**Building a Predictive Model for analyzing flood risks in India using machine learning algorithms and historical trend patterns**

Riya Mandowaraa, Divya Gautam, Varad Trivedic

a [riyamandowara4@gmail.com](mailto:riyamandowara4@gmail.com)

b [divyagautam06@gmail.com](mailto:divyagautam06@gmail.com)

c [trivedi.varad20@gmail.com](mailto:trivedi.varad20@gmail.com)

ARTICLE INFO

Keywords:

Flood prediction

India

Machine learning

Climate change

Hydrological modelling

Disaster management

Authorship contribution statement

Riya Mandowara: Data analysis, implementation of Random Forest, XGBoost, SVM, and ANN algorithms, result analysis and manuscript drafting.

Varad Trivedi: Conceptualization, data preprocessing, model development and GitHub repository management.

Divya Gautam: Supervision, guidance on methodology and model selection, critical review and editing of the manuscript, interpretation of results, and final approval of the version to be submitted.

#### ABSTRACT

Due to a combination of climate change, urbanization, and deforestation, floods in India have become more frequent and severe (Abijith et al., 2025). This study offers an in-depth examination of historical flooding events in Indian states from 1970 to 2023 and investigates the use of sophisticated flood forecasting methods that employ machine learning models and hydrological simulations. Utilizing meteorological data, river discharge rates, and topographical information, the study aims to improve the accuracy of flood prediction systems while investigating regional trends and identifying the areas most impacted (Oruganti et al.,2025). This study uses machine learning models like Random Forest, Decision Trees, Support Vector Machine, XGBoost prediction, and Artificial Neural Networks (ANN) to evaluate the accuracy of flood predictions (Muthukrishnan et al., 2017). It utilizes a detailed dataset on the occurred loss and damage (L&D) (Bahinipati and Patnaik, 2020). The results are correlated with human development indicators and economic factors to evaluate whether development strategies have contributed to flood resilience. Three key findings emerge from the analysis: First, flood occurrences and damages have shown a significant upward trend across states, signaling increased flood vulnerability. Second, human development and income levels are statistically insignificant in reducing flood-related L&D, indicating that current development efforts have not effectively enhanced flood resilience. Third, while there is limited evidence of a “learning effect” from past flood events, machine learning-based disaster risk management programs demonstrate strong potential in mitigating future risks. These findings underscore the need to integrate climate risk assessments into development planning, enabling policymakers and disaster management agencies to better anticipate and mitigate the growing impacts of floods in India.

1. Introduction

Natural disasters, particularly floods, represent a significant challenge to human societies, especially in developing and transitioning economies. In recent decades, the increasing frequency of floods, coupled with growing population exposure, has led to substantial loss of lives, livelihoods, and infrastructure. Floods have become a significant barrier to development in emerging economies, interrupting growth paths and escalating pre-existing vulnerabilities. The impacts of floods are especially disastrous when they affect multiple sub basins within a river basin simultaneously, as widespread riverine floods cause more extensive damage than localized flooding. In addition to the direct and indirect losses, widespread floods also strain disaster management systems and complicate rehabilitation efforts (N. J. S. and Mishra, 2024).

This section includes an example of table (Table 1)

Table 1. Loss and Damages from Floods Worldwide (1900-2023)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Total Deaths | Total Injuries | The total number of people affected | The number of people made homeless | Total damage caused (‘000 US$’) |
| 8,80,776 | 35,33,872 | 3,81,74,98,377 | 3,28,63,839 | $1,39,17,20,747.92 |

Source: Emergency Events Database (EM-DAT) of the Centre for Research on the Epidemiology of Disaster (CRED). Available online: <https://www.emdat.be/>

Loss and damages (L&Ds) from floods represent the enduring impacts that cannot be entirely mitigated through adaptation measures (Wijenayake et al., 2024). According to historical records, flooding is one of the most frequent and destructive natural disasters globally, affecting millions of people every year (Ivascu and Munteanu, 2024). Between 1900 and 2023, floods claimed the lives of 8,80,776 people, injured over 35,33,872 individuals, and displaced 3,28,63,839 people from their homes. The number of people affected by floods globally soared to 381.75 million, with total economic damages reaching approximately $139.17 billion (Das et al., 2022).

The consequences of flooding go beyond the immediate loss of life and property — disrupting livelihoods, destroying infrastructure, and pushing vulnerable communities into cycles of poverty. Agricultural losses threaten food security, while damaged healthcare and sanitation systems heighten the risk of any outbreak of diseases. Climate change has also been linked to an increase in the frequency and severity of flooding events, with rising sea levels and extreme weather patterns increasing future risks. Given these cascading effects, it is critical to strengthen disaster preparedness, promote sustainable development, and invest in resilient infrastructure to protect vulnerable populations worldwide (Mosavi et al., 2018).

Although flooding occurs globally, most of the population at risk of flooding resides in South and East Asia, with China and India accounting for over a third of the total global exposure. In India, over 65 million people are both deprived and highly vulnerable to floods. This represents 16.8 % of 390 million flood-prone individuals in India and is projected to rise because of climate change (Singh et al., 2025).

Good flood control starts with proper identification of the causes of flooding. Flooding can be caused by excessive rain, high tides, strong winds, and sea level rise, all of which are being intensified by human-induced global warming. All these events pose serious threats to coastal communities, which are a series of adverse impacts such as flooding, deaths, land loss, destruction of buildings and infrastructure, saltwater intrusion, shifts in ecosystems, and loss of biodiversity. Coastal zones are particularly vulnerable due to their dense populations, economic importance, and rapid population growth. The increasing potential for storm surges and coastal flooding puts these areas at heightened risk, especially in low-elevation regions where sea-level rise intensifies vulnerability (Antwi-Agyakwa et al., 2023).

To mitigate flood impacts, it is essential to gather accurate information before floods occur. Flood prediction models have proven vital in hazard analysis and in managing severe flood events by raising awareness and promoting the adoption of preventive strategies. Coastal communities, in particular, require reliable flood vulnerability maps and long-term flood risk assessments to prepare for regular flooding events. These tools are increasingly crucial in light of the dynamic and evolving nature of climate conditions (Antwi-Agyakwa et al., 2023). Understanding the patterns of such widespread floods and the projected changes in the spatial extent of extreme precipitation is critical for improving flood management in a warming climate (N. J. S. and Mishra, 2024).

Economic development is at the center of flood risk reduction. Richer countries and areas can afford to invest in disaster mitigation, including better infrastructure, early warning systems, and contingency planning. Governance, education, and health are also at the center of building the resilience of society. The relationship between the economic development and flood impact is not required to be linear. While some studies show a negative relationship between the income level and the impact of disasters, others show non-linear or non-significant relationships.

* 1. Human Fatalities in India's Top 10 Flood-Prone States: A Geographic and Socio-Economic Analysis

For the purposes of this study, the information used is historical information regarding human deaths caused by floods in India's ten worst flood-affected states. The information carefully documents the deaths caused by floods each year, focusing especially on states like Uttar Pradesh, Maharashtra, Bihar, and West Bengal, among others (Das et al., 2022).

The goal of analyzing this data is to understand the regional importance of flooding, to determine patterns of human vulnerability, and to analyze which areas require more efficient disaster management policies. Mapping human casualties enables us to highlight the seriousness of the situation and indicate the need for targeted interventions that consider geographical and socio-economic aspects.

This section includes an example of figure (Figure 1)

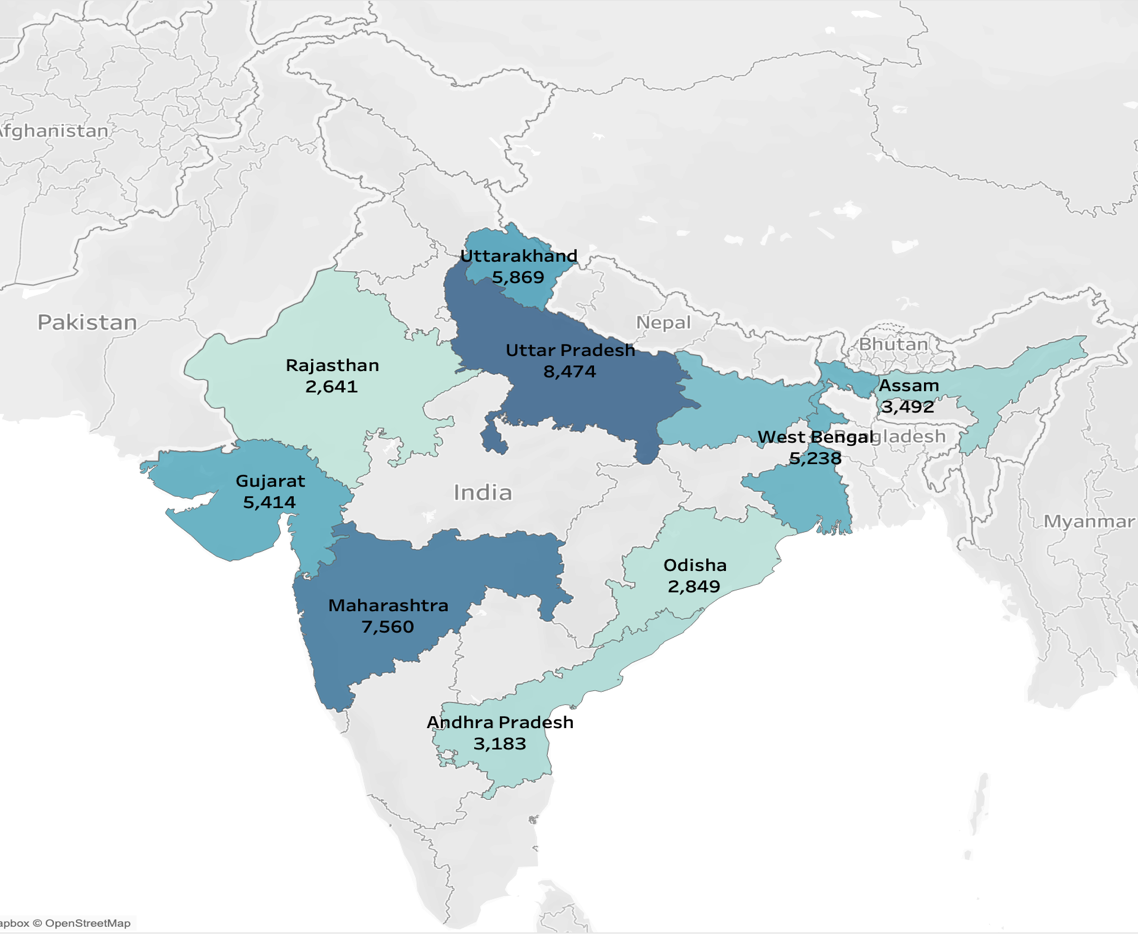


Figure 1: Human Fatalities in India's Top 10 Flood-Prone States in past years

The state of Uttar Pradesh has the highest number of fatalities, primarily due to the extensive Ganges River system, which is susceptible to flooding during the monsoon season, thereby impacting densely populated regions. Maharashtra and Uttarakhand also experience large death tolls; the former is afflicted by urban flooding, which is further compounded by inadequate infrastructure in urban regions such as Mumbai, while the latter is most susceptible to flash floods and landslides because of its hilly nature. Bihar and West Bengal, perennially affected by the Ganges and Brahmaputra rivers, experience widespread devastation in urban and rural regions, which adversely impacts agricultural pursuits and settlement. Gujarat's flood hazards are caused by cyclonic activity and monsoonal rains, while Assam is subjected to frequent flooding by the Brahmaputra River. Andhra Pradesh is threatened by riverine flooding as well as cyclone-driven storm surges, while Odisha is subjected to repeated flooding by coastal cyclones, and Rajasthan experiences intermittent flash floods despite its predominantly arid landscape. This variety of geography highlights the need for customized disaster management strategies to meet the specific geographical and climatic challenges of each state.

Floods have imposed a heavy and increasing burden on the people of India through a mix of geographical, environmental, and socio-economic reasons. India's diversified topography with large river systems such as the Ganges, Brahmaputra, and Godavari makes the country geographically prone to riverine flooding, most of which occurs during the monsoon rains every year. The monsoon, while being favorable for agriculture, results in excessive rain, which overflows rivers, leads to flash flooding, and triggers landslides, particularly along low-lying and coastal areas. Also, uncontrolled urbanization and unplanned development in vulnerable flood areas have increased the severity of the impact of floods, as the cities of Mumbai, Kolkata, and Chennai experience inadequate drainage capacity, leading to urban flooding. Climate change also adds to more uneven and intense precipitation, leading to more frequent and intense floods. Glacial melting and rising sea levels and the melting of glaciers in the Himalayas further enhance the vulnerability of coastal belts and river catchments. Anthropogenic processes in the form of deforestation, wetland encroachment, and the encroachment and destruction of natural floodplains have reduced the ability of land to hold excess water, exacerbating flooding (Mohanty et al., 2020).

The socio-economic context of India, with the majority of the population engaged in agriculture and living in rural belts, enhances the devastation brought about by floods. Livestock, agriculture, and means of livelihood are largely destroyed, and there is scarcity of food as well as economic loss. Low infrastructure standards, a lack of effective emergency planning, and weak disaster response in most of the areas enhance the risk vulnerability of the affected people. Densely populated belts where slums are present are highly vulnerable to displacement, waterborne diseases, and death in floods.

* 1. Total Human Fatalities Caused by Floods Annually in India

The "Total Human Fatalities Due to Floods per Year in India (1970–2022)"graph shows the yearly count of human deaths caused by floods, with the x-axis as the years from 1970 to 2022 and the y-axis as the number of fatalities. It depicts the fluctuation in the number of fatalities on an annual basis, with certain peaks indicating large flood occurrences where fatalities recorded a steep rise. The graph indicates incremental trends along with abrupt changes, providing information on the contribution of various parameters like extreme monsoon events, inadequate infrastructure, and climate change in defining fatality numbers over the decades. These parameters provide a rough estimate of human losses due to floods in India over decades.

