FLOOD SUSCEPTIBILITY MODELLING USING REMOTE SENSING – MACHINE LEARNING APPROACH AND OPTICAL WATER QUALITY ANALYSIS OF VEMBANAD LAKE SYSTEM IN KERALA, INDIA

Thesis

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

PARTHASARATHY K. S. S.

DEPARTMENT OF WATER RESOURCES AND OCEAN ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALURU – 575 025

OCTOBER, 2023

FLOOD SUSCEPTIBILITY MODELLING USING REMOTE SENSING – MACHINE LEARNING APPROACH AND OPTICAL WATER QUALITY ANALYSIS OF VEMBANAD LAKE SYSTEM IN KERALA, INDIA

Thesis

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

PARTHASARATHY K. S. S.

Under the Guidance of
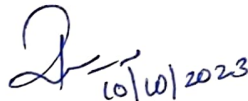
Dr. SUBRAHMANYA KUNDAPURA

DEPARTMENT OF WATER RESOURCES AND OCEAN ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALURU – 575 025

OCTOBER, 2023

DECLARATION

By the Ph.D. Research Scholar

I hereby *declare* that the Research Thesis entitled **Flood Susceptibility Modelling using Remote Sensing – Machine Learning Approach and Optical Water Quality Analysis of Vembanad Lake System in Kerala, India** which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfilment of the requirements for the award of the Degree of **Doctor of Philosophy** in **Water Resources Engineering** is a *bonafide report of the research work* carried out by me. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

10/10/2023

177089AM007, Parthasarathy K. S. S.

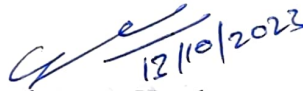(Register Number, Name & Signature of the Research Scholar)

Department of Water Resources and Ocean Engineering

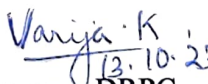Place: **NITK-Surathkal**

Date: **10/10/2023**

CERTIFICATE

This is to certify that the Research Thesis entitled **Flood Susceptibility Modelling using Remote Sensing – Machine Learning Approach and Optical Water Quality Analysis of Vembanad Lake System in Kerala, India** submitted by **Parthasarathy K. S. S. (177089AM007)** as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of degree of **Doctor of Philosophy**.

12/10/2022

Dr Subrahmanya Kundapura

Research Guide

(Name and Signature with Date and Seal)

Varija . K

13. 10. 23

Chairman – DRPC

(Signature with Date and Seal)

Chairman (DRPC)

Dept. of Water Resources & Ocean Engineering

With love and affection

Parthasarathy K. S. S.

ABSTRACT

Wetlands are essential ecosystems that play a significant role in mitigating the impacts of climate change. Wetlands store large amounts of carbon and help to regulate the climate by reducing the amount of carbon dioxide in the atmosphere. They also help to reduce the impacts of extreme weather events, such as floods and hurricanes, by absorbing and retaining water. However, wetlands are also vulnerable to the effects of natural and anthropogenic factors, which can alter their hydrology and lead to the loss of wetland habitats. It is crucial to protect and preserve wetlands to maintain their vital role in mitigating the impacts of climate change. The wetland functions, commodities, and services are lost due to upland land use activities. Hence, accurate and up-to-date information on the upland regions around wetlands is essential. The present research considers the Vembanad Lake System (VLS) in Kerala, India, which is specifically affected by challenging issues to its health and survival. The study area faces threats like encroachment and climate change resulting in floods and alteration in the precipitation patterns. Further, the lake system is endangered by the deteriorating quality of incoming water. Thus, the overall spatio-temporal analysis is critical in protecting and managing water resources in the study region.

Anthropogenic activities result in a massive Land Use and Land Cover (LULC) change, and it has become a prominent issue for decision planners and conservationists due to inappropriate growth and its effect on natural ecosystems. As a result, the change in LULC for the short term, i.e., within a decade, is carried out using three Machine Learning (ML) approaches, Random Forest (RF), Classification And Regression Trees (CART), and Support Vector Machine (SVM), on the Google Earth Engine (GEE) platform. When comparing the three techniques, SVM performed poorly at an average accuracy of around 82.5%, CART being the next at 87.5%, and the RF model being good at an average of 89.5%. The RF outperformed the SVM and CART in almost identical spectral classes, such as barren land and built-up areas. As a result, RF-classified LULC is considered to predict the Spatio-temporal distribution of LULC transition analysis for 2035 and 2050. This analysis was conducted in Idrisi TerrSet software using the Cellular Automata (CA) - Markov chain analysis. The model's efficiency is evaluated by comparing the projected 2019 image to the actual 2019

classified image. The model efficiency obtained was good, with more than 94.5% accuracy for the classes except for barren land, which might have resulted from the recent natural calamities and the accelerated anthropogenic activity in the study area. Floods have claimed the lives of countless people and caused significant property damage, putting their livelihoods in jeopardy. The study area faced adverse mishappening during the 2018, 2019, and 2021 floods due to the torrential rainfall events. Estimations of flood-inundated areas are prepared from 2018, 2035, and 2050 LULC maps. The extent of flood inundation during the 2018 floods and the possible flood inundation region for the projected LULC in 2035 and 2050 are determined. From the analysis of the 2018 classified image, 14.7 km^2 of built-up area was found inundated during the year 2018 floods. The scenario of the 2018 flood event is used to quantify the flood that may occur and inundate the projected LULC 2035 and 2050 scenarios. It is found that the flood will affect about 19.87 km^2 and 23.32 km^2 of the built-up region, majorly for the 2035 and 2050 projected scenarios, respectively. The goal of this research is to construct effective decision tree-based ML models such as Adaptive Boosting (AdaBoost), RF, Gradient Boosting Machines (GBM), and Extreme Gradient Boosting (XGBoost) for integrating data, processing and generating flood susceptibility maps. Eighteen conditioning parameters, including seven categorical and eleven numerical data, are used for flood modelling using ML. These seven categorical data are converted into 50 numerical data, resulting in a total input data of 61. The Recursive Feature Elimination (RFE) is utilized as the feature selection technique, and 22 layers are chosen to feed into the ML models to generate the flood susceptibility maps. The efficiencies of the models are evaluated using Receiver Operating Characteristic – Area Under Curve (ROC-AUC), F1 score, Accuracy, and Kappa. According to the results obtained, all four ML models demonstrated fairly good performance. However, XGBoost fared well in terms of the model's metrics. The ROC-AUC values of XGBoost, GBM, and AdaBoost for the testing dataset are 0.90, whereas 0.89 for RF. The accuracy varied significantly among the four models, with XGBoost scoring 0.92, followed by GBM (0.88), RF (0.87), and AdaBoost (0.87). The resulting flood susceptibility map can be utilized for early mitigation actions during future floods and for land use planners and emergency managers, assisting in reducing flood risk in regions prone to this hazard.

Water quality is one of the essential parameters of environmental monitoring; even a slight variation in its characteristics may significantly influence the ecosystem. The water quality of Vembanad Lake is affected by anthropogenic effects such as industrial effluents and tourism. The optical parameters representing water quality, such as diffuse attenuation (K_d), turbidity, Suspended Particulate Matter (SPM), and Chlorophyll-a (Chl-a), are considered in this study to evaluate the water quality of the Vembanad Lake. As this lake is regarded as of ecological importance by the Ramsar Convention and has faced severe concerns over recent years, there was a substantial change in the water quality during the lockdowns of the COVID-19 pandemic. This research aimed to examine the change in water quality using optical data from Sentinel-2 satellites in the ACOLITE processing software from 2016 to 2021. The analyses showed a 2.5% decrease in the values of K_d, whereas SPM and turbidity show a reduction of about 4.3% from the year 2016 to 2021. The flood and the COVID lockdown had an impact on the improvement in the quality of water from 2018 to 2021. The findings indicated that the reduction in industrial activities and tourism had a more significant effect on the improvement in the water quality of the lake. There was no substantial change in the Chl-a until 2020, whereas an average decrease of 12% in Chl-a values was observed throughout 2021. This decrease can be attributed to the reduction in the lake's Hydrological Residence Time (HRT).

The outcome of this research depicts augmentation of the change in the LULC pattern and its prediction, future flood-inundation regions, flood susceptibility mapping, and the lake's water quality. The findings of this research work will be a valuable reference to help the government and Non-Government Organisations (NGOs) during strategic planning.

Keywords: Google Earth Engine; Machine Learning; Kerala floods; LULC prediction; CA-Markov chain analysis; Flood susceptibility mapping; Optical property; ACOLITE processing; Chlorophyll-a.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
ACOLITE	Atmospheric correction for OLI 'lite'
AdaBoost	Adaptive Boosting
API	Application Program Interface
AUC	Area Under Curve
BSI	Bare Soil Index
C	Cost
CA	Cellular Automata
CART	Classification And Regression Trees
CDOM	Coloured Dissolved Organic Matter
cHABs	Harmful cyanobacterial blooms
CHIRPS	Climate Hazards group InfraRed Precipitation with Station
Chl-a	Chlorophyll-a
DEM	Digital Elevation Model
DN	Digital Number
DSF	Dark Spectrum Fitting
ERTS	Earth Resources Technology Satellite
FMA	Flood Mapping Algorithm
FN	False Negative
FNU	Formazin Nephelometric Unit
FP	False Positive
FPAR	Fraction of absorbed Photosynthetically Active Radiation
FRL	Full Reservoir Level
FSM	Flood Susceptibility Modelling
GBM	Gradient Boosting Machines
GEE	Google Earth Engine
GHG	Greenhouse Gas
GIS	Geographical Information System
GPS	Global Positioning System
GRD	Ground Range Detected

HABs	Harmful Algal Blooms
HRT	Hydrological Residence Time
IDE	Interactive Development Environment
IOPs	Inherent Optical Properties
IW	Interferometric Wide
IWMI	International Water Management Institute
K_d	Diffuse Attenuation
l2gen	Level 2 data processing algorithm
LCM	Land Change Modeller
LULC	Land Use and Land Cover
LULCC	Land Use and Land Cover Changes
MCB	Mini Cloud Burst
MDA	Mean Decrease in Accuracy
MDG	Mean Decrease in Gini
ML	Machine Learning
MNDWI	Modified Normalized Water Index
MsCB	Mesoscale Cloudburst
MSL	Mean Sea Level
NAP	Non-Algal Particles
NDBI	Normalized Difference Built-up Index
NDCI	Normalized Differential Chlorophyll Index
NDVI	Normalized Difference Vegetation Index
netCDF	network Common Data Form
OA	Overall Accuracy
plan_c	Plan Curvature
prof_c	Profile Curvature
QAA	Quasi-Analytical Algorithm
RBINS	Royal Belgian Institute of Natural Sciences
RF	Random Forest
RFE	Recursive Feature Elimination
ROI	Region Of Interest

SAR	Synthetic Aperture Radar
SDG	Sustainable Development Goals
Sen2Cor	Sentinel-2 Correction
SLC	Single Look Complex
SPI	Stream Power Index
SPM	Suspended Particulate Matter
STI	Sediment Transport Index
SVM	Support Vector Machine
TDS	Total Dissolved Solids
TN	True Negative
TOA	Top of Atmosphere
TP	True Positive
TRI	Topographic Ruggedness Index
TWI	Topographic Wetness Index
VLS	Vembanad Lake System
WoS	Web of Science
XGBoost	Extreme Gradient Boosting
κ	Kappa statistics

CHAPTER 1

INTRODUCTION

1.1   GENERAL

Climate plays a critical function in sustaining the ecosystem; climate change and global warming threaten humanity's social and economic structure (Kocsis et al. 2020). Climate change has garnered international attention as it influences every facet of society. It is reasonable to anticipate that climate change, in addition to its inherent variability, will impact meteorological parameters, which will in turn, cause changes in the global and regional hydrological cycle (Mohammed and Scholz 2019). These shifts can affect the region's ability to maintain its ecology and water resources. It is becoming one of the most significant obstacles to attaining food, energy, and water security on a global scale, and the effect is greatest in developing nations owing to their weak adaptive ability and inadequate natural resource management (Adhikari et al. 2015). It also significantly affects earth systems, resulting in unexpected weather phenomena such as precipitation and high temperatures that are more frequent (Gebrechorkos et al. 2020). This causes environmental issues, such as localized downpours, flooding, droughts, and heat waves, which may result in numerous fatalities. The increase in the occurrence and intensity of severe events is one of the most significant effects of global warming. According to Wang et al. (2019), the direct influence of climate change on the hydrological cycle is owing to an increase in temperature and its high variability and a shift in rainfall at various scales. One of the major causes of climate change is the rapid change in land use and land cover (LULC), industrialization, and urbanization. Over the last few decades, economic prosperity and population growth have resulted in unplanned urbanization and industrialization to meet livelihood and job needs (Jose and Dwarakish 2020). Wetlands can reduce greenhouse gas (GHG) production by controlling processes like decomposition because of their dense vegetation, algal activity, and soil composition. The impact of climate change on wetland ecosystems could be irreversible, but wetland ecosystems can help mitigate the effects of climate change.

1.2 IMPORTANCE OF WETLANDS

Wetlands are important in providing beneficial activities for humans and wildlife. They provide a large volume of food that attracts a large volume of animal species, remain a "Biological supermarket", and help improve the ground and surface water quality, as well as release the vegetative matter into the rivers; this helps in feeding the fish. They help to rejuvenate the river that got polluted and helps in balancing the ecosystem (Figure 1.1). Wetlands filter out the sedimentation, decompose the vegetative matter and purify the water. They also help to recycle nutrients balancing the nutrient cycle and creating a productive ecosystem. Wetlands are recognized globally as important ecological infrastructure because they offer many critical ecosystem services to people. Due to their significance, many wetlands have been exploited or utilized in an unsustainable manner, resulting in rapid worldwide wetland reductions (Mitsch and Gosselink 2000; Rebelo et al. 2017; Russi et al. 2013; Simonit and Perrings 2011). According to scientific estimates, 64% of the world's wetlands have vanished since 1900 (Wetlands International 2014). Thus in the 1970s, the international treaty called the "Convention on Wetlands of International Importance", popularly known as the "Ramsar Convention of 1971", has globally come into force for the conservation and wise use of wet or watery lands (Tatu and Anderson 2017). This intergovernmental treaty defines the term wetlands as "areas of marsh, fen, peatland or water, whether natural or artificial; standing or flowing; fresh, brackish or salt, including areas of marine water, the depth of which at low tide does not exceed six meters" (Navid 1989). As a result of the inclusion of the word "water" in the definition, the Ramsar Convention of 1971's definition of the wetland is almost all-encompassing. All lakes, rivers, subterranean aquifers, swamps and marshes, wet grasslands, peatlands, oases, estuaries, deltas and tidal flats, mangroves, and other coastal habitats, coral reefs, and artificial wetlands are included. The terms "wetlands" and "waterbodies" are interchangeable, with the exception that in marine/coastal regions, the word "wetland" refers to only those places where the depth of the water does not exceed 6 m at low tide. The 6 m depth restriction was most likely devised to represent the approximate greatest depth to which sea ducks descend to feed (Anderson and Davis 2013).
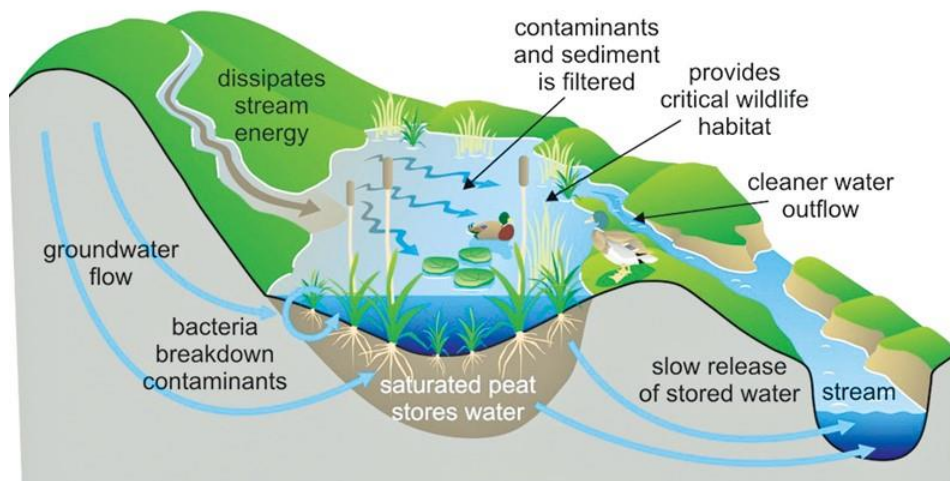
Figure 1.1 Functioning of wetlands

(Source: Wetlands International (2011))

Countries that have accepted the Ramsar Convention must undertake planning to encourage the sustainable use of wetlands and create management and conservation strategies (Bott 1994). Even though 169 nations have signed this agreement, there are still negative tendencies; wetlands are being destroyed or degraded, and wetland species populations are diminishing. Furthermore, the Convention on Biological Diversity (CBD 1992) requires contracting parties to rehabilitate and repair damaged ecosystems and manage biological resources critical to biological diversity protection (Glowka et al. 1994). On a worldwide scale, wetland surface area is almost twice that of lakes and reservoirs (Lehner and Döll 2004), and their size is anticipated to fluctuate seasonally (Papa et al. 2010). Wetland responses are therefore an important part of sustainable water resource management, and their involvement in regional water balances and ecosystem preservation must be assessed and understood.

Sustainable Development Goals (SDG) are defined by United Nations under goal 6 as "clean water and sanitation" to ensure clean water is accessible for all as it is a basic human need. SDG 6.6 states, "By 2020, protect and restore water-related ecosystems, including mountains, forests, wetlands, rivers, aquifers and lakes". The goal is expected to be met by 2030, which can be achieved effectively by regularly monitoring and assessing the changes in the lake due to anthropogenic and other external factors. Due to the remoteness and vastness of many wetlands, frequent data gathering to understand

3

their eco-hydrology is extremely costly and impossible (Finlayson 2003). The threat posed by humans to wetland ecosystems is global (Mitsch and Gosselink 2000). Due to the fast growth in world population and the need for natural resources for food, fuel, and fodder, there has been a significant degradation and reduction in all kinds of wetlands during the last few decades (Brij and Krishnamurthy 1993). Wetlands are used as cesspools for the disposal of solid waste as well as the discharge of wastewater. The loss is significantly greater in certain areas since wetlands' degradation and loss are global issues.

1.3 LAND USE AND LAND COVER

1.3.1 Impacts of change in land use and land cover

The natural and anthropogenic activities worldwide influence the land cover, modifying its landscapes and the subsequent dynamics of natural processes (Silva et al. 2020). Monitoring and assessing urban growth aid in planning and utilizing natural resources for the near future (Mandal et al. 2023). Anthropogenic processes have altered almost half of the Earth's land surfaces (Tayyebi and Pijanowski 2014). These changes are called Land Use and Land Cover Changes (LULCC). The enormous need increased the demand for critical infrastructures such as water supplies, sewage services, and recreational activities. It also causes road congestion, pollution, climate change–related problems, urban floods, and Urban Heat Island (UHI) impacts (Jose and Dwarakish 2022; Saxena et al. 2021). As a result, LULC change is regarded as a critical environmental issue of global significance. Major causes of biodiversity loss and related habitat loss are human-induced effects (Elias et al. 2021), such as urbanization, erosion, overgrazing, and land degradation (Abijith et al. 2021). Besides, nature also contributes to this alteration (Halmy et al. 2015a; Lambin 1997). The complex interaction of the factors like policy management, human needs, environment, culture, and economics results in changing LULC (AlDousari et al. 2023). Alteration in the LULC can significantly alter the water quality as its need increases in urban and agricultural activities resulting in nitrates and phosphates in the freshwater (Álvarez-Cabria et al. 2016; Krishnaraj and Deka 2020).

Rapid urbanization in different parts of cities has been a severe threat to developing countries such as India, Indonesia, Malaysia, and Sri Lanka. India is the most populous

country in the world, with 1.38 billion people. The Composite Water Management Index (CWMI) report released by the National Institution for Transforming India (NITI) Aayog in August 2019 under the Government of India (GOI) stated that many cities in India, such as Delhi, Bengaluru, and Chennai, may face "Day Zero" in the upcoming years (Abijith et al. 2020). Due to the country's rapid population growth and subsequent urbanization, policies for planning, analysing, and tracking land use transitions are required to meet the people's basic needs (Darem et al. 2023). As a result, enormous areas of forest cover are being converted to other land uses, causing severe soil erosion. Rapid soil erosion will lead to catastrophic floods and landslides affecting downstream residents. Thus, a sustainable LULC is critical for long-term livelihood and environmental improvement (Mishra et al. 2020).

1.3.2 Cellular Automata–Markov chain analysis

As the population grows, so does the need for land. As the demand for land increases, it eventually increases in urban (Kumar and Agrawal 2023). This further increases urban areas on LULC, causing a disturbance in the ecosystem and affecting sustainability (Aburas et al. 2018). Thus, the Cellular Automata (CA)–Markov chain model is used to understand the factors affecting the spatio-temporal distribution of LULC and to predict future LULC changes. It is one of the most widely used models to predict LULC change (Ozturk 2015). This model was created using Idrisi TerrSet's Land Change Modeller (LCM). To forecast the change in LULC, this hybrid model combines the CA with the Markov chain model (Aburas et al. 2018). As the Markov chain model is stochastic, it predicts one cell transformation to another, i.e., the transition probability of the cell. However, the disadvantage of the Markov chain model is that it does not consider the effect of the neighbouring cells over the other. Thus, it lacks spatial modelling capability (Ozturk 2015). CA model considers only the neighbouring cell of interest to estimate the future. Thus, both models combined to make the CA–Markov model to analyse the spatio-temporal changes in the land cover. It takes two time periods of LULC (i.e., earlier and later dates) as the input to analyse the change in trend. This software aids in analysing and developing models in the event of stable land cover as opposed to a rapidly changing environment. LCM facilitates the comparison of LULC categories and the net change observed by every class and the

contributor to the net change observed by every other LULC category (Hamad et al. 2018a). Predicting LULC transition has been used in several applications, including environmental planning by modelling rural development and urban growth, identifying conservation target areas and establishing alternate conservation strategies, analysing the dynamics of changing agriculture, and simulating rangeland dynamics under various climate change scenarios (Halmy et al. 2015a).

1.4 FLOODS

1.4.1 Impact of Flood

Floods are a common natural disaster that has a massive influence on natural and human ecosystems all over the globe. It results from unfavourable climatic, hydrological, and physical factors. At the lower magnitudes, it is advantageous to ecosystems in several ways, such as providing water and nutrients to riparian corridors, removing contaminants from floodplain zones, recharging the groundwater table, and improving soil fertility (Abijith et al. 2020; Kraus et al. 2019; Lyubimova et al. 2016; Mahato et al. 2021; Zhang et al. 2017). When the flood tolerance limit is surpassed, it will create colossal damage to lives, agricultural activities, residential areas, properties, highways, and natural habitats (George et al. 2022; Pradhan and Youssef 2011). Because of the large amounts of property damage and economic loss associated with flooding, flood is classified as a major natural hazard (Papaioannou et al. 2015). Floods cause tens of thousands of deaths and millions of dollars in damage to the economy each year (Aerts et al. 2018; Pham et al. 2021). Over the last two decades, the frequency of floods has increased globally (Alfieri et al. 2017). Floods are typically classified into five kinds based on their location, cause, and influencing factors: urban drainage, riverine flooding, ground failures, fluctuating lake levels, and coastal flooding and erosion (Karamouz and Zahmatkesh 2016; Wright 2008). According to several studies, floods affect roughly 200 million people worldwide yearly (Bui et al. 2019; Towfiqul Islam et al. 2021). Between 1995 and 2015, the United Nations Office for Disaster Risk Reduction (UNDRR) stated that 150,061 flood occurrences occurred globally, accounting for 11.1% of global damage (Chakrabortty et al. 2022). Flood-induced hazards are likely to increase as the processes in the hydrological cycle get altered because of climate change leading to the risk to lives and properties around the globe.

Moreover, owing to the change in climate, increase in population, and unsustainable land use development, it is expected that the flood occurrence rates and severity may increase by 2050, potentially causing a massive loss of about ≈1 trillion USD (Alexander et al. 2019; Ali et al. 2019; Bubeck and Thieken 2018; Koc and Işık 2021; Tehrany et al. 2019). The increased urbanization causes the surface to be sealed, resulting in less water seepage into the subsurface. Furthermore, the construction of dykes leads to the loss of natural storage by blocking off the river from the floodplain, putting strain on the ecosystem and perhaps leading to a catastrophe and deterioration (Abijith et al. 2020). The deterioration of the environment is the reduction of its resources, such as air, water, and soil, as well as the destruction of ecosystems, habitats, and natural balance (Parthasarathy et al. 2022). Landslides and flash floods are difficult to anticipate and manage, but well-planned mitigation measures may greatly minimize the danger to lives and property (Pal et al. 2018). Natural catastrophes are becoming more frequent and intense on the earth's surface, causing significant economic damage to property and services (Mohanty et al. 2019).

1.4.2 Floods in Indian Scenario

Floods are the most common among all other natural calamities in India. Over 40 million hectares of land are at risk of flooding in one-eighth of the country's geographical region (Chowdhuri et al. 2020). India has several perennial and non-perennial rivers, and its climate is impacted by monsoons, despite being a subtropical country (Abijith and Saravanan 2022; Chakrabortty et al. 2022; Kumar and Bhattacharjya 2020). The seasonal and temporal variation in the rainfall patterns causes floods during the monsoon seasons (Rakesh Kumar et al. 2005). Furthermore, siltation on riverbeds, inadequate capacity on riverbanks to sustain large flows, changes in the river's course, dam failure, inadequate drainage in flood-prone locations, and glacial outbursts are all sensitive elements contribute to flood occurrence (Witzany et al. 2008). Due to its monsoonal rainfall variance, large sediment flow, and other factors, India ranks second in floods, only after Bangladesh (Sarkar and Mondal 2020). According to the World Resources Institute (WRI), India tops the list (Figure 1.2), with 4.84 million people affected by river floods yearly (Luo et al. 2015). Due to their geomorphological positions and climatic circumstances, several states, including Assam, West Bengal,

7

Uttar Pradesh, Bihar, and Kerala, are regularly flooded (Bhattacharjee and Behera 2017).

15 Countries Account for 80% of Population Exposed to River Flood Risk Worldwide

Country	Value
India	4.84
Bangladesh	3.48
China	3.28
Vietnam	0.93
Pakistan	0.71
Indonesia	0.64
Egypt	0.46
Myanmar	0.39
Afghanistan	0.33
Nigeria	0.29
Brazil	0.27
Thailand	0.25
Congo, Dem. Rep.	0.24
Iraq	0.19
Cambodia	0.19
Rest of the World	4.24

The top 15 countries with greatest population exposed to river flood risk are either least developed or developing countries, which are the most vulnerable to natural disasters and climate change

Annual Expected Population Affected by River Floods (millions)

Figure 1.2 Annual expected population affected by floods

(Source: WRI India)

According to the Flood Damage Statistics (2018), 347,581 crore rupees of crops, houses, and public utilities were damaged by floods and extreme rainfall in India from 1953 to 2016. Eastern and Western Ghats in South India are among the most affected regions due to floods during the Northeast and Southwest monsoon periods. In August 2018 flood that occurred in the Indian state of Kerala took away the lives of 504 people, and almost 23 million people got affected due to the flood. Subsequently, during the monsoon seasons of 2019, 2020, and 2021, the state has also seen significant flooding. Hence, there is a need for Flood Susceptibility Modelling (FSM) to understand the areas that are inundated frequently during above-average rainfall. To make flood management easier, there is a significant necessity for regular flood power estimations (Abijith et al. 2021; George et al. 2021; Madsen et al. 2014). Most researchers emphasize peak discharge in analysing floods' effectiveness, frequency, and amplitude (Abrishamchi et al. 2011; Cameron et al. 2000; Das 2019; Elias et al. 2022). However, multiple elements such as flow competency, stream power, sequence of events, channel

geometry, and others influence the geomorphic efficacy of an intense flood (Flax et al. 2002; Tang et al. 2021a). The consideration of these factors helps in better prediction of susceptibilities and consequently reduces the fatalities and major economic losses.

1.5 WATER QUALITY ANALYSIS

1.5.1 Importance of Water

Water is essential in many facets of life, and its quality is deteriorating due to increased pollution caused by urbanization, industrialization, and population growth. It is necessary to identify water constituents causing pollution to preserve the water quality. Water quality is a key indicator of environmental health since even small changes in their properties may have far-reaching consequences for the ecosystem (Bhuyan et al. 2020). More than 40% of the world's population lives near lakes, rivers, or coastal waters, which is still increasing. Continuous water quality monitoring may provide light on the human-induced pressures and stresses on an ecosystem. Assessing the change in water quality is often time-consuming and requires on-field measurements, laboratory and temporal statistical analysis. The projected effect of water shortage and water quality would be the major concern in this twenty-first century, especially in developing countries (Krishnaraj and Honnasiddaiah 2022; UN WWDR 2022). The International Water Management Institute (IWMI 2014) reports that more than a quarter of the world's population does not have access to fresh water. Therefore, over 3 million people, primarily children, suffer yearly from polluted water. Providing people in underdeveloped countries with easy access to freshwater improves their literacy, equality, and health; it also helps them minimize hunger and boosts their economies (Hope 2011). In-situ ways of water monitoring are limited, and implementation is resource-intensive, requiring time, labour, and funds (Andres et al. 2018). Remote Sensing techniques overcome these obstacles due to their synoptic coverage, allowing the monitoring of waterbodies with high temporal accuracy (Antonini et al. 2017; Codd et al. 2005; Gholizadeh et al. 2016; Krishnaraj and Honnasiddaiah 2022). Anthropogenic causes such as increased water consumption and climate change will exacerbate water shortages, jeopardizing the accomplishment of the majority of the Millennium Development Goals (Ahmed et al. 2020; Schewe et al. 2014).

Cyanobacterial Harmful Algal Blooms (cHABs) have a severe social and ecological effect (Carmichael and Boyer 2016; Falconer 1999).

1.5.2   Influencing factors of water quality

The optical property is analysed to detect the change in the lake's water quality pattern. The biophysical property such as Chl-a, an indicator of phytoplankton biomass, is commonly used to assess the quality of the wetlands. Knowledge about these enables us to improve the monitoring of these unique areas, which are otherwise non-accessible using site visits. This can be studied by analysing the optical properties of the water. Morel and Prieur (1977) classified waterbodies into two cases. Case 1 waters have optical qualities predominantly determined by phytoplankton and related Coloured Dissolved Organic Matter (CDOM) and detritus by-products. Case 2 waters are those whose optical properties are considerably impacted by other constituents such as mineral particles, CDOM, or microbubbles, the quantities of which do not fluctuate with the abundance of phytoplankton. Two significant factors exist in the irradiance over water as measured by multispectral imagery. They are (i) the water itself, owing to sunlight's interaction with optically active components such as clear water, CDOM, Non-Algal Particles (NAP), phytoplankton, and suspended sediments; and (ii) the atmosphere, which is mainly composed of molecules and gases; it thus accounts for roughly ninety percentage of the signal (IOCCG 2010; Pereira-Sandoval et al. 2019). The duration of the water spent in any linked network segment is known as the Hydrological Residence Time (HRT). The residency time of a body of water increases with the duration it spends in a particular system (such as a river, lake, pond, etc.). Numerous hydrological domains, such as water quality analysis, stratification, wetland ecology, age dating, water mixing and circulation, microbiological pollutants, etc., have significant applications for HRT (Avtar et al. 2020; Leray et al. 2016; Zwart et al. 2017). Chlorophyll-a (Chl-a), for example, was linked to the HRT, as León et al. (2016) reported. Stumpner et al. (2020) found that Chl-a concentrations are highest in areas with long HRT (15-60 days) and lowest in areas with short HRT (1-14 days).

1.6   ROLE OF GEOSPATIAL TECHNOLOGY

Geospatial technology is a significant finding of humanity that has evolved since prehistoric times. The main types of geospatial technology involve remote sensing,

Geographical Information System (GIS), and Global Positioning System (GPS). Geospatial technology benefits the user by offering greater synoptic coverage of the area of interest and remains a viable technique for evaluating multi-temporal satellite data to estimate periodical changes (Saravanan et al. 2015). Remote Sensing applications are critical in monitoring hazard-related activities in real-time. Because of its very long spectral bands and high spatial resolutions, remote sensing imagery may detect even minor changes in land use. GIS is a computer system that captures, stores, queries, analyses, and displays geographic data for inventory and decision-making purposes. Using GIS, all hazard-related data may be converted into visual information as a map (Thomas et al. 2007). These satellite imageries have attracted much interest in mapping the vulnerabilities due to the eagle-eye perspective in conjunction with a multispectral and temporal resolution (Sudha Rani et al. 2015). With remote sensing and GIS, better forecasting of disaster-prone areas and security measures may be accomplished (Lawal et al. 2011). The primary goal of GIS is to make it helpful to the general public so that their participation may improve the requirements of investors. This is referred to as Participatory GIS (PGIS). Including commercial and open-source software in this technology enables the user to access a specific degree of programming interface (Jankowski 2009).

1.6.1   Google Earth Engine

Google Earth Engine (GEE) is a multi-petabyte geospatial data collection co-located with high-performance and intrinsically parallel computation service. In recent years, it has been in the remote sensing big data processing spotlight. GEE is an Interactive Development Environment (IDE) to enable rapid visualization and analysis controlled through an Application Program Interface (API) accessed via the internet. It contains large publicly available satellite and aerial imagery datasets in both optical and non-optical wavelengths. It comprises the most freely available remote sensing imagery of the entire Landsat archives, Sentinel-1 and Sentinel-2 (Gorelick et al. 2017). The major advantage of GEE is that it provides large data sets, including land cover, environmental variables, weather, and climate forecast, which is already pre-processed to access the data. Besides, the raw data is also pre-processed, cloud removed, and mosaicked in the GEE to reduce computational time. The client libraries handle the Earth engine code

editor via JavaScript and Python (Gomes et al. 2020). GEE utilizes the parallel processing technique using MapReduce architecture. It is a method to process a large amount of data into several smaller chunks in several machines. Thus, the data processed as several chunks were recompiled for the result. As the data can be accessed through the API, it is less labour-intensive and requires less storage space to save the data. It possesses simple yet effective architecture which does not require high-power computing machines (Noi Phan et al. 2020). Thus, the use of GEE has significantly increased within the remote sensing community (Tamiminia et al. 2020).

1.6.2 Application in LULC

Remote Sensing is an excellent tool for regularly monitoring and assessing the LULC change in the natural ecosystem and landforms' evolution by analysing the geomorphological changes. Likewise, it has played a major role in studying geography, geology, and the environment for researchers and scientists (Ghosh et al. 2017). It is also more cost-effective and less time-consuming than the traditional method. It has very long spectral bands, and good spatial resolution helps distinguish a significant change in land cover (Abijith et al. 2020). One of the most extensively utilized sources for analysis is satellite images. In 1972, Earth Resources Technology Satellite (ERTS)-1 was launched and renamed "Landsat" in 1975. Landsat has launched nine series of satellites and contains data contiguity for almost 50 years. As of now, Landsat-7, 8, and 9 are functional. This series of satellites has become one of the important long-term freely available data for the civilian purpose and has been used widely in fields like coastal monitoring (Parthasarathy et al. 2022), LULC (Shi and Yang 2015), vegetation phenology (Senf et al. 2017), and hydrology (Abijith et al. 2020). Thus, Landsat offers a deeper understanding of LULC changes for better decision-making and resource management. Several methods for detecting change using remotely sensed data have been established during the last three decades (Hua et al. 2014; Jat et al. 2017; Rienow and Goetzke 2015; Saxena and Jat 2020; Serasinghe Pathiranage et al. 2018).

1.6.3 Application in Flood Analysis

Flood inundation and spatially distributed hydrological processes in urban environments can be difficult to model using hydraulic and hydrological-based models, typically data-scarce and ungauged, especially in highly varied urbanized areas (Darabi

et al. 2022). Remote Sensing and GIS are excellent tools for computing massive amounts of geographical data for spatial management and analysis (Yang et al. 2020; Zhang et al. 2019b; a). New flood prediction models have been developed due to recent advances in GIS and remote sensing techniques. Its application is not only limited to flood analysis but also helps analyse landslides, forest fires, and cyclones (Pal et al. 2017). The Synthetic Aperture Radar (SAR) in remote sensing is a primary tool for real-time assessment of flooded areas as it can penetrate the cloud, haze, and rain (Jacinth Jennifer et al. 2020). In contrast, traditional ground and aerial surveys are time-consuming and require skilled labourers for flood mapping. On the other hand, ground surveys may also pose a great threat to human life during extreme events (Rahman and Thakur 2018). Besides its penetration capacity, the most important advantage of employing SAR data is that land and water contrast can be easily identified (Dewan et al. 2006).

1.6.4 Remote Sensing of water quality

Remote Sensing observations help us better understand ecological processes, severe weather events, and natural hazards. This could be because the technology is getting better, and therefore more data becomes accessible, which can enhance research and environmental monitoring products systematically. The remote sensing approach identifies differences in the spectral characteristics backscattered from water and also aids in relating observed variations to a water quality parameter using empirical or analytical models. It has been utilized to reflect the spatial pattern and temporal changes in water quality components due to spatial and temporal coverage benefits. Evaluating the quality of waterbodies in inaccessible regions is one of the benefits of remote sensing (Ansper and Alikas 2019). Satellite remote sensing is a key tool to analyse wetland changes by continuous monitoring of phytoplankton and cyanobacterial pigments such as Chl-a and phycocyanin (PC), SPM, CDOM, and the diffuse attenuation coefficient (K_d) (Bhuyan et al. 2020; Caballero et al. 2020, 2022; Vanhellemont and Ruddick 2016b). The optimum wavelength for sensing a water quality parameter is determined by the tested material, its concentration, and the sensor's features (Yang et al. 2022). Various space-borne sensors at different wavelengths can monitor water quality due to high-frequency data collection and

synoptic coverage. SAR is useful in estimating surface temperature and water salinity in the microwave region. However, water quality retrieval has been effectively and extensively employed in the visible and infrared domains of remotely sensed data (Zhang et al. 2003). Optical satellite data can be used to manage water resources, especially during and after a natural disaster, by measuring the influence on the environment and comprehending the processes that could enable sustainable development in data-scarce areas. The presence of certain substances in surface water can considerably alter the features backscattering of surface water.

1.7   SCOPE OF THE WORK

Wetlands face issues such as siltation, pollution, cultural eutrophication, choking of open waterways by abundant hydrophytic vegetation, encroachment of waterways and basins for commercial agriculture, and deterioration caused by filling, dredging, and draining. This is despite the fact that wetlands provide a variety of ecological and socioeconomic services that benefit humanity. As a result, wetland protection is critical. Current data on the upland areas around wetlands is crucial because land use activities on uplands result in the loss of wetland functions, commodities, and services. This research has explored the efficiency of different Machine Learning (ML) algorithms in classifying LULC and FSM applications. The study area considered for the research is Vembanad Lake System (VLS) in Kerala, India. Major cities in the state of Kerala border the wetland of VLS. Hence, change in the LULC is predicted to analyse the urbanization trends through CA-Markov analysis with the help of various driving variables. The region has also experienced repeated flooding in the last five years, making it more vulnerable to flooding. Thus, the FSM is formulated using eighteen variables to understand the region prone to flooding. In addition, the number of regions submerged in various land use classes has also been identified to understand the amount of flood inundation in the region. The optical characteristic of the lake was studied from 2016 to 2021 using Sentinel-2 satellite images. This aids in investigating changes in the optical water quality parameters such as SPM, turbidity, K_d, and Chl-a in the lake during floods and changes during the COVID-19 lockdowns.

1.8   ORGANIZATION OF THE THESIS

The structure of this thesis is comprised of seven chapters.

Chapter 1: This chapter provides an introduction to the topic of wetlands and a brief overview of LULC, floods, water quality analysis, and remote sensing applications related to these analyses. The chapter also delves into the scope of work.

Chapter 2: This chapter extensively deals with the literature review to cover the entire research domain. A bibliographic analysis of the relevant past research on this topic and the key texts underpinning this research are discussed.

Chapter 3: This chapter provides supporting information about the study area, such as its location, geographical extent, physical characteristics, and history of disastrous floods.

Chapter 4: This chapter provides an in-depth description of the methodology and data used for LULC analysis and prediction, followed by its outcomes.

Chapter 5: This chapter examines the methodology and outcomes of flood delineation from previously inundated regions, possible future flood inundation regions, and Flood Susceptibility Modelling.

Chapter 6: This chapter examines the methodology and results of the optical characteristics of water to determine Spatio-temporal variations in water quality.

Chapter 7: The summary of the findings and conclusions drawn from the study are presented in this last chapter. It also comprises the limitations of the study and future scope.

CHAPTER 2

LITERATURE REVIEW

2.1   GENERAL

Detecting a water body beneath vegetation is a significant challenge in remote sensing of wetlands. For the ability to map flooded vegetation, several satellite systems have been examined. Satellite-based optical and radar land surface imagery are two common data sources with distinct strengths and weaknesses. Numerous inundation classification procedures identify open waterbodies but ignore flooded vegetation because the vegetation cover tends to dominate the spectral signal in optical imagery (Soti et al. 2009). Optical imagery has been used to map flood-affected vegetation; the algorithms heavily rely on in-situ data and are often calibrated using single occurrences in time (Ordoyne and Friedl 2008). Radar images, on the other hand, are preferred since they can see through cloud cover. Typically, radar systems assess reflectance at a particular wavelength, and each wavelength has a special success rate for identifying floods under various kinds of vegetation (Whitcomb et al. 2009). Using the double-bounce phenomenon, longer wavelengths have been utilized to identify flooded forests since they can penetrate canopies more deeply (Hess et al. 2003). Thus, development in earth observation technology leads many studies using remote sensing to assess wetland regions (Farda et al. 2016; Ji et al. 2015; Ordoyne and Friedl 2008; Rebelo et al. 2009; Soti et al. 2009; Di Vittorio and Georgakakos 2018).

2.2   LULC ANALYSIS

Significant changes have occurred in the LULC of India during the last 140 years, such as the decline in forest cover, change in farmland, and rise in urbanization. According to UN predictions, major cities will encompass 60% of the world's rural communities by 2050 (Parvinnezhad et al. 2020) due to inadequate planning and development, resulting in unplanned urban expansion. The intricacies of future agricultural output, accompanying LULC change, and environmental effects may be investigated using spatial modelling (Abijith et al. 2022). Anthropogenic and environmental processes relating to temporal dynamics and potential changes in the land cover must be properly understood. The development of remote sensing has resulted in various land use studies,

such as the evolution of LULC in a global context and assessing the transition of diverse land types (Abijith and Saravanan 2022).

A case study was conducted by Rebelo et al. (2009) in Muthurajawela Marsh in Sri Lanka and Lake Chilwa in Malawi, South Africa. A decadal analysis of the shift in land use was carried out and disclosed that wetland loss and degradation occurred due to land use and population changes. The specific reasons for wetland conversion may be determined via remote sensing and modelling of change in land use; in this case, the growth of settlements was brought on by the abandonment of paddy fields and the eradication of coastal vegetation. The research results in South Africa, especially using wetland inventory data for planning and management reasons, have shown the relationship between wetland conditions and human livelihoods. To gather information on the distribution, size, and condition of wetlands worldwide, the research addressed the results of wetland management.

For residential and other anthropogenic uses, urban development is transforming and replacing many natural land covers (Kleemann et al. 2017; Rawat and Kumar 2015; Wu et al. 2006). LULCC is key to modifying the global environment, which significantly influences diversity, biological cycles, and ecosystem dynamics (Halmy et al. 2015a; Schwalm et al. 2017). These modifications immediately degrade the landscape and impact the land surface, resulting in greenhouse gas emissions, biodiversity loss, degraded soil resources, and a shift in the global climate system (Hamad et al. 2018b; a; Liu et al. 2020b). To model future LULCC in this situation, it is essential to pinpoint the causes of these changes (Giri et al. 2003; Ruben et al. 2020).

2.2.1   Bibliographic analysis on LULCs

The Web of Science (WoS) database search is used for bibliographic analysis, and the keyword "LULC Change" is used to find the number of journals published in the database pertaining to the study of the LULC change. According to the database, a total of 188 countries have published a total of 35,101 articles as of January 2023. Figure 2.1 shows that the number of articles published each year increases steadily, showing more studies being carried out to study the change in LULC in recent years. In 2021, a maximum of 3924 articles has been published, accounting for 11.17% of the total articles published.

Figure 2.2 depicts the number of articles published in the first 25 countries in WoS. The United States published 10,790 articles, accounting for 30.74% of all published articles. They were followed by China, Germany, England, Canada, and so forth. India ranks eighth with 1617 published articles, accounting for 4.6% of the total articles in the LULC change in the WoS database.
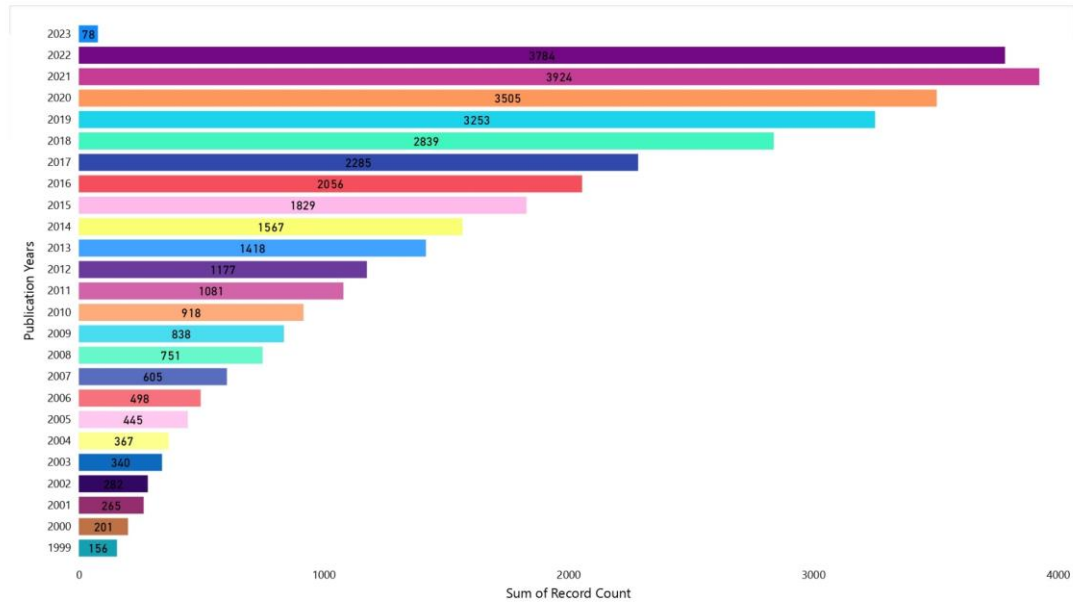
Figure 2.1 Number of articles published each year on the keyword LULCC
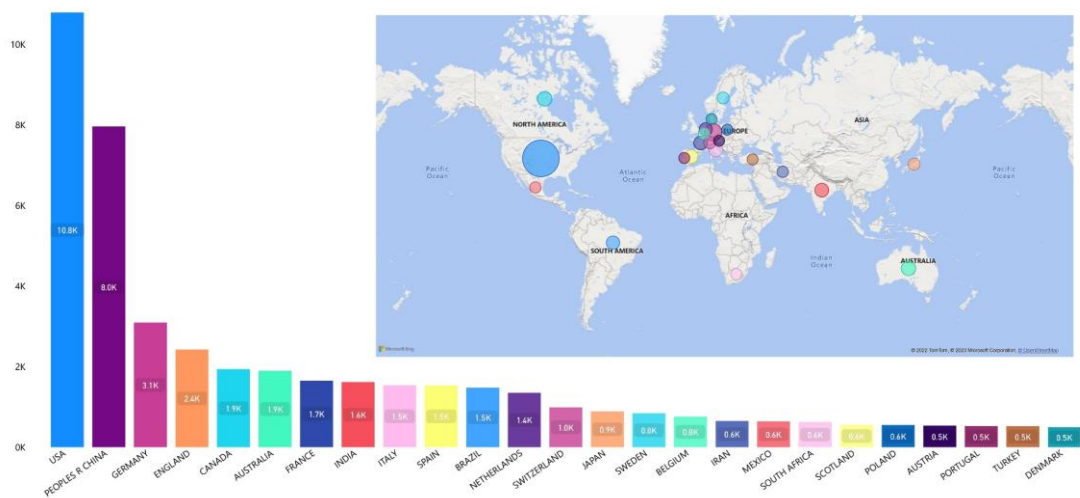
Figure 2.2 Number of articles from countries using the keywords LULCC

2.2.2 Review on LULCC

LULCC modelling is crucial for understanding the LULCC process and its alterations in urbanized regions in the future (Ruben et al. 2020). It was found that GEE aids in the

study of LULC shift in a cost-effective and time-consuming manner, and it is commonly used in the literature (Agarwal and Nagendra 2019; Gomes et al. 2020; Noi Phan et al. 2020; Sidhu et al. 2018; Tamiminia et al. 2020; Tassi and Vizzari 2020; Xing et al. 2021). Several models, including analytical equation-based, statistical, evolutionary, cellular, Markov, hybrid, expert, and multi-agent models, are used for LULCC analysis and predictions. These models concentrate on analysing the causes and effects of LULCC (Aitkenhead and Aalders 2009; Hyandye et al. 2015; Shamsi 2010; Singh et al. 2015; Stefanov et al. 2001; Subedi et al. 2013; Verburg et al. 2004; Yang et al. 2012). Various researchers used CA – Markov chain model to determine the future land cover change and their land-use patterns for decision-making (Aburas et al. 2018; Ansari and Golabi 2019; Bose and Chowdhury 2020; Faichia et al. 2020; Fu et al. 2018; Ghosh et al. 2017; Gidey et al. 2017; Halmy et al. 2015b; Hamad et al. 2018a; Leta et al. 2021; Ozturk 2015). Table 2.1 summarizes some of the findings of the past study on the LULC and its future predictions. The classifier package in GEE allows the use of four ML algorithms, i.e. RF, SVM, CART and NB. The studies also shows that the ML algorithms RF, CART and SVM performed well when compared to the other models. They also need fewer input compared to the other models (Darem et al. 2023; Pan et al. 2022).

Table 2.1 Review of literature on LULCC

Author(s)	Study area	Data used	Model used for classification	Model used for forecasting LULC	Key Conclusion
Pijanowski et al. (2002)	Michigan's Grand Traverse Bay Watershed.	Michigan Resource Information System, agricultural density, distance to road, lakeshore distance, county road distance, river distance, urban distance, quality views, and residential street distance.	-	Land Transformative model with ANN	Each contribution of the predictor variable was investigated and found to differ across regional scales. Quality views were the most powerful predictive variable at the lowest scales. They also analysed multi-scale implications of land use change, demonstrating the relative influences of site and situation factors at various scales.
Rienow and Goetzke (2015)	Federal state of North Rhine-Westphalia.	LULC, elevation, slope, socio-economic variables, population density, distance related	SVM and Binomial Logistic	CA-based SLEUTH model	Compared to the BLR model, the SVM technique requires fewer input features and works robust. Only when the BLR model is connected

		variables such as distance from airport, highway, city, and river.	Regression (BLR)		can the effective direction of those chosen driving factors be understood.
Jat et al. (2017)	Ajmer city of Rajasthan, India.	Survey Of India (SOI) Toposheet, GeoEye.	Maximum Likelihood Classifier method	CA-based SLEUTH model	The model's findings also suggested that SLEUTH cannot account for small unit-size development, such as the fragmented growth seen in rural regions that are typical in developing nations.
Ansari and Golabi (2019)	Meighan Wetland in Iran.	Landsat images, Toposheet, elevation, and slope.	Maximum Likelihood Classifier method	LCM in Idrisi TerrSet	The total model effectiveness in predicting future LULC images was nearly 80%. Due to the accessibility, most land transformations occurred due to human interventions.
Bose and Chowdhury (2020)	Siliguri in West Bengal, India.	Landsat images, elevation, distance to road, stream, school, and railways.	Maximum Likelihood Classifier method	Multi-layer Perceptron (MLP) using LCM in Idrisi TerrSet	The Markov chain model simulates potential LULC scenarios. However, it occasionally falls short in terms of anticipating the spatial

					manifest. It is also crucial to note that urban growth is dynamic, making it extremely challenging to predict the future precisely.
Liu et al. (2020a)	Gannan Prefecture in Northeastern Tibetan Plateau, China.	Landsat images, elevation, slope, aspect, population density, GDP, temperature, annual rainfall, vegetation, and soil type.	RF and CART algorithm	Land use transfer matrix	The performance of RF and CART are equally good. Due to the fragile biological environment, lack of investment, and lack of transportation, factor detection revealed that natural causes had a bigger impact on LULC changes in Gannan Prefecture than anthropogenic factors.
Silva et al. (2020)	Taperoa River basin in Northeastern Brazil.	Landsat images, elevation, slope, economic indicators, distance to roads, urban centre, and river	Maxver classification	MLP using LCM in Idrisi TerrSet	After 10,000 iterations, the MLP neural network used for the dynamic modelling of the land cover provided extremely good results with good accuracy.

Tassi and Vizzari (2020)	Trasimeno Lake, in Umbria, Central Italy.	Landsat, Sentinel-2, PlanetScope satellite images,	Simple Non-Iterative Clustering algorithm, RF, and SVM classifier		Developed and tested an Object-Oriented classification strategy that combines the Simple Non-Iterative Clustering algorithm to find spatial clusters, the Gray-Level Co-occurrence Matrix (GLCM) to produce cluster textural indices, and two ML algorithms, RF and SVM, to do the final classification.
Abijith and Saravanan (2021)	Northern Tamil Nadu coast, India	Landsat images, elevation, slope, and built-up maps	RF algorithm	CA-Markov analysis using LCM in Idrisi TerrSet	The magnitude and direction of possible change on the spatial and quantitative effects of urban sprawl are the primary components of the model simulation. Rapid industrialization and transition are important determinants of how land use will develop. Climate factors

	indirectly influence changes in LULC.

2.3 FLOOD ANALYSIS

Remote Sensing and GIS development in the past few decades has helped researchers gain much attention to mapping the vulnerabilities. They are an excellent tool for computing detailed geographical data for spatial management and analysis (Yang et al. 2020; Zhang et al. 2019b; a). The continuous development of computational models and the ease with which they can be integrated with remote sensing and GIS techniques has improved the forecasting of extreme flood scenarios and the model's dependability at the time of action (Costache et al. 2021). Longer execution times and a high processing load are two drawbacks of using ML algorithms (Aydin and Iban 2023). As the ML models are robust, these drawbacks may be eliminated by lowering the number of independent elements (Jacinth Jennifer and Saravanan 2022). Recursive Feature Elimination (RFE) technique helps select a subset of important characteristics from a larger dataset based on pre-defined criteria like classification performance and eliminating inconsistent data (Gholami et al. 2012), which is very important in ML applications. A better feature selection method should always deliver advantages such as better data understanding, a better classifier model, improved generalization, and the detection of irrelevant features. It should also aid in comprehending the relationship between features and target variables, lowering the computational requirements for solving a specific problem, efficient dimensionality reduction in high-dimensional datasets where the number of data points is less than the number of features, and improving the predictor performance used to complete a specific task in terms of cost and time (Chen and Jeong 2007).

