Flood Analysis with satellite images using Advanced Machine learning techniques

A Project Report

Submitted by

Sourabh Bohare	**111508038**
Swapnil kadam	**111508032**

in partial fulfillment for the award of the degree

of

B.Tech Information Technology

Under the guidance of

Mr. Suraj Sawant

College of Engineering, Pune

DEPARTMENT OF COMPUTER ENGINEERING

AND

INFORMATION TECHNOLOGY,

COLLEGE OF ENGINEERING, PUNE-5

May, 2019

DEPARTMENT OF COMPUTER ENGINEERING

AND

INFORMATION TECHNOLOGY,

COLLEGE OF ENGINEERING, PUNE

CERTIFICATE

Certified that this project, titled FLOOD ANALYSIS WITH SATELLITE IMAGES USING ADVANCED MACHINE LEARNING TECHNIQUES has been successfully completed by

Sourabh Bohare	**111508018**
Swapnil Kadam	**111508032**

and is approved for the partial fulfillment of the requirements for the degree of "B.Tech.Information Technology".

SIGNATURE	SIGNATURE
Mr. Suraj Sawant	**Dr. Vahida Attar**
Project Guide	**Head**
Department of Computer Engineering	**Department of Computer Engineering**
and Information Technology,	**and Information Technology,**
College of Engineering Pune,	**College of Engineering Pune,**
Shivajinagar, Pune - 5.	**Shivajinagar, Pune - 5.**

Abstract

Floods have disastrous effect on humans who suffer from it. The livestock, crops, property, land, etc get devastated and predicting the exact loss is highly difficult with normal and primitive methods of accessing, hence satellite imagery comes into picture.

In this project we intend to compute approx damage cost by floods to property, crops, infrastructure with the help of satellite images by studying pre-flood and post-flood conditions of particular area and analyze the loss in different areas and fields with the help of images of all spectrums including those beyond visible spectrum like infrared, Microwave, ultra-violet. The images for this purpose will be taken from satellites like Sentinel-2(ESA), Landsat 9(NASA), Digital globe, IRNSS (India) and the database of USGS. The study area is going to be on recent Kerala floods

Our approach is to analyze the damage using advanced machine learning algorithms.

Contents

List of Tables

List of Figures

List of Abbreviations

GIS : Geographic Information System

SAR : Synthetic Apperture Radar

CART : Classification and Regression Tree

IMD : Indian Meterological Department

KCN : Kohonen Clustering Network

DEM : Digital Elevation Model

Chapter 1

Introduction

1.1 Purpose

Flood disaster is one of natural hazard that affected human safety and economic growth in the world. As inferred by white in 1945, Floods are act of God, but flood losses are largely act of Man. In recent years, strategies for the mitigation and prevention of flood disasters have shifted from a flood defense approach, aimed at controlling the hazard by means of structural measures, to a flood management approach, based on comprehensive risk assessment studies and costs and benefits analyses .Following this trend, an increasing number of flood risk modeling tools are becoming available, ranging from simpler approaches to sophisticated methodologies involving hydrological and hydraulic models. A key part of risk assessment studies consists in the estimation of the vulnerability of flood-prone areas, that is, the consequences that could be caused, in terms of damage and loss, by flood events of different magnitudes.

However, while considerable efforts are often put into flood hazard evaluation, much less elaborated methods are usually applied for damage and loss assessment. Second, consistent flood risk assessment requires taking into account all the possible sources of hazard in order to estimate the likelihood of a given area of being hit by a flood of a certain intensity, which is referred to, in the context of the present study, as flood susceptibility. However, in many cases, for instance in flood plain areas protected by multiple embankments, this would require the evaluation of a large number of flood scenarios, each one characterized by its magnitude and probability .As such, flood susceptibility analysis for complex scenarios is generally performed using Geographic Information System (GIS) approaches that take into account different layers of information, such as land use, elevation or existing flood maps. Although simple and fast, the results cannot be compared with hydraulic models, as the level of information provided is often only qualitative (i.e. floodno flood). For this purpose we have taken Kerala floods as our basic source.

1.2 Objective

In this project we will be analyzing the effect which the flood had on Kerala. Kerala experienced an abnormally high rainfall from 1 June 2018 to 19 August 2018. This resulted in severe flooding in 13 out of 14 districts in the State. As per IMD data, Kerala received 2346.6 mm of rainfall from 1 June 2018 to 19 August 2018 in contrast to an expected 1649.5 mm of rainfall. This rainfall was about 42% above the normal. Further, the rainfall over Kerala during June, July and 1st to 19th of August was 15%, 18% and 164% respectively, above normal.

-To analyze the damage caused by flood by comparing pre flood and post flood images.

-To provide details of the location where heath needs to be provided.

-To develop a efficient machine learning program which can be used to analyze damages caused by other floods.

