EMPOWERING GEO-BASED AI ALGORITHM TO AID COASTAL FLOOD RISK ANALYSIS: A REVIEW AND FRAMEWORK DEVELOPMENT

T. Atmajal and K. Fukushi

1 Dept. of Urban Engineering, The University of Tokyo, 7-3-1 Hongo, Tokyo, Japan - atmajal@env.t.u-tokyo.ac.jp
2 Institute of Future Initiative (IFI), The University of Tokyo, 7-3-1 Hongo, Tokyo, Japan - fukushi@ifi.u-tokyo.ac.jp

ICWG III/IVa - Disaster Assessment, Monitoring and Management

KEYWORDS: Literature, Machine Learning, Vulnerability, Ecological Solution.

ABSTRACT:
Climate change and current susceptibilities exacerbated the coastal flood loss and damage resulting in livelihoods and property damage. Urban areas in the Low to Lower-Middle Income Countries are expected to be disproportionately impacted by the disaster, given a higher share of citizens living in the Low Elevation Coastal Zone, limited financial resources, and poorly constructed disaster protection. Documentation of historical coastal floods, population, and property affected, could advance the assessment by considering those parameters in risk analysis. Besides, incorporating such geographic features e.g., mangroves as the ecological solution for alternative coastal flood protection in the prediction is also essential. Mangrove is considered fit for the LLMIC primarily situated in the tropical zone. The prediction utilizing spatial Machine Learning (ML) could aid climate-related disaster risk analysis and contribute to risk reduction and policy suggestions to improve disaster resilience. The research aims to archive recent studies on the application of geospatial science empowering Artificial Intelligence, notably ML in coastal flood risk assessment, so-called GIS-based AI. Another aim is to document population, property, and mangrove distribution across the LLMIC. Artificial Neural Networks were mostly utilized for disaster risk assessment in past research. The number of 58 historical coastal flood events and 908 expected coastal flood hotspots for 2006 to 2021 has been documented. Over 1.2 million Km² falls under vulnerable areas toward coastal flood in LLMIC under different settlement types where Large City (urban areas) dominates it. Mangrove distribution is mainly distributed across tropical regions mostly distributed along the Southeast Asia coast.

1. INTRODUCTION

1.1 Background

The coastal cities have experienced and been exposed to a range of coastal hazards, notably due to extreme Sea Level Rise (SLR) with its four significant impacts: coastal flood; coastal erosion; exacerbated land subsidence; and saltwater intrusion (Azevedo de Almeida and Mostafavi, 2016). According to the Special Report on the impacts of global warming of 1.5°C by IPCC, coastal flood has the highest risk of a severe impact associated with climate change. Each degree of increasing temperature is a matter of coastal flood risk (IPCC, 2018). The risk is projected to increase, associated with rising temperature and triggered by current susceptibilities, resulting in population exposure, property damage, and disruption of economic activities in the coastal zone.

Previous documentation revealed that nearly 10% of the world's population (618 million) and 2.3% (2,599 thousand km²) of the world's land area of coastal countries resided and situated in Low Elevation Coastal Zone (LECZ) in 2000, defined as the contiguous area along the coast that is less than 10 meters above sea level (McGranahan et al., 2007; Neumann et al., 2015). By looking at the urban boundary, it accounted for 13% of the total urban population (352 million) living within LECZ, which covered 8% of the whole world's urban land area (275 thousand km²). Another documentation indicated that in 2015 the urban area that falls under LECZ was estimated to have 10% of the world's population and 13% of the world's urban population, equal to 815 million (MacManus et al., 2021). By 2060, the total LECZ population was projected to reach 1.4 billion inhabitants (534 people/km²) or equal to 12% of the world's population of 11.3 billion under the highest-end of forecast scenarios (Neumann et al., 2015). On the other hand, the situation is getting worst, exacerbated by the increase of such disasters in the future followed by their damage. According to historical global data by EM-DAT, the event tends to increase in upcoming years, which is associated with increasing loss and damage (Kirezci et al., 2020). Average global flood losses in 2005 are estimated to be approximately US$ 6 billion per year, rising to US$ 52 billion by 2050 with projected socio-economic change alone (Hallegatte et al., 2013). In short, coastal flood is expected to have the highest risk of a severe impact of loss and damage on livelihood and properties damage (Chan et al., 2018; Hallegatte et al., 2013; Kirezci et al., 2020; Neumann et al., 2015; Nicholls et al., 2008). It is, therefore, essential to carry out a coastal flood risk analysis to better grasp the disaster across the coastal zone.

