Flood Prediction Alert System in the state of Odisha, India:

An Explainable and Actionable AI based predictive model

By

Prajnajit Mohanty

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

September, 2024

Flood Prediction Alert System in the state of Odisha, India:

An Explainable and Actionable AI based predictive model

by

Prajnajit Mohanty

Under the Guidance from

Prof. (Dr.) Kishore Kunal

APPROVED BY

____

Dissertation chair

RECEIVED/APPROVED BY:

____

Admissions Director

DECLARATION

I hereby declare that the thesis entitled **"Flood Prediction Alert System in the state of Odisha, India: An Explainable and Actionable AI based predictive model "** submitted to SSBM, Geneva for the award of degree of Global Doctor of Business Administration is my original research work. This thesis or any part thereof has not been submitted partially or fully for the fulfillment of any degree of discipline in any other University/Institution.

(PRAJNAJIT MOHANTY)

Table of Contents

List of Figures

ACKNOWLEDGEMENTS

This journey has been one marked by perseverance, support, and unwavering encouragement from many individuals. First and foremost, I would like to extend my deepest gratitude to my advisor and mentor Prof. Dr. Kishore Kunal, whose expertise, wisdom, and steadfast guidance have been pivotal in the completion of this thesis. Your insights and constructive feedback have significantly shaped the direction and quality of this work, and your support has been instrumental in not only my academic but also in personal growth.

I dedicate the result of this work to my whole family. First of all, to my father Late Prof. Dr. Hemanta Kumar Mohanty and to my mother Late Smt. Indira Das, who taught me the power of education. To my wife Priyanka, for her relentless pursuit to encourage me to dedicate adequate time and energy to my research work whilst she takes care of our family. Words cannot express my appreciation for your unyielding belief in me, you always stand rock solid whom I can always count on. To my son Prajnaansh, your little word of encouragement always motivates me to go for that extra mile. To my mentor, friend and colleague Mikael Andersson for the encouragement to complete the research and always checking in on progress. Micke, you are someone like a guardian angel to me. To Abhijit Nanda, my mentor and guide who always pushes me to do my best and suggested to go for the research.

Finally, I am immensely grateful to the esteemed faculty members and staff of SSBM, whose dedication to excellence in education has provided me with a robust foundation and an inspiring environment for learning and research.

.

ABSTRACT

Floods are a recurrent and devastating natural disasters, cause immense loss of life, infrastructure damage, and economic hardship. Mitigating these impacts requires accurate and timely flood prediction. Traditional hydrological models, while valuable, often rely on complex physics and extensive data. This thesis explores the application of explainable machine learning (XAI) for flood prediction, offering a data-driven alternative with interpretable results.

Black-box machine learning models excel at prediction but lack transparency. XML techniques bridge this gap, enabling us to understand the model's decision-making process. This is crucial in flood prediction, where understanding the factors contributing to a predicted flood is critical for effective risk mitigation strategies.

Thesis Key Objectives:

I. To develop an explainable machine learning model for flood prediction. This model will utilize historical data on factors influencing floods, such as rainfall patterns, river discharge measurements, land use data, and digital elevation models (DEMs).

II. To evaluate the model's accuracy and explainability. The model's performance will be assessed using standard flood prediction metrics like Root Mean Squared Error (RMSE) and Nash-Sutcliffe Efficiency (NSE). XML techniques like SHAP (SHapley Additive exPlanations) will be employed to understand feature importance and gain insights into the model's predictions.

III. To compare the XAI model with traditional approaches. This will involve comparing the prediction accuracy, explainability, and computational efficiency of the XAI model with existing hydrological models.

IV. To open up new avenues for further research into this space. The idea is also to open new opportunities and avenues in this space of Flood Prediction using AI and with each passing day creating newer horizons for AI and generative AI, probably this research will open and inspire new researchers to come up with new, modern and more sophisticated techniques in future.

Expected Benefits:

I. Enhanced Flood Prediction Accuracy: The XML model is designed to deliver precise flood forecasts, providing critical lead time for disaster response and preparedness.

II. Improved Decision-Making Capabilities: By elucidating the key factors influencing flood forecasts, stakeholders can devise focused mitigation plans, ensuring the protection of at-risk areas and essential infrastructure.

III. Greater Transparency and Trust: XML promotes confidence in the model's forecasts by offering interpretable results, enabling stakeholders to make well-informed decisions based on transparent data.

Thesis Structure:

Chapter 1 introduces the topic of floods and their significant impact on society and the environment and highlights that among natural disasters, floods are the most frequent and affect over half a billion people annually worldwide, with projections suggesting this could rise to two billion by 2050, particularly impacting Asia. Chapter 2 provides a comprehensive literature review on flood prediction methods, highlighting the limitations of traditional models and the potential of XAI in this domain. Chapter 3 delves into the chosen XAI algorithm, explaining its theoretical background and suitability for flood prediction. Chapter 4 details the data acquisition, preprocessing, and feature engineering techniques employed. Chapter 5 focuses on model development, training, and hyperparameter tuning. Chapter 6 presents the model evaluation, including accuracy metrics and explainability analysis using XAI techniques, compares the XAI model with traditional approaches, analyzing their strengths and weaknesses including the challenges faced, potential future improvements, and the broader suggested implementation of the XAI for flood prediction.

Conclusion:

This thesis explores the potential of XAI for accurate and interpretable flood prediction. By leveraging XAI, we can move beyond black-box models and gain valuable insights into the factors driving flood events. This knowledge empowers stakeholders to make informed decisions and develop targeted mitigation strategies, ultimately contributing to a safer future in flood-prone regions.

KEYWORDS

Artificial Intelligence

Machine Learning

Explainable Machine Learning

Flood Prediction

Artificial Neural Network

Odisha

Flood Warning System

Riverbed Flooding

Reservoir

Ground Water Level

Hydrometeorology

Rainfall Variability

LIST OF ABBREVIATIONS

AI: Artificial Intelligence

XAI: Explainable Artificial Intelligence

RMSE: Root Mean Squared Error

NSE: Nash-Sutcliffe Efficiency

SHAP: SHapley Additive exPlanations

DEMs: Digital Elevation Models

IMD: Indian Meteorological Department

IOT: Internet of Things

ANN: Artificial Neural Network

DNN: Deep Neural Network

SVM: Support Vector Machines1

NARX NN: Nonlinear Autoregressive Network with Exogenous Inputs2

DAX: Data Analysis Expressions

EDA: Exploratory Data Analysis

GFDS: The Global Flood Detection System

OFDA/CRED: Overseas Development Institute/Centre for Research on the Epidemiology of Disasters

UNU: United Nations University

GFDS: Global Flood Detection System

JRC: Joint Research Centre (European Commission)

CHAPTER 1 – INTRODUCTION

1.1. An Overview:

We are all aware of ‘flood’ and the impact it can have on society & environment. Some of the devasting floods have proved to have left an irreparable and irreversible loss and impact on humankind, society, ecological & geographical balance.

As per the study done by the European Commission led Joint Research Centre (2009) as a part of Developing a The Global Flood Detection System (GFDS) - Of all natural disasters the floods are most frequent (46%) and cause most human suffering and loss (78% of population affected by natural disasters). They occur twice as much and affect about three times as many people as tropical cyclones. While earthquakes kill more people, floods affect more people (20000 affected per death compared to 150 affected per death for earthquakes) (OFDA/CRED, 2006).

A study of the United Nations University (2004) shows that floods impact over half a billion people every year worldwide and might impact two billion by 2050, of which a disproportionate number live in Asia (44% of all flood disasters worldwide and 93% of flood-related deaths in the decade 1988-1997).

In general, one third of humanitarian aid goes to flood related disasters.

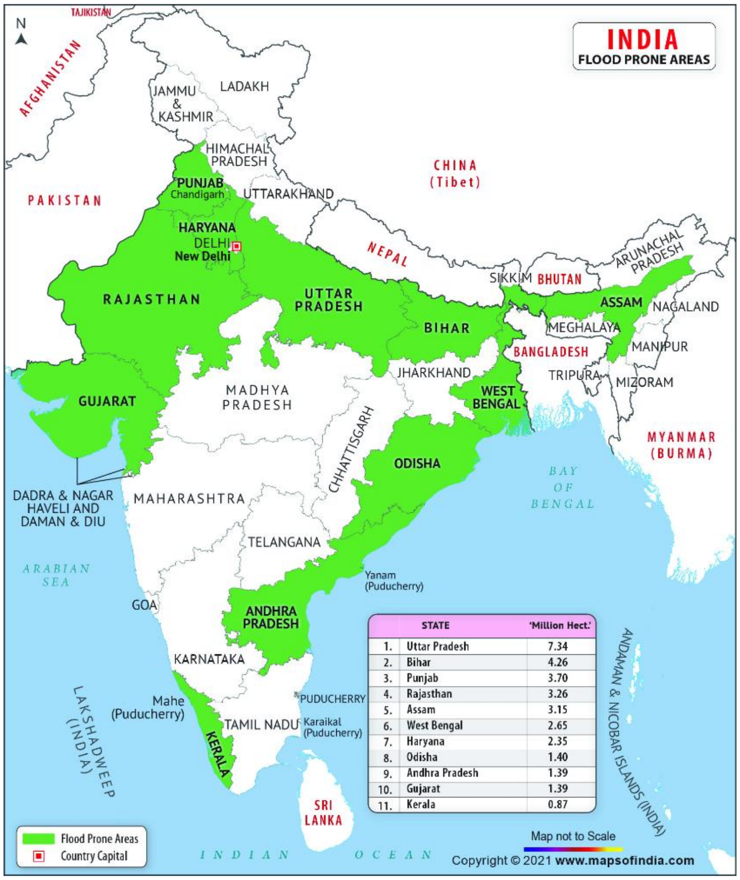

Figure 1.1: Top 10 Flood Prone Areas of India

1.2. Geographical Introduction of Indian Sub-continent: A Land of Diverse Landscapes

The Indian sub-continent, a vast expanse stretching from the snow-clad peaks of the Himalayas in the north to the coastal plains in the south, is renowned for its diverse geographical features and weather patterns. This region is characterized by a variety of landscapes, including extensive river basins, arid deserts, fertile plains, and dense forests.

The northern boundary of the sub-continent is dominated by the Himalayan mountain range, which not only serves as a natural barrier but also significantly influences the climate of the region. The Himalayas intercept the monsoon winds, causing heavy rainfall on the southern slopes while creating a rain shadow effect on the northern side. This results in a stark contrast between the lush, rain-fed valleys of the southern Himalayas and the arid, windswept plateaus of Tibet.

During the monsoon season, which typically spans from June to September, the Indian sub-continent experiences the majority of its annual rainfall. The southwest monsoon winds, originating from the Indian Ocean, bring torrential rains to the western coast, the Western Ghats, and the northeastern states. This period of intense precipitation is crucial for the agricultural sector, replenishing water sources and sustaining crops. However, the monsoon can also lead to severe flooding, particularly in low-lying areas and along major river systems like the Ganges, Brahmaputra, and Mahanadi.

The variability in weather conditions across the sub-continent is further influenced by other mountain ranges such as the Western and Eastern Ghats, the Aravalli Range, and the Vindhya Range. These ranges create microclimates, affecting local weather patterns and contributing to the overall climatic complexity of the region. For instance, the Western Ghats receive heavy rainfall, supporting lush evergreen forests, while the interior Deccan Plateau remains relatively dry.

Flooding in the Indian sub-continent is a multifaceted issue, often exacerbated by the interaction of geographical and climatic factors. The convergence of monsoon rains, snowmelt from the Himalayas, and the region's extensive river networks can lead to catastrophic floods, affecting millions of people and causing widespread damage to infrastructure and agriculture. The

Brahmaputra and Ganges river basins are particularly prone to flooding due to their expansive catchment areas and the high volume of water flow during the monsoon season.

In summary, the geographical and climatic diversity of the Indian sub-continent plays a pivotal role in shaping the weather patterns and flood dynamics of the region. From the towering Himalayas to the coastal plains, each geographic feature contributes to the intricate mosaic of conditions that characterize this part of the world.

1.3. Geographical Introduction of Odisha: A land of rivers and deltas

Odisha, a state nestled in the eastern part of the Indian peninsula, boasts a rich tapestry of geographical regions and climates. Spanning roughly between 17°49'N and 22°36'N latitudes and 81°36'E and 87°18'E longitudes, it covers an impressive 1,55,707 sq.km and is administratively divided into 30 districts.

Geographical Tapestry

Odisha's diverse landscape can be broadly divided into four distinct regions:

i. The Northern Plateau:

   This region, encompassing districts like Mayurbhanj, Keonjhar, and Sundargarh, is an undulating upland frequently intersected by hill ranges. The elevation gradually slopes southward, with the central section reaching an average of 900 meters above sea level.

ii. The Central River Basins:

Nestled between the northern plateau and the eastern hills lie the fertile central river basins. This region includes districts like Balangir, Sambalpur, and Angul, where major rivers like Mahanadi and Brahmani flow, nourishing the land.

iii. The Eastern Ghats:

Forming the final stretch of the majestic Eastern Ghats mountain range, these elevated hills (around 900 meters) traverse districts like Koraput and Kalahandi in a northeast-southwest direction for about 250 km.

iv. Coastal Delights:

Bordering the Bay of Bengal, the coastal plains offer a contrasting landscape. This region, comprising districts like Balasore, Puri, and Cuttack, experiences a more moderate climate due to the maritime influence. The Bay of Bengal's presence also contributes to the region's fertile soil, making it agriculturally productive.

Climatic Variations:

The coastal plains are not just picturesque but also vital to the state's economy, mainly due to their agricultural productivity. The moderate climate, characterized by mild winters and moderately hot summers, allows for the cultivation of various crops, including rice, which is a staple in the region. The monsoon season, occurring from June to September, brings significant rainfall to the coastal areas, essential for agriculture but also occasionally causes flooding.

In recent years, the coastal regions have also experienced the impacts of climate change. Rising sea levels and increased frequency of cyclonic storms have become critical concerns. The coastal districts, particularly Balasore, Puri, and Cuttack, have had to adapt to these changes, implementing advanced warning systems and constructing cyclone shelters to mitigate the effects of such natural disasters.

Odisha's varied geography ensures a rich tapestry of climatic conditions, with the coastal plains standing out for their relatively stable weather patterns. The Bay of Bengal's maritime influence acts as a buffer, moderating temperatures and providing the region with its characteristic fertility. This makes the coastal plains a critical area for both ecological balance and economic sustenance.

Odisha's climate varies across these regions. The interior experiences distinct seasons, with hot summers and cold winters. In contrast, the coastal plains enjoy a more equable climate throughout the year, thanks to the Bay of Bengal's influence.

Unveiling Recent Climate Trends: A Focus on the Deltaic Region

Understanding climate patterns is crucial for effective management and adaptation strategies. While numerous studies have analyzed rainfall patterns and trends in India over the past 100 years or more (Guhathakurta & Rajeevan, 2008), limited research has focused specifically on Odisha's recent climate changes.

Odisha shares a coastline of approximately 480 KMs with the Bay of Bengal. The state of Odisha has witnessed a number of extreme climatological events in the last few decades. After the 1999 Super cyclone, the state witnessed major floods in 2003, 2008, 2011 and 2013 apart from many smaller ones. The deltaic area which is home to three major rivers namely the Mahanadi, the Brahmani, and the Baitarani, has been a cause of major worry and where flood waters intermingle, and when in spate simultaneously, wreak considerable havoc. This problem becomes even more acute when floods coincide with high tide. The water level rises due to deposits of silt on the riverbed. Rivers often overflow their banks or water rushes through new channels causing heavy damage. Floods and drainage congestion also affect the lower reaches along the Subarnarekha. The rivers Rusikulya, Vansadhara and Budhabalanga also cause occasional floods. Over the past few decades, Odisha has witnessed a series of extreme climatological events, underscoring the need for a comprehensive understanding of its climate patterns, particularly in the deltaic region. This region, home to three major rivers—the Mahanadi, the Brahmani, and the Baitarani—has been particularly vulnerable to climate-induced hazards such as floods.

Historical Context and Recent Climatic Trends

Odisha's climate history is marked by significant events that have shaped its current landscape. The devastating Super Cyclone of 1999 is a notable example, followed by major floods in 2003, 2008, 2011, and 2013. These events have had profound impacts on the deltaic region, where the convergence of major rivers exacerbates flood risks when they overflow or change their channels.

Recent research has highlighted the rising frequency and intensity of extreme weather events in Odisha. Studies on rainfall patterns and trends over the past century have shown significant variations, but specific research on Odisha’s recent climate changes remains limited. A deeper focus on the deltaic region is crucial, given its unique geographical and hydrological characteristics.

Deltaic Region: A Hotspot for Climatic Changes

The deltaic region of Odisha is characterized by its flat terrain and the presence of the Mahanadi, Brahmani, and Baitarani rivers. This area is highly susceptible to flooding, particularly when heavy rainfall coincides with high tides. The silt deposits on the riverbeds raise water levels, often leading to the rivers overflowing their banks or creating new channels, causing extensive damage.

The confluence of the Mahanadi, Brahmani, and Baitarani rivers exacerbates the flood situation. During periods of high tide, the floodwaters find it difficult to drain into the Bay of Bengal, leading to prolonged waterlogging and drainage congestion. This is particularly evident in the lower reaches along the Subarnarekha river and the occasional floods caused by the Rusikulya, Vansadhara, and Budhabalanga rivers.

One of the simple and most common reasons for flooding has been the overbank flow of water due to extreme rainfall at the upper end and catchment areas of these major rivers. All the major rivers of Odisha attain their old stage in the coastal flat region, contributing to the flooding problems.

Changing Rainfall Patterns

One of the most significant climatic changes observed in Odisha is the alteration in rainfall patterns. While the state has always experienced seasonal monsoons, recent data suggests an increase in the intensity and variability of rainfall. This has led to more frequent and severe flooding events, particularly in the deltaic region.

Studies have shown that the monsoon season is becoming more unpredictable, with periods of intense rainfall followed by dry spells. This pattern disrupts the traditional agricultural practices in the region, as farmers struggle to cope with the changing conditions. The increased rainfall also contributes to the siltation of rivers, further aggravating the flood risk.

The unpredictability of the monsoon season has led to a series of challenges for the local population, particularly farmers who rely on consistent weather patterns for their crops. This unpredictability also affects water management systems, making it difficult to anticipate and prepare for extreme weather events.

Efforts to Mitigate Climatic Impacts

Recognizing the growing threat of climate change, the state government of Odisha has taken several measures to mitigate the impacts of extreme weather events. These include the construction of embankments and flood control structures, as well as the implementation of early warning systems to alert communities about impending floods.

In addition, efforts are being made to improve the resilient of the deltaic region through sustainable land and water management practices. Reforestation projects and the restoration of wetlands are being undertaken to enhance the natural buffering capacity of the landscape. These initiatives aim to reduce the vulnerability of the region to floods and other climate-related hazards.

There has also been a push towards community-based disaster management, where local communities are trained and equipped to respond effectively to climatic events. This includes the establishment of local task forces and the dissemination of information on best practices for disaster preparedness.

Adaptive Strategies for the Future

Moving forward, it is essential to adopt adaptive strategies that can help communities in the deltaic region cope with the changing climate. This includes investing in climate-resilient infrastructure, such as elevated roads and flood-resistant buildings, to reduce the impact of floods. Additionally, promoting diversified livelihoods can help reduce the dependence on agriculture, which is highly vulnerable to climatic variations.

Public awareness campaigns and capacity-building programs are also crucial to empowering local communities to take proactive measures in response to climatic changes. By fostering a culture of preparedness and resilience, Odisha can better navigate the challenges posed by a changing climate.

Furthermore, collaboration with academic institutions and international organizations can provide access to cutting-edge research and technologies that can aid in better understanding and managing the climatic changes in the region. This collaboration can also lead to the development of innovative solutions tailored to the unique challenges faced by the deltaic region.

The recent climatic changes in Odisha, particularly in the deltaic region, highlight the urgent need for comprehensive and adaptive strategies to manage the increasing risks of extreme weather events. By understanding the shifting climate patterns and implementing effective mitigation and adaptation measures, Odisha can safeguard its communities and ensure a sustainable future for generations to come.

The state's experience serves as a valuable case study for other regions facing similar climatic challenges, demonstrating the importance of resilience and proactive planning in the face of a changing climate.

In conclusion, the path forward requires a multi-faceted approach that includes scientific research, government intervention, community participation, and international collaboration. Only by working together can we hope to mitigate the impacts of climate change and build a more resilient and sustainable future for Odisha and beyond.

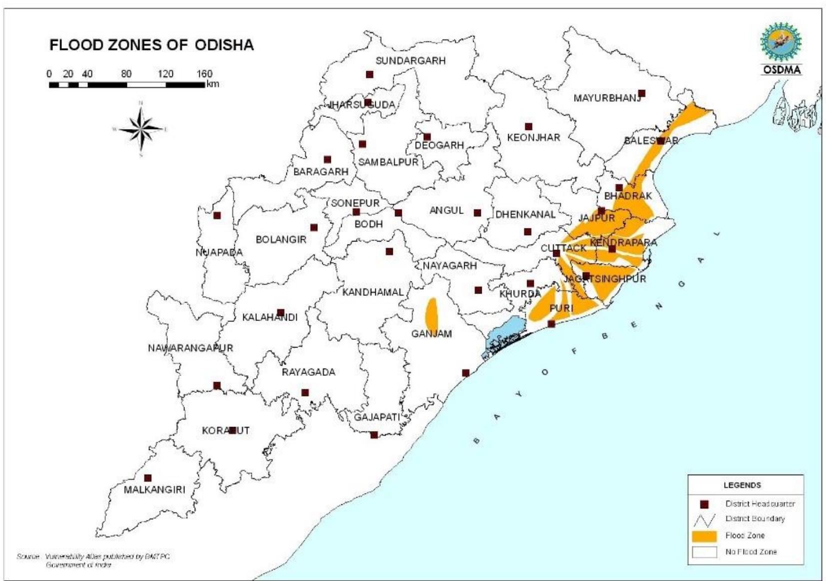

Figure 1.2: Flood zones of Odisha

One of the simple and most common reasons has been the overbank flowing of water due to extreme rainfall at the upper end and catchment areas of the major rivers like Mahanadi. All the major rivers of Odisha attain their old stage in the coastal flat region

There have been many major attempts since 1855 by various central governmental agencies to find the cause and to diminish the impacts by way of recommendations but unfortunately, the floods are a common phenomenon in this region which could only be dealt in form of aftermath but not in terms of early major reactive measures.

With the use of technology and IMD (Indian Meteorological Department) Forecasts, which also sometime may not be able to accurately time the flood, the after effect of the floods are reduced as compared to the recent history but still the loss in terms of human lives, livestock, agriculture, property etc. has been nothing short of disasters.

Figure 1.3: Wind & Cyclone zones -Odisha 1

Looking at the Wind & Cyclone map of Odisha, in the above figure #3, which clearly shows a correlation with the Flood Map of Odisha in figure 2, as often the flood follows the cyclone in many past instances. Hence it is imperative to look at the cyclone forecasting as well in order to come up with the flood forecasting model.

Hence this calls for greater use of sophisticated forecasting techniques using the various heterogenous data at our disposal, to come up with early warning system that would help get the alerts in time and help to prepare to deal with the flood which is inevitable. The more we get the time before the flood hits, the more time effective it becomes and the lesser the loss of human lives and other irreversible losses.

1.4. Artificial Intelligence – A game changer

Artificial intelligence (AI) stands at the forefront of technological innovation, a driving force that is reshaping industries, economies, and even our daily lives. From its humble beginnings as a theoretical concept explored by visionaries and computer scientists, AI has evolved into a powerful tool that can emulate human cognitive functions. The capabilities of AI range from simple automation processes to complex decision-making and predictive analytics.

In its essence, AI is designed to process information in a way that mimics human intelligence. This involves learning from experiences, adapting to new inputs, and performing tasks that typically require human intellect. The versatility of AI systems allows them to be applied in various domains, from healthcare and finance to transportation and entertainment.

In recent years, the advent of Artificial Intelligence (AI) has revolutionized various sectors, offering unprecedented capabilities and transforming traditional methods. AI refers to the simulation of human intelligence in machines, enabling them to perform tasks that typically require human intellect, such as learning, reasoning, problem-solving, and decision-making.

One of the most groundbreaking aspects of AI is its ability to process vast amounts of data with incredible speed and accuracy. Leveraging techniques such as machine learning and deep learning, AI systems can identify patterns, make predictions, and even improve their performance over time without explicit programming. This adaptability makes AI particularly valuable in complex and data-intensive fields.

AI's game-changing potential lies in its versatility and scalability. It can integrate and analyze diverse data sources, including structured, unstructured, and real-time streaming data from sensors and IoT devices. This capability is crucial for developing robust models that can foresee and mitigate natural disasters, like floods, by providing timely and accurate predictions.

1.5. Why Artificial Intelligence in Flood Prediction

With the evolution of sophisticated hardware in recent times, achieving supreme computing power and processing data to the tune of gigabytes or petabytes are getting very common these days. Also, with the popularity of cloud computing soaring high, the affordability and scalability of Artificial Intelligence is significantly democratizing.

This also makes it very easy to combine, blend and analyze not only structured or unstructured data but also streaming data like data coming out of sensor or IOT devices. This makes it more special in our case when we are going to leverage huge data which are images or coming out of sensor.

Artificial Intelligence (AI) helps to establish the correlation between a variety of causes and the impact with the help of data and intensive patterns. There are already variety of related data available to us, some of them are very much evident as obvious for example- satellite image data , weather forecast in terms of rains or storms, underground water levels, wireless sensor data from the riverbed etc. but there may be some other data which may not seem to have a direct correlation but could have a long term and strong causal impact to the frequent occurring of floods , for example – urbanization near deltaic regions, periodic cultivation in the region, in terms of satellite images etc. Hence these could be very important from a disaster management perspective and for a better effective solution.

With the advancement of various Artificial Intelligence tools and techniques including the very recent Generative Artificial Intelligence, it will be a great boon if we are able to establish the root cause of the floods, their frequencies and device an early warning system which can prove a history changing move!

1.6. Explainable Artificial Intelligence Model- What and Why:

It is perceived in general and many of the Artificial Intelligence model are 'black box' that is, the model takes inputs and provides the output with no way for a human to understand the justification and rationale behind the output. This is especially significant in case of Deep Learning which is an Artificial Neural Network (ANN) based models which is an advanced study in the field of Artificial Intelligence, which is used for processing huge heterogeneous data, where there is no way to figure out the outcome with the inputs provided.

