Estimation of Change in LU/LC Mapping with Identification of Digital Signature using AI/ML Techniques over Google Earth Engine

Thesis submitted in partial fulfillment

for the Award of Degree of

Doctor of Philosophy

By

Anubhava Srivastava

RAJIV GANDHI INSTITUTE OF PETROLEUM TECHNOLOGY

JAIS, 229304

PGI18001 ***2023***

CERTIFICATE

It is certified that the work contained in the thesis titled **Estimation of Change in LULC Mapping with Identification of Digital Signature using AI/ML Techniques over Google Earth Engine** by **Anubhava Srivastava** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive, Candidacy, and Open Seminar.

Dr. Susham Biswas

(Thesis Supervisor)

DECLARATION BY THE CANDIDATE

I, ***Anubhava Srivastava***, certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of ***Dr.Susham Biswas*** from *July 2018* to *July 2023,* at the *Department of Computer Science and Engineering,* Rajiv Gandhi Institute of Petroleum Technology, Jais. The matter embodied in this thesis has not been submitted for the award of any other degree. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and have not included them in this thesis and have not cited as my own work.

Date:

Place: Jais, Amethi **Anubhava Srivastava**

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of our knowledge.

_ _ _ _ _ _ _ _ _ _ _ _ _ _   _ _ _ _ _ _ _ _ _ _ _ _ _ _

(Dr. Susham Biswas)   (Head of Department)

Supervisor   **Computer Science and Engineering**

CERTIFICATE

CERTIFIED that the work contained in the thesis titled **Estimation of Change in LU/LC Mapping with Identification of Digital Signature using AI/ML Techniques over Google Earth Engine** by **Mr. Anubhava Srivastava** has been carried out under my supervision. It is also certified that he fulfilled the mandatory requirement of TWO quality publications that arose out of his thesis work.

It is further certified that the two publications (copies enclosed) of the aforesaid Mr. Anubhava Srivastava have been published in the Journals indexed by –

(a) SCI

(b) SCI Extended

(c) SCOPUS

Dr. Susham Biswas   **Dr.Kalka Dubey**

(Thesis Supervisor)   (Convener, DPGC)

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Estimation of Change in LULC Mapping with Identification of Digital Signature using AI/ML Techniques over Google Earth Engine

Name of the Student: Anubhava Srivastava

Copyright Transfer

The undersigned hereby assigns to the Rajiv Gandhi Institute of Petroleum Technology Jais all rights under copyright that may exist in and for the above thesis submitted for the award of the "Doctor of Philosophy".

Date: **Anubhava Srivastava**

Place: Jais, Amethi Roll No : PGI18001

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for the author's personal use provided that the source and the Institute's copyright notice are indicated.

Dedicated

to

My Loving Son

ACKNOWLEDGEMENT

I would like to start my thesis by showing gratitude towards the people who pave a path directly and indirectly for my PhD journey. This journey is like a roller coaster, fill of ups and downs with lots of emotions, happiness and cherishable memories.First and foremost, I would like to express my deepest and sincere gratitude towards my research supervisor Dr. Susham Biswas for his constant encouragement and guidance. He has been helping me out and supported me throughout the course of this research work. I also thank him for his trust and patience on me and giving me scientific freedom to explore various research ideas.

I express my earnest gratitude to Prof. A.S.K. Sinha, Director, RGIPT for unconditional support during challenging time.I want to thank all the faculty members of the Department of Computer Science & Engineering, RGIPT for their continuous support.

I would like to thank my parents, two strong pillars in my life for their constant support. Their love, care, trust and all the sacrifices made me the person who I am today.. I thank my wife Akanksha for making me believe in myself and for his continuous support and encouragement.

I would like to thank:Shruti, Rakesh and Vinarm for making such a stressful journey smooth and memorable one. They have endured encouragement, camaraderie and the wonderful memories that they have provided for a lifetime. For my entire PhD life I have been so fortunate for these seniors, friends and juniors. I have missed few names also but I thank all of them whom I have interacted, developed a friendship during my RGIPT days. I duly acknowledge each and every nonteaching staff at RGIPT for indirectly contributing to my research by saving my time.

....Anubhava Srivastava

TABLE OF CONTENTS

ACKNOWLEDGEMENT			v
Table of contents			vi
List of figures			ix
List of tables			xiii
Abbreviations/notations			xvi
Preface			xviii
Chapter 1	Introduction		1
	1.1	Doon Valley (Dehradun Region)	4
		1.1.1 Climate and climate change	5
		1.1.2 Water Availability and Sustainability	7
	1.2	Eastern Himalayan Region (East Sikkim)	8
		1.2.1 Climate and climate change	9
		1.2.2 Water Availability and Sustainability	11
	1.3	Urban Area (Lucknow)	11
		1.3.1 Climate and climate change	12
		1.3.2 Water Availability and Sustainability	13
	1.4	Motivation and Rationale for This Study	14
	1.5	Problem Statement	15
	1.6	Research Questions and Hypotheses	18
		1.6.1 Research Questions	18
		1.6.2 Hypotheses	19
	1.7	The Objectives	20
	1.8	Thesis Outline	22
Chapter 2	Performance Assessment of Sentinel 2 MSI and Landsat OLI Data for Land Use Land Cover Classification across Diverse Geospatial Study Area Using a Comparison between Supervised Machine Learning Classifiers		
	2.1	Abstract	25
	2.2	Introduction	26
		2.2.1 Types Of data	26
		2.2.2 Data Processing and Analysis	27
		2.2.3 Integration and Analysis	28
	2.3	Satellite Data	28
		2.3.1 Landsat Data Set	31
		2.3.2 Sentinel DataSet	33
	2.4	Classification Algorithm	36

	2.4.1	Classification and Regression Tree (CART) Algorithm	39
	2.4.2	Random Forest	41
	2.4.3	Gradient Tree Boost	42
	2.4.4	Support Vector Machine	44
	2.5	Accuracy	45
	2.5.1	Confusion Matrix	45
	2.5.2	Kappa Coefficient	47
	2.6	Methodology	50
	2.7	Result & Discussion	51
	2.7.1	Dehradun (Doon Valley):	51
	2.7.2	Lucknow (Urban Area)	54
	2.8	Conclusion	60
Chapter 3	Geospatial Analysis of Land Cover and Land Use Change Detection in Doon Valley for the Period 2018-2022: A Comparative Analysis Utilizing Classification Algorithms and Different Satellite Datasets		
	3.1	Abstract	65
	3.2	Introduction	66
	3.3	Methodology	69
	3.4	Result	73
	3.5	Discussion	85
	3.6	Conclusion	89
Chapter 4	Comprehensive Assessment of Long-Term Forest Transformation by Analyzing Spatial-Temporal Vegetation Indices from 2000 to 2022		
	4.1	Abstract	91
	4.2	Introduction	92
	4.3	Related Work	94
	4.4	Study Area	95
	4.5	Method & Data	97
	4.6	Result	100
	4.7	Conclusion	107
Chapter 5	Detecting Short-Term Forest Cover Changes Caused by Fires in Sentinel-2 Data using Spectral Indices and Machine Learning Algorithms Over Ernakulam Fire Incident		
	5.1	Abstract	109
	5.2	Introduction	110
	5.3	Related Work	112
	5.4	Study Area	116
	5.5	Method & Data	119
	5.6	Result	124

	5.7	Conclusion	131
Chapter 6	Development of User Interface for detecting Changes over Different Multispectral Band and indices.		
	6.1	Abstract	132
	6.2	Introduction	134
		6.2.1 Background	135
	6.3	The Challenge: Managing Complex Remote Sensing Data	136
	6.4	The Solution: Spectral Indices Interfaces for Land Cover Change Detection	137
	6.5	Method And Data	138
		6.5.1 Atmospherically Resistant Vegetation Index	140
		6.5.2 Enhanced Vegetation Index 2	142
		6.5.3 Normalized Difference Vegetation Index	144
		6.5.4 Normalized Difference Water Index	147
		6.5.5 Soil-Adjusted Vegetation Index	148
	6.6	Result	151
	6.7	Conclusion	156
Chapter 7	Conclusion and Future Scope		
	7.1	Conclusion	158
	7.2	Future scope	163
		7.2.1 Hyper-Resolution Mapping	163
		7.2.2 Uncertainty and Validation Metrics	164
		7.2.3 User-Friendly Interfaces and Decision Support Systems	164
		7.2.4 Environmental Justice and Equity Considerations	164
		7.2.5 Case Studies in Specific Domains	164
References			166
List of publications			

LIST OF FIGURES

Figure 1.1: Geographical Location of Study Area_Dehradun 5

Figure 1.2 : Daily Temperature over the Study Area_Dehradun in Kelvin 6

Figure 1.3: Monthly rain level over the Study Area_Dehradun in mm 7

Figure 1.4: Percentage of Water in Study area _Dehrdun 7

Figure 1.5: Geographical Location of Study Area_East Sikkim 8

Figure 1.6 : Daily Temperature over the study area East Sikkim in Kelvin in year 2023 10

Figure 1.7: Monthly rain level over the study area East Sikkim in mm mm between year (2018-2023) 10

Figure 1.8: Percentage of Water in Study area _East Sikkim 11

Figure 1.8: Geographical Location of Study Area_Lucknow 11

Figure 1.10 : Daily Temperature over the study area_Lucknow in Kelvin 12

Figure 1.11: Monthly rain level over the study area_Lucknow in mm 12

Figure 1.12: Percentage of Water in Study area_Lucknow 13

Figure 2.1 : Comparison of Landsat and Sentinels data sets over wavelength and Atmospheric Transmission 36

Fig 2.2 : Methodology used in comparing data sets and algorithm for finding better dataset and classification algorithm 50

Fig2.3: Output generated by Sentinel data sets with resolution 10m, , with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 over Dehradun 52

Fig2.4: Output generated by Landsat data sets with resolution 30m, , with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 over Dehradun 52

Fig2.5: Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 using Confusion Matrix Over Dehradun 53

Fig2.6: Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 Using Kappa Over Dehradun	54
Fig2.7: Output generated by Sentinel data sets with resolution 10m, , with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 over Lucknow	55
Fig2.8: Output generated by Landsat data sets with resolution 10m, , with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 over Lucknow	56
Fig2.9: Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 using Confusion Matrix Over Lucknow	57
Fig2.10: Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 Using Kappa Over Lucknow	57
Figure 3.1 : Methodology used in comparing data sets and algorithm for finding land cover change over study area Data Sets	70
Figure 3.2 : Output generated by Sentinel-2 data set, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2018	74
Figure 3.3 : Output generated by Sentinel-2 data set, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2018	74
Figure 3.4: Output generated by Sentinel-2 data set, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2020	76
Figure 3.5: Output generated by Landsat-8 data sets, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2020	76
Figure 3.6: Output generated by Sentinel-2 data set with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022	78
Figure 3.7: Output generated byLandsat-9 data sets, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022	79

Figure 3.8: Comparative Accuracy calculated by Confusion Matrix over Sentinel Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022 80

Figure 3.9: Comparative Accuracy calculated by Confusion Matrix over Landsat Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022 81

Figure 3.10: Comparative Accuracy calculated by Kappa over Sentinel Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022 82

Figure 3.11 : Comparative Accuracy calculated by Kappa over Landsat Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022 82

Figure 3.12: Land Cover Class Area computed by classification Algorithms (a) CART (b) Random Forest (c) Gradient Tree Boost (d) Support Vector Machine over Sentinel data sets in the year 2018, 2020, and 2022 83

Figure 3.13: Land Cover Class Area computed by classification Algorithms (a) CART (b) Random Forest (c) Gradient Tree Boost (d) Support Vector Machine over Landsat data sets in the year 2018, 2020, and 2022 84

Figure 4.1 : Geographical Location of Study Area 85

Figure 4.2 : Methodlogy Used in Processing of data for finding Change 98

Figure 4.3 :Forest Cover Area of East Sikkim in the Year 2000 101

Figure 4.4 :Forest Cover Area of East Sikkim in the Year 2010 102

Figure 4.5 :Forest Cover Area of East Sikkim in the Year 2023 102

Figure 4.6 : Change in NDVI between the Year 2000- 2023 103

Figure 4.7: Change in Enhanced Vegetation Index between the Year 2000- 2023 103

Figure 4.8:Change in NDWI between the Year 2000- 2023 104

Figure 4.9: Yearly Forest loss between the years 2000- 2023 over the study area 105

Figure 4.10: Change in LULC areas between the years 2000- 2023 106

Figure 5.1: Geographical Location of Study Area 116

Figure 5.2: Daily Average rain in Study Area between May 2022 - March 2023 117

Figure 5.3: Daily temperature in Study Area between May 2022 - March 2023 118

Figure 5.4: Percentage of water over study area 119

Figure 5.5: Methodology used in research 121

Figure 5. 6 : Output Generated over Landsat-9 data sets before fire incident 124

Figure 5.7 : Output Generated over Landsat-9 data sets after fire incident 125

Figure 5.8: Fire Incident happen between January- February Month 126

Figure 5.9: Fire Incident happens between February - March Month 126

Figure 5.10 : Change in Study Area during Fire Incident 127

Figure 5.11: Green Normalize Difference Vegetation Index Calculated on Study Area during Forest Fire Time over Landsat-9 and Sentinel 2 data set 128

Figure 5.12: Normalize Difference Water Index Calculated on Study Area during Forest Fire over Landsat -9 and Sentinel 2 data set 128

Figure 5.13 : Adjusted Transformed Soil Adjusted Vegetation Index Calculated on Study Area during Forest Fire over Landsat -9 and Sentinel 2 data set 129

Figure 5.14 : Enhanced Vegetation Index Calculated on Study Area during Forest Fire over Landsat -9 and Sentinel 2 data set 129

Figure 6.1: Work Flow Of Spectral Indicies Interface 151

Figure 6.2: Snip of User Interface available at Spectral Indices Interface (earthengine.app) 153

Figure 6.3: Google Earth Engine View of User Interface with selecting any area by using drawing tool available at Spectral Indices Interface (earthengine.app) 154

Figure 6.4: Google Earth Engine View of User Interface with GEE Repository URL available at Spectral Indices Interface (earthengine.app) 155

LIST OF TABLES

Table 2.1 :Bands, Wavelength, Spatial Resolution, Equatorial Crossing Time (E.C.T.), and the available date range for Landsat-8 satellite 32

Table2.2:Bands, Wavelength, Spatial Resolution, Equatorial Crossing Time (E.C.T.), and the available date range for Sentinel-2 satellite 35

Table 2.3 : Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2022 Over Dehradun 53

Table 2.4 : Accuracy of classification algorithm over both data sets using Kappa in the Year 2022 Over Dehradun 53

Table 2.5: Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2022 over Lucknow 56

Table 2.6 : Accuracy of classification algorithm over both data sets using Kappa in the Year 2022 over Lucknow 57

Table 3.1: Date of acquisition, Path , Row and Cloud Cover(%) over Satellite data 73

Table 3.2: Land cover Id and Sample Point for land cover classes 73

Table 3.3 : Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2018 75

Table 3.4 : Accuracy of classification algorithm over both data sets using Kappa in the Year 2018 75

Table 3.5: Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2020 77

Table 3.6 : Accuracy of classification algorithm over both data sets using Kappa in the Year 2020 77

Table 3.7: Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2022 80

Table 3.8: Accuracy of classification algorithm over both data sets using Kappa in the Year 2022 80

Table 4.1 : Band Uses for calculating changes in Forest Cover 97

Table 5. 1: Band Uses for calculating changes in Forest Cover 121

ABBREVIATIONS/NOTATIONS

ADEOS	Advanced Earth Observation System
AIRS	Atmospheric Infrared Sounder
ALI	Advanced Land Imager (on EO-1 satellite)
ALOS	Advanced Land Observing Satellite
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High-Resolution Radiometer
BRDF	Bidirectional Reflectance Distribution Function
CEOS	Committee on Earth Observation Satellites
CIR	Colour Infrared
DEM	Digital Elevation Model
DN	Digital Number (pixel value)
DTM	Digital Terrain Model
ENVISAT	Environmental Satellite
EOS	Earth Observing System
ERS-1	Earth Remote Sensing Satellite
ERTS	Earth Resources Technology Satellite
ETM	Enhanced Thematic Mapper
GIS	Geographic Information System
GMS	Geostationary Meteorological Satellite
GOES	Geostationary Operational Environmental Satellite
GPS	Global Positioning System
LAI	Leaf Area Index
LIDAR	Light Detection and Ranging

LUT	Look-up Table
LWIR	Long-Wave Infrared
MISR	Multi-Angle Imaging Spectro Radiometer
MLS	Microwave Limb Scanner
MODIS	Moderate Resolution Imaging Spectro radiometer
MSS	Multispectral Scanner
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
RADAR	Radio Detection and Ranging
Radarsat	Radar Satellite
RGB	Red, Green, Blue
SAR	Synthetic Aperture Radar
SIR-C	Shuttle Imaging Radar-C
SRTM	Shuttle Radar Topography Mission
SWIR	Short-Wave Infrared
TIMS	Thermal Infrared Multispectral Scanner
TIR	Thermal Infrared
TM	Thematic Mapper
TRMM	Tropical Rainfall Measurement Mission
UNEP-WCMC	UNEP World Conservation Monitoring Centre
UV	Ultraviolet
VIS	Visible Spectrum
NDVI	Normalized Difference Vegetation Index
NBR	Normalize Burn Rate

EVI	Enhanced Vegetation Index
NDWI	Normalize Difference Water Index
GEE	Google Earth Engine
RF	Random Forest
CART	Classification & Regression Tree
SVM	Support Vector Machine
GTB	Gradient Tree Boost

PREFACE

The realm of Earth observation, geospatial analysis, and environmental monitoring is undergoing a transformative revolution, one that is underpinned by a symphony of technological innovation, machine learning prowess, and a collective endeavor to better understand and protect our dynamic planet. This journey into "Estimation of Change in Land Use/Land Cover (LU/LC) Mapping with Identification of Digital Signature using Artificial Intelligence and Machine Learning Techniques over Google Earth Engine" stands as a testament to our unwavering curiosity, our commitment to scientific exploration, and our dedication to addressing some of the most pressing challenges facing our world today.

The initial chapter of the thesis serves as the introductory segment of the thesis, offering a comprehensive overview of the study area's geographical location, rainfall patterns, temperature variations, and water percentage. The research is conducted in a specific region whose geographical coordinates are elucidated, outlining its topographical and geological features, thereby setting the stage for the ensuing investigations. An extensive analysis of historical rainfall data is presented, emphasizing the variability and distribution of precipitation across different seasons and years. Concurrently, temperature patterns are explored, dissecting both diurnal and seasonal temperature trends, which play an instrumental role in shaping the local environment. Furthermore, the chapter delves into the water percentage of the study area, encompassing information on surface water bodies, groundwater availability, and land use practices. These foundational insights serve as the bedrock for the subsequent chapters of the thesis, contributing to a holistic understanding of the study area's environmental context.

The second chapter of thesis provide introductory of each embarks on a journey to explore the realms of land use/land cover (LU/LC) mapping, focusing on the application of machine learning and artificial intelligence (AI) techniques using Google Earth Engine. This chapter serves as a foundational framework by meticulously describing, comparing, and evaluating various classification algorithms and datasets essential for LU/LC mapping.

In the arena of classification algorithms, we delve into the intricacies of the methods that play a pivotal role in deciphering satellite imagery data. This includes, but is not limited to, Classification And Regression Tree (CART), Random Forest (RF), Gradient Tree Boost (GTB), and Support Vector Machines (SVM). We dissect each method's underlying principles, strengths, and limitations to provide a comprehensive understanding of their applicability in the context of LU/LC mapping.

Furthermore, this chapter pays heed to the selection and comparison of datasets that form the bedrock of the research. Datasets, such as Sentinel-2 and Landsat, are scrutinized for their spatial and temporal resolutions, spectral bands, and cloud cover percentages. The objective is to identify the most suitable datasets that align with the research goals and the specific geographic areas of study.

Chapter three focuses on the core objective of the research, which is the estimation and analysis of land cover changes over the Dehradun area. Leveraging advanced AI and ML techniques, this chapter engages in a rigorous examination of historical satellite imagery data to track, quantify, and understand the dynamics of LU/LC changes over time.

Urbanization, a significant factor in land cover transformation, is examined with a focus on Dehradun. The application of classification algorithms, as discussed in the second chapter, helps delineate the shifts in urban landscapes, urban expansion, and the resultant land cover changes. Additionally, the agricultural landscape is scrutinized to reveal the patterns and alterations in land use practices. Through a detailed investigation of land cover classification, we gain insights into agricultural dynamics and their impact on land cover.

This chapter forms the first of several regional case studies, providing a detailed assessment of LU/LC changes in the Dehradun area. The insights garnered here have the potential to influence regional planning and development policies by shedding light on evolving land cover patterns.

Moving forward, the research shifts its lens to the forested regions of Sikkim, a critical ecological zone. The fourth chapter employs AI and ML algorithms to identify and classify changes in forest cover over a specified time period.

In this chapter, we dissect the unique challenges posed by forested areas, such as dense canopies and complex ecosystems. Techniques for forest change detection are explored, including the use of multi-spectral and radar imagery for improved forest cover mapping. The chapter also dives into the identification of deforestation and afforestation events, which are vital in understanding the ecological health of the Sikkim forests.

The findings of this chapter serve a dual purpose: contributing to the knowledge of biodiversity conservation and enabling sustainable forest management through the informed analysis of land cover changes in forested areas.

Chapter five extends the research's purview to investigate the impact of fire incidents on forest cover within the Ernakulam area. Through the utilization of advanced AI and ML techniques, we analyze satellite data to identify areas affected by fire damage and assess the extent of this impact.

Detection of fire-affected regions is critical for understanding the ecological and environmental implications of fire events. The chapter delves into methodologies for fire detection in remote sensing data, discussing thermal bands, spectral indices, and normalized burn rate (NBR) as indicators. The results of this analysis not only contribute to fire prevention strategies but also inform post-fire ecological restoration efforts.

By examining forest cover changes due to fire incidents, this chapter extends the scope of the research to disaster monitoring and mitigation, playing a significant role in understanding the resilience of ecosystems to disturbances.

The sixth chapter presents the culmination of this research project by designing a user-friendly interface for detecting LU/LC changes across the studied regions. This interface is developed to cater to a diverse set of users, including researchers, policymakers, and the general public, providing an intuitive platform for interactive exploration and analysis of the data.

The user interface incorporates various indices and visualization tools to facilitate efficient exploration of LU/LC changes. These indices may include the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and the Enhanced Vegetation Index (EVI), among others. Users can customize their analysis by choosing different indices based on their specific objectives.

The interface serves as a practical tool for informed decision-making, providing users with a visual representation of LU/LC changes and their underlying causes. It enables users to explore the spatial and temporal dynamics of land cover in the regions of Dehradun, Sikkim, and Ernakulam and also any non covered area by his geometry area selection tool available in Google Earth Engine.

Conclusive remarks are addressing in final chapter, In conclusion, this research is a multifaceted exploration into the estimation of land use/land cover changes using AI and ML techniques applied to satellite imagery available through Google Earth Engine. It encompasses a thorough examination of classification algorithms and datasets, the analysis of LU/LC changes in specific regions, including Dehradun and Sikkim, and the investigation of forest cover changes caused by fire incidents in Ernakulam. The development of a user-friendly interface adds a practical dimension to the research, facilitating the dissemination of information and insights to various stakeholders. The outcomes of this study have practical applications in regional planning, ecological conservation, disaster management, and informed decision-making, making it a valuable contribution to the field of geospatial analysis and land cover mapping.

Chapter 1

Introduction:

Natural resources are facing significant challenges and transformations on a global scale [1], including within the United States [2]. The development and intensive utilization of these resources are recognized as major drivers of environmental changes occurring on Earth [3]. Land, in particular, is a critical resource that significantly impacts the livelihood and well-being of people. One of the most prominent features related to land is its use and land cover, which reflects the physical characteristics of the land surface and its environmental attributes. Changes in land use and land cover are indicative of alterations in natural resources and the evolving trends associated with these changes [4]. Such changes have profound impacts on the functioning of socioeconomic and environmental systems, influencing factors like sustainability, food security, biodiversity, human vulnerability, and global ecosystems [5]. Consequently, the assessment of land use and land cover change is a crucial tool for studying global transformations on various spatial and temporal scales. The sustainable management of the Earth's surface [6], encompassing land use and land cover changes, remains a critical environmental challenge that humanity must quantify and address.

Traditional approaches to environmental management often fail to capture the full spectrum of benefits derived from natural resources or consider the diverse stakeholders who depend on these resources. An ecosystem-based approach to resource management seeks to identify the most effective strategies for optimizing the use of natural resources in a comprehensive and flexible manner [7]. Water is another vital resource for human existence, as well as for the well-being of animals, plants, and ecosystems [8]. Changes in water resources serve as significant indicators of environmental, meteorological, and anthropogenic interactions [9], [10]. The degradation of water resources can lead to increased poverty, insecurity, and the deterioration of biological diversity [11]. Accurate information on surface water quantity and distribution is essential for tasks like surface water mapping, quantifying water supplies for drinking and irrigation, evaluating land use and land cover changes [12], and monitoring environmental transformations [13]. Documenting changes in surface water dynamics also provides essential metrics for safeguarding the environment and its components. The intensification of water use over the past century and the first decades of the current century has resulted in severe water scarcity in numerous regions worldwide. Additionally, climate change has the potential to negatively impact the availability of resources, potentially leading to decreased environmental sustainability. However, long-term trends indicate that climate change [14], population growth, and rising demands for food, energy, and water are inextricably linked. Remote sensing imagery [15], a valuable data source, plays a significant role in assessing land cover changes and surface water dynamics. The wide array of remote sensing data, featuring various spatial and temporal coverage and resolutions [7], allows numerous researchers and professionals to address diverse issues across various fields and scales with a high degree of precision. These data sources facilitate the identification and analysis of complex problems, yielding confident results to comprehensively understand these issues, identify appropriate solutions, and make informed decisions. Remote sensing data, in conjunction with geographic information systems (GIS) [16] technologies, can aid stakeholders in mapping regions undergoing changes, comprehending development patterns and seasonal land transformations [17], and evaluating existing management activities and policies. Remote sensing methodologies are particularly advantageous for cross-border studies [18] where logistical constraints may be present and for investigations that focus on projecting future changes. Furthermore, remote sensing data sources provide cost-effective and readily accessible data for the development of land use maps. These maps are instrumental in calculating trends related to water use and urbanization, thereby enabling more effective resource management within targeted areas and facilitating sustainable growth and economic development [19]. The transition from dispersed populations to densely inhabited communities dominated by non-agricultural economic activities is commonly referred to as urbanization. Moreover, satellite imagery [20]facilitates the assessment of temporal changes[21] and leverages plant phenological attributes to distinguish between various vegetation types. This research project, which utilizes the capabilities of remote sensing and GIS technologies, addresses environmental challenges[22] and reflects the true nature of these challenges, ultimately leading to informed decision-making. The research focuses on four distinct study areas, each characterized by varying climatological conditions. Various techniques and temporal resolutions[23] are applied to data collected by Landsat and Sentinel satellites. The study also evaluates the performance of these satellite data and machine learning algorithms within the selected study regions. Of the four study areas, two regions, Doon Valley and Sikkim, share several commonalities. They are both situated in Himalayan valleys, experiencing semi-arid to arid conditions, and confronting similar climatic challenges despite having different surface characteristics and water resources. The other two regions, Lucknow Region and Ernakulam, are characterized by intensive urbanization, which affects land use and land cover features, causing similar changes in urban areas, surrounding agricultural lands, and disruptions to ecosystem services. Urban growth patterns in these two regions generally involve horizontal expansion with fewer tall buildings, as shown by previous research. Both regions face significant water scarcity issues and increasing demands for water in both human and environmental sectors. For instance, horizontal expansion often leads to higher per capita rates of outdoor water consumption, while high-density urban development patterns tend to result in lower outdoor water use. Moreover, climate change influences urban water demand, with rising temperatures expected to increase water requirements. Water scarcity is further exacerbated by rapid urbanization and changing climate patterns.

In addition to water quantity, water quality degradation is a pressing concern, which increasingly restricts water usage. The following sections provide detailed information on the primary study regions, including their geographic locations, water resources, climates, populations, vegetation, and land use and land cover patterns.

1.1 : Doon Valley (Dehradun Region)

In the Doon Valley of the Himalayan foothills, the study area is located between a tributary of the Ganga, the Song River, on the east and, a tributary of the Yamuna, the Asan River on the west. With the Main Boundary Thrust (MBT) to the north, the Himalayan Frontal Thrust (HFT)[24] to the south, the Yamuna Tear Fault (YTF) to the west, and the Ganga Tear Fault (GTF) to the east [25] . Dehra Dun is surrounded on all sides by significant faults [26], [4]The city is admittance to the surrounding area and is renowned for its picturesque surroundings and benevolent climate. It is vulnerable to a variety of natural hazards due to its geomorphologic and climatic features. Earthquakes, landslides, cloudbursts, and flash floods frequently cause devastation in the area. The city is located at a height of 2100 feet above sea level at coordinates 30.3165° N and 78.0322° E . The study area experiences a generally temperate climate. [24]Summers are comfortable in hilly areas, but in the Doon Valley, the heat is frequently oppressive. Even at low elevations, such as Dehradun, where the higher peaks are covered in snow throughout the winter, the temperature drops below freezing [27].

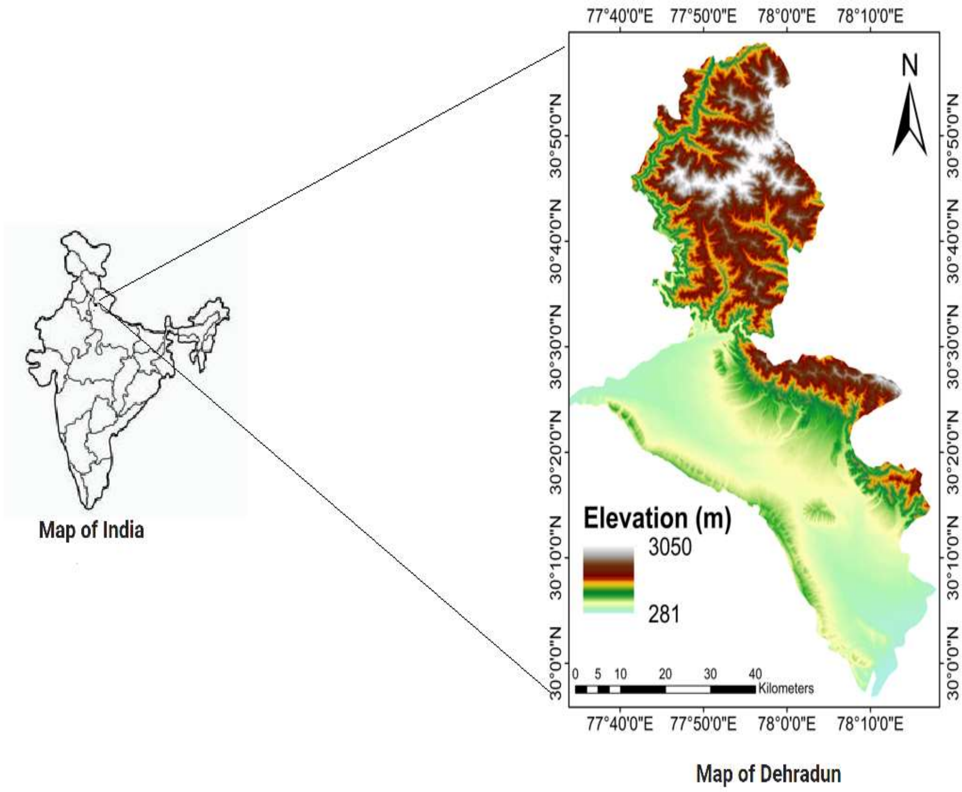

Figure 1.1 Geographical Location of Study Area_Dehradun

1.1.1: Climate and climate change

The study area experiences a generally temperate climate. [24]Summers are comfortable in hilly areas, but in the Doon Valley, the heat is frequently oppressive. Even at low elevations, such as Dehradun, where the higher peaks are covered in snow throughout the winter, the temperature drops below freezing [27]. The monsoon arrives around the middle of June during the summer, which begins in March. The hottest months are often May and the first half of June when mean temperatures might reach 40°C. The wintertime average temperature at Dehradun was 5°C.

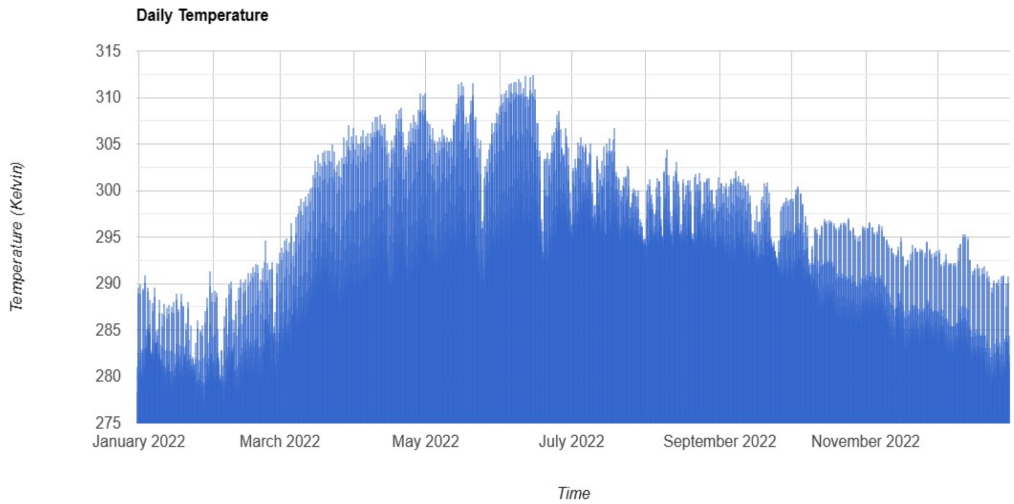

Figure 1.2 : Daily Temperature over the Study Area_Dehradun in Kelvin

The district has 2073.3 mm of rain each year on average. The months of July and August are the wettest, while the majority of the rainfall falls between the months of June and September. From June to September, there is an average yearly rainfall of about 87% of total year. The sub-montane soil in the area ranges in texture from sandy loam to silty clay loam and sand, and its depth ranges from very shallow (0–8 cm) to deep (4–9 cm). The soil can be seen all over the steep slope, river bank[28], and low elevated places. It is often brown to greyish brown and dark grey in colour, has a high percentage of organic matter, and is coarse in texture.

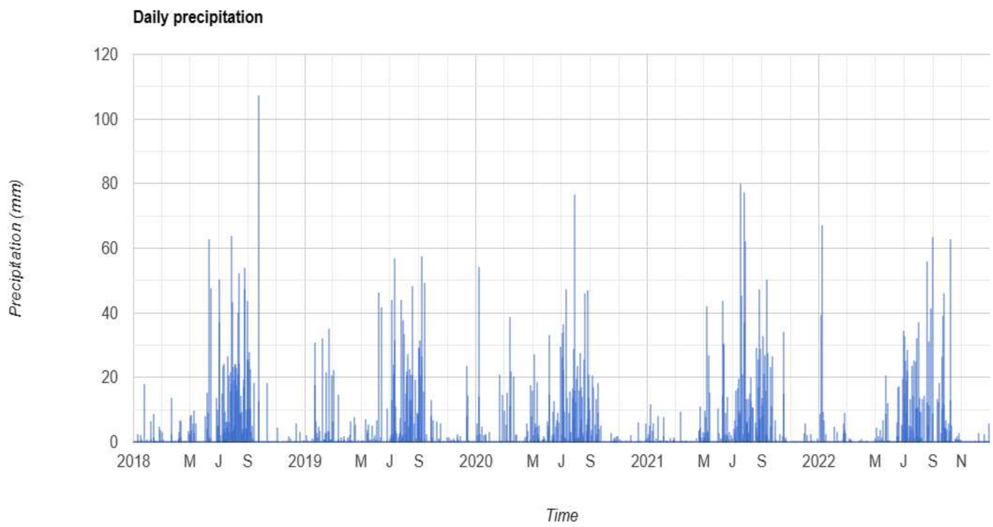

Figure 1.3: Monthly rain level over the Study Area_Dehradun in mm

1.1.2: Water Availability and Sustainability

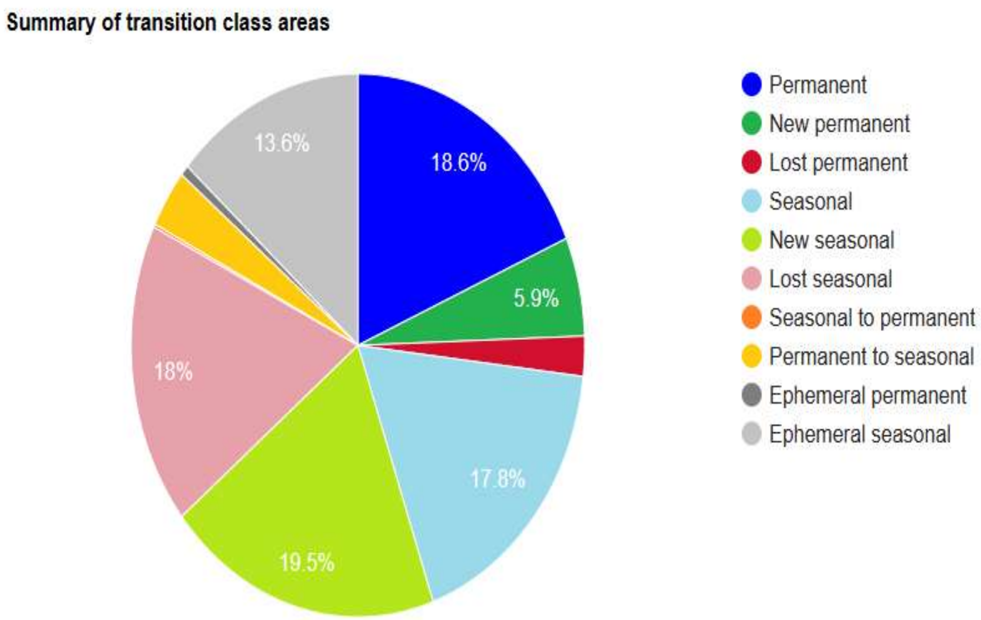

Figure 1.4: Percentage of Water in Study area _Dehrdun

1.2 : Eastern Himalayan Region (East Sikkim) :

Sikkim is one of the largest forest heaps in the northeast of India. Sikkim state covers a total area of 7096 sq Km. The geographical area we have targeted as the study area is the East Sikkim region, which is approximately 964 sq. km in size and located at 27.3084° N, 88.6724° E in figure 1.5

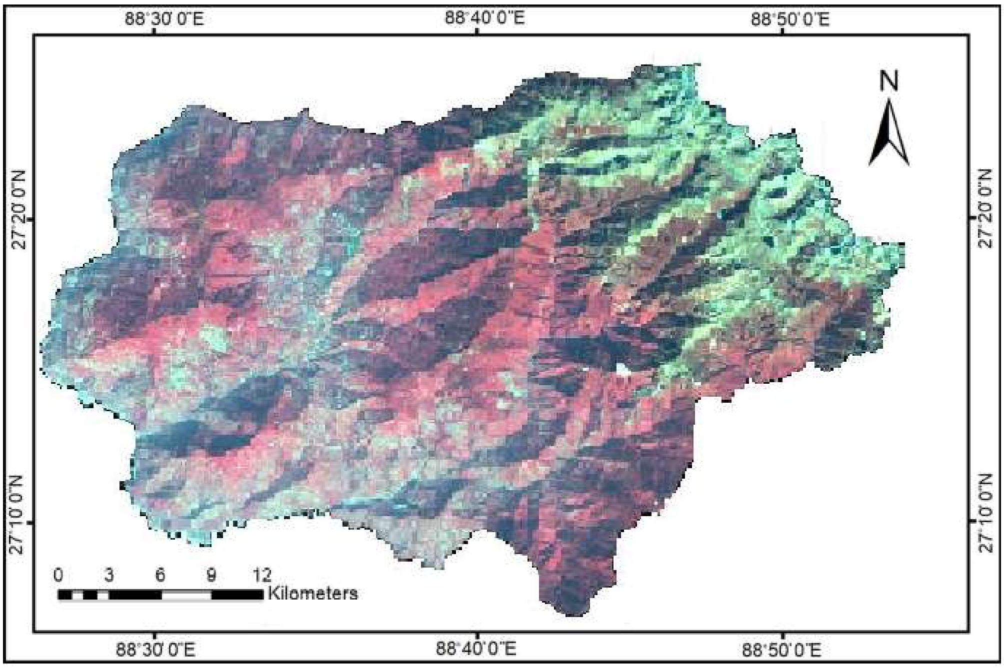

Figure 1.5: Geographical Location of Study Area_East Sikkim

Sikkim is known for its significant forest cover, and a large part of the state's total land area is under forested regions. The exact forest area of Sikkim can fluctuate over time due to factors like deforestation [29], afforestation[30] , and natural changes[31]. Therefore, it's essential to refer to the latest data from official sources for the most accurate and up-to-date information. According to the State of Forest Report (SFR) 2019, published by the Forest Survey of India (FSI) [32], the forest cover in Sikkim was approximately 3,221 square kilometers, which accounted for around 47% of the state's geographical area at that time. The state government of Sikkim has been proactive in promoting sustainable forestry practices and conservation efforts, which have contributed to maintaining a substantial forest cover in the region

1.2.1 Climate and climate change:

East Sikkim experiences a temperate climate with distinct seasonal variations throughout the year. The climate is influenced by its location in the eastern Himalayas and its varying elevations, ranging from lower subtropical regions to higher alpine areas. Here's a general overview of the climate conditions in East Sikkim: Summer (March to June):

During the summer months, East Sikkim experiences a pleasant climate with mild to moderately warm temperatures[33]. The temperature in the lower elevations, including Gangtok, the capital of Sikkim, can range from 15°C to 25°C (59°F to 77°F). As you go higher into the mountains, temperatures decrease, providing a refreshing escape from the summer heat.

Monsoon (July to September): The monsoon season[34] brings heavy rainfall to the region. The southern and eastern parts of East Sikkim receive more rainfall compared to the northern areas. Monsoon showers are essential for the lush greenery and the state's vibrant ecosystem. However, heavy rainfall may lead to landslides and disrupt transportation in some areas.

Autumn (October to November): After the monsoon season, East Sikkim experiences a relatively drier and milder period. The skies are clearer, and the temperatures begin to drop gradually. This is an excellent time for travelers to visit, as the weather is pleasant, and the landscapes are vibrant.

Winter (December to February): Winters in East Sikkim are cold, especially in the higher elevations. The temperature can drop to freezing levels, particularly in places like Nathu La and Tsomgo Lake. The lower regions, including Gangtok, experience cool temperatures, with the mercury often hovering around 4°C to 10°C (39°F to 50°F).