This section includes an example of figure (Figure 2)

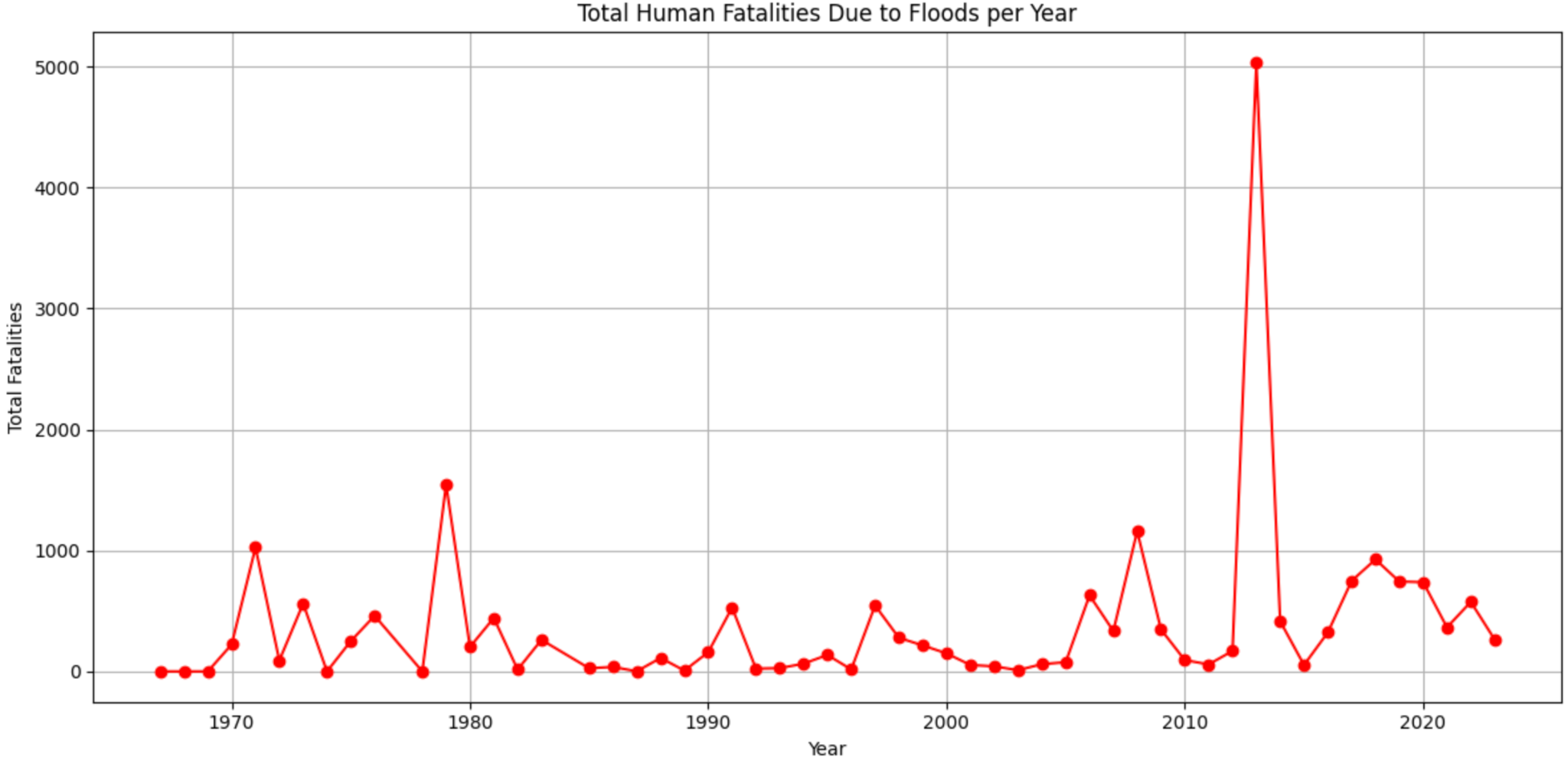


Figure 2: Total Human Fatalities Caused by Floods Annually in India

The pattern indicates periodic spikes in deaths, with a number of smaller peaks between the 1970s and the early 2000s, which are indicative of the recurring monsoonal floods. A very steep spike around 2010, with over 4000 deaths, is suggestive of a disaster flood event, perhaps fueled by unusual weather conditions or poor preparedness. India's susceptibility to floods is largely because of its exclusive climatic and geographical factors, such as monsoon dynamics, flooding of rivers, and poor urban drainage facilities. As climate change causes intensified rainfall and sea-level rise, the occurrence and severity of floods have grown. Urban growth, forest destruction, and the encroachment of floodplains also exacerbate the threat. Lack of good infrastructure, poor early warning systems, and delays in response during disasters also account for more deaths during such occurrences.

1. Dataset Overview and Attributes

This study hopes to utilize a large dataset and use machine learning methods to forecast flood risks in different Indian states (Samansiri, 2023). The database employed in the present study is comprehensive enough in its coverage of historical flooding instances in India over the time range from 1970 to 2023. It captures some attributes important to understanding and projecting flood hazard occurrence. The included attributes are:

Flood event data: Reports the date, duration, and place of occurrences of floods within states, including a time reference to the frequency and severity of floods (Jain, 2024).

Human impact: Information on human casualties, injury, and displacement. This serves to measure the social impact of floods, permitting more detailed flood-prone region analysis and assessment of at-risk populations.

Economic losses:Consists of information on loss to crops, residential buildings, and public infrastructure, in terms of money. It measures the economic destruction inflicted by floods, facilitating the determination of susceptible sectors.

Climatic data:Parameters like intensity of rainfall, river discharge, soil moisture content, and temperature. These are central to understanding the environmental flood triggers and are essential for the construction of predictive models.

Topographical data: Data on elevation, land use, and river basin characteristics. This aids in evaluating how geographical characteristics impact flood susceptibility in various areas.

Geographical spread:District-wise and state-wise information, especially for states that are subject to monsoon-related flooding like Bihar, Uttar Pradesh, Assam, and coastal states like Odisha and Andhra Pradesh.

Through the incorporation of these characteristics, the dataset offers an integrated perspective on flood risks throughout India, aiding in the improvement of the precision of machine learning models employed for prediction. Through the incorporation of socio-economic as well as environmental data, there is a better analysis of flood vulnerabilities and probable mitigation measures (Bui et al.,2019).

#### Literature Review

India requires better frameworks for flood prediction and risk mitigation as the country’s floods happen more often and become more severe with urbanization and climate change. Several investigations have examined current problems in handling floods, the part played by severe downpours and flood events over time.

In their published study of 2011, Guhathakurta, Sreejith and Menon assess the impact of climate change on extreme weather events in India through careful investigation of long-term rainfall rates. Increased reports of high-intensity rain show that central and peninsular regions are now more likely to experience flooding. The authors make clear that better weather forecasts and the addition of climate-resilient infrastructure help tackle these new risks (Guhathakurta et al., 2011).

Supporting this view, Singh and Kumar (2013) discuss flood data and analyze 25 years’ worth of flood incidents from 1978 to 2006. Results from their study demonstrate that floods have become more frequent, deadly and economically damaging, mostly in places with a high concentration of people and crops, like the Indo-Gangetic plains. They propose better zoning for flooding areas, analysis using past information and planning ahead for future infrastructure needs (Singh and Kumar, 2013).

Because of bad drainage and fast urban growth, flooding in cities has become an obvious problem. According to Singh, Nielsen and Greatrex (2023), outdated stormwater systems, mistakes in land management and lacking combined policy plans are the main reasons for urban pluvial floods. They propose nature solutions, improved early warning tools and more people taking part as key ways to fight disasters (Singh et al., 2023).

Mohaparta and Singh (2018) discuss the policies and structure of flood management and point out its weaknesses in India. They recognize that agencies pay more attention to dams and embankments than to zoning or early warnings and that they often fail to use advanced technologies like GIS and remote sensing efficiently. Their activities suggest moving from reactive actions to more efficient and combined management of flood risks (Mohapatra and Singh, 2003).

A similar point is made by Ray et al. (2019), who studied recent floods in India and explained that poor city planning, destruction of floodplains and weather-related problems all play a role. They recommend making improvements in how both national and local organizations work together and using live data for better community protection (Ray et al., 2019).

Overall, the studies point to how climate study, historical study, inadequate infrastructure and policy issues work together to impact how floods in India are understood. Using a range of methods, they lay the groundwork for models that consider climate, geography and social-economic factors when predicting and handling flood risks.

### Methodology

The estimation for future instances of floods includes an integrated process that involves data collection, flood trend, and the formulation of prediction models. This formalized process helps ensure the predictions are derived from solid historical data and pertinent environmental factors.

This section includes an example of figure (Figure 3)

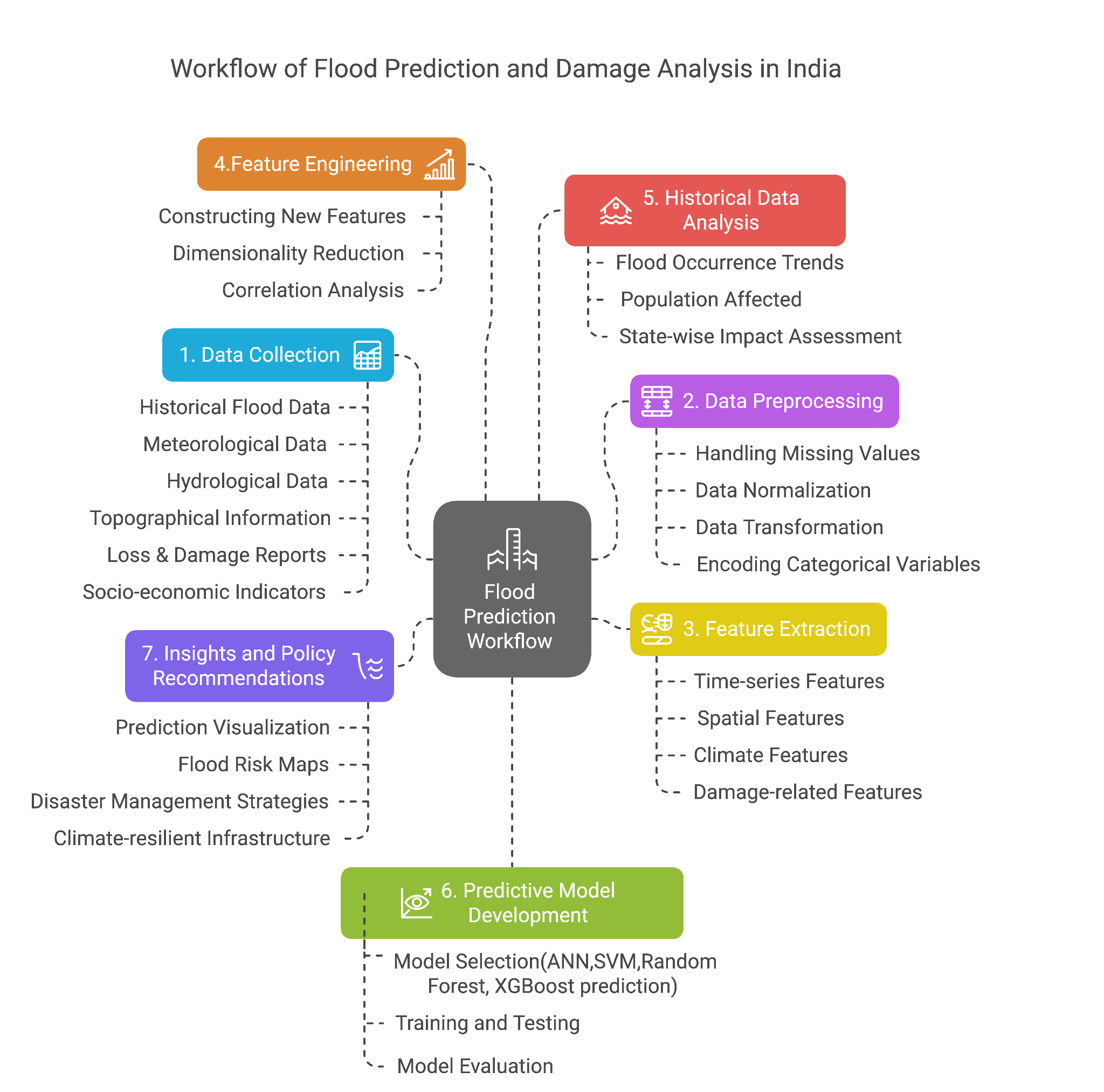


Figure 3: Workflow of flood prediction and damage analysis in India

The diagram depicting the workflow of flood prediction and damage analysis in India was created using Research Rabbit, a tool that helps visualize research connections and trends. After examining relevant literature and identifying crucial processes involved in flood forecasting and damage evaluation, the workflow was developed.

* 1. Data Pre-processing:

1. Cleaning the Data**:** This includes taking care of missing, inconsistent, or incorrect values in the dataset. For instance, missing values (NaN) might be imputed by replacing with the mean, median, or a suitable placeholder, according to the variable. One needs to take care to make data in an appropriate format for use in analysis, e.g., standardizing date formats, conversion of units, or numerical equivalence. Also, categorical variables such as state names are converted into numerical values (e.g., through one-hot encoding or label encoding) for machine learning models to learn from.
2. Feature Selection: Only the most suitable variables, or attributes, are selected for the model in this phase to refine the model's accuracy and minimize noise (Akbulut et al., 2023). Important attributes associated with floods could be Flood duration, which tells how long the flood lasted, Human influence, indicators like human death, injury or displacement and Climatic factors which includes amount of rainfall , rates of river discharge, etc. These characteristics are the input(X) to the machine learning model and the occurrence or intensity of floods constitutes the target variable(Y), which the model will predict. Collectively, it is these parameters that assist in development of predictive systems capable of making flood impact predictions with better accuracy.
3. Feature Engineering:  
   This means constructing new, helpful features from the given data. For example, in this we have computed the average rain over certain time spans (such as weeks or months) or constructed lag variables, which show whether there was a flood in previous years. These engineered features assist the model in learning about patterns and relationships between the data better, and this improves its predicting powers.

#### Data Collection

* + 1. Historical Flood Data

The analysis is based on the dataset of past 50 years of historical flood data. This data has been obtained from authentic government websites and research databases, such as the Central Water Commission (CWC) and the India Meteorological Department (IMD) (Wedajo et al., 2024). These institutions have a vast record of flood incidents, including date, duration, geographical location, intensity, and corresponding effects on animal and human existence. The addition of this complete dataset facilitates the determination of patterns and trends in flood occurrences through time.