The soul objective of an Explainable Artificial Intelligence (XAI) model is to provide that rationale of the output with a function of the input(s) or to make the XAI model a 'White Box'. This makes it easy to understand the causes and helps to create the 'TRUST' on the model, albeit sometimes with a trade off with accuracy.

This is an important aspect in this topic, especially most of the advanced research which suggested machine learning based model for flood alert system, are based on Artificial Neural Network.

Hence, there were dual objectives of using an XAI model for this research:

I. This would present a 'White Box' or transparent machine learning model, which should clearly be able to not only establish the root causes but also can create the 'Trust' on the model and variations on input data on the predicted result.

II. The second and one of the unique features of this study will be able to open the doors for next set of future researches in Artificial Intelligence, using the input parameters in a transparent manner and with the evolution of technology paradigms, this should lead to further increase in terms of accuracy of model and lead time in terms of prediction.

1.7. Artificial Intelligence in Flood Prediction- A pragmatic approach with Actionable Insights

A key aspect of this research was to come up with clear actionable insights. This is important from the usability and deriving value from the research. Actionable insight would include the key variables contributing to the flood prediction in a measurable way so that it would be possible to identify the pragmatic mitigating actions. For any AI based model to be successful, clearly articulated actionable insights are required. Actionable insights are vital because they translate complex data and predictions into concrete steps that can be taken to mitigate risks and manage disasters effectively. They help in:

I. To minimize the aftermath impact of floods in terms of resources.

II. Contain critical factors responsible for the floods in a staggered way so that the very occurrences of the flood would tend to slow down.
III. Making informed decisions based on clear, understandable data.
IV. Prioritizing resources and efforts in areas that will have the most significant impact.
V. Ensuring that strategies and actions are based on verifiable data, increasing their likelihood of success.
VI. Creating a feedback loop where the outcomes of actions inform future predictions and strategies, continuously improving the system.

There were some attempts already to leverage the machine learning techniques to predict the floods well in advance, taking into consideration of various potential factors. But none of them have so far has been satisfactorily successful, although some of them have been able to open next door in this direction.

It is worth mentioning here that – the reasons of floods can differ from geography to geography and hence the technique and approach. It will be very improbable to generalize the causes of floods happening around the world, though some may seem obvious but predicting the obvious will be a difficult question.

CHAPTER 2 - LITERATURE REVIEW

2.1. Introduction

Flood Prediction with the help of Artificial Intelligence has been on the fray for quite some time, especially since last decade. There have been some great attempts made to develop an early warning system with various sophisticated techniques. The author did a literature review to have a better knowledge of the current state of research into developing a flood early warning system with the help of Machine Learning & Artificial Intelligence. Some of the past attempts were mostly generic and some of them have been able to open up the next steps in this direction. The usual approach was mostly to identify the template based on the some of the generic and mostly obvious reasons, for example rainfall, ground water levels, water levels in riverbeds, reservoirs capacity etc. This approach was also helpful to some extent, but the veracity of prediction and post impact of floods definitely warrants a fresh approach, especially as a society we have embarked into a new era of Artificial Intelligence, more towards an Artificial General Intelligence (AGI) era. It will be improbable to generalize the causes of floods happening around the world, though some may seem obvious but predicting the obvious will be a difficult ask.

The new innovations both in terms of data availability with better quality and the newer techniques in Machine Learning, would support the new research methodologies but it is definitely worth reflecting on the existing works already done so that a fresh perspective can be taken based on the past works and would help in identifying the approach for the new direction. This literature review intended to serve as a solid foundation for understanding the variety of possible factors for floods across geographies, their variability to the outcome along with shortcomings. This also provides how diverse the causes and potential reasons can be, in different geo and their demographics. It is worth mentioning here that – the reasons of floods can differ from geography to geography and so the availability of different potential data source. Hence the technique and approach also differ as per data and geography.

The literature review becomes more effective when the literature which is processed and reviewed is in the same line of thoughts and scope. So, a clear thoughtful approach as mentioned was devised and the literature were carefully selected as per the relevance and criteria. It was also important not to reinvent the wheel rather than to build on already the great work done and to take a leaf or two from the past.

Several critical factors were considered to ensure that the literature selected was relevant and of high quality. By systematically reviewing existing literature, researchers are able to:

Identify Gaps in Knowledge:

A thorough literature review helps in uncovering the areas where research is lacking or where there are inconsistencies in the findings. This identification of gaps is crucial as it directs future research efforts towards addressing these unknowns or resolving discrepancies.

Build on Existing Work:

Rather than starting from scratch, a literature review allows researchers to build upon the work that has already been done. This not only saves time and resources but also enables the integration of previous findings into new research, fostering cumulative knowledge growth.

Avoid Reinventing the Wheel:

By understanding what has already been studied, researchers can avoid duplicating efforts. This ensures that time and resources are utilized more efficiently, focusing on novel contributions rather than redundant studies.

Establish a Theoretical Framework:

A well-conducted literature review provides a solid theoretical foundation for the research. It helps in formulating hypotheses, defining research questions, and choosing appropriate methodologies by drawing insights from past studies.

Contextualize Findings:

Literature reviews place new research within the context of existing knowledge, allowing for a better understanding of how new findings fit into the broader academic landscape. This contextualization is essential for interpreting results and drawing meaningful conclusions.

Support Evidence-Based Practice:

In fields such as medicine, education, and social sciences, literature reviews are crucial for evidence-based practice. They synthesize findings from multiple studies to provide practitioners with a reliable basis for decision-making and policy formulation.

Facilitate Academic Dialogue:

By engaging with existing literature, researchers contribute to the ongoing academic dialogue. They critique, challenge, and complement previous work, advancing the collective understanding of the field.

The below important aspects were considered whilst looking at the existing studies -

Relevance to the Research Question

The primary criterion was the relevance of literature to the research question at hand. Only those studies that directly addressed or contributed to the topic of flood prediction and warning systems were included.

Credibility of Sources

Scholarly articles, peer-reviewed journals, and recognized conference papers were prioritized to ensure the credibility and reliability of the information. Sources from reputable institutions and authors with a strong background in the field were given preference.

Recency of Publications

To keep the review current and reflective of the latest advancements, emphasis was placed on recent publications. However, seminal works and influential earlier studies were also included to provide a comprehensive historical context.

Methodological Rigor

Studies employing robust and scientifically sound methodologies were selected. This included a focus on those utilizing advanced data analysis, machine learning models, and empirical validation.

Diversity of Perspectives

To capture a broad spectrum of insights, literature from different geographical regions, and encompassing various flood prediction methods and technologies, was considered. This ensured a more holistic understanding of the topic.

Interdisciplinary Approach

Given the multifaceted nature of flood prediction, literature that integrated perspectives from hydrology, meteorology, data science, and environmental engineering was included. This interdisciplinary approach enriched the depth and breadth of the review.

Employing these criteria not only ensured a thorough and relevant literature review but also facilitated the development of a well-rounded and informed perspective on flood prediction systems.

2.2. Flood Prediction Warming System based on Historical Data Analysis

In one of the earlier studies by Hill, J. M., et al(1987), a computerized geographic information system (GIS) was created in support of data requirements by a hydrologic model designed to predict the runoff hydrograph from ungagged basins. Using simple data analysis technique, in their research paper by Kugler, Z., and De Groeve, T (2007), they suggested the user of ultrasonic sensors to monitor the river water levels which can be used as an indicator for an early detection of flood. In their study by Wu, H. et al (2012), they spoke of an improved global flood detection system using Satellite Rainfall data and Hydrologic model. Research by Damle, C., & Yalcin, A. (2007) talks about flood prediction using time series data mining which combines chaos theory and data mining to characterize and predict events in complex, nonperiodic and chaotic time series. With the help of infrared-based satellite rainfall remote sensing data, Hossain, F. et al (2004) did an assessment flood prediction. Yoshimura, K. et al (2008) provided a statistical analysis approach based on 29 years of river discharge simulation data which puts the focus on river flows trends as one of the primary obvious causes which was great work of creating a correlation with the floods based on simulation. Sankaranarayanan, S. et al (2020) used Deep Neural Network (DNN) for predicting the occurrence of flood based on temperature and rainfall intensity, which also in addition, used comparison with other machine learning models (support vector machine (SVM), K-nearest neighbor (KNN) and Naïve Bayes) in terms of accuracy and error and concluded that indicate that the deep neural network can be efficiently used for flood forecasting with highest accuracy based on monsoon parameters only before flood occurrence. In their renowned paper Flood Forecasting with Machine Learning Technique on Hydrological Modeling, Noymanee, J., & Theeramunkong, T. (2019) presented hydrological modeling augmented with alternative five machine learning techniques; linear regression, neural network regression, Bayesian linear regression and boosted decision tree regression which used the testbed system, the so-called MIKE-11 hydrologic forecasting model, developed by Danish Hydraulic Institute (DHI), Denmark. This was an important step in using the strength of Machine Learning modeling and Hydrological modeling which was tested scientifically in a sophisticated test system thus paving the way for future studies in this area. In another multi-fold study, Maspo, N.A. et al (2020), proposed the approach in two clear steps; the first part was to identify flood prediction approaches specifically using Machine Learning methods and the second part is to identify flood prediction parameters that have been used as input parameters for flood prediction models. The conclusion of this paper was a steppingstone for the future studies in this direction as the main contribution of this paper was to determine the most recent ML techniques in flood prediction and identify the notable parameters used as model input so that researchers and/or flood managers can refer to the prediction results as the guideline in considering ML method for early flood prediction. In a unique approach of blending the crowd sourced data with available big data, Puttinaovarat, S., & Horkaew, P. (2020), proposed a Flood Forecasting System using Machine Learning Techniques - this paper proposed a novel flood forecasting system based on fusing meteorological, hydrological, geospatial, and crowdsource big data in an adaptive machine learning framework. This attempt focused on Data intelligence which was driven by state-of-the-art learning strategies and subjective and objective evaluations indicated that the developed system was able to forecast flood incidents, happening in specific areas and time frames. In their paper, Real-time flood disaster prediction system by applying machine learning technique, Keum, H. J. et al (2020), showcased a classification-based approach for real-time flood prediction model for urban areas, by combining a numerical analysis model based on hydraulic theory blending with a machine learning model. In this study, flood databases were constructed in advance for different rainfall scenarios using the Environmental Protection Agency-Storm Water Management Model (EPA-SWMM) and a two-dimensional inundation model and the flood depth data for each map grid were divided into five categories based on the average flood depth using the Latin hypercube sampling (LHS) and probabilistic neural network (PNN) classification techniques for higher-precision flood range prediction. In another attempt to build a flood forecasting system with machine learning models in an operational framework, Khalaf, M. et al (2019) showcased a Data Science approach to Flood Severity Prediction using Artificial Neural Network (ANN), the model had a feature to add the severity of the predicted flood. In another attempt using various machine learning modeling techniques, Syeed, M. M. A. et al (2022), with an objective to reduce the extreme risks of this natural disaster and also contributing to policy suggestions by providing a prediction for floods using different machine learning techniques namely Binary Logistic Regression, K-Nearest Neighbor (KNN), Support Vector Classifier (SVC) and Decision tree Classifier to provide an accurate prediction. Han, D. et al (2007) focused on Support Vector Machine (SVM) as an effective Machine Learning technique on flood forecasting and showcased that an optimum selection among a large number of various input combinations and parameters is a real challenge for any modelers in using SVMs, in this respect a comparison with some benchmarking models had been made, i.e. Transfer Function, Trend and Naive models. This study demonstrated that SVM is able to surpass all of them in the test data series, at the expense of a huge amount of time and effort. In another first to present a manifold model as a machine-learning alternative to hydraulic modeling of flood inundation, Nevo, S. et al (2021), showcased a forecasting system consists of four subsystems: data validation, stage forecasting, inundation modeling, and alert distribution, where Machine learning was used for two of the subsystems, Stage forecasting is modeled with the Long Short-Term Memory (LSTM) networks and the Linear models. Flood inundation was computed with the Thresholding and the Manifold models, where the former computes inundation extent and the latter computes both inundation extent and depth. It was a great attempt in this new approach and paved the way for new research in this field of using the best manifold modeling techniques with innovative data approach. In an innovative approach in the flood forecasting space, Khalaf, M. et al (2020) to create an IoT-Enabled Flood Severity Prediction via Ensemble Machine Learning Models, this approach explained a new approach for the prediction of water level in association with flood severity using the ensemble model which leverages the latest developments in the Internet of Things (IoT) and machine learning for the automated analysis of flood data. Dodangeh, E. et al (20209 proposed a very interesting approach for flood prediction integrating machine learning and resampling algorithms, where multi-time resampling approaches, random subsampling (RS) and bootstrapping (BT) algorithms, integrated with machine learning models: generalized additive model (GAM), boosted regression tree (BTR) and multivariate adaptive regression splines (MARS). In this study the RS and BT algorithms provided 10 runs of data resampling for learning and validation of the models which provided effective outcome in flood prediction, this was a new approach in terms of data sampling feeding into Machine Learning model and paved the way for newer research in this space. Another innovative technique integrating the long short-term memory (LSTM) machine learning model along with reduced order model (ROM) framework proposed by Hu, R. et al (2019) which was found to have the capability of representing the spatio-temporal distribution of floods since it takes advantage of both ROM and LSTM. This was a unique combination involving the trustfulness of LSTM technique in another innovative way. In a very interesting approach combining using bivariate statistics and machine learning techniques, Costache, R. (2019) used Fuzzy Support Vector Machine ensemble for the first time and the validation of the model output was carried away using ROC Curve model. In their attempt in coming up with a Flood Probability Maps (FPMs) , Avand, M. et al (2023) used Digital Elevation Model (DEM) for analysis and assess the influence of the spatial resolution of the DEMs 12.5 m (ALOS PALSAR) and 30 m (ASTER) on the accuracy of flood probability prediction using three machine learning models (MLMs), including Random Forest (RF), Artificial Neural Network (ANN), and Generalized Linear Model (GLM). Allahbakhshian-Farsani, P. et al (2020) in their research experimented with multiple machine learning models (MLMs), including support vector regression (SVR), multivariate adaptive regression spline (MARS), boosted regression trees (BRT), and projection pursuit regression (PPR) and the results were compared to traditional method i.e. nonlinear regression (NLR) in regional flood frequency analysis (RFFA). Based on the result, the generalized normal (GNO) probability distribution function (PDF) is selected by the L-moment method among five PDFs, including the GNO, generalized Pareto (GP), generalized logistic (GL), generalized extreme value (GEV) and Pearson type 3 (P III) to estimate flood discharge for the expected return periods. The overall results proved that the SVR, PPR, and MARS models in comparison to the NLR and BRT models have a better performance to estimate flood discharge with the expected return periods. This was an important milestone in the approach of using Machine Learning ensembles in the flood prediction framework. In another excellent attempt on using ensemble machine learning models Wu, W. et al (2020) used numerical weather and climate prediction, expansion in high performance computing, growing interest in shifting from deterministic to risk-based decision-making that accounts for forecast uncertainty, and the efforts of communities such as the international Hydrologic Ensemble Prediction Experiment (HEPEX), which focuses on advancing relevant ensemble forecasting capabilities and fostering its adoption. With this shift, comes the need to understand the current state of ensemble flood forecasting, in order to provide insights into current capabilities and areas for improvement, thus identifying future research opportunities to allow for better allocation of research resources. Future research directions include opportunities to improve technical aspects of ensemble flood forecasting, such as data assimilation techniques and methods to account for more sources of uncertainty, and developing ensemble forecasts for more variables, for example, flood inundation, by applying techniques such as machine learning. Mosaffa, H et al (2022) in their paper, experimented machine learning in four subfields of hydrology, including flood, precipitation estimation, water quality, and groundwater and concluded that machine learning performs better in flood prediction than traditional data-driven and physical hydrology modeling, particularly in short-term flood forecasting. Also, they were able to showcase that Machin Learning techniques help to estimate precipitation from satellite datasets in a way which is statistically proven and verified. This study paved the way for application of Machine Learning in water quality and groundwater modeling. The study also showcases the use of optimization algorithms in an effective way for parameter selection can improve the performance of machine learning approach. Kan, G et al (2020) used Artificial Neural Network (ANN) modeling technique with the K-nearest neighbor (KNN) modeling method to propose a novel hybrid machine learning (HML) hydrological model for flood forecasting where the advantage of the proposed model over traditional neural network models was that it could predict discharge continuously without accuracy loss owed to its specially designed model structure.

2.3. Flood Prediction based on Artificial Intelligence models in Specific Geographies

The generic nature of the flood prediction warning system based on the data analysis and Artificial Intelligence Models was put forward in the form of an interesting study by Dodangeh, E. et al (2020) where they presented an overview of machine learning models used in flood prediction and develops a classification scheme to analyze the existing literature. Liong, S. Y. et al (2000) shows the Advance flood forecasting system for flood stricken Bangladesh with a fuzzy reasoning method. This geo has similarities in terms of geographical and demographic profiles with the current study in focus. Using simple Artificial Intelligence based classification modeling namely Random Forest, Janizadeh, S. et al (2019) showed Spatial prediction of flood susceptibility in Seoul metropolitan city, Korea. In their study on Yangtze river caused floods, which rises in the mountains of Qinghai province of China, on the Tibetan plateau, and flows 6,300km to reach the East China Sea at Shanghai, Liu, D. et al (2020) predicted the streamflow using deep learning neural network. In another study on floods in Dhaka, Bangladesh, which bears similarity on the rivers network with the study in focus that is Odisha Deltaic region, Liong, S. Y. and Sivapragasam, C. (2002) did a flood stage forecasting with support vector machines which is another popular classification machine learning techniques. Similarly in their paper Mane, P. et al (2020) have suggested the use of machine learning on the Real-time river-level monitoring data for an early detection of flood. A study by Pravin, A. et al (2021) where they proposed use of Machine Learning on the IOT Sensors data on water level of the waterbodies and the ground water level in and around the mostly effected areas. Research by Subeesh, A. et al (2018) suggested the use of Artificial Neural Network (ANN) on the IoT data which are again based on the sensor data from the river bed but uses ANN effectively on these data. Gude, V. et al (2020) demonstrated with a deep learning (Dl) model using river water levels, this study explored deep learning techniques for predicting gauge height and evaluating the associated uncertainty. Gauge height data for the Meramec River in Valley Park, Missouri was used to develop and validate the model. In their paper, Motta, M. et al (2021) tried to develop a flood prediction system using a combination of Machine Learning classifiers along with GIS techniques to be used as an effective tool for urban management and resilience planning specifically in the areas of Lisboa, Portugal. This approach was an attempt to establish sensible factors and risk indices for the occurrence of floods at the city level, which could be instrumental for outlining a long-term strategy for Smart Cities. In very focused study on flood prediction using rainfall data, in the state of Kebbi in Nigeria, Lawal, Z. K. et al (2021) used three machine learning techniques on historical rainfall dataset of thirty-three years (33), so that it can be used in other Nigerian states with high flood risk. In this article, the Accuracy, Recall, and Receiver Operating Characteristics (ROC) scores of three machine learning algorithms, namely Decision Tree (DT), Logistic Regression (LR), and Support Vector Classification (SVC) were used and compared for the best classification model. Nguyen, D. T. & Chen, S. T. (2020) in their paper to create a real-time Flood Forecasting alert model using various Machine Learning models including support vector regression (SVR), a fuzzy inference model (FIM), and the k-nearest neighbors (k-NN) method, to establish a probabilistic forecasting model which was necessarily is a combination of a deterministic forecast produced using SVR and a probability distribution of forecast errors determined by the FIM and k-NN method. Another similar attempt to device an early flood warning system Razali, N., Ismail, S., & Mustapha, A. (2020) using Bayesian network (BN) and other Machine Learning (ML) techniques such as Decision Tree (DT), k-Nearest Neighbors (kNN) and Support Vector Machine (SVM) based on the data in the region of Kuala Krai, Kelantan, Malaysia. In their study, Munawar, H. S. et al (2022) used data from remote sensing devices to forecast flood prediction and focuses on the pre-disaster phase of the disaster management process for the past 20 years, this was an important study which also tried to correlate the pre-disaster phase preparation with the post flood actions. Another first in the country of Egypt to predict the flash flood in the Central Eastern Desert of Egypt, El-Magd, S. A. A., Pradhan, B., & Alamri, A. (2021) used two machine learning techniques, namely extreme gradient boosting (XGBoost) and k-nearest neighbor (KNN) to generate flood susceptibility maps for flash flood. In their paper Ganguly, K. et al (2019) focused both on prediction and determination of influencing factors because both of these are effective in flood damage reduction, where they used machine learning driven approach for prediction using three algorithms namely linear regression, random forest and artificial neural network: it was worth mentioning that that in this particular case linear regression performed better than the other two because the assumptions made to use considering the local demographics. This was important as the use of the machine learning methods can be effective based on the assumptions and conditions of local parameters. In the paper by Ke, Q. et al (2020) which used data of floods & rainfall in the region of Shenzhen in China, using Machine Learning Classifications models to categorize the rainfall conditions leading to binary outcome that is flood or non-flood event, this was quite interesting to prepare for the post-rainfall events. For the country of China using past rainfall and flood data, Hou, J. et al (2021) attempted to create a forecasting model using Machine Learning models and a hydrodynamic-based urban flood model where the Machine Learning model was obtained by training the simulation results of the hydrodynamic model and rainfall characteristic parameters. Chen, C. et al (2022), proposed an incredibly significant and interesting study that combines geographical coordinates with rainfall and subsequent rainfall discharge stations sensor data, resulting in tensors by station coordinates. Thus, creating an input feature which was a two-dimensional time series with spatial information that feeds into a machine learning model combining a Convolutional Neural Network (CNN) with a Long Short Term Memory Network (LSTM) to forecast flood and flood like events. Li, X. et al (2021) in their paper which aims to conduct the risk assessment of global watersheds based on multiple machine learning models, the post evaluation results of the model showed that the random forest classification model was superior in the test, and thus proved an efficient and reliable tool in flood susceptibility assessment. Also, this study, using sensitivity analysis of the conditioning factors, showed that precipitation concentration degree and Manning coefficient were the main factors affecting flood risk in the watersheds. In their study Tayfur, G. et al (2018) showed the application of the artificial neural network (ANN), the genetic algorithm (GA), the ant colony optimization (ACO), and the particle swarm optimization (PSO) methods for flood hydrograph predictions. In this approach, the flow field data recorded on an equipped reach of Tiber River, central region of Italy, were used for training the ANN with an objective to find the optimal values of the parameters of the rating curve method (RCM) by the GA, ACO, and PSO methods. In another very innovative attempt to Flood susceptibility modelling using advanced ensemble machine learning models, Islam, A. R. M. T. et al (2021), showcased a flood susceptibility modelling using advanced ensemble machine learning models – this study they applied and assessed two new hybrid ensemble models, namely Dagging and Random Subspace (RS) coupled with Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machine (SVM) which are the other three state-of-the-art machine learning models for modelling flood susceptibility maps at the Teesta River basin, the northern region of Bangladesh. In an attempt to forecast floods in the Pattani Basin, Thailand, Noymanee, J. et al (2017) explored opportunities of machine learning methods for forecasting of flooding phenomena using open data and the study factors included data collection period and location and configuration of prediction models. In their interesting comparison of Machine Learning models Mosavi, A et al (2018) demonstrated the state of the art of ML models in flood prediction and give an insight over the most suitable models. Machine learning models are benchmarked through a qualitative analysis of robustness, accuracy, effectiveness, and speed have been investigated to provide an extensive overview on various machine learning algorithms usage in the field. In a similar study but specific to four machine learning models, Adnan, M. S. G. et al (2023) focused on creating a framework for reducing spatial disagreement among four standalone and hybridized ML-based FSMs: random forest (RF), k-nearest neighbor (KNN), multilayer perceptron (MLP), and hybridized genetic algorithm-gaussian radial basis function-support vector regression (GA-RBF-SVR), and then they developed an optimized model by combining the outcomes of those four models. This was a very effective way of creating some very powerful ensemble machine learning models using geospatial approaches for flood prediction. In the approach of using images and using Machine Learning, Munawar, H. S. et al (2021) proposed a fresh approach for classification framework for flood management, those have been proposed and studied till date and aimed to group the various technologies in the area of image process approach for flood prediction. In their study, Luu, C. et al (2021) used the flood data in the river basin of Buzau in Romania in the six machine learning models namely : Support Vector Machine (SVM), Decision Tree (DT), Adaptive Neuro-Fuzzy Inference System (ANFIS), Random Forest (RF), Artificial Neural Network (ANN) and Alternating Decision Tree (ADT), the effectiveness of this approach lies in the fact that they were able to identify 205 flooding points of interest and later special interest were explored around these flooding points to reduce the aftermath impact. Focusing on the two fold machine learning approach, Felix, A. Y., & Sasipraba, T. (2019) used Gradient Boost (GB) Algorithm to classify the data sets and perform regression on it to produce the best result from the datasets and they used that train and proposed a prediction model based on a Decision Tree (DT) algorithm. In another very interesting study focusing on the urban drainage systems and rainfall data to predict flood or coming up with an early warning system (EWS), Duncan, A. et al (2013) used single multi-output Artificial Neural Network (ANN) model that could provide a considerable lead time to prepare for the flood in specific areas, that paper also provided a sensitivity analysis and demonstrated that the - predictive capability of such a system based on actual rainfall is limited to a maximum of the Time of Concentration (ToC) of each node being modelled. To achieve operationally useful prediction times, predictions of rainfall as input signals are likely to be needed for most urban drainage networks. This was indeed a pragmatic approach to minimizing the losses due to flood in some areas. In a study that focused on South Korea geography, Kim, H. I., & Han, K. Y. (2020) used effectively a Deep Neural Network (DNN) for predicting the total accumulative overflow, and because of the insufficiency of observed rainfall data, 6 h of rainfall were surveyed nationwide in Korea. Statistical characteristics of each rainfall event were used as input data for the DNN. The target value of the DNN was the total accumulative overflow calculated from Storm Water Management Model (SWMM) simulations, and the methodology of data augmentation was applied to increase the input data. The SWMM is one-dimensional model for rainfall-runoff analysis. The data augmentation allowed enrichment of the training data for DNN. Taromideh, F. et al (2022) used the data the city Rasht in Iran in order to come up with a flood risk map, they used data measuring the consistency of decision makers' judgments, generating pairwise comparisons for choosing a solution, and considering criteria and sub-criteria to evaluate all possible execution options where the map included the vulnerabilities and hazards of different urban areas using classification and regression-tree models, and the map also served both as a first stage in advancing flood-risk mitigation approaches and in allocating warning and forecasting systems. This very interesting study also proved that machine-learning methods are efficient in creating urban flood zones and also it showed that distance to rivers, urban drainage density, and distance to vulnerable areas are the most significant parameters that influence flood hazards. In their paper In this study which deals with flood prediction on the coastal areas of South Korea with high flood risks, Park, S. J., & Lee, D. K. (2020) focused on the climatic and environmental parameters and used various machine learning algorithms to produce a flood prediction model. They forecasted flooding hazards under different representative concentration pathway climate change scenarios and regional climate models while considering ratios of sea level rise. Ahmad, M. et al (2022) did an interesting study on various Artificial Neural Networks (ANNs) models built on the rainfall parameters for the Vésubie River, Nice, France and compared with other classic machine learning models like Random Forest and Decision Tree. They trained and tested the model with discharge, precipitation, temperature, and evapotranspiration data collected over a period of four years (2011–2014) and then did a comparative investigation to compare the performance of the different models by various statistical methods such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and a correlation coefficient (R). They concluded that Feed Forward Neural Network (FFNN) (a type of ANN) models are less efficient than NAM models and in all machine learning models, the decision tree model is the one which performed the best with the given datasets. Furquim, G. et al (2016) published in their analysis which they conducted on the use of data gathered in the in São Carlos, São Paulo State, Brazil region which were collected on the rainfall and the river level data. They used machine learning time series farecasting modeling techniques using Chaos Theory to produce a flood forecasting model for the São Carlos region. Karyotis, C. et al (2019) proposed an urban flood forecasting and monitoring platform developed as part of a UK Newton Fund project in Malaysia which was used in monitoring and forecasting platform for Urban Deployment that was developed to deliver an effective and low cost urban flood forecasting solution, which should be able to accurately forecast flood risk at street level, and deliver optimized recommendations to the relevant authorities as well as an early warning alerts to members of the public. The model was based on a hybrid Deep Learning and Fuzzy Logic based architecture and could demonstrate proven ability of the model with the experimental results to account for factors that were not included in other modern flood forecasting systems for that region and served a good basis for averting the flood related losses. Avand, M. et al (2021) in their paper, published the analysis of effects of climate and land use changes on flood susceptibility areas in the Tajan watershed in Iran. This was done with an eye on future land use and to do this, land use changes over the next 20 years (2019–2040) were predicted from land use changes of the past 29 years (1990–2019) using the land change modeler (LCM) method. They developed a machine learning model based on random forest (RF) model and a Bayesian generalized linear model (GLMbayes) to predict areas susceptible to flooding. The results showed some key factors such as elevation, distance from river, land use, slope, and rainfall were the most important factors affecting flooding in this particular basin. The factors were modified according to land use changes and climate changes and the models were revised for other areas and the area-under-the-curve (AUC) evaluation of the models indicated that the RF model more accurately predicted flood probability than the Bayesian generalized linear model. Tang, Y. et al (2023) in their study with the objective of creating a localized flood prediction mode, used the around 100 floods related data in the Jingle sub-basin, a tributary of the Yellow River basin, China, for the floods that occurred between 1971 and 2014, analyzed using dynamic clustering and random forest machine learning modeling techniques to identify flood types and select appropriate model parameters. The Xin'anjiang model was then used for real-time flood forecasting and the results indicated that the rainfall characteristic indicators developed by the model can effectively identify potential flood types, and the model parameters determined by historical flood rates can be adapted and utilized for new forecasting tasks based on similarities. It was emphasized that forecasting results given b ensemble models with a probabilistic nature were more accurate than the results from single fractal forecasting model. The recommended model could identify extreme flood events, facilitate flood classification and prediction, promote basin disaster mitigation, and enable the efficient use of water resources efficiently.

