**FloodWatch: Devising an Autonomous Cyber Physical System for Real-Time Flood in an Operational Framework**

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

**Abhir Karande**

Fall, 2023

Technical Project Team Members

Andrew Ma

Siddardh Burre

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments Advisor

N. Rich Nguyen, Associate Professor Department of Computer Science

ABSTRACT
Flooding has devastating impacts in Southeast Asia, in part due to the lack of appropriate infrastructure. As a part of my involvement in the FloodWatch research group, I am devising an autonomous cyber physical system for real-time flood intelligence that can be leveraged by citizens in a progressive webapp. FloodWatch is a smart city service centered around the development of cyber physical system infrastructure for disaster intelligence and forecasting. The service provides a frontend progressive web application (PWA), live sensor readings through LoRAWAN router technology, and predictive insights on flood susceptibility. To produce insights on susceptibility, I am working on several machine learning pipelines to ensure accuracy, explainability and corroboration. The current approach combines an existing hydrological model called HEC-RAS with a customized Artificial Neural Network (ANN) to produce geographical flood inundation that is then interpreted and overlaid on the map of Vietnam using a region tileset automation script. Additionally, I am devising a corroboration mechanism for crowdsourced flood reports by using LiDAR water level interpretation, computer vision for flood label detection of live camera feeds.

## INTRODUCTION

Floods are the most frequent type of natural disaster in the world, and they are a serious problem in Vietnam and other Southeast Asian countries. Floods are caused by a variety of factors, including heavy rainfall, river bank overflows, inadequate drainage systems, and land use changes. They can have a devastating impact on communities, causing loss of life, damage to property, and displacement of people. The risk of flooding is increasing in Vietnam and other Southeast Asian countries due to population growth, climate change, and urbanization. In the past three years, flooding in Southeast Asia has killed over 300 people and displaced over 200,000. As flooding is a major challenge, some steps to mitigate the risk and impact of floods include investing in flood prediction and prevention infrastructure, improving drainage systems, implementing land use policies, and educating the public about flood risks.

Flood prediction, in particular, is extremely crucial due to its ability to spur on appropriate infrastructure preparation. Two prominent classes of flood prediction models include deterministic hydrological models based on physics, and machine learning models based on historical data. Recently, machine learning approaches to forecasting have gained popularity due to availability of historical data as well as adaptibility to various prediction scenarios. In order to produce accurate, replicable, and explainable predictions, it is necessary to combine these two paradigms of flood forecasting for robustness.

## 1. RELATED WORKS

One of the leading physical modeling tools used for mapping inundation is called HEC RAS. One study uses an Artificial Neural Network (ANN) to predict runoff, which was used as a prior/input to HEC-RAS for the granular inundation mapping of the areas surrounding the Baro River Basin in Ehtiopia (Tamiru, 2021).This study focuses on one location in a controlled environment. The goal of my work is to apply similar hybrid modeling to any given location within the country of Vietnam. To the point of being able to forecast for multiple locations at a lower granularity, Google's operational framework flood forecasting system makes use of time-series long short-term memory

(LSTM) and linear models for first forecasting river stage and then using threshold and manifold models to compute inundation extent and depth through satellite imagery (Nevo, et al. 2018). Satellite imagery is not necessarily the most scalable and is highly dependent on available hardware and river stage data, so I am focused on the ability to have highly specific flood forecasting within an operational framework that is not too intensive on available resources outside of sensors.

Another aspect of flood intelligence that is important in devising a forecasting system is corroboration. With the FloodWatch project, the use of sensors for various data acquisition allows for different modalities of data, whether it is water level increase through LiDAR or accumulated precipitation through rain gauges. Work has been done on using a multimodal fusion of LiDAR data and vision (cameras) to detect targets on bodies of water. Through the use of LiDAR point clouds and bounding boxes, targets are able to be detected on a constant body of water (Wang, et al. 2023). A similar mechanism could certainly be applied to detecting changing water levels through the use of LiDAR and live camera feed through a multimodal fusion. Such a mechanism serves as one example of corroboration in devising our flood intelligence system.

## 2. PROPOSED DESIGN

The proposed design is currently in development. The below subsections detail the data used and the methodology for development. The data and workflow are both in respect to devising the ensemble method for flood susceptibility prediction and the corroboration mechanism with LiDAR and camera feeds.

### 2.1 Data Used

For the purposes of training a time-series model that predicts precipitation, or runoff for a specified time window, there is a need for a large corpus of historical precipitation data. Furthermore, as the goal is to predict in an operational framework that is responsive to location, it is important to collect data for all the provinces and districts within Vietnam. To collect this data, an automated script that would parse through districts of a GeoJSON file (containing the bounding boxes that represent districts and provinces at different granularities) would then make calls to the WeatherAPI, with appropriate dates in order to get historical precipitation readings ranging over 10 years.

