

**Prediction Model for Flood-affected Areas in River Basin: A Case Study of the Attanagalu Oya River Basin, Sri Lanka.**

**By**

R. R. C. M. Rathnayake, R. D. M. L Weerasinghe, G. G. D. K Jayathilaka

Supervisor:Mr. Chamila Karunatilake

Senior Lecturer

Faculty of Technology

University Of Sri Jayewardenepura

Table of Contents

[Team Members 3](#_heading=h.l9jl2df7bxck)

[Introduction 4](#_heading=h.r1c5z0it1031)

[Background 5](#_heading=h.yltb6bozwhcz)

[Problem Statement 6](#_heading=h.y5o0zb34mohy)

[Research Questions 7](#_heading=h.s317ytyeq08s)

[Objectives 8](#_heading=h.auvmyxsgcnql)

[1.](#_heading=h.2lhj6g5933lt) General Objective 8

[2.](#_heading=h.b0rrb8jvq7pl) Specific Objectives 8

[2.1](#_heading=h.xknim3uf8sx) Data Collection and Preprocessing. 8

[2.2](#_heading=h.6hrvehrzkuud) Feature Identification and Selection. 8

[2.3](#_heading=h.xc5vc03biqb7) Model Implementation. 8

[Literature Review 9](#_heading=h.4waxdpv1j6ds)

[Methodology 15](#_heading=h.wef44xco4jyy)

[1.](#_heading=h.5lgfhyo9d6ya) Research Design 15

[2.](#_heading=h.dmoyz7lz8glt) Data Collection and Preprocessing 15

[2.1](#_heading=h.nizbnduvq49) Data Collection 15

[2.2](#_heading=h.7v45pivqlwuq) Data Preprocessing 15

[3](#_heading=h.1y8ky4mezn08) Feature Identification and Selection 16

[4](#_heading=h.w4bk858gwgp8) Model Design and Training 16

[5](#_heading=h.6fnc5a5iqshf) Model Validation 16

[Project Time Line 17](#_heading=h.wqhxpmafnhmm)

[References 18](#_heading=h.5b2gxrt275dc)

# Team Members

1. **R. R. C. M. Rathnayake**

University of Sri Jayewardenepura, Faculty of Technology, Information & Communication Technology Department, Software Technology Focus area.

071 104 3325

[ict20916@fot.sjp.ac.lk](mailto:ict20916@fot.sjp.ac.lk)

1. **R. D. M. L Weerasinghe**

University of Sri Jayewardenepura, Faculty of Technology, Information & Communication Technology Department, Network Technology Focus area.

071 362 1617

[ict20957@fot.sjp.ac.lk](mailto:ict20957@fot.sjp.ac.lk)

1. **G. G. D. K Jayathilaka**

University of Sri Jayewardenepura, Faculty of Technology, Information & Communication Department, Software Technology Focus area.

071 1426725

[ict20859@fot.sjp.ac.lk](mailto:ict20859@fot.sjp.ac.lk)

**Supervisor:**

Mr. Chamila Karunatilake

Senior Lecturer

Faculty of Technology, University of Sri Jayewardenepura

Email: [chamilakarunathilake@sjp.ac.lk](mailto:chamilakarunathilake@sjp.ac.lk)

**Co-Supervisor:**

# Introduction

Environmental circumstances that may have the capacity to affect human environments and societies are referred to as natural hazards. Natural or technical disasters are defined as natural phenomena that went awry and led to the hardware, infrastructure, and human losses. For instance, calamities include destruction in occurrences such as earthquake, landslides or other related attributes occurring on or beneath the surface of the earth. A combination of several different forces may cause some phenomena that produce natural disasters. They tend to occur repeatedly in the same geographical locations because they are related to weather patterns or physical characteristics of an area. Disaster management is a strategic and multi-faceted procedure for mitigation, preparedness, response, and recovery to protect the vulnerable community and critical intrastate from any disaster. According to the terminology of the United Nations Office for Disaster Risk Reduction (UNISDR), a disaster is a “serious disruption of the functioning of a community or a society involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its resources Floods are the deadliest natural hazards, striking numerous regions in the world each year. Increasing population pressure, degradation of ecosystems, and climate variability and change contribute to a further increase in flood risks worldwide. Floods have the potential to wreak havoc across large areas, killing people and damaging private property as well as vital public health facilities.