2.4. Flood Prediction Models for India and Odisha

In their paper Ghorpade, P. et al (2021) demonstrated the machine learning-based models trained using climatic parameters' historical data are increasingly useful for forecasting tasks and the recent advancements in the flood forecasting field using machine learning algorithms. In an interesting work intended for mass usages using mobile application, Kunverji, K. et al (2021) worked on developing a flood determining model based on AI calculations and a hearty, productive and precise flood expectation framework offering the fundamental aid and assistance needed to the residents and government based on a Decision Tree Model which actualizes various calculations on datasets with a scope of accuracy. The model utilizes an AI calculation which predicts floods, sending alerts to the local and government authorities using an Android Application. In their paper Sahoo, A. et al (2021) showcase the model on Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: To model complex nature of hydrologic processes artificial neural network (ANN) tool is effectively being utilized for modelling different nonlinear relationships, and the model also showed the efficiency to be an appropriate method for flood prediction in that region.

In their paper Goel, R. (2020), using Machine Learning techniques based on rainfall data in India, using. Regression analysis, a predictive model was built using models like Gaussian Naïve Bayes, tree-based approaches and K Nearest Neighbor, The analysis was run on three states Bihar, Uttar Pradesh and Kerala. It is interesting to note that among all the selected supervised learning technique Random Forest and KNN performed best. In a very interesting work, Rani, D. et al (2020) showcased a low cost IoT based flood monitoring system using machine learning and neural networks, for flood alerting and rainfall prediction, where they used effectively used Neural networks to alert on abnormality rainfall thus alerting a possible flood scenario, the method was indeed effective, but the lead time looked less for any preventive measures. On a similar work, Bande, S. et al (2017) used IoT data for building a machine learning model using Artificial Neural Network (ANN), designed with the aim of enhancing the scalability and reliability of flood management system, where the main objective was to monitor humidity, temperature, pressure, rainfall, river water level and to find their temporal correlative information for flood prediction analysis. This was one of the first models to have incorporated a wide range of data systems from IoT for feeding into an Artificial Neural Network (ANN) system based on inter correlation values. Arora, A. et al (2021) used Fuzzy logic along with various machine learning to quantitatively test and compare novel advanced-machine learning algorithms in achieving the goal of predicting flood susceptible areas in a low altitudinal range, sub-tropical floodplain environmental setting, in the Middle Ganga Plain (MGP), India which is also the hotbed of annual flood disaster. In their attempt to come up with a flood prediction model with lesser number of variables, Yeditha, P. K. et al (2020) developed a machine learning model by forecasting method using the satellite precipitation product and wavelet-based machine learning models especially in data scarce regions. Saha, T. K. et al (2021) in their study, worked on fifteen flood conditioning parameters which were generated from both coarse and high-resolution datasets, then they used multiple machine learning models such as the ANN-multilayer perceptron (MLP), random forest (RF), bagging (B)-MLP, B-gaussian processes (B-GP) and B-SMOreg algorithms to integrate the flood conditioning parameters for generating the flood susceptible model and they also proposed an index of flood vulnerability model to validate flood susceptibility models along with conventional statistical techniques, such as the ROC curve. In their study for the floods in the Mahanadi Basins, which is also the focus of this study, Mohapatra, S. et al (2023) used multiple machine Learning techniques on a 30-year historical rainfall dataset, demonstrated Logistic regression (LR) as the best performing model found as the best model against decision tree (DT) and random forest (RF) whilst comparing the results in receiver operating characteristics (ROC), accuracy (ACC) and recall (R) scores. Mittal, V. et al (2023a) in their attempt to build a flood prediction model used rainfall-runoff data and other meteorological data parameters like temperature, evapotranspiration, crop evapotranspiration, cloud cover, wet day frequency, and vapor pressure and analyzed ten flood affected districts of Odisha to create ML-based flood forecasting models to predict floods with larger lead times which is essential to minimize the impact. They used the outcome from the machine learning model and compared on parameters such as accuracy, precision, recall, F-measure, and AUC-ROC and concluded that random forest-based flood forecasting model performed comparatively better than Naïve Bayes, logistic regression, k-nearest neighbor, artificial neural networks, and support vector machine. In a very recent study again on the floods induced in the Mahanadi deltaic region, Pradhan, S., et al (2024), used rainfall data along with other meteorological parameters used Logistic Regression (LR) against other classical machine learning models on several parameters to propose the flood prediction model. In an innovative study on the river basins on the branches of Mahanadi like Daya and Bhargavi that flow across Odisha, Nayak, M. et al (2022) attempted to showcase the positive impact of barrage construction in key points in the river basins. They used Deep Learning (DL) techniques like Deep Belief Network (DBN) where the flood forecasting was carried out for 1 day, 1 week and 2 weeks of time period using both the rivers, and they compared the model performance using Teaching Learning-Based Optimization method (TLBO) with the use of many parameters like Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Balamurugan, R. et al (2022) used Machine Learning techniques such as k-nearest neighbors (KNNs), support vector machines (SVMs), random forests (RFs), and decision trees (DTs) to build an ensemble of ML models, they effectively used some data sampling techniques like stacking classifier to overcome the issue of oversampling and low accuracy. They used accuracy, f1-scores, recall, and precision for comparing the performance outcome of the various models. They showcased that stacked models are best for predicting floods due to real-time rainfall in the eight coastal states including the state of Odisha. The accuracy from the model was highest for Andhra Pradesh, which was at 97.91%, whereas Orissa achieved an accuracy of 92.36%, lowest among the eight coastal states. This also indicated how variability the data points where for the state of Odisha and how important it is to have a more accurate forecasting model in this state which is pounded by the floods almost every alternate year since 2010. Treated in a customized way the rainfall in Monsoon and non-monsoon season for the state of Odisha, Zhang, X. et al (2020), used data from 1991 to 2015 from Department of Forest and Environment Govt. of Odisha used various machine learning techniques like Support Vector Regression and Multilayer perception implemented to predict the maximum rainfall in annual and non-monsoon seasons which has a clear correlation with the flood both in monsoon and non-monsoon seasons. They also considered meteorological parameters, namely average temperature in month, wind velocity, humidity, and cloud cover along with the rainfall data and then measured the performance of the model outcome with MSE (mean squared error), correlation coefficient, coefficient of efficiency and MAE (mean absolute error). The outcome of models showed MLP being a computationally intensive method, SVR could be used as an efficient alternative for runoff and sediment yield prediction under comparable accuracy in predictions and SVR-MLP may be used as efficient prediction tool with considerable high accuracy in flood prediction with rainfall data. Misra, R. K. et al (2021) used various data-driven machine learning techniques for predicting the rainfall of an average rainy season (June to September, 1901-2018) in Odish which was based on three aspects: pre-processing techniques input, modeling and modeling methods. Outcome of the various machine learning models namely: linear regression analysis (LR), random forest method (RF) and Artificial Neural network (ANN) to come up with the best performing mode. Some of the interesting observations that came up in this study were the maximum rain fall 385.3mm occurred in year 1961and minimum197.2mm rain fall occurred in year 1974 and based on the model outcome it was established that Random forest regression had the best accuracy amongst the models at 91% and mean absolute error was 2.3%, for ANN accuracy was 89.64% and mean absolute error was 3.2% and for Linear regression accuracy was 87.28% and mean absolute error was 3.8%. Hence it was concluded that Random forests were very effective in the rainfall prediction in the state of Odisha. Anupam, S., & Pani, P. (2020) in their study on creating a Flood forecasting model where they experimented with a hybrid extreme learning machine-particle swarm optimization algorithm (ELM-PSO) model, by using the data on gauge, rainfall, and discharge data of 4 decades from the Jenapur flood-forecasting station (Brahmani river, Odisha) and the Anandpur station (Baitarani river, Odisha) were used to create models for mean gauge height prediction and then the models were validated using tenfold cross-validation, with mean-squared error (MSE) and the coefficient of determination (R-squared) as parameters for evaluation of the models. In their study to improve the existing classification models mostly on the imbalanced datasets on the river basins of Odisha, Mittal, V., et al (2023b), they established a more balanced dataset to help in improving the efficiency of classification models which they did using SMOTE technique and some of its variants. Sahoo, B., & Bhaskaran, P. K. (2019) in their study proposed an alternate approach using soft computing techniques such as Artificial Neural Network (ANN) for the prediction of storm surge and onshore flooding which tested to be viable and highly cost-effective method consistently maintaining high level of computational accuracy (>92%) thereby finding potential real-time application. To prove the efficiency of the approach, they used as a test case the onshore flooding associated with the 1999 Odisha Super cyclone and tested this Artificial Neural Network (ANN) model which was trained on pre-computed scenarios of storm-tide and inundation data for the entire Odisha coast with a success rate of 99%. They could show the results which were quite encouraging in demonstrating the efficiency of Artificial Neural Network model for real-time application and effectiveness for disaster risk reduction during tropical cyclone activity.

2.5. Summary

This chapter concludes with the literature reviews of studies on various perspectives of Flood Prediction in form of an early warning system with the help of Data Analysis and Artificial Intelligence. Multiple research projects in direction of flood prediction have shifted from mathematical models or hydrological models to algorithmic based approaches. Flood data is dynamic data and non-linear in nature. To predict floods, techniques such as artificial neural networks are used to devise prediction algorithms.

There are some which are referred to, but they are based on data which is mostly on the direct correlation with the flood and are very much anticipated. The accuracy of the suggested models may also be increasing with the evolution of modern tools and technologies but most of them do not talk about the long-term avoidance of the floods and how the floods could be anticipated much earlier, especially in populous regions like Coastal cities of Odisha.

Hence these models could not be applied in the deltaic region of Odisha and hence demands more geo-specific research and causal analysis.

After carefully studying the above great research and some more in this area, there are some points which come out very clearly and they are:

I. Artificial Intelligence can help establish the various causes with a probabilistic outcome.
II. The reasons are geography specific.
III. There is still a long way to go in terms of identification of the symptoms early enough and taking actionable remedies so that we can help diminish the impact to a greater level especially in densely populous regions.

But one thing which is very evident is that these studies mostly focus on a specific type of data or related data types. For example: some studies try to infer and dig deep into the satellite images and in some, the underground water level along with the river water levers along with the periodic seasonal data. Hence there was already some degree of correlation already existing between the set of data.

So, all these pointers suggest to have a new approach with a holistic view combining all the data at our disposal from various aspects, for example – Satellite Imaging Data, Groundwater level data, sensor data from river beds, level of urbanization, seasonal farming, historical cyclone data, data on deforestation etc., and come up with a new model with the help of machine learning for this specific geography that is the Deltaic region of Odisha.

CHAPTER 3 - RESEARCH METHODOLOGY

3.1. Introduction

The effectiveness of any data-based research objective has a lot of stake lies on the adopted methodology. So, a robust research methodology plays a pivotal role in ensuring the validity, reliability, and generalizability of findings. Irrespective of the nature of objective and research, a well-structured research approach holds a key to success. A well-defined methodology ensures the research is conducted in a systematic and objective manner, fostering the generation of reliable and credible results. The different sections describe the research methodologies and processes planned to achieve the research objectives. It starts with the research strategy, which talks about how the research was carried out, how the potential datasets along with the variables under study were discussed & zeroed upon, the data analysis planned, the machine learning experiments carried out to establish the causal relationship, the ethical consideration the potential recommended actions those can be impactful to the predicted outcome. Being data-based research, the methodology also sheds a lot of attention to the Data Quality when the analysis was carried out, as datasets with lot of noise can impact not only the recommended actions but also the evaluation process of machine learning carried out throughout the hypothesis testing iterations.

3.2. Overview of Research Strategy & Phases

The research was divided into multiple phases. The objective was to come up with an explainable predictive Machine Learning Model indicating an advance alert for the floods in the deltaic region of the state of Odisha, India. This was planned to have multiple phases with each phase having its own objectives which are described below-

Phase – I Objective: To capture the Historical Data and analytical study on this

The first phase was analysis of the historical data on the floods involving the rivers and the impact on the of these floods on the human society including the losses in agriculture, livestock & more importantly any changes to the geographical and ecological if any, be reversible or irreversible or slowly reversible changes.

Phase – II Objective: To do the Literature review & identification of related studies in similar condition or similar demographics

The second phase and a large part of the study was devoted to the earlier studies done in this area, be in in Machine Learning based flood alert and monitoring system Or studies involving observing some of the obvious factors like water level in the rivers using Internet Of Things (IoT) devices. This phase encapsulates a lot of literature in the field of Flood Alert space but in this space a lot of variables are so dependent on the local conditions or demographics and hence it was important to segregate the earlier research done in this space. It is like 'not same size fits all' approach! The objective is quite important from this research perspective.

Phase – III Objective: To deep dive into the identified related research, finetune and then come up with all potential datasets for analysis.

The idea is to bring all the potential candidates or variables to the table which may have a correlation or impact to the geography under study. This is very important from this applicability of the research objective. The list of all potential candidates will help clarity in data gathering and the next phases of the study, which is data & impact analysis.

Some of the Datasets under study are given the Appendix.

Phase – IV Objective: To analyze the identified datasets, do exploratory analysis and data engineering for Machine Learning Algorithms

This phase will involve deeper analysis of identified datasets, do exploratory analysis and the find a measurable correlation with potential outcome. This will also involve data engineering on the identified datasets and help prepare the data for Machine Learning Model Consumption.

Phase – V Objective: Come up with Machine Learning Models & Performance Matrix

In this phase, it is planned to create & come up with the Explainable Machine Learning Models. The idea is to come up with multiple Machine Learning models with multiple techniques & algorithms. This is a pragmatic way for coming up with a final model with better accuracy as we know that some of the Machine Learning Techniques & algorithms may be skewed towards a particular set of data types, and we are not sure about the type of data which will have a better correlation with the outcome. Hence, it is a better approach adopted by this kind of research where the variability in 'unknowns' data is not yet established. This also involves the iteration and testing of the final set of Machine Learning models against the identified historical data to test the accuracy & performance matrix.

Phase – VI Objective: Analysis of Explainable Machine Learning Model & Come up with Actionable Insights

This will be the final phase & will include the explainability of the Machine Learning model and the actionable insights recommendation. This will pave the path for further research in this approach and possibilities for more advanced studies with the available of more data and evolution of Machine Learning Algorithms.

3.3. Sources of Data

There are various categories of the data sources used for the purpose of analysis. The various categories are explained below.

Primary data: The study will involve the data available in the public domain and mostly government agencies data available in public domain. Some data under study will be requested from various government agencies and may be available on request

Secondary Data: Secondary data would be gathered by conducting a comprehensive review of the subject's literature. References were made to research papers published in reputable journals and databases, as well as to published articles, reviews, and websites pertaining to the hospitality industry, as well as to books, manuals, library resources, Newspapers, Business Magazines, and Textbooks etc.

3.4. Research Objectives

The key objectives were:

i. To come up with an explainable predictive Machine Learning Model indicating an advance alert for the floods in the deltaic region of the state of Odisha, India.

ii. To explain the Machine Learning model so that the same can be trusted and also can be used in further studies with evolution of more sophisticated machine learning techniques.

iii. To come up with actionable insights to minimize the impact of the flood on the society & environment

3.5. Scope of the Study

This study gives a solid basis and a wealth of data to expand this research into additional demographic areas. Given the explainability of the Machine Learning model, it is intended to aid in understanding distinct causes and patterns, particularly in the deltaic region of Odisha in the context of frequent floods which have become more frequent in last one hundred years, suggesting the rapid industrialization and population growth. Further research can be conducted in other Indian regions or may be other parts of world bearing similar geographical or demographics features, which do face similar floods. Given the actionable insights planned under the study will definitely come handy in application of this study and help minimize the impact of the floods not only in the specific geography but also might find applicability in other parts of the world.

3.6. Data Analysis Strategies

There are numerous factors that may seem to have an impact on the floods in general, some are direct, and some are indirect. Also, this space of Flood Prediction with the help of Machine Learning can have many factors which are local or specific to the geographical co-ordinates or demographical in nature. Hence, it was very important to categorize those factors in the first place. Then the identification of datasets available to this area are being studied. The range of data ranges from river floods data to oceanic or sub-oceanic data, to urbanization, industrial and census data. The details of the data are given in Appendix.

After the data collection, it is planned to do exploratory data analysis. Then it is planned to deep dive the analysis using Azure Databricks tool in Azure Cloud. Once the data is loaded in the Azure Databricks Lakehouse, it will be used to prepare for the machine learning studies using SQL and PYTHON Scripts. Post to this, it is planned to use various Machine Learning Algorithms, namely DCNN (Deep Convolutional Neural Network), SVM (Support Vector Machines), NARX NN (nonlinear autoregressive network with exogenous inputs) etc. These are some of the quite advanced Machine Learning Algorithms which are extremely efficient in identifying the correlation with multiple data types especially in our case some of the data are expected to be images and large streaming data (like data from IoT Sensor devices). All the machine learning modeling will be done using Azure Databricks Data Science Workbench.

3.7. Limitations of the Study and Directions for Future Research

The study only looked at predicting flood in advance in the state of Odisha, India, using Machine Learning which can be also explained meaning this model can be trusted with similar set of data and can have expected outcome. The study also aims to cater for some actionable insights to minimize the after impact of the flood in the Deltaic region of Odisha. Further research can be conducted in other Indian regions or may be other parts of world bearing similar geographical or demographics features, which do face similar floods, especially places where the floods have become more frequent in last one hundred years, suggesting the rapid industrialization and population growth. There is also constant innovation in the space of capturing the data with more and more sophisticated devices being innovated and being used. This is definitely going to change the way we are going to do the analysis and help in the right direction of coming up with multiple innovative approaches to manage this problem in terms of new preventive mechanism.

3.8. Data Analysis Process

Data analysis is the cornerstone of effective flood prediction models. It uncovers patterns, trends, and relationships within complex datasets. By exploring historical rainfall, river levels, soil moisture, and other relevant factors, analysts identify key indicators of flood events. Data cleaning and preprocessing ensure data quality and consistency, crucial for model accuracy. Exploratory data analysis helps visualize data distribution, detect outliers, and identify missing values. Feature engineering transforms raw data into meaningful features that capture flood-related information. In-depth analysis of these features aids in understanding their impact on flood occurrence. Ultimately, data analysis provides insights for model selection, training, and evaluation, leading to more accurate and reliable flood predictions. The key phases that were adopted in the study are explained in the figure 3.1

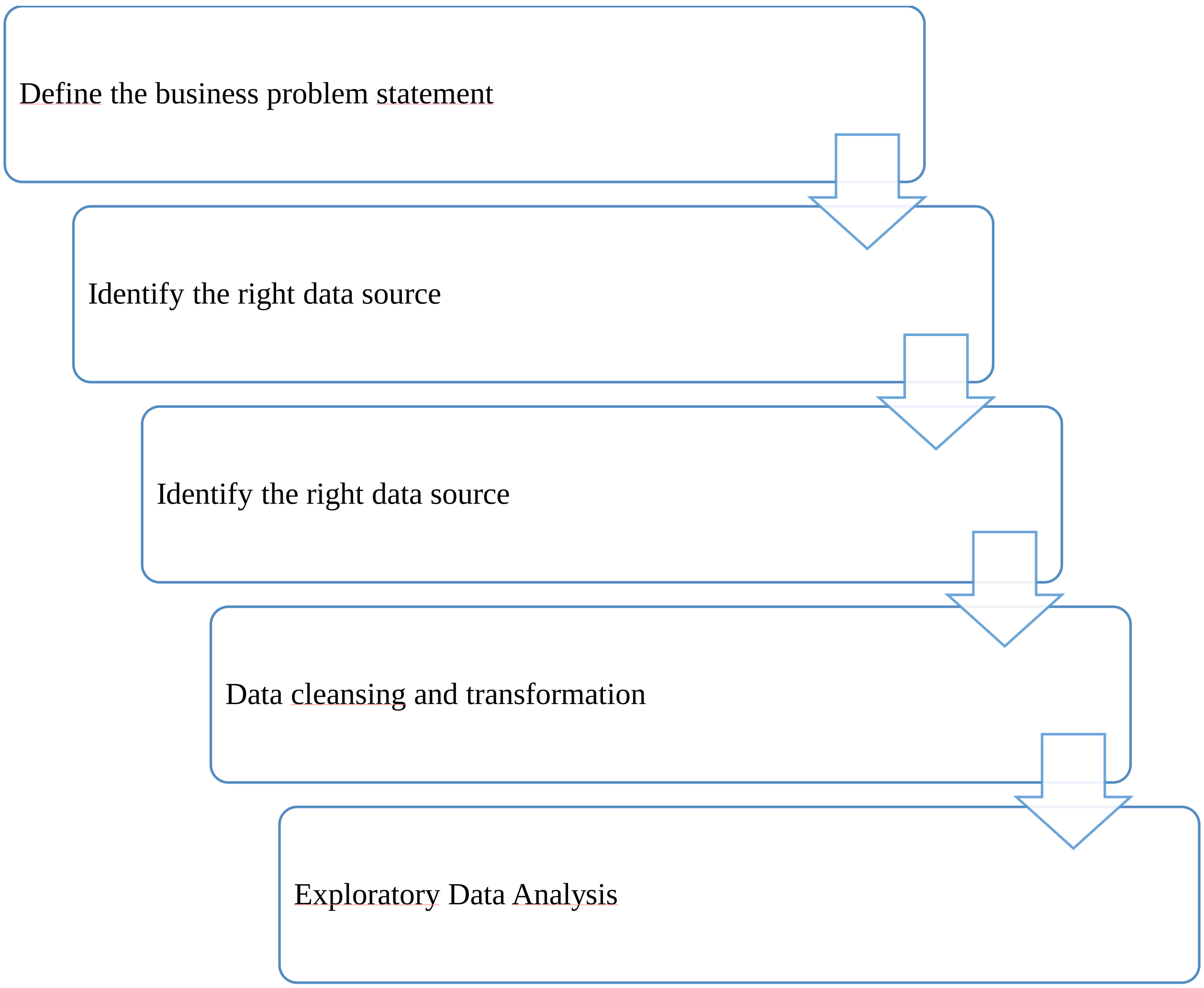

Figure 3.1: Key Steps in Data Analysis

Having already defined the problem statement, the next major step was to identify the key data sources.

Data Identification and filtering:

The author did extensive research on the possibility of a variety of sources and possible correlation with the impact on floods. The approach was to consider all possible data sources in the first step and the do a correlation study to find the impact and then on this basis eliminate the ones with lesser impact.

Principle followed during Data Identification:

I. The idea was to check for relevance that is how well the data source addresses the research question and objectives

II. The source and quality of data has been reliable, that is it has to consistent and trustworthy.

III. It is also about the validity of the data that means the accuracy and credibility of the data source should be of high standards.

IV. Also, it was looked for timeliness that is whether the data source is current and updated.

V. It was important to look at the availability of the data for any future research purpose so that any future researcher can validate and improve on top of the findings of this research.

VI. It was also important to look at the ethical part of the data collection that is at whether the data source respects the rights and dignity of participants or subjects, complies with relevant laws and regulations, and acknowledges any limitations or conflicts of interest.

VII. Started with Datasets which are kind of obvious like Rainfall, weather, riverbed etc. as we are sure of the integrity and reliability and those could be validated as well from multiple sources.

VIII. It was also important to check for the co-relation amongst the datasets so that we do not bias with the results.

IX. Since the scope of the research was on the Deltaic region of the state of Odisha in India, it was important to consider data specific to this geography. But having said that, some of the data were at the India level so it was important to consider if the data could be drilled down to this region.