2.3.1 Bibliographic Analysis on Flood Modelling

The WoS database search is used for bibliographic analysis, and the keyword "Flood Susceptibility" is used to find the number of journals published in the database pertaining to the study of flood modelling. According to the database, 117 countries have published 1711 articles till January 2023. The publications started increasing exponentially after 2019 and recorded the maximum number of publications in 2021, having 260 journals (Figure 2.3). In particular, the keywords "Flood Susceptibility" and "Machine Learning" have only 253 articles out of 1711 regarding flood susceptibility. The flood susceptibility using ML emerged after 2018 and started increasing rapidly,

with a maximum of 84 journals published in 2021, accounting for 33.2% of the journals published on the said topic (Figure 2.4).
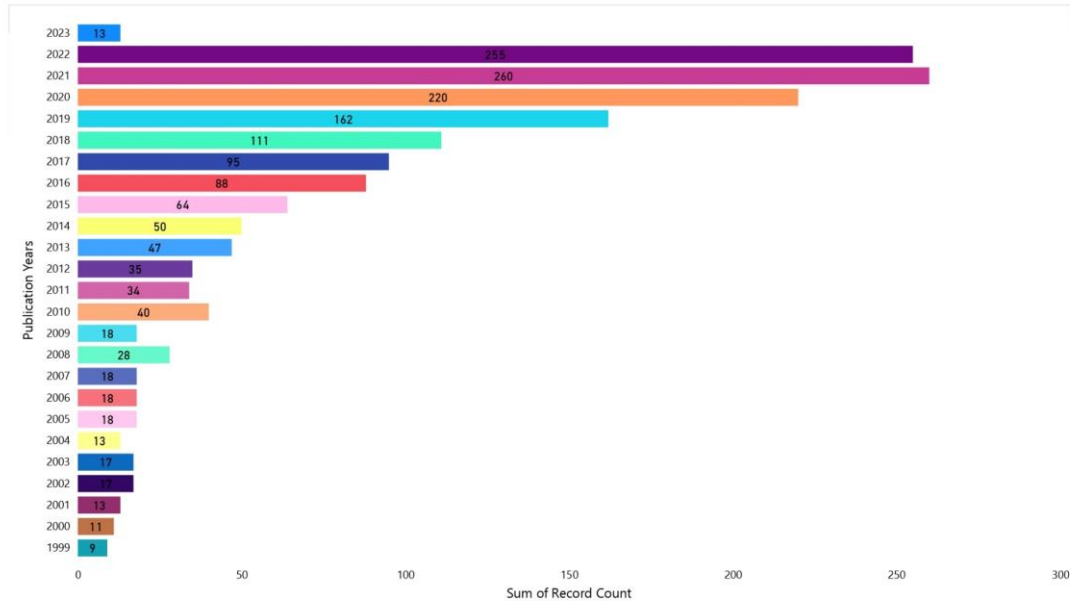
Figure 2.3 Number of articles published each year on the keyword Flood Susceptibility
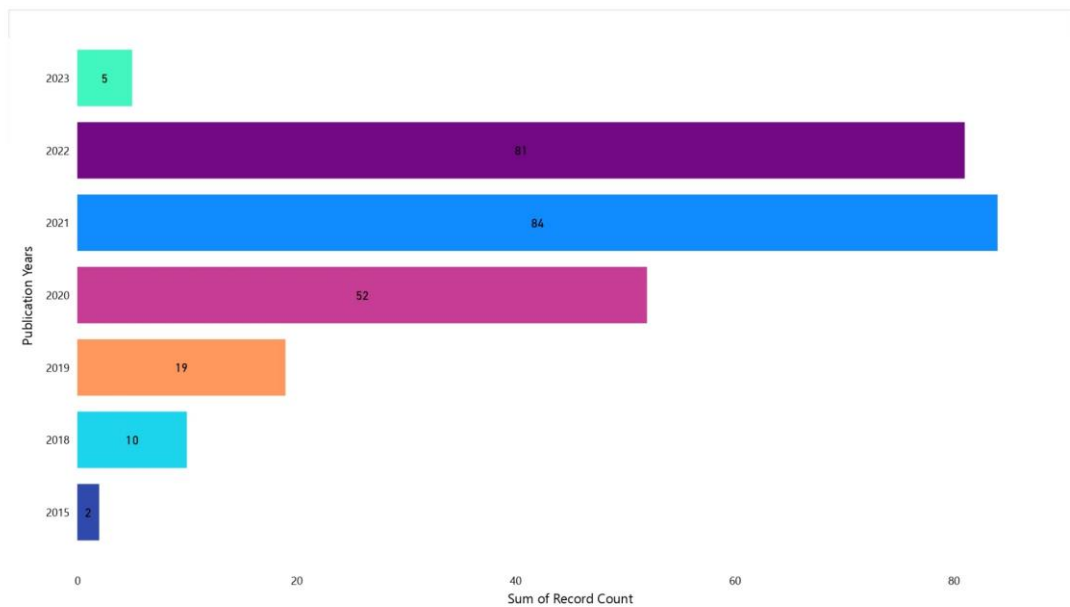
Figure 2.4 Number of articles published each year on the keywords Flood Susceptibility and Machine Learning
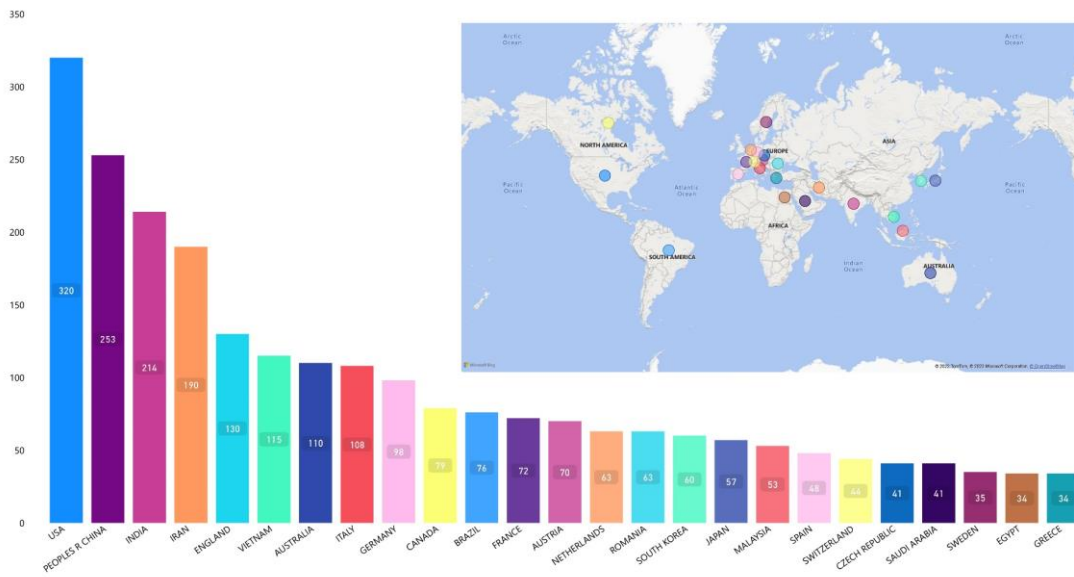
Figure 2.5 Number of articles from countries using the keyword Flood Susceptibility
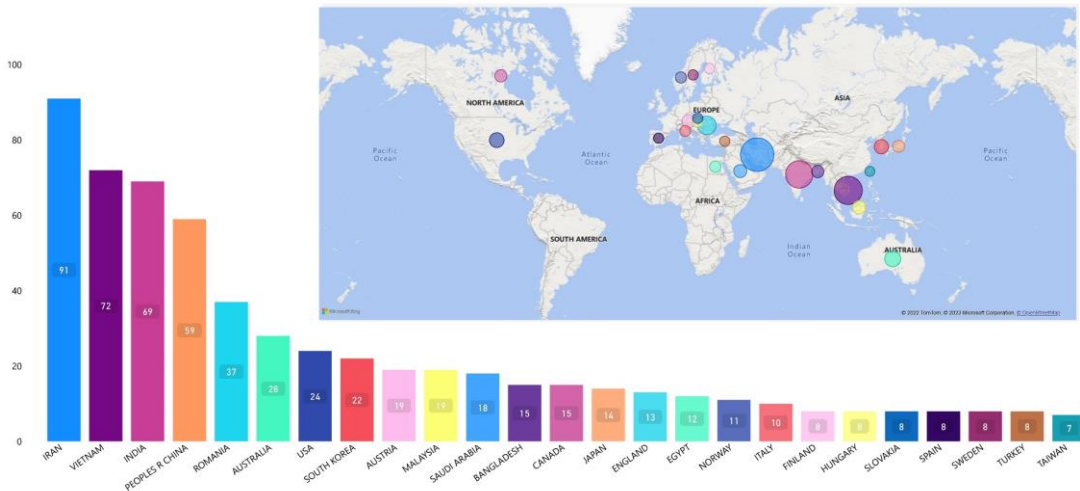
Figure 2.6 Number of articles from countries using the keywords Flood Susceptibility and Machine Learning

Figure 2.5 shows that the USA, China, and India published 320, 253, and 214 articles, respectively, leading the lists of articles published on flood susceptibility. Iran, Vietnam, and India have published a maximum of 91, 72, and 69 articles regarding flood susceptibility using ML (Figure 2.6). According to the analysis, many flood-related studies have been published in India. The reason is due to numerous events of

floods affecting India greatly. As a result, research into the nature and magnitude of floods has increased in the near past (Ward et al. 2013; Winsemius et al. 2013).

2.3.2 Review on Flood Susceptibility

The modelling approaches can be broadly classified into four basic groups based on the available research on flood susceptibility models. They are (i) hydrological models, including Hydrologic Engineering Centre-River Analysis System (HEC-RAS) (Jha and Afreen 2020), Hydrologic Engineering Centre's Hydrologic Modelling System (HEC-HMS) (Romali et al. 2018; Sarchani et al. 2021), Soil and Water Assessment Tool (SWAT) (Sufiyan and Magaji 2019; Yu et al. 2018), HYDROTEL (Aissia et al. 2012), HSAMI (Gbambie et al. 2017), Hydrologiska Byråns Vattenbalansavdelning model (HBV-light) (Sarchani et al. 2021). (ii) data-driven and statistical approaches such as Monte Carlo (Garrote et al. 2021), Evident Belief Function (EBF) (Chowdhuri et al. 2020), Frequency Ratio (FR) (Sarkar and Mondal 2020), Weight of Evidence (WoE) (Rahmati et al. 2015a). (iii) Multi-criteria decision-making models such as Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) (Rafiei-Sardooi et al. 2021), Vlsekriterijumska Optimizacija I Kompromisno Resenje (VIKOR) (Malekian and Azarnivand 2016), Simple Additive Weighting (SAW) (Nawindah 2017), Analytical Hierarchical Process (AHP) (Patrikaki et al. 2018; Rahmati et al. 2015b), Analytical Network Process (Yariyan et al. 2020), Fuzzy-AHP (Ekmekcioğlu et al. 2021). (iv) ML approaches such as Artificial Neural Network (ANN) (Chakrabortty et al. 2022; Kia et al. 2012), Adaptive Boosting (AdaBoost) (Mahdizadeh Gharakhanlou and Perez 2023), Decision Trees (Sachdeva and Kumar 2022), Classification And Regression Trees (CART) (Rahman et al. 2021), Random Forest (RF) (Tang et al. 2021a), Logistic Regression (LR) (Chowdhuri et al. 2020), boosted trees (Lee et al. 2017), Support Vector Machine (SVM) (Saha et al. 2021; Towfiqul Islam et al. 2021), Naïve Bayes (NB) (Tang et al. 2021b), Multi-Layer Perceptron Neural Network (MLPNN) (Darabi et al. 2022), Extreme Gradient Boosting (XGBoost) (Nguyen et al. 2021), Gradient Boosting Machines (GBM) (Felix and Sasipraba 2019), and Deep Learning Neural Network (DLNN) (Li and Hong 2023; Panahi et al. 2021). ML models are rapidly being created and employed in flood susceptibility because of their high performance, accuracy, and predictive potential, and also to overcome the challenges

faced in mountainous regions (Hasan et al. 2023; Shahabi et al. 2021). ML techniques are gaining popularity because they can forecast flood variation from past flood information, eliminating the requirement for sophisticated mathematical formulations of physical processes and basin behaviour (Mosavi et al. 2018). It is less expensive to compute and less complicated than physical and traditional models (Jacinth Jennifer and Saravanan 2022; Jose et al. 2022). Another effective method for measuring a specific region's flood susceptibility is combining many algorithms to generate ensemble models (Abedi et al. 2022).

In recent years, ML algorithms have emerged as serious challengers to traditional statistical models in susceptibility analysis. Because of their capacity to extract correlations rapidly utilizing previous occurrences and topography, ML techniques in flood susceptibility assessment are becoming more popular. The functional link between the major factors may be crucial in predicting floods and other natural disasters. ML models have been investigated because they can detect flood-prone areas based on previous events without necessarily knowing the physical mechanisms that generate them. Table 2.2 summarizes some of the findings of past studies on flood susceptibility and the performance of various ML models in identifying likely flooding zones. The review of literatures shows that the ensemble based models such as RF, GBM, XGBoost and AdaBoost performs comparatively better than the other ML algorithms.

Table 2.2 Review of literature on Flood Susceptibility using Machine Learning

Author(s)	Study area	Parameters	ML model used	Key Conclusion from the Literature
Chen et al. (2020)	Quannan County in Jiangxi Province, China	Aspect, slope angle, curvature, STI, SPI, TWI, elevation, LULC, NDVI, distance to rivers, rainfall, soil, and lithology	NB, Alternating Decision Tree (ADTree), and RF	The RF model has been recognized as the best appropriate ensemble approach. The ADTree and NBTree models are essentially two base classifiers with little generalization. Their performance is strongly related to the characteristics of the initial datasets.
Chowdhuri et al. (2020)	Koiya river basin, India	LULC, soil, rainfall, NDVI, distance to river, elevation, TWI, and SPI.	EBF, LR, and ensemble of EBF and LR (EBF-LR) model	The results demonstrate that the ensemble EBF-LR method is the best, with a 91.56% success rate for the flood susceptibility model and an 88.20% prediction rate for the flood susceptibility map.

Abedi et al. (2021)	Basca Chiojdului River Basin, Romania	Aspect, LULC, soil, lithology, slope, TWI, TPI, profile curvature, convergence index, and SPI.	CART, RF, Boosted Regression Tree (BRT), and XGBoost	According to all models, the slope is the most crucial element influencing the likelihood of flash floods. The accuracy of all tree-based ensemble models was approximately the same, but the RF model gained out because it only needs a limited number of tuning hyperparameters, whereas other models, especially the XGBoost, need many more.
Chakrabortty et al. (2021)	Kangsabati River Basin, India	Aspect, elevation, slope, plan curvature, profile curvature, TRI, TWI, SPI, distance to river, drainage density, distance to stream, rainfall, LULC, and geology	ANN and DLNN	Regions with lower altitudes and slopes are more prone to flash floods with shorter durations, and the distance to river appears to have a significant role in the spatial distribution of flooding. PSO displayed the best performance of

				these three models, with an AUC value of 0.942.
Costache et al. (2021)	Izvorul Dorului, Romania	Slope angle, aspect, convergence index, profile curvature, plan curvature, topographic position index (TPI), TWI, LULC, lithology, and soil map	Fuzzy logic algorithm with the following four ML models: CART, DLNN, XGBoost, and NB.	Lithology has a significant impact on the water runoff-infiltration process. Its value in predicting flash floods is larger than other predictors. The optimal model in terms of success rate is Fuzzy-XGBoost (AUC - 0.886), whereas the optimal model in terms of prediction rate is Fuzzy-DLNN (AUC - 0.84).
Darabi et al. (2021)	Amol City, Iran	Elevation, slope, distance from river, distance from channel, Curve Number, and Precipitation.	Standalone MLPNN, BRT, RF, Hybrid Multi-Boosting Multilayer Perceptron Neural Network (MultiB-MLPNN)	This study combined the multi-boosting method with the MLPNN model to develop a unique hybridized model (MultiB-MLPNN) for mapping urban flood susceptibility. MultiB-MLPNN outperformed RF and BRT.

Mirzaei et al. (2021)	Talar watershed, Iran	Elevation, slope, profile curvature, TWI, distance from rivers, NDVI, plan curvature, rainfall, LULC, SPI, and lithology	XGBoost, Frequency Ratio (FR), RF, and Generalized Additive Model (GAM)	This study compares the performance of the FR, RF, GAM, and XGBoost. The RF model and XGBoost perform the best, with Area Under Curve (AUC) of 0.985 and 0.980, respectively. It was observed that GBM lacks a strong regulatory parameter and is susceptible to overfitting.
Pham et al. (2021)	Nghe An Province, Vietnam	Slope, elevation, aspect, curvature, river density, distance from rivers, flow direction, geology, soil, and LULC	Three ensemble models based on the Best First Decision Tree (BFT) are the Bagging (BBFT), Decorate (DBFT), and Random Subspace (RSS-BFT) ensemble learning techniques	The ensemble BBFT, DBFT, and RSSBFT models outperformed the single BFT model regarding future flood likelihood. The efficacy of the ensemble modelling approach was proved when the single BFT model's high-level training performance declined significantly in the validation phase to levels

				substantially lower than those of the three ensemble models. Although DBFT is not as popular as Bagging and RSS, this ensemble learning technique demonstrated its efficacy for predicting flash flood susceptibility.
Saha et al. (2021)	Koiya River basin, India	LULC, soil type, rainfall, NDVI, distance to river, elevation, TWI, and SPI.	HyperPipes (HP), Support Vector Regression (SVR); and the novel ensemble of HP-SVR	The ensemble model of HP-SVR is the ideal model for assessing flood susceptibility spatial prediction analysis, followed by HP and SVR.
Shahabi et al. (2021)	Haraz watershed in Iran	Slope, elevation, curvature, TWI, SPI, distance to river, river density, rainfall, lithology, land use, and NDVI.	Deep Belief Network with Back Propagation algorithm optimized by the Genetic Algorithm (DBPGA), LR, Logistic Model Trees, Bayesian logistic regression, Alternating Decision Tree, NB, and Reduced Error	The slope angle is the most crucial component, even if all 11 flood conditioning parameters impact flood occurrence. DBPGA has the potential for usage in other flash-flood-prone areas.

			Pruning Tree, Adaptive Neuro-Fuzzy Inference System (ANFIS) –BAT Algorithm, ANFIS - Imperialistic Competitive, ANFIS – Invasive Weed Optimization, ANFIS – Firefly.	
Towfiqul Islam et al. (2021)	Teesta sub-catchment of the northern region of Bangladesh	Elevation, curvature, aspect, slope, TRI, TWI, SPI, STI, LULC, distance to the river, soil type, and rainfall	Dagging and RSS coupled with ANN, RF, and SVM	The Dagging model had the greatest adaptability and predictive powers, followed by the RF, ANN, SVM, and RS models. Compared to other methods, the Dagging model performs best (AUC = 0.863-training stage; AUC = 0.873-validating stage). However, the standard of all models for mapping flood susceptibility was satisfactory and dependable.

Saravanan and Abijith (2022)	The northern coastal area of Tamil Nadu, India	Slope, aspect, elevation, plan curvature, profile curvature, SPI, TWI, STI, TRI, rainfall, distance to road, distance to stream, distance to coast, LULC, soil, geology, geomorphology, wind, and NDVI	GBM, XGBoost, RTF, SVM, and NB	The study compared different algorithms to improve the capability of the individual methods due to the need for an accurate and reliable method to detect flood-prone areas. According to the investigation results, GBM and XGBoost outperforms the other three models.

2.4 ANALYSIS OF THE OPTICAL STUDY OF THE LAKE

Data on water quality at various spatio-temporal scales is required to manage contaminated sites and control environmental pollution. In reality, assessing river water's physical, chemical, and biological characteristics at more precise spatio-temporal scales is essential for effectively managing wastewater facilities and sectoral riverine water supplies. The proper treatment of the waterbodies can be suggested after examining spatio-temporal variations in the water quality of the study. Numerous studies have demonstrated that anthropogenic activities are one of the main causes of pollution in all areas of the environment (Abraham and Kundapura 2022; Caballero et al. 2020; Carmichael and Boyer 2016; Hassan et al. 2016; Krishnaraj and Deka 2020; Krishnaraj and Honnasiddaiah 2022). The inland and coastal water quality are monitored using data products generated by different remote sensing sensors. The particles in the water column influence its IOPs, which remote sensing sensors may measure because they tend to alter the spectral composition of the water column, according to the principle regulating optical remote sensing (Balasubramanian et al. 2020; Mabit et al. 2022; Toming et al. 2017). It is very hard to get spatial and temporal fluctuations of quality indices in large waterbodies using *in situ* data collections, which can only reflect point assessments of the quality of water conditions in time and space. As well as, *in situ* data collection is costly, labour-intensive, and time-consuming. Due to the topographic situations, the regions may be inaccessible for data collection. Thus, remote sensing as an effective tool can overcome these limitations (Gholizadeh et al. 2016; Xing et al. 2021). Remote Sensing has demonstrated a potent ability to track and assess the calibre of inland waterways for more than 4 decades (Jafar-Sidik et al. 2017). To obtain strong correlations between water column reflection (or emission) and physical and biogeochemical constituents, such as transparency, chlorophyll concentration (phytoplankton), organic matter, and mineral-suspended sediments in various waterbodies, many researchers frequently use the visible and near-infrared bands of the solar spectrum (mostly from the blue to near-infrared region) (Ritchie et al. 2003). During the lockdown imposed by COVID-19, water and air pollution significantly decreased (Ma et al. 2020; Ogen 2020; Venter et al. 2020). All industrial and human mobility is prohibited for many weeks, and it is believed that environmental

contamination would also diminish. As anticipated, the level of carbon emissions has sharply decreased in a short period of time (Yunus et al. 2020).

2.4.1 Bibliographic analysis on Optical study of water

The WoS database search is used for bibliographic analysis, and the keyword "Remote Sensing", "SPM", and "Turbidity" are used to find the number of journals published in the database pertaining to the study of optical properties of the water. According to the database, 194 countries have published 63,624 articles till January 2023. The number of papers published on this topic has increased exponentially since 2006 and recorded the maximum number of publications in 2021, having 4258 journals in that year (Figure 2.7). Figure 2.8 shows that USA, China, and England have published 12529, 9325, and 3659 papers on water quality using remote sensing techniques. India has published 3394 articles ranking fourth in the list. In particular, the keyword "ACOLITE Processing" is used to refine the journals published in the database pertaining to the ACOLITE software. Only 30 countries have published 61 papers on this topic, with a maximum of 15 articles published in 2021 (Figure 2.9). Spain, USA, and Belgium have published a maximum of 10, 10, and 9 articles, respectively (Figure 2.10). India ranks 10th in the list, with only 4 articles published on this topic.
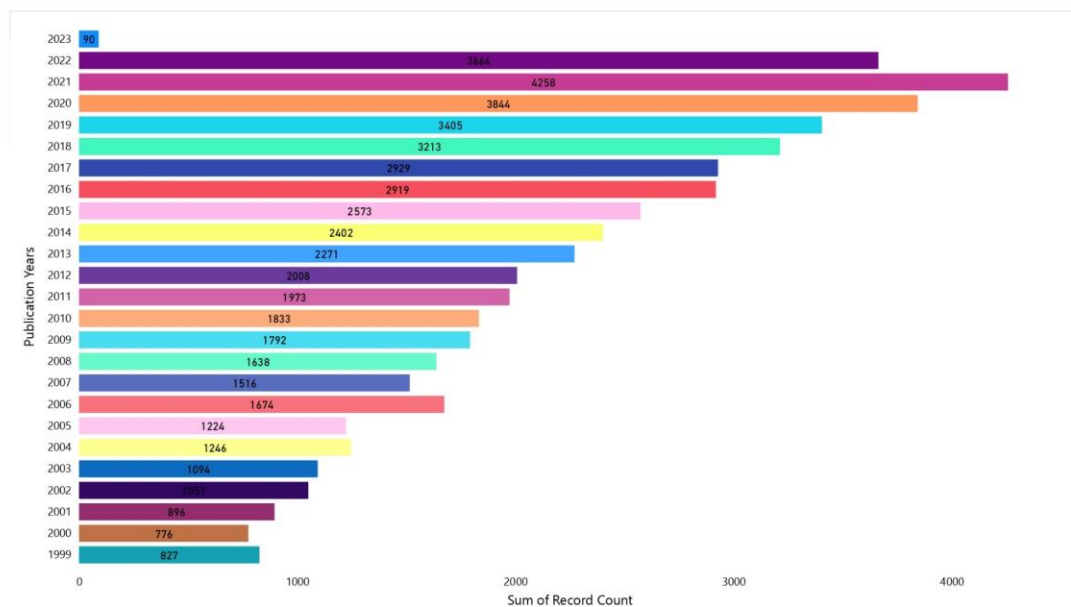
Figure 2.7 Number of articles published each year on the keywords Remote Sensing and SPM and Turbidity
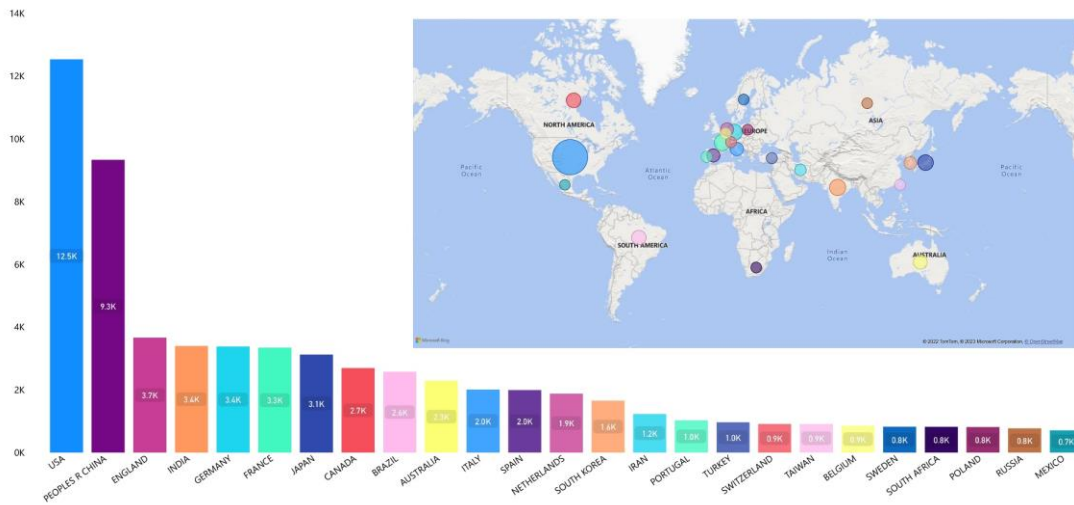
Figure 2.8 Number of articles from countries using the keywords Remote Sensing and SPM and Turbidity
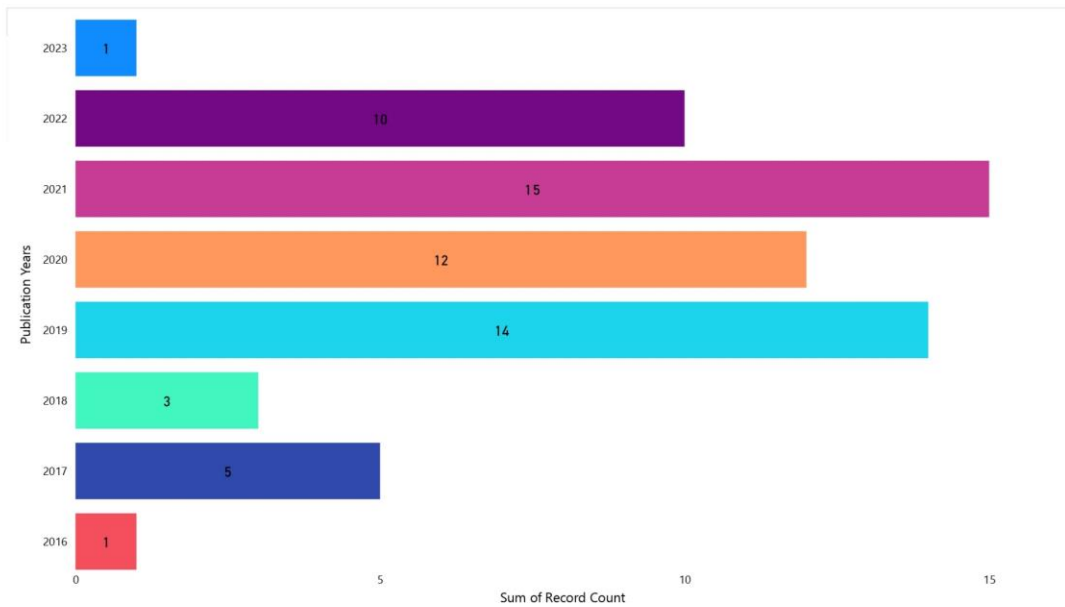
Figure 2.9 Number of articles published each year on the keyword ACOLITE
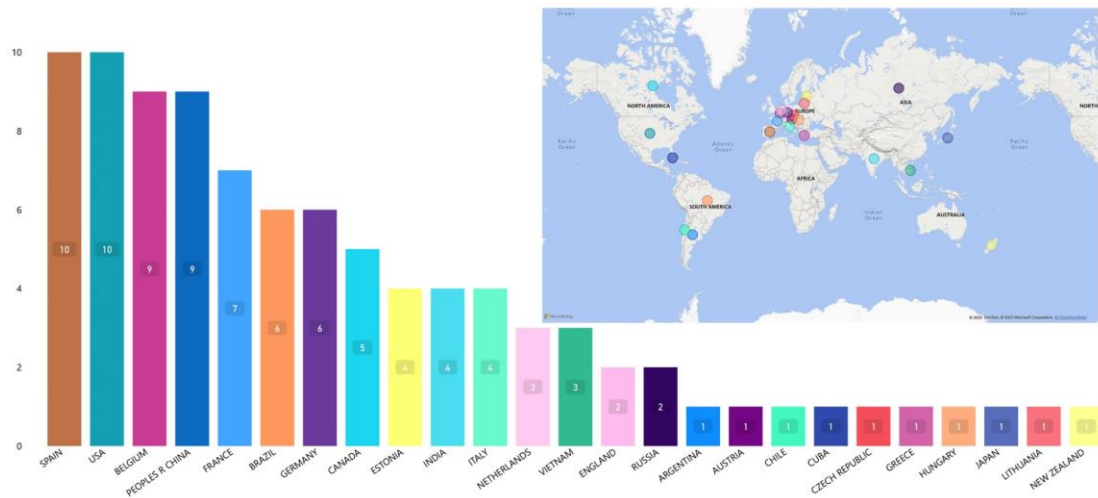
Figure 2.10 Number of articles from countries using the keyword ACOLITE

2.4.2 Review of the Optical studies of the water

Remote Sensing techniques provide a better understanding of the spatiotemporal patterns of WQ at finer scales. It uses various optical characteristics of surface water by altering the reflected energy spectrum or by having it emit heat radiation. The amounts of optically active water constituents may be inferred from satellite images by analysing the received radiance at the sensor at various wavelengths. Various physical parameters are analysed using NDVI, Fraction of absorbed Photosynthetically Active Radiation (FPAR), Normalized Difference Water Index (NDWI), and the Photochemical Reflectance Index (PRI) are widely used for the analysis of biophysical parameters on the lakes and waterbodies (Barducci et al. 2009). The evolution of atmospheric correction algorithms by comparing the *in situ* data and deriving the relationship between the water constituents and the reflectance of various spectral bands and their combinations has gained huge popularity (Nechad et al. 2010). Some of the widely used atmospheric correction processors include Atmospheric correction for OLI 'lite' (ACOLITE) (Caballero et al. 2020; Vanhellemont and Ruddick 2016a), Case 2 Regional Coast Colour (C2RCC) (Nazirova et al. 2021; Pereira-Sandoval et al. 2019), Case 2 Regional Coast Colour for Complex waters (C2RCCCX) (Pereira-Sandoval et al. 2019), Case 2 extreme (C2X) (Mabit et al. 2022), Image correction for atmospheric effects (iCOR) (de Keukelaere et al. 2020), Polynomial-based algorithm applied to MERIS (Polymer) (Pereira-Sandoval et al. 2019), Level 2 data processing algorithm (l2gen) (Warren et al. 2019) and Sentinel-2 Correction (Sen2Cor) (Louis et al. 2018).

Table 2.3 summarizes some of the findings of the past study on the optical properties of water. The review of literatures shows that the ACOLITE software performs better in the eutrophic waterbodies.

Table 2.3 Review of literature on Optical studies of water

Author(s)	Parameter used	Satellite data	Software used	Key Conclusion
Kari et al. (2016)	SPM and turbidity	MERIS	ACOLITE	Despite the significant yearly and regional fluctuation in SPM within the two studied sub-regions of the Baltic Sea, the link between the significant environmental and optical variables remains mostly stable. Regarding SPM concentration for the Baltic Sea, the validation datasets were demonstrated to be representative. It is suggested that remote sensing data be used to monitor the Baltic Sea, particularly in coastal waters regularly.
Toming et al. (2017)	Chl-a, SPM, and CDOM	Sentinel-3A	CR2CC	The study demonstrated the critical requirement for creating a C2RCC processor specialized in the Baltic Sea. The existing one works rather well when there are no cyanobacterial blooms but fails when there are blooms since the IOP and other products do not correlate with *in situ* measurements. On the other hand, the consistency of *in situ* data utilized in the calibration and validation of satellite data products has to be improved.

Pereira-Sandoval et al. (2019)	Chl-a, Secchi disk depth (Z_{sd}), CDOM and SPM	Sentinel-2	ACOLITE, C2RCC, C2RCCCX, iCOR, Sen2Cor and PolyMER	According to a performance assessment based on water types, single bands categorization, *in situ* Chl-a concentration ranges, and Z_{sd}, the atmospheric correction processors perform better for relatively diverse waters. Compared to oligotrophic waters, ACOLITE, iCOR, and Sen2Cor performed better in meso- and hyper-eutrophic waters.
Avtar et al. (2020)	Chl-a and SPM	Landsat-8 and Sentinel-2	ACOLITE	The increased Chl-a shows an enhanced HRT in the lakes of Wuhan. It is also possible that a prolonged HRT might lead the surface phytoplankton to settle down, explaining the declining trend followed by the increase. Throughout the lockdown, there were no appreciable changes in the Chl-a concentration in Vembanad Lake, India. One explanation is that, unlike the lakes in Wuhan, the Vembanad Lake is not closed. It might be that the HRT is insufficient to increase primary production in the Vembanad Lake due to the flow from the rivers (without pollution) and tidal action. During the lockdown, SPM levels significantly dropped in both Vembanad and Wuhan Lake.

| Bhuyan et al. (2020) | Chl-a, CDOM, turbidity, and Sea Surface Temperature (SST) | Landsat-8/OLI | ACOLITE | The extremely changeable water quality in the region is a result of various factors, the two most significant of which are local runoff and precipitation levels. Microbial activity, and anthropogenic causes, in addition to high temperatures and irradiance, impact the CDOM concentration. Chl-a concentration showed a direct one-to-one relationship with the runoff in the area where the curve mostly followed the runoff curve. This can be mostly attributed to the increased nutrient availability in the estuary during periods of high runoff, which in turn impacts the Chl-a cycle found in the natural environment of the estuary. However, because several other factors impact turbidity, there was no correlation between it and the region's runoff in the instance of that variable. |
| Garg et al. (2020) | NDVI and NDWI | Sentinel-2 | - | The reflectance in the visible range increases when turbidity increases and vice versa. Due to interference from the bottom, the blue and green bands could not depict the spatial heterogeneity in the turbidity variation. However, even in these bands, there was a modest decrease in turbidity or reflectance in the deep water. Red |

				and NIR bands were shown to be more sensitive to turbidity estimates.
Yunus et al. (2020)	SPM	Landsat-8	ACOLITE	This study found that when industries and boating were shut down during the lockdown, SPM levels dropped. The SPM readings decreased in 18 of 20 zones in the lake on average by 15.9%. According to the results, pollution from industry and tourists had a bigger impact than non-industrial sources.
Kulk et al. (2021)	Chl-a, SPM, CDOM, turbidity, and Forel-Ule classification	Sentinel-2 and Landsat-8	ACOLITE	The aquatic environment in the lake may have benefited from the observed drop in TSM and CDOM by lowering turbidity and improving light penetration, which may have increased phytoplankton photosynthesis. No obvious variations in Chl-a during the lockdown, indicating that other variables were much more crucial in regulating phytoplankton development in Lake.
Caballero et al. (2022)	Turbidity and Chl-a	Sentinel-2A/B and Landsat-8	OC3 and ACOLITE	The turbidity and Chl-a products employed were proxies for the biogeochemical characteristics in the coastal waters, despite the fact that satellite products were not validated with local *in situ* data to address their quality and uncertainty. These procedures,

				which provide a universal application, are proving reliable and consistent in investigating turbidity and TSM.
Mabit et al. (2022)	CDOM and SPM	Landsat-8 and Sentinel-2	C2RCC, C2X, SeaDAS, iCOR and ACOLITE	The CDOM red/green algorithms in ACOLITE perform better, with the point distributions indicating the quasi-null systematic error retrieval. Even while C2RCC provides the most precise absolute data for remote sensing reflectance, ACOLITE produces the best results for CDOM estimates.

2.5 RESEARCH GAP

Understanding the influence of LULC and changes aids in the investigation of the cause of the floods and their magnitude. The prediction of LULC is based on historical LULC maps and other driving factors. The study area encompasses significant districts in the state of Kerala which has a very high population density and experiencing rapid growth. Due to the increasing occurrences of disaster-related events, the inhabitants of the region are being significantly impacted. Therefore, it is crucial to conduct an analysis of the alterations in LULC and to forecast the future. Thus, the LULCC and prediction play a larger role in understanding changing trends in land cover and help to understand the rate of change in different land use types in the region. Integrating the predicted LULC and the extreme flood scenario will provide insight into the likely inundation region in different land use types and predict the likely magnitude of the flood in the region.

Many studies have been conducted to identify flood inundation zones in the Vembanad Lake System (VLS) and other areas of Kerala. However, no study has previously carried out FSM to demarcate the flood prone regions in VLS. Due to heavy monsoon precipitation, this region has been continuously affected by numerous flooding events over the past half a decade. This flooding is also caused by unplanned urbanization and poor drainage in cities. The FSM aids in analysing the flood vulnerability region and the steps that must be taken to reduce future flooding. It also aids in taking the necessary precautions during a flood to prevent loss of life.

The construction of the Thannermukkom barrage in 1976 created several water quality problems, such as reducing the flushing action in the lake, thereby causing a proliferated growth of weeds and water hyacinth in the water body. Considering the fragile ecosystem of the wetland, deterioration of water quality and consequent damage to aquatic organisms, and the shrinkage of Vembanad Lake, this wetland system was included in the National Lake Conservation Plan by the National River Conservation Authority, chaired by the Prime Minister under the Ministry of Environment and Forest (MoEF) in 2003. The sewage effluents and the heavy load of organic material released from the neighbouring areas are responsible for decreased dissolved oxygen content in the water body. Few studies on the lake suggested that the lake's water quality has improved during the COVID-19 lockdown. However, no studies have considered the

spatio-temporal variation in water quality during and after the floods in the lake. This aids in understanding the change in lake water quality over time and its effects on various events such as floods and COVID lockdowns.

Based on the research gaps, the following research questions have been formulated.

➢ How will LULC patterns in VLS evolve over the period of time, and what are the key driving factors influencing these changes?

➢ How can remote sensing techniques effectively integrate the accuracy and timeliness of flood inundation mapping?

   ➢ What would be the impact of a flood of equivalent intensity on the LULC in future?

   ➢ What are the data-driven techniques that can be leveraged to develop an accurate flood susceptibility mapping?

➢ What are the effects of rapid evolution and frequent flooding on the water quality in Vembanad Lake?

Thus, these research questions are formulated as the objectives of the study.

2.6   RESEARCH OBJECTIVES

The objectives of this study are formulated based on a comprehensive review of the existing literature and identification of research gaps. These objectives aim to establish a solution for the current research work.

➢ Land Use and Land Cover Change

   ➢ To map the Land Use and Land Cover Change (LULCC) using Machine Learning techniques like Random Forest (RF), Support Vector Machine (SVM), and Classification And Regression Trees (CART) in Google Earth Engine (GEE).

   ➢ To develop the future trends of the LULC change from the best working model using CA-Markov analysis.

➢ Flood Modelling

   ➢ To delineate the frequently flooded areas in Vembanad Lake System (VLS) using SAR images and determine the degree of flooding in each LULC class during the 2018 floods.

   ➢ To quantify the possible flood inundation region in the future LULCs of 2035 and 2050 based on the severity of the 2018 flood.

> ➢ Geospatial analysis of flood susceptibility modelling using the decision tree-based machine learning algorithms
➢ Water quality analysis
> ➢ To retrieve and analyse the optical properties of water in the Vembanad Lake from the satellite optical images such as Diffuse attenuation coefficient (K_d), Suspended Particulate Matter (SPM) concentration, Turbidity, Chlorophyll-a (Chl-a) concentration.

In order to bridge the literature gap, the above objectives for this research are mapped in Figure 2.11 in a nutshell. The likely outcomes of the study include three vital results. They are possible flood inundation regions in the future LULCs, Flood Susceptibility Modelling for the VLS, and changes in the optical properties of the lake due to floods and COVID lockdowns. This research will examine the quantitative and qualitative hydrological aspects of the waterbody (VLS), such as the extent of the region inundated by the floods, their susceptibility, and the quality of the water that is ultimately drained into the lake system, respectively. Various software has been used to carry out the analysis, such as GEE, Idrisi TerrSet, ArcGIS, R program, Python, ACOLITE, and SeaDAS.
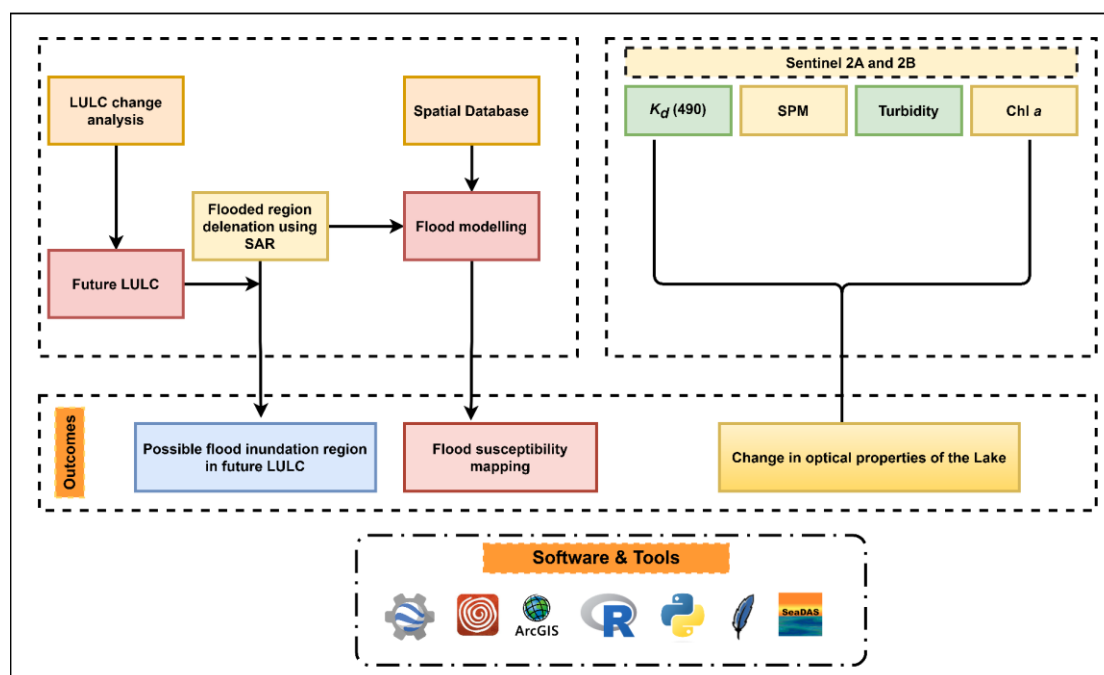
Figure 2.11 Mapping of the objectives

CHAPTER 3

STUDY AREA

3.1   GENERAL

The study area is the Vembanad Lake System (VLS) in the state of Kerala, located in the extreme south-western part of India. The VLS comprises of Vembanad wetland and six watersheds. These six watersheds drain into the Vembanad wetland, which is connected to the Arabian Sea in the west. The wetland covers an area of 1512 km^2, making it India's second-largest wetland system. This lake faces several problems that threaten its health and survival. Some of the main issues include:

➤ Encroachment and land reclamation: Vembanad wetland has been facing a rapid loss of habitat due to encroachment by human settlements and agricultural activities.

➤ Climate change: Climate change is also affecting the wetland, causing changes in precipitation patterns and rising sea levels that can alter its hydrology and lead to heavy flooding resulting loss of habitats.

➤ Water pollution: The wetland is subjected to high pollution levels from agricultural runoff, sewage, and industrial discharge, which negatively impacts the health of its flora and fauna.

These problems must be addressed to conserve and protect the Vembanad wetland and its unique biodiversity. The following sections narrate the geographical extent, climate, geology, and geomorphology. Also, the explanation about the 2018, 2019, and 2021 Kerala floods which are considered for analysis in the present research for arriving at flood inundation and region of flood susceptibility, are provided.

3.2   GEOGRAPHICAL EXTENT AND CLIMATE

The field of study comprises six river watersheds, including Periyar, Muvattupuzha, Meenachil, Manimala, Pamba, and Achencoil, draining into the Vembanad wetland (Figure 3.1). Also, these are the main rivers in the central part of Kerala that originates from the Western Ghats and joins the Vembanad wetland. The VLS spreads across six districts: Ernakulam, Idukki, Kottayam, Alappuzha, Pathanamthitta, and Kollam, and is located between the latitude of 9° 1' 9" N to 10° 20' 22" N and longitude of 76° 16' 47" E and 77° 24' 43" E, comprising an area of 12,183 km^2. Of this, 398.12 km^2 is

below the MSL, and 763.23 km^2 is below 1 m of MSL. And these are the prominent regions affected by the floods. The eastern part of the study area consists of the Western Ghats, whereas the Arabian Sea binds the western part. Figure 3.2 shows the keymap of the study area. The lake is considered of international importance under the Ramsar Convention of Wetlands.
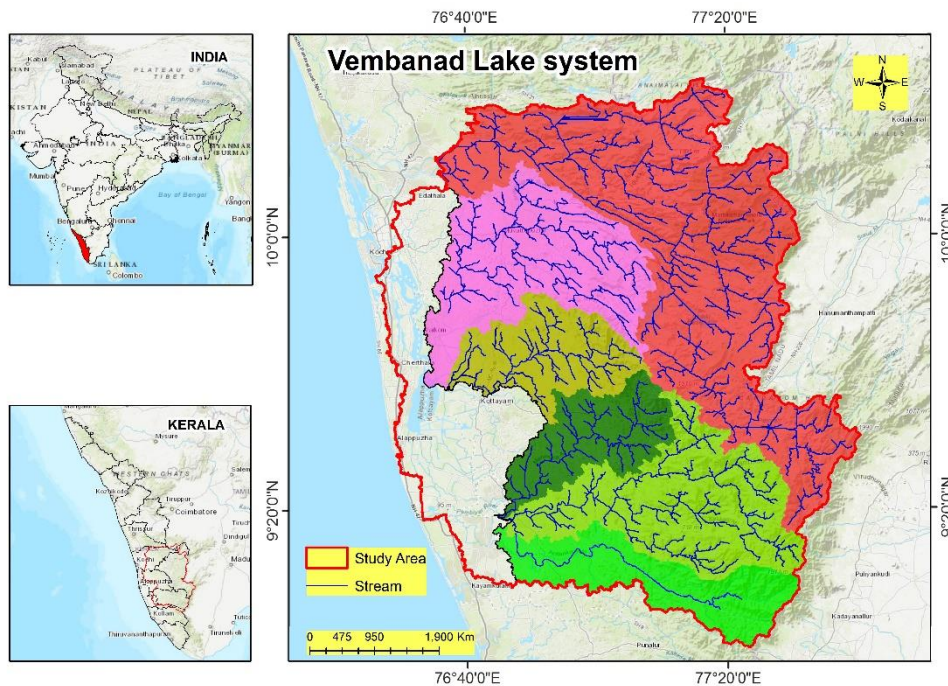
Figure 3.1 Study area

The state has a wet and maritime tropical climate with an average of 150 rainy days yearly during the monsoon season. It has a higher mean annual temperature in the coastal lowlands of about 25°C-27.5°C than the eastern highlands that range from 20°C-22.5°C. An average rainfall of 3000 mm/yr is received from the southwest and northeast monsoons, of which 65% accounts for the former (Figure 3.3). When the monsoon hits the Western Ghats situated on the eastern side of the study area, the state of Kerala receives the first rainfall in India. From July to September, heavy rainfall is common in the study area. Whereas between September and November, the Northeast winter monsoon arrives, bringing cool winds. According to the 2011 census, India has a population density of approximately 382 people/km^2 and is found to be increasing. Kerala is a state in the Indian sub-continent with a population density of 859 people/km^2. A disaster affecting the state can result in huge disturbances in its socio-

economic status. Thus to overcome the disaster, resilient cities with critical infrastructure should be embedded in order to provide fast recovery and normal functioning of the affected communities (Andrić and Lu 2017).
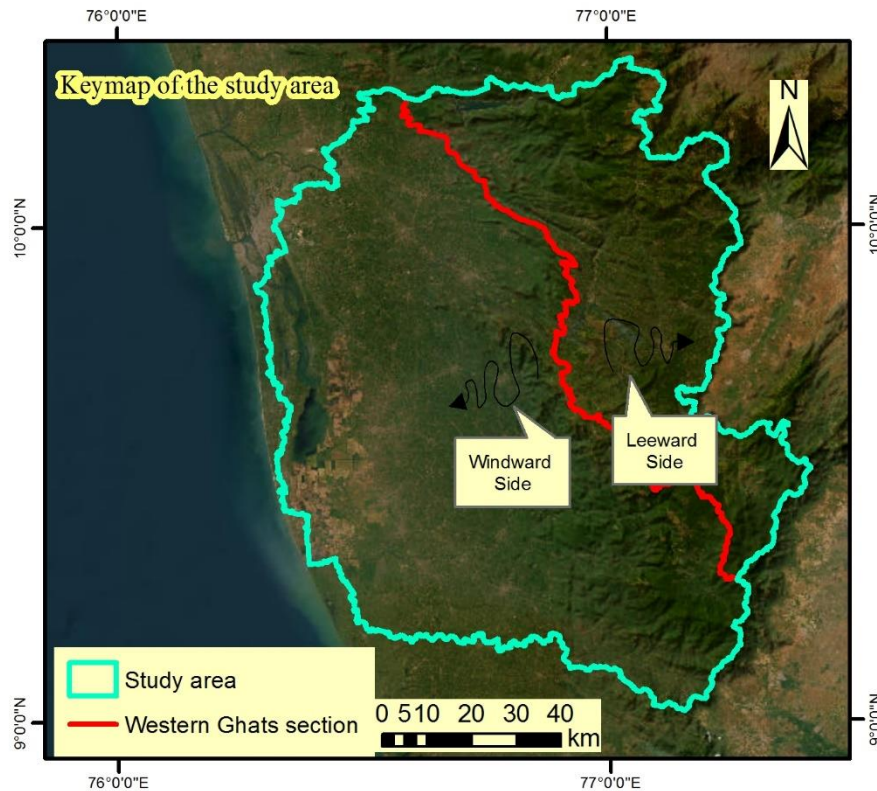
Figure 3.2 Keymap of the study area

The region around the VLS faced repeated flooding during 2018, 2019, and 2021 due to the unprecedented rainfall and the runoff generated from the rivers resulting in heavy flooding across the state. This resulted in considerable damage to the buildings and infrastructure in the state. Lots of agricultural lands flooded completely, causing lots of damage to the resources. The state took quite a lot of damage due to this unprecedented rainfall.
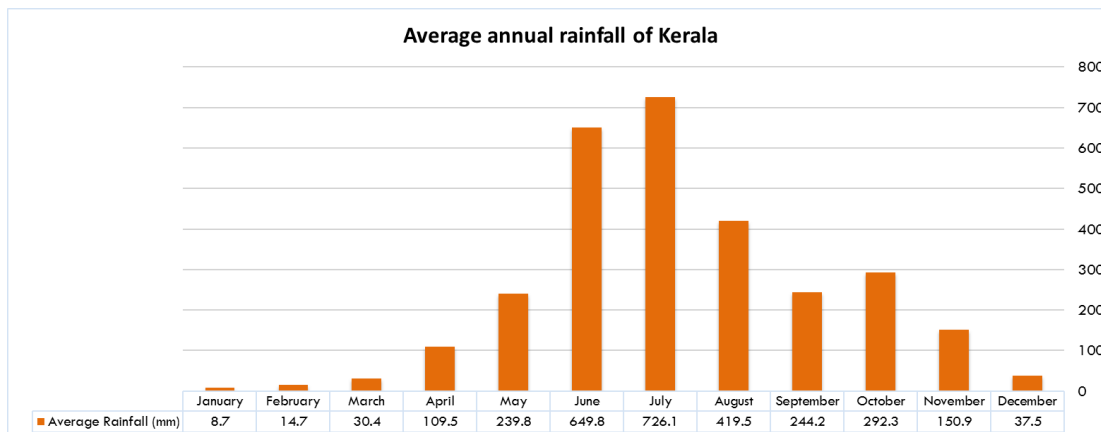
Figure 3.3 Average rainfall of Kerala

(Data Source: CWC (2018))

3.3 GEOLOGY AND GEOMORPHOLOGY

The study region lies along the windward slopes of the Western Ghats, covering the range of highlands with geomorphic features such as denudational hills, dissected middle plateaus, lateritic plateaus, valleys, narrow flood plains, and river channels. The physiographical features of the district constitute low-lying plains, mid-land plateaus, and hill range with a slope gradient of 10-80%. Most of the study area comprises vegetation and forest, accounting for approximately 75% of the total area. The geological forms present in the study area are the Archaean basement complex, which is dominated by low-grade metamorphic and gneiss rocks. The area forms a part of the Precambrian metamorphic shield of South India, comprising rocks of the Wayanad, Khondalite, Charnockite, and Migmatite Groups. The geology of the study is characterized by charnockites, charnockitic gneisses, garnet biotite gneiss, hornblende biotite gneiss, pink granite gneiss, and pyroxene-bearing granulite in the Western Ghats and central areas of the study. The sedimentary deposits of the Neogene and Quaternary periods dominate the western sections of the sample. Alluvial deposits from the recent past can be found along the coast. The uplands have soils of the Ultisols order, and the soil is under an iso-hyperthermic temperature regime that varies from summer to winter with a variability of ±5 °C.

