principles and should be strengthened where possible. That ensures that silviculture and managing should not reduce biodiversity, soil erosion should be managed, no loss of soil fertility, maintain water quality on-and off-and preserve forest health and vitality.

2.3.3.5. Increase the area and standard of management of protected areas

For any effort to maintain biological diversity it is important to have protected areas (Myers, et al., n.d.). Furthermore, protected areas are not enough to preserve biodiversity alone. In addition to and as part of a more holistic biodiversity conservation plan.

2.3.4. Technologies Associated with Deforestation Prediction

As we have discussed about the impacts of deforestation already it is essential that we take immediate remedies to control this. When we consider Sri Lanka last few years' deforestation created many new on every media. It became a political and social problem at the end. Therefore, as a student engage in technological world this attempt is to find a solution for deforestation in Sri Lanka.

In this part of the literature review It will intend to discuss about the technologies that we are going to use in making our research project successful. Also It will discuss about the existing projects have been done around the world and how we are going to introduce a unique algorithm for our research project in latter part of this section. When we talk about technologies machine learning concepts are considered in making the prediction algorithm. To find the deforestation details Geographic Information Systems (GIS) and remote sensing techniques with the use of satellite images will be taken in to consideration. By using these, deforestation can be mapped considering various factors which depends the relationship between deforestation and environmental and socioeconomic factors. Accordingly, from next chapter onwards we will be discussing in details about the machine learning methods and other technologies we are planning to use.

2.3.4.1. Introduction to Prediction Algorithms

Today with the accessibility of the earth observation data, open data and open source analysis tools it has been easy to figure out forest cover change more accurately that in the past. However, with the increasing volume and complication of data needs necessary techniques that can handle large multi-source and multi scare datasets. During the past there have been many researched went on to estimate the deforestation risk using several underlying drivers. They were based on many approaches like statistical approaches, machine learning(ML) and spatial modelling (Mayfield, 2015). Traditional methods are usually based on logistic regression, generalized linear or generalized additive models unlike machine learning techniques like Maxtent (Amuchastegui, et al., 2014). Although conventional methods are very useful in

estimating deforestation risk, this needs a lot of technical expertise which often lacks in developing countries (Romijn, et al., 2015). Sometimes it's not possible to know all the drivers and their relative influence, but in particular machine learning algorithms have capability to depict and generalize the relationships of data. Therefore, it has ability to even work with large datasets. These machine learning approaches have been found to be suitable for situations where data concerning observed pattern are available but it's not abundant for theory concerning process. (Anon., 2014) Therefore, compared to statistical approaches such algorithms do not need strong mathematical assumptions to describe the relationship with target variable and predictor variable. In this proposed research we suggest and find a new approach to figure out high risk deforestation areas and causes for it.

2.3.4.2. Bayesian Model for Deforestation

This is machine learning method which identifies the key factors on its own and makes predictions using available spatial data. When we go through previous projects that has been done it's the efficient Bayesian multivariate classifiers (EBMC) are used to learn feature based Bayesian networks(BN) of deforestation where multisource data are available. This model learns using the algorithms K2 and BDeu, which were used to predicts deforestation and also it gives us a directed acyclic graphical view of the main factors that it will interact on. These were then compared with constraint and knowledge based BN's which has been developed using common EBMC selected variables. Considering the previous projects which has been conducted, knowledge-based and constraint based BN's have performed well by interacting with variables to develop a casual structure. (Dlamini, 2016)

a) Data for the model

It will expect to obtain a publicly available dataset for the country which is based on analysis of satellite images for the period between 2005 to 2019. Binary Maps of deforestation can be created by identifying pixels with a change from non-zero to zero percent tree cover. Conversely growth of the forest too can be mapped. Through a review of literature on deforestation we expect to find the variable that will be directly effect on deforestation in Sri Lanka. Then with use of QGIS/ArcGIS software we expect to do all the spatial analysis and visualization.

b) Bayesian networks

The aim of using Bayesian Network in modelling the deforestation is to derive a directed acyclic graph showing the relationship of target variable(deforestation) and predictor variables represented as nodes connected by edges following dependencies. Bayesian Networks are based on interpretation of Bayes' theorem which represents deforestation by D and evidence by e such that P(e) = 0 and P(D) = 0, then:

$$P(D|e) = \frac{P(e|D)P(D)}{P(e)}$$

Probability of deforestation event D is conditioned upon evidence e is considered as it's likelihood P(e|D) times its initial probability to any evidence P(D) and it's divided (e) to normalize. Therefore, conditional probabilities of al hypothesis adds to 1. P(D|e) is usually known as the conditional probability or posterior probability of deforestation D, when P(e|D) is likelihood of e given in an event of deforestation. In situations where P(D|e) = P(D) and P(e|D)=P(e) the events D and e can be considered independent.

Bayesian network learning includes of studying both the parameters and the structure or directed acyclic graph(DAG). Therefore, using this flexible BN's method it is possible to have a probabilistic graphical modelling of highly interacting variables. In Bayesian network the algorithm approximates the probable graphical model by searching the space of possible networks through single-arc changes that enhance some score. (Dlamini, 2016)

c) Efficient Bayesian Multivariate Classifier (EBMC) Approach

Usually there could be several potential predictors of complex deforestation which it makes challenging to understand which of them are better predictors. Therefore, it's needed to find highly predictive variable while also modelling the interaction among them. EBMC algorithm is a good solution for that as it selects predictors from high dimensional data, and then those elected predictors uses in a Bayesian Network classifier for prediction purposes.

By this algorithm it searches for number of BN models and it finds the best one to train the data and given prior probabilities (Cooper, et al., 2010). It will initiate by scoring all models where single variable is the parent of the deforestation node D and by using either a K2 or a BDeu score. The model scores the highest will be the initial model. After that algorithm will decide which variable when added as a parent of D to the first model having the highest scoring second model. If the second model gets a higher score than the first model, then a new model is taken. This process will continue to repeat by using a greedy search and adding variables to the model while there's an increase in K2 or BDeu score. If no longer any increase is happening, then algorithm will search for a variable which is when removed increase the score and by that way it will find and exclude the variable whose removal increases the score most. Then variables in the final model children of D will be made to create edges or links between them by the algorithm. The EBMC algorithm thereby create the resultant BN which is an augmented naïve BN which shows the interactive between the edges or probabilistic relationships among the variables that is being selected by the algorithm. (Cooper, et al., 2010)

Therefore, EBMC algorithm searches over hybrids of Bayesian rules and augmented naïve by the scoring the method with use of BDeu or K2 scoring. Hence that is clear this method will always be improving the prediction of target variable with the use of selected predictor variable. Also by this greedy search it will increase the score of a Markov blanket of the target node D, which is well known for making highly effective probabilistic predictions. This method also seems to unique compared to other BN-learning algorithms as the ordering of the features in not required.

Also there is inductive causation (ICS) algorithms which will consider the latent variables to produce a BN with undirected and bi-directed arcs. This algorithm is varies compared to other algorithms like Spirtes, Glymour, Scheines as those algorithms first creates an undirected graph that models dependencies among variables. Therefore, ICS algorithm is improved for recovering the casual structure and not finding an optimal classifier. (Dlamini, 2016)

2.3.4.3. Artificial Neural Networks(ANN)

These are nonlinear mapping structures and it is based on functions of human brain. ANNs' are advantageous when we consider its capability to nonlinear functions, perform model free function estimation, to learn from data relationship unseen situations and to generalize them. ANNs have ability to work with any data, therefore it is flexible and universal function. ANN can create powerful tools for models even when data relationships are unknown. ANNs have been one of the biggest interest during last decade and have been used for variety of domains such as image recognition, medicine, molecular biology and in recently the technology has exposed to ecological and environmental sciences. In recent times ANNs have been used in environmental sciences to predict species distribution and to find diversity as a function of environmental variable, predict floods, rice crop damages, air quality parameters and eco system characteristics from remotely sensed data. (J.F. Mas, 2003)

Multi-layer perceptron(MLP) could be considered as one of the most popular ANN architecture which is used in currently for many purposes which is based on a supervised procedure which means the model can create a model based on examples of data with known outputs. These examples will do the training and they are together assumed to contain information needed to build relation. MLP is capable of building complex relationships with variables. Therefore, MLP can successfully give out predications for given set of input objects or for an object. (J.F. Mas, 2003)

MLP has a layered feed forward neural network which consists of nonlinear elements(neurons) that arranged in successive layers. In this methods information is flowing from input layers to output layer in a hidden layer unidirectional. Input variable values will be placed on input units once the network is executed and after that hidden and output layers will begin to execute. By the considering the weighted sum of the outputs of the in the earlier layers, each of those layers will calculates its activation value. This value will be passed through the activation function to give out the outputs of the neuron. Finally, when entire network has been executed, the final outputs in the output layers will be consider as the output of entire network.

If this learning procedure give a wrong output, weights are corrected to give out a more accurate answer by reducing the error. A well-known neural network training algorithm is backpropagation. Levenberg–Marquadt, a faster and second order algorithm is an advanced

non-linear optimization algorithm. But it is limited to smaller networks using sum-squared error function (Bishop, 1995).

In minimizing errors, the biggest problem is over-learning or over-fitting. A networks having more weights will produce a complex function and thereby it will be disposed to over-fitting. (Bishop, 1995) To avoid this cross-verification is used. Some verifications sets are not used for training, but it will keep for an independent check. If the verification error starts to rise, that implies network is trying to over-fit the data and thereby training will be stopped.

In this output units' task is to give out a strong signal if that case belong to the class or otherwise a weak signal. This activation value studies as a fuzzy membership which depicts as the measure of probability in belonging to the class. Therefore, these activation values show the propensity of deforestation when ANNs are used to model deforestation. (J.F. Mas, 2003)

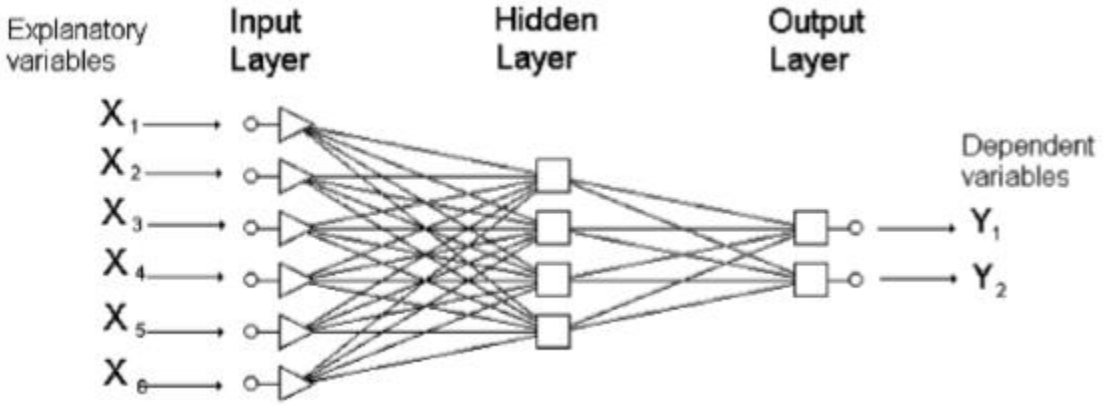

Figure 4: Illustration of a three-layered perceptron

There are seven different types of Artificial Neural Networks (Davydova, 2017).

They are;

1. Multilayer perceptron (MLP)
2. Convolutional neural network (CNN)
3. Recursive neural network (RNN)
4. Recurrent neural network (RNN)
5. Long short-term memory (LSTM)
6. Sequence-to-sequence models
7. Shallow neural networks

2.3.4.4. Logistic Regression model(LRM)

Another important method used to predict the deforestation is by logistic regression. This method also has been used for a long period and it has used for many number research projects. This is a variation of ordinary regression which is used when dependent variable is a dichotomous variable. In this the dependent variable is either binary presence or null as it considers 1 = forest change and 0 = no change, for the period of 2005 to 2019. Using this method gives the probability of deforestation as a function of the explanatory variable. i.e probability of deforestation for each pixel is a function of values in which the other variables that will affect have for the same pixel. It is mentioned this function as a monotonic curvilinear response which is bounded between 0 and 1. This can be defined in the form of; (Arekhi , 2011)

$$p = E(Y)\frac{exp^{(\beta_0+\beta_1X_1+\beta_2X_2+\cdots)}}{1+exp^{(\beta_0+\beta_1X_1+\beta_2X_2+\cdots)}}$$

Where:

P – probability of deforestation in the cell
E(Y) – The expected value of binary dependent variable Y
β_0 – constant to be estimated
β_i's – coefficient to be estimated for each independent variable X_i

The logistic function can be transformed into a linear response as show below:

$$p' = \log_e \left(\frac{p}{1-p} \right)$$

Hence:

$$p' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots$$

This transformation show in the figures Is known as logit or logistic transformation. This transformation allows linear regression to estimate each β_i. Thereby we can obtain a probability score for each pixel as the final result since the observation is a pixel.

In Logistic Regression method the significance is of the coefficient β_i is tested using the Wald test which taken from comparing the maximum tendency estimate of every β_i estimated standard error. Coefficient will be divided by its standard error to find it. Therefore, when the relative error is high the Wald statistic is low. So it tells how important the each and every variable. It will be more significant when the absolute value is greater. (Eastman, 2006)

To depicts the correct meaning of the first equation it is needed to use the coefficients as a power to the natural log(e). By doing this it will shows the probability that an event could occur divided by the probability that it will not occur. If this gives a positive coefficient it says more probable to occur the event as its transformation to log value is greater than one. Also if it gives a negative coefficient the event occurrence decreases and the transformed log value will be less than one. If the coefficient is positive, the probability plotted against the values of an independent variable gives an S shaped curves and a mirror image will be produced it's a negative coefficient. (Arekhi , 2011)

2.4. Evaluation of the Model

In this research evaluation has to be done in two phases. First we have find the Normalized difference vegetation index(NDVI) which will have discussed further in methodology and then we have to check obtained data are true or not. Then we can evaluate the prediction model. Evaluation of NDVI can be done by using taking the stats which have been published by the Forest department and other authorities. For the evaluation of the model we can take the methods we have been used by the previous researches. For model validation it is used the relative operation curve (ROC) which asses the accuracy and gives out the probability of deforestation. It is considered as a very effective method of evaluating a statistical model (Hu & Lo, 2007). This curve represents by true positive and false positive fractions. In the vertical axis there will be true positives and rate of false positives on the horizontal axis. The ROC is equal to 1 when the sequence of suitability values matches. When the model performance improves this will move upper left corner and there by area under ROC also increases. ROC can have two situations, one is when there's change in both actual and model scenarios it is a positive value and if there's no change in actual and but in the model scenario it will be false positive.

2.5. Personal Reflection

There have been several projects been done using the above techniques in several parts of the world and they have been served the countries in many ways. Here in this research it is supposed to analyse the deforestation by means of NDVI values that we are going to find in the research. Based on the accuracy satellite images NDVI values can be found. This data will be the input data what we are going to apply for the prediction model. Therefore, successful completion of this phase is necessary to work on the prediction model.

When we take the topic deforestation it is a broad topic and has different sides of it to discuss about. In this research it is mentioned some direct causes of deforestation, but their indirect causes too which are relevant. Some of them are colonization, debt burden of a country, land rights and unbalanced land distribution and resources. Also we can take into consideration the

economic causes like development/land conversion value, fiscal policies, markets etc. Corruptions within a government also can be considered as political cause for deforestation. So when we consider all of these it is very broad topic to discuss about although we consider only direct causes which can be easily recognized.

Therefore, to identify deforestation and some of it causes we can use modern technology. As we have discussed above satellites images and GIS along with remote sensing can be used. By studying these satellite images thorough, we can come to very appropriate conclusions. This is really very important, because we have seen general folk in the society talks about deforestation in Sri Lanka even without seen what is going on. In the past few years it led to the racial problem as well. So we can understand how important it is and also we can supply these details to Non-government organization which stand against deforestation to protect the environment.

When consider prediction model the use of these machine learning methods can be differ from project to project based on the interest and available amount of input data. For example there have been projects done using artificial neural networks in Gulf of Mexico (J.F. Mas, 2003), Guatemala (Renno, et al., 2006), Amazon (Liang & Tang, 2017) (Xu & Zeng, 2017) etc. Also Bayesian network has been used to study about deforestation in Swaziland (Dlamini, 2016). Similarly, logistic regression techniques also been used for long period of time for different purposes and it can be considered as traditional method. When we compare Artificial Neural Networks and Logistic Regression methods, ANNs are new methodological tool based on nonlinear model. Therefore, they appear better at making predictions with the data provided. Through the literature review we have discussed in detail differences between each of the machine learning algorithms. As said before traditional methods require a lot of technical expertise. Also regression methods have limitations, it's solely depend on the sample data set and they cannot adjust the missing data sets. But when you take Bayesian networks they are capable to overcome the limitation problem in regression methods. Also these Bayesian networks are capable to work with quantitative data and qualitative data simultaneously. An advantages of this method compared to regression method is, no priori hypothesis is made about the modelled relationship. When we compare Bayesian Network and Artificial Neural Networks, they both are similar and giving out directional graphs. However, when we compare the two structure major difference we can observe is Bayesian networks have an intrinsic meaning while neural networks do not.