We are not defenceless against disasters: deaths from disasters have fallen significantly over the last century as a result of early warning systems, better infrastructure, more productive agriculture, and coordinated responses. As climate change increases the risks of more extreme events, making societies even more resilient will be crucial to prevent our recent progress from reversing. In order to address the above issues, it is essential to explore how changes in disaster events occur, which individuals are at the highest risk, and how it is possible to promote their safety. Projects undertaken for disaster management have become global issues as they endanger the lives of people and put buildings and even economies at risk. Different aspects such as population growth, urban development, climatic changes, heavy investments in disaster perilous zones, and destruction of the environment are some of the reasons as to why there is a high rate of occurrence of natural calamities. Disaster risk management is the design and implementation of policy and strategy for prevention, preparedness, response, recovery, and rehabilitation from natural hazards, climate change, conflict, or any other emergency or disaster.

# Background

Attanagalu Oya basin Located in the wet zone of Sri Lanka between the longitudes 79o 50’E and 80o 07'E and latitudes 60o 59’N and 70o 17’N (Perera, 2012), covering an area of 727 km2 (Chathurani, 2022) It has four streams, Attanagalu Oya, Diyaeli Oya, Kimbulapitiya Oya, and Uruwal Oya, which discharge into the Negombo lagoon as Dandagamuwa Oya (Wijesekara, n.d.). Attanagalu Oya flows across Gampaha, Katunayake, and Negombo, and the longest length of Oya is about 76 km. (Anuruddhika, 2022)

It experiences both the southwest monsoon (May to September) and the northeast monsoon (November to February), with annual rainfall ranging from 1400 to 2500 mm. Peak rainfall occurs from October to November and again in May to June. Therefore, every year has to be faced with flooding hazards in this area. (Chathurani, 2022) The lower basin, which contains flat terrain and inadequate drainage systems, is particularly exposed to seasonal and flash floods.

As well, the basin is densely populated, and it provides homes for about 5% of Sri Lanka's total population. (Chathurani, 2022) Due to rapid urbanization in the region, surface runoff has increased, and natural drainage has reduced. It may cause exacerbating flood risks.

Giving these challenges, predicting flood-prone areas in the Attanagalu Oya basin is important to implement an early flood warning system, disaster preparedness, sustainable water resource management, and urban planning. An AI-based system can provide an important observation to forecast floods and identify flood inundation areas by analyzing meteorological, hydrological, and topographical data. (Mosavi, 2018).

# Problem Statement

This is truly one of the most devastating natural disasters in Sri Lanka, and it causes the most damage to lives, properties, and infrastructures. Floods have only intensified and multiplied over the past few years, especially in the Northern and Eastern regions. In recent years, the frequency and severity of floods have increased, but the Northern and Eastern regions have been particularly hard-hit. For instance, late November 2024 saw widespread flooding in these areas, with the resultant loss of life, displacement, and massive damage to homes and public infrastructure. These recurrent disasters call for effective flood risk mitigation measures.

Among those at the damaged destination is the Gampaha district in the Western Province, which in fact has a very high susceptibility to flooding. The Attanagalu River, a prominent river in the area, is responsible for flash floods, especially in urbanized regions. The ACAPS Anticipatory Briefing Note established that floods between May and June 2024 left about 27,000 people affected in the Gampaha district, affecting an area of nearly 124 square kilometers. An uncommonly severe flooding incident occurred in October 2024; about 130,000 people in total were affected, of which three lost their lives and two suffered injuries. Historical data further reveals that floods in Sri Lanka are most frequent in May and during some parts of December, affecting districts, among others, Jaffna, Kalutara, Ratnapura, Gampaha, and Ampara. Among these, Gampaha remains one of the most flood-prone areas and highly affected regions.

Though existing flood prediction and management models have shown a lot of promises, still, there are quite a few lacunae. For instance, hydrological models like those in the Attanagalu Oya basin conduct flood simulations but are pretty rudimentary because of their static methodologies. Just in the same manner, the technologies like ultrasonic sensors and real-time image processing have been studied for flood monitoring and detection; These models, though, do not take into account several other important parameters like the intensity of rainfall, river levels, and even geographical slopes to predict accurately flood-affected zones.

This research proposes an AI-based predictive model by which the flood-affected areas can be identified as per main parameters such as rainfall, river levels, and topographical slopes. The model was developed using machine learning advances in historical and real-time data that would present an accurate, specific location prediction for better disaster management and preparedness.