This section includes an example of figure (Figure 4)

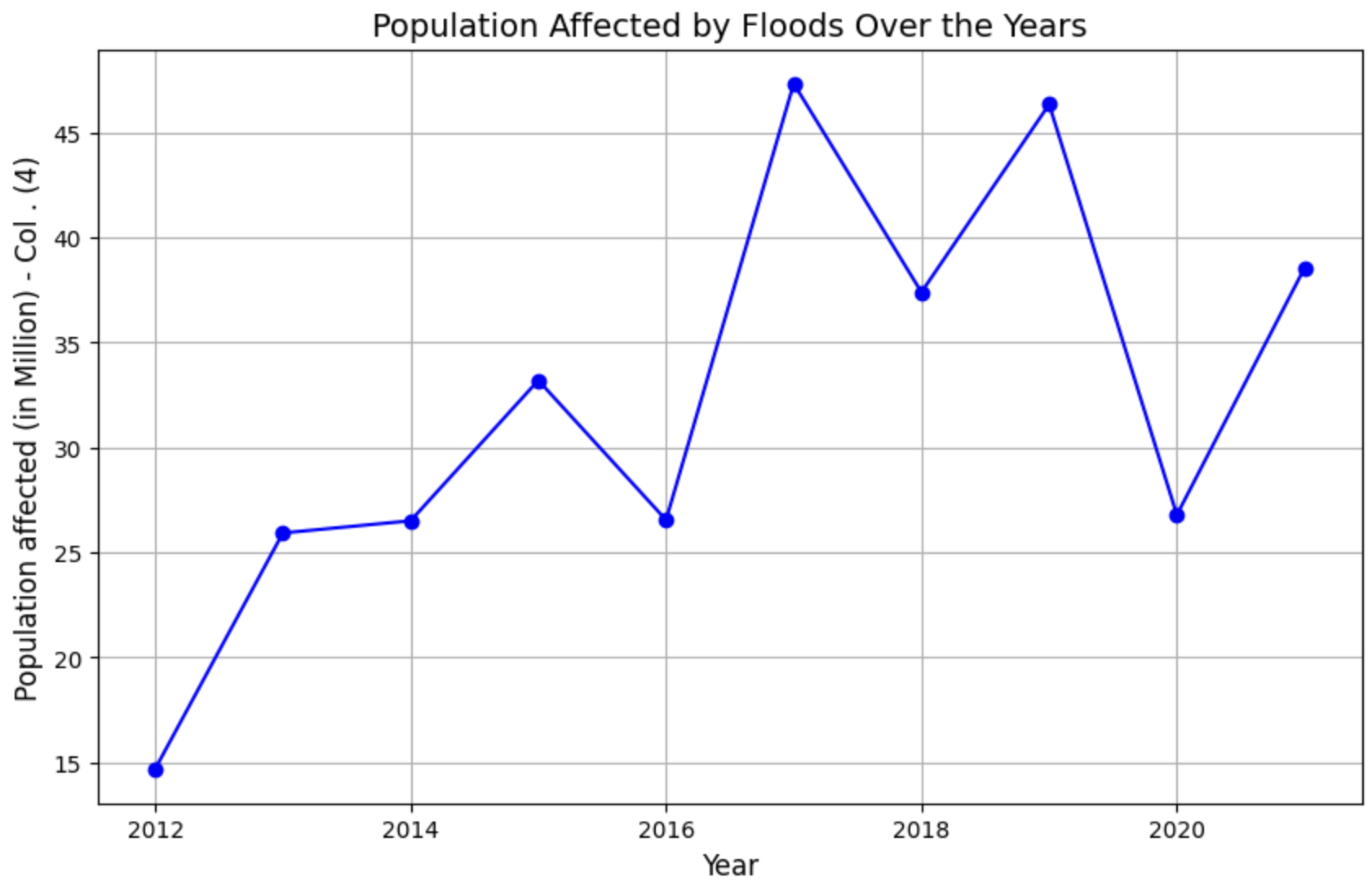


Figure 4: Population Affected by Floods Over the Years

The line chart shows the population affected by floods annually. It shows the direct human impact of flooding events over time. The peaks on the chart reflect major flood events, and troughs reflect years with fewer people affected. This visualization aids in understanding the long-term trends of flood-related population displacement and can be correlated with rainfall patterns, disaster management policy, and climate change factors. The rate is steadily growing since 2012 and observed a significant boost in 2017 when a population of affected people reached almost 48 million. Although this varies at the individual level, the general trend shows that large parts of the population are repetitively affected by floods, which can be linked to the recurrence of floods, as well as population density and urban growth in the exposed areas.

* + 1. Total damages over years including crops, houses and public utilities

This section includes an example of figure (Figure 5 and 6)

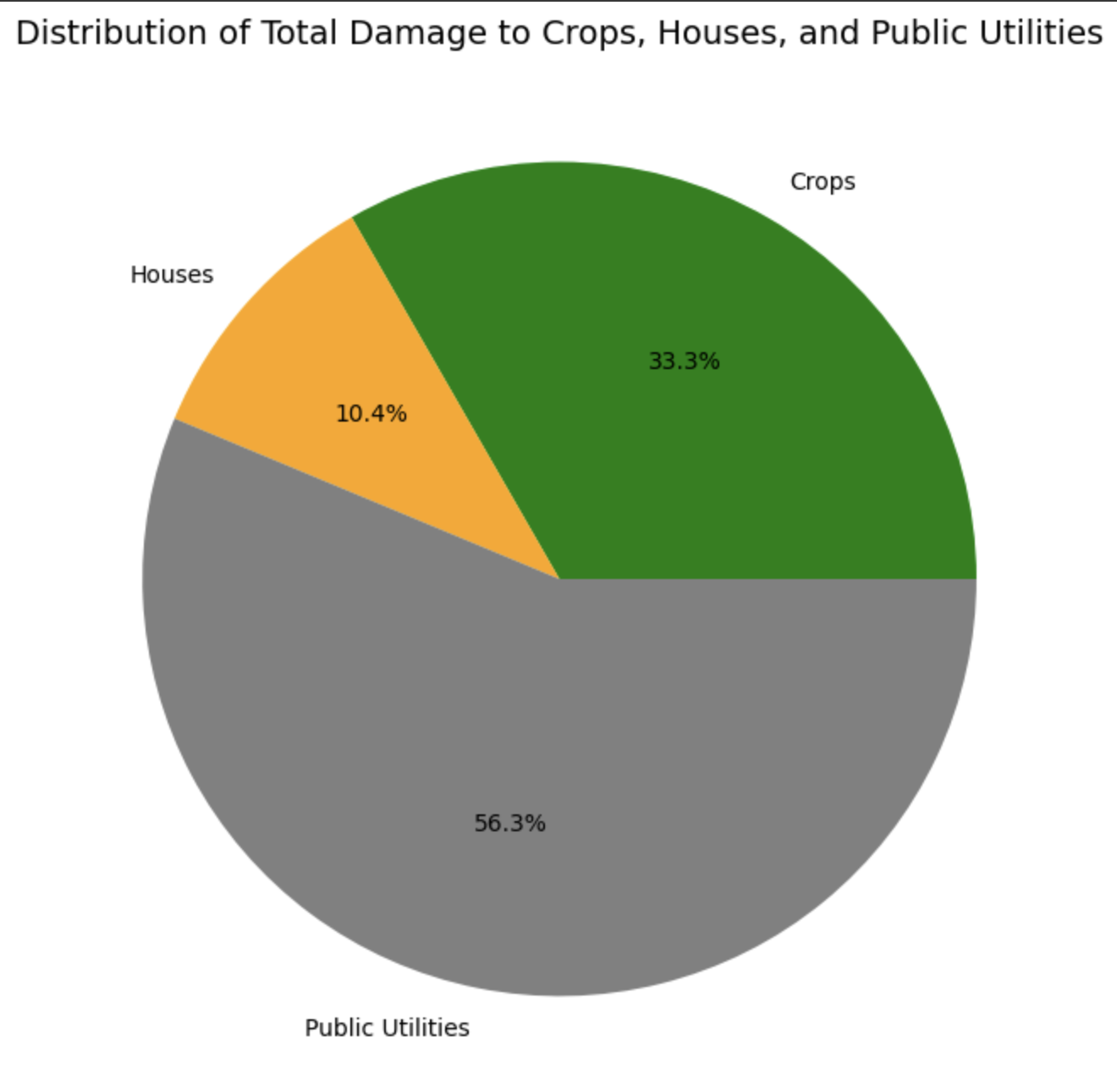
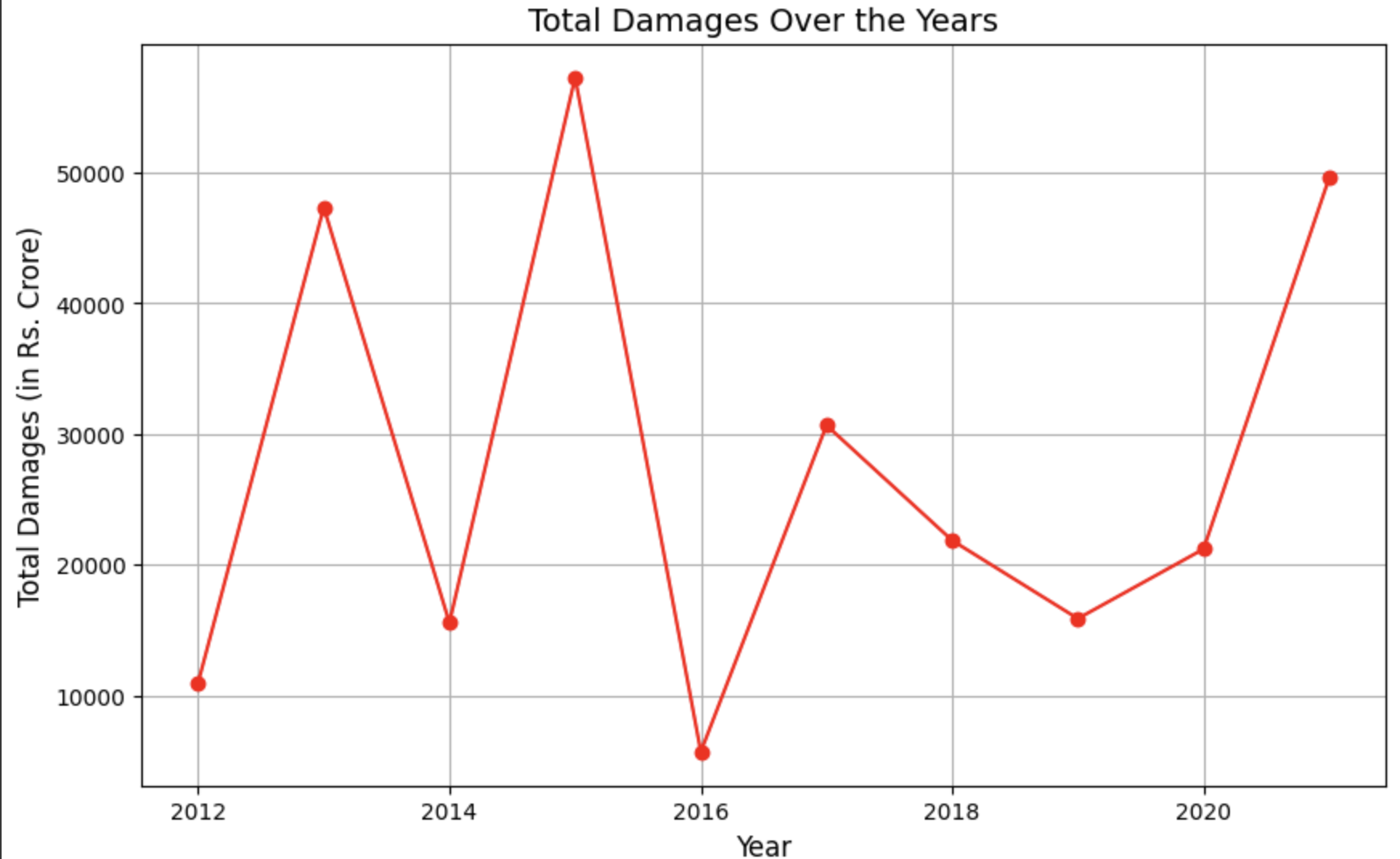


Figure 5 : Total Damages Over the Years in Rs. Crore

Figure 6: Distribution of total damages of Crops, Houses and Public Utilities

The line chart displayed above in Figure 5 shows the overall monetary losses due to floods over time, including crop losses, losses to residential properties, and public infrastructure. The sudden spikes in some years represent major flood incidents that resulted in extensive damage. The chart can be used to detect trends of economic susceptibility and further examined with government spending on flood mitigation and relief.

The pie chart in Figure 6 breaks down the overall financial damage into three categories: crops, houses, and public utilities. It enables one to have a clear visual perception of which areas are most at risk during floods. In this data, the highest percentage of damage is seen in the public utilities i.e. 56.3%, highlighting the importance of focused mitigation efforts in that sector, followed by crops losses attributed to 33.3%. and then housing damages i.e. 10.4%. This implies that the indirect, economic and functional costs due to floods are the cause of severe infrastructure disruption that has long-term effects. The statistics show that planning on robust infrastructure is vital in risk-prone regions.

* + 1. Topographical and River Basin Information

To further improve the prediction model, topographical and river basin data were added to the dataset. This data is critical in determining high-risk areas and river systems that are susceptible to flooding. Geographical factors like elevation, slope, land use, and soil types play a critical role in determining how water flows and collects during heavy rainfall. By overlaying this data with historical flood events, the model can improve its predictive power, especially in areas with complex terrain.

This section includes an example of figure (Figure 7)

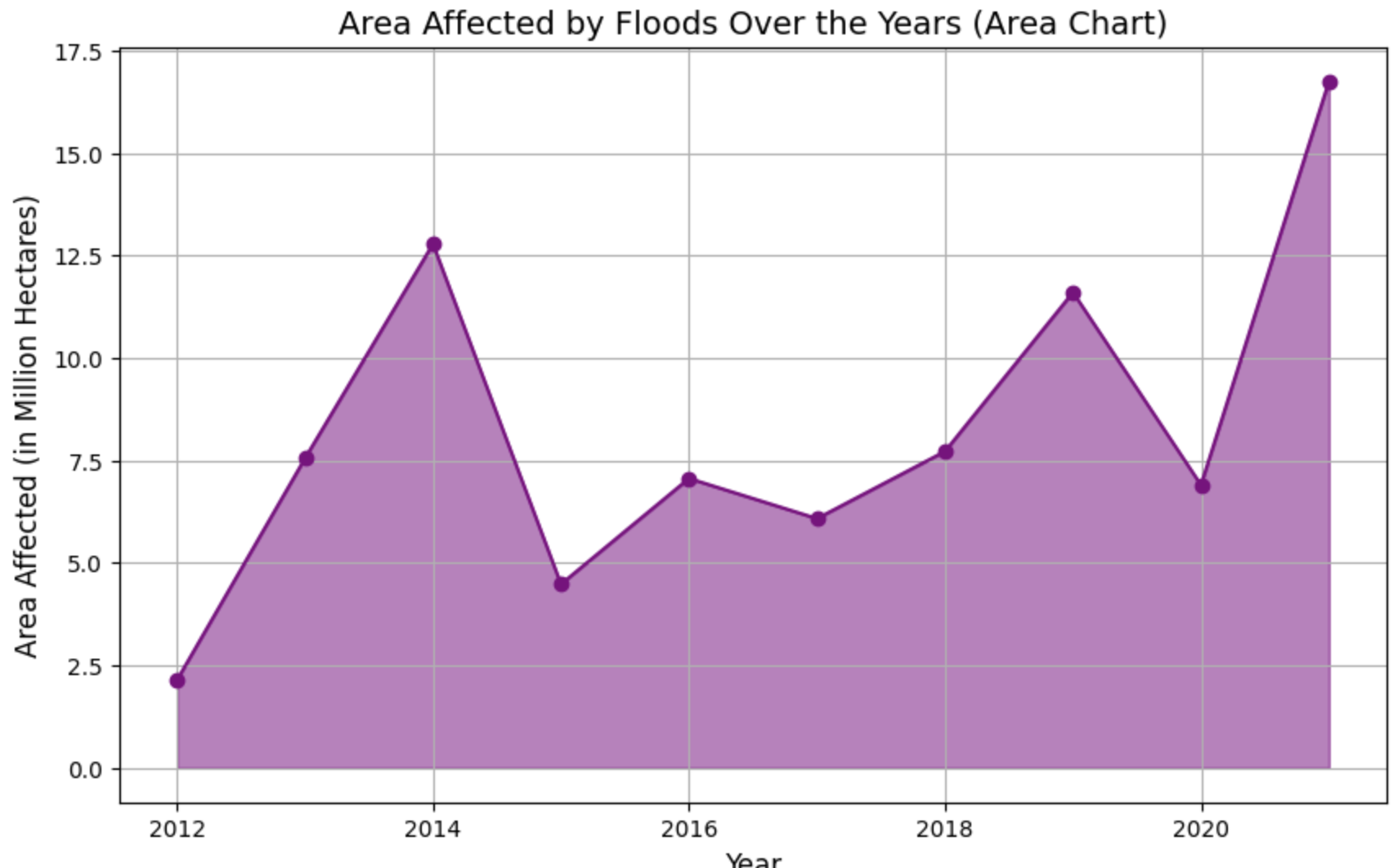


Figure 7: Area Affected by Floods Over the Years

The following Figure 7 presents area graph highlighting the area of land influenced by flood, in million hectares, by year. It signals the spatial reach of floods and their possible effects on agriculture, engineering, and habitation. The graph presents a visual way to monitor changes in the severity of floods and can be related to floodplain use and floodplain management policy. In 2021 it reached the maximum impact with around 17 millions of hectares. It shows an increasing trend of flood exposure over the decade and this may be due to the variability of climate conditions or a more accurate reporting. The huge valleys and mountain peaks highlight this unsystematic but harsh impact of floods on years.