3.4   2018 KERALA FLOODS

Kerala experienced heavy rainfall in August 2018, which was about 164% higher than the average rainfall of the state in that month. In 2018, the intensity of rain started gradually from June, 15% excess than the average rainfall accounting for 749.6 mm. In July, an 18% excess than the average rainfall of about 857.4 mm was received. Whereas the first 20 days of August received a rainfall of 758.6 mm instead of the average 287.6 mm of rain, which is 164% excess than the average rainfall of the month. As the rainfall exceeded and continued in June, all the reservoirs were filling fast. A severe spell of rain was experienced on the days of 8th and 9th August 2018 in the Malappuram, Wayanad, Idukki, and Palakkad districts, leading to floods in Wayanad and Idukki districts. Several reservoirs reached their Full Reservoir Level (FRL) due to the runoff from all the catchments. Again, another severe spell of rain occurred from 14th August and continued for a week till 19th August, as the state recorded 758.6 mm of rainfall in the first 20 days of August. Among which, 415 mm of rain occurred on 15-17 August 2018, leading to severe flooding in 13 out of the 14 districts in Kerala. In addition to the high intensity of the rain, the dams in FRL opened, leading to massive flooding in the region. According to the report submitted by CWC (2018), dams in Kerala did not add to or reduce the amount of flood in the state. Due to the above-average rainfall in the previous months, the dams were already to the FRL or close to the FRL by 14th August 2018. Even if the dams were less than the FRL, the situation would not be different, as the rainfall was so intense for 3 days and even the 4th day in some locations. It is observed that all parts of Kerala except Kasaragod and Thiruvananthapuram, situated in the extreme north and south of Kerala state, respectively, received unusually heavy rainfall, of which 50% of precipitation was observed in just 20 days, leaving the scene to the worst case. As the Idukki dam opened, Cheruthoni town was the first to be severely flooded. The Periyar and Pamba Rivers were flooded with Peerumade as the eye centre. The Periyar River flooded regions like Cheruthoni town and Kanjikuzhy upstream and Aluva, North Paravoor, Vypin, and Kochi downstream. The Pamba River flooded Ranny, Chengannur, Thiruvalla, Mavellikara, and Alappuzha. The Kuttanad region in the Alappuzha district below the MSL is more prone to flood even during slightly heavy precipitation. The Kakki reservoir in the River Kakki, a tributary of River Pamba, was already full with the massive spell in July. Most of the Kuttanad was

inundated with floodwater during the release of water from the reservoir. The Pamba River basin has a control structure, namely the Kakki dam, whereas the other rivers, like Achencoil, Meenachil, and Manimala Rivers, do not have any dams to control the flow of the river. To control further inundation in the Kuttanad region, the Thottapally barrage spillway at the Vembanad Lake is the only way to discharge the waters from the above four rivers. This Thottapally spillway has a discharging capacity of only about 630 cumecs, which takes a long time for the water to be discharged, and meanwhile can inundate the Kuttanad region. Therefore, opening the Kakki dam also remained challenging for the officials. As the region of Kuttanad was facing severe inundation, any further release would have affected it in the worst way. On the other hand, the Kakki reservoir was also at its FRL anticipating a possible future flood. If Kuttanad had not flooded before the second spell, the water level could have been brought down in the reservoir, saving the region from the extreme flood.

The total runoff generated from 14th - 19th August 2018 rainfall in Pamba, Manimala, Achencoil, and Meenachil was about 1.63 BCM. The Vembanad Lake has a carrying capacity of only about 600 MCM, and the Thaneermukkom barrage exists across the lake. The Thaneermukkom and the Thottappally barrages have a discharge capacity of about 1706 cumecs and 630 cumecs, respectively. Thus, from the 1.63 BCM runoff, only 605 MCM runoff could be drained out from the Vembanad Lake remaining 1 BCM runoff was stored in the lake, creating the rise in water level. Thus, the severe river runoff accumulated water in the Vembanad Lake, shrinking its capacity. The reduced discharge in the Thottappally barrage worsened the flooding in the Kuttanad region. This was why the low-lying region near the Vembanad Lake, such as the districts of Kottayam, Alappuzha, and Pathanamthitta, were flooded.

3.5   2019 KERALA FLOODS

Three active spells—one in July, one in August, and one in September—occurred in 2019 following their commencement, with the strongest spell occurring between the sixth and eleventh of August. After a somewhat drier than typical June and July, it is highly uncommon that the State got more than 150 mm of rain on one day, namely on August 8, which resulted in floods in various sections of the state. The daily rainfall deviation across Kerala exceeds 998% of normal as of August 8, 2019. While the 2018

floods saw scattered rainfall with less cumulative rainfall that fell fairly equally over a 24-hour period, A few places in Kerala, most notably Kozhikode, Malappuram, and Idukki, have had rainfall surpassing 50 mm/2 h, according to 2 hours of cumulative rainfall reported on August 8, 2019. The region affected by this occurrence was bigger than that commonly affected by cloudbursts or Mini Cloud Burst (MCB) events, typically occurring in fifty to hundred square kilometres. A Mesoscale Cloudburst (MsCB) is a geographically dispersed collection of pixels with high rainfall values. This Mesoscale convective storm delivered rainfall so strong that it earned the nickname "mini cloudburst", in contrast to the typical systems developing near India's west coast. In Kerala's documented meteorological history during the monsoon season in August, this occurrence may be the first of its sort. After this MsCB occurrence, numerous areas of Malappuram and Kozhikode experienced flash floods (Vijaykumar et al. 2021).

3.6   2021 KERALA FLOODS

Kerala has witnessed high-intensity floods due to the incessant rains over October. The state was hit by torrential monsoon rains in September that contributed to the floods. Due to the excessive rainfall, rivers and other water bodies overflowed, resulting in extensive flooding and landslides. Hundreds of thousands of people were forced to flee their homes due to the flood, and they had to be rescued and relocated to relief camps. Residents in parts of the coastal state of Kerala were cut off as the rain began to intensify on 15th October 2021. Kottayam and Idukki, which received 180 mm and 300 mm of rain over two days, are the most severely impacted areas. The MsCB that occurred in Kerala during the floods of 2021 resembled that of the floods of 2019. The year 2021 was the wettest in Kerala in the last six decades, with the state receiving 110% more rain than usual from the northwest monsoon. Pathanamthitta district in Kerala experienced a maximum of 186% rainfall during the northwest monsoon season. During the north-west monsoon season, excess rainfall was also recorded in the following places: Kannur (143%), Alappuzha (57%), Ernakulam (102%), Idukki (119%), Kasaragod (141%), Kollam (100%), Kottayam (119%), Kozhikode (135%), Malappuram (76%), Palakkad (104%), Thiruvananthapuram (80%), Thrissur (90%), and Wayanad (80%). Roads are washed away, homes are damaged, and trees are

uprooted due to the state-wide landslides sparked by the heavy rains. The flooding has cut off a number of villages in hilly areas.

CHAPTER 4

CLASSIFICATION AND PREDICTION OF LULC IN THE VEMBANAD LAKE SYSTEM

4.1 GENERAL

The intensified population movement from rural areas and smaller towns to larger cities, often in developing nations, is the main cause of population acceleration. This migration is motivated by the greater quality of life and work prospects (Güneralp and Seto 2008). Even though this phenomenon appears normal, studies have shown that population growth will have numerous adverse effects, particularly on the environment and natural surroundings, without proper planning and management. This will result in a decline in the quality of infrastructure, a reduction in agricultural productivity, and a loss of land use due to microclimatic changes and environmental pollution (Parry et al. 2018). The effects of industrialization-related urbanization have a dynamic relationship between urban development and the environment (Wahap and Shafri 2020).

This chapter explains the classification of LULC in the GEE platform using three nonparametric classification ML algorithms: (i) Random Forest (RF); (ii) Support Vector Machine (SVM), and (iii) Classification And Regression Trees (CART). This analysis is carried out to understand the performance of the models using the same collection of training and validation points. Further, using the CA − Markov chain model, the future Spatio-temporal LULC transition analysis was forecasted, and the changes in LULC for the years 2035 and 2050 were comprehended.

4.2 MATERIALS AND METHODS

4.2.1 Machine Learning Models

In recent years, ML techniques for high-precision classification have developed in remote sensing. LULC is categorized using three ML techniques in the GEE environment: RF, SVM, and CART. These techniques are explained in detail as follows.

4.2.1.1 Random Forest classifier

An RF classifier is a multi-decision tree ensemble classifier that creates many decision trees using a random selection of training samples and variables. In recent years, these ensemble-learning approaches have been frequently applied in remote sensing applications. It was introduced by Breiman (2001), who combined the output of individual decision trees as a non-parametric classifier. So no statistical assumptions must be made before the data distribution. RF trees are different from Decision Trees in the way of selecting each node of a subgroup among the input variables in a random manner and are built without pruning (Pelletier et al. 2016). Hence, this is a recursive process until the samples are similar or the splitting no longer enriches the model in each subgroup. Random-decision Forests control the over-fitting of a training set.

A set of predictive variables is chosen individually and influences the response of each tree; it has the same distribution for all of the trees in the forest and is a subset of the original dataset's predictor values (Tang et al. 2020; Wang et al. 2015; Zhu and Zhang 2022). Kim and Kim (2020) explained the RF method in four steps: (a) N bootstrap samples are extracted, i.e., they are chosen at random while accounting for overlap in the training dataset; (b) Each bootstrap sample is being used to train a decision tree, and d features are chosen at random, with no duplicate data allowed at any node. The tree nodes are separated in order to create the optimal partition based on an objective function such as information gain; (c) repeat the steps n times; (d) get all predictions out of each decision tree and assign the class using dominant votes. The ensemble of simple trees determines the reaction, the most common class. The average of the findings is utilized in a regression procedure to estimate the dependent variable. The usage of tree ensembles might result in a significant improvement in prediction accuracy (Band et al. 2020). It makes no assumptions about the connections between the dependent and independent variables, and it is an appropriate tool for analysing hierarchical and non-linear interactions in huge datasets (Lee et al. 2017; Zhu and Zhang 2022). The RF algorithm is based on tree-structured classifiers and has been expressed as follows:

$$h(x, {}^{i}k), k = 1, 2, ...n \qquad (4.1)$$

Where ${}^{i}k$ represents conditioning factors, and $1, 2…n$ are input vectors x.

4.2.1.2 Classification And Regression Trees (CART)

Classification And Regression Trees (CART) is a simple binary decision tree classifier developed by Breiman et al. (1984). A sequence of nodes linked to one another, each node is divided into two branches, ultimately these branches leading to leaf nodes representing class labels in classification trees and continuous variables in regression trees. The node splitting continues until a threshold condition is reached. Using the Gini Impurity Index, CART determines which input features will provide the best split at each node. Despite a general tendency to overfitting, it is one of the most extensively used LULC classifiers because of its classification accuracy and performance. Compared to multilayer neural networks, the fundamental advantage of this architecture is that classification choices can be considered a white box system with understood and interpreted input-output relations (Mather and Tso 2016). CART uses the cross-validation technique for pruning, which eliminates branches, and their removal will not affect the results beyond a certain threshold. This may result in a fall in accuracy for training data classification and the loss of certain information, but it also improves accuracy for unknown data for testing (Shao and Lunetta 2012).

4.2.1.3 Support Vector Machine

Support Vector Machine (SVM), introduced by Cortes and Vapnik (1995), is based on statistical learning theory. By minimizing the empirical risk and confidence interval of learning derived from the systemic risk minimization hypothesis, it aimed to achieve strong generalization capability. It is an efficient and robust algorithm for both classification and regression. The SVM concept uses the support vectors at the class domain's edges to create hyperplanes between classes in feature space. The model seeks an optimal hyperplane to separate the classes at maximum margins (Shi and Yang 2015). SVM was designed to handle linearly separable classes by bilinear classification, but hyperplane may not be located between the two in most cases. In such cases, the model converts the highly inseparable data into a higher dimension or infinite-dimensional feature space to separate it linearly. Theoretically, the error penalty, which allows for misclassification, substantially impacts SVM classification accuracy (Ustuner et al. 2015).

4.2.2 Data Sources and Preparation

The data prepared from various sources have shown in Table 4.1. Landsat - 7 and Landsat - 8 Top of Atmosphere (TOA) images are used for the classification in the GEE environment. The cloud cover is set to less than 30% in the imagery, and the classification is predominantly focused on the pre-monsoon season (February – May). Landsat - 7 was chosen for 2009, while Landsat - 8 was chosen for 2013, 2015, 2017, and 2019. Digital Elevation Model (DEM) is prepared using the ASTER data in the ArcGIS environment. Slope and stream networks are derived from the DEM using hydrology tools in the ArcGIS environment. OpenStreetMap is an open licensed map that creates a free editable map worldwide that allows users to download all the features like roads, rivers, and streams to the desired vector file. Hence, the road layer is downloaded from OpenStreetMap. The built-up and forest are extracted from the classified 2019 land cover using the Raster tool in ArcGIS. Public and Industrial areas are prepared from google earth pro, comprising schools, grounds, hospitals, hotels, religious places, industrial and commercial places, railway stations, bus stations, airports, and other frequently accessible places by the public like markets, theatres, etc. The population density map is derived from the Census of India (2011). According to the Census of India (2011), the Kerala state has the highest population density in the country. In the GEE environment, mean annual rainfall is collected from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). All the data are converted to the Projected Coordinate System WGS 1984 UTM Zone 43N and fed into the CA-Markov model.

Table 4.1 Data Sources for LULCC Modelling

Sl. No	Parameter	Data Source	Software used	Resolution (m)	Feature type
1	LULC	Landsat - 7 & 8 Path: 143 and144; Row: 53 and 54	GEE	30	-
2	DEM	ASTER	ArcGIS	30	-
3	Slope	ASTER	ArcGIS	30	-

4	Road	OpenStreetMap	ArcGIS	-	Line
5	Built-up	LULC	ArcGIS	30	-
6	Stream	ASTER	ArcGIS	-	Line
7	Public and Industrial areas	Google data	Google Earth Engine	-	Point

4.2.3 Land Use and Land Cover classification

GEE helps analyse the data at a planetary scale and helps the researchers and developers detect changes in the map trends and quantify differences on the Earth's surface. Landsat - 7 and 8 TOA reflectance images are available in the environment, where they are calibrated using Chander et al. (2009) proposed rescaling parameters. The images before and after the specified duration considered in the research were used to replace and supplement the images obscured by clouds and fog, and most pixel image composites were created. In the GEE, median *ee.Reducer* function determines the median value of each pixel in an image collection for a specified timeline to reduce the collection to a single image for each year. For a specific year, the best image is obtained from the set of images. The obtained image is combined with the Normalized Difference Vegetation Index (NDVI) (Figure 4.1), Modified Normalized Water Index (MNDWI) (Figure 4.2), Normalized Difference Built-up Index (NDBI) (Figure 4.3), and Bare Soil Index (BSI) (Figure 4.4) to further enhance and differentiate the classes in the classification on GEE platform.
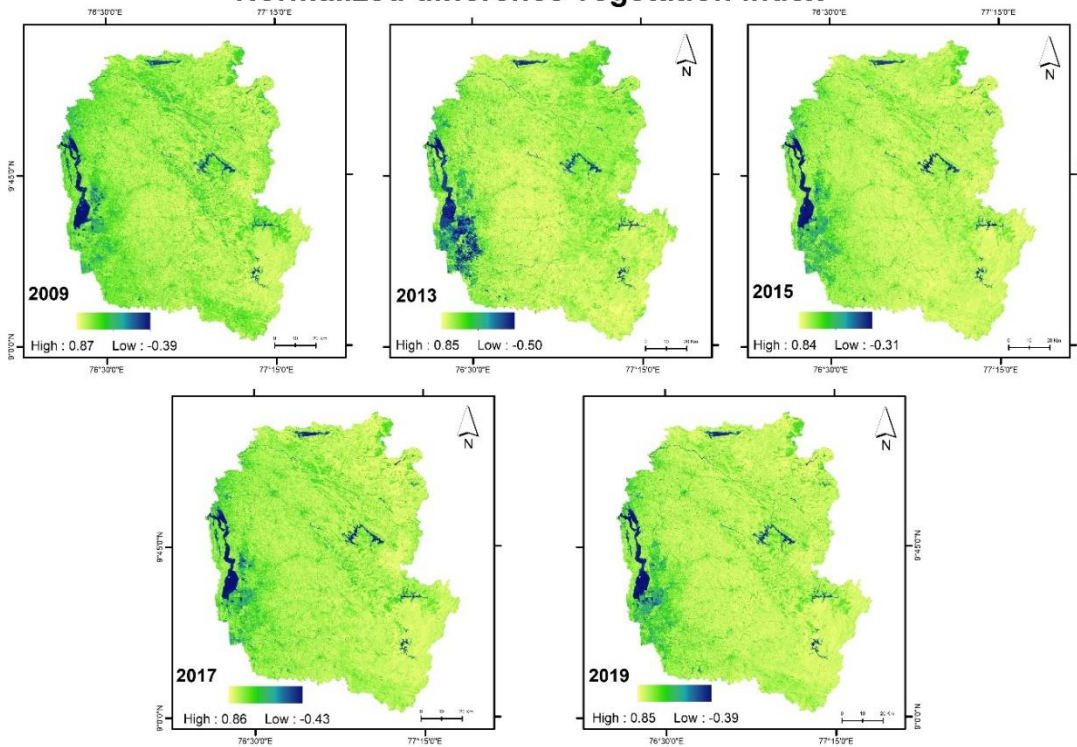
Figure 4.1 NDVI map for LULC classification
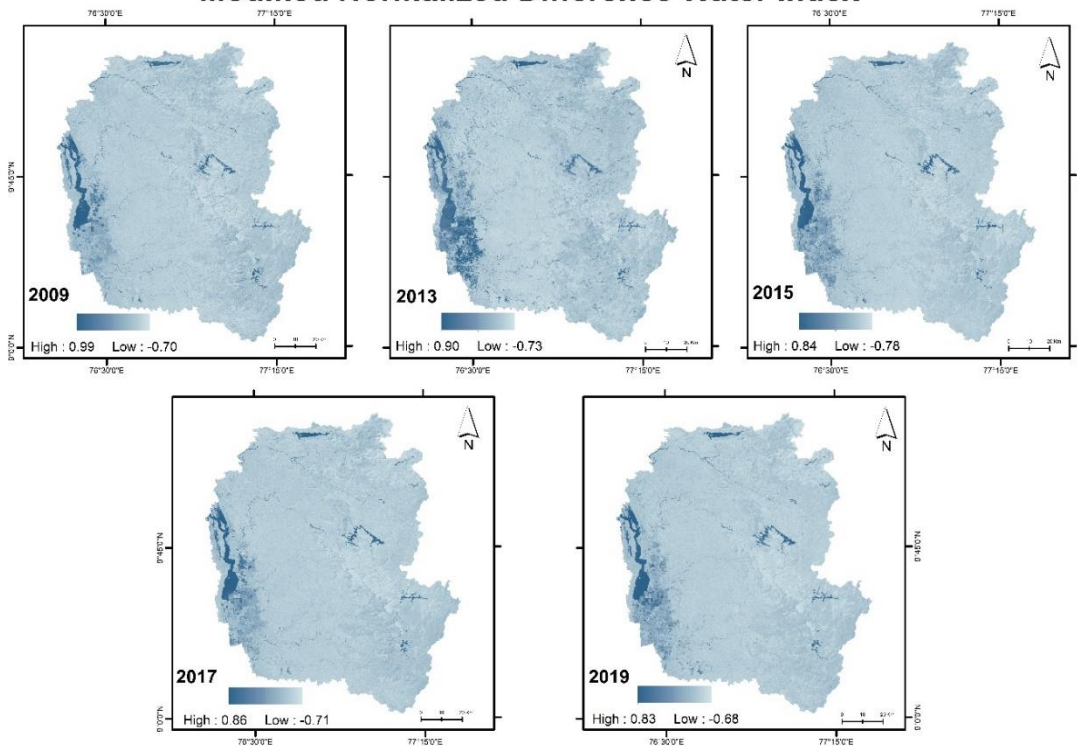
Figure 4.2 MNDWI map for LULC classification

64

Normalized Difference Built-up Index
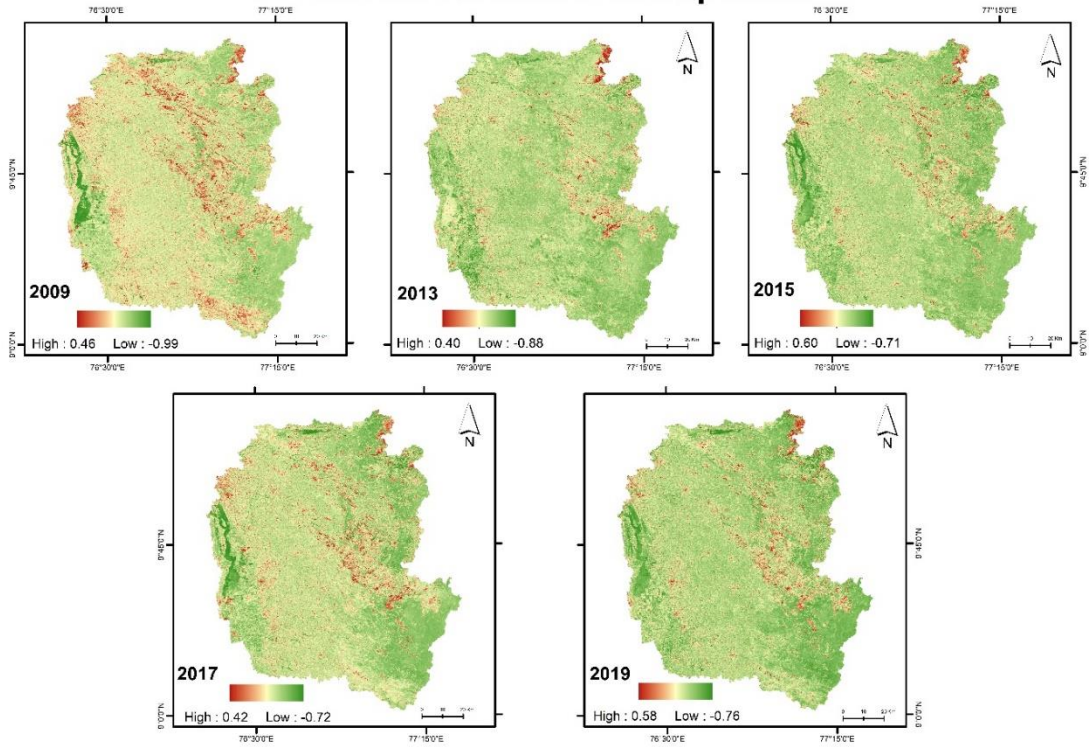
Figure 4.3 NDBI map for LULC classification

Bare Soil Index
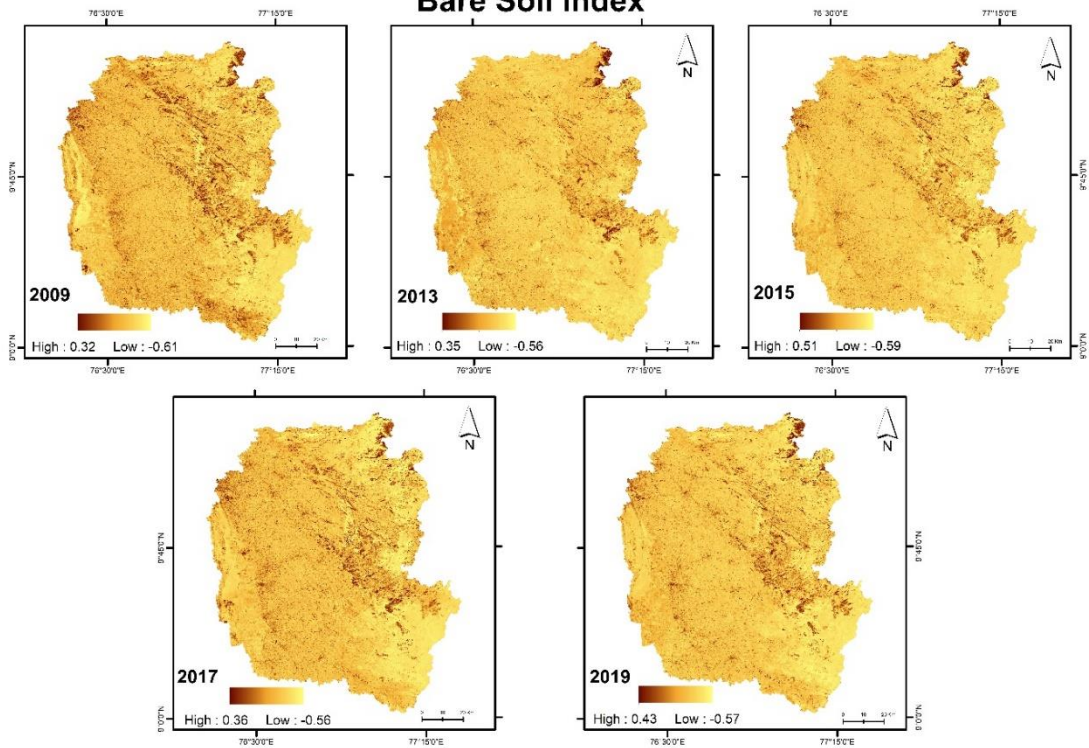
Figure 4.4 BSI map for LULC classification

4.2.3.1 Tuning Parameters for Machine Learning Models

The RF classifier has attracted the interest of the remote sensing community (Abijith and Saravanan 2022; Belgiu and Drăgu 2016; Gislason et al. 2006; Hamad et al. 2018a; Mahdianpari et al. 2018). RF algorithm in GEE allows the users to set the following arguments: number of decision trees to create (ntree), number of variables per split, a fraction of input to bag per tree, and maximum number of leaf nodes in each tree. Ntree is the essential input (Noi Phan et al. 2020). The RF algorithm is given as input using the following code in GEE *ee.Classifier.smileRandomForest(numberOfTrees, variablesPerSplit, minLeafPopulation, bagFraction, maxNodes, seed)* where an empty Random Forest classifier is created.

CART is one of the most widely used supervised ML models for evaluating the effectiveness of the detailed and automated LULC classification approach (Pan et al. 2022; Shetty et al. 2021). The important tuning parameter for CART in GEE is the maximum number of leaf nodes in each tree. The CART algorithm is given as the input using the code *ee.Classifier.smileCart(maxNodes, minLeafPopulation),* where an empty CART classifier is created.

The SVM, a novel algorithm based on statistical learning theory, has not been exploited much within the remote sensing community (Adam et al. 2014; Adelabu et al. 2014; Shi and Yang 2015). SVM provides many tuning parameters in the model, but the most important parameters are kernel type, gamma value in the kernel function, and cost (C) parameter (Yang 2011). An SVM classifier is created using the code *ee.Classifier.libsvm(decisionProcedure, svmType, kernelType, shrinking, degree, gamma, coef0, cost, nu, terminationEpsilon, lossEpsilon, oneClass)* in GEE. As a result, we aimed to test, investigate, and compare the performance of these ML algorithms in the classification of LULCs.

4.2.4 Classification Accuracy

The classification is prepared by giving the Region Of Interest (ROI) for the selected image in points and polygons. To provide a better ROI, 15-35 pixels in each sample were selected. All three models were given the same collection of training and validation points as input to examine their results closely. It is classified into 5 major classes: waterbody, built-up, vegetation, barren land, and forest. Every class was trained

with 75 - 95 ROI for classification and 60 - 75 ROI for validation. It also ensured that the information was normally distributed and spectrally pure. According to Lillesand and Kiefer (1979), a minimum of 50 samples per class is required as a rule of thumb. Random stratified sampling is used to determine accuracy, with minimal observations put in each segment at random.

Once the classification was done using the above ML techniques, an error matrix was generated for each year to identify the accuracy of the classifications. The degree to which the results are close to values accepted as true is referred to as accuracy. Thus, various matrices such as Overall Accuracy (OA), consumer accuracy, producer accuracy, and kappa statistics (Damtea et al. 2020) are identified from the error matrix. Consumer accuracy of each class depends on the number of correctly classified pixels in the class to the number of pixels belonging to this class in the classification. Simultaneously, the producer accuracy depends on the number of correctly classified pixels to that of the number of pixels belonging to each class in the reference data. The notation is similar to Cohen (1960), and the kappa coefficient κ is estimated. The reduction of errors by the classification classes is proportional to the error of the completely random class (Forghani et al. 2007; Tassi and Vizzari 2020). The magnitude of κ usually lies between -1 to +1. Values above +0.5 indicate it is in good agreement with the classification (Taati et al. 2015). The best-performing model is analysed. Then the classified model is further future-predicted to analyse the spatio-temporal change in the model.

4.2.5   CA-Markov model

CA-Markov analysis has been carried out using the Idrisi TerrSet software in the LCM module. The LULC map 2009 and 2015 are given as the earlier and later land cover images, respectively. The transition probabilities from 2009 to 2015 and the 2015 LULC base map were used to model the 2019 LULC using the CA-Markov approach. CA assesses the contiguity configuration as well as transition probabilities (Hamad et al. 2018a). The CA-Markov model, which consists of the basic LULC layer, transition potential areas for future change formed by the Markov Chain model, and potential transitional layers for LULC, such as road network, built-up regions, and stream network, are used to determine the appropriate transition from one class to the other (Halmy et al. 2015a).

The potential for transition probability is calculated based on the area of each LULC class (preferably barren land, forest, and vegetation) that can be converted into the built-up (Figure 4.5). The transition of these areas was divided into the number of timesteps (as an iteration) given by the simulation, which provided the areas to be transitioned each iteration. Land suitability maps show the most suitable pixel for each LULC class.
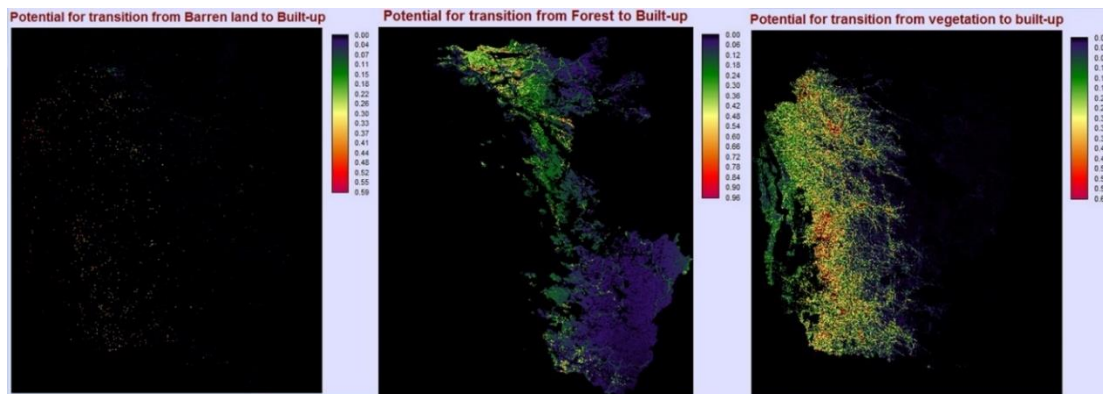
Figure 4.5 Potential for the transition from barren land (1), forest (2), and Vegetation (3) to built-up
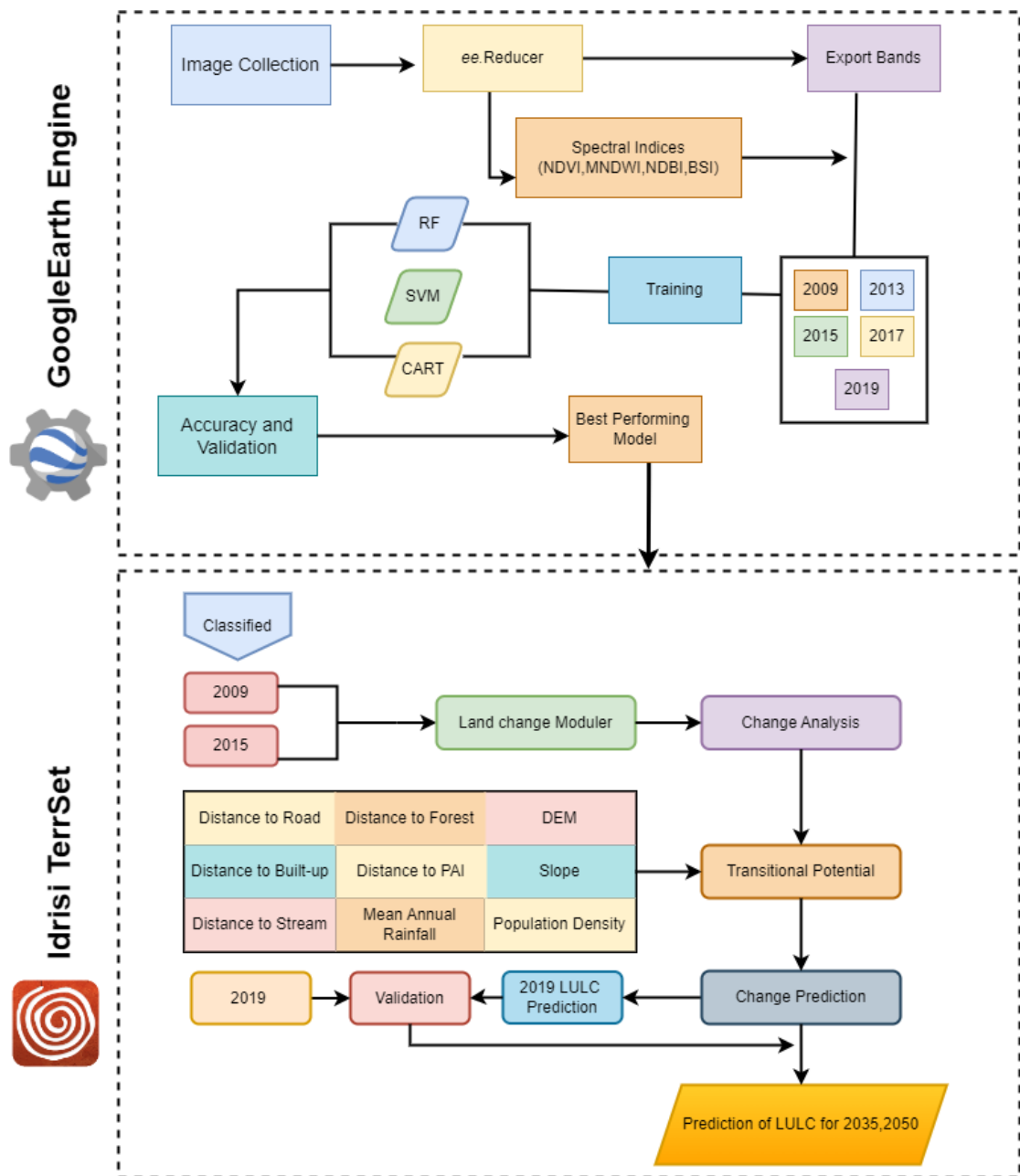
Figure 4.6 Methodology for the classification and prediction of LULC change

Down-weighted pixels are further apart from the pixel to be converted. These neighbourhood rules were defined using the contiguity filter of 5x5 pixels. As a result, each cell is surrounded by a 5x5 - matrix space that defines the neighbourhood of each land class. The standard contiguity filter is given in the equation (4.2). Assignment of the pixels in the future depends on the suitability of the pixel to a specific LULC class. This simulation continues till each pixel of the LULC class is iterated. The LULC layer of 2009 and 2015 is the input to predict the LULC for 2019. This is performed to prove

the model's accuracy under the given conditions. Then the model is further used to predict future spatio-temporal LULC changes. The Overall methodology for the study is shown in Figure 4.6.

$$\text{contiguity filter } 5 \times 5 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad (4.2)$$

4.3 RESULTS AND DISCUSSION

4.3.1 LULC using GEE

The LULC for 2009, 2013, 2015, 2017, and 2019 were classified using RF, SVM, and CART in GEE, as shown in Figure 4.7, Figure 4.8, and Figure 4.9, respectively. The indices help classify the images accurately by different means. The NDVI and MNDWI are widely used to extract vegetation, forest, and waterbodies, whereas NDBI and BSI aid in the classification of built-up and barren land, respectively (Liu et al. 2020a). Previous studies by Cánovas-García et al. (2017); Ghimire et al. (2012); Noi Phan et al. (2020) stated the performance of the RF model is good at ntree of 100. However, according to the current study, the ntree of 30 functioned better and had greater accuracy than the ntree of 100. For 2009, this model accurately categorized the waterbody, built-up, and barren land than the forest and vegetation type. Using Landsat - 8 images, the differentiation between forest and vegetation was clear and precise. This may be attributed to the Landsat - 8's (level 1 products) higher radiometric resolution, i.e., twice the predecessor Landsat - 7 (Abuzar et al. 2014; Mancino et al. 2020). Table 4.2 shows the consumers' and producers' accuracy for RF classification. The consumers' and producers' accuracy is relatively high for the forest classes, whereas in the built-up areas, it is low.
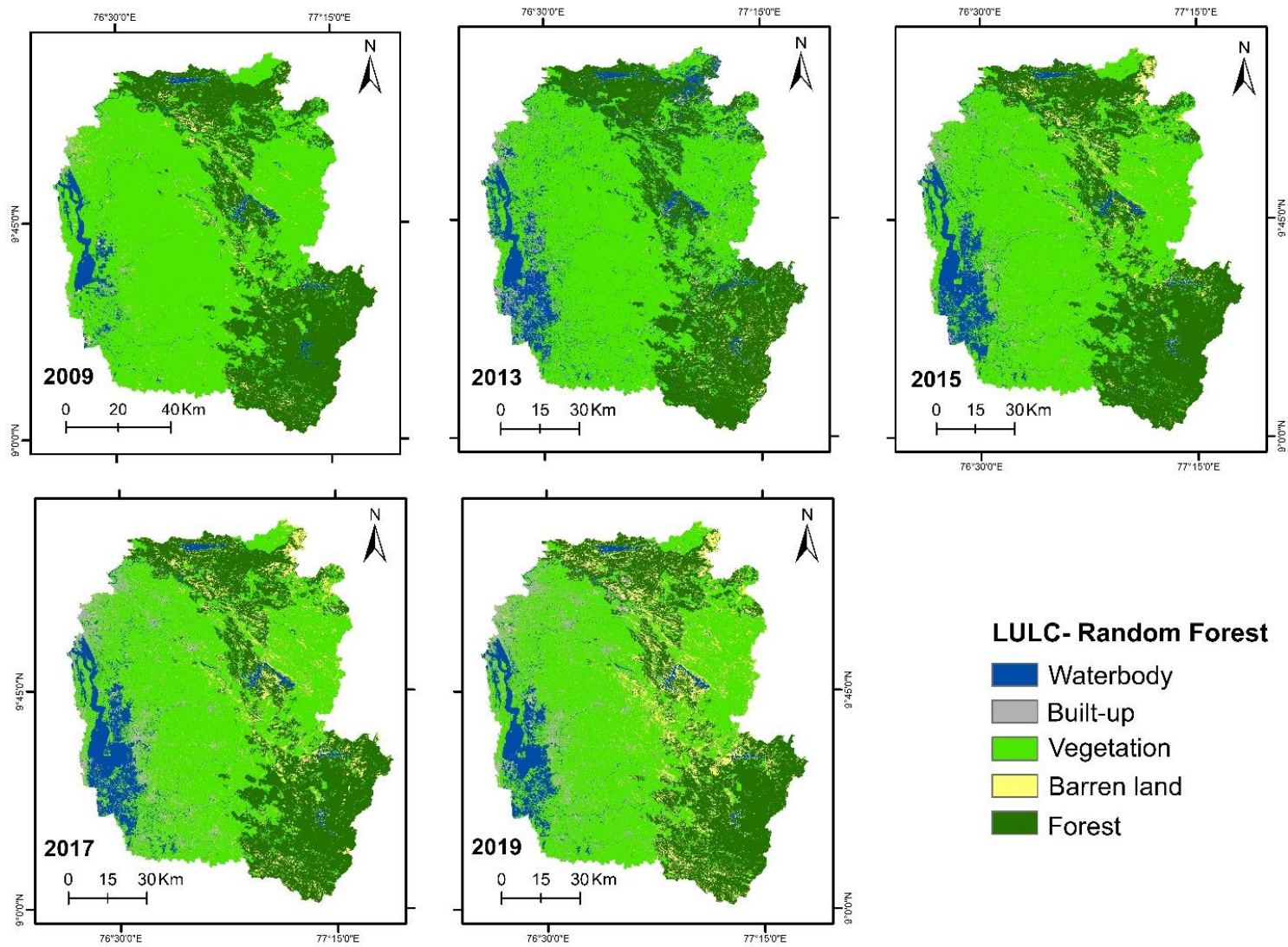
Figure 4.7 Classification of LULC using RF in GEE
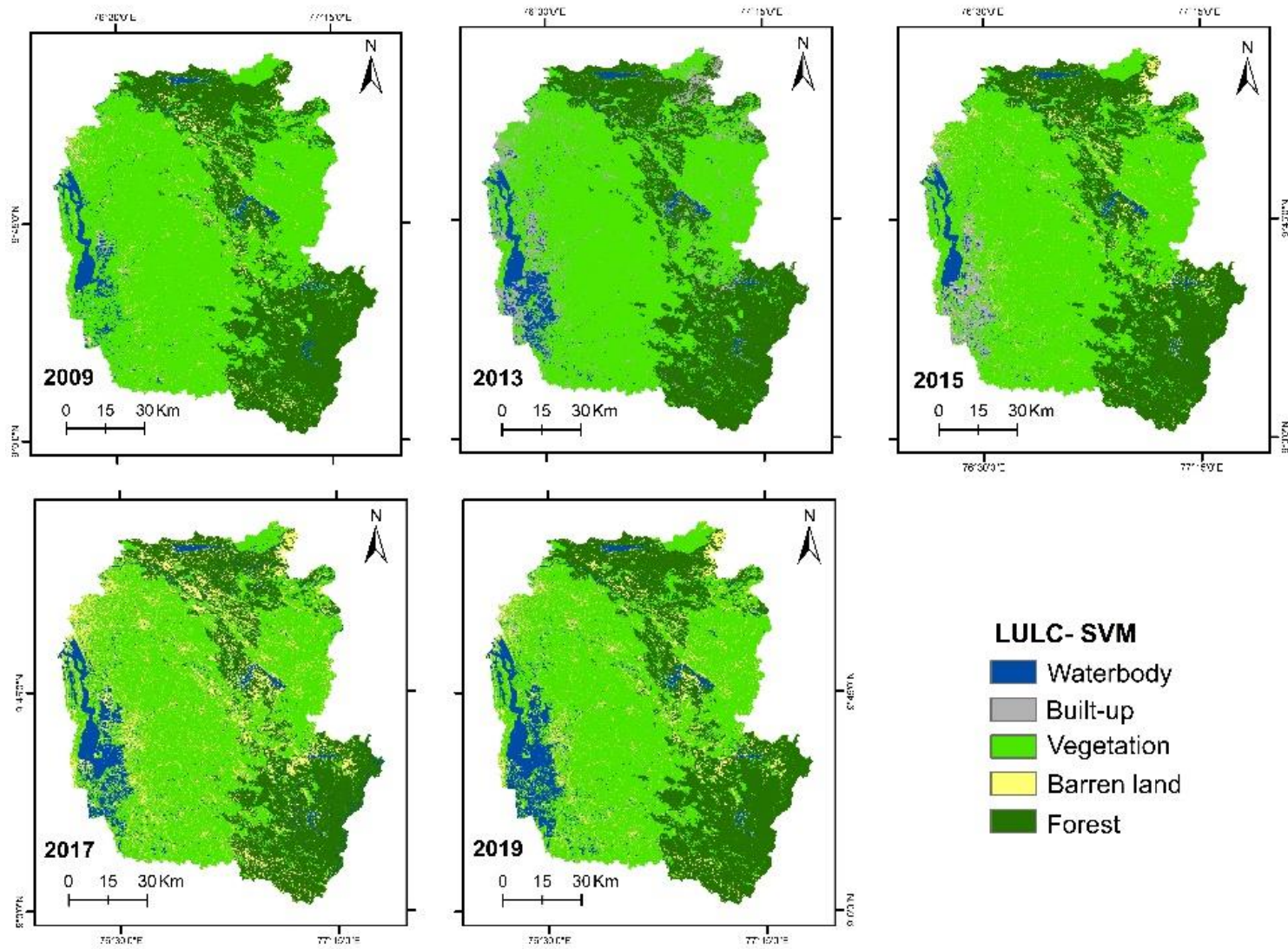
Figure 4.8 Classification of LULC using SVM in GEE
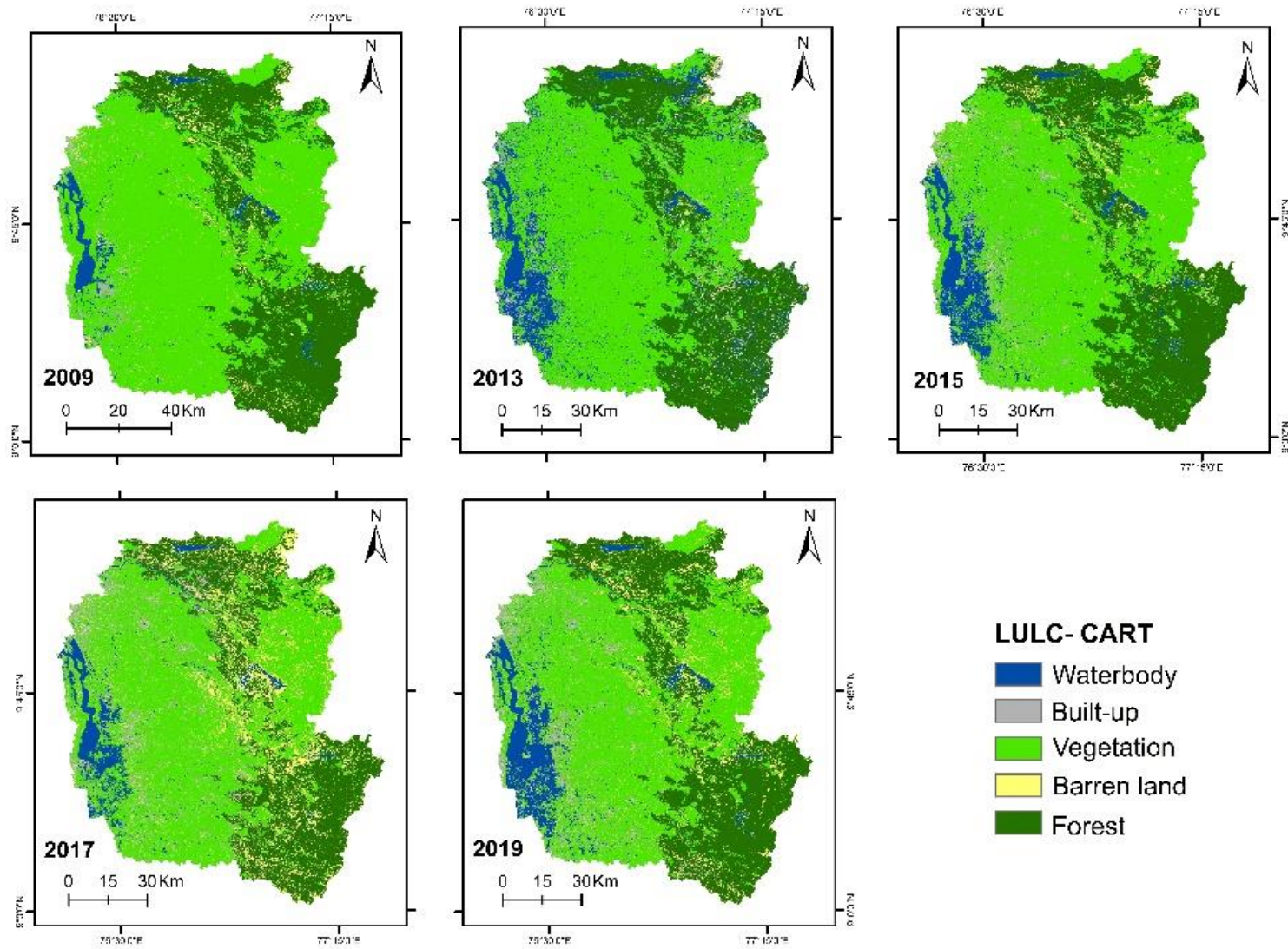
Figure 4.9 Classification of LULC using CART in GEE

Over the image, an SVM model is trained with the same collection of ROI and validation points. The importance of the SVM settings significantly impacts classification results (Huang et al. 2002; Shi and Yang 2015). The kernel type, cost parameter, and gamma value are the critical parameters in the model. Hence kernel type is assigned with Radial Basis Function (RBF), cost (C) is assigned with the value of 100, and gamma value is given as 0.143 (Kavzoglu and Colkesen 2009; Yang 2011). Table 4.3 shows the consumers' and producers' accuracy in SVM. In the built-up and barren land classes, the consumers' accuracy is so low that the former is classified as the latter.

The CART model with the same training and testing samples for the respective years performed considerably better than the SVM. The maximum number of leaf nodes is set to default. In 2009, the shallow waterbodies on the southern side of the lake were misclassified as built-up due to the identical pattern in the Landsat - 7 images. Table 4.4 reveals that, except for built-up areas, the CART model's consumers' and producers' accuracy is comparable to that of the RF classification all the year. Table 4.5 shows the performance of RF, SVM, and CART algorithms for each year. In 2009, both RF and CART performed equally well in the classification. Conversely, CART did not perform better in the other years, resulting in a modest change in the OA and κ.

Table 4.2 Consumers' accuracy and Producers' accuracy using RF for each year of LULC

LULC	Consumers' accuracy (%)					Producers' accuracy (%)				
	2009	**2013**	**2015**	**2017**	**2019**	**2009**	**2013**	**2015**	**2017**	**2019**
Waterbody	100	88.46	96.51	94.7	100	100	80.7	87.37	96.43	94.8
Built-up	95	76.66	72.37	89.55	83.33	91.57	92	87.3	80	85.53
Vegetation	75.6	94.82	93.33	87.18	97.53	97	93.21	96	100	97.53
Barren land	88.78	92.54	97.8	85.29	85.77	72.04	85.87	81.48	85.29	86
Forest	89.35	90.14	89.36	91.12	92.65	88.23	94.58	93.82	88.25	91.35

Table 4.3 Consumers' accuracy and Producers' accuracy using SVM for each year of LULC

LULC	Consumers' accuracy (%)					Producers' accuracy (%)				
	2009	**2013**	**2015**	**2017**	**2019**	**2009**	**2013**	**2015**	**2017**	**2019**
Waterbody	100	82	94	87.5	94.66	100	71.92	98.36	97	92.2
Built-up	90.18	63.8	91.17	89.65	82.09	83.13	92	58.49	64.67	68
Vegetation	76.86	80	92.1	95.77	76.95	94	89	90.3	92	75.76
Barren land	81.82	84.3	55.81	60.19	74	67.74	53.22	80.57	89.17	88
Forest	83.85	83.34	79.41	87.39	84.14	89.55	86.11	84.25	81.88	83.6

Table 4.4 Consumers' accuracy and Producers' accuracy using CART for each year of LULC

LULC	Consumers' accuracy (%)					Producers' accuracy (%)				
	2009	**2013**	**2015**	**2017**	**2019**	**2009**	**2013**	**2015**	**2017**	**2019**
Waterbody	100	76.27	96.51	95.14	100	100	78.95	87.37	91.64	92.2
Built-up	100	72.88	69.62	92.16	84.4	90.36	86	87.3	75.67	85.53
Vegetation	76.85	91.67	95.89	83.31	92.59	100	98.21	100	92	92.59
Barren land	89.04	93.62	90.9	82.13	80	69.89	70.97	74.07	89.64	88
Forest	85.28	82.66	86.46	84.32	91.22	89.24	82.94	88.89	88.36	89.68

Table 4.5 Comparison of OA and κ for RF, SVM, and CART models in their respective years

Year	RF		SVM		CART	
	OA (%)	**κ (%)**	**OA (%)**	**κ (%)**	**OA (%)**	**κ (%)**
2009	89.47	85.9	86.45	81.77	89.47	85.9
2013	88.44	84.6	78.22	71.14	83.11	77.52
2015	89.36	85.63	82.76	77.11	87.94	83.72
2017	89.89	86.49	84.4	79.65	87.01	84.98
2019	91.19	88.16	81.69	75.64	89.79	86.28
Average	89.67	86.16	82.70	77.06	87.46	83.68

RF is more suited to multiclass problems to handle small differences in classification. The SVM model shows some misclassification in the region around the water bodies. During the dry season, the classification is accurate despite RF's outperforming SVM in terms of efficiency. When the three approaches were compared, SVM performed poorly, with an average OA of about 82.7%; CART came in second with 87.46%, and the RF model performed well, with an average OA of 89.67% (Table 4.5). Similarly, the average kappa for SVM, CART and RF algorithms are 77.06, 83.68, and 86.16, respectively. The RF outperformed the SVM and CART in almost comparable spectral classifications, such as barren land and built-up regions. As a result, the RF algorithm-based classification is used for further analysis.