X. All the datasets are free to use and publicly available even though some were exclusive on request.

The datasets considered in the first phase for the study are mentioned in the table 3.1:

Table 3.1

First set of datasets

Sl#	Data Set Description
1	All Odisha Monthly, Seasonal and Annual Temperature Series (Maximum, Minimum and Mean Temperatures)

2	Rainfall data – All Odisha and five homogeneous regions
3	Demographic data
4	Flood sensor data
5	Weather Statistical data for Odisha
6	Geomorphology map data
7	Crowd density
8	Groundwater Prospects map data for Odisha
9	Water Spread Area map data
10	Forest Fire Alerts data
11	Heatwave Data
12	Potential Evapotranspiration Data
13	Flood Hazard zonation map Data
14	Ocean eddies Data
15	Ocean surface currents Data
16	Ocean surface wind vector data
17	Rainfall (from HE and IMSRA methods) Data
18	Odisha Drainage Data
19	Primary census abstract Data
20	Groundwater Levels Data
21	Groundwater Quality Data
22	Surface water Quality Data
23	Reservoir Level & Storage Data

24	Flood Layers Data

3.9. Data Preparation for Analysis

Post identification of the shortlisted data sources, the approach was to consider all the possible data sources with aggregable correlation with the outcome. The author did a co-relation analysis of each of the shortlisted data sources and used **Pearson's correlation coefficient** as a measure of selection.

Pearson's correlation coefficient

The method of correlation analysis is widely used, and the outcome of the test is defined by the Pearson Correlation Coefficient- This coefficient measures varies between -1 to +1. Here the value -1 denotes the strongest 'negative correlation' and +1 indicates the strongest 'positive correlation'.

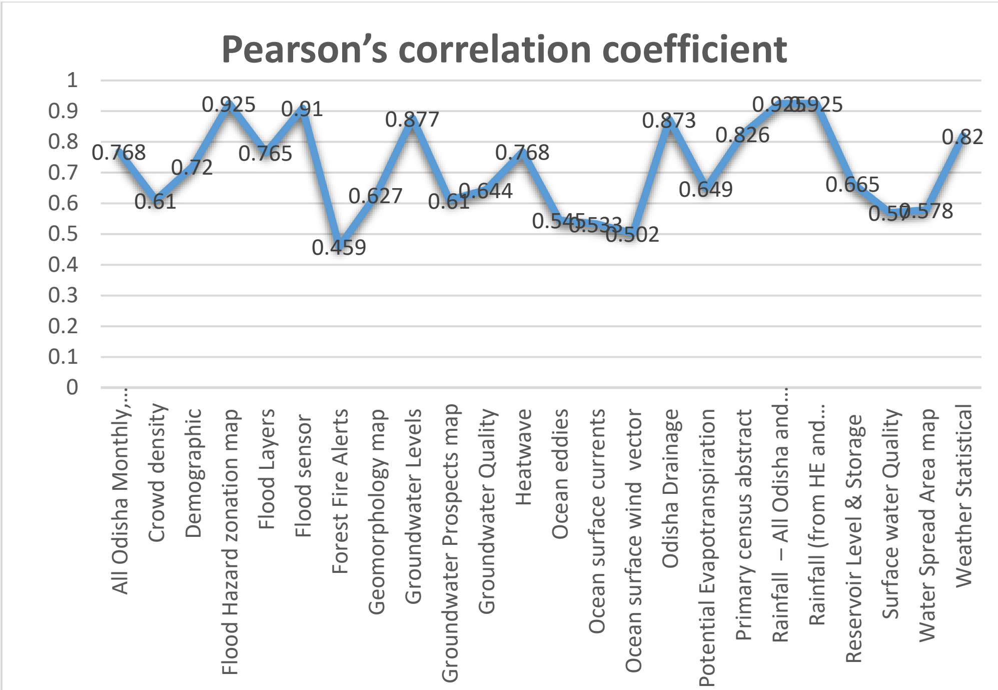

Figure 3.2: Pearsons Correlation

Based on this approach the identified datasets for the analysis were considered and mentioned in the table 3.2:

Table 3.2

List of shortlisted datasets as per Pearson's correlation coefficient

Sl#	Dataset
1	Rainfall – All Odisha and five homogeneous regions
2	Flood Hazard zonation map
3	Rainfall (from HE and IMSRA methods)
4	Flood sensor

5	Groundwater Levels
6	Odisha Drainage
7	Primary census abstract
8	Weather Statistical
9	All Odisha Monthly, Seasonal and Annual Temperature Series (Maximum, Minimum and Mean Temperatures)
10	Heatwave
11	Flood Layers
12	Demographic
13	Reservoir Level & Storage
14	Potential Evapotranspiration

3.10. Data Cleansing & Transformation

The next important task was to make sure the dataset of interest is cleaned and transformed for analysis and to be feed into the machine learning model. Since most of the data sources are publicly available from multiple sources, it is one of the times taking and very important step in the process. Any dataset which is not of good quality can have an adverse impact on the outcome of the research.

Some of the principles considered for data cleansing are:

I. Removing major errors, duplicates, and outliers—all of which are inevitable problems when aggregating data from numerous sources especially when collected at large from publicly available systems.

II. Removing unwanted data points—extracting irrelevant observations that have no bearing on the intended analysis.

III. Bringing structure to the data—general ‘housekeeping’, i.e. fixing typos or layout issues, which will help you map and manipulate your data more easily.

IV. Filling in major gaps—when started tidying up, it was noticed that some of the important data points of interest were missing. So, it was important to fill the gap by making sure the datapoint of interest is filled in.

Post data cleansing, the next important step was to make sure the cleaned data has the right kind of attributes to contribute to the analytical needs. So, Data transformation is another important step in the process and goes beyond cleaning. This typically involves manipulating the data to suit the analytical need of the research objective. The below high-level steps were performed to transform the data –

I. Deriving New Features: Created new variables based on existing ones. For example, some rainfall data were at the daily level which needed to be featured into other variables.

II. Encoding Categorical Data: Converted categorical data (like colors) into numerical values for analysis by algorithms.

III. Normalization: Scaled numerical data to a common range to prevent certain features from dominating analysis.

IV. Aggregation & de-aggregation: Based on the available data, in some cases it was needed to summarize data by grouping and combining values. On the other hand, in some cases the data were at the Country Level or state level, which needed to be de-aggregated to the region of interest level.

By following the above mentioned steps, it was ensured that the data is clean, accurate, and ready to be fed into the machine learning modeling so that it would enable us to extract valuable insights

3.11. Exploratory Data Analysis (EDA)

This is one of key steps in the process of analysis which was done alongside the Data Cleansing and Transformation process.

The key objective of this step is explained below:

I. It helped to clean up the dataset.

II. It gave a better understanding of the variables and the relationships between them.

III. Helped identify initial trends and characteristics, and even refined the hypothesis

IV. It was like an early stage analysis, which led to creation of “final data set “for analysis

Some of the Exploratory Data Analysis performed on the Rainfall data led to key identification of potential factors.

3.12. Machine Learning Modeling Approach

The Classical Data Preparation and Analysis phases were thoroughly followed from Data Identification to Machine Learning Model creation including the Data Visualization with the help of sophisticated tools like Azure Databricks and Microsoft Power BI. There were some calculations done in Power BI Analysis tools using Data Analysis Expressions (DAX) which were crucial to Data Visualization.

The high-level phases are mentioned in the Figure 3.2 below.

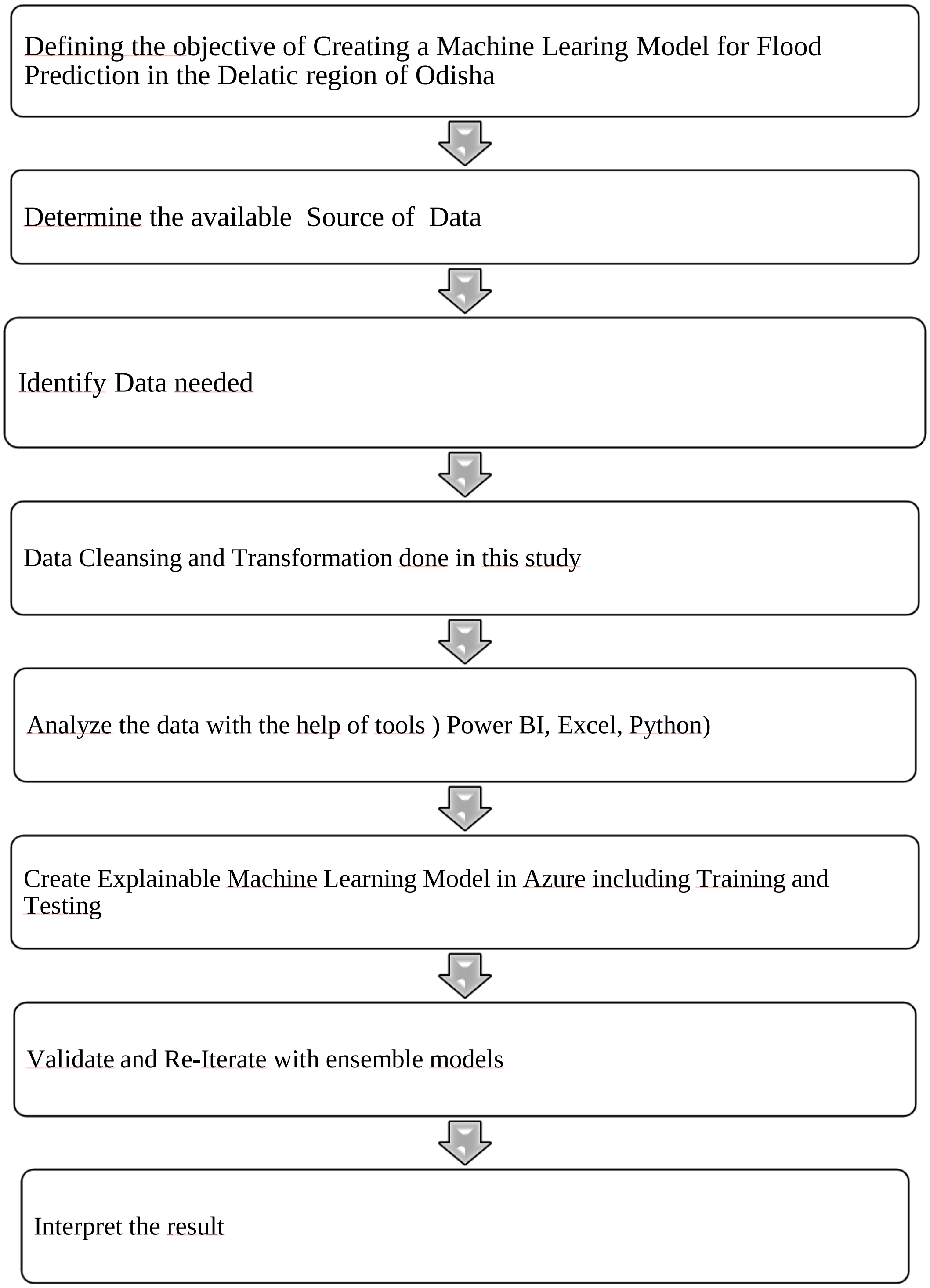

Figure 3.3: Machine Learning Phases

The entire process was grouped and distributed into some key phases and summarized below:

I. *Defining the object*

- Data Availability: Identified the relevant datasets, including historical rainfall, river water levels, soil moisture, land use, topography, meteorological data, water resources etc. The entire dataset details are provided in Appendix 1

- Flood Types: It was important to determine the specific flood types to focus on (e.g., riverine, coastal, urban etc.) as the study aims to analyze and create a solution around the deltaic region of Odisha for any type of floods but to correlate with the potential impactful causes.

- Prediction Timeframes: It was also important to define desired prediction horizons (e.g., short-term, medium-term, long-term). This is part of the solution to recommend time frame based preventive actions in order to minimize the aftermath impacts of the floods in the specified geographical region.

- Evaluation Metrics: Since the number of data points are not huge, it was even more critical to choose appropriate metrics to assess model performance in terms of accuracy, precision, recall, F1-score etc., so that the model can be sufficient to withstand the test of time.

II. *Data Collection and Preprocessing done in this study*

- Gathering Data: Collection of historical data from various sources (government agencies, meteorological departments, remote sensing) based on initial understanding of the flood data pertaining to the geographical area of interest and also on the macroeconomic and demographical data. The data was explored from diverse data sources, considering both structured and unstructured formats.

- Data Cleaning: This is of paramount importance to have excellent quality data for input to any machine learning model for an effective impactful outcome. Manage missing values, outliers, inconsistencies etc. are some of the tasks involved in this phase. This is even more important, especially with the public domain data as usually the data accountability is a time taken process to establish. Sometimes it is important to ensure the data diversity to prevent bias and improve model generalizability. In the specific context of Odisha, India, this process is particularly challenging due to several factors:

III. Data Availability and Quality Issues

- Data Scarcity: Historical data, especially for parameters like soil moisture, land use, and high-resolution meteorological data, might be limited or unavailable for extensive periods.

- Data Inconsistency: Different data sources might use varying formats, units, and time resolutions, making integration and standardization difficult.

- Data Accuracy: Errors in data collection, transcription, or storage can introduce noise and bias into the dataset.

- **Missing Values:** Gaps in data due to equipment failures, human errors, or data loss can significantly impact model performance.

IV. *Data Complexity and Heterogeneity*

- Multiple Data Sources: Integrating data from diverse sources (meteorological, hydrological, remote sensing, etc.) requires careful handling and alignment.

- Spatial and Temporal Variability: Flood events are influenced by complex spatial and temporal patterns, demanding sophisticated data preprocessing techniques.

- Data Format Issues: Dealing with different data formats (CSV, NetCDF, GeoTIFF, etc.) can be time-consuming and error prone.

Specific Challenges in Odisha

- Data Infrastructure: The data collection and storage infrastructure have some challenges like storage of historical data, consistency and integrity of data leading to data accessibility and quality issues.

- Extreme Weather Events: Odisha is prone to extreme weather events like cyclones, which can disrupt data collection and introduce outliers.

- Data Privacy: Adhering to data privacy regulations while accessing and processing sensitive information can be challenging.

- Limited Computational Resources: Managing large datasets and complex data cleaning processes might require significant computational power.

V. *Feature Engineering*

Feature engineering is a critical step in building an effective machine learning model for flood prediction. It involves transforming raw data into informative features that can be used by the model to learn patterns and make accurate predictions. It is important to elaborate this as this is an important step in this context, hence listing out the important feature engineering carried out with data categories –

Meteorological Data

- Rainfall**:**
    - Intensity, duration, and distribution
    - Cumulative rainfall over different time periods
    - Antecedent rainfall index
- Temperature:
    - Air temperature and its impact on evaporation rates
    - Ground temperature and its influence on soil moisture
- Humidity**:**
    - Relative humidity and its correlation with rainfall
- Wind speed and direction:
    - Influence on rainfall patterns and potential storm surges

- Atmospheric pressure:
    - Correlation with weather systems and potential for heavy rainfall

Hydrological Data

- River discharge:
    - Water flow rate at different river stations
    - Peak discharge and time to peak
- Water level:
    - River and reservoir water levels
    - Rate of change in water levels
- Soil moisture:
    - Soil moisture content in different layers
    - Infiltration capacity of the soil
- Land use and land cover:
    - Impact of urbanization, deforestation, and agriculture on runoff
- Topography**:**
    - Elevation, slope, and drainage basin characteristics

- Reservoir levels:
    - Storage capacity and release rates

Additional Features Considered in this study

- Time-based features:
    - Day of the week, month, year, and season
    - Festive and special events which have a direct impact on mobility and environment
- Lagged features:
    - Historical values of variables to capture temporal dependencies
- Derived features:
    - Combinations of existing features (e.g., rainfall intensity * duration)

Feature Engineering Techniques Adopted in this study

- Scaling: Normalize features to a common scale (e.g., Min-Max scaling, Standardization)
- Transformation: Apply mathematical transformations (e.g., log, square root) to manage skewed distributions
- Binning: Group continuous features into categorical bins for better model interpretability

- Aggregation: Combine multiple features into a single feature (e.g., mean, sum, standard deviation)
- Feature selection: Identify the most relevant features using techniques like correlation analysis, feature importance, or dimensionality reduction

Challenges and Considerations in the approach of Feature Engineering

- Data availability and quality: It was ensured to have data completeness, accuracy, and consistency across multiple cross domain data.
- Feature relevance: The selected features were filtered that have a strong correlation with flood events.
- Computational efficiency: Created features that can be processed efficiently by the model.
- Overfitting: Avoided creating too many features to prevent overfitting.

It was really important to have effective features by carefully selecting and engineering features, as this significantly has potential to improve the performance of the flood prediction model. It is essential to iterate on feature engineering and evaluate the impact of different features on model accuracy.

VI. Model Selection and Training

With the variability in data and the existing of skewness in various machine learning model algorithms, selecting the right model and training them become the essence for the findings of the study results. Model selection involves choosing algorithms suited for the prediction problem. Training involves feeding cleaned data to the chosen model. Hyperparameter tuning optimizes performance. In this study, an ensemble of statistical models with machine learning for enhanced accuracy. Addressing the data imbalance and overfitting issues was also looked into.

- Algorithm Selection: While advanced machine learning algorithms can offer superior performance, their complexity often makes them difficult to interpret. Explainability in machine learning is essential for building trust and ensuring the practical utility of flood prediction models. This research approach experimented with multiple modeling approaches and also explained below the insights into an ensemble modeling approach.

1) Decision Trees:

Purpose: Decision Trees are among the simplest and most interpretable machine learning models. They are used to make predictions based on a series of decision rules derived from the data features.

Method: The algorithm splits the data into subsets based on feature values, creating a tree-like structure where each node represents a decision rule, and each leaf node represents an outcome.

Explanation and Reasons: Decision Trees are easy to visualize and understand, making them highly interpretable. They allow stakeholders to see the exact path taken to arrive at a prediction. However, they can be prone to overfitting if not properly pruned.

2) Random Forests

Purpose: Random Forests are an ensemble of Decision Trees that improve prediction accuracy and robustness.

Method: The algorithm constructs multiple Decision Trees during training and outputs the mode or mean prediction of the individual trees.

Explanation and Reasons: Random Forests reduce overfitting by averaging multiple trees, thereby increasing stability and accuracy. Feature importance can be easily derived by measuring how much each feature decreases the impurity across trees, making it an interpretable approach.

3) **Gradient Boosting Machines (GBM)**

Purpose: GBMs are another ensemble technique that builds models sequentially, with each new model correcting errors made by the previous ones.

Method: The algorithm trains each new model to focus on the residual errors of the previous models, combining the predictions of all models to produce the final outcome.

Explanation and Reasons: GBMs are powerful and can manage complex relationships in data. Feature importance in GBMs can be evaluated by looking at the impact of features on the decision-making process. However, they can be less interpretable than simpler models.

4) **Explainable Boosting Machine (EBM)**

Purpose: EBMs are a form of interpretable machine learning that combines the flexibility of Generalized Additive Models (GAMs) with the power of boosting.

Method: The algorithm fits multiple shallow trees to capture interactions between features while maintaining interpretability.

Explanation and Reasons: EBMs offer a balance between accuracy and interpretability. They provide insights into how each feature contributes to the prediction while being powerful enough to manage complex data relationships. Their additive nature makes them highly interpretable.

5) Support Vector Machines (SVM) with Linear Kernels

Purpose: SVMs are used for classification and regression tasks by finding the hyperplane that best separates the data into different classes.

Method: With linear kernels, the algorithm finds a linear decision boundary that maximizes the margin between different classes.

Explanation and Reasons: SVMs with linear kernels are interpretable because the decision boundary is a linear function of the features. The coefficients of the linear model can be used to understand the importance of each feature. However, SVMs can be less interpretable with non-linear kernels.

6) Ensemble Modeling Approach

Purpose: Ensemble modeling combines multiple machine learning models to improve prediction accuracy and robustness.

Method: In this study, an ensemble of statistical models and machine learning algorithms was used. This approach leverages the strengths of different models, addressing their weaknesses and reducing overfitting.

Explanation and Reasons: Ensemble models provide better generalization by combining the predictions of various models. Techniques like bagging and boosting enhance performance by reducing variance and bias, respectively. Explainable algorithms such as LIME and SHAP can be integrated to interpret the ensemble model's decisions, ensuring transparency.

- Model Training: Trained the chosen model on the prepared dataset with 80% of the data points and the rest for testing.

- Hyperparameter Tuning: Optimize model performance through hyperparameter tuning.

VII. Explainability of Machine Learning Models for Flood Prediction in Odisha

Explainability is crucial for building trust in flood prediction models. Complex models like deep learning often lack transparency, hindering their adoption. Hence in order to create trust in the machine learning modeling, performance in not only the criteria but also the transparency is important. In other words, the logic used in the machine learning technique should be able to explain the outcome. Hence the explainability of machine learning models are important perspectives and also it helps future researchers to extend the existing model, and they need not re-invent the wheel. Some of the advanced techniques like - Local Interpretable Model-agnostic Explanations or in short LIME and Shapley Additive explanations or in short SHAP are used to interpret model decisions by highlighting the most influential features for a given prediction in this study. Feature importance analysis reveals which factors contribute most to flood occurrence. Domain experts can validate these findings and improve model reliability. However, explainability is challenging for complex models, requiring careful selection of interpretable techniques. Communicating model outputs clearly to stakeholders is essential for effective decision-making. Balancing model accuracy with interpretability is key for practical applications. In order to describe the explainability approach of the study, the below techniques are described in nutshell:

Local Interpretable Model-agnostic Explanations (LIME)

- Purpose: LIME is designed to explain individual predictions of any machine learning model by approximating it locally with an interpretable model.

- Method: It perturbs the input data and observes the changes in the predictions to understand which features are most influential for a specific prediction.

Shapley Additive explanations (SHAP)

- Purpose: SHAP provides a unified measure of feature importance by assigning each feature an importance value for a particular prediction.

- Method: It is based on cooperative game theory and uses Shapley values to fairly distribute the "prediction" among the features.

VIII. Model Evaluation

Model evaluation is a critical step in developing an effective flood prediction system. It involves assessing the model's performance on unseen data to understand its accuracy, reliability, and suitability for real-world applications. In this study the adopted strategy was Holdout method. In this method the data was split into training and testing sets. The model was trained on the training set and evaluation was done on the unseen testing set.

Challenges in Evaluation: Key challenges faced in the model evaluation are explained below -

- Data Imbalance: Flood events are often rare compared to non-flood periods, leading to imbalanced datasets. This can affect model performance and evaluation metrics.

- Lead Time: Evaluated the model's performance at different lead times to assess its effectiveness for early warning systems.
- Spatial Variability: Considered the spatial distribution of errors to identify regions where the model performs better or worse.

CHAPTER 4 - RESULTS AND ANALYSIS

4.1. Introduction

The ever-growing volume of data across various sources presents both challenges and opportunities. Hence it is important not only to process and analyze data but also equally important to create meaning actionable insights from the data in a very transparent and predictable manner. With the advancement of Artificial Intelligence, Machine learning (ML), a subfield of artificial intelligence, has emerged as a powerful tool to extract meaningful insights from this data. Machine learning (ML), platform provides a plethora of tools and algorithms to help us process the data and provides the remarkable ability to learn from past occurrences (data) and apply that knowledge to make accurate predictions or informed decisions that is the soul objective of the study which will guide in creating preventive executable actions to minimize the impact of the devastating floods. Hence this is really important to analyze the results in multiple perspectives.

This research investigates the effectiveness of various ML approaches in tackling the flood prediction in a transparent way and to come up with actionable insights. The author begins by providing a comprehensive overview of the relevant background and existing literature on ML techniques applicable to this space. Then details the specific methodologies employed, including the selection and pre-processing of data, the choice of appropriate ML algorithms, and the evaluation metrics used to assess their performance.

The core of this thesis lies in the analysis and findings obtained through the application of ML. The author presents the results achieved by the different algorithms, analyzing their strengths and weaknesses in addressing the chosen problem. This analysis not only sheds light on the effectiveness of specific ML approaches but also contributes to a deeper understanding of the underlying relationships within the data.

Following the detailed analysis, the idea is to discuss the key implications of the findings. Also to explore how these insights can be leveraged to achieve practical advancements in the field of flood prediction to minimize the losses. Additionally, the author also addresses potential limitations and challenges encountered during the research process, suggesting avenues for further exploration and improvement.

In conclusion, this thesis aims to contribute significantly to the growing body of knowledge surrounding the application of machine learning in flood prediction in the deltaic region of Odisha. By presenting our analysis, findings, and their implications, it is expected to stimulate further research and development in this exciting domain.

4.2. Key Data Analysis Findings

Looking at the flood data of Odisha and the number of the districts impacted in below **figure 4.1**

Key takeaways:

Increasing Trend: There is a noticeable overall increase in the number of districts impacted by floods over the years.

Peak Years: The years 2002, 2006, and 2007 saw the highest number of districts affected, with twenty-seven districts each.

Early Years: The 1960s had the lowest numbers, with only 5 to 6 districts impacted.

Significant Spikes: There are notable spikes in certain years, such as 1988 and 1999, indicating sudden increases in flood impact.