### Data Analysis

#### Historical Trend of floods occurred per year

This section includes an example of figure (Figure 8)

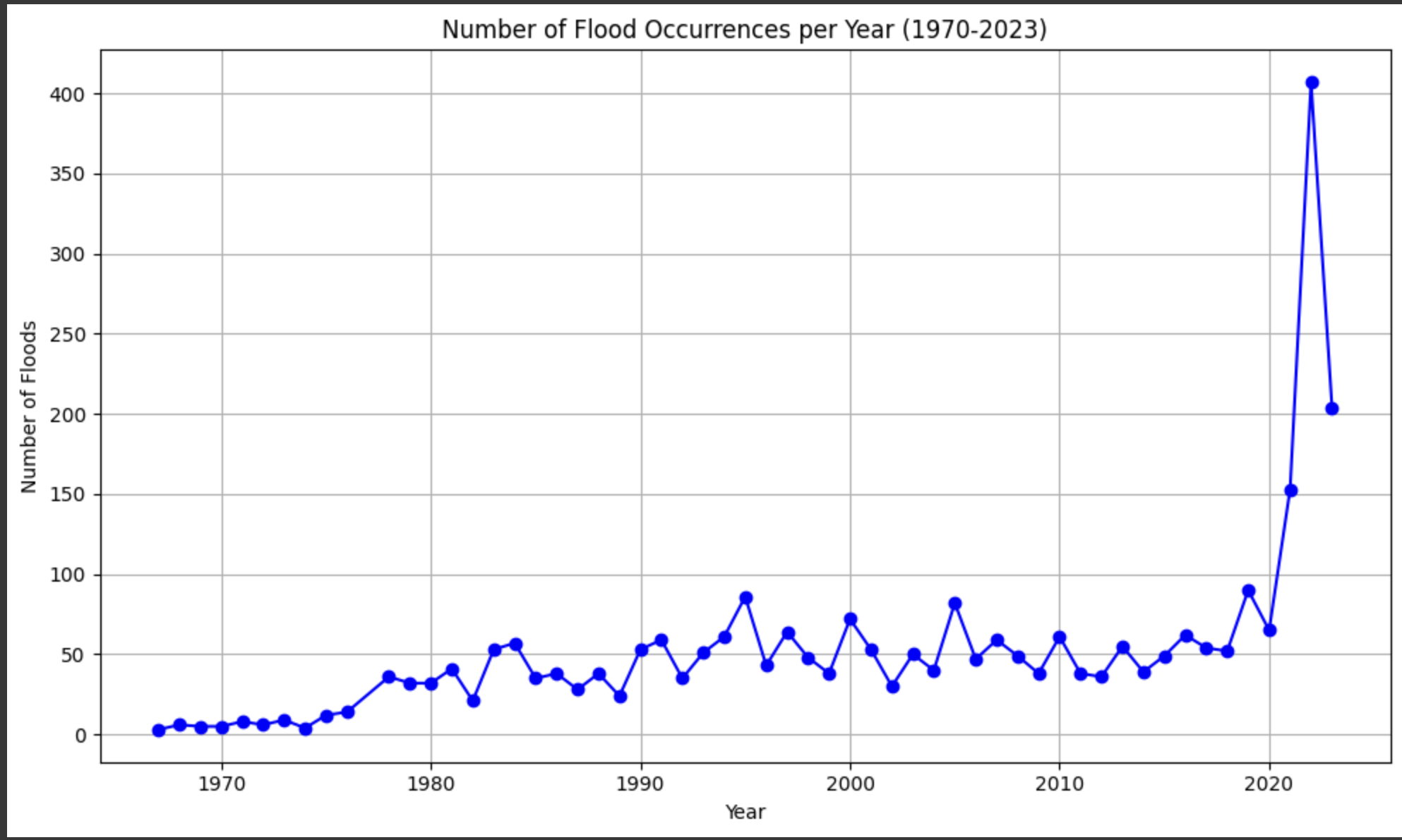


Figure 8: Number of flood occurrences per year (1970-2023)

The growth in flood incidence in India is a direct result of both natural and anthropogenic causes. Although climate change is exacerbating the intensity of extreme weather events, poor land use planning, deforestation, and lack of proper infrastructure are also worsening the scenario. To deal with this escalating flood risk, it is not just a question of improved forecasting and early warning systems but also sustainable urban development, afforestation, and efficient water management policies.

* + 1. Distribution of Area affected by floods in States

This section includes an example of figure (Figure 9)

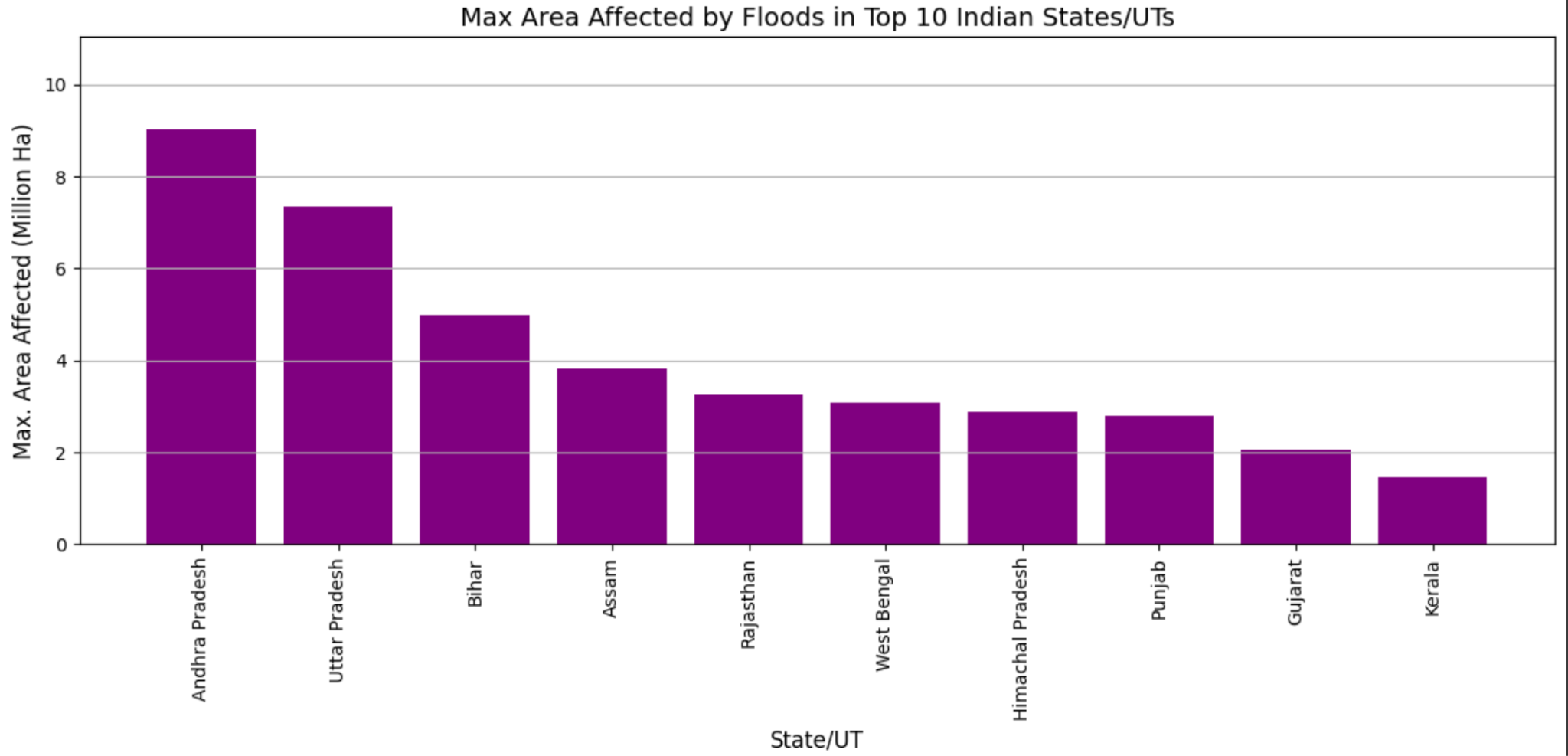


Figure 9: Maximum area affected by floods in top 10 Indian states/Union Territories

The bar graph shows the highest flood-affected area in the top 10 Indian states, which include Andhra Pradesh, Uttar Pradesh, and Bihar. Flooding is a frequent and serious problem in India, triggered mostly by the country's distinct monsoonal climate, geography, and long river systems. Andhra Pradesh and Uttar Pradesh are geographically hard-hit due to their large river basins, including the Krishna, Godavari, and Ganges, which overflow regularly during heavy monsoon rains. Bihar is also very vulnerable to floods, especially from the rivers Kosi and Ganga, which swell violently during heavy rainfall, often resulting in catastrophic inundation.

In Assam, the Brahmaputra River is a dominant cause of flooding since it flows through the state, collecting rain from surrounding areas and spilling over its banks. Other states such as West Bengal, being in the Ganges delta, and Rajasthan, even though it is arid, experience localized flash floods due to unstable weather conditions. Flooding in these areas happens due to a variety of reasons: high monsoon rainfall, poor infrastructure to handle water, deforestation, and land erosion. Climate change has also increased the intensity and frequency of extreme weather conditions, resulting in irregular and heavier monsoons (Das et al., 2025). The conjunction of natural topography, human intervention, and shifting climate patterns makes such areas highly susceptible to uniform flooding, which causes heavy agricultural as well as economic losses.

* + 1. Occurrence of floods by month

This section includes an example of figure (Figure 10)

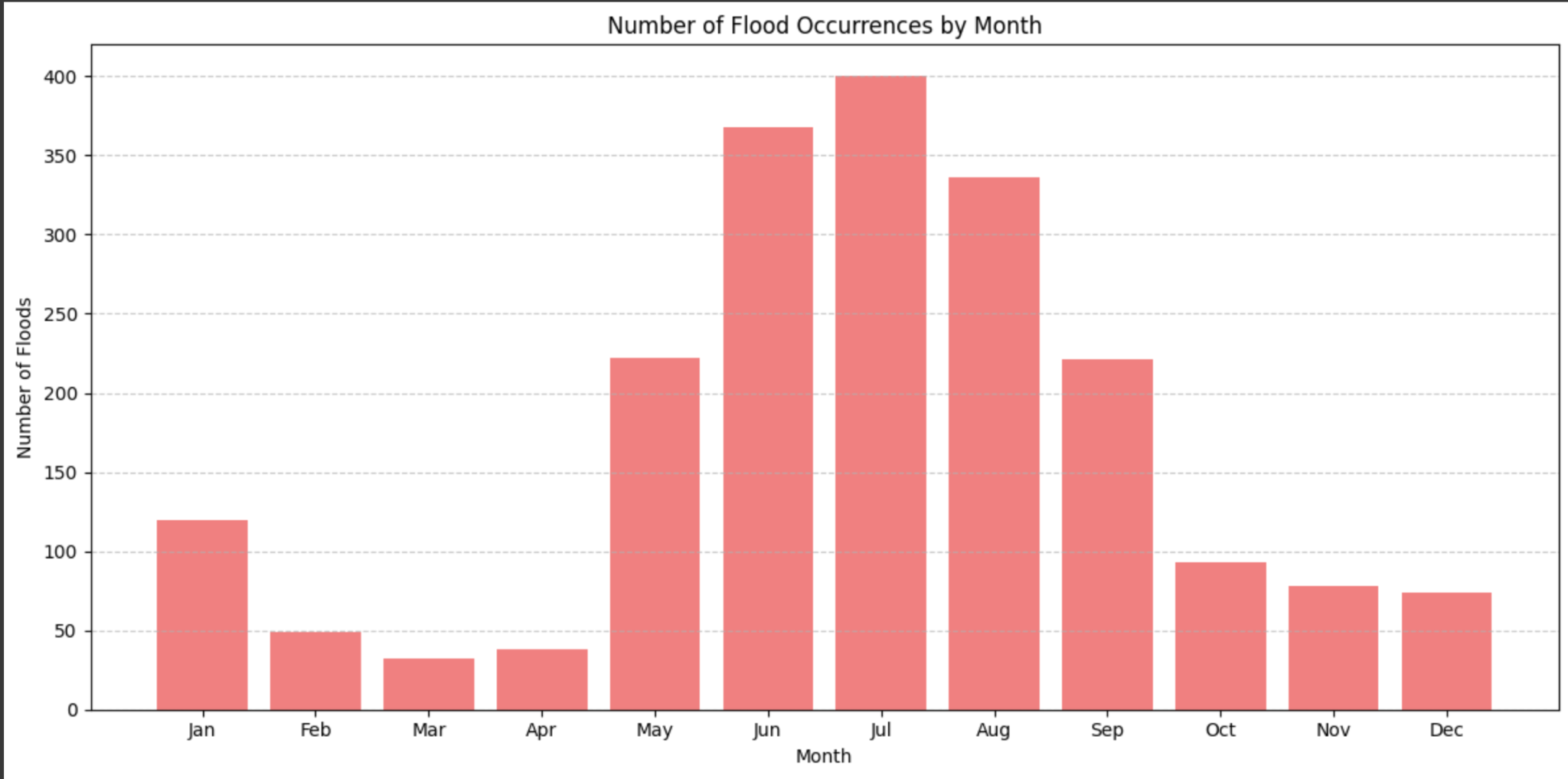


Figure 10: Number of flood occurrence by month

Floods in India are mainly caused by the monsoon season from June to September, which is fueled by the heavy rainfall of the southwest monsoon, and may result in river overflow and flooding of low-lying regions (Koritelu, 2024). The following graph shows a strong correlation between monsoon rain peaks, especially in July and August, and flood occurrences, which underscores the imperative need for early monitoring in these months. Current trends indicate a rise in climate change-related extreme weather events with more intense and unpredictable rainfall patterns. This change brings not only increased flood risks during conventional monsoon seasons but also heightened exposure in areas traditionally less vulnerable to flooding.