2.6. Summarization of gap

First thing about this research project can be considered as every data on this project is unique. Analysis of the images are unique to the project. Therefore, the input data that we are going to introduce in to the algorithm will be unique. Also we are giving approach to the algorithm based on the NDVI results that we obtain.

Also when we consider Sri Lanka still there are no any project been conducted to predict deforestation in Sri Lanka although there's a big problem in the country. Therefore, during this project, we take prominent forests in Sri Lanka into consideration and hope elaborate how deforestation is happening around those forest areas in Sri Lanka. For predictions there are several machine learning techniques we can use as we discussed in the literature review. For this research project Artificial Neural Networks will be used to make the predictions.

2.7. Chapter Summary

As the deforestation have become a serious problem in Sri Lanka it is very clear that we need to find solutions using new technologies how to control it by identifying the causes for it. Through this chapter we have identified the main concepts that have been used in the world so far to overcome the issue. We have identified the common foundation of each the models in statistical pattern recognition and got a brief idea about the procedure how it processes. We also described how Bayesian network, Artificial Neural network and logistic regression can be evaluated. Now we have to look ways and mean in which these methods are applicable to Sri Lanka's deforestation. As we discussed at the beginning Sri Lanka has unique causes for deforestation when we consider last few years. With the use of modern day technologies above we can find the other causes for it and take necessary remedies for deforestation.

3. Methodology

Content

3.1. Chapter Introduction

In this chapter we are going to discuss about methods that are using in this research project to achieve our research objectives. By studying the previous research projects related to the topic it had been possible to commence the work and learning more about related technologies and It had been possible find many ways to implement it. Therefore, during this chapter, it is expected to discuss more on technologies which have used in working on the implementation. Also it will take into account other probable methods that we can used in making this research project successful.

3.2. Research Execution

As this research is based on assumptions the existing findings related to the research can be verified mathematically, experimentally and on philosophical basic is positivisms. Hence this research follows experimental Research Methodology, which will be collecting and interpret data based on the it's experiments and observation. The validation will be done by experimentation and findings of the research will be observable and quantifiable.

3.3. Research Design

This research is based on Experimental methodology and it needs to follow the below steps in building the final solution. First of all, it is necessary to do a proper literature review to identify what are the existing deforestation predicting systems already have developed in different parts of the world. Then it is necessary to understand the key findings of those methods that can take into consideration. There we would be able to understand the existing methods of measuring deforestation in by means of mathematical formulas and technical approaches. Thereafter before the implementation we can find information on what is the present situation of forests in Sri Lanka by finding facts from different sources including Forest Department in Sri Lanka. Then we have to identify the areas in Sri Lanka that are vulnerable to deforestation as the study area. Firstly, we have to find out the historical data in the selected area and download the available satellite images accordingly from available sources. After completing that we can introduce these satellite images to QGIS software to render satellite images using different bands by remote sensing techniques. Once they are processed, it is needed to create database by storing set of rendered images of the events. After completing that, we should analyse stored data, identify the necessary parameters and develop the model. IR image data here can be considered as one of the parameters. As we have discussed in literature review we can use forest cover index and NDVI (Normalized difference Vegetation Index) for calculation of deforestation that has occurred in the relevant areas. Once the results from these methods are obtained we have to get them validated by comparing the data provided by authorities engaged in the relevant field (Forest Department in Sri Lanka).

The model should develop with improvements until the validations become true. When the model gets valid store satellite data for future map predictions and from there we can obtain the training data for ANN model. Then as we discussed in the literature review, it is necessary to categorize the data in to three as The training data, verification set and Test sets and proceed with developing the model. Once when the training data is processed we have to assess the model by using several ways like we have discussed in the research methodology.

3.3.1. Satellite Images

In this research satellite images are the main source of input to conduct the research. Therefore, it is needed to obtain satellite images from 2000 to 2020. For this purpose, satellite images from sentinel satellites and Landsat satellites are used.

In the past few decades, satellites have been used for collecting a broad variety of knowledge about the Earth's atmosphere, from military applications to global weather records, tectonic activities, atmosphere ecosystems, ocean currents and temperatures, polar ice variations and emissions, etc. The development of hyperspectral and multispectral sensors worldwide has contributed to the satellite imaging technology, a device that can be used to map complex materials by identifying the chemical and material bonds which are from satellite and airborne sensors. Remote sensing is the art and science of measuring an object or the environment without physical contact. Remote satellite sensing allows data collection on dangerous or inaccessible areas. Two major types exist: passive remote and active remote sensing. Passive sensors detect naturally occurring radiation emitted by the object or its surroundings. The most common radiation source detected by passive sensors is reflected sunlight. Types include videos, infra-root, load-coupled devices, and radiometers for passive Remote Sensors. At the other hand, active collects emit energy for the scanning of objects and areas where a sensor senses and tests the reflections or spreads of radiation from the target. RADAR and LIDAR can be considered as the examples of active remote sensing and it is capable of measuring the time delay and determine the location, height, speed and direction of the object. (Upadhyay & Gupta, n.d.). For this research project satellite images were obtained mainly from Copernicus open Access Hub and USGS earth explorer. We have acquired Sentinel and Landsat Images from those resources. Hence It will be discussed briefly in the coming topics about these resources.

3.3.1.1. Copernicus open Access Hub

This is one source we could obtain the satellite images. Copernicus is an EU initiative, offering practical information on land, ocean and atmosphere around the world, supporting environmental and health policies and meeting the needs of people and providers of services. Under the Copernicus Space Portion, the ESA is developing a dedicated satellite family named Sentinels to meet the Earth Observation requirements of the program. The data obtained from these activities is systematically linked down and processed by the Sentinel ground segments to operational consumer items. The ESA Sentinel Data Access Program is designed to retrieve products from the related ground section of the Sentinel-1,-2,-3 (land) and -5P and to make available to users products to be downloaded from dedicated access points. (Castriotta, n.d.)

3.3.1.2. USGS Earth Explorer

This is another source that we obtained satellite images. Earth Explorer (http:/earthexplorer.usgs.gov) provides online search, browsing, export of metadata, and downloading of earth science data from the U.S. archives. Geological Survey. Geological Survey, USA. Earth Explorer offers the most up to date user interface with JavaScript, Hypertext Preprocessor (PHP) and Oracle's advanced space engine, state of the art libraries. (USGS, n.d.)

Earth Explorer's main features include:

• Fast, geospatial search engine

• Map viewer can be used to find overlay footprints and browse overlays

• Simple, fast Graphical User Interface (GUI)

• Data access tool to search and discover data

• Textual query capability

• Save or export queries and results

• Request on-demand products

3.3.1.3. Sentinel Data

Sentinels are a family mission launched by Earth Observing System(ERS) and each of these missions will focus on different aspects on earth including Land monitoring, oceanic and Atmospheric. It has six mission where oldest is the Sentinel 1 and latest the Sentinel 6. Since this research engages in land monitoring, images from sentinel 1 and sentinel 2 are considered. Sentinel 1 has earth observing data from April 2014 onwards and sentinel 2 consists of earth observing data from June 2015 onwards. These sentinel data can be obtained from Copernicus Open Access Hub and USGS Earth Explorer. For this research paper it is expected to use Sentinel 2 images. Sentinel 2 is a combination of two satellite which was launched by European Space Agency (ESA) and it can map the geophysical features of the land surface. Sentinel 2 satellite consists of Multispectral Imager (MSI) and this sensor will capture and deliver 13 spectral bands. Each of these bands are with pixel size 10,20 or 60 meters. Among these bands B2(blue), B3(Green), B4 (Red) and B8 which consists of infra-red have 10-meter resolution. The bands B5 of red edge, B6, B7 and B8A of near-infrared (NIR) and set of B11 and B12 bands of short-wave infrared (SWIR) have 20-meter resolution. The bands B1 and B10 which of that represent coastal aerosol cirrus band respectively have 60-meter resolution.

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

Figure 5: The bands of each Landsat satellite and descriptions of how each band is best used (Anon., n.d.)

3.3.1.4. Landsat data

Landsat mission is the longest running enterprise to collect satellite imagery of Earth. It launched the first Landsat satellite on July 23rd 1972 which was under the name of Earth Resources Technology Satellite and later renamed to Landsat. The last Landsat satellite was launched on February 11, 2013 which is considered as Landsat 8. Landsat has millions of images and they are unique resource for global change research and used for application is agriculture, cartography, regional planning, geology etc. These images can be obtained from USGS Earth Explorer. For this research we will be using satellite images Landsat 7 and Landsat 8. These images have been used by government, commercial, industrial and educational communities all over world for different purposes. They are widely used for research projects in agriculture, mining, forestry and land cover change detection.

Landsat 7 satellite images contains Enhanced Thematic Mapper Plus (ETM+). In these images contains seven bands like in previous mission with a newly added panchromatic band 8 which gives higher resolution of 15m. Compared to Landsat 7 the Landsat 8 images are narrower spectral bands, relatively more precise geometry, improved 12-bit radiometric resolution and improved calibration.

Uses of Landsat bands

Band name	L8 OLI/TIRS	L7 ETM+	L4-5 TM	L4-5 MSS	L1-3 MSS	Description of use
Coastal/Aerosol	Band 1	--	--	--	--	Coastal areas and shallow water observations; aerosol, dust, smoke detection studies.
Blue (B)	Band 2	Band 1	Band 1	--	--	Bathymetric mapping; soil/vegetation discrimination, forest type mapping, and identifying manmade features.
Green (G)	Band 3	Band 2	Band 2	Band 1	Band 4	Peak vegetation; plant vigor assessments.
Red (R)	Band 4	Band 3	Band 3	Band 2	Band 5	Vegetation type identification; soils and urban features.
Near-Infrared (NIR)	Band 5	Band 4	Band 4	Band 3	Band 6	Vegetation detection and analysis; shoreline mapping and biomass content.
	--	--	--	Band 4	Band 7	
Shortwave Infrared-1 (SWIR-1)	Band 6	Band 5	Band 5	--	--	Vegetation moisture content/drought analysis; burned and fire-affected areas; detection of active fires.
Shortwave Infrared-2 (SWIR-2)	Band 7	Band 7	Band 7	--	--	Additional detection of active fires (especially at night); plant moisture/drought analysis.
Panchromatic (PAN)	Band 8	Band 8	--	--	--	Sharpening multispectral imagery to higher resolution.
Cirrus	Band 9	--	--	--	--	Cirrus cloud detection.
Thermal (T)	Band 10	Band 6	Band 6	--	--	Ground temperature mapping and soil moisture estimations.
	Band 11			--	--	

Figure 6: The bands of each Landsat satellite and descriptions of how each band is best used (U.S. Department of the Interior, 2016)

3.3.1.5. Landsat vs sentinel data

When we consider both these satellites, they have same performance in terms of accuracy, with accuracies of 96.7% and 95.7% for sentinel-2 and Landsat 8 respectively, which is considered to be the latest of their relevant missions. Nevertheless, when compare these Landsat-8 has 36.9% more area of selective logging compared to sentinel-2. (Lima, et al., 2019)

3.3.2. Remote Sensing Phenology

This is the main technology that is used in satellites and it is by which we will be able to find the NDVI of images. Remote sensing can monitor the process occurring on the earth by using satellite images and analysing them. Remote sensing can be used for both long time and short time processes that are happening on earth. In this research we are focusing on deforestation which is considered as a long term process. Because we are expecting to analyse the degradation of vegetation or change in land use for during past two decades. (Meneses-Tovar, 2011)

Remote sensing is business for people every day therefore. Look at a speaker during class, watch cars drive in front of you, both of them are the sensing behaviours of the human eye. Human eyes record the solar light that represents certain objects and the colours, grey shades and changes in strength of your brains perceive it. The next move is to turn these data into usable knowledge. Nevertheless, the human eye is limited to a small portion, i.e. about 400 to 700 nm, of the total electromagnetic spectrum. Various types of instruments and equipment are used in remote sensing in order to make electromagnetic radiation noticeable for the human eye outside of this range. Remote sensing has now become an important part of a wide range of environmental disciplines including geography, geology, zoology, agriculture, forestry, meteorology, oceanography and civil engineering. Since the first earth-watching satellite remote sensing has been used increasingly to collect information on environmental processes such as agriculture. (Jong, et al., 2004)

3.3.3. QGIS

QGIS was previously called as Quantum GIS and this will be the software that we are going to work through this whole project. This tool supports both raster and vector layers. Vector data is stored as a point, line or a polygon. In QGIS multiples formats of raster images can be used. This software supports shapefiles, coverages, personal geodatabases, dxf, MapInfo, PostGIS and other formats. To make use of data from external sources web services including Web Map Service and Web Feature Service are used. QGIS can act as an alternative Graphical User Interface for collection of GRASS GIS geospatial processing functions that will be depending on the host system configuration. Also this is with a "plug- in" architecture that give extensions to the core functionalities with already available plug-ins including GPS integration support, accessibility to OpenStreetMap data servers and data transformation tools. (Anon., n.d.)

3.3.3.1. Semi-Automatic Classification(SCP) Plugin

This is main plugin or tool that will help in the process of analysing the images. It is available in QGIS to download and by this tool we can do atmospheric correction and it is responsible many of the processes like pre-processing of images, post processing of classification and raster calculation. SCP is a free open plugin and it is available in QGIS. This tool includes supervised and unsupervised classification of the remote sensing images.

3.3.4. Normalized Difference Vegetation Index (NDVI)

NDVI stands for Normalized Difference Vegetation Index. Multi Spectral Sensing data techniques is used by NDVI to find vegetation index, water resources, hilly areas, open areas, agricultural areas, thin and thick forest areas. It is done by using band combination in remote sensing methods. Hence land resources can be interpreted by using Normalized Difference Vegetation Index. This can be done by collecting remote sensing data from Satellite images along with NDVI and DEM layers and processing them to perform multi-source classification. The method of change detection is called as NDVI differencing. (Gandhi.G, et al., 2015).It is estimated by analysing the satellite images to give indication about the green-ness. That is, it will give us information about the degradation of vegetation by analysing the satellite images and it will indicate by taking the green-ness as the vegetation. If there's a degradation in vegetation, the green-ness of the images have decreased and it will show a decrease in NDVI value. The NDVI value is a dimensionless index, so it's values ranges from -1 to +1. (Meneses-Tovar, 2011). This simulation helps detecting the surfaces of visible area which highly useful for authorities that makes policies and for decision makers. This vegetation analysis can be further used to predict future effects like disasters that could originate due to deforestation and also it can be furthermore use to decide protecting strategies.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

Figure 7: Equation for NDVI calculation

3.3.5. Evaluating NDVI Maps

Once the NDVI maps are found we have to evaluate them before implementing for the predictions. This can be done by using the previous researches that have been done in the relevant forest areas. Also we can obtain the forest maps from the forest departments and evaluate them. This can be considering as a core part before moving in to predictions.

3.3.6. Artificial Neural Network for Deforestation

This is by which that we are going to have an algorithm for the prediction model. According to our research we are going to use NDVI as our input data for the model. Also we are going to use spatial variables as the input weights and that will be the parameters which predictions based on. For examples spatial variables can be area of areas which are less green.

Around the world ANNs have been widely used for deforestation finding work with different parameters. By studying previous researches it was very clear how these techniques have been use to find susceptible places for deforestation and to find its main determinants or deforestation variables. These researches have helped to improve local management and conservation forests. Due to the methodological stability of the long application history of these models, it is ideal for decision makers to analyse about the matters. As we discussed (Upadhyay & Gupta, n.d.)ANNs are used in 1990s for related work in ecology and environmental sciences. One of the objective of this research is to construct an artificial neural network, evaluate and then find the best model to find deforestation in Sri Lanka in coming decades. Through the literature review of the research it is very clear ANNs are the ideal for this purpose compared to other techniques as it can produce a nonlinear model. Therefore, it is expected to use ANN techniques in making the predictions.

a) Method for Building the Prediction Model

These are the five step that are included in making this model; (J.F. Mas, 2003)

1) Explanation of deforestation maps which are obtained by overlaying maps of forest cover from more than one point in time
2) Finding the best predictor variables by quantifying the relationships between deforestation and related causes
3) Calibration of the model
4) Simulation (elaboration of the maps which predicts the deforestation for the particular period)
5) Check the model performance and reliability by comparing the actual and predicted deforestation

3.3.7. Calibration

Variables which we have selected in the previous step is used to elaborate the ANN training data. We've to divide the data into three sections.