Among the world's countries, urban areas in Low and Lower Middle-Income Countries (LLMIC) are expected to be the most vulnerable to coastal floods, given a higher share of the population living in the LECZ and limited financial resources for disaster management. The majority (83%) of the global LECZ population lived in less developed countries (Neumann et al., 2015). Accounted 28% of the urban area of the LLMIC population lives in the LECZ (McGranahan et al., 2007), which makes it vulnerable. Dasgupta et al. (2009) assessed that approximately 0.3% (194,000 km²) of the territory in the 84 developing countries would be impacted by a 1-m SLR. It would

* Corresponding author: atmaja@env.t.u-tokyo.ac.jp
equal 56 million people (1.28% of the population) exposed. LLMIC also tends to have non-engineered or poorly constructed coastal protection due to financial resources (Takagi, 2019).

Despite their budgetary limitations, many developing countries have the advantage of considering an ecological solution because they are often situated in tropical and subtropical regions. The ecological solution, i.e., the mangrove ecosystem for Eco-DDR, is currently acknowledged as one of the alternative strategies for coastal flood protection. Besides, it provides co-benefits for carbon sequestration both in the soil and the plant. The strategies also allow stewardship supporting people surrounding livelihood. Therefore, understanding to what extent ecological solutions can be applied and beneficial for LLMIC will be advantageous for coastal flood management.

An advancement in coastal flood risk simulation could contribute to risk reduction, policy suggestions, minimizing the loss of livelihood, and property damage associated with coastal floods. Nowadays, Artificial Intelligence (AI), notably Machine Learning (ML) approach for flood risk simulation, has emerged in the past few years (Chang et al., 2019; Mosavi et al., 2018). The paper is trying to document the utilization of AI, especially ML, for aiding coastal flood risk analysis. ML utilization is encouraged to aiding climate-related disaster analysis and advance disaster risk prediction for handling big spatial data and high spatial-temporal data (Huntingford et al., 2019). In addition, looking at LLMIC which is considered a vulnerable region, it is essential to document the population, property, and coastal strategies, especially mangroves as Eco-DDR along its coastal zone. This documentation could enrich an advancement in coastal flood risk prediction by incorporating those parameters into the simulation. Besides, it is expected that the significance of ecological solutions in averting loss and damage from a coastal flood is also revealed.

1.2 Objective
The study aims to archive recent geospatial science studies that empower Artificial Intelligence, notably ML, in aiding disaster risk analysis, so-called GIS-based AI. Another aim is to document the population and property in LLMIC, which is vulnerable to coastal floods and the distribution of mangroves along their coastal zone.

2. METHODOLOGY
2.1 Research Boundary
Case Study: The study selects cases in the Low to Lower Middle-Income Countries, or LLMIC, given their vulnerability toward coastal floods, i.e., a higher share of citizens living in their urban area, limited financial resources for coastal flood management, and poorly constructed coastal protection. Urban areas along the coastal zone are selected as essential locations for livelihood and economic activities within LLMIC.

LLMIC: countries that fall under the category where their GNI per capita is lower than $4,095 (World Bank 2021). Forty-five countries fall under this category worldwide.

Coastal Flood Risk Terminology: The study defines the risk as to the potential occurrence and impact of the coastal flood in terms of loss and damage, including population exposure, property damage, and economic loss. A coastal flood is water that penetrates onto land, especially within Low Elevation Coastal Zone (LECZ), areas lower than 10 m above sea level and hydrologically connected to the coast.

2.2 Mapping on Coastal Flood Events, Population, Property, and Mangrove Distribution across LLMIC
The study utilized the ArcGIS Pro to document historical and projected coastal flood events, population, property, and mangrove distribution across LLMIC. Historical coastal flood data were collected based on a dataset of global spatial dataset for flood studies, which is well explained by previous research (Kireczki et al., 2020; Lindersson et al., 2020).

2.3 Documentation of Past Studies on Empowering Geo-based AI on Disaster Risk Analysis
The study concerns on application of Artificial Intelligence (AI), both Machine Learning (ML) and Deep Learning (DL), for disaster risk analysis, especially flood risk. The study focuses on recent research for the period 2016-2022 at any level and coverage. It emphasizes what kind of analysis is used in terms of temporal or spatial machine learning, ML algorithm used, and feature variables incorporated in the simulation.