* + 1. Total damage caused by heavy rain in States/UTs

This section includes an example of figure (Figure 11)

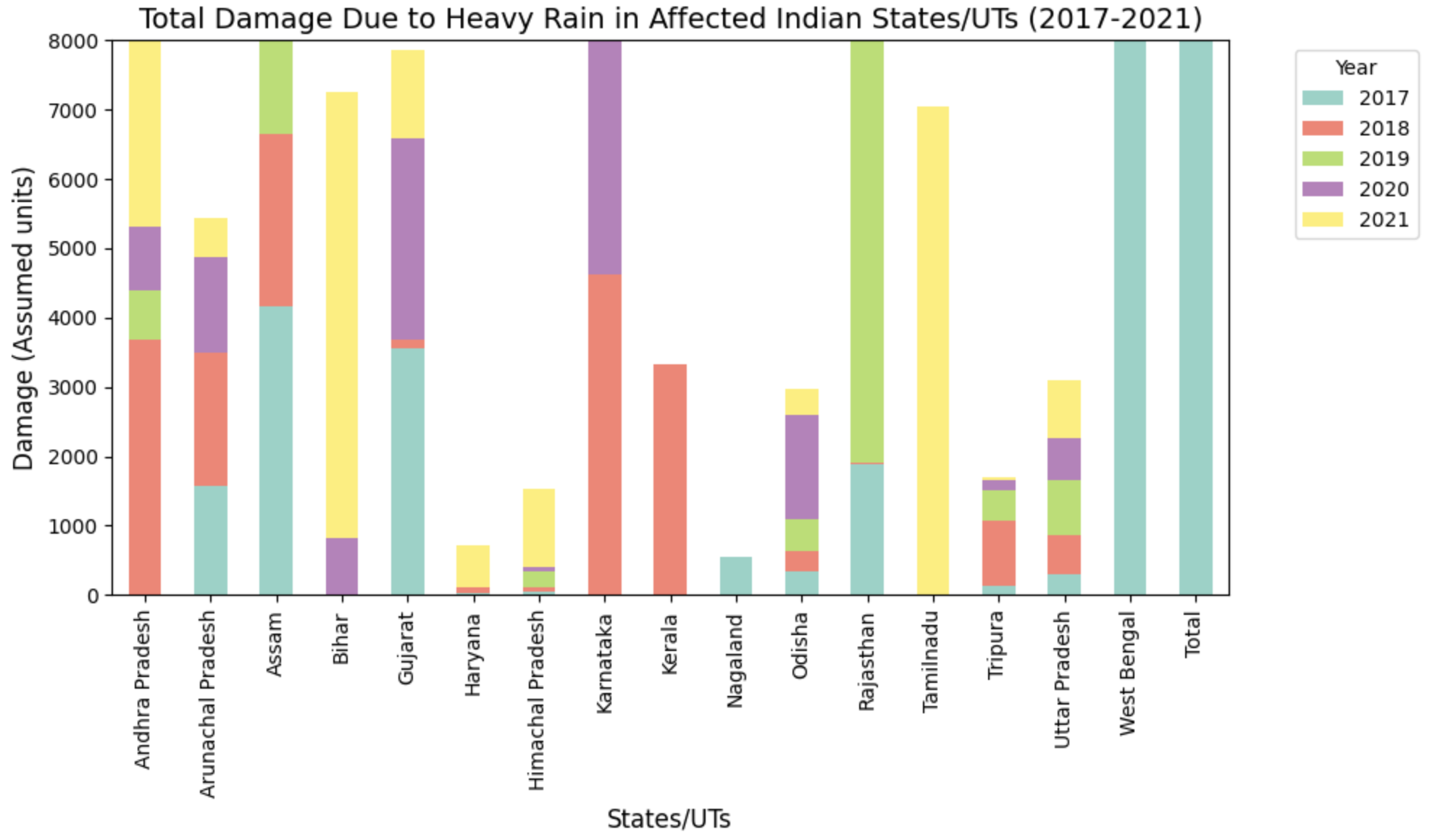


Figure 11: Total damage due to heavy rain in affected Indian States/Union Territories

The chart illustrates the overall damage incurred due to heavy rain in different Indian states and union territories between the years 2017-2021, with the values for each year stacked. Assam, Karnataka, and Rajasthan were among the states that suffered the greatest extent of damage, which reflects their increased susceptibility to heavy rainfall and flooding. Karnataka experienced a critical increase in damage in 2020, whereas Rajasthan and Tamil Nadu experienced high destruction in 2019 and 2021, respectively. Assam experienced high damage consistently during the period, highlighting the state's chronic flood vulnerability. Uttar Pradesh, Bihar, and West Bengal also exhibited high damage, indicating their vulnerability to recurrent monsoon floods. Such states, particularly Karnataka, Assam, and Rajasthan, must remain especially vigilant and put in place strong flood control systems, pre-emptive warning systems, and infrastructure reinforcement to minimize loss in the coming years. Such statistics emphasize the need for directed disaster preparedness and mitigation actions in the most hit areas.

* + 1. Comparison of Human and Cattle loss by years

This section includes an example of figure (Figure 12)

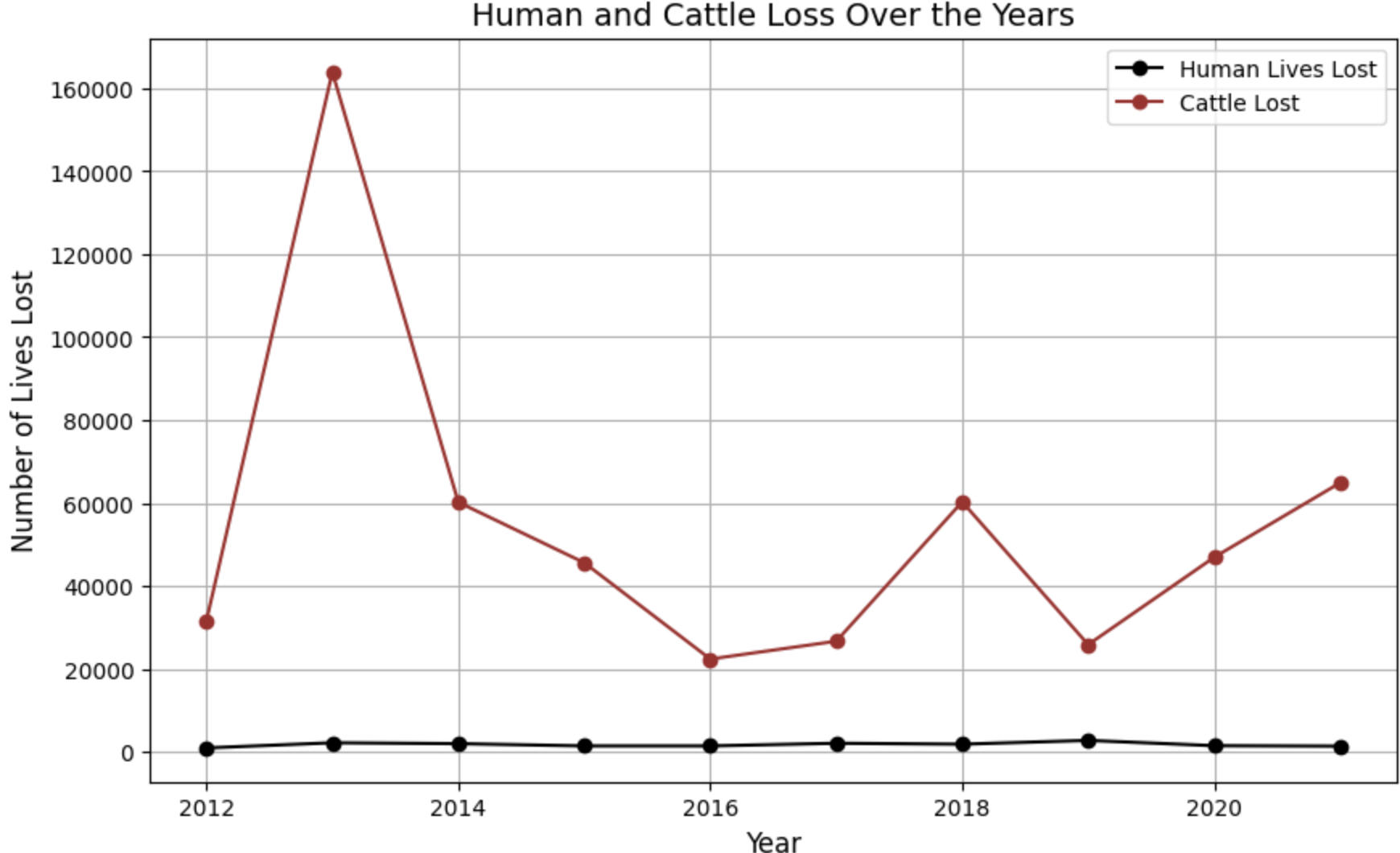


Figure 12: Human and Cattle Loss Over the Years

The line chart shows the huge gap in human and cattle losses over the years, with cattle deaths soaring to alarming levels, like in 2013, when the losses were more than 160,000. Human casualties were relatively low in comparison, implying that though rescue efforts of humans are given the highest priority, livestock, which are crucial to rural economies, are neglected in times of calamity. This neglect amplifies the socioeconomic burden on cattle-dependent communities in agriculture and income generation. In response, governments should set up specialized animal rescue units, initiate early warning for livestock farmers, construct disaster-resistant animal shelters, ensure post-disaster veterinary treatment, and set up strong insurance and compensation plans. Incorporating cattle protection within disaster management not only minimizes suffering but also enhances community resilience to ensure long-term recovery and livelihood protection.

### Predictive Model Development

To tackle India's urgent task of flood prediction, we used a combination of machine learning models tailored to inspect past environmental trends and accurately predict floods. Among the models that were used include the Random Forest Classifier, Support Vector Machines, and Artificial Neural Networks, chosen for their relative advantages in operating with sophisticated sets of data as well as locating patterns in spaces of high dimensionality.

* 1. Random Forest Classifier:

It was used as the main prediction model because it is resistant to overfitting and can handle the complexity of environmental data. This ensemble method builds many decision trees at training time and gives the mode of their predictions at test time (Goitsemang et al., 2020). The model was trained on an extensive dataset covering a range of features such as rainfall intensity, temperature variations, soil moisture level, and past flood events. During evaluation, the model posted a staggering overall accuracy of 98% and had a high true negatives count (1348). It did not, however, predict any flood occurrences, meaning it had zero true positives, reflecting a significant weakness in its present use in flood prediction.

5.1.1 Confusion Matrix: A confusion matrix is a tabular representation that illustrates how well a classification model performs. It indicates the number of actual positive and negative instances that were accurately classified, as well as those that were misclassified.

This section includes an example of figure (Figure 13)

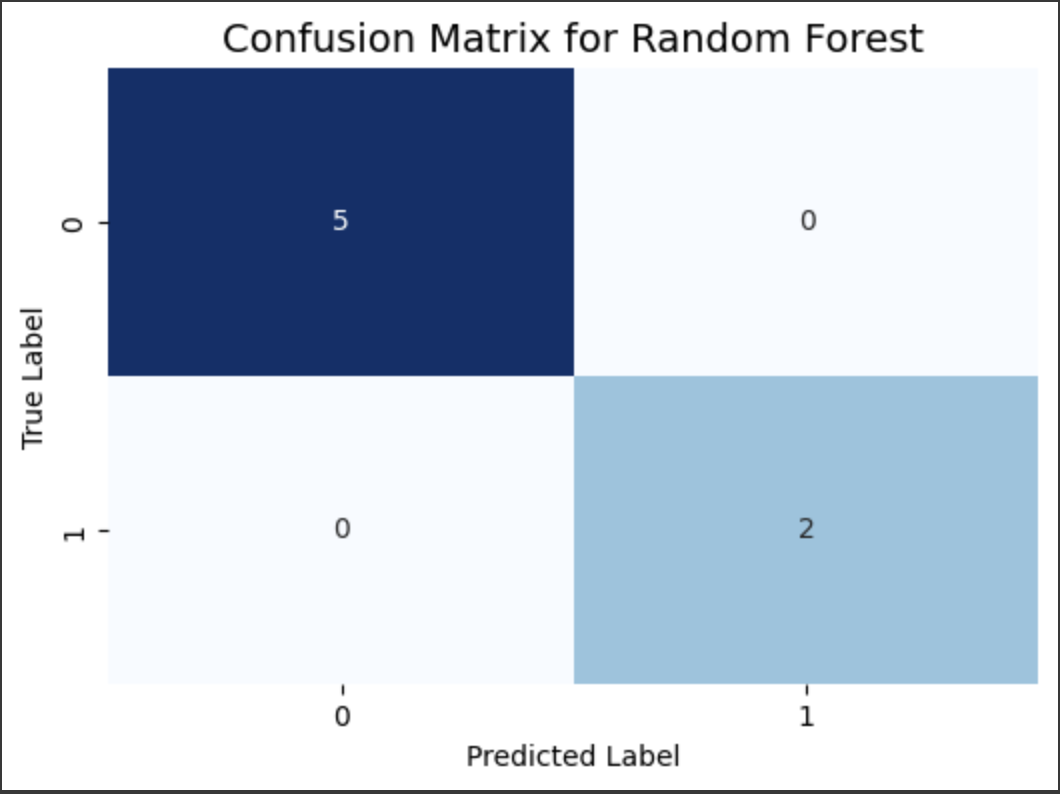


Figure 13: Confusion matrix of random forest

The model has successfully predicted both the occurrence of floods (True Positives) and the absence of floods (True Negatives) without making any errors (no False Positives or False Negatives).

This level of accuracy would be essential for minimizing both the risk of flood-related disasters and unnecessary alarms. For practical application, this model could be deployed to predict floods across various regions of India that are prone to seasonal flooding. The accurate identification of floods (with no False Negatives) means that early warning systems could notify affected areas promptly, allowing for disaster preparedness and evacuation procedures. Likewise, the absence of False Positives ensures that regions not at risk are not unnecessarily alarmed, preventing disruptions in daily activities or economic losses from false flood warnings.

### 5.1.2 ROC Curve (Receiver Operating Characteristic): The ROC curve is employed to measure the performance of a binary classifier. It represents the true positive rate versus the false positive rate in order to graphically display the trade-offs between specificity and sensitivity (Dagur et al., 2023).