1) The training data
2) Verification set
3) Test sets

Test sets are not used to adjust weights by training algorithms. Test sets are used to track the networks error performance and if there's any over learning it will stop training. Test sets are used to have independent assessment of network's performance when entire design is done and they are not used for training purposes at all.

3.3.8. Elaborating & Evaluating prediction maps

By these NDVI maps, the model and the trained data from the 2000-2020 are planned to use to predict the deforestation maps of highly vulnerable areas for the period of 2020-2030. The output of the MLP is an activation value that will show you for each pixel the probability of deforestation. From that map we can identify the gradation of getting deforested.

For the evaluation of prediction maps, it has planned to consider maps from 2000-2010 and predict deforestation maps for 2020. So once we found the maps we can evaluate them by taking current satellites images and as we discussed the literature review ROC curve can be considered. Also for more clarification maps from Forest Department can be obtained.

3.4. Assumptions

In this research we are considering Normalized Difference Vegetation Index(NDVI) as input data for the prediction model. This index contains information about live green vegetation. Therefore, we have to make an assumption as we only considering the forest area, the greenness shown in the images are all about the forest density.

Also for easier understanding and lack of resources made us to divide project into five years' periods. Hence we can assume during these periods the greenness that we have found kept unchanged.

3.5. Chapter Summary

During this chapter we have gone through the methods we are expected to follow. At first technologies that are going to use have been described. Then a description about the materials that are going to use and how they are obtained are given. After discussing about the materials tools which will used to process the materials have been explained. Hence from this chapter mythology of this research are project are elaborated.

4. Project Management

Content

4.1. Chapter Introduction

This chapter will bring you about the planning and management of this research project. First it will show you how the dates are allocated to complete the tasks on time. It has further elaborated through graphical representation using a Gantt Chart. Then we discuss about the risk assessment of the project and finally the following of the BCS code of conduct will explained.

4.2. Time Management

The below figure shows you the time plan for the research project. It contains all the details about the durations for the tasks with start and ending date.

ID	Task	Duration	Start	End
1	**Intial meeting with supervisor**	0	03/09/2019	03/09/2019
2	**Project Idea Discussion and finalization**	19	07/09/2019	26/09/2019
3	**Project Initiation Document**	17	03/10/2019	20/10/2019
	Research on project background	3	03/10/2019	06/10/2019
	Research on problem	3	06/10/2019	09/10/2019
	Research on Existing Systems	2	09/10/2019	11/10/2019
	Preparation of PID	14	05/10/2019	19/10/2019
	Submission of PID	0	20/10/2019	20/10/2019
4	**Literature Review**	24	21/10/2019	14/11/2019
	Comprehensive Research on existing systems	15	21/10/2019	05/11/2019
	Preparation of Literature Review	8	05/11/2019	13/11/2019
	Submission of Literature Review Document	0	14/11/2019	14/11/2019
5	**System Requirements Specification**	60	15/11/2019	14/01/2020
	Preparation of Requirement Specification Document	20	15/11/2019	05/12/2019
	Evaluation of Requirement Specification Document	10	06/12/2019	16/12/2019
	Submission of Requirement Specification Document	0	14/01/2020	14/01/2020
6	**Interim Report**	30	15/01/2020	14/02/2020
	Preparation of Document	29	15/01/2020	13/02/2020
	Submission of Document	0	14/02/2020	14/02/2020
7	**Development of Prototype**	60	14/01/2020	14/03/2020
	Design Prototype	10	14/01/2020	24/01/2020
	Development of Prototype	45	24/01/2020	09/03/2020
	Testing and Evaluation	5	09/03/2020	14/03/2020
	Finalize the prototype	0	14/03/2020	14/03/2020
8	**Prototype Report**	10	15/03/2020	25/03/2020
	Preparation of Document	9	15/03/2020	24/03/2020
	Submission of Document	0	25/03/2020	25/03/2020
9	**Final Report**	46	25/03/2020	10/05/2020
	Preparation of Draft document	6	25/03/2020	31/03/2020
	Evaluation of Draft Document	10	31/03/2020	10/04/2020
	Preparation of Final Document	30	10/04/2020	10/05/2020
	Submission of Document	0	10/05/2020	10/05/2020

Table 3: Table for Time Management

4.3. Gantt Chart

The below Gantt chart will show you the time graphical representation of the time plan of this project which is based on the time plan we discussed above.

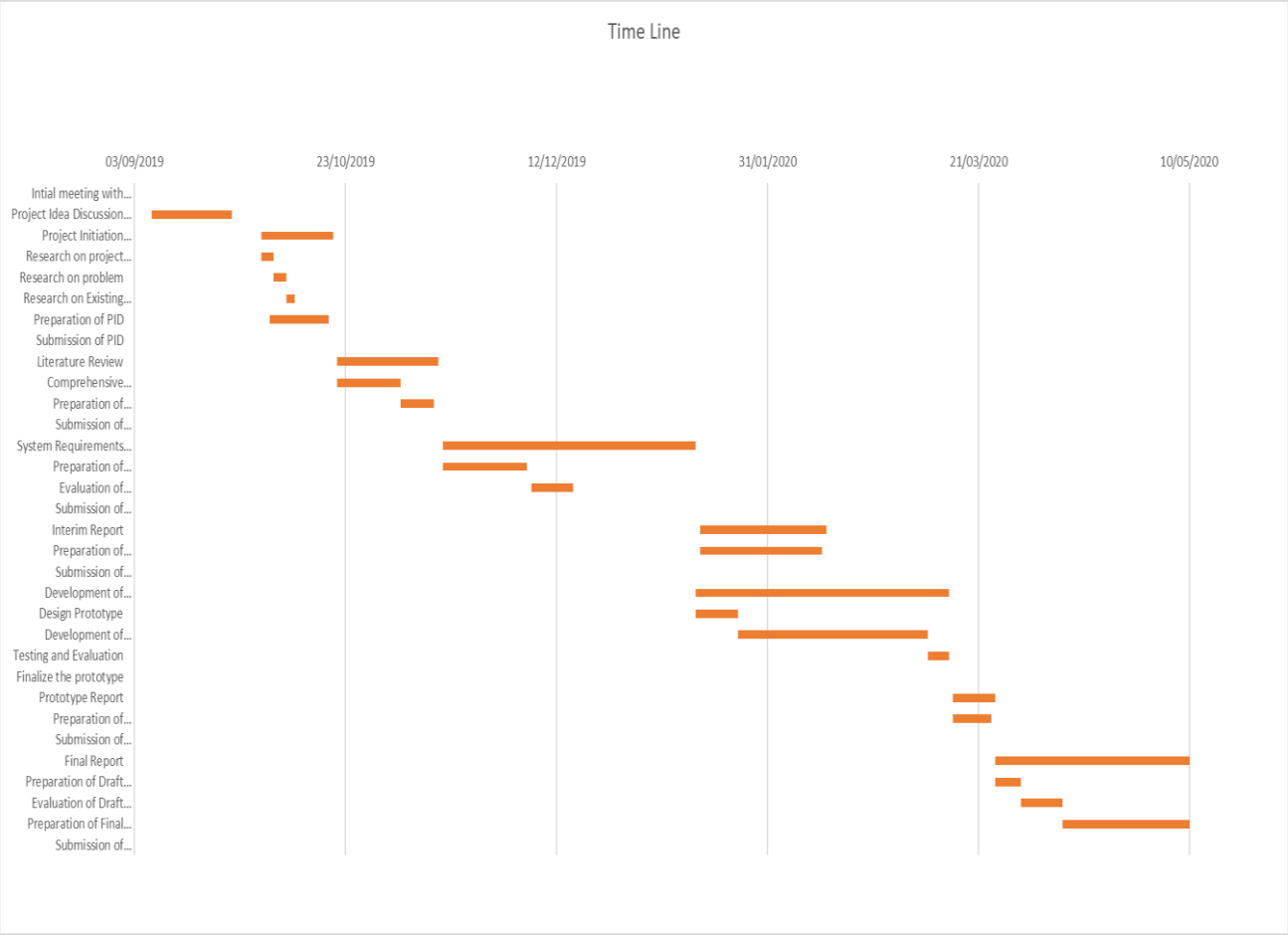

Figure 8: Gantt Chart for the time plan

4.4. Risk Assessment

Risk	Details of Risk	How to avoid risk
Weather conditions	in most of the satellite images weather condition has been problem in processing with satellite images	Using only images cloud coverage lesser than 10%
Social matters	Deforestation had been ethnic problem in Sri Lanka for some time	Not mentioning about ethnic in this project
Political Problem	Deforestation had been very hot topic in Sri Lankan politics	not mentioning any politicians or period of specific government

Table 4: Risk Assessment Table

4.5. BCS code of conduct

1. All the satellites images for this project is taken by USGS earth explorer and Copernicus Open Access, therefore all the images are obtained from official websites and have legitimate rights
2. During this research projects there has no any discrimination upon any person/organization
3. All the details obtained are legitimate, therefore no active or passive bribery occurred while carrying out the research
4. We have done the research in best effort, therefore due care and diligence is maintained.

5. During the research there's no anyone's opinion or interest challenged, therefore no conflict of interest is minimized
6. All the images and details are obtained through websites which are available for public, therefore confidential information is disclosed
7. There's no information withhold about the carried out project

4.6. Chapter Summary

At the beginning of the chapter we have identified the tasks and specified dates of this research project. Then risk assessment for the specific research topic have been discussed and finally to show this research is successfully achieved according to the required protocols, it has shown the following the BCS code of conduct for the specified research project.

5. System Requirement Specification

Content

5.1. Chapter Introduction

During this chapter it will discussed how the necessary information for the research is gathered by which mainly studying related researches and carrying questionnaire. This chapter contains about requirement engineering process that we have undertaken to find the need requirements, analyse requirements for incompleteness and inconsistency and lastly to negotiate final requirements. Hence within this chapter information found from 102 persons asking from the public has shown to you. Thereafter diagrams to depict the research project is shown.

5.2. Requirement Elicitation Process Methods

Method 01	**Literature Review**
Literature review was the main method that we could gather information about. We studied previous researches that had been done and their technologies. It put the base for understanding how to develop a neural network.	
Advantages	1. It has been helpful to identify the technologies straightaway and based on those we could identify the best method for us. 2. We could get a thorough idea about the limitations in implementing the research. 3. Identification of key questions Forest Management 4. The researches were done based on parameters that were unique to their lands. So by reading them it helped us to identify what parameters should be included in our model.
Disadvantages	1. Some of the researches were not done, therefore it was doubtful to follow the particular research. 2. It was very much time consuming process as we have go through whole research to identify the technologies.
Method 02	**Online Survey**
Online survey was conducted mainly to find out the value of the project. Therefore, we have identified how important conducting this research and we have gained information about what Sri Lankan society think of deforestation happening in the country.	
Advantages	1. Through survey we could prove the information that we have discussed in the literature review are true. It clearly shows there's a burning issue of deforestation in Sri Lanka.

	2. Online survey added a value to our project. Because it confirms again how important this project and implies the need of forest management.
Disadvantages	1. Reliability We cannot guarantee the identity of the person who filled the online form. Hence we cannot say some people have given the reliable information. 2. Time Consumption This had been a time consuming process as our target was to collect hundred forms.

Table 5: Table for requirement elicitation process methods

5.2.1. Requirement Elicitation by Questionnaire

From this section it will bring details about a survey that conducted to collect some information about how important doing this research project. 102 persons have participated in the survey and results are as below. Each of the question that were asked in the questionnaire is shown below pie charts and bar diagrams.

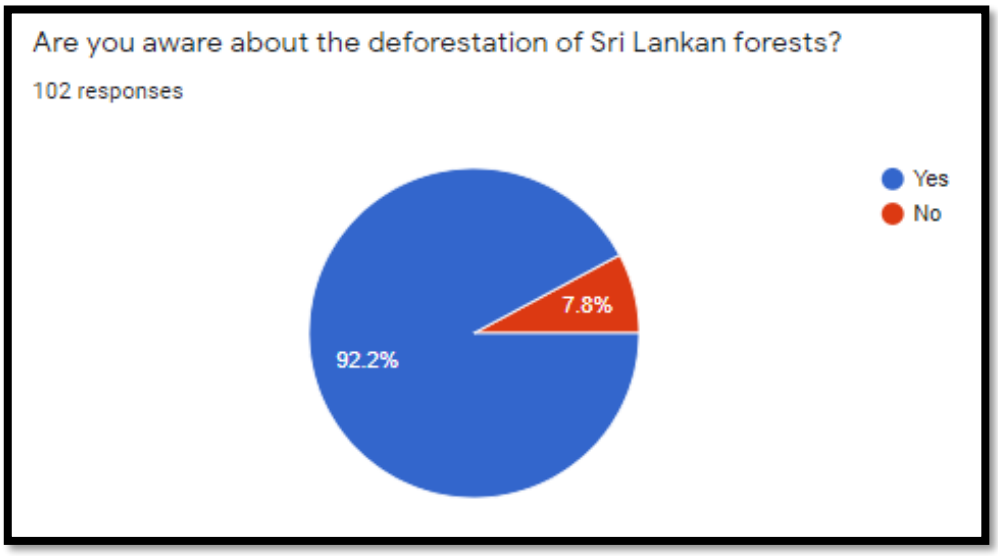

Figure 9: First Question of the questionnaire

This shows that Sri Lankan society is very much aware about the deforestation in Sri Lanka. It has very high parentage and it proves deforestation has been a hot topic in the country recent years. Only a few percentages of 7.8 answered, not aware about deforestation. Therefore, this will prove deforestation is a topic Sri Lankans are very much aware.

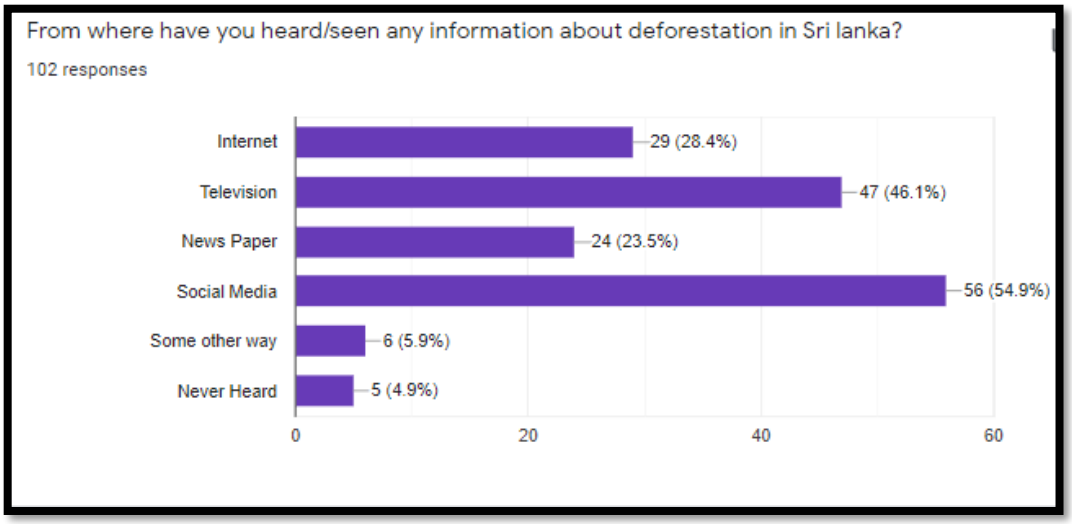

Figure 10: Second Question of the questionnaire

This image shows by which ways and means people have got aware about the deforestation in Sri Lanka. Social media has taken a prominent place in this case and it shows social media can also be used to prevent deforestation. According to the diagram it's a very less percentage people who doesn't get the news from any of the media. Hence this diagram proves again the message of deforestation has conveyed to the society through different platforms like social media, television, newspaper etc.

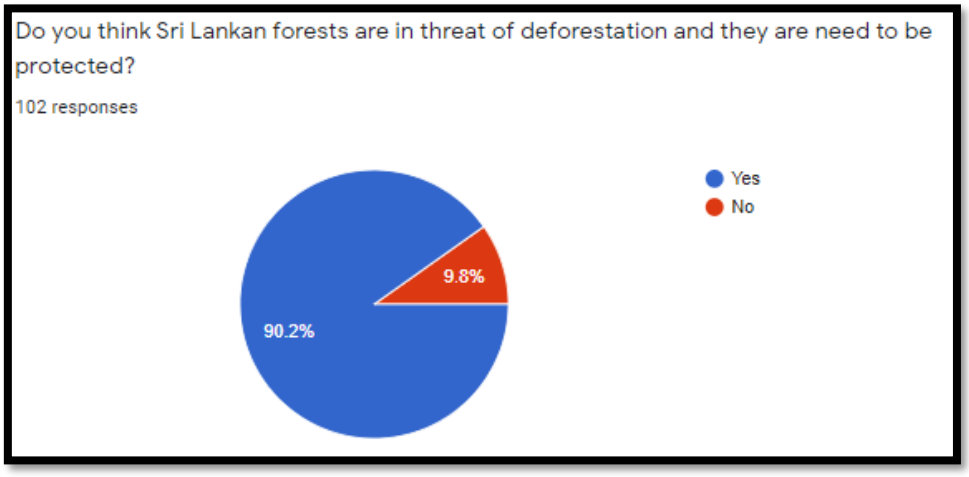

Figure 11: Third question of the questionnaire

For this question everyone has said yes and they believe Sri Lankan forests are in a threat and need to be protected. This is a strong response, because everyone knows there were some matters happening with Sri Lankan forests for some time and they believe there's a threat definitely. So this question emphasizes even though some people do not know there's actually deforestation is happening.