Figure 4.10 shows the change in the area of each LULC class from the year 2009 to 2019. The area of waterbody decreased steadily from 2013 to 2017, but after the flooding, the area of waterbody increased significantly (Table 4.6). It has also been observed that the vegetation cover declined by 8% between 2017 and 2019. There was an increase in the built-up area from the year 2009 to the year 2019. In a decade, the built-up area has grown by almost 97%. This alarming change is also the main driver for the present research, as the cities of Kerala state exhibit the highest urbanization rate. Figure 4.10 shows the increasing trend of barren land between 2009 and 2017, whereas the trend shows a rapid shoot-up beyond 2017 till 2019. This may be attributed to the combined effect of anthropogenic activities and the state's floods in 2018 and 2019, which caused landslides in the region. It accounts for a 125% increase in barren land from 2009 to 2019, resulting in decreased vegetation and forest cover combined. All the above changes in the LULC are predominantly observed in the RF - classified image in GEE over a decadal period (2009 - 2019) in the study region. The nature of the change of LULC is represented in Figure 4.7, and the changes are quantified in Table 4.6.
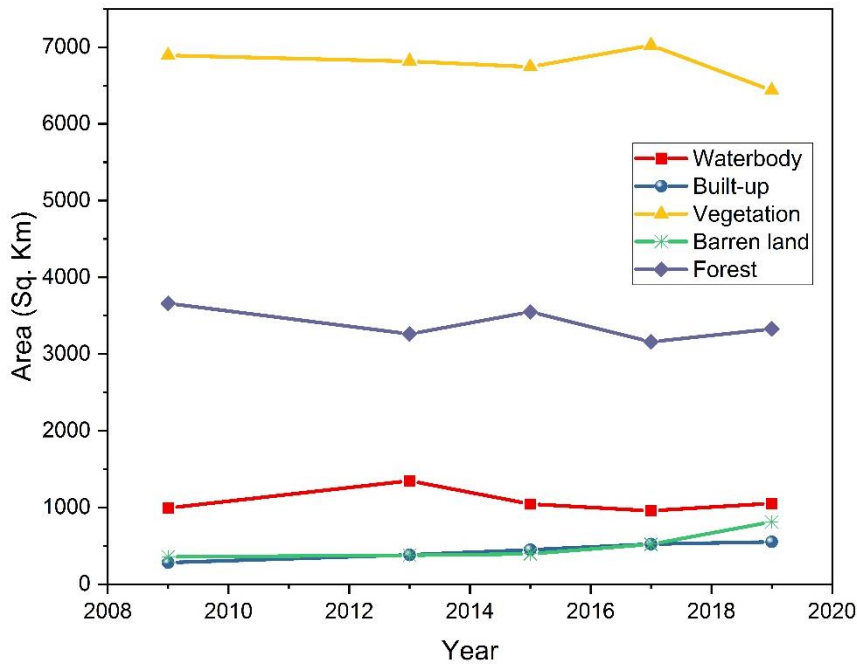
Figure 4.10 Graphical representation of change in the area of LULC from 2009 to 2019

The classified image based on the RF algorithm (Figure 4.7) is then used to forecast and interpret the LULC change in 2019, 2035, and 2050 using CA-Markov analysis in LCM using Idrisi TerrSet software.

Table 4.6 Change in the area of LULC classes from 2009 to 2019

Year	Area (km^2)				
	Waterbody	Built-up	Vegetation	Barren land	Forest
2009	992.50	280.78	6890.39	358.50	3661.08
2013	1349.44	382.33	6814.76	376.12	3260.59
2015	1044.93	449.05	6743.43	396.29	3549.55
2017	956.43	523.93	7023.94	521.36	3157.59
2019	1054.38	552.36	6437.12	812.70	3326.70

4.3.2 Future prediction of LULC using CA-Markov

Changes in LULC are influenced by various driver variables such as DEM, slope, population density, mean annual rainfall, distance from the road, built-up, stream network, forest, and Public and Industrial areas. These driver variables are shown in Figure 4.11. The DEM and slope reveal that the built-up class has a higher probability

78

in areas with a lower elevation and a lesser degree of slope. Distance from the road, distance from the built-up, and distance from the stream are the driver variables that signify the likelihood of a rise in the built-up classes that are very close to the road, current built-up, and stream, respectively, than those that are farther apart. In this analysis, the distance from the road, DEM, slope, and distance from built-up areas are generated and used to predict the increase in built-up areas, but the resulting map failed to perform better. Including the distance from the stream to the model increased the accuracy of the forecasted map for 2019. Thus, it reveals that dwellings in urban areas rely heavily on the river valley, as many residents depend on agriculture. In addition, to the model, some parameters like distance from the forest and Public and Industrial areas, mean annual rainfall and population density made the model still more reliable in predicting the future trends of the LULC, and the resultant map shows high accuracy in all the classes except barren land.

The association between the driver variables and the distribution of LULC classes in later year image (2015) has been derived by finding Cramer's V. Although it is an imprecise fishing tool, higher Cramer's V shows that the variable's possible explanatory value is good. However, it does not guarantee good results because it does not account for the mathematical constraints of the modelling technique and the relationship's complexity. The Cramer's V for the driver variables is given in Table 4.7. V<0.3 shows a weak association, V∈ [0.3-0.5] shows a medium association and V>0.5 shows a strong association. Mean annual rainfall shows a weak association with the classification. Distance from built-up, Public and Industrial areas, forest, and slope shows a medium association, whereas DEM, distance from the road, stream, and population density, shows potentially high Cramer's V.

The model's accuracy is evaluated with the crosstab module in the Idrisi TerrSet to analyse the agreement of the predicted 2019 image to that of the classified 2019 image. The predicted image of 2019 is in very good agreement with the classified image of 2019, with a kappa value of 84.90%. Hence, Table 4.8 compares the predicted map of 2019 to the classified map of 2019, and the accuracy for each class in LULC is evaluated. The LULC classes have a high correlation, with more than 95% confidence precision in the waterbody, built-up area, and forest cover.
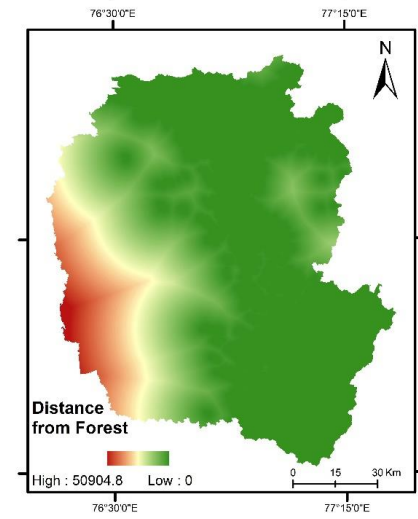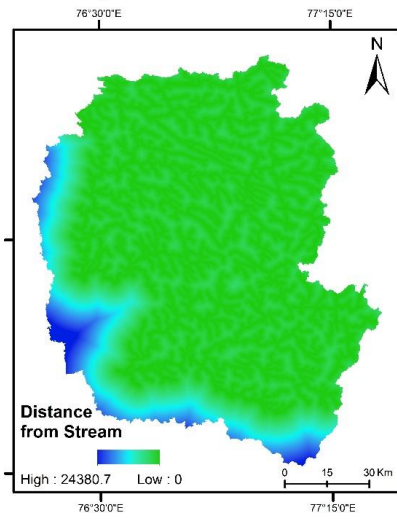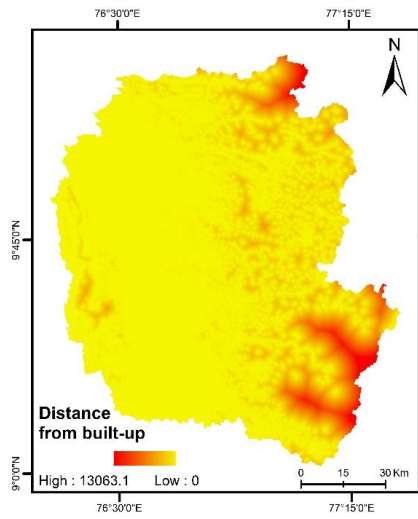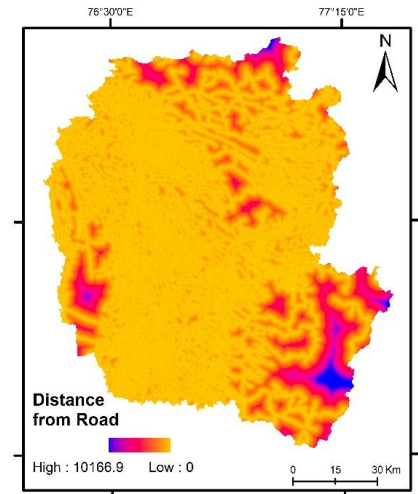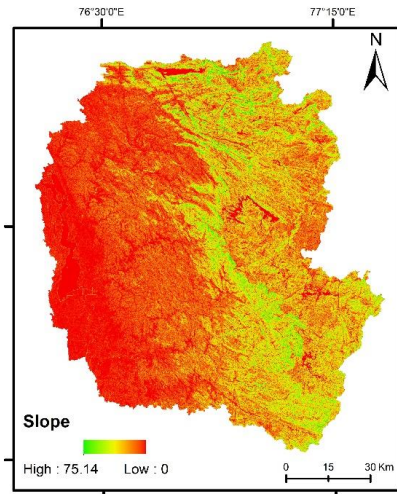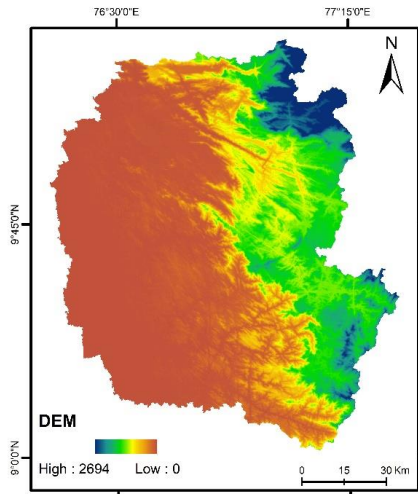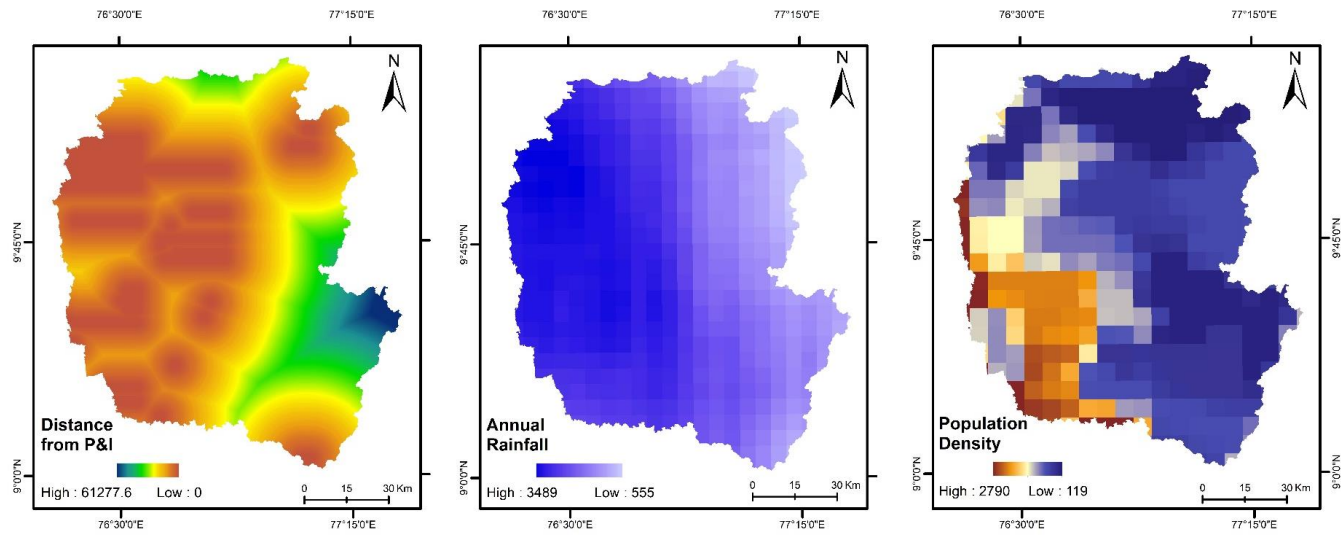
Figure 4.11 Driver variables for the CA-Markov model

Table 4.7 Cramer's V for driver variables

Driver variable	Cramer's V
DEM	0.58
Slope	0.37
Distance from road	0.60
Distance from built-up	0.43
Distance from stream	0.56
Population density	0.61
Distance from Public and Industrial areas	0.46
Distance from forest	0.40
Mean annual rainfall	0.22

The accuracy of the vegetation cover is 94.63% compared to the barren land, which is 47.21%, the least accurate compared to the classified 2019 LULC map. The study area was one of the most severely damaged and devastated during the 2018 and 2019 Kerala floods. These heavy floods caused landslides and erosion along the river banks in the Western Ghats section (Jacinth Jennifer and Saravanan 2022). This is one of the key explanations for more barren land than predicted by the model and is visible in Figure 4.12.

Table 4.8 Accuracy assessment of the predicted 2019 LULC with actual 2019 LULC

LULC	2019 actual (km²)	2019 predicted (km²)	Accuracy (%)
Waterbody	1054.38	1034.14	98.08
Built-up	552.36	555.16	99.49
Vegetation	6437.12	6782.62	94.63
Barren land	812.70	383.70	47.21
Forest	3326.70	3427.63	96.97
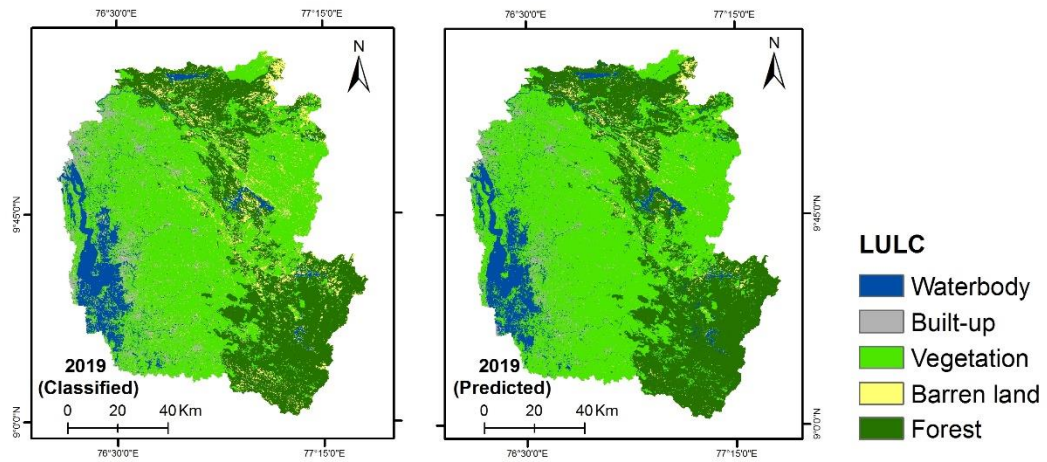
Figure 4.12 Classified and Predicted 2019 LULC

Table 4.9 Change in the area of the predicted LULC 2019, 2035 and 2050

Year	Area (km²)				
	Waterbody	**Built-up**	**Vegetation**	**Barren land**	**Forest**
2019 Predicted	1034.14	555.16	6782.62	383.70	3427.63
2035 Predicted	1024.66	633.64	6743.83	446.67	3334.45
2050 Predicted	1018.99	684.01	6665.16	448.84	3366.25

The change in the area of the predicted LULC of 2019, 2035, and 2050 are shown in Figure 4.13. With respect to the 2019 predicted image, there is a gradual growth in built-up and barren land in 2035 and 2050 (Table 4.9). The predicted barren land area for 2019 is 383.7 km², whereas the total area of the classified image is 812.7 km² (Table 4.6). The forecast shows the annual rise in barren land accounts for 446.67 km² and 448.84 km² for 2035 and 2050, respectively. There is a slight reduction in built-up area from 555.16 km² in predicted 2019 to 552.36 km² in classified 2019 LULC (Table 4.6). The estimated growth in a built-up area for 2035 and 2050 is 633.64 km² and 684.01 km², respectively. The analysis clearly showed a considerable decrease in vegetation cover from 2019 to 2050.
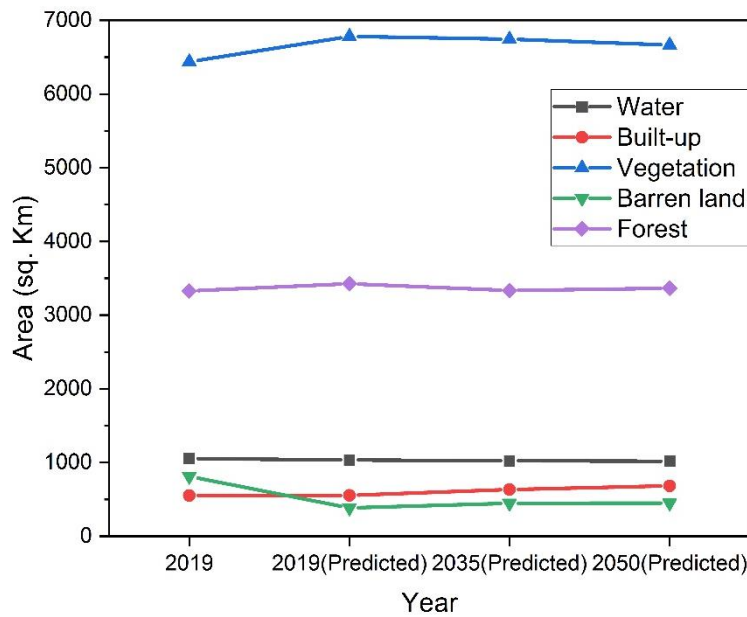
Figure 4.13 Graphical visualization of change in the area of the actual LULC 2019 with that of the predicted LULC for 2019, 2035 and 2050
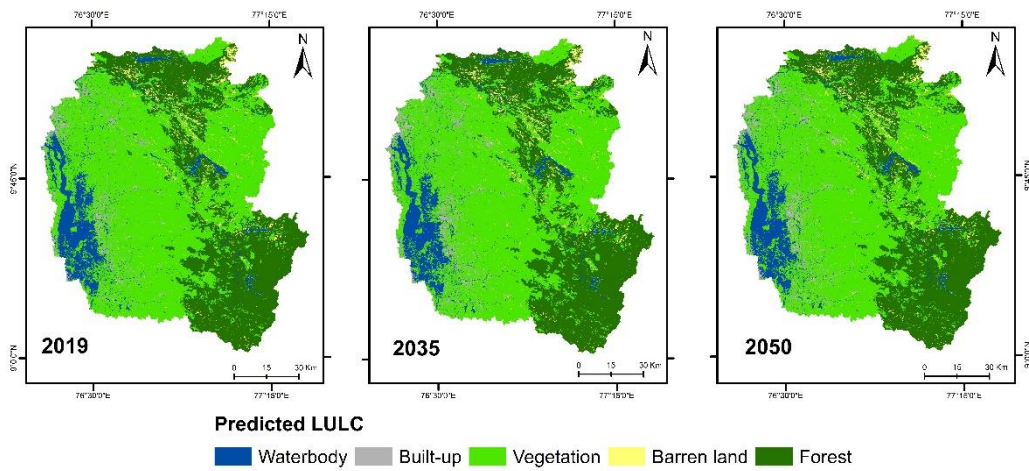
Figure 4.14 Predicted LULC for the years 2019, 2035 and 2050

Figure 4.14 shows the spatio-temporal prediction of the LULC maps for 2019, 2035, and 2050. The map depicts the rise in the built-up area along the streams from 2019 till 2035; For example, the common specific images show the predicted change in the built-up region. In 2050, only the density of built-up rises rather than moving spatially. This is evident from Table 4.9, wherein the built-up area changes from 555.16 km^2 in 2019 to 633.64 km^2 in 2035 and 684.01 km^2 in 2050. Thus, the spatio-temporal trend

indicates that by 2050, almost 684.01 km^2 of land will be converted into a built-up area, with a density significantly higher than the current LULC.

To simulate the transitions, all driving variables were included in this study. Table 4.7 shows that Cramer's V for distance to road and population density is above 0.6, showing the close reliability of two layers with the transition potential maps. Population density has a direct relationship with the LULC change. Also, regions where the rainfall is moderate in magnitude, will tend to attract the population to settle down in those regions. Distance from Public and Industrial areas showed a significant improvement in the model as it is one of the most accessible regions by the people in their daily life, such as schools, colleges, and religious places. The elevation and slope are the prime topographic factors affecting LULC change. Topography affects the spread and extent of urban distribution, forest, and barren land conversion to agricultural land. It is also found that deforestation decreases with the increase in the slope. Other driving factors like distance from the stream, distance from built-up, and distance from roads also play a role in land use change since they make it easier for inhabitants to access basic commodities.

Further, these drivers are not only limited to local-specific issues; rather, they are regional, national, and global. The anticipated LULC transition has been impacted by more than just the trend of anthropogenic interventions on the land. Natural calamities like the 2018 and 2019 floods, which increased the area of barren land, might have resulted in variations of LULC. The current pattern of LULC transition is primarily due to development practices, and it may lead to land deterioration, pollution, a decline in groundwater quality, a perturbed coastal environment, and so on. These predicted maps could be seen as a prospectus for stakeholders to understand better the impact of land use patterns on land cover. It can help them demonstrate an interest in environmental planning and decision-making for potential land use management and sustainable land cover utilization.

4.4 CLOSURE

GEE is used to classify LULC using three ML techniques: RF, SVM, and CART. The ML application in the CA-Markov model of Idrisi TerrSet software was used to analyse the spatio-temporal change prediction of the LULC using the LCM module for the years 2035 and 2050. From the above analyses, it is inferred that the change process in the

VLS and the potential future change in LULC is detrimental to the environment in the region. When comparing the three techniques, SVM performed poorly at an average accuracy of around 82.5%, CART being the next at 87.5%, and the RF model being good at an average of 89.5%. The RF outperformed the SVM and CART in almost identical spectral classes, such as barren land and built-up areas. As a result, RF-classified LULC is considered to predict the Spatio-temporal distribution of LULC transition analysis for 2035 and 2050. The 2019 prediction is carried out to analyse the model's performance. Except for barren land, this model had more than 94% accuracy for all LULC classes. The lower prediction accuracy of barren land may be influenced by an increase in anthropogenic activities and natural disasters.

From 2013 to 2017, the waterbody area decreased gradually, but after the floods, the same was found to increase considerably in the study. Additionally, it has been noted that between 2017 and 2019, the vegetation decreased by 8%. From 2009 to 2019, there was an increase in the built-up area. Since cities of Kerala state have the greatest rate of urbanization, this is also the key motivation for the research. The urban area has increased by about 97% in ten years. The floods that occurred in the state during 2018 and 2019, which triggered landslides and the combined impact of human activities, maybe to account for the rapid growth in barren land in 2019. Between 2009 and 2019, the combined effects of landslides and human activities were found to be responsible for a 125% rise in bare land, which caused a reduction in both plant cover and forest cover.

The driver variables play an important role in predicting future changes in the LULC. From the obtained Cramer's V values, it is found that population density, distance from the road, DEM, and distance from the stream greatly influence the LULC pattern change. The other parameters, such as distance from Public and Industrial areas, built-up, and forest and slope, are moderately associated with the change in land use patterns. From the prediction, it is observed that up to 2035, the built-up area along the stream is predicted to grow; by 2050, however, the density of built-up areas was predicted to increase. Thus, according to the spatio-temporal pattern, over 685 km^2 of land will be converted to a built-up area by 2050, with much greater built-up density than the present scenario.

CHAPTER 5

FLOOD ANALYSIS AND SPATIAL MAPPING OF FLOOD SUSCEPTIBILITY

5.1 GENERAL

Climate change has added to precipitation uncertainty in many regions of the world and occasional urban floods. Flooding, landslides, erosion, and damage to transportation amenities are also anticipated because of increased precipitation and the resulting stream flows. These urban floods negatively impact the economy, livelihoods, and everyday activities. Flooding is a complicated and location-dependent phenomenon that has captivated the interest of researchers in trying to investigate, analyse, and able to comprehend its causative factors, owing mainly to sizable environmental damage caused by urban expansion, riverside floodplain invasion, habitat destruction, as well as other factors (Hong et al. 2018). The degradation of the environment is the depletion of resources such as air, water, and soil, as well as the destruction of ecosystems, habitats, and natural balance.

Remote Sensing and GIS technology have made it possible to capture, monitor, and assess the calamity. The availability of free data in remote sensing has become a valuable advantage in assessing the temporal and spatial disturbances caused by catastrophes. Optical and microwave sensors play an important role in flood management (Jacinth Jennifer et al. 2020). The constant monitoring facility sets the way for a warning system and plays an important role in catastrophe preparedness efforts. The rapid development and accessibility of SAR data provide the potential for expanded study in various fields. Because the SAR can penetrate clouds, continual flood occurrence surveillance is possible through SAR (Rahman and Thakur 2018). Although satellite-imaging platforms capture permanent surface water extent, microwave sensors utilizing SAR can map flooding in small to medium-sized catchment areas under all-weather circumstances.

Furthermore, multi-date images give researchers an additional tool for monitoring change or recreating the evolution of earlier flood-affected scenarios (Saravanan and Abijith 2022). The present study uses SAR images to evaluate the flood-inundated region during the 2018 and 2019 floods. Using the flood-inundated regions obtained from the SAR images, and the prepared LULC, each flooded class in the LULC is

estimated. Preparing the Flood Susceptibility Map using decision tree-based methods and analysing the potential flood vulnerability zones are attempted in the study.

5.2 MATERIALS AND METHODS

5.2.1 Processing of SAR imagery

The flooded regions are delineated using multi-temporal Sentinel - 1 - Level 1 - Ground Range Detected (GRD) digital data in Interferometric Wide (IW) Swath mode. The Sentinel - 1 mission consists of two polar-orbiting satellites that perform C-band SAR imaging at 5.405 GHz. The level-1 GRD product is a SAR dataset detected, multi-looked, and projected to the ground range using an Earth ellipsoid model. The employed terrain height changes in azimuth but remains constant in range. Due to the multi-look processing, the resultant output has nearly square spatial resolution and square pixel spacing, with decreased speckle. The pixel values represent the observed magnitude, but the phase information is lost. Sentinel - 1 mission pioneer applications include maritime surveillance, land monitoring, and emergency management (ESA 2000). GEE makes it easier to specify various modes of operation for integrating input data and efficiently creating composite data, cloud-free, and multi-temporal datasets. To determine flood inundation extents, the Flood Mapping Algorithm (FMA) in GEE is utilized, which is widely used by various researchers to determine the flooded region (Saravanan and Abijith 2022; Tiwari et al. 2020; Zhang et al. 2020). It helps reduce time consumption and high processing load as the processing is carried out completely in the GEE cloud infrastructure. The image change detection approach has been utilized to delineate flooded zones in multi-temporal SAR data. Typically, change detection is accomplished by comparing pre or post-disaster reference data with flood images (Manavalan 2017).

The images in Sentinel - 1 suffer from erroneous noise (i.e., rigorous geometric, radiometric corrections, thermal, and speckle). Any satellite data requires rigorous pre-processing before being utilized for any application. Pre-processing comprises (a) orbital file correction to remove orbital noise; and (b) thermal noise correction to remove noise in data generated by sensors during the data-collecting process aboard the satellite. Thermal noise can degrade data quality in areas with low mean signal response detected by the SAR system, such as lakes, standing water, rivers, and so on; (c) radiometric calibration, which is used to calibrate RADAR reflectivity Digital Number

(DN) to backscattering coefficient (physical units) and is primarily used to compare SAR images acquired at different times; and (d) Terrain correction, which involves converting Sentinel - 1 Single Look Complex (SLC) data from slant range geometry to a map coordinate system and correcting distortions such as foreshortening, layover, and shadowing effects. After pre-processing Sentinel - 1 VV and VH SAR images, the speckle filter of smoothing radius 50 is applied to smoothen the acquired images. The filtered image visualized differentiates the water and land before and after the occurrence of flood in 2018 (Figure 5.1 and Figure 5.2), 2019 (Figure 5.3 and Figure 5.4), and 2021 (Figure 5.5 and Figure 5.6).
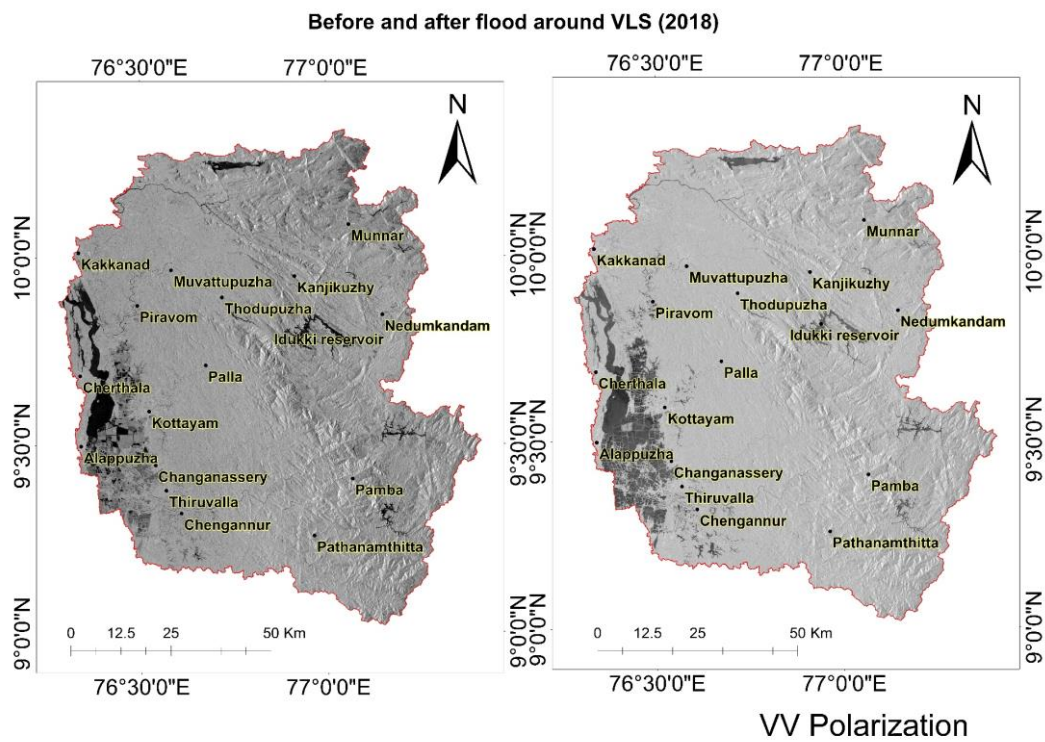
Figure 5.1 Backscattering of VV Polarization before and after 2018 flood

Before and after flood around VLS (2018)
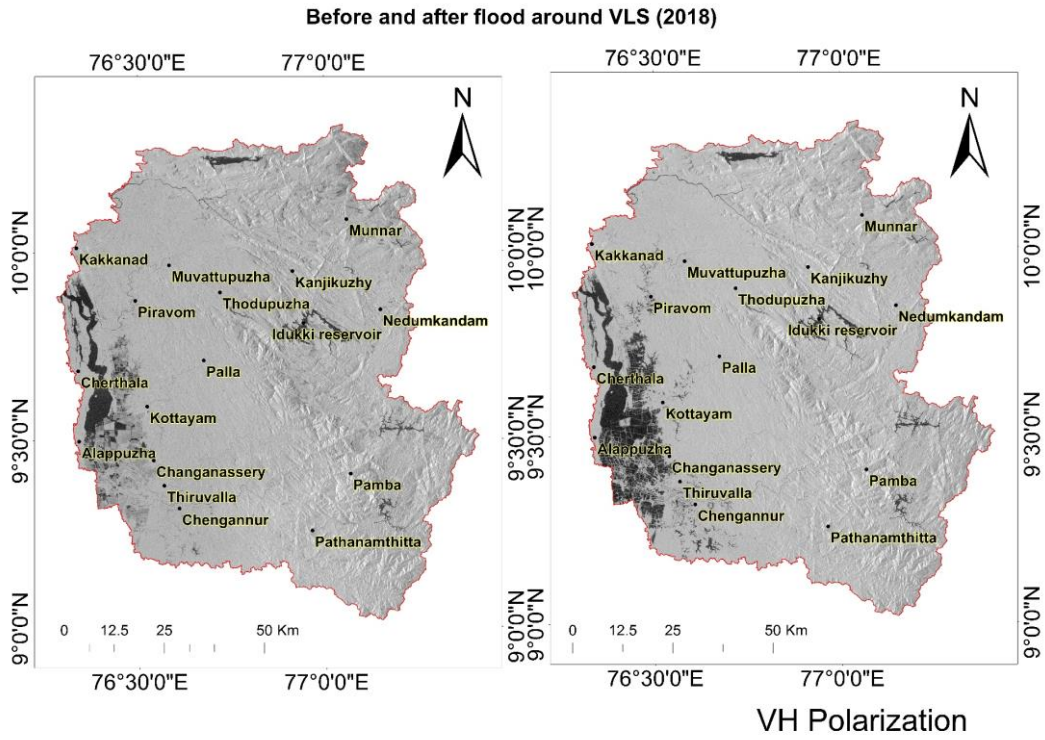
VH Polarization

Figure 5.2 Backscattering of VH Polarization before and after 2018 flood

Before and after flood around VLS (2019)
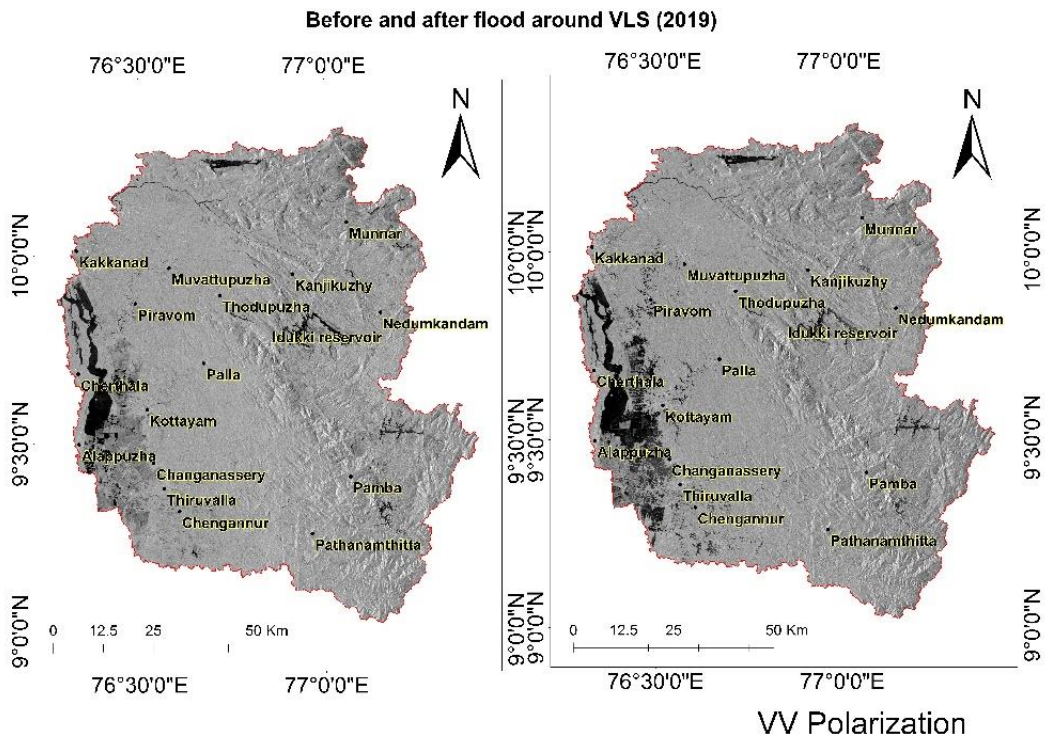
VV Polarization

Figure 5.3 Backscattering of VV Polarization before and after 2019 flood
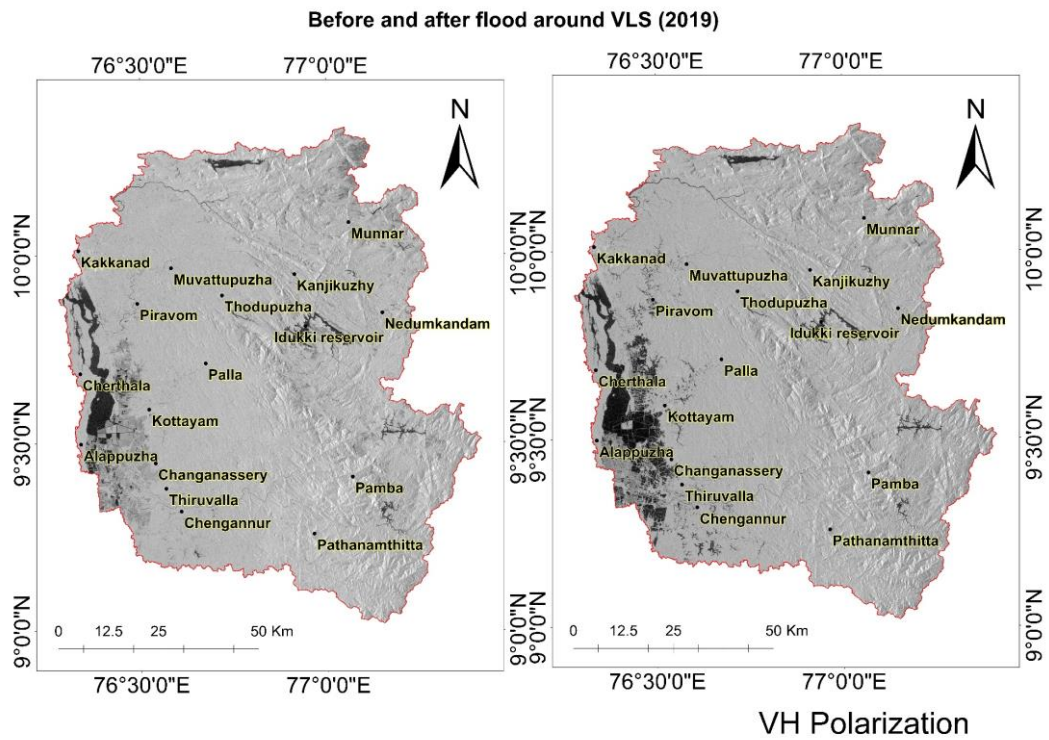
VH Polarization

Figure 5.4 Backscattering of VH Polarization before and after 2019 flood

Before and after flood around VLS (2021)
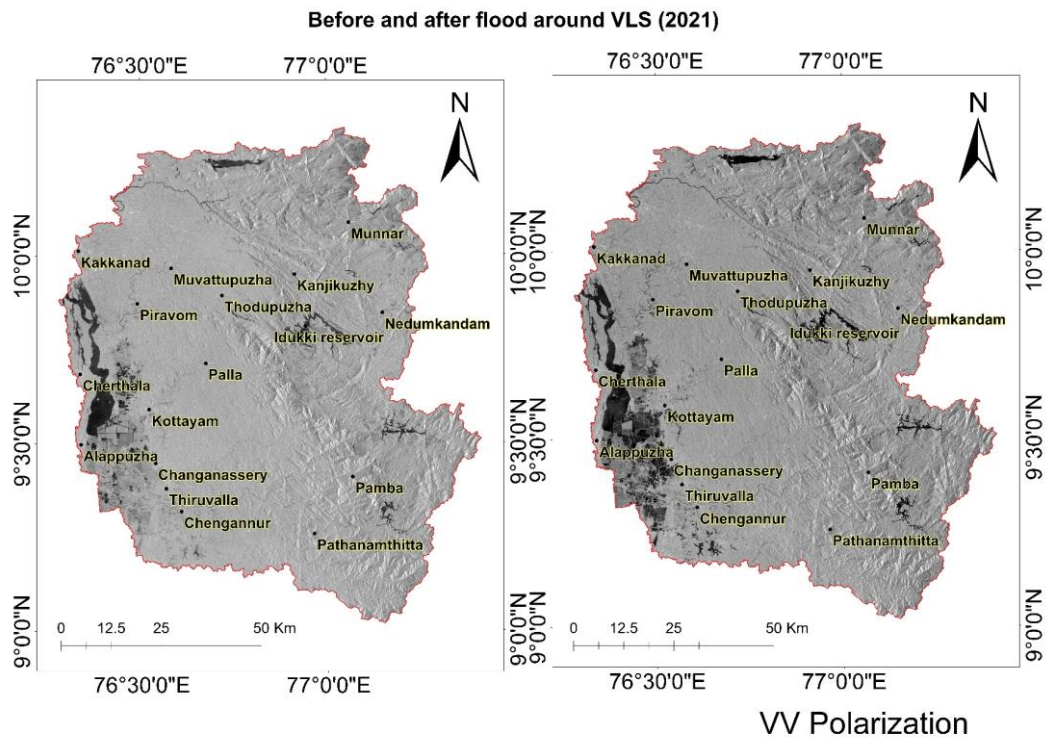
VV Polarization

Figure 5.5 Backscattering of VV Polarization before and after 2021 flood
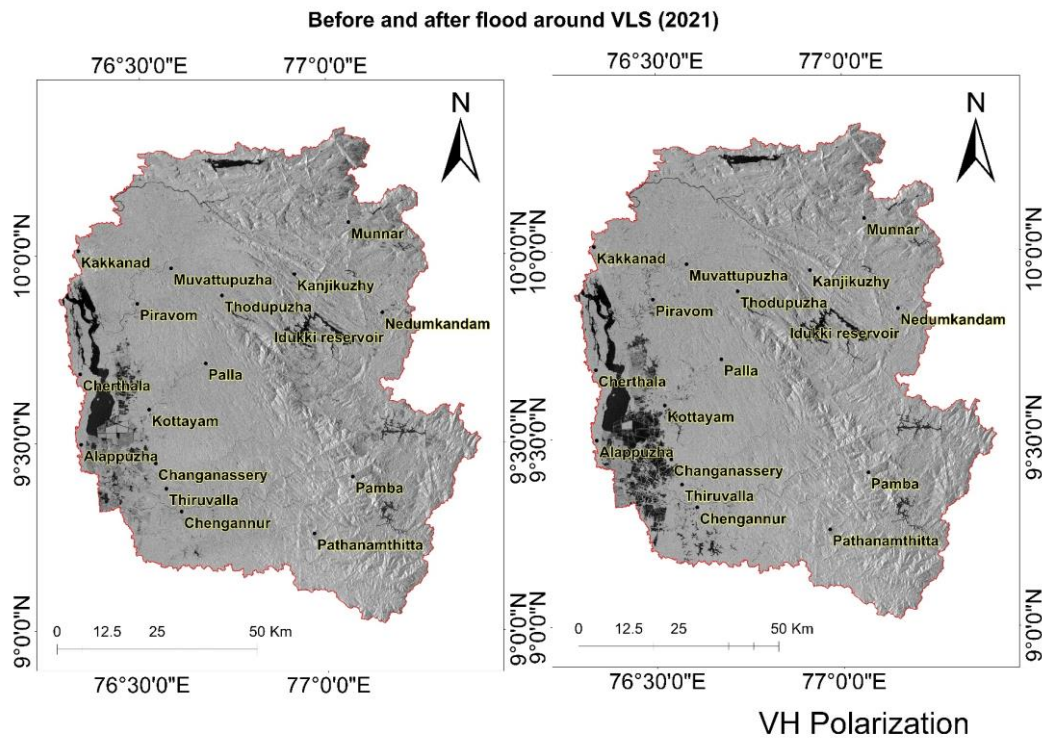
Figure 5.6 Backscattering of VH Polarization before and after 2021 flood

The suitable polarization is utilized for the study, and the image ratio technique is carried out to delineate the flood inundation region. The 2018 flood-damaged regions are validated through the field study. The 2019 and 2021 flood maps are validated with the flood hazard maps published by Kerala State Disaster Management Authority (KSDMA 2022). Therefore, the accuracy of the output flood-inundated regions obtained from the SAR imagery is validated.

5.2.2 Flood inundation in future LULC

The study region is frequently prone to flooding, with severe flood events as discussed in sections 3.4, 3.5, and 3.6. Apart from these floods, the state of Kerala experienced floods in 2020, but the severity of the floods and inundation had no significant impact on the socioeconomic system of the state. According to previous flood analysis, the extent of inundation and severity was greater in 2018. Hence, the 2018 flood scenario is used in the analysis. This section attempts to assess the amount of flood inundation that occurred during the 2018 LULC. Furthermore, the possible flood inundation region for the same severity as the 2018 Kerala floods using future predicted maps from 2035 and 2050 is analysed. This is accomplished in ArcGIS software, where the amount of

each class in the LULC that will flood during the severe flood scenarios is calculated. The four classifications in the LULC, i.e., built-up, vegetation, barren land, and forest, are analysed to quantify the flood inundation area. The area inundated during the flood in each class is determined by overlaying each class of LULC with the flooded area map. The same approach is used to evaluate the extent of land inundated in future LULCs, such as 2035 and 2050, if a flood of the same severity occurs again.

5.2.3 Flood Susceptibility Modelling (FSM)

Five essential phases are included in FSM such as (1) Flood inventory data, as well as non-flood location data; (2) Data gathering and creation of a spatial database for the flood conditioning factors; (3) Selection of sensitive parameters for the flood by using RFE technique; (4) Establishing the relationship between the flood and the conditioning parameters to model the flood susceptibility; (5) Validation of the flood susceptibility map.

5.2.3.1 Flood inventory data

Flood inventory mapping (Figure 5.7) is prepared by delineating the frequently flooded regions from the Sentinel - 1 SAR imagery during 2018, 2019, and 2021 floods, as explained in section 5.2.1. The most frequently flooded areas determined are fed into the ArcGIS environment to create inputs for the ML models. Thus, the 1500 flood points are distributed over the flooded region by establishing spatially balanced points throughout the study. 70% (1050 flooded points) of the data are fed into the model for training, whereas 30% (450 flooded points) are fed into the model for evaluation. In addition, 1500 non-flood points are chosen randomly from the areas that are not prone to flood and distributed over the study before being entered into the model in the same 7:3 ratio.
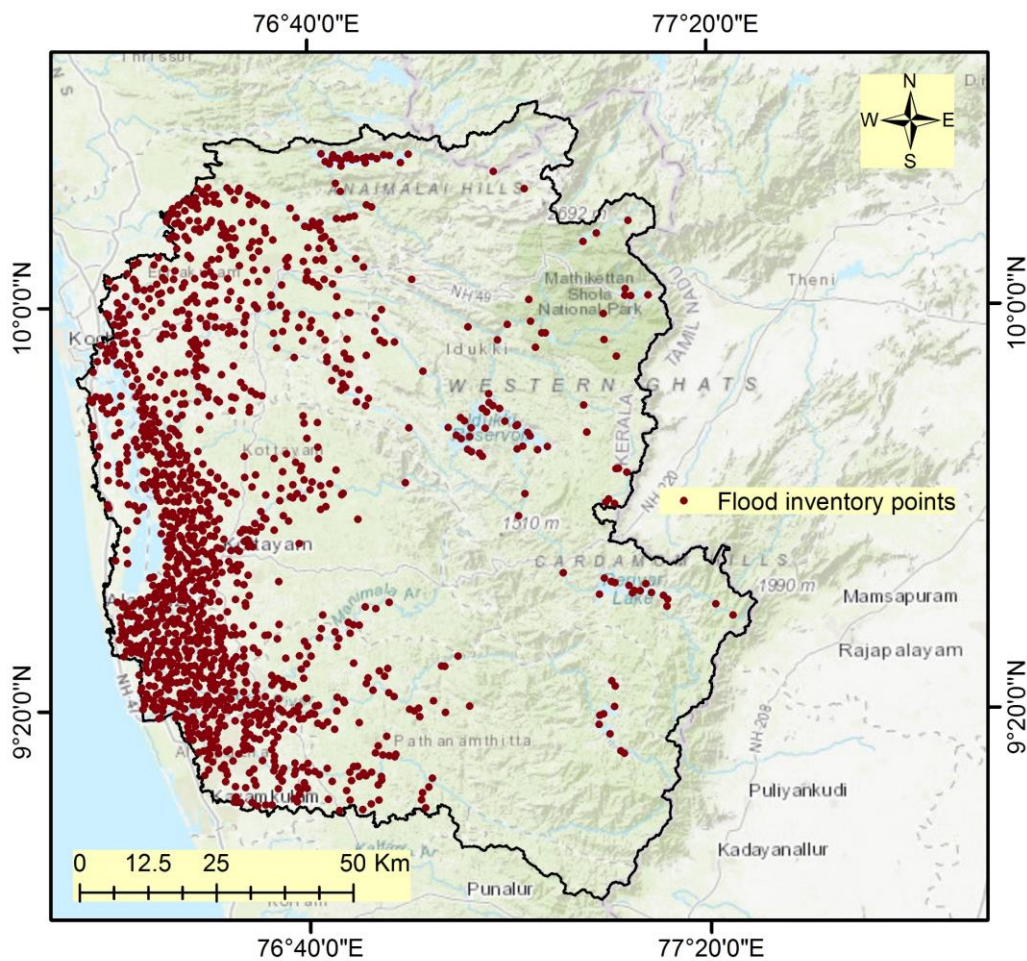
Figure 5.7 Flood Inventory map

5.2.3.2 Conditioning factors for flood

The selection of flood conditioning factors directly influences the accuracy of the mathematical model. The predisposing parameter selection is preliminarily based on the components of the flood, the scale of the analysis, the morphology of the study (Jacinth Jennifer and Saravanan 2022; Shahabi et al. 2021), and the knowledge obtained from the literature. Hence, eighteen parameters are considered for the susceptibility analysis, and the spatial database of these factors is compiled. Those factors are elevation, slope, plan curvature (plan_c), profile curvature (prof_c), aspect, LULC, Stream Power Index (SPI), Topographic Ruggedness Index (TRI), Topographic Wetness Index (TWI), Sediment Transport Index (STI), geology, geomorphology, NDVI, rainfall, soil, distance to road, distance to stream, and wind speed. The basic parameters such as slope, profile curvature, plan curvature, aspect, SPI, TRI, TWI, and

94

STI are the derivatives of the DEM collected from ASTER with a resolution of 30 m, which itself represents the variation in elevation. As an amplifying factor, topography greatly impacts flood intensity and determines flood-prone regions. Topographic parameters, on the one hand, directly impact flow size and runoff velocity, whereas river flood-prone locations, on the other hand, often have low relief and a modest topographic slope. The elevation represents altitude with respect to the Mean Sea Level (MSL) and ranges from 0 to 2700 m in the study region (Figure 5.8a). The elevation has an opposite relationship with flooding; as the elevation increases, the accumulation of the flood consecutively decreases. The flood accumulation is higher in the lower elevation as low-lying areas are susceptible to flood. Slope can be termed as the angle between the ground and a horizontal datum. It implies that gravity has a role in causing accelerated runoff (Yang 2017). Floods are more likely to accumulate on gentle slopes, as steeper slopes yield faster disposal of floods. The slope in the study region varies from 0° to 75° and is categorized into five classes (Figure 5.8b).

Plan curvature is perpendicular to the maximum slope direction, also known as contour curvature. It plays a major role in converging or diverging flow, soil water content, and the characteristics of soil (Wilson and Gallant 2000). The sideward convex curvature gives the positive value, the sideward concave curvature gives the negative value, and the linear surface gives the zero. In the study, the value of plan curvature ranges from -27.4 to 22.1 (Figure 5.8c). Profile curvature is parallel to the maximum slope direction, i.e., curving downwards to the direction of the slope (Taşoğlu and Abujayyab 2022). It has a significant role in accelerating the flood and soil erosion rate. Upwardly concave curvature is positive, upwardly convex curvature is negative, and linear curvature is zero. The value of profile curvature ranges from - 23 to 21 (Figure 5.8d). Another important parameter is the aspect, as rainfall is more likely on windward slopes. The intensity of solar radiation, which influences surface vegetation and soil moisture, is also related to the aspect. It shows which way the actual slopes are facing. It is classified into nine classes, as shown in Figure 5.8e. The SPI measures discharge erosive power concerning a particular region within a catchment (Shahabi et al. 2021). It demonstrates the flood's abrasive strength. A high SPI number indicates rapid downstream water flow, which indicates a lesser risk of flooding, whereas a low value indicates sluggish or stationary water, which indicates greater inundation potential (Chowdhuri et al.

2020; Tehrany and Kumar 2018). SPI values range from - 3.51 to 17.9 (Figure 5.8f) calculated from the equation (5.1).

$$SPI = A_s \tan \beta \qquad (5.1)$$

Where A_s implies the specific catchment area and β is the slope gradient.

STI shows the intensity of sediment movement generated by water and depicts the effect of topography on erosion. It is a widely used factor in flood susceptibility analysis as floods may become more prevalent when sediment transport becomes more dynamic (Fang et al. 2021). Its value ranges from 0 to 1610 in the study (Figure 5.9a). The TWI is a set of topographic parameters that describe the geographical distribution of wetness, aquifer depth, soil humidity, saturation zone, and topographic flow accumulation and is employed as an important flood component. TWI reflects the moisture content of the soil. Hence a greater number implies saturated terrain, which can lead to floods. Beven and Kirkby (1979) first developed the TWI, assuming steady-state conditions and uniform soil properties. TWI values vary from 2.1 to 26, indicating that the study has a very high TWI value (Figure 5.9b). Thus the STI and TWI values are calculated using the equation (5.2) and (5.3) respectively.

$$STI = \left(\frac{A_s}{22.13} \right)^{0.6} \left(\frac{\sin \beta}{0.0896} \right)^{1.3} \qquad (5.2)$$

$$TWI = \ln \left(\frac{\alpha}{\tan \beta} \right) \qquad (5.3)$$

Where α is the upslope area per unit contour length.

At the watershed scale, TRI impacts stream energy, surface storage capacity, runoff velocity, and routing. It represents the difference in elevation between adjoining cells. It gives an objective quantitative measure of topographic variation, and the value ranges from 0.0001 to 0.97 in the study (Figure 5.9c). As the study discusses rainfall-induced floods, rainfall intensity is an operational parameter. Flooding can be caused by short-duration heavy rain or long-duration, lower-intensity rain (Shahabi et al. 2021; Towfiqul Islam et al. 2021). The ground surface gets saturated in both cases resulting in the impermeable surface causing severe flooding. The rainfall is collected from the CHIRPS using the GEE platform. The amount of rainfall ranges from 1670 to 4700 mm/yr in the study, as shown in Figure 5.9d.

LULC is one of the critical parameters where the influence of human intervention or anthropogenic activities leads to the change in the land cover and, consequently, alters the infiltration rate in the area. This is very common in the built-up region resulting in low infiltration and generation of surface flow. The forest and the vegetation region have an inverse relationship with flooding due to the high infiltration and less flooding (Kia et al. 2012). The LULC map is prepared from the Landsat-8 for 2020 in GEE using the RF algorithm and classified into five basic classes: waterbody, built-up, vegetation, forest, and barren land (Figure 5.9e). The NDVI is a measure for determining the greenness of a landmass and the presence of water bodies. NDVI variations over time reflect changes in vegetation and surface water cover and can reveal the link between floods and vegetation within a watershed. Higher vegetation concentrations are thought to reduce the likelihood of floods in the study region. NDVI is derived from the Landsat - 8 image by normalizing the spectral reflectance of the near - infrared and red bands using the equation (5.4). The value of NDVI ranges from - 1 with the least density of vegetation to +1 with a high density of vegetation (Figure 5.9f).

$$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}} \tag{5.4}$$

Distance to the stream determines the effect of flood inundation, as the regions near the stream are severely more vulnerable to flooding than those away from the stream. The frequency and magnitude of the flood are also very severe near the streams. Hence, the Euclidean distance is applied to the streams derived from the DEM and is classified into five classes such as <100 m, 100-200 m, 200-300 m, 300-400 m, and >400 m (Figure 5.10a). The development of road networks results in an increase in impermeable surface, which results in less groundwater recharge and more flow. This causes a considerable change in topography, resulting in excessive runoff and flow accumulation in low - lying areas adjacent to the roads. The road network is downloaded from the OpenStreetMap tool, and the Euclidean distance is applied to categorize the distance to the roads into five classes, <100 m, 100-200 m, 200-300 m, 300-400 m, and >400 m (Figure 5.10b). Due to geological variance, the flood frequency of a region has considerable power to modify the stream profile. The rock's permeability and infiltration rate has a strong relationship (Chen et al. 2016). As a result, impermeable rocks induce surface runoff to the maximum extent, which can cause

severe flooding (Das 2019). The major geology in the study comprises charnockite, khondalite groups, and peninsular gneiss (Figure 5.10c). Geomorphology is a major contributor to flood hazards; it is probably more prevalent than hydrology, although it is less crucial (Slater et al. 2015). The combination of tectonics, denudational, and fluvial geomorphic processes has sculpted the geomorphology of Kerala's Western Ghats (Ramasamy et al. 2020). The geomorphology of the study area comprises mainly highly and moderately dissected valleys, pediment pediplain, coastal, deltaic, and flood plains (Figure 5.10d). The geology and geomorphology data are collected from the Geological Survey of India (GSI).

