Application of Machine Learning Algorithms for Flood Susceptibility Assessment in the state of Kansas

Prashant Rimal, Zelalem Demissie, Glyn Rimmington

Department of Geology, Wichita State University, Wichita, Kansas

Abstract

Flooding has been a significant problem over the past century in the United States (US), causing growing threats to human lives and socio-economic damage. In the state of Kansas, since 1996, more than 1,500 flood events were recorded, resulting in an economic loss of between US$2b and US$5b. Many factors influence flood-susceptibility at a local scale. It may be helpful and timely to improve community resilience to flood disasters in Kansas. Our initial step was to assess factors that trigger flooding using Machine Learning (ML). Six ML algorithms: 1) Logistic Regression (LR); 2) Random Forest (RF); 3) Support Vector Machine (SVM); 4) K-nearest neighbor (KNN); 5) Adaptive Boosting (Ada Boost); 6) Extreme Gradient Boosting (XG boost) were used to evaluate their ability to classify locations in terms of flood-susceptibility. The learning data for these ML algorithms comprised a geo-spatial database of twelve flood-susceptibility factors from 1,528 flood inventories since 1996. The susceptibility factors comprised: rainfall, elevation, slope, aspect, flow direction, flow accumulation, Topographic Wetness Index (TWI), distance from the nearest stream, evapotranspiration, land cover, land surface temperature, and hydrographic soil type. The ML algorithms were compared, and the best algorithm was selected to estimate flood-susceptibility for each location in the geodatabase resulting in a flood-susceptibility map. A sensitivity analysis of flood-susceptibility factors indicated that the intensity or magnitude of the rainfall, land cover and soil type were the most significant factors for Kansas during this period.

Rationale

The changes in land use and climate change trends affect the number of floods and intensity, and the scale of damage; resulting in inundation, death, infrastructure damages, and chaos in societies' social and economic trends. Moreover, rising flood frequencies and the intensity in the Midwest states in recent decades [1] are of major concern. Identification of places that are susceptible to floods might helps to reduce the human casualties, social and economic costs associated with the floods . ML algorithms has ben used in past for the susceptibility analysis for river basins, however studies related to its use for larger administrative regions are limited.

Objective

The overall objective of our study was to create a flood susceptibility map of Kansas state and evaluate the ML algorithms response.

Methodology

Study Area:
Kansas State, USA

- Flood and Non-flood Control Points (1,528 points each)
- Train Set (80%-2364) Test (20% - 592)
- Reclassification of thirteen flood-influencing factors and value extraction for train and test sets
- Training and testing of the Six ML algorithms (LR, RF, SVM, KNN, Ada Boost, XG boost)
- Feature Importance
- Creation of Susceptibility Map form ArcGIS pro (Using Multi criteria evaluation)

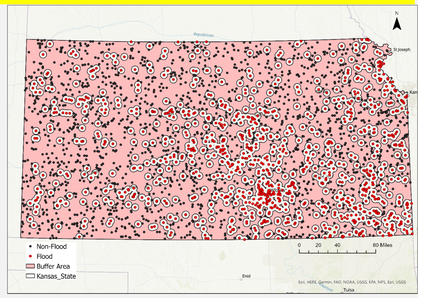

Figure 1: Map of study area with control points (flood [2]and non flood) and buffer area

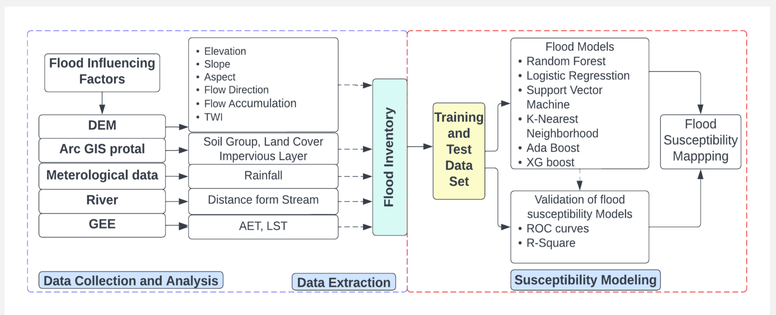

Fig. 2: Overall Methodology

Results and Discussion

GIS-based multi-criteria evaluation showed several places around the state that are susceptible to flood. Most of these regions are concentrated in the states' central and eastern portion (as shown in fig 3). The areas around the Wichita and the Kansas City are the areas highly susceptible to flood.

