Imperial College London

Department of Earth Science and Engineering

MSc in Environmental Data Science and Machine Learning

Independent Research Project
Project Plan

Utilizing Geospatial Data and Machine Learning Algorithms for Constructing an Analytical Tool to Evaluate Urban Flood Risk

by

Chaofan Wu

Email: chaofan.wu22@imperial.ac.uk
GitHub username: edsml-cw522
Repository: https://github.com/ese-msc-2022/irp-cw522

Supervisors:

Sesinam Dagadu(MEng)

Dr. Yves Plancherel

Company'name: SnooCODE
Company's address: 18 Royal Palm Avenue, McCarthy Hill,
Accra-Ghana

June 2023

1 Abstract

This project focuses on developing an analytical tool to evaluate urban flood risk in Accra, Ghana. The tool is designed to provide valuable insights for urban planners, emergency response teams and residents to mitigate and management the frequent and severe flooding problems in Accra. This Accra urban flood risk analytical tool utilises a combination of geographic spatial data, machine learning models (ML), and geographic information systems (GIS) to generate a comprehensive visualisation of the entire city's flood risk. The tool may also consider different land use and climatic scenarios, providing potential strategies and inspirations for flood prevention. Recently, advanced flood modelling techniques, including one/two-dimensional shallow water equation models and ML-based models, each have their own advantages and drawbacks. For project management and time allocation, a whole problem has first been split into 7 sub-objectives to be achieved and a detailed flowchart has been created for the most basic 4 of these sub-objective to describe the development process, then a detailed Gantt chart has been created based on this detailed flowchart and all the sub-objectives outlining the future plans. To date, project progress includes a comprehensive literature review and preliminary selection of model development approaches and user interface (UI) design techniques. The potential data sources have been identified, including Google Earth Engine (GEE) and other satellite map data providers.

2 Introduction

2.1 Problem Description

The present study utilises a variety of geospatial data, including flood maps and topographic data, to develop an analytic tool for evaluating urban flood risk in Accra, a major city in Ghana. This analytic tool is expected to provide a comprehensive depiction of the distribution of flood risk throughout the city. Additionally, this analytic tool might be employed to investigate the impact of diverse land use and climate scenarios on flood risk, where appropriate. Furthermore, this analytic tool possibly has the potential to provide recommendations and adaptation flood prevention strategies. These flood prevention strategies probably includes the implementation of green infrastructure and the construction of flood walls to aid users in the mitigation of flood risk.

2.2 Literature Review

In the rainy season between May and July, seasonal floods occur almost every year in the capital city of Accra, Ghana, posing a serious threat to human life and property [1]. The city has also experienced some of the most severe flood events in its history, such as those in 1955, 1960, 1963, 1973, 1986, 1991, 1995, and 2001, resulting in significant damage to infrastructure, industrial areas, and residential areas [2, 3]. Therefore, assessing the urban flood risk in the Accra region is a necessary and meaningful task. Reports indicate that the city's flooding is mainly caused by three factors: 1. climate factors (including large-scale rainfall, especially during the wet season from May to July); 2. geographical factors (including the low altitude of the city and the clay nature of the soil); 3. human factors (including insufficient and inadequate drainage systems, dumping garbage into drainage channels and water bodies, and unreasonable development and use of environmentally sensitive areas)[1]. If there is a method that can visualise and evaluate flood-prone areas in Accra on a map based on these three factors, it will help urban planners and managers to develop measures to manage the recurring flood problem [1]. In addition, it will also help emergency teams to identify areas for concentrated action during flood events, thereby reducing future damage and risks to residents and mitigating the long-standing flood problem in Accra[1].

Currently, one-dimensional(1D) shallow water equation models(1D-SWEs), two-dimensional shallow water equation models(2D-SWEs), and machine learning-based models(ML) are the mainstream methods for modelling and evaluating flood risk [4]. They all have their own advantages and disadvantages. For example, the 1D-SWEs was initially used to simulate floods. It is relatively intuitive and easy to calculate, but its performance is poor for complex urban flood environments (such as multiple river intersections or obstacles such as bridges and dams) [4]. This is because the model assumes that the velocity of water flow is uniform across the entire cross section and simplifies the terrain into a one-dimensional line [4]. The 2D-SWEs was optimised based on the 1D-SWEs, improving the limitations of the 1D-SWEs and better reflecting the complex situations in urban flood environments, such as nonlinear water flow paths [4]. In addition, the 2D-SWEs can now use high-precision digital terrain models (DTMs) obtained from Light Detection and Ranging(LIDAR), which can clearly represent small-scale structural elements and small terrain changes [4]. Furthermore, the use of high-performance computing (HPC) technology can improve the computational efficiency of the 2D-SWEs [4]. Although the computational burden of 2D-SWEs can be reduced by using coarser grid and sub-grid models, these models are still relatively heavy to compute [4]. ML-based models are very effective when quick results are needed or when the model needs to be run with input parameters that change multiple times [4]. The input parameters can also come from different dimensions, such as climate, geography, and human factors, and can be simultaneously input into the model for training. When ML models are properly calibrated and tuned, ML method can produce prediction results comparable to the physical models [4, 5].