1.3 Background

Period	Normal Rainfall	Actual Rainfall	Departure from Normal
	(mm)	(mm)	(%)
June, 2018	649.8	749.6	15
July, 2018	726.1	857.4	18
1-19 August	287.6	758.6	164
Total	1649.5	2349.6	42

Fig 1.3.1.1 Rainfall Analysis

1.4 Earlier Floods in Kerala

In 1924 Kerala experienced huge loss of life, vegetation, property, crops, etc due to the largest flood. In 1961 too Kerala experienced huge loss due to flooding, entire state was under water for 9-10 days the rainfall this time was about 56% above normal. And now in 2018 was another big flood which caused loss of many lives. The analysis of its rainfall is provided in table and map below

Districts	Normal Rainfall(mm)	Actual Rainfall(mm)	Departure from normal	
Kerala	1701.4	2394.1	41	Excess
Alappuzha	1380.6	1784	29	Excess
Kannur	2333.2	2573.3	10	Normal
Ernakulam	1680.4	2477.8	47	Excess
Idukki	1851.7	3555	92	Large Excess
Kasargod	2609.8	2287	-12	Normal
Kollam	1038.9	1579.4	52	Excess
Kottayam	1531.1	2307	51	Excess
Kozhikode	2250.4	2898.2	29	Excess
Malappuram	1761.9	2637	50	Excess
Palakkad	1321.7	2285.6	73	Large Excess
Pathanamthitta	1357.5	1968	45	Excess
Thiruvananthpuram	672.1	966.7	44	Excess
Thrissur	1824.2	2077.5	14	Normal
Wayanad	2281.3	2884.6	26	Excess