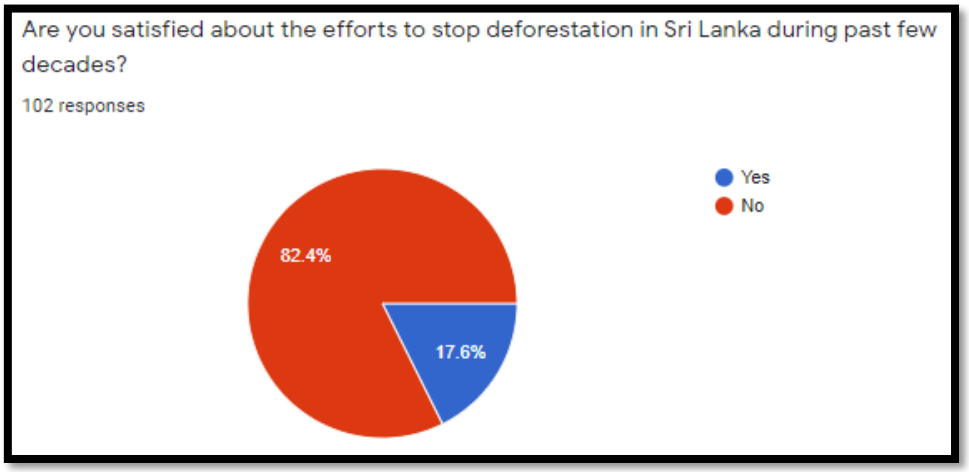

Figure 12: Fourth question of questionnaire

It is mainly the government's responsibility to take necessary legal actions and remedies to fight back deforestation. According to the answers given in this question it is clear that people are not happy with the actions that has been taken so far. Government needs to take strict action to protect the forests. With the results of this research government can make a management plan to protect these lands.

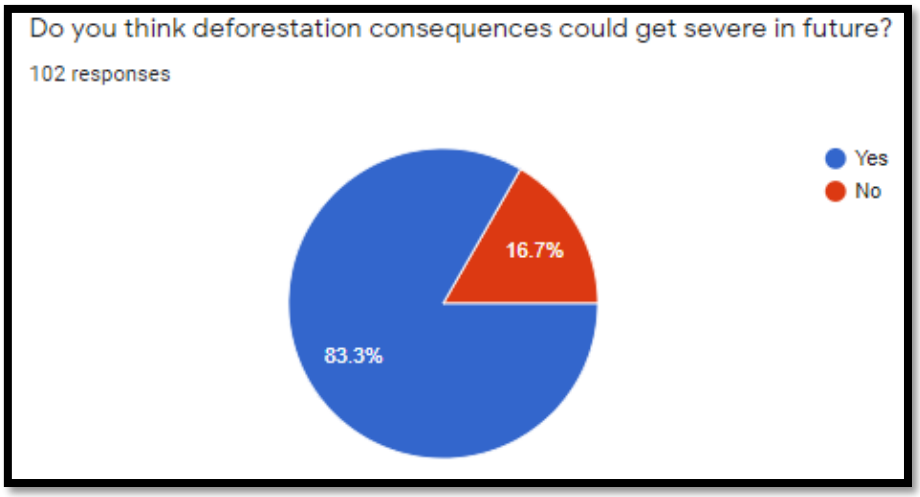

Figure 13: Fifth question of questionnaire

Majority of the people who answered the survey believe if we didn't address the problem now, it could get severe in the future. Approximately 16.7% do not believe it could get severe in future. Nevertheless, it is very important that we do a research about the problem now and investigate what can this problem cause.

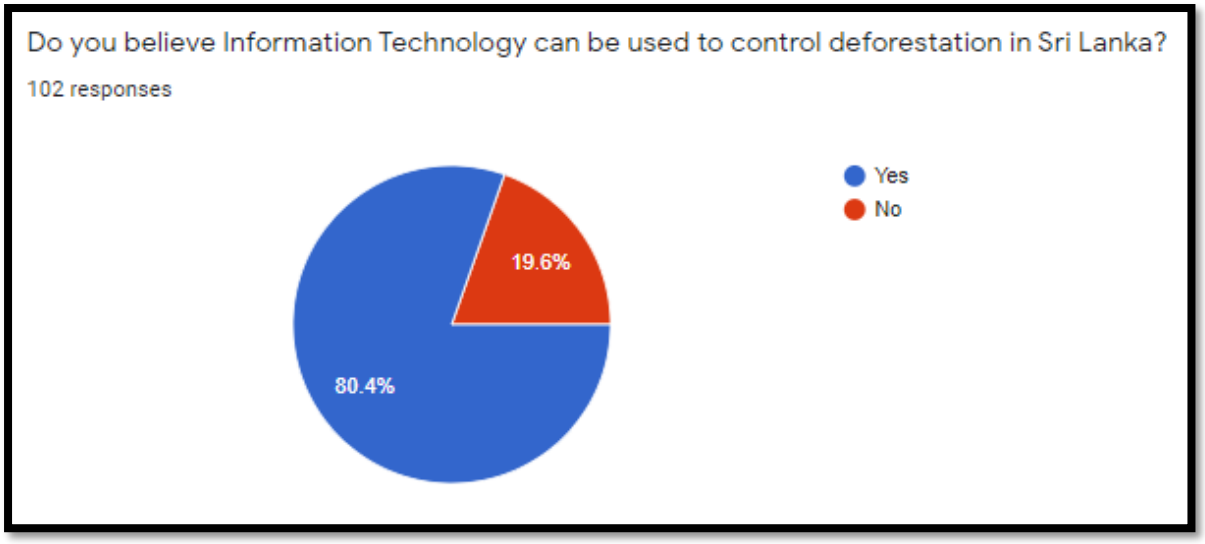

Figure 14: Sixth question of the questionnaire

This is the most important part of the questionnaire. Because it is needed to confirm that our project will be useful for this problem or not. Many people believe IT solution would be helpful to solve this matter. It may be because sometimes people do not exactly get the extent of deforestation in Sri Lanka. So by using technology we can show with images and graphs to which extent it has been damaged.

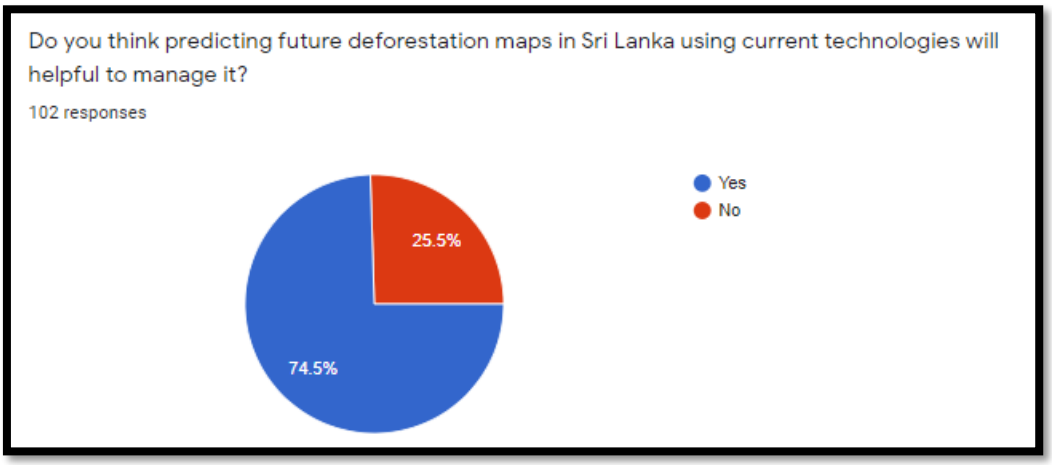

Figure 15: Eighth question of the questionnaire

This has confirmed the project again. Majority of the people who took the online questionnaire believe an IT solution that gives a predicting model for deforestation in Sri Lanka would help to overcome the current issues. Only minor percentage of 25.5% do not believe prediction model won't be helpful. Since the majority are saying 'yes' to the project it will be great fact this project going to achieve the targeted objectives towards to the society.

5.3. Quantitative and Qualitative Analysis of Data

For the analysis of the satellite images, it was taken Sentinel 2, Landsat 7 and Landsat 8 images. As we have discussed in literature review these are satellites which were launched by Earth Observing organizations with special tasks. Therefore, all these images which were taken from those websites are for professional use. In this remote sensing plays an important role in finding accurate details about deforestation. The different band lengths in satellite images help us to identify exact details about existence of physical features on earth. The satellite images taken for this project are with very low cloud cover percentage (less than 10 percent) and that been a clear factor which helped in identifying the physical features very clearly.

5.4. Context Diagram

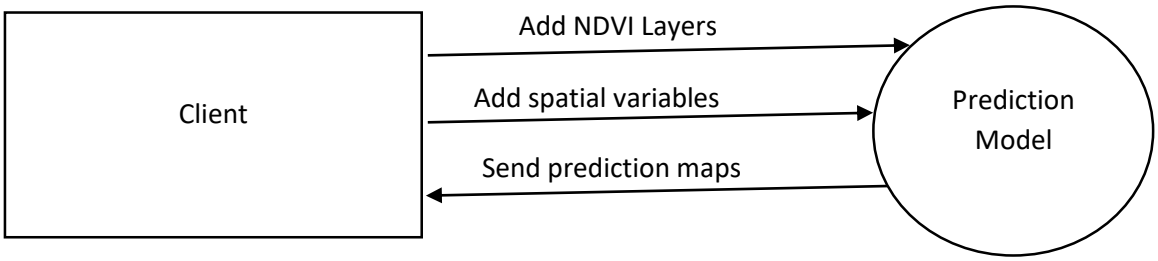

Figure 16: Context Diagram for the proposed model

Given above is the context diagram for the proposed system. It shows how the information flows in the system. The client will add NDVI maps and spatial variables which needed by prediction model to give out prediction maps. Once the findings (NDVI maps) have went through process by means of the algorithms the prediction maps will be given out.

5.5. Use case Diagram

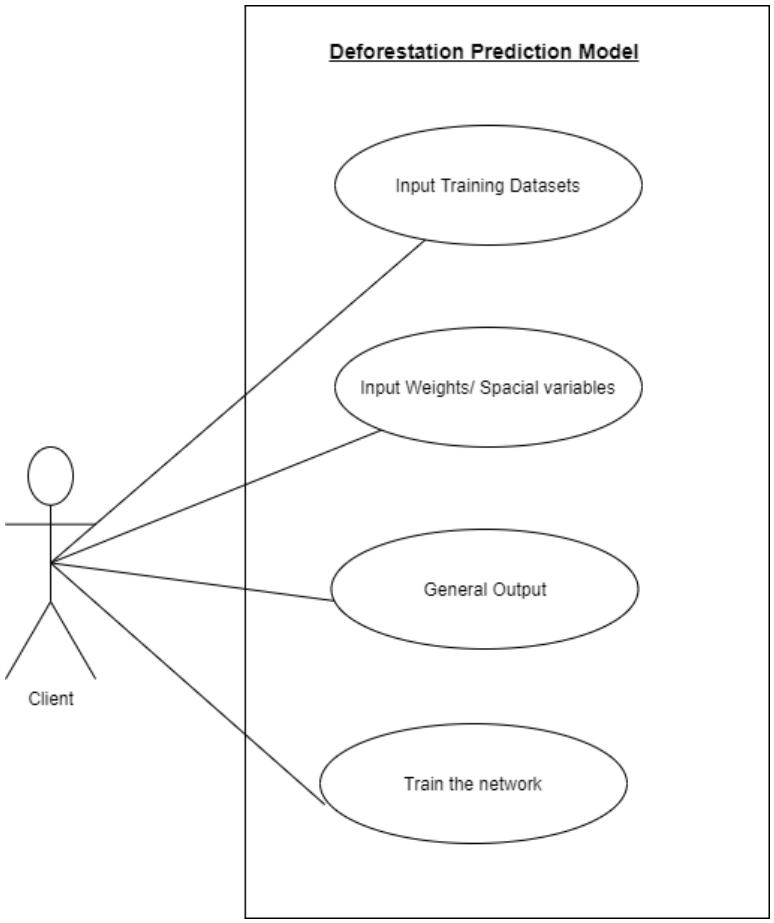

Figure 17: Use case Diagram for prediction model

The use case diagram for the proposed system is given above. According this diagram, it allows user to input training data and input weights. As input variables we will assign spatial variables that have been found. Thereby the training neural network needs initialization of input weights and it will then train the network to give out predictions which consists predicted NDVI images.

5.5.1. Use Case Specification

Use case name	Generate Output
Description	This use case will used to represent Generate the predictions Maps
Participating actors	Client
Pre – conditions	Having processed NDVI maps
Included use cases	None
Main flow	Client will add NDVI maps Client will add spatial variables. Prediction model accepts client data if valid. Prediction model generates prediction maps accordingly
Alternative flow	None
Exceptional flow	If the NDVI maps and special variables are not matching the algorithm will not proceed. .
Post conditions	None

Table 6: Use Case specification table

5.5.2. Functional requirement

No	Description	Priority level	Use case mapping
1	Prediction Model should produce prediction maps	Core	Generate Output
2	NDVI maps should be inserted to prediction model	Core	Input training data set
3	Relevant spatial variables should be added	Core	Input weights

Table 7:Functional Requirements Table

5.5.3. Non-Functional Requirements

NO	User's concern	Non-functional requirement
1	Ease of use	Usability
2	Likelihood if Failure	reliability
3	Resource utilization	Efficiency

Table 8:Non-Functional requirements Tables

5.6. Chapter Summary

In this chapter main objective is to clarify the necessary requirements to carry out the project. In the first stage of this chapter steps that were taken to collect data were discussed. It was done mainly through online questionnaire and raw data collected has shown to you by using pie charts. By bringing out the data collection methods we have justified the fact collection process that we have taken in this research.

6. System Architecture & Design

Content

6.1. Chapter Introduction

In this design chapter we are focussing on building the architecture of the systems and its design. As we have discussed about requirement engineering in the previous chapter, now we can build the design using the details of requirements engineering. Hence the main purpose of this chapter is to have clear idea about the structural components in the system.

6.2. High Level Design

6.2.1. Layered Structure

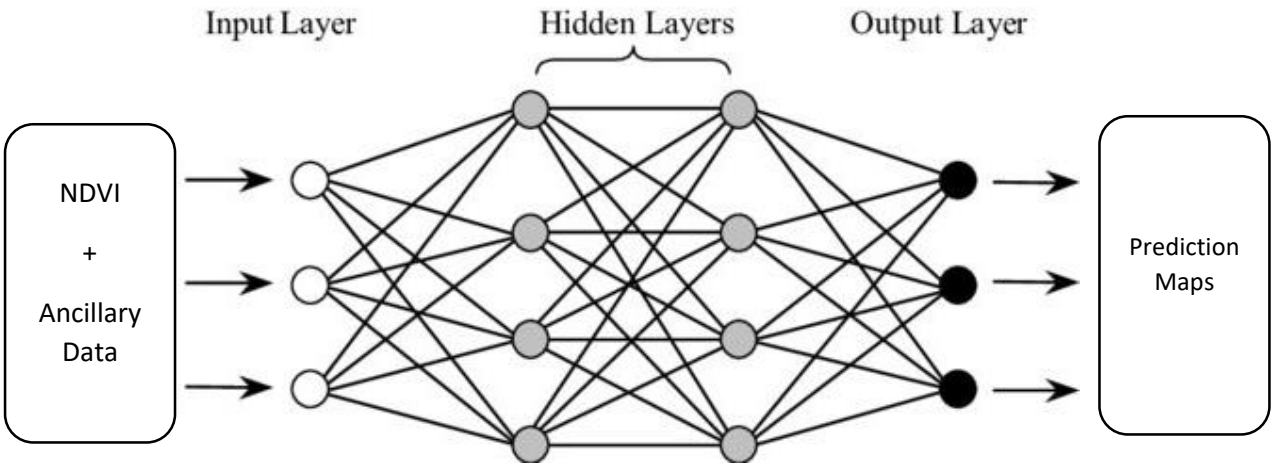

Figure 18: Layered Structure of proposed prediction model

The above picture depicts a layer diagram of the research project. As shown in the diagram first we need to collect the data that we need to introduce to the artificial neural network. That will the NDVI images that found by processing the satellite raster images and then ancillary data like dates of images and prediction date and spatial variables. That will be the input layer for the system. Then during the hidden layer images will be processed by artificial neural network. Once the process is done the it will show the areas that could get deforested that particular period we have given. That will the output layer of the project which is with produced prediction maps.