ML models have advantages in prediction accuracy and computational efficiency [4]. However, this method requires a large amount of data for training, and may lack the theoretical support and interpretation that physical models possess. Therefore, choosing to use machine learning as a modeling method for this project will mean that the collection of data in the early stages and the interpretation of the model in the later stages will need special attention. Methods for modelling and evaluating flood risk based on ML have become a mainstream direction in recent years, and these methods are often using Geographic Information Systems (GIS) to extract geographic information (such as slope, distance from rivers and altitude, etc.) from geospatial Data (such as satellite maps) [6, 7]. For those who wish to predict the impact of different spatial geographical locations on flood risk, ML based approaches have been proven to be effective and accurate multiple times [1, 6, 7].

2.3 Objectives

Based on the description of the problem, this project is broken down into a total of seven goals to be completed have been extracted, of which four are mandatory goals, while the remaining three are optional (see below a list of 7 goals). In future implementation, the four mandatory goals will be given priority.

1. (Mandatory) A back-end machine learning model shall be developed to predict the distribution of flood risk in Accra.
2. (Mandatory) A front-end analytic and interactive tool shall be developed to visualise the prediction results from back-end model.
3. (Mandatory) This analytic tool allows user interaction, such as displayed flood risk map zooming in and out.
4. (Mandatory) This analytic tool allows users to display different information in a flood risk map, such as altitude, population density, etc.

5. (Optional) This analytic tool may allow users to select different back-end model results such as Logistic Regression, SVM, KNN, XGBoost, etc. (note: this needs to be done, if several back-end machine learning models are implemented successfully).
6. (Optional) This analytic tool might give the user a results of how distinct land use and climate scenarios are affecting the flood risk.
7. (Optional) This analytic tool might give the user some flood prevention strategies, such as the implementation of green infrastructure and the construction of flood walls.

In regards to the four mandatory objectives that require completion, it is plausible that a detailed development plan could potentially facilitate a more streamlined and straightforward approach. Therefore, in Figure 1 below, a detailed flowchart designed to develop the aforementioned four mandatory objectives is created. In this flowchart(Figure 1), each objective is further divided into several smaller steps to be completed.

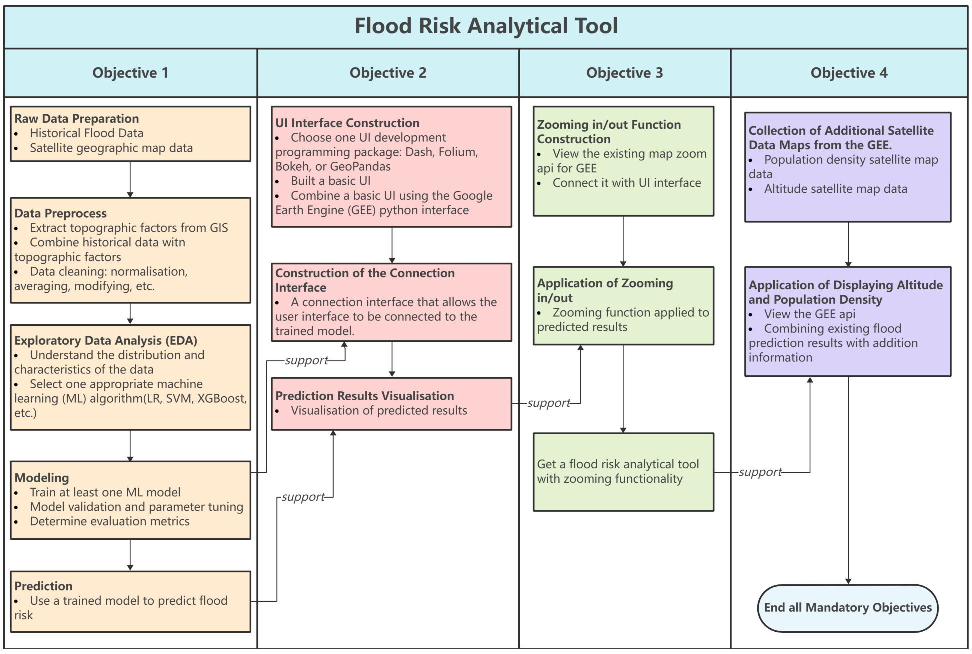

Figure 1: The flowchart of development 4 mandatory objectives

3 Progress to Date

To date, extensive literature review and study of similar application platforms (such as the Global Flood Database, a platform that visualises historical flood data by searching any country names of the world[8]) have been conducted. It is now possible to determine that the model portion of the project will be built using GIS techniques combined with ML algorithms. And the choice of User Interface(UI) development tools could be Dash, Folium, Bokeh, or GeoPandas, all of which are excellent interactive visualisation tools for construction in Python. Acquisition of map data may be done by using Google Earth Engine (GEE), which provides a wealth of satellite map data and image processing capabilities [8, 9]. The GEE also provides third-party Python packages such as earthengine-api for interaction, making it an excellent choice [9].

