Daily Flooding Risk Assessment Map for Medical Infrastructures using Machine Learning Algorithm

Jihoon Chung, Kazuki Neo, Muhammad Khan, Ethan Cruz,
Zhi Zheng, Megan Yu, and Thilanka Munasinghe

Rensselaer Polytechnic Institute, Troy, NY, United States

What is the Problem?

- Sea-level rise has a significant influence on the frequency of coastal floods, and it is expected to show **a two-fold increase every five years in the odds of extreme coastal flooding** in the near future. It means that, by 2100, today's 'one-in-a-lifetime' flood occurrence will be shifted to a near-daily occurrence (Taherkhani et al., 2020)
- Medical facilities are critical infrastructures to provide local communities with vital health services; however, they are **especially vulnerable to the effect of natural disasters** and are key components of the emergency response (Melnychuk et al., 2022)
- In the case of Superstorm Sandy on October 29, 2012, **10.4% of the hospitals reported flooding** as a result of the storm, and **16.9% had a significant challenge for hospital operation** due to structural damage by flood and wind during the storm (Levinson and General, 2014)

Fig 1. An elderly patient waits to be rescued from the Gulf Health Care Center when hurricane Harvey slams Texas

Fig 2. St. Joseph set up a triage center to care for the influx of patients when hurricane Harvey slams Texas

References | Christensen, J. (2017, August 31). Some hospitals hang on as others close amid Harvey's floods. *CNN*. Available at https://www.cnn.com/2017/08/30/health/harvey-houston-hospitals/index.html

What is the Problem?

- Currently, FEMA provides information on flood hazard areas for encouraging the hospitals in the area to plan emergency preparedness for the disaster; however, it is not accurate to predict all the hazard areas and there are **technical challenges when it comes to predicting the hazards accurately**

- In the case of Harris County in Texas, 16 (24%) of 66 non-psychiatric hospitals experienced flood impact during Hurricane Harney in August 2017, and **five (31%) hospitals were located outside of the flood hazard area** (Hines and Reid, 2020)

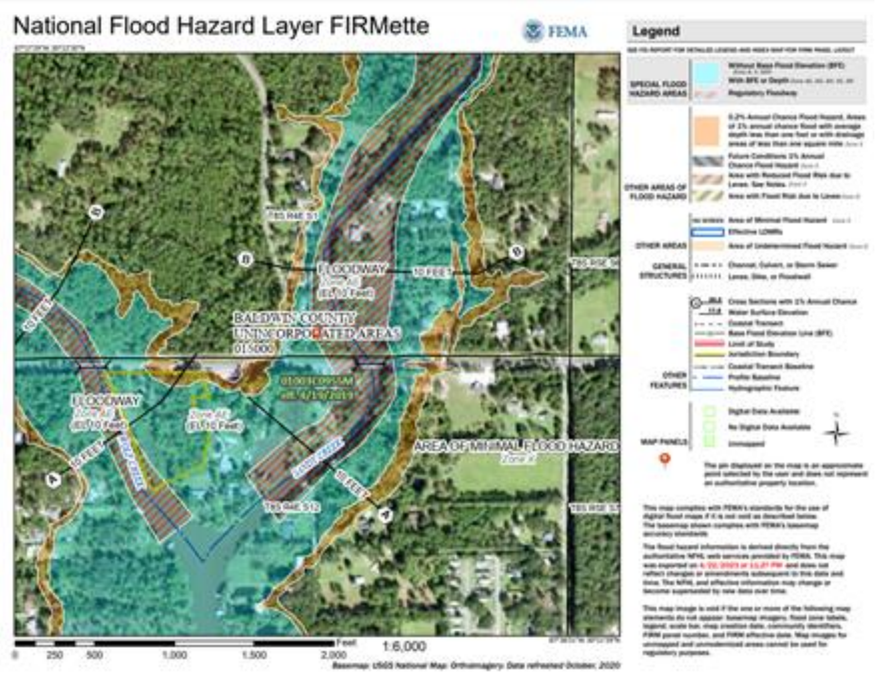


Fig. 3: Example of Flood Hazard Layer provided by FEMA

* **FEMA:** Federal Emergency Management Agency

References | FEMA (2023). FEMA's National Flood Hazard Layer (NFHL) Viewer. Available at https://hazards-fema.maps.arcgis.com/apps/webappviewer/