Fig 1.4.1.1 District-wise Rainfall

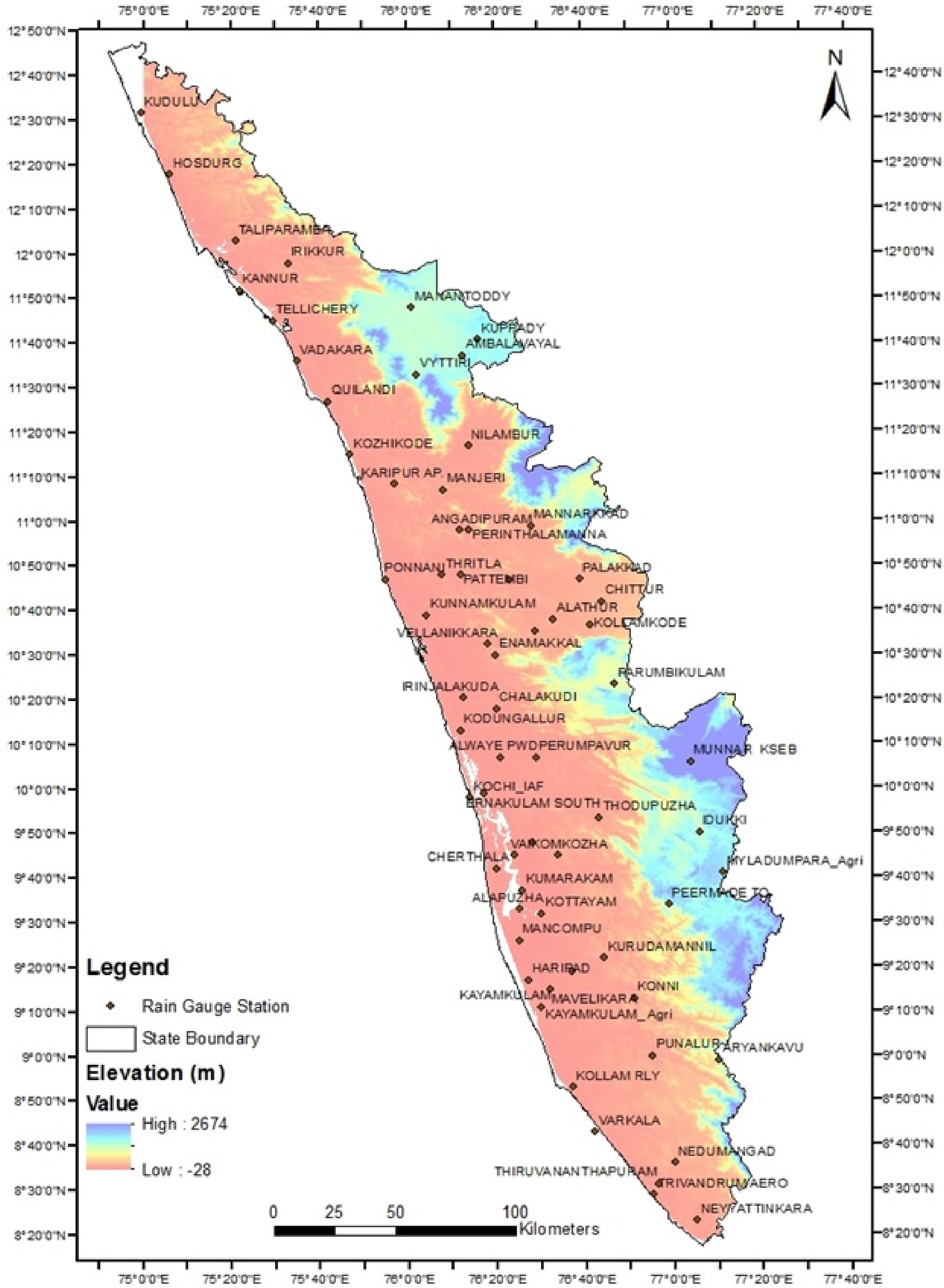

Fig 1.4.1.2 Rainfall in Kerala

1.5 Further into Subject

Remote sensing studies of flooding dynamics to date have focused almost exclusively on rural reaches. Urban areas, where most assets are located and where flood risk is thus very high, have been given relatively little attention, mainly due to lack of distributed data. Kerala Flood had Impact on rural as well as urban area.

Several European projects were developed in order to analyze flood and other natural hazards, e.g. EUROflood, FLOODsite 12, ConHAZ 13, and CapHAZ 14. These projects contributed to the development of several methods to evaluate flood risk and damage potential. Following Methods of flood assessment are currently available

1. Actual and Potential damage assessment

a. Actual damage refers to the damage that has occurred in past and Potential damage means which could occur in future

b. Posteriori evaluations serve to estimate actual economic damage caused by one or many flooding events occurred in the past; and a priori or ex ante estimations are used to forecast the potential damage related to theoretical future flood scenarios, i.e. estimation of potential flood damage (Hubert and Ledoux, 1999).

This Assessment was done by Central Water Commission. 2. Conceptual Methods to estimate damage potential

a. Based on the capacity of the market to associate a monetary value to the flood risk. The fundamental hypothesis of this analysis is that the reduction of flood risk is expected by the society, which is ready to pay for it (Nascimento et al., 2007).

b. Eg-: Benefit Cost Analysis

3. Deterministic method to estimate damage potential

a. Based on a detailed estimation of the flood damage, passing by a precise description of the vulnerability of assets to suffer damage. These methods serve to evaluate direct and indirect damage; however, they generally concentrate on direct tangible damage due to the complexity linked to indirect damaging processes (Hubert and Ledoux, 1999; D4E, 2007).

b. E.g -: GIS Based Method

We will be using Classification and Regression Trees (CaRT) Algorithm, Random Forest Algorithm, K-Means grid Clustering Algorithm along With Extended Learning Machine and Artificial Intelligence. The dataset would be from Latest LANSAT8, Sentinel 2, which will be processed using above mentioned algorithm.

Chapter 2

Literature Survey

There are many methods for analyzing floods; few may be by survey, by satellite images, and by AI implementation in particular area. Of these methods Survey is a primitive technique and AI implementation is not a cost effective technique in this time. The most efficient and effective method comes out to be analyzing satellite images using data mining algorithms. The base study in this domain was done by Bates et al in 2006 on the UK town of Tewkesbury. This led to further advancements in study. Next groundbreaking study was done by Ismail Elkhrachy on Najran City of Saudi Arabia, it gave detailed report about percentage of people affected in this area and sector-wise loss was calculated. Later studies were done by same methods using different algorithms. Most efficient study till now was by wavDAE. Given are different approaches for study and their advantages along with their drawbacks.

2.1 Sequential aerial photography and SAR data Method

This was first effort taken to analyse flood and its effected areas in Urban Area.

Advantages:

1. Developed a unique data set consisting of a series of eight space borne

SAR and aerial photographic images of flooding.

2. All Day and Night analysis can be done which help find more accurate results.

Disadvantages:

1. Areas that are unseen by SAR due to shadow are caused by layover effects .The physical an interaction between emergent vegetations was not shown as well as the radar signal was difficult to stimulate.

2. The accuracy that they obtained is only 92% which also can be only acquired if there are only good quality images available.

3. This study is not detailed because of data inadequacies and absent of record.

4. They used GO GO Model analysis instead of GO PO model would be advantageous to deal with the problem.

2.2 Satellite supported flood forecasting in river networks.

Used Ensemble Kalman (EnKF) Filters as it is very promising technique for assimilation. Satellite-based (e.g., Synthetic Aperture Radar [SAR]) water level observations (WLOs) of the floodplain can be sequentially assimilated into a hydrodynamic model to decrease forecast uncertainty.

Advantages: -

1. Ensemble Kalman Filters used in it is very promising techniques for assimilation.

2. Also used Satellite based WLO of the floodplain can be sequentially assimilated into a hydrodynamic model to decrease forecast uncertainty.