6.3. Low Level Design

6.3.1. Class Diagram

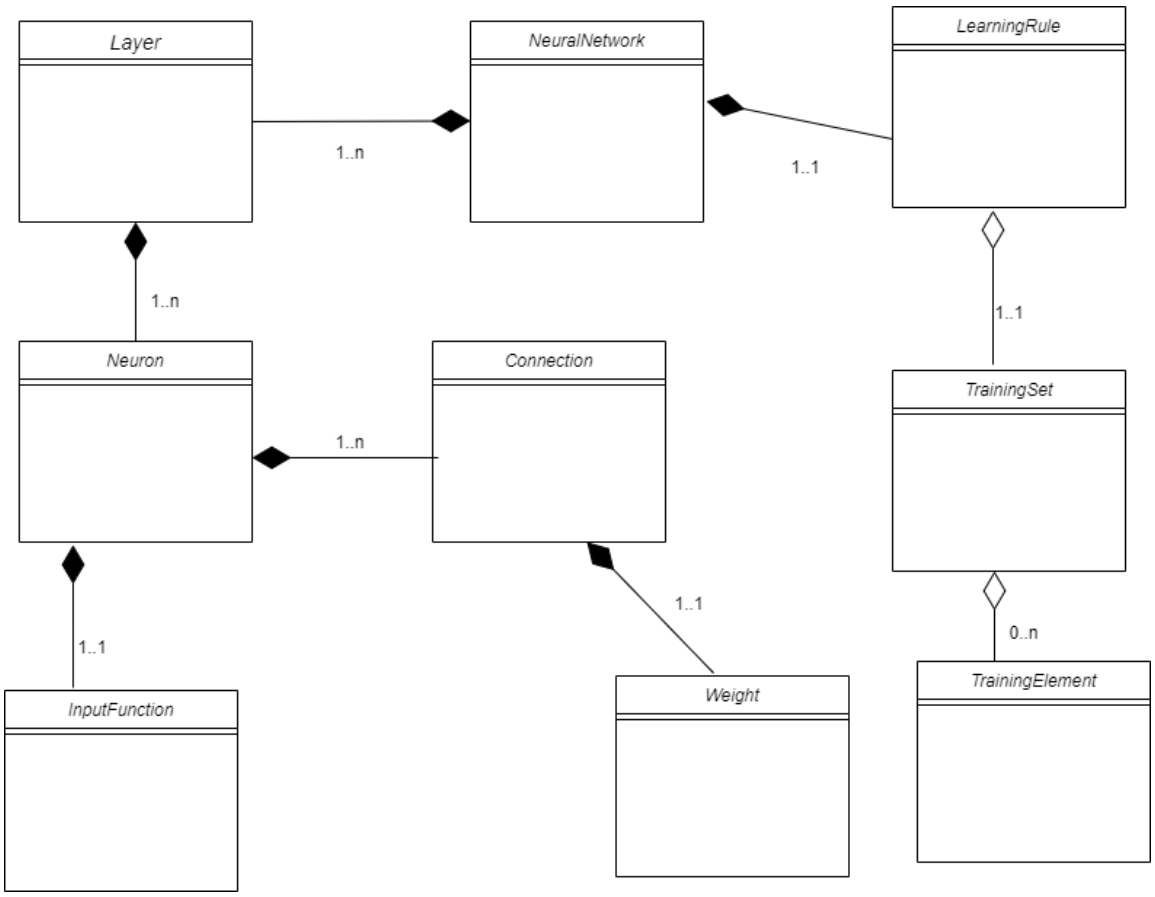

Figure 19:Architecture of Class Diagram for the proposed System

The above diagram shows you the architecture for the proposed system. This is developed based on the previous researches that has been done using artificial neural networks. It has shown you only the classes and their relationships among them. Detailed classless can be found in the next page with a clear view.

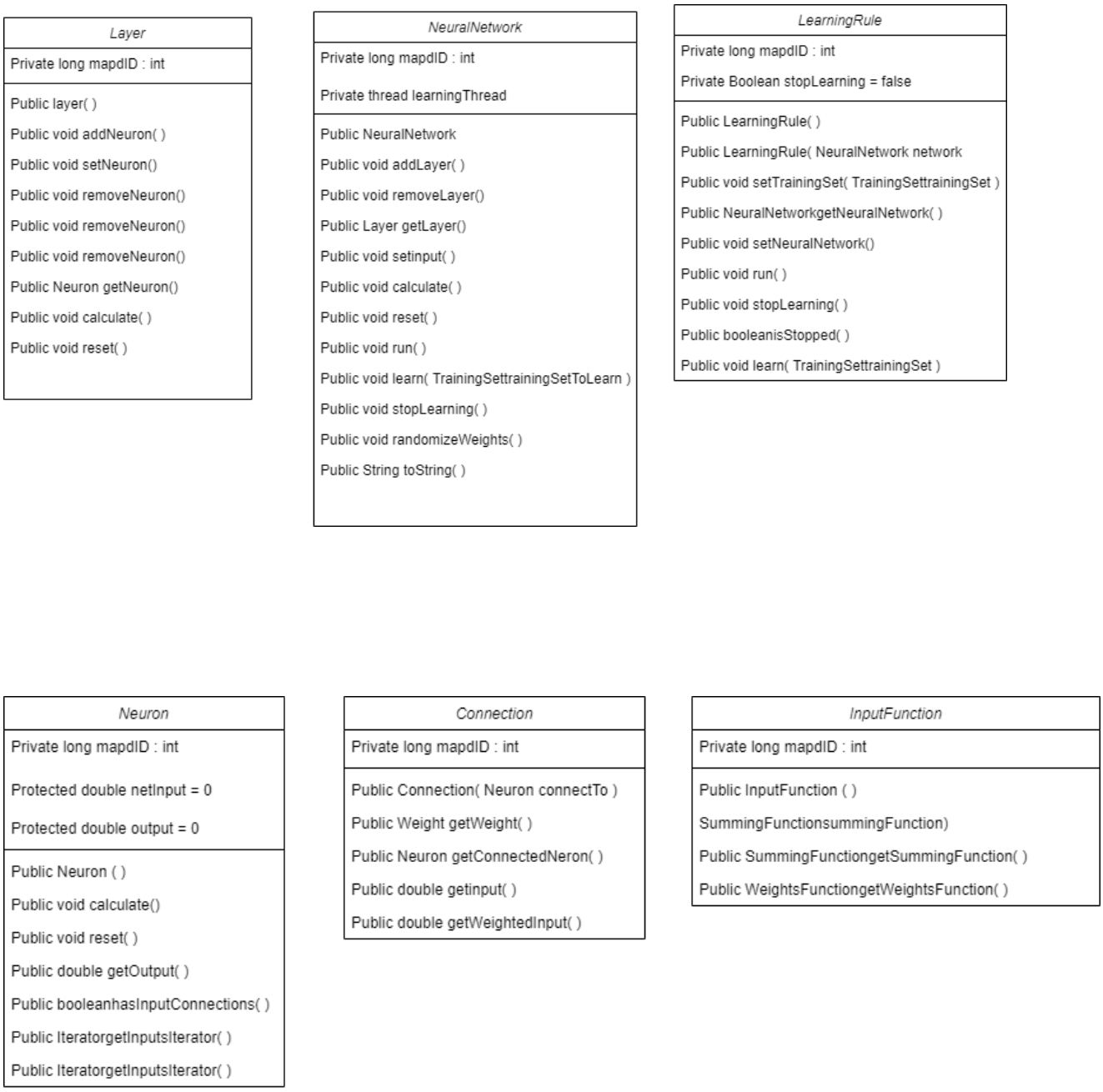

Figure 20: Classes for the Class Diagram

6.3.2. Activity Diagram

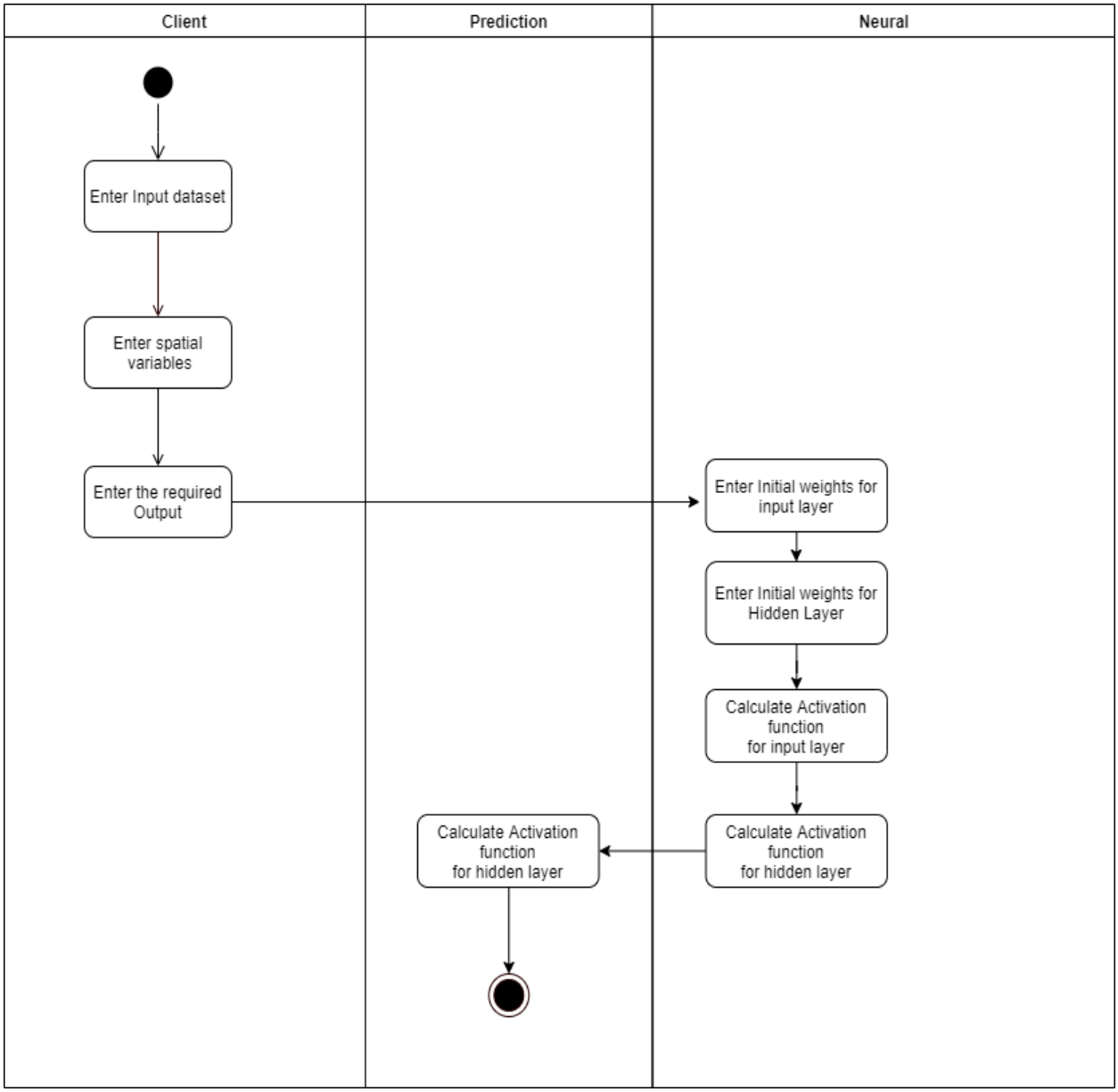

Figure 21: Activity Diagram for the proposed model

According to the activity diagram above the client is expected to enter the input dataset, spatial variables and details of required output details like dates. So client will select the input variables the neural network will initialize the weight for input layer and hidden layer. The prediction will be based on these variables the client will input. As we have planned it will be spatial variables that we will be given to the neural network to process.

6.3.3. Sequence Diagram

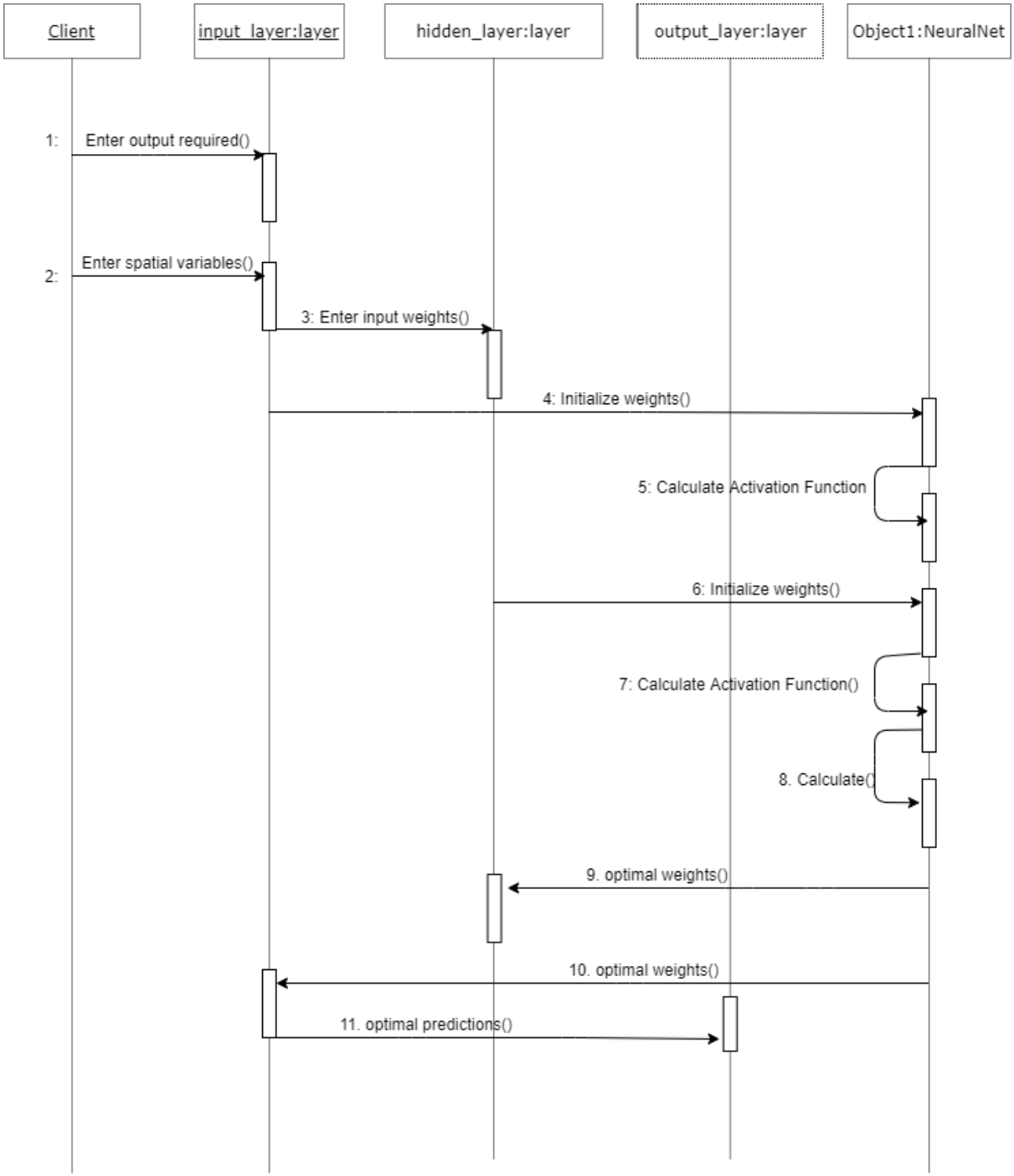

Figure 22: Sequence diagram for proposed model

The above diagram is the sequence diagram for the proposed system. All the activities have been numbered here from the input of data to the prediction state. This will give you a clear image of how the prediction model will work.

6.4. Chapter Summary

During this chapter we have shown the system architecture and designs of the research project. For the purpose we have used high level diagrams and low level diagrams with proper explanation which makes it easy for understanding. In this high level designs which comprised of layered diagrams which show you system functionalities while low level diagrams such as class diagrams and activity diagrams will show you the functionalities step wise. Hence by studying this chapter it will help you to understand design goals of this research project.

7. Implementation

Content

7.1. Chapter Introduction

This chapter includes all the findings of the research. The results obtained by conducting the research has been shown out in this section and it has graphically demonstrated and described. Results have categorized according to the forest areas.