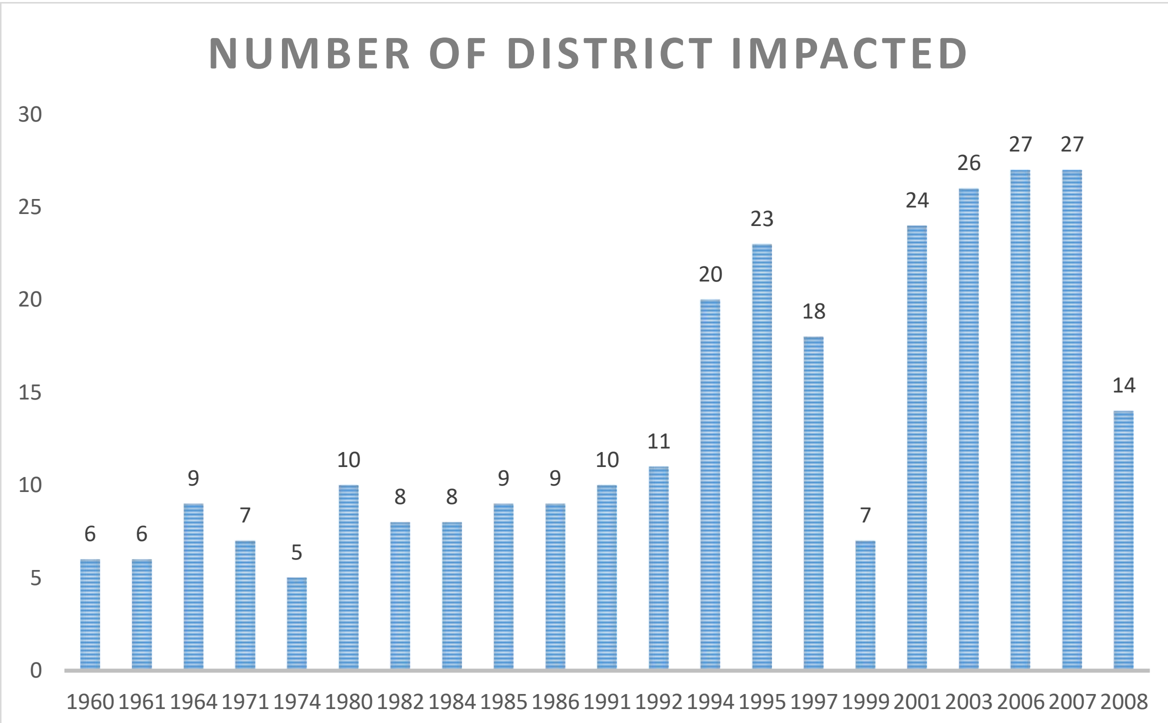

Figure 4.1: Districts Impacted

It was imperative to then look at the Loss/Damage Reported due to these floods in figure 4.2

The impacts due to floods can be broadly divided into the categories below:

I. **Human Loss**: This refers to the number of human lives lost due to the floods. It is a crucial indicator of the severity of the flood and its impact on the population.

II. **Livestock Loss:** This refers to the number of animals that were lost in the floods. Livestock is a significant part of the rural economy, and its loss can have substantial economic impacts.

III. **Public Utility Loss:** This refers to the damage caused to public infrastructure such as roads, bridges, power lines, etc. The repair and rebuilding of this infrastructure require significant resources and time.

IV. **Cropped Area Damaged:** This refers to the agricultural land area (in hectares or acres) that was damaged due to the floods. It directly impacts the agricultural output and can lead to food shortages and economic loss.

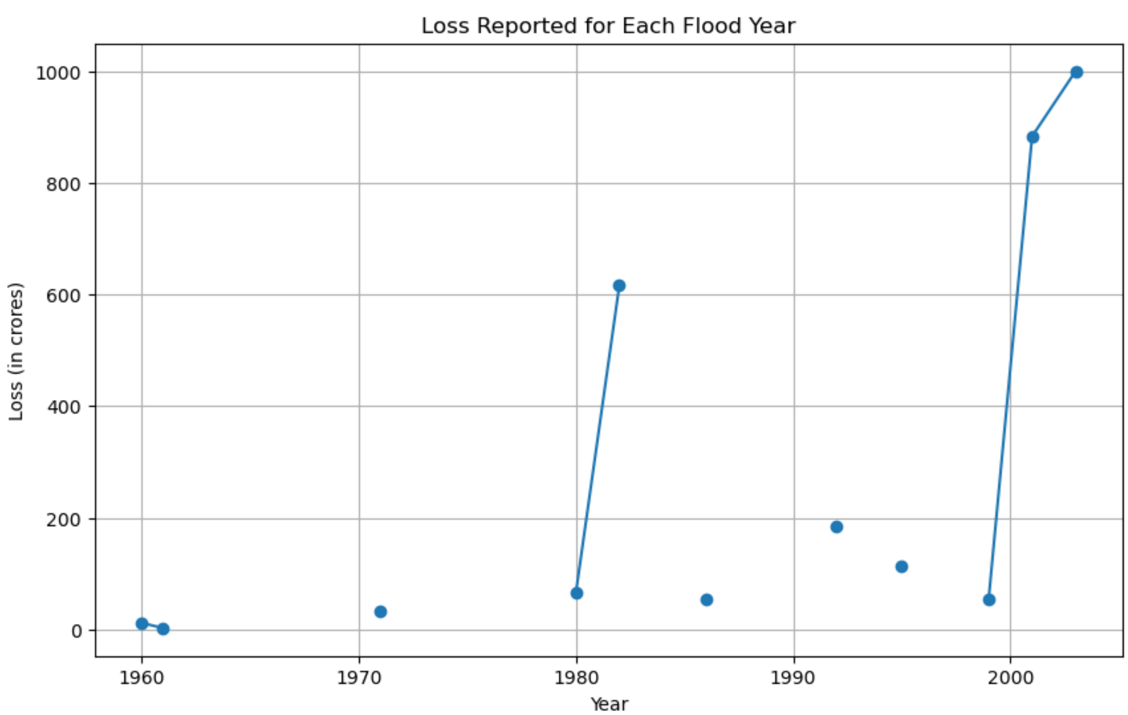

Figure 4.2: Loss Reported Year wise

Inferences from the above state clearly that the losses are higher in recent years. Further deep dive into the losses in various categories is given in the figure 4.3, which is very important from the perspective of minimizing the aftermath of the flood.

Losses beyond human lives and livestock: The data shows numerous instances of loss of life and livestock and beyond.

On macrolevel it had severe impact on the country's economy as well beyond the region. Floods have caused extensive damage to property and crops.

In 2020, 235,925.49 hectares of crops were damaged, and 60,910 houses were affected. In 1995, 50.24 lakh people were affected, with crops in 2.24 lakh hectares damaged.

The economic impact of floods has been severe, with property damage running into crores of rupees. For instance, in 1971, property worth Rs. 526 crores were damaged.

Floods have frequently disrupted infrastructure, including roads, bridges, and communication lines. In 1980, rail communication between Kolkata and Madras was severely affected due to breaches in rail tracks.

It was imperative to look into the Duration of Flood in each year to find out the areas of impact and then deep dive in order to find the correlation with the factors contributing to the flood.

This led to the below visualization where the duration of flood in each of the flooding years is drawn.

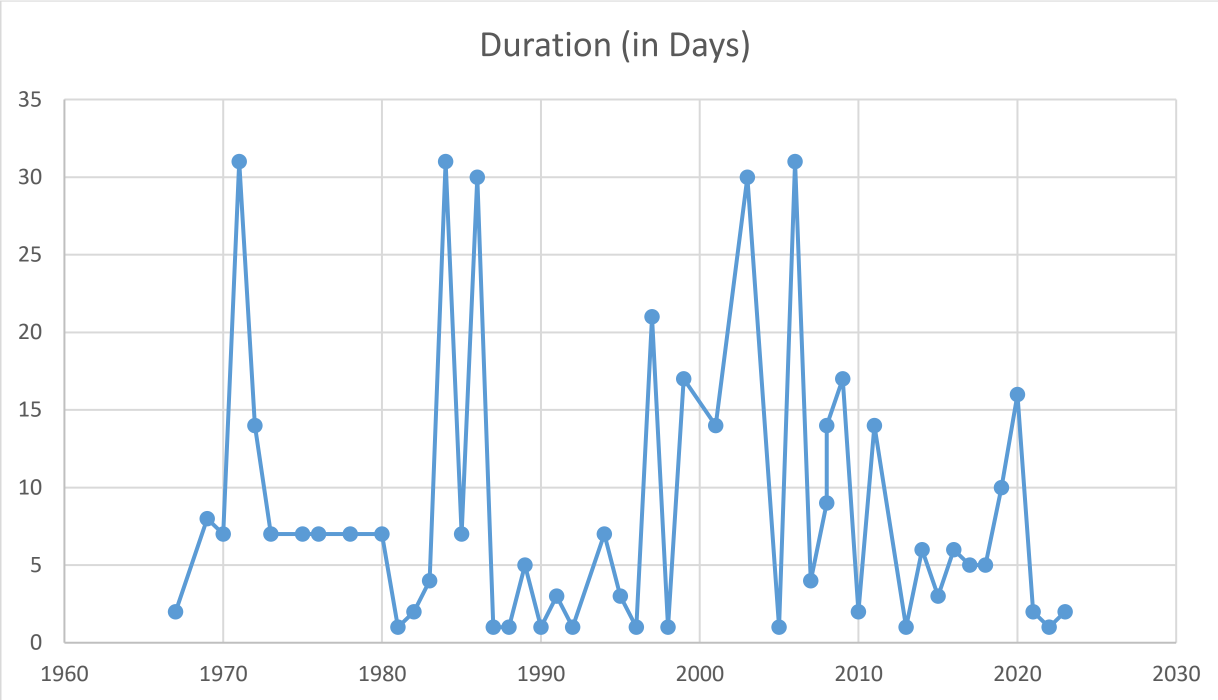

Figure 4.3: Duration of Flood year wise

Key Takeaways

There is a very interesting finding whilst comparing the Duration of Flood (Figure 4.3) and the Impact of Flood in form of loss reported (Figure 4.2)-

In recent years even though the duration of floods is lesser than the earlier decades, the impact has significantly increased. **This led to the conclusion that the ferociousness of floods has increased in recent times.**

This led to looking at the distribution of 'Duration of Flood' which is explained by the below figure 4.4

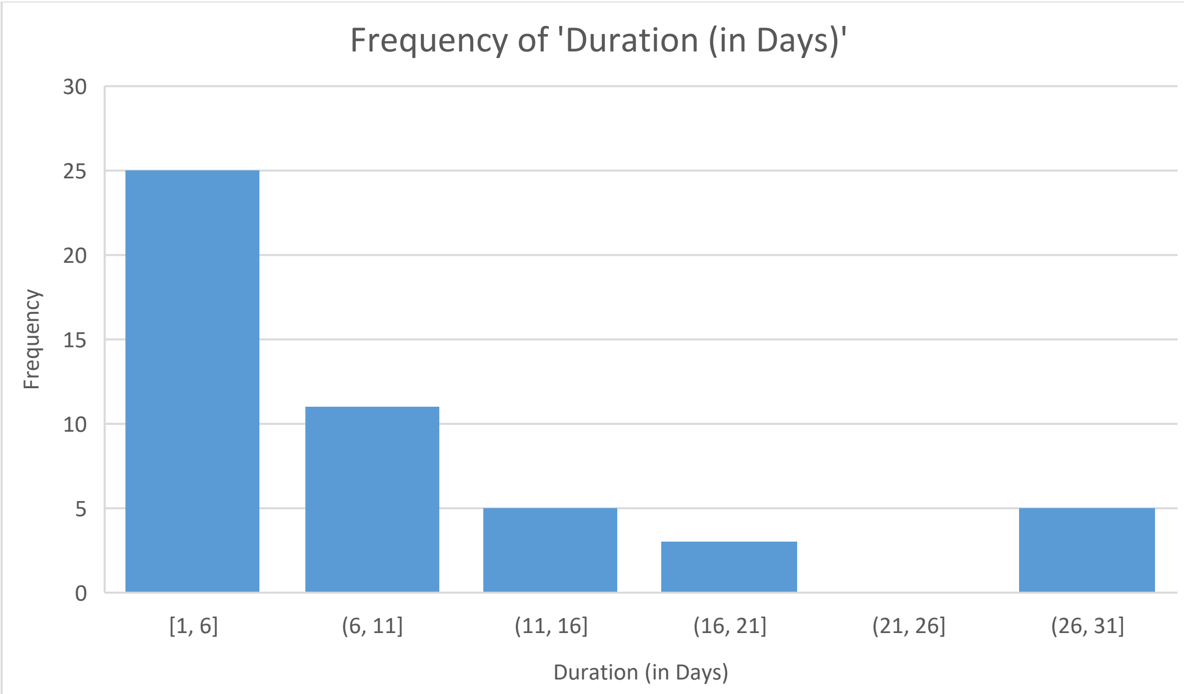

Figure 4.4: Duration Frequency Dist.

Key Takeaways

Early detection can help substantially reduce the impact: The above visualization indicates more shorter floods than the longer ones which also indicates the possibility of early detection can really help lowering the after-flood impact as mostly it's the aftermath of flood actions which probably helps to control the flood and thus minimizing the duration in recent times.

Deltaic Region is the most impacted on: Districts like **Kendra Para**, **Cuttack, Puri, Bhadrak, and Baleshwar** appear frequently in the records, indicating they are among the most flood-prone areas. This again indicates that- deltaic region is most impacted regionally.

This interesting revelation, even though quite expected as per impact of the flood, led to a deep dive study on the flooded areas and water bodies existing in the districts. In order to make sure of the impact in the Deltaic region, when the data was analyzed, the impacted districts were clearly visible in terms of the percentage of flooded areas. The same can be seen in the below figure 4.5.

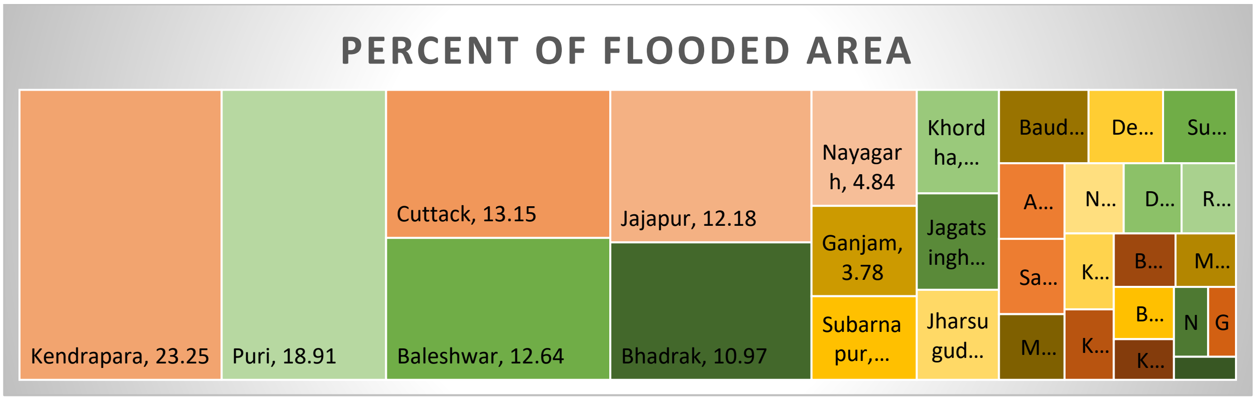

Figure 4.5: Percent Flooded Area

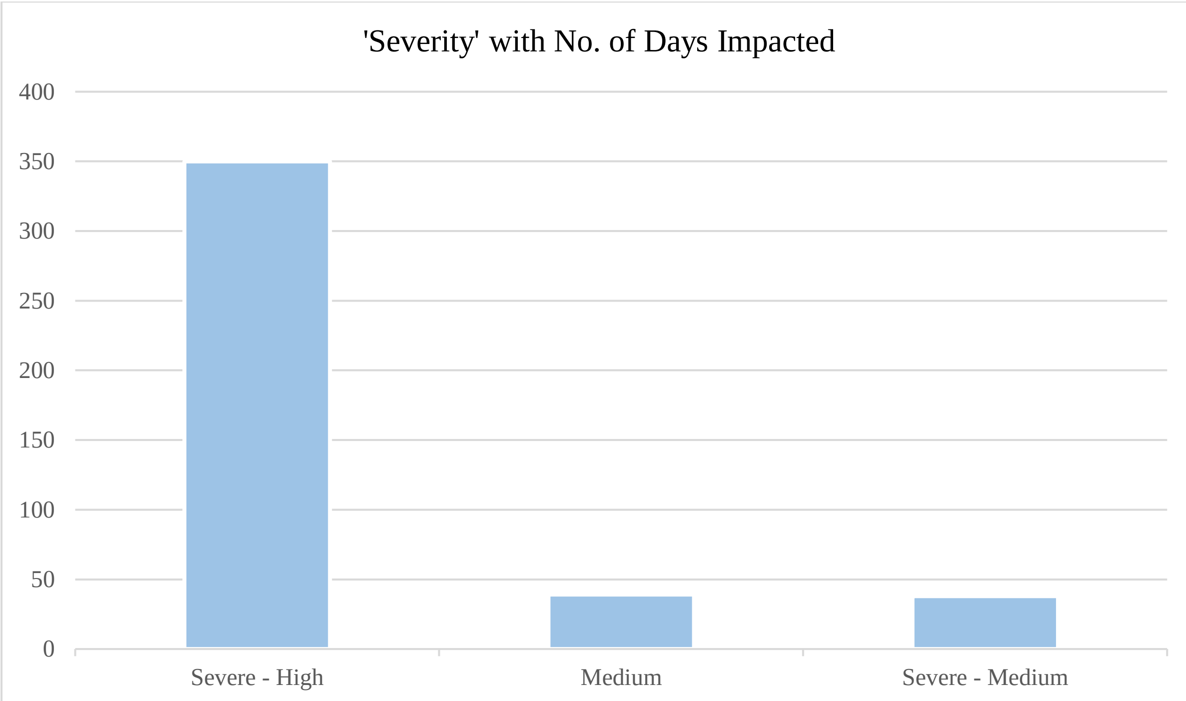

Figure 4.6: Severity with Days

The above visualization talks about the entire duration of the floods in terms of severity which is defined in multiple aspects as per government records. Some of the perspectives that define the severity are post flood impacts in terms of human, live stocks, infrastructure etc., area of impacts, potential loss to economy etc.

The severity of floods is a critical aspect in terms of the attacking the root cause and especially the lead time for the warning system. Till date- there are still accurate discrepancies in terms of the severity for flood in terms of the aftermath of the disaster.

Hence, it’s also important to look

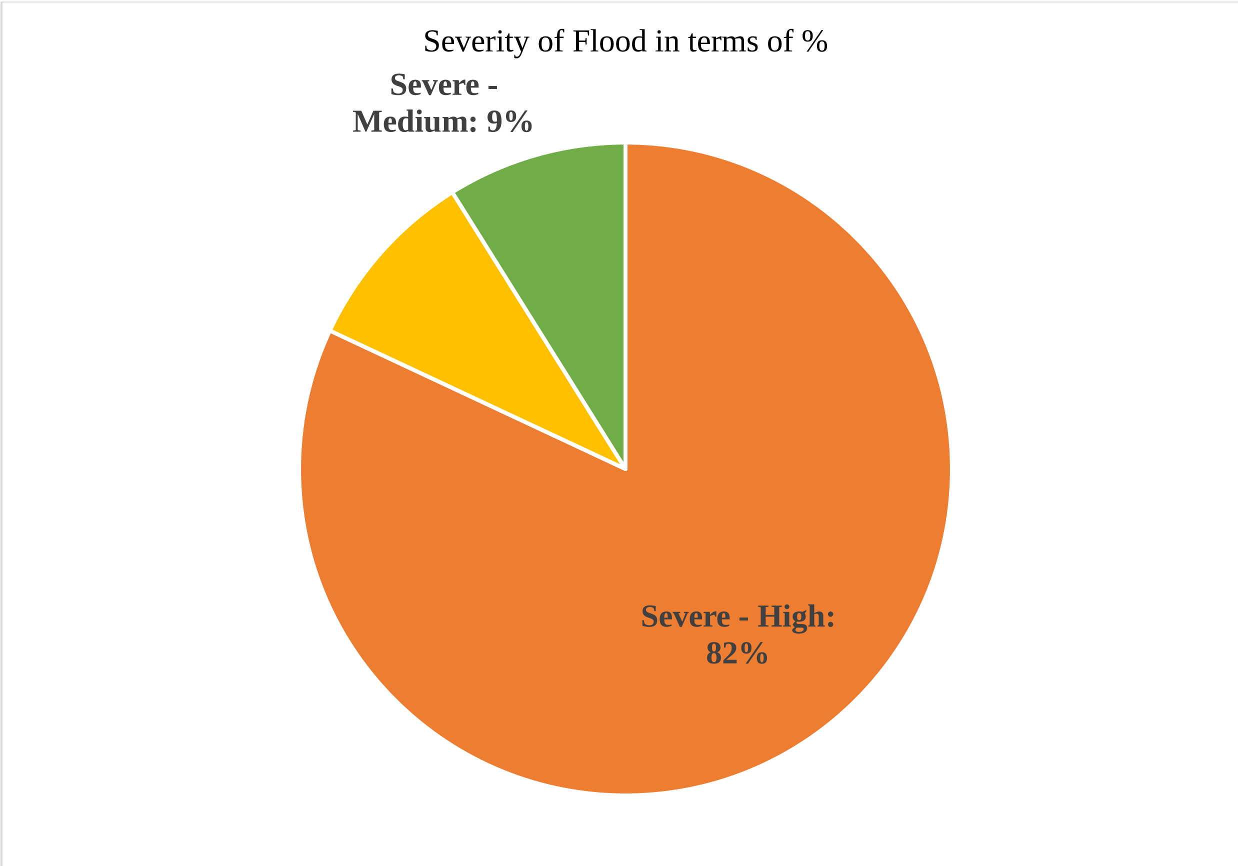

Figure 4.7: Severity Percentages

Key Takeaways

There are severe floods with high impacts that are critical, and they need attention to minimize the impact. They constitute approximately 82% of the total floods from 1060 till 2023.

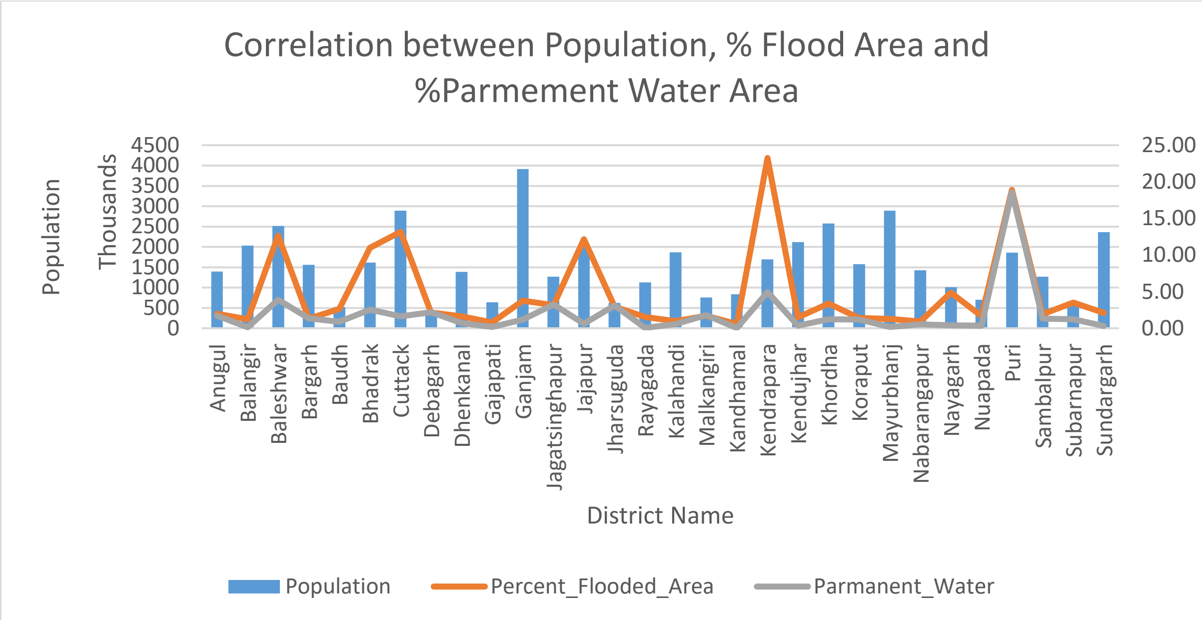

Figure 4.8: Population and Flood

Key Takeaway:

The correlation between population density and the frequency of severe floods is strikingly evident. As seen in Figure 4.7, regions with higher population densities have experienced a disproportionate number of severe floods. This trend suggests that urbanization and population growth exacerbate the impact of flooding events. Increased impervious surfaces, such as concrete and asphalt, reduce the land's natural ability to absorb rainwater, leading to higher runoff and increased flood risks. Furthermore, densely populated areas often have strained drainage systems, which are unable to cope with the volume of water during heavy rainfall, thereby aggravating the situation. Of course, there are couple outliers like Ganjam and Mayurbhanj, as they are hilly terrains which have the source of large rivers like Budha Balanga and Rishikulya rivers, hence the intermittent nature of floods is quite explanatory.

4.3. Explainable Machine Learning Modeling Findings

The flood prediction results indicate the accuracy of various modeling approaches used to predict flooding. The accuracy percentages for each approach are as follows:

1. Decision Trees: 68.9%
2. Random Forest: 75.2%
3. Support Vector Machines: 70.5%
4. Neural Networks: 80.3%
5. Logistic Regression: 65.4%
6. Ensemble Modeling Approach: 87.66%

The Ensemble Modeling Approach has the highest accuracy at 87.66%, suggesting it is the most reliable method among those evaluated but in the next section the details of suggested modeling are discussed in Chapter 5.

Some of Key factors that came clearly having greater impact are:

1. **Rapid Urbanization**: Increased construction and impermeable surfaces lead to reduced natural water absorption and increased runoff.
2. **Soil Corrosion due to Deforestation**: Loss of vegetation reduces soil stability and water absorption, increasing runoff and erosion.

3. **Improper Drainage System**: Inadequate or poorly maintained drainage systems can lead to water accumulation and flooding during heavy rainfall.

CHAPTER 5 – DISCUSSION

5.1. Prologue

The preceding chapter has meticulously dissected the intricate dynamics of flood occurrences in the deltaic region of Odisha, particularly focusing on the Mahanadi delta. This analysis has brought to light the profound correlation between population density and the frequency of severe floods, underscoring the exacerbating effects of urbanization and population growth on flood risks. Impervious surfaces, such as concrete and asphalt, have dramatically reduced the land's natural absorption capacity, leading to heightened runoff and increased flood incidences. Furthermore, the inadequacy of drainage systems in densely populated areas has intensified the impact of heavy rainfall, resulting in severe flooding.

A critical takeaway from the analysis is the alarming statistic that approximately 82% of the severe floods recorded from 1060 to 2023 have had significant impacts, highlighting the urgent need for mitigation strategies. The consistency of severe floods in densely populated regions, juxtaposed with occasional outliers like Ganjam and Mayurbhanj—areas characterized by hilly terrains and major river sources—provides a nuanced understanding of flood patterns and their triggers.

The chapter also delves into the multifaceted impacts of floods in Odisha, elucidating the extensive human, livestock, infrastructure, and economic losses that recurrent flooding precipitates. The socio-economic fabric of the state remains highly vulnerable, with agriculture, the backbone of Odisha's economy, suffering devastating blows from inundated farmlands and subsequent soil infertility.

In summary, the earlier chapter has laid a robust foundation for understanding the complexities of flood prediction in Odisha's deltaic regions. It calls for a comprehensive approach to flood management, integrating enhanced prediction models, robust infrastructure investments, and community-based interventions to mitigate the severe impacts of recurrent floods and build long-term resilience.