This research intends to focus on disaster risk reduction in terms of the geographic area of the Gampaha district, thus greatly focusing on highly vulnerable communities. The results concerning the creation and validation of this model in this area will serve as a guiding parameter for extending the methodology to other flood-prone districts in the country, leading to a nationwide reduction of flood disaster impacts.

# Research Questions

1. What historical flood-related data is available for the Gampaha District, and how can it be collected effectively?
2. What preprocessing techniques are required to clean, normalize, and prepare the dataset for machine learning?
3. Which machine learning algorithms are best suited for flood prediction tasks?
4. What steps are required to implement the selected ML model?
5. What metrics should be used to evaluate the performance of the model?

# Objectives

## General Objective

Develop an AI model to predict flood-prone areas in Gampaha District based on the Attanagalu Oya basin.

* Use machine learning techniques to identify and predict areas prone to flooding based on rainfall, slope, elevation, land use, geology, river density, distance to river, and rainfall intensity.

## Specific Objectives

### Data Collection and Preprocessing.

* Gather historical flood data, rainfall patterns, topographic maps, river flow data, and land use information.
* Clean and preprocess the data

### Feature Identification and Selection.

* Determine critical factors influencing flooding such as rainfall, slope, land use, elevation, rainfall intensity, and distance to the river.
* Select features to improve the accuracy of the AI model.

### Model Implementation.

* Build and train predictive models using suitable algorithms.
* Optimize the model for better prediction accuracy and efficiency.
* Evaluate the model’s performance using cross-validation techniques.
* Validate results against actual flood events.

# Literature Review

Flood Prediction Using Rainfall-Flow Pattern in Data-Sparse Watersheds research proposed the multifeatured algorithm (spatial-temporal dynamic time warping (ST-DTW)) for real-time forecasting using a rainfall-flow pattern. This extends one-dimensional features to multi-dimensional features for rainfall-flow pattern matching. It has a rainfall-flow pattern matching method. to construct rainfall-flow patterns they used hydrological data from wet and dry watersheds. Using original flood data, which includes historical flow and rainfall data, then preprocessing the data they constructed rainfall-flow patterns. The experimental results based on the datasets of various models in wet and drought-ridden watersheds show that the proposed models offer considerable advantages in accurately predicting the peak time of floods in real-time. (YUELONG ZHU, March 4, 2020). Flood prediction using rainfall – runoff spatial variation: an overview of flood prediction models Research about flood forecasting with rainfall-runoff models underscores the rise in rainfall occurrences in cities such as Port Harcourt in Nigeria. The examination consolidates discoveries from different research works and underscores the challenges of precisely anticipating floods caused by the unpredictable patterns of rainfall and runoff mechanisms. The study highlighted the consequence of methods for predicting floods using time series analysis and advanced modeling tools like Artificial Neural Networks (ANN) which greatly improve forecast precision when compared to conventional approaches, & the research addresses the complications that emerge from the lack of historical hydrological records in the Niger Delta region, where numerous waterways have not been monitored. (ORUPABO, April 21, 2015 ). Another research study offers an extensive analysis of coastal flooding risks in South Korea, using some ML algorithms to predict coastal flooding risk under climate change. used multiple algorithms, including k-Nearest Neighbor (kNN), Random Forest (RF), and Support Vector Machine (SVM), and developed a risk probability map that quantifies flooding risks, discloses that southern coastal regions face higher threats compared to eastern and western areas. A novel aspect of this study is its probabilistic framework, which also incorporates rainfall data – usually excluded from past studies – and uses it together with tidal elevations for improved future hazard assessments. Results suggested that the averaged receiver operating characteristic (ROC) scores for the KNN algorithm came out approximately equal to 0.946, proving its effectiveness in risk prediction. (Lee2, 2020). Saravi et al. (2019) explore the integration of AI to improve resilience and preparedness against flooding events, emphasizing the importance of historical data in understanding flood behaviours. Their research employs various ML models, including Random Forest and J48 decision trees, to classify different types of floods based on weather forecasts and historical data from 1994 to 2018. Moreover, the authors highlight that while existing literature predominantly focuses on social media data for flood modeling, there is a growing need for comprehensive methods that incorporate diverse data types. By leveraging AI, stakeholders can better allocate resources, ultimately improving community resilience against natural disasters. (Sara Saravi 1, 2019). The research presents different ML models, including the random forest algorithm, which was the most efficient in classifying flood events with the aid of weather forecasts and historical data patterns. Also, the authors clearly indicate the importance of adding different types of data, such as social networks and imagery, to get an insight into the complex nature of floods. They use this flood data aggregation tool (FDAT) for collecting flood data, sorting it, and preparing it systematically and efficiently. (Duminda Perera a, 2020) Flood prediction using rainfall – runoff spatial variation: an overview of flood prediction models Research about flood forecasting with rainfall-runoff models underscores the rise in rainfall occurrences in cities such as Port Harcourt in Nigeria. The examination consolidates discoveries from different research works and underscores the challenges of precisely anticipating floods caused by the unpredictable patterns of rainfall and runoff mechanisms.   
The study highlighted the consequence of methods for predicting floods using time series analysis and advanced modeling tools like Artificial Neural Networks (ANN), which greatly improve forecast precision when compared to conventional approaches, & the research addresses the complications that emerge from the lack of historical hydrological records in the Niger Delta region, where numerous waterways have not been monitored. (ORUPABO, April 21, 2015 ).

