Predictive Deep Learning for Flash Flood Management

PREPARED BY:

Steven M. Corns, PhD

Suzanna K. Long, PhD, PEM, F.ASEM

Jacob Hale

Bhanu Kanwar

Lauren Price

Missouri University of Science and Technology

PREPARED FOR:

Missouri Department of Transportation

Construction and Materials Division, Research Section

Predictive Deep Learning for Flash Flood Management

Dr. Steven Corns, PhD
Associate Professor
Engineering Management and Systems
Engineering Department
Missouri University of Science and
Technology

Dr. Suzanna Long, PhD
Professor and Department Chair
Engineering Management and Systems
Engineering Department
Missouri University of Science and
Technology

Lauren Price
Undergraduate Research Assistant
Engineering Management and Systems
Engineering Department
Missouri University of Science and
Technology

Jacob Hale
Graduate Research Assistant
Engineering Management and Systems
Engineering Department
Missouri University of Science and
Technology

Bhanu Kanwar
Graduate Research Assistant
Engineering Management and Systems
Engineering Department
Missouri University of Science and
Technology

1. Report No. cmr 21-001	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Predictive Deep Learning for Flash Flood Management		**5. Report Date** December 2020 Published: January 2021
		6. Performing Organization Code
7. Author(s) Steven M. Corns, PhD ORCID: 0000-0002-3685-2892; Suzanna K. Long, PhD, PEM, F.ASEM, ORCID: 0000-0001-6589-5528; Jacob Hale; Bhanu Kanwar; Lauren Price		**8. Performing Organization Report No.**
9. Performing Organization Name and Address Department of Engineering Management and Systems Engineering Missouri University of Science and Technology 600 W. 14th Street, Rolla, MO 65409		**10. Work Unit No.**
		11. Contract or Grant No. MoDOT project # TR202023
12. Sponsoring Agency Name and Address Missouri Department of Transportation (SPR-B) Construction and Materials Division P.O. Box 270 Jefferson City, MO 65102		**13. Type of Report and Period Covered** Final Report (April 10, 2020-December 31, 2020)
		14. Sponsoring Agency Code

15. Supplementary Notes

Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration. MoDOT research reports are available in the Innovation Library at https://www.modot.org/research-publications.

16. Abstract

This research was completed in tandem as a project funded through MoDOT and the Mid-America Transportation Center. It used deep learning methods, along with weather information from NOAA/National Weather Service and geospatial data from the USGS National Map and other public geospatial data sources, to develop forecasting tools capable of assessing the probability of flash flooding in high risk areas. These tools build on existing models developed by the USGS, FEMA, and others and were used to determine evacuation routing and detours to mitigate the potential for loss of life during flash floods. The project scope included analysis of publicly available data in Greene county in and around Springfield, MO as part of a pilot project in Missouri. This data was then used to determine the probability of flash flooding in order to model evacuation or detour planning modules that can be implemented to assure the safety of the community and highway personnel. These modules used existing rainfall data and weather forecasts in a three-day sliding window to include soil moisture in the flash flood predictions. The transportation safety or disaster planner can use these results to produce planning documents based on geospatial data and information to develop region-specific tools and response methods to potential flash flood events.

17. Key Words Deep learning; Flood management; Evacuation routing and planning	18. Distribution Statement No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.		
19. Security Classif. (of this report) Unclassified.	**20. Security Classif. (of this page)** Unclassified.	**21. No. of Pages** 51	**22. Price**
Form DOT F 1700.7 (8-72)	Reproduction of completed page authorized		

Table of Contents

List of Exhibits

List of Abbreviations

Alternating Decision Tree (ADT)
Artificial Neural Network (ANN)
Functional Tree (FT)
Geographic Information System (GIS)
Kernel Logistic Regression (KLR)
Light Detection and Ranging (Lidar)
Multilayer Perceptron (MLP)
National Hydrography Dataset (NHD)
National Oceanic and Atmospheric Administration (NOAA)
National Transportation Dataset (NTD)
National Weather Service (NWS)
Quadratic Discriminant Analysis (QDA)
S-band Doppler Weather Radar (SPOL)
Short Message Service (SMS)
Simulation of Urban Mobility (SUMO)
Support Vector Machine (SVM)
Support Vector Regression (SVR)
United States Geological Survey (USGS)

Acknowledgments

Disclaimer

The opinions, findings, and conclusions expressed in this document are those of the investigators. They are not necessarily those of the Missouri Department of Transportation, U.S. Department of Transportation, or Federal Highway Administration. This information does not constitute a standard or specification.