7.2. Operational Steps

1) Downloading images from the sources
   As we have discussed in previous chapters it is possible to downloaded Landsat images from USGS earth explorer

2) Atmospheric Correcting

   By doing atmospheric correct it will clear the surface area of a satellite image. Because it will remove the scattering and absorption effects from the atmosphere.

   For the research it was taken four types of Landsat images

   - Landsat 7 ETM+C1 C1 Level 1
   - Landsat 7 ETM+C1 Level 2
   - Landsat 8 OLI/TIRS C1 Level 1
   - Landsat 8 OLI/TIRS C1 Level 2

   This is need to be done only to level 1 images. Level 2 images are already atmospheric corrected and you have to request them from USGS earth explorer to obtain them. You can do it by using the Semi-Automatic classification plugin.

3) Correcting scan line error

In the Landsat 7 images which was collected after May 31st 2003 Scan Line Collector has failed which mean they have data gaps. Therefore, it was needed to fill the gaps in the images first.

```
Input parameters:
{ 'BAND' : 1, 'DISTANCE' : 50, 'EXTRA' : '', 'INPUT' : 'D:/FYP/Project Implementation/
Sinharaja/2000-2005/2004.02.19 - No atm/LE07_L1TP_141056_20040219_20170122_01_T1_B3.TIF',
'ITERATIONS' : 0, 'MASK_LAYER' : '/vsigzip/D:/FYP/Project Implementation/Sinharaja/
2000-2005/2004.02.19 - No atm/gap_mask/
LE07_L1TP_141056_20040219_20170122_01_T1_GM_B3.TIF.gz', 'NO_MASK' : True, 'OPTIONS' : '',
'OUTPUT' : 'TEMPORARY_OUTPUT' }

GDAL command:
python3 -m gdal_fillnodata -md 50 -b 1 -nomask -mask "/vsigzip/D:\\FYP\\Project
Implementation\\Sinharaja\\2000-2005\\2004.02.19 - No atm\\gap_mask\
\LE07_L1TP_141056_20040219_20170122_01_T1_GM_B3.TIF.gz" -of GTiff "D:\\FYP\\Project
Implementation\\Sinharaja\\2000-2005\\2004.02.19 - No atm\
\LE07_L1TP_141056_20040219_20170122_01_T1_B3.TIF" C:/Users/banuk/AppData/Local/Temp/
processing_924fff3766f14a2a8e4f6447b3086145/0fc36054e08b4e27a0b227dc6f6e03bc/OUTPUT.tif
GDAL command output:
0...10...20...30...40...50...60...70...80...90...100 - done.

Execution completed in 26.03 seconds
Results:
{'OUTPUT': 'C:/Users/banuk/AppData/Local/Temp/
processing_924fff3766f14a2a8e4f6447b3086145/0fc36054e08b4e27a0b227dc6f6e03bc/OUTPUT.tif'}
```

Figure 23: Python Script for filling gaps

4) Merge bands using for the NDVI classification

For the NDVI classification, when it was Landsat 7 images were chosen bands 2,3,4 and 5 are considered. When it was using Landsat 8 images bands 3,4,5 and 6 are considered which are needed to obtain data about deforestation.

```
Input parameters:
{ 'DATA_TYPE' : 8, 'EXTRA' : '', 'INPUT' : ['D:/FYP/Project Implementation/Wilpattu/
2000-2005/Filled/Band_2.tif','D:/FYP/Project Implementation/Wilpattu/2000-2005/Filled/
Band_3.tif','D:/FYP/Project Implementation/Wilpattu/2000-2005/Filled/Band_4.tif','D:/FYP/
Project Implementation/Wilpattu/2000-2005/Filled/Band_5.tif'], 'NODATA_INPUT' : None,
'NODATA_OUTPUT' : None, 'OPTIONS' : '', 'OUTPUT' : 'TEMPORARY_OUTPUT', 'PCT' : False,
'SEPARATE' : True }

GDAL command:
python3 -m gdal_merge -separate -ot CInt32 -of GTiff -o C:/Users/banuk/AppData/Local/Temp/
processing_924fff3766f14a2a8e4f6447b3086145/f4d54637fa4e44bdb5d060dfb9568189/OUTPUT.tif --
optfile C:/Users/banuk/AppData/Local/Temp/processing_924fff3766f14a2a8e4f6447b3086145/
eadc7c6be1314305bd68ed5dbfc2c1b1/mergeInputFiles.txt
GDAL command output:
0...10...20...30...40...50...60...70...80...90...100 - done.

Execution completed in 190.21 seconds
Results:
{'OUTPUT': 'C:/Users/banuk/AppData/Local/Temp/processing_924fff3766f14a2a8e4f6447b3086145/
f4d54637fa4e44bdb5d060dfb9568189/OUTPUT.tif'}
```

Figure 24: Python script for Merging bands

5) Create a mask

Now we have to create a mask to cut the forest area with the boundary, so that we can use that mask to cut all the Landsat images which we are going to analyze for separate periods. It can be done by using the 'New Shapefile' option.

6) Cut the Forest Area from Landsat images

We have the mask file and relevant merged Landsat images for NDVI. By using the mask file, we can cut the forest area that we need from the merged Landsat images.

```
Input parameters:
{ 'ALPHA_BAND' : False, 'CROP_TO_CUTLINE' : True, 'DATA_TYPE' : 0, 'EXTRA' : '', 'INPUT' :
'D:/FYP/Project Implementation/Wilpattu/2000-2005/Filled/NDVI_2004.tif', 'KEEP_RESOLUTION'
: True, 'MASK' : 'D:/FYP/Project Implementation/Wilpattu/2000-2005/Willpattu_shapecut.dbf',
'MULTITHREADING' : False, 'NODATA' : None, 'OPTIONS' : '', 'OUTPUT' : 'TEMPORARY_OUTPUT',
'SET_RESOLUTION' : True, 'SOURCE_CRS' : None, 'TARGET_CRS' : None, 'X_RESOLUTION' : None,
'Y_RESOLUTION' : None }

GDAL command:
gdalwarp -of GTiff -tr 30.0 -30.0 -tap -tr 30.0 -30.0 -tap -cutline "D:/FYP/Project
Implementation/Wilpattu/2000-2005/Willpattu_shapecut.dbf" -cl Willpattu_shapecut -
crop_to_cutline "D:/FYP/Project Implementation/Wilpattu/2000-2005/Filled/NDVI_2004.tif" C:/
Users/banuk/AppData/Local/Temp/
processing_924fff3766f14a2a8e4f6447b3086145/37a68c0f18074ea39aeb47310cb400e2/OUTPUT.tif
```

Figure 25 :Python Script for extracting raster map by mask layer

7) Calculation NDVI

Now we have to find the Normalized Difference Vegetation Index using the equation we have discussed in previous chapters. You can use raster calculator for this purpose.

8) Raster Image Analysis

We have the NDVI maps with us now. We can analyze it by different options given by QGIS software. We've considered giving raster images green colour, so the vegetation can be identified easily with the changing frequency of green colour. Also it has analyzed with two interpolations linear and discrete. In linear, the resulting colour is linearly interpolated from the colour map entries and in discrete the colour is used from the colour maps entry and it is with equal or higher value.

7.3. Explanation of the test results

In this section it will explained the founding of the research by taking out results of each forest separately.

7.3.1. NDVI Results of Wilpattu Forest and Surrounding Forest Areas

Surrounding Areas Include;

1. Tabbowa Sanctuary
2. Veppal Reserve Forest
3. Periyamurippu Reserved project
4. Mavillu Reserved project
5. Vilanthimulam Reserved Forest
6. Weerakulicholai-Elavankulam Forest Reserve
7. Maraichukaddi/Karadikkuli reserved Forest

7.3.1.1. Period: 2000-2005

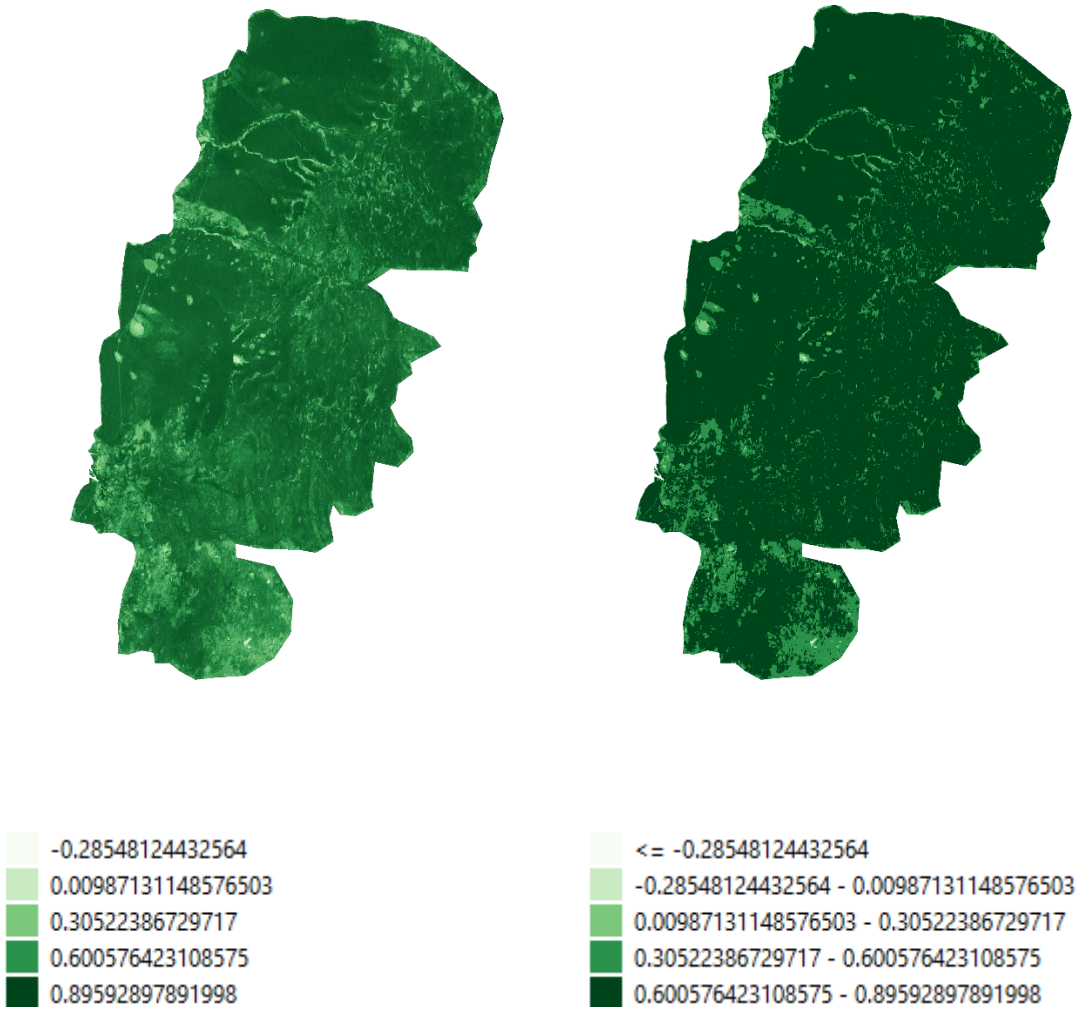

Figure 26:Linear Image of NDVI and colour ramp 2004

Figure 27-Discrete Image of NDVI and colour ramp 2004

The above results are obtained by analysing the Landsat images captured on 26/02/2004. By analysing greenness in linear and discrete colour images we can see there is thick dense greenness during the period 2000-2005.

7.3.1.2. Period: 2005-2010

Figure 28: Linear Image of NDVI and colour ramp 2008

Figure 29: Discrete Image of NDVI and colour ramp 2008

The above images are obtained by analysing the Landsat images which were captured on 03/11//2008. These images were used to find the vegetation during the period of 2005-2010. Considering with the previous images it doesn't show much drastic change.

7.3.1.3. Period: 2010-2015

Figure 30: Linear Image of NDVI and colour ramp 2015

Figure 31:Discrete Image of NDVI and colour ramp 2015

The above pictures are obtained by Landsat images that had been captured on 15/01/2015. When you compared with the images of previous period there are visible places that deforestation has taken places. The extent of deforestation has analysed with histograms in the next chapter.

7.3.1.4. Period: 2015-2020

Figure 32: Linear Image of NDVI and colour ramp 2018

Figure 33: Discrete Image of NDVI and colour ramp 2018

The above images were obtained by analysing the satellite images which was taken on 07.01.2018. when you compare it with the vegetation which was during the period of 2000-2005 there's a big change. When we observer the top part of the images we can see there has been deforestation happened in those areas.

7.3.2. NDVI's of Yala Forest

This section will show you the results of Yala forest. Analyzed NDVI images will take according to the period. Anyhow it had been tough task in finding images for this area. So following the five year periods were not possible in this case.

7.3.2.1. Period: 2000-2005

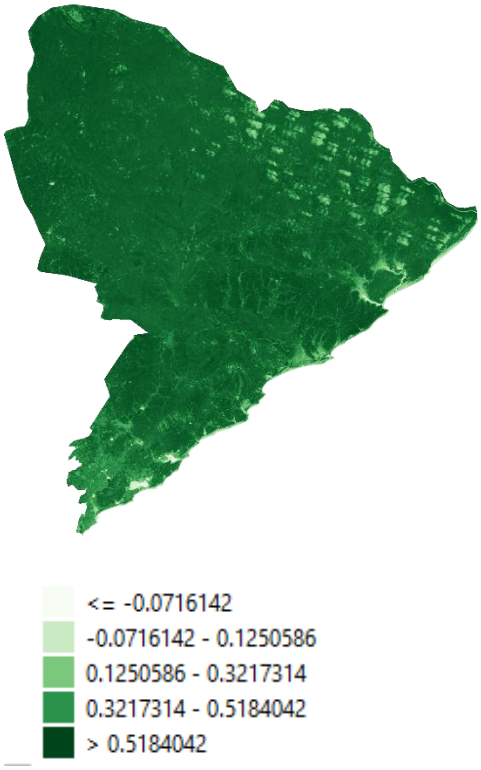

Figure 34: Yala Forest Linear Image of NDVI and colour ramp 2001

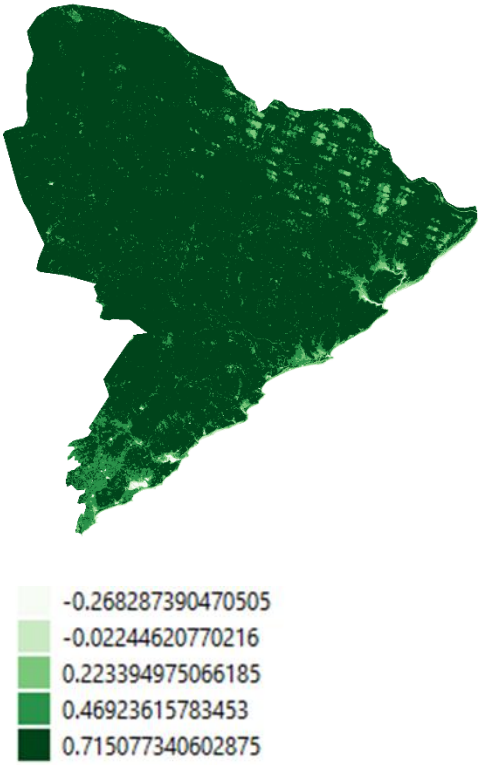

Figure 35: Yala Forest Discrete Image of NDVI and colour ramp 2001

The above images of Yala forest are obtained from analysing the Landsat images that had been captured on 07/03/2001. In this image we can observe there had been an obstruct from the cloud coverage even though the images had cloud coverage less than 10 percent. Also it had been unable to find an image during this period for the Yala forest.