Disadvantages: -

1. The analysis can be only done up to couple of hours.

2. It cannot also work for more than 24 hours.

2.3 Identification of flooded area from satellite images using Hybrid Kohonen Fuzzy C-Means sigma classifier

Approach-: A novel neuro fuzzy classifier Hybrid Kohonen Fuzzy C-Means-r (HKFCM-r) is proposed in this paper. The proposed classifier is a hybridization of Kohonen Clustering Network (KCN) with FCM-r clustering algorithm. The network architecture of HKFCM-r is similar to simple KCN network having only two layers, i.e., input and output layer. However, the selection of winner neuron is done based on FCM-r algorithm. Thus, embedding the features of both, a neural network and a fuzzy clustering algorithm in the classifier.

Advantages-:

1. This hybridization results in a more efficient, less complex and faster classifier for classifying satellite images.

2. The error matrix was computed to test the performance of the method.

3. The method yields high producers accuracy, consumers accuracy and kappa coefficient value indicating that the proposed classifier is highly effective and efficient.

Disadvantages-:

1. The result was that it detected changes in vegetation, the total loss of

forest area the total area of flood. Though the results of this study were efficient, it could not analyze the loss due at various heights of Kashmir as the region is full of terrians the result would have been more accurate if varying height would have been considered.
2. Digital Elevation Model(DEM) should have been used. Also it gives only the details of changes in vegetations and does not predict the exact loss due to floods.

2.4 A methodology for flood susceptibility and vulnerability analysis in complex flood scenarios

Approach-: The methodology is based on a GIS-based index called Flood Intensity Index (Iw), which can be considered as a trade-off between morphometric indexes and physically based two-dimensional (2D) hydraulic models. It synthesizes the spatial distribution of the index values into maps and curves that can be used to rank the susceptibility and implement a vulnerability analysis in the area of interest.
Advantages-:
1. Used Water level as a measure of Hazard.
2. Water level in rivers can easily be related with discharge and magnitude of flow, and can be estimated both by on-site measurements and simulations of 1D hydraulic models.
Disadvantages-:
1. As expected, the results showed limitations due to the approximations applied (absence of temporal evolution of flow, no draining processes, no estimation of velocity), and it should be clearly understood that the analysis

performed with It cannot reach the level of detail and accuracy of 2D hydraulic models.

2. Nevertheless the results show that the algorithm is able to represent reasonably well flow dynamics in lowland areas.

2.5 Flood Hazard mapping using satellite images and GIS tools

Though it is the most latest and most efficient method, more study needs to be done in this field to get accurate results.

Advantages: -

1. Uses analytical hierarchical process to determine relative impact weight of flood factors to get a composite flood hazard index.

2. Solves Mixed Pixel problem and large handling data present in these images.

Disadvantages: -

1. This study is not detailed because of data inadequacies and absent of record.

2. RBFN is efficient for regular peaks and values. The network is also preferable as it gives more efficient results.

Chapter 3

Methodology

We analyzed pre-flood and post-flood spatio-temporal images of Kerala. These images were from satellites like Landsat8 and Sentinel 2. The Images were subjected to various algorithms to find the difference and to get maximum accuracy of the difference. The analysis was done using two different methods, one was with the Google earth engine and other was using Q-Gis tool. In both methods the images were subjected to same algorithms and it produced the results.

For Google Earth Engine particular area of Kerala was selected and seed-points of data like Land, water, Forest, etc were provided and the algorithms like CART, Random Forest, and K-Means were applied to produce results. The Seedpoints were supported with Imagery from Landsat8 and Sentinel 2 which provides good quality of data. The data was sided with a good code in Python.

The GIS tool was used to analyze the images from Sentinel 2. Many tiles of imagery for whole Kerala state (Before and After Flood) were downloaded first and then they were processed. The process included selecting BAND 1-10 of the images, then converting each image from Vector to Raster form, then combining those images to one single image. The later process was mosaicing all the Combined tiles to form one single image of Kerala state. The

Image was then subjected to various processes and algorithms like Watershed Segmentation, Terrain Analysis, Random Forest Algorithm, K-Means Algorithm, etc.

The methodology for second will be as follows.

1. Downloading of data of Sentinel 2 from USGS Earth explorer
2. Pre Processing using Q-GIS of the obtained Spatio-Temporal Images
3. Atmospheric Correction and Image Mosaicing
4. Layer Stacking
5. Analysing Difference of two images using ELM