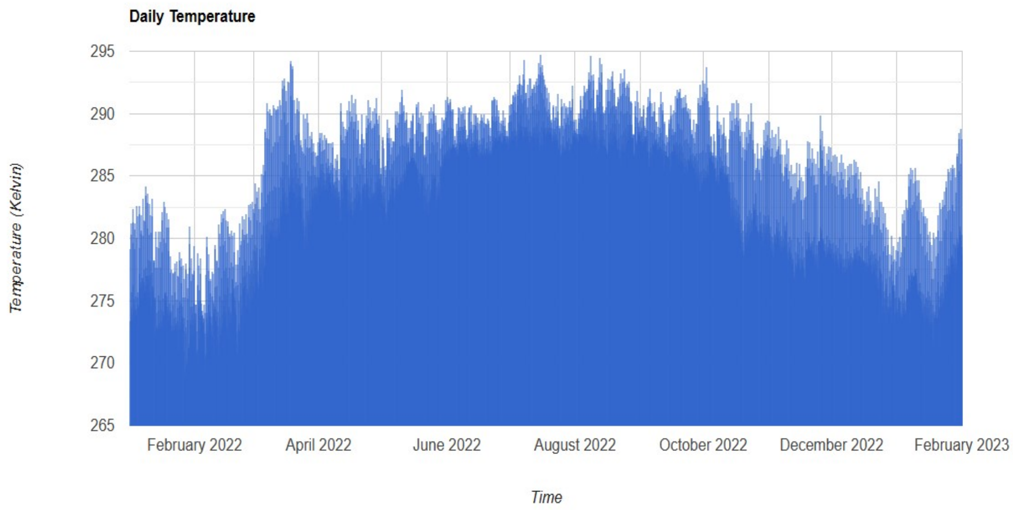

Figure 1.6 Daily Temperature over the study area East Sikkim in Kelvin in year 2023

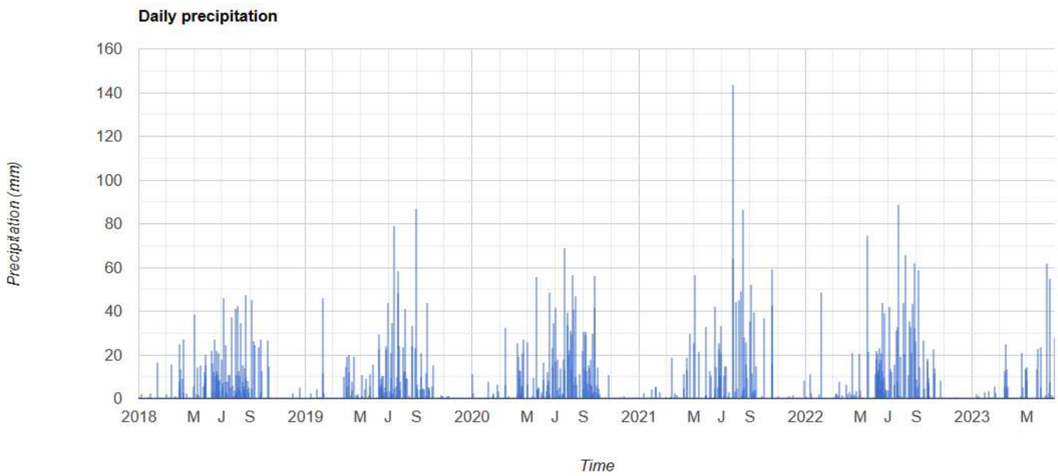

Figure 1.7 Monthly rain level over the study area East Sikkim in mm between year (2018-2023)

1.2.2 Water Availability and Sustainability:

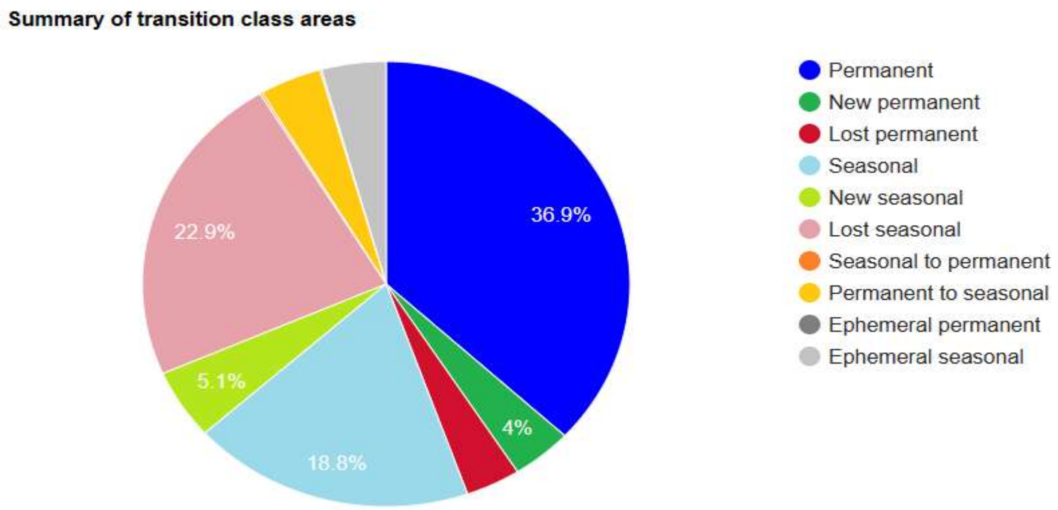

Figure 1.8 Percentage of Water in Study area _East Sikkim

1.3: Urban Area (Lucknow)

The study area has a surface area of nearly 2528 sq-km and is situated at 26°51'N and 80°56'E is above 123m mean sea level. It is India's eleventh most populous city. Figure 1.8 shows the geographical location of the study area.

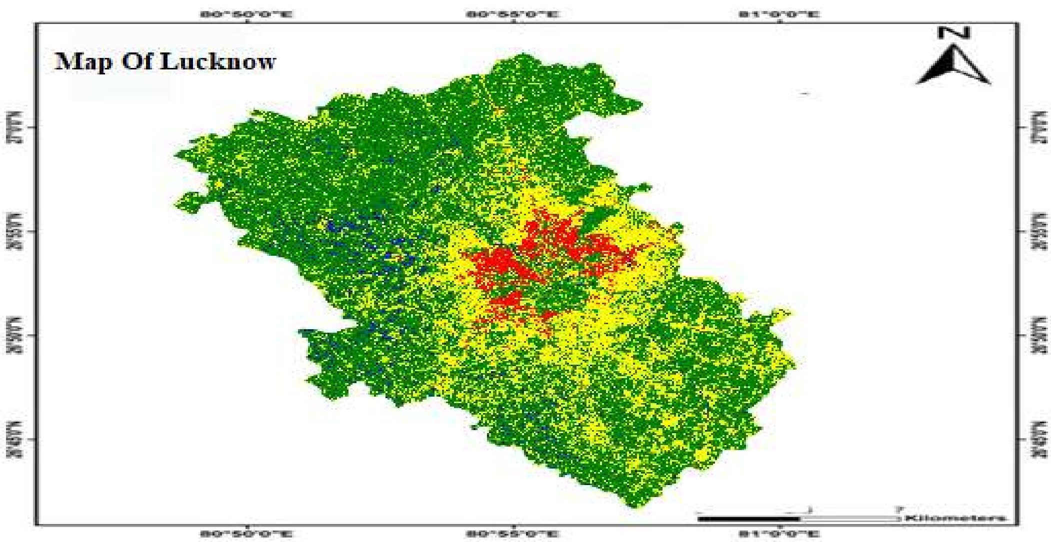

Figure 1.8 Geographical Location of Study Area_Lucknow

1.3.1 Climate and climate change:

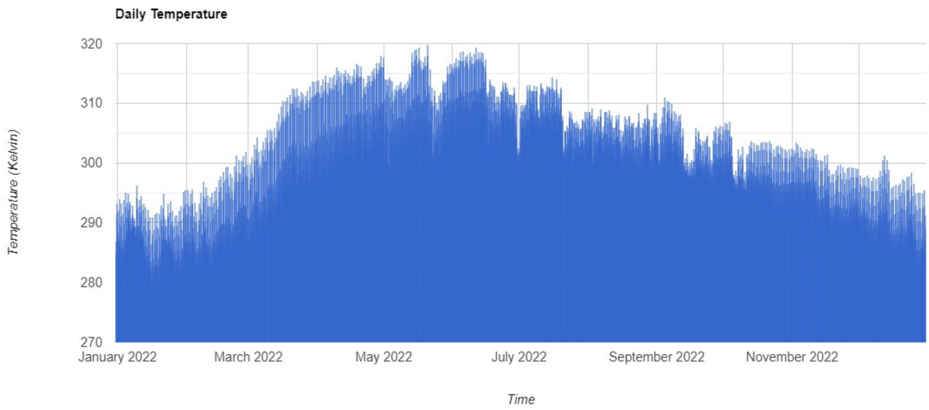

Figure 1.10 Daily Temperature over the study area_Lucknow in Kelvin

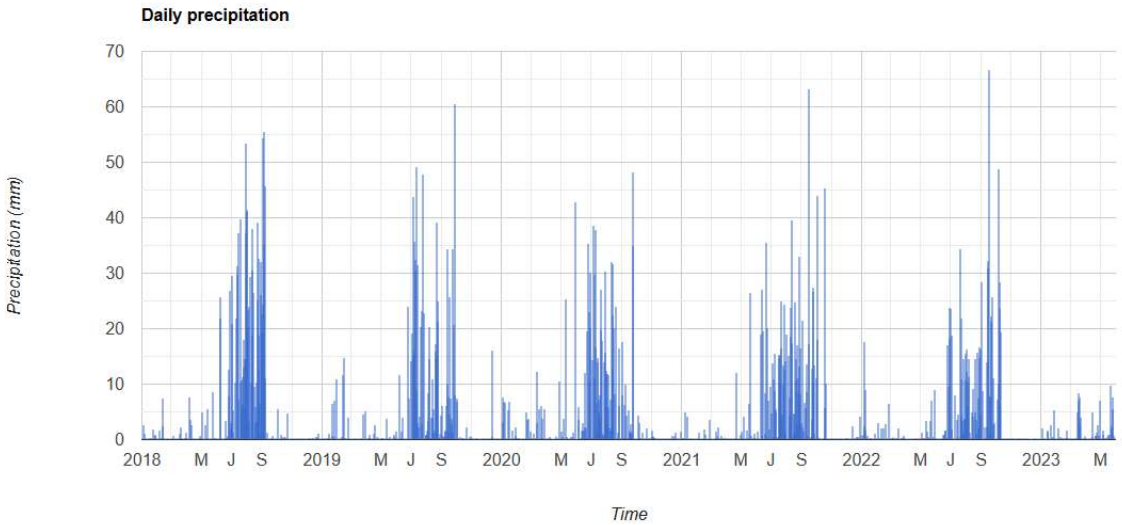

Figure 1.11: Monthly rain level over the study area_Lucknow in mm

Study area experiences a humid subtropical climate with four distinct seasons. Summers are characterized by scorching heat, often reaching temperatures exceeding 40°C. The monsoon season provides relief from the heat with heavy downpours, leading to cooler temperatures. Autumn in Lucknow is a pleasant season with mild temperatures, perfect for outdoor activities. Winters are cool and dry, typically ranging from 8°C to 20°C, offering a respite from the extremes of summer.Rainfall in Lucknow is a crucial aspect of its climate. The city receives the majority of its annual precipitation during the monsoon season, which spans from July to September. This period witnesses heavy rainfall, occasional flooding, and cooler weather. Throughout the rest of the year, Lucknow experiences relatively low rainfall, particularly during the winter months. On average, the city receives around 1000-1100 millimeters of rain annually, with significant variations in rainfall intensity between the seasons.

1.3.2 Water Availability and Sustainability:

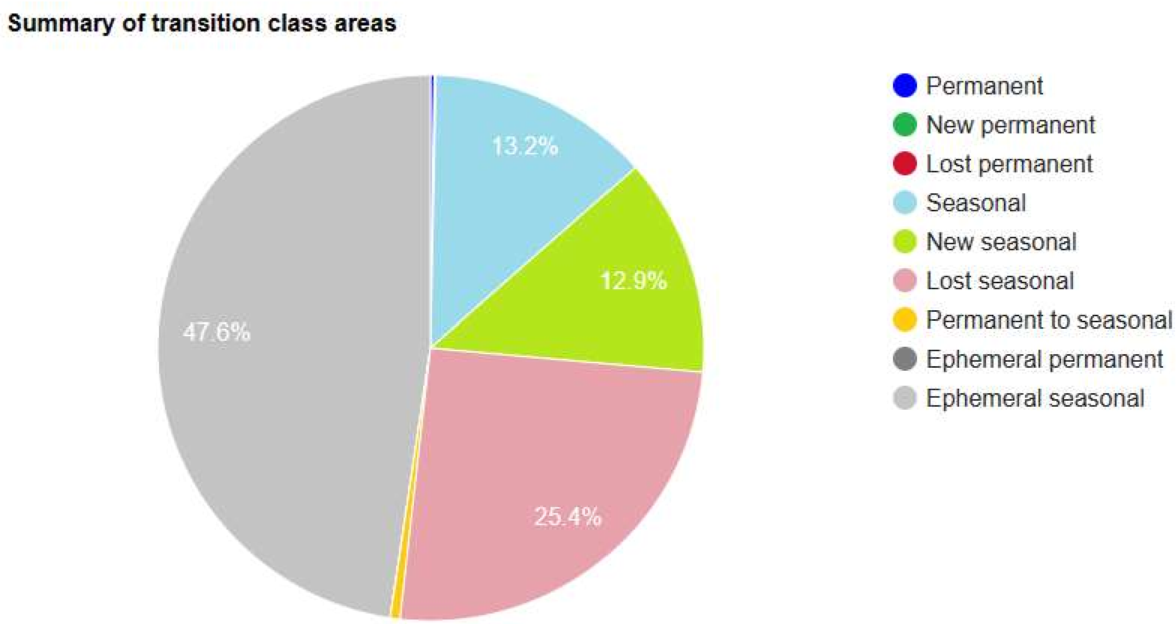

Figure 1.12 Percentage of Water in Study area_Lucknow

Study area faces challenges related to permanent water levels due to its geographic and hydrological characteristics. The city is situated in a region with a high water table, which means that the permanent water level is relatively shallow in many areas. This shallow water table can have both positive and negative implications. On the one hand, it can make groundwater more accessible for various uses, including domestic, agricultural, and industrial. However, it also increases the risk of waterlogging, especially during the monsoon season, which can lead to infrastructure and sanitation issues in certain parts of the city.

1.4 Motivation and Rationale for This Study:

Accurate estimation of changes in land use/land cover (LU/LC) mapping is of paramount importance for a wide range of applications, including environmental monitoring [23], urban planning, agriculture, and natural resource management. Traditional methods for LU/LC mapping often rely on manual interpretation of satellite imagery or coarse resolution data, leading to limitations in accuracy, consistency, and efficiency. Therefore, there is a growing need to leverage advanced technologies, such as AI/ML techniques, to improve the accuracy and efficiency of LU/LC mapping and change detection[35].

The utilization of AI/ML techniques offers several advantages in the context of LU/LC mapping. Firstly, these techniques have the ability to learn complex patterns and relationships from large-scale datasets, enabling the development of more accurate and robust models for land cover classification. AI/ML algorithms can effectively process and analyze high-resolution satellite imagery, extracting valuable features and information that are often difficult to capture through traditional methods.

Google Earth Engine, a cloud-based platform for geospatial analysis, provides access to an extensive collection of satellite imagery and geospatial data, that's make it an ideal platform for conducting large-scale LU/LC mapping and change detection studies. By combining AI/ML techniques with the computational power and data resources of Google

Earth Engine, researchers can harness the potential of these technologies for more efficient and scalable analysis.

Additionally, the classification of digital signatures derived from satellite imagery plays a crucial role in accurate land cover mapping. AI/ML techniques can be leveraged to effectively classify digital signatures, taking into account the complexity and variability of land cover patterns. By developing models that can accurately differentiate between different land cover types, researchers can improve the overall accuracy and reliability of LU/LC mapping and change detection.

The motivation for this research stems from the need to address the limitations of traditional methods, enhance the accuracy and efficiency of LU/LC mapping, and provide valuable insights for decision-making processes. By exploring the potential of AI/ML techniques and Google Earth Engine, this research aims to advance the field of LU/LC mapping and change detection, enabling better understanding and management of our changing environment.

1.5 Problem Statement:

Urban expansion in many parts of the India is a significant concern [36]. In the second decade of the twenty-first century, urban sprawl is consistently defined as a chaotic shift in the spatial structure of suburban communes that occurred because of the deepening of suburbanization, with little control over these processes by spatial policy . This sprawl takes crucial areas from the agricultural lands around the district[37]. The urban sprawl also takes important areas from the native ecosystems[38]. The urban sprawl also encroached into the Himalayan region, with possible devastating environmental impacts on marine resources [1] The native ecosystems lost substantial areas during agricultural and urban lands expansion. These changes in the Himalayan region cause great destruction to the fragile environment of the district and the surrounding areas and exacerbate desertification indications[1] Urban sprawl is also a real problem in the non Himalayan Region also like we have studied over Lucknow and Kocchi. Rapid urban growth and climate change in the study areas have increased water resource demand. The predominant upland mixed vegetation land cover category has steadily declined, giving up land to urban and agricultural development. The urban sprawl negatively affects the area by losing native vegetation and agricultural lands. It also causes environmental deterioration like flood and fire occurrences also. The accurate estimation of changes in land use/land cover (LU/LC) mapping is crucial for effective environmental monitoring and land management. However, existing methods for LU/LC mapping and change detection often suffer from limitations in accuracy and efficiency.

This thesis addresses several critical challenges in the field of remote sensing and geospatial analysis for environmental monitoring and decision-making. The research focuses on the optimization of land cover and land use classification and change detection methods across various study areas, including Doon Valley (Dehradun), Lucknow, and Sikkim.

1. **Data and Algorithm Selection:** The first challenge involves the comparison and selection of the most suitable datasets and machine learning algorithms for these study areas. This includes determining which combination of data sources, such as Landsat and Sentinel, and machine learning algorithms, including CART, Random Forest, Gradient Boosting, and SVM, yield the most accurate and reliable results for land cover and land use classification.

2. **Temporal Change Analysis:** The research extends to a detailed temporal analysis of land cover changes in Dehradun between the years 2018 and 2022. It aims to quantify and understand how the landscape has evolved during this period, providing insights into urban expansion, land use alterations, and environmental changes.

3. **Long-Term and Short-Term Change Assessment:** To gain a comprehensive understanding of environmental dynamics, the study also delves into long-term change analysis in Sikkim. This investigation aims to identify and quantify gradual alterations, such as natural deforestation, over an extended period. Simultaneously, the research includes short-term change computation in the Ernakulam area, specifically focusing on changes resulting from natural deforestation and fire incidents. This dual approach offers insights into both gradual and abrupt changes in the landscape.

4. **User Interface Design:** In recognition of the need for accessibility and usability in remote sensing and geospatial analysis, the research endeavors to design a user-friendly interface. This interface will facilitate the computation of changes in land cover and land use using various indices and bands, empowering a broader range of users, from researchers to decision-makers, to leverage these valuable insights for environmental monitoring and management.

By addressing these challenges, the thesis aims to contribute to the advancement of remote sensing and geospatial analysis methodologies and provide practical tools for environmental assessment and decision support.

1.6: Research Questions and Hypotheses

1.6.1 Research Questions :

The following research questions (RQs) will be addressed to identify, quantify, and measure the urban growth and the land use/ land cover change:

1. Data and Algorithm Selection:

- "Which combination of datasets, such as Landsat and Sentinel, and machine learning algorithms (CART, Random Forest, Gradient Boosting, SVM) yields the most accurate and reliable results for land cover and land use classification in different study areas, including Doon Valley, Lucknow, and Sikkim?"

2. Temporal Change Analysis in Dehradun (2018-2022):

- "How has the land cover in Dehradun evolved between 2018 and 2022, and what are the key drivers of change, including urban expansion, land use alterations, and environmental shifts?"

3. Long-Term Change Analysis in Sikkim:

- "What are the long-term patterns of change, particularly natural deforestation, in Sikkim, and how have these changes evolved over an extended period?"

4. Short-Term Change Computation in Ernakulam (Fire Incidents and Natural Deforestation):

- "How can short-term changes in the Ernakulam area be quantified, distinguishing between alterations caused by natural deforestation and those resulting from fire incidents, and what insights can be derived from these assessments?"

5. User Interface Design:

- "What design principles and functionalities are essential in developing a user-friendly interface for computing changes in land cover and land use using various indices and bands, and how can such an interface enhance accessibility and usability in remote sensing and geospatial analysis for environmental monitoring and decision support?"

1.6.2 Hypotheses

Based on results from prior research covered in the literature review presented in chapter one that covers the introduction of this research, thesis proposed the following hypotheses (Hs), which correspond with each of the research questions outlined immediately above:

1. Data and Algorithm Selection:

- **Hypothesis:** "Certain combinations of datasets and machine learning algorithms will demonstrate superior performance for land cover and land use classification in different study areas. The choice of dataset and algorithm will significantly impact classification accuracy and efficiency."

2. Temporal Change Analysis in Dehradun (2018-2022):

- **Hypothesis:** "The analysis of land cover changes in Dehradun between 2018 and 2022 will reveal significant alterations, with urban expansion, land use shifts, and environmental changes being key drivers. Understanding the temporal patterns of change is pivotal for informed decision-making."

3. Long-Term Change Analysis in Sikkim:

- **Hypothesis:** "Long-term analysis in Sikkim will uncover patterns of natural deforestation and other gradual changes. Environmental and climatic factors are likely to influence these long-term alterations, offering insights for conservation and management strategies."

4. Short-Term Change Computation in Ernakulam (Fire Incidents and Natural Deforestation):

- **Hypothesis:** "Short-term change computations in the Ernakulam area will distinguish between changes resulting from natural deforestation and those induced by fire incidents. It is expected that fire incidents will lead to rapid and abrupt changes, while natural deforestation will exhibit slower alterations."

5. User Interface Design:

- **Hypothesis:** "The development of a user-friendly interface for computing changes in land cover and land use using various indices and bands will enhance accessibility and usability in remote sensing and geospatial analysis. Such an interface is anticipated to accommodate users with varying levels of expertise, facilitating data analysis for research and decision-making."

1.7 The Objectives

The objective of this research is to develop an AI/ML-based approach for estimating changes in LU/LC mapping using Google Earth Engine. The specific problem areas to address include:

1. Accuracy of LU/LC Mapping: Current methods for LU/LC mapping often rely on manual interpretation or coarse resolution satellite imagery, leading to inaccurate and inconsistent results. There is a need to leverage AI/ML techniques to improve the accuracy and reliability of LU/LC mapping.

2. Change Detection: Identifying changes in land use and land cover over time is critical for understanding the dynamics of ecosystems and assessing the impact of human activities. However, existing change detection algorithms lack robustness and often struggle to distinguish between different types of land cover changes. Developing advanced AI/ML techniques to enhance change detection accuracy is essential.

3. Identification of Digital Signatures: Digital signatures, such as spectral reflectance patterns derived from satellite imagery, provide valuable information for land cover classification. However, the classification of digital signatures is challenging due to the complexity and variability of land cover patterns, as well as the presence of cloud cover, noise and artifacts in the data. Developing user interface that can effectively classify digital signatures and handle these challenges is a key objective.

4. Efficiency and Scalability: Processing large-scale satellite imagery datasets requires efficient and scalable algorithms and infrastructure. Developing AI/ML techniques that can leverage the computational power of Google Earth Engine and efficiently process large volumes of data is necessary for practical implementation.

1.8 Thesis Outline:

Chapter 2 explains the basics of different satellite data, data processing techniques, and techniques to integrate the available satellite data with urban application modeling. The chapter continues with simple understanding of implementation of different satellite data with number of machine learning algorithm and maps them for finding better data sets and machine learning algorithm over diverse geospatial study area .

Chapter 3 describes computation of change detection in land use land cover over different time zone for understanding the dynamics of ecosystem and assessing the impact of human activities. Chapter continues with computation of land use land cover over different Himalayan, foothills of Himalaya "Doon Valley" for finding different dynamics of ecosystem and growth in urbanization.

Chapter 4 In this chapter, we dissect the unique challenges posed by forested areas, such as dense canopies and complex ecosystems. Techniques for forest change detection are explored, including the use of multi-spectral and radar imagery for improved forest cover mapping. The chapter also dives into the identification of deforestation and afforestation events, which are vital in understanding the ecological health of the Sikkim forests. The findings of this chapter serve a dual purpose: contributing to the knowledge of biodiversity conservation and enabling sustainable forest management through the informed analysis of land cover changes in forested

Chapter 5 perform study over Ernakulam fire incident and compute forest area changes due incident, this study delves into the specific case of the Ernakulam fire incident, employing a multifaceted approach to detect short-term forest cover changes caused by fires. Leveraging Sentinel-2 satellite data, spectral indices, and machine learning algorithms, this research seeks to offer a comprehensive assessment of the impact of the Ernakulam fire on local forest ecosystems. By combining the power of remote sensing technologies and advanced analytical methods, the study not only sheds light on the dynamic nature of forest cover changes in the aftermath of fire incidents but also aims to provide invaluable insights for swift response and effective management of such disturbances in the Ernakulam region. The research contributes to the broader field by showcasing a case study that underscores the practical applications of cutting-edge technology in environmental monitoring and disaster response, highlighting the importance of addressing short-term changes in forest cover to mitigate the impact of fires on natural landscapes.

Chapter 6 delves into the multifaceted task of developing a user interface tailored for the detection of changes over various multispectral bands and indices. As remote sensing technology continues to play a pivotal role in applications ranging from land cover analysis to disaster management, the need for accessible and efficient tools becomes increasingly apparent. This study addresses the critical intersection of spectral data, indices, and user interface design, aiming to enhance the accessibility and usability of change detection for both experts and non-experts. By laying a strong theoretical foundation and employing robust methodologies, this research seeks to contribute to the broader field of remote sensing, with implications for resource management, disaster response, and environmental conservation. Through a well-structured methodology, the study examines the creation of the user interface, encompassing data integration, visualization, and interaction aspects, with a focus on facilitating the analysis of multispectral data and their derived indices.

Chapter 7, the main contributions of this thesis work are summarized in this chapter. The potential future implications of these studies along with the future research scope are also highlighted in this chapter.

Chapter 2:

Performance Assessment of Sentinel 2 MSI and Landsat OLI Data for Land Use Land Cover Classification across Diverse Geospatial Study Area Using a Comparison between Supervised Machine Learning Classifiers

2.1 Abstract :

Land cover change detection plays a crucial role in monitoring and understanding the dynamic transformation of the Earth's surface. This study aims to identify the best-suited dataset and machine learning algorithm for accurate and efficient land cover change detection. Three satellite image datasets, namely Landsat, Sentinel, and LiDAR-derived data, were utilized to compare their performance. Additionally, four popular machine learning algorithms— Classification and Regression Tree (CART), Random Forest, Gradient Tree Boost (GTB), and Support Vector Machine (SVM) were implemented to analyze their effectiveness in detecting land cover changes. The study involved preprocessing the datasets to enhance their quality and extracting relevant features for input to the machine learning models. Training and testing data were carefully selected to represent diverse land cover change scenarios. Evaluation metrics such as accuracy, precision, recall, and F1-score were used to quantify the performance of each model. Results demonstrated that the combination of optical imagery and Random Forest algorithm exhibited the highest accuracy in land cover change detection, achieving an average accuracy of over 90%. CART and GTB also showed promising results but required more extensive training data and computational resources. Landsat, although valuable in

certain scenarios, exhibited lower accuracy in general due to its sensitivity to environmental conditions. In conclusion, this study suggests that using sentinel data sets in combination with the Random Forest algorithm is a robust and efficient approach for detecting land cover changes in areas where forest and vegetation region are more and landsat data sets in combination with the Gradient Tree Boost perform well where urban areas are dense . The findings will aid in informed decision-making for land management, environmental conservation, and urban planning applications, ultimately contributing to sustainable development and better land use practices. However, further research is recommended to explore the integration of multi-sensor data and advanced deep learning architectures to improve accuracy and adaptability in diverse landscapes.

2.2 Introduction:

Satellite data [39] refers to the information and imagery collected by satellites orbiting the Earth. These satellites are equipped with sensors and instruments that capture various types of data about the Earth's surface, atmosphere, and other environmental parameters. Satellite data is widely used for scientific research, environmental monitoring, weather forecasting, disaster management, urban planning, and many other applications.

2.2.1 Types of Data:

1. Optical Imagery: Optical sensors [40] capture visible and near-infrared light reflected from the Earth's surface. This type of data provides high-resolution images and is useful for applications such as land cover mapping, vegetation analysis, and urban development monitoring.

2. Radar Imagery: Radar sensors emit microwave signals and measure the reflected signals, allowing for the capture of data in all weather conditions. Radar data is particularly useful for applications like topographic mapping, monitoring sea ice, detecting changes in land surfaces, and assessing forest structure.

3. Thermal Imagery: Thermal sensors measure the infrared radiation emitted by objects on the Earth's surface. Thermal imagery is valuable for monitoring temperature variations, identifying hotspots, studying urban heat islands, and detecting fires.

4. LiDAR Data: LiDAR (Light Detection and Ranging)[41], [42] sensors use laser pulses to measure the distance between the satellite and the Earth's surface, creating highly accurate elevation models and 3D point clouds. LiDAR data is utilized for applications such as terrain mapping, flood modelling, urban planning, and forestry.

2.2.2 Data Processing and Analysis:

1. Pre-processing: Satellite data often requires pre-processing to correct for sensor-specific distortions, such as atmospheric effects or sensor noise. Pre-processing steps may include radiometric and geometric corrections, calibration, and image enhancement.

2. Image Classification: Image classification techniques involve categorizing pixels or objects within satellite images into different classes based on their spectral properties. This allows for land cover classification, identification of specific features, and change detection analysis.

3. Change Detection: Change detection techniques compare satellite images captured at different times to identify and analyze changes that have occurred on the Earth's surface.

This is useful for monitoring urban growth, deforestation, natural disasters, and environmental changes.

2.2.3 Integration and Analysis :

1. Data Fusion: Data fusion techniques[43], [44] combine multiple sources of satellite data, such as optical imagery and radar data, to create a more comprehensive and accurate representation of the Earth's surface. Data fusion enables more detailed analysis and interpretation of complex phenomena.

2. Spatial Analysis: Satellite data can be integrated with other geospatial data, such as geographic information system (GIS) data, to perform spatial analysis. This involves overlaying and analyzing different data layers to understand spatial relationships, patterns, and trends.

3. Machine Learning and Artificial Intelligence: Satellite data can be leveraged with machine learning algorithms and artificial intelligence techniques to extract patterns, make predictions, and automate analysis tasks. This allows for efficient processing of large volumes of data and discovery of hidden insights.

2.3 Satellite Data:

Satellite data provides a valuable source of information for understanding and monitoring our planet. By processing, analyzing, and integrating satellite data, scientists, researchers, and decision-makers gain valuable insights into various environmental processes, enabling informed decision-making and sustainable development. Satellite data plays a crucial role in monitoring land use and land cover changes over time. Here are some key uses of satellite data in this context:

1. Land Cover Mapping: Satellite imagery is used to create detailed land cover maps that classify different types of land cover, such as forests, croplands, urban areas, water bodies, and barren lands. These maps provide baseline information about the distribution and extent of land cover classes.

2. Change Detection: By comparing satellite images captured at different time points, land cover changes can be detected and quantified. Change detection techniques analyze the differences between images to identify areas that have undergone changes, such as deforestation, urban expansion, agricultural expansion, or natural disasters.

3. Urban Growth Monitoring: Satellite data enables the monitoring and analysis of urban growth and expansion. It helps track changes in urban areas, identify patterns of urbanization, assess the rate of urban growth, and monitor encroachments into natural or protected areas.

4. Deforestation Monitoring: Satellite imagery is widely used to monitor deforestation, especially in remote or inaccessible areas. By comparing images from different time periods, changes in forest cover can be detected, and deforestation rates can be estimated. This information is vital for conservation efforts, forest management, and assessing the impacts of deforestation on ecosystems and climate change.

5. Agricultural Monitoring: Satellite data helps in monitoring agricultural land use and crop dynamics. It can provide information on crop types, health, yield estimation, and changes in agricultural practices. This data supports agricultural planning, resource management, and early warning systems for crop diseases and pests.

6. Ecosystem Monitoring: Satellite data aids in monitoring changes in natural ecosystems, including wetlands, grasslands, and biodiversity hotspots. It helps assess habitat loss, land degradation, and impacts on ecological processes. This information is critical for conservation planning, ecological restoration, and assessing the effectiveness of protected areas.

7. Land Use Planning: Satellite data assists in land use planning and decision-making processes. It provides information on land suitability, land capability, and potential areas for development or conservation. Satellite data supports sustainable land management practices, urban planning, and infrastructure development.

8. Climate Change Analysis: Satellite data helps in studying the impacts of climate change on land cover and land use patterns. It enables the assessment of changes in glaciers, permafrost, coastal zones, and other vulnerable areas affected by climate-related processes. This data aids in understanding the drivers and consequences of climate change and supports adaptation and mitigation strategies.

By leveraging satellite data, land use and land cover changes can be systematically monitored, providing valuable insights into environmental dynamics, informing policy decisions, and promoting sustainable land management practices.

Landsat and Sentinel satellite missions are two prominent Earth observation programs that provide valuable data for a wide range of applications. While both missions aim to monitor the Earth's surface, there are differences in terms of their sensors, spatial and spectral resolutions, revisit times, data availability, and mission objectives. Here's a comparison of Landsat and Sentinel satellite data:

2.3.1 Landsat Data Set:

The Landsat data set is a collection of satellite imagery captured by the Landsat series of satellites. It is one of the longest and most comprehensive Earth observation programs, providing a consistent record of the Earth's surface since the launch of the first Landsat satellite in 1972. The data set is widely used for various applications; including land cover mapping, environmental monitoring, agriculture, urban planning, and natural resource management.

Satellite Missions: The Landsat program consists of multiple satellite missions, including Landsat 1 to Landsat 9. Each mission represents an individual satellite launched at different times with improved technologies and capabilities.

Sensors: The Landsat satellites use different sensors to capture multispectral data. Landsat 7 and Landsat 8 employ the Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager (OLI) sensors, respectively. These sensors capture data in the visible, near-infrared, short-wave infrared, and thermal infrared spectral ranges.

Spatial Resolution: The spatial resolution of Landsat data is 30 meters for the visible, near-infrared, and short-wave infrared bands. The thermal infrared band has a resolution of 100 meters. This resolution allows for detailed analysis of land cover patterns, vegetation health, and changes in the Earth's surface.

Revisit Time: The Landsat satellites have a repeat cycle of approximately 16 days, meaning they revisit the same location on the Earth's surface once every 16 days. This revisit time allows for monitoring changes and capturing images of different seasons and conditions.

Data Availability: Landsat data is openly accessible and freely available to the public. The United States Geological Survey (USGS) is responsible for managing and distributing Landsat data. The USGS Earth Explorer or other data portals provide access to the Landsat archive, which includes a vast collection of images covering different regions and time periods.

Data Products: Landsat data is available in different levels and formats. Level-1 data consists of ortho-rectified and calibrated imagery, while Level-2 data includes additional atmospheric correction and surface reflectance products. These products facilitate accurate and consistent analysis of land cover, vegetation indices, and other applications.

Table 2.1 Bands, Wavelength, Spatial Resolution, Equatorial Crossing Time (E.C.T.), and the available date range for Landsat satellite

Satellite	Bands	Type	Wave Length(µm)	Resolution (m)	E.C.T	Date Range
Landsat-8 (OLI & TIRS)	SR_B1	Coastal	0.43-0.45	30m	10:00 A.M (16 Day)	11 April 2013 to Present
	SR_B2	Blue	0.45-0.51			
	SR_B3	Green	0.53-0.59			
	SR_B4	Red	0.64-0.67			
	SR_B5	NIR	0.85-0.87			
	SR_B6	SWIR-1	1.56-1.65			
	SR_B7	SWIR-2	2.107-2.294			
	SR_B8	Pan	0.50-0.67	15m		
	SR_B9	Cirrus	1.36.-1.384	30m		
	SR_B10	TIR-1	10.60-11.19	100m		
	SR_B11	TIR-2	11.50-12.91	100m		

Multi temporal Analysis: One of the significant strengths of the Landsat data set is its long-term record, allowing for multi temporal analysis and trend monitoring. The archive of Landsat images enables the study of changes in land cover, urban growth, deforestation, and other environmental phenomena over several decades.

Applications: Landsat data has a wide range of applications, including land cover mapping, forest monitoring, agriculture, water resource management, disaster assessment, and urban planning. The long history of Landsat imagery enables researchers, scientists, and policymakers to study long-term environmental trends and make informed decisions.

2.3.1 Sentinel DataSet:

The Sentinel data set is a collection of satellite imagery and other Earth observation data acquired by the Sentinel series of satellites. It is part of the European Union's Copernicus program, which aims to provide free and open access to environmental data for a wide range of applications. The Sentinel missions offer a diverse set of sensors and data products that contribute to monitoring and understanding various aspects of the Earth's environment.

Satellite Missions: The Sentinel program consists of several satellite missions, including Sentinel-1, Sentinel-2, Sentinel-3, and Sentinel-5. Each mission is designed to capture specific types of data and focuses on different aspects of Earth observation.

Sensors: The Sentinel satellites are equipped with different sensors to capture various types of data. For example, Sentinel-1 carries a synthetic aperture radar (SAR) sensor that provides all-weather, day-and-night imaging for applications such as land and ocean monitoring, disaster mapping, and ice monitoring. Sentinel-2 uses the Multi Spectral Instrument (MSI) sensor, which captures data in multiple spectral bands for applications like land cover classification, vegetation monitoring, and change detection. Other sensors in the Sentinel missions include the Ocean and Land Colour Instrument (OLCI), the Sea and Land Surface Temperature Radiometer (SLSTR), and the Tropospheric Monitoring Instrument (TROPOMI).

Spatial Resolution: The spatial resolution varies depending on the specific Sentinel mission and sensor used. For example, Sentinel-1 SAR data has different resolution modes ranging from 5 meters to 40 meters, while Sentinel-2 data has a spatial resolution of 10 meters for most visible, near-infrared, and short-wave infrared bands.

Revisit Time: The Sentinel missions are designed to provide frequent revisit times for monitoring purposes. For instance, Sentinel-2 satellites have a revisit time of 5 days at the equator, enabling more frequent imaging for time-sensitive applications.

Data Availability: The Sentinel data is freely available to the public as part of the Copernicus program's open data policy. The data can be accessed through the Copernicus Open Access Hub, the SciHub, or other data portals. The availability and accessibility of Sentinel data support widespread use in research, commercial applications, and policy-making.

Data Products: Sentinel data comes in different levels and formats, including Level-1, Level-2, and Level-3 products. Level-1 products consist of calibrated and geo referenced imagery, while Level-2 and Level-3 products involve additional processing steps such as atmospheric correction, surface reflectance, or higher-level data fusion.

Applications: The Sentinel data set serves a wide range of applications across various sectors, including environmental monitoring, agriculture, forestry, water resource management, disaster response, and urban planning. The diverse sensors and data products cater to different needs and enable users to analyze and understand the Earth's environment from multiple perspectives.

Table2.2 Bands, Wavelength, Spatial Resolution, Equatorial Crossing Time (E.C.T.), and the available date range for Sentinel-2 satellite

Satellite	Used Bands	Type	Wave Length(nm)	Resolution (m)	E.C.T	Date Range
Sentinel-2 (MSI)	*B1*	Aerosols	*443.9nm (S2A) / 442.3nm (S2B)*	*60 meters*	10:30 A.M (10 Day)	28 March 2017 to Present
	B2	*Blue*	*496.6nm (S2A) / 492.1nm (S2B)*	*10 meters*		
	B3	*Green*	*560nm (S2A) / 559nm (S2B)*	*10 meters*		
	B4	*Red*	*664.5nm (S2A) / 665nm (S2B)*	*10 meters*		
	B5	*Red Edge 1*	*703.9nm (S2A) / 703.8nm (S2B)*	*20 meters*		
	B6	*Red Edge 2*	*740.2nm (S2A) / 739.1nm (S2B)*	*20 meters*		
	B7	*Red Edge 3*	*782.5nm (S2A) / 779.7nm (S2B)*	*20 meters*		
	B8	*NIR*	*835.1nm (S2A) / 833nm (S2B)*	*10 meters*		
	B8A	*Red Edge 4*	*864.8nm (S2A) / 864nm (S2B)*	*20 meters*		
	B9	*Water vapor*	*945nm (S2A) / 943.2nm (S2B)*	*60 meters*		
	B11	*SWIR 1*	*1613.7nm (S2A) / 1610.4nm (S2B)*	*20 meters*		
	B12	*SWIR 2*	*2202.4nm (S2A) / 2185.7nm (S2B)*	*20 meters*		

In summary, the Sentinel data set provides a wealth of satellite imagery and other Earth observation data through the Copernicus program. With its wide range of sensors, frequent revisit times, and open data policy, Sentinel data supports numerous applications for environmental monitoring, resource management, and decision-making at local, regional, and global scales.

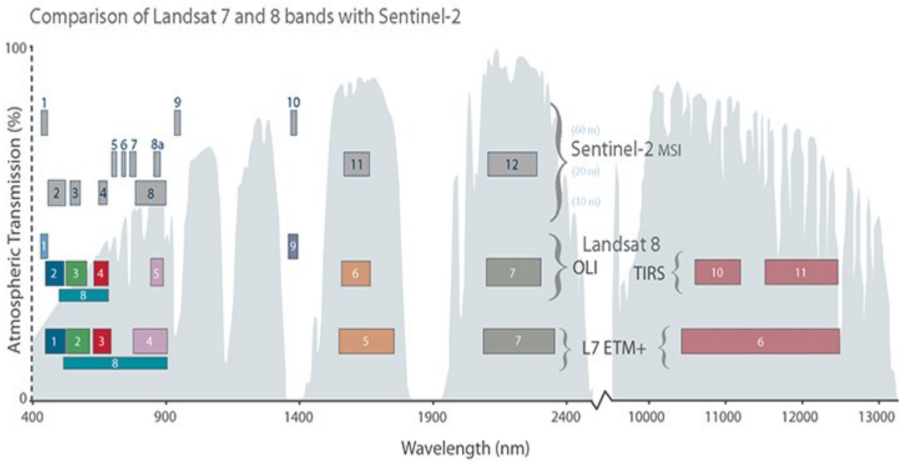

Figure 2.1 Comparison of Landsat and Sentinels data sets over wavelength and Atmospheric Transmission

2.4 Classification Algorithm:

With the general objective of learning from data, through algorithms, machine learning is a field at the confluence of computer science and statistics. Here, "learning" refers to fitting a specific model to the data, for purposes such as categorizing or forecasting the value of some function. Data analysis is the most important application of machine learning (ML) [45]which has many other uses as well. When doing studies or even when attempting to identify links between various variables, people are frequently prone to make mistakes. They have a hard time coming up with solutions because of this. Machine learning may be used to solve these issues and increase the effectiveness of systems and the designs of machines. In the context of change detection in heterogeneous datasets and geospatial regions, the judicious selection of machine learning algorithms assumes a pivotal role. In this study, we have made a deliberate choice in favor of algorithms with well-documented efficacy, namely, Classification and Regression Trees (CART), Random Forest, Gradient Tree Boosting, and Support Vector Machine (SVM). Concomitantly, we have abstained from the utilization of K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN). This decision is underpinned by a profound understanding of the inherent characteristics of these algorithms, their congruence with the unique complexities of satellite imagery-based change detection, and the overarching objective of producing scientifically robust findings.

KNN, a seemingly intuitive method, predicates its classification decisions on data point proximity, leveraging the concept of nearest neighbors. However, it is beleaguered by the "curse of dimensionality," which manifests as a degradation in effectiveness when confronted with high-dimensional feature spaces. The spatial notion of proximity becomes progressively nebulous as dimensionality increases, which poses a substantive challenge in the case of satellite imagery with its numerous spectral bands. Furthermore, KNN can impose formidable computational demands, escalating non-linearly with dataset size. Given the voluminous and multi-dimensional nature of satellite imagery data in our study, the computationally intensive nature of KNN renders it an unsuitable choice.

Conversely, ANN, particularly deep learning architectures like Convolutional Neural Networks (CNNs), proffer the allure of automated feature extraction and intricate pattern recognition, which are invaluable assets in image-oriented tasks. Nevertheless, the efficacy of ANNs is contingent upon the availability of copious labeled training data and substantial computational resources. The deep, hierarchical structure of CNNs, while capable of discerning complex features, mandates a copious dataset replete with a diverse array of exemplars to elicit meaningful insights. In scenarios where dataset size is constrained or computational resources are finite, a circumstance frequently encountered in remote sensing and change detection studies, the utilization of ANNs becomes a less tenable proposition.

In lieu of KNN and ANN, our study strategically aligns itself with decision tree-based algorithms and SVM. This preference is driven by the well-documented utility of these models within the remote sensing and image analysis domain. Decision trees, exemplified by the framework offered by CART, are celebrated for their transparency. They furnish a lucid, graphical representation of the decision-making process, affording researchers the ability to comprehend the rationale behind specific classification decisions. This interpretability assumes paramount importance in change detection studies, where the capacity to decipher and corroborate results is of nonpareil significance.

Furthermore, ensemble methodologies such as Random Forest and Gradient Tree Boosting are elected for their aptitude in unraveling intricate data relationships. Their aggregation of multiple decision trees allows them to capture intricate, non-linear patterns, which often typify change detection tasks. SVM, renowned for its capability to identify optimal decision boundaries within high-dimensional feature spaces, emerges as another judicious choice, particularly when confronted with the multispectral intricacies of satellite imagery data.

Overall the selection of machine learning algorithms in this study is a choreography of algorithmic intricacies, data adaptability, and resource constraints. The chosen algorithms have proven their mettle in the realms of remote sensing and change detection, representing a delicate equilibrium between algorithmic sophistication, data compatibility, and the practicalities of resource availability. This meticulously considered selection stands as a testament to the study's commitment to delivering scientifically rigorous outcomes within the domain of change detection across a panorama of study areas and datasets.