Another research study offers an extensive analysis of coastal flooding risks in South Korea, using some ML algorithms to predict coastal flooding risk under climate change. used multiple algorithms, including k-Nearest Neighbor (kNN), Random Forest (RF), and Support Vector Machine (SVM), and developed a risk probability map that quantifies flooding risks discloses that southern coastal regions face higher threats compared to eastern and western areas. A novel aspect of this study is its probabilistic framework, which also incorporates rainfall data—usually excluded from past studies—and uses it together with tidal elevations for improved future hazard assessments. Results suggested that the averaged receiver operating characteristic (ROC) scores for the KNN algorithm came out approximately equal to 0.946, proving its effectiveness in risk prediction. (Lee2, 2020). Saravi et al. (2019) explore the integration of AI to improve resilience and preparedness against flooding events, emphasizing the importance of historical data in understanding flood behaviours. Their research employs various ML models, including Random Forest and J48 decision trees, to classify different types of floods based on weather forecasts and historical data from 1994 to 2018. Moreover, the authors highlight that while existing literature predominantly focuses on social media data for flood modeling, there is a growing need for comprehensive methods that incorporate diverse data types. By leveraging AI, stakeholders can better allocate resources, ultimately improving community resilience against natural disasters. (Sara Saravi 1, 2019). The research presents different ML models, including the random forest algorithm, which was the most efficient in classifying flood events with the aid of weather forecasts and historical data patterns. Also, the authors clearly indicate the importance of adding different types of data, such as social networks and imagery, to get an insight into the complex nature of floods. They use this flood data aggregation tool (FDAT) for collecting flood data, sorting it, and preparing it systematically and efficiently. (Duminda Perera a, 2020) Another research study offers an extensive analysis of coastal flooding risks in South Korea, using some ML algorithms to predict coastal flooding risk under climate change. used multiple algorithms, including k-Nearest Neighbour (kNN), Random Forest (RF), and Support Vector Machine (SVM), and developed a risk probability map that quantifies flooding risks and discloses that southern coastal regions face higher threats compared to eastern and western areas. A novel aspect of this study is its probabilistic framework, which also incorporates rainfall data—usually excluded from past studies—and uses it together with tidal elevations for improved future hazard assessments. Results suggested that the averaged receiver operating characteristic (ROC) scores for the KNN algorithm came out approximately equal to 0.946, proving its effectiveness in risk prediction. (Lee2, 2020). Saravi et al. (2019) explore the integration of AI to improve resilience and preparedness against flooding events, emphasizing the importance of historical data in understanding flood behaviors. Their research employs various ML models, including Random Forest and J48 decision trees, to classify different types of floods based on weather forecasts and historical data from 1994 to 2018. Moreover, the authors highlight that while existing literature predominantly focuses on social media data for flood modeling, there is a growing need for comprehensive methods that incorporate diverse data types. By leveraging AI, stakeholders can better allocate resources, ultimately improving community resilience against natural disasters. (Sara Saravi 1, 2019). The research presents different ML models including the random forest algorithm, which was the most efficient in classifying flood events with the aid of weather forecasts and historical data patterns. Also, the authors clearly indicate the importance of adding different types of data such as social networks and imagery to get an insight into the complex nature of floods. They use this flood data aggregation tool (FDAT) for collecting flood data, sorting it, and preparing it systematically and efficiently. (Duminda Perera a, 2020)