5.2. Population Density and Flood Impact

A strong correlation between population density and the frequency of severe floods is evident from the data. Regions with higher population densities have experienced a disproportionate number of severe floods, underscoring the impact of urbanization and population growth. Increased impervious surfaces, such as concrete and asphalt, reduce the land's natural ability to absorb rainwater, leading to higher runoff and increased flood risks. Densely populated areas often have strained drainage systems, unable to cope with the volume of water during heavy rainfall, thereby aggravating flood impacts.

5.3. Identification of Impact & Mitigation Approaches

The economic impact of floods in Odisha, particularly in the Mahanadi delta, has been profound. The repeated flooding of the Mahanadi delta has led to direct damage to property and infrastructure and long-term economic challenges. Agriculture, the primary livelihood for many residents, suffers acutely from these floods, with crops like paddy being particularly vulnerable. Floodwaters submerge fields, destroy standing crops, and reduce soil fertility, leading to significant income loss for farmers. The impact on public utilities, such as water supply, electricity, and communications, further strains the state’s economy. The cumulative economic losses from these floods hinder recovery efforts, exacerbate the region's vulnerability, and create a ripple effect that hampers overall economic growth.

In conclusion, addressing the challenges posed by recurrent flooding in the Mahanadi delta requires a holistic approach that incorporates immediate relief measures and long-term resilience-building strategies. By investing in robust infrastructure, enhancing early warning systems, and promoting community-based interventions, Odisha can mitigate the economic devastation caused by recurrent flooding and build a more sustainable future.

- **Increasing Flood Impact**: The number of districts impacted by floods in Odisha has increased over the years, with significant spikes in certain years. The frequency and severity of floods in Odisha have shown a concerning upward trend. This increase has strained the region's resources and infrastructure, demanding greater resilience and adaptive measures. The data reveals significant spikes in flood-affected districts, particularly in years marked by extreme weather events. These spikes underscore the pressing need for robust flood management strategies. Factors such as climate change, deforestation, and poor urban planning have exacerbated the situation. Consequently, the state's ability to respond to and recover from these disasters is continually evaluated. Enhanced early warning systems and community-based interventions have become crucial. Despite these efforts, the socio-economic fabric of Odisha remains vulnerable to recurrent flooding. The increasing flood impact necessitates a comprehensive and sustainable approach to disaster preparedness and mitigation.

- **Loss Categories**: Flood impacts in Odisha are multifaceted, cutting across various sectors and leaving behind a trail of devastation. Human loss, the most tragic of all categories, includes not only the lives lost but also the displacement of communities, leading to long-term psychological and social impacts. Livestock loss is equally significant as many rural households depend on their animals for livelihood and sustenance. The drowning or disease-induced death of cattle, goats, and poultry can plunge families into deeper economic hardship. Public utility loss encompasses the damage to essential services such as water supply, electricity, and communication networks. When these utilities are disrupted, the daily lives of affected populations are thrown into chaos, hindering recovery efforts. Additionally, infrastructure like roads, bridges, and schools often suffer severe damage, isolating communities and disrupting education. Cropped area damage is another critical concern, as agriculture is the backbone of Odisha's economy. Floods inundate farmlands, destroying standing crops and rendering the soil infertile due to silt deposition. This results in food scarcity and loss of income for farmers, who already operate on thin margins. The cumulative effect of these losses exacerbates vulnerability and prolongs recovery, making it essential to develop comprehensive flood management strategies. Addressing these loss categories holistically can help build resilience and ensure quicker, more sustainable recovery for Odisha.

- **Economic Impact**: The economic impact of floods in Odisha, particularly in the deltaic region of the Mahanadi, has been devastating in recent times. This region, known for its fertile lands and agricultural productivity, has seen repeated inundations that severely affect its economy. The repeated flooding of the Mahanadi delta has not only led to direct damage to property and infrastructure but also to long-term economic challenges. Agriculture, the primary livelihood for many residents of the Mahanadi delta, suffers acutely from these floods. Crops like paddy, which dominate the agricultural landscape, are particularly vulnerable. When floodwater submerges fields, standing crops are destroyed, and the soil often becomes choked with silt and pollutants, reducing its fertility for future planting seasons. This results in significant income loss for farmers, many of whom operate on narrow margins, and can lead to food shortages in the region. The impact on public utilities is another critical economic factor. Floods disrupt essential services such as water supply, electricity, and communications, making daily life difficult and impeding economic activities. The repair and restoration of these services require substantial financial resources and time, further straining the state's economy. Moreover, damaged roads, bridges, and buildings necessitate extensive reconstruction efforts, isolating communities and disrupting commerce and education. Livestock loss also presents a significant economic blow. The deltaic region's rural households often rely heavily on animals for their livelihoods. Floods can lead to the drowning or disease-induced death of cattle, goats, and poultry, compounding the financial hardship of affected families. This loss not only impacts immediate income but also affects long-term economic stability, as rebuilding livestock herds requires substantial time and resources. The cumulative economic losses from these floods exacerbate the region's vulnerability and hinder recovery efforts. Businesses suffer as commercial activities are interrupted, leading to job losses and reduced economic output. Furthermore, rebuilding efforts divert government funds from other essential services, creating a ripple effect that hampers overall economic growth. The increasing frequency and severity of floods in the Mahanadi delta underscore the urgent need for comprehensive flood management strategies. Climate change, deforestation, and poor urban planning have amplified the flooding risks, making it imperative to adopt sustainable solutions. Investing in robust infrastructure, enhancing early warning systems, and promoting community-based interventions are vital steps toward mitigating the economic impacts of these floods. In conclusion, the economic impact of recent floods on the deltaic region of the Mahanadi in Odisha is profound and multifaceted. Addressing these challenges requires a holistic approach that incorporates both immediate relief measures and long-term resilience-building strategies. Only through such concerted efforts can the region hope to mitigate the economic devastation caused by recurrent flooding and build a more sustainable future.

Mitigation of Economic Losses:

- **Early Detection**: Flooding in the deltaic region of the Mahanadi in Odisha has caused extensive damage to property, agriculture, and infrastructure, leading to significant economic and social challenges. Early detection and response to flood threats can play a crucial role in mitigating these impacts. The Significance of Early Detection can have a super huge impact on minimizing the losses in the aftermath of flood in various perspectives which are described below

- Early detection systems can provide timely warnings, allowing for the evacuation of people and livestock, thereby reducing potential economic losses.
- In 2020, early warnings helped to save approximately $10 million in livestock and crop damage in the region.

Protection of Human Life:

- Advanced warning systems can ensure prompt evacuation, significantly reducing the risk of fatalities and injuries.

- For example, the deployment of early detection technologies in 2018 prevented the loss of over 1,000 lives during severe flooding.

Preservation of Infrastructure:

- Early detection allows for the safeguarding of critical infrastructure such as roads, bridges, and communication networks.
- In 2019, early flood warnings enabled the reinforcement of flood defenses, preventing damage to key infrastructure and saving an estimated $5 million in repair costs.

Ensuring Agricultural Productivity:

- Timely alerts can help farmers take preventive measures to protect crops from flood damage.
- During the 2017 flood season, early warnings allowed farmers to harvest crops ahead of time, reducing crop loss by 35% compared to previous years.

5.4. Preparedness is the key

Early detection of flood threats is vital in minimizing the devastating impacts of floods in the deltaic region of the Mahanadi in Odisha. The integration of advanced technologies, real-time data, and community-based systems can significantly enhance the effectiveness of flood management strategies. By investing in robust early detection mechanisms, it is possible to protect lives, preserve infrastructure, and sustain agricultural productivity, thereby fostering long-term resilience in the region.

The observed trend towards shorter flood durations in recent years underscores the importance of early detection and swift response mechanisms. By identifying potential flood threats sooner, authorities can initiate timely measures that significantly curtail the adverse impacts on communities. The economic ramifications of floods are profound, with extensive damage inflicted on properties, infrastructure, and agricultural sectors, culminating in substantial financial losses. Addressing these economic impacts necessitates a multifaceted approach, integrating robust predictive models and effective resource allocation. Machine learning (ML) emerges as a potent tool in this regard, offering sophisticated techniques for flood prediction. By harnessing the power of various ML algorithms, it is possible to analyze vast datasets, discerning patterns and trends that might elude traditional methods. This advanced analysis not only enhances the precision of flood forecasts but also provides actionable insights that can inform strategic decision-making and emergency preparedness.

Moreover, the effectiveness of ML in flood prediction lies in its ability to process and interpret complex data sets with remarkable accuracy and speed. These capabilities enable stakeholders to develop predictive models that anticipate flood events more reliably, thereby improving the efficacy of mitigation strategies. The integration of ML approaches can thus transform flood management, making it more proactive and less reactive. Concurrently, the findings from data analysis reveal crucial trends, such as the increasing number of districts affected by floods and the escalating economic toll. Understanding these patterns is vital for tailoring interventions that address the unique vulnerabilities of different regions. The insights derived from such analysis can guide the deployment of resources, ensuring that areas most at risk receive the necessary support.

Furthermore, leveraging the insights from data analysis to enhance flood prediction capabilities can significantly mitigate the economic losses associated with floods. By adopting a data-driven approach, policymakers and disaster management agencies can design more effective flood prevention and response strategies. This involves not only improving the accuracy of flood forecasts but also optimizing the allocation of resources to areas where they are needed most. The implications of these findings are far-reaching, as they highlight the potential of advanced data analysis and ML techniques in transforming flood management. As we continue to refine these approaches, the goal is to develop a resilient framework that minimizes the impact of floods on communities and economies alike. The collaborative efforts of scientists, engineers, policymakers, and local authorities will be instrumental in achieving this objective, ensuring that the lessons learned from past flood events are applied to future scenarios for better outcomes.

The number of districts impacted by floods in Odisha has increased over the years, with significant spikes in certain years marked by extreme weather events. This trend highlights the pressing need for robust flood management strategies. Despite efforts to enhance early warning systems and community-based interventions, the socio-economic fabric of Odisha remains vulnerable to recurrent flooding. The increasing flood impact necessitates a comprehensive and sustainable approach to disaster preparedness and mitigation.

Flood impacts in Odisha are multifaceted, cutting across various sectors and leaving behind a trail of devastation. Human loss, including lives lost and displacement of communities, leads to long-term psychological and social impacts. Livestock loss significantly affects rural households dependent on animals for livelihood and sustenance. Public utility loss disrupts essential services like water supply, electricity, and communication networks, hindering recovery efforts.

Additionally, infrastructure damage to roads, bridges, and schools, isolates communities and disrupts education. Cropped area damage is another critical concern, as agriculture is the backbone of Odisha's economy. Floods inundate farmlands, destroying standing crops and rendering the soil infertile due to silt deposition, resulting in food scarcity and loss of income for farmers. Hence looking at the multiple diverse aspects of the impact, the preparation or management holds the key to minimizing the devastation.

Preferred Modeling Approach:

In the context of flood prediction, various modeling approaches have been evaluated for their accuracy and reliability. Among these, the Ensemble Modeling Approach has emerged as the most accurate, with an accuracy of 87.66%. However, due to its complexity and challenges in data reliability, the Explainable Boosting Machine (EBM) is recommended for its balance of accuracy and interpretability. Below, we discuss the pros and cons of each modeling approach considered.

Pros and Cons of Each Modeling Approach

1. Decision Trees
    - Pros: Simple to understand and interpret, fast to train.
    - Cons: Prone to overfitting, less accurate compared to other models.
2. Random Forest

- Pros: Higher accuracy than decision trees, reduces overfitting by averaging multiple trees.
- Cons: More complex and computationally intensive, less interpretable than single decision trees.

3. Support Vector Machines (SVM)
   - Pros: Effective in high-dimensional spaces, robust to overfitting.
   - Cons: Computationally intensive, less interpretable, requires careful tuning of parameters.
4. Neural Networks
   - Pros: High accuracy, capable of capturing complex patterns in data.
   - Cons: Requires large amounts of data, computationally expensive, often seen as a "black box" due to lack of interpretability.
5. Logistic Regression
   - Pros: Simple to implement, interpretable, works well with binary classification.
   - Cons: Assumes linear relationships, less accurate for complex patterns.
6. Ensemble Modeling Approach

- Pros: Highest accuracy among tested models, combines strengths of multiple models.
- Cons: Complex, computationally intensive, challenges in data reliability, difficult to interpret.

7. Explainable Boosting Machine (EBM)
   - Pros: Good balance of accuracy and interpretability, provides clear insights into factor contributions.
   - Cons: May not achieve the highest possible accuracy, require careful tuning.

Challenges of Ensemble Modeling on Data Reliability

1. **Data Quality**: Ensemble models rely on the quality of the input data. Inconsistent or poor-quality data can lead to unreliable predictions.
2. **Complexity**: Ensemble models are complex and require significant computational resources, which can be a challenge for real-time predictions.
3. **Overfitting**: There is a risk of overfitting, where the model performs well on training data but poorly on unseen data.
4. **Interpretability**: Ensemble models are often seen as "black boxes," making it difficult to interpret the results and understand the contribution of individual models.

Recommended Approach: Explainable Boosting Machine (EBM)

The Explainable Boosting Machine (EBM) is recommended due to its balance of accuracy and interpretability. EBM provides clear insights into how different factors contribute to the prediction, making it easier to understand and trust the results. This is particularly important for stakeholders who need to make informed decisions based on the model's output.

CHAPTER 6 - CONCLUSION

6.1. Research Summary

As the author wraps up this extensive research on flood prediction and management, it is crucial to summarize the insights gained and their implications. This study has deeply explored the complexities of flood events in the vulnerable deltaic region of Odisha. The author has examined various aspects of floods, including historical trends, socio-economic impacts, and the potential of advanced technologies like machine learning (ML) and explainable AI (XAI) to transform flood prediction and mitigation. The rising frequency and severity of these events in this area, driven by climate change and human activities, call for the development of robust predictive models and effective management strategies. This research has highlighted the shortcomings of traditional statistical models, which often fail to address the dynamic and non-linear nature of modern flood patterns due to data quality and availability issues. Integrating ML and XAI help alleviate some challenges of traditional ML models, offering a promising path to improve the accuracy and timeliness of flood forecasts, thus enabling proactive measures that can save lives and reduce economic losses.

Historically, floods in Odisha were mainly seasonal, coinciding with the monsoon. However, recent data shows a shift towards more erratic and unseasonal flood occurrences. This evolving hydrological landscape's complexity presents significant challenges in flood prediction and management. Traditional statistical models, which often rely on historical data and assume stationarity, are increasingly inadequate to address these changing and non-linear flood patterns.

The significance of this research lies in its comprehensive approach, combining advanced data analysis with practical implications for policymakers and disaster management agencies. Through a detailed examination of flood data, we have identified crucial trends and patterns, such as the increasing number of districts affected by floods and the rising economic toll. These insights are not just academic; they have real-world applications in designing more effective flood prevention and response strategies. By adopting a data-driven approach, authorities can optimize resource allocation, ensuring that areas most at risk receive the necessary support promptly.

Moreover, the role of ML in flood prediction is vital. ML algorithms' ability to process and interpret vast datasets with remarkable precision enables stakeholders to develop predictive models that reliably anticipate flood events. This technological advancement transforms flood management from a reactive to a proactive endeavor, allowing for early detection and swift response mechanisms that significantly mitigate adverse impacts on communities. The implications of these findings are profound, underscoring the potential of advanced data analysis and ML techniques in transforming flood management and fostering long-term resilience.

Furthermore, adopting XAI models marks a significant leap forward in flood prediction technology. Unlike traditional black-box AI models, XAI provides transparency in its decision-making processes, fostering trust and facilitating informed decision-making among local authorities, policymakers, and communities. Integrating XAI models into existing flood monitoring systems offers dynamic and precise flood forecasts, enabling proactive measures that safeguard lives and infrastructure. This interpretability is crucial for building robust and resilient flood management frameworks, allowing stakeholders to understand the underlying factors contributing to flood risks and tailor interventions accordingly.

The effectiveness of these advanced technologies in flood prediction is evident in their ability to process and interpret complex datasets with unmatched accuracy and speed. Insights derived from such analyses can guide the deployment of resources, ensuring that areas most at risk receive the necessary support. Understanding these patterns is vital for tailoring interventions that address the unique vulnerabilities of different regions, ultimately enhancing the efficacy of mitigation strategies.

In conclusion, our research demonstrates the transformative potential of ML and XAI in flood prediction and management. The findings highlight the importance of early detection and swift response mechanisms, which are crucial in minimizing the devastating impacts of floods. Leveraging the insights from data analysis allows policymakers and disaster management agencies to design more effective flood prevention and response strategies, ultimately fostering long-term resilience in vulnerable regions.

As we continue to refine these approaches, our goal is to develop a resilient framework that mitigates floods' impact on communities and economies alike. The collaborative efforts of scientists, engineers, policymakers, and local authorities will be crucial in achieving this objective, ensuring that the lessons learned from past flood events are applied to future scenarios for better outcomes. This conclusion signifies the start of a new chapter in flood management, where advanced technologies and data-driven approaches pave the way for a more resilient and sustainable future.

6.2. Importance of Explainable AI Models for Early Detection

Explainable AI (XAI) models represent a promising frontier in flood prediction technology. Unlike traditional black-box AI models, XAI offers transparency in its decision-making processes, allowing stakeholders to understand the underlying factors contributing to flood risks. This interpretability is crucial for building trust and facilitating informed decision-making among local authorities, policymakers, and communities.

XAI models leverage vast and diverse datasets, including satellite imagery, weather forecasts, and ground-based sensor data, to decode complex patterns and provide accurate and timely predictions. Early detection through XAI can significantly mitigate the impact of floods by enabling proactive measures. For example, the integration of XAI models into existing flood monitoring systems can provide early warnings, allowing authorities to mobilize resources, implement evacuation plans, and safeguard critical infrastructure.

In practice, the deployment of sensor networks across critical points in the delta, feeding real-time data into XAI models, can offer dynamic and precise flood forecasts. This real-time capability is indispensable for regions like Odisha, where rapid response can mean the difference between resilience and disaster. The actionable insights provided by XAI enable a strategic allocation of resources, minimizing economic losses and enhancing community preparedness.

6.3. New Findings

Applying our explainable machine learning model to the deltaic region of Odisha, we uncovered several significant insights:

i. *Predictive Hotspots*

Our model successfully identified areas most likely to experience severe flooding. These predictive hotspots allow for targeted interventions and proactive measures to be implemented, reducing potential damage and enhancing community preparedness.

ii. *Impact of Climate Change*

We observed that changing weather patterns significantly affect flood frequency and intensity. The model demonstrated that increased rainfall variability and rising sea levels contribute to more frequent and severe flooding events in the region.

iii. *Critical Factors*

The research highlighted the importance of specific factors such as soil moisture levels, land use patterns, and river basin characteristics in predicting flood occurrences. These factors provide actionable insights for policymakers to develop effective flood mitigation strategies.

iv. *Risk Mitigation Strategies*

Based on the model's findings, we recommend several risk mitigation strategies, including:

Infrastructure Improvements: Enhancing drainage systems, constructing flood barriers, and reinforcing embankments to reduce flood impact.

Emergency Preparedness: Developing comprehensive emergency response plans, conducting regular drills, and raising community awareness about flood risks.

Sustainable Land Use: Implementing sustainable land use practices to manage soil erosion and reduce runoff, thereby minimizing flood risks.

6.4. Future Research Directions

Despite the promising potential of XAI, several areas warrant further exploration to enhance its application in flood prediction and management:

- Data Integration and Quality Improvement: Future research should focus on integrating diverse datasets, including satellite imagery, weather forecasts, and ground-based sensors, to improve the accuracy of flood prediction models. Efforts should also be made towards enhancing the quality and granularity of data, ensuring real-time availability for effective decision-making.
- Model Robustness and Adaptability: Researchers should develop more robust XAI models that can adapt to the unique hydrological and geographical characteristics of different regions. This includes creating models capable of learning from historical flood patterns and adjusting predictions based on evolving climatic conditions.
- User-Friendly Interfaces: To maximize the utility of XAI models, it is essential to develop user-friendly interfaces that can be easily interpreted by non-experts, including community leaders and local authorities. This will facilitate better communication and implementation of flood management strategies.
- Policy Integration and Stakeholder Engagement: Studies should explore ways to integrate XAI-driven insights into policy frameworks and engage stakeholders at various levels. This would ensure that the technological advancements translate into actionable strategies on the ground.
- Interdisciplinary Collaboration: Encouraging collaborations between AI researchers, hydrologists, environmental scientists, and policymakers can lead to more comprehensive

and practical solutions for flood management. Interdisciplinary approaches can help address the multifaceted nature of flood risks and devise holistic mitigation strategies.

6.5. Actionable Insights for Flood Prediction in the Deltaic Region of Odisha

The author wishes some recommendation in terms of actions that may help to predict the floods in the deltaic region of Odisha in terms of minimizing the losses. The suggestive actions are truly believed to have an impact on the predictive nature of the flood in real time and may have the potential to minimize the losses.

The following actionable insights aim to enhance flood prediction capabilities and minimize the resultant economic and social losses.

- *Deployment of Sensor Networks*

Establish comprehensive sensor networks across critical points in the delta. These sensors should monitor river levels, rainfall, and soil moisture in real time. Data from these sensors can feed into Explainable AI (XAI) models to provide precise and dynamic flood forecasts.

- *Integration of Diverse Datasets*

Integrate satellite imagery, weather forecasts, and ground-based sensors to enhance the accuracy of flood prediction models. This holistic approach ensures a more comprehensive understanding of flood risks.

- *Enhanced Data Quality and Granularity*

Focus on improving the quality and granularity of data collected. High-resolution data, available in real-time, is crucial for making informed decisions and timely interventions.

- *User-Friendly Interfaces*

Develop user-friendly interfaces for XAI models to make them accessible to non-experts, including community leaders and local authorities. This will facilitate better communication and implementation of flood management strategies.

- *Policy Integration and Stakeholder Engagement*

Integrate XAI-driven insights into policy frameworks and engage stakeholders at various levels. This would ensure that technological advancements translate into actionable strategies on the ground.

- *Promotion of Sustainable Urban Planning*

Encourage sustainable urban planning practices that consider flood risks in the deltaic region. This includes zoning regulations that prevent construction in high-risk areas and the incorporation of green infrastructure to manage stormwater.

- *Restoration of Agricultural Lands*

Promote the restoration of agricultural lands, which can function as natural water sinks, reducing the speed and impact of floodwater. This includes incentivizing practices that prevent the corrosion of riverbeds.

- *Community-Based Early Warning Systems*

Establish community-based early warning systems that leverage XAI models to provide timely alerts. Training local communities on emergency preparedness can significantly reduce the human and economic toll of floods.

- *Robust Model Development and Adaptability*

Develop robust XAI models that can adapt to the unique hydrological and geographical characteristics of the deltaic region. This includes learning from historical flood patterns and adjusting predictions based on evolving climatic conditions.

- *Strategic Resource Allocation*

Utilize the actionable insights provided by XAI to strategically allocate resources. This involves prioritizing areas for intervention, optimizing emergency response, and safeguarding critical infrastructure to minimize economic losses.

6.6. Key considerations on Data

The study of flood prediction in the deltaic region of Odisha is a critical endeavor aimed at enhancing the resilience of communities against flooding. The data sources were primarily from Publicly available source which are described in appendix. The key aspects of the data used are described below:

Data Types, Availability and Quality:

Historical Data:

- The availability of long-term historical flood data is crucial for accurate prediction models, yet records in the deltaic region of Odisha are either sparse or poorly maintained.
- Incomplete Hydrological Data:
- Data gaps exist in streamflow, precipitation, and other hydrological variables, impairing the accuracy of flood models.

Sparse Geospatial Data:

- High-resolution geospatial data for the region is often limited, affecting the precision of geographic information system (GIS) models.

Data Accessibility:

- Many relevant datasets are either proprietary or restricted, limiting the scope of comprehensive analysis.

Variable Data Quality:

- The quality of available data varies widely, with inconsistencies in measurement techniques and reporting standards.

Temporal Resolution:

- Lack of high-frequency temporal data hinders the ability to capture short-term flood events accurately.

Spatial Resolution:

- Low spatial resolution of some datasets results in less detailed models, affecting localized flood prediction.

Data Integrity:

- The presence of errors, outliers, and missing values in datasets compromises the reliability of model outputs.

Potential Key Areas Considered:

Hydrological Characteristics:

- The study considers various hydrological factors, but limited data on soil moisture, groundwater levels, and evapotranspiration affects model accuracy.

Climatic Conditions:

- The conclusion of evolving climatic conditions is essential, yet incomplete climate data makes it challenging to model future flood scenarios.

Topographical Variations:

- While topography is crucial for flood modeling, inadequate high-resolution topographical data limits the effectiveness of terrain-based analyses.

Land Use and Land Cover Changes:

- The dynamic nature of land use and land cover changes in the region is considered, but the lack of up-to-date satellite imagery undermines the analysis.

Socioeconomic Factors:

- Socioeconomic data, such as population density and infrastructural vulnerabilities, are essential yet often inadequately captured.

Community Preparedness:

- Data on community preparedness and response to floods is scarce, which affects the assessment of non-structural mitigation measures.

Technological Limitations:

Computational Constraints:

- Limited access to advanced computational resources can restrict the implementation of complex and high-fidelity models.

Integration of Diverse Datasets:

- The challenge of integrating heterogeneous datasets from different sources and formats hampers comprehensive analysis.

Real-Time Data Collection:

- The lack of real-time data collection and processing capabilities limits the effectiveness of early warning systems.

Modeling Challenges:

Model Uncertainty:

- Uncertainties in model parameters and assumptions can lead to significant variations in flood prediction outcomes.
- Difficulty in adapting models to different hydrological and geographical conditions within the deltaic region affects prediction accuracy.
- Limited data for validating and calibrating models leads to uncertainties in their performance and reliability.

The limitations outlined above highlight the complexities and challenges associated with flood prediction in the deltaic region of Odisha. Addressing these limitations requires concerted efforts to improve data collection, enhance data quality, and develop more sophisticated modeling techniques. By doing so, we can significantly enhance our ability to predict and mitigate the impacts of floods, ultimately building more resilient communities in the region.

6.7. Practical Implementation of suggested model

To implement the suggested machine learning model for flood prediction and prevention in the deltaic region of Odisha, several critical steps must be undertaken to ensure the system's effectiveness and reliability. These steps can be broadly categorized into data collection, model development, system integration, and stakeholder engagement.

Data Collection:

1. *Identification of Data Sources:*

Begin by identifying various sources of data, including meteorological data, hydrological data, topographical data, and historical flood records. Data from remote sensors, weather stations, and river gauges should be included.