4 Future Plan

In the subsequent development, in order to better allocate time and manage project processes, the following Gantt chart (Figure 2) has been selected as a display tool for project management.

ID	Task name	Start	Finish	Duration	Complete
1	Raw Data preparation	19/6/2023	30/6/2023	10.0 d.	0%
2	GIS study	19/6/2023	30/6/2023	10.0 d.	0%
3	UI interface study	19/6/2023	7/7/2023	15.0 d.	0%
4	Data preprocess	3/7/2023	7/7/2023	5.0 d.	0%
5	EDA + Modeling	10/7/2023	21/7/2023	10.0 d.	0%
6	UI interface construction	14/7/2023	27/7/2023	10.0 d.	0%
7	Connection interface	27/7/2023	4/8/2023	7.0 d.	0%
8	Visualise the prediction results	4/8/2023	17/8/2023	10.0 d.	0%
9	Zooming in/out function construction	4/8/2023	17/8/2023	10.0 d.	0%
10	Collect addtional satellite data	17/8/2023	23/8/2023	5.0 d.	0%
11	Adding additonal data into predition results	17/8/2023	30/8/2023	10.0 d.	0%
12	Testing + Optimise	23/8/2023	29/8/2023	5.0 d.	0%
13	Completeing the final report	15/8/2023	1/9/2023	14.0 d.	0%

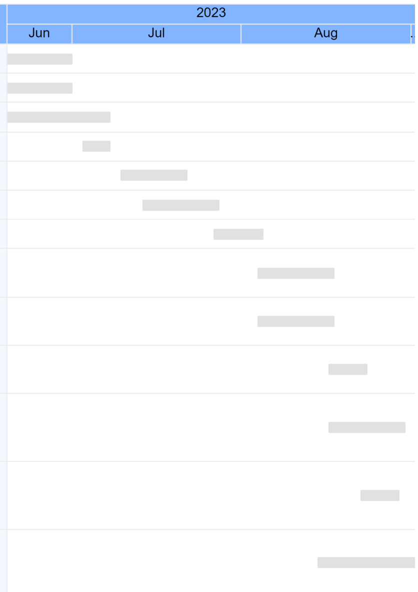

Figure 2: The gantt chart of anticipated future development plans

5 Supplementary Information

This is a summer internship project supported and provided by SnooCODE. The project is under the joint leadership of Sesinam Dagadu (MEng) from SnooCODE and Dr. Yves Plancherel from Imperial College London. Additionally, my colleague Wenxin Ran has been assigned to work on this project with a focus on model development and optimisation. If the project progresses smoothly, Wenxin Ran and I will compare and select the best Flood Risk Prediction Model in the final stage. And then the selected model will be used as the back-end prediction model for my Flood Risk Analytical Tool, thereby enhancing its predictive ability and making it a final product delivered to SnooCODE.

References

[1] Y.A. Twumasi and R. Asomani-Boateng. Mapping seasonal hazards for flood management in accra, ghana using gis. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 5, pages 2874–2876 vol.5, 2002.

[2] S.A. Adinku. Disaster preparedness: A sociological study of the flood problem in the odaw catchment in accra. unpublished Bachelor's thesis, Department of Sociology, University of Ghana, Legon, Accra,Ghana, 1994.

[3] R. Asomani-Boateng. Planning and managing urban organic solid waste in an african city: linking organic solid waste composting to urban cultivation in accra, ghana. unpublished PhD thesis, School of Planning, University of Waterloo, Waterloo, Ontario, Canada.

[4] Luís Cea and Pierfranco Costabile. Flood risk in urban areas: Modelling, management and adaptation to climate change. a review. *Hydrology*, 9(3), 2022.

[5] B. Jamali, E. Haghighat, A. Ignjatovic, J.P. Leitão, and A. Deletic. Machine learning for accelerating 2d flood models: Potential and challenges. *Hydrological Processes*, 35, 2021. e14064.

[6] Z. Wang, X. Chen, Z. Qi, and et al. Flood sensitivity assessment of super cities. *Scientific Reports*, 13:5582, 2023.

[7] W. Fenglin, I. Ahmad, M. Zelenakova, and et al. Exploratory regression modeling for flood susceptibility mapping in the gis environment. *Scientific Reports*, 13:247, 2023.

[8] Global flood database, 2023. (Accessed on: 4, June, 2023).

[9] Google earth engine datasets, 2023. (Accessed on: 4, June, 2023).