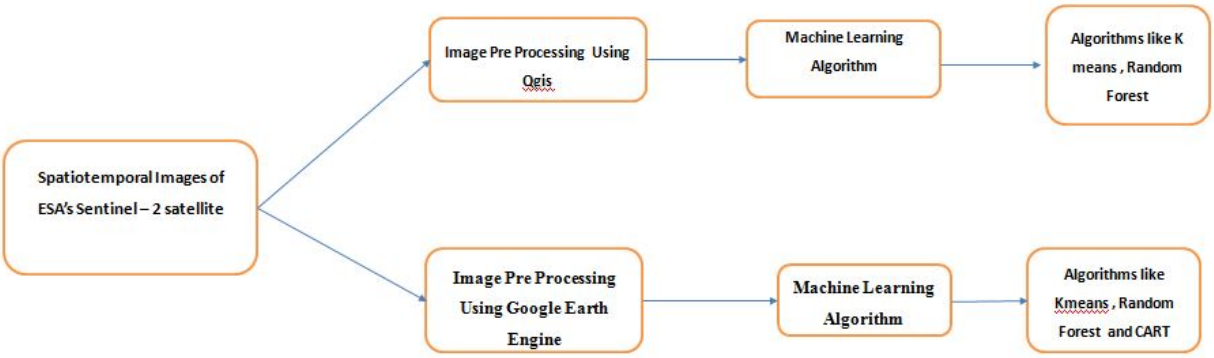

Fig 3.1.1 Methodology

Chapter 4

Implementation

4.1 Datasets

Dataset of Sentinel 2 and Landsat 8 was used.

4.2 Algorithms

Following algorithms were used:

For Classification

1.Random Forest

2.CART

3.K-Means

For Water Recognition

1.Watershed Segmentation

2.K-means

3.Terrain Analysis

4.3 Why GIS Based Analysis?

- **Most Recent**

  Q-gis is a free and Open source platform which supports viewing, editing, and analysis of GoeSpatial Data. It Uses Python for coding which gives it an edge over other methods of Analysis.

4.4 Random Forest Algorithm

It is a Flexible, easy to use ML Algorithm that produces great results for most of the times. It is one of the most used algorithms, as it is simple and can be used for both classification and regression tasks.

How it Works?

Random Forest is a supervised ML algorithm. It creates a forest and makes it somehow random. The forest it builds, is an ensemble of Decision Trees, most of the time trained with the bagging method. The general idea of the bagging method is that a combination of learning models increases the overall result.