Another aspect of flood forecasting that needs to be considered is the land use, vegetation, change in elevation, and topical wetness index. For producing inundation extent and coverage at scale, a digital elevation map (DEM) of Vietnam as a whole country was used.

The data required for the proposed corroboration mechanism includes positive and negative examples of flood images (flood/no-flood).

### 2.2 Workflow

The main workflow involving flood forecasting starts with the prediction of runoff. To do so, LSTM networks are incorporated to predict for a window of 24 hours following the current hour. Runoff readings are then used as an input to the HEC-RAS simulation, which is previously prepared and fitted with the appropriate DEM and stream drawings. Upon the completion of the inundation simulation, the mapping is overlaid on the region tile-sets of the PWA.

With available inundation mapping on the PWA, the closest available live camera/ LiDAR station is to be queried for both flood image classification and water level detection. The mechanism by which the appropriate/closest devices are chosen is through geographic clustering.

## 3. ANTICIPATED RESULTS

The implementation of the autonomous cyber-physical system for real-time flood intelligence, as described above, is expected to yield several anticipated results. For one, the design offers improved flood prediction accuracy through the combination of LSTM networks for runoff prediction with a HEC RAS simulation for inundation mapping afterwards. The inclusion of historical precipitation data from the WeatherAPI spanning a decade will contribute to fine tuning the predictive capabilities of the model. This improved accuracy will enable early warnings and preparedness for potential flooding events.

In regards to the real-time flood inundation mapping, the integration of the digital elevation maps and HEC-RAS's inundation mapping simulation allows for the visualiztion of potential flood areas in Vietnam. The overlay of this map(?) on the PWA will provide users with a clear understanding of the flood risk in their specific regions.

The corroboration mechanism that works through the incorporation of live camera feeds, LiDAR, and crowdsourced flood reports for flood image classification and water level detection should be expected to ensure robust and accurate flood predictions. By utilizing geographic clustering to select the closest devices, the system aims to validate flood susceptibility insights and enhance the accuracy of predictions.

Last, the successful impementation of this flood intelligence system is expected to contribute to a reduction in flood-related losses, including loss of life, property damage, and displacement of people. By providing accurate and timely flood predictions, the system aims to enable better disaster preparedness and response, which ultimately will save lives and resources.

These anticipated results underscore the potential benefits of the autonomous cyber physical system for real-time flood intelligence, which seeks to address the pressing issue of flooding in Southeast Asia and similar regions.

## 4. CONCLUSION

This work with machine intelligence in the setting of cyber physical system infrastructure is very useful for municipalities at various scales. The aspects of multimodal intelligence and forecasting ability provide robust insights for such municipalities to use for infrastructure allocation prior to floods and other natural disasters occuring.

Multimodal intelligence, the ability to process and analyze information from multiple sources, plays a crucial rôle in enhancing situational awareness. By combining data from sensors, weather forecasts and historical records, machine learning algorithms can identify patterns and trends that indicate impending flooding and other natural disasters.

The other key aspect in this intelligent system is forecasting ability, which as mentioned previously, allows for municipalities to anticipate future conditions and make informed decisions. By analyzing historical data and incorporating real-time information, we are able to use machine learning models to predict the likelihood and severity of flooding at an extremely granular level. Futhermore, through the use of sensor hardware and cameras, granular forecasts become available and a corroboration mechanism is in place to ensure accuracy.

## 5. FUTURE WORK

Other aspects of intelligence that are yet to be developed in this system involve safety checks to make sure that there is not anomalistic behavior amongst sensors. It is not uncommon for there to be mechanical defects, environmental factors, poor signal strenght, or battery issues that prevent sensors from providing proper measurements and readings. A sensor verification service will need to be constructed that makes use of time-series anomaly detection in order to identify sensors with anomalistic behavior.

This system intelligence also needs to be delivered to consumers in a digestable format and as other sections of the FloodWatch team are developing the PWA, there will need to be further integration of previously discussed mechanisms.

## REFERENCES


Belay, M., Blakseth, S., Rasheed, A., Rossi, P. (2023) Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series : Existing Solutions, Performance Analysis and Future Directions. MDPI. 2023.

Javed, N. (2012). Automated Sensor Verification Using Outlier Detection in the Internet of Things. IEEE. 2012.

Malhotra, P., Ramakrishnan, A. (2016) LSTM-based Encoder-Decoder for Multi sensor Anomaly Detection. ICML 2016. https://arxiv.org/pdf/1607.00148.pdf

Nguyen, H., & Hong, H. (2023). Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in the DuPage River Basin, Illinois. Water, 9(7), 117.

Tamiru, H., & Wagari, M. (2022). Machine learning and HEC-RAS integrated models for flood inundation mapping in Baro River Basin, Ethiopia. Modeling Earth Systems and Environment, 8(2), 2291-2303. https://doi.org/10.1007/s40808-021-01175-8