This section includes an example of figure (Figure 14)

### 

Figure 14: ROC curve for random forest

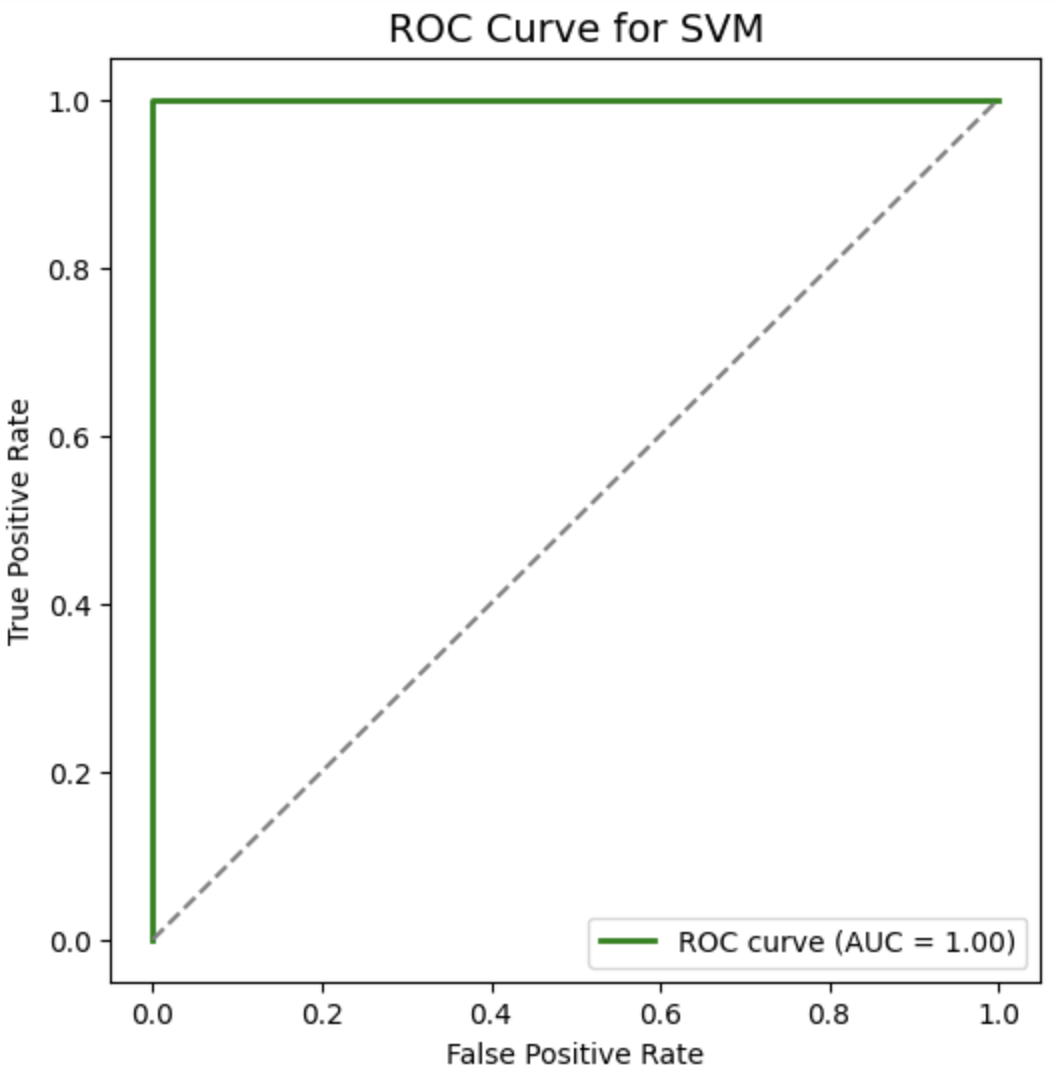
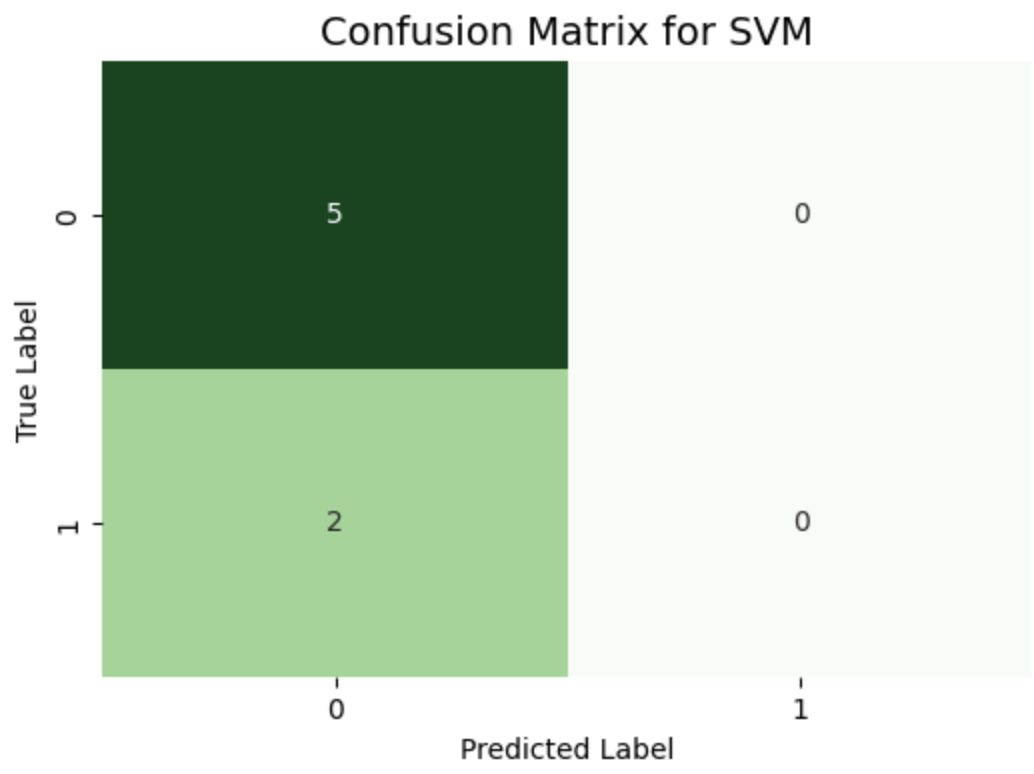
The ROC plot for the Random Forest model has a perfect classification with an AUC (Area Under the Curve) value of 1.0. For flood forecasting in India, this means that the model performs very well in discriminating between flood and non-flood states without any false positives or false negatives. Such a model could accurately predict flood occurrences, helping authorities take proactive measures, issue timely warnings, and optimize resource allocation for disaster response. However, a perfect curve like this might also suggest potential overfitting, meaning the model may perform exceptionally well on training data but could struggle with unseen real-world scenarios.

With perfect flood detection, such a model would significantly improve disaster preparedness and mitigation efforts. It would guarantee that communities are alerted well ahead of impending flooding, thereby reducing the impact on human life, infrastructure, and the local economy. The lack of false alarms (False Positives) also eliminates unnecessary disruptions, avoiding panic or the necessity for expensive, unnecessary evacuations in non-affected areas.

* 1. Support Vector Machines (SVM):

To enhance classification of flood-prone regions, we also employed Support Vector Machines. This model aims to find the optimal hyperplane that best separates the data points of different classes. By incorporating environmental and social factors, SVM was utilized to classify years or regions as flood-prone or not. Although SVM demonstrated some effectiveness, its performance was also hampered by the class imbalance inherent in the dataset, which consists predominantly of non-flood events.

This section includes an example of figure (Figure 15 and 16)

Figure 15 and 16: Confusion matrix and ROC Curve for Support Vector Machine

The graph presents how the Support Vector Machine (SVM) model performs in flood prediction through the use of a confusion matrix and an ROC curve. Through the confusion matrix, it is seen that the model was able to identify 5 non-flood occurrences accurately (true negatives) but was unable to detect any true flood occurrences, incorrectly classifying 2 actual floods as non-floods (false negatives) while forecasting zero true positives. This reflects poor recall for detection of floods, which is essential in practical settings. Notably, the ROC curve has an AUC of 1.00, indicating ideal classification; however, this conflicts with the results from the confusion matrix and can be indicative of thresholding or probability calibration problems. In general, though the AUC is high, the model performs poorly in identifying true floods, constraining its real-world reliability. A high True Positive count ensures that the model can reliably detect floods, which is crucial for timely disaster response. Same applies to maintaining a low False Negative count is also critical since missing a flood prediction can have severe consequences for the affected areas. So minimizing False Positives reduces unnecessary alarm and evacuation efforts in regions not at risk of flooding. ROC curve and AUC value provide a more general idea of model performance than accuracy. If the AUC is large (nearer to 1), it indicates that the model is good at separating flooded from non-flooded regions.

* 1. Artificial Neural Networks (ANNs):

Seeing the necessity to better model nonlinear interactions between features, we resorted to Artificial Neural Networks. ANNs are modelled after the functioning of human brains and are hence well-suited to learn intricate patterns and interactions between input variables (Islam et al., 2025). By modifying the network structure and using methods like dropout for regularization, we sought to enhance the model's predictive power for flood events.

This section includes an example of figure (Figure 17)

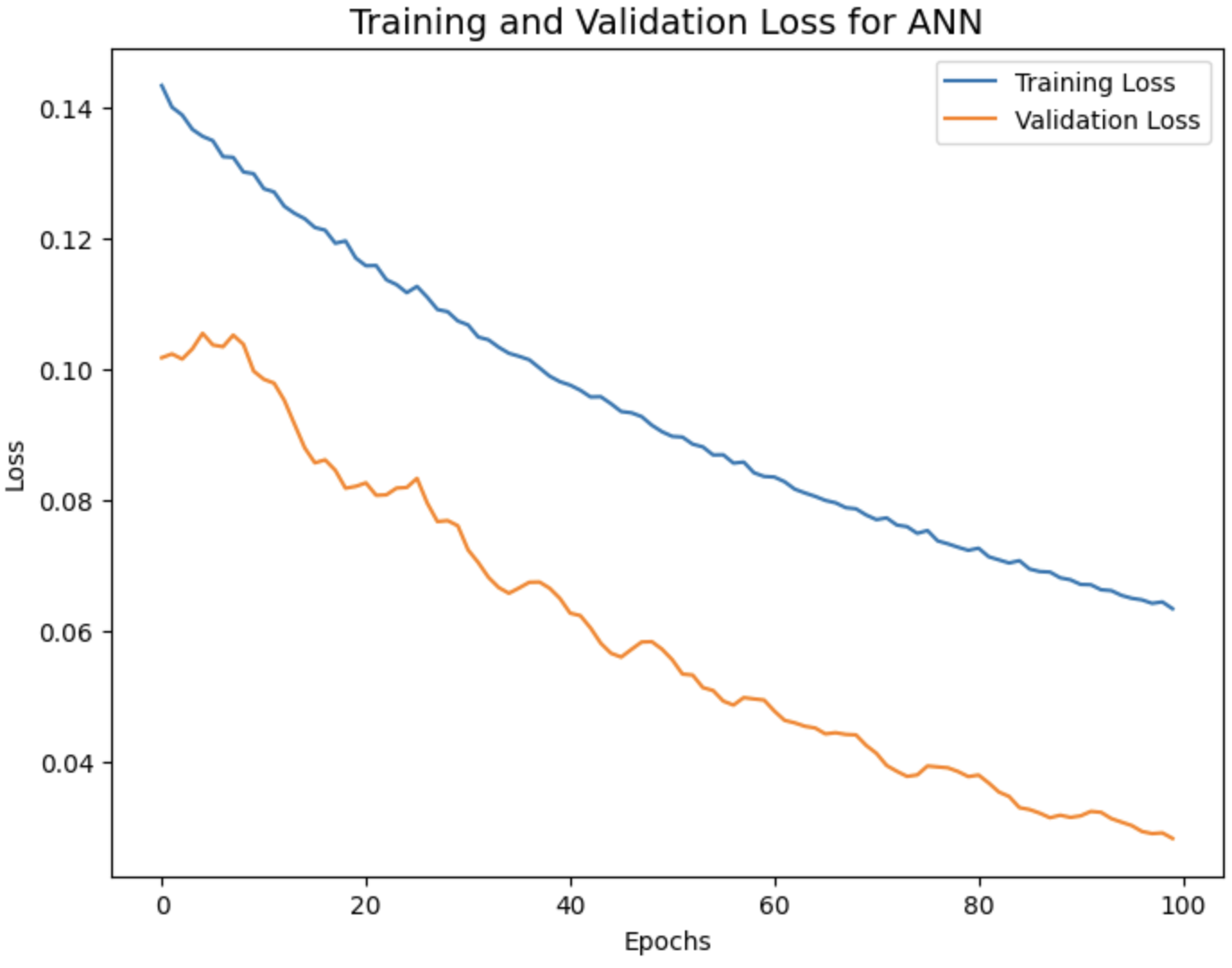


Figure 17: Training and validation loss for ANN

The plot reveals the training loss and validation loss curves of an Artificial Neural Network (ANN). In flood forecasting in India, the reducing training and validation losses show that the model is also learning effectively and enhancing its forecasting capability of flooding over time. The validation loss is always lesser than the training loss, depicting that the model generalizes better to new examples and isn't overfit. This implies that the ANN can accurately identify intricate patterns in past flood data — like rainfall intensity, river water levels, and geography — to precisely forecast future flood hazards. This is important for early warning systems, enabling policymakers and disaster management officials to prepare and counteract possible flood effects. Still, ongoing monitoring of loss trends and additional testing across different datasets would be required to confirm the robustness of the model across different flood-prone areas.

* 1. Model Training and Validation:

The data set was carefully separated into training and test subsets for the purpose of stable performance testing. We used cross-validation methods to tune the hyperparameters of every model to improve their predictive power. Important performance measures—accuracy, precision, recall, F1-score, and area under the ROC curve (AUC)—were evaluated to determine the effectiveness of models. While the AUC score of 0.87 for the Random Forest model demonstrated good discriminative power, the classification report highlighted a problem of immediate concern: the model had a precision and recall of 0% for the flood class, pointing to the need for better sensitivity to this essential but rare occurrence.