3. RESULT AND DISCUSSION
3.1 State of the Art Geo-AI Approach in Aiding Disaster Risk Analysis
Artificial Intelligence, especially the Machine Learning (ML) approach, has emerged in the past few decades (Mosavi et al., 2018), as shown in Figure 1. This approach allows for various purposes, especially for resilience and preparedness against flooding (Faraji et al., 2019). According to the documentation, the research revealed that researchers mainly utilized ANNs (Artificial Neural Networks) followed by the SVM (Support Vector Machine), which gradually increased in use. Aiyelokun et al. (2021) predicted flood risk and drought through Naïve Bayes (NB) approach for traditional ML using wind, rainfall, temperature, and Relative Humidity (RH) dataset (Aiyelokun et al., 2021). Park and Lee (2020) assessed coastal flood risk under climate change impacts in South Korea using multiple machine learning algorithms (KNN-k-Nearest Neighbor; RF-Random Forest; SVM) spatially (Park and Lee, 2020). They have included geographic features such as Tide, DEM, and urban characteristics for the analysis despite lacking in considering the population and coastal protection in the simulation. At the same time, other researchers assessed the flood risk using traditional machine learning through flood datasets only (area, location, duration, etc.) for flood classification or prediction of the inundation (Chang et al., 2019; Faraji et al., 2019; Tayfur et al., 2018).

In short, based on the documentation, the most general ML algorithms for flood prediction were Adaptive Neuro-Fuzzy Inference System (ANFIS), Multilayer Perceptron (MLP), Wavelet Neural Network (WNN), Ensemble Prediction Systems (EPSs), Decision Tree (DT), Random Forest (RF), classification and regression trees (CART), Support Vector Machine (SVM), Naïve Bayes (NB), and Artificial Neural Networks (ANNs) (Aiyelokun et al., 2021; Ganguly et al., 2019; Manandhar et al., 2020; Park and Lee, 2020; Ruckelshaus et al., 2020; Saravi et al., 2019). They were widely used in flood modeling and provide robust and efficient ML algorithms for flood prediction.
Although most of the researchers have acknowledged the robustness of ML in flood prediction, they were still analyzing through traditional ML and neglected the geographic features which essential and may influence the assessment. In addition, the researchers have limited estimates of risk in terms of loss and damage such as area flooded, the property affected, and community exposed, including the projection analysis considering climate change and population scenarios. Previous research limited their risk assessment to engineered ones instead of focusing on ecological solutions, i.e., Eco-DRR mainly through mangrove ecosystems that hypothesized fit for LLMIC primarily situated in the tropical or subtropical region. Table 1 shows the comparison recommended for future research, originality, and novelty.

Table 1. Review on coastal flood risk simulation through ML

<table>
<thead>
<tr>
<th>Reference</th>
<th>Level - Coverage</th>
<th>Focus</th>
<th>Analysis</th>
<th>Approach</th>
<th>Feature Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Aiyelokun et al., 2021)</td>
<td>Urban – province</td>
<td>Flood, drought</td>
<td>TML</td>
<td>NB</td>
<td>Wind, Rainfall, Temp., RH</td>
</tr>
<tr>
<td>(Park and Lee, 2020)</td>
<td>All – country</td>
<td>Coastal flood</td>
<td>SML</td>
<td>KNN, RF, SVM</td>
<td>Tide, DEM, Urban geo-characteristics</td>
</tr>
<tr>
<td>(Saravi et al., 2019)</td>
<td>All – global</td>
<td>Flood Classification</td>
<td>TML</td>
<td>RF, ANN, NB, LR</td>
<td>Flood data (area, location, duration)</td>
</tr>
<tr>
<td>(Chang et al., 2019)</td>
<td>Urban – city</td>
<td>Flood Inundation</td>
<td>TML</td>
<td>ANN</td>
<td>Flood data (area, location, duration)</td>
</tr>
<tr>
<td>(Tayfur et al., 2018)</td>
<td>All – country</td>
<td>Coastal Flood</td>
<td>TML</td>
<td>ANN</td>
<td>Flood data (area, location, duration)</td>
</tr>
<tr>
<td>Future Research</td>
<td>Urban - global (LLMIC)</td>
<td></td>
<td>SML</td>
<td>Ensemble Model</td>
<td>SLR, Tide, Surge, Waves, Wind</td>
</tr>
</tbody>
</table>

3.2 Distribution of Urban Settlement across LLMIC

Cattaneo et al. (2021) has identified and divided the catchment areas of urban centers of different sizes called Urban-Rural Catchment Areas (URCAs), varied by the total population and time travel to the city. URCA is a raster dataset of the 30 urban-rural catchment areas showing different sizes of catchment areas around cities and towns. As explained by the authors, each rural pixel is assigned to a specific category. In this study, it is adapted and modified into ten categories of urban settlement types to make it simpler as follows:

1. Large city (> 5 million)
2. Large city (1 – 5 million)
3. Intermediate city (500,000 – 1 million)
4. Intermediate city (250,000 – 500,000)
5. Small city (100,000 – 250,000)
6. Small city (50,000 – 100,000)
7. Town (20,000 – 50,000)
8. Rural (beyond other types)
9. Dispersed towns (>3 hours to any city)
10. Hinterland (>3 hours to any city)

The study concerns the urban areas in the Low Elevation Coastal Zone (LECZ) across Low to Lower-Middle Income Countries (LLMIC) located in tropical and subtropical regions. The selection of the areas is mainly motivated due to higher vulnerability zone among other zones toward coastal floods. This vulnerability means that the location has a higher risk of coastal flood occurrence, followed by the potential impact on population exposure and property damage (MacManus et al., 2021; McGranahan et al., 2007; Neumann et al., 2015). Besides, they also have limited financial resources for disaster management and poorly constructed coastal protection fields(Takagi, 2019).

Figure 1. Primary ML methods are used for flood prediction in the literature. Reference year: 2008 (source: Scopus adapted from Mosavi et al., 2018).

Figure 2. Distribution of Urban Settlement along LECZ across LLMIC

As explained by the authors, each rural area flooded, the property affected, and community exposed, including the projection analysis considering climate change and population scenarios. Previous research limited their risk assessment to engineered ones instead of focusing on ecological solutions, i.e., Eco-DRR mainly through mangrove ecosystems that hypothesized fit for LLMIC primarily situated in the tropical or subtropical region. Table 1 shows the comparison recommended for future research, originality, and novelty.
3.3 Distribution of Historical Coastal Flood Events and Coastal Floods Hotspot

Historical coastal flood event was compiled based on Dartmouth Flood Observatory (DFA) specifically for coastal flood event from 2006 to 2021. DFA provides the flood information from 1985, but due to lacking coordinated information, it is only started from 2006. There are compiled 109 cases of coastal floods worldwide caused by high tides, storm surges, cyclones, and typhoons but only 58 events across the LLMIC areas, as shown in Figure 3. Some of these events will be validated by the news if available. Each point comprises information such as country, detailed location including coordinate, duration of the flood, people exposed and damage, primary cause, and flooded area.

Figure 3. Global distribution of historical coastal floods (2006–2021 June) across LLMIC

In addition to the historical coastal flood event, the study recommends considering the global episodic coastal flood hotspot due to SLR, as shown in Figure 4. Coastal flood "hotspot" regions of changes in episodic coastal flooding in 2100 for RCP8.5 were estimated based on projected episodic flooding in 2100 minus present-day episodic flooding (Kirezci et al., 2020).

There were 908 total cases of extreme sea-level rise projected in the future (2100), mostly under 1.5 m. Furthermore, there is a vast distribution of coastal hotspots, especially in southeast Asia, i.e., Indonesia. On average, the coastal hotspots range from 0.7-0.8 m asl. Based on the historical information and distribution of coastal hotspots, there is a case match in a city currently exposed by the storm surge and, in the future, expected to have a 1.5-2.5 m extreme sea-level rise.

Figure 4. Global distribution of historical (2006-2021), projected Extreme Sea Level (ESL) for RCP8.5 in 2100, and Coastal ’Hotspot’ across LLMIC

3.4 Distribution of Mangrove across LLMIC

The study concerns ecological solutions for coastal flood countermeasure in the LLMIC region. The distribution of ecosystem types for coastal protection benefits has been documented in the SNAPP project (Science for Nature and People Partnership). It shows various coastal protection and its benefits, as indicated in Figure 5 (Li et al., 2017). In addition, mangroves as massive coastal protection applied in the tropical zone were documented. Figure 6 shows the majority of mangrove distribution globally. Indonesia occupied almost one-fourth of global mangroves, equal to 3,244 thousand ha of (Giri et al., 2011; Kusmana, 2014).