The study demonstrated a predictive model for flood-prone areas using logistic regression and Geographic Information Systems (GIS). Logistic regression analysis is used to find the relationship between categorical response variables with a nominal or ordinal scale and one or more continuous and categorical explanatory variables. The model has used environmental factors like rainfall, elevation, slope, land use, flow accumulation, and proximity to rivers to assess and predict flooding. To identify key factors for flood events, they have used historical flood data, satellite-based measurements, and geographical parameters The results mention that regions with low elevation and gentle slopes are more vulnerable to flooding (Hidayat Jati, 2019). Salvati et al. (2023) has been done to create flood susceptibility maps for the Haraz watershed in northern Iran using machine learning. The models, including support vector regression, linear kernel, base classifier, and hyper-parameter optimization, have been tested to identify flood-prone areas. Ten effective factors (slope, elevation, curvature, rainfall, stream power index [SPI], topographic wetness index [TWI], distance to river, river density, land use, and geology) have been selected for flood zoning, and three of the factors were most important for predicting flood-sensitive areas: slope, distance to the river, and river (Salvati, 2023). Another study describes a model for evaluating flood risk in Kuala Krai, Kelantan, Malaysia, utilizing Bayesian Networks and machine learning algorithms like decision trees, kNN, and support vector machines. The performance of the four models (BN, DT, KNN, and SVM) has been compared using both normal data and SMOTE-balanced data. BN has performed well with an accuracy of 99.94% with the unbalanced data set compared to others. However, after using the Synthetic Minority Oversampling Technique (SMOTE), DT, KNN, and SVM also got high accuracy. So, this research shows that the SMOTE method is highly useful in combating with imbalanced dataset (Razali, 2020). Anuruddhika et al, (2022) explain a method for forecasting flood-prone areas in the Attanagalu Oya basin, Sri Lanka, to simulate water flow and analyze flood inundation areas, they used the Hydrologic Engineering Centre-Hydrologic Modeling System (HEC-HMS) and the Hydrologic Engineering Centre-River Analysis System (HEC-RAS). Hydrologic Engineering Centre (HEC) models are one of the computer-based software programs that is widely used for many hydrological and hydraulic simulations such as flooding, flood frequency, stream restoration, etc. The additive Holt-Winters' method has been used for rainfall and evaporation forecasting. This research showed a 72% similarity between the simulated flood extent and actual flood extent, while showing a 68% similarity between the forecasted flood extent and actual flood extent (Anuruddhika, 2022). Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors research employs GIS to map flood susceptibility and logistic regression to model flood probability based on environment factors like slope, flow accumulation DEM and rainfall. ArcGIS and EViews were used for data analysis. While the model identifies flood-prone areas, it lacks site-specific details that can be critical for small-scale infrastructure planning. (Al-Juaidi, 2018).

Real-Time Flood Detection with Image Processing - designed a real-time flood monitoring and alarm system integrating image processing and sensors like rain gauges and flow meters. The system captures images at regular intervals, processes them to analyze flood levels, and sends updates via an Android application. With a prediction accuracy of 91.81%, it offers a cost-efficient solution for rural flood management. Powered by solar energy, this system minimizes environmental impact and operating costs, addressing the needs of resource-constrained areas. The system uses multiple linear regression as the mathematical model (Karlo Tolentino, 2022) AI-Enhanced Flood Monitoring and Rescue System introduced a novel approach combining IoT for flood monitoring and AI for rescue operations in western Maharashtra, India. The system detects flood levels using BMP180 sensors and coordinates rescue efforts via drones equipped with YOLO object detection algorithms. Real-time updates are provided to authorities through a web portal and mobile application. This solution effectively bridges flood detection and emergency response, demonstrating how technology can be leveraged to reduce disaster-related fatalities and damages (Pathan, 2020) Flood Monitoring and Early Warning Using Ultrasonic Sensors - proposed a flood monitoring system utilizing ultrasonic sensors, Arduino, and GSM modules to detect and alert authorities and residents about rising water levels. This system is tailored for communities in the northern Philippines near the Cagayan River. In addition to automated alerts, it features an SMS inquiry system enabling individuals to check water levels remotely. By addressing challenges of manual monitoring methods, the study highlights the importance of technology in improving preparedness in flood-prone areas (Natividad, 2018) Detection with Machine Learning and Deep Features introduces an innovative approach for disaster detection, focusing on water disasters among other disaster types, using machine learning (ML) and deep learning techniques. The study utilizes datasets from social media and other sources to classify disaster types into nine categories, including water disasters. Convolutional Neural Networks (CNNs) such as SqueezeNet, VGG16, and VGG19 were employed for feature extraction, and machine learning algorithms like Neural Networks (NN), Random Forests (RF), and Logistic Regression (LR) were applied for classification. Notably, the study achieved a classification accuracy of 90.2% with NN using features extracted by SqueezeNet, demonstrating the potential of ML models for rapid and reliable disaster type identification.