Talking about the response of the algorithms in the train and test set, the XG boost showed the best response in the train set and it was used in the test set where it was able to classify flood locations (with F1 score of .99). The table showing the score and the hyperparameters associated with it is shown as follows:

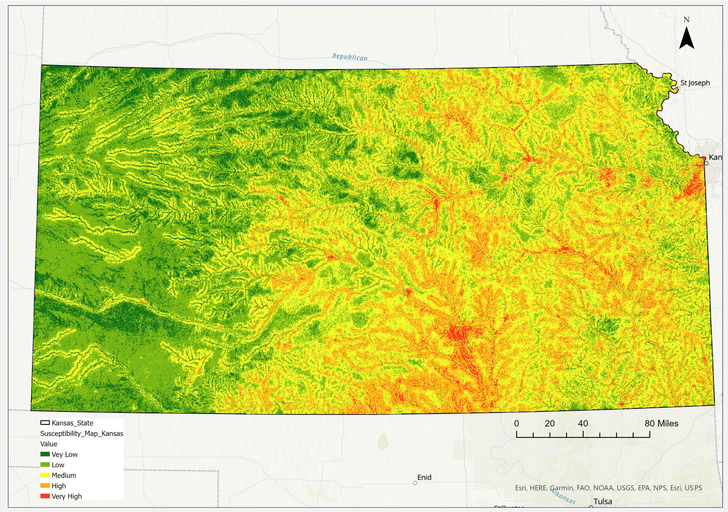

Figure 3. Susceptibility Map of the state of Kansas created using ArcGIS pro (500m resolution)

Table 1. Algorithms F1 score and associated hyperparameters

ML Algorithms	Score	Hyperparamaters
LR	0.73	{'C': 2}
RF	0.86	{'n_estimators': 15}
KNN	0.77	{'n_neighbors': 8}
SVM	0.81	{'C':3, 'rbf'}
Ada Boost	0.83	{'learning_rate': 1.5, 'n_estimators': 25}
XG boost	0.88	{'colsample_bytree': 0.7, 'gamma': 0.4, 'learning_rate': 0.2, 'max_depth': 10, 'min_child_weight': 1}

The permutation-based feature importance showed that the hydrographic soil type, Rainfall, and land cover were the major factors influencing the floods. *Validation and the map generation using models is in progress.*

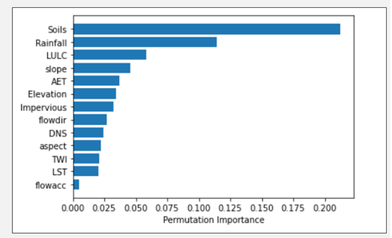

Figure 4. Permutation based factor importance

Conclusion and Recommendation

- The areas centered in the central and eastern areas are susceptible to the flood.
- Potential areas are identified in five levels (Very low, Low, Medium, High, and Very High) but field ground-truthing is recommended for flood vulnerable groups; and finer resolution should be targeted as per feasibility.

References

[1] I. Mallakpour and G. Villarini, "The changing nature of flooding across the central United States," *Nat. Clim. Chang.*, vol. 5, no. 3, pp. 250–254, 2015

[2] NOAA National Centers for Environmental Information (NCEI), "Storm Events Database," Noaa.gov. [Online]. Available: https://www.ncdc.noaa.gov/stormevents/. [Accessed: 09-Jun-2022].