7.3.2.2. Period: 2010-2015

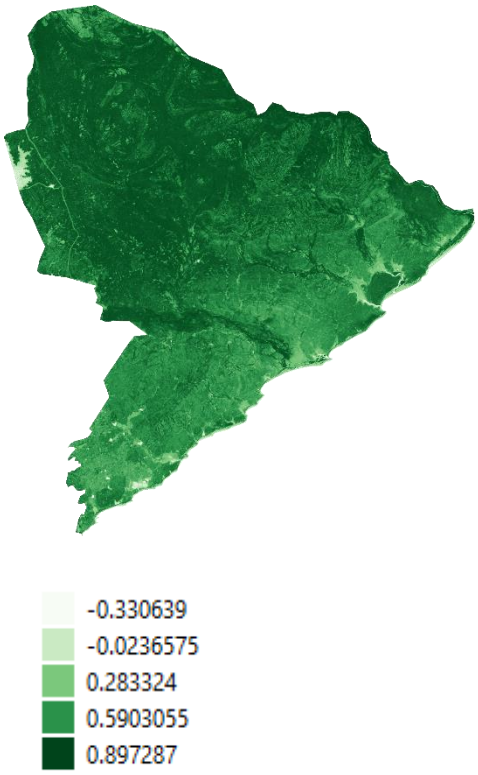

Figure 36:Yala Forest Linear Image of NDVI and color ramp 2013

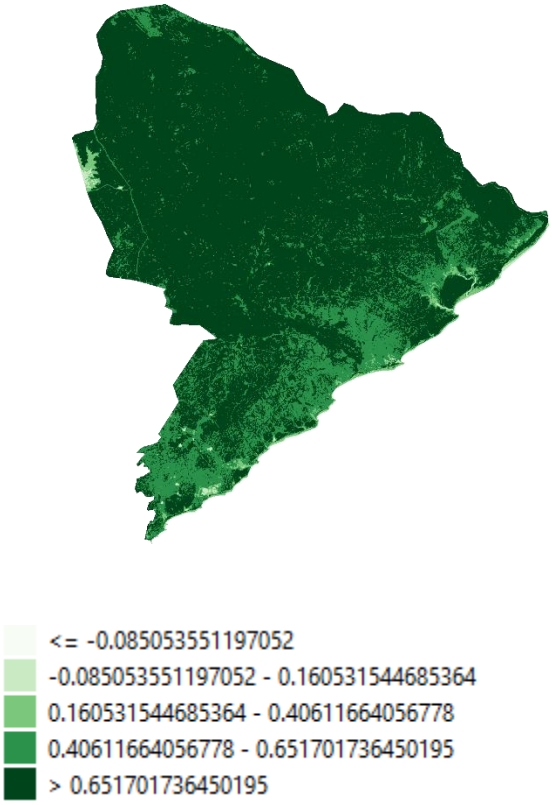

Figure 37: Yala Forest Linear Image of NDVI and colour ramp 2013

These images are obtained by analysing the Landsat images which have been captured on 07/08.2013. when we compare it with the previous images of the time period, you can see how the vegetation has lost its greenness. The graphical analysis of the greenness can be found in next section.

7.3.2.3. Period: 2015-2020

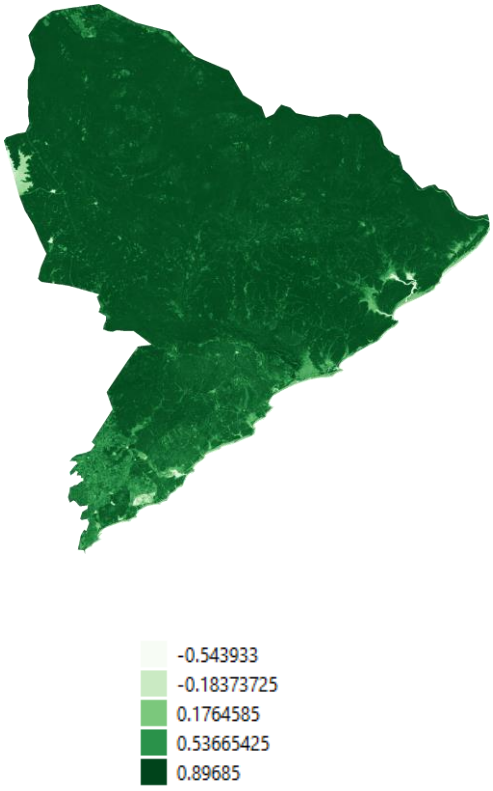

Figure 38:Yala Forest Linear Image of NDVI and colour ramp 2019

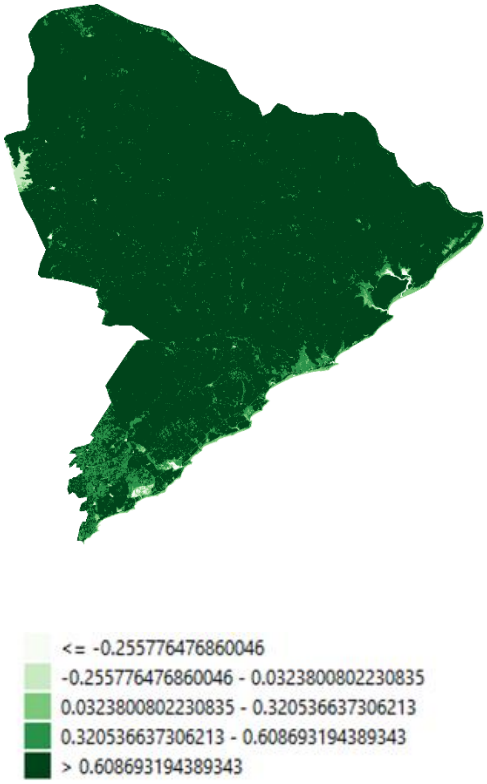

Figure 39: Yala Forest Linear Image of NDVI and colour ramp 2019

The above images are obtained by analysing the Landsat images of Yala forest which has captired on 28/01/2019. With observations you can identify there has some vegetation growth during that period. For further analysis, we can see NDVI histograms which can be found in next section

7.4. Chapter Summary

During this chapter we have discussed about the findings of our research through the implementation. In next chapter all these results are used for evaluations to make using raster histograms.

8. Testing

Content

8.1. Chapter Introduction

During this chapter graphical analysis of the findings will be done. It will test how accurate the findings of this research are. Also this chapter includes details about white box and black box test. Therefore, it is expected to prove the forest change during this chapter.

8.2. Applicability of Testing Types

Black Box Testing	White Box Testing
This carried out without any idea of the flows in the system	This carried after knowing the flows
Errors can be found, in data structures, algorithms and in functionalities	Test cases exercises a set of conditions
Advantages	
Errors in classes can be identified	It is checked all the logical decision in the model
Disadvantages	
It's no easy to create test cases for this	It is possible to look in to every minute details

Table 9: Compare the applicability of test types

8.3. Test Cases

Test Case	Status
Found the Wilpattu Forest Area NDVI for 2000-2005	Yes
Found the Wilpattu Forest Area NDVI for 2005-2010	Yes
Found the Wilpattu Forest Area NDVI for 2010-2015	Yes
Found the Wilpattu Forest Area NDVI for 2015-2020	Yes
Found the Yala Forest NDVI for 2000-2005	Yes
Found the Yala Forest NDVI for 2005-2010	Not Available Details
Found the Yala Forest NDVI for 2010-2015	Yes
Found the Yala Forest NDVI for 2015-2020	Yes
Prediction Maps for Wilpattu Forest Area	No
Prediction Maps for Yala Forest	No

Table 10: Table for Test Cases

8.4. Graphical Representation of Test Results

In this section we will discuss about NDVI changes graphically to show you the test results clearly and it is done methodically according to the specific periods.

8.4.1. Histogram Analysis of NDVI for Wilpattu & Surrounding forests

In this section it will explain the raster histograms to accomplish the research objectives.

8.4.1.1. Period: 2000 - 2005

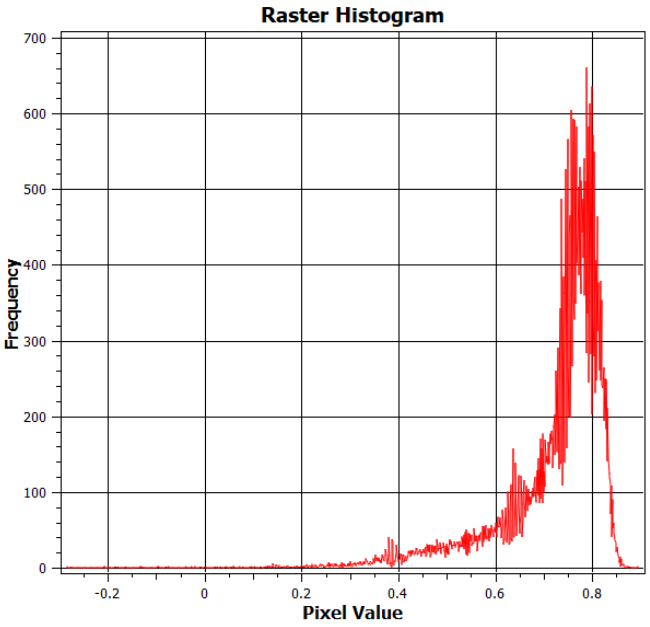

Figure 40:Histogram for the NDVI calculate for the period of 2000-2005

The above diagram shows the histogram of NDVI colour ramp for the period of 2000-2005. Pixel values represent the NDVI values and when it goes toward positive direction forest density increases and when it goes towards negative direction forest density decreases. In this you can observe there's huge variation between 0.6 to 0.85 which we can obtain from horizontal axis. If we look at the frequency (vertical axis) we can see frequency ranges from 0 to 700. Considering this is as the beginning, we can now compare rest of the results.

8.4.1.2. Period: 2005-2010

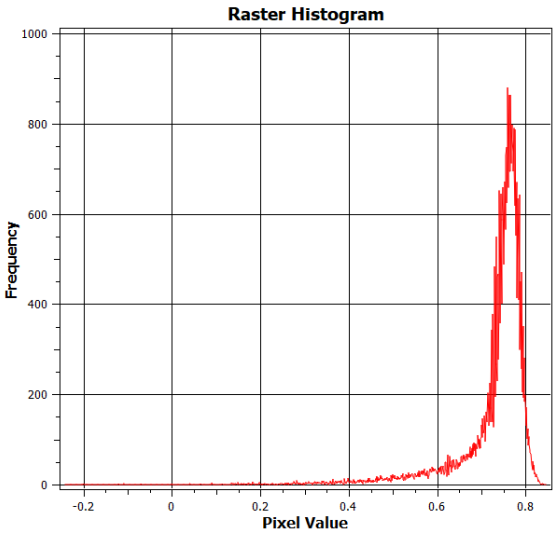

Figure 41: Histogram for the NDVI calculate for the period of 2005-2010

The above diagram depicts the NDVI variation for the time period 2005-2010 in Wilpattu forest area according to the analysed images. From the vertical axis we can see frequency ranges from 0 to 1000 which is higher than previous period. Also again the frequency ranges between 0.4 and 0.85 with hype from 0.6 to 0.8. So we can conclude from the above diagram forest density has been increased from 2005 – 2010.

8.4.1.3. Period: 2010-2015

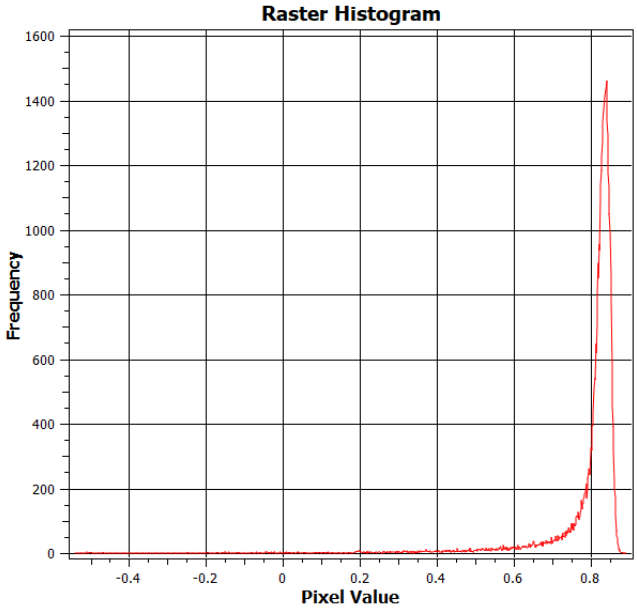

Figure 42: Histogram for the NDVI calculate for the period of 2010-2015

The above diagram of histogram shows frequency variation for the period of 2010-2015. When you compare it with the previous period, you can observe frequency has increased further. It ranges from 0 to 1600. Also the pixel value stays between 0.6 and 0.8 mainly. Therefore, we can conclude during the period of 2010 to 2015 forest density has increased further.

8.4.1.4. Period: 2015-2020

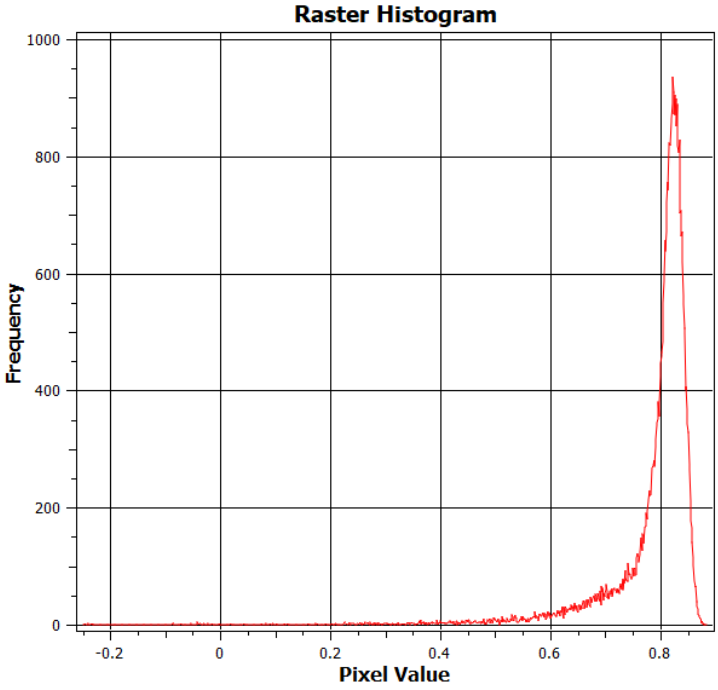

Figure 43:Histogram for the NDVI calculate for the period of 2015-2020

The above histogram diagram shows the frequency variation for the period of 2015-2020. When you compared it with the previous time period, you observe the frequency has decreased which mean during that period forest density could have decreased comparatively. But still the pixel value takes a higher place.

8.4.2. Histogram Analysis of NDVI for Yala Forest

In this section a clear description about test results for Yala forest will be given. Raster histograms obtained will show the variation of NDVI in Yala forest.

8.4.2.1. Period: 2000-2005

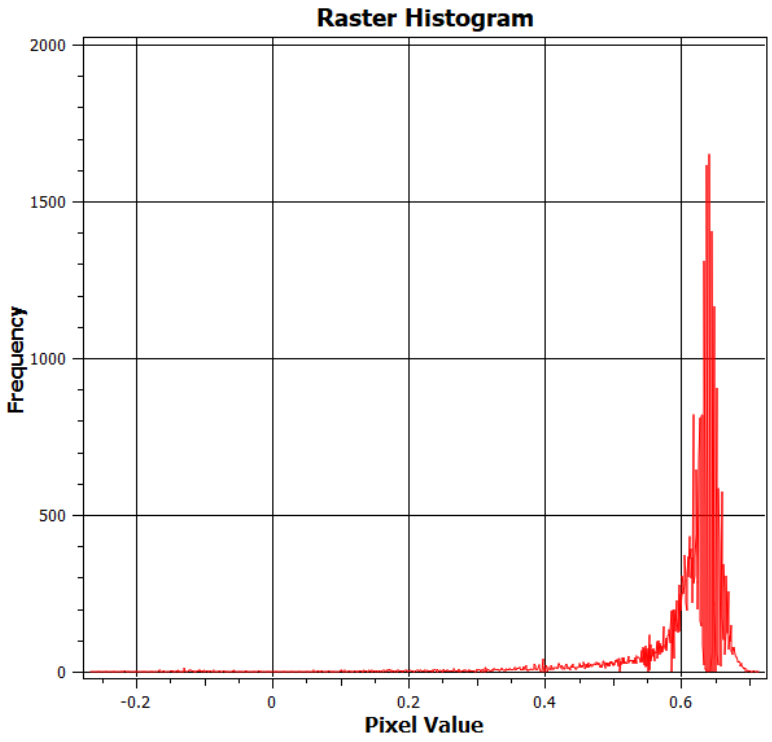

Figure 44:Histogram for the NDVI calculate for the period of 2000-2005

The above diagram explains the frequency variation of Yala forests during the period of 2000-2005 which was obtained by analysing the images. You can observe during 2000-2005 period the frequency ranges from 0 to 2000 while pixel value ranges between 0.45 and 0.7. Considering this as the basis now we can talk about the forest densities.