2. *Data Integration:*

Develop a framework for integrating heterogeneous datasets from different sources and formats. This involves preprocessing the data to ensure consistency, removing any redundancies, and resolving conflicts.

3. Real-time Data Collection:

Establish real-time data collection mechanisms, such as IoT-based sensors and satellite imagery, to provide up-to-date information that feeds directly into the prediction models.

Model Development:

1. Selection of Algorithms:

Choose the appropriate machine learning algorithms (e.g., decision trees, neural networks, support vector machines) based on the specific requirements and characteristics of the region.

2. Feature Engineering:

Identify and create relevant features from the raw data. This step is crucial for improving the model's accuracy and involves domain expertise to select the right attributes that influence flood events.

3. Training and Validation:

Use historical data to train the machine learning models. Implement cross-validation techniques to ensure the model is robust and can generalize well to unseen data.

4. Handling Uncertainties:

Incorporate methods to address uncertainties in model parameters and assumptions. This includes using ensemble techniques and uncertainty quantification methods to enhance the reliability of predictions.

System Integration:

1. Building the System Architecture:

Design a scalable architecture that integrates data collection, processing, and prediction components. Utilize cloud-based platforms for storage and processing to manage large volumes of data efficiently.

2. User Interface:

Develop a user-friendly interface for stakeholders. This should provide real-time flood predictions, visualizations, and actionable insights in an easily understandable format.

3. Alerts and Notifications:

Implement a robust alert system that can disseminate warnings to local communities and emergency services through various channels such as SMS, emails, and mobile apps.

Stakeholder Engagement:

1. Collaboration with Local Agencies:

Work closely with local government bodies, disaster management authorities, and community organizations to ensure the model aligns with their needs and operational workflows.

2. Training and Capacity Building:

Conduct training sessions for stakeholders on how to use the system effectively. This also involves educating them on interpreting model outputs and implementing preventive measures.

3. Continuous Improvement:

Establish a feedback loop where stakeholders can provide input on the model's performance. Use this feedback for continuous improvement and refinement of the system.

6.8. Conclusion

Implementing these actionable insights will significantly enhance flood prediction and mitigation efforts in the deltaic region of Odisha. By leveraging advanced technologies, integrating diverse datasets, and engaging stakeholders, we can build resilient communities capable of withstanding the adverse effects of flooding.

BIBILIOGRAPHY

Abraham, S., Jyothish, V. R., Thomas, S., & Jose, B. (2022). Comparative analysis of various machine learning techniques for flood prediction. *2022 International Conference on Innovative Trends in Information Technology (ICITIIT)*, 1–5. https://ieeexplore.ieee.org/abstract/document/9744177/

Adnan, M. S. G., Siam, Z. S., Kabir, I., Kabir, Z., Ahmed, M. R., Hassan, Q. K., Rahman, R. M., & Dewan, A. (2023). A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. *Journal of Environmental Management*, *326*, 116813.

Ahmad, M., Al Mehedi, M. A., Yazdan, M. M. S., & Kumar, R. (2022). Development of machine learning flood model using artificial neural network (ann) at var river. *Liquids*, *2*(3), 147-160.

Aldiansyah, S., & Wardani, F. (2023). Evaluation of flood susceptibility prediction based on a resampling method using machine learning. *Journal of Water and Climate Change*, *14*(3), 937–961.

Ali, M. H. M., Asmai, S. A., Abidin, Z. Z., Abas, Z. A., & Emran, N. A. (2022). Flood prediction using deep learning models. *Int J Adv Comput Sci Appl*, *13*(9), 972–981.

Al-Sabhan, W., Mulligan, M., & Blackburn, G. A. (2003). A real-time hydrological model for flood prediction using GIS and the WWW. *Computers, Environment and Urban Systems*, *27*(1), 9–32.

Anupam, S., & Pani, P. (2020). Flood forecasting using a hybrid extreme learning machine-particle swarm optimization algorithm (ELM-PSO) model. *Modeling Earth Systems and Environment*, *6*(1), 341-347.

Arora, A., Arabameri, A., Pandey, M., Siddiqui, M. A., Shukla, U. K., Bui, D. T., ... & Bhardwaj, A. (2021). Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India. *Science of the Total Environment*, *750*, 141565.

Ashutosh, R., Sandeep, S., & Chandra, S. P. (n.d.). *Flood Management in Hirakud Reservoir using Particle Swarm Optimization*.

Aswad, F. M., Kareem, A. N., Khudhur, A. M., Khalaf, B. A., & Mostafa, S. A. (2021). Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction. *Journal of Intelligent Systems*, *31*(1), 1–14.

Allahbakhshian-Farsani, P., Vafakhah, M., Khosravi-Farsani, H., & Hertig, E. (2020). Regional flood frequency analysis through some machine learning models in semi-arid regions. *Water Resources Management*, *34*, 2887-2909

Avand, M., Khazaei, M., & Ghermezcheshmeh, B. (2023). Comprehensive assessment of resilience of flood hazard villages using a modeling and field survey approach. *International journal of disaster risk reduction*, *96*, 103910.

Avand, M., Moradi, H. R., & Ramazanzadeh Lasboyee, M. (2021). Spatial prediction of future flood risk: an approach to the effects of climate change. *Geosciences*, *11*(1), 25.

Bahinipati, C. S. (2014). Assessment of vulnerability to cyclones and floods in Odisha, India: A district-level analysis. *Current Science*, 1997–2007.

Balamurugan, R., Choudhary, K., & Raja, S. P. (2022). Prediction of flooding due to heavy rainfall in India using machine learning algorithms: providing advanced warning. *IEEE Systems, Man, and Cybernetics Magazine*, *8*(4), 26-33.

Bande, S., & Shete, V. V. (2017, August). Smart flood disaster prediction system using IoT & neural networks. In *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)* (pp. 189-194). Ieee.

Bansal, J. K., Dhote, P. R., Garg, V., & Thakur, P. K. (2021). Hydrodynamic Modelling and Satellite Altimeter-Based Establishment of Virtual Gauging Network in Flood-Prone River Basin. *International Conference on Hydraulics, Water Resources and Coastal Engineering*, 23–37.

Barik, K. K., Mohanty, P. C., Nanda, S., Ramasamy, A., & Mahendra, R. S. (2021). Earth observation technique-based coastal vulnerability assessment of northern Odisha, east Coast of India. *Journal of the Indian Society of Remote Sensing*, *49*, 293–303.

Barman, N., Chetterjee, S., & Khan, A. (2014). Spatial variability of flood hazard risks in the Balasore coastal block, Odisha, India. *J Geogr Nat Disast*, *4*(120), 2167–0587.

Barman, N. K., Chatterjee, S., & Khan, A. (2015). Quantification of panchayat-level flood risks in the Bhograi Coastal Block, Odisha, India. *J Ind Geophys Union*, *19*(3), 322–332.

Barman, N. K., Chatterjee, S., Khan, A., & Bisai, D. (2014). Determining the degree of flood hazard risks in the Baliapal coastal block, Odisha, India: A quantitative approach. *Open Journal of Ocean and Coastal Sciences*, *1*(1), 1–11.

Barman, N. K., Paul, A. K., & Khan, A. (2014). *Physical Sensitivity and Social Exposure of Flood Hazard Risks in Subarnarekha Delta Plain, Odisha, India*.

Bhatt, C., Rao, G., Diwakar, P., & Dadhwal, V. (2017). Development of flood inundation extent libraries over a range of potential flood levels: A practical framework for quick flood response. *Geomatics, Natural Hazards and Risk*, *8*(2), 384–401.

Bhuyan, M. K., Mohanty, S., Jena, J., Kumar, P., & others. (2016). Hydrologic analysis for river Diversion scheme of Kanupur Dam Project, odisha–a case study. *Water and Energy International*, *58*(11), 52–57.

Chakrabortty, R., Pal, S. C., Janizadeh, S., Santosh, M., Roy, P., Chowdhuri, I., & Saha, A. (2021). Impact of climate change on future flood susceptibility: An evaluation based on deep learning algorithms and GCM model. *Water Resources Management*, *35*, 4251–4274.

Chatterjee, S., & Biswas, B. (2023). Cyclone Yaas: A Curse to Coastal People of Odisha and West Bengal (India). *National Academy Science Letters*, *46*(4), 321–324.

Chen, C., Jiang, J., Liao, Z., Zhou, Y., Wang, H., & Pei, Q. (2022). A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China. *Journal of Hydrology*, *607*, 127535.

Choudhari, K., Panigrahi, B., & Paul, J. C. (2014). Simulation of rainfall-runoff process using HEC-HMS model for Balijore Nala watershed, Odisha, India. *International Journal of Geomatics and Geosciences*, *5*(2), 253–265.

Choudhury, B., & Kar, A. K. (2021). Rain-Gauge Network Design and Rainfall Estimation—Case Study of Odisha Basins. *International Conference on Hydraulics, Water Resources and Coastal Engineering*, 537–546.

Chowdhury, J. R., & Parida, Y. (2023). Flood shocks and post-disaster recovery of households:

An empirical analysis from rural Odisha, India. *International Journal of Disaster Risk Reduction*, *97*, 104070.

Costache, R. (2019). Flash-flood Potential Index mapping using weights of evidence, decision Trees models and their novel hybrid integration. *Stochastic Environmental Research and Risk Assessment*, *33*(7), 1375-1402.

Dalei, N. N., Gupta, A., & Anand, N. (2020). Empirical nexus between global temperature, local weather and agriculture: Evidence from the Indian state of Odisha. *Energy, Environment and Globalization: Recent Trends, Opportunities and Challenges in India*, 143–156.

Damle, C., & Yalcin, A. (2007). Flood prediction using time series data mining. *Journal of Hydrology*, *333*(2–4), 305–316.

Dar, M. H., Chakravorty, R., Waza, S. A., Sharma, M., Zaidi, N. W., Singh, A. N., Singh, U. S., & Ismail, A. M. (2017). Transforming rice cultivation in flood prone coastal Odisha to ensure food and economic security. *Food Security*, *9*, 711–722.

Das, A., & Tripathy, B. (n.d.). *Assessing Climatic Change Impacts on Water Resource of Different Districts of Odisha*.

Das, A., & Tripathy, B. (2020). Long term assessment of rainfall over different districts of Odisha: A case study. *IOSR Journal of Mechanical and Civil Engineering*, *17*(2), 11–22.

Das, S., & Das, S. (2024). Cyclone Vulnerability Assessment of Coastal Odisha: A Sub-district Level Analysis. *IDRiM Journal*, *14*(1), 142–168.

Das, T., Shahfahad, & Rahman, A. (2024). Assessing tropical cyclone risk for improving mitigation strategies in Coastal Odisha, India. *Environmental Science and Pollution*

Research, 1–21.

Das, T., Talukdar, S., Shahfahad, Baig, M. R. I., Hang, H. T., Siddiqui, A. M., & Rahman, A. (2024). Assessing vulnerability to cyclones in coastal Odisha using fuzzy logic integrated AHP: towards effective risk management. *Spatial Information Research*, *32*(3), 277–295.

Directorate, F. F. M. (2012). *Flood Forecasting and Warning Network Performance Appraisal Report 2011*. New Delhi.

Dodangeh, E., Choubin, B., Eigdir, A. N., Nabipour, N., Panahi, M., Shamshirband, S., & Mosavi, A. (2020). Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Science of the Total Environment*, *705*, 135983.

Dtissibe, F. Y., Ari, A. A. A., Abboubakar, H., Njoya, A. N., Mohamadou, A., & Thiare, O. (2024). A comparative study of Machine Learning and Deep Learning methods for flood forecasting in the Far-North region, Cameroon. *Scientific African*, *23*, e02053.

Dube, A., Ashrit, R., Ashish, A., Sharma, K., Iyengar, G., Rajagopal, E., & Basu, S. (2014). Forecasting the heavy rainfall during Himalayan flooding—June 2013. *Weather and Climate Extremes*, *4*, 22–34.

Duncan, A., Keedwell, E., Djordjevic, S., & Savic, D. (2013). Machine learning-based early warning system for urban flood management.

El-Haddad, B. A., Youssef, A. M., Pourghasemi, H. R., Pradhan, B., El-Shater, A.-H., & El-Khashab, M. H. (2021). Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. *Natural Hazards*, *105*(1), 83–114. https://doi.org/10.1007/s11069-020-04296-y

El-Magd, S. A. A., Pradhan, B., & Alamri, A. (2021). Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita and surroundings, Central Eastern Desert, Egypt. *Arabian Journal of Geosciences*, *14*(4), 323. https://doi.org/10.1007/s12517-021-06466-z

Flood Prediction. (n.d.).

Furquim, G., Pessin, G., Faiçal, B. S., Mendiondo, E. M., & Ueyama, J. (2016). Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory: A case study involving a real wireless sensor network deployment in Brazil. *Neural Computing and Applications*, *27*(5), 1129–1141. https://doi.org/10.1007/s00521-015-1930-z

Felix, A. Y., & Sasipraba, T. (2019, December). Flood detection using gradient boost machine learning approach. In *2019 International conference on computational intelligence and knowledge economy (ICCIKE)* (pp. 779-783). IEEE.

Ganguli, P., Nandamuri, Y. R., & Chatterjee, C. (2023). Understanding flood regime changes of the Mahanadi River. *ISH Journal of Hydraulic Engineering*, *29*(3), 389–402.

Ganguly, K. K., Nahar, N., & Hossain, B. M. (2019). A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh. *International Journal of Disaster Risk Reduction*, *34*, 283–294.

Gharakhanlou, N. M., & Perez, L. (2023). Flood susceptible prediction through the use of geospatial variables and machine learning methods. *Journal of Hydrology*, *617*, 129121.

Ghorpade, P., Gadge, A., Lende, A., Chordiya, H., Gosavi, G., Mishra, A., Hooli, B., Ingle, Y. S., & Shaikh, N. (2021). Flood forecasting using machine learning: A review. *2021 8th International Conference on Smart Computing and Communications (ICSCC)*, 32–36.

https://ieeexplore.ieee.org/abstract/document/9528099/

Ghosh, A., & Dey, P. (2021). Flood Severity assessment of the coastal tract situated between Muriganga and Saptamukhi estuaries of Sundarban delta of India using Frequency Ratio (FR), Fuzzy Logic (FL), Logistic Regression (LR) and Random Forest (RF) models. *Regional Studies in Marine Science*, *42*, 101624.

Giri, M., Panigrahi, B., & Paul, J. (2019). Flood frequency analysis in middle reach of Mahanadi River Basin, Odisha. *Agricultural Engineering Today*, *43*(4), 6–17.

Giri, R., Panda, J., Rath, S. S., & Kumar, R. (2016). Validating quantitative precipitation forecast for the Flood Meteorological Office, Patna region during 2011–2014. *Journal of Earth System Science*, *125*, 709–723.

Goel, R. (2020). Flood damage analysis using machine learning techniques. *Procedia Computer Science*, *173*, 78–85.

Gude, V., Corns, S., & Long, S. (2020). Flood prediction and uncertainty estimation using deep learning. *Water*, *12*(3), 884.

Guhathakurta, P., Ratan, R., Chattopadhyay, R., Kulkarni, D., Bile, L. S., & Prasad, A. (2021). Prediction of monthly malaria outbreaks in districts of Odisha, India with meteorological parameters using statistical and artificial neural network techniques. *Climate and Health Journal*, *1*(3), 122–136.

Guru, N. (2016). *Flood Frequency Analysis of Partial Duration Series Using Soft Computing Techniques for Mahanadi River Basin in India* [PhD Thesis].

Guru, N., & Jha, R. (2015). Flood frequency analysis of Tel Basin of Mahanadi river system, India using annual maximum and POT flood data. *Aquatic Procedia*, *4*, 427–434.

Haldar, D., Nigam, R., Patnaik, C., Dutta, S., & Bhattacharya, B. (2016). Remote sensing-based assessment of impact of Phailin cyclone on rice in Odisha, India. *Paddy and Water Environment*, *14*, 451–461.

Halder, B., & Bandyopadhyay, J. (2022). Monitoring the tropical cyclone 'Yass' and 'Amphan'affected flood inundation using Sentinel-1/2 data and Google Earth Engine. *Modeling Earth Systems and Environment*, *8*(3), 4317–4332.

Han, D., Chan, L., & Zhu, N. (2007). Flood forecasting using support vector machines. *Journal of Hydroinformatics*, *9*(4), 267–276.

Hayder, I. M., Al-Amiedy, T. A., Ghaban, W., Saeed, F., Nasser, M., Al-Ali, G. A., & Younis, H. A. (2023). An intelligent early flood forecasting and prediction leveraging machine and deep learning algorithms with advanced alert system. *Processes*, *11*(2), 481.

Hazra, S., Ghosh, A., Ghosh, S., Pal, I., & Ghosh, T. (2022). Assessing coastal vulnerability and governance in Mahanadi Delta, Odisha, India. *Progress in Disaster Science*, *14*, 100223.

Hill, J. M., Singh, V. A., & Aminian, H. (1987). A COMPUTERIZED DATA BASE FOR FLOOD PREDICTION MODELING 1. *JAWRA Journal of the American Water Resources Association*, *23*(1), 21-27.

Hossain, F., & Anagnostou, E. N. (2004). Assessment of current passive-microwave-and infrared-based satellite rainfall remote sensing for flood prediction. *Journal of Geophysical Research: Atmospheres*, *109*(D7).

Hou, J., Zhou, N., Chen, G., Huang, M., & Bai, G. (2021). Rapid forecasting of urban flood inundation using multiple machine learning models. *Natural Hazards*, *108*(2), 2335–2356. https://doi.org/10.1007/s11069-021-04782-x

Hu, R., Fang, F., Pain, C. C., & Navon, I. M. (2019). Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. *Journal of Hydrology*, *575*, 911–920.

Ighile, E. H., Shirakawa, H., & Tanikawa, H. (2022). Application of GIS and machine learning to predict flood areas in Nigeria. *Sustainability*, *14*(9), 5039.

Islam, A. R. M. T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., ... & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, *12*(3), 101075.

Jain, S., Gautam, A., Chaudhary, A., Soni, C., & Sharma, C. (2022). Flood Mapping Using Sentinel-1 GRD SAR Images and Google Earth Engine: Case Study of Odisha State, India. In *Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2021* (pp. 455–464). Springer.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, B. T., & Lee, S. (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. *Sustainability*, *11*(19), 5426.

Jena, P. P. (2018). Climate change and its worst effect on coastal Odisha: An overview of its impact in Jagatsinghpur District. *IOSR Journal of Humanities and Social Science*, *23*(1), 1–15.

Jena, P. P., Chatterjee, C., Pradhan, G., & Mishra, A. (2014). Are recent frequent high floods in Mahanadi basin in eastern India due to increase in extreme rainfalls? *Journal of Hydrology*, *517*, 847–862.

Kabir, S., Patidar, S., & Pender, G. (2021). A machine learning approach for forecasting and visualising flood inundation information. *Proceedings of the Institution of Civil Engineers - Water Management*, *174*(1), 27–41. https://doi.org/10.1680/jwama.20.00002

Kan, G., Liang, K., Yu, H., Sun, B., Ding, L., Li, J., He, X., & Shen, C. (2020). Hybrid machine learning hydrological model for flood forecast purpose. *Open Geosciences*, *12*(1), 813–820. https://doi.org/10.1515/geo-2020-0166

Karyotis, C., Maniak, T., Doctor, F., Iqbal, R., Palade, V., & Tang, R. (2019). Deep learning for flood forecasting and monitoring in urban environments. *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 1392–1397. https://ieeexplore.ieee.org/abstract/document/8999292/

Kaur, G., & Bala, A. (2019). An efficient automated hybrid algorithm to predict floods in cloud environment. *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 1–4.

Kaur, S., Das, A. K., & Agarwala, A. (n.d.). *Performance of WRF (ARW) over River Basins in Odisha, India During Flood Season 2014*.

Ke, Q., Tian, X., Bricker, J., Tian, Z., Guan, G., Cai, H., Huang, X., Yang, H., & Liu, J. (2020). Urban pluvial flooding prediction by machine learning approaches–a case study of Shenzhen city, China. *Advances in Water Resources*, *145*, 103719.

Keum, H. J., Han, K. Y., & Kim, H. I. (2020). Real-Time Flood Disaster Prediction System by

Applying Machine Learning Technique. *KSCE Journal of Civil Engineering*, *24*(9), 2835–2848. https://doi.org/10.1007/s12205-020-1677-7

Kewat, N., & Sonekar, S. V. (2020). *A Survey on Flood Prediction through Wireless Sensor Network*.

Khalaf, M., Alaskar, H., Hussain, A. J., Baker, T., Maamar, Z., Buyya, R., Liatsis, P., Khan, W., Tawfik, H., & Al-Jumeily, D. (2020). IoT-enabled flood severity prediction via ensemble machine learning models. *IEEE Access*, *8*, 70375–70386.

Khalaf, M., Hussain, A. J., Al-Jumeily, D., Baker, T., Keight, R., Lisboa, P., Fergus, P., & Al Kafri, A. S. (2019). A data science methodology based on machine learning algorithms for flood severity prediction. *2018 IEEE Congress on Evolutionary Computation (CEC)*, 1–8. https://ieeexplore.ieee.org/abstract/document/8477904/

Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., & Ho, H. L. (2019). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology*, *573*, 311–323.

Kugler, Z., & De Groeve, T. (2007). The global flood detection system. *JRC scientific and technical reports*, 1-45.

Khristodas, P., Palanivelu, K., Ramachandran, A., Anushiya, J., Prusty, B., & Guganesh, S. (2022). Assessment of climate-induced sea-level rise scenarios and its inundation in coastal Odisha, India. *Applied Ecology and Environmental Research*, *20*(4), 3393–3409.

KHRISTODAS, P., PALANIVELU, K., RAMACHANDRAN, A., & PRUSTY, B. (2023). PREDICTION OF CLIMATE CHANGE-INDUCED SEA LEVEL RISE IN CHILIKA-

PURI COAST OF ODISHA, INDIA: WITH SPECIAL PROMINENCE ON ADAPTATION ACTION STRATEGY FRAMEWORK. *Applied Ecology & Environmental Research*, *21*(5).

Kim, H. I., & Han, K. Y. (2020). Linking hydraulic modeling with a machine learning approach for extreme flood prediction and response. *Atmosphere*, *11*(9), 987.

Kinage, C., Kalgutkar, A., Parab, A., Mandora, S., & Sahu, S. (2019). Performance evaluation of different machine learning based algorithms for flood prediction and model for real time flood prediction. *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, 1–7. https://ieeexplore.ieee.org/abstract/document/9128379/

Kumar, A., & Patra, K. C. (2021). Flood Hazard Assessment of Baitarani River Basin using One-Dimensional Hydrodynamic Model. In *Advanced Modelling and Innovations in Water Resources Engineering: Select Proceedings of AMIWRE 2021* (pp. 157–171). Springer.

Kumar, B., Baliarshingh, A., Jain, S., & Sahu, K. (2018). Analysis of rainfall probability for strategic crop planning in Puri district of Odisha. *Int. J. Chem. Stud*, *6*(4), 1–5.

Kumar, M. (2015). Remote sensing and GIS based sea level rise inundation assessment of Bhitarkanika forest and adjacent eco-fragile area, Odisha. *International Journal of Geomatics and Geosciences*, *5*(4), 674–686.

Kumar, M. K. (n.d.). *Early Warning System For Flood Prediction In The River Basins Of India*.

Kumar, P., Dasgupta, R., Dhyani, S., Kadaverugu, R., Johnson, B. A., Hashimoto, S., Sahu, N., Avtar, R., Saito, O., Chakraborty, S., & others. (2021). Scenario-based hydrological modeling for designing climate-resilient coastal water resource management measures:

Lessons from Brahmani River, Odisha, Eastern India. *Sustainability*, *13*(11), 6339.

Kumar, V., Azamathulla, H. M., Sharma, K. V., Mehta, D. J., & Maharaj, K. T. (2023). The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. *Sustainability*, *15*(13), 10543.

Kumar, V., Guganesh, S., Babu, D., & Kumaresan, P. (2024). Flood Risk Assessment for an Irrigation Project in Odissa, India. *Indian Journal of Science and Technology*, *17*(13), 1304–1314.

Kundu, S., Aggarwal, S., Kingma, N., Mondal, A., & Khare, D. (2015). Flood monitoring using microwave remote sensing in a part of Nuna river basin, Odisha, India. *Natural Hazards*, *76*, 123–138.

Kunverji, K., Shah, K., & Shah, N. (2021). A flood prediction system developed using various machine learning algorithms. *Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021)*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3866524

Lawal, Z. K., Yassin, H., & Zakari, R. Y. (2021). Flood prediction using machine learning models: A case study of Kebbi state Nigeria. *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1–6. https://ieeexplore.ieee.org/abstract/document/9718497/

Liong, S. Y., Lim, W. H., & Paudyal, G. N. (2000). River stage forecasting in Bangladesh: neural network approach. *Journal of computing in civil engineering*, *14*(1), 1-8.

Liong, S. Y., & Sivapragasam, C. (2002). Flood stage forecasting with support vector machines

1. *JAWRA Journal of the American Water Resources Association*, *38*(1), 173-186.

Li, X., Hu, Q., Wang, R., Zhang, D., & Zhang, Q. (2021). Influences of the timing of extreme precipitation on floods in Poyang Lake, China. *Hydrology Research*, *52*(1), 26-42.

Li, Z., Liu, H., Luo, C., & Fu, G. (2021). Assessing surface water flood risks in urban areas using machine learning. *Water*, *13*(24), 3520.

Liu, D., Cui, Y., Jin, W., Wang, H., & Tang, H. (2023). Channel aggradation triggered by dam failure amplifies the damage of outburst flood. *Landslides*, *20*(7), 1343-1362.

Luu, C., Bui, Q. D., Costache, R., Nguyen, L. T., Nguyen, T. T., Van Phong, T., Van Le, H., & Pham, B. T. (2021). Flood-prone area mapping using machine learning techniques: A case study of Quang Binh province, Vietnam. *Natural Hazards*, *108*(3), 3229–3251. https://doi.org/10.1007/s11069-021-04821-7

Mahapatra, A., Mahammood, V., & DurgaRao, K. (2021). Development of real-time flood forecast model for Vamsadhara river through hydrological approach. *Int. J. Adv. Technol. Eng. Explor.*, *81*(8), 1059–1079.

Mahapatra, A., Mahammood, V., & Venkatesh, K. (2022). Unsteady flow analysis using hydrological and hydraulic models for real-time flood forecasting in the Vamsadhara river basin. *Journal of Hydroinformatics*, *24*(6), 1207–1233.

Mahato, S., Mandal, G., Kundu, B., Kundu, S., Joshi, P., & Kumar, P. (2023). Comprehensive Drought Vulnerability Assessment in Northwestern Odisha: A Fuzzy Logic and Analytical Hierarchy Process Integration Approach. *Water*, *15*(18), 3210.