Figure 5. Various coastal protection for specific hazard (n = 52). (Li et al., 2017)

Figure 6. Global Mangrove Forests Distribution in 2000 (Giri et al., 2011; Kusmana, 2014)

3.5 Population and Property Documentation across LLMIC

The documentation of population and property across LLMIC is addressed to reconsider that these areas are prioritized zone for the assessment. The documentation of the people with high-resolution was documented by CIESIN (Facebook Hub) called High-Resolution Settlement Layer (HRS). HRSL is an estimation of human population distribution at a resolution of 1 arc-second (approximately 30m) for the year 2015. The population estimates are based on recent census data and high-resolution (0.5m) satellite imagery from DigitalGlobe. The Connectivity Lab at Facebook developed the settlement extent data using computer vision techniques to classify blocks of optical satellite data as settled (containing buildings) or not. CIESIN used proportional allocation to distribute subnational census data to the settlement extent. World Bank Living Standards Measurement Study (LSMS) program was used to validate the final dataset against anonymized "ground-truth" household surveys. While the World Settlement Footprint (WSF) 2015 is a 10m (0.32 arcsec) resolution binary mask outlining the 2015 global settlement extent derived by jointly exploiting multitemporal Sentinel-1 radar and Landsat-8 optical satellite imagery. Settlements are associated with value 255; all other pixels are associated with value 0. Figure 7 indicates the documentation of population and property with high spatial resolution potentially used for the coastal flood risk simulation.
3.6 Framework Development for Coastal Flood Risk Assessment by Employing Geo-based AI

The coastal flood simulation approach was developed, as shown in Figure 8. The study presents a novel use of the ensemble Spatial Machine Learning Algorithm (SMLA) for coastal flood simulation, harnessing big spatial data of high temporal and horizontal resolution globally. Once the key parameters documentation is finished, all data will be transformed and harmonized into grid-based data and compiled into (A) Data Table of Key Parameters as variable features for flood modeling. The parameters include Sea Level Rise (SLR); Tide; Surge; Wave Setup; Wind/tornado/cyclone; MERIT-DEM; Slope; Population; DRR applied; (*) Coastal Flood Occurrence and Duration; (*) Flooded Area (Km²); (*) Property Damage ($); (*) Injuries or Death Direct/Indirect (number).

The Data Table of Key Parameter will be split into a training dataset (70%) and a testing dataset (30%) contained by data feature as a predictor and feature target (indicated in asterisk (*) above). Subsequently, these data will be (B) modeled through GIS-based Machine Learning (ML) using ArcGIS API for Python. Following ML approaches will be compared to figure out the best accuracy, such as Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naïve Bayes (NB), and Logistic Regression (LR). Result evaluation such as confusion matrix, ROC, and F1 score (>75%) will be employed. Previous research has compared these approaches for flood simulation but is limited to coastal floods especially considering such DRR in the simulation (Faizollahzadeh Ardabili et al., 2019; Park and Lee, 2020).

Sea level rise is defined as the height of water over the mean sea surface in a given time and region. In this analysis, the dataset of sea level anomalies is computed concerning a twenty-year mean reference period (1993-2012) using up-to-date altimeter standards. Tide and surge are easily defined as the difference between seawater anomaly and mean sea level rise (high tide), or the contrast of the rise in water level above the average tidal level, and do not include waves.

For further simulation, variable targets (Y) are required, such as flooded areas, the property affected, and the population exposed. The following is impact documentation in the case study (Pekalongan northern coast, central java, Indonesia).

The study developed the flooded areas due to coastal floods (Figure 10). It was generated using Landsat 8. There are some limitations where Landsat did not cover some areas during coastal floods. Therefore, the study assumed the most extended duration of the flood. Landsat 8 issues on cloud cover are also challenging. Here, Modified Normalized Different Water Index was used to estimate flooded areas despite overestimating compared with the news. The flooded area is defined as a temporary water index by MNDWI (> 0.5). The Modified Normalized Difference Water Index (MNDWI) uses green and SWIR bands to enhance open water features. It also diminishes other indices’ built-up area features that are often correlated with open water Fields(Muis et al., 2020).
A study on utilizing Machine Learning (ML) to aid disaster risk analysis has emerged recently. It is vastly improved to advance disaster prediction both spatially and temporally. Despite a limited study on spatial disaster risk assessment using the ML, there is a trend on this, and expected to utilize it shortly extensively. The application of this approach in the case study shows that ML has promising results to advance risk prediction in loss and damage.

Considering LLMIC, especially urban zone, for the focus study is essential given they are expected to have a severe impact from the coastal flood. The documentation of population and property has revealed that it is crucial to consider this region for assessment. The study recommends that those information and geographic features parameters such as ecological solutions applied to climate change and population scenarios enrich the risk forecast under various conditions.

**ACKNOWLEDGMENTS**

This work was supported by JST SPRING, Grant Number JPMJSP2108.

**REFERENCES**


Learning in Hydrological Processes, Climate Change and Earth Systems: A Systematic Review. MDPI AG. https://doi.org/10.20944/preprints201908.0166.v1


This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper. https://https://doi.org/10.5194/isprs-annals-V-3-2022-517-2022 | © Author(s) 2022. CC BY 4.0 License.