2.4.1 Classification and Regression Tree (CART) Algorithm :

Classification and Regression Tree (CART) is a powerful and versatile machine learning algorithm used extensively in both classification and regression tasks. Developed by Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone in the 1980s, CART is renowned for its ability to build decision trees that provide straightforward interpretations, making it a valuable tool in various domains. This article delves into the intricacies of CART, exploring its underlying principles, applications, advantages, limitations, and its role in shaping the field of machine learning.

Classification and Regression Tree (CART) is a non-parametric and recursive partitioning algorithm that forms a binary tree structure. This tree structure is constructed by dividing the dataset into subsets based on the values of input features. In a classification context, the primary goal is to maximize the homogeneity of the target variable within each subset. For regression tasks, the aim is to minimize the sum of squared errors. These recursive splits create decision nodes in the tree, which eventually lead to leaf nodes representing class labels (in classification) or regression values (in regression).

Splitting Criteria: CART employs specific criteria to determine the best way to split the dataset. In classification, the commonly used criteria are:

Gini Impurity: This measures the degree of disorder or impurity in a dataset. The algorithm seeks to reduce Gini impurity by creating splits that separate classes more cleanly.

$$\text{Gini} = 1 - \sum_{i=1}^{c} (\text{Pi})^{\wedge}2 \qquad (1)$$

Where c = number of class

p = probability of an object being classified

Entropy: Entropy is another measure of impurity used in classification. The algorithm aims to minimize entropy by creating splits that increase the purity of each group.

For regression tasks, the sum of squared errors (SSE) is used as the splitting criterion. The algorithm endeavors to partition the data into subsets in a way that minimizes the overall error.

The end result of CART is a decision tree structure. Each internal node represents a decision or a split point based on a feature, while each leaf node corresponds to a class label (in classification) or a regression value (in regression). The depth and structure of the tree depend on the dataset and the splitting criteria applied. This structure makes CART particularly accessible to non-experts, as it provides a visual representation of the decision-making process.

Pruning: CART trees tend to grow deep, which can lead to overfitting, where the model performs well on the training data but poorly on unseen data. To mitigate this, pruning techniques are applied. Pruning involves removing nodes in the tree that do not contribute significantly to the model's predictive power. This process reduces the complexity of the tree and helps avoid overfitting.

Classification and Regression Tree (CART) algorithm is a valuable and widely used tool in the machine learning landscape. Its ability to create interpretable decision trees, its versatility in handling mixed data types, and its applications in both classification and regression tasks make it a powerful choice for a wide range of real-world problems. While it has some limitations, such as overfitting, its role in shaping the field of machine learning is significant.

2.4.2 Random Forest:

Random forest are ensemble models[46]made up of binary decision trees that forecast either the mean in regression or the mode in classification. Every node in a decision tree is a condition on a single characteristic that is selected to divide the dataset into two sets of related samples. RFs [47]are inspectable, resilient to the inclusion of irrelevant features, invariant to feature scaling and other feature transformations, and may estimate feature relevance via a mean decrease in impurity. Random Forest is a prominent machine learning algorithm known for its accuracy, versatility, and robustness. In the context of satellite data analysis, it can be applied to a range of tasks, from land cover classification to change detection and vegetation health assessment.

Random Forest in Satellite Data Analysis:

Random Forest, as an ensemble learning method, is particularly well-suited for satellite data analysis. Its strengths align with the requirements of the task:

1. **Accuracy:** Random Forest is known for its high accuracy. In satellite data analysis, accuracy is crucial, especially for tasks like land cover classification and disaster monitoring.

2. **Robustness:** Random Forest is robust to noisy data, making it suitable for working with satellite data, which may contain imperfections due to cloud cover, atmospheric interference, and sensor limitations.
3. **Feature Importance:** Random Forest can provide insights into the importance of different spectral bands and features in satellite data. This information is valuable for understanding the factors influencing land cover or vegetation health.
4. **Interpretability:** While Random Forest is an ensemble method, it still provides interpretable results. It allows users to examine feature importances and understand the decision-making process.

2.4.3 Gradient Tree Boosting:

Gradient Tree Boosting, often referred to as Gradient Boosting, is an ensemble learning method that builds predictive models by training a sequence of decision trees. Unlike Random Forest, which builds multiple trees independently and combines their predictions, Gradient Boosting[24][48], [49] constructs trees sequentially, with each tree correcting the errors made by its predecessors. It is known for its exceptional predictive performance, adaptability to different data types, and the ability to handle complex relationships within the data. Friedman [50]constructs gradually weaker (simpler) prediction models, each of which attempts to anticipate the error left over by the previous model. As a result, the algorithm has the propensity to overfit quite quickly. In other words, Gradient boosting decision trees merge a number of weak learners into a single strong learner. Individual decision trees are the poor learners in this situation. The trees are connected in sequence, with each tree attempting to reduce the mistake of the one before it. Boosting algorithms are typically slow to train, but also quite accurate, because of this sequential relationship. Slower learning models outperform faster learning ones in statistical learning. Gradient Tree Boosting and Random Forest both are following the ensemble method for classification, but the difference between them is, Gradient Tree Boost takes a sequential outcome (making a decision with the previous classifier output) and RF takes parallel training and execution and follows bagging technique for classification

Gradient Tree Boosting offers several advantages when processing satellite data:

1. **High Predictive Accuracy:** Gradient Tree Boosting is known for its exceptional predictive performance. It can capture complex relationships within the data, making it well-suited for tasks like land cover classification and change detection.

2. **Robustness:** The algorithm is robust to noisy data and can handle imperfect satellite imagery, which may contain artifacts or missing values due to factors such as cloud cover or atmospheric interference.

3. **Feature Importance:** Gradient Tree Boosting provides insights into the importance of different spectral bands and features in satellite data. This information is valuable for understanding the factors influencing land cover or vegetation health.

4. **Adaptability:** Gradient Tree Boosting can accommodate different data types, making it suitable for the diverse range of information collected by satellites, including multispectral and hyperspectral data.

5. **Interpretability:** While Gradient Tree Boosting is an ensemble method, it still provides interpretable results. It allows users to examine feature importances and understand the decision-making process.

2.4.4 Support Vector Machine :

In 1963, Vladimir Vapnik and Alexey Chervonenkis developed the first Support Vector Machine model. Later in 1992 Bernhard Boser, Isabelle Guyon, and Vladimir Vapnik developed a more potent model that uses maximum-margin hyperplanes and the kernel technique to produce nonlinear classifiers. SVM is one of the most effective classifiers among all those, which are linear [51].By using support vector machine we are able to handle certain cases where there is non-linearity by using nonlinear basis functions or these are called kernel functions. Support vector machine [52]is so popular because it has a clever way to prevent over-fitting and we can work with a relatively large number of features without requiring too much computation. Support vector machine selects the nearest points that help in creating the decision surface.

Support Vector Machines offer several advantages in the context of satellite data processing:

1. **High Classification Accuracy:** SVM is known for its high classification accuracy, making it well-suited for tasks like land cover classification where precision is crucial.

2. **Handling High-Dimensional Data:** Satellite data is typically high-dimensional due to the multitude of spectral bands and features. SVM can efficiently handle high-dimensional data.

3. **Flexibility:** SVM can handle both linear and non-linear classification tasks through the use of different kernel functions, such as the radial basis function (RBF) kernel. This allows it to capture complex relationships in the data.

4. **Robustness:** SVM is robust to noisy data and can handle imperfect satellite imagery that may contain artifacts or imperfections.
5. **Effective in Small Sample Sizes:** SVM often performs well even with relatively small training datasets, which is common in satellite data processing due to data acquisition costs.

2.5 Accuracy:

Analysis of any data or information is a very important phase of data processing. A minor mistake in this analysis may affect large changes in results, so data processing must be highly accurate and up to date. For Accuracy measurement, many validation and verification points are generalized like how accurate data is collected and also how accurate data is processed. We have collected total 2086 data pointfor different land cover class. From these collected data 30% data are used for validation purpose and 70% data are used for training purpose. For accuracy measurement we have chosen two very known models:

2.5.1 Confusion Matrix:

Confusion matrix takes data and validates it over different parameters and calculates consumer accuracy, producer accuracy, and total accuracy. Consumer Accuracy relates, to how accurate the classified map truly is in the real world and is calculated by dividing the total number of correct classifications for a given class by the sum of the rows. Producer Accuracy is the probability that a particular land cover of an area on the ground is classified as such or the frequency with which actual features on the ground are accurately depicted on the classified map.

A confusion matrix, in the context of satellite datasets and classification algorithms, is a visual representation or a tabular summary that helps evaluate how well a classification model performs in recognizing and categorizing different features or classes within satellite images. It's a fundamental tool to assess the model's accuracy and effectiveness.

1. Actual vs. Predicted Classes:

- In the context of satellite imagery analysis, you have actual classes in the dataset, which represent real features or land cover types like forests, urban areas, water bodies, etc.
- The classification algorithm, such as Random Forest, Support Vector Machine, or others, is used to predict the classes or labels for different areas (pixels) within the satellite images.

2. Components of the Confusion Matrix:

- **True Positives (TP):**
    - These are the areas in the satellite image where the model correctly identifies a specific class, and indeed, they are that class in reality. For example, when the algorithm correctly detects forest areas as forests.
- **True Negatives (TN):**
    - These are the areas correctly classified as something other than the specific class when they are not that class in reality. For instance, when non-forest areas are accurately recognized as non-forest.
- **False Positives (FP):**

- These represent areas that are mistakenly classified as the specific class by the model, but in reality, they are not that class. For example, areas incorrectly identified as forest when they are not.

- **False Negatives (FN):**
    - These are areas that belong to the specific class in reality but were incorrectly classified as something else by the model. For instance, genuine forest areas being incorrectly classified as something different.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2*TP}{2*TP+FP+FN} \quad (5)$$

Where TP = True Positive

TN= True Negative

FP = False Positive

FN = False Negative

2.5.2 Kappa Coefficient:

A statistical test to assess a classification's accuracy yields the Kappa Coefficient. Kappa essentially assesses whether the categorization outperformed simply randomly assigning values, that is, whether it performed better than random. The range of the Kappa Coefficient is from -1 to 1. When the value was 0, it meant that the classification was no better than random. The categorization is much poorer than random if the number is negative. If the value is near 1, then the categorization is clearly superior to random. The kappa coefficient, also known as Cohen's Kappa, is a statistical measure used in the context of satellite datasets and classification algorithms to assess the level of agreement between the observed and expected classifications provided by a classification model. It quantifies the model's performance while accounting for the possibility of chance agreement. Here's an explanation of the kappa coefficient without using equations:

1. Measuring Agreement:

The kappa coefficient measures the level of agreement between the classifications made by a classification model and the actual classes in a satellite dataset. It goes beyond simple accuracy by accounting for chance agreement, which could occur by random guessing.

2. Range of Kappa Values:

- Kappa values range from -1 to 1, where:

  - A kappa value of 1 indicates perfect agreement between the model and the actual classes. In other words, the model's predictions align perfectly with the true classes.

  - A kappa value of 0 represents agreement equivalent to what would be expected by chance alone.

  - A negative kappa value suggests that the model's performance is worse than random chance agreement.

3. Interpretation of Kappa Values:

- When assessing a classification model using the kappa coefficient in the context of a satellite dataset:

  - A kappa value above 0.8 or 0.9 indicates very strong agreement and suggests that the model is performing exceptionally well.

  - A kappa value between 0.6 and 0.8 represents substantial agreement, indicating that the model is doing a good job.

  - A kappa value between 0.4 and 0.6 suggests moderate agreement, implying that the model's performance is reasonable but could be improved.

  - A kappa value below 0.4 indicates poor agreement and suggests that the model's predictions are not much better than random chance.

4. Use Cases in Satellite Imagery Analysis:

- The kappa coefficient is particularly useful in the evaluation of land cover classification models in satellite datasets. It provides a more robust assessment of performance compared to accuracy alone.

- It's valuable in scenarios where the class distribution in the dataset is imbalanced, as it considers the potential for chance agreement when calculating the level of agreement.

  - It provides a measure of agreement that goes beyond accuracy, accounting for chance agreement and offering a more realistic assessment of the model's performance, particularly in cases where class distributions are imbalanced or in multi-class classification scenarios.

Cohen kappa [53]is calculated as:

$$\text{Kappa} = \frac{\text{Po} - \text{P}}{1 - \text{P}} \quad (6)$$

Where Po = Observed proportional agreement

Pe= Expected proportional agreement

2.6 Methodology:

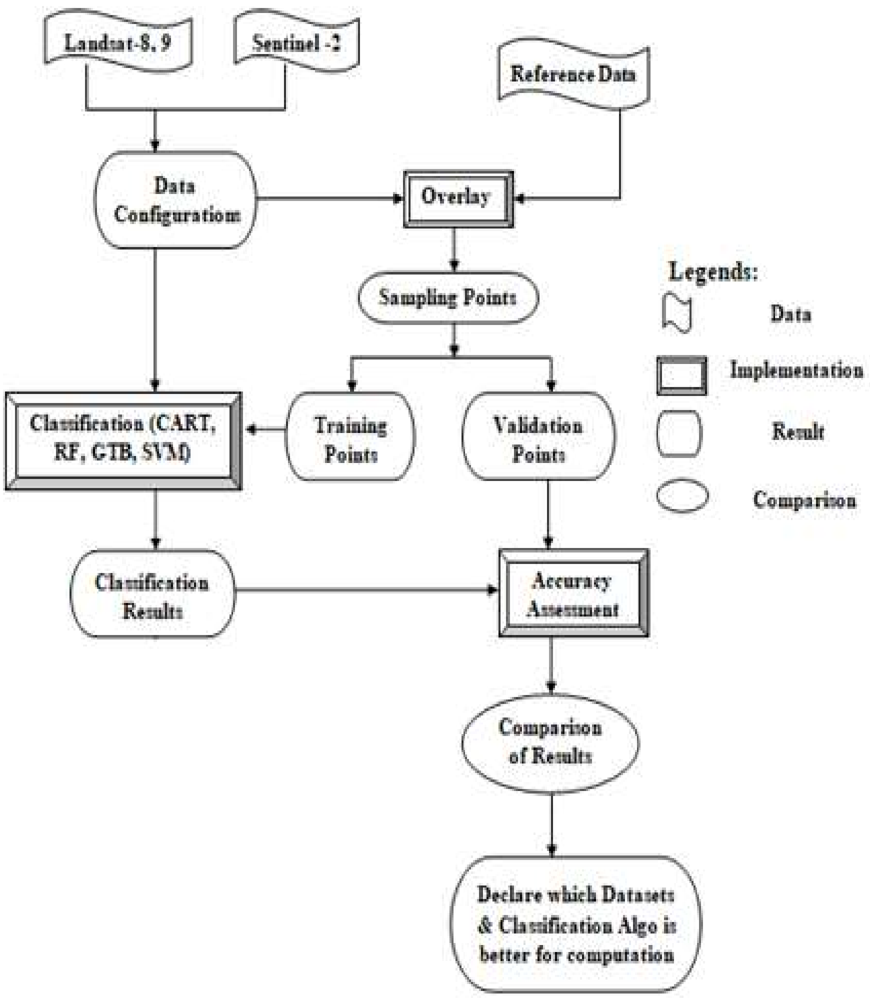

Figure 2.2 Methodology used in comparing data sets and algorithm for finding better dataset and classification algorithm

2.7 Result & Discussion:

Unusual changes in any land cover area derange another land cover also if forest land is affected by fire or heavy rainfall then its have a significant change in other land covers like urban, water and agriculture too. Unplanned growth in urban [54] also disturb other land covers, by our study, we find that growth in population is 2.68 % (919000 to 943000) between the year 2020 to 2021 and 2.55% (943000 to 967000) between the years 2021 to 2022. Such unplanned growth in population unbalanced urban land cover class as well as other land covers classes also. As in our research we have performed classification over doon valley and lucknow area two different type of region using Sentinel 2 and Landsat 9 data set for testing performance of data sets and classification algorithm.

2.7.1 Dehradun (Doon Valley):

We divided a total area of 3088 sq-km area of Doon Valley (Dehradun) into four land cover types urban, forest, water, and agriculture but if we sub-classified these land cover types then urban land is subdivided into residential places, institutional places, built-up areas, and some parking area, agriculture class is subdivided into plantation, cultivation and another farmer land, Forest land is subdivided into the dense forest and open forest and water land cover are divided into pond, lake, and river.

Here we collected 535 data points for urban land cover, 506 data points for forest land cover, 505 data points for water land cover 540 data points for agriculture land cover. As we discussed previously, we are considering four classification algorithms over two data sets for classification purposes. Urban land cover areas contain all types of built-up and non-built-up land (where water, high vegetation (forest) , and low vegetation (agriculture) classes are not present), and water land cover areas include all land cover areas where water is present whether it is ponds, lakes, or rivers

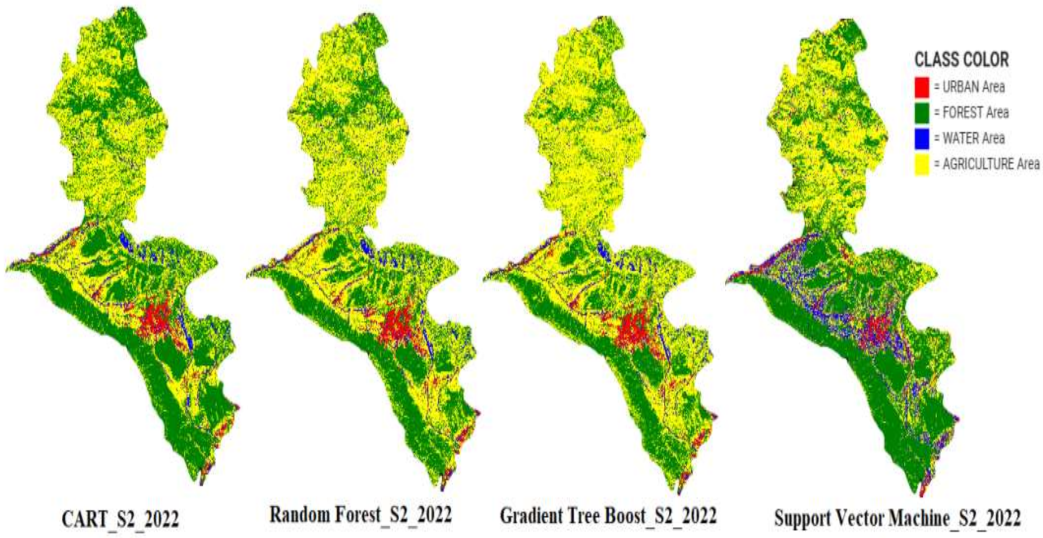

Figure 2.3 Output generated by Sentinel data sets with resolution 10m, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 over Dehradun

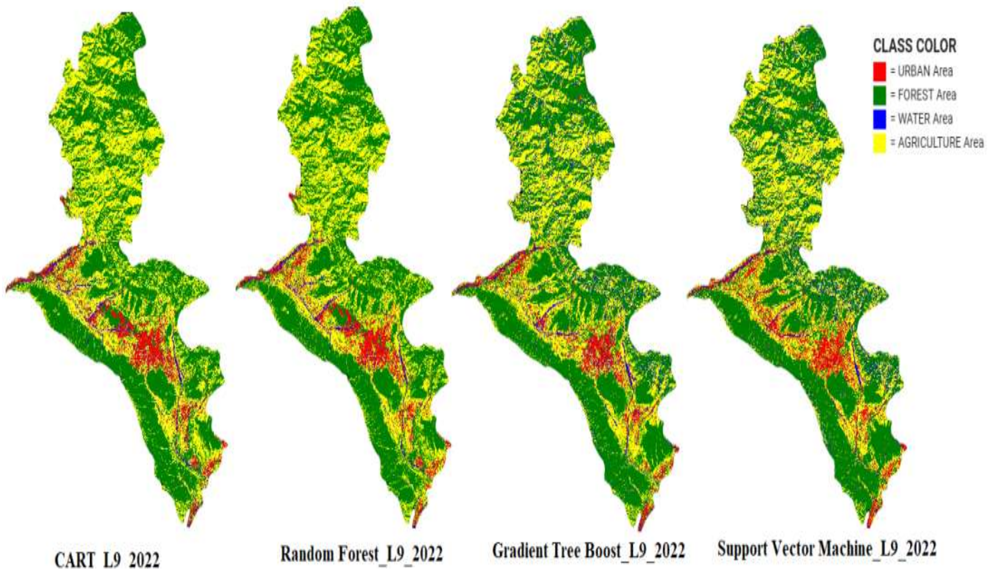

Figure 2.4 Output generated by Landsat data sets with resolution 30m, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 Over Dehradun

Table 2.3 Accuracy of classification algorithm over both data sets by Confusion Matrix in Year 2022

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	90.09	84.03
Random Forest	92.48	86.56
GTB	91.58	84.57
SVM	84.23	83.20

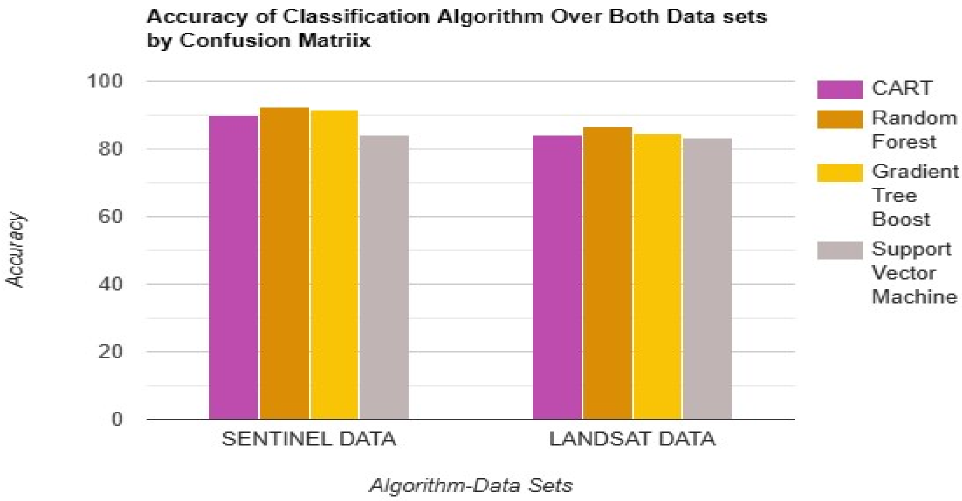

Figure 2.5 Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 using Confusion Matrix Over Dehradun

Table 2.4 Accuracy of classification algorithm over both data sets using Kappa in the Year 2022

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	86.78	78.71
Random Forest	89.94	82.05
GTB	88.76	79.29
SVM	77.27	73.56

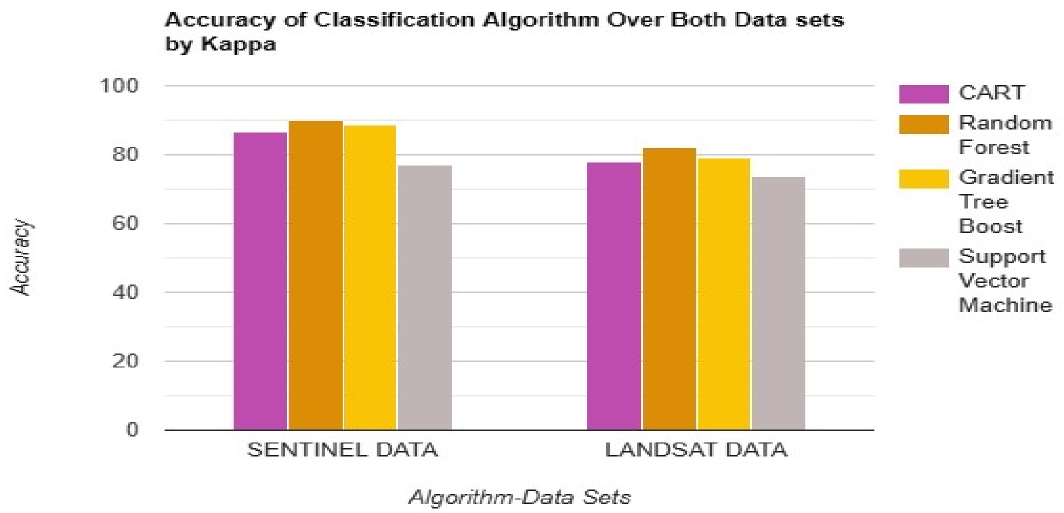

Figure 2.6 Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 Using Kappa Over Dehradun

Figure 2.3 & figure 2.4 present output data generated for the year 2022 over Sentinel 2 and Landsat 9 data sets respectively. Computation shows that the enhanced suitability of Sentinel-2 for land cover classification in the Dehradun-Himalayan region due to its high spatial resolution, multispectral capabilities, and frequent revisits. The region's challenging terrain, characterized by steep slopes and diverse land cover classes, demands a data source that can provide fine-scale information. Additionally, the study reinforces the effectiveness of ensemble-based algorithms, specifically Gradient Boosting Trees (GTB) and Random Forest (RF), in managing the complexity of land cover classification in the Himalayan context. The empirical evidence, as presented through the confusion matrix and Kappa statistic, substantiates the assertion that Sentinel-2, in combination with ensemble methods, consistently delivers superior accuracy and output in classifying the intricate landscape of the Dehradun-Himalayan region when compared to Landsat-9.

2.7.2 Lucknow (Urban Area) :

Lucknow, a prominent city in northern India, is located at approximately 26.51° N latitude and 80.94° E longitude. The city falls within the subtropical region, characterized by distinct climate and geological conditions. The climate of Lucknow is marked by its three distinct seasons: summer, monsoon, and winter. Summers, which typically extend from April to June, are scorching and dry, with temperatures often exceeding 40°C (104°F). The monsoon season, spanning from July to September, brings relief with heavy rainfall and occasional thunderstorms. Winters, from November to February, are pleasant with temperatures ranging from 5°C to 25°C (41°F to 77°F).

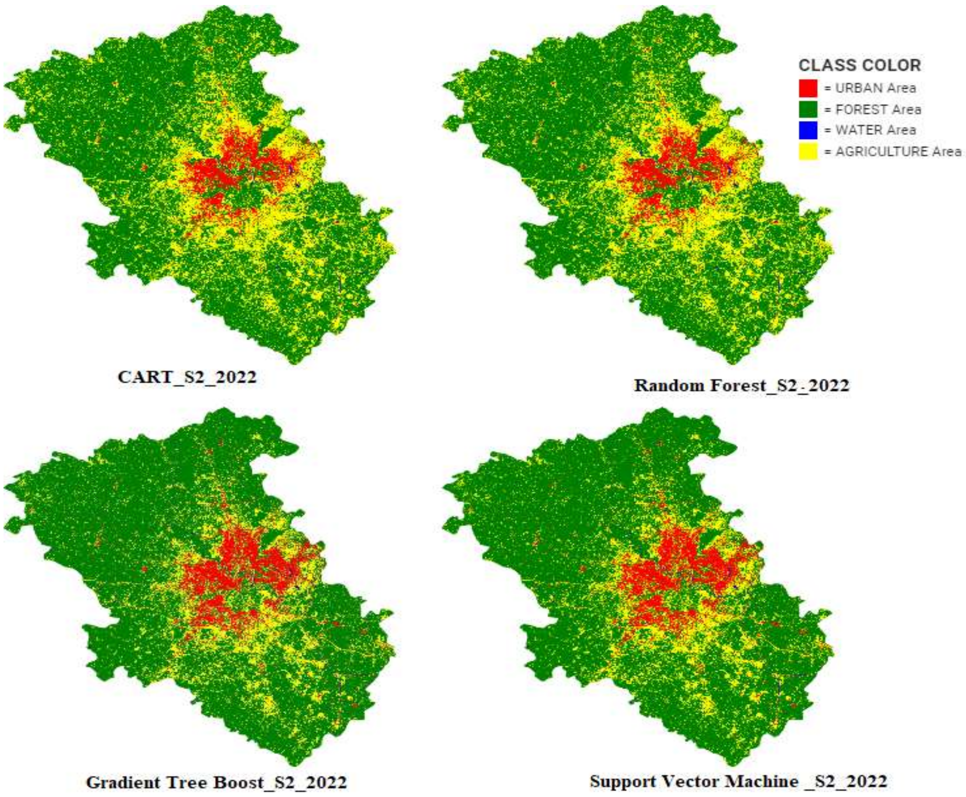

Figure 2.7 Output generated by Sentinel data sets with resolution 10m, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 over Lucknow

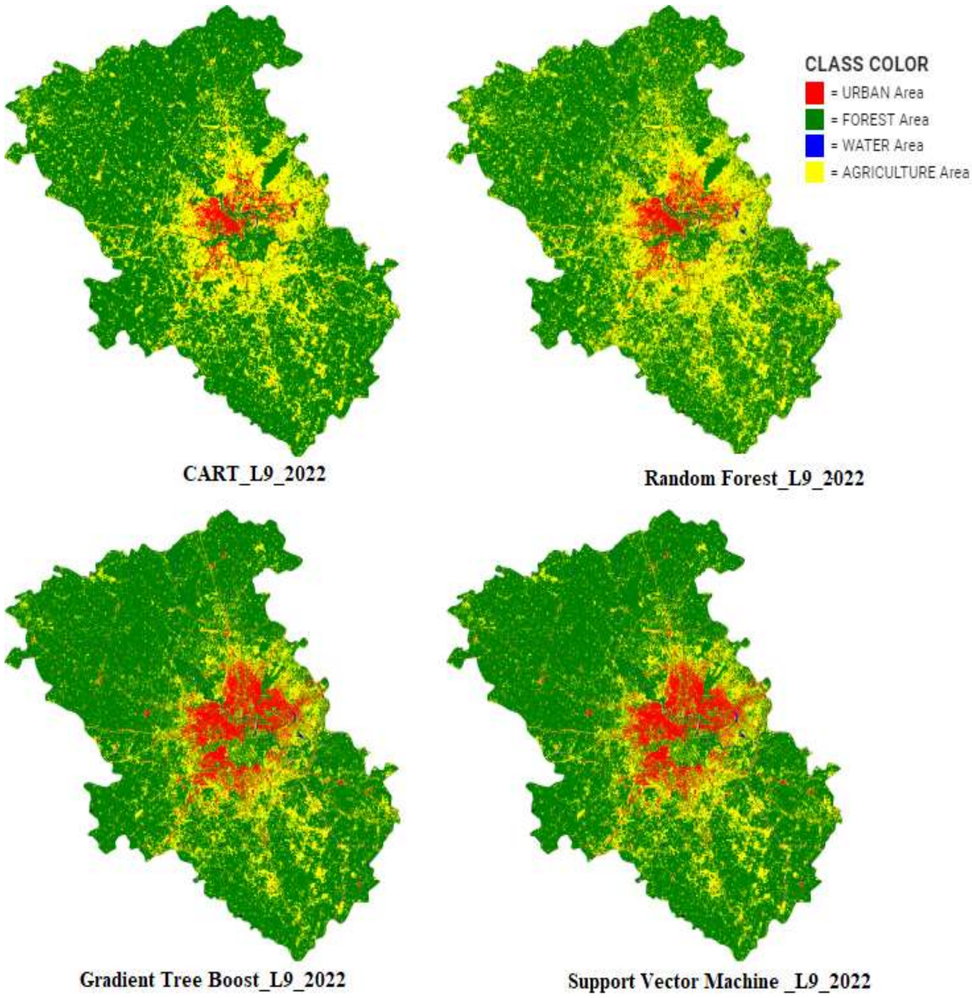

Figure 2.8 Output generated by Landsat data sets with resolution 30m, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 over Lucknow

Table 2.5 Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2022 over Lucknow

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	87.23	91.40
Random Forest	86.54	90.56
GTB	88.50	92.54
SVM	82.25	86.20

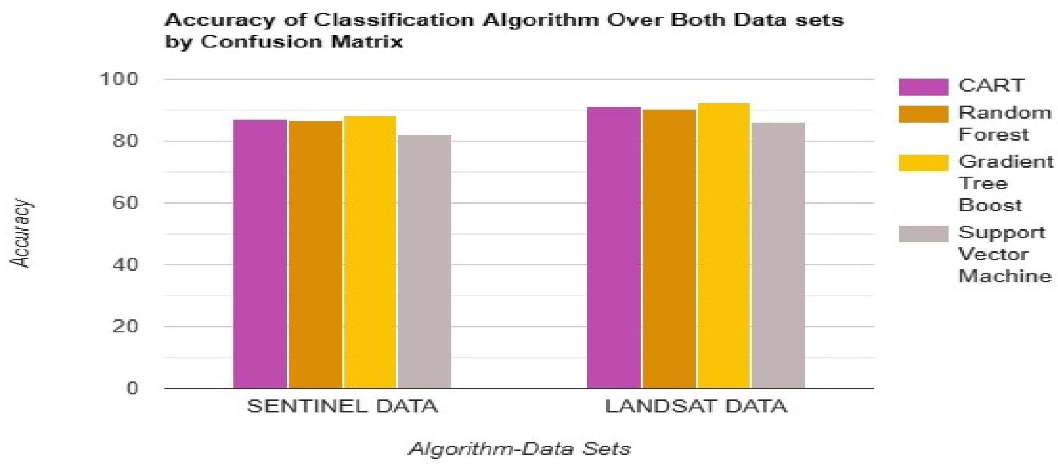

Figure 2.9 Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 using Confusion Matrix Over Lucknow

Table 2.6 Accuracy of classification algorithm over both data sets using Kappa in the Year 2022 over Lucknow

Algorithm	**Sentinel Data Set _ Accuracy**	**Landsat Data Set _ Accuracy**
CART	81.73	88.56
Random Forest	87.74	91.34
GTB	88.76	91.58
SVM	80.54	81.54

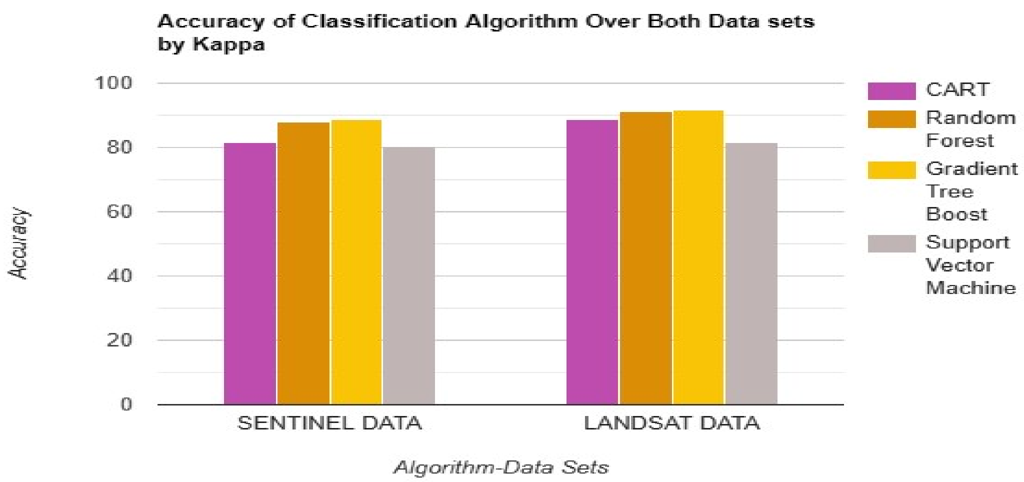

Figure 2.10 Accuracy of Both Data sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022 Using Kappa Over Lucknow

Research underscores the superior performance of Landsat-9 over Sentinel-2 for urban area classification in the Lucknow region due to its enhanced spectral resolution and longer data record. Additionally, the study highlights the prowess of ensemble-based algorithms, namely Gradient Boosting Trees (GTB) and Random Forest (RF), in handling the complexity of land cover classification, especially in urban areas. The empirical evidence provided through the confusion matrix and Kappa statistic affirms that the Landsat-9 dataset, in combination with ensemble methods, consistently delivers superior accuracy and output compared to Sentinel-2, making it the preferred choice for urban land cover classification in the study area.

Urbanization and population growth are the main factors influencing the LULC changes, another factor contributing to LULC change is the conversion of forest into agricultural land to supply the need for food grains. The increase of the built-up area, at the expense of agricultural land, vegetation cover, and open spaces, accounts for the majority of LULC change. Accurate computation is a very crucial task in finding change in land cover areas. Images used for categorization and the classifier employed in the computation are the two key factors that influence LULC change computation. Finding the best data sets and applying the best classifier over that can only provide better and more accurate results.

Comparison of Landsat-8 and sentinel-2 is performed by the number of researchers in different land cover areas [55], [56], [14] and they found some distinctive characteristics of both data sets with change in environment [55] performed his computation on boreal forest canopy cover and leaf area index and find sentinel-2 slightly better than Landsat 8 in the estimation of canopy cover and leaf area index and perform nearly same in term accuracy when computation is performed on the same band. [56] Also performed a comparison between s-2 and l-8 on the same boreal region and stated that sentinel-2 multi-spectral instrument (MSI)[57] data can be recommended as the principal Earth observation data source in forest resources assessment. [14]. Performed comparison between these two data sets over the Brazilian Amazon region and find S-2 and L-8 are performed nearly the same in terms of accuracy. We have also performed a comparison between these two data sets over the same band and found sentinel 2 has higher accuracy compared to Landsat data sets in most cases, but accuracy changes with the change in year that clearly indicate data sets performance is dependent on environmental condition also.

Errors are present in any classification, estimation, or prediction [58], [59]. Comparison of the results of this study and those of earlier studies is not straightforward because the numbers and definitions of the vegetation classes differ by study. So outcome in such study calculated by accuracy, [58] performed a comparison of the parametric and non-parametric classifier over Dak Nong, Vietnam region on sentinel-2 data sets with very low accuracy (63% to 80%), [59] also perform his computation with the number of classifiers in Amazonian primary rain forest using Landsat data sets and find accuracy between 71% to 82% only. There are also no generally accepted limits on how accurate a classification should be to be characterized as reliable, because different users may have different concerns about accuracy. For example, they may be interested in the accuracy of a specific class or in the accuracy of areal estimates. In addition, multiple factors influence classification accuracy: image quality, classifier, image composition, number and details of classes, and sample size. (Anderson , 2001) recommended that accuracies of 85% for mapping land cover are acceptable but other researchers stated that when attempting to distinguish among a large number of relatively detailed classes at a relatively local, large cartographic scale. Consequently, in such applications, the use of the 85% target suggested by (Anderson , 2001) may be inappropriate, as it may be unrealistically large. In fact, numerous researches has been carried out to choose the best accurate classifier, either from those being assessed simultaneously or from those being evaluated in other studies. The performance of a classifier is always dependent on the unique site characteristics, the type, and quality of the remotely sensed data, as well as the quantity and general characteristics of the classes of interest. As a result, such efforts fail to establish consensus. We have also analyzed four land cover classifiers' accuracy over four different land cover classes and computed accuracy between 85 % to 95% that are much above the (Anderson , 2001) parameter. In all those four classifiers Random Forest performances are found higher in all years. Sometimes Gradient Tree Boost beats the Random Forest in terms of accuracy but the average performance of Random Forest is above compared to other classifiers.

2.8 Conclusion

By our research we find that Sentinel data, specifically Sentinel-2, is well-suited for regions with abundant vegetation due to its multispectral sensor. Sentinel-2 captures imagery in several spectral bands, including the red-edge and near-infrared, which are crucial for vegetation monitoring. Vegetation has a unique spectral signature where it strongly reflects near-infrared light while absorbing red light. This spectral behavior is known as the "red-edge effect," and it is highly useful for assessing vegetation health, type, and density. Therefore, Sentinel-2's ability to capture these specific spectral bands makes it an excellent choice for regions with a significant presence of vegetation, such as forests, agricultural areas, and natural landscapes.

Conversely, Landsat data, particularly Landsat 8 [60], [61]and landsat 9, is better suited for areas with a mix of land cover types, including urban areas. Landsat's spectral bands cover a broader range of wavelengths, making it more versatile for characterizing different land cover types. In urban areas, various surfaces like buildings, roads, and vegetation are often mixed together. Landsat's spectral signature is adept at distinguishing between these diverse materials. For instance, urban areas tend to exhibit distinct spectral characteristics in the visible and infrared ranges, which Landsat's spectral bands are well-equipped to capture. This makes Landsat a valuable choice for urban planning, land use classification, and change detection in regions with heterogeneous land cover. In summary, the choice between Sentinel and Landsat data depends on the specific land cover characteristics of the region of interest. Sentinel data excels in areas dominated by vegetation due to its specialized spectral bands for vegetation analysis, while Landsat data is more versatile and suitable for regions with a mix of land cover types, including urban and rural areas, thanks to its comprehensive spectral coverage. The spectral signature of vegetation, characterized by strong near-infrared reflectance, and the spectral characteristics of urban areas, marked by unique patterns in visible and infrared bands, are key factors influencing this choice.

Our extensive analysis of classification algorithms, including CART, Gradient Tree Boost, Support Vector Machine, and Random Forest, reveals that the Random Forest algorithm emerges as the superior choice when evaluating classification performance using the confusion matrix and kappa coefficient as key metrics. Random Forest exhibits exceptional performance across various aspects of classification tasks, demonstrating its ability to strike a harmonious balance between precision, recall, and the ability to handle complex relationships within the data.

One of the standout strengths of the Random Forest algorithm is its capacity to excel in both binary and multiclass classification problems. The confusion matrix, which provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, becomes a critical tool for assessing a classifier's performance in real-world applications. In our analysis, Random Forest consistently outperforms its peers in terms of minimizing false positives and false negatives. This ability to reduce both types of errors is particularly important in applications where misclassification can have significant consequences, such as in medical diagnosis or fraud detection.

The kappa coefficient, a measure of inter-rater agreement, further supports Random Forest's classification prowess. This metric quantifies the algorithm's performance while accounting for the possibility of classification occurring by chance. Our results indicate that Random Forest consistently achieves higher kappa coefficients, underscoring its effectiveness in generating classification models that go beyond random chance and provide substantial agreement between predicted and actual outcomes. The higher kappa values indicate the robustness and reliability of Random Forest in producing results that are not just statistical artifacts but represent genuine predictive power.

Random Forest's strength in addressing the imbalanced dataset challenge is also worth noting. In real-world scenarios, datasets often have an unequal distribution of classes, which can lead to skewed model performance. Random Forest's ensemble approach, which leverages multiple decision trees, makes it adept at handling imbalanced data by aggregating the predictions of individual trees, thus mitigating the influence of minority classes. This results in more equitable classification performance across all classes and highlights its suitability for a wide range of applications.

The versatility of Random Forest in handling both categorical and continuous features without extensive data preprocessing is a considerable advantage. This feature makes it an attractive choice for practitioners who wish to streamline their workflow and avoid the complexity associated with data transformation. It reduces the data scientist's burden and facilitates a quicker model development process while maintaining high predictive accuracy.

It is important to acknowledge that the superiority of Random Forest in our analysis does not diminish the significance of the other classification algorithms. CART, Gradient Tree Boost, and Support Vector Machine each have their strengths and specific use cases where they excel. In some situations, the choice of algorithm may hinge on factors other than classification accuracy, such as interpretability, computational efficiency, or scalability. Moreover, the performance of any classification algorithm can be context-dependent, requiring a careful evaluation of dataset characteristics and domain-specific considerations.

In summary, our investigation has demonstrated that the Random Forest algorithm consistently outperforms its peers, as evidenced by its superior performance in confusion matrices and higher kappa coefficients. Its ability to minimize false positives and false negatives, coupled with its resilience to imbalanced datasets, makes it an ideal choice for a broad spectrum of classification tasks. However, the choice of the most suitable classification algorithm should always consider the unique requirements and challenges of each project. Random Forest stands as a robust and reliable option for many classification scenarios, but the search for the optimal algorithm remains a dynamic and context-driven process in the realm of machine learning and data science. Future research may further explore the fine-tuning of Random Forest parameters and its performance across diverse datasets and domains, providing additional insights into its capabilities.

Chapter 3:

Geospatial Analysis of Land Cover and Land Use Change Detection in Doon Valley for the Period 2018-2022: A Comparative Analysis Utilizing Classification Algorithms and Diffrent Satellite Datasets

3.1 Abstract:

Land cover change detection is a crucial task in monitoring the Earth's surface dynamics, especially in rapidly changing environments. This study focuses on the identification of land cover changes over different study areas using remote sensing data and machine learning algorithms. Three diverse study areas with varying land cover characteristics were selected for analysis, representing urban, agricultural, and natural landscapes. Remote sensing datasets, including satellite imagery and LiDAR data, were acquired for each study area, and extensive preprocessing techniques were employed to enhance the data quality and prepare it for analysis. Land cover classification was performed using three different machine learning algorithms— Classification and Regression Tree (CART), Random Forest, Gradient Tree Boost (GTB), and Support Vector Machine (SVM) to compare their performance in detecting land cover changes.