Flood susceptibility is determined by topsoil, which plays a major role in floods. Soil texture is one of the most important physical qualities influencing water infiltration, runoff, and inundation. Soil types determine water storage, drainage, and permeability, which causes water inundation (Tehrany and Kumar 2018). The clay, gravelly clay, and loamy soil dominate most of the study area (Figure 5.10e). Sandy soil is present near the beaches, i.e., along the western part of the study area.

Wind speed indirectly affects flooding by influencing the rainstorm kinetic energy for a short-span high-intensity precipitation event. The percentage of raindrop energy produced from wind speed accounted for around one-fourth of overall kinetic rain energy on average. Wind speed immediately affects rain energy-related soil processes such as surface sealing, runoff, and erosion (Helming 2001). The wind speed data is collected from NASA's POWER data access viewer. The wind speed in the study area varies from 4.69 to 6.92 m/s (Figure 5.10f).
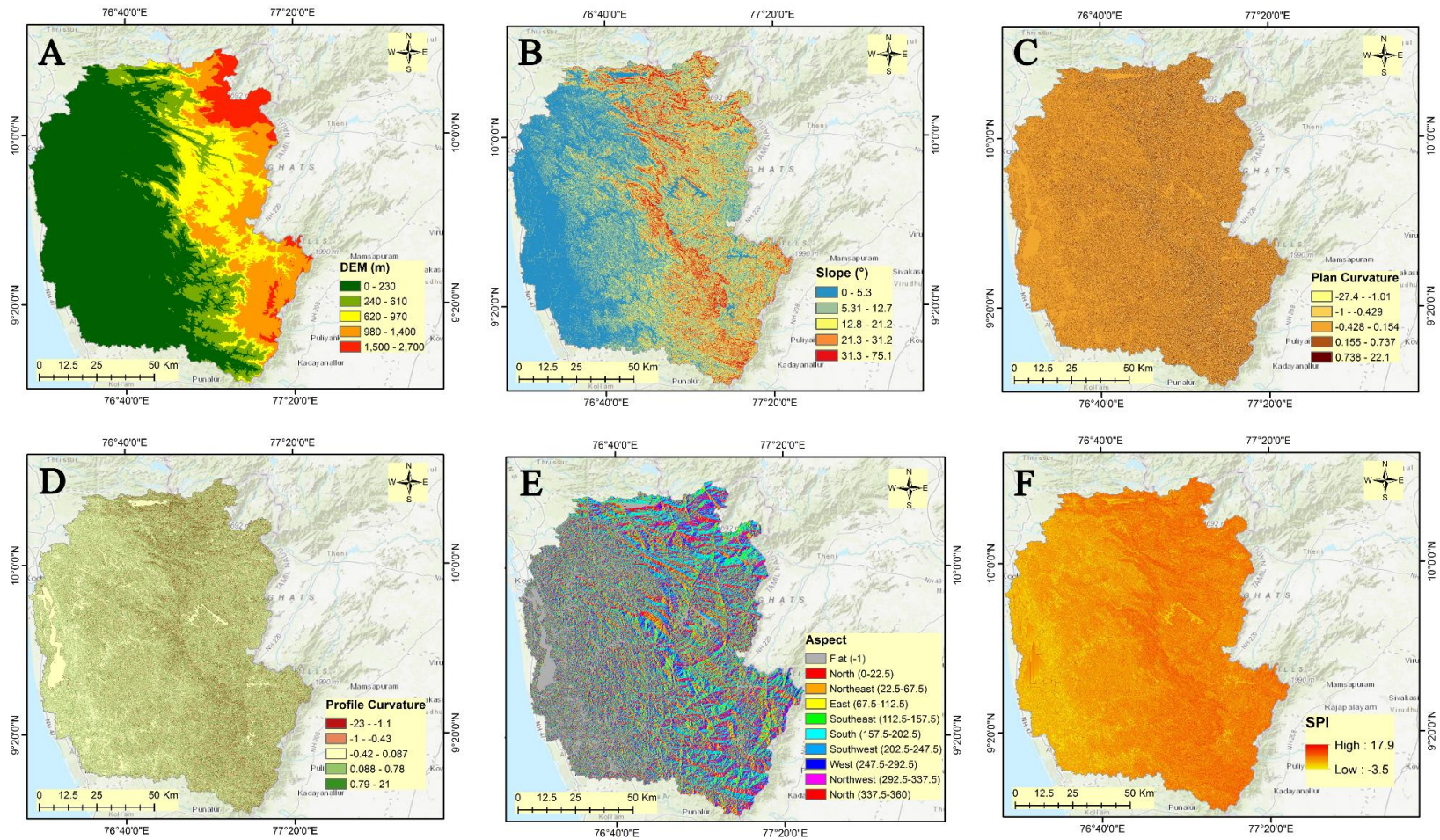
Figure 5.8 (a) DEM (b) Slope (c) Plan Curvature (d) Profile Curvature (e) Aspect (f) SPI
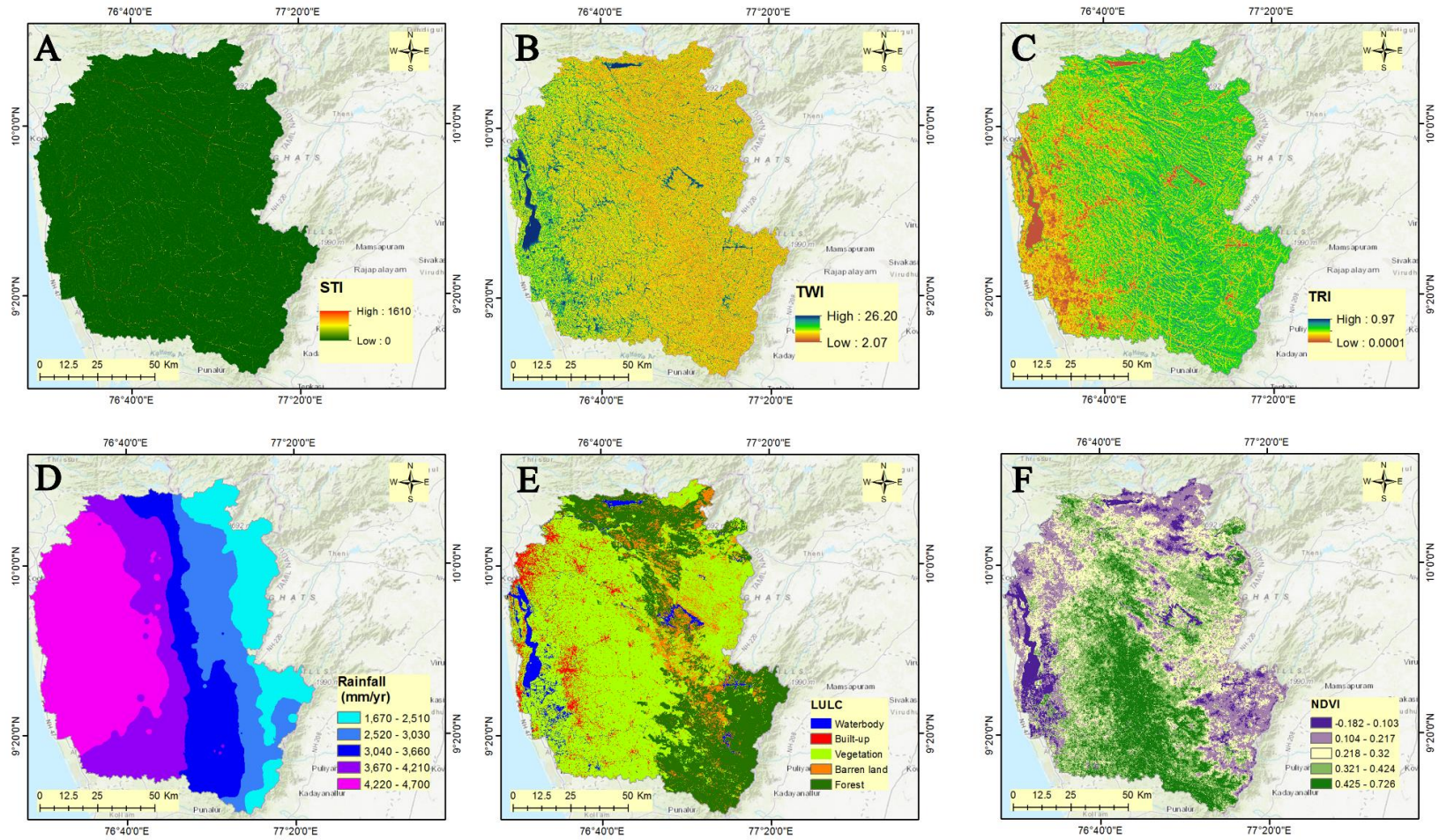
Figure 5.9 (a) STI (b) TWI (c) TRI (d) Rainfall (e) LULC (f) NDVI
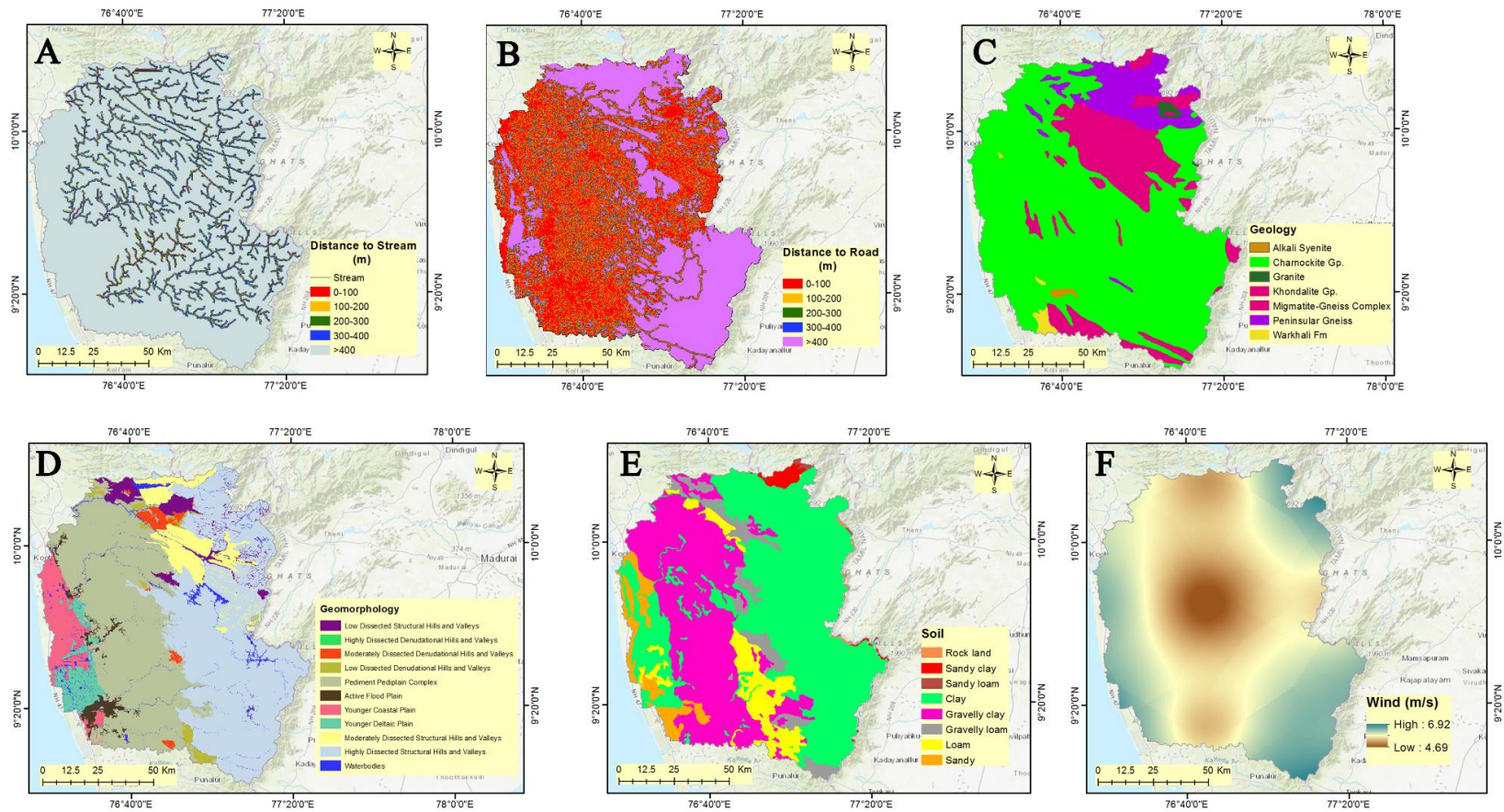
Figure 5.10 (a) Distance to stream (b) Distance to road (c) Geology (d) Geomorphology (e) Soil (f) Wind

5.2.3.3 Recursive Feature Elimination (RFE)

Bellman (1957) coined the term "Curse of Dimensionality" for the difficulty produced by the exponential rise in volume associated with adding more dimensions to Euclidean space. In other words, as the number of features grows, the inaccuracy grows as well. It refers to the notion that high-dimensional algorithms are more difficult to build and have a running time exponentially proportional to the dimensions. Theoretically, a larger number of dimensions allows more information to be stored, but in practice, it seldom helps since real-world data contains more noise and redundancy. In addition, not all of the features are equally important; some are even unrelated to the problem. So high-dimensional data processing is a huge challenge for data scientists, and the role of feature selection or feature engineering comes in. Feature selection provides an efficient way to address difficulties such as overfitting, learning accuracy, computational time, and enabling enhanced model learning since ML algorithms are especially prone to the curse of dimensionality. As a result, in ML and data mining, feature reduction is defined as selecting a subset of relevant features from a dataset to utilize in subsequent model formulation (Chen and Jeong 2007; Gholami et al. 2012). RFE is a wrapper method of feature selection. In the feature selection phase, this approach uses a supervised learning algorithm. It uses the subset evaluation approach to rank features. When choosing characteristics, correlation and relationships are considered (Khaire and Dhanalakshmi 2022). It removes redundant and weak features with the least impact on the training error while keeping the independent and strong features to enhance the model's generalization performance. It employs an iterative feature ranking technique, an example of backward feature removal. This method began by creating a model based on the whole collection of characteristics and then ranking each feature according to its relevance. The model is rebuilt, and the feature importance is recalculated after removing the least significant feature (Misra and Singh Yadav 2020). The researcher examines and determines the number of predictor subsets and the size of every subset. As a result, the subset size is an RFE tuning parameter. The predictors are chosen based on the rankings according to the importance, using the subset size that maximizes the performance requirements. The final model is then trained using the optimal subset.

A representative bagging algorithm, RF (an ensemble approach), has been shown to perform well in terms of predicted accuracy (Lee et al. 2017). Hence, RF-based RFE is used for the feature selection process in this study. RF model typically employs backward selection. Variables are seldom excluded from the prediction equation in RF (Degenhardt et al. 2019), as this is the nature of the ensemble models. Increased ensemble performance is linked to the variety of constituent models; averaging models that are practically the same does not reduce model prediction variance (Darst et al. 2018). As a result, utilizing a random sample of predictors, RF coerces the trees to contain sub-optimal splits of the predictors (Svetnik et al. 2003). The RF algorithm uses two approaches to estimate variable importance and calculates criteria weights from the training phase. The Mean Decrease in Accuracy (MDA) indicates how much the model's accuracy degrades as the value of each variable is modified. The Mean Decrease in Gini (MDG) is the average of a variable's total decrease in node impurity in each unique decision tree, weighted by the proportion of data reaching that node (Jeon and Oh 2020). This study uses the MDA approach in which B represents a tree's Out-Of-Bag (OOB) observations, and VI represents the importance of variable X_i in tree t. The MDA approach is calculated using the equation (5.5).

$$W_R(X_i) = \frac{\sum_{t \in B} VI^{(t)}(X_i)}{ntree} \qquad (5.5)$$

The model's inputs are the eighteen flood conditioning factors, and the model's output is the flood inventory map. These eighteen parameters are categorized into seven categorical and eleven numerical data. Each class in the categorical data using one-hot encoding is separated into each layer to convert it into numerical data, where the number of layers is equal to the sum of the total number of classes in all seven categorical data. Seven layers are converted into numerical data as follows. The aspect layer consists of nine classes grouped namely a ((-1°)-0°), b (0°-22.5° & 337.5°-360°), c (22.5°-67.5°), d (67.5°-112.5°), e (112.5°-157.5°), f (157.5°-202.5°), g (202.5°-247.5°), h (247.5°-292.5°), and i (292.5°-337.5°). Geology is classified into seven layers, namely g1 (khondalite), g2 (migmatite-gneiss complex), g3 (peninsular gneiss), g4 (warkhali), g5 (alkali syenite), g6 (granite), and g7 (charnockite). Geomorphology is classified into eleven layers such as ge1 (highly dissected denudational hills and

valleys), ge2 (moderately dissected denudational hills and valleys), ge3 (low dissected denudational hills and valleys), ge4 (pediment pediplain complex), ge5 (flood plain), ge6 (coastal plain), ge7 (deltaic plain), ge8 (moderately dissected structural hills), ge9 (highly dissected structural hills), ge10 (waterbodies), and ge11 (low dissected structural hills). Soil is classified into eight classes such as s1 (clay), s2 (gravelly loam), s3 (loam), s4 (sandy), s5 (sandy loam), s6 (sandy clay), s7 (rock land), and s8 (gravelly clay). LULC is classified into five classes such as l1 (built-up), l2 (vegetation), l3 (barren land), l4 (forest), and l5 (waterbodies). Distance to the road is classified into five classes such as D2R1 (<100 m), D2R2 (100-200 m), D2R3 (200-300 m), D2R4 (300-400 m), and D2R5 (>400 m). Distance to stream is also classified into five classes based on the proximity distance of <100 m, 100-200 m, 200-300 m, 300-400 m, and>400 m as D2S1, D2S2, D2S3, D2S4, and D2S5 respectively. The categorical and non-categorical data are shown in Table 5.1. Thus, the layers add up to 61 conditioning factors, which are given into the model for feature selection using the RF algorithm.

Table 5.1 Classes in categorical and non-categorical data

Data type	Parameters	No. of Classes
	Aspect	9
	Geology	7
	Geomorphology	11
Categorical data	Soil	8
	LULC	5
	Distance to the road	5
	Distance to stream	5
	Elevation	1
	Slope	1
	Plan Curvature	1
Non-categorical data	Profile Curvature	1
	Stream Power Index	1
	Topographic Ruggedness Index	1

Topographic Wetness Index	1
Sediment Transport Index	1
NDVI	1
Annual Rainfall	1
Wind	1
Total input data	61

5.2.3.4 Machine Learning Models

The selected feature obtained from the RFE technique is fed into the ML models for the FSM. Four standalone ML algorithms are used to model the severity of the flood, and the performance of the models is compared. These four models are Adaptive Boosting (AdaBoost), Random Forest (RF), Gradient Boosting Machines (GBM), and eXtreme Gradient Boosting (XGBoost). The employed algorithm is discussed as follows.

AdaBoost

AdaBoost, or Adaptive Boosting, is a classifier for ensemble boosting developed by Freund and Schapire (1997). In AdaBoost, each sample is assigned a weight that influences its likelihood of being included in a base classifier training set (Dou and Chen 2017). AdaBoost is comparable to the forward stage-wise multiplicative function method, which decreases the failure rate. It trains the classification system repeatedly based on the training samples. Each training process focuses on different examples of those samples using adaptive weights. This approach differs from other ensemble methods, such as bagging, which does not adjust the weights. The main goal of AdaBoost is to maximize the distribution and train the data set (t) in each iteration (i) to accurately assess the maximum level of unexpected results. AdaBoost initially selects a small portion of the training sample $D = (x_i, y_i)$ where the individual x_i instance is the vector of character values connected to the area *X,* and each y_i category mark is connected to the x_i belonging to the *Y*. Then, consecutively trains the AdaBoost predictor by selecting a training set that belongs to a specific prediction of the final training set. It assigns a higher weight (w) to incorrectly classified data to ensure these results to classify correctly in the next iteration. It also allocates weight to the trained

classifier based on its accuracy in each iteration. In addition, a more accurate classifier will be optimized for maximum weights. This procedure iterates until the entire training data set matches without error or until the maximum number of estimators specified is attained. Furthermore, at each stage of classifier training, the algorithm will loop through all possible features and assess the error of each feature on each instance. As the initial weak classifier, the best feature is selected. The task of the weak learner is to identify a weak hypothesis $(h_t)\, ht : X \rightarrow \{-1,+1\}$. That is suitable for distribution d_t. The aim is to set h_t to minimize error ϵ_t.

$$\epsilon_t = Pr_i \sim_{d_t} (h_t(x_t) \neq y_i) \qquad (5.6)$$

$$\alpha_t = \frac{1}{2}\ln(\frac{1-\epsilon_t}{\epsilon_t}) \qquad (5.7)$$

The final equation for the AdaBoost classifier is shown by modifying the D distribution and highlighting the incorrectly classified areas.

$$H(x) = Sign\left(\sum_{i=1}^{T} \alpha_t h_t(x)\right) \qquad (5.8)$$

Where h_t is a weak beginner, the coefficient is α_t, and the final hypothesis is the outcome $H(x)$.

Random Forest

RF classifier is explained in detail in section 4.2.1.1

Gradient Boosting Machines (GBM)

Friedman (2000) introduced GBM, which can resolve classification and regression problems. GBM is based on the concept of combining multiple weak learners to increase efficiency (Janizadeh et al. 2021). The main heuristic underlying the strategy is to build new base learners that are most correlated with the negative gradient of the ensemble's loss function. The important variables in the GBM tree model are the number of trees and the learning rate. GBM is based on three elements: a loss function that will probably be optimized, a predicted weak learner, and a loss function optimizer that acts as an additive function to add weak learners. It also includes three critical tuning parameters: ntree, tree depth, and shrinkage (Kuhn and Johnson 2013; Sahin 2022; Shahzad et al. 2022; Wang et al. 2020). It is difficult to obtain an estimate using

the loss function $\Psi(y, f)$ and weak learner $h(x, \theta)$. Therefore, a new function $h(x, \theta_t)$ was proposed with a negative gradient $\{g_t(x_i)\}_{i=1}^{N}$ to use alongside the observed data:

$$g_t(x) = \mathrm{E}_y \left[\frac{\partial \psi(y, f(x))}{\partial f(x)} \mid x \right]_{f(x) = f^{t-1}(x)} \qquad (5.9)$$

Where $E_y \left(\Psi \left[y, f(x) \right] \right)$ is the expected loss function over the response variable, and $f(x)$ is true functional dependence. Thus, the new function is highly correlated with $-g_t(x)$. This permits the least square minimization of the model by using the algorithm (5.10) (Natekin and Knoll 2013).

$$(\rho_t, \theta_t) = \underset{\rho, \theta}{\arg\min} \sum_{i=1}^{N} \left[-g_t(x_i) + \rho h(x_i, \theta) \right]^2 \qquad (5.10)$$

Where ρ is the step size at each iteration.

Extreme Gradient Boosting (XGBoost)

Chen and Guestrin (2016) proposed the XGBoost algorithm, which is currently one of the most popular and effective decision tree algorithms. It is a highly versatile and flexible tool capable of solving most regressions, classifications, and user-defined decision variables (Ma et al. 2021). The algorithm employs the CART as the base classifier rather than grouping independent trees. It generates a series of subsequent decision trees based on the prediction errors or residuals of the previous tree model (Krishnaraj and Honnasiddaiah 2022). Thus, the algorithm prioritizes higher uncertainty samples and produces the final result (Abedi et al. 2022; El-Magd et al. 2021; Taghizadeh-Mehrjardi et al. 2020). The XGBoost parameters, such as nrounds, help to calculate the maximum number of iterations; max depth, which controls the depth of the tree; eta controls the learning rate to capture patterns in data; gamma controls regularisation to prevent overfitting; and colsample bytree, which controls the number of variables provided for a tree. In the classification task, the tree splitting is stopped when the leaf node has a minimum sum of instance weight that is less than min child weight; the final tuning parameter subsample controls the number of observations provided for a tree (Boehmke and Greenwell 2019). The objective of the XGBoost algorithm is to minimize the regularized objective function, which is expressed as follows.

$$L(\Phi) = \sum_i l(y_i, y_i) + \sum_k \Omega(f_k) \qquad (5.11)$$

The initial component of Eq. (5.11) represents a loss function that quantifies the dissimilarity between the predicted class y_i and the target class y_i. The subsequent component, as delineated in the equation (5.12), serves as a penalty term that regulates the model's complexity to prevent overfitting.

$$\Omega(f) = \gamma T = \frac{1}{2} \|\lambda w\|^2 \qquad (5.12)$$

The formula is expressed as follows: T represents the total number of leaves in the tree, w denotes the score assigned to each leaf, and γ and λ represent the degrees of regularization. The method employs an iterative strategy to minimize the objective function, as specified in the equation (5.13), at each step t.

$$L^{(t)} = l\left(yi, y_i^{(t-1)} + f(x_i)\right) + \Omega(f_t) \qquad (5.13)$$

5.2.3.5 Model Validation

Validation is perhaps the most important phase in any prediction modelling since it determines the model's scientific significance. Validation is done by comparing the prediction findings to an unknown target, the region that the future flood will impact. As a result, a portion of a previous flood is employed to represent the goal pattern. The data is divided into two subsets, generating the prediction image and validating the prediction outcomes (Meliho et al. 2022). The training and testing samples are used to build and validate the results. The prediction rule's performance was evaluated using the ROC-AUC curve, sensitivity, specificity, F1 score, and accuracy. The metrics are evaluated using the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) scores. The sensitivity, specificity, and accuracy metrics are calculated using the equation (5.14), (5.15), and (5.16) respectively (Shahabi et al. 2021).

$$Sensitivity \ (True \ Positive \ Rate) = \frac{TP}{TP + FN} \qquad (5.14)$$

$$Specificity \ (True \ Negative \ Rate) = \frac{TN}{TN + FP} \qquad (5.15)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (5.16)$$

The ROC curve is a graphical tool for evaluating the performance of a model. The y-axis represents the true positive rate (sensitivity), while the false positive rate (1-specificity) is represented by the x-axis (Hanley 1989). For each point on the ROC curve, a single decision criterion may be retrieved to forecast the model's accuracy. AUC is a quantitative metric for evaluating model performance; the greater the value of AUC (i.e., closer to 1) better the model's performance. The overall methodology flowchart framed for the study is shown in Figure 5.11.
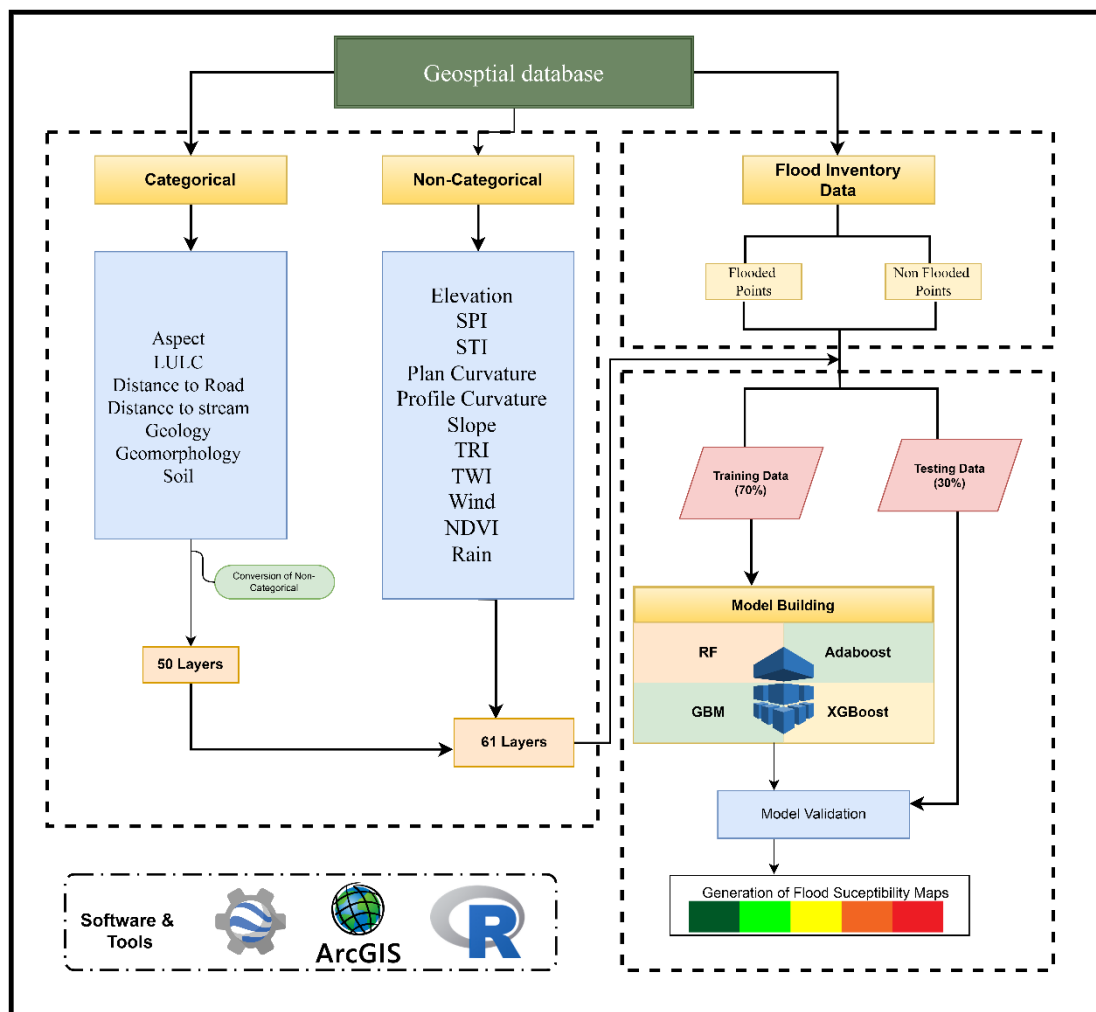
Figure 5.11 Methodology flowchart for FSM

5.3   RESULTS AND DISCUSSION

5.3.1   Delineation of flooded regions

A correct polarization choice optimizes the discrimination of flooded areas (Klemas 2015). The results obtained in our polarization configurations and the contributions of Conde and De Mata Muñoz (2019); Martinis and Rieke (2015); Matgen et al. (2007) in

the comparative studies of polarizations to monitor flood areas reaffirm that the VH polarization is more suitable for delineating flooded areas. It generates well-defined and correctly defined surfaces, results that VV polarization cannot offer. The roughness and heterogeneity of the terrain influence VV polarization. Hence, VH polarization is further used to delineate the flooded region. The image ratio technique in VH polarization is carried out to delineate the flooded region (Figure 5.12a, 5.13a and 5.14a) from the non-flooded region. In the figures, the white patches show the flood-inundated regions, and the remaining portion shows the general terrain of the region.
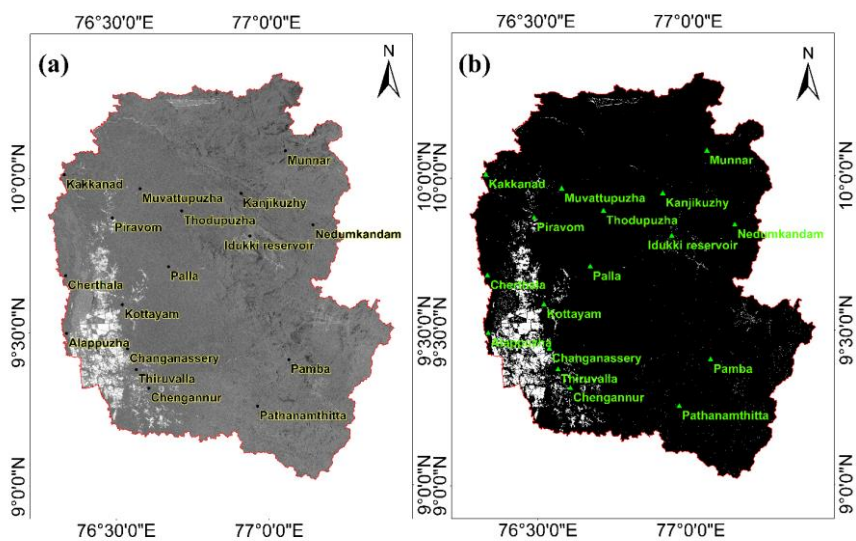
Figure 5.12 Flooded region for 2018 flood
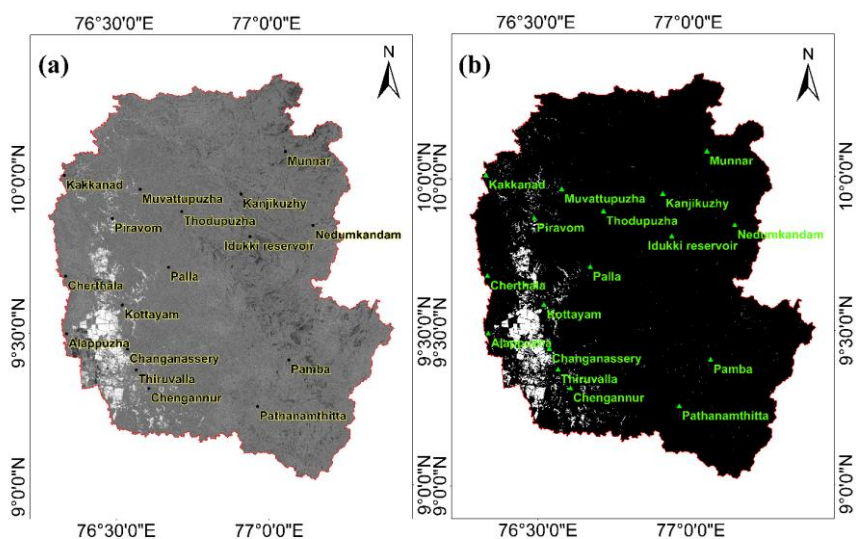
Figure 5.13 Flooded region for 2019 flood
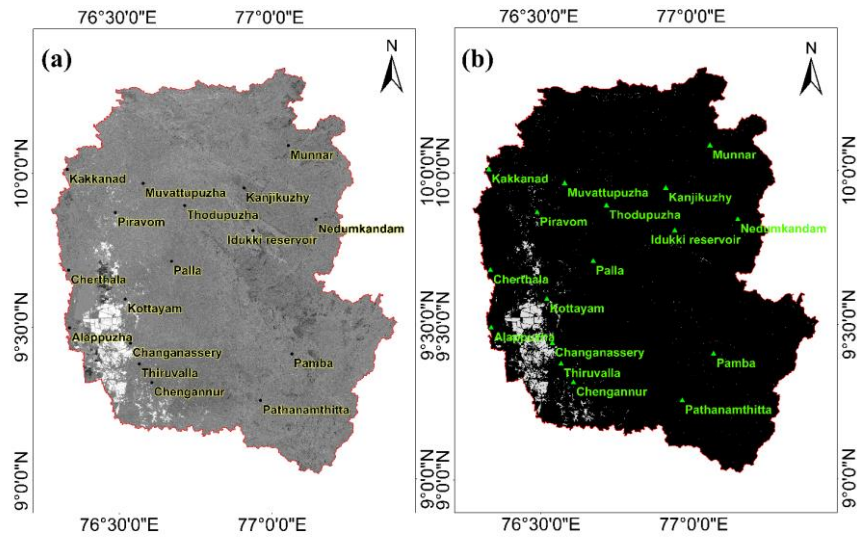
Figure 5.14 Flooded region for 2021 flood

In the process of the image ratio technique, a threshold of 1.20 is applied to extract the flooded region properly. The outputs of the threshold applied image is a binary of 0 and 1, where 0 represents the non-flooded region (black in the shade), whereas 1 represents the flooded region (white patches) in Figure 5.12b, 5.13b and 5.14b. These figures show the clear differentiation between the flooded and non-flooded regions in 2018, 2019, and 2021 respectively. During the 2018 floods, the region surrounding the VLS comprised mainly Cherthala, Kottayam, Alappuzha, Changanassery, Thiruvalla, Piravom, and Chengannur are the most flood-affected regions, which comprises 4% of the total study area as shown in Figure 5.12b. These are some of the important cities situated in the study area and comprise a huge population. During the 2019 floods, the region comprising Kottayam, Alappuzha, Changanassery, Thiruvalla, Piravom, and Chengannur are affected by the floods, and it comprises 3.21% of the study region, as shown in Figure 5.13b. The floods during 2021 comprised major parts of Kottayam, Changanassery, Thiruvalla, Chengannur, Alappuzha, and Piravom, where 2.36% of the total study area was flooded, as shown in Figure 5.14b. The above analysis shows the impact of the 2018, 2019, and 2021 floods on the important urban regions in the study area. These flooded areas are combined using GIS software and compiled to create the flood inventory mapping, which is used as an input to the FSM in the study.

5.3.2　Flood inundation region

The LULC map created using the RF algorithm for 2019 is used to calculate the amount of inundation in the region during the 2018 floods. The area inundated by each LULC class for 2018 is determined using the ArcGIS environment. The analysis shows that 14.7 km^2 of the built-up region, 103.4 km^2 of vegetation, 14.6 km^2 of barren land, and 10.4 km^2 of forest were flood inundated during the 2018 floods (Figure 5.15). From the flood-inundated region derived from SAR imagery (Figure 5.12b), it is concluded that the 2018 floods have been the worst in the state in terms of magnitude. Thus, this scenario is treated as the basis for the likely flood-affecting regions for the future LULC projections of years 2035 and 2050. The 2035 and 2050 LULC maps obtained using the CA-Markov analysis are considered to determine the intensity of the flood for these years. It has been observed that 19.87 km^2 of the built-up region, 99.18 km^2 of vegetation, 13.7 km^2, and 9.68 km^2 of barren land and forest, respectively, will get inundated during the year 2035 in the study region (Figure 5.15). This shows a 34.99% increase in the inundation of the built-up and a 4% decrease in the inundation of the vegetation when compared with the 2018 LULC map. From the 2050 predicted map, it has been analysed that 23.32 km^2 and 99.31 km^2 of built-up and vegetation, respectively, will be inundated. Whereas 13.78 km^2 and 9.58 km^2 of barren land and forest, respectively, will be inundated by the flood in these regions. This is 58.4% of the increase in inundation of the build-up compared with the 2018 flood-inundated region. Between 2035 and 2050, the increase in the inundation of the built-up region is about 17.36%. The area of inundation for 2018 and for years 2035 and 2050 (future predicted) is shown in Figure 5.15. Considering the above facts, floods of similar intensity as 2018 will cause severe damage in the future years 2035 and 2050.
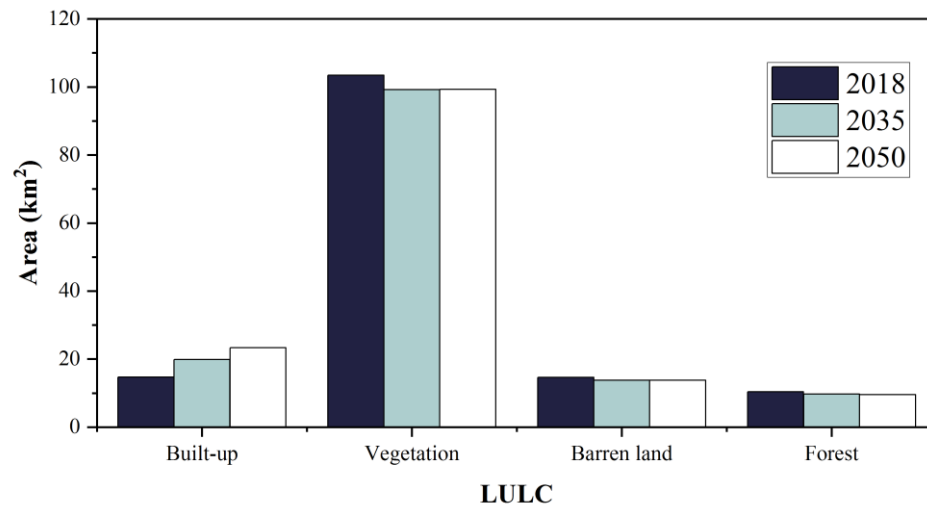
Figure 5.15 Flood inundation for the years 2018, 2035 and 2050

5.3.3 Flood Susceptibility Mapping

The sub-setting of data into training and testing datasets is the priority following the data gathering. The flooded and non-flooded regions are the primary attributes in the data classifying them as 1 and 0, respectively. The dataset is prepared using the flood inventory points and the flood conditioning factors and is fed into the R program for further data modelling. The eighteen parameters comprise both categorical (seven) and numerical (eleven), in which each category in the categorical layer is converted into a single layer using one hot encoding. Thus, it helps in converting these categorical layers into numerical values. The other eleven numerical layers are normalized using the scale function in the R program. So all the input data are normalized between 0 and 1. The RFE technique is used for the feature selection process with RF as a base model. This assists in reducing model dependencies and collinearity by iteratively eliminating a modest number of features in each iteration. The feature selection is carried out by calculating the MDA approach for all the data combinations. The maximum accuracy of 0.94 and Kappa of 0.88 is obtained for the twenty-two variable combinations out of the sixty-one variables. These twenty-two variables include slope, elevation, TRI, TWI, SPI, NDVI, plan_c, wind, rain, prof_c, ge9, STI, l2, b, l5, s1, ge6, g7, ge7, l4, e and f. The variable importance of each variable is shown in Figure 5.16. Thus, these twenty-two variables are fed into the ML models, and the performance of the models is evaluated.

113
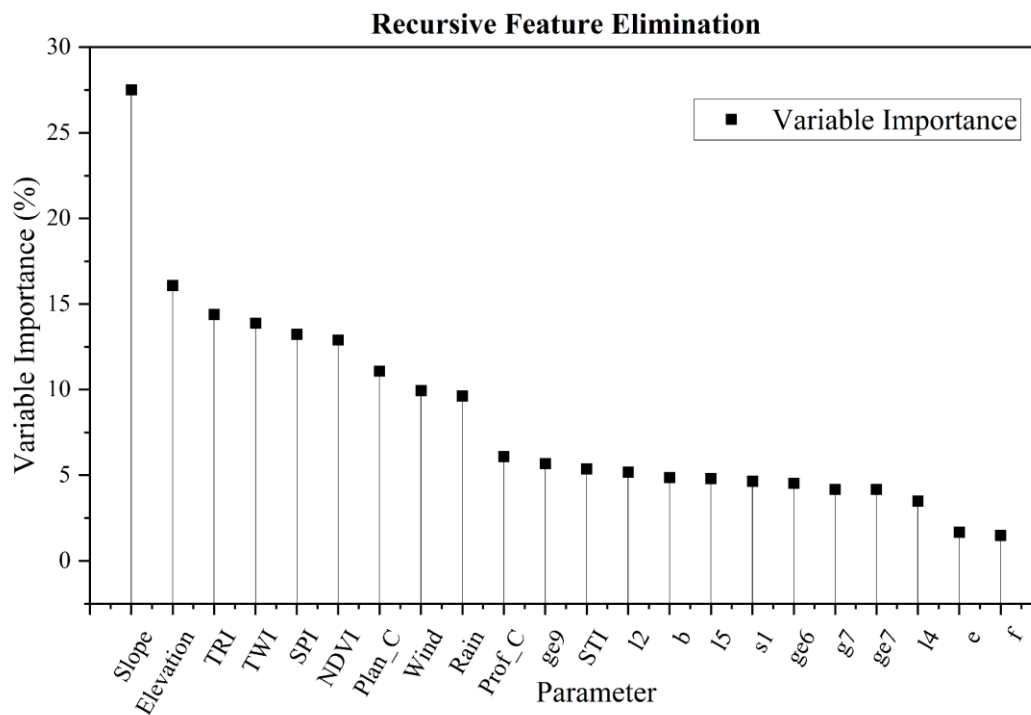
Figure 5.16 Variable importance of RFE

The FSM is created by merging the twenty-two factors with the four stand-alone models, and their performance is assessed. The values of flood susceptibility range from 0 to 1. The natural break algorithm is used to reclassify the susceptibility maps into five categories very low (0-0.2), low (0.2-0.4), medium (0.4-0.6), high (0.6-0.8), and very high (0.8-1.0) classes. This algorithm helps to identify the trends in the representation of the datasets and the accurate classification of the data. The parameter tuning and the performance of each algorithm are discussed as follows.

5.3.3.1 Performance of AdaBoost

The performance of the algorithm is weighed by performing hyperparameter tuning. The model uses n iterations to train and build the base classifiers to combine them to form a final output. The optimal values are obtained in the grid search approach and the 10-fold cross-validation approach in the tuning. The highest ROC-AUC value of 0.9476 is obtained for the training data with n iteration of 100 and the model AdaBoost.M1 when multiple combinations of permutations are used (Figure 5.17a). The model AdaBoost.M1 is a discrete AdaBoost model that is used for classification rather than regression. The slope has the full significance of 100%, followed by elevation

114

(96.96%), TRI (92.86%), TWI (88.28%), and SPI (79.99%) in the model's variable importance plot (Figure 5.17b). STI is ranked last, with a score of 28.02%. The testing dataset has a ROC-AUC value of 0.90, an accuracy of 0.87, and a Kappa value 0.72. The AdaBoost model has an F1 score of 0.897 for the testing dataset. The sensitivity and specificity values are 0.75 and 0.95, respectively. The flood susceptibility map obtained from AdaBoost is shown in Figure 5.18. Around 48% of the region, or 5843 km^2, is very low susceptible. 25% of the study area is in the moderate and low susceptibility zone, while the remaining 27% is in the high and very high susceptibility zone.
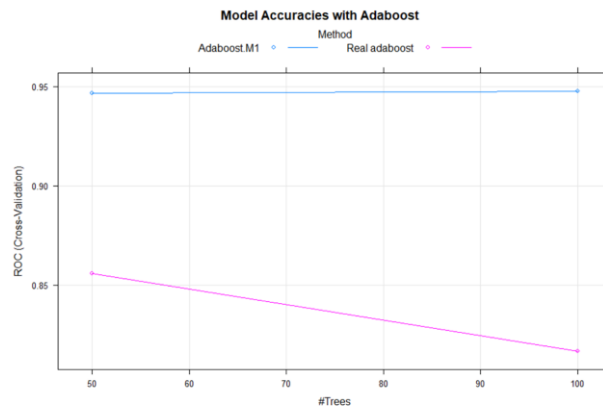
(a)
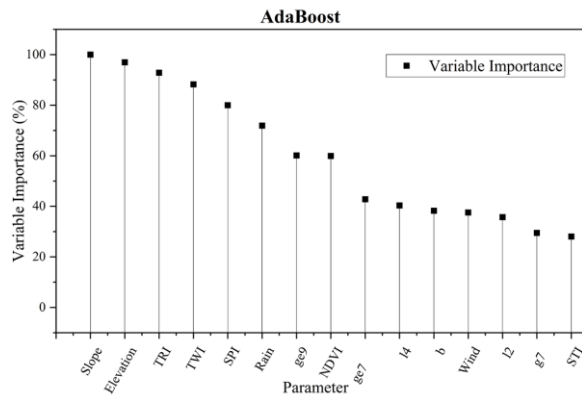
(b)

Figure 5.17 (a) Model accuracy and (b) Variable importance of AdaBoost
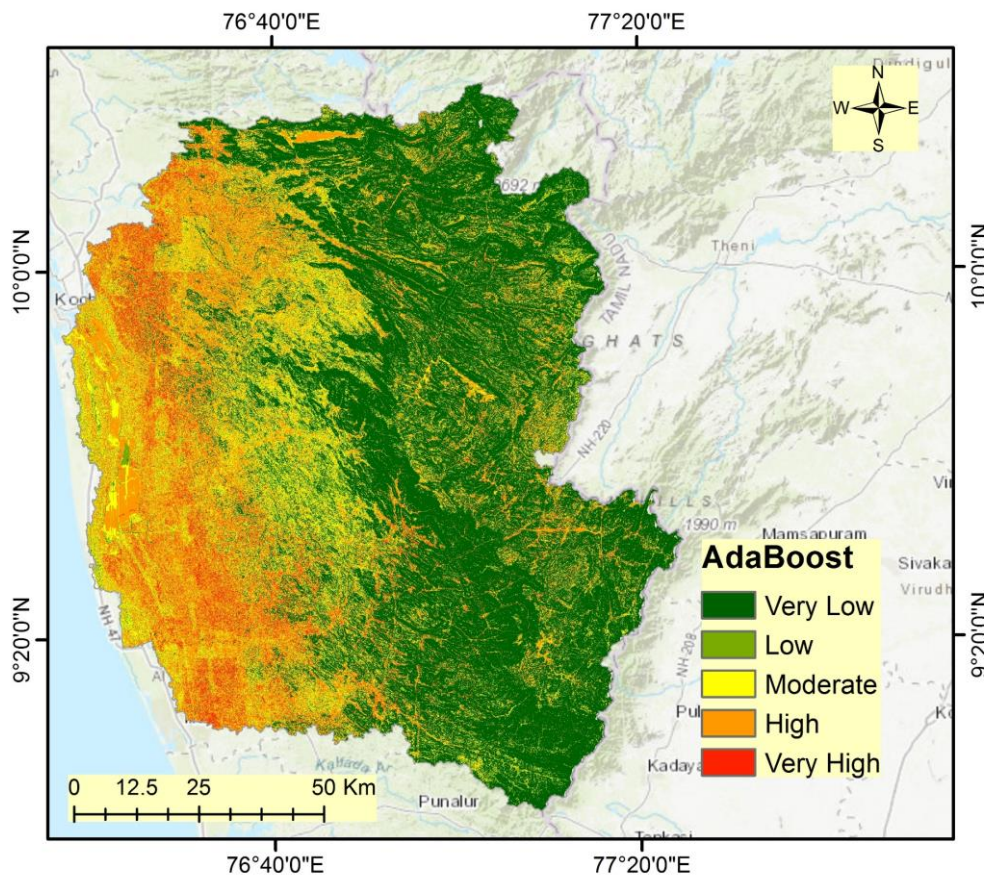
Figure 5.18 Flood Susceptibility Map using AdaBoost

5.3.3.2 Performance of Random Forest

The optimal values by resampling the results across tuning parameters resulted in the best ROC-AUC value of 0.9464 for the training data with a mtry of 25 obtained from grid search and a 10-fold cross-validation approach (Figure 5.19a). The number of variables randomly picked as candidates at each split is defined by mtry. The fifteen most significant variables for the model are ranked in which slope has a maximum variable importance of 100%, followed by elevation (24.85%), TRI (9.01%), wind (4.38%), and NDVI (4.16%). SPI, TWI, rain, prof_c, plan_c, STI, f, e, and ge6 has an importance of less than 4%, and the least important parameter is l5, i.e., waterbodies with a significance of 0.23% (Figure 5.19b). For the testing dataset, the ROC-AUC value is 0.89, while the accuracy and Kappa values are 0.87 and 0.73, respectively. The RF model has an F1 score of 0.899 for the testing dataset. The sensitivity and specificity values for the model are 0.75 and 0.96, respectively. The flood susceptibility map obtained from the RF is shown in Figure 5.20. It is found that 42% of the region is

classified as a very high and high susceptible zone, while 48% is classified as a very low susceptible region. Only 10% of the region is classified as low and moderately susceptible.