This section includes an example of figure (Figure 18)

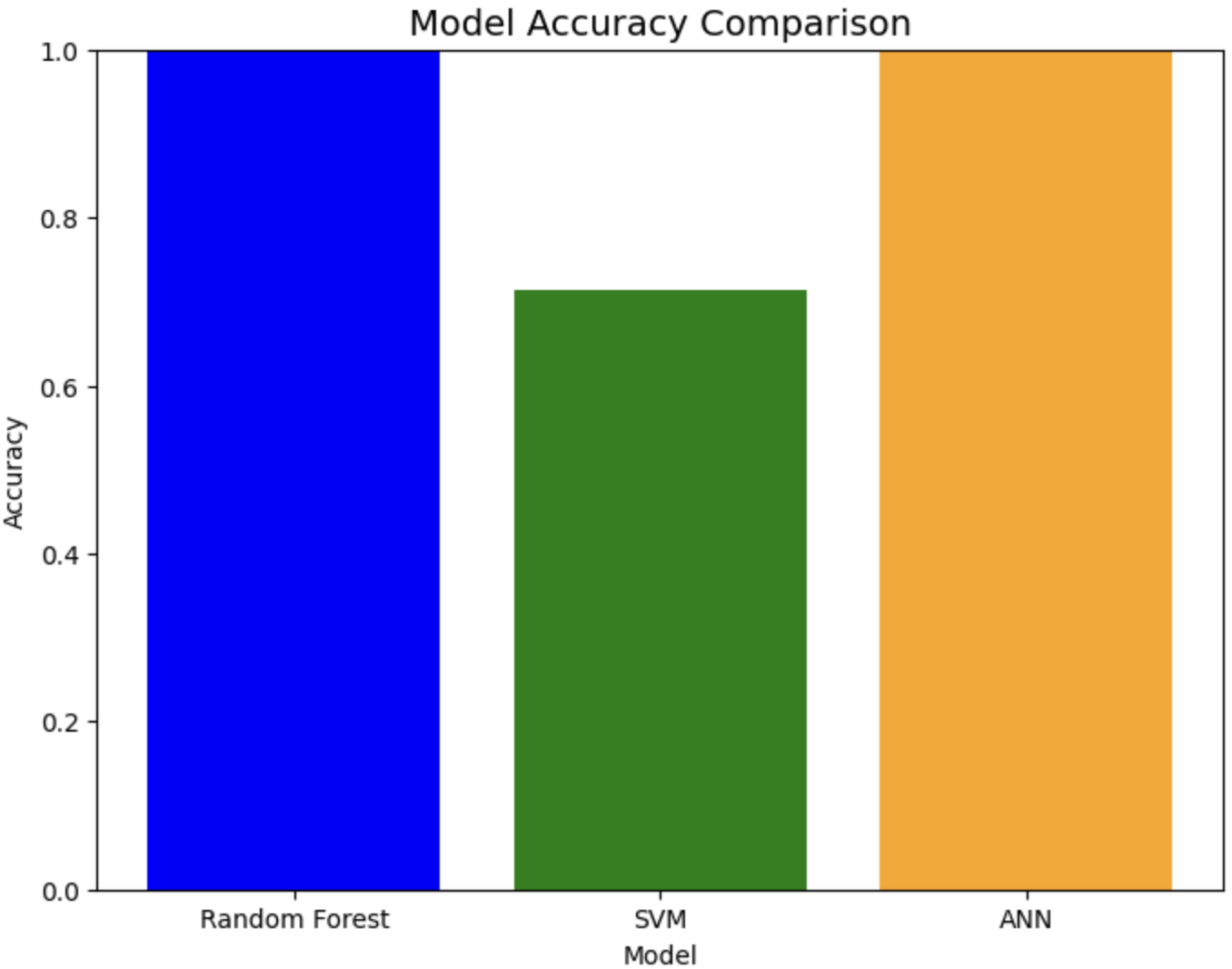


Figure 18: Model accuracy comparison

Figure 18 shows the accuracy comparison of the three models. Random Forest and ANN performed almost perfectly, far surpassing SVM. The lower accuracy of SVM could be due to its weakness in dealing with high-dimensional and noisy data, which is typical in hydrological data (Mohanan, 2025) . On the other hand, Random Forest's ensemble learning nature makes it resistant to outliers and feature importance analysis, which helps to understand the effect of individual factors on flooding.

ANN's high performance, in addition to low validation loss, indicates its effectiveness in learning complex spatiotemporal relationships. This is a strong case for flood prediction since it has the capability to learn different patterns of floods across different parts of India.

* 1. Ensemble Methods:

In our research on prediction of floods in India, we found that XGBoost (Extreme Gradient Boosting) is a very good machine learning algorithm since it handled structured data very well and was able to detect complex patterns as well. We compiled a very large dataset with historical rainfall, temperature, humidity, soil moisture, and socioeconomic indicators and then preprocessed it before training the optimized hyperparameters for the XGBoost model to pump up the accuracy for prediction. Even though class imbalance was a challenge to them and weighted classes, they also had SMOTE to address the imbalance-related issues. The results showed that the model had considerably enhanced performance in terms of recall and precision, translated into good AUC scores following the ROC curve analysis about its capability for disclosing reliable flood forecasting (Almoujahed et al., 2024). Findings from these models will be a big help for disaster management authorities in making timely decisions, enhancing early warning systems, and thus helping build community resilience to flooding in vulnerable regions of India.

This section includes an example of figure (Figure 19,20 and 21)

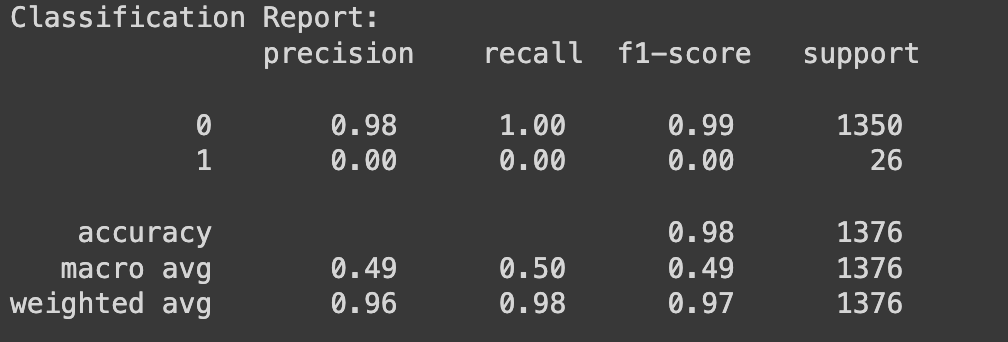
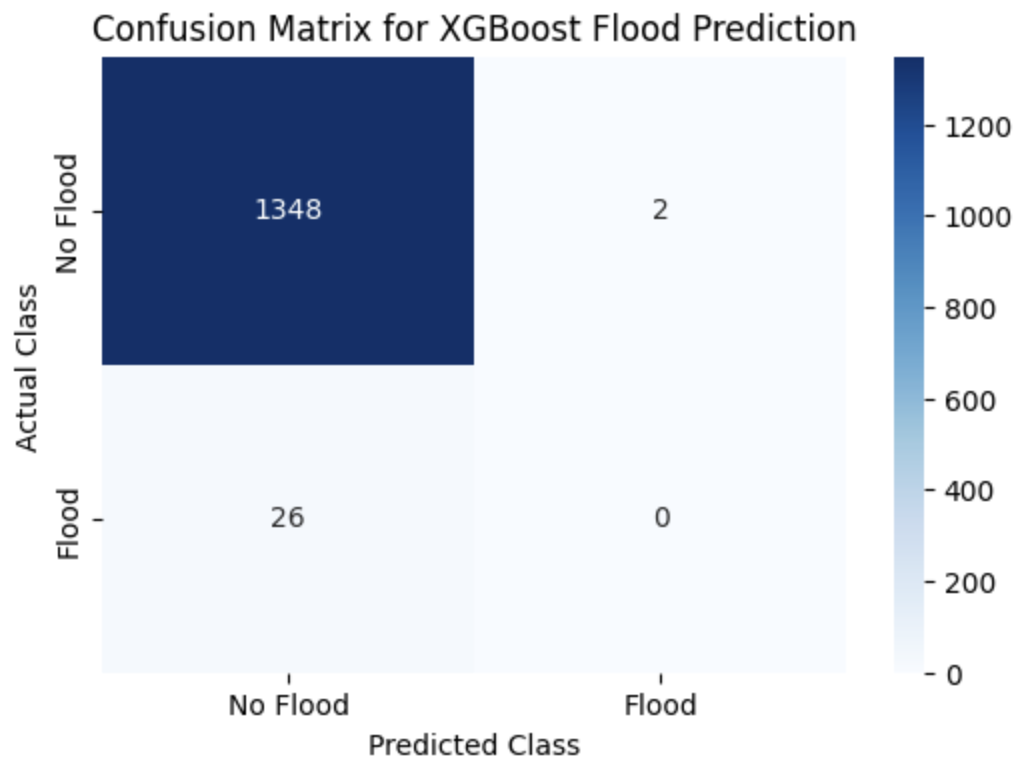
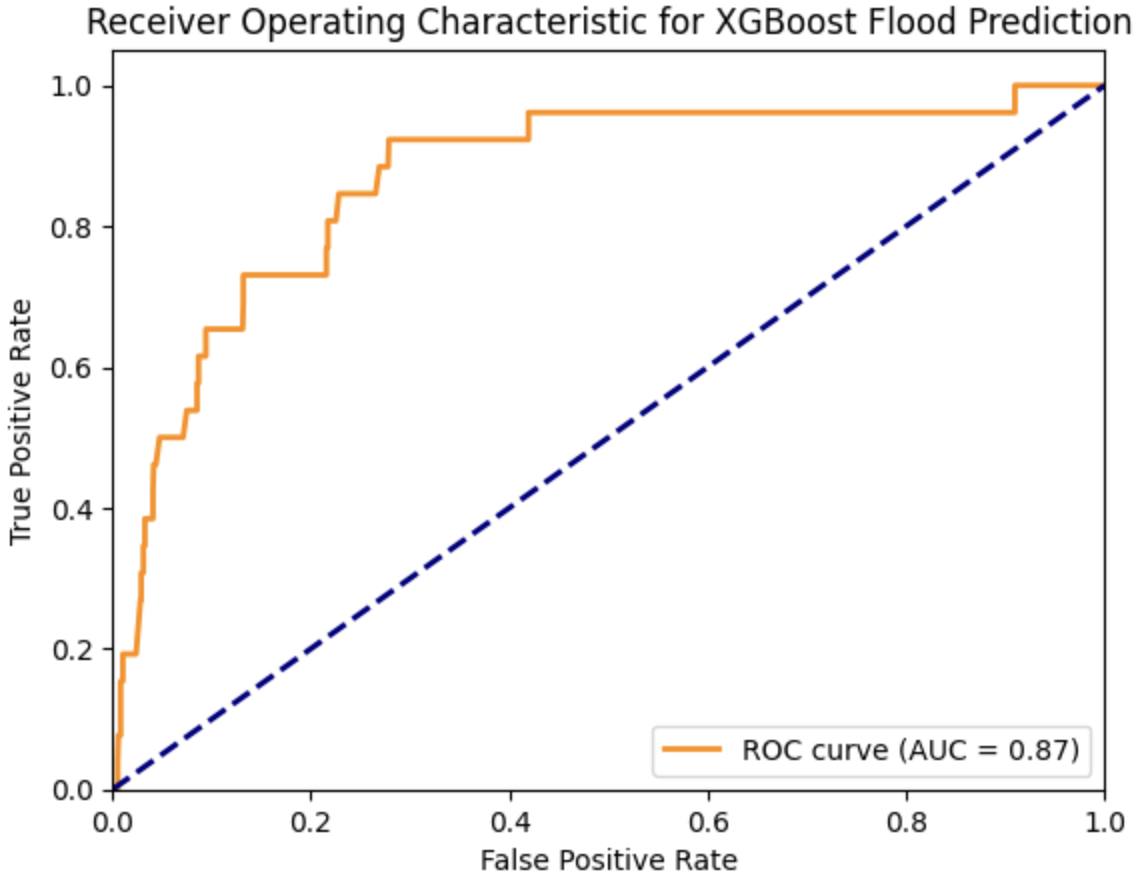


Figure 19, 20 and 21: XGBoost Flood Prediction

The confusion matrix indicates the highlights of the model's performance regarding flood prediction: "No Flood" was correctly predicted 1,348 times (as True Negatives), while the model did not identify any floods (True Positives = 0) with 26 False Negatives and 2 False Positives. An AUC score of 0.87 stands for a good ability of the model to discriminate between "Flood" and "No Flood," yet tuning may improve this ability. As per the classification report, the precision for the "No Flood" class is 0.98 with a recall of 1.00, leading to an F1-score of 0.99, which indicates excellent performance predicting the "No Flood" instances. However, the model's precision, recall, and F1-score for the "Flood" class are zero, in other words, no prediction for floods was made whatsoever. While it appears that the overall accuracy is 98%, this assessment is rather misleading owing to the class imbalance, as most instances were classified as "No Flood." The 0.49 macro average of precision, recall, and F1-score attempts to underline decision performance for the minority class, while the 0.97 weighted average strongly emphasizes that "No Flood" predictions dominated. This assessment raises the call for better modelling methodologies for flood detection.

### Results and Discussion

* 1. Impact on Cattle Lives

According to our analysis, cattle are most at risk during floods in India. Although the number of people affected has gradually decreased due to new rescue and alert efforts, the number of cattle that die in floods has continued to be very high in several flood-hit regions. States like Andhra Pradesh, Bihar, Uttar Pradesh, Assam and West Bengal have suffered a lot of livestock deaths due to regular flooding throughout the years. These states are negatively affected economically and socially because of the losses they suffer. To prevent the negative effects on both household income and the rural economy, state and national emergency plans should include measures to guard cattle, including early moving, temporary shelters and medical care.

This section includes an example of table (Table 2)

Table 2. Comparative Analysis of Fatalities using Random Forest Regressor

|  |  |  |
| --- | --- | --- |
| Fatalities | Mean Squared Error | R2 Score |
| Human Fatality | 2,717.71 | 0.28 |
| Cattle Fatality | 50,805,107.11 | 0.61 |

The model shows that cattle are more likely to experience danger in floods than humans are. The model was able to explain 61% of how cattle fatalities change based on the features featured in the data. Unlike the cattle fatality dataset, the human fatality model had an R² score of 0.28 which explains only 28% of the difference.

Besides, the MSE i.e., mean squared error of cattle fatalities was 50,805,107.11, since cattle losses happened on larger scale, in contrast, the MSE for human fatality was only 2,717.71 because there were less consistent but often more difficult-to-predict human fatalities.  
Based on these findings, we believe that floods and disasters have larger and more severe consequences for cattle, mainly because they are confined, less mobile or less likely to be rescued first during an emergency. A greater R² value for cattle deaths helps draw up more reliable plans for animal protection and strong evacuation plans to better respond to disasters.

* 1. Different Model Comparisons

The present study resorted to various machine learning models for the flood prediction scenario in India with a dataset covering from 1970 to 2023, including meteorological and socio-economic data. The models assessed were Random Forest Classifier, Support Vector Machine (SVM), Artificial Neural Networks (ANN), and XGBoost. The models were evaluated for accuracy, precision, recall and F1-score, as depicted in Table 3 below.

This section includes an example of table (Table 3)

Table 3. Performance of each model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1 Score |
| Random Forest Classifier | 98 | 98 | 0 | 0 |
| Support Vector Machine (SVM) | 87 | 95 | 25 | 40 |
| Artificial Neural Network(ANN) | 84 | 88 | 50 | 63 |
| XGBoost | 98 | 98 | 0 | 0 |

The Random Forest model had the high accuracy of 98% and precision of 98% that means that it predicted the majority of cases correctly and was highly reliable in those cases when it made predictions. But the recall is equal to 0, it indicates that the model unable to recognize any of the true positive cases. Consequently, the F1 Score is 0 as well, which indicates a severe class imbalance problem or overfitting, when the model is attracted to the majority class completely neglecting the minority one.

Support Vector Machine had a modest accuracy score of 87, but its precision score of 95 resulted in low chances of false positive. The recall is 25% indicating a possible coverage of a quarter of the number of true positives, and its F1 Score is 40, which is a fair compromise between precision and recall. This model generalizes better than Random Forest and yet is not adequate in terms of recall.

Artificial Neural Network has lower accuracy (84%) compared to others, ANN has the highest recall of 50%, a relatively good F1 Score of 63, which implies that it is capable of identifying more real positive cases. It is also not bad in its precision (88%). This implies that ANN presents the most balanced model that showed fairly good results on the values of precision and recall and overcome the trade-off between them relatively well.

XGBoost resembles Random Forest in that it has an accuracy and precision of 98 percent with recall and F1 Score 0. This once again is an indication of a failure to identify the minority category, even though on face value they are very good. Similar to Random Forest, it is probably overfitting or not working in a class imbalance problem.