Research Objectives

- We aim to develop a **prediction model for flooding fractions in critical medical infrastructure locations** using the integrated database of explanatory variables and machine learning algorithms.
- As a pilot project, this work mainly focuses on 11,521 medical infrastructures that are located on the **Southeastern coast of the United States (US)**, including Alabama, Florida, Louisiana, Mississippi, and Texas states

Fig 4. Region of Interest

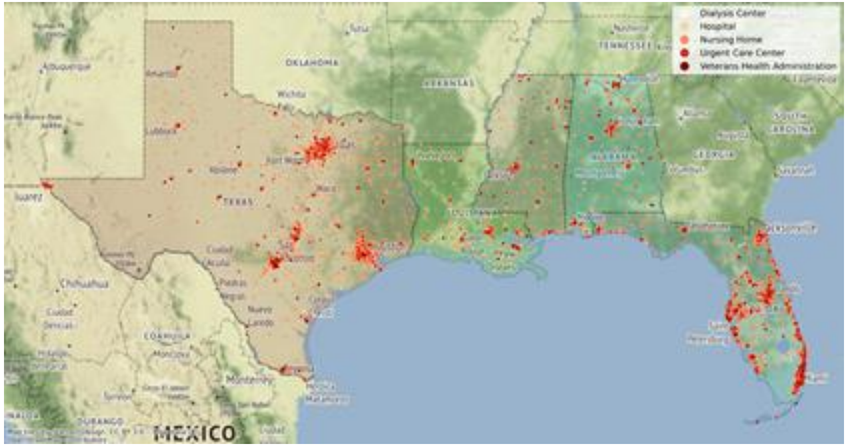


Fig 5. Locations of Medical Infrastructure on the southeastern coast, USA

Related Works – Input Parameters

Input Parameter	Examples	References
Meteorological parameter	Precipitation, temperature, humidity, wind speed/direction, sea level pressure	Coulibaly et al. (2005), Schoof and Pryor (2001), Šaur (2017), Bande and Shete (2017), Chai et al. (2016), and Keong et al. (2016)
Hydrological parameter	Water level, Stream flow, River discharge, Runoff depth	Widiasari and Nugroho (2017), Garcia et al. (2015), Phitakwinai et al. (2016), Wan et al. (2015), Jaafar et al. (2016), Furquim et al. (2016), Harun et al. (2017), Li et al. (2016), El-Telbany (2017), Kulkarni and Shete (2014), Liu et al. (2014), Young et al. (2015), and Panigrahi et al. (2018)
Topographic parameter	Elevation, Ground slope	Merkuryeva et al. (2015), Puttinaovarat and Horkaew (2020), Hauer et al. (2021), and Edmonds et al. (2020)
Geological parameter	Land cover, Land use	Otago et al. (2020), Puttinaovarat et al. (2014) and Ramly et al. (2015), Emerton et al. (2016), Tang et al. (2021), Santos and Fragoso (2016), and Ogato et al. (2020)
Historical Flood Records	Flood return period, Flood Zones	Harun et al. (2017), Santos and Fragoso (2016), Dottori et al. (2017), and Puttinaovarat and Horkaew (2020)

Related Works – Prediction Approach

Approach	Description	Pros & Cons	References
Physical Principle-based Approach	It **considers principles of physical processes**, such as storm, rainfall-runoff, or hydrodynamic models	● This approach has great capability for flood prediction in different scenarios; however, it **requires in-depth knowledge and expertise to formulate the physical process** regarding hydro-geomorphological parameters ● The models are **too complex and require a long computational time**	Maspo et al. (2020) and Mosavi et al. (2018)
Statistical Approach	It predicts the flood **based on stochastic processes** with certain probability distributions from historical data.	● This data-driven approach has a less computational cost and better generalization; however, it is **inaccurate for short-term prediction**	
Machine Learning Techniques	It **identifies the relationship or pattern** between input and output parameters for data-driven decision-making.	● This advanced approach is able to numerically formulate the nonlinear flooding mechanism without the need for in-depth knowledge about physical processes, and it requires **less complexity and computational cost with minimal inputs** ● Several scientific research articles have proved that ML models **outperform the two conventional approaches and are suitable for flood prediction**	