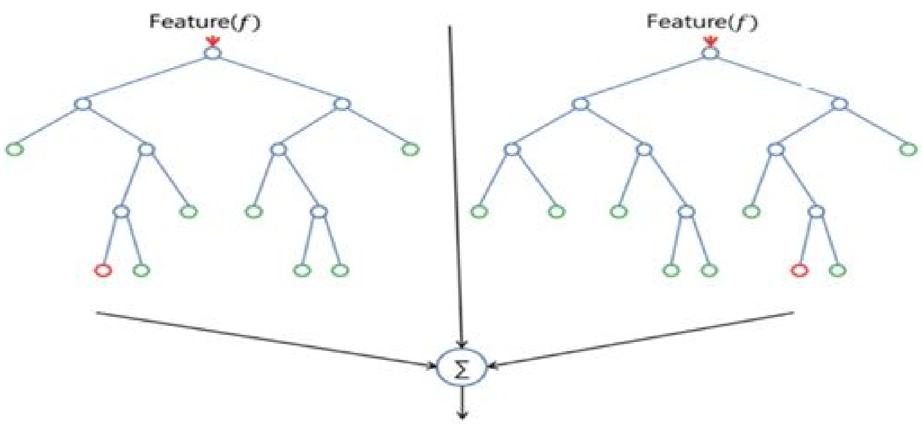

Fig 4.1.1 Random Forest

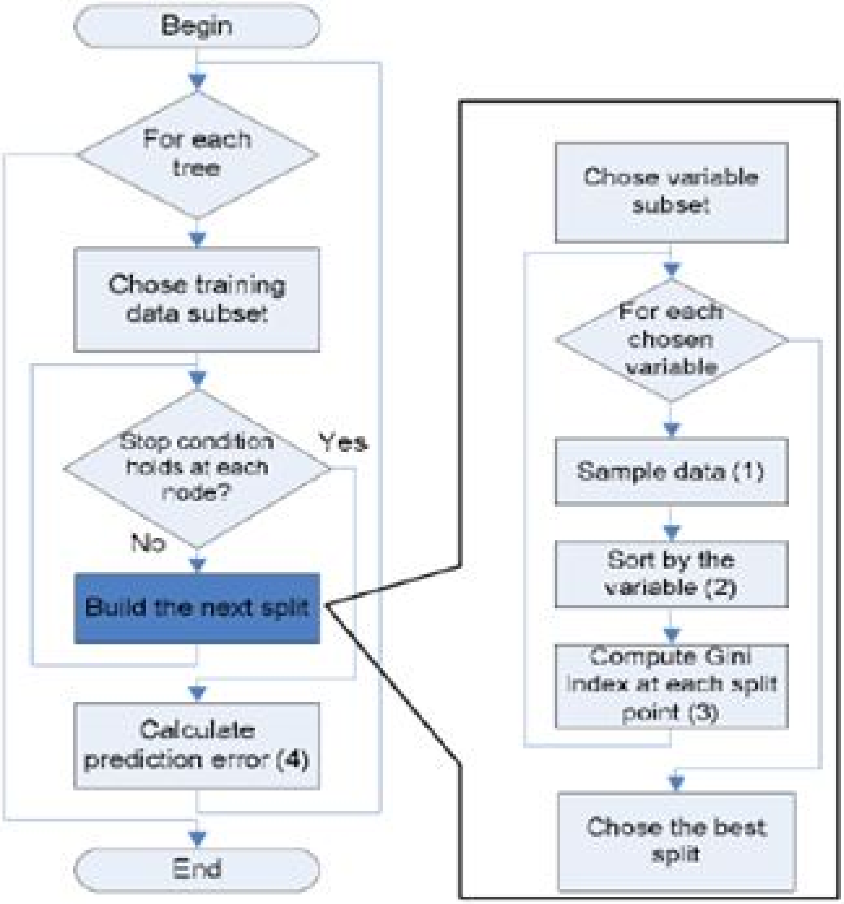

Fig 4.1.1 Random Forest Algorithm Flow Chart

Chapter 5

Results, Statistics and Inferences

5.1 Result and Statistics

5.1.1 Pre Flood Analysis Using Random Forest Algorithm

	Points	Water	Land Non-Forest	Mild Forest	Shrubs and Grass	Dense Forest
Water	130	90.45%	1.65%	2.35%	1.42%	4.13%
Land Non-Forest	25	4.76%	95.24%	0	0	0
Mild Forest	32	0%	0%	90.63%	0%	9.38%
Shrubs and Grass	27	1.48%	0%	0%	98.52%	0%
Dense Forest	96	4.55%	0%	2%	0%	93.18%

Fig 5.1.1 Random Forest

5.1.2 Post Flood Analysis Using Random Forest Algorithm

	Points	Water	Land Non-Forest	Mild Forest	Shrubs and Grass	Dense Forest
Water	130	97.66%	0.78%	0	0.78%	0.78%
Land Non-Forest	25	15%	75%	5%	0%	5%
Mild Forest	32	0%	0%	96.43%	0%	3.57%
Shrubs and Grass	27	3.70%	0%	0%	96.30%	0%
Dense Forest	96	2.08%	0%	4.17%	3.13%	90.63%

Fig 5.1.2 Random Forest Post Flood

5.1.3 Post Flood Analysis Using CART Algorithm

	Points	Water	Land Non-Forest	Mild Forest	Shrubs and Grass	Dense Forest
Water	130	99.23%	0	0	0.77%	0%
Land Non-Forest	25	9.52%	90.48%	0	0	0
Mild Forest	32	0	0	81.25%	6.25%	12.50%
Shrubs and Grass	27	11.11%	0	0	81.48%	7.41%
Dense Forest	96	2.08%	5.21%	1.04%	1.04%	90.63%

Fig 5.1.3 CART Algorithm Post Flood

5.1.4 Post Flood Analysis Using K-Means Classifier

	Points	Water	Land Non-Forest	Mild Forest	Shrubs and Grass	Dense Forest
Water	130	100%	0%	0%	0%	0%
Land Non-Forest	25	4.76%	90.48%	4.76%	0%	0%
Mild Forest	32	0%	0%	90.63%	9.38%	0%
Shrubs and Grass	27					
Dense Forest	96	6.67%	0%	0%	0%	93.33%