Malakar, K. D. (2020). Flood frequency analysis using Gumbel's method: A case study of Lower

Godavari River Division, India. *Journal of Scientific Computing*, *2*(9), 33–51.

Mane, P., Katti, M., Nidgunde, P., & Surve, A. (2020). Early flood detection and alarming system using machine learning techniques. *International Journal of Research in Engineering, Science and Management*, *3*(10), 29-32.

Manisha, K., & Chettry, V. (2022). Mapping Cyclone and Flood Hazard Vulnerability in Puri District, Odisha, India, Using Geoinformatics. *International Conference on Variability of the Sun and Sun-like Stars: From Asteroseismology to Space Weather*, 595–603.

Maspo, N.A., Harun, A. N. B., Goto, M., Cheros, F., Haron, N. A., & Nawi, M. N. M. (2020). Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review. *IOP Conference Series: Earth and Environmental Science*, *479*(1), 012038. https://iopscience.iop.org/article/10.1088/1755-1315/479/1/012038/meta

Mishra, S. P., & Jena, A. (2017). Management of Probabilistic Peak Flood: Mahanadi Branches draining into Chilika Lagoon, Odisha, India. *International Conference on" Global Civil Engineering Challenges in Sustainable Development and Climate Change"(ICGCSC-March 2017)*.

Mishra, S. P., & Jena, J. (2015). Analytical Study of Monsoon Rainfall South Mahanadi Delta and Chilika Lagoon, Odisha. *International Journal of Engineering and Technology (IJET)*, *7*, 985–996.

Mishra, S. P., & Ojha, A. C. (2020). Appraising the impact of Naraj barrage on sedimentation of Chilika Lagoon; the soft computing model for prediction. *Archives of Current Research International*, *20*(6), 31–41.

Mishra, S. P., Ojha, A. C., Mishra, S., & Sahu, D. K. (2022). *Cyclogenesis and Odisha Coast, the Hotbed*.

Misra, R. K., Panda, P., Sahu, A., Sahoo, S., & Behera, D. (2021). Rainfall prediction using machine learning approach: A case study for the state of odisha. *Indian Journal of Natural Sciences*.

Mitra, P., Ray, R., Chatterjee, R., Basu, R., Saha, P., Raha, S., Barman, R., Patra, S., Biswas, S. S., & Saha, S. (2016). Flood forecasting using Internet of things and artificial neural networks. *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 1–5. https://ieeexplore.ieee.org/abstract/document/7746363/

Mittal, V., Kumar, T. V., & Goel, A. (2023a). Ascertaining the impact of balancing the flood dataset on the performance of classification-based flood forecasting models for the river basins of Odisha. *International Journal of Global Warming*, *30*(3), 233–254.

Mittal, V., Kumar, T. V., & Goel, A. (2023b). Forecasting Floods in the River Basins of Odisha Using Machine Learning. *International Conference on IoT, Intelligent Computing and Security: Select Proceedings of IICS 2021*, 91–101.

Mohanty, B., Sarkar, R., & Saha, S. (2023). Preparing coastal erosion vulnerability index applying deep learning techniques in Odisha state of India. *International Journal of Disaster Risk Reduction*, *96*, 103986.

Mohanty, L. K., Panda, B., Samantaray, S., Dixit, A., & Bhange, S. (2024). Analyzing water level variability in Odisha: Insights from multi-year data and spatial analysis. *Discover Applied Sciences*, *6*(7), 363.

Mohanty, M. P., Sherly, M. A., Karmakar, S., & Ghosh, S. (2018). Regionalized design rainfall estimation: An appraisal of inundation mapping for flood management under data-scarce situations. *Water Resources Management*, *32*, 4725–4746.

Mohanty, S., Mustak, S., Singh, D., Van Hoang, T., Mondal, M., & Wang, C.-T. (2023). Vulnerability and risk assessment mapping of Bhitarkanika national park, Odisha, India using machine-based embedded decision support system. *Frontiers in Environmental Science*, *11*, 1176547.

Mohanty, S., Nadimpalli, R., Mohanty, U., & Pattanayak, S. (2024). Storm surge prediction improvement using high resolution meso-scale model products over the Bay of Bengal. *Natural Hazards*, *120*(2), 1185–1213.

Mohapatra, S., Tudu, K., Sahoo, A., Mohanty, S., & Marandi, C. (2023). MLFP: Machine Learning Approaches for Flood Prediction in Odisha State Check for updates. *Intelligent Systems: Proceedings of 3rd International Conference on Machine Learning, IoT and Big Data (ICMIB 2023)*, *728*, 195.

Mohite, A. R., Chatterjee, C., & Singh, R. (2020). Development of Flood Forecasting System for Middle Mahanadi Basin. *Roorkee Water Conclave*, 26–28.

Moon, H., Yoon, S., & Moon, Y. (2023). Urban flood forecasting using a hybrid modeling approach based on a deep learning technique. *Journal of Hydroinformatics*, *25*(2), 593–610.

Mosaffa, H., Sadeghi, M., Mallakpour, I., Jahromi, M. N., & Pourghasemi, H. R. (2022). Application of machine learning algorithms in hydrology. In *Computers in earth and environmental sciences* (pp. 585-591). Elsevier.

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536.

Motta, M., de Castro Neto, M., & Sarmento, P. (2021). A mixed approach for urban flood prediction using Machine Learning and GIS. *International Journal of Disaster Risk Reduction*, *56*, 102154.

Munawar, H. S., Hammad, A., Ullah, F., & Ali, T. H. (2019). After the flood: A novel application of image processing and machine learning for post-flood disaster management. *Proceedings of the 2nd International Conference on Sustainable Development in Civil Engineering (ICSDC 2019), Jamshoro, Pakistan*, 5–7. https://www.academia.edu/download/71649796/ICSDC2019_paper_229_1_.pdf

Munawar, H. S., Hammad, A. W., & Waller, S. T. (2021). A review on flood management technologies related to image processing and machine learning. *Automation in Construction*, *132*, 103916.

Munawar, H. S., Hammad, A. W., & Waller, S. T. (2022). Remote sensing methods for flood prediction: A review. *Sensors*, *22*(3), 960.

Nageswararao, M., Sinha, P., Mohanty, U., Panda, R., & Dash, G. (2019). Evaluation of district-level rainfall characteristics over Odisha using high-resolution gridded dataset (1901–2013). *SN Applied Sciences*, *1*, 1–24.

Nath, T. K., Tripathy, B., & Das, A. (2018). Climatic change on different Districts of Odisha. *International Journal of Engineering Research & Technology*, *7*(07), 192–204.

Nayak, M., Das, S., & Senapati, M. R. (2022). Improving Flood Prediction with Deep Learning Methods. *Journal of The Institution of Engineers (India): Series B*, *103*(4), 1189–1205.

https://doi.org/10.1007/s40031-022-00720-y

Nayak, P., Baliarsingh, A., Swain, C. K., & Rath, B. (2022). *Validation of extended range forecasts of rainfall and temperature in Cuttack district of Odisha*.

Nayak, P., Wagh, P., Venkatesh, B., Thomas, T., & Srivastav, R. (2024). Statistical Downscaling of Precipitation for Mahanadi Basin in India—Prediction of Future Streamflows. In *Modern River Science for Watershed Management: GIS and Hydrogeological Application* (pp. 281–307). Springer.

Nevo, S., Morin, E., Rosenthal, A. G., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Levi, N. P., … Matias, Y. (2021). *Flood forecasting with machine learning models in an operational framework* (arXiv:2111.02780). arXiv. http://arxiv.org/abs/2111.02780

Nguyen, D. T., & Chen, S.-T. (2020). Real-time probabilistic flood forecasting using multiple machine learning methods. *Water*, *12*(3), 787.

Noymanee, J., Nikitin, N. O., & Kalyuzhnaya, A. V. (2017). Urban pluvial flood forecasting using open data with machine learning techniques in pattani basin. *Procedia Computer Science*, *119*, 288–297.

Noymanee, J., & Theeramunkong, T. (2019). Flood forecasting with machine learning technique on hydrological modeling. *Procedia Computer Science*, *156*, 377–386.

Padhan, N., & Madheswaran, S. (2023). An integrated assessment of vulnerability to floods in coastal Odisha: A district-level analysis. *Natural Hazards*, *115*(3), 2351–2382.

Pal, S. C., Chowdhuri, I., Das, B., Chakrabortty, R., Roy, P., Saha, A., & Shit, M. (2022). Threats of climate change and land use patterns enhance the susceptibility of future floods in India. *Journal of Environmental Management*, *305*, 114317.

Park, S. J., & Lee, D. K. (2020). Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms. *Environmental Research Letters*, *15*(9), 094052.

Panda, G. (2015). Assessment of Human Vulnerability and Risk of Flood Hazards in Orissa, India. *Spatial Diversity and Dynamics in Resources and Urban Development: Volume 1: Regional Resources*, 325–340.

Panda, R., Mohanty, U., Dash, S., & Parhi, C. (2023). Flash drought in Odisha-prediction, impact assessment, coping strategies: Current status and future strategies. *Journal of Agrometeorology*, *25*(4), 491–497.

Panda, S. P., & Sethy, K. M. (2016). Human perception and response to flood hazards: A case study of Baitarani basin in Odisha. *International Journal of Applied Research*, *2*(5), 511–516.

Pandey, R. P., Desai, M., & Panwar, R. (2023). Hybrid deep learning model for flood frequency assessment and flood forecasting. *Multidisciplinary Science Journal*, *5*.

Panigrahi, B., Giri, M., & Paul, J. (2018). Development of Stage-Discharge Relationship for Middle Reach of Mahanadi River Basin, Odisha. *Agricultural Engineering Today*, *42*(2), 45–51.

Panigrahi, B., Paramjita, D., Giri, M., & Paul, J. (2020). Frequency analysis for prediction of maximum flood discharge in Mahanadi River Basin. *Int. J. Curr. Microbiol. App. Sci*, *9*(8), 3626–3639.

Parida, B. R., Tripathi, G., Pandey, A. C., & Kumar, A. (2022). Estimating floodwater depth using SAR-derived flood inundation maps and geomorphic model in kosi river basin (India). *Geocarto International*, *37*(15), 4336–4360.

Parihar, R. S., Bal, P. K., Saini, A., Mishra, S. K., & Thapliyal, A. (2022). Potential future malaria transmission in Odisha due to climate change. *Scientific Reports*, *12*(1), 9048.

Patel, P., & Sarkar, A. (2023). Streamflow Estimation Using Entropy-Based Flow Routing Technique in Brahmani River, Odisha. In *River, Sediment and Hydrological Extremes: Causes, Impacts and Management* (pp. 167–182). Springer.

Patel, S. K. (2018). *Community-level assessment of floods and cyclones in coastal Odisha, India: Impact, resilience, and implications*.

Patra, J. P., Kumar, R., & Mani, P. (2019). Flood hazard assessment for a dam failure. *International Journal of Advance and Innovative Research*, *6*(2), 34–39.

Pattanaik, D., & Das, A. K. (2015). Prospect of application of extended range forecast in water resource management: A case study over the Mahanadi River basin. *Natural Hazards*, *77*, 575–595.

Paul, A., & Das, P. (2014). Flood prediction model using artificial neural network. *International Journal of Computer Applications Technology and Research*, *3*(7), 473–478.

Piadeh, F., Behzadian, K., Chen, A. S., Campos, L. C., Rizzuto, J. P., & Kapelan, Z. (2023).

Event-based decision support algorithm for real-time flood forecasting in urban drainage systems using machine learning modelling. *Environmental Modelling & Software*, *167*, 105772.

Pradhan, P. P., Pradhan, B., & Khatua, K. K. (2022). *Flood Risk Assessment of Subarnarekha River Using Adaptive Neural-Based Fuzzy Inference System (ANFIS)*.

Pradhan, S., Sahoo, J. M., & Bhusan, B. (2024). Application of machine learning models in Mahanadi River basin, Odisha for prediction of flood. In *Industry 4.0 with Modern Technology* (pp. 108–112). CRC Press.

Prasad, B., Shreya, S., Sinha, T., & Priya, S. (2021). Ensemble Model-Based Prediction for River Management: A Case Study on River Kaveri and Coastal Karnataka. *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 1–6.

Pravin, A., Jacob, T. P., & Rajakumar, R. (2021, July). Enhanced flood detection system using IoT. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 507-510). IEEE.

Priyadarshinee, I., Sahoo, K., & Mallick, S. (2015). A model for flood prediction and prevention using wireless sensor network. *Int. J. Comput. Appl*, *1*, 22–30.

Puttinaovarat, S., & Horkaew, P. (2020). Flood forecasting system based on integrated big and crowdsource data by using machine learning techniques. *IEEE Access*, *8*, 5885-5905.

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

Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. *Earth Systems and Environment*, *3*(3), 585–601. https://doi.org/10.1007/s41748-019-00123-y

Rahman, T., Syeed, M. M. A., Farzana, M., Namir, I., Ishrar, I., Nushra, M. H., & Khan, B. M. (2023). Flood prediction using ensemble machine learning model. *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–6. https://ieeexplore.ieee.org/abstract/document/10156673/

Rajni, R., & Patra, S. (2022). Time series analysis of rainfall for the state of Odisha. *AIP Conference Proceedings*, *2435*(1).

Rakshita, C., Priya, M. G., & Krishnaveni, D. (2023). Flood Mapping and Damage Assessment of Odisha During Fani Cyclone Using HSR Data. In *Futuristic Communication and Network Technologies: Select Proceedings of VICFCNT 2021, Volume 1* (pp. 401–410). Springer.

Raman, R. V., Jeyakanthan, V., & Rao, Y. S. (2023). 2D FLOOD SIMULATION OF NAGAVALI RIVER BASIN USING HMS AND RAS. *Journal of Water Resources Development (JWRD)*, *1*(01).

Rani, D. S., Jayalakshmi, G. N., & Baligar, V. P. (2020). Low cost IoT based flood monitoring system using machine learning and neural networks: Flood alerting and rainfall prediction. *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 261–267. https://ieeexplore.ieee.org/abstract/document/9074928/

Ranjan, R. (2017). Flood disaster management. *River System Analysis and Management*, 371–

417.

Rao, M. U. M., Patra, K. C., & Sasmal, S. K. (2022). *Monthly Average Rainfall Forecasting Based On An Adaptive Neuro-Fuzzy Inference System in Upper Brahmani Basin, Odisha, India*.

Rays, K., Mohapatra, M., Chakravarthy, K., Ray, S., Singh, S., Das, A., Kannan, B., & Bandyopadhyay, B. (2017). Hydro-meteorological aspects of tropical cyclone phailin in bay of bengal in 2013 and the assessment of rice inundation due to flooding. *Tropical Cyclone Activity over the North Indian Ocean*, 29–43.

Razali, N., Ismail, S., & Mustapha, A. (2020). Machine learning approach for flood risks prediction. *IAES International Journal of Artificial Intelligence*, *9*(1), 73.

Saha, S., Kundu, B., Paul, G. C., Mukherjee, K., Pradhan, B., Dikshit, A., Abdul Maulud, K. N., & Alamri, A. M. (2021). Spatial assessment of drought vulnerability using fuzzy-analytical hierarchical process: A case study at the Indian state of Odisha. *Geomatics, Natural Hazards and Risk*, *12*(1), 123–153.

Saha, T. K., Pal, S., Talukdar, S., Debanshi, S., Khatun, R., Singha, P., & Mandal, I. (2021). How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region. *Journal of Environmental Management*, *297*, 113344.

Sahoo, A., Samantaray, S., & Ghose, D. K. (2021). Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: A Case Study. *Journal of the Geological Society of India*, *97*(2), 186–198. https://doi.org/10.1007/s12594-021-1650-1

Sahoo, B., & Bhaskaran, P. K. (2019). Prediction of storm surge and coastal inundation using

Artificial Neural Network–A case study for 1999 Odisha Super Cyclone. *Weather and Climate Extremes*, *23*, 100196.

Sahoo, B., Nanda, T., & Chatterjee, C. (2022). Flood Forecasting Using Simple and Ensemble Artificial Neural Networks. *Geospatial Technologies for Land and Water Resources Management*, 429–456.

Sam, A. S., Kumar, R., Kächele, H., & Müller, K. (2017). Vulnerabilities to flood hazards among rural households in India. *Natural Hazards*, *88*, 1133–1153.

Samantaray, D., Chatterjee, C., Singh, R., Gupta, P. K., & Panigrahy, S. (2015). Flood risk modeling for optimal rice planning for delta region of Mahanadi river basin in India. *Natural Hazards*, *76*, 347–372.

Samantaray, S., Das, S. S., Sahoo, A., & Satapathy, D. P. (2022). Monthly runoff prediction at Baitarani river basin by support vector machine based on Salp swarm algorithm. *Ain Shams Engineering Journal*, *13*(5), 101732.

Samantaray, S., Tripathy, O., Sahoo, A., & Ghose, D. K. (2020). Rainfall forecasting through ANN and SVM in Bolangir Watershed, India. *Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 1*, 767–774.

Sampurno, J., Vallaeys, V., Ardianto, R., & Hanert, E. (2022). Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta. *Nonlinear Processes in Geophysics*, *29*(3), 301–315.

Sankalp, S., & Panda, P. K. (2023). A comparative evaluation of machine learning and ARIMA models for forecasting relative humidity over Odisha districts. *Developments in*

Environmental Science, *14*, 91–105.

Sankalp, S., Rao, U. M., Patra, K. C., & Sahoo, S. N. (2023). Modeling gated recurrent unit (GRU) neural network in forecasting surface soil wetness for drought districts of Odisha. *Developments in Environmental Science*, *14*, 217–229.

Sankaranarayanan, S., Prabhakar, M., Satish, S., Jain, P., Ramprasad, A., & Krishnan, A. (2020). Flood prediction based on weather parameters using deep learning. *Journal of Water and Climate Change*, *11*(4), 1766–1783.

Santos, C. A. G., Neto, R. M. B., do Nascimento, T. V. M., da Silva, R. M., Mishra, M., & Frade, T. G. (2021). Geospatial drought severity analysis based on PERSIANN-CDR-estimated rainfall data for Odisha state in India (1983–2018). *Science of the Total Environment*, *750*, 141258.

Sen, S., Nayak, N. C., Mohanty, W. K., & Keshri, C. K. (2023). Vulnerability and risk perceptions of hydrometeorological disasters: A study of a coastal district of Odisha, India. *GeoJournal*, *88*(1), 711–731.

Sethi, R. R., Dandapat, A. K., Jena, S., SANKALP, S., Panda, D., & Patra, P. (2022). *Flood Susceptibility Mapping in the Baitarani River basin, implementing cloud-based IS-AHP Technique*.

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

Sharma, O., Trivedi, D., Pattnaik, S., Hazra, V., & Puhan, N. B. (2023). Improvement in district scale heavy rainfall prediction over complex Terrain of North East India using deep learning. *IEEE Transactions on Geoscience and Remote Sensing*.

Sherpa, S. F., Shirzaei, M., Ojha, C., Werth, S., & Hostache, R. (2020). Probabilistic mapping of August 2018 flood of Kerala, India, using space-borne synthetic aperture radar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 896–913.

Sinam, R. (2019). At Site Flood Frequency Analysis of Baitarani River at Champua Watershed, Odisha. *International Journal of Scientific Research in Science and Technology*, *6*(6), 54.

Subeesh, A., Kumar, P., & Chauhan, N. (2019). Flood early detection system using internet of things and artificial neural networks. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 1* (pp. 297-305). Springer Singapore.

Subudhi, C., Suryavanshi, S., Jena, N., & others. (2019). Rainfall probability analysis for crop planning in Bargarh district of Odisha, India. *Biomettrics and Biostatistics International Jounal*, *8*(5), 178–182.

Surwase, T., Manjusree, P., Prakash, S., & Kuntla, S. (2020). Development of algorithms for evaluating performance of flood simulation models with satellite-derived flood. *H2oj*, *3*(1), 222–235.

Swain, J. B. (2013). *Hydrological modeling of a typical ungauged basin Of Odisha* [PhD Thesis].

Swain, J., Jha, R., & Patra, K. (2015). Stream flow prediction in a typical ungauged catchment using GIUH approach. *Aquatic Procedia, 4,* 993–1000.

Swain, M., Pattanayak, S., Mohanty, U., & Sahu, S. (2020). Prediction of extreme rainfall associated with monsoon depressions over Odisha: An assessment of coastal zone vulnerability at district level. *Natural Hazards*, *102*, 607–632.

Swain, M., Sinha, P., Pattanayak, S., Guhathakurta, P., & Mohanty, U. (2020). Characteristics of observed rainfall over Odisha: An extreme vulnerable zone in the east coast of India. *Theoretical and Applied Climatology*, *139*, 517–531.

Swain, S. (2014). Impact of climate variability over Mahanadi river basin. *International Journal of Engineering Research and Technology*, *3*(7), 938–943.

Swain, S., Mishra, S., Pandey, A., & Dayal, D. (2021). Identification of meteorological extreme years over central division of Odisha using an index-based approach. *Hydrological Extremes: River Hydraulics and Irrigation Water Management*, 161–174.

Swain, S. S. S. K. K. (n.d.). *PREDICTION OF FLOOD, SEDIMENT & WATER QUALITY FOR HIRAKUD RESERVOIR ON MAHANADI RIVER* [PhD Thesis]. Department of Civil Engineering, Veer Surendra Sai University of Technology ….

Syeed, M. M. A., Farzana, M., Namir, I., Ishrar, I., Nushra, M. H., & Rahman, T. (2022). Flood prediction using machine learning models. *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–6. https://ieeexplore.ieee.org/abstract/document/9800023/

Talukdar, G., Swain, J. B., & Patra, K. C. (2021). Flood inundation mapping and hazard assessment of Baitarani River basin using hydrologic and hydraulic model. *Natural Hazards*, *109*(1), 389–403.

Tang, Y., Sun, Y., Han, Z., Wu, Q., Tan, B., & Hu, C. (2023). Flood forecasting based on machine learning pattern recognition and dynamic migration of parameters. *Journal of Hydrology: Regional Studies*, *47*, 101406.

Taromideh, F., Fazloula, R., Choubin, B., Emadi, A., & Berndtsson, R. (2022). Urban flood-risk assessment: integration of decision-making and machine learning. *Sustainability*, *14*(8), 4483.

Tayfur, G., Singh, V. P., Moramarco, T., & Barbetta, S. (2018). Flood hydrograph prediction using machine learning methods. *Water*, *10*(8), 968.

Tehrany, M. S., Jones, S., & Shabani, F. (2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena*, *175*, 174–192.

Thakur, D. A., Mohanty, M. P., Mishra, A., & Karmakar, S. (2024). Quantifying flood risks during monsoon and post-monsoon seasons: An integrated framework for resource-constrained coastal regions. *Journal of Hydrology*, *630*, 130683.

Tripathy, S. S., Bhatia, U., Mohanty, M., Karmakar, S., & Ghosh, S. (2021). Flood evacuation during pandemic: A multi-objective framework to handle compound hazard. *Environmental Research Letters*, *16*(3), 034034.

Trivedi, D., Sharma, O., & Pattnaik, S. (2024). Spatio-attention-based network to improve heavy rainfall prediction over the complex terrain of Assam. *Neural Computing and Applications,* 1–17.

Trivedi, D., Sharma, O., Pattnaik, S., Hazra, V., & Puhan, N. B. (2024). Improving rainfall forecast at the district scale over the eastern Indian region using deep neural network. *Theoretical and Applied Climatology*, *155*(1), 761–777.

Uthaman, N. (2024). Drought Analysis and Forecasting in Odisha using Machine Learning Techniques. *Communications on Applied Nonlinear Analysis*, *31*(3s), 29–43.

Vahida, S., & Sahoo, S. N. (2022). *Urban Flood Estimation in Bhubaneswar city using Storm Water Management Model (SWMM)*.

Venkata Rao, G., Nagireddy, N. R., Keesara, V. R., Sridhar, V., Srinivasan, R., Umamahesh, N., & Pratap, D. (2024). Real-time flood forecasting using an integrated hydrologic and hydraulic model for the Vamsadhara and Nagavali basins, Eastern India. *Natural Hazards*, 1–29.

Wu, H., Adler, R. F., Hong, Y., Tian, Y., & Policelli, F. (2012). Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *Journal of Hydrometeorology*, *13*(4), 1268-1284.

Wu, W., Emerton, R., Duan, Q., Wood, A. W., Wetterhall, F., & Robertson, D. E. (2020). Ensemble flood forecasting: Current status and future opportunities. Wiley Interdisciplinary Reviews: Water, 7(3), e1432.

Yeditha, P. K., Kasi, V., Rathinasamy, M., & Agarwal, A. (2020). Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*(6).

Yoshimura, K., Kanamitsu, M., Noone, D., & Oki, T. (2008). Historical isotope simulation using reanalysis atmospheric data. *Journal of Geophysical Research: Atmospheres*, *113*(D19).

Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., & Behl, M. (2020). Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal

Community. *Water Resources Research*, *56*(10), e2019WR027038. https://doi.org/10.1029/2019WR027038

Zeng, C., & Bertsimas, D. (2023). *Global Flood Prediction: A Multimodal Machine Learning Approach* (arXiv:2301.12548). arXiv. http://arxiv.org/abs/2301.12548

Zhang, X., Mohanty, S. N., Parida, A. K., Pani, S. K., Dong, B., & Cheng, X. (2020). Annual and non-monsoon rainfall prediction modelling using SVR-MLP: an empirical study from Odisha. *IEEE Access*, *8*, 30223–30233.

Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., & Xu, C.-Y. (2022). Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *Journal of Hydrology*, *604*, 127255.

Appendix 1:

Datasets under research were:

1) All Odisha Monthly, Seasonal and Annual Temperature Series (Maximum, Minimum and Mean Temperatures)
2) Rainfall data – All Odisha and five homogeneous regions
3) Demographic data
4) Air Quality from vehicles
5) Flood sensor data
6) Weather data
7) Emissions and pollutants data
8) Crowd density
9) Geomorphology map data
10) Groundwater Prospects map data
11) Water Spread Area map data
12) Forest Fire Alerts data
13) Heatwave Data
14) Landslide Early Warning System Data
15) Flood Hazard zonation map Data
16) Flood Layers Data
17) Reservoir Level & Storage Data
18) Surface water Quality Data
19) Groundwater Levels Data

20) Groundwater Quality Data

21) Primary census abstract Data

22) India: Drainage Data

23) Potential Evapotranspiration Data

24) Rainfall (from HE and IMSRA methods) Data

25) Ocean surface wind vector data

26) Ocean surface currents Data

27) Ocean eddies Data