Research Method – Overview

- This project deploys four machine learning techniques to predict flood fractions at the medical infrastructures, and the the predicted flood fraction values are visualized on a website

Data Collection
- HIFLD infrastructure data
- NASA Land Surface Model data
- FEMA National Risk Index

Data Preprocessing
- Reformatting data & replace null values
- Extract relevant training data through analysis and preprocessing

Machine Learning Model
- Outlier elimination and feature selection
- Prediction of fractions of flooded areas at medical facilities

Result Analysis
- Model evaluation and analysis based on Mean Squared Error

Data Visualization
- Development of daily flood fraction map on a website
- Visualization of predicted flood fraction values

ML Models	Description
Random Forest (RF)	Utilizing an ensemble of decision trees, Random Forest Regression processes subsets of the training dataset. Each decision tree separates and organizes data based on the dependent variable to enhance accuracy. The amalgamation of these decision trees forms the random forest. Resulting predictions are obtained by averaging outputs from individual trees.
XGBoost	XGBoost employs decision trees to predict outcomes. The model differs in that each new decision tree aims to reduce the error left by its predecessors. Unlike isolated trees, they don't learn independently but work collaboratively to refine predictions.
Multivariate Linear Regression	This linear learning method predicts the target variables by weighing and summing independent variables (features). Each feature is assigned a unique weight to minimize the error between predicted and actual values. By optimizing these weights, the model discerns the impact and significance of each feature on predictions.
Neural Network (NN)	Neural networks utilize layers of interconnected nodes to learn patterns and relationships within the data. This model is employed to predict the target variable by iteratively adjusting parameters across layers, enabling the model to capture complex dependencies within the dataset.

Fig 6. Research Method (left) and developed Machine Learning Models (right) We Used in this Project

Feature Selection

- Based on the correlation analysis, we selected 16 variables out of 50 variables from the data resources.
- Permutation Feature Importance is implemented to identify main contributors on the performance of the models