(a)

(b)
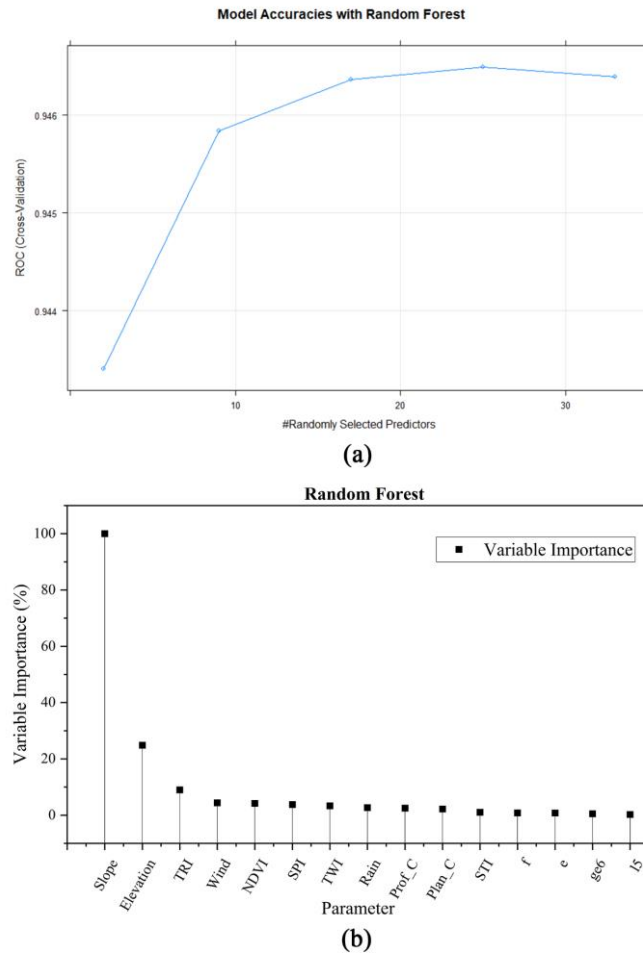
Figure 5.19 (a) Model accuracy and (b) Variable importance of Random Forest
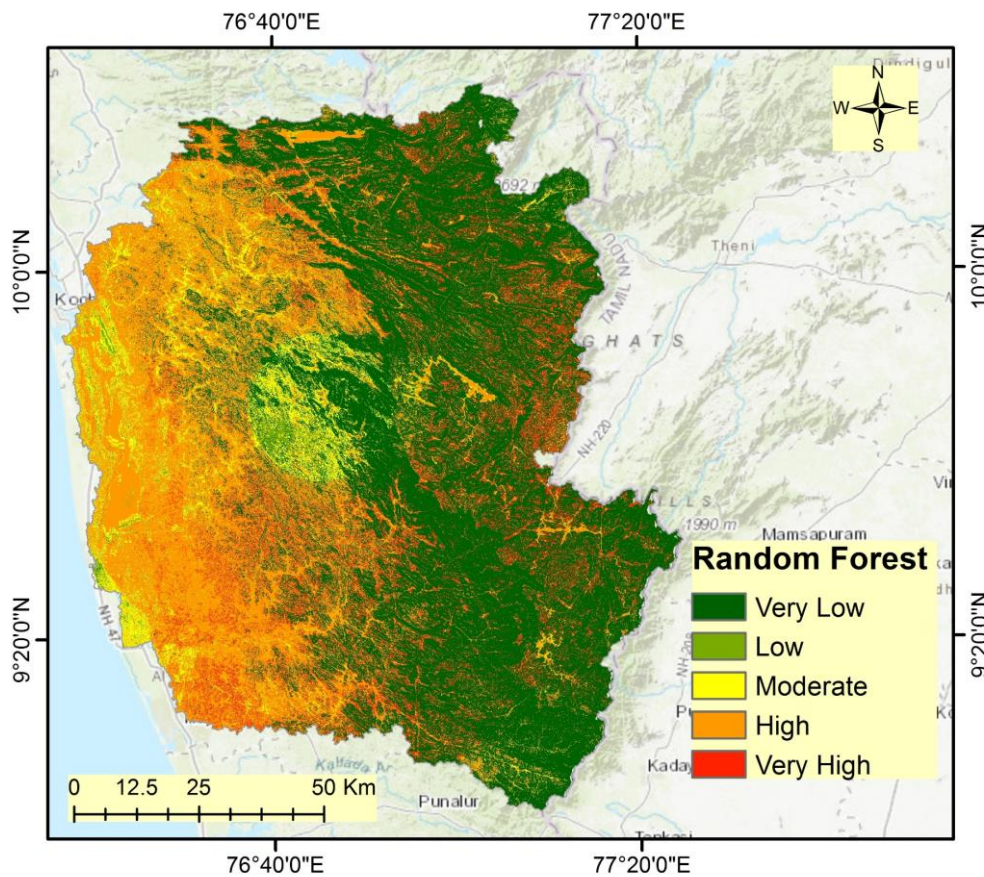
Figure 5.20 Flood Susceptibility Map using Random Forest

5.3.3.3 Performance of Gradient Boosting Machines

The GBM model uses significant tuning parameters such as n.trees, shrinkage, interaction.depth, and n.minobsinnode. n.tree specifies the number of gradient boosting iterations performed in the model. Shrinkage is the learning rate of an algorithm; interaction.depth helps in the number of splits that the tree has to perform from the starting node. n.minobsinnode denotes the minimal number of observations in the terminal nodes of trees. The hyperparameter tuning obtained from grid search and ten-fold cross-validation approach resulted in the optimal GBM model with the ROC-AUC values of 0.9494 for the training datasets. The parameter values are set as follows: n.trees = 150, interaction.depth = 3, shrinkage = 0.1 and n.minobsinnode = 10 (Figure 5.21a). The fifteen most important variables are ranked as the slope (100%) being the highest importance, followed by elevation (4%), TRI (2.2%), NDVI (2.18%), and wind (2%) (Figure 5.21b). TWI, plan_c, SPI, prof_c, rain, e, f, STI, l4, and the least important

118

variable s1, scores less than 2% importance, are among the other variables. The ROC-AUC value, accuracy, and Kappa of the testing datasets are 0.90, 0.88, and 0.74, respectively. The GBM has an F1 score of 0.903 for the testing dataset. The sensitivity and specificity for the model are 0.89 and 0.95, respectively. According to the flood susceptibility map (Figure 5.22), 42.5% of the region has a high and very high risk of flooding. Low and moderately sensitive areas comprise 8.5% of the region, whereas very low susceptibility areas comprise 49%.
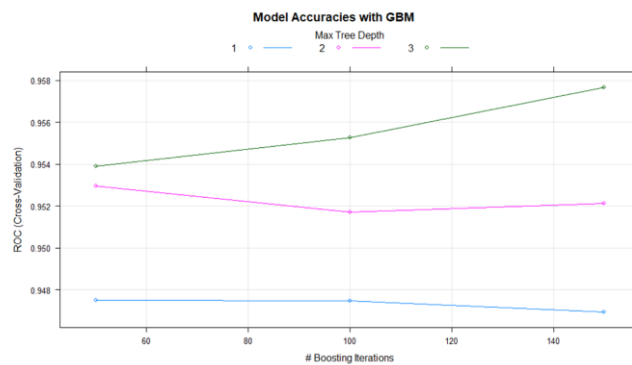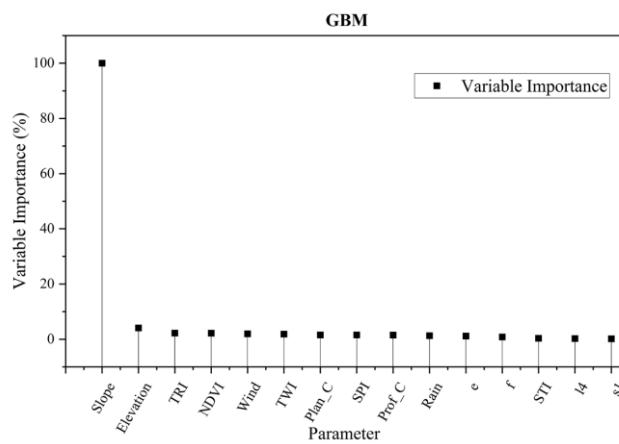
Figure 5.21 (a) Model accuracy and (b) Variable importance of GBM

119
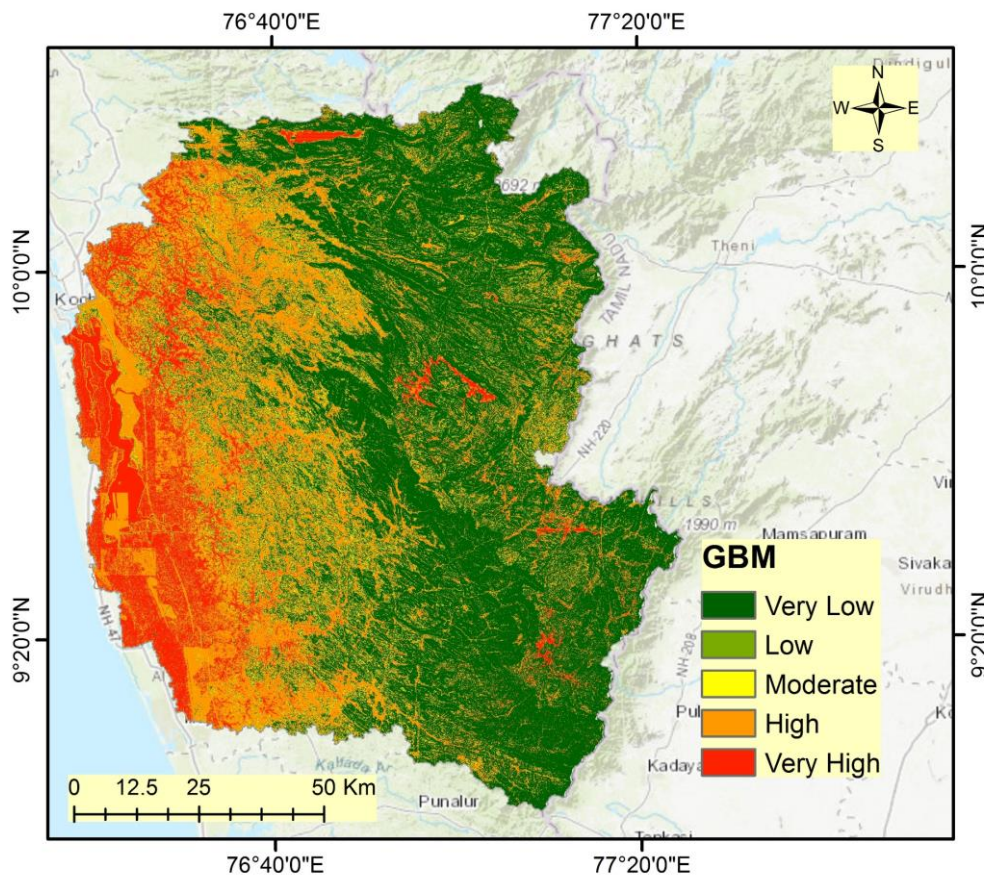
Figure 5.22 Flood Susceptibility Map using GBM

5.3.3.4 Performance of Extreme Gradient Boosting

The XGBoost is a distributed gradient-boosting library optimized for efficiency, flexibility, and portability between languages (Chen and Guestrin 2016). The highest ROC-AUC value of 0.9494 is obtained from the hyperparameter tuning using grid search and a ten-fold cross-validation approach for the training dataset. The hyperparameter values for the model tuning are nrounds=150, max_depth=2, eta=0.4, colsample_bytree=0.8, and subsample=0.625. The gamma and min_child_weight are held constant at 0 and 1, respectively (Figure 5.23a). The importance of the slope is the highest, being 100%, followed by TRI (30.05%), NDVI (14.20%), wind (13.72%), and TWI (10.87). The other parameters have scored less than 10% of importance, including SPI, elevation, rain, prof_c, plan_c, STI, f, e, l5, and the least being ge6 (Figure 5.23b). The ROC-AUC value of the testing dataset is 0.90. The accuracy and Kappa for the testing datasets are 0.92 and 0.84, respectively. The XGBoost has an F1 score of 0.934 for the testing dataset. The sensitivity and specificity values are 0.90 and 0.94,

respectively. The flood susceptibility map obtained from the XGBoost model is shown in Figure 5.24. From the susceptibility map, it was found that 35% of the study area falls in the very high and high susceptibility zones. 55% of the study area falls under the very low susceptibility zone. Only 14% of the region falls in the low and medium susceptibility zone.
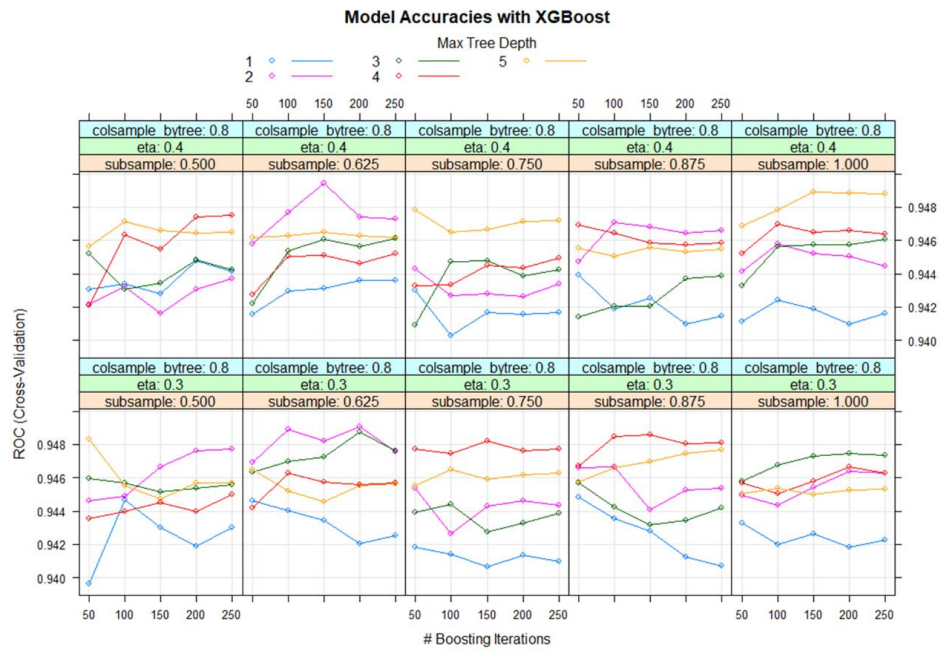
(a)
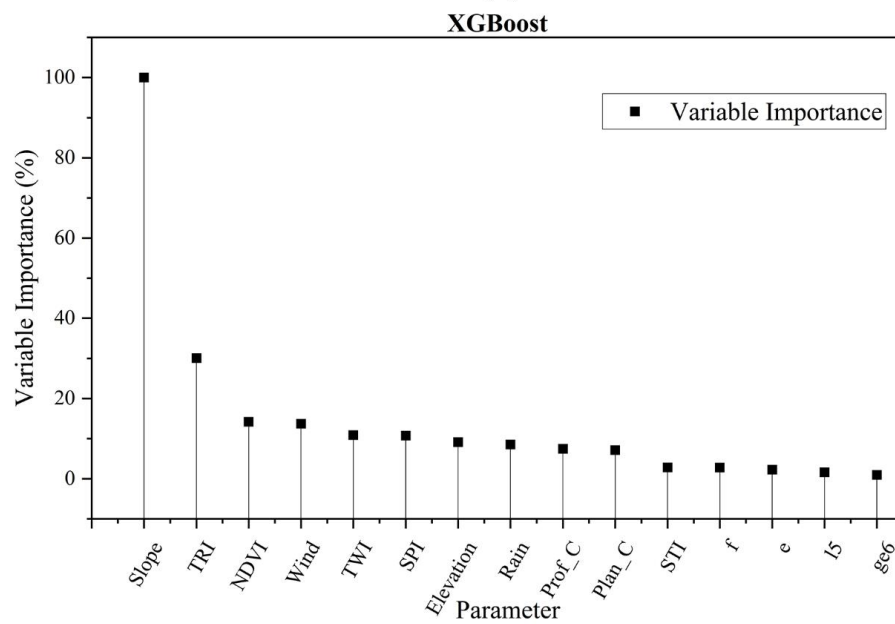
(b)

Figure 5.23 (a) Model accuracy and (b) Variable importance of XGBoost
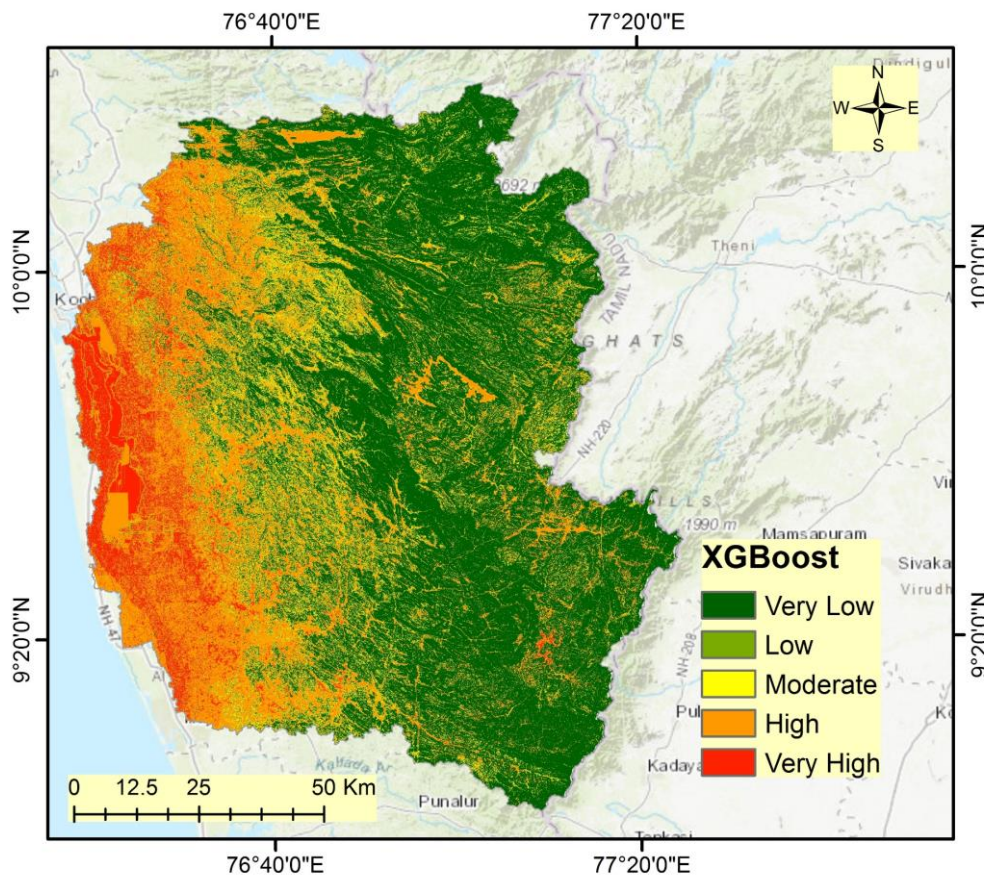
Figure 5.24 Flood Susceptibility Map using XGBoost

5.3.3.5 Discussion on Flood Susceptibility Mapping

Floods are the most catastrophic natural disasters, causing massive financial losses and human deaths worldwide. The important influencing factors are not only human or anthropogenic activities but also meteorological factors and physical characteristics of the watershed, as it plays an important role in the intensity of the flood (Pati and Sahoo 2021). Because of the changing climate, the frequency of floods and interruptions has increased dramatically (Abraham and Kundapura 2022; Jose and Dwarakish 2022). Moreover, recurrent floods in India are likely to occur with the increasing air temperature and rainfall maxima during the monsoon months (June–September) (Chakrabortty et al. 2022; Kumar et al. 1999). This was the case during the 2018 Kerala floods, where the excess rainfall led to the rapid filling of the reservoir, and all the reservoirs in the state attained their FRL by the end of July 2018. This was a matter of concern in the state, prompting officials to release water from reservoirs, resulting in a catastrophic situation due to floods and landslides (Sudheer et al. 2019). Apart from the

122

pre-flood, during-flood, and post-flood actions performed, producing a flood susceptibility map for the region is critical, which will be the most important step taken to protect the people from the deadly flood, with the evolution of remote sensing, GIS, ML and cloud computing technologies in the recent decades changed the perception of the data visualization and the ability to simulate data with a large number of parameters (Wang et al. 2019a). There is also great reliability in the simulations and their practical applicability, which are essential to ML models. Many research has been carried out to utilize the capability of learning algorithms, statistical, and data-mining models and to improve their predictive performance, as modelling and simulation systems are valuable decision-support tools for analysing past and present flood events and providing precise flood susceptibility maps (Balica et al. 2012; Das 2019; Dodangeh et al. 2020; Mahato et al. 2021; Meliho et al. 2022; Saha et al. 2021; Shafizadeh-Moghadam et al. 2018).

One of the most influential parameters in flood susceptibility studies is the slope, which is evident in the variable importance chart of RFE and all four models. The slope is an important factor that controls the water infiltration rate, where the surface and sub-surface runoff increases with respect to an increase in the slope (Mu et al. 2015). At the same time, it increases the velocity of the runoff with an increase in the slope resulting in a faster accumulation of water at the end of the basin. The susceptibility maps show that the terrain slope between 0 to 5° is the most vulnerable to floods. It is also noted that VLS as a whole constitutes a nearly circular-shaped watershed. Thus, the above conditions may result from the peak flow in the discharge value becoming sharper in the hydrograph (Saghafian et al. 2002). In the AdaBoost, RF, and GBM models, elevation is the second most weighted parameter, whereas it is placed seventh in the XGBoost model. In the eastern part of the study lies the Western Ghats, and most regions are significantly 500 m above the MSL. The lowland is towards the western part of the study area covering the coastal regions. The elevation significantly impacts floods because the water flows from higher to lower elevations, resulting in less water storage in highly elevated areas. The susceptibility maps generated by all the model shows that the region at higher altitudes, i.e., in the eastern part of the study area, is less vulnerable to flooding than the region at lower elevations. The model shows the very high and highly susceptible region in and around the Kochi and Alappuzha, where the

elevation is less than 50 m above the MSL. Rainfall is one of the important parameter as the floods discussed are rainfall-induced floods. In the study, the main cause for the flood is both prolonged rainfall i.e. lasting for days or weeks as well as heavy downpour in the region. This significantly results in the saturation of the terrain resulting in the more surface and subsurface flow. Thus apart from the other parameters, the conjunction of these three distinguishing traits such as low elevation, less slope and heavy rainfall has a direct impact on the flood susceptibility of the region.

TRI and SPI are negatively associated with the flood susceptibility zones where the increase in the ruggedness of terrain poses less threat to flooding. The SPI shows the abrasive power of the stream, and the possibility of flooding is high in the area where the SPI values are less. This occurs because when the abrasive power declines, water stagnates, resulting in floods (Chowdhuri et al. 2020). Whereas the TWI shows a positive correlation with the flooding scenarios as the wetness of the terrain increases, it results in less seepage of the water to the ground and results in huge surface flow. Higher STI values show high runoff and less stagnation. Hence, the area with low STI values increases the risk of flooding (Tehrany et al. 2017). Profile and plan curvature values that are moderate and near zero influence the risk of flood in the region. NDVI negatively correlates with the flood, as the high NDVI values denote the vegetation and forest. Hence the risk of flooding is less, whereas the low NDVI values of less than 0.2 comprising of barren land, waterbody, and impervious layers like roads and settlements may lead to less water seepage and increased surface flow. In the geomorphology parameters, ge9 negatively correlates with flooding. Whereas ge6 and ge7 positively correlate for the flood scenario as they are the low-lying region, and the rivers converge together to drain into the lake, increasing the possibility of flooding in the region. Clay soil has an impact on floods since it has a poor infiltration rate due to its low permeability. Because of their high porosity, they retain more soil moisture, resulting in increased runoff (Taylor and Krüger 2019). The key influencing land use factors for flooding are forest and vegetation types, which have a negative impact on flooding because they absorb water and minimize the impact of floods. The flood susceptibility map of all four algorithms shows that the very high flood susceptibility spans from 4.7% to 10.17% of the total area in the study area across all models, while the high

susceptibility ranges from 22.5% to 33%. The total area of the susceptibility in percentage is shown in Figure 5.25.
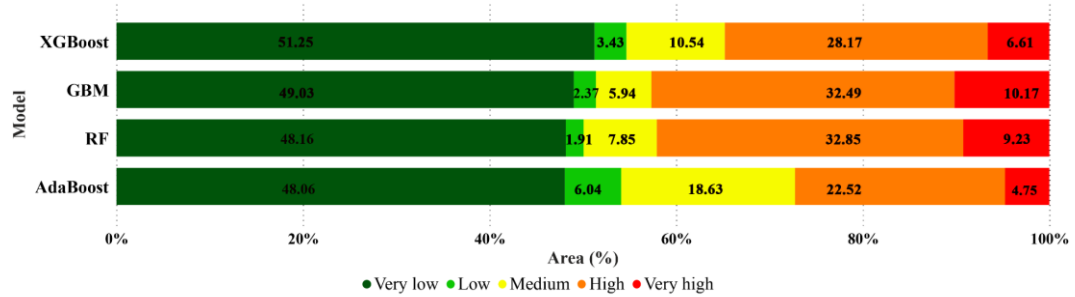
Figure 5.25 Percentage area susceptible to flood

The plot of the accuracy and Kappa values in Figure 5.26 illustrates the overlap of the boxes and the mean values of the four models. It is arranged in descending order of mean performance in terms of model accuracy. The mean accuracy values for XGBoost, GBM, RF, and AdaBoost are 0.944, 0.943, 0.941, and 0.939, respectively. The XGBoost model has a distribution of 0.90 to 0.97. The interquartile range is from 0.933 to 0.955 as well. The interquartile range for the RF model is relatively narrow, ranging from 0.933 to 0.944. The GBM model has a distribution between 0.90 and 0.97. However, the outliers of the AdaBoost are from 0.894; the lower and the upper whisker lies at 0.917 and 0.966, respectively. The interquartile range for the model is from 0.933 to 0.948. The findings show that the XGBoost outperforms the other three models, with a ROC-AUC value of 0.90 and accuracy, Kappa, and F1 scores of 0.92, 0.84, and 0.934, respectively. Followed by GBM, RF, and AdaBoost, with ROC-AUC values of 0.90, 0.89, and 0.90, respectively (Figure 5.27), and F1 scores of 0.903, 0.899, and 0.897, respectively.
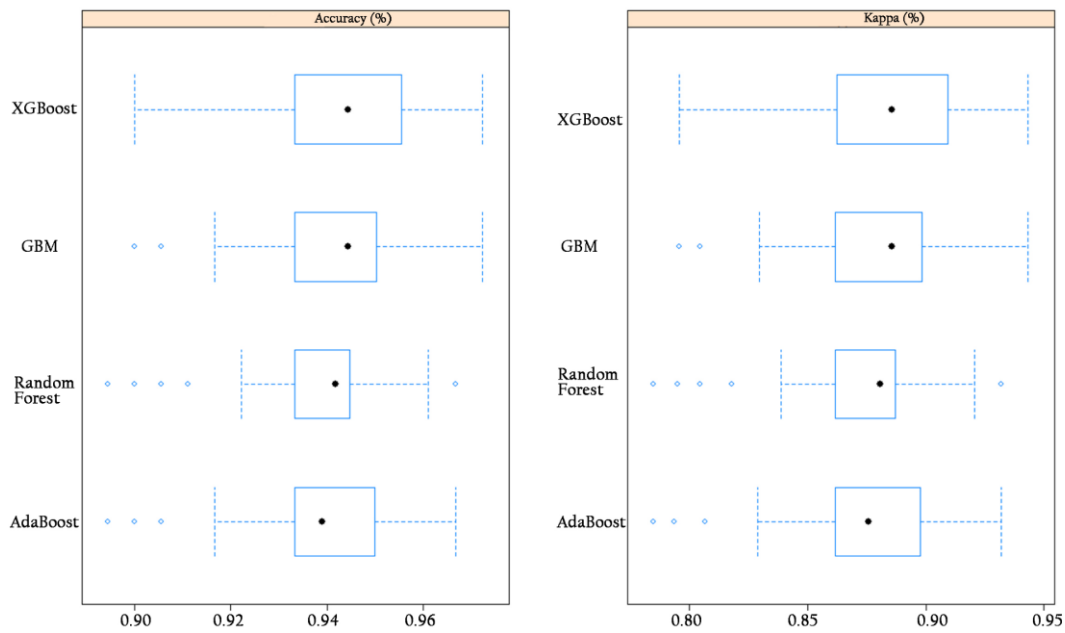
Figure 5.26 Plot of accuracy and Kappa
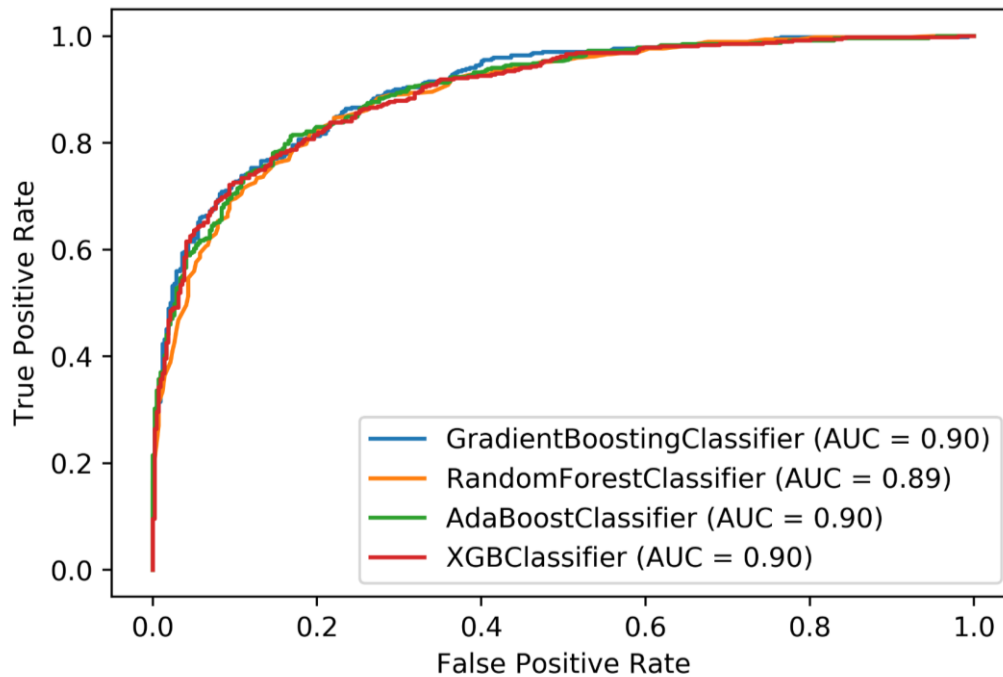
Figure 5.27 ROC-AUC of the models

The purpose of the study is that the VLS has been facing catastrophic flooding for the previous half a decade, but no substantial research has been conducted in the region for this lake system. Hence, flood susceptibility mapping is carried out using ML algorithms to analyse the susceptibility of the region. The parameters employed in this

work and the models may be widely accepted to map the flood susceptibility zone for rainfall-induced flood occurrences, providing that the main influencing parameters and geographical topography are the same as in Kerala. All four models considered for the study are decision tree-based models. One significant advantage of the tree-based model is that it considers interaction effects between independent variables and ranks the influential input factors depending on their contribution to the modelling process (Johnson et al. 2002).

5.4 CLOSURE

This chapter focuses on delineating flood-inundated areas using SAR imagery for 2018, 2019, and 2021. The extent of flood inundation during the 2018 floods and the possible flood inundation region for projected future LULC in 2035 and 2050 is determined. The flood inventory map is created using SAR data for 2018, 2019, and 2021. The flood susceptibility map for the VLS is prepared using four different decision tree-based algorithms AdaBoost, RF, GBM, and XGBoost.

The flooded regions are delineated using the Sentinel-1 SAR imagery in the GEE platform by applying the threshold for the pre and during-flood scenario. It was found that almost 4%, 3.21%, and 2.36% of the study area was flooded during 2018, 2019, and 2021 respectively. The affected regions during these floods are Kottayam, Changanassery, Thiruvalla, and Piravom. According to the analysis, to quantify the extent of flood inundation, 14.7 km^2 of built-up area was inundated during the 2018 floods. Using the 2018 flood event to quantify the amount of flood that may inundate the future LULC 2035 and 2050, it is found that the flood will severely affect 19.87 km^2 and 23.32 km^2 of the built-up region, respectively.

The generation of FSM and the map creation process is carried out in R and ArcGIS platforms, respectively. The application of decision tree models reveals the adaptability of tree models in classification problems and aids in developing highly accurate FSM. The RFE technique plays a vital role in the feature selection processes as it reduces the model dependencies and collinearity by iteratively eliminating a modest number of features in each iteration. This aids in reducing the dimensionality of the data from the sixty-one layers to twenty-two, resulting in a more accurate prediction of the model. The performance of all four models shows their practical applicability. However, XGBoost performed well in terms of the metrics of the model. According to the

findings, an average of 30-40% of the region falls into the high and very high susceptibility category, which is an alarming range, as evidenced by the number of persons affected in the region. Because several of the state's important cities are on the shore, they are quite close to the MSL, leading to floods in the region. Apart from rainfall, one of the major causes of the flood is topography, which causes the five major rivers in the state to converge and drain into the lake, increasing the amplitude of the flood in the surrounding areas.

CHAPTER 6

ANALYSIS OF THE OPTICAL PROPERTIES OF THE LAKE WATER

6.1 GENERAL

Spatial remote sensing analysis aids in identifying spatio-temporal patterns, which is one of the most important factors to examine in conjunction with *in situ* monitoring and numerical modelling. Satellite remote sensing has recently been utilized to monitor water surface biogeochemical indicators such as Chl-a and SPM in coastal areas. The Inherent Optical Properties (IOPs) of offshore waters are primarily driven by phytoplankton and its associated material (Case 1 waters). The coastal and inland waters are optically more complex due to the diverse presence of suspended, dissolved organic and inorganic substances and multiple combined physical/biological/chemical/geological processes occurring over a wide range of time scales. In contrast to open ocean waters, coastal and inland surface waters have substantial regional variation in terms of bio-optical characteristics, necessitating sensors with medium to high spatial resolution. The SPM, which contains organic and mineral-suspended particles, is one of the most studied characteristics, owing to its importance in various processes like sediment dynamics, coastal erosion processes, and water quality monitoring. Several algorithms have been created to measure SPM using remote sensing reflectance, $R_{rs}(\lambda)$, in which the radiometric parameter is determined after the atmospheric corrections. Among these several techniques, entirely empirical algorithms based on a single band, numerous bands, or band ratios (Doxaran et al. 2002; Forget and Ouillon 1998; Siswanto et al. 2011; Tassan 1994; Yepez et al. 2018), and semi-analytical algorithms (Balasubramanian et al. 2020; Han et al. 2016; Nechad et al. 2010) utilizing IOPs are widely used.

6.2 MATERIALS AND METHODS

6.2.1 Data Preparation

Water quality is influenced by a variety of biological, chemical, and physical factors. Aquatic humus and suspended matter concentrations in surface water are calculated using reflectance and attenuation spectra obtained in the optical atmospheric window. The IOPs of these components must be understood to analyse recorded spectra in terms

of concentrations (Buiteveld et al. 1994). Copernicus Sentinel-2 satellites are a constellation of two polar-orbiting satellites in sun-synchronous orbit phased at 180° to each other. It is an open-access dataset developed under the European Commission and European Space Agency to meet the operational needs of the Copernicus program. Both Sentinel-2A & 2B satellites together form a revisiting capability of 5 days at the equator capturing images at both visible and infrared wavelengths using 13 spectral bands. Though the Sentinel mission was primarily designed for land-based studies and its changes, because of its high spatial resolution (10m, 20m, and 60m) and good radiometric resolution (Pereira-Sandoval et al. 2019), it is an ideal instrument for remote monitoring of the water quality of lakes, reservoirs, and coastal waters (Avtar et al. 2020; Caballero et al. 2020, 2022; Kulk et al. 2021). Level - 1C raw Sentinel - 2 images were downloaded from Copernicus Open Access Hub (https://scihub.copernicus.eu/) for the period of 2016 to 2021. In the present study, the ACOLITE (Atmospheric Correction for OLI lite) generic processor is used to perform the atmospheric correction and derive the optical properties of Lake Vembanad. ACOLITE was created by the Royal Belgian Institute of Natural Sciences (RBINS) specialized in the application of inland and coastal water bodies. It also supports a wide range of sensors, including Sentinel-3 OLCI, Landsat 5/7/8/9, EO1/HYPERION, PlanetScope, HICO, PRISMA, SPOT 6/7, WorldView-2/3, and CHRIS. The Dark Spectrum Fitting (DSF) algorithm for atmospheric correction of the Sentinel-2 and Landsat series satellite images (Vanhellemont and Ruddick 2016a; b) has demonstrated its potential for monitoring the changes in water, and this is possible due to its spectral resolution (Caballero et al. 2022). Since sunglint interference is noted at Level-1C, the optional image-based sunglint correction of the surface reflectance is also considered. Finally, the bottom-of-atmosphere reflectance is derived for visible, NIR, and SWIR bands.

6.2.1.1 Diffuse attenuation coefficient (K_d)

The diffuse attenuation coefficient K_d is one of the most widely used parameters to represent the availability of light in the water column. The optical properties of absorption and backscattering coefficients are intrinsic. K_d is an apparent optical property affected by the Raman scattering or chlorophyll fluorescence caused by the

angular distribution of light in the water column and inelastic scattering mechanisms (Mishra et al. 2005). They influence the colour of the water, which is often determined by the water-leaving radiance, together with downwelling Sunlight and sky (ratio of water-leaving radiance to above-surface downwelling irradiance). Since a water's IOPs are directly related to its constituents, their values are used to identify the water type, the intensity of subsurface light, the depth at which solar heat fluxes, pigment concentration, and sediment loading. K_d is derived from the absorption outputs from the Quasi-Analytical Algorithm (QAA) of Lee et al. (2002) at 490 nm using ACOLITE software.

6.2.1.2 Suspended Particulate Matter (SPM) concentration

SPM is an important factor in determining the water quality in and around the riverine and coastal areas as the natural light will be disturbed by the high concentrations of suspended sediments in estuaries and inland lakes, causing it to affect the development of aquatic species and the primary productivity of water bodies. These suspended particles can contain minerals, humic, or phytoplankton resulting in cyanobacterial growth. Hence, SPM is composed of both inorganic and organic material and is maintained in suspension in the upper mixed layer by physical force-like currents. The majority of the mineral particles in the inorganic fraction come from stream flow and erosion (Garg et al. 2020). These particles backscatter a part of UV rays from the Sun, resulting in favourable conditions for bacterial growth. The cyanobacteria once developed are difficult to control and will result in eutrophication, which is a great threat to both human and animal life. As well, extreme human activity along the river's course and runoff with the high-suspended matter may cause high Total Dissolved Solids (TDS) content in the rivers (Jafar-Sidik et al. 2017). Mapping of SPM from satellites and airborne imagery has become a valuable tool for assessing and monitoring suspended particle distribution (Alem et al. 2021). Ritchie et al. (1976) found a curvilinear relationship between the SPM and the remote sensing reflectance or radiance. The 1 to 50 gm^{-3} concentration shows a linear relationship between the radiance and the SPM, whereas the 50 to 500 gm^{-3} range necessitates a higher curvilinear relationship (Ritchie et al. 2003). SPM concentration is computed using the algorithm for turbid waters proposed by Nechad et al. (2010) and recalibrated by

Nechad et al. (2016) specifically for Sentinel-2 red band using the ACOLITE processing software.

6.2.1.3   Turbidity

Turbidity is another key parameter for analysing the quality of the natural water in lakes and rivers (Dogliotti et al. 2015), and remote sensing is an effective way of retrieving data temporally. It determines the optical property of the water where the suspended sediments in the lake attenuate the light rather than transmit it along the water depth. High quantities of these substances may cause turbidity and reduce the amount of light accessible to phytoplankton for photosynthesis. Higher trophic levels may experience cascade effects from a decline in phytoplankton biomass, and there have also been reports of direct impacts of suspended matter on zooplankton and fish survival (Kulk et al. 2021). The turbidity is usually measured in the field using the Secchi disk, where the disk is lowered in the water column, and the visibility of the disk is measured accordingly. The information obtained from the Secchi disk is point data rather than spatial, but the concentration of the suspended sediments varies temporally and spatially. Despite the accuracy of the field measurements, they are more time-consuming, costly, and labour-intensive (Kari et al. 2016). It has been stated that the visible part of the spectrum has a good relationship with the concentration of suspended sediments, but the red and NIR sections of the spectrum are more sensitive to turbidity (Caballero et al. 2020; Garg et al. 2020). The turbidity is measured using a bio-optical approach that correlates the red spectral reflectance of water derived by Nechad et al. (2016) in ACOLITE, and it is measured in Formazin Nephelometric Unit (FNU).

6.2.1.4   Chlorophyll-a (Chl-a) concentration

Deriving Chl-a concentration acts as a crucial water quality indicator, a predictor for biomass, and a public health measure for cyanobacteria monitoring is a successful strategy for keeping track of Harmful Algal Blooms (HABs) in inland and coastal waterbodies (Caballero et al. 2020). Monitoring the Chl-a level is essential to determine the eutrophic state of the water bodies. Numerous remote bio-optical techniques have been used to assess concentrations of the typical phytoplankton pigment Chl-a. The red-NIR bands, which use the Chl-a absorption peak at 665 nm and red-edge location, are the primary basis for Chl-a algorithms for coastal and inland water bodies. The red-NIR

band ratio algorithms produce more accurate retrievals of Chl-a concentration in turbid inland and coastal waters because they are less impacted by the absorption caused by other water components (Dev et al. 2022). The algorithm proposed by Mishra and Mishra (2012) is used in this study. It comprises two bands to derive the Normalized Differential Chlorophyll Index (NDCI), demonstrating its ability to estimate the Chl-a concentration in turbid productive waters. The red band at the wavelength of 665 nm is more sensitive and absorbs the Chl-a pigment, whereas the NIR band at 708 nm is less sensitive to the Chl-a pigment. Hence, the spectral band difference of 708 and 665 nm is taken and normalized using their sum of the differences. The ACOLITE software is used to estimate the Chl-a concentration using NDCI proposed by Mishra and Mishra (2012).

6.2.2 Data Visualization

The processed data from ACOLITE comprises L2W parameters in a network Common Data Form (netCDF) format. These data are loaded into the SeaDAS software, which extracts the values of each parameter within the study region as ASCII files. The average values for each parameter are calculated from these ASCII files, and those values are visualized in a graphical format to analyse the trends that those parameters follow, as discussed in the following section.

6.3 RESULTS AND DISCUSSION

The stream flow of the six rivers reaches the Vembanad Lake. The study period considered is from January 2016 to December 2021. Only five scenes of data were evaluated during 2016 due to the presence of the cloud in the data and since Sentinel-2B was launched later in 2017. Thus a total of 109 scenes are evaluated from 2016 to 2021. As illustrated in Figure 6.1, the lake has been split into three segments to discuss the outcomes: Northern (A), Central (B), and Southern (C). In order to analyse and summarize the results, the mean values of the pixels for the entire image are calculated and evaluated. The river streamflow reaching the lake is graphically represented with the mean values of water quality parameters. As a result, understanding the changes in total water quality parameter values becomes convenient. Changes in each sector of the lake are also examined and explained for a better understanding of the changes happening.
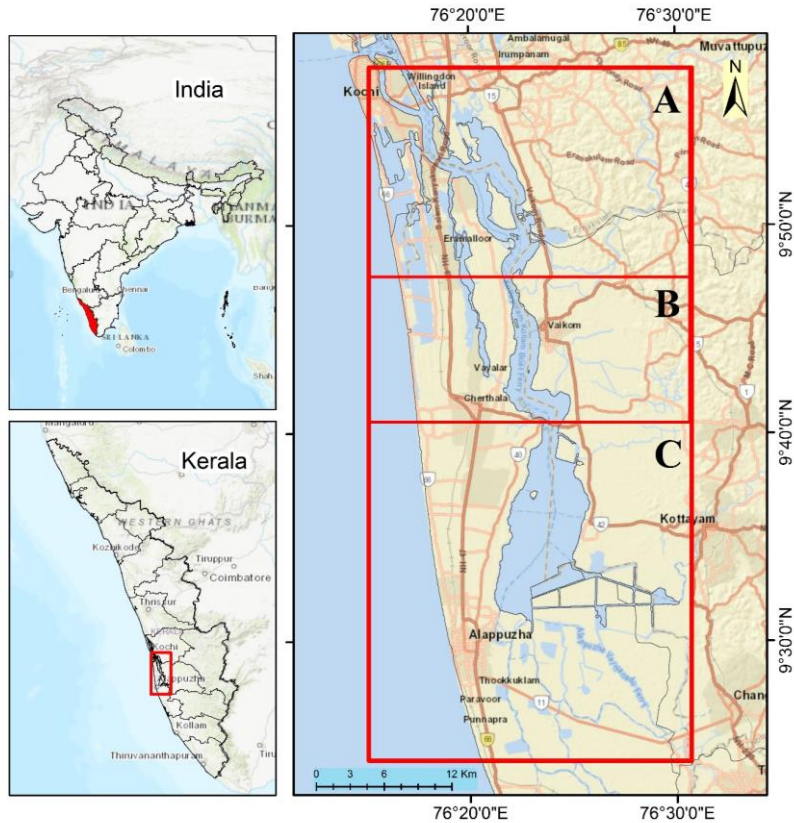
Figure 6.1 Study considered for the analysis

6.3.1 Variability in water quality parameters

The observations reveal that the mean values of K_d are at their lowest during the pre-monsoon season, i.e., from January to May, when the streamflow generated by the rivers reaching the lake is the least. From 2016 to 2021, the least mean K_d (±standard deviation) observed is on the 14th of April 2019 with a value of 0.915±0.15 (Figure 6.2). The highest mean value of K_d is found during the monsoon season. During December 2017, the K_d value increased by more than 1.3±0.1 in the central and southern regions of the lake. The 2018 flood influenced the increase in K_d value to be more than 1.5 in the lake's southernmost region, where the average streamflow value of the rivers is about 2250 m^3/s (Figure 6.2), and its effects can be observed from September to November (Figure 6.3). The rest of the southern lake has comparatively less K_d values of around 1±0.2. There is a significant increase in the value of K_d (greater than 1.5) in the northern part of the lake during the post-2019 floods during October and November, as shown in Figure 6.3. The COVID-19 breakout caused a nationwide lockdown in mid-March 2020. Due to the first COVID-19 lockdown, industries and tourism halted,

134

resulting in less pollution. The K_d value drops significantly during this lockdown. According to Figure 6.3, the southern part of the lake has a very low diffuse attenuation value, i.e., less than one, during April 2020. The second wave of COVID-19 hit the country in April 2021, leading to the lockdown by the end of April. This has also halted the normal functioning of industries and tourism. Due to the presence of the cloud in the Sentinel-2 data, a higher mean value of K_d (1.62) is displayed in Figure 6.2 during May, which may be treated as an outlier. The month of June recorded the lowest K_d value during the monsoon season, with a mean value of 1.19±0.2 which might be the influence of the second lockdown. It is observed that the K_d value decreased by overall 5.15% and 2.5% for 2019 and 2021, respectively, from 2017.

Figure 6.2 Variation of average K_d from 2016 to 2021
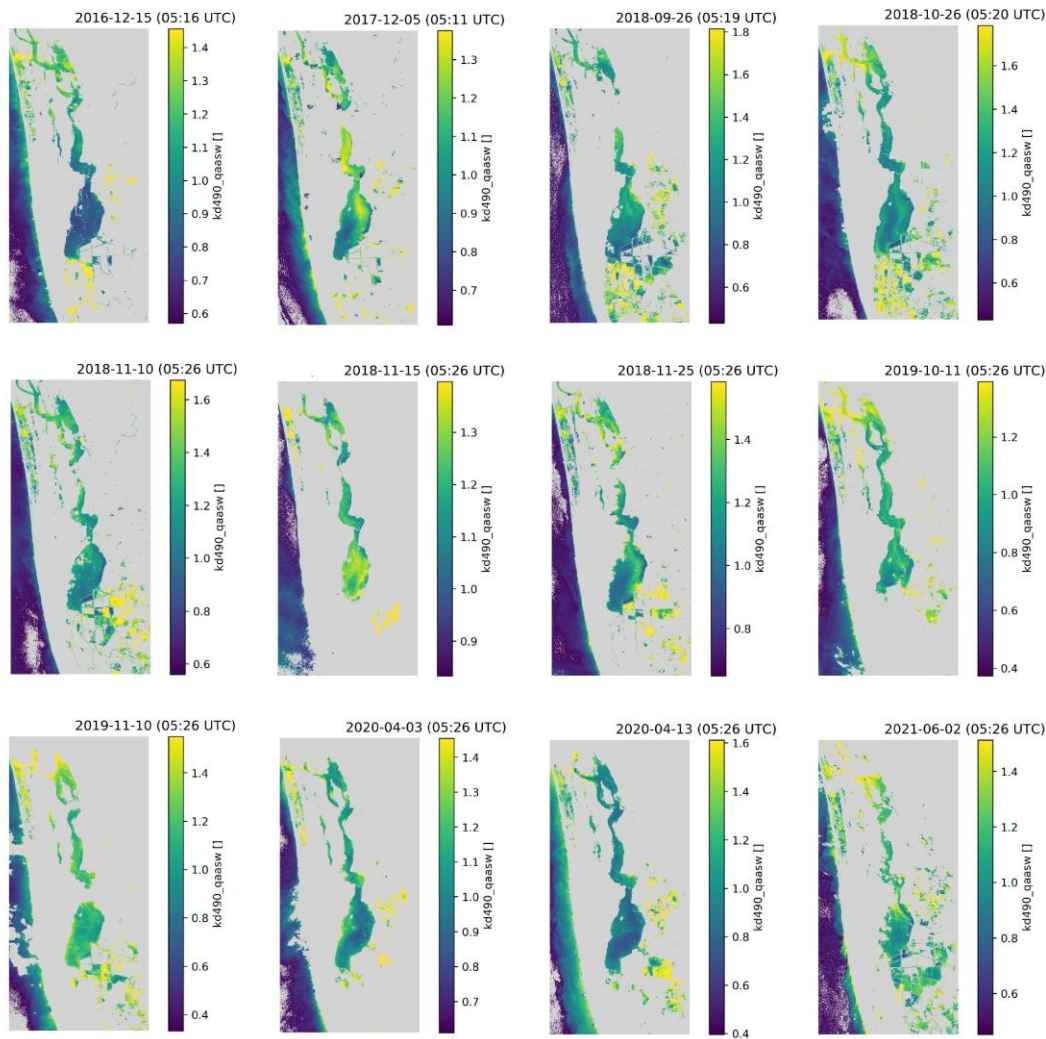
Figure 6.3 Time series of K_d concentration (2016-2021) for the Vembanad Lake

The SPM values decrease during the pre and post-monsoon seasons and reach their peak during the monsoon season. This is due to the increase in the streamflow resulting in high inflow during the monsoon seasons. The results from the analysis carried out in the lake from 2016 to 2021 show several change patterns in the values of SPM due to the floods and the COVID-19 lockdown measures. It is found that there is an average decreasing trend in the SPM level of 15.5 to 13.5 gm^{-3} from the year 2016 to 2021, respectively. The Figure 6.4 shows that the SPM values decrease to the minimum during the post-monsoon season of the post-flood scenarios. After the 2018 floods, the SPM values in the lake's southern part have fallen to 8±1.3 gm^{-3} during September and October (Figure 6.5). It further decreased to less than 5 gm^{-3} during November. The

same trend is found during the post - 2019 floods. The southern part of the lake falls up to 8±2 gm^{-3} during November 2019. During December, these values decrease up to 4±0.7 gm^{-3} in the central and southern regions of the lake. These trends may be due to the flushing action of the lake during the flood resulting in less SPM during the post-monsoon seasons. The effects of the first COVID-19 lockdown decreased the SPM values to 9±1.3 gm^{-3} on April 03, 2020. It further decreased to less than 4 gm^{-3} in the Southern part of the lake by the mid of April (Figure 6.5). These are the lowest recorded SPM values during April in the entire span of the study period. The lockdown imposed due to the second wave of COVID-19 has also influenced the SPM values in the lake up to 7±1.5 gm^{-3} in the southern part of the lake during June. This is the lowest mean value observed during the monsoon season in the lake. The SPM values decreased by 12.1% and 11.4% in 2019 and 2020, respectively, and recorded a 4.3% decrease in 2021 from 2017.
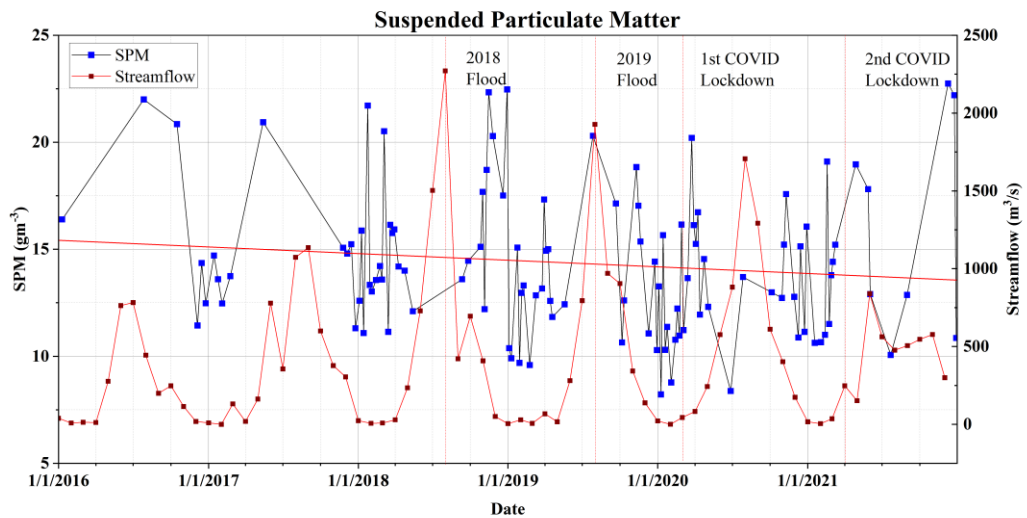
Figure 6.4 Variation of average SPM from 2016 to 2021
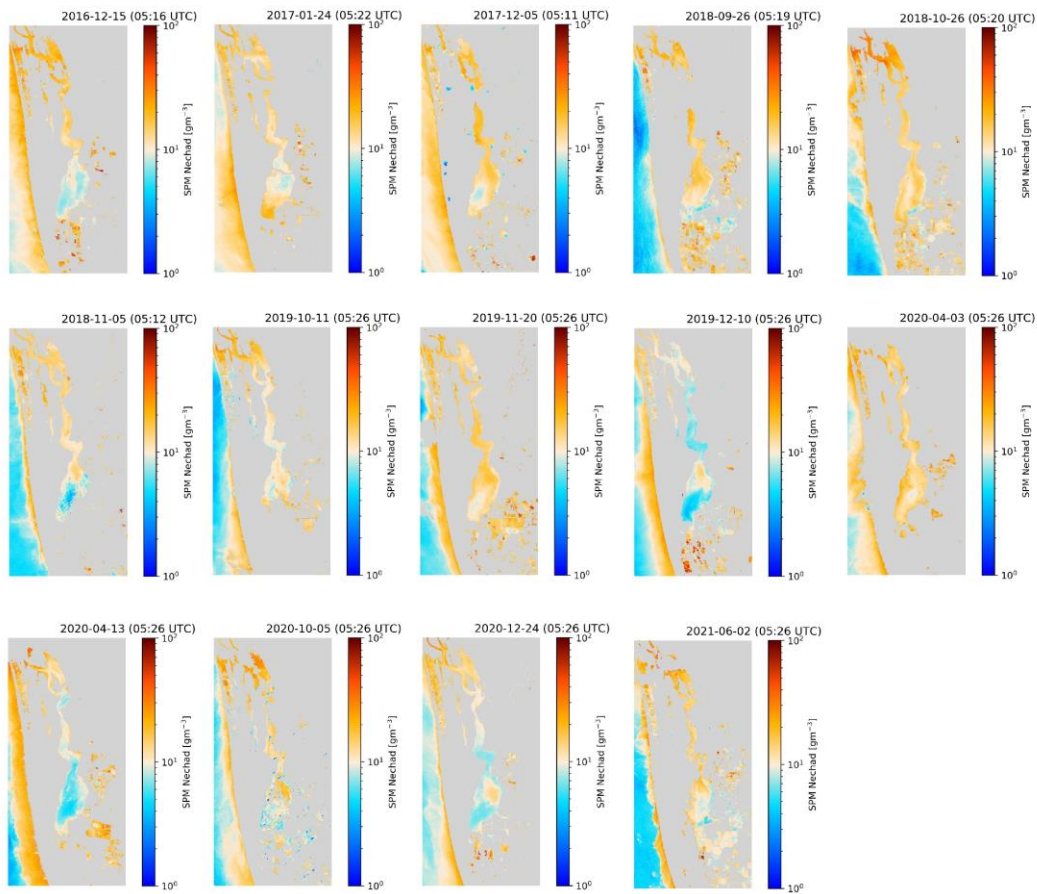
Figure 6.5 Time series of SPM concentration (2016-2021) for the Vembanad Lake