Fig 5.1.4 K-Means Classifier Post Flood

5.1.5 Algorithm and Their Accuracy

Algorithm	Accuracy
Random Forest	96.76%
CART	92.31%
Kmeans	96.31%

Fig 5.1.5 Algorithm vs Accuracy

5.1.6 Algorithm and Their Accuracy

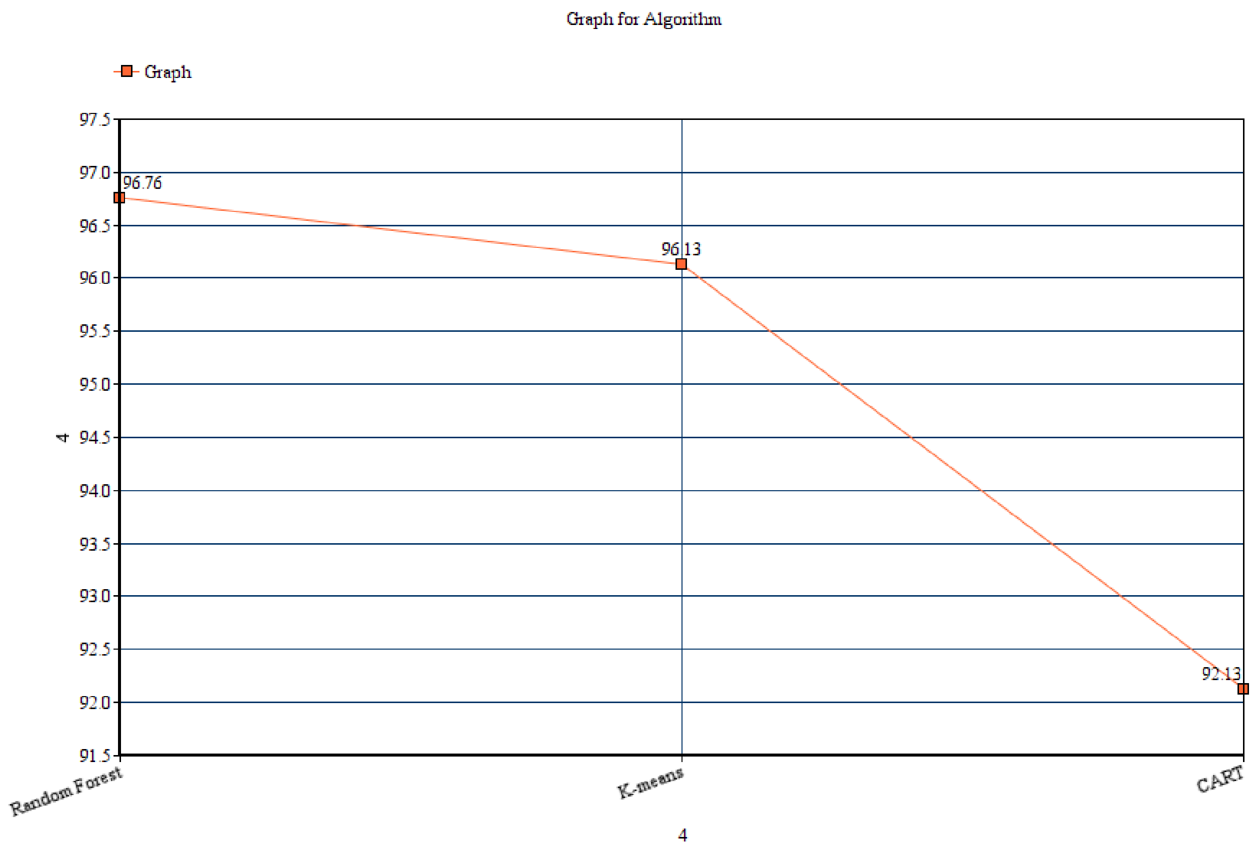

Fig 5.1.5 Algorithm vs Accuracy Graph

5.1.7 Pre Flood And Post Flood Kerala using Google Earth Engine

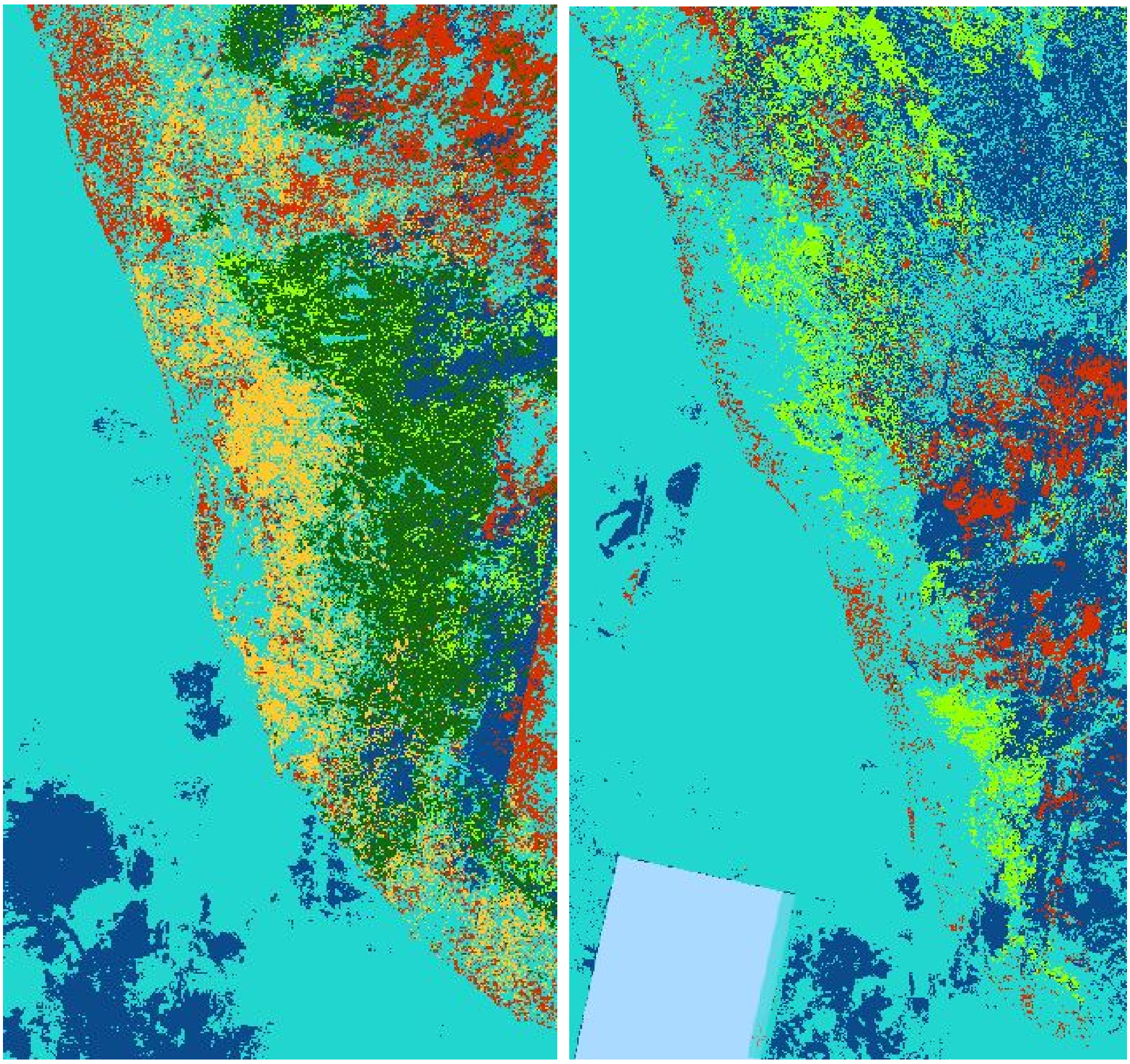

Fig 5.1.6 Before Flood, After Flood

5.2 Inferences

We have used three algorithms to find out the accuracy of the land damage caused in Kerala. The list of algorithms is as follows:

1. Random forest algorithm:

We have considered 5 entities for classification, they are water , Land non forest , mild forest ,shrubs and grass and dense forest. The preflood analysis of Sentinel 2 satellite shows us that there is 90.45% water available in water bodies or water like bodies but in post flood analysis shows us that there is 97.66% water available in water bodies. This shows us that the water has increased by 7.21% during floods. Even after the water was reduced after floods still there was lot of water present in Non forest land. The preflood results shows that there was 4.67% water present on Land non forest before flood but the same rised up to 15% water on land non- forest. Also due to heavy floods the Dense forest was reduced to 90.63% from 93.18%. This shows us that kerala faced a heavy loss in Non forest land as well as forest land. The overall accuracy of Random forest preflood was 93.66% and Post flood was 96.76%.

The second algorithm that we used was CART (Classification and regression tree) in this algorithm we got 92.31% accuracy which was very less than Random forest .The third algorithm that we used was K-means algorithm here also we got accuracy of 96.31% which was a good result but still the accuracy was less than Random forest algorithm.

Chapter 6

Conclusion

- Out of total area of 38863 Sqkm area of Kerala, Approximately 18500 SqKm area was affected by Flood.
- Random forest algorithm gives more accuracy compared to K-means clustering and CART algorithm.