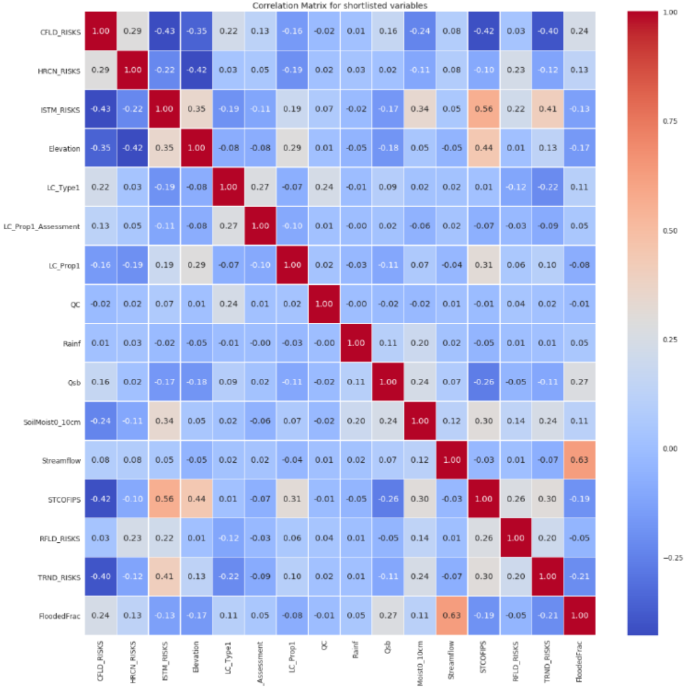


Fig 7. Correlation Matrix

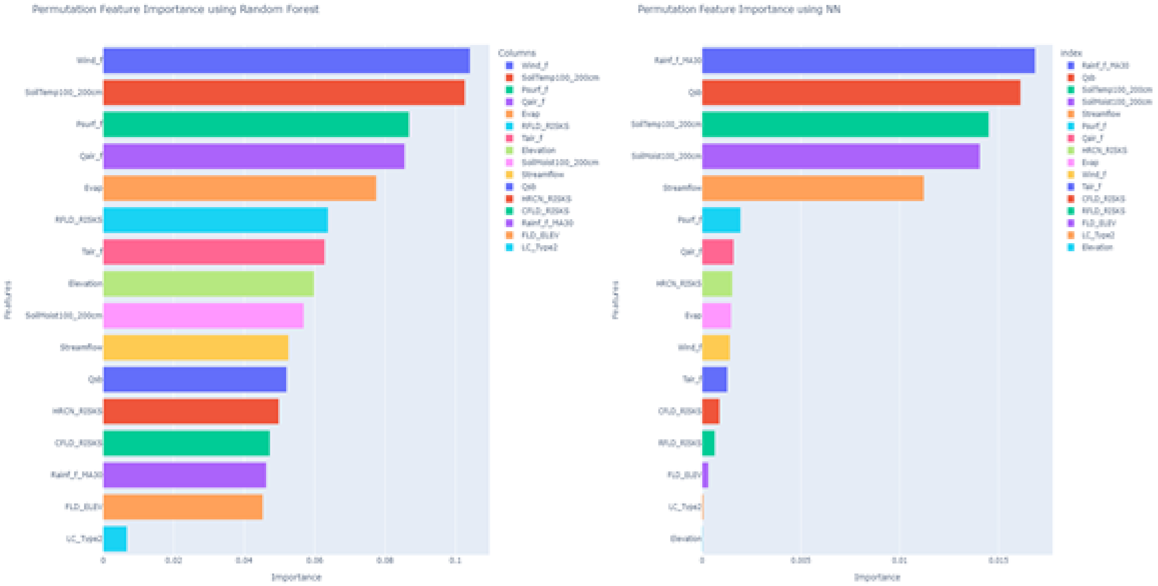
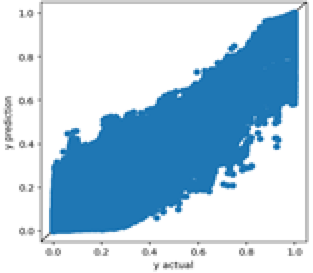
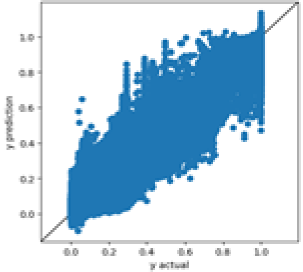
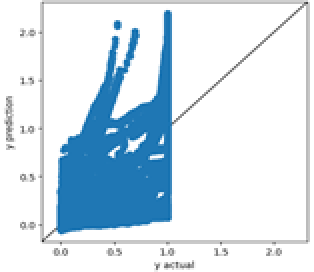
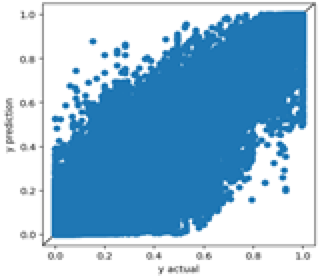
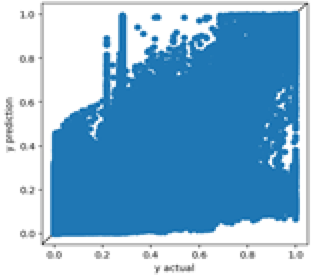
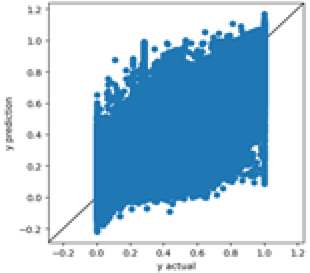
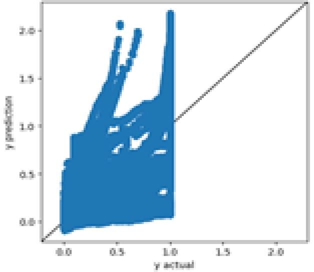
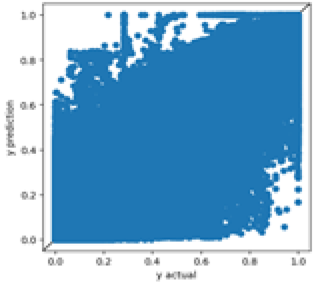


Fig 7. Results of Permutation Importance

Performance of the Developed ML Models

- To evalute importance of each feature, we implemented Permutation Feature Importance for the Random Forest and Neural Network models