Abstract

This research was completed in tandem as a project funded through MoDOT and the Mid-America Transportation Center. It used deep learning methods, along with weather information from NOAA/National Weather Service and geospatial data from the USGS National Map and other public geospatial data sources, to develop forecasting tools capable of assessing the probability of flash flooding in high risk areas. These tools build on existing models developed by the USGS, FEMA, and others and were used to determine evacuation routing and detours to mitigate the potential for loss of life during flash floods. The project scope included analysis of publicly available data in Greene county in and around Springfield, MO as part of a pilot project in Missouri. This data was then used to determine the probability of flash flooding in order to model evacuation or detour planning modules that can be implemented to assure the safety of the community and highway personnel. These modules used existing rainfall data and weather forecasts in a three-day sliding window to include soil moisture in the flash flood predictions. The transportation safety or disaster planner can use these results to produce planning documents based on geospatial data and information to develop region-specific tools and response methods to potential flash flood events.

Executive Summary

This research uses publicly available geospatial data to create a historic flash flood database that is then used as an input for a deep learning model capable of classifying flash flood risk for discrete locations within an area of interest (AOI) as presented in Exhibit E-1. The project scope includes analysis of publicly available flash flood data for a subwatershed in Greene County, Missouri that frequently experiences flash floods. This data was procured from USGS, NOAA, and NWS. A framework is presented that extends the utility of standard flash flood susceptibility maps by adding a dynamic predictive component based on potential rainfall events.

Flash flood susceptibility maps are not created with transportation-specific use cases involved. Consequently, there is no methodology available that provides both risk quantification and optimal rerouting guidance. The algorithms used in this research capture the complex relationship between geospatial characteristics and rainfall data to classify locations on the basis of their flash
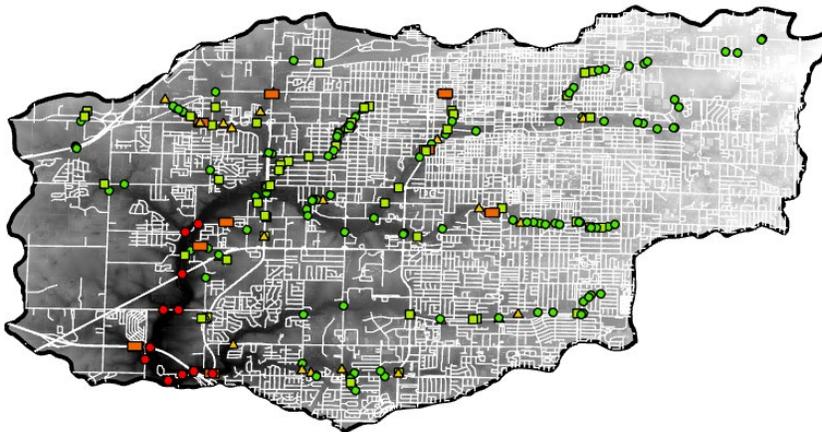
Exhibit E-1. Flash Flood Risk for AOI in Greene County, MO

flood risk. Elevation, slope, aspect, and curvature constitute the geospatial data whereas day-of and prior rainfall observations represent the latter. Three machine learning models were used: artificial neural network, logistic regression, and support vector machine. The artificial neural

network exhibited superior performance with a prediction accuracy of 85.23%. At present, there are no flash flood prediction models being used by practitioners or local decision makers.

An additional component of the framework is the determination of optimal rerouting protocols that takes into account in-route traffic and road segments at high risk for flash flood events. This feature provides transportation officials with critical information that can guide the deployment of resources in a timely manner to minimize risk exposure to motorists. Collectively, the framework presented here provides a suite of tools that are not currently in use at any level throughout the state.