8.4.2.2. Period: 2010-2015

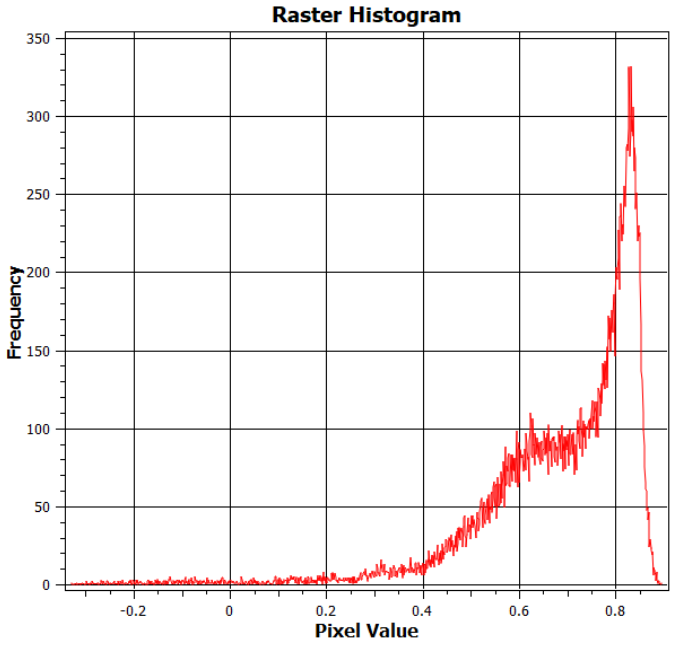

Figure 45: Histogram for the NDVI calculate for the period of 2010-2015

The above diagram depicts the frequency variation of Yala Forest during the period of 2010-2015. According to the above diagram you can observe forest density has reduced immensely during 2010 to 2015 as frequency ranges between 0 and 350. Sometimes it could be weather condition has been affected on it.

8.4.2.3. Period: 2015-2020

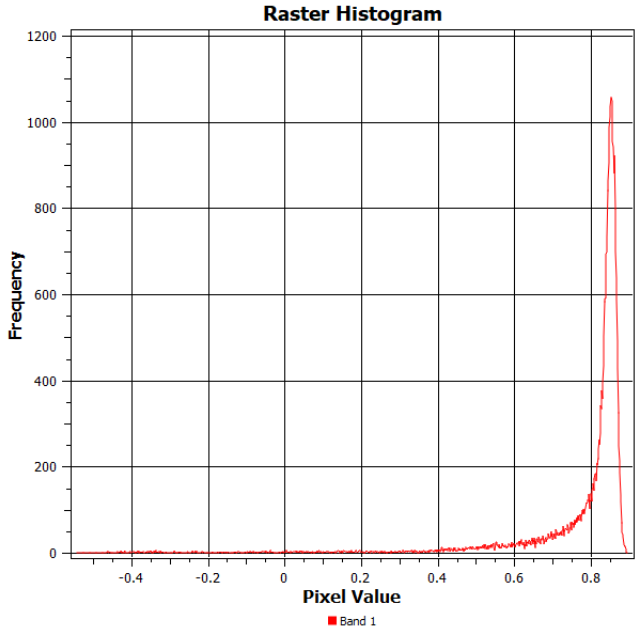

Figure 46: Histogram for the NDVI calculate for the period of 2015-2020

According to the histogram above, we can summaries during the period 2015 to 2020 the vegetation has increased in Yala forest. Because the frequency has again increased up to 1200 and also pixel value ranges majorly between 0.6 and 0.85. Therefore, we can conclude the forest density has increased in the Yala Forest by the start of 2020.

8.5. Chapter Summary

This chapter will summarise the all findings of the previous chapter. It will take out raster histograms for that purpose. Also White box and Black testing information can be found in this chapter. Therefore, now we can evaluate the research.

9. Evaluation

Content

9.1. Chapter Introduction

This chapter will evaluate all the research findings. It will discuss about strategies used to evaluate the research first and justifications for it. Also it will give a brief information on justification of the tools that had been used for evaluation.

9.2. Evaluation Strategy Used and Justification

In doing this research QGIS software was used and depending on the needs, some of its plugins had to be installed. It has been discussed in methodology discussed. There were many reasons to choose QGIS software over the other famous softwares like ARGIS in the market. QGIS and ArcGIS are geographic information systems(GIS) are used to analyse and edit spatial information. Therefore, both this software can be used for the research. But to work with this research successfully QGIS has chosen, as it is convenience. QGIS is an open software which is freely available on internet. Therefore, it will reduce unnecessary costs in completing this research project. This software can be installed on different operating systems without any barrier. Also there's no access barrier to the software as this is not a licensed software. In developing QGIS many contributions from programmers all over the world has been done.Hence considering all these factors above using QGIS for the research has been convenient.

9.3. Justification Accomplishment of Research Objectives

During the first phase of the project it was aim to find the current deforestation of the Sri Lanka. Accordingly, When It was looking for what have been done in the previous researches, people had followed different methods. As this research going to be totally depend on the data that will be gathered by remote sensing, finding the NDVI values of satellite images was one of the best option that could take for the completion of the project. As per the research objectives we could obtain the vegetation of forest area from it. According to the information we could obtain there has been deforestation happened in particular forest areas from the period of 2000 to 2020.

For further evaluation of NDVI maps we took previous researches that had been done in Sri Lanka. One of the best research we could tally is research by B.E. S Bandara, which was done in Postgraduate Institute of Agriculture, University of Peradeniya. From that aspect we could verify our NDVI maps for the prediction model.

Also by that we could identify the areas that have been deforested inside the Sri Lankan forest. It was one of the main research objective. For this purpose, QGIS software was taken as the tool along with the plugin semi-automatic classification. It helped to find the raster calculations and also to map out specific land areas from the satellite images.

9.4. Chapter Summary

In this chapter we have discussed about the evaluation process as a whole. All the justification for the using software and tools are given in this chapter. Also finally we discussed how it did help to achieving the research objectives. Hence now we can conclude the research.

10. Conclusion

Content

Chapter Introduction
Research Objective Accomplishment
Problem faced during the Research
Limitations
Practical Functionality of the Research Project
Future Improvements
Overall Conclusion

10.1. Chapter Introduction

As the end chapter it will discuss about how the objectives mentioned in the first chapter is accomplished, what problems that underwent in achieving the objectives and how further this model can be improved for predictions. Hence this chapter will conclude and justify about all the findings that was gathered in making the research a successful one.

10.2. Research Object Accomplishment

In doing this research project one of the important research objective was to clearly identify the areas which are in threat of deforestation. It has been clearly shown during the research by taking out most important forest areas in Sri Lanka and given how it has been threat by deforestation. As we have taken it in separate four periods it's very clear to observe how the threat level has increased each of the period. Therefore, this fulfils the one of the main objective of the research.

Another research object was to acknowledge people about the current situation and reduce the problems in society arisen due to deforestation. When you clearly go through this research project you can see the difference from 2000 to 2020. This will give you much enough information that how big the deforestation happened inside the Sri Lankan forests. So if there's anyone interested in the subject, he/she can further study about the legal and can act upon the justice. Therefore, this research project will help to resolve a significant problem in Sri Lanka to a certain extent.

10.3. Problems faced during the Research

One of the main thing that was needed for the completion of this project was finding the necessary satellite images. Because the whole project was depending on the amount of satellite images or the data that we can gather. At the beginning of the project it was planned to find deforestation in all the major forests in Sri Lanka. But it was not being possible to find enough images for some forest areas. Hence the project had to be limited to few number of forests areas. Even after finding the relevant data, some images were not rendering properly due to bandwidth problems. Therefore, those images were not giving possible NDVI values and such all the images were had to remove. It is also need to mention the storage capacity of these images are high. Therefore, sometimes keeping a set of satellite had been a problem. For that I used external storage to store the relevant images.

There were times the QGIS algorithm kept on running when we set to process and it never stopped and then the whole machine will get stuck. It was seen when raster images were trying to convert to a vector images to have cut the area of interest and have a shape file. Therefore, I had follow another way to get a mask created to cut the relevant area that is needed to analyse.

10.4. Limitations

One of the challenge in doing this research project was to find the required data to analyse. It was first planned to collect data from each year from 2000 to 2020, but the task was not being easy as it was scarce to find the images each year. Therefore, collection of images had to be collected according to certain period as 2000-2005, 2005-2010, 2010-2015 and 2015-2020. When processing the images as some of the images didn't give proper results. Hence these periods couldn't be followed in some forest areas.

In making algorithms it's accuracy could have been increased and more variables could have been introduced if we had a site visit. But as it is needed to stick to the time period the data only taken by satellite images are considered in preparing the algorithm.

Also it was planned first to take some information from the Forest Department Sri Lanka to clarify and prove our findings further. But unfortunately due to the pandemic situation It was not been possible as most of the public sector offices were not working until the submission of this research project. Therefore, it was taken the previous researches that had been done to evaluate NDVI maps which we have discussed in previous chapter.

10.5. Practical Functionality of the Research Project

When we consider this research project it had two phases. During the first phase we identified the areas that have been deforested during the past two decades and in next phase we have tried to produce a prediction model to find future deforestation. Finding these forest information is very much important as we were having very burning social problems among ethnic groups due to deforestation. People were in doubt whether it is within the forest areas or not. So by observing the conclusions we have made in this project anyone can get a clear idea about it. By identifying factor, we have made an approach to predication model. Hence by developing it further we can aware the authorities about these deforesting areas prior to any further issues can happen.

10.6. Future Improvements

During this research project we focussed on a few forest areas of Sri Lanka. Therefore, this project can be extended more covering larger area in future work. When we consider the technical side of the project our prediction model is based on the normalized difference vegetation index (NDVI). So this model can be improved more with different approach which uses different index to calculate forest density.

Also the accuracy of this model can be increase in several ways. One method is identifying the more factors that could affect the deforestation in the specific areas This means you can have more data. Because when we have more data the neural network can generalize well. Hence having more data will affect in the quality of the training data as that will provide a broad picture of the problem.

10.7. Overall Conclusion

In this research paper deforestation of Sri Lanka has been taken in to consideration. By doing this research we could observe, in different areas there's a trend of land is clearing for different purposes. To analyse it we have used several technologies of Remote sensing and machine learning techniques. Specially this research discusses about the applicability of Artificial Neural Networks for the deforestation prediction. By using this model Land cover changes are analysed and investigated. However, it is difficult to predict deforestation exactly as it depends on many environmental factors, effect of climatic patters, social and political situations. But we can improve the accuracy of the model more by introducing the factors as much as we can. Hence when we consider this model it can be used in Sri Lanka by environmentalists, politicians and other policy makers to control economic and social impacts from deforestation.

References

1. Ahmadi, V., 2018. *Using GIS and Artificial Neural Network for Deforestation Prediction,* s.l.: Preprints.
2. Bishop, C. M., 1995. *Neural Networks for Pattern Recognition.* s.l.:Oxford University Press, Inc. New York, NY, USA.
3. Gandhi.G, . M., Thummalu , N., S. & . C. A., 2015. Ndvi: Vegetation change detection using remote sensing and gis – A case study of Vellore District. *3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)*.
4. Hooper, D. U., Chapin , F. S., Ewel, J. & Wardle, D., 2005. Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge. *Ecological Monographs*.
5. Hurtt, G. C. et al., 2007. The Effects of Deforestation on the Hydrological Cycle in Amazonia: A Review on Scale and Resolution. *International Journal of Climatology.*
6. Liang , C. & Tang, M., 2017. *Understand Amazon Deforestation using Neural Network,* s.l.: s.n.
7. Renno, F., Paegelow, M., Follador, M. & Bruno, R., 2006. Integrating Remote Sensing, GIS and Prediction Models to Monitor the Deforestation and Erosion in Peten Reserve, Guatemala. *Society for Mathematical Geology .*
8. Tra, H. H., 2014. Hydropower and deforestation: An unexpected linkage or an obvious one?. *Forest Asia.*
9. Amuchastegui, A., Riveros, J. C. & Forrest, J. L., 2014. Identifying areas of deforestation risk for REDD+ using a species modeling tool. *Carbon Balance and Management.*
10. Anon., 2014. *Committee on Needs and Research Requirements for Land Change Modeling,* s.l.: THE NATIONAL ACADEMIES PRESS.
11. Anon., 2018. Conservation of Biodiversity for Sustainable Development in Sri Lanka. *Daily Mirror online*, 22 May.
12. Anon., n.d. Quantum GIS (QGIS). *DataONE.*
13. Anon., n.d. Sentinel-2A (10m) Satellite Sensor.
14. Arekhi , S., 2011. Modeling spatial pattern of deforestation using GIS and logistic regression: A case study of northern Ilam forests, Ilam province, Iran. *African Journal of Biotechnology.*
15. Barraclough, S. L. a. K. B. G., 2000. *Agricultural Expansion and Tropical Deforestation: Poverty, International Trade and Land Use. Earthscan, Sterling, Virginia, USA,* s.l.: s.n.
16. Bennett, L., 2017. *Deforestation and Climate Change,* s.l.: s.n.
17. Bogahawatte, C., 1999. *Forestry Policy, Non-timber Forest Products and the Rural Economy in the Wet Zone Forests in Sri Lanka ,* s.l.: EEPSEA .
18. Bryant, D., Nielsen, D. & Tangley, . L., 1997. Last Frontier Forests. *Ecosystems and Economies on the Edge.*

19. Castriotta, A. G., n.d. Sentinels Data Sentinels Data Access Annual Report. *SERCO PUBLIC.*

20. Chomitz, W. A., Minasian, S. G., Sutton, A. D. & Arnold, J., 2007. *Synthesis and Reactivity of Metal Complexes Supported by the Tetradentate Monoanionic Ligand Bis(2-picolyl)(2-hydroxy-3,5-di-tert-butylbenzyl)amide (BPPA),* s.l.: s.n.

21. Christensen, N. L. & Pfaff, A., 2008. Road Impact on Deforestation and Jaguar Habitat Loss in the Mayan Forest.

22. Cooper, G. F., Hennings-Yeomans, P., Visweswaran, S. & Barmada, . M., 2010. An Efficient Bayesian Method for Predicting Clinical Outcomes from Genome-Wide Data. *AMIA Annual Symposium Proceedings,* p. 127–131.

23. Davydova, O., 2017. *7 types of Artificial Neural Networks for Natural Language Processing,* s.l.: s.n.

24. Dias, K. & Liyanawatte, C., 2017. Analysis on Deforestation and Environmental Law in Sri Lanka.

25. Dlamini, W. M., 2016. Analysis of deforestation patterns and drivers in Swaziland using efficient Bayesian multivariate classifiers. *Modeling Earth Systems and Environment.*

26. Eastman, J., 2006. IDRISI Andes Guide to GIS and Image Processing.

27. Fernando, P., 2019. Inception Workshop How critical is the human- elephant conflict ?. *Daily Mirror Online*, 16 January.

28. Fernando, P., Fernando, K. & Kumarasiri, M., 2009. Forced to Move : Involuntary Displacement and Resettlement — Policy and Practice. *9th Annual Symposium on Poverty Research in Sri Lanka.*

29. G. Cornelius van, K. & Bulte, E. H., 2000. *The Economics of Nature: Managing Biological Assets,* s.l.: s.n.

30. Hu, Z. & Lo, C., 2007. Modeling urban growth in Atlanta using logistic regression. *Computers Environment and Urban Systems .*

31. J.F. Mas, H. P. J. P. ,. A. S.-L., 2003. Modelling deforestation using GIS and artificial neural networks. *Environmental Modelling & Software .*

32. Jayasuriya, R., 2019. Wilpattu, Islamophobia. *Daily Mirror.*

33. Jong, S. M. d., Meer, F. D. v. d. & Clevers, J. G., 2004. Basics of Remote Sensing. *Remote Sensing Image Analysis: Including the Spatial Domain,.*

34. Lentin, M. et al., 2013. *Selective Logging and Its Relation to Deforestation,* s.l.: s.n.

35. Mayfield, H., 2015. *Making the most of machine learning and freely available datasets:,* s.l.: The University of Queensland .

36. Myers, N. et al., n.d. *Biodiversity hotspots for conservation priorities,* s.l.: s.n.

37. Okia, C. A., 2012. *Global Perspectives on Sustainable Forest Management,* s.l.: s.n.

38. Okia, D. C. A., 1990. *Global Perspectives on Sustainable Forest Management,* s.l.: s.n.

39. Packiam, D., 2015. Deforestation: Causes And Consequences. *Valley International Journal.*

40. Pinker, R. T., Thompson, O. E. & Eck, T. F., n.d. *The albedo of a tropical evergreen forest,* s.l.: s.n.

41. Romijn, E. et al., 2015. Assessing change in national forest monitoring capacities of 99 tropical. *Forest Ecology and Management.*

42. Schmink, M., 1992. *The socioeconomic matrix of deforestation,* s.l.: s.n.

43. Sheram & Katherine, n.d. *The environmental data book : a guide to statistics on the environment and development,* s.l.: 1993.

44. Sooriyagoda, L., 2019. Deforestation of Wilpattu national park; writ petition fixed for judgment. *Daily News.*

45. Upadhyay, P. & Gupta, S., n.d. *Introduction To Satellite Imaging Technology And Creating Images Using Raw Data Obtained From Landsat Satellite,* s.l.: s.n.

46. USGS, n.d. EarthExplorer Help Documentation. *USGS science for a changing world.*

47. Xu, E. & Zeng, O., 2017. *Predicting Amazon Deforestation with Satellite Images,* s.l.: s.n.

48. Zeraatpisheh, M. et al., 2013. The Effect of Deforestation on Soil Erosion, Sediment and Some Water Quality Indicators. *Conference: First International Conference on Environmental Crisis and its Solution.*