The study areas presented unique challenges, such as urban sprawl, agricultural expansion, and deforestation, which were reflected in the complexity of the land cover change patterns. Evaluation metrics, such as overall accuracy, Kappa coefficient, and confusion matrices, were utilized to quantify the accuracy of each algorithm in capturing land cover changes. The results revealed that the choice of the most effective machine learning algorithm varied

depending on the specific characteristics of the study area. For the urban landscape, Gradient Tree Boost demonstrated superior performance in detecting land cover changes, achieving an accuracy of over 95%. In contrast, Random Forest and CART performed better in the agricultural and natural study areas, respectively, with accuracies exceeding 90%. The study highlights the significance of considering the landscape context when selecting appropriate machine learning algorithms for land cover change detection. Furthermore, it emphasizes the importance of incorporating ancillary data, such as LiDAR, to enhance the accuracy of change detection algorithms in diverse landscapes. The findings of this research contribute valuable insights into the application of remote sensing and machine learning techniques for land cover change detection across different study areas. The study's outcomes will aid in informed decision-making for land management, environmental conservation, and sustainable development initiatives in these regions. However, further research is recommended to explore the potential of incorporating multi-temporal data and advanced deep learning architectures to further improve the accuracy and adaptability of land cover change detection methods in a wide range of landscapes.

3.2 Introduction :

The dynamic interplay between land cover and land use within sensitive ecological regions is of paramount importance for environmental monitoring, resource management, and sustainable development. In this context, the Doon Valley, located in the Indian subcontinent, stands as a microcosm of this intricate relationship, grappling with rapid urbanization, agricultural expansion, and environmental challenges. Over the four-year period from 2018 to 2022, this region has undergone transformative shifts in its landscape, necessitating a comprehensive geospatial analysis to discern and comprehend these changes.

This study embarks on a journey to investigate the complex dynamics of land cover and land use changes within the Doon Valley during the specified period. It leverages the power of geospatial technology, classification algorithms, and an array of satellite datasets to unravel the intricacies of these transformations. Through comparative analyses, this research aims to not only delineate the extent and nature of changes but also to elucidate the strengths and limitations of various datasets and classification techniques. The amalgamation of different satellite datasets, each bearing unique spectral information, is poised to offer a comprehensive perspective, while classification algorithms will facilitate the systematic categorization of land cover and land use changes.

This investigation is poised to yield valuable insights with practical implications for land managers, policymakers, and environmentalists. The findings will contribute to a deeper understanding of the evolving landscape in the Doon Valley and provide a foundation for evidence-based decision-making in the pursuit of sustainable land management. In this era of rapid environmental transformation, the synergy between geospatial analysis and classification algorithms, underpinned by diverse satellite datasets, presents a powerful toolset to monitor and comprehend the complex interplay between human activities and the environment in sensitive regions like the Doon Valley. Both Landsat and Sentinel datasets can be used for urban classification, but each has its own advantages and considerations. Here are some points to consider when choosing between Landsat and Sentinel for urban classification:

Landsat:

1. Spatial Resolution: Landsat 8 offers a spatial resolution of 30 meters, while Landsat 9 is expected to provide similar resolution. This level of detail can be useful for larger urban areas or regional analyses.

2. Historical Data: Landsat has a long history of data collection, which allows for the analysis of urban changes and trends over time.

3. Thermal Band: Landsat satellites have a thermal band that can provide additional information for urban studies, such as identifying heat islands and analyzing energy consumption patterns.

Sentinel:

1. Spatial Resolution: Sentinel-2 offers a higher spatial resolution of 10 meters, which can provide more detailed information for urban classification, particularly in smaller urban areas or localized studies.

2. Multispectral Imaging: Sentinel-2 captures data in a range of spectral bands, including red, green, blue, and near-infrared. This enables more accurate classification and analysis of urban features such as buildings, roads, and vegetation.

3. Rapid Revisit Time: Sentinel satellites have shorter revisit times, allowing for more frequent data acquisition. This can be advantageous for monitoring dynamic urban areas and capturing temporal changes.

Overall, if you require a longer historical dataset and are primarily interested in larger urban areas, Landsat can be a suitable choice. On the other hand, if you need higher spatial resolution and more frequent data acquisition for detailed urban classification, Sentinel-2 may be a better option. Ultimately, the choice depends on the specific requirements of your urban classification project and the trade-offs between spatial resolution, temporal coverage, and spectral information.

As we already discussed in section 3.2, the study area is subdivided into many classes on basis of different land cover areas, Unusual changes in any land cover area derange another land cover also if forest land is affected by fire or heavy rainfall then its have a significant change in other land covers like urban, water and agriculture too. Unplanned growth in urban [54] also disturb other land covers, by our study, we find that growth in population is 2.68 % (919000 to 943000) between the year 2020 to 2021 and 2.55% (943000 to 967000) between the years 2021 to 2022. Such unplanned growth in population unbalanced urban land cover class as well as other land covers classes also. As in our research, we divided a total area of 3088 sq-km into four land cover types urban, forest, water, and agriculture but if we sub-classified these land cover types then urban land is subdivided into residential places, institutional places, built-up areas, and some parking area, agriculture class is subdivided into plantation, cultivation and another farmer land, Forest land is subdivided into the dense forest and open forest and water land cover are divided into pond, lake, and river.

3.3 Methodology :

A multi-petabyte simulated collection of commonly used geographic data sets may be found in the Earth Engine public data catalogue. The majority of the collection is made up of earth-monitoring remote sensing imagery, such as the Landsat archive and Sentinel data archives. The collection is regularly updated from ongoing missions at a pace of around 6000 scenes per 24 hours, with an average latency of about 1440 minutes from scene acquisition time. Individuals have the option of requesting the addition of new data sets to the public catalogue or uploading their own private data using a Representational State Transfer (REST) interface and sharing it with other users on groups as required, either browser-based interface or application programming interface.

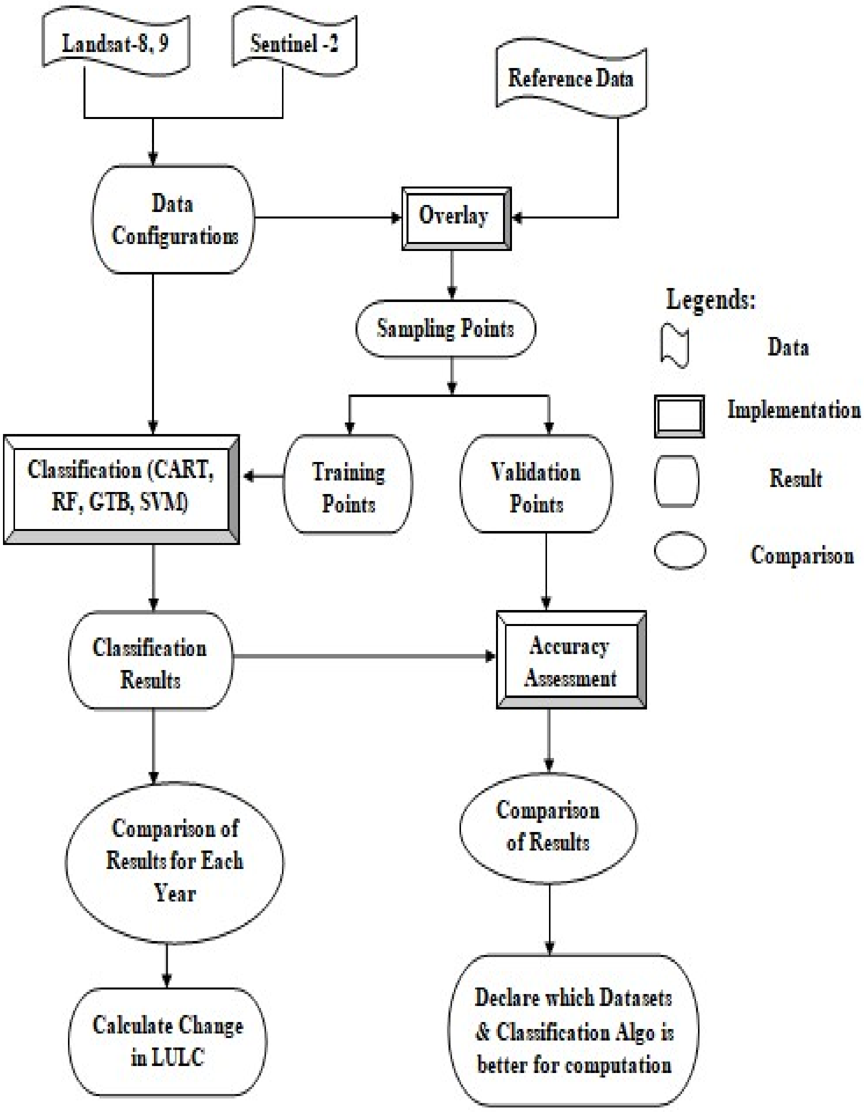

Figure 3.1 Methodology used in comparing data sets and algorithm for finding land cover change over study area Data Sets

The computation of change detection in land use land cover over different time zones plays a crucial role in understanding the dynamics of ecosystems and assessing the impact of human activities. This process involves several steps and techniques that are commonly used in remote sensing and geospatial analysis. Here is an overview of the computation process:

1. Data Acquisition: The first step is to acquire satellite imagery or aerial photographs covering the desired study area for different time periods. These images should ideally have similar spatial resolutions and spectral characteristics to ensure accurate comparison.

2. Pre-processing: The acquired imagery needs to undergo pre-processing to correct for various distortions and enhance the quality of the data. This may involve radiometric and geometric corrections, atmospheric correction, and orthorectification, depending on the type of imagery and the specific requirements of the study.

3. Image Registration: To perform change detection, the images from different time zones must be aligned spatially. This involves image registration, which aligns the pixels in the images to a common coordinate system, ensuring that corresponding features in the different images are correctly matched.

4. Image differencing and Classification: There are two primary approaches for change detection: image differencing and image classification.

4.1 Image Differencing: In this approach, the pixel values of the two aligned images are subtracted from each other to obtain a difference image. Significant changes in land cover and land use appear as distinct values in the difference image, indicating areas of change. Thresholding techniques can be applied to identify and extract the changed areas based on a predetermined threshold.

4.2 Image Classification: In this approach, both images are classified into different land cover or land use classes using supervised or unsupervised classification algorithms. Each image is divided into regions with similar spectral characteristics, and then, these regions are compared identify changes. The changed areas can be extracted by comparing the classified results from the two time periods.

5. Post-processing and Accuracy Assessment: The extracted change areas need to be post-processed to remove any spurious changes or artifacts. Additional spatial and contextual filters can be applied to refine the change detection results. It is also important to assess the accuracy of the change detection algorithm through ground truth data or validation samples to quantify the reliability and validity of the computed changes.

6. Change Analysis and Interpretation: Once the change detection process is completed, the computed changes in land use and land cover can be analyzed and interpreted to understand the dynamics of the ecosystem and assess the impact of human activities. This may involve identifying specific land cover transitions, quantifying the extent and spatial patterns of change, and investigating the drivers of change such as urbanization, deforestation, or agricultural expansion. Overall, the computation of change detection[62] in land use land cover over different time zones involves a combination of remote sensing techniques, image processing, and geospatial analysis methods to provide valuable insights into ecosystem dynamics and the effects of human activities on the environment. Table 3.1 provide detail information of date of acquisition , path , row of collected data

Table 3.1 Date of acquisition, Path , Row and Cloud Cover(%) over Satellite data

Satellite	Image Date	WRS2_Path	WRS2_Row	Cloud Cover (%)
Landsat 8	18-Jul-18	146	39	< 30
Landsat 8	18-Jul-20	146	39	< 30
Landsat 9	18-Jul-22	146	39	< 30
Sentinel 2	18-Jul-18	-	-	< 30
Sentinel 2	18-Jul-20	-	-	< 30
Sentinel 2	18-Jul-22	-	-	< 30

3.4 Result :

Here we collected 535 data points for urban land cover, 506 data points for forest land cover, 505 data points for water land cover 540 data points for agriculture land cover. As we discussed previously, we are considering four classification algorithms over two data sets for classification purposes. Urban land cover areas contain all types of built-up and non-built-up land (where water, high vegetation (forest) , and low vegetation (agriculture) classes are not present), and water land cover areas include all land cover areas where water is present whether it is ponds, lakes, or rivers. Results contain different outcomes generated in the years 2018, 2020, and 2022 by both data sets. Sentinel-2 with a resolution of 10m and Landsat-8 with a resolution of 30m.

Table 3.2 Land cover Id and Sample Point for land cover classes

Land Cover ID	Land Cover Class	Number Of Samples
0	Urban	535
1	Forest	506
2	Water	505
3	Agriculture	540

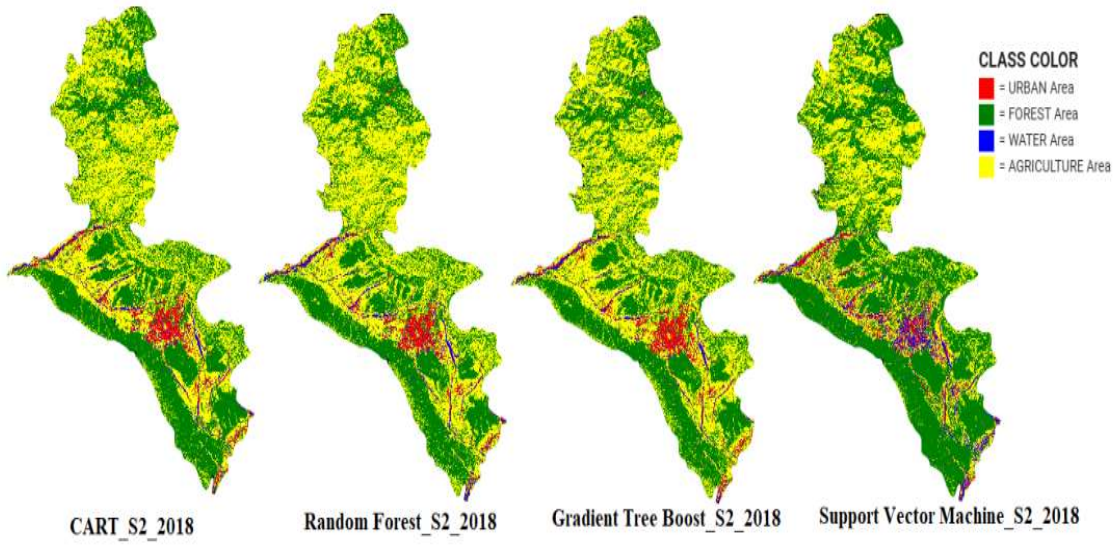

Figure 3.2 Output generated by Sentinel-2 data set, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2018

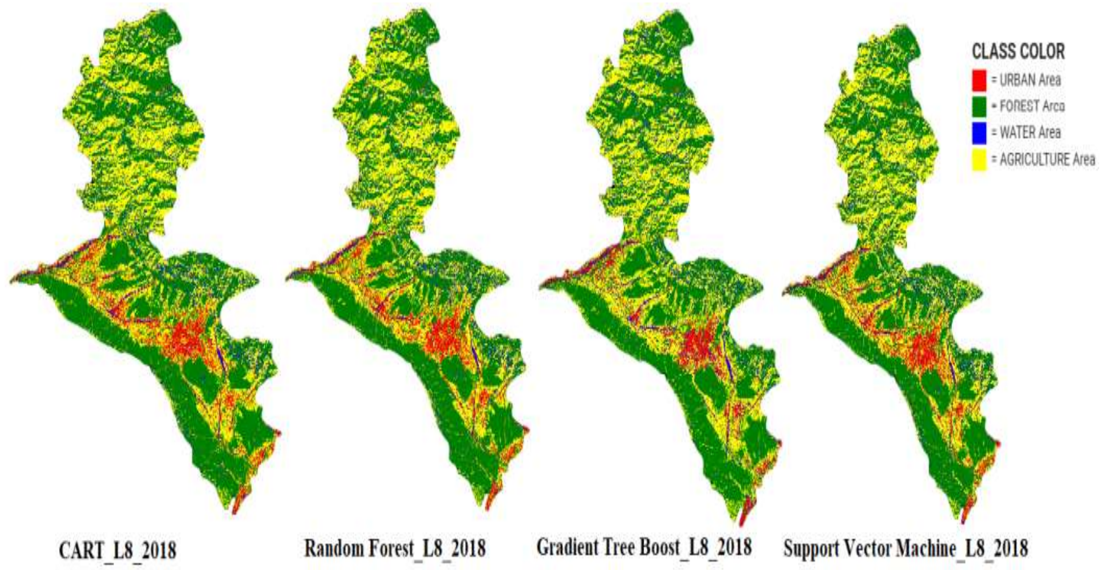

Figure 3.3 Output generated by Sentinel-2 data set, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2018

Figure 3.2 contains data generated over Sentinel dataset after processing classification algorithms and figure 3.3 contains data generated by Landsat 8 data sets, by the processing of these two data sets we found classification algorithm, those are based on ensemble technique are performed better. Here Random Forest algorithm has better accuracy in mostly all cases 89.90% and 86.28% using confusion matrix and kappa over sentinel data sets and 86.53% and 81.69% using confusion matrix and kappa over Landsat-8 data sets respectively.

Table 3.3 Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2018

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	87.15	83.51
Random Forest	89.9	86.28
GTB	89.79	85.41
SVM	78.79	77.19

Table 3.4 Accuracy of classification algorithm over both data sets using Kappa in the Year 2018

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	82.87	77.95
Random Forest	86.53	81.69
GTB	87.15	80.58
SVM	71.45	69.61

Total urban area 126 sq-km, forest area 1645 sq -km agriculture area 1158 sq-km, and water area 135 sq-km computed by Random Forest. The accuracy of all classifications over both data sets using the confusion matrix is listed in table 3.3 and using kappa is listed in table 3.4

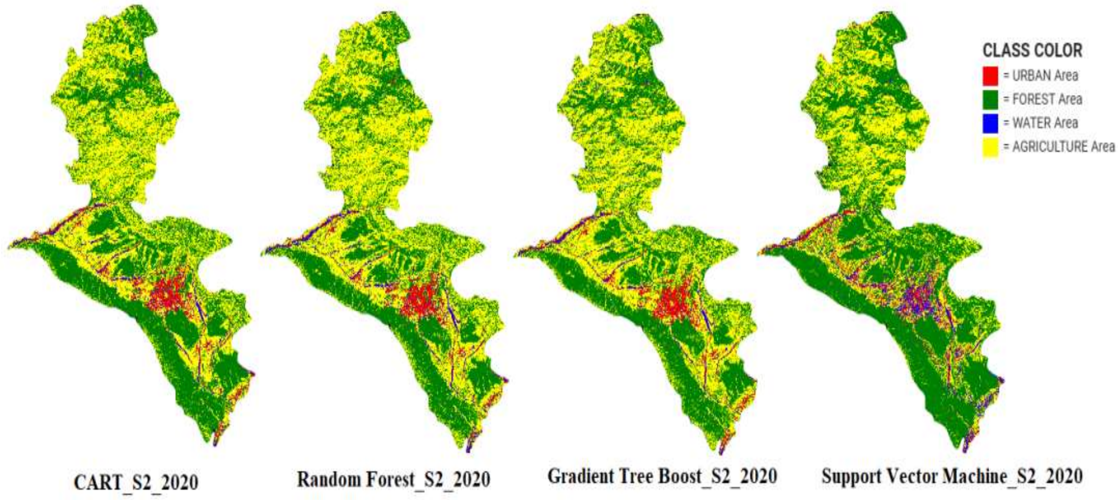

Figure 3.4 Output generated by Sentinel-2 data set, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2020

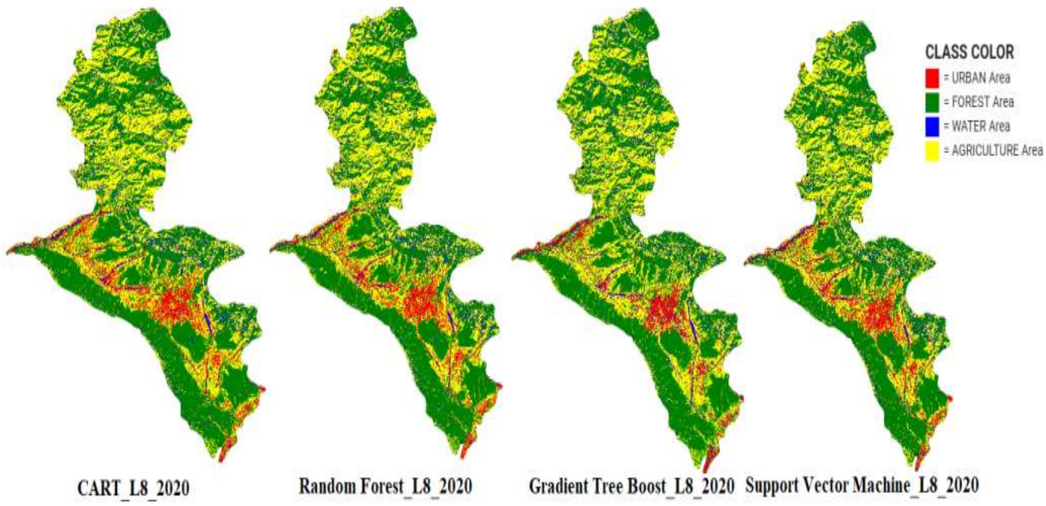

Figure 3.5 Output generated by Landsat-8 data sets, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2020

Figure 3.4 and figure 3.5 contain the output data generated over Sentinel-2 and Landsat-8 data sets in year 2020 respectively, by the processing of these two data sets we found there is a little bit of change in all land cover classes if we compare these output from previous output generated over year 2018 in figure 3.4 and figure 3.5 then it easily concluded that, forest area is decreases and agriculture area is proportionally increased, and some change also in other two land cover areas also, after computation total area computed by analysis of data over the year 2020 is urban area 145 sq-km, forest area 1545 sq-km agriculture area 1173 sq-km and water area 205 sq-km.

Table 3.5: Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2020

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	87.07	89.21
Random Forest	88.42	84.32
GTB	88.65	89.21
SVM	71.35	73.31

Table 3.6 : Accuracy of classification algorithm over both data sets using Kappa in the Year 2020

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	82.73	85.6
Random Forest	83.4	79.1
GTB	84.81	85.6
SVM	61.8	64.39

So for finding the exactness in computation we calculate the accuracy of our algorithms over both data sets Sentinel-2 and Landsat-8 listed in table 3.5 by confusion matrix and by kappa in table 3.6 Gradient tree Boost has a maximum accuracy of 88.65% and 84.81%

using the confusion matrix and 89.32% and 85.60% using kappa over Sentinel-2 and Landsat-8 data sets respectively.

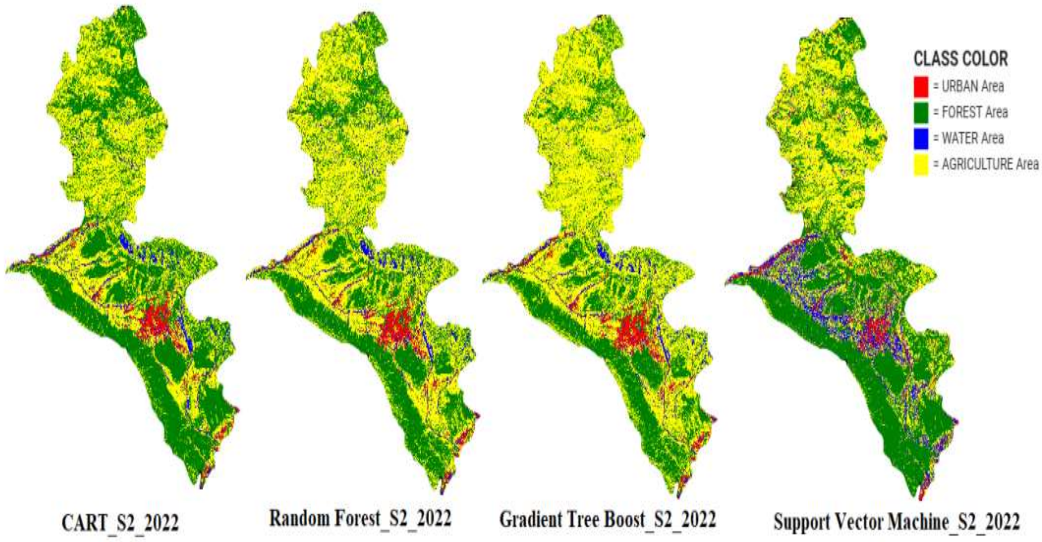

Figure 3.6 Output generated by Sentinel-2 data set with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022

Figure 3.6 & figure 3.7 present output data generated for the year 2022 over Sentinel 2 and Landsat 9 data sets respectively. On comparing the output generated by both data sets with the previous output we found urban land cover class have minor changes but the forest, water, and agriculture land cover class have a huge change, if we focus on figure 3.6 output generated by Sentinel-2 data sets and figure3.7, output generated by Landsat-9 data sets then it is clearly visible that areas that have forest land cover class in figure 3.5 and in figure 3.6 are now converted into agriculture or urban land cover class.

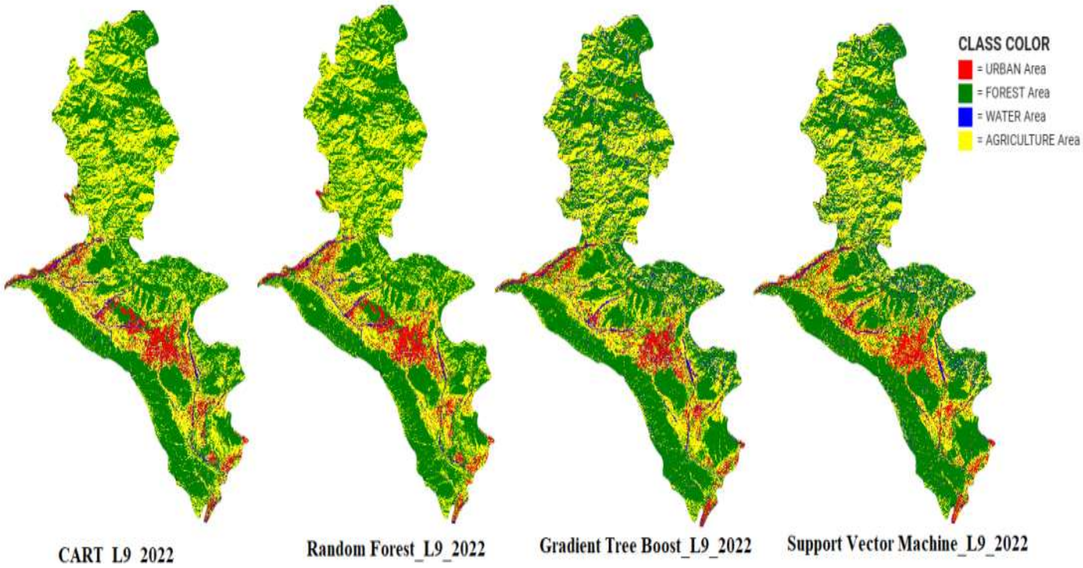

Figure 3.7 Output generated byLandsat-9 data sets, with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) in the year 2022

By the processing of these two data sets we found urban area capture 188 sq-km, forest area 1321 sq-km, agriculture area 1366 sq-km and water area 212 sq-km. In the Computation of land cover class, random forests have better accuracy compared to all other classification algorithms 92.48% and 86.56% using the confusion matrix and 89.94% and 82.05% using kappa over Sentinel-2 and Landsat-8 data sets respectively. A detailed description of accuracy computed over both data sets by confusion matrix and kappa is listed in table 3.7 and table 3.8 respectively. Figure 3.7 shows the comparative accuracy calculated by the confusion matrix over sentinel data sets, here random forest have maximum accuracy compared to the other three classification model, support vector machine have less accuracy, and the result obtained by the support vector machine also have a large number of overlapped data in above figures

Table 3.7 Accuracy of classification algorithm over both data sets using Confusion Matrix in the Year 2022

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	90.09	84.03
Random Forest	92.48	86.56
GTB	91.58	84.57
SVM	84.23	83.2

Table 3.8 Accuracy of classification algorithm over both data sets using Kappa in the Year 2022

Algorithm	Sentinel Data Set _ Accuracy	Landsat Data Set _ Accuracy
CART	86.78	78.71
Random Forest	89.94	82.05
GTB	88.76	79.29
SVM	72.27	77.56

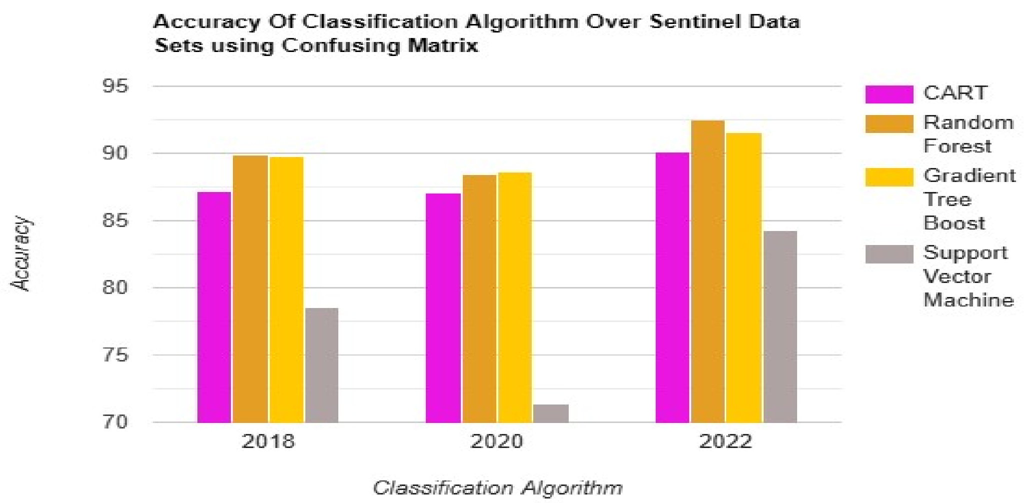

Figure 3.8 Comparative Accuracy calculated by Confusion Matrix over Sentinel Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022

Figure 3.8 shows the comparative accuracy calculated by the confusion matrix over Landsat data sets, here Gradient Tree Boost and Random Forest have maximum accuracy compared to the other two classification model, support vector machine have less accuracy compared to other classification models, in Landsat output also support vector machine have overlapped data in above figures. But in compare to sentinel data sets it to have better accuracy for support vector machine classification model.

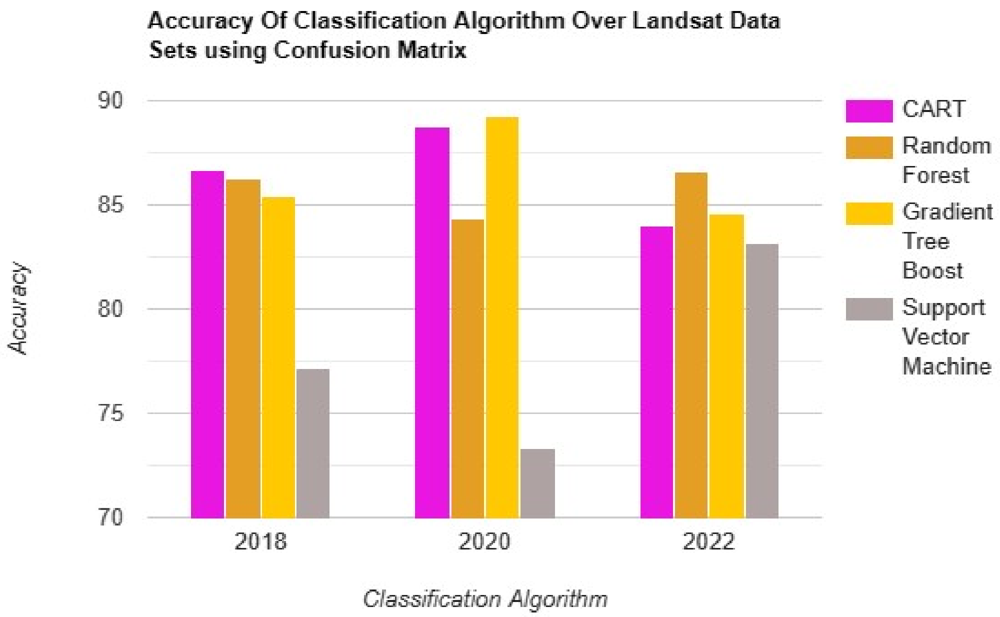

Figure 3.9 Comparative Accuracy calculated by Confusion Matrix over Landsat Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022

Later for validation of this output generated by the confusion matrix, we calculate accuracy by kappa result obtained by kappa is listed in table 3.2, table 3.4 and table 3.6. The comparative result obtained by sentinel and Landsat by kappa is shown in figure 3.9.

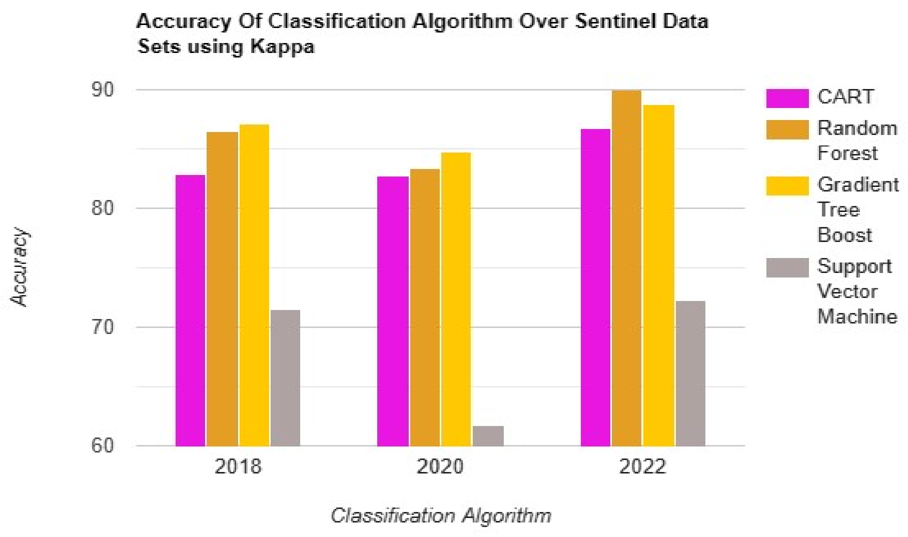

Figure 3.10 Comparative Accuracy calculated by Kappa over Sentinel Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022

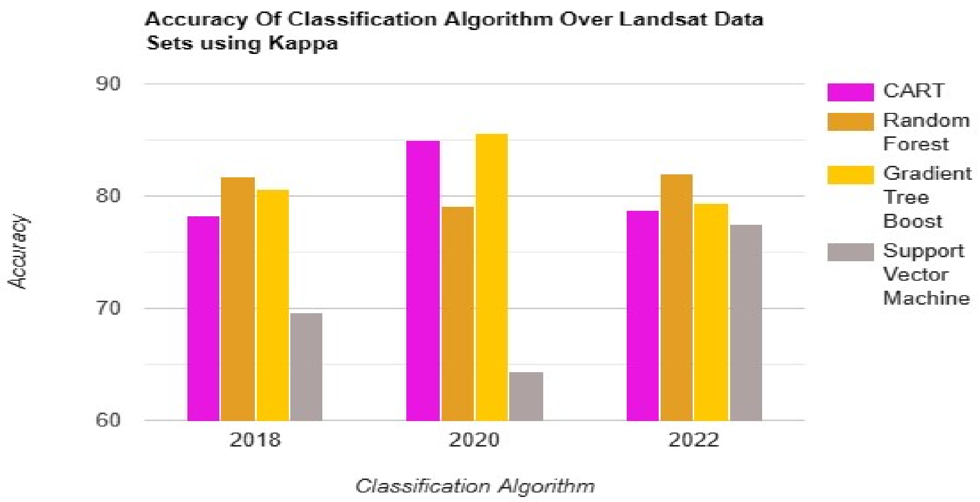

Figure 3.11 Comparative Accuracy calculated by Kappa over Landsat Data Sets with classification algorithm (Classification And Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) using Confusion Matrix in the year 2018, 2020, and 2022

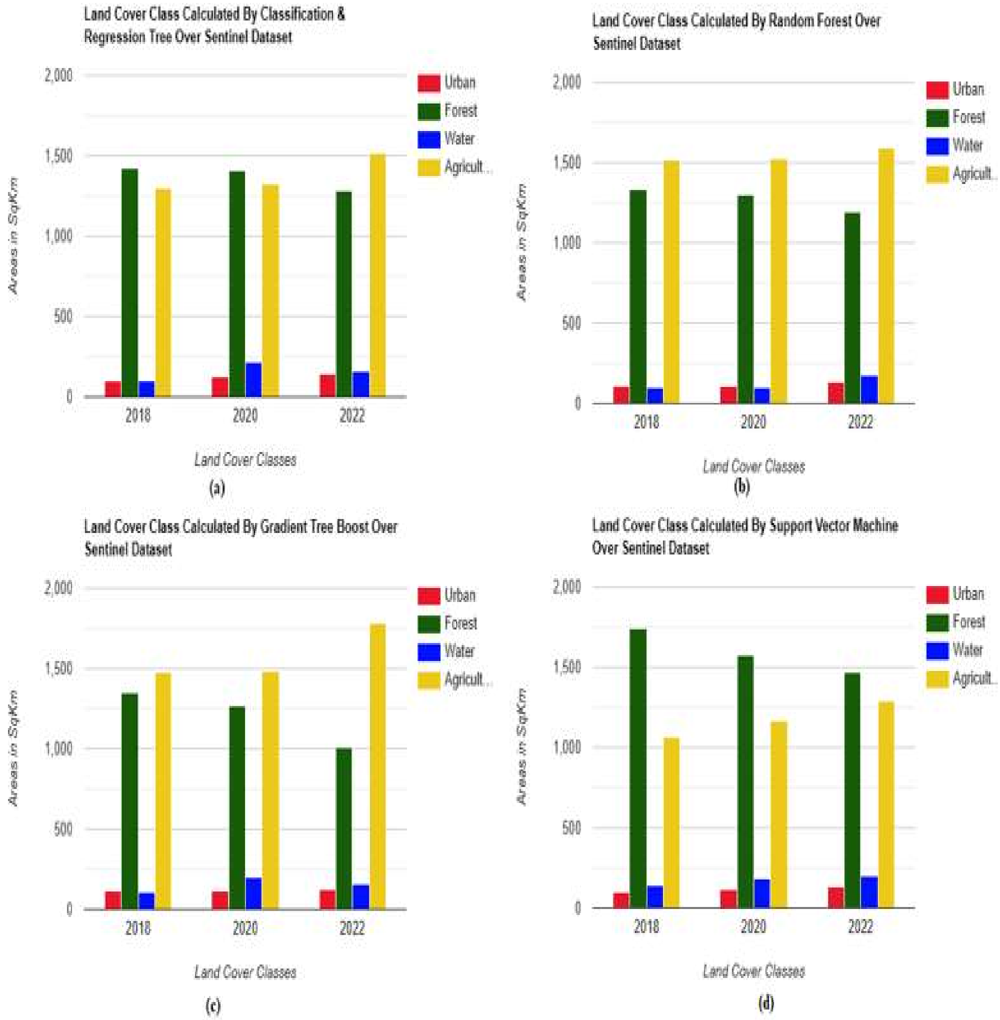

Figure 3.12 Land Cover Class Area computed by classification Algorithms (a) CART (b) Random Forest (c) Gradient Tree Boost (d) Support Vector Machine over Sentinel data sets in the year 2018, 2020, and 2022

Each classification algorithm has different accuracy so land cover areas calculated by the classification algorithm are different. Figure 3.11 shows land cover areas computed by classification algorithm in years 2018, 2020, and 2022 by sentinel data sets and figure 3.12 shows land cover areas over Landsat data sets in years 2018, 2020 and 2022. All classification models have different accuracy but have a common scenario that forest is being decreased and agriculture land cover class is increased. After analyzing land cover class in years 2018, 2020, and 2022, we found that lots of changes happen over the study area and classification algorithms random forest perform better in most of the computation, in some computation gradient tree boost also has high accuracy sometimes, after comparing these three algorithms with support vector machine we found output generated by support vector machine has a large number of overlapped data so it has less accuracy.

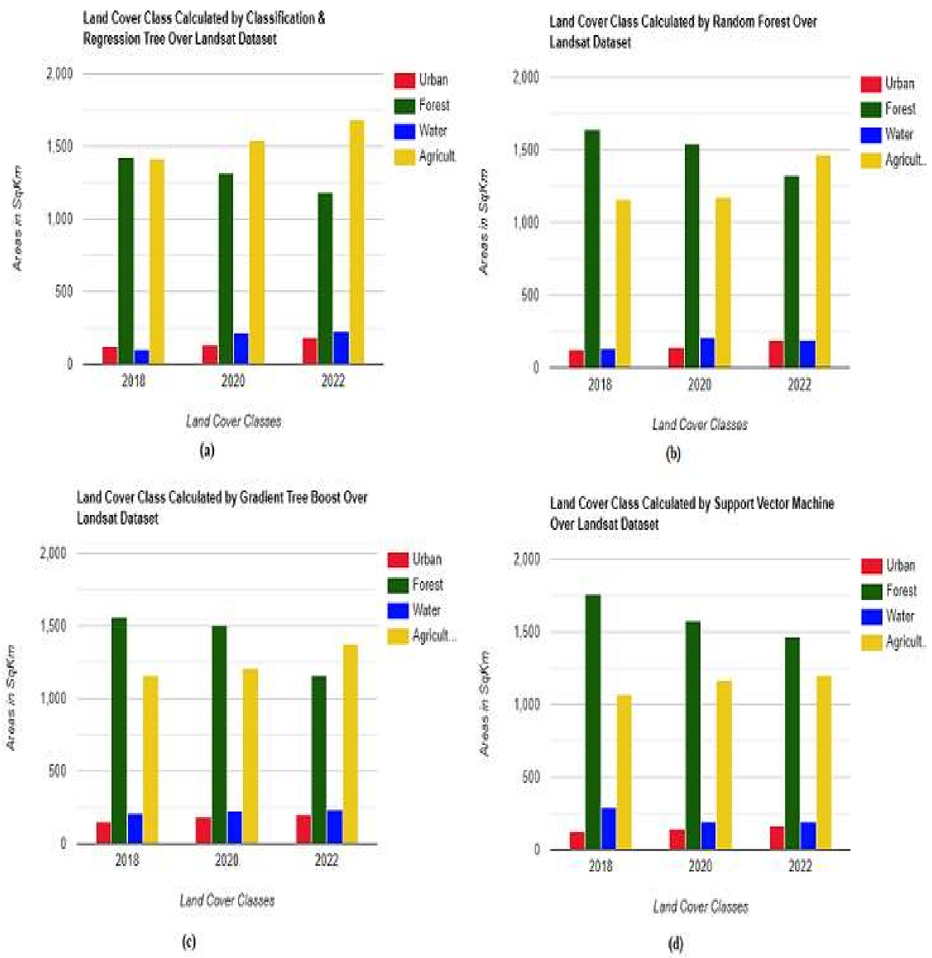

Figure 3.13 Land Cover Class Area computed by classification Algorithms (a) CART (b) Random Forest (c) Gradient Tree Boost (d) Support Vector Machine over Landsat data sets in the year 2018, 2020, and 2022

We have used here .jpg/.png image format for showing different land cover classes, anyone can see point-wise data description by following tif image link where all classification result is stored in .tif(Tag Image File Format). When we go with the point-wise analysis we found the study area changes at the same time on both data sets, on focusing output images of the support vector machine we found, algorithms have little bit different colour than Random Forest, Gradient Tree Boost, and Classification And Regression Tree classifiers in maximum output images but when we did deep analysis point by point then we find these changes are same at each geo-location.

3.5 Discussion:

Urbanization and population growth are the main factors influencing the LULC changes, another factor contributing to LULC change is the conversion of forest into agricultural land to supply the need for food grains. The increase of the built-up area, at the expense of agricultural land, vegetation cover, and open spaces, accounts for the majority of LULC change. Accurate computation is a very crucial task in finding change in land cover areas. Images used for categorization and the classifier employed in the computation are the two key factors that influence LULC change computation. Finding the best data sets and applying the best classifier over that can only provide better and more accurate results.