- K- means algorithm shows us more accurately the location of water in Qgis toolbox with compared to other algorithms.
- Watershed segmentation analysis gave accurate results of location of water

Chapter 7

Future Scope and Work

- This technology also can be used to predetermine the floods using weather change broadcast.
- Rescuing affected people by sending help to their location.
- Live tracking of damages caused during floods.
- Providing Instant medical services to Victims of floods.
- Flood alerts to all the villages that are on that particular riverbanks.
- DEM data if used will give the result for area where help is to be provided.

Chapter 8

Bibliography

- Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation P. L. HOUTEKAMER Meteorology Research Division, Environment and Climate Change Canada, Dorval, Qubec,Canada

- The accuracy of sequential aerial photography and SAR data for observing urban flood dynamics, a case study of the UK summer 2007 floods Guy J.-P. Schumann, Jeffrey C. Neal a, David C. Mason, Paul D. Bates

- Flood risk analysis : impact of uncertainty in hazard modelling and vulnerability assessments on damage estimations by Julian Eleutrio

- Detection of flooded urban areas in high resolution Synthetic Aperture Radar images using double scattering by D.C. Masona, L. Giustarinib, J. Garcia-Pintadoa, H.L. Clokec

- A fully automated TerraSAR-X based flood service Sandro Martinis, Jens Kersten, Andr Twele

- Flash Flood Hazard Mapping Using Satellite Images and GIS Tools: A case study of Najran City, Kingdom of Saudi Arabia (KSA) Ismail Elkhrachy

- Green cover change detection using a modified adaptive ensemble of extreme learning machines for North-Western India Madhu Khurana, Vikas Saxena

- Satellite image classification using Genetic Algorithm trained radial basis function neural network, application to the detection of flooded areas Akansha Singh , Krishna Kant Singh

- Performing Google Earth Engine Analysis: https://earthengine.google.com/