The work emphasizes the utility of real-time image data from social media during crises, highlighting the importance of accurate classification for effective disaster response. Water disaster images were categorized alongside other disaster types, with confusion matrices revealing common misclassifications, such as overlaps between water disasters and infrastructure damages. The study underscores the relevance of ML in disaster management, suggesting improvements in response times and resource allocation (Cinar, n.d.) Disaster Detection Using Machine Learning and Bluetooth employed CNNs to analyze disaster images, achieving a detection accuracy of 94.3% after fine-tuning. They also proposed a Bluetooth-based scatternet communication system to facilitate connectivity in disaster-hit areas, ensuring data transfer and coordination among rescue teams even without internet access. This approach is relevant for water-related disasters, providing robust detection and communication solutions (Gupta & Kumar Rana, n.d.) Use of AI for Flood Resilience and Preparedness explored the application of machine learning (ML) for classifying flood types, including flash floods and coastal floods, based on historical data. Random Forest emerged as the best-performing model, achieving 80.49% accuracy. This research demonstrates the potential of ML to predict flood types and improve disaster preparedness. The study also highlights the role of public awareness and emergency responses in mitigating flood impacts and building community resilience. (Saravi, 2019)

# 

# 

# 

# 

# 

# Methodology

## Research Design

The proposed study plan is to develop a flood-affected area prediction model to predict the affected areas in Gampaha near the Attanagalu Oya Basin, considering its frequent flooding impact on local communities. The ML model will analyze several parameters like rainfall, area slope, rainfall intensity, elevation, and the river water level to predict the affected areas. Major phases of the work include data collection, feature identification and selection, development of a machine learning model for predicting the flood, and validation.

## Data Collection and Preprocessing

### Data Collection

In this phase we gather comprehensive and valid data in the Gampaha area and planned to collect this data from the Irrigation Department, Meteorology Department, and Disaster Management Centre in Sri Lanka.

Data about the flooding in the Gampaha area,

* Rainfall: Historical and real-time rainfall data for the Gampaha area.
* Attanagalu Oya water level: Measurements of water level to detect the effect that water level has on flooding.
* Slope of the affected areas.
* Records of flood history, distribution, and severity.

### Data Preprocessing

The collected data will be prepared for analysis, handling missing values and normalization of continuous variables like rainfall and elevation.

## Feature Identification and Selection

This phase will identify critical factors influencing flooding, such as rainfall, slope, land use, elevation, rainfall intensity, river water level, and distance to the river, based on literature and domain knowledge. Using statistical techniques like correlation analysis, the relationship between features and flood occurrence can be determined.

## Model Design and Training

In this phase, we focus on designing and implementing the ML model. We identified some models that are suitable for our model. So, we are reviewing these existing models and identifying the best model.

Ex: Decision Tree Algorithm, Support Vector Machine, Random Forest.

Develop the new ML model tailored to the unique parameters of the Gampaha District Attanagalu Oya basin. After identifying the most critical factors influencing flooding in this area, this model will use those factors as primary input features.

After the preprocessing, use 70-80% of that processed data to train the new ML model.

## Model Validation

It will be tested with different sets of data (20-30% of the data) to ensure the flood-prone areas. We will then verify its performance by measuring its metrics like accuracy, precession and recall. We will then also compare model prediction with historical flood maps to assess spatial accuracy.

# Project Time Line

|  | Sep | Oct | Nov | Dec | Jan | Feb | Mar | | April |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Planning & preparation |  |  |  |  |  |  |  | |  |
| Data Collection |  |  |  |  |  |  |  | |  |
| Data Preprocessing |  |  |  |  |  |  |  | |  |
| Feature Identification |  |  |  |  |  |  |  | |  |
| Model Design & Training |  |  |  |  |  |  |  |  |  |
| Validation & Testing |  |  |  |  |  |  |  |  |  |
| Report & Dissemination |  |  |  |  |  |  |  | |  |

# References

Al-Juaidi, A. E. ,. N. A. M. ,. A.-J. O. E., 2018. Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arabian Journal of Geosciences,* 11(24).