Comparison of Landsat-8 and sentinel-2 is performed by the number of researchers in different land cover areas [55], [56], [14] and they found some distinctive characteristics of both data sets with change in environment [55] performed his computation on boreal forest canopy cover and leaf area index and find sentinel-2 slightly better than Landsat 8 in the estimation of canopy cover and leaf area index and perform nearly same in term accuracy when computation is performed on the same band. [56] Also performed a comparison between Sentinel -2 and Landsat on the same boreal region and stated that sentinel-2 multi-spectral instrument (MSI) data can be recommended as the principal Earth observation data source in forest resources assessment. [14]. Performed comparison between these two data sets over the Brazilian Amazon region and find Sentinel-2 and Landsat are performed nearly the same in terms of accuracy. We have also performed a comparison between these two data sets over the same band and found sentinel-2 has higher accuracy compared to Landsat data sets in most cases, but accuracy changes with the change in year that clearly indicate data sets performance is dependent on environmental condition also.

Errors are present in any classification, estimation, or prediction [58], [59]. Comparison of the results of this study and those of earlier studies is not straightforward because the numbers and definitions of the vegetation classes differ by study. So outcome in such study calculated by accuracy, [58] performed a comparison of the parametric and non-parametric classifier over Dak Nong, Vietnam region on sentinel-2 data sets with very low accuracy (63% to 80%), [59] also perform his computation with the number of classifiers in Amazonian primary rain forest using Landsat data sets and find accuracy between 71% to 82% only. There are also no generally accepted limits on how accurate a classification should be to be characterized as reliable, because different users may have different concerns about accuracy. For example, they may be interested in the accuracy of a specific class or in the accuracy of areal estimates. In addition, multiple factors influence classification accuracy: image quality, classifier, image composition, number and details of classes, and sample size. [63] recommended that accuracies of 85% for mapping land cover are acceptable but other researchers stated that when attempting to distinguish among a large number of relatively detailed classes at a relatively local, large cartographic scale.

Consequently, in such applications, the use of the 85% target suggested by [63] may be inappropriate, as it may be unrealistically large. In fact, numerous researches has been carried out to choose the best accurate classifier, either from those being assessed simultaneously or from those being evaluated in other studies. The performance of a classifier is always dependent on the unique site characteristics, the type, and quality of the remotely sensed data, as well as the quantity and general characteristics of the classes of interest. As a result, such efforts fail to establish consensus. We have also analyzed four land cover classifiers' accuracy over four different land cover classes and computed accuracy between 85 % to 95% that are much above the [14]parameter. In all those four classifiers Random Forest performances are found higher in all years. Sometimes Gradient Tree Boost beats the Random Forest in terms of accuracy but the average performance of Random Forest is above compared to other classifiers.

The study area is highly vulnerable to natural causalities due to its geographical location. [13] perform his computation over the study area and find shifts between each land cover class from the year 2009 to 2019. This research identifies the land cover changes of Doon valley in different seasons using supervised classification. [35] Showing the application of remote sensing and GIS for LULC change using supervised classification in Basrah province, southern Iraq. [64]Perform LULC change in Zhejiang Province using a supervised classification approach. However, there have been lots of papers that reported the use of supervised classification to detect land cover changes such as Tirupati [65] India 2013, western Nile delta of Egypt 2011 [66]

Dipanwita [67] performed their computation over study area and stated that about 20% built-up growth between year 2000 to 2010. But researchers calculate the change in only built-up and vegetation (green) areas, they did not provide a change in most greenery areas to fewer greenery areas. If the built-up area has a growth of 20% then how much it affected the other areas, [54]also computed change in built-up land between the year 1987 to 2008 and provide a change in built-up land with direction considering city center as a mid-point but not elaborate about change in land cover class. [27],[24] also computed climate change and rising of temperature over study area and declare these change may create environmental damage over study area. So in our research, we computed changes in each class and find shifting from one class to another class.

By the analysis in figure 3.2 to figure 3.7 we found some changes in the land cover class from the year 2018 to the year 2022 these changes are shifted from forest to agriculture and agriculture to urban area. Total changes in land cover areas happening from January 2018 to December 2022 in Urban, forest, agriculture, and water area are 9.89%, 8.16%, 5.66%, and 13.25% respectively. As we already discussed that urban areas have a collection of built-up land, park, and residential, and non-residential built land. Here water land cover area is showing too much deviation but as we already discussed this is due to flooding happening in the year 2021 in the study area. The area of concern is changing in Forest and Agriculture land cover areas.

In previous studies [64] provide a change in urbanization in Urban Extension across Zhejiang Province using NPP-VIIRS Nighttime Light Data, [61]. Study over land use land cover[68] classification and state growth in urbanization by analyzing different land cover classes [20] also done their study on how land cover classes are changes day to day. So till now in the land use land cover classification model single point of computation is not provided by any researcher, estimating changes in land cover classes and confirming which data sources are more accurate for capturing these changes. The previous researcher either applied their research on calculating change detection in any study area or did their research for finding better accuracy models for classification or comparing two data sets. We have performed all three comparisons over a single point and by our analysis able to declare the best classification model for classifying land use land cover area with comparative better data sets. After analyzing outcome from figure 10, we find a lot of shift has happened in the forest area to urban and low vegetation areas. The outcome of this study clearly validates that the performance of data sets is dependent on the environmental condition and the classifier, the non-parametric classifier is better for landsat data and the parametric classifier is better for sentinel data sets. In regions where forests and dense vegetation are present, sentinel data provide better accuracy. For computation over built-up land, Landsat data sets provide comparative better accuracy.

3.6 Conclusion:

The goal of this study was to compare the two existing data sets applying four machine learning techniques and analyze changes in the different land cover classes for each month of 2021 in the Dehradun study region. We find here land cover classes are shifted from one to another from sets of months due to human conduct activity or natural activity. Figure 10 shows the changes that happened in each land cover class. Classification performance index table 3.2 to table 3.7 easily shows the accuracy as well as wellness of each classification algorithm. Output describes that each data sets have some unique feature and behave according to that feature with a different classification algorithm. Landsat (USGS) and Sentinel (COPERNICUS) are two different data sets providers and provide images of different behaviour Landsat images have thermal infrared sensors but Sentinel does not have these bands. In most cases (Gradient Tree Boost) GTB, Classification and regression tree (CART), and Random Forest(RF) performed slightly better than Support Vector Machine (SVM), Compared to SVM, decision tree-based algorithms are better at handling co-linearity and categorical data, it constructs hyper-rectangles in the input space to solve the problem, SVM uses the kernel method to address non-linear problems. Differences in CART, GTB, and RF accuracy estimates were generally statistically insignificant; the accuracy of Sentinel-2 and Landsat-8 is almost the same in all the seasons while the classification algorithm which uses the kernel method to solve the non-linear problem has higher accuracy for sentinel-2 data set. There are the following outcomes generated for future research also: We have covered four land cover areas and found overlapped data after classification, especially in forest and agriculture land cover areas. So in the future, we need to do a close observation of these land cover areas by using different vegetation analysis techniques. If a user uses the aforementioned classification approach to compute land use land cover change, they don't need to adopt a separate classification algorithm; alternatively, they only need to choose the accurate classification model and data set for better accuracy and perform their analysis.

Chapter 4:

Comprehensive Assessment of Long-Term Forest Transformation by Analyzing Spatial-Temporal Vegetation Indices from 2000 - 2022

4.1 : Abstract

The world's ecosystem and environment are rapidly deteriorating with an increase in the depletion of forest conditions due to forest fires. In recent past years, wildfire incidents in Sikkim have increased due to severe climatic changes such as turbulent rainfall, untimely summers, extreme droughts in winters, and a reduction in the percentage of yearly rainfall. Forest fires are one of the numerous kinds of disasters that impose disastrous changes on the entire environment and disrupt the complex correspondence of the flora and fauna. The research's goal is to examine the vegetation indices based on different climates to know why forest vegetation is decreasing day by day from 2000 to 2023. The frequent changes in forest vegetation are extensively studied by using satellite images. This data has been collected by three satellites Landsat-5, Landsat-7, and Landsat-9 on different vegetation indices NDVI, EVI, and NDWI. East Sikkim area is chosen to compute forest vegetation indices based on the heap's landmass this region is unexplored yet and also studied about the forest changes by using different spatial temporal indices in the range of the entire district in the future. The authors of this paper have used Landsat multi-spectral data to assess changes in the area of vegetation in a sub-tropical region like a dense forest region in east Sikkim. The analysis depicts space images, computes vegetation indices (NDVI, EVI, NDWI), and accomplishes mathematical computation of findings. The proposed method will be helpful to discuss the variance of vegetation in the entire East Sikkim region at the

time span of 2000–2023. In the analysis, we find that mean and standard deviation values change over the years in all indices. Later, we also calculated changes by using a classification model and find a total 10% change in forest areas in approximately 22 years.

4.2: Introduction

It is found that climate is very much affected by global warming throughout the world. Spectral Indices detected different factors like rainfall and temperature within 1 to 15 years [69]. Based on the increasing extremes caused by human-induced climate change, as well as the limited progress made towards finding climate change solutions, the National Academies of Science and Engineering recently recommended that the USA develop a trans-disciplinary research program into proposed climate intervention techniques. [69]. Earth quacks also deteriorate the vegetation area because of their frequent occurrence as earlier we used to hear about earth quacks occurring in 4 to 5 years but now every month it probably happens. Due to earthquakes lots of forest decaying problem arises like plant community shift, Species loss, and productivity reduction of alpine grasslands [47]. Flood is also one crucial factor to decrease the vegetation area but it is also found that initially, floods affected crops but later crop productivity and fertile land improved and resulted in dense vegetation area[70]. By mapping vegetation cover before and after floods, spacecraft images, rainfall data, a tool used for analyzing the geographical area, and a rain gauge were used to evaluate post-flood loss or benefit. The deforestation and degradation of forests alone contribute between 20% and 25% of global greenhouse gas emissions [32]. District-wise Sikkim climate data is analyzed over the period 1901 to 2007 to predict rainfall, precipitation, and temperature followed by mean 17.82 mm per day, Standard Deviation 3.55 mm per day.

The environmental influence on vegetation increases as more regions of the world become forested. Inside the city limits, there are forests, which causes changes in the vegetation of the forests. Computing the quality of forest areas and the surrounding environment and making decisions to ensure the population's sanitary and environmental safety depends on understanding where these changes occur, in what quantities, and in response to what variables [71].

The dynamics of forest vegetation may be studied in great detail using satellite photos over Google Earth Engine [72]. Satellite photos enable us to get factual data about the temporal and spatial variations in vegetation. The commonly uses type of satellite data is Landsat, Sentinel, and MODIS from various very famous data archive providers like USGS, Copernicus Programme, NASA, etc. [61]. Landsat itself has various development, from Landsat 1 to Landsat 9. Sentinel also has various versions from Sentinel 1 to Sentinel 5. [73]

A number of methods have been employed to determine the area occupied by vegetation [74]. These methods involve classifying the outcome based on supervised and unsupervised learning , automatic image processing, the generation of index pictures, as well as visual interpretation [47]. The dynamics of plant cover can be analyzed using spectral & temporal indices obtained from time-lapse images.

The research's goal is to use remote sensing data from the year 2000 to the year 2023 to examine the forest vegetation in the east Sikkim forest area. Principal research goals: (1) Building a database of Landsat satellite photos for the years 2000 to 2023; (2) Creating an algorithm along with software development acquired from several vegetation indices. (3) The identification of the primary mechanisms that brought about the observed alterations in the east Sikkim forest areas.

4.3 : Related Work

The most common origin for the remote sensing data of Earth for a variety of examinations is Landsat pictures. The benefit of Landsat data is related to the policy of free picture access and the continuity of observations over a 50-year period. As per the analysis of spatiotemporal characteristics, the moderate forest conquered approximately 46% in 1985 and 57% in 2005, and 58% of the total land is occupied by open forests which were a replacement for these [75]. In addition to this, we have analyzed spatial and temporal indices in the East Sikkim area. The majority of India's forests are degrading due to forest fires. In East Sikkim, forest fire is a frequent process as the water stress level is very high in the summers. It is a tedious task to figure out the statistical data on the occurrence of forest fires in a year, but statistics cleared that estimated that 33% of a few states and more than 90% of other states are exposed to forest fires annually [76], [77]. The burnt areas could be easily seen in the SWIR band when using band(3 2 1), and band(4 3 1)[78].

Landsat data are useful for tracking forest regions. Many different techniques are used for analyzing and monitoring the Landsat input data. Several image processing techniques created in remote sensing are used in extracting area-covering details with the help of satellite images.

Using Landsat to map forests, presents a variety of challenges [79] [80][81]. Landsat images of forested areas typically include a combination of data about anthropogenic items and vegetation in their pixels. When recognizing forest areas, more vegetation cover from the image must be retrieved.

To distinguish between forest and non-forest regions, one method uses spectral vegetation indices such as the Normalized Difference Vegetation Index and vegetation index sensitive to the water content of plants normalize difference water index. Another spectral technique makes the assumption that forest pixels are linear mixtures of the three common land cover, vegetation, impervious surface, and soil components (the so-called V–I–S model). In the paper, it is suggested that mapping of the forest area be done using a mix of spectral and spatial data. The technique comprises these two distinct categories of elementary coverage classes based on pixels and segments (segment-based).

4.4 : Study Area

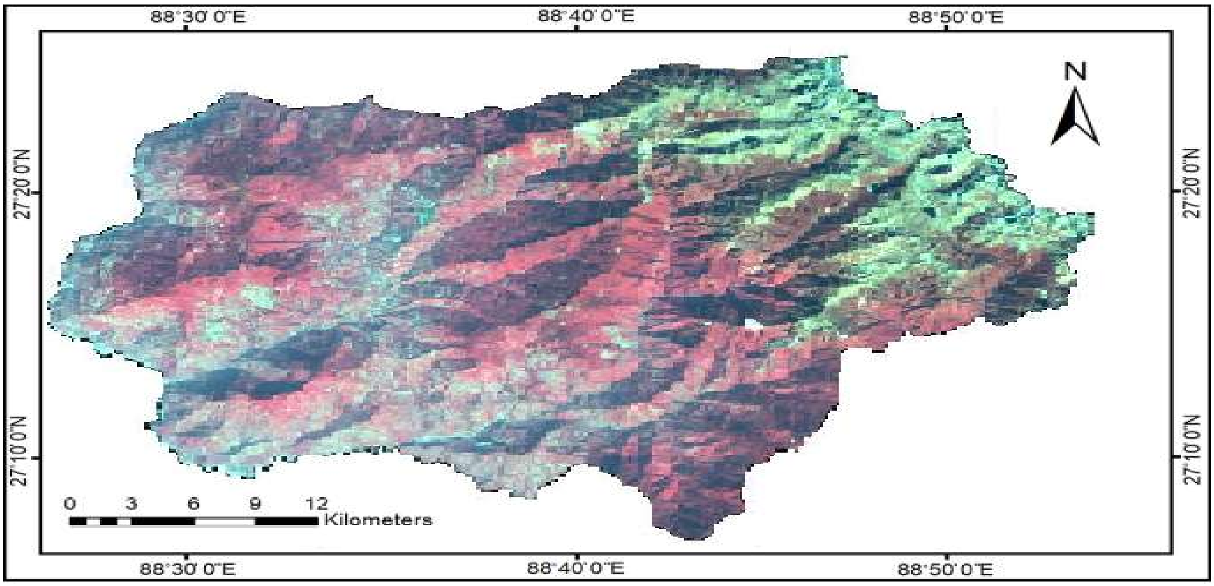

Figure 4.1 Geographical Location of Study Area

Sikkim is one of the largest forest heaps in the northeast of India. Sikkim state covers a total area of 7096 sq Km. The geographical area we have targeted as the study area is the East Sikkim region, which is approximately 964 sq. km in size and located at 27.3084° N, 88.6724° E in figure 4.1.

As per the Forest Management Department, 14.44 % area of Sikkim is covered under scrub and alpine pasture, and 29.5 % area is occupied by perpetual snow cover. Remote Sensing Data for the year 1988 depicts that the vegetation area for crops which may be Terraced/Semi Terraced is 604.85 sq. km and this cropland is mixed with dense forest of capacity 603.34 sq. km. The district area is 173.19 sq km which is 2.44 % of the total area. As stated by the Forest Survey of India (FSI) [82], the reported Forest State covers an area of 5841.39 sq km which is equal to 82.32% and 0.8% of the whole nation's forest region [83]. According to the State of Forest Report of the Forest Survey of India, Ministry of Environment & Forest, Government of India, the status of Forest cover evaluation is gradually increasing which was 2756 sq km in 1987 and 3262 sq km in 2003. In 2021, the number of trees and forests in India covered 80.9 million hectares, which is equal to 24.62% of the country's total land area. Areas that are part of a biosphere reserve's buffer zone are excluded from the Protected Area Network [61] [84].

The amount of area covered by the protected Area Network of State is 2177.10 sq km. (i.e.30.68% of the entire geographical area) whereas the amount of land covered by the protected Area Network and biosphere reserve in the State is 3013.10 sq km (i.e.42.46% of the entire geographical area). There is mainly five types of forest present in East Sikkim: Wet temperate forests find generally in hilly areas, subtropical or moist broad-leaf forests, which are areas of forests where half of the world species are living in different zones, moist mixed forests are types where greenery increases or decreases with the season these are also known as dry deciduous forest, other types of forest are conifer forest and sub-alpine forest, conifer forests are also deciduous forest but the property of these forests are these are always green but sub-alpine forests are primary factors of nature and environmental disturbance these are basically prone to fire and 80% forest fire incident are happened due to these indices in India these are basically found at eastern middle Himalayas. Sikkim is India's greenest state but with time and change in climate and environmental conditions, we find that there are high changes are occur from one type of forest to another type of forest which also leads to deforestation.

4.5 Method & Data:

Landsat data has been a reliable choice for forest classification for many years. It provides moderate spatial resolution (30 meters) and captures imagery in several spectral bands, including the near-infrared (NIR) band, which is particularly useful for assessing vegetation health. The historical archives of Landsat data are extensive,

Table 4.1 Band Uses for calculating changes in Forest Cover

Landsat Datasets	Blue	Green	Red	Near Infrared	SWIR 2
Landsat 5 TM (Band number: Wavelength)	B1: 0.45–0.52	B2: 0.52–0.60	B3: 0.63–0.69	B4: 0.76–0.90	B7: 2.08–2.35
LANDSAT 7 ETM+ (Band number: Wavelength)	B1: 0.45–0.52	B2: 0.52–0.60	B3: 0.63–0.69	B4: 0.77–0.90	B7: 2.08–2.35
Landsat 9 OLI (Band number: Wavelength)	B2: 0.45–0.51	B3: 0.53–0.59	B4: 0.64–0.67	B5: 0.85–0.88	B7: 2.11–2.29

making it a valuable resource for long-term forest monitoring and change detection. Landsat is suitable for larger forested areas or regions with diverse land cover types.

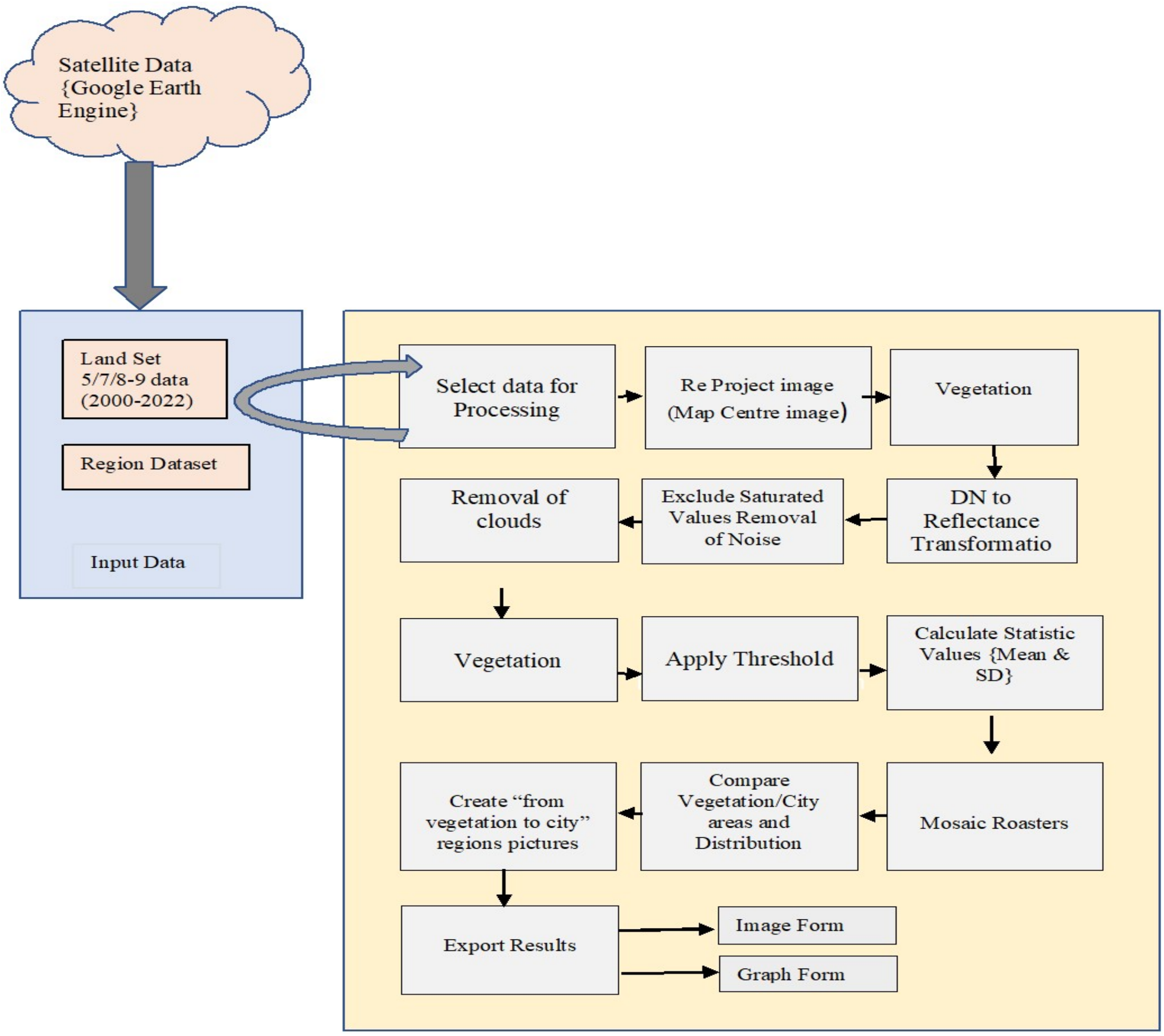

Figure 4.2 Methodlogy Used in Processing of data for finding Change

The Landsat program has played a crucial role in Earth observation for decades, with different satellites in the series carrying various sensors equipped with distinct spectral bands. Landsat 5, which operated from 1984 to 2013, featured five primary spectral bands. The BLUE band (Band 1) centered around 0.45-0.52 micrometers captured blue light, while the GREEN band (Band 2, 0.52-0.60 micrometers) and RED band (Band 3, 0.63-0.69 micrometers) recorded green and red light, respectively. The Near Infrared (NIR, Band 4, 0.76-0.90 micrometers) and SWIR 2 (Shortwave Infrared 2, Band 5, 1.55-1.75 micrometers) bands were essential for vegetation health assessment, land cover classification, and the study of water bodies and geology.

Landsat 7, launched in 1999, inherited similar spectral bands from its predecessor, Landsat 5. The BLUE, GREEN, RED, NIR, and SWIR 2 bands continued to serve as vital tools for remote sensing applications, capturing light in the blue, green, red, near-infrared, and shortwave infrared parts of the electromagnetic spectrum, respectively. Landsat 7 also introduced the Panchromatic band (Band 8), centered around 0.52-0.90 micrometers, providing high-resolution black-and-white imagery suitable for a wide array of applications.

Landsat 9, launched in 2021, carries on the legacy of its predecessors with spectral bands that closely mirror those of Landsat 5 and Landsat 7. These include the BLUE, GREEN, RED, NIR, SWIR 2, and Panchromatic bands. Landsat 9's sensors, however, feature enhancements in terms of radiometric and geometric calibration. These bands continue to be indispensable for monitoring land cover changes, vegetation health, land use classification, and geological assessments.

For analysis of change in East Sikkim, we are using google earth engine (GEE) satellite-based planetary tool [72]. It has a collection of many satellites based real-time data sets like Landsat, Sentinel, and MODIS as well as provides users, a tool (code editor) for analyzing these data. Here we are using Landsat data for finding changes in the forest area as well as East Sikkim from the year 2000 to 2023. Landsat has a collection of data from 1972 to 2023. Landsat has different versions based on time of availability, spatial resolution, and wavelength. Till now USGS has launched nine versions of Landsat data sets from Landsat

1 to Landsat 9. We are using, data from the Landsat-5 (Thematic Mapper), Landsat-7 (Enhanced Thematic Mapper Plus), and Landsat-9 (Operational Land Imager-II) satellites to examine the dynamics of the study area's forest vegetation. Many scenes are present in the chosen location. The work uses five bands (Short Wave Infrared, Green, Red, Near Infrared and Blue), which have a spatial resolution of 30m. Landsat uses a worldwide reference system (WRS) that catalog Landsat data by path and row. Table 4.1 represent used datasets and band for analysis of change in the study area. Figure 4.2 shows the technique used in research for the analysis of change using different data sets and vegetation indices. Later for verification of the result, it also calculates changes in land cover areas of the study area, East Sikkim using the supervised classification model Random forest.

4.6 Result:

Low vegetation and high vegetation land cover class contain all the land cover areas which having some greenery or dense forest, for proper differentiating between these two land cover classes we are using Normalize Difference Vegetation Index (NDVI) [85] Normalized Difference Water Index (NDWI) [86], and Enhance Vegetation Index (EVI) [87], After calculating the ground reference point we observed the study area and calculated its daily average temperature and rain and peak months as August to October rain goes above 90mm. From the observation, we find the average temperature is high in the month between May to July and the average rain is maximum in the month of August and September.

$$\text{NDWI} = \frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{R}} \qquad \text{...................1}$$

$$\text{NDWI} = \frac{\text{Green} - \text{NI}}{\text{Green} + \text{NIR}} \quad \text{....................2}$$

In Landsat 8 and Landsat 9, Enhanced Vegetation Index is calculated as

$$\text{EVI} = \frac{2.5*(Band\ 5-Ba \quad 4)}{Band\ 5+6*Band\ 4-7.5*Band\ 2+1} \quad \text{....................3}$$

Vegetation Indices (VIs) combine surface reflectance at two or more wavelengths to emphasize a specific characteristic of vegetation. They are created using vegetation's reflective qualities. Each VI is intended to calculate a specific characteristic of the vegetation. Every VI needs accurate reflectance readings from multispectral or hyperspectral sensors. The spectral bands sampled in the input dataset dictate which VIs can be generated on that dataset. A VI is accessible for the dataset if it has all the spectral bands necessary for that index. Some place is used to calculate density by using the normalized difference vegetation index.

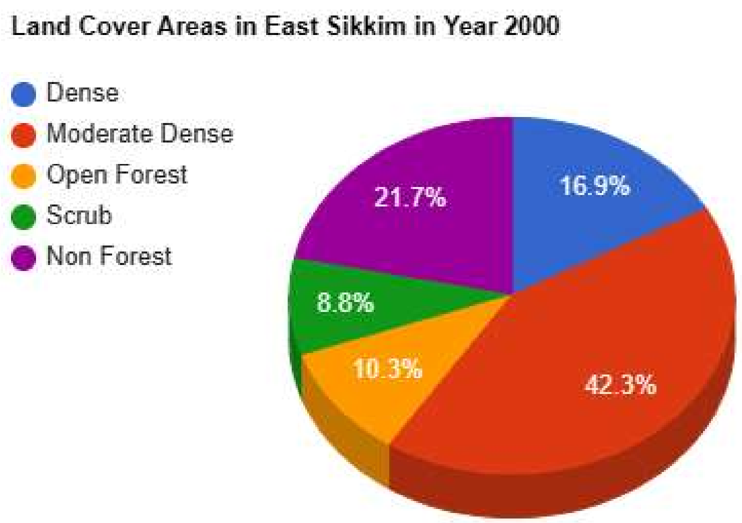

Figure 4.3 :Forest Cover Area of East Sikkim in the Year 2000

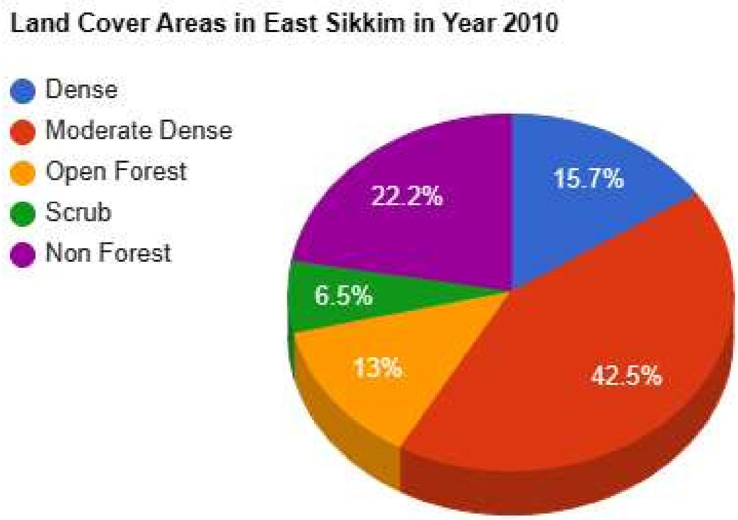

Figure 4.4 :Forest Cover Area of East Sikkim in the Year 2010

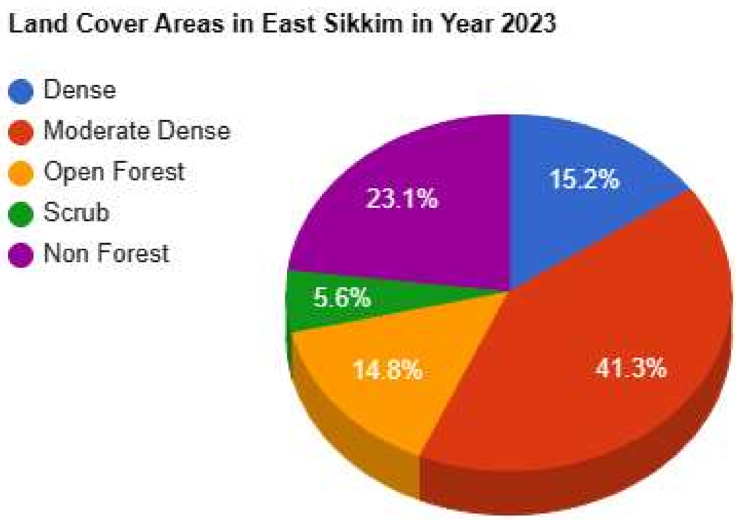

Figure 4.5 :Forest Cover Area of East Sikkim in the Year 2023

NDVI is frequently used in agriculture, forestry, and the environment to track the development and well-being of vegetation as well as to spot stressed or damaged areas. In addition to mapping and categorizing different vegetation types, NDVI values can be used to track changes in vegetation cover over time

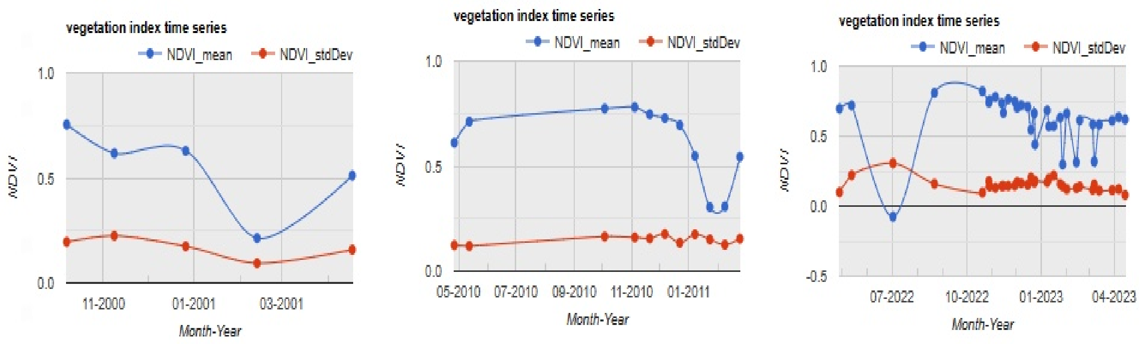

Figure 4.6 : Change in NDVI between the Year 2000- 2023

. Figure 4.6 shows the change occurring in NDVI value on the study area over the period of time. Every time it is changed with time but in the year 2022- 2023, this data is changed more number times.

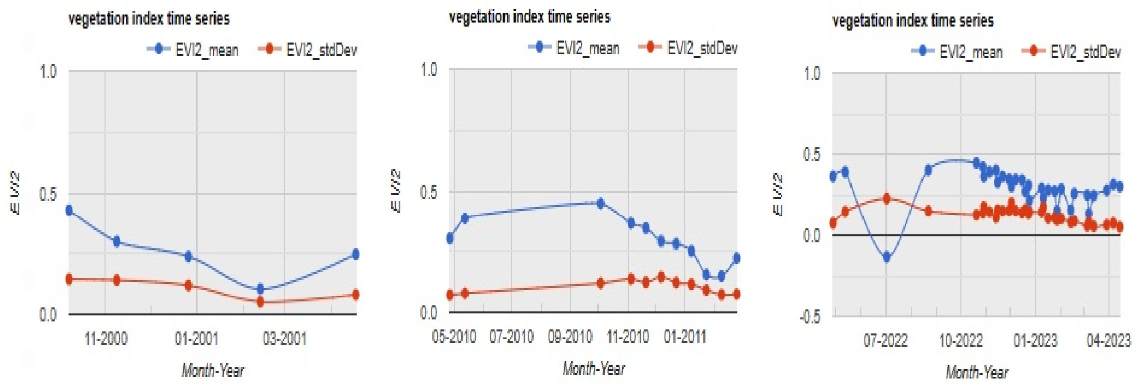

Figure 4.7: Change in Enhanced Vegetation Index between the Year 2000- 2023

Enhanced Vegetation Index (EVI) is also a mechanism for measuring vegetation greenness that is similar to the Normalised Difference Vegetation Index (NDVI). EVI, on the other hand, compensates for some atmospheric factors and background noise from the canopy and is more sensitive in regions with dense vegetation. EVI also provides producers the ability to precisely compare data and monitor changes. These comparisons are quick and simple thanks to the use of our vigor items scaled to an absolute standard. Figure provides a change in the study between the years 2000 to 2023. EVI values are changed rapidly between these years. Later for verification purposes, we calculate the change in the study area using the ensemble-based classification algorithm random forest and find nearly 10% of land cover classes are changing.

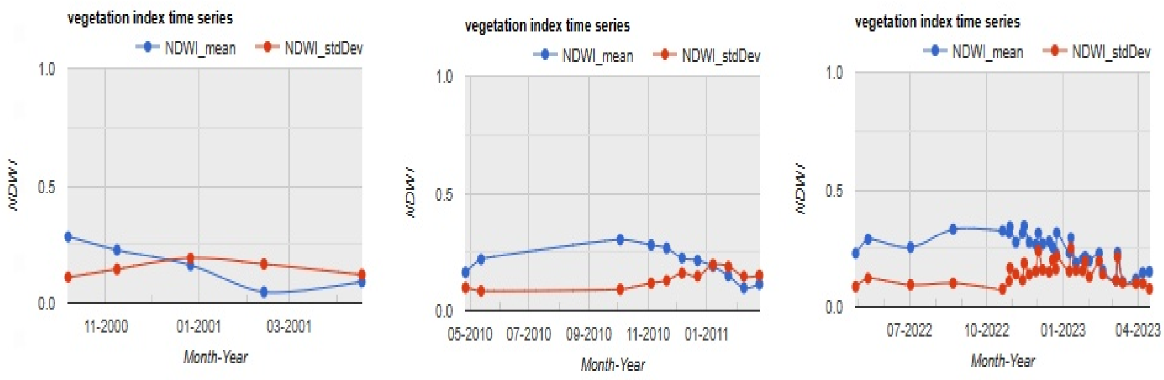

Figure 4.8:Change in NDWI between the Year 2000- 2023

Later we calculate the change in water content over vegetation in the study area using the normalized difference water index (NDWI) . The Near-Infrared (NIR) and Short Wave Infrared (SWIR) channels are used to create the Normalised Difference Water measure, which is a satellite-derived measure. While the NIR reflectance is influenced by changes in leaf internal structure and dry matter content but not water content, the SWIR reflectance reflects changes in both vegetation water content and the spongy mesophyll structure in vegetation canopies. The accuracy of determining the water content of vegetation is increased when the NIR and SWIR are combined because they eliminate changes brought on by the internal structure and dry matter content of leaves. The spectral reflectance in the SWIR region of the electromagnetic spectrum is substantially governed by the quantity of water present in the interior leaf structure. Therefore, leaf water content has a negative relationship with SWIR reflectance. Figure 4.6, figure 4.7 and figure 4.8 shows the change that arises over the study area mainly in the forest region.

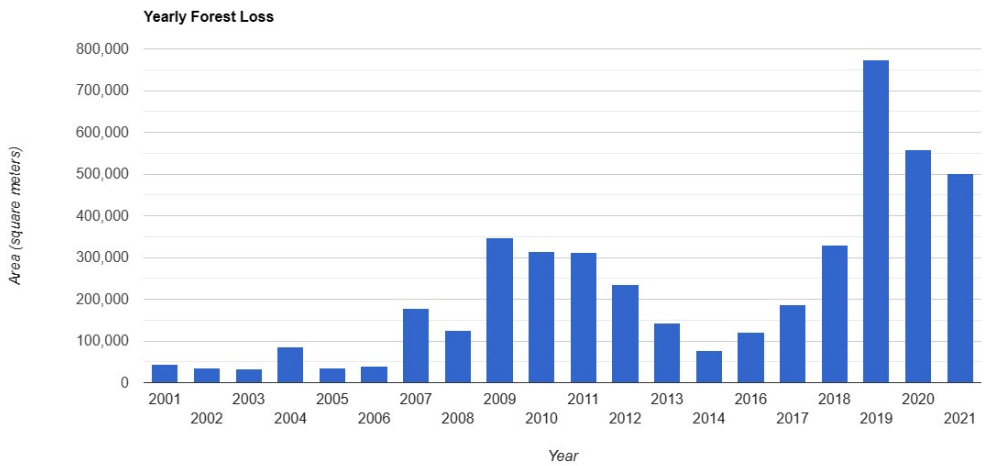

Figure 4.9: Yearly Forest loss between the years 2000- 2023 over the study area

From figure 4.3, figure 4.4, and figure 4.5, it is quite clear that deforestation happened over the study area. Still, due to good climate and environmental conditions, this deforestation is not converted into non-forest areas. From the calculation we find about 10% of forest loss present over the study area figure 4.9, shows the yearly forest in square meters from the year 2000 to the year 2022. Here if we focus on the output generated by vegetation indices figure 4.6, figure 4.7 and figure 4.8, overall mean and standard deviation values are not changed but the pattern of these changes are frequent in the current year

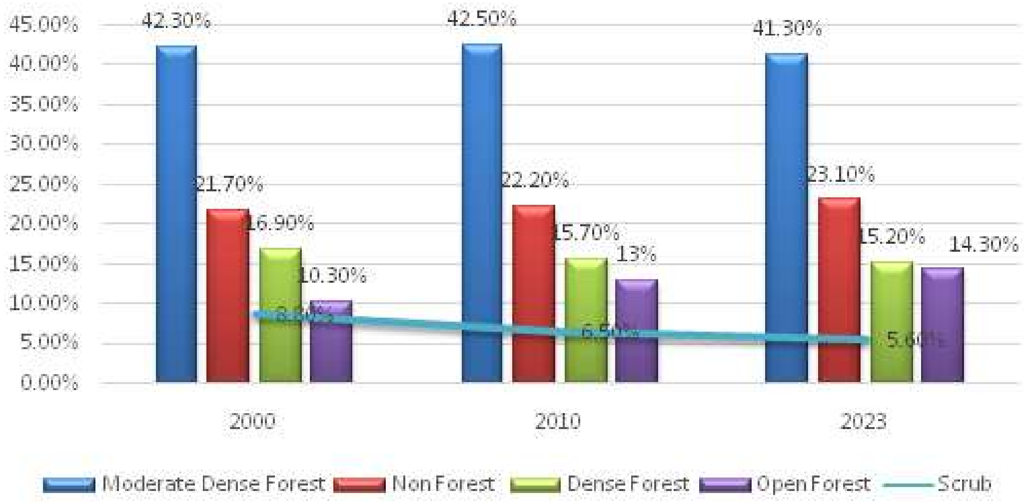

Figure 4.10: Change in LULC areas between the years 2000- 2023

By the above findings, it is figured out that moderate dense forest is higher in comparison to the non-forest area, dense forest area, open forest area, and scrub which is 42.3 %, 21.7 %, 16.09%, 10.3 %, and 8.8% in the year 2000 respectively and it is also analyzed that moderate dense forest area is decreasing year by year due to forest fire which is 42.5 % in the year 2010 which gradually decreasing after decades too. It became 41.3 % in 2023 and for other land areas transformation can also be seen. Dense forests were 16.9 % in the year 2000 which decreased by 15.70% in 2010 and 15.20 % in 2023. The non-forest area is decreasing by these fire incidents and it is clearly shown by the graph that it was 21.7 % in 2000 which increased to 22.20 % in 2010 and 23.10 in the year 2023. Open forest area is also increasing by the incident of forest fire while scrubs are continuously decreasing. By the figure 4.10, it is concluded that open forest area is gradually increasing i.e. 10.3% (2000), 13 %(2010), and 14.3 % (2023). In this proposed study area it is observed that we can generate land cover area loss or affected parameters by forest fire by using this developed tool or framework on any land with approximately similar spectral indices.

4.6 : Conclusion

The ability to perform high-frequency time series analyses using new-generation multi-spectral sensors aboard the Landsat 5, Landsat 7, and landsat 9 satellite platforms opens up previously unheard-of possibilities for multi-temporal change detection studies on phenomena with significant dynamic behavior (for example, high-frequency mapping for disaster management) or on regions with recurring cloud cover issues. These new sensors' radiometric properties, while similar, are not equivalent, and this might result in noticeable variations in the radiometric amounts that are received. the spectral signatures identified in the computation of forest transformation between the years 2000 and 2023 in East Sikkim represent vital components of remote sensing analysis. Spectral signatures are essentially graphical representations of how different land cover features reflect and absorb electromagnetic radiation at various wavelengths across the electromagnetic spectrum. These signatures are instrumental in classifying and monitoring land cover changes over time. The spectral signature for the year 2000 reveals that the forested areas in East Sikkim exhibited distinctive characteristics associated with healthy vegetation, characterized by strong reflectance in the Near-Infrared (NIR) band, a hallmark of flourishing vegetation, and considerable reflectance in the Red band. However, as we move forward in time to 2023, the spectral signatures divulge specific changes. Notably, there's a decline in NIR reflectance, which can be attributed to a decrease in vegetation health, potentially due to deforestation, land use shifts, or natural disturbances. This alteration in NIR reflectance is a unique signature of significant change. Simultaneously, there's a notable increase in Red band reflectance, signifying the rise in non-vegetated surfaces or modifications in land cover. Additionally, variations in the Shortwave Infrared (SWIR) region of the spectral signature point to shifts in land surface properties. Furthermore, the results obtained through this assessment have been validated and enriched by referencing various reports and surveys conducted by India's forest and environmental authorities and verified by using supervised classification results and global forest loss computation technique. Long-term changes in land use and land cover classification in East Sikkim have been marked by a gradual but significant transformation of the region's landscape. Urbanization has led to the expansion of built-up areas at the expense of natural land covers such as forests and grasslands. This has not only reduced green spaces but also increased impervious surfaces, contributing to higher temperatures and decreased biodiversity. Deforestation, driven by agricultural expansion and timber extraction, has further eroded natural land covers. This loss of forests, in particular, has had far-reaching consequences, impacting local wildlife, altering the hydrological cycle, and making the region more susceptible to landslides. Agriculture expansion, including terraced farming, has transformed traditional land covers like meadows and natural grasslands into croplands, affecting soil quality and water resources.