ML Models	Random Forest	XGBoost	Multivariable Linear Regression	Neural Network
Test MSE (without PCA)	6.7×10^{-4} 	1.8×10^{-4} 	8.8×10^{-3} 	6.4×10^{-4} 
Test MSE (With PCA)	4.2×10^{-3} 	1.7×10^{-3} 	8.8×10^{-3} 	1.7×10^{-3} 

Daily Flood Fraction Map

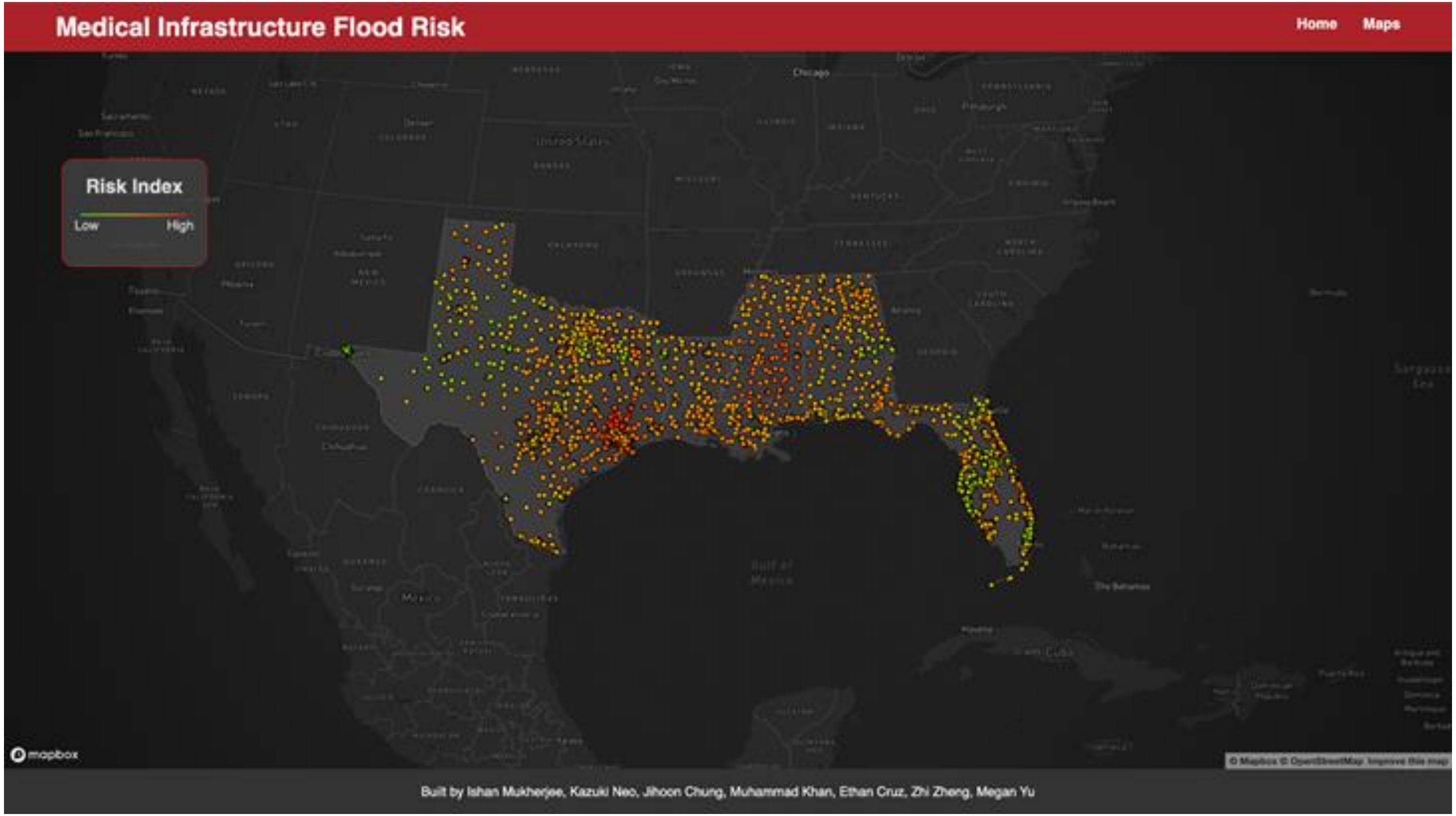


Fig 8. Developed Daily Flooding Risk Assessment Map