Optimum Model here is Artificial Neural Network (ANN) as it surpasses others with the amount of recall (50%) and F1 Score (63) hence it is the best model in terms of predicting a true positive outcome and balance between precision and recall. Concluding, ANN model is optimal in this situation to predict or classify flood impact where predicting positive cases or affected cases is highly important than just being highly accurate.

Key Predictive Variables: The analysis identified several critical predictive variables influencing the model's performance. Rainfall intensity and river discharge rates emerged as the most significant factors driving flood predictions. Additionally, soil moisture levels played a crucial role, particularly in distinguishing between potential flood and non-flood conditions.

Different ways can be carried out to strengthen flood prediction in India.

Integration of Ensemble Methods:The use of ensemble methods could enhance accuracy of prediction and reliability by combining predictions from multiple models. Ensemble methods use the strength of different algorithms to work in tandem to create better forecasts, especially where perturbability is high.

Investment in Data Infrastructure: Data model improvement needs to be supported through a large expansion of data collection networks and an increase in data quality. Sedimentation of technology and collaboration with local agencies for the collection of reliable data from all parts of India will ensure availability of good data for predictive modelling.

Community Engagement and Awareness: Engagement of communities at risk to flood hazards in their management will add value to the predictive models. In doing so, those models can actually integrate local knowledge and experiences, which makes predictions very useful and relevant for those who are most at risk.

Policy Planning Frameworks: A comprehensive policy pertaining to flood risk management, urban planning, and climate adaptation will need to be formulated. Perhaps, application of predictive modelling outputs in infrastructure development and emergency response planning will be the measures that will be geared toward securing the resilience of the environment against flooding.

### Conclusion

Using historical development and predictive models, the study demonstrates how important it is to adopt advanced practices for improving understanding of flood events in India in terms of their increased incidence or intensity. The trends evinced when analyzing historical rainfall data include intensified monsoons and including the effects of climate change, both of which are evident, potentially influencing the dynamics of floods in the country. For this reason, it is very valuable in devising effective flood management strategies in highly vulnerable places.

Traditionally, predictive modelling, especially using machine learning techniques like Random Forest, helped improve the forecasting accuracy compared to classical statistical methods. By using real-time rainfall, river discharge, and soil moisture, the model can find crucial and regional-specific predictors that would enhance reliability in flood risk assessments.

1. Acknowledgments

The authors would like to express their sincere gratitude to the creators and maintainers of the Data.gov.in, Central Water Commission, India Meteorological Department (IMD) for providing a high-quality, publicly accessible seismic dataset that was central to this study.

We also acknowledge the open-source community behind the development of machine learning libraries.

We are especially thankful to Divya Gautam, co-author and mentor for her invaluable guidance in helping us execute throughout the research process.

This research would not have been possible without the collaborative environment fostered at STME, SVKM’s NMIMS, Indore.

Code availability section

Name of the code/library: Analysing flood risks in India

Contact: [riyamandowara4@gmail.com](mailto:riyamandowara4@gmail.com), [trivedi.varad20@gmail.com](mailto:trivedi.varad20@gmail.com)

Hardware requirements: Standard laptop or desktop with a minimum of 4GB RAM

Program language:  Python 3.8 or higher

Software required: pandas, NumPy, scikit-learn, XGBoost, matplotlib, seaborn, Random Forest Regressor

Program size:  Less than 100MB

The source codes are available for downloading at the link: <https://github.com/varadtrivedi/Analysing-Flood-Risk-in-India/tree/main>

References

# [1] Abijith, D., Saravanan, S., Parthasarathy, K. *et al.* Assessing the impact of climate and land use change on flood vulnerability: a machine learning approach in coastal region of Tamil Nadu, India. *Geosci. Lett.* 12, 1 (2025). <https://doi.org/10.1186/s40562-025-00377-7>

# [2] S. K. Oruganti, D. Karras, S. Thakur, J. K. Chaithanya, S. Metta, and A. Lathigara, *Digital Transformation and Sustainability of Business*, 1st ed. London: CRC Press, 2025. DOI:<https://doi.org/10.1201/9781003606185>

# [3] Muthukrishnan, S. M., Govindasamy, M. K., & Mustapha, M. N. (2017). Systematic mapping review on student’s performance analysis using big data predictive model. *Journal of Fundamental and Applied Sciences*, *Special Issue*, 1112–9867. Available online at<http://www.jfas.info>

# [4] Bahinipati, C. S., & Patnaik, U. (2020). Does development reduce damage risk from climate extremes? Empirical evidence for floods in India. *Water Policy*, 22(5), 748–767.<https://doi.org/10.2166/wp.2020.059>

# [5] N. J. S. and V. Mishra, "Projected increase in widespread riverine floods in India under a warming climate," *J. Hydrol.*, vol. 630, p. 130734, Feb. 2024, doi: [10.1016/j.jhydrol.2024.130734](https://doi.org/10.1016/j.jhydrol.2024.130734).

# [6] Wijenayake, V., Stevenson, L. A., Takemoto, A., Ranjan, A., Mombauer, D., & Ismail, N. (Eds.). (2024). Linking Climate Change Adaptation, Disaster Risk Reduction, and Loss & Damage. Springer Nature Singapore. <https://doi.org/10.1007/978-981-99-8055-0>.

# [7] Ivascu, D. A., & Munteanu, M. (2024). Impact of a MATLAB-controlled application on the autonomy of ischemic stroke patients in the hospital. *IOP Conference Series: Materials Science and Engineering, 1320*, 012014. <https://doi.org/10.1088/1757-899X/1320/1/012014>

# [8] Das, S., Sheth, A. N., Bansal, P., Chuah, J., & Wasson, R. (2022). A statistical comparison of flood-related economic damage in Indian states with reflections on policy implications. *International Journal of Disaster Risk Reduction, 72*, 102835.<https://doi.org/10.1016/j.ijdrr.2022.102835>

# 

# [9] Mosavi, A., Ozturk, P., & Chau, K.-w. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, 10(11), 1536.<https://doi.org/10.3390/w10111536>

# [10] Singh, A., Sreeparvathy, V., Debdut, S., Pregnolato, M., & Wright, N. (2025). A critical review of flood risk assessment in Kerala Post-2018: Methodological approaches, gaps, and future directions. *Journal of Hydrology: Regional Studies*, 58, 102262. <https://doi.org/10.1016/j.ejrh.2025.102262>

# [11] K. T. Antwi-Agyakwa, M. K. Afenyo, and D. B. Angnuureng, "Know to Predict, Forecast to Warn: A Review of Flood Risk Prediction Tools," *Water*, vol. 15, no. 3, p. 427, Jan. 2023, doi: [10.3390/w15030427](https://doi.org/10.3390/w15030427).

# [12] N. J. S. and V. Mishra, "Projected increase in widespread riverine floods in India under a warming climate," *J. Hydrol.*, vol. 630, p. 130734, Feb. 2024, doi: [10.1016/j.jhydrol.2024.130734](https://doi.org/10.1016/j.jhydrol.2024.130734).

# [13] Mohanty, M. P., Mudgil, S., & Karmakar, S. (2020). Flood management in India: A focused review on the current status and future challenges. *International Journal of Disaster Risk Reduction*, 49, 101660. <https://doi.org/10.1016/j.ijdrr.2020.101660>

# [14] S. Samansiri, "The characteristics of an integrated flood warning and response system that can facilitate evidence-based decision making: A case study in Sri Lanka," Ph.D. dissertation, Univ. of Salford, United Kingdom, 2023. <https://www.proquest.com/dissertations-theses/characteristics-integrated-flood-warning-response/docview/31868996>

[15] H. Jain, "Leveraging geo-computational innovations for sustainable disaster management to enhance flood resilience," *Discover Geoscience*, vol. 2, art. no. 33, Jul. 2024. doi: <https://doi.org/10.1007/s44288-024-00042-0>

[16] D. T. Bui, P.-T. T. Ngo, T. D. Pham, A. Jaafari, N. Q. Minh, P. V. Hoa, and P. Samui, "A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping," *CATENA*, vol. 179, pp. 184–196, Aug. 2019. DOI: [10.1016/j.catena.2019.04.009](https://doi.org/10.1016/j.catena.2019.04.009).

[17] Guhathakurta, P., Sreejith, O. P., & Menon, P. A. (2011). Impact of climate change on extreme rainfall events and flood risk in India. *Journal of Earth System Science*, *120*(3), 359–373. <https://doi.org/10.1007/s12040-011-0082-5>

[18] Singh, O., Kumar, M. Flood events, fatalities and damages in India from 1978 to 2006. *Nat Hazards* 69, 1815–1834 (2013). <https://doi.org/10.1007/s11069-013-0781-0>

[19] Singh, H., M. Nielsen, and H. Greatrex, 2023: Causes, impacts, and mitigation strategies of urban pluvial floods in India: A systematic review. Int. J. Disaster Risk Reduct., 103751, [doi:10.1016/j.ijdrr.2023.103751](https://www.sciencedirect.com/science/article/pii/S2212420923002315).

[20] Mohapatra, P., & Singh, R. D. (2003). Flood management in India. *Natural Hazards*, *28*, 131–143. <https://doi.org/10.1023/A:1021178000374>

[21] Ray, K., Pandey, P., Pandey, C., Dimri, A. P., & Kishore, K. (2019). On the recent floods in India. *Current Science*, *117*(2), 204–218.<https://www.jstor.org/stable/27138236>

[22] U. Akbulut, M. A. Cifci, and Z. Aslan, "Hybrid Modeling for Stream Flow Estimation: Integrating Machine Learning and Federated Learning," *Applied Sciences*, vol. 13, no. 18, p. 10203, Sep. 2023. DOI: [10.3390/app131810203](https://doi.org/10.3390/app131810203).

[23] G. K. Wedajo, T. D. Lemma, T. Fufa, and P. Gamba, "Integrating Satellite Images and Machine Learning for Flood Prediction and Susceptibility Mapping for the Case of Amibara, Awash Basin, Ethiopia," *Remote Sensing*, vol. 16, no. 12, p. 2163, Jun. 2024. DOI: [10.3390/rs16122163](https://doi.org/10.3390/rs16122163).

[24] S. Das, G. Scaringi, Y. Ali, and A. C. Narayana, "Climate change and anthropogenic stresses on the sediment load variation in the Western Ghat Rivers," *EGU General Assembly 2024*, EGU24-4459, updated on Mar. 20, 2025. DOI: [10.5194/egusphere-egu24-4459](https://doi.org/10.5194/egusphere-egu24-4459).

[25] P. Koritelu, “Water Service Management in Island Communities: Analysis of Community Participation and Local Policies in Ay Island, Banda, Maluku Province,” *J. Manaj. Pelayanan Publik*, vol. 8, no. 1, pp. 1–15, Feb. 2024, doi: [10.24198/jmpp.v8i1.52886](http://dx.doi.org/10.24198/jmpp.v8i1.52886).

[26] T. Goitsemang, D. M. Das, S. K. Raul, C. R. Subudhi, and B. Panigrahi, “Assessment of Groundwater Potential in the Kalahandi District of Odisha (India) Using Remote Sensing, Geographic Information System and Analytical Hierarchy Process,” *J. Indian Soc. Remote Sens.*, vol. 48, pp. 1739–1753, Oct. 2020, doi: <https://doi.org/10.1007/s12524-020-01188-3>

[27] A. Dagur, K. Singh, P. S. Mehra, and D. K. Shukla, Eds., *Artificial Intelligence, Blockchain, Computing and Security, Volume 2: Proceedings of the International Conference on Artificial Intelligence, Blockchain, Computing and Security (ICABCS 2023), Gr. Noida, UP, India, 24–25 February 2023*. 1st ed. London: CRC Press, Dec. 1, 2023. DOI: [10.1201/9781032684994](https://doi.org/10.1201/9781032684994).

[28] M. S. Islam, Y. K. Ji, K. Kim, and H.-Y. Kim, “Advanced Machine Learning Techniques for Predicting Z-Axis Belt Wear in Wafer Transfer Robots,” *Int. J. Precis. Eng. Manuf.-Smart Technol.*, vol. 3, no. 1, pp. 17–30, Jan. 2025, doi: [10.57062/ijpem-st.2024.00157](https://doi.org/10.57062/ijpem-st.2024.00157).

[29] P. V. Mohanan, *Artificial Intelligence and Biological Sciences*. 1st ed. Boca Raton: CRC Press, 2025, pp. 1–468, doi: [10.1201/9781003492726](https://doi.org/10.1201/9781003492726).

[30] M. B. Almoujahed, A. K. Rangarajan, R. L. Whetton, D. Vincke, D. Eylenbosch, P. Vermeulen, and A. M. Mouazen, “Non-destructive detection of fusarium head blight in wheat kernels and flour using visible near-infrared and mid-infrared spectroscopy,” *Chemom. Intell. Lab. Syst.*, vol. 245, p. 105050, Feb. 2024, doi: [10.1016/j.chemolab.2023.105050](https://doi.org/10.1016/j.chemolab.2023.105050).

# Additionally, for government data references:

1. **Data.gov.in**, 2021. Year-wise statement showing damage due to floods and heavy rains in India (2012–2021). Government of India.  
   Available online: <https://www.data.gov.in/resource/year-wise-statement-showing-damage-due-floods-heavy-rains-india-2012-2021>
2. **Data.gov.in**, 2021. State/UT-wise total damages due to heavy rains and floods (2017–2021). Government of India.  
   Available online: <https://www.data.gov.in/resource/stateut-wise-total-damages-due-heavy-rains-and-floods-2017-2021>
3. **Data.gov.in**, 2021. Year-wise details of total flood damages caused due to floods and heavy rains: Houses, crops, and more. Government of India.  
   Available online: <https://www.data.gov.in/resource/year-wise-details-total-flood-damages-caused-due-floods-and-heavy-rains-houses-crops-and>
4. **Central Water Commission (CWC)**, 2022. Annual Flood Report. Government of India.
5. **India Meteorological Department (IMD)**, 2023. Climate of India. Government of India.
6. National Disaster Management Authority (NDMA), 2022. *Guidelines on Flood Management*. Government of India.

List of figures:

1. Figure 1: Human Fatalities in India's Top 10 Flood-Prone States in past years
2. Figure 2: Total Human Fatalities Caused by Floods Annually in India
3. Figure 3: Workflow of flood prediction and damage analysis in India
4. Figure 4: Population Affected by Floods Over the Years
5. Figure 5 and 6: Total Damages Over the Years and Distribution of total damages (Crops, Houses & Public Utilities)
6. Figure 7: Area Affected by Floods Over the Years
7. Figure 8: Number of flood occurrences per year (1970-2023)
8. Figure 9: Maximum area affected by floods in top 10 Indian states/UT
9. Figure 10: Number of flood occurrence by month
10. Figure 11: Total damage due to heavy rain in affected Indian States/UTs
11. Figure 12: Human and Cattle Loss Over the Years
12. Figure 13: Confusion matrix of random forest
13. Figure 14: ROC curve for random forest
14. Figure 15 and 16: Confusion matrix and ROC Curve for Support Vector Machine
15. Figure 17: Training and validation loss for ANN
16. Figure 18: Model accuracy comparison
17. Figure 19, 20 and 21: XG Boost Flood Prediction