Turbidity rises during the monsoon season due to the large number of sediments transported into the lake by the high inflow from the rivers. The turbidity value decreases during the pre and post-monsoon season as the inflow into the lake is less, and the sediments settle to the bottom of the lake. The scenario of post - 2018 floods shows a considerable decrease in the turbidity values (Figure 6.6) during September, October, and November 2018, where the southern part of the lake falls up to 8±2.2 FNU during September and October, whereas below 7±0.6 FNU in November (Figure 6.7). The post-flood conditions in the year 2019 also resulted in a decrease in the turbidity of the water during the post - monsoon months. There is a considerable decrease in the central and the southern part of the lake up to 8±1.2 FNU during October. By December, the turbidity value further decreased below 7 FNU in the southern part and up to 8±1.4 FNU in the northern and central parts of the lake. It is observed that the effect of the

138

first COVID-19 lockdown resulted in the decrease in turbidity of the region, where the values decreased by less than 6 FNU in the southern part of the lake. The turbidity of the lake is found to be up to 7±2.2 FNU during July, October, and November 2020. During the second COVID-19 lockdown, there was a significant decrease in turbidity values up to 8±0.9 FNU in June 2021. The turbidity values have an overall decrease of about 12.4%, 11.6%, and 4.28% for 2019, 2020, and 2021 respectively, compared to 2017.
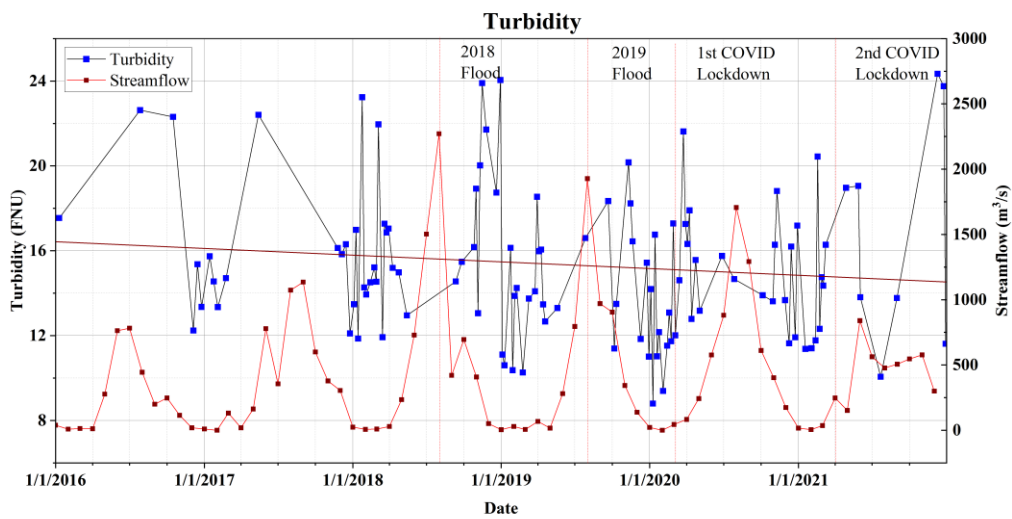
Figure 6.6 Variation of average Turbidity from 2016 to 2021
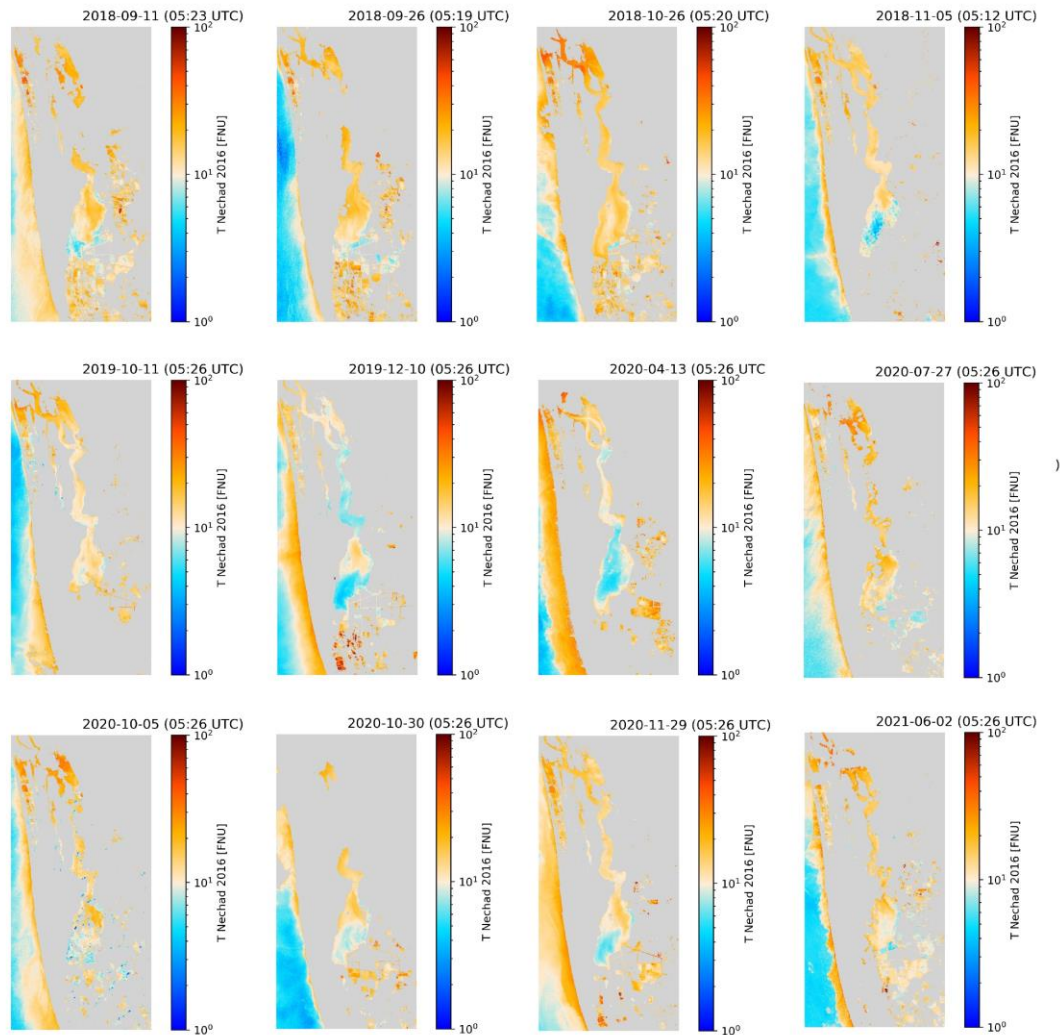
Figure 6.7 Time series of turbidity concentration (2016-2021) for the Vembanad Lake

Chl-a concentration increases during the pre-monsoon months (Figure 6.8), when there is less inflow into the lake resulting in water stagnation. This phenomenon is owing to an increase in HRT in the lake, and the pattern is observed clearly in Figure 6.9 during February 2017, where the Chl-a of the lake increased by about 15±2.1 µg/l in the southern part. During the 2018 floods, there was a considerable increase in the Chl-a concentration due to the runoff generated during the flood flushing the region, which led to the increase in Chl-a concentration in the lake. This phenomenon was observed from September to November 2018 in the southernmost region of the lake, where the Chl-a concentration increased up to 25±3.4 µg/l in those regions. The Chl-a during January 2019 increased up to 30±4 µg/l in the southern part of the lake owing to the

less streamflow from the river. Post-2019 flood shows an abrupt increase in the overall Chl-a values up to 25±1.4 μg/l, and the southernmost part of the lake has reached up to 30±2.1 μg/l during October and November 2019. During January and February 2020, the southernmost region of the lake had an increased Chl-a concentration of up to 40±5.2 μg/l, which may be due to the effect of the 2019 flood. There is no substantial change in the concentration of Chl-a during the first and second COVID-19 lockdown periods. The average Chl-a concentration remains unchanged for 2018, 2019, and 2020. Whereas there is a 12% decrease in the concentration for 2021 compared to the previous years.
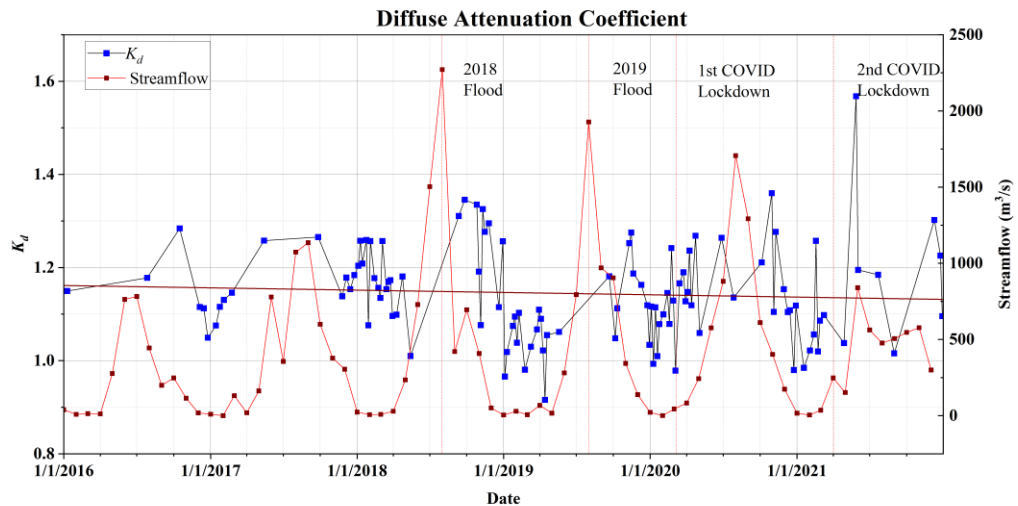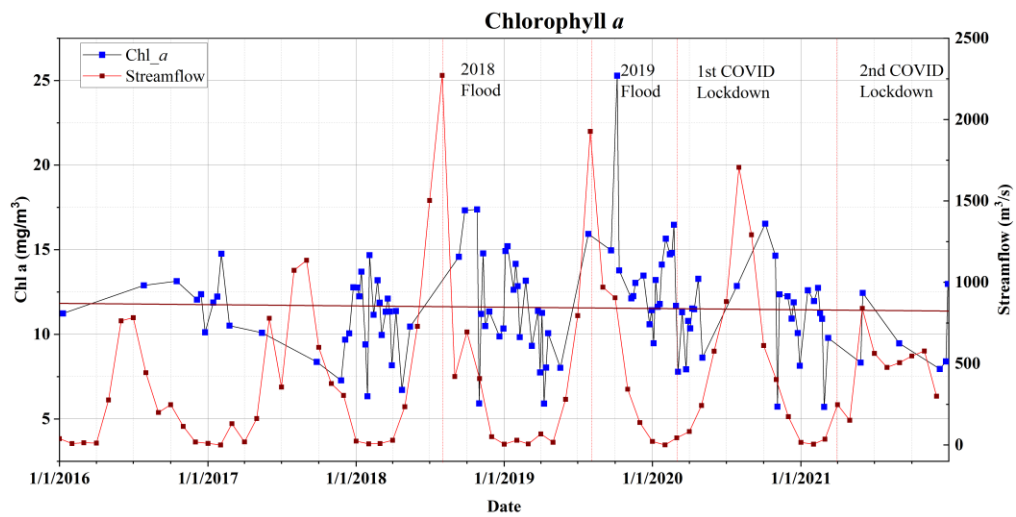

Figure 6.8 Variation of average Chl-a from 2016 to 2021
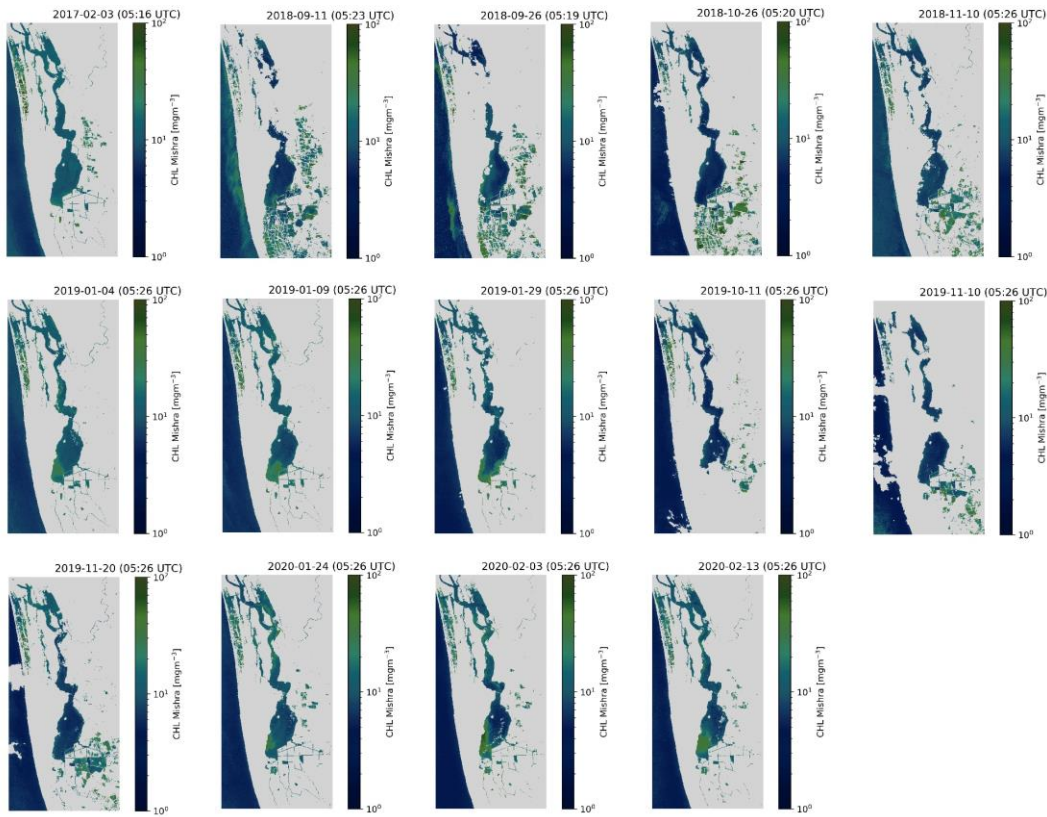
Figure 6.9 Time series of Chl-a concentration (2016-2021) for the Vembanad Lake

6.3.2 Discussion on change in the water quality

The effect of streamflow from the rivers during the monsoon has a direct or indirect relationship with the water quality parameters. The value of K_d and Chl-a increased considerably during the 2018 flood. The increase in K_d may be due to the increase in streamflow into the lake, and the southernmost region of the lake is connected with the agricultural fields, and this may be the prime reason for the increase in the Chl-a where the agricultural residues remain in the lake as the paddy fields adjacent to the southern part of the lake are also flooded due to excessive river streamflow. The values of SPM and turbidity were decreased due to the flushing action in the lake. There are six rivers draining into the lake. River Periyar drains into the northern part of the lake, whereas Muvattupuzha, Meenachil, Manimala, Pamba, and Achencoil drain in the southern part. The river-suspended sediment concentration showed a significant seasonal change in sediment supply, and its variations are strongly related to the hydrodynamics across the wetland (Vinita et al. 2017). As a result, the SPM, turbidity, and K_d values rise

142

dramatically during the monsoon season. Whereas the flushing action of the lake during the flood results in a reduction of these values during the post-flood scenarios. When compared to the main body of the lake, the southernmost region of the lake has very high values of the parameters mentioned above. This may be due to the presence of agricultural fields in the area and the channel pattern, which restricts the free movement of water to flush this part of the lake. The partially treated or untreated effluents from the industries along the banks of the river Periyar and the untreated municipal sewage from Kochi city may account for the lack of the northern half of the lake's noticeable alterations (Kulk et al. 2021). The dredging activities carried out during the first COVID-19 lockdown period in the main shipping channel might have also resulted in the increase in turbidity and SPM values in the northern part of the lake during this time (HCSL 2020). Turbidity increase impacts both top-down and bottom-up processes in the lake system because they render less light accessible, limiting the lake's primary productivity. Raising the inorganic SPM compared to phytoplankton may also lower the nutritive benefits of the zooplankton (Kari et al. 2016). Chl-a is a measure of phytoplankton biomass and controls the purity of the water, which reveals how effectively the lakes are functioning normally. The cHABs spurred on by anthropogenic activities impact the water, causing several environmental issues that worsen the water quality and limit swimming, fishing, and water supplies. A considerable increase of Chl-a during the time of flooding is observed. This could also be caused by an increase in turbidity/SPM by mineral/inorganic particles, as the NDCI algorithm can be affected by non-algal particle backscattering that does not covary with Chl-a (Maciel et al. 2023).

During the first COVID-19 lockdown in 2020, there was a significant decrease in the K_d, SPM, and turbidity values. This trend is observed during the second COVID-19 lockdown in the year 2021, showing a significant decrease in the values of these parameters. The findings indicated that the pollution from industry and tourists had a greater impact on the lake than non-industrial pollution, which continued to exist over this time (Yunus et al. 2020). The region adjacent to the lake is densely populated as it comprises the major cities of Kerala; the resuming of the industries, tourism, and other activities around the lake increased the K_d, SPM, and turbidity values. But Chl-a values are unchanged for both the lockdowns, and this phenomenon was also observed by

Avtar et al. (2020) during the first lockdown in the Vembanad Lake. This might be owing to lower HRT due to river discharge and tidal currents, resulting in poor primary productivity.

The ACOLITE program is chosen for its simplicity of deployment and the ability to rectify intrinsic effects such as sunglint and land pixel adjacency (Pereira-Sandoval et al. 2019). It employs the DSF method for atmospheric correction by default; the exponential extrapolation method can also be employed (Vanhellemont 2019; Vanhellemont and Ruddick 2016b). The ACOLITE algorithm is designed to work in turbid water environments where traditional atmospheric correction algorithms may not perform well. This is because these algorithms assume a homogeneous atmosphere, which is not the case in turbid waters where the presence of water vapour and aerosols can lead to significant variations in atmospheric properties. One limitation of the study is the lack of *in situ* data to verify the results, but the streamflow generated by the rivers is compared and evaluated. Apart from this, several studies both globally (Ansper and Alikas 2019; Caballero et al. 2020, 2022; Dogliotti et al. 2015; Mabit et al. 2022; Nazirova et al. 2021; Nechad et al. 2010; Ngoc et al. 2020; Pereira-Sandoval et al. 2019) and locally (Avtar et al. 2020; Kulk et al. 2021) in the lake, have previously examined these generally used semi-analytical methods for estimating biophysical parameters. The *in situ* data in the lake yielded a better comparison with the ACOLITE processor than the other satellite retrieval algorithms (Avtar et al. 2020; Kulk et al. 2021; Yunus et al. 2020). Even though satellite products were not evaluated using local *in situ* data to address quality and uncertainty, the products utilized are proxies for the biophysical features of the lake. These models emerge as consistent and reliable techniques for a wide range of global applications (Caballero et al. 2022; Nechad et al. 2009, 2016). Even in confined inlets and ports, Sentinel-2's red (665 nm) and NIR (842 nm) bands at 10 m spatial resolution enable the recovery of turbid or SPM concentration. Sentinel-2's incorporation of a 20 m resolution red edge band at about 705 nm, which overcomes the limitations of open ocean blue-green ratio algorithms, is one of its key benefits over Landsat-8 in terms of determining Chl-a absorption in turbid and productive waters. Previous studies (Sravanthi et al. 2013; Vinita et al. 2017; Yunus et al. 2020) done in the lake region showed the concentration of SPM in the lake as <110 gm^{-3}, considering the range it is assumed that the model (Nechad et al. 2010) can retrieve the SPM values

144

with an overall Root Mean Square (RMS) error of less than 10 gm^{-3}. Although the *in situ* readings have uncertainties, which may be fairly considerable, they are always assumed in remote sensing investigations to represent the truth (Toming et al. 2017). Pereira-Sandoval et al. (2019) emphasized the use of ACOLITE in Case II waters, as their performance is good in the visible spectrum of the region. Additionally, Vanhellemont and Ruddick (2018) state that the DSF algorithm in ACOLITE performed well in murky coastal waters. One major benefit of remote sensing data is that it allows for the examination of previous water quality as well as changes in water quality during times of flood or a lockdown without the need for *in situ* data, with the sole need that the data collected for the analysis be cloud-free.

Thus, the overall decreasing trend in K_d, SPM, and turbidity indicates an improvement in the lake's water quality over the past six years, which may benefit the lake's dependent flora and fauna as well as the increase in phytoplankton photosynthesis. But the Chl-a also shows a slightly decreasing trend in the past six-year analysis, which shows the various other factors that influence the growth in the phytoplankton (Avtar et al. 2020; Kulk et al. 2021). Some important reasons that may reduce the decrease in the phytoplankton growth are due to the increase in the surface temperature of the lake, which has been higher in the past few decades, the flushing action of the lake during the flood, and the anthropogenic activities resulting in the eutrophication of the lake.

6.4   CLOSURE

The bio-optical algorithms have been analysed using the Sentinel-2 images in the ACOLITE software to evaluate the water quality of the Vembanad Lake and its temporal changes during the floods and the COVID-19 lockdown. The parameter K_d shows a substantial decrease from 2016 to 2021, accounting for an overall decrease of 2.5%. The values of SPM and turbidity also show an overall decrease of about 4.3% and 4.28% from 2016 to 2021. There was no substantial change in the Chl-a content from 2016 to 2020. For 2021, there is an overall decrease of 12% in the Chl-a content compared to the other years. From the study, it is observed that the flooding and lockdown decreased the values of K_d, SPM, and turbidity. It is anticipated that the values of Chl-a content will eventually rise with an improvement in water quality, but in 2021, they began to fall. Though the precise cause of the Chl-a decline could not be

determined, previous studies in the lake suggested that the Chl-a content declines due to a reduction in HRT. To develop a long-term solution for the lake environment, precise standards based on the framework must be followed within the allocated period because the Ramsar Convention defines the lake as being of international importance. The socioeconomic position of the inhabitants who depend on the lake could be impacted if a long-term solution is not found to decrease the environmental harm caused to the lake, as it is highly dependent on tourism.

CHAPTER 7

SUMMARY AND CONCLUSION

Water resources in coastal regions and wetlands are incredibly significant since any development activity will rely heavily on the availability of fresh water to support household, industrial, and agricultural needs. However, owing to the overexploitation of water resources, discharging wastewater into the lake, and sea level rise, fresh water supplies in these wetlands are anticipated to suffer devastating and permanent consequences in the future years. The construction of the Thanneermukkom barrage in the lake led to the intervention, disrupting the physical and biological continuity of the lake with coastal waters. This has resulted in the decline of fish production and species diversity of the lake. The area around the lake is heavily populated since the major cities in the state are situated nearby. VLS is also severely affected by floods in the past few years due to the unprecedented rainfall during the monsoon seasons.

As a result, this study is being conducted to map the change in LULC and predict the projected LULC. The mapping of flood-inundated areas, frequently flood-prone areas, and the possible impact of floods on projected LULCs that inundate the area are analysed. The spatio-temporal change in lake water quality, its consequences during flooded seasons, and the COVID-19 lockdowns are investigated.

7.1 LULC CHANGE DETECTION AND PREDICTION

As urbanization in developing countries becomes uncontrollable and unsustainable, determining the spatial and urban development patterns is a crucial problem for achieving a sustainable environment. As a result, there is a need to raise public consciousness about the dramatic improvements and degradation of natural environments. In this regard, the research helped to examine the transition of the urban environment. Since the study area was seriously affected by the floods in 2018, 2019, and 2021, this research will help society understand the LULC changes due to anthropogenic and flood-related factors. This will assist planners in making the requisite precautionary plans for the current and potential urban growth trends for sustainable development.

GEE has a powerful scripting language that works in tandem with its cloud infrastructure and user-friendly API. It aided in the efficient study of the classification based on ML. From the analysis, the RF algorithm outperformed the CART and SVM models in accuracy. This demonstrates the RF algorithm's ability to tackle multiclass classification and outperform the other two models. The ML application in the CA-Markov model of Idrisi TerrSet software helps to analyse the spatio-temporal change prediction of the LULC using the LCM module. Except for barren land, this model had more than 94% accuracy for all LULC classes. The transition study in the LULC groups reveals a dramatic growth in the built-up area and barren land and a steady decline in vegetation and forest cover.

7.2 FLOOD ANALYSIS AND MODELLING

The primary ideas of flood susceptibility mapping are risk assessment and primary prevention measures. Evidence currently reveals that unrestricted land-use planning and policy execution, as well as climate change, have dramatically increased flood occurrences worldwide. The focus of risk evaluation and prevention efforts focus no longer on controlling floods but on local governments' obligations to lessen flood impacts. Residents in flood-prone areas should be warned about the hazards and possibilities. Land-use planners and government entities are obligated to inform local communities about the most recent flood susceptibility evaluations and the rules prohibiting new projects in areas with a high risk of flooding. As a result, this study will be extremely useful to the government and non-governmental organizations in terms of risk assessment and mitigation during times of crisis, as well as early people rescue and worthwhile preparation.

The mapping of the flood-inundated region demonstrates the ability of SAR images to penetrate clouds and map during disasters. It also demonstrates the evolution of remote sensing and its ability to monitor temporally, even during critical times. The 2018 flood scenario is utilized to understand the likely flood affecting the region in future predicted LULCs. By the year 2050, nearly 23.32 km^2 of the region, representing a 58.4% increase in the built-up area compared to 2018, will be inundated by floods of the same intensity as that of 2018. The flood inventory map is prepared from the flooded data of 2018, 2019, and 2021. The flooded regions are delineated using the Sentinel-1 SAR imagery in the GEE platform by the image ratio technique and applying the threshold

for the pre and during-flood scenario. The flood susceptibility map is generated for the VLS using the four decision tree-based algorithms: AdaBoost, RF, GBM, and XGBoost. Based on the research results, 30-40% of the region falls under high and very high vulnerable regions to flooding. This is a concerning range, given that the study includes some of the most urbanized areas of Kerala, including Ernakulam, Alappuzha, and Kottayam, which have population densities of 1072, 1504, and 895 km^2, respectively.

7.3 OPTICAL PROPERTY OF THE LAKE

The Sentinel-2 has aided in the continuous monitoring of the lake over the last seven years. With the launch of Sentinel-2B, the temporal resolution of the satellite has been cut in half because the satellites are monitoring at 180° out of phase. The only drawback of multispectral imagery is that it cannot pass through clouds. Because of this, there is very little data availability during the monsoon and post-monsoon seasons, when the region experiences heavy rainfall. Henceforth, remote sensing can aid open-source data in continuously observing the lake's water quality changes without collecting *in situ* data during natural or man-made crises.

The ACOLITE software was selected because of its ease of deployment and capability to correct inherent issues like sunglint and pixel adjacency on land. Thus, the effect of change in water quality during various seasons and during the flood and COVID-19 lockdowns are attempted to understand. The Vembanad Lake is one of the most severely polluted lakes in the country. The analysis revealed a 2.5% decrease in K_d values between 2016 and 2021. From 2016 to 2021, the SPM and turbidity decreased by 4.3% and 4.28%, respectively. The drop in K_d, SPM, and turbidity shows improved water quality over the years, observed due to the combined effects of the floods and the COVID lockdowns. The Chl-a concentration remained unchanged from the year 2016 to 2020. In contrast, it drops by 12% during the year 2021. This could be due to lower HRT due to river discharge and tidal currents, decreasing primary productivity. This demonstrates that continuous monitoring of the water quality and taking precautionary measures in the lake will aid in achieving the SDG, which promotes the recovery, protective measures, and sustainable management of marine and aquatic ecosystems. Thus, this study provides insights to the planners, government, and NGOs about the

change in the water quality and can serve as a valuable reference during strategic planning.

7.4   LIMITATIONS

- The increase in the influence of anthropogenic activities and natural disasters may result in a rapid change in LULC, which will not be accounted for during the prediction. This is because spatio-temporal variation is limited to the current trend of transition in the LULC, which can differ over time.

- In the CA-Markov model, only two LULC maps can be used as a base map to forecast possible LULC.

- The multispectral images suffer from severe cloud cover because the study area experiences rainfall during the southwest and northeast monsoons. This poses a potential barrier to continuous monitoring of changes in the lake's water quality.

7.5   SCOPE OF FUTURE WORK

- The classification using Sentinel-2 may help generate the fine-resolution LULC map, but the Sentinel-2 collection of satellite data is not available for more than seven years.

- Rainfall-runoff modelling coupled with the optical properties of the lake may also provide more insight into the lake's water quality change.

- Integrating the influence of climate change into the projected LULC data for the analysis of flood susceptibility has the potential to facilitate the development of future flood susceptibility models.

REFERENCES

Abedi, R., Costache, R., Shafizadeh-Moghadam, H., and Pham, Q. B. (2022). "Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees." *Geocarto Int.*, 37(19), 5479–5496.

Abijith, D., and Saravanan, S. (2022). "Assessment of land use and land cover change detection and prediction using remote sensing and CA Markov in the northern coastal districts of Tamil Nadu, India." *Environ. Sci. Pollut. Res.*, 29(57), 86055–86067.

Abijith, D., Saravanan, S., Jennifer, J. J., Parthasarathy, K. S. S., Singh, L., and Sankriti, R. (2021). "Assessing the impact of damage and government response toward the cyclone Gaja in Tamil Nadu, India." *Disaster Resil. Sustain.*, Elsevier, 577–590.

Abijith, D., Saravanan, S., Jennifer, J. J., Singh, L., Saranya, T., Sankriti, R., Selvaraj, A., and Parthasarathy, K. S. S. (2022). "GIS Based Groundwater Potential Zone Identification Using AHP for Ponnaniyaru Watershed, Tamil Nadu, India." *Groundw. Water Qual. Water Sci. Technol. Libr.*, R. Jha, V. P. Singh, V. Singh, L. B. Roy, and R. Thendiyath, eds., Springer, Cham, 313–323.

Abijith, D., Saravanan, S., Singh, L., Jennifer, J. J., Saranya, T., and Parthasarathy, K. S. S. (2020). "GIS-based multi-criteria analysis for identification of potential groundwater recharge zones - a case study from Ponnaniyaru watershed, Tamil Nadu, India." *HydroResearch*, 3, 1–14.

Abraham, A., and Kundapura, S. (2022). "Evaluating the long-term trends of the climatic variables over three humid tropical basins in Kerala, India." *Arab. J. Geosci.*, 15, 811.

Abrishamchi, A., Dashti, M., Alamdari, N., and Salavitabar, A. (2011). "A GIS-Google Earth Based Approach to Estimating the Flood Damage Function in Large River Basins." *World Environ. Water Resour. Congr. 2011 Bear. Knowl. Sustain. - Proc. 2011 World Environ. Water Resour. Congr.*, 3811–3821.

Aburas, M. M., Abdullah, S. H., Ramli, M. F., Ash'aari, Z. H., and Ahamad, M. S. S. (2018). "Simulating and monitoring future land-use trends using CA-Markov and LCM models." *IOP Conf. Ser. Earth Environ. Sci.*, 169(1), 012050.

Abuzar, M., Sheffield, K., Whitfield, D., O'Connell, M., and McAllister, A. (2014). "Comparing Inter-Sensor NDVI for the Analysis of Horticulture Crops in South-

Eastern Australia." *Am. J. Remote Sens.*, 2(1).

Adam, E., Mutanga, O., Odindi, J., and Abdel-Rahman, E. M. (2014). "Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers." *Int. J. Remote Sens.*, 35(10), 3440–3458.

Adelabu, S., Mutanga, O., and Adam, E. (2014). "Evaluating the impact of red-edge band from Rapideye image for classifying insect defoliation levels." *ISPRS J. Photogramm. Remote Sens.*, 95, 34–41.

Adhikari, U., Nejadhashemi, A. P., and Woznicki, S. A. (2015). "Climate change and eastern Africa: a review of impact on major crops." *Food Energy Secur.*, 4(2), 110–132.

Aerts, J. C. J. H., Botzen, W. J., Clarke, K. C., Cutter, S. L., Hall, J. W., Merz, B., Michel-Kerjan, E., Mysiak, J., Surminski, S., and Kunreuther, H. (2018). "Integrating human behaviour dynamics into flood disaster risk assessment." *Nat. Clim. Chang.*, 8(3), 193–199.

Agarwal, S., and Nagendra, H. (2019). "Classification of Indian cities using Google Earth Engine." *J. Land Use Sci.*, 14(4–6), 425–439.

Ahmed, U., Mumtaz, R., Anwar, H., Mumtaz, S., and Qamar, A. M. (2020). "Water quality monitoring: from conventional to emerging technologies." *Water Supply*, 20(1), 28–45.

Aissia, M. A. B., Chebana, F., Ouarda, T. B. M. J., Roy, L., Desrochers, G., Chartier, I., and Robichaud, É. (2012). "Multivariate analysis of flood characteristics in a climate change context of the watershed of the Baskatong reservoir, Province of Québec, Canada." *Hydrol. Process.*, 26(1), 130–142.

Aitkenhead, M. J., and Aalders, I. H. (2009). "Predicting land cover using GIS, Bayesian and evolutionary algorithm methods." *J. Environ. Manage.*, 90(1), 236–250.

AlDousari, A. E., Kafy, A. Al, Saha, M., Fattah, M. A., Bakshi, A., and Rahaman, Z. A. (2023). "Summertime Microscale Assessment and Prediction of Urban Thermal Comfort Zone Using Remote-Sensing Techniques for Kuwait." *Earth Syst. Environ.*, 7(2), 435–456.

Alem, A. El, Lhissou, R., Chokmani, K., and Oubennaceur, K. (2021). "Remote Retrieval of Suspended Particulate Matter in Inland Waters: Image-Based or Physical

Atmospheric Correction Models?" *Water*, 13(16), 2149.

Alexander, K., Hettiarachchi, S., Ou, Y., and Sharma, A. (2019). "Can integrated green spaces and storage facilities absorb the increased risk of flooding due to climate change in developed urban environments?" *J. Hydrol.*, 579, 124201.

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., Roo, A. de, Salamon, P., Wyser, K., and Feyen, L. (2017). "Global projections of river flood risk in a warmer world." *Earth's Futur.*, 5(2), 171–182.

Ali, R., Kuriqi, A., Abubaker, S., and Kisi, O. (2019). "Long-Term Trends and Seasonality Detection of the Observed Flow in Yangtze River Using Mann-Kendall and Sen's Innovative Trend Method." *Water*, 11(9), 1855.

Álvarez-Cabria, M., Barquín, J., and Peñas, F. J. (2016). "Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors." *Sci. Total Environ.*, 545, 152–162.

Anderson, J. T., and Davis, C. A. (2013). *Applications and Management. Wetl. Tech.*

Andres, L., Boateng, K., Borja-Vega, C., and Thomas, E. (2018). "A Review of In-Situ and Remote Sensing Technologies to Monitor Water and Sanitation Interventions." *Water*, 10(6), 756.

Andrić, J. M., and Lu, D. G. (2017). "Fuzzy methods for prediction of seismic resilience of bridges." *Int. J. Disaster Risk Reduct.*, 22, 458–468.

Ansari, A., and Golabi, M. H. (2019). "Prediction of spatial land use changes based on LCM in a GIS environment for Desert Wetlands – A case study: Meighan Wetland, Iran." *Int. Soil Water Conserv. Res.*, 7(1), 64–70.

Ansper, A., and Alikas, K. (2019). "Retrieval of chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes." *Remote Sens.*, 11(1).

Antonini, K., Langer, M., Farid, A., and Walter, U. (2017). "SWEET CubeSat – Water detection and water quality monitoring for the 21st century." *Acta Astronaut.*, 140, 10–17.

Avtar, R., Kumar, P., Supe, H., Jie, D., Sahu, N., Mishra, B. K., and Yunus, A. P. (2020). "Did the COVID-19 Lockdown-Induced Hydrological Residence Time Intensify the Primary Productivity in Lakes? Observational Results Based on Satellite Remote Sensing." *Water*, 12(9), 2573.

Aydin, H. E., and Iban, M. C. (2023). "Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations." *Nat. Hazards*, 116, 2957–2991.

Balasubramanian, S. V., Pahlevan, N., Smith, B., Binding, C., Schalles, J., Loisel, H., Gurlin, D., Greb, S., Alikas, K., Randla, M., Bunkei, M., Moses, W., Nguyễn, H., Lehmann, M. K., O'Donnell, D., Ondrusek, M., Han, T. H., Fichot, C. G., Moore, T., and Boss, E. (2020). "Robust algorithm for estimating total suspended solids (TSS) in inland and nearshore coastal waters." *Remote Sens. Environ.*, 246, 111768.

Balica, S. F., Wright, N. G., and Meulen, F. van der. (2012). "A flood vulnerability index for coastal cities and its use in assessing climate change impacts." *Nat. Hazards*, 64(1), 73–105.

Band, S. S., Janizadeh, S., Pal, S. C., Saha, A., Chakrabortty, R., Melesse, A. M., and Mosavi, A. (2020). "Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms." *Remote Sens.*, 12(21), 3568.

Barducci, A., Guzzi, D., Marcoionni, P., and Pippi, I. (2009). "Aerospace wetland monitoring by hyperspectral imaging sensors: A case study in the coastal zone of San Rossore Natural Park." *J. Environ. Manage.*, 90(7), 2278–2286.

Belgiu, M., and Drăgu, L. (2016). "Random forest in remote sensing: A review of applications and future directions." *ISPRS J. Photogramm. Remote Sens.*, 114, 24–31.

Bellman, R. (1957). *Dynamic Programming*. Princeton, New Jersey: Princeton University Press.

Beven, K. J., and Kirkby, M. J. (1979). "A physically based, variable contributing area model of basin hydrology." *Hydrol. Sci. J.*, 24(1), 43–69.

Bhattacharjee, K., and Behera, B. (2017). "Forest cover change and flood hazards in India." *Land use policy*, 67, 436–448.

Bhuyan, M., Jayaram, C., Menon, N. N., and Joseph, K. A. (2020). "Satellite-Based Study of Seasonal Variability in Water Quality Parameters in a Tropical Estuary along the Southwest Coast of India." *J. Indian Soc. Remote Sens.*, 48, 1265–1276.

Boehmke, B., and Greenwell, B. (2019). *Hands-On Machine Learning with R. Hands-On Mach. Learn. with R*, New York: Chapman and Hall/CRC.

Bose, A., and Chowdhury, I. R. (2020). "Monitoring and modeling of spatio-temporal

urban expansion and land-use/land-cover change using markov chain model: a case study in Siliguri Metropolitan area, West Bengal, India." *Model. Earth Syst. Environ.*, 6(4), 2235–2249.

Bott, R. (1994). "Convention on Wetlands of International Importance especially as Waterfowl Habitat." *Ramsar*, (1), 1–5.

Breiman, L. (2001). "Random forests." *Mach. Learn.*, 45(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification And Regression Trees. Classif. Regres. Trees*, Boca Raton: Routledge.

Brij, G., and Krishnamurthy, K. (1993). "Wetlands of South Asia." *Wetl. World*, 345–414.

Bubeck, P., and Thieken, A. H. (2018). "What helps people recover from floods? Insights from a survey among flood-affected residents in Germany." *Reg. Environ. Chang.*, 18(1), 287–296.

Bui, D. T., Khosravi, K., Shahabi, H., Daggupati, P., Adamowski, J. F., M.Melesse, A., Pham, B. T., Pourghasemi, H. R., Mahmoudi, M., Bahrami, S., Pradhan, B., Shirzadi, A., Chapi, K., and Lee, S. (2019). "Flood Spatial Modeling in Northern Iran Using Remote Sensing and GIS: A Comparison between Evidential Belief Functions and Its Ensemble with a Multivariate Logistic Regression Model." *Remote Sens.*, 11(13), 1589.

Buiteveld, H., Hakvoort, J. H. M., and Donze, M. (1994). "Optical properties of pure water." *Ocean Opt. XII Proc Soc Photoopt Inst Eng*, 2258, 174–183.

Caballero, I., Fernández, R., Escalante, O. M., Mamán, L., and Navarro, G. (2020). "New capabilities of Sentinel-2A/B satellites combined with in situ data for monitoring small harmful algal blooms in complex coastal waters." *Sci. Rep.*, 10(1), 8743.

Caballero, I., Román, A., Tovar-Sánchez, A., and Navarro, G. (2022). "Water quality monitoring with Sentinel-2 and Landsat-8 satellites during the 2021 volcanic eruption in La Palma (Canary Islands)." *Sci. Total Environ.*, 822, 153433.

Cameron, D., Beven, K., and Naden, P. (2000). "Flood frequency estimation by continuous simulation under climate change (with uncertainty)." *Hydrol. Earth Syst. Sci.*, 4(3), 393–405.

Cánovas-García, F., Alonso-Sarría, F., Gomariz-Castillo, F., and Oñate-Valdivieso, F. (2017). "Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery." *Comput. Geosci.*, 103, 1–11.

Carmichael, W. W., and Boyer, G. L. (2016). "Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes." *Harmful Algae*, 54, 194–212.

Carreño Conde, F., and Mata Muñoz, M. De. (2019). "Flood Monitoring Based on the Study of Sentinel-1 SAR Images: The Ebro River Case Study." *Water*, 11(12), 2454.

CBD. (1992). "Convention on biological diversity." *Estud. Avançados*, 6(15), 193–233.

Census. (2011). "Census of India." <https://censusindia.gov.in/2011-Common/CensusData2011.html> (Aug. 15, 2021).

Chakrabortty, R., Chandra Pal, S., Rezaie, F., Arabameri, A., Lee, S., Roy, P., Saha, A., Chowdhuri, I., and Moayedi, H. (2022). "Flash-flood hazard susceptibility mapping in Kangsabati River Basin, India." *Geocarto Int.*, 37(23), 6713–6735.

Chander, G., Markham, B. L., and Helder, D. L. (2009). "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors." *Remote Sens. Environ.*, 113(5), 893–903.

Chen, J., Li, Y., Zhou, W., Iqbal, J., and Cui, Z. (2016). "Debris-Flow Susceptibility Assessment Model and Its Application in Semiarid Mountainous Areas of the Southeastern Tibetan Plateau." *Nat. Hazards Rev.*, 18(2).

Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 785–794.

Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., Pradhan, B., and Ahmad, B. Bin. (2020). "Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods." *Sci. Total Environ.*, 701, 134979.

Chen, X., and Jeong, J. C. (2007). "Enhanced recursive feature elimination." *Sixth Int. Conf. Mach. Learn. Appl. (ICMLA 2007)*, 429–435.

Chowdhuri, I., Pal, S. C., and Chakrabortty, R. (2020). "Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India." *Adv. Sp. Res.*, 65(5), 1466–1489.

Codd, G. A., Morrison, L. F., and Metcalf, J. S. (2005). "Cyanobacterial toxins: risk management for health protection." *Toxicol. Appl. Pharmacol.*, 203(3), 264–272.

Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Educ. Psychol. Meas.*, 20(1), 37–46.

Cortes, C., and Vapnik, V. (1995). "Support-vector networks." *Mach. Learn. 1995 203*, 20(3), 273–297.

Costache, R., Arabameri, A., Moayedi, H., Pham, Q. B., Santosh, M., Nguyen, H., Pandey, M., and Pham, B. T. (2021). "Flash-flood potential index estimation using fuzzy logic combined with deep learning neural network, naïve Bayes, XGBoost and classification and regression tree." *Geocarto Int.*, 37(23), 6780–6807.

CWC. (2018). "Kerala Floods of August 2018." (August), 46.

Darabi, H., Rahmati, O., Naghibi, S. A., Mohammadi, F., Ahmadisharaf, E., Kalantari, Z., Torabi Haghighi, A., Soleimanpour, S. M., Tiefenbacher, J. P., and Tien Bui, D. (2022). "Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood." *Geocarto Int.*, 37(19), 5716–5741.

Darem, A. A., Alhashmi, A. A., Almadani, A. M., Alanazi, A. K., and Sutantra, G. A. (2023). "Development of a map for land use and land cover classification of the Northern Border Region using remote sensing and GIS." *Egypt. J. Remote Sens. Sp. Sci.*, 26(2), 341–350.

Darst, B. F., Malecki, K. C., and Engelman, C. D. (2018). "Using recursive feature elimination in random forest to account for correlated variables in high dimensional data." *BMC Genomic Data*, 19, 65.

Das, S. (2019). "Geospatial mapping of flood susceptibility and hydro-geomorphic response to the floods in Ulhas basin, India." *Remote Sens. Appl. Soc. Environ.*, 14, 60–74.

Degenhardt, F., Seifert, S., and Szymczak, S. (2019). "Evaluation of variable selection methods for random forests and omics data sets." *Brief. Bioinform.*, 20(2), 492–503.

Dev, P. J., Sukenik, A., Mishra, D. R., and Ostrovsky, I. (2022). "Cyanobacterial pigment concentrations in inland waters: Novel semi-analytical algorithms for multi- and hyperspectral remote sensing data." *Sci. Total Environ.*, 805, 150423.

Dewan, A. M., Kankam-Yeboah, K., and Nishigaki, M. (2006). "Using Synthetic Aperture Radar (SAR) Data for Mapping River Water Flooding in an Urban Landscape: A Case Study of Greater Dhaka, Bangladesh." *J. Japan Soc. Hydrol. Water Resour.*, 19(1), 44–54.

Dodangeh, E., Choubin, B., Eigdir, A. N., Nabipour, N., Panahi, M., Shamshirband, S., and Mosavi, A. (2020). "Integrated machine learning methods with resampling

algorithms for flood susceptibility prediction." *Sci. Total Environ.*, 705, 135983.

Dogliotti, A. I., Ruddick, K. G., Nechad, B., Doxaran, D., and Knaeps, E. (2015). "A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters." *Remote Sens. Environ.*, 156, 157–168.

Dou, P., and Chen, Y. (2017). "Remote sensing imagery classification using AdaBoost with a weight vector (WV AdaBoost)." *Remote Sens. Lett.*, 8(8), 733–742.

Doxaran, D., Froidefond, J. M., Lavender, S., and Castaing, P. (2002). "Spectral signature of highly turbid waters: Application with SPOT data to quantify suspended particulate matter concentrations." *Remote Sens. Environ.*, 81(1), 149–161.

Ekmekcioğlu, Ö., Koc, K., and Özger, M. (2021). "District based flood risk assessment in Istanbul using fuzzy analytical hierarchy process." *Stoch. Environ. Res. Risk Assess.*, 35(3), 617–637.

El-Magd, S. A. A., Pradhan, B., and Alamri, A. (2021). "Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita and surroundings, Central Eastern Desert, Egypt." *Arab. J. Geosci.*, 14, 323.

Elias, T., Geetha, T., and Shirlal, K. G. (2022). "Effect of armour unit layers and placement mode in the determination of stability of geotextile sand container (GSC) breakwaters." *Geotext. Geomembranes*, 50(3), 444–454.

Elias, T., Shirlal, K. G., and E.V., K. (2021). "Physical model studies on damage and stability analysis of breakwaters armoured with geotextile sand containers." *Geotext. Geomembranes*, 49(3), 604–618.

ESA. (2000). *Sentinel: User-guides*.

Faichia, C., Tong, Z., Zhang, J., Liu, X., Kazuva, E., Ullah, K., and Al-Shaibah, B. (2020). "Using rs data-based ca–markov model for dynamic simulation of historical and future lucc in Vientiane, Laos." *Sustain.*, 12(20), 1–20.

Falconer, I. R. (1999). "An Overview of problems caused by toxic blue–green algae (cyanobacteria) in drinking and recreational water." *Environ. Toxicol.*, 14, 5–12.

Fang, Z., Wang, Y., Peng, L., and Hong, H. (2021). "Predicting flood susceptibility using LSTM neural networks." *J. Hydrol.*, 594, 125734.

Farda, N. M., Danoedoro, P., Hartono, and Harjoko, A. (2016). "Image Mining in Remote Sensing for Coastal Wetlands Mapping: from Pixel Based to Object Based Approach." *IOP Conf. Ser. Earth Environ. Sci.*, 47(1), 012002.

Felix, A. Y., and Sasipraba, T. (2019). "Flood Detection Using Gradient Boost Machine Learning Approach." *Int. Conf. Comput. Intell. Knowl. Econ. ICCIKE 2019*, UAE: IEEE Xplore, 779–783.

Finlayson, C. M. (2003). "The challenge of integrating wetland inventory, assessment and monitoring." *Aquat. Conserv. Mar. Freshw. Ecosyst.*, 13(4), 281–286.

Flax, L. K., Jackson, R. W., and Stein, D. N. (2002). "Community Vulnerability Assessment Tool Methodology." *Nat. Hazards Rev.*, 3(4), 163–176.

Flood Damage Statistics. (2018). *State Wise and for the Country as a Whole for the Period 1953 to 2016. Central Water Commission, Government of India*.

Forget, P., and Ouillon, S. (1998). "Surface suspended matter off the Rhone river mouth from visible satellite imagery." *Oceanol. Acta*, 21(6), 739–749.

Forghani, A., Cechet, B., and Nadimpalli, K. (2007). "Object-based classification of multi-sensor optical imagery to generate terrain surface roughness information for input to wind risk simulation." *IEEE Int. Geosci. Remote Sens. Symp.*, 3090–3095.

Freund, Y., and Schapire, R. E. (1997). "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." *J. Comput. Syst. Sci.*, 55(1), 119–139.

Friedman, J. H. (2000). "Greedy function approximation: A gradient boosting machine." *Ann. Stat.*, 29(5), 1189–1232.

Fu, X., Wang, X., and Yang, Y. J. (2018). "Deriving suitability factors for CA-Markov land use simulation model based on local historical data." *J. Environ. Manage.*, 206, 10–19.

Garg, V., Aggarwal, S. P., and Chauhan, P. (2020). "Changes in turbidity along Ganga River using Sentinel-2 satellite data during lockdown associated with COVID-19." *Geomatics, Nat. Hazards Risk*, 11(1), 1175–1195.

Garrote, J., Peña, E., and Díez-herrero, A. (2021). "Probabilistic Flood Hazard Maps from Monte Carlo Derived Peak Flow Values—An Application to Flood Risk Management in Zamora City (Spain)." *Appl. Sci.*, 11(14), 6629.

Gbambie, A. S. B., Poulin, A., Boucher, M. A., and Arsenault, R. (2017). "Added Value of Alternative Information in Interpolated Precipitation Datasets for Hydrology." *J. Hydrometeorol.*, 18(1), 247–264.

Gebrechorkos, S. H., Bernhofer, C., and Hülsmann, S. (2020). "Climate change impact assessment on the hydrology of a large river basin in Ethiopia using a local-scale

climate modelling approach." *Sci. Total Environ.*, 742, 140504.

George, S. L., Balasubramani, K., Shekhar, S., Venkatesham, E., Prasad, K. A., Grover, A., Mahata, D., Kumar, A., Ashique, V. V., Libina, R. S., Swaminathan, D. R., Balasundareshwaran, A., and Annaidasan, K. (2021). "A comprehensive study on preparedness, impacts, response and recovery from tropical severe cyclonic storm 'GAJA': lessons for the future." *J. Coast. Conserv.*, 25(6), 1–13.

George, S. L., Kantamaneni, K., V, R. A., Prasad, K. A., Shekhar, S., Panneer, S., Rice, L., and Balasubramani, K. (2022). "A Multi-Data Geospatial Approach for Understanding Flood Risk in the Coastal Plains of Tamil Nadu, India." *Earth*, 3(1), 383–400.

Ghimire, B., Rogan, J., Galiano, V. R., Panday, P., and Neeti, N. (2012). "An Evaluation of Bagging, Boosting, and Random Forests for Land-Cover Classification in Cape Cod, Massachusetts, USA." *GIScience Remote Sens.*, 49(5), 623–643.

Gholami, B., Norton, I., Tannenbaum, A. R., and Agar, N. Y. R. (2012). "Recursive Feature Elimination for Brain Tumor Classification using Desorption Electrospray Ionization Mass Spectrometry Imaging." *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, NIH Public Access, 5261.

Gholizadeh, M. H., Melesse, A. M., and Reddi, L. (2016). "A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques." *Sensors*, 16(8), 1298.

Ghosh, P., Mukhopadhyay, A., Chanda, A., Mondal, P., Akhand, A., Mukherjee, S., Nayak, S. K., Ghosh, S., Mitra, D., Ghosh, T., and Hazra, S. (2017). "Application of Cellular automata and Markov-chain model in geospatial environmental modeling- A review." *Remote Sens. Appl. Soc. Environ.*, 5, 64–77.

Gidey, E., Dikinya, O., Sebego, R., Segosebe, · Eagilwe, and Zenebe, · Amanuel. (2017). "Cellular automata and Markov Chain (CA_Markov) model-based predictions of future land use and land cover scenarios (2015-2033) in Raya, northern Ethiopia." *Model. Earth Syst. Environ.*, 3, 1245–1262.

Giri, C., Defourny, P., and Shrestha, S. (2003). "Land cover characterization and mapping of continental Southeast Asia using multi-resolution satellite sensor data." *Int. J. Remote Sens.*, 24(21), 4181–4196.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2006). "Random forests for

land cover classification." *Pattern Recognit. Lett.*, 27(4), 294–300.

Glowka, L., Burhenne-Guilmin, F., and Synge, H. (1994). *A guide to the convention on biological diversity. Environ. Policy Law Pap.*

Gomes, V. C. F., Queiroz, G. R., and Ferreira, K. R. (2020). "An overview of platforms for big earth observation data management and analysis." *Remote Sens.*, 12(8), 1–25.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote Sens. Environ.*, 202, 18–27.

Güneralp, B., and Seto, K. C. (2008). "Environmental impacts of urban growth from an integrated dynamic perspective: A case study of Shenzhen, South China." *Glob. Environ. Chang.*, 18(4), 720–735.

Halmy, M. W. A., Gessler, P. E., Hicke, J. A., and Salem, B. B. (2015a). "Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA." *Appl. Geogr.*, 63, 101–112.

Halmy, M. W. A., Gessler, P. E., Hicke, J. A., and Salem, B. B. (2015b). "Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA." *Appl. Geogr.*, 63, 101–112.

Hamad, R., Balzter, H., and Kolo, K. (2018a). "Predicting Land Use/Land Cover Changes Using a CA-Markov Model under Two Different Scenarios." *Sustainability*, 10(10), 3421.

Hamad, R., Kolo, K., and Balzter, H. (2018b). "Land Cover Changes Induced by Demining Operations in Halgurd-Sakran National Park in the Kurdistan Region of Iraq." *Sustainability*, 10(7), 2422.

Han, B., Loisel, H., Vantrepotte, V., Mériaux, X., Bryère, P., Ouillon, S., Dessailly, D., Xing, Q., and Zhu, J. (2016). "Development of a Semi-Analytical Algorithm for the Retrieval of Suspended Particulate Matter from Remote Sensing over Clear to Very Turbid Waters." *Remote Sens.*, 8(3), 211.