Chapter 5:

Detecting Short-Term Forest Cover Changes Caused by Fires in Sentinel-2 Data using Spectral Indices and Machine Learning Algorithms Over Ernakulam Fire Incident

5.1 Abstract

For the ecosystem to maintain a balance between the social and environmental spheres, forests play a crucial role. The greatest threat to forests for this significance is fires and natural disasters caused by several factors. It is crucial to assess the genesis and behavioral characteristics of fires in forest areas. The discovery of the forest fire areas and the intensity of the fire affected are greatly facilitated by the satellite image obtained by different sensors and data sets. We are suggesting a novel approach to compute changes using spectral indices, using landsat-9 and sentinel-2 satellite datasets for measuring the change in forest areas affected by fire accidents over Kochi areas on March 2023. Kochi is a city in Kerala, South India, and is located at 9° 50' 20.7348" N and 77° 10' 13.8828" E. coordinates. Computation is performed by calculating forest area before the fire incident (pre-fire) and after the fire incident (post-fire) and total loss is calculated by the difference between pre-fire and post-fire incident. The proposed work uses Sentinel-2 and Landsat-9 satellite images to recover burn scars using several vegetation indicators. We have identified the fire locations using the object-based classification approach. For verification of results computed by vegetation indices, we have also performed land use land cover classification and calculated the changes in forest areas. Accuracy is computed by the

confusion matrix with an accuracy of 89.45 % and the kappa coefficient with an accuracy of 87.68%. In particular, there was a strong correlation between forest loss and the burned area in the subtropical evergreen broadleaf forest zone (6.9 %) and the deciduous coniferous forest zone (18.9 % of the lands). These findings serve as a foundation for future forecasts of fire-induced forest loss in regions with similar climatic and environmental conditions.

Keywords: GEE; Remote Sensing; Classification; Landsat; Sentinel, Forest Fire;

5.2 Introduction

The forest is a foundation for all living things and is crucial in influencing global climate change. The forest is divided into basically three types: Reserved Forest (RF), Protected Forest (PF), and Unclassified Forest (UF). Under federal or state forest legislation, a place is known as a Reserved Forest (RF), which is completely protected. Unless otherwise permitted, no activity is allowed in the reserved forest. The State Forest Act or the Federal Forest Act may designate a territory as a Protected Forest (PF). An unclassified Forest (UF) is a place that has been given a forest designation but is not a restricted or protected forest. Depending on the state, these woodlands may be owned differently. Being the primary categories of natural landscapes, forests are the most priceless natural resources on earth. Unlike other natural resources like minerals, mineral oils, and natural gas, which are finite and cannot be replenished, forests have the amazing advantage of being renewable. However, the productivity of forests depends heavily on human activity. More than four billion acres of forest cover the entire earth, according to a survey by the Global Forest Resource Assessment (GFRA) in 2020 [88] the size of these forests cover varies, and it covers the entire surface of the earth. With around 45% of the world's total forest cover, Europe has the largest forest cover, according to another report from 2011. Considering percentages, South America is the top continent in terms of the area that is covered by forests, with almost half of its land mass falling under forests. If we consider per capita forest area, Oceania is placed first. But a percentage of these forest areas are degraded year to year due to several natural factors or man-made activity.

According to the Forest Survey of India in the Year 2023 between the dates 1 March 2023 to 15 March 2023, a total of 772 large fire accidents happened. Mizoram has been affected by a maximum total of 110 fire incidents, and Meghalaya, Manipur, and Assam forest areas have been affected a total of 59, 52, and 43 fire incidents respectively. The reasons behind these fire incidents in forest areas are less precipitation level in February 2023. The country recorded only a 7.2 mm rain level in February which is the sixth lowest from the year 1901, central India recorded 77% rain deficiency while the northwest, southern, and northeast area recorded 76%, 54%, and 43% rain deficiency respectively. Total 38 forest fire incidents were reported in Kerala between January and April of 2018. Kerala had numerous fire events, from January to March 2017; 2,100 hectares were destroyed in 440 fire incidents. The reason behind this incident is climate change also because one of the factors that contribute to fires is the expansion of dryness in the environment. Forest Fires can be identified and characterized by several indicators. The spectral reflectance characteristics of healthy vegetation and burnt scar vegetation can be used to identify forest fires. For fire detection, thermal variations between burning and background pixels are also frequently used. The thermal infrared bands on satellite sensors like ASTER, MODIS, and VIIRS also can be used to detect forest fires. Since the majority of these sensors are categorized as having coarser spatial resolutions, it is challenging to monitor fires on a regional scale. For instance, burn area products from the Moderate Resolution Imaging Spectroradiometer (MODIS) are available globally in 500 m (MCD64A1) since 2001 [89]. There are a number of researchers who performed change detection due to fire incidents over different land cover areas like [90] performed forest fire area computation over Mugla, Turkey using different vegetation indices. [77] performed their analysis over the Pacific Northwest, USA for finding the effects of climate change on fire regimes.

The rest of the paper is structured as follows. Related work is presented in Section II. Section III illustrates the study area, where we discuss the geo-morphological, and climate conditions of the study area, section IV presents a methodology and materials like data sets, vegetation indices, and classification algorithms used in the computation. The result is discussed in section V and in the last section VI presents the concluding remarks.

5.3: Related Work

Forests fire occur in recent years on a very large manner and remote sensing techniques mainly used for the finding reason of these fires incidents, there are a number of plate-form available in geographic information systems (GIS) which are used to calculate these changes like QGIS, ArcGIS, and ELWIS using satellite images. But these GIS software required large processing of data before computation [76] have used such geographic information system-based software and produced a framework for computing change analysis. In his analysis data sampling was put into place in two primary stages: maintaining a binary map for the burned area and unburned areas and mapping the burned areas of various land cover classes. Primary stage focuses on mainly five steps: removal of cloud cover and other noise (prepossessing), spatial and spectral feature extraction for pre-fire and post-fire, analysis of change in forest area in pre-fire and post-fire computation, extraction of feature, mapping of burned areas based on the selection of feature from VIIRS hotspots, but extraction of burned area and unburned area takes a very large amount of time and computation so some researcher used Google Earth Engine for computation and pre-analysis. [91] have performed their analysis using Google Earth Engine (GEE)[72], over two satellites Landsat 8 and Sentinel 2. The researcher performed their computation by partitioning into four-level, performed the supervised burned area over the cartography tool for stratified random sampling, after the selection of data change in the area over a given time period. [92] performed his study by use of 14 different sets of fire variables that are taken from spectral vegetation indices, environmental variables, climate factors, and other spatial features. The training and testing validation is performed using a classification algorithm later bootstrap, and optimistic bootstrap approaches to evaluate the models' accuracy and estimate the variance and bias of the estimate. But these researchers are focusing on the only sample and collected data set not on real-time data sets. Later Andrea Tassi and Marco Vizzari [93] proposed point-based (PB) and object-based (OB) classification and categorization techniques, that may be implemented on the spatial cloud platform, Google Earth Engine (GEE). Ramita Manandhar [94] also studied the framework-based technique of classification to enhance the accuracy of the classification algorithm by combining other data, such as different land cover classes [95], spatial features, spectral indices, and digital elevation models. [96] discuss the other category of classification, he performed categories based on the study of light detection and ranging (LIDAR) was used as part of an innovative strategy for mapping the danger of forest fires. The hierarchy process was used to calculate the criteria weights that affect fire risk and was applied to two different data sets over different location of Spain. The findings indicate that about 50 % -

65% of study area are classified as 3-moderate fire risk zones. Researcher will be able to choose the best vegetation management strategies based on the danger of forest fires according to the technique given in this .[96] study. The forest fires problem does not occur only in the global region but also causes in many places in small areas like the valley of the Himalayas. Over the course of the whole research region, the forest fire burn area dramatically fluctuates with time and space. [97] concluded that fire problems are arises in west-to-south Himalayan region multiple times between year 2000 -2010. Higher burn area fraction patches (0.7sq-km to 1 sq-km) were discovered over the southeast regions of the eastern Himalayas, the central Himalayas, and the western Himalayas. Over the past two decades (2001–2020), the average yearly burn area was 5557.35 sq-km, with a large amount of variability (standard deviation 2661.71 sq-km). The yearly burn area increased significantly by 755.7 sq-km per year between 2001 and 2010 but declined in the following decade. [98] provides the three models anticipated forest fire susceptibility maps (FFSM) using the google earth engine (GEE) platform and then maps it over geographic information system software. The researcher uses a range of probability mechanisms for finding change due to fire incidents value lie between 1 to 0 if it surrounds near 1 then occurrence and loss is high and in other case occurrence and loss is very less. Deforestation is the long-term loss of trees as a result of anthropogenic or natural activity. It happens everywhere as a result of intricate socioeconomic processes like population and housing increase, agricultural expansion, and wood exploitation in underdeveloped nations. Deforestation is also made worse by economic, political, technological, and cultural causes. Deforestation contributes to a number of serious issues, such as soil erosion, disruptions to the water cycle, and possibly global repercussions. Deforestation has caused the land to change too quickly, resulting in the loss of vegetation, and wildlife, which hinders the functioning of ecosystems [99], [100] computed the growth in urbanization and loss of forests between the years 2009 and 2019, and find the current trend of urban growth is continuing at a pace of 0.16% each year. For environmentalists and land planners to comprehend the effects of land use land cover to provide recommendations for effective policy strategies to manage growth in Cameron Highlands.

[78] discussed the forest management systems in Finland and Scandinavia. Satellite images are used in Finland and Sweden's national forest inventories. Researchers [101] witnessed fast changes in land use and cover, with the area covered by vegetation dropping from roughly 46% in 2009 to 28% in 2019. Area of about 27.54 percent 4.60 percent of the area covered by vegetation is converted to urban/built-up areas, 4.60 percent to agriculture, and 6 percent to arid terrain. The amount of agriculture and urban/built-up area has greatly expanded. [48] analyzed the change in different classes of land cover using a number of sample points and training and testing points for each class. Points are split into subgroups for training (70%) and evaluation (30%). Metrics obtained from an error matrix were used to measure accuracy. [102] describes global service on SUHI monitoring is available to provide helpful cues for our cities' increasingly sustainable urban design. [49] discusses different spatial resolution and wavelength of satellite sentinel-2 and Landsat-8 and describe that both satellites are useful and provide good accuracy in measuring changes that occur in a fire-prone area. Sentinel-2 satellites have a spatial resolution of 10 meters while landsat data sets have a spatial resolution of 20 meters.

[59] states that planning for conservation and management, as well as ecological study, is made more difficult in tropical rainforests due to a dearth of spatially and thematically precise vegetation maps. Such maps have a great deal of potential to be produced by remote

sensing, but the categorization accuracy within primary rain forests has typically fallen short of practical uses. Here, we investigate the ability of remote sensing data Landsat ETM+ to distinguish between lowland tropical rain forests in Peru's Amazonia that have different floristically defined forest types based on their and length and parametric value.

5.4 Study Area:

The study area shown in figure 1 is located at 9.50°N and 77.10°E. covering about 2407 square kilometres (1,171sq mi) and bordered on the north by the district of Thrissur, on the south by the districts Kottayam and Alappuzha, on the east by the district Idukki, and on the west by the Arabian Sea.

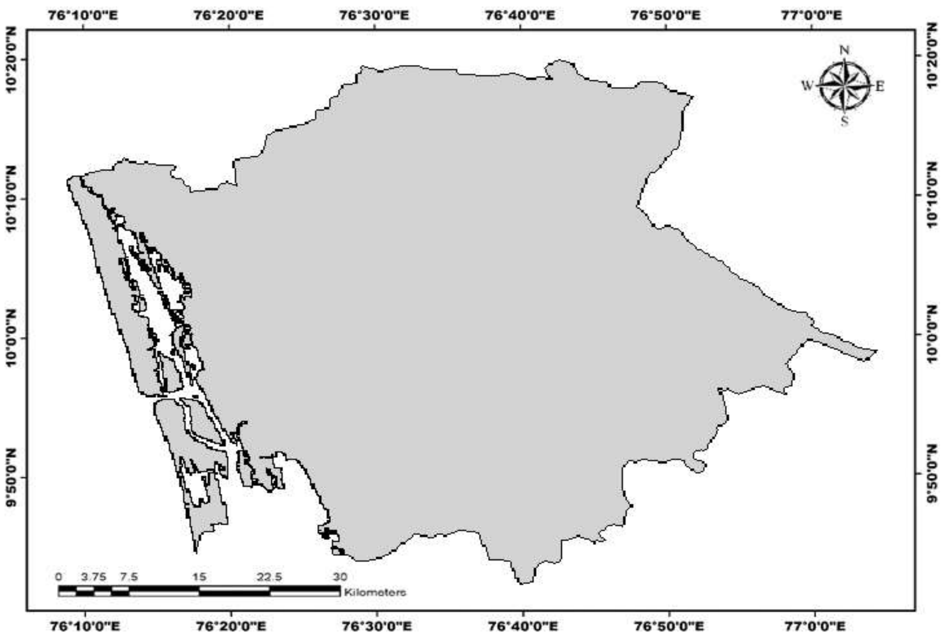

Figure 5.1 Geographical Location of Study Area

Three separate sections make up the district: Hills and woods, plains, and the seashore, respectively, make up the highland, midland, and lowland regions. The highlands are located at a height of around 300 m. (980 ft). Except for Muvattupuzha, the Periyar

River, Kerala's longest, traverses every taluk. The district is traversed by the Chalakkudy River and the Muvattupuzha River.

It consists of several land cover classes like urban, water, and forest (high vegetation), agriculture (low vegetation) but forest alone consists of 45.20% of total land. In the year 2022-23, a total of nearly 430.75 hectares is affected by 391 fire incidents. According to data from the state pollution control board, the mean air quality index remained above 300 $PM_{2.5}$ (particulate matter) concentrations in the air for five days before the fire occurrence.

It was 441 $PM_{2.5}$ concentrations on March 5; 445 $PM_{2.5}$ concentrations on March 6; 465 $PM_{2.5}$ concentrations on March 7; 324 $PM_{2.5}$ concentrations on March 8; and 380 $PM_{2.5}$ concentrations on March 9. Good breathable air quality has an index value of less than 50 $PM_{2.5}$ concentrations, while before the dump yard fire, the city's average air quality index was below 100 $PM_{2.5}$ concentrations.

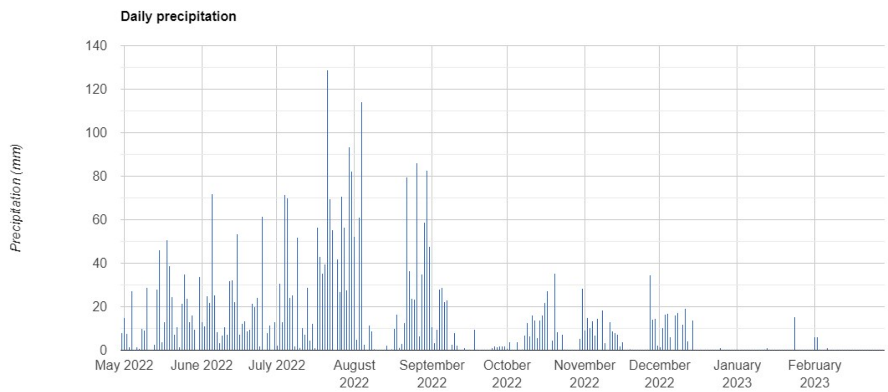

Figure 5.2 Daily Average rain in Study Area between May 2022 - March 2023

The study area is computed as 3,432 mm of rain falls on average in the district each year. The average rain between May 2022 to March 2023 is shown in figure 2. In figure 2 mean precipitations also indicates that the study area has very less rain between October 2022 to

March 2023. It is achieving its highest value of nearly 130mm in the month of August 2022. So it may also be a reason behind these fire occurrences in forest regions. Later we calculate the daily mean temperature also finding its relevance over the atmospheric condition.

The district has a mild temperature average temperature between May 2022 to March 2023 as shown in figure 5.3.

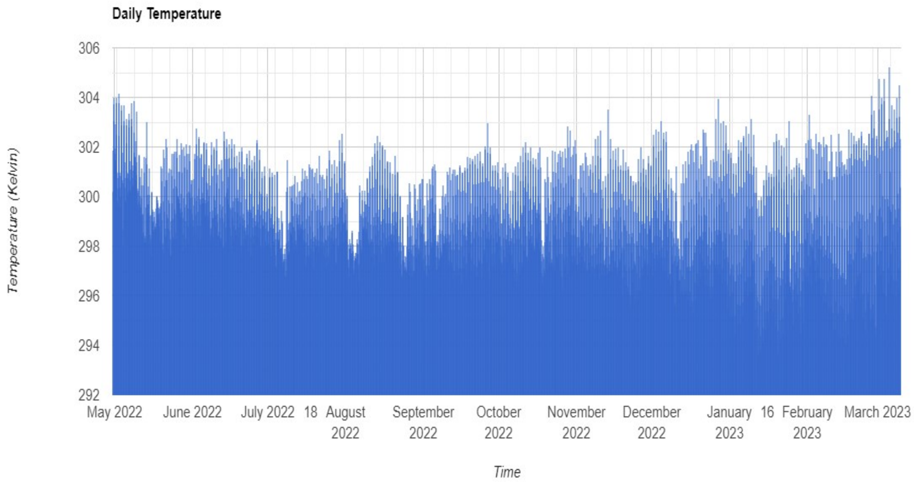

Figure 5.3 Daily temperature in Study Area between May 2022 - March 2023

The temperature rapidly increases during fire time and is largely located in the Malabar Coast moist forests ecoregion, while the highlands are a part of the south-western Ghats moist deciduous forests ecoregion. On the border of the districts of Ernakulam and Idukki, the Anamudi is the tallest mountain in South India. Sholas can be found in some areas of the Mankulam Forest Division and Idamalayar Reserve Forest, however, these areas cannot be reached by road. Edamalakkudy and the Idamalayar Protected Forest, have different kinds of rocks, silt, and sand. The majority of the district's eastern forests are secluded and are a portion of the Anamalais. Temperature is also a very significant factor after the forest fire incident it changed moderately. In March 2023 it went to its peak and affected the burned area and nearby areas as well. Transition class is also affected maximum in the case of water in figure 4.

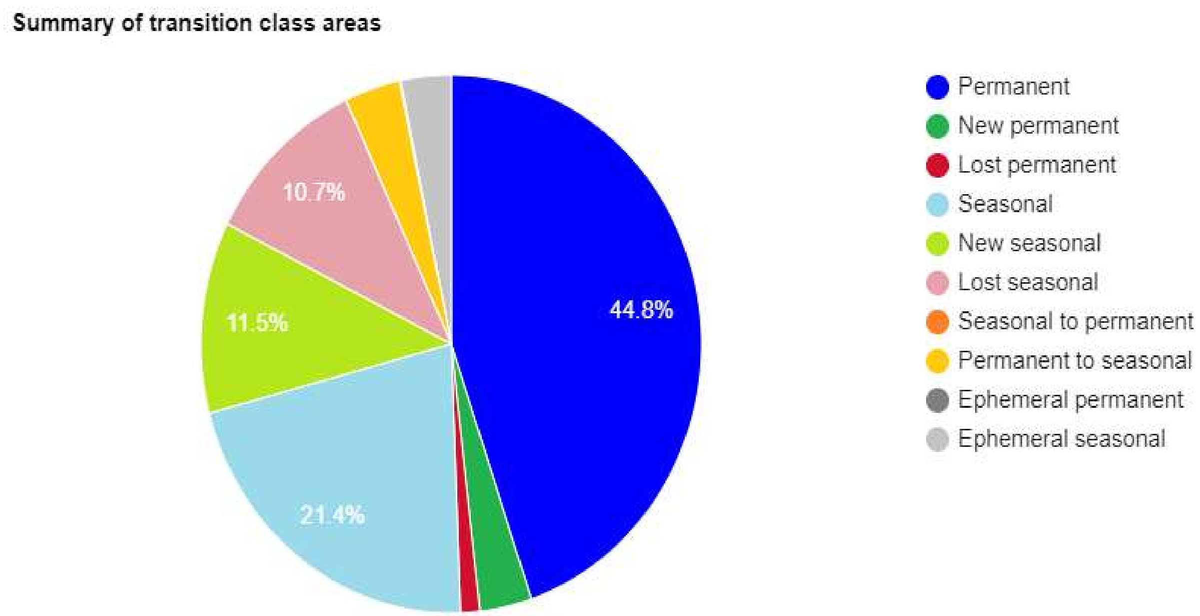

Figure 5.4: Percentage of water over study area

5.5 Method & Data:

Since the 1970s, surface soil moisture (SSM) and change in the surface area has been determined by remote sensing. The primary benefit of remote sensing is that it offers geographically diverse data, as surface variables with spatial information are necessary for many applications, including evapotranspiration evaluation, soil erosion mitigation, irrigation scheduling, drought monitoring, and forest management.

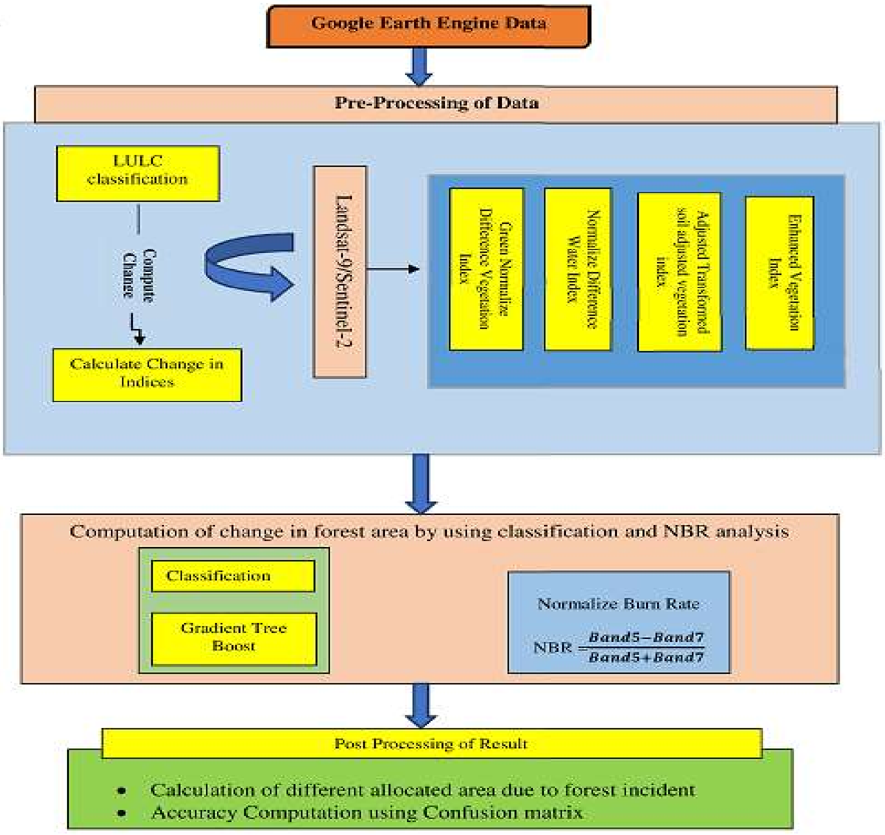

Figure 5.5 Methodology used in research

Table 5. 1 Band Uses for calculating changes in Forest Cover

COLOR	LANDSAT		SENTINEL	
	BAND	WAVE LENGTH	BAND	WAVE LENGTH
BLUE	BAND - 2	0.45 - 0.51	BAND - 2	0.492 - 0.496
GREEN	BAND - 3	0.53 - 0.59	BAND - 3	0.559 - 0.560
RED	BAND - 4	0.63 - 0.67	BAND - 4	0.664 - 0.665
NIR	BAND - 5	0.85 - 0.87	BAND - 8	0.833 - 0.835
SWIR	BAND - 7	2.1 - 2.2	BAND - 12	2.18 - 2.20

The study is performed by using the satellite data set Sentinel and Landsat over the cloud platform Google Earth Engine (GEE). GEE is a planetary platform that has access to different satellite data like Sentinel, Landsat, MODIS, etc. Each satellite has the unique characteristic as wavelength, band combination, and resolution. Here we are accessing sentinel-2 and Landsat-9 for observing changes in the forest area. A detailed description of the analysis is shown in the figure 5.5.

Research is subdivided into two parts. We calculate the change in forest area from pre-fire and post-fire accidents. Later we perform observation over the burn area and calculate the different changes in the forest area. These changes are calculated by computing the change in a land cover class by the gradient tree boost classification model and also by computing the change in mean and standard deviation value of different indices like Green Normalized Difference Vegetation [103], Adjusted Transformed Soil Adjusted Vegetation Index (ATSAVI) [104], [105], Normalize difference water index (NDWI) and Enhance vegetation index (EVI). The methodology used in the research is illustrated in figure 5. GNDVI (Green Normalized Difference Vegetation) [106], [107]measures the "greenness" or photosynthetic activity of plants. While it saturates later than NDVI, it is a chlorophyll index [108]that is utilized during later phases of development. It is one of the most popular vegetation indices for calculating crop canopy water and nitrogen uptake. The values for Normalize difference water index (NDWI), like other indices, range from -1 to 1, with high values denoting high plant water content and coverage of a significant portion of the plant and low values denoting low vegetation water content and sparse cover.

Pre-fire and post-fire[109] results are calculated by using the normalized burn ratio mechanism and calculated by equation 1.

$$\text{NBR} = \frac{NIR - SWIR}{NIR + SWIR} \qquad \ldots\ldots\ldots\ldots\ldots\ldots 1$$

Where: NIR : Near Infra-Red

SWIR: Short Wave Infra-Red

GNDVI has a greater saturation point than NDVI and is more sensitive to changes in the crop's chlorophyll content. While NDVI is useful for predicting crop vigor in the early stages, it can be used in crops with dense canopies or in more mature phases of growth.

$$\text{GNDVI}= \frac{NIR-GREE}{NIR+\ GREEN} \qquad \text{....................2}$$

Where: NIR : Near Infra-Red

GREEN: GREEN Wavelength

A water body can "stand out" against the land and vegetation by using the Normalized Difference Water Index (NDWI) to emphasize open water features in a satellite picture.

$$\text{NDWI} = \frac{GREEN-NIR}{GREEN+NIR} \qquad \text{....................3}$$

The normal reflectivity of the sea surface is maximized in the visible green wavelengths. The near-infrared wavelengths maximize the high reflectance of terrestrial vegetation and soil components while minimizing the low reflection of aquatic features. The NDWI equation yields positive values for water features and negative ones (or zero) for soil and terrestrial vegetation.

$$\text{ATSAVI} = \frac{a*(NIR-a*RED-b)}{(a*NIR)+\ (RED-a*b)+\ (X*(1+a^2))} \qquad \text{...................4}$$

In dense vegetation, EVI is more sensitive and can compensate for some atmospheric conditions and canopy background noise. While ATSAVI is used where vegetation cover is very low.

$$\text{EVI} = \frac{2.5*(Band\ 5-Band\ 4)}{Band\ 5+6*Band\ 4-7.5*Band\ 2+1} \qquad \text{....................5}$$

Classification is performed using the Gradient tree boost machine learning classifier algorithm [50]. It provides a hypothetical model in the form of an ensemble of decision trees. Mainly it is a collection of nodes of weak prediction models. The resulting technique, called gradient-boosted trees, typically beats random forest when a decision tree is a weak learner. The construction of a gradient-boosted trees [49]model follows the same stage-wise process as previous boosting techniques, but it generalizes other techniques by optimizing any differentiable loss function, later we computed accuracy by using the confusion matrix by equation 6 and kappa coefficient by equation 7. Cohen recommended the following interpretation of the Kappa result: values 0 as showing no agreement and 0.01-0.20 as none to the partial agreement, 0.21-0.40 as fair agreement, 0.41- 0.60 as moderate agreement, 0.61-0.80 as significant agreement, and 0.81-1.00 as almost perfect agreement

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{.....................6}$$

Where TP = True Positive

TN= True Negative

FP = False Positive

FN = False Negative

$$\text{Kappa}= \frac{Po-Pe}{1-Pe} \qquad \text{.......................7}$$

Where Po = Observed proportional agreement

Pe= Expected proportional agreement

5.6: Result:

Deforestation is taking place day to day in a very large manner, sometimes it is being done by human activity, and sometimes it is by natural activity, like flood, and fire. We are calculating a change in the forest area of Kochi district from all the above mention activity. The primary reason for the change in forest area over the study area is some fire accidents that occurred between January 2023 to March 2023.

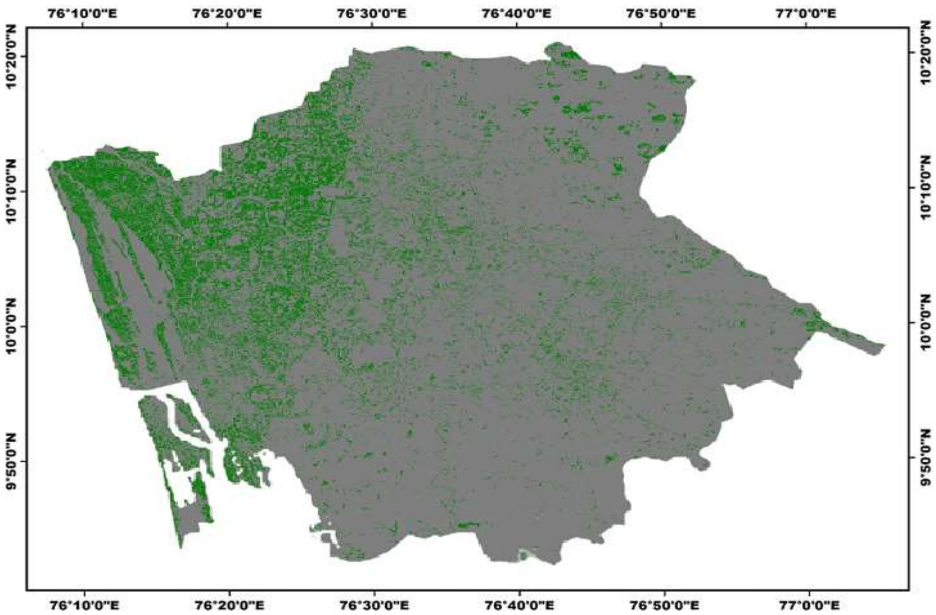

Figure 5. 6 Output Generated over Landsat-9 data sets before fire incident

But we are not only rigid for these fire issues we also calculate all other possible reasons also for deforestation. So here we calculated vegetation indices like NDVI (Normalize Difference Vegetation Index), GNDVI (Green Normalize Difference Vegetation Index), ATSAVI (Adjusted Transformed Soil Vegetation Index), and EVI (Enhanced Vegetation Index) these vegetation indices are used for the calculation of

Change of Vegetation indices and also for evaluating the effect on soil moisture and water due to these changes.

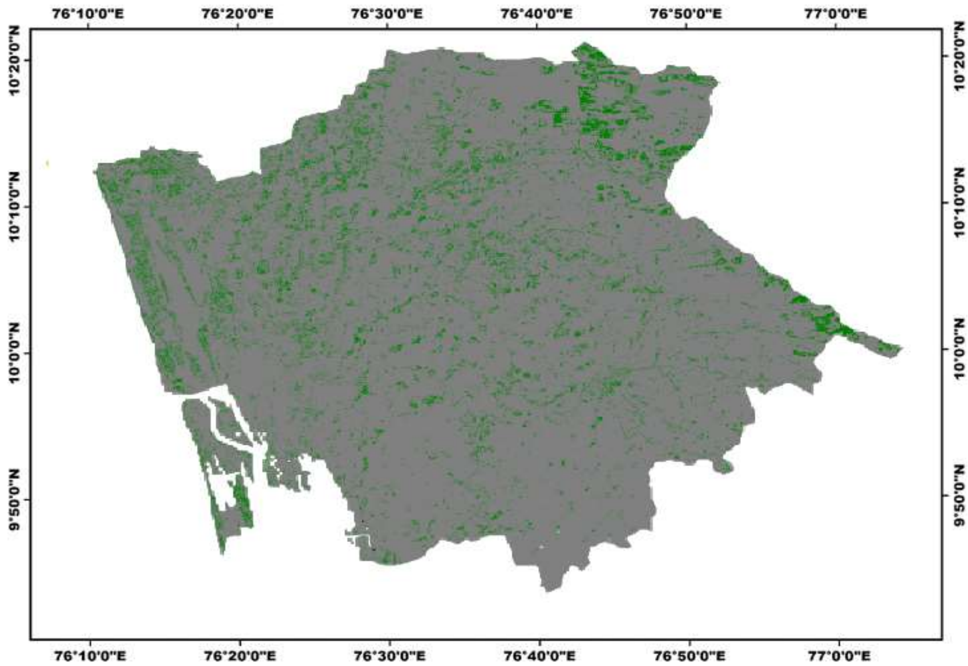

Figure 5.7 Output Generated over Landsat-9 data sets after fire incident

For the calculation of forest change in the study area we first calculate the change as shown in figure 5.6 and figure 5.7 using a supervised classification technique Gradient Tree Boosting. For this, we calculated about 1200 data points of different land cover classes and took 70% data as training points and 30% data as testing points. Results of this classification are shown in figure 6 and figure 7 and these changes are described that a large amount of forest loss occurred by these fire incidents.

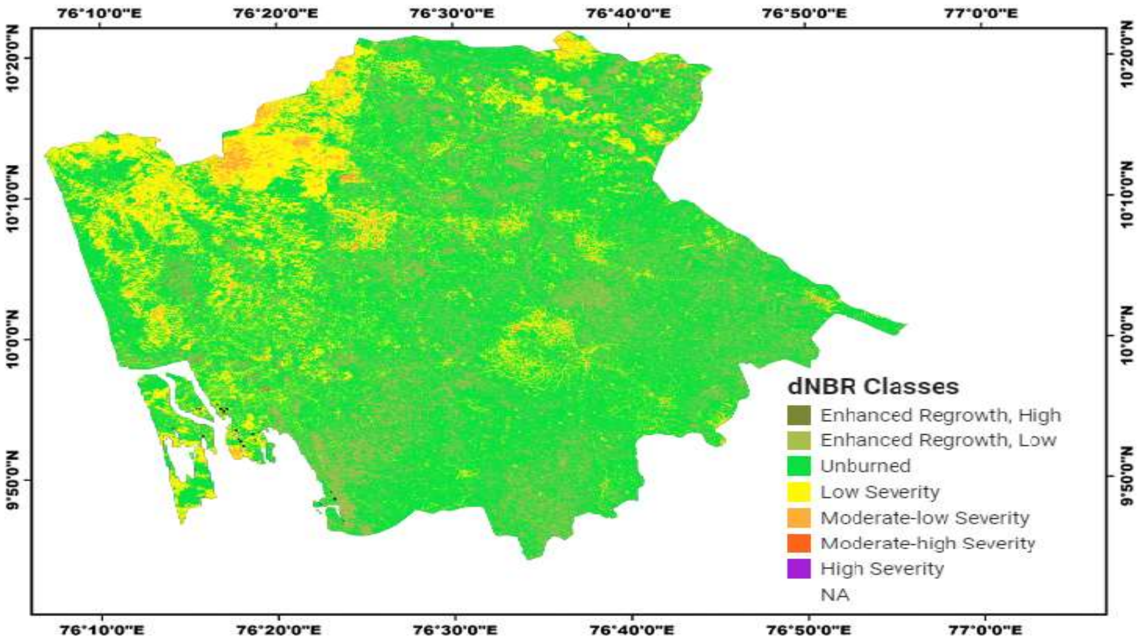

Figure 5.8 Fire Incident happen between January- February Month

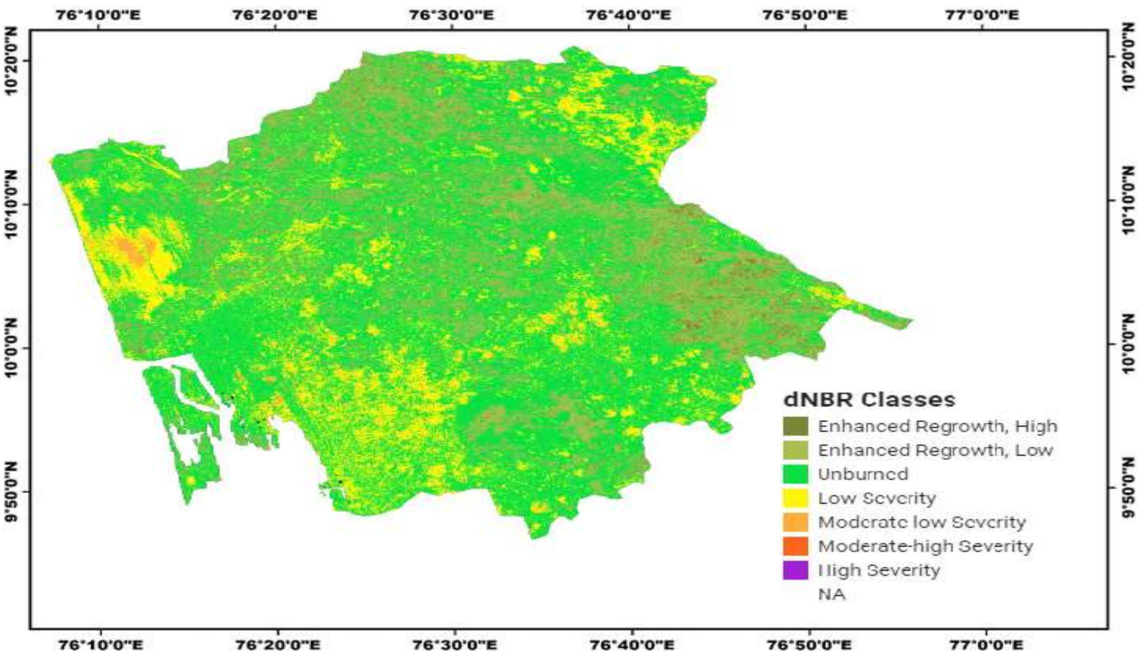

Figure 5.9 Fire Incident happens between February - March Month

So for calculating these losses here, we computed the percentage of the forest before the fire incident (pre-fire) and after the fire incident (post-fire). After classification, we computed the accuracy of the classification result by confusion matrix and kappa coefficient, total accuracy computed by confusion matrix is 89.45% and by kappa coefficient is 87.68%.

On focusing output generated by gradient tree boost in figure 6 and figure 7, there is a lot of change happening in the east-south region of the study area. The reason behind these changes is fire accidents in these areas. So we calculate fire accidents in this area, mainly two major fire accidents found at two places, and many vegetation changes occur due to this fire accident. Figure 8 and figure 9 shows the change in vegetation due to these fire accidents in the study area.

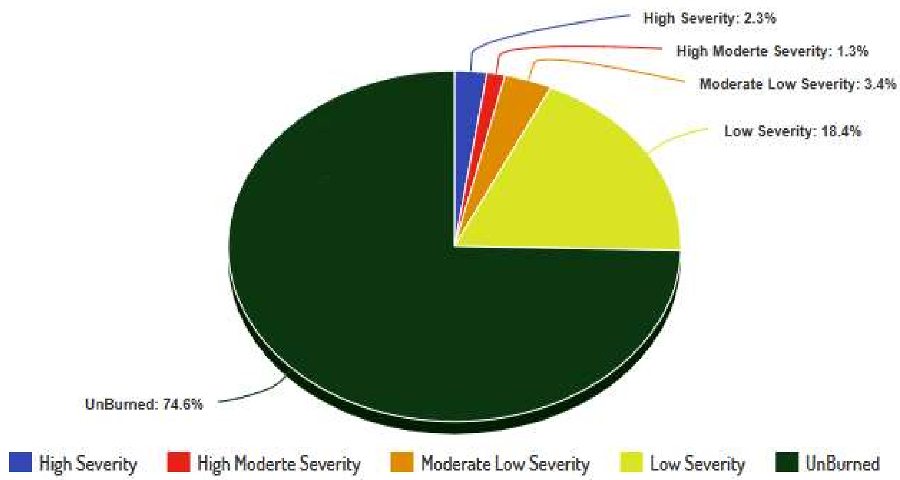

Figure 5.10 Change in Study Area during Fire Incident

The calculation of burned area is sub-categorized into seven different layers and by doing this we can easily calculate the level of loss and growth as well as the type of loss and growth. The level of loss represents how many areas are highly affected by the fire incident due to fire accidents and the low area affected by the fire incident.

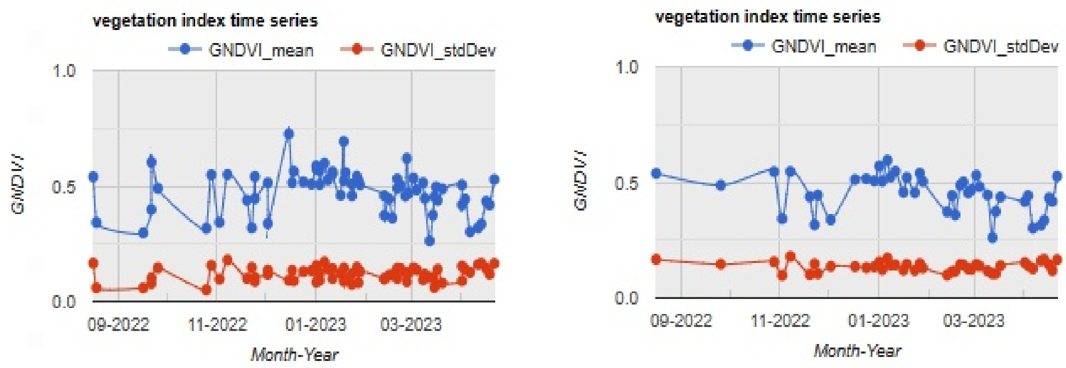

Figure 5.11 Green Normalize Difference Vegetation Index Calculated on Study Area during Forest Fire Time over Landsat-9 and Sentinel 2 data set

Later for extensive classification we calculated some vegetation indices. The main active forest fire episodes occurred between February and March in this region. A study is done on the burned areas in severity class before the fire and after the fire incident and found gradual change, which is shown in figure 5.10.

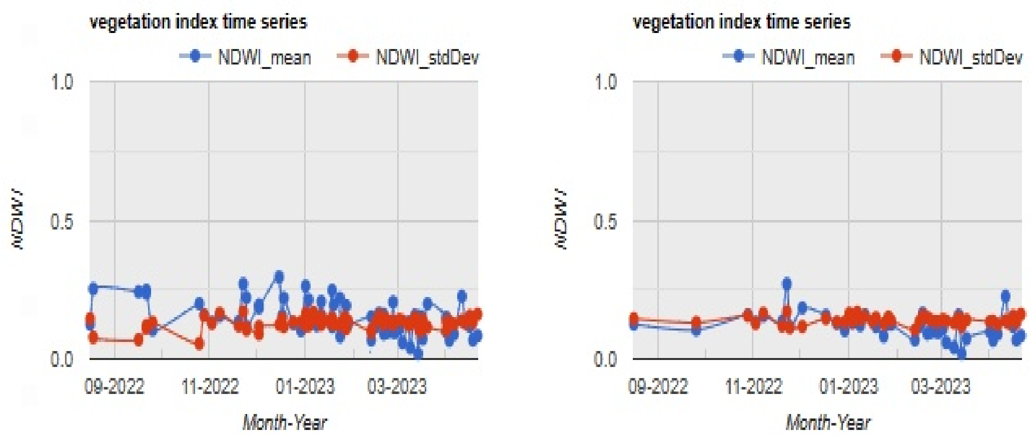

Figure 5.12 Normalize Difference Water Index Calculated on Study Area during Forest Fire over Landsat -9 and Sentinel 2 data set

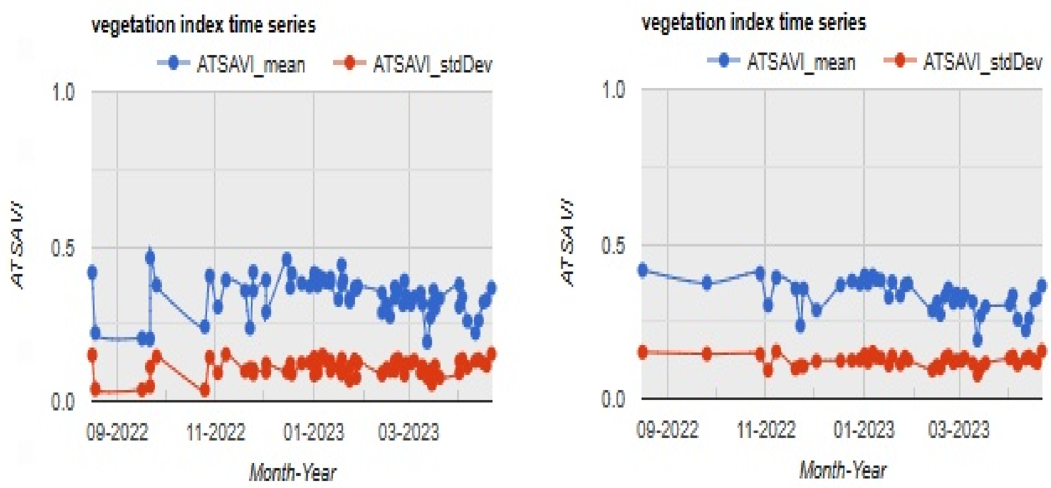

Figure 5.13 Adjusted Transformed Soil Adjusted Vegetation Index Calculated on Study Area during Forest Fire over Landsat -9 and Sentinel 2 data set

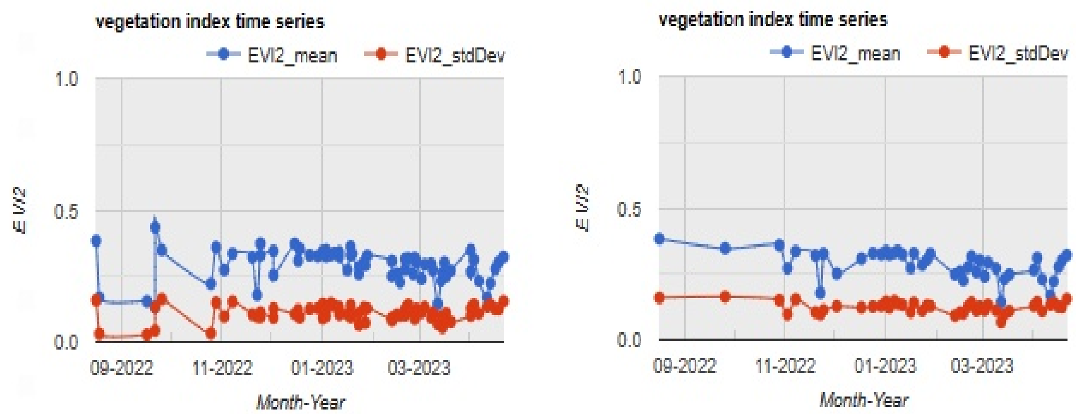

Figure 5.14 Enhanced Vegetation Index Calculated on Study Area during Forest Fire over Landsat -9 and Sentinel 2 data set

Green Normalize Difference vegetation index, Normalize Difference Water Index, Adjusted Transformed Soil Adjusted Vegetation Index, and Enhanced Vegetation Index is calculated over Ernakulum during the forest fire time from September 2022 to March 2023 and we found that time when rapid fire arises its mean and standard deviation value change.

For verification purposes, we have computed the mean GNDVI over two data sets Landsat-9 and Sentinel-2. These two data sets are easily available in google earth engine directory basic difference between both data sets in term of resolution Landsat -9 data sets having resolution 30 meter and Sentinel data sets having resolution 10 meter, these data sets also have a collection of nine different bands each band having unique attributes and the ratio of these bands providing a different signature for each vegetation indices. Figure 11 shows the change in green normalizes difference vegetation index over both data sets and both data sets having nearly same result over the study area. Later we calculate mean and standard deviation value over normalize difference water index (NDWI), adjusted transformed soil adjusted vegetation index (ATSAVI) and enhanced vegetation index (EVI). For observing effects of fire incident over water, soil and natural vegetation indices. Figure 12, Figure 13 and Figure 14 shows the result obtained from NDWI, ATSAVI [105], and EVI over Landsat -9[110] and Sentinel-2 data sets [15], these result shows that change in these indices occurred at the time of fire incident over study area and also after the fire incident also.. From the above observation, we find that fire forest reduces greenery as well as affects water and soil also. Equations (1 to 5) are used for the calculation of these indices and finding change during fire incidents. Equation 1 is used for calculating a change in the study area due to fire incidents and providing details results which are shown in figure 8 and figure 9. Areas affected by these fire incidents are described in figure 10. From the result, it is shown that a large area of forest loss nearly happened over the study area due to fire incidents. These fire incidents not only disrupted climate conditions and global warming but also had a huge effect on soil and other vegetation indices.

5.7: Conclusion

The proposed model of analysis is used to compute forest loss in Ernakulum (Kochi) using Landsat-9 OLI, and Sentinels 2 satellite data due to the occurrence of the fire incident. Through our analysis, we found deciduous forests are more vulnerable to fires. About 18.2% of vegetation area was affected by fires in 2023. Among all types of vegetation classes, vegetation that has higher density is affected most by the fire incident. Fire not only destroys vegetation cover but also has an impact on soil and water as well as on climate conditions also. We are also focused on finding factors behind these multiple fire occurrences over study and find an increase in temperature always increases the probability of fire occurrence. The study has provided valuable insights into the dynamic nature of forest ecosystems in response to fire incidents. The application of spectral indices and machine learning algorithms to Sentinel-2 data has enabled precise and rapid detection of short-term forest cover changes, particularly in the aftermath of fire events in the Ernakulam area. The significance of this research lies in its potential to inform timely decision-making for forest conservation and disaster management. By accurately identifying and characterizing short-term changes [111]caused by fires, it aids in assessing the impact of such incidents on the environment, facilitating a swift response to mitigate further damage and restore affected areas.Furthermore, the methodology employed in this study offers a valuable framework that can be applied in a broader context to monitor and manage forest ecosystems in the face of increasing fire incidents. The integration of spectral indices and machine learning algorithms with Sentinel-2 data showcases a versatile and effective approach for environmental monitoring, with applications extending beyond Ernakulam.

Chapter 6:

Development of User Interface for detecting Changes over Different Multispectral Band and indices.

6.1 Abstract:

The rapid urbanization and environmental changes occurring globally have necessitated the accurate and timely monitoring of land cover changes. Remote sensing technology, with its capacity to provide extensive and consistent data over various spectral bands and indices, plays a pivotal role in this endeavor. This chapter explores the critical need for an effective user interface in the context of detecting land cover changes using remote sensing data.

The study begins by addressing the significance of monitoring land cover changes, emphasizing its relevance in climate change mitigation, natural resource management, disaster response, and urban planning. To effectively utilize the wealth of data generated by remote sensing instruments, an intuitive and user-friendly interface is essential. The paper presents a detailed analysis of the challenges associated with handling complex data and explores the implications of inefficient user interfaces in land cover change detection processes.

The next section delves into the fundamentals of remote sensing, explaining the principles of data acquisition, spectral bands, and vegetation indices commonly used in land cover change analysis. It highlights the diversity and complexity of remote sensing data, including multispectral and hyperspectral images, synthetic aperture radar (SAR) data,

and various indices like the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). Understanding these data sources and their attributes is crucial for building an effective user interface for land cover change detection.

The core of the paper focuses on the need for a user interface [99] that simplifies the process of data selection, preprocessing, analysis, and visualization. The challenges of data management and manipulation are discussed, and the paper provides insights into how a well-designed user interface can streamline these tasks. Examples of existing remote sensing software and platforms are examined to identify their strengths and limitations in terms of user-friendliness.

The study further explores the role of machine learning and artificial intelligence in automating land cover change detection processes. It highlights the potential of user interfaces in facilitating human-computer collaboration, allowing domain experts to harness the power of machine learning algorithms. The paper discusses the integration of user interfaces with machine learning models for improved accuracy and efficiency in land cover change detection.

In addition to discussing the technical aspects of user interfaces, the paper considers the importance of user experience (UX) design principles in creating effective tools for land cover change detection. It discusses the need for user interfaces that are not only functional but also aesthetically pleasing and easy to navigate. This section also addresses the importance of user feedback and iterative design processes to refine the user interface for maximum usability.

The paper concludes by highlighting the critical role of user interfaces in enhancing the accessibility and utility of remote sensing data for land cover change detection. It emphasizes the potential benefits of such interfaces, including improved data interpretation, increased efficiency, and broader adoption of remote sensing technology in environmental and urban planning applications. The study underscores the need for interdisciplinary collaboration between remote sensing experts, computer scientists, and UX designers to develop user interfaces that meet the specific needs of end-users.

In a world increasingly influenced by rapid land cover changes, having user interfaces that bridge the gap between complex remote sensing data and end-users is essential. This paper provides a comprehensive exploration of the need for such interfaces and offers valuable insights into their design and implementation.

6.2 Introduction:

The Earth's surface is constantly undergoing transformation, driven by factors such as urbanization, deforestation, agricultural expansion, and climate change. Monitoring these changes is vital for understanding and managing our environment, natural resources, and the impacts of human activities. Remote sensing technology, with its ability to capture vast amounts of data through satellite imagery and other sensors, has become an invaluable tool for tracking changes in land cover over time. However, as the volume and complexity of remote sensing data continue to increase, the need for efficient tools and user interfaces to analyze and interpret this data has become paramount. This paper explores the imperative need for user interfaces in the context of detecting land cover changes using different spectral bands and indices, shedding light on their role in making remote sensing technology accessible and effective for a broader audience.

6.2.1: Background

Land cover changes, including urban expansion, deforestation, agricultural intensification, and climate-induced alterations, have wide-ranging consequences on ecosystems, biodiversity, and human societies. In the era of climate change and rapid urbanization, tracking and understanding these changes is critical for informed decision-making. Traditional methods of monitoring land cover changes relied heavily on ground surveys and aerial photography. While these methods provided valuable information, they were often limited in their spatial and temporal coverage. The advent of remote sensing technology revolutionized our ability to capture data on land cover changes on a global scale.

Remote sensing involves the use of satellite, airborne, or ground-based sensors to acquire data from the Earth's surface. These sensors capture data in various spectral bands, providing valuable information on different aspects of the Earth's surface, such as vegetation health, water content, and land use. Moreover, remote sensing data can be processed to derive various vegetation indices, which offer insights into vegetation cover, health, and stress. The data generated by remote sensing instruments are rich and diverse, comprising multispectral and hyperspectral images, synthetic aperture radar (SAR) data, and a variety of indices such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and many more.

The potential for extracting critical information from remote sensing data has propelled the use of these technologies across various domains. In environmental science, remote sensing is indispensable for monitoring deforestation, land degradation, and habitat changes. Agriculture benefits from remote sensing to assess crop health, optimize irrigation, and detect diseases. Urban planners use remote sensing to track urban growth and plan for sustainable development. Disaster management agencies rely on it for assessing the impacts of natural calamities, such as wildfires and floods. Remote sensing data has also found applications in climate change research, water resource management, and conservation efforts, among others.

6.3 The Challenge: Managing Complex Remote Sensing Data

While the potential of remote sensing data is vast, managing, processing, and extracting meaningful insights from this data can be a complex and challenging endeavor. Remote sensing data is characterized by its multi-dimensional nature, as it spans a wide range of spectral bands and indices, each with specific attributes and uses. Additionally, the sheer volume of data generated by satellites and sensors can be overwhelming. The complexities associated with remote sensing data can act as barriers, preventing many potential users, including scientists, land managers, and policymakers, from effectively harnessing the technology's capabilities.

One of the primary challenges in using remote sensing data is the need for data preprocessing. This often involves tasks like atmospheric correction, radiometric calibration, geometric correction, and mosaicking. These preprocessing steps are crucial to ensure the accuracy and consistency of the data, but they require expertise and specific software tools. Users without the necessary background and resources may struggle to perform these tasks effectively.

Furthermore, the diverse nature of remote sensing data necessitates knowledge about the spectral bands, indices, and their interpretation. For instance, the NDVI is a commonly used vegetation index that provides insights into vegetation health. Users need to understand that NDVI values range from -1 to 1, with higher values indicating healthier vegetation. Similarly, SAR data is complex and requires knowledge of radar principles and image interpretation. Understanding and working with such data can be a barrier for those who are not specialized in remote sensing.

6.4 The Solution: Spectral Indices Interfaces for Land Cover Change Detection

To address the challenges associated with remote sensing data, there is a pressing need for user interfaces that simplify the process of data selection, preprocessing, analysis, and visualization. An effective user interface can bridge the gap between complex data and users with varying levels of expertise, making remote sensing technology accessible to a broader audience. Such interfaces can empower non-experts and domain specialists alike to explore, analyze, and interpret remote sensing data without requiring an in-depth understanding of the underlying technicalities.

1. Data Selection and Retrieval: Users should be able to easily access and select remote sensing data for their specific region and time of interest. User interfaces can provide tools to browse, search, and retrieve relevant data from various sources.

2. Preprocessing and Analysis: The user interface should offer tools for data preprocessing, allowing users to perform necessary corrections and enhancements without advanced technical knowledge. Additionally, the interface can integrate algorithms for land cover change detection, making it easier to analyze and interpret the data.

3. Visualization and Interpretation: Effective visualization is crucial for conveying information. User interfaces should provide options for generating meaningful and informative visual representations of land cover changes, such as maps, time-series plots, and 3D reconstructions.

4. Machine Learning Integration: Machine learning and artificial intelligence (AI) play a significant role in automating the detection of land cover changes. User interfaces can facilitate the integration of machine learning models, enabling users to apply AI techniques to remote sensing data with ease.

5. User Experience (UX) Design: The design of user interfaces should prioritize user experience, ensuring that the tools are intuitive, efficient, and aesthetically pleasing. UX design principles can make a significant difference in the accessibility and effectiveness of the interface.

6.5 Method And Data :

Vegetation indices[112][113] are extremely useful for monitoring change detection in various environmental and land use applications. Here's how these indices contribute to change detection:

1. **Assessing Vegetation Health and Density:** Vegetation indices like NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), and others provide a quantifiable measure of vegetation health and density. By comparing reflectance in different spectral bands, these indices can detect changes in the condition of vegetation, whether it's thriving, stressed, or degraded. These changes in vegetation health are often the first signs of broader environmental changes.

2. **Detecting Land Cover Changes:** Vegetation indices are widely used to monitor land cover changes, such as deforestation, urban expansion, and reforestation. As vegetation is replaced or altered by land use changes, these indices can detect these shifts over time. Reduced NDVI values, for instance, might indicate deforestation, while an increase in NDVI could signify reforestation.

3. **Identifying Water Bodies and Wetlands:** Indices like NDWI (Normalized Difference Water Index) are specifically designed to detect water bodies. Change detection with NDWI is crucial for monitoring water resources, flood events, and changes in wetland areas.

4. **Monitoring Agricultural Practices:** In agriculture, vegetation indices[114] are used to assess crop health, growth stages, and yield predictions. Change detection through these indices can reveal the impact of various farming practices, such as irrigation, fertilization, and pest control.

5. **Evaluating Ecosystem Health:** In ecological studies, vegetation indices help monitor ecosystem health and changes in biodiversity. Tracking variations in vegetation density and health can provide insights into how environmental changes impact ecosystems over time.

6. **Environmental Impact Assessment:** Vegetation indices are essential for assessing the environmental impact of infrastructure development, land use changes, and urbanization. They help policymakers and planners understand how these activities affect the natural environment.

7. **Drought and Climate Change Monitoring:** Vegetation indices can detect the early signs of drought by assessing reduced vegetation health. Long-term monitoring can also help in assessing the impact of climate change on ecosystems and vegetation patterns.

8. **Flood and Disaster Monitoring:** In disaster management, vegetation indices play a crucial role in flood detection. Satellite imagery with vegetation indices can show changes in water bodies during and after flood events.

9. **Precision Agriculture:** Farmers use these indices to optimize irrigation, fertilization, and other practices, thus increasing crop yield and sustainability.

Here we detailed some Vegetation Indices that we are using in our spectral Interface:

6.5.1 Atmospherically Resistant Vegetation Index (ARVI) :

Atmospherically Resistant Vegetation Index is a vegetation index used in remote sensing and environmental monitoring to assess vegetation health while minimizing the impact of atmospheric conditions. ARVI [97] is particularly useful in areas where atmospheric interference, such as haze, aerosols, and varying illumination, can affect the accuracy of other vegetation indices like NDVI [85], [115] (Normalized Difference Vegetation Index).

ARVI is calculated using the following formula:

ARVI = (NIR - (1 - L) * Red) / (NIR + (1 + L) * Red)

- **NIR (Near-Infrared):** Reflectance in the near-infrared portion of the electromagnetic spectrum.
- **Red:** Reflectance in the red portion of the spectrum.

- **L:** A parameter to adjust for atmospheric conditions, often set to a specific value based on the characteristics of the study area.

Components:

1. **NIR - Red:** This part of the formula represents the difference between the reflectance in the near-infrared (NIR) and red bands. Healthy vegetation typically reflects more NIR and absorbs more red light, resulting in a positive value for this difference.
2. **NIR + Red:** The sum of NIR and red reflectance. This denominator helps normalize the index, making it less sensitive to varying illumination conditions and atmospheric effects.
3. **(1 - L) and (1 + L):** These parameters account for atmospheric conditions. They are often chosen based on the characteristics of the study area to reduce the impact of atmospheric interference on the index.

Key Characteristics and Uses:

1. **Vegetation Health Assessment:** Like other vegetation indices, ARVI is used to assess the health and density of vegetation. Healthy and dense vegetation will yield higher ARVI values.
2. **Change Detection:** ARVI is employed in change detection studies, such as monitoring land cover changes, deforestation, reforestation, and urban expansion. It is particularly valuable in regions where atmospheric conditions may otherwise distort the results of other vegetation indices.

3. **Environmental Management:** It is used in environmental impact assessments to evaluate the effects of infrastructure development and land use changes on vegetation and ecosystems.

4. **Precision Agriculture:** ARVI is applied in precision agriculture to monitor crop health and optimize farming practices. It assists in detecting areas of stress or nutrient deficiencies in crops.

6.5.2 Enhanced Vegetation Index 2:

Enhanced Vegetation Index 2[116] is a vegetation index used in remote sensing and environmental monitoring. It is a simplified version of the Enhanced Vegetation Index (EVI) and the Enhanced Vegetation Index 2 (EVI2), making it computationally more efficient while still providing a reliable measure of vegetation health and density. EVI2 is particularly useful in regions with varying atmospheric conditions where the presence of aerosols or clouds can affect the accuracy of vegetation assessments.

EVI2 is calculated using the following formula:

EVI2 = 2.5 * ((NIR - Red) / (NIR + 2.4 * Red + 1))

- **NIR (Near-Infrared):** Reflectance in the near-infrared portion of the electromagnetic spectrum.
- **Red:** Reflectance in the red portion of the spectrum.

Components:

1. **NIR - Red:** This part of the formula represents the difference between the reflectance in the near-infrared (NIR) and red bands. Healthy vegetation typically

reflects more NIR and absorbs more red light, resulting in a positive value for this difference.

2. **NIR + 2.4 * Red + 1:** The denominator of the formula helps normalize the index. The addition of 1 prevents division by zero and makes the index less sensitive to varying illumination conditions.

Key Characteristics and Uses:

1. **Vegetation Health Assessment:** EVI2, like other vegetation indices [74], is used to assess the health and density of vegetation. Higher EVI2 values typically correspond to healthier and denser vegetation.

2. **Change Detection:** EVI2 is employed in change detection [19] studies, including monitoring land cover changes, deforestation, reforestation, and urban expansion. Its atmospheric correction properties make it useful for regions with challenging atmospheric conditions.

3. **Environmental Monitoring:** EVI2 is used in environmental impact assessments to evaluate the effects of land use changes on vegetation and ecosystems. It contributes to the assessment of ecosystem health.

4. **Precision Agriculture:** In agriculture, EVI2 is applied to monitor crop health, growth stages, and yield predictions. It assists in detecting areas of stress or nutrient deficiencies in crops.

6.5.3 Normalized Difference Vegetation Index (NDVI) :

Normalized Difference Vegetation Index is one of the most widely used vegetation indices in remote sensing and environmental monitoring. NDVI quantifies the density and health of vegetation in a given area by comparing the reflectance of two different spectral bands from satellite or aerial imagery. It is a fundamental tool for assessing changes in vegetation cover, health, and land use over time.

Here's a detailed explanation of NDVI:

Formula: NDVI is calculated using the following formula:

NDVI = (NIR - Red) / (NIR + Red)

- **NIR (Near-Infrared):** Reflectance in the near-infrared portion of the electromagnetic spectrum.
- **Red:** Reflectance in the red portion of the spectrum.

Components:

1. **NIR - Red:** This part of the formula represents the difference between the reflectance in the near-infrared (NIR) and red bands. Healthy vegetation typically reflects more NIR and absorbs more red light, resulting in a positive value for this difference.
2. **NIR + Red:** The sum of NIR and red reflectance. This denominator helps normalize the index, making it less sensitive to varying illumination conditions and atmospheric effects.

Key Characteristics and Uses:

1. **Vegetation Health Assessment:** NDVI is used to assess the health and density of vegetation. Healthy and dense vegetation will yield higher NDVI values, while barren or non-vegetated areas will have lower or even negative NDVI values.

2. **Change Detection:** NDVI is a key tool for change detection studies. It is applied in monitoring land cover changes, such as deforestation, reforestation, urbanization, and other land use transformations.

3. **Environmental Monitoring:** NDVI is employed in environmental impact assessments to evaluate the effects of infrastructure development, mining, and land use changes on vegetation and ecosystems.

4. **Precision Agriculture:** In agriculture, NDVI is used for precision farming. It helps monitor crop health, growth stages, and the effectiveness of irrigation and fertilization practices.

5. **Ecosystem Studies:** NDVI contributes to the monitoring of ecosystem health, especially in assessing the impact of climate change and other environmental factors on vegetation.

6. **Drought and Water Stress Detection:** NDVI is used to identify areas suffering from water stress or drought. Decreased vegetation health is reflected in lower NDVI values.

7. **Urban Planning:** NDVI is also used for urban planning and management, helping assess green spaces, tree cover, and the urban heat island effect.

8. **Forestry:** It plays a vital role in forest management, aiding in monitoring forest health and changes in forest cover.

Interpretation of NDVI Values:

- NDVI values range from -1 to 1.
- Values close to 1 represent healthy, dense vegetation.
- Values close to 0 indicate sparse or stressed vegetation, barren land, or urban areas.
- Negative values often represent non-vegetated surfaces such as water bodies or clouds.

6.7.4 Normalized Difference Water Index (NDWI):

Normalized Difference Water Index is a remote sensing vegetation index used to detect and quantify the presence of water in various environments. NDWI is particularly valuable for monitoring changes in water bodies, wetlands, and aquatic ecosystems. It's based on the differences in reflectance between the near-infrared (NIR) and short-wave infrared (SWIR) bands of remote sensing data.

Here's a detailed explanation of NDWI:

Formula: NDWI is calculated using the following formula:

NDWI = (NIR - SWIR) / (NIR + SWIR)

- **NIR (Near-Infrared):** Reflectance in the near-infrared portion of the electromagnetic spectrum.
- **SWIR (Short-Wave Infrared):** Reflectance in the short-wave infrared portion of the spectrum.

Components:

1. **NIR - SWIR:** This part of the formula represents the difference between the reflectance in the near-infrared (NIR) and short-wave infrared (SWIR) bands. Water absorbs SWIR radiation, while vegetation and other surfaces reflect it. So, areas with water will yield positive NDWI values.
2. **NIR + SWIR:** The sum of NIR and SWIR reflectance. This denominator helps normalize the index, making it less sensitive to varying illumination conditions and atmospheric effects.

Key Characteristics and Uses:

1. **Water Detection:** NDWI is primarily used to identify the presence of water in a landscape. Water bodies, such as lakes, rivers, and ponds, typically have positive NDWI values, whereas non-water surfaces, like land or built-up areas, tend to yield lower or negative NDWI values.
2. **Change Detection:** NDWI [86]is essential for monitoring changes in water bodies over time, including assessing changes in water levels, identifying flood events, and tracking alterations in wetland areas.
3. **Wetland Mapping:** NDWI is widely used in wetland mapping and conservation efforts to monitor the health and extent of wetland ecosystems. Changes in NDWI values can indicate changes in wetland conditions.
4. **Flood Monitoring:** NDWI is valuable for flood monitoring and management. It aids in identifying flood extents and changes in water bodies during flood events.
5. **Hydrological Studies:** NDWI is used in hydrological studies to analyze changes in water resources and assess the impact of land use changes on water bodies.

6. **Agriculture:** NDWI can be applied in agriculture to monitor soil moisture and assess irrigation needs. It helps in optimizing water usage in farming.
7. **Ecosystem Studies:** It contributes to the study of aquatic ecosystems and can be used to track changes in aquatic habitats and ecosystems over time.

Interpretation of NDWI Values:

- Positive NDWI values indicate the presence of water.
- Values close to 1 represent open water bodies with high reflectance in the NIR and absorption in the SWIR.
- Values close to 0 indicate surfaces with limited water content.
- Negative values may represent clouds or other non-vegetated surfaces.

6.5.5 Soil-Adjusted Vegetation Index (SAVI):

Soil-Adjusted Vegetation Index is a vegetation index used in remote sensing and environmental monitoring to assess vegetation health and density, particularly in areas where exposed soil or non-photosynthetic vegetation may affect traditional vegetation indices. SAVI[87] was developed to account for variations in soil brightness, which can have a significant impact on the accuracy of vegetation assessments[117].

SAVI = ((NIR - Red) / (NIR + Red + L)) * (1 + L)

- **NIR (Near-Infrared):** Reflectance in the near-infrared portion of the electromagnetic spectrum.
- **Red:** Reflectance in the red portion of the spectrum.

- **L:** A parameter used to adjust for soil brightness. Typically, L is set to a specific value based on the characteristics of the study area.

Components:

1. **NIR - Red:** This part of the formula represents the difference between the reflectance in the near-infrared (NIR) and red bands. Healthy vegetation reflects more NIR and absorbs more red light, resulting in a positive value for this difference.
2. **NIR + Red + L:** The sum of NIR, red, and the L parameter. The denominator helps normalize the index and reduce the influence of soil brightness.
3. **(1 + L):** A multiplier that helps adjust the index based on the soil brightness parameter L.

Key Characteristics and Uses:

1. **Soil Brightness Correction:** SAVI is designed to minimize the impact of exposed soil or non-photosynthetic vegetation in areas with low vegetation cover. It helps provide more accurate assessments of vegetation health by reducing the influence of soil brightness.
2. **Vegetation Health Assessment:** SAVI is used to assess the health and density of vegetation. It is particularly useful in areas where soil brightness can distort traditional vegetation indices like NDVI.

3. **Change Detection:** SAVI is employed in change detection studies to monitor changes in vegetation cover, assess land degradation, and track ecological changes in regions with challenging conditions, such as arid and semi-arid areas.
4. **Environmental Impact Assessment:** It is used in environmental impact assessments to evaluate the effects of land use changes on vegetation and ecosystems, particularly in regions where sparse vegetation is prevalent.
5. **Agriculture:** SAVI can be applied in agriculture to monitor crop health and optimize farming practices. It helps in assessing the effectiveness of irrigation and fertilization practices.

Interpretation of SAVI Values:

- SAVI values typically range from -1 to 1.
- Negative values often represent barren or non-vegetated surfaces.
- Values close to 0 indicate sparse or stressed vegetation, barren land, or urban areas.
- Positive values represent healthy and dense vegetation.

6.6 Result:

The development of the "Spectral Indices Interface" was guided by the need to create a versatile and user-friendly tool for extracting valuable information from

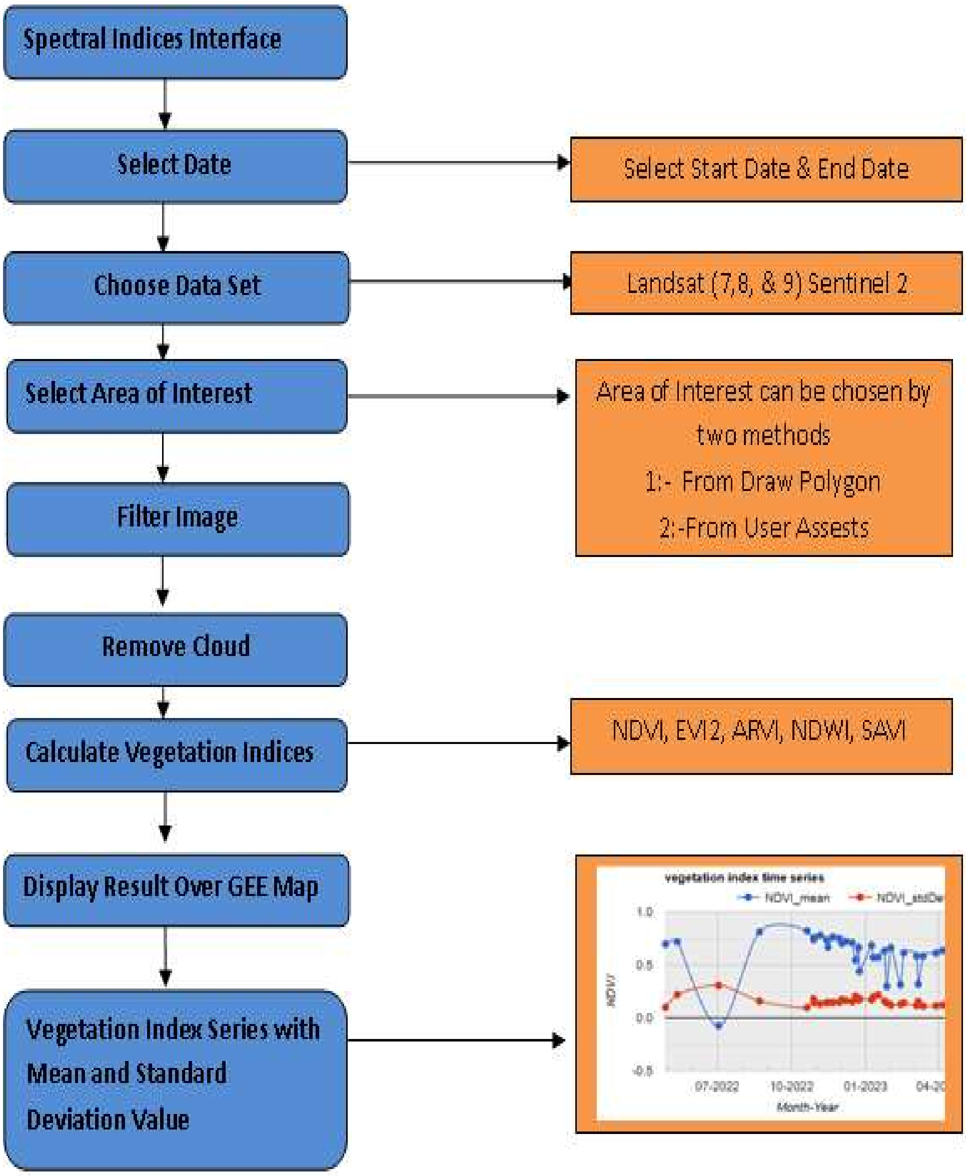

Figure 6.1 Work Flow Of Spectral Indicies Interface

satellite imagery. This interface was designed to empower users to estimate various Vegetation Indices (VIs) and visualize their temporal evolution within user-defined areas. The supported imagery sources include Landsat 7, Landsat 8, Landsat 9, and Sentinel-2, ensuring comprehensive coverage for a wide range of applications.

The key design principles underlying the Spectral Indices Interface were centered on accessibility, functionality, and adaptability. The primary objectives were to provide users with the means to calculate and analyze VIs, particularly in agriculture, and to offer a convenient way to plot time series of these indices for specific regions of interest defined by the user.

A fundamental aspect of the Spectral Indices Interface is that it should be user-friendly, eliminating the need for additional software installations. Users can access the interface through a web-based platform, where they can define the area of interest through polygons. The interface leverages the capabilities of Google Earth Engine (GEE), a cloud-based platform for geospatial data analysis, enabling efficient processing and analysis of large-scale satellite imagery datasets.

The Spectral Indices Interface facilitates the extraction and visualization of VIs, which are essential tools for assessing vegetation health, land cover changes, and water resource distribution. Vegetation Indices are calculated using mathematical formulas that take advantage of the differential reflectance characteristics of healthy vegetation, stressed vegetation, and non-vegetated surfaces in various spectral bands.

The general methodology of the Spectral Indices Interface is outlined in Figure 6.1. The process involves collecting and analyzing atmospherically corrected land surface reflectance images from Landsat (missions 7, 8, and 9) and Sentinel-2. These images are available from 2003 to the present, ensuring an extensive historical dataset for analysis. Users can select the appropriate satellite source (Landsat or Sentinel-2) based on their specific requirements and data availability. This flexibility allows users to tailor their analysis to different research and application needs. Two dataset selection option are available for calculation within the Spectral Indices Interface:

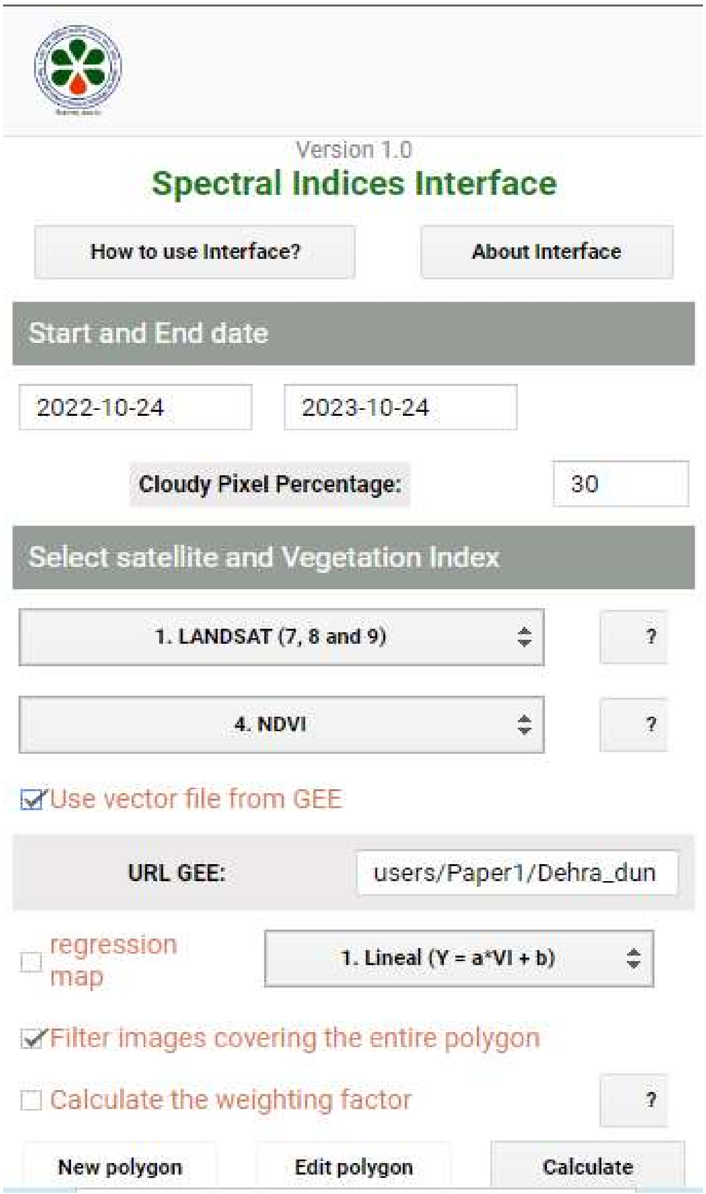

Figure 6.2: Snip of User Interface available at [Spectral Indices Interface (earthengine.app)]

1. Landsat (7, 8, and 9): This option provides a comprehensive dataset from the Landsat missions, offering long-term observations and historical data for analysis.

2. Sentinel-2: Sentinel-2 imagery is available for users who prefer this data source. It is known for its high spatial resolution and frequent revisits, making it suitable for applications requiring more recent and detailed information.

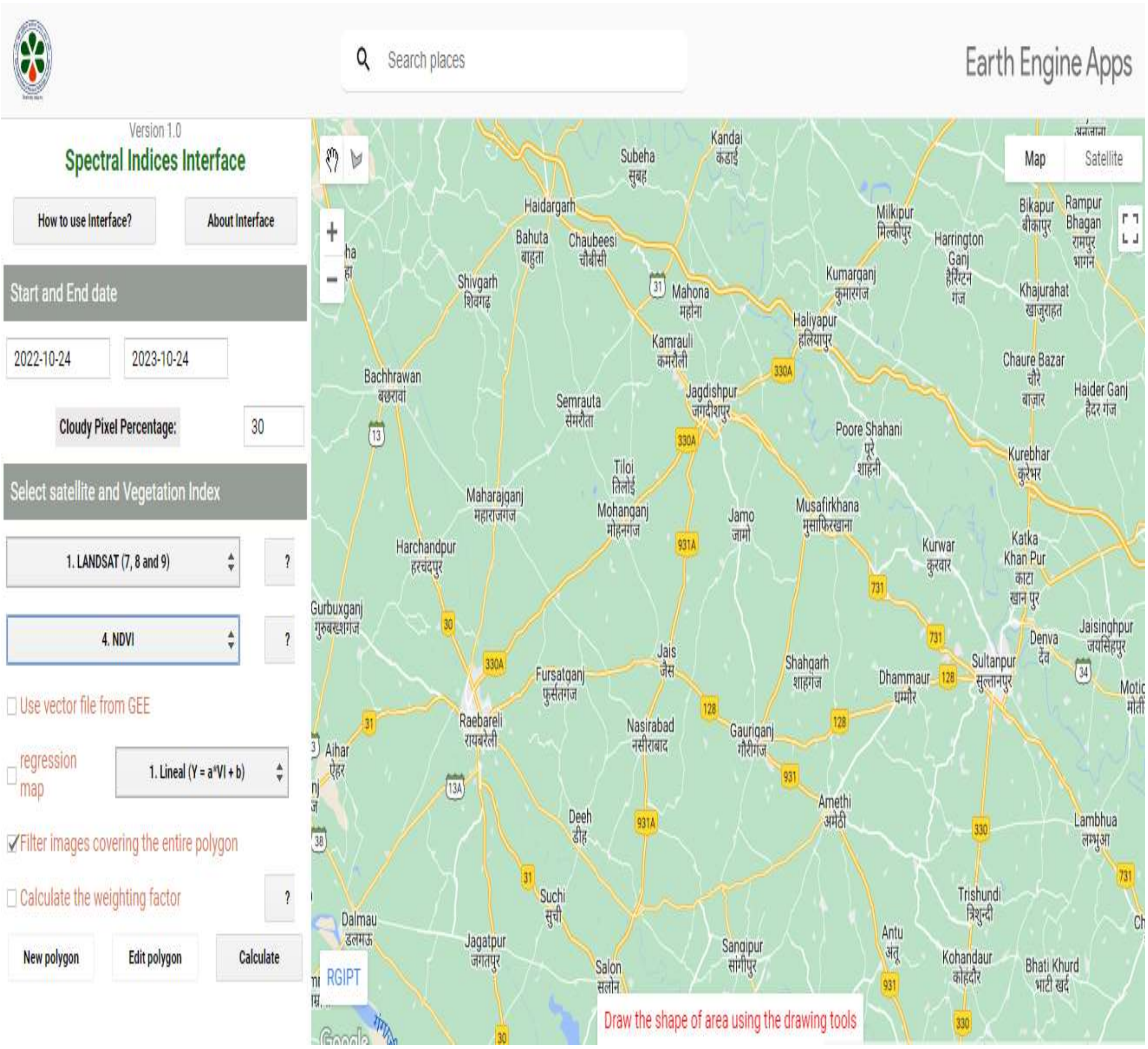

Figure 6.3 Google Earth Engine View of User Interface with selecting any area by using drawing tool available at [Spectral Indices Interface (earthengine.app)]

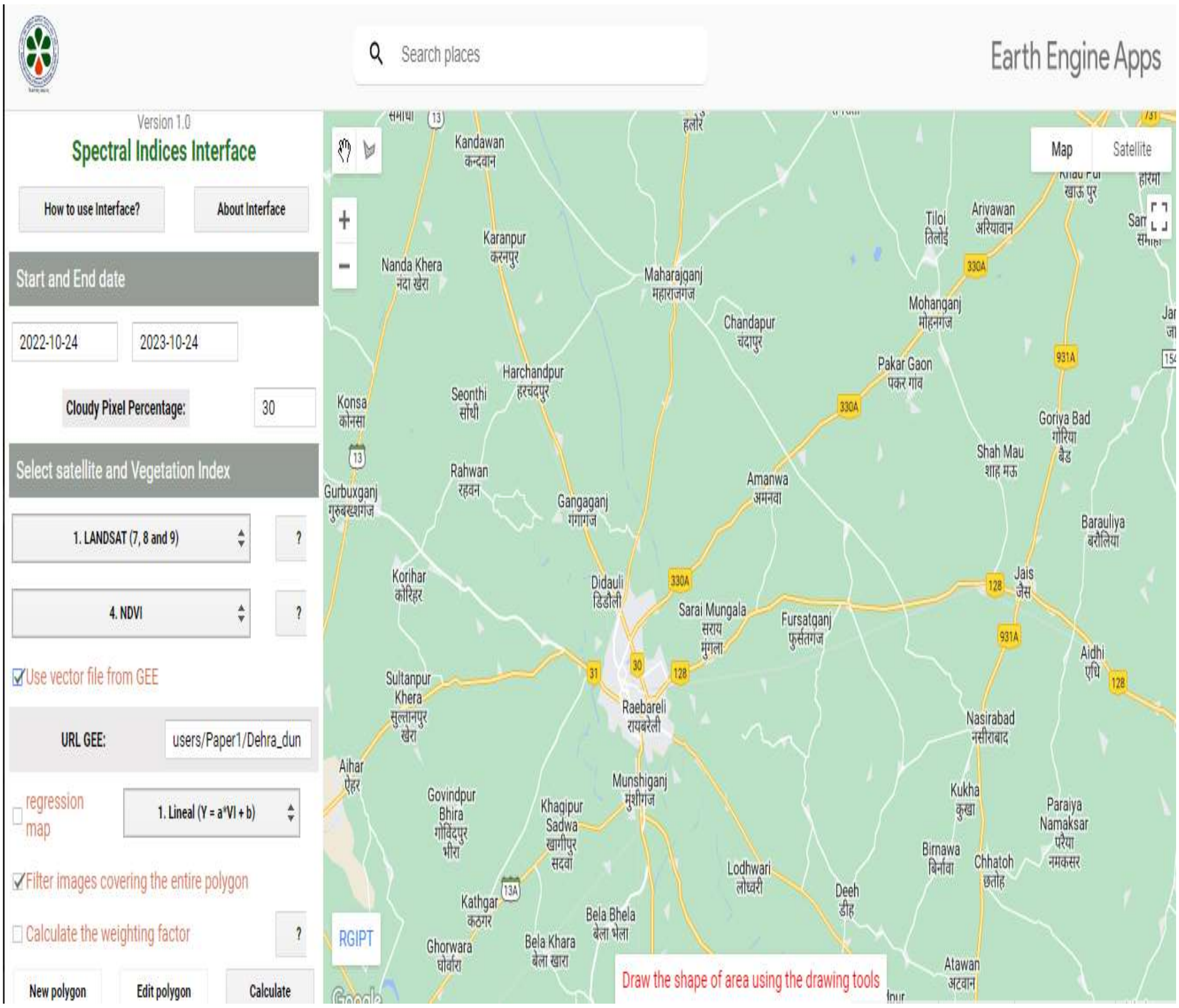

Figure 6.4 Google Earth Engine View of User Interface with GEE Repository URL available at [Spectral Indices Interface (earthengine.app)]

The versatility of these options ensures that users can make informed choices based on the specific objectives of their analysis. Whether they are monitoring changes in agricultural fields, assessing land cover changes, or studying the effects of land use changes on vegetation health, the Spectral Indices Interface provides the tools needed to perform detailed and comprehensive analysis.

Furthermore, the Spectral Indices Interface is adaptable. While it does not require additional software installations, users with programming knowledge can modify or incorporate code into Spectral Indices Interface for more advanced or specialized analysis. This adaptability allows researchers and analysts to extend the functionality of the interface to meet their unique research requirements.

In summary, the Spectral Indices Interface is a valuable resource for researchers, environmental scientists, and land managers. It simplifies the process of estimating and visualizing Vegetation Indices for user-defined regions, and it leverages the power of the Google Earth Engine for efficient data processing. By accommodating various satellite sources and offering adaptability through programming, the interface ensures that users can effectively and comprehensively monitor changes in vegetation, land cover, and water resources, supporting a wide range of applications in environmental monitoring and research.

6.7 Conclusion:

This chapter has delved into the design of a user-friendly interface for processing and analyzing satellite imagery data from Landsat 7, Landsat 8, Landsat 9, and Sentinel-2 satellites. The primary goal of this interface is to compute the mean and standard deviation of various vegetation indices, namely NDVI, ARVI, EVI 2, NDWI, and SAVI. The user is provided with two distinct options for data selection – they can either upload their own repository of satellite imagery or define an area of interest using a polygon. The calculations are performed within the Google Earth Engine environment, ensuring efficiency and scalability for handling large-scale datasets.

This user interface serves as a valuable tool for researchers, environmental scientists, and land managers by simplifying the process of obtaining critical information about vegetation health and water resources from satellite imagery. It empowers users to access and process data from multiple satellite sources, enabling the assessment of trends and changes in vegetation cover, health, and water distribution over time.

The ability to compute the mean and standard deviation of various vegetation indices facilitates a comprehensive analysis of the chosen regions, offering insights into vegetation dynamics, land cover alterations, and the impact of environmental factors. The interface's flexibility allows users to tailor their analysis to their specific research needs, whether it be related to agriculture, forestry, ecosystem monitoring, land use change detection, or environmental impact assessments.

Furthermore, the integration with Google Earth Engine ensures the efficiency, scalability, and accessibility of data processing and analysis. This is particularly important in the context of modern remote sensing, where vast datasets are generated and quick analysis is essential for informed decision-making.

As we move forward in our exploration of remote sensing and environmental monitoring, this user interface will continue to be a valuable resource, aiding researchers and land managers in their efforts to understand and address the complex challenges related to vegetation health, land use changes, and water resource management. The ability to harness the power of Earth observation data and compute vegetation indices' statistics with ease is a significant step towards a more sustainable and data-driven approach to environmental stewardship.

Chapter 7

Conclusion & Future Scope

7.1 Conclusion:

In conclusion, the estimation of changes in land use/land cover (LU/LC) mapping using AI/ML techniques over Google Earth Engine holds significant promise for improving the accuracy, efficiency, and reliability of environmental monitoring and land management. By leveraging the computational power and extensive data resources of Google Earth Engine, combined with advanced AI/ML algorithms, researchers can overcome the limitations of traditional methods and extract valuable insights from satellite imagery.

The integration of AI/ML techniques with Google Earth Engine enables more accurate and robust LU/LC mapping by utilizing the learning capabilities of AI models to capture complex patterns and relationships in the data. Through the classification of digital signatures derived from satellite imagery, AI/ML models can effectively differentiate between different land cover types, enhancing the accuracy of LU/LC mapping and change detection.

The research questions proposed in this context address the key challenges and areas of focus for this field, including accuracy improvement, change detection, digital signature classification, efficiency, scalability, and the practical implementation of AI/ML techniques over Google Earth Engine.

By testing the hypotheses outlined, researchers can evaluate the effectiveness and performance of AI/ML-based approaches for LU/LC mapping and change detection, comparing them with traditional methods. Additionally, addressing the limitations and challenges associated with these techniques is crucial to ensure their practical applicability.

Overall, the research in this field aims to advance the accuracy, efficiency, and reliability of LU/LC mapping, providing valuable information for environmental monitoring, land management, and decision-making processes. By leveraging AI/ML techniques and the capabilities of Google Earth Engine, we can gain a deeper understanding of our changing environment and contribute to more informed and sustainable land management practices.

In our research, we embarked on a journey to estimate and analyze changes in land use and land cover (LU/LC) by leveraging advanced machine learning (ML) and artificial intelligence (AI) techniques over Google Earth Engine. The culmination of this study reveals a multifaceted exploration into the intricacies of LU/LC mapping, including the selection of classification algorithms and datasets, regional case studies in Dehradun, Sikkim, and Ernakulam, and the development of a user-friendly interface for LU/LC change detection using various indices. This extensive investigation has yielded valuable insights and implications across various domains, ranging from environmental monitoring to sustainable land management and disaster mitigation.

The journey began with a thorough examination and comparison of classification algorithms and datasets. This chapter provided an essential foundation for the subsequent research, as it allowed us to make informed decisions regarding the tools and data sources that would be most effective for our analysis. We discussed various ML and AI techniques, such as Support Vector Machines (SVM), Random Forest, CART (Classification & Regression Tree) and GTB (Gradient Tree Boosting) each offering distinct advantages and limitations.

Our analysis also extended to the selection of datasets, such as Sentinel-2 and Landsat, , where we considered factors like spatial and temporal resolution, spectral bands, and cloud cover percentages. The combination of the right algorithms and datasets was crucial for the success of the subsequent chapters.

The insights from this chapter revealed the dynamic landscape of classification algorithms and the vital role played by data selection in the context of LU/LC mapping. These considerations are essential for researchers and practitioners in the field, emphasizing the need for a tailored approach based on the specific objectives of each study.

Chapter three focused on the application of AI and ML techniques to estimate and analyze land cover changes within the Dehradun area. The results of this analysis provided a comprehensive understanding of the changing land use patterns in this region, offering critical insights into urbanization and shifts in agricultural practices.

The application of classification algorithms, as described in the initial chapter, played a pivotal role in identifying and quantifying changes in urban and agricultural landscapes. The findings highlighted the extent of urban expansion and its associated impacts on land cover, as well as the evolving agricultural dynamics that are integral to the region's economy.

The information gleaned from this chapter has far-reaching implications for urban planning, agricultural policy, and environmental management in the Dehradun area. It underscores the importance of using advanced technology to monitor and assess land cover changes, particularly in regions experiencing rapid transformation.

Chapter four extended the research to the forested regions of Sikkim, a critical ecological zone. Here, we applied AI and ML algorithms to detect and classify changes in forest cover over long period of time. This chapter enabled us to assess the conservation status and ecological health of Sikkim's forests, which are vital for biodiversity and environmental balance.

The unique challenges of working in forested areas were addressed, including the complexities of canopy cover and the intricate ecological interactions within these ecosystems. The use of multi-spectral and radar imagery proved invaluable for accurate forest cover mapping.

The insights from this chapter contribute to sustainable forest management and biodiversity conservation. By understanding the dynamics of forest cover changes, we can formulate better-informed policies for preserving the delicate ecosystems of Sikkim and other forested regions worldwide.

Chapter five expanded our focus to the impact of fire incidents on forest cover within the Ernakulam area. Using AI and ML techniques, we analyzed satellite data to identify fire-affected regions and assessed the extent of damage caused by these incidents.

The detection of fire-affected areas is pivotal in understanding the ecological and environmental repercussions of such events. The methodologies discussed in this chapter, including the use of thermal bands, spectral indices, and fire radiative power (FRP), enable us to pinpoint the effects of fire on forest cover in very short period of time.

The findings of this chapter are of utmost significance for fire prevention and ecological restoration efforts. By understanding the extent and impact of forest fires, we can better equip ourselves to mitigate future disasters and support ecosystems in their recovery.

The final chapter introduced a user-friendly interface developed to visualize and access the detected LU/LC changes across the studied regions. This interface serves as a practical tool for researchers, policymakers, and the general public, allowing them to explore and analyze LU/LC changes using various indices, including the Normalized Difference Vegetation Index (NDVI)[118], Normalized Difference Water Index (NDWI), and Enhanced Vegetation Index (EVI).

The development of this interface highlights the significance of effectively communicating research findings to a broader audience. By providing a user-friendly tool, we bridge the gap between advanced AI/ML techniques and practical decision-making. This interface can support stakeholders in making informed choices related to land management, environmental conservation, and disaster response.

Overall, this research project underscores the importance of harnessing AI and ML techniques for the estimation and analysis of LU/LC changes. By leveraging Google Earth Engine, we have been able to explore dynamic landscapes, from urban areas to forested regions and disaster-prone zones. The insights gained are invaluable for regional planning, ecological conservation, disaster management, and informed decision-making.

In conclusion, this research project represents a significant contribution to the field of geospatial analysis and land cover mapping. The findings emphasize the power of advanced technology in understanding and managing dynamic landscapes, and they call for continued innovation and research in the realm of AI and ML applied to Earth observation data. As technology and data availability continue to advance, the possibilities for monitoring and managing our changing world become even more promising.

7.2 Future Scope :

The research conducted on the "Estimation of Change in LU/LC Mapping with Classification of Digital Signature using AI/ML Techniques over Google Earth Engine" has laid a strong foundation for advancing the field of geospatial analysis and environmental monitoring. This comprehensive study explored various aspects, including classification algorithms, data sets, regional case studies, and the development of a user interface. As technology continues to evolve, new opportunities and challenges arise, offering a wide range of future research possibilities and avenues for expansion. This section outlines an extensive future scope that can guide researchers, policymakers, and practitioners in further advancing the field.

7.2.1 Hyper-Resolution Mapping:

The study primarily worked with medium to high-resolution imagery. Future research can delve into hyper-resolution mapping, utilizing imagery from drones, unmanned aerial vehicles (UAVs), or high-resolution satellites like WorldView-3 and WorldView-4. Hyper-resolution mapping allows for more detailed and accurate monitoring of land cover changes at a smaller scale. This can be particularly valuable in urban planning, agriculture, and ecosystem monitoring.

7.2.2 Uncertainty and Validation Metrics:

There is a growing need to quantify the uncertainty associated with AI/ML-based LU/LC mapping. Future research can focus on developing robust uncertainty estimation methods and standardized validation metrics to assess the accuracy of classification models. This is essential for building trust in AI/ML-derived results and making informed decisions based on these outputs.

7.2.3 User-Friendly Interfaces and Decision Support Systems:

The development of user-friendly interfaces, as presented in this study, can be extended to include advanced features, such as augmented reality (AR) and virtual reality (VR) for immersive exploration and scenario testing. Additionally, decision support systems that integrate AI/ML-based LU/LC mapping into the decision-making process can be further refined and customized for specific applications.

7.2.4 Environmental Justice and Equity Considerations:

Future research can address environmental justice and equity issues related to LU/LC changes. Researchers can explore how AI/ML can be used to identify and mitigate disparities in land use decisions and their impacts on marginalized communities.

7.2.5 Case Studies in Specific Domains:

Focusing on specific domains, such as agriculture, water resources, urban planning, or conservation, can provide in-depth insights into the unique challenges and opportunities associated with AI/ML-based LU/LC mapping in those areas. Case studies can inform tailored solutions and strategies.

In conclusion, the future of LU/LC mapping with AI/ML techniques over Google Earth Engine is promising, with a wealth of opportunities for innovation and development. The continuous evolution of technology, data availability, and environmental challenges demands ongoing research and adaptation. Researchers, policymakers, and practitioners should remain agile and open to exploring these future research directions to address the complex environmental and land management issues of our time. The potential benefits, in terms of improved environmental conservation, sustainable land use, and data-driven decision-making, are substantial and can contribute significantly to addressing global challenges.

Reference:

[1] O. S. Belhaj and S. T. Mubako, "Land use land cover change and urban growth in Khoms district, Libya, 1976-2015," *Int. J. Appl. Geospatial Res.*, vol. 11, no. 2, pp. 42–58, 2020, doi: 10.4018/IJAGR.2020040103.

[2] X. Liu *et al.*, "High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform," *Remote Sens. Environ.*, vol. 209, no. February, pp. 227–239, 2018, doi: 10.1016/j.rse.2018.02.055.

[3] K. Islam, M. Jashimuddin, B. Nath, and T. K. Nath, "Land use classification and change detection by using multi-temporal remotely sensed imagery: The case of Chunati wildlife sanctuary, Bangladesh," *Egypt. J. Remote Sens. Sp. Sci.*, vol. 21, no. 1, pp. 37–47, 2018, doi: 10.1016/j.ejrs.2016.12.005.

[4] S. Sinha and R. Sinha, "Geomorphic evolution of Dehra Dun, NW Himalaya: Tectonics and climatic coupling," *Geomorphology*, vol. 266, pp. 20–32, 2016, doi: 10.1016/j.geomorph.2016.05.002.

[5] H. M. Mwangi, S. Julich, S. D. Patil, M. A. McDonald, and K. H. Feger, "Relative contribution of land use change and climate variability on discharge of upper Mara River, Kenya," *J. Hydrol. Reg. Stud.*, vol. 5, pp. 244–260, 2016, doi: 10.1016/j.ejrh.2015.12.059.

[6] A. C. Guzha, M. C. Rufino, S. Okoth, S. Jacobs, and R. L. B. Nóbrega, "Impacts of land use and land cover change on surface runoff, discharge and low flows: Evidence from East Africa," *J. Hydrol. Reg. Stud.*, vol. 15, no. December 2017, pp. 49–67, 2018, doi: 10.1016/j.ejrh.2017.11.005.

[7] R. Goldblatt and J. Burney, "High Spatial Resolution Visual Band Imagery Outperforms Medium Resolution Spectral Imagery for Ecosystem Assessment in the Semi-Arid Brazilian Sertão," *Remote Sens.*, vol. 9, no. 12, p. 1336, 2017, doi: 10.3390/rs9121336.

[8] M. L. Berihun *et al.*, "Hydrological responses to land use/land cover change and climate variability in contrasting agro-ecological environments of the Upper Blue Nile basin, Ethiopia," *Sci. Total Environ.*, vol. 689, pp. 347–365, 2019, doi: 10.1016/j.scitotenv.2019.06.338.

[9] N. Yang, B. H. Men, and C. K. Lin, "Impact analysis of climate change on water resources," *Procedia Eng.*, vol. 24, pp. 643–648, 2011, doi: 10.1016/j.proeng.2011.11.2710.

[10] R. Ky, "Impact of Climate Change on Water Resources," *J. Earth Sci. Clim. Change*, vol. 05, no. 03, 2014, doi: 10.4172/2157-7617.1000185.

[11] G. Fernandino, C. I. Elliff, and I. R. Silva, "Ecosystem-based management of coastal zones in face of climate change impacts: Challenges and inequalities," *J. Environ. Manage.*, vol. 215, pp. 32–39, 2018, doi: 10.1016/j.jenvman.2018.03.034.

[12] M. He *et al.*, "Regional crop gross primary productivity and yield estimation using fused landsat-MODIS data," *Remote Sens.*, vol. 10, no. 3, 2018, doi: 10.3390/rs10030372.

[13] A. Agarwal, K. K. Soni, and M. Rawat, "Monitoring Land Use Land Cover Change for Dehradun District of Uttarakhand from 2009-2019," *Int. J. Adv. Remote Sens. GIS*, vol. 8, no. 1, pp. 3106–3113, 2019, doi: 10.23953/cloud.ijarsg.431.

[14] C. P. de Oliveira *et al.*, “Dynamic Modeling of Land Use and Coverage Changes in the Dryland Pernambuco, Brazil,” *Land*, vol. 11, no. 7, 2022, doi: 10.3390/land11070998.

[15] M. Viana and J. Rocha, “Long-Term Satellite Image Time-Series for Land Use / Land Cover Change Detection Using Refined Open Source Data in a Rural Region,” 2019.

[16] B. E. Mennecke and M. D. Crossland, “Geographic information systems: Applications and research opportunities for information systems researchers,” *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 3, no. May, pp. 537–546, 1996, doi: 10.1109/HICSS.1996.493249.

[17] H. Chen, L. Fleskens, J. Baartman, F. Wang, S. Moolenaar, and C. Ritsema, “Impacts of land use change and climatic effects on streamflow in the Chinese Loess Plateau: A meta-analysis,” *Sci. Total Environ.*, vol. 703, p. 134989, 2020, doi: 10.1016/j.scitotenv.2019.134989.

[18] S. Fuentes and J. Chang, “Methodologies Used in Remote Sensing Data Analysis and Remote Sensors for Precision Agriculture,” *Sensors*, vol. 22, no. 20, pp. 4–6, 2022, doi: 10.3390/s22207898.

[19] E. L. Birch and S. M. Wachter, *Global urbanization*. 2011. doi: 10.2307/j.ctv307fgfg.6.

[20] M. Campos-Taberner *et al.*, “Global estimation of biophysical variables from Google Earth Engine platform,” *Remote Sens.*, vol. 10, no. 8, pp. 1–17, 2018, doi: 10.3390/rs10081167.

[21] S. Jin *et al.*, “Spatial-temporal changes of land use/cover change and habitat quality in Sanjiang plain from 1985 to 2017,” *Front. Environ. Sci.*, vol. 10, no. October, pp. 1–12, 2022, doi: 10.3389/fenvs.2022.1032584.

[22] F. N. Kogan, “Global Drought Watch from Space,” *Bull. Am. Meteorol. Soc.*, vol. 78, no. 4, pp. 621–636, 1997, doi: 10.1175/1520-0477(1997)078<0621:GDWFS>2.0.CO;2.

[23] S. H. Erena and H. Worku, “Dynamics of land use land cover and resulting surface runoff management for environmental flood hazard mitigation: The case of Dire Daw city, Ethiopia,” *J. Hydrol. Reg. Stud.*, vol. 22, no. November 2018, p. 100598, 2019, doi: 10.1016/j.ejrh.2019.100598.

[24] R. Thadani, V. Singh, D. . Chauhan, V. Dwivedi, and A. Pandey, “Climate change in Uttarakhand: Current state of knowledge and way forward,” no. December 2020, pp. 1–66, 2015, [Online]. Available: http://cedarhimalaya.org/images/pdf/Climate Change in Uttarakhand.pdf

[25] D. K. Mukhopadhyay and P. Mishra, “The Main Frontal Thrust (MFT), northwestern Himalayas: Thrust trajectory and hangingwall fold geometry from balanced cross sections,” *J. Geol. Soc. India*, vol. 64, no. 6, pp. 739–746, 2004.

[26] P. K. Sahoo, S. Kumar, and R. P. Singh, “Neotectonic study of Ganga and Yamuna tear faults, NW Himalaya, using remote sensing and GIS,” *Int. J. Remote Sens.*, vol. 21, no. 3, pp. 499–518, 2000, doi: 10.1080/014311600210713.

[27] O. Singh, P. Arya, and B. S. Chaudhary, “On rising temperature trends at Dehradun in Doon valley of Uttarakhand , India,” no. June, 2013, doi: 10.1007/s12040-013-

0304-0.

[28] H. J. Jumaah, M. H. Ameen, and B. Kalantar, “Surface Water Changes and Water Depletion of Lake Hamrin, Eastern Iraq, Using Sentinel-2 Images and Geographic Information Systems,” *Adv. Environ. Eng. Res.*, vol. 04, no. 01, pp. 1–11, 2023, doi: 10.21926/aeer.2301006.

[29] M. A. Brovelli, Y. Sun, and V. Yordanov, “Monitoring Forest Change in the Amazon Using Multi-Temporal Remote Sensing Data and Machine Learning Classification on Google Earth Engine,” pp. 1–21, 2020.

[30] C. Yağcı and F. İşcan, “Turkish Journal of Geographic Information Systems GIS-based site suitability analysis of afforestation in Konya province, Turkey,” *Turkish J. Geogr. Inf. Syst.*, vol. 3, no. 2, pp. 89–95, 2021, [Online]. Available: https://dergipark.org.tr/en/pub/tucbis

[31] “978-3-031-42056-6_4.”

[32] K. D. Hengaju and U. Manandhar, “Analysis on causes of deforestation and forest degradation of Dang district: using DPSIR framework,” *Nepal J. Environ. Sci.*, vol. 3, no. December 2015, pp. 27–34, 2015, doi: 10.3126/njes.v3i0.22732.

[33] H. Jamal Jumaah, A. Jasim, A. Rashid, and Q. Ajaj, “Air Pollution Risk Assessment Using GIS and Remotely Sensed Data in Kirkuk City, Iraq,” *J. Atmos. Sci. Res.*, vol. 6, no. 3, pp. 41–51, 2023, doi: 10.30564/jasr.v6i3.5834.

[34] A. M. Hadi *et al.*, “Gis-based rainfall analysis using remotely sensed data in Kirkuk province, Iraq,” *Tikrit J. Eng. Sci.*, vol. 29, no. 4, pp. 48–55, 2022, doi: 10.25130/tjes.29.4.6.

[35] A. S. Hadeel, “Application of Remote Sensing and GIS to the Study of Land Use / Cover Change and Urbanization Expansion in Basrah Province , Southern Iraq,” vol. 12, no. 2, pp. 135–141, 2009, doi: 10.1007/s11806-009-0244-7.

[36] M. Sharma and R. K. Abhay, “Urban growth and quality of life: inter-district and intra-district analysis of housing in NCT-Delhi, 2001–2011–2020,” *GeoJournal*, vol. 87, no. s4, pp. 797–819, 2022, doi: 10.1007/s10708-021-10570-8.

[37] M. Qiao, C. Wong, and W. Zheng, “Sustainable Urbanisation and Community Well-Being in Suburban Neighbourhoods in Beijing, China,” *Int. J. Community Well-Being*, vol. 2, no. 1, pp. 15–39, 2019, doi: 10.1007/s42413-019-00019-9.

[38] M. D. N. Hadi Mousavi, “A new decade for social changes,” *Tech. Soc. Sci. J.*, vol. 6, no. December, pp. 101–105, 2020.

[39] R. Balstad, “Satellite data,” *Limits Mark. Organ.*, no. September, pp. 294–317, 2005, doi: 10.1002/9781118445112.stat07755.pub2.

[40] S. Ghosh and B. Lohani, “EXPERIMENTAL EVALUATION of LIDAR DATA VISUALIZATION SCHEMES,” *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 1, no. August, pp. 135–140, 2012, doi: 10.5194/isprsannals-I-2-135-2012.

[41] S. Ghosh and B. Lohani, “Development and comparison of aerial photograph aided visualization pipelines for LiDAR datasets,” *Int. J. Digit. Earth*, vol. 8, no. 8, pp. 656–677, 2015, doi: 10.1080/17538947.2014.897386.

[42] B. Lohani and S. Ghosh, “Airborne LiDAR Technology: A Review of Data Collection and Processing Systems,” *Proc. Natl. Acad. Sci. India Sect. A - Phys. Sci.*, vol. 87, no. 4, pp. 567–579, 2017, doi: 10.1007/s40010-017-0435-9.

[43] T. Ranchin and S. Antipolis, “Data fusion in remote sensing: examples,” *Proc. 4th Annu. Conf. Inf. fusion*, no. August 2001, 2001.

[44] A. Karagiannopoulou, A. Tsertou, G. Tsimiklis, and A. Amditis, “Data Fusion in Earth Observation and the Role of Citizen as a Sensor: A Scoping Review of Applications, Methods and Future Trends,” *Remote Sens.*, vol. 14, no. 5, 2022, doi: 10.3390/rs14051263.

[45] U. S. Shanthamallu and A. Spanias, “Introduction to Machine Learning,” in *Synthesis Lectures on Signal Processing*, 2022, pp. 1–8. doi: 10.1007/978-3-031-03758-0_1.

[46] P. Pintelas and I. E. Livieris, “Special issue on ensemble learning and applications,” *Algorithms*, vol. 13, no. 6. 2020. doi: 10.3390/A13060140.

[47] H. A. Zurqani, J. S. Allen, C. J. Post, C. A. Pellett, and T. C. Walker, “Mapping and quantifying agricultural irrigation in heterogeneous landscapes using Google Earth Engine,” *Remote Sens. Appl. Soc. Environ.*, vol. 23, no. December 2020, p. 100590, 2021, doi: 10.1016/j.rsase.2021.100590.

[48] A. Srivastava, S. Bharadwaj, R. Dubey, V. B. Sharma, and S. Biswas, “Mapping Vegetation and Measuring the Performance of Machine Learning Algorithm in Lulc Classification in the Large Area Using Sentinel-2 and Landsat-8 Datasets of Dehradun As a Test Case,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 43, no. B3-2022, pp. 529–535, 2022, doi: 10.5194/isprs-archives-XLIII-B3-2022-529-2022.

[49] G. Forkuor, K. Dimobe, I. Serme, and J. E. Tondoh, “Landsat-8 vs. Sentinel-2:

examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso," *GIScience Remote Sens.*, vol. 55, no. 3, pp. 331–354, 2018, doi: 10.1080/15481603.2017.1370169.

[50] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[51] E. E. Osuna, R. Freund, and F. Girosi, "Support Vector Machines : Training and Applications," vol. 9217041, no. 1602, 1999.

[52] C. El Morr, M. Jammal, H. Ali-Hassan, and W. El-Hallak, "Support Vector Machine," in *International Series in Operations Research and Management Science*, vol. 334, 2022, pp. 385–411. doi: 10.1007/978-3-031-16990-8_13.

[53] R. G. Pontius and M. Millones, "Death to Kappa : Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment," no. August, 2011, doi: 10.1080/01431161.2011.552923.

[54] K. Gupta, "UNPRECEDENTED GROWTH OF DEHRADUN URBAN AREA : A SPATIO-TEMPORAL ANALYSIS," no. October, 2015.

[55] L. Korhonen, Hadi, P. Packalen, and M. Rautiainen, "Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index," *Remote Sens. Environ.*, vol. 195, no. May 2019, pp. 259–274, 2017, doi: 10.1016/j.rse.2017.03.021.

[56] H. Astola, T. Häme, L. Sirro, M. Molinier, and J. Kilpi, "Remote Sensing of Environment Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region," *Remote Sens. Environ.*, vol. 223, no. October 2017, pp.

257–273, 2019, doi: 10.1016/j.rse.2019.01.019.

[57] N. Pahlevan, S. Sarkar, B. A. Franz, S. V. Balasubramanian, and J. He, “Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations,” *Remote Sens. Environ.*, vol. 201, no. August, pp. 47–56, 2017, doi: 10.1016/j.rse.2017.08.033.

[58] H. T. T. Nguyen, T. M. Doan, E. Tomppo, and R. E. McRoberts, “Land Use / Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification,” *Remote Sens*, vol. 12, pp. 1–27, 2020, [Online]. Available: www.mdpi.com/journal/remotesensing

[59] K. J. Salovaara, S. Thessler, R. N. Malik, and H. Tuomisto, “Classification of Amazonian primary rain forest vegetation using Landsat ETM + satellite imagery,” vol. 97, pp. 39–51, 2005, doi: 10.1016/j.rse.2005.04.013.

[60] D. P. Roy *et al.*, “Landsat-8: Science and product vision for terrestrial global change research,” *Remote Sens. Environ.*, vol. 145, pp. 154–172, 2014, doi: 10.1016/j.rse.2014.02.001.

[61] P. S. Roy *et al.*, “Development of decadal (1985-1995-2005) land use and land cover database for India,” *Remote Sens.*, vol. 7, no. 3, pp. 2401–2430, 2015, doi: 10.3390/rs70302401.

[62] A. M. Abdi, “Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data,” *GIScience Remote Sens.*, vol. 57, no. 1, pp. 1–20, 2020, doi: 10.1080/15481603.2019.1650447.

[63] A. Poortinga *et al.*, “An operational before-after-control-impact (BACI) designed

platform for vegetation monitoring at planetary scale," *Remote Sens.*, vol. 10, no. 5, 2018, doi: 10.3390/rs10050760.

[64] Y. Yan, H. Lei, Y. Chen, and B. Zhou, "Article Analyzing the Dynamic Spatiotemporal Changes in Urban Extension across Zhejiang Province Using NPP-VIIRS Nighttime Light Data," *Remote Sens.*, vol. 14, no. 13, 2022, doi: 10.3390/rs14133212.

[65] P. K. Mallupattu, J. Reddy, and S. Reddy, "Analysis of Land Use / Land Cover Changes Using Remote Sensing Data and GIS at an Urban Area , Tirupati , India," vol. 2013, no. Figure 1, pp. 1–7, 2013.

[66] O. R. A. El-kawy, J. K. Rød, H. A. Ismail, and A. S. Suliman, "Land use and land cover change detection in the western Nile delta of Egypt using remote sensing data," *Appl. Geogr.*, vol. 31, no. 2, pp. 483–494, 2011, doi: 10.1016/j.apgeog.2010.10.012.

[67] D. Dutta, A. Rahman, and A. Kundu, "Growth of Dehradun city: An application of linear spectral unmixing (LSU) technique using multi-temporal landsat satellite data sets," *Remote Sens. Appl. Soc. Environ.*, vol. 1, no. July, pp. 98–111, 2015, doi: 10.1016/j.rsase.2015.07.001.

[68] S. Talukdar, P. Singha, S. Mahato, and S. Pal, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations — A Review," 2020.

[69] G. M. Gandhi, S. Parthiban, N. Thummalu, and A. Christy, "Ndvi: Vegetation Change Detection Using Remote Sensing and Gis - A Case Study of Vellore District," *Procedia Comput. Sci.*, vol. 57, no. December, pp. 1199–1210, 2015, doi:

10.1016/j.procs.2015.07.415.

[70] R. T. Vilasan and V. S. Kapse, *Evaluation of the prediction capability of AHP and F-AHP methods in flood susceptibility mapping of Ernakulam district (India)*, vol. 112, no. 2. 2022. doi: 10.1007/s11069-022-05248-4.

[71] R. A. Sedjo and B. Sohngen, “Carbon Credits for Avoided Deforestation,” *Discuss. Pap. Resour. Futur.*, vol. RFF DP 07-, no. April, p. 12, 2007.

[72] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google Earth Engine: Planetary-scale geospatial analysis for everyone,” *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017, doi: 10.1016/j.rse.2017.06.031.

[73] M. Amani *et al.*, “Application of google earth engine cloud computing platform, sentinel imagery, and neural networks for crop mapping in Canada,” *Remote Sens.*, vol. 12, no. 21, pp. 1–18, 2020, doi: 10.3390/rs12213561.

[74] R. S. Ayyangar, P. P. N. Rao, and K. R. Rao, “Crop cover and crop phenological information from red and infrared spectral responses,” *J. Indian Soc. Photo-Interpretation Remote Sens.*, vol. 8, no. 1, pp. 23–29, 1980, doi: 10.1007/BF02990665.

[75] “A Land Use And Land Cover Classification System For Use With Remote Sensor Data,” vol. 2001, 2001.

[76] S. T. Seydi, M. Akhoondzadeh, M. Amani, and S. Mahdavi, “Wildfire damage assessment over australia using sentinel-2 imagery and modis land cover product within the google earth engine cloud platform,” *Remote Sens.*, vol. 13, no. 2, pp. 1–30, 2021, doi: 10.3390/rs13020220.

[77] J. E. Halofsky, D. L. Peterson, and B. J. Harvey, “Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA,” *Fire Ecol.*, vol. 16, no. 1, 2020, doi: 10.1186/s42408-019-0062-8.

[78] H. Astola, T. Häme, L. Sirro, M. Molinier, and J. Kilpi, “Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region,” *Remote Sens. Environ.*, vol. 223, no. January, pp. 257–273, 2019, doi: 10.1016/j.rse.2019.01.019.

[79] J. T. Schoof and S. M. Robeson, “Projecting changes in regional temperature and precipitation extremes in the United States,” *Weather Clim. Extrem.*, vol. 11, pp. 28–40, 2016, doi: 10.1016/j.wace.2015.09.004.

[80] J. E. Hay, D. Easterling, K. L. Ebi, A. Kitoh, and M. Parry, “Introduction to the special issue: Observed and projected changes in weather and climate extremes,” *Weather Clim. Extrem.*, vol. 11, pp. 1–3, 2016, doi: 10.1016/j.wace.2015.08.006.

[81] L. V. Alexander, “Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond,” *Weather Clim. Extrem.*, vol. 11, pp. 4–16, 2016, doi: 10.1016/j.wace.2015.10.007.

[82] ISFR, “India State of Forest Report 2021,” *Minist. Environ. For. Clim. Chang.*, vol. 2021, 2022.

[83] A. Chakraborty, M. V. R. Seshasai, C. S. Reddy, and V. K. Dadhwal, “Persistent negative changes in seasonal greenness over different forest types of India using MODIS time series NDVI data (2001–2014),” *Ecol. Indic.*, vol. 85, no. February, pp. 887–903, 2018, doi: 10.1016/j.ecolind.2017.11.032.

[84] A. R. O'Neill, H. K. Badola, P. P. Dhyani, and S. K. Rana, "Integrating ethnobiological knowledge into biodiversity conservation in the Eastern Himalayas," *J. Ethnobiol. Ethnomed.*, vol. 13, no. 1, 2017, doi: 10.1186/s13002-017-0148-9.

[85] P. Waylen, J. Southworth, C. Gibbes, and H. Tsai, "Time series analysis of land cover change: Developing statistical tools to determine significance of land cover changes in persistence analyses," *Remote Sens.*, vol. 6, no. 5, pp. 4473–4497, 2014, doi: 10.3390/rs6054473.

[86] S. K. McFeeters, "Using the normalized difference water index (ndwi) within a geographic information system to detect swimming pools for mosquito abatement: A practical approach," *Remote Sens.*, vol. 5, no. 7, pp. 3544–3561, 2013, doi: 10.3390/rs5073544.

[87] Z. Zhen *et al.*, "Using the negative soil adjustment factor of soil adjusted vegetation index (Savi) to resist saturation effects and estimate leaf area index (lai) in dense vegetation areas," *Sensors*, vol. 21, no. 6, pp. 1–15, 2021, doi: 10.3390/s21062115.

[88] C. D. R. Foundation, *Main report*. 2020. doi: 10.4324/9781315184487-1.

[89] S. T. Piralilou *et al.*, "A Google Earth Engine Approach for Wildfire Susceptibility Prediction Fusion with Remote Sensing Data of Different Spatial Resolutions," *Remote Sens.*, vol. 14, no. 3, pp. 1–26, 2022, doi: 10.3390/rs14030672.

[90] Kurnaz, "Forest Fire Area Detection by Using Landsat-8 and Sentinel-2 Satellite Images: A Case Study in Mugla, Turkey," vol. 1, no. 2004, pp. 2234–2239, 2007.

[91] E. Roteta, A. Bastarrika, M. Franquesa, and E. Chuvieco, "Landsat and sentinel-2 based burned area mapping tools in google earth engine," *Remote Sens.*, vol. 13, no.

4, pp. 1–30, 2021, doi: 10.3390/rs13040816.

[92] B. Kalantar, N. Ueda, M. O. Idrees, S. Janizadeh, K. Ahmadi, and F. Shabani, “Forest fire susceptibility prediction based on machine learning models with resampling algorithms on remote sensing data,” *Remote Sens.*, vol. 12, no. 22, pp. 1–24, 2020, doi: 10.3390/rs12223682.

[93] A. Tassi and M. Vizzari, “Object-oriented lulc classification in google earth engine combining snic, glcm, and machine learning algorithms,” *Remote Sens.*, vol. 12, no. 22, pp. 1–17, 2020, doi: 10.3390/rs12223776.

[94] R. Manandhar, I. O. A. Odehi, and T. Ancevt, “Improving the accuracy of land use and land cover classification of landsat data using post-classification enhancement,” *Remote Sens.*, vol. 1, no. 3, pp. 330–344, 2009, doi: 10.3390/rs1030330.

[95] A. Srivastava, S. Umrao, and S. Biswas, “Exploring Forest Transformation by Analyzing Spatial-temporal Attributes of Vegetation using Vegetation Indices,” vol. 14, no. 5, 2023.

[96] A. Novo, N. Fariñas-álvarez, J. Martínez-Sánchez, H. González-Jorge, J. M. Fernández-Alonso, and H. Lorenzo, “Mapping forest fire risk—a case study in Galicia (Spain),” *Remote Sens.*, vol. 12, no. 22, pp. 1–21, 2020, doi: 10.3390/rs12223705.

[97] S. Bar, B. R. Parida, A. C. Pandey, and N. Kumar, “Pixel-Based Long-Term (2001–2020) Estimations of Forest Fire Emissions over the Himalaya,” *Remote Sens.*, vol. 14, no. 21, 2022, doi: 10.3390/rs14215302.

[98] Y. Piao, D. Lee, S. Park, H. G. Kim, and Y. Jin, “Forest fire susceptibility

assessment using google earth engine in Gangwon-do, Republic of Korea," *Geomatics, Nat. Hazards Risk*, vol. 13, no. 1, pp. 432–450, 2022, doi: 10.1080/19475705.2022.2030808.

[99] M. A. Brovelli, Y. Sun, and V. Yordanov, "Monitoring forest change in the amazon using multi-temporal remote sensing data and machine learning classification on Google Earth Engine," *ISPRS Int. J. Geo-Information*, vol. 9, no. 10, 2020, doi: 10.3390/ijgi9100580.

[100] D. H. J. Aik, M. H. Ismail, and F. M. Muharam, "Land use/land cover changes and the relationship with land surface temperature using landsat and modis imageries in Cameron Highlands, Malaysia," *Land*, vol. 9, no. 10, pp. 1–23, 2020, doi: 10.3390/land9100372.

[101] A. Srivastava and S. Biswas, "Analyzing Land Cover Changes over Landsat-7 Data using Google Earth Engine," *Proc. 3rd Int. Conf. Artif. Intell. Smart Energy, ICAIS 2023*, no. Icais, pp. 1228–1233, 2023, doi: 10.1109/ICAIS56108.2023.10073795.

[102] R. Ravanelli *et al.*, "Monitoring the Impact of Land Cover Change on Surface Urban Heat Island through Google Earth Engine : Proposal of a Global Methodology , First Applications and Problems," pp. 1–21, doi: 10.3390/rs10091488.

[103] X. X. Yang *et al.*, "A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform," *GIScience Remote Sens.*, vol. 57, no. February, pp. 85–91, 2020, doi: 10.1016/j.isprsjprs.2018.07.017.

[104] L. Shen, Y. He, and X. Guo, "Exploration of Loggerhead Shrike Habitats in

Grassland National Park of Canada Based on in Situ Measurements and Satellite-Derived Adjusted Transformed Soil-Adjusted Vegetation Index (ATSAVI)," *Remote Sens.*, vol. 5, no. 1, pp. 432–453, 2013, doi: 10.3390/rs5010432.

[105] G. Zhou, "Determination of green aboveground biomass in desert steppe using litter-soil-adjusted vegetation index," *Eur. J. Remote Sens.*, vol. 47, no. 1, pp. 611–625, 2014, doi: 10.5721/EuJRS20144734.

[106] E. R. Hunt, W. Dean Hively, S. J. Fujikawa, D. S. Linden, C. S. T. Daughtry, and G. W. McCarty, "Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring," *Remote Sens.*, vol. 2, no. 1, pp. 290–305, 2010, doi: 10.3390/rs2010290.

[107] C. Zhao, W. Chen, Q. Wang, Y. Wu, and B. Yang, "A comparative study of statistical index and certainty factor models in landslide susceptibility mapping: a case study for the Shangzhou District, Shaanxi Province, China," *Arab. J. Geosci.*, vol. 8, no. 11, pp. 9079–9088, 2015, doi: 10.1007/s12517-015-1891-7.

[108] C. Liu *et al.*, "Use of a leaf chlorophyll content index to improve the prediction of above-ground biomass and productivity," *PeerJ*, vol. 2019, no. 1, pp. 1–15, 2019, doi: 10.7717/peerj.6240.

[109] R. Arsada, E. Surojo, D. Ariawan, N. Muhayat, and W. W. Raharjo, "Effect of NBR (Nitrile Butadiene Rubber) on flexural strength of composite friction brake," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, 2018, doi: 10.1088/1757-899X/420/1/012057.

[110] A. Srivastava, S. Umrao, S. Biswas, R. dubey, and M. I. Zafar, "FCCC: Forest Cover

Change Calculator User Interface for Identifying Fire Incidents in Forest Region using Satellite Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 7, pp. 948–959, 2023, doi: 10.14569/IJACSA.2023.01407103.

[111] J. Muro *et al.*, "Short-term change detection in wetlands using Sentinel-1 time series," *Remote Sens.*, vol. 8, no. 10, pp. 1–14, 2016, doi: 10.3390/rs8100795.

[112] P. Engineering, A. Javed, I. Ara, and N. Saleem, "HYLH RI SHFWUDO , QGLFHV IRU," no. July, 2021, doi: 10.14358/PERS.87.7.1.

[113] S. Note, "S a t e l l i t e - b a s e d a s s e s s m e n t of a g r i c u l t u r a l d r o u g h t in karnataka state," vol. 33, no. 3, 2005.

[114] D. Montero, C. Aybar, M. D. Mahecha, F. Martinuzzi, M. Söchting, and S. Wieneke, "A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research," *Sci. Data*, vol. 10, no. 1, pp. 1–20, 2023, doi: 10.1038/s41597-023-02096-0.

[115] R. P. Singh, S. Roy, and F. Kogan, "Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India," *Int. J. Remote Sens.*, vol. 24, no. 22, pp. 4393–4402, 2003, doi: 10.1080/0143116031000084323.

[116] Z. Yang, L. Di, G. Yu, and Z. Chen, "Vegetation condition indices for crop vegetation condition monitoring," *Int. Geosci. Remote Sens. Symp.*, no. July, pp. 3534–3537, 2011, doi: 10.1109/IGARSS.2011.6049984.

[117] H. Ullah *et al.*, "Vegetation assessments under the influence of environmental variables from the Yakhtangay Hill of the Hindu-Himalayan range, North Western Pakistan," *Sci. Rep.*, vol. 12, no. 1, pp. 1–17, 2022, doi: 10.1038/s41598-

022-21097-4.

[118] K. C. Niraj, A. Singh, and D. P. Shukla, "Effect of the Normalized Difference Vegetation Index (NDVI) on GIS-Enabled Bivariate and Multivariate Statistical Models for Landslide Susceptibility Mapping," *J. Indian Soc. Remote Sens.*, vol. 51, no. 8, pp. 1739–1756, 2023, doi: 10.1007/s12524-023-01738-5.

PUBLICATIONS

Journal Paper

1. **Srivastava, A.,** Umrao, S., and Biswas, S.: 2023. "Exploring Forest Transformation by Analyzing Spatial-Temporal Attributes of Vegetation Using Vegetation Indices." *International Journal of Advanced Computer Science and Applications* 14(5):1110–17. doi: 10.14569/IJACSA.2023.01405114 .**(ESCI, Scopus, Impact factor 0.90)**
2. **Srivastava, A.,** Umrao, S., and Biswas, S., Dubey, R., and Zafar, M.I.; 2023. "FCCC: Forest Cover Change Calculator User Interface for Identifying Fire Incidents in Forest Region Using Satellite Data." *International Journal of Advanced Computer Science and Applications* 14(7):948–59. doi: 10.14569/IJACSA.2023.01407103. **(ESCI, Scopus, Impact factor 0.90)**
3. **Srivastava, A.,** Bharadwaj, S., Dubey, R., Sharma, V. B., and Biswas, S.: Mapping Vegetation And Measuring The Performance Of Machine Learning Algorithm In Lulc Classification In The Large Area Using Sentinel-2 And Landsat-8 Datasets Of Dehradun As A Test Case, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3-2022, 529–535, https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-529-2022, **(ISPRS) (citation score: 1.2)(h-index: 71)**
4. Zafar, M.I.; Dubey, R.; Bharadwaj, S.; **Srivastava, A.;** Kumar, A.; Paswan, K.K.; Tiwary, S.K.; Biswas, S. GIS Based Road Traffic Noise Mapping and Assessment of Health Hazards for a Developing Urban Intersection. *Acoustics* **2023**, *5*,87-119. https://doi.org/10.3390/acoustics5010006 **(ESCI, Scopus, Impact factor 2.1)**
5. Dubey, R.; Bharadwaj, S.; Zafar, M.I.; **Srivastava, A.**; Biswas, S. GIS Mapping of Short-Term Noisy Event of Diwali Night in Lucknow City. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 25. https://doi.org/10.3390/ijgi11010025.**((SCIE, Scopus, Impact factor 3.4))**
6. Sharma, V.B.; Dubey, R.; Bharadwaj, S.; **Srivastava, A**.; Biswas, S. Review of Structural Health Monitoring Techniques in Pipeline and Wind Turbine Industries. *Appl. Syst. Innov.* **2021**, *4*, 59. https://doi.org/10.3390/asi4030059.**(ESCI, Scopus, Impact factor 3.8)**
7. Bharadwaj, S., Dubey, R., Zafar, M.I., **Srivastava, A.,** Bhushan Sharma, V., Biswas, S., 2020. Determination of Optimal Location for Setting Up Cell Phone Tower in City Environment Using Lidar Data. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*. XLIII-B4-2020, 647–654. https://doi.org/10.5194/isprs-archives-xliii-b4-2020-647-2020 **(ISPRS) (citation score: 1.2)(h-index: 71)**
8. Sharma, V. B., Dubey, R., Bhatt, A., Bharadwaj, S., **Srivastava, A.,** and Biswas, S.: A Method For Extracting Deformation Features From Terrestrial Laser Scanner 3d Point Clouds Data In Rgipt Building, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLIII-B2-2022, 267–272, https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-267-2022, **(ISPRS) (citation score: 1.2)(h-index: 71)**.

Conference:

9. **Srivastava, Anubhava,** and Susham Biswas. 2023. “Analyzing Land Cover Changes over Landsat-7 Data Using Google Earth Engine.” Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2023 1228–33. doi: 10.1109/ICAIS56108.2023.10073795.

11. **Srivastava, A.**, Dubey, R., Biswas, S. (2023). Comparison of Sentinel and Landsat Data Sets over Lucknow Region Using Gradient Tree Boost Supervised Classifier. In: Noor, A., Saroha, K., Pricop, E., Sen, A., Trivedi, G. (eds) Proceedings of Third Emerging Trends and Technologies on Intelligent Systems. ETTIS 2023. Lecture Notes in Networks and Systems, vol 730. Springer, Singapore. https://doi.org/10.1007/978-981-99-3963-3_18

12. **Anubhava Srivastava,** Susham Biswas. Computation of environmental deterioration by using annual NDVI observation, Taylor & Francis,CRC Press , International conference on Advances in Computational Intelligence and its Applications, (ICACIA-2023).