Anuruddhika, M. L. P. ,. P. K. K. K. R. ,. P. L. P. N. D. ,. H. W. P. T. ,. W. V. P. A., 2022. Forecasting flood inundation areas of Attanagalu Oya. *Ceylon Journal of Science,* 51(3), p. 217.

Chathurani, K. L. N. A. H. A. S. K. B. A. S. P. M. U. I. G. A. A. N., 2022. Case Study on Identification of Flood Hazard in the Lower Catchment Area of the Attanagalu Oya River Basin. *Journal of Geoscience and Environment Protection,* 10(7), pp. 305-318.

Duminda Perera a, b., 2020. Identifying societal challenges in flood early warning systems. *ELSEVIER,* Issue 15 August 2020.

Hidayat Jati, M. I. S. S. P. B., 2019. *Prediction of flood areas using the logistic regression method (case study of the provinces Banten, DKI Jakarta, and West Java).* s.l., Institute of Physics Publishing.

Karlo Tolentino, L. S. B. R. E. A. B. C. C. M. A. J. D. C. D. G. D. E. B. F. J. A. V. R. C. F. Q. J. C. V. P. M. C. A. M. G. M. V., 2022. Real Time Flood Detection, Alarm and Monitoring System Using Image Processing and Multiple Linear Regression. *Journal of Computational Innovations and Engineering Applications,* pp. 12-23.

Lee2, S.-J. P. a. D.-K., 2020. Prediction of coastal flooding risk under climate change impacts. Issue 25 August 2020.

Mosavi, A. O. P. C. K. W., 2018. *Flood prediction using machine learning models: Literature review,* s.l.: MDPI AG.

Natividad, J. G. M. J. M., 2018. *Flood Monitoring and Early Warning System Using Ultrasonic Sensor.* s.l., Institute of Physics Publishing.

ORUPABO, S. H. L. A. O. A., April 21, 2015 . FLOOD PREDICTION USING RAINFALL – RUNOFF SPATIAL VARIATION: AN. *Ethiopian Journal of Environmental Studies and Management.*

Pathan, A. I. K. A. P. G. N. L. P. P. M. S. A. R., 2020. *An IoT and AI based Flood Monitoring and,* s.l.: International Journal of Engineering Research & Technology (IJERT).

Perera, K. R. J. ,. W. N. T. S., 2012. Potential on the Use of GIS Watershed Modeling for River Basin Planning \_ Case Study of Attanagalu Oya Basin, Sri Lanka.

Razali, N. I. S. M. A., 2020. Machine learning approach for flood risks prediction. *IAES International Journal of Artificial Intelligence,* 9(1), pp. 73-80.

Salvati, A. N. A. M. S. A. G. K. A. D. T. A.-A. N. S. F. C. J. J., 2023. Flood susceptibility mapping using support vector regression and hyper-parameter optimization. *Journal of Flood Risk Management,* 16(4), p. 18.

Sara Saravi 1, \*. ,. R. K. 1., 2019. Use of Artificial Intelligence to Improve Resilience. Issue : 9 May 2019.

Saravi, S. K. R. J. D. C. M. R. F. G. M. F., 2019. Use of artificial intelligence to improve resilience and preparedness against adverse flood events. *Water (Switzerland),* 11(5).

Wijesekara, R. S. ,. K. C., n.d. *Preliminary Groundwater Assessment and Water Quality Study in the Shallow Aquifer System in the Attanagalu Oya,* s.l.: s.n.

YUELONG ZHU, J. F. ,. L. Y. ,. T. G. ,. A. X. L., March 4, 2020. Flood Prediction Using Rainfall-Flow Pattern in.

Cinar, I. (n.d.). *Disaster Detection Using Machine Learning Methods with Deep Features DisasterrDetectionnUsinggMachineeLearninggMethodss withhDeeppFeatures IlkayyCINAR*. <https://www.researchgate.net/publication/370599940>

Gupta, N., & Kumar Rana, K. (n.d.). *Disaster Prediction And Post Disaster Management Using Machine Learning And Bluetooth* (Vol. 18, Issue 5). <http://www.webology.org>