Hanley, J. A. (1989). "Receiver operating characteristic (ROC) methodology: the state of the art." *Crit. Rev. Diagnostic Imaginf*, 29(3), 307–335.

Hasan, A. H., Anbar, M., and Alamiedy, T. A. (2023). "Deep learning approach for detecting router advertisement flooding-based DDoS attacks." *J. Ambient Intell. Humaniz. Comput.*, 14(6), 7281–7295.

Hassan, Z., Shabbir, R., Ahmad, S. S., Malik, A. H., Aziz, N., Butt, A., and Erum, S. (2016). "Dynamics of land use and land cover change (LULCC) using geospatial techniques: a case study of Islamabad Pakistan." *Springerplus*, 5(1), 812.

HCSL. (2020). *Hooghly Cochin Shipyard Limited*. Cochin, India.

Helming, K. (2001). "Wind Speed Effects on Rain Erosivity." *Sustain. Glob. Farm*, D. E. Stott, R. H. Mohtar, and G. C. Steinhardt, eds., West Lafayette, IN: ISCO, 771–776.

Hess, L. L., Melack, J. M., Novo, E. M. L. M., Barbosa, C. C. F., and Gastil, M. (2003). "Dual-season mapping of wetland inundation and vegetation for the central Amazon basin." *Remote Sens. Environ.*, 87(4), 404–428.

Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A. X., Chen, W., Kougias, I., and Kazakis, N. (2018). "Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution." *Sci. Total Environ.*, 621, 1124–1141.

Hope, R. (2011). "Clean water is not just a health issue: it is critical for education | University of Oxford." *Univ. Oxford*.

Hua, L., Tang, L., Cui, S., and Yin, K. (2014). "Simulating urban growth using the SLEUTH model in a coastal peri-urban district in China." *Sustain.*, 6(6), 3899–3914.

Huang, C., Davis, L. S., and Townshend, J. R. G. (2002). "An assessment of support vector machines for land cover classification." *Int. J. Remote Sens.*, 23(4), 725–749.

Hyandye, C., Mandara, C. G., and Safari, J. (2015). "GIS and Logit Regression Model Applications in Land Use/Land Cover Change and Distribution in Usangu Catchment." *Am. J. Remote Sens.*, 3(1), 6.

IOCCG. (2010). *Atmospheric Correction for Remotely-Sensed Ocean-Colour.* (M. Wang, ed.), Dartmouth, NS, Canada: International Ocean Colour Coordinating Group (IOCCG).

IWMI. (2014). *Global Water Demand Projections: Past, Present and Future*. Colombo, Sri Lanka.

Jacinth Jennifer, J., and Saravanan, S. (2022). "Artificial neural network and sensitivity analysis in the landslide susceptibility mapping of Idukki district, India." *Geocarto Int.*, 37(19), 5693–5715.

Jacinth Jennifer, J., Saravanan, S., and Abijith, D. (2020). "Integration of SAR and multi-spectral imagery in flood inundation mapping–a case study on Kerala floods

2018." *ISH J. Hydraul. Eng.*, 28(1), 480–490.

Jafar-Sidik, M., Gohin, F., Bowers, D., Howarth, J., and Hull, T. (2017). "The relationship between Suspended Particulate Matter and Turbidity at a mooring station in a coastal environment: consequences for satellite-derived products." *Oceanologia*, 59(3), 365–378.

Janizadeh, S., Chandra Pal, S., Saha, A., Chowdhuri, I., Ahmadi, K., Mirzaei, S., Mosavi, A. H., and Tiefenbacher, J. P. (2021). "Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future." *J. Environ. Manage.*, 298, 113551.

Jankowski, P. (2009). "Towards participatory geographic information systems for community-based environmental decision making." *J. Environ. Manage.*, 90(6), 1966–1971.

Jat, M. K., Choudhary, M., and Saxena, A. (2017). "Application of geo-spatial techniques and cellular automata for modelling urban growth of a heterogeneous urban fringe." *Egypt. J. Remote Sens. Sp. Sci.*, 20(2), 223–241.

Jeon, H., and Oh, S. (2020). "Hybrid-Recursive Feature Elimination for Efficient Feature Selection." *Appl. Sci.*, 10(9), 3211.

Jha, M. K., and Afreen, S. (2020). "Flooding Urban Landscapes: Analysis Using Combined Hydrodynamic and Hydrologic Modeling Approaches." *Water*, 12(7), 1986.

Ji, L., Geng, X., Sun, K., Zhao, Y., and Gong, P. (2015). "Target Detection Method for Water Mapping Using Landsat 8 OLI/TIRS Imagery." *Water*, 7(2), 794–817.

Johnson, M. A., Brown, C. H., and Wells, S. J. (2002). "Using classification and regression trees (CART) to support worker decision making." *Soc. Work Res.*, 26(1), 19–29.

Jose, D. M., and Dwarakish, G. S. (2020). "Uncertainties in predicting impacts of climate change on hydrology in basin scale: a review." *Arab. J. Geosci.*, 13(19), 1037.

Jose, D. M., and Dwarakish, G. S. (2022). "Bias Correction and Trend Analysis of Temperature Data by a High-Resolution CMIP6 Model over a Tropical River Basin." *Asia-Pacific J. Atmos. Sci.*, 58(1), 97–115.

Jose, D. M., Vincent, A. M., and Dwarakish, G. S. (2022). "Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques." *Sci. Rep.*, 12(1).

Karamouz, M., and Zahmatkesh, Z. (2016). "Quantifying Resilience and Uncertainty in Coastal Flooding Events: Framework for Assessing Urban Vulnerability." *J. Water Resour. Plan. Manag.*, 143(1).

Kari, E., Kratzer, S., Beltrán-Abaunza, J. M., Harvey, E. T., and Vaičiūtė, D. (2016). "Retrieval of suspended particulate matter from turbidity − model development, validation, and application to MERIS data over the Baltic Sea." *Int. J. Remote Sens.*, 38(7), 1983–2003.

Kavzoglu, T., and Colkesen, I. (2009). "A kernel functions analysis for support vector machines for land cover classification." *Int. J. Appl. Earth Obs. Geoinf.*, 11(5), 352–359.

Keukelaere, L. de, Sterckx, S., Adriaensen, S., Bhatia, N., Monbaliu, J., Toorman, E., Cattrijsse, A., Lebreton, C., Zande, D. Van der, and Knaeps, E. (2020). "Coastal turbidity derived from PROBA-V global vegetation satellite." *Remote Sens.*, 12(3).

Khaire, U. M., and Dhanalakshmi, R. (2022). "Stability of feature selection algorithm: A review." *J. King Saud Univ. - Comput. Inf. Sci.*, 34(4), 1060–1073.

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., and Moradi, A. (2012). "An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia." *Environ. Earth Sci.*, 67(1), 251–264.

Kim, H. Il, and Kim, B. H. (2020). "Flood Hazard Rating Prediction for Urban Areas Using Random Forest and LSTM." *KSCE J. Civ. Eng.*, 24, 3884–3896.

Kleemann, J., Baysal, G., Bulley, H. N. N., and Fürst, C. (2017). "Assessing driving forces of land use and land cover change by a mixed-method approach in north-eastern Ghana, West Africa." *J. Environ. Manage.*, 196, 411–442.

Klemas, V. (2015). "Remote Sensing of Floods and Flood-Prone Areas: An Overview." *J. Coast. Res.*, 314, 1005–1013.

Koc, K., and Işık, Z. (2021). "Assessment of Urban Flood Risk Factors Using Monte Carlo Analytical Hierarchy Process." *Nat. Hazards Rev.*, 22(4).

Kocsis, T., Kovács-Székely, I., and Anda, A. (2020). "Homogeneity tests and non-parametric analyses of tendencies in precipitation time series in Keszthely, Western Hungary." *Theor. Appl. Climatol.*, 139, 849–859.

Kraus, C. N., Bonnet, M. P., Nogueira, I. de S., Lobo, M. T. M. P. S., Marques, D. da M., Garnier, J., and Vieira, L. C. G. (2019). "Unraveling Flooding Dynamics and

Nutrients' Controls upon Phytoplankton Functional Dynamics in Amazonian Floodplain Lakes." *Water*, 11(1), 154.

Krishnaraj, A., and Deka, P. C. (2020). "Spatial and temporal variations in river water quality of the Middle Ganga Basin using unsupervised machine learning techniques." *Environ. Monit. Assess.*, 192(744).

Krishnaraj, A., and Honnasiddaiah, R. (2022). "Remote sensing and machine learning based framework for the assessment of spatio-temporal water quality in the Middle Ganga Basin." *Environ. Sci. Pollut. Res.*, 29(43), 64939–64958.

KSDMA. (2022). *Quantifying historic and climate change induced Flood Hazard Probability of Kerala & Assessing the exposure of schools & hospitals to Floods in Kerala*. Thiruvananthapuram.

Kuhn, M., and Johnson, K. (2013). *Applied predictive modeling*. *Appl. Predict. Model.*, New York: Springer.

Kulk, G., George, G., Abdulaziz, A., Menon, N., Theenathayalan, V., Jayaram, C., Brewin, R. J. W., and Sathyendranath, S. (2021). "Effect of Reduced Anthropogenic Activities on Water Quality in Lake Vembanad, India." *Remote Sens.*, 13(9), 1631.

Kumar, D., and Bhattacharjya, R. K. (2020). "Estimation of Integrated Flood Vulnerability Index for the Hilly Region of Uttarakhand, India." *J. Hazardous, Toxic, Radioact. Waste*, 24(4).

Kumar, K. K., Rajagopalan, B., and Cane, M. A. (1999). "On the Weakening Relationship Between the Indian Monsoon and ENSO." *Science (80-.).*, 284(5423), 2156–2159.

Kumar, V., and Agrawal, S. (2023). "A multi-layer perceptron–Markov chain based LULC change analysis and prediction using remote sensing data in Prayagraj district, India." *Environ. Monit. Assess.*, 195(5), 1–27.

Lambin, E. F. (1997). "Modelling and monitoring land-cover change processes in tropical regions." *Prog. Phys. Geogr. Earth Environ.*, 21(3), 375–393.

Lawal, D. U., Matori, A. N., Hashim, A. M., Chandio, I. A., Sabri, S., Balogun, A. L., and Abba, H. A. (2011). "Geographic information system and remote sensing applications in flood hazards management: A review." *Res. J. Appl. Sci. Eng. Technol.*

Lee, S., Kim, J. C., Jung, H. S., Lee, M. J., and Lee, S. (2017). "Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan

city, Korea." *Geomatics, Nat. Hazards Risk*, 8(2), 1185–1203.

Lee, Z., Carder, K. L., and Arnone, R. A. (2002). "Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters." *Appl. Opt.*, 41(27), 5755.

Lehner, B., and Döll, P. (2004). "Development and validation of a global database of lakes, reservoirs and wetlands." *J. Hydrol.*, 296(1–4), 1–22.

León, J. G., Beamud, S. G., Temporetti, P. F., Atencio, A. G., Diaz, M. M., and Pedrozo, F. L. (2016). "Stratification and residence time as factors controlling the seasonal variation and the vertical distribution of chlorophyll-a in a subtropical irrigation reservoir." *Int. Rev. Hydrobiol.*, 101(1–2), 36–47.

Leray, S., Engdahl, N. B., Massoudieh, A., Bresciani, E., and McCallum, J. (2016). "Residence time distributions for hydrologic systems: Mechanistic foundations and steady-state analytical solutions." *J. Hydrol.*, 543, 67–87.

Leta, M. K., Demissie, T. A., and Tränckner, J. (2021). "Modeling and Prediction of Land Use Land Cover Change Dynamics Based on Land Change Modeler (LCM) in Nashe Watershed, Upper Blue Nile Basin, Ethiopia." *Sustain. 2021, Vol. 13, Page 3740*, 13(7), 3740.

Li, Y., and Hong, H. (2023). "Modelling flood susceptibility based on deep learning coupling with ensemble learning models." *J. Environ. Manage.*, 325, 116450.

Lillesand, T. M., and Kiefer, R. W. (1979). *Remote sensing and image interpretation. Wiley, New York*.

Liu, C., Li, W., Zhu, G., Zhou, H., Yan, H., and Xue, P. (2020a). "Land Use/Land Cover Changes and Their Driving Factors in the Northeastern Tibetan Plateau Based on Geographical Detectors and Google Earth Engine: A Case Study in Gannan Prefecture." *Remote Sens.*, 12(19), 3139.

Liu, L., Liang, Y., and Hashimoto, S. (2020b). "Integrated assessment of land-use/coverage changes and their impacts on ecosystem services in Gansu Province, northwest China: implications for sustainable development goals." *Sustain. Sci.*, 15(1), 297–314.

Louis, J., Pflug, B., Main-Knorn, M., Debaecker, V., Mueller-Wilm, U., and Gascon, F. (2018). "Integration and assimilation of meteorological (ECMWF) aerosol estimates into Sen2COR atmospheric correction." *Int. Geosci. Remote Sens. Symp.*, 18, 1894–

1897.

Luo, T., Maddocks, A., Iceland, C., Ward, P., and Winsemius, H. (2015). "World's 15 Countries with the Most People Exposed to River Floods." *World Resour. Inst.*, <https://www.wri.org/insights/worlds-15-countries-most-people-exposed-river-floods> (Jan. 26, 2023).

Lyubimova, T., Lepikhin, A., Parshakova, Y., and Tiunov, A. (2016). "The risk of river pollution due to washout from contaminated floodplain water bodies during periods of high magnitude floods." *J. Hydrol.*, 534, 579–589.

Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., and Wang, Z. (2021). "XGBoost-based method for flash flood risk assessment." *J. Hydrol.*, 598, 126382.

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., and Luo, B. (2020). "Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China." *Sci. Total Environ.*, 724, 138226.

Mabit, R., Araújo, C. A. S., Singh, R. K., and Bélanger, S. (2022). "Empirical Remote Sensing Algorithms to Retrieve SPM and CDOM in Québec Coastal Waters." *Front. Remote Sens.*, 3.

Maciel, F. P., Haakonsson, S., Ponce de León, L., Bonilla, S., and Pedocchi, F. (2023). "Challenges for chlorophyll-a remote sensing in a highly variable turbidity estuary, an implementation with Sentinel-2." *Geocarto Int.*, 38(1), 2160017.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T. R. (2014). "Review of trend analysis and climate change projections of extreme precipitation and floods in Europe." *J. Hydrol.*, 519(PD), 3634–3650.

Mahato, S., Pal, S., Talukdar, S., Saha, T. K., and Mandal, P. (2021). "Field based index of flood vulnerability (IFV): A new validation technique for flood susceptible models." *Geosci. Front.*, 12(5), 101175.

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., and Zhang, Y. (2018). "Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery." *Remote Sens.*, 10(7), 1119.

Mahdizadeh Gharakhanlou, N., and Perez, L. (2023). "Flood susceptible prediction through the use of geospatial variables and machine learning methods." *J. Hydrol.*, 617, 129121.

Malekian, A., and Azarnivand, A. (2016). "Application of Integrated Shannon's

Entropy and VIKOR Techniques in Prioritization of Flood Risk in the Shemshak Watershed, Iran." *Water Resour. Manag.*, 30(1), 409–425.

Manavalan, R. (2017). "SAR image analysis techniques for flood area mapping - literature survey." *Earth Sci. Informatics*, 10(1), 1–14.

Mancino, G., Ferrara, A., Padula, A., and Nolè, A. (2020). "Cross-Comparison between Landsat 8 (OLI) and Landsat 7 (ETM+) Derived Vegetation Indices in a Mediterranean Environment." *Remote Sens.*, 12(2), 291.

Mandal, S., Bandyopadhyay, A., and Bhadra, A. (2023). "Dynamics and future prediction of LULC on Pare River basin of Arunachal Pradesh using machine learning techniques." *Environ. Monit. Assess.*, 195(6), 709.

Martinis, S., and Rieke, C. (2015). "Backscatter Analysis Using Multi-Temporal and Multi-Frequency SAR Data in the Context of Flood Mapping at River Saale, Germany." *Remote Sens.*, 7(6), 7732–7752.

Matgen, P., Schumann, G., Henry, J.-B., Hoffmann, L., and Pfister, L. (2007). "Integration of SAR-derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management." *Int. J. Appl. Earth Obs. Geoinf.*, 9(3), 247–263.

Mather, P., and Tso, B. (2016). *Classification Methods for Remotely Sensed Data*. Boca Raton: CRC Press.

Meliho, M., Khattabi, A., Driss, Z., and Orlando, C. A. (2022). "Spatial prediction of flood-susceptible zones in the Ourika watershed of Morocco using machine learning algorithms." *Appl. Comput. Informatics*.

Mirzaei, S., Vafakhah, M., Pradhan, B., and Alavi, S. J. (2021). "Flood susceptibility assessment using extreme gradient boosting (EGB), Iran." *Earth Sci. Informatics*, 14(1), 51–67.

Mishra, D. R., Narumalani, S., Rundquist, D., and Lawson, M. (2005). "Characterizing the vertical diffuse attenuation coefficient for downwelling irradiance in coastal waters: Implications for water penetration by high resolution satellite data." *ISPRS J. Photogramm. Remote Sens.*, 60, 48–64.

Mishra, P. K., Rai, A., and Rai, S. C. (2020). "Land use and land cover change detection using geospatial techniques in the Sikkim Himalaya, India." *Egypt. J. Remote Sens. Sp. Sci.*, 23(2), 133–143.

Mishra, S., and Mishra, D. R. (2012). "Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters." *Remote Sens. Environ.*, 117, 394–406.

Misra, P., and Singh Yadav, A. (2020). "Improving the Classification Accuracy using Recursive Feature Elimination with Cross-Validation." *Int. J. Emerg. Technol.*, 11(3), 659–665.

Mitsch, W. J., and Gosselink, J. G. (2000). "The value of wetlands: importance of scale and landscape setting." *Ecol. Econ.*, 35(1), 25–33.

Mohammed, R., and Scholz, M. (2019). "Climate Variability Impact on the Spatiotemporal Characteristics of Drought and Aridityin Arid and Semi-Arid Regions." *Water Resour. Manag.*, 33, 5015–5033.

Mohanty, A., Hussain, M., Mishra, M., Kattel, D. B., and Pal, I. (2019). "Exploring community resilience and early warning solution for flash floods, debris flow and landslides in conflict prone villages of Badakhshan, Afghanistan." *Int. J. Disaster Risk Reduct.*, 33(January), 5–15.

Morel, A., and Prieur, L. (1977). "Analysis of variations in ocean color." *Limnol. Oceanogr.*, 22(4), 709–722.

Mosavi, A., Ozturk, P., and Chau, K. W. (2018). "Flood Prediction Using Machine Learning Models: Literature Review." *Water*, 10(11), 1536.

Mu, W., Yu, F., Li, C., Xie, Y., Tian, J., Liu, J., and Zhao, N. (2015). "Effects of Rainfall Intensity and Slope Gradient on Runoff and Soil Moisture Content on Different Growing Stages of Spring Maize." *Water*, 7(6), 2990–3008.

Natekin, A., and Knoll, A. (2013). "Gradient boosting machines, a tutorial." *Front. Neurorobot.*, 7, 21.

Navid, D. (1989). "The International Law of Migratory Species: The Ramsar Convention." *Nat. Resour. J.*, 29(4), 1001–1016.

Nawindah. (2017). "Simple Additive Weighting (SAW) Mathematics Method for Warehouse Disaster Location Selection In Central Jakarta, Indonesia." *Int. J. Pure Appl. Math.*, 117(15), 795–803.

Nazirova, K., Alferyeva, Y., Lavrova, O., Shur, Y., Soloviev, D., Bocharova, T., and Strochkov, A. (2021). "Comparison of in situ and remote-sensing methods to determine turbidity and concentration of suspended matter in the estuary zone of the mzymta river,

black sea." *Remote Sens.*, 13(1), 1–29.

Nechad, B., Dogliotti, A. I., Ruddick, K. G., and Doxaran, D. (2016). "Particulate Backscattering and Suspended Matter Concentration Retrieval From Remote-Sensed Turbidity in Various Coastal and." *Proc. ESA Living Planet Symp.*, 2015(May), 9–13.

Nechad, B., Ruddick, K. G., and Neukermans, G. (2009). "Calibration and validation of a generic multisensor algorithm for mapping of turbidity in coastal waters." *Remote Sens. Ocean. Sea Ice, Large Water Reg. 2009*, 7473(September), 74730H.

Nechad, B., Ruddick, K. G., and Park, Y. (2010). "Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters." *Remote Sens. Environ.*, 114(4), 854–866.

Ngoc, D. D., Loisel, H., Vantrepotte, V., Xuan, H. C., Minh, N. N., Verpoorter, C., Meriaux, X., Minh, H. P. T., Thi, H. Le, Hong, H. L. V., and Van, T. N. (2020). "A Simple Empirical Band-Ratio Algorithm to Assess Suspended Particulate Matter from Remote Sensing over Coastal and Inland Waters of Vietnam: Application to the VNREDSat-1/NAOMI Sensor." *Water*, 12(9), 2636.

Nguyen, D. H., Hien Le, X., Heo, J. Y., and Bae, D. H. (2021). "Development of an Extreme Gradient Boosting Model Integrated with Evolutionary Algorithms for Hourly Water Level Prediction." *IEEE Access*, 9, 125853–125867.

Noi Phan, T., Kuch, V., and Lehnert, L. W. (2020). "Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition." *Remote Sens.*, 12(15), 2411.

Ogen, Y. (2020). "Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality." *Sci. Total Environ.*, 726, 138605.

Ordoyne, C., and Friedl, M. A. (2008). "Using MODIS data to characterize seasonal inundation patterns in the Florida Everglades." *Remote Sens. Environ.*, 112(11), 4107–4119.

Ozturk, D. (2015). "Urban growth simulation of Atakum (Samsun, Turkey) using cellular automata-Markov chain and Multi-layer Perceptron-Markov chain models." *Remote Sens.*, 7(5), 5918–5950.

Pal, I., Ghosh, T., and Ghosh, C. (2017). "Institutional framework and administrative systems for effective disaster risk governance – Perspectives of 2013 Cyclone Phailin in India." *Int. J. Disaster Risk Reduct.*, 21, 350–359.

Pal, I., Tularug, P., Jana, S. K., and Pal, D. K. (2018). *Risk assessment and reduction measures in landslide and flash flood-prone areas: A case of Southern Thailand (Nakhon Si Thammarat Province). Integr. Disaster Sci. Manag. Glob. Case Stud. Mitig. Recover.*, Elsevier Inc.

Pan, X., Wang, Z., Gao, Y., Dang, X., and Han, Y. (2022). "Detailed and automated classification of land use/land cover using machine learning algorithms in Google Earth Engine." *Geocarto Int.*, 37(18), 5415–5432.

Panahi, M., Jaafari, A., Shirzadi, A., Shahabi, H., Rahmati, O., Omidvar, E., Lee, S., and Bui, D. T. (2021). "Deep learning neural networks for spatially explicit prediction of flash flood probability." *Geosci. Front.*, 12(3), 101076.

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B., and Matthews, E. (2010). "Interannual variability of surface water extent at the global scale, 1993-2004." *J. Geophys. Res.*, 115(12), 1–17.

Papaioannou, G., Vasiliades, L., and Loukas, A. (2015). "Multi-Criteria Analysis Framework for Potential Flood Prone Areas Mapping." *Water Resour. Manag.*, 29(2), 399–418.

Parry, J. A., Ganaie, S. A., and Sultan Bhat, M. (2018). "GIS based land suitability analysis using AHP model for urban services planning in Srinagar and Jammu urban centers of J&K, India." *J. Urban Manag.*, 7(2), 46–56.

Parthasarathy, K. S. S., Saravanan, S., Deka, P. C., and Devanantham, A. (2022). "Assessment of potentially vulnerable zones using geospatial approach along the coast of Cuddalore district, East coast of India." *ISH J. Hydraul. Eng.*, 28(sup1), 422–432.

Parvinnezhad, D., Delavar, M. R., Pijanowski, B. C., and Claramunt, C. (2020). "Integration of adaptive neural fuzzy inference system and fuzzy rough set theory with support vector regression to urban growth modelling." *Earth Sci. Informatics*, 17–36.

Pati, A., and Sahoo, B. (2021). "Effect of Low-Impact Development Scenarios on Pluvial Flood Susceptibility in a Scantily Gauged Urban–Peri-Urban Catchment." *J. Hydrol. Eng.*, 27(1).

Patrikaki, O., Kazakis, N., Kougias, I., Patsialis, T., Theodossiou, N., and Voudouris, K. (2018). "Assessing Flood Hazard at River Basin Scale with an Index-Based Approach: The Case of Mouriki, Greece." *Geosciences*, 8(2), 50.

Pelletier, C., Valero, S., Inglada, J., Champion, N., and Dedieu, G. (2016). "Assessing

the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas." *Remote Sens. Environ.*, 187, 156–168.

Pereira-Sandoval, M., Ruescas, A., Urrego, P., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J., and Moreno, J. (2019). "Evaluation of Atmospheric Correction Algorithms over Spanish Inland Waters for Sentinel-2 Multi Spectral Imagery Data." *Remote Sens.*, 11(12), 1469.

Pham, B. T., Jaafari, A., Phong, T. Van, Yen, H. P. H., Tuyen, T. T., Luong, V. Van, Nguyen, H. D., Le, H. Van, and Foong, L. K. (2021). "Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques." *Geosci. Front.*, 12(3), 101105.

Pijanowski, B. C., Brown, D. G., Shellito, B. A., and Manik, G. A. (2002). "Using neural networks and GIS to forecast land use changes: a Land Transformation Model." *Comput. Environ. Urban Syst.*, 26(6), 553–575.

Pradhan, B., and Youssef, A. M. (2011). "A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan River Corridor, Malaysia." *J. Flood Risk Manag.*, 4(3), 189–202.

Rafiei-Sardooi, E., Azareh, A., Choubin, B., Mosavi, A. H., and Clague, J. J. (2021). "Evaluating urban flood risk using hybrid method of TOPSIS and machine learning." *Int. J. Disaster Risk Reduct.*, 66, 102614.

Rahman, M., Ningsheng, C., Mahmud, G. I., Islam, M. M., Pourghasemi, H. R., Ahmad, H., Habumugisha, J. M., Washakh, R. M. A., Alam, M., Liu, E., Han, Z., Ni, H., Shufeng, T., and Dewan, A. (2021). "Flooding and its relationship with land cover change, population growth, and road density." *Geosci. Front.*, 12(6), 101224.

Rahman, M. R., and Thakur, P. K. (2018). "Detecting, mapping and analysing of flood water propagation using synthetic aperture radar (SAR) satellite data and GIS: A case study from the Kendrapara District of Orissa State of India." *Egypt. J. Remote Sens. Sp. Sci.*, 21, S37–S41.

Rahmati, O., Pourghasemi, H. R., and Zeinivand, H. (2015a). "Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran." *Geocarto Int.* , 31(1), 42–70.

Rahmati, O., Zeinivand, H., and Besharat, M. (2015b). "Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis." *Geomatics, Nat. Hazards*

Risk, 7(3), 1000–1017.

Rakesh Kumar, R. D. Singh, and K. D. Sharma. (2005). "Water resources of India." *Curr. Sci.*, 89(5), 794–811.

Ramasamy, S., Gunasekaran, S., Saravanavel, J., Joshua, R. M., Rajaperumal, R., Kathiravan, R., Palanivel, K., and Muthukumar, M. (2020). "Geomorphology and Landslide Proneness of Kerala, India A Geospatial study." *Landslides*, 18(4), 1245–1258.

Rawat, J. S., and Kumar, M. (2015). "Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India." *Egypt. J. Remote Sens. Sp. Sci.*, 18(1), 77–84.

Rebelo, A. J., Scheunders, P., Esler, K. J., and Meire, P. (2017). "Detecting, mapping and classifying wetland fragments at a landscape scale." *Remote Sens. Appl. Soc. Environ.*, 8, 212–223.

Rebelo, L. M., Finlayson, C. M., and Nagabhatla, N. (2009). "Remote sensing and GIS for wetland inventory, mapping and change analysis." *J. Environ. Manage.*, 90(7), 2144–2153.

Rienow, A., and Goetzke, R. (2015). "Supporting SLEUTH - Enhancing a cellular automaton with support vector machines for urban growth modeling." *Comput. Environ. Urban Syst.*, 49, 66–81.

Ritchie, J. C., Schiebe, F. R., and McHenry, J. R. (1976). "Remote Sensing of Suspended Sediments in Surface Waters." *J. Am. Soc. Photogramm.*, 42(12).

Ritchie, J. C., Zimba, P. V., and Everitt, J. H. (2003). "Remote Sensing Techniques to Assess Water Quality." *Photogramm. Eng. Remote Sensing*, 69(6), 695–704.

Romali, N. S., Yusop, Z., and Ismail, A. Z. (2018). "Hydrological Modelling using HEC-HMS for Flood Risk Assessment of Segamat Town, Malaysia." *IOP Conf. Ser. Mater. Sci. Eng.*, 318(1), 012029.

Ruben, G. B., Zhang, K., Dong, Z., and Xia, J. (2020). "Analysis and Projection of Land-Use/Land-Cover Dynamics through Scenario-Based Simulations Using the CA-Markov Model: A Case Study in Guanting Reservoir Basin, China." *Sustainability*, 12(9), 3747.

Russi, D., Brink, P. ten, Farmer, a., Badura, T., Coates, D., Förster, J., Kumar, R., Davidson, N., and Russi, Daniela; ten Brink, Patrick; Farmer, Andrew; Badura, Tomas;

Coates, David; Förster, Johannes; Kumar, Ritesh; Davidson, N. (2013). "The Economics of Ecosystems and Biodiversity for Water and Wetlands." *Vasa*, 84.

Sachdeva, S., and Kumar, B. (2022). "Flood susceptibility mapping using extremely randomized trees for Assam 2020 floods." *Ecol. Inform.*, 67.

Saghafian, B., Julien, P. Y., and Rajaie, H. (2002). "Runoff hydrograph simulation based on time variable isochrone technique." *J. Hydrol.*, 261(1–4), 193–203.

Saha, A., Pal, S. C., Arabameri, A., Blaschke, T., Panahi, S., Chowdhuri, I., Chakrabortty, R., Costache, R., and Arora, A. (2021). "Flood Susceptibility Assessment Using Novel Ensemble of Hyperpipes and Support Vector Regression Algorithms." *Water*, 13(2), 241.

Sahin, E. K. (2022). "Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping." *Geocarto Int.*, 37(9), 2441–2465.

Saravanan, S., and Abijith, D. (2022). "Flood susceptibility mapping of Northeast coastal districts of Tamil Nadu India using Multi-source Geospatial data and Machine Learning techniques." *Geocarto Int.*, 37(27), 15252–15281.

Saravanan, S., Parthasarathy, K. S. S., and Kumaresan, P. R. (2015). "Shoreline Change Detection for Chennai Coast Using Geospatial." *20th Int. Conf. Hydraul. Water Resour. River Eng.*, IIT Roorkee, India, 17–19.

Sarchani, S., Awol, F. S., and Tsanis, I. (2021). "Hydrological Analysis of Extreme Rain Events in a Medium-Sized Basin." *Appl. Sci.*, 11(11), 4901.

Sarkar, D., and Mondal, P. (2020). "Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region." *Appl. Water Sci.*, 10(1), 1–13.

Saxena, A., and Jat, M. K. (2020). "Land suitability and urban growth modeling: Development of SLEUTH-Suitability." *Comput. Environ. Urban Syst.*, 81(February), 101475.

Saxena, A., Jat, M. K., and Clarke, K. C. (2021). "Development of SLEUTH-Density for the simulation of built-up land density." *Comput. Environ. Urban Syst.*, 86, 101586.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P. (2014).

"Multimodel assessment of water scarcity under climate change." *Proc. Natl. Acad. Sci. U. S. A.*, 111(9), 3245–3250.

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., Litvak, M., Ogle, K., Shaw, J. D., Wolf, A., Huntzinger, D. N., Schaefer, K., Cook, R., Wei, Y., Fang, Y., Hayes, D., Huang, M., Jain, A., and Tian, H. (2017). "Global patterns of drought recovery." *Nature*, 548(7666), 202–205.

Senf, C., Seidl, R., and Hostert, P. (2017). "Remote sensing of forest insect disturbances: Current state and future directions." *Int. J. Appl. Earth Obs. Geoinf.*, 60, 49–60.

Serasinghe Pathiranage, I. S., Kantakumar, L. N., and Sundaramoorthy, S. (2018). "Remote Sensing Data and SLEUTH Urban Growth Model: As Decision Support Tools for Urban Planning." *Chinese Geogr. Sci.*, 28(2), 274–286.

Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., and Shirzadi, A. (2018). "Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping." *J. Environ. Manage.*, 217, 1–11.

Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B. T., Mansouripour, F., Geertsema, M., Clague, J. J., and Bui, D. T. (2021). "Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm." *Geosci. Front.*, 12(3), 101100.

Shahzad, N., Ding, X., and Abbas, S. (2022). "A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan." *Appl. Sci.*, 12(5), 2280.

Shamsi, S. . F. (2010). "Integrating Linear Programming and Analytical Hierarchical Processing in Raster-GIS to Optimize Land Use Pattern at Watershed Level." *J. Appl. Sci. Environ. Manag.*, 14(2), 75–80.

Shao, Y., and Lunetta, R. S. (2012). "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points." *ISPRS J. Photogramm. Remote Sens.*, 70, 78–87.

Shetty, S., Gupta, P. K., Belgiu, M., and Srivastav, S. K. (2021). "Assessing the Effect of Training Sampling Design on the Performance of Machine Learning Classifiers for Land Cover Mapping Using Multi-Temporal Remote Sensing Data and Google Earth Engine." *Remote Sens.*, 13(8), 1433.

Shi, D., and Yang, X. (2015). "Support Vector Machines for Land Cover Mapping from Remote Sensor Imagery." *Monit. Model. Glob. Chang. A Geomatics Perspect.*, Li J. and Yang. X, eds., Dordrecht: Springer, 265–279.

Sidhu, N., Pebesma, E., and Câmara, G. (2018). "Using Google Earth Engine to detect land cover change: Singapore as a use case." *Eur. J. Remote Sens.*, 51(1), 486–500.

Silva, L. P. e., Xavier, A. P. C., Silva, R. M. da, and Santos, C. A. G. (2020). "Modeling land cover change based on an artificial neural network for a semiarid river basin in northeastern Brazil." *Glob. Ecol. Conserv.*, 21, e00811.

Simonit, S., and Perrings, C. (2011). "Sustainability and the value of the 'regulating' services: Wetlands and water quality in Lake Victoria." *Ecol. Econ.*, 70(6), 1189–1199.

Singh, S. K., Mustak, S., Srivastava, P. K., Szabó, S., and Islam, T. (2015). "Predicting Spatial and Decadal LULC Changes Through Cellular Automata Markov Chain Models Using Earth Observation Datasets and Geo-information." *Environ. Process.*, 2(1), 61–78.

Siswanto, E., Tang, J., Yamaguchi, H., Ahn, Y. H., Ishizaka, J., Yoo, S., Kim, S. W., Kiyomoto, Y., Yamada, K., Chiang, C., and Kawamura, H. (2011). "Empirical ocean-color algorithms to retrieve chlorophyll-a, total suspended matter, and colored dissolved organic matter absorption coefficient in the Yellow and East China Seas." *J. Oceanogr.*, 67(5), 627–650.

Slater, L. J., Singer, M. B., and Kirchner, J. W. (2015). "Hydrologic versus geomorphic drivers of trends in flood hazard." *Geophys. Res. Lett.*, 42(2), 370–376.

Soti, V., Tran, A., Bailly, J. S., Puech, C., Seen, D. Lo, and Bégué, A. (2009). "Assessing optical earth observation systems for mapping and monitoring temporary ponds in arid areas." *Int. J. Appl. Earth Obs. Geoinf.*, 11(5), 344–351.

Sravanthi, N., Ramana, I. V, Ali, Y., Ashraf, M., Ali, M. M., and Narayana, A. C. (2013). "An Algorithm for Estimating Suspended Sediment Concentrations in the Coastal Waters of India using Remotely Sensed Reflectance and its Application to Coastal Environments." *Int. J. Environ. Res*, 7(4), 841–850.

Stefanov, W. L., Ramsey, M. S., and Christensen, P. R. (2001). "Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers." *Remote Sens. Environ.*, 77(2), 173–185.

Stumpner, E. B., Bergamaschi, B. A., Kraus, T. E. C., Parker, A. E., Wilkerson, F. P.,

Downing, B. D., Dugdale, R. C., Murrell, M. C., Carpenter, K. D., Orlando, J. L., and Kendall, C. (2020). "Spatial variability of phytoplankton in a shallow tidal freshwater system reveals complex controls on abundance and community structure." *Sci. Total Environ.*, 700, 134392.

Subedi, P., Subedi, K., and Thapa, B. (2013). "Application of a Hybrid Cellular Automaton − Markov (CA-Markov) Model in Land-Use Change Prediction: A Case Study of Saddle Creek Drainage Basin, Florida." *Appl. Ecol. Environ. Sci.*, 1(6), 126–132.

Sudha Rani, N. N. V., Satyanarayana, A. N. V., and Bhaskaran, P. K. (2015). "Coastal vulnerability assessment studies over India: a review." *Nat. Hazards*, 77(1), 405–428.

Sudheer, K. P., Murty Bhallamudi, S., Narasimhan, B., Thomas, J., Bindhu, V. M., Vema, V., and Kurian, C. (2019). "Role of dams on the floods of August 2018 in Periyar River Basin, Kerala." *Curr. Sci.*, 116(5), 780–794.

Sufiyan, I., and Magaji, J. (2019). "Modeling Flood Hazard Using Swat and 3d Analysis in Terengganu Watershed." *J. Clean Was*, 2(2), 19–24.

Svetnik, V., Liaw, A., Tong, C., Christopher Culberson, J., Sheridan, R. P., and Feuston, B. P. (2003). "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling." *J. Chem. Inf. Comput. Sci.*, 43(6), 1947–1958.

Taati, A., Sarmadian, F., Mousavi, A., Pour, C. T. H., and Shahir, A. H. E. (2015). "Land use classification using support vector machine and maximum likelihood algorithms by landsat 5 TM images." *Walailak J. Sci. Technol.*, 12(8), 681–687.

Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., and Scholten, T. (2020). "Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space." *Remote Sens.*, 12(7), 1095.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., and Brisco, B. (2020). "Google Earth Engine for geo-big data applications: A meta-analysis and systematic review." *ISPRS J. Photogramm. Remote Sens.*, 164, 152–170.

Tang, T., Liang, Z., Hu, Y., Li, B., and Wang, J. (2020). "Research on flood forecasting based on flood hydrograph generalization and random forest in Qiushui River basin,

China." *J. Hydroinformatics*, 22(6), 1588–1602.

Tang, X., Machimura, T., Liu, W., Li, J., and Hong, H. (2021a). "A novel index to evaluate discretization methods: A case study of flood susceptibility assessment based on random forest." *Geosci. Front.*, 12(6), 101253.

Tang, X., Shu, Y., Liu, W., Li, J., Liu, M., and Yu, H. (2021b). "An Optimized Weighted Naïve Bayes Method for Flood Risk Assessment." *Risk Anal.*, 41(12), 2301–2321.

Taşoğlu, E., and Abujayyab, S. K. M. (2022). "Comparison of the frequency ratio, index of entropy, and artificial neural networks methods for landslide susceptibility mapping: A case study in Pınarbaşı/Kastamonu (North of Turkey)." *Comput. Earth Environ. Sci.*, H. R. Pourghasemi, ed., Elsevier, 491–508.

Tassan, S. (1994). "Local algorithms using SeaWiFS data for the retrieval of phytoplankton, pigments, suspended sediment, and yellow substance in coastal waters." *Appl. Opt.*, 33(12), 2369–2378.

Tassi, A., and Vizzari, M. (2020). "Object-Oriented LULC Classification in Google Earth Learning Algorithms." *Remote Sens.*, 2020(12), 3776.

Tatu, K. S., and Anderson, J. T. (2017). "An Introduction to Wetland Science and South Asian Wetlands." *Wetl. Sci. Perspect. From South Asia*, 3–30.

Taylor, M., and Krüger, N. (2019). "Changes in salinity of a clay soil after a short-term salt water flood event." *Geoderma Reg.*, 19, e00239.

Tayyebi, A., and Pijanowski, B. C. (2014). "Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools." *Int. J. Appl. Earth Obs. Geoinf.*, 28(1), 102–116.

Tehrany, M. S., and Kumar, L. (2018). "The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods." *Environ. Earth Sci.*, 77(13).

Tehrany, M. S., Kumar, L., and Shabani, F. (2019). "A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia." *PeerJ*, 7, e7653.

Tehrany, M. S., Shabani, F., Neamah Jebur, M., Hong, H., Chen, W., and Xie, X. (2017). "GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques."

Geomatics, Nat. Hazards Risk, 8(2), 1538–1561.

Thomas, D. S. K., Ertuˆgay, K., and Kemec‚, S. (2007). "The Role of Geographic Information Systems/Remote Sensing in Disaster Management." Springer, New York, NY, 83–96.

Tiwari, V., Kumar, V., Matin, M. A., Thapa, A., Ellenburg, W. L., Gupta, N., and Thapa, S. (2020). "Flood inundation mapping-Kerala 2018; Harnessing the power of SAR, automatic threshold detection method and Google Earth Engine." *PLoS One*, 15(8 August).

Toming, K., Kutser, T., Uiboupin, R., Arikas, A., Vahter, K., and Paavel, B. (2017). "Mapping Water Quality Parameters with Sentinel-3 Ocean and Land Colour Instrument imagery in the Baltic Sea." *Remote Sens.*, 9(10), 1070.

Towfiqul Islam, A. R. M., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A., and Linh, N. T. T. (2021). "Flood susceptibility modelling using advanced ensemble machine learning models." *Geosci. Front.*, 12(3), 101075.

UN WWDR. (2022). *The United Nations World Water Development Report 2022: groundwater: making the invisible visible*. *United Nations World Water*, Paris: UNESCO.

Ustuner, M., Sanli, F. B., and Dixon, B. (2015). "Application of support vector machines for landuse classification using high-resolution rapideye images: A sensitivity analysis." *Eur. J. Remote Sens.*, 48, 403–422.

Vanhellemont, Q. (2019). "Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives." *Remote Sens. Environ.*, 225, 175–192.

Vanhellemont, Q., and Ruddick, K. (2016a). "ACOLITE processing for Sentinel-2 and Landsat-8: atmospheric correction and aquatic applications." *Ocean Opt. Conf.*

Vanhellemont, Q., and Ruddick, K. (2016b). "ACOLITE For Sentinel-2: Aquatic Applications of MSI imagery." *2016 ESA Living Planet Symp.*, Prague, Czech Republic: ESA Special Publication, 740.

Vanhellemont, Q., and Ruddick, K. (2018). "Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications." *Remote Sens. Environ.*, 216, 586–597.

Venter, Z. S., Aunan, K., Chowdhury, S., and Lelieveld, J. (2020). "COVID-19

lockdowns cause global air pollution declines with implications for public health risk." *medRxiv*.

Verburg, P. H., Nijs, T. C. M. de, Eck, J. R. van, Visser, H., and Jong, K. de. (2004). "A method to analyse neighbourhood characteristics of land use patterns." *Comput. Environ. Urban Syst.*, 28(6), 667–690.

Vijaykumar, P., Abhilash, S., Sreenath, A. V, Athira, U. N., Mohanakumar, K., Mapes, B. E., Chakrapani, B., Sahai, A. K., Niyas, T. N., and Sreejith, O. P. (2021). "Kerala floods in consecutive years - Its association with mesoscale cloudburst and structural changes in monsoon clouds over the west coast of India." *Weather Clim. Extrem.*, 33, 100339.

Vinita, J., Revichandran, C., and Manoj, N. T. (2017). "Suspended sediment dynamics in Cochin estuary, West Coast, India." *J. Coast. Conserv.*, 21(1), 233–244.

Vittorio, C. A. Di, and Georgakakos, A. P. (2018). "Land cover classification and wetland inundation mapping using MODIS." *Remote Sens. Environ.*, 204, 1–17.

Wahap, N. A., and Shafri, H. Z. M. (2020). "Utilization of Google Earth Engine (GEE) for land cover monitoring over Klang Valley, Malaysia." *IOP Conf. Ser. Earth Environ. Sci.*, 540(1), 012003.

Wang, L., Yan, J., and Ma, Y. (2019a). *Cloud Computing in Remote Sensing. Cloud Comput. Remote Sens.*, Boca Raton: Chapman and Hall/CRC.

Wang, Y., Yang, X., Zhang, M., Zhang, L., Yu, X., Ren, L., Liu, Y., Jiang, S., and Yuan, F. (2019b). "Projected effects of climate change on future hydrological regimes in the upper yangtze river basin, China." *Adv. Meteorol.*

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., and Bai, X. (2015). "Flood hazard risk assessment model based on random forest." *J. Hydrol.*, 527, 1130–1141.

Wang, Z., Liu, Q., and Liu, Y. (2020). "Mapping Landslide Susceptibility Using Machine Learning Algorithms and GIS: A Case Study in Shexian County, Anhui Province, China." *Symmetry (Basel).*, 12(12), 1954.

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., Beek, R. Van, Bierkens, M. F. P., Ligtvoet, W., and Winsemius, H. C. (2013). "Assessing flood risk at the global scale: model setup, results, and sensitivity." *Environ. Res. Lett*, 8, 44019–44029.

Warren, M. A., Simis, S. G. H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., and Ansper, A. (2019). "Assessment of atmospheric

correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters." *Remote Sens. Environ.*, 225, 267–289.

Wetlands International. (2011). "Wetlands." *Microbe Wiki*, <https://microbewiki.kenyon.edu/index.php/Wetlands> (Apr. 26, 2023).

Wetlands International. (2014). "Wetlands: why should I care?" *Ramsar Fact Sheet*, <http://www.ramsar.org/sites/default/files/documents/library/factsheet1_why_should_i_care_0.pdf>.

Whitcomb, J., Moghaddam, M., McDonald, K., and Podest, E. (2009). "Mapping Canadian wetlands using L-band radar satellite imagery S." *Int. Geosci. Remote Sens. Symp.*

Wilson, J. P., and Gallant, J. C. (2000). "Digital terrain analysis." *Terrain Anal. Princ. Appl.*, J. P. Wilson and J. C. Gallant, eds., New York: John Wiley & Sons, Inc, 1–27.

Winsemius, H. C., Beek, L. P. H. Van, Jongman, B., Ward, P. J., and Bouwman, A. (2013). "A framework for global river flood risk assessments." *Hydrol. Earth Syst. Sci*, 17, 1871–1892.

Witzany, J., Cejka, T., and Zigler, R. (2008). "Failure Resistance of Historic Stone Bridge Structure of Charles Bridge. II: Susceptibility to Floods." *J. Perform. Constr. Facil.*, 22(2), 83–91.

Wright, J. M. (2008). *FLOODPLAIN MANAGEMENT PRINCIPLES AND CURRENT PRACTICES*. TN, USA.

Wu, Q., Li, H. qing, Wang, R. song, Paulussen, J., He, Y., Wang, M., Wang, B. hui, and Wang, Z. (2006). "Monitoring and predicting land use change in Beijing using remote sensing and GIS." *Landsc. Urban Plan.*, 78(4), 322–333.

Xing, H., Hou, D., Wang, S., Yu, M., and Meng, F. (2021). "O-LCMapping: a Google Earth Engine-based web toolkit for supporting online land cover classification." *Earth Sci. Informatics*, 14, 529–541.

Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., and Chen, F. (2022). "A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges." *Remote Sens.*, 14(8), 1770.

Yang, S.-R. (2017). "Assessment of Rainfall-Induced Landslide Susceptibility Using GIS-Based Slope Unit Approach." *J. Perform. Constr. Facil.*, 31(4), 04017026.

Yang, X. (2011). "Parameterizing Support Vector Machines for Land Cover

Classification." *Photogramm. Eng. Remote Sens.*, 77(1), 27–37.

Yang, X., Zheng, X. Q., and Lv, L. N. (2012). "A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata." *Ecol. Modell.*, 233, 11–19.

Yang, Y., Hou, C., Lang, Y., Sakamoto, T., He, Y., and Xiang, W. (2020). "Omnidirectional Motion Classification with Monostatic Radar System Using Micro-Doppler Signatures." *IEEE Trans. Geosci. Remote Sens.*, 58(5), 3574–3587.

Yariyan, P., Avand, M., Abbaspour, R. A., Torabi Haghighi, A., Costache, R., Ghorbanzadeh, O., Janizadeh, S., and Blaschke, T. (2020). "Flood susceptibility mapping using an improved analytic network process with statistical models." *Geomatics, Nat. Hazards Risk*, 11(1), 2282–2314.

Yepez, S., Laraque, A., Martinez, J. M., Sa, J. De, Carrera, J. M., Castellanos, B., Gallay, M., and Lopez, J. L. (2018). "Retrieval of suspended sediment concentrations using Landsat-8 OLI satellite images in the Orinoco River (Venezuela)." *Comptes Rendus Geosci.*, 350(1–2), 20–30.

Yu, D., Xie, P., Dong, X., Hu, X., Liu, J., Li, Y., Peng, T., Ma, H., Wang, K., and Xu, S. (2018). "Improvement of the SWAT model for event-based flood simulation on a sub-daily timescale." *Hydrol. Earth Syst. Sci.*, 22(9), 5001–5019.

Yunus, A. P., Masago, Y., and Hijioka, Y. (2020). "COVID-19 and surface water quality: Improved lake water quality during the lockdown." *Sci. Total Environ.*, 731, 139012.

Zhang, G., Feng, G., Li, X., Xie, C., and Pi, X. (2017). "Flood Effect on Groundwater Recharge on a Typical Silt Loam Soil." *Water*, 9(7), 523.

Zhang, K., Wang, Q., Chao, L., Ye, J., Li, Z., Yu, Z., Yang, T., and Ju, Q. (2019a). "Ground observation-based analysis of soil moisture spatiotemporal variability across a humid to semi-humid transitional zone in China." *J. Hydrol.*, 574, 903–914.

Zhang, K., Wang, S., Bao, H., and Zhao, X. (2019b). "Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China." *Nat. Hazards Earth Syst. Sci.*, 19(1), 93–105.

Zhang, M., Chen, F., Liang, D., Tian, B., and Yang, A. (2020). "Use of Sentinel-1 GRD SAR Images to Delineate Flood Extent in Pakistan." *Sustainability*, 12(14), 5784.

Zhang, Y., Pulliainen, J. T., Koponen, S. S., and Hallikainen, M. T. (2003). "Water

quality retrievals from combined Landsat TM data and ERS-2 SAR data in the Gulf of Finland." *IEEE Trans. Geosci. Remote Sens.*, 41(3), 622–629.

Zhu, Z., and Zhang, Y. (2022). "Flood disaster risk assessment based on random forest algorithm." *Neural Comput. Appl.*, 34, 3443–3455.

Zwart, J. A., Sebestyen, S. D., Solomon, C. T., and Jones, S. E. (2017). "The Influence of Hydrologic Residence Time on Lake Carbon Cycling Dynamics Following Extreme Precipitation Events." *Ecosystems*, 20(5), 1000–1014.

PUBLICATIONS

Journals

Parthasarathy Kulithalai Shiyam Sundar and Paresh Chandra Deka (2021). "Spatio-temporal classification and prediction of land use and land cover change for the Vembanad Lake system, Kerala: a machine learning approach." *Environmental Science and Pollution Research*, 29(57), 86220–86236. https://doi.org/10.1007/s11356-021-17257-0

Parthasarathy, K. S. S., and Subrahmanya Kundapura (2023). "Spatial Mapping of Flood Susceptibility Using Decision Tree–Based Machine Learning Models for the Vembanad Lake System in Kerala, India." *Journal of Water Resources Planning and Management (ASCE),* 149(10), 04023052. https://doi.org/10.1061/JWRMD5/WRENG-5858

Parthasarathy, K. S. S., and Subrahmanya Kundapura (2023). "Spatiotemporal variation in the water quality of Vembanad Lake, Kerala, India: a remote sensing approach." *Environmental Monitoring and Assessment*, 195(9). https://doi.org/10.1007/s10661-023-11746-0

Conference

Parthasarathy K.S.S., and Subrahmanya Kundapura (2022). "Mapping of flood inundated urban regions using Sentinel-1 SAR imagery for the 2018 and 2019 Kerala floods." Proceedings of 2nd International Conference on River Corridor, Research & Management, Organized by IIT Guwahati and IIT Jammu on 30th May – 1st June 2022. Pp. 88.

Presented in Conference

Parthasarathy K.S.S., and Subrahmanya Kundapura (2023). "Mapping of 2018 flood and estimation of future flood inundation region for Vembanad Lake System in Kerala, India using Sentinel-1 SAR imagery." *Sustainable Infrastructure: Innovation, Opportunities and Challenges*, Organized by National Institute of Technology Karnataka Surathkal on 20th-21st April 2023.

Book Chapters

Parthasarathy, K. S. S., and Kundapura, S. (2023). "Mapping of Flood-Inundated Urban Regions Using Sentinel-1 SAR Imagery for the 2018 and 2019 Kerala Floods." *Recent Dev. River Corridor Manag. Lect. Notes Civ. Eng.,* S. Dutta and V. Chembolu, eds., Springer, Singapore, 279–292. https://doi.org/10.1007/978-981-99-4423-1_20

Parthasarathy K.S.S., and Subrahmanya Kundapura (2023). "Mapping of 2018 flood and estimation of future flood inundation region for Vembanad Lake System in Kerala, India using Sentinel-1 SAR imagery." **(Under Review)**.

Award

Parthasarathy K.S.S., and Subrahmanya Kundapura (2022). "Mapping of flood inundated urban regions using Sentinel-1 SAR imagery for the 2018 and 2019 Kerala floods." 2nd International Conference on River Corridor, Research & Management, Organized by IIT Guwahati and IIT Jammu on 30th May – 1st June 2022 - **BEST PRESENTATION AWARD.**

BIODATA

Name	**Parthasarathy K. S. S.**
Date of Birth	12.06.1993
Address	Plot 11, Renga nagar,
	Vaishnavi Avenue,
	M. Gudalur,
	Kambarasasmpettai P.O.
	Tiruchirappalli – 620 101
	Tamil Nadu
Qualification	M. Tech (Geo-Technology and Geo-Informatics)
Mobile	+91-8973533057
E-mail	kssparthasarathy@gmail.com
Publications	Journals: 11
	Book Chapters: 7
	Conferences: 5
Awards/achievements	Best presentation award for 2nd International Conference on River Corridor, Research & Management.
	JALVIGYAN PURASKAR 2022 award for best paper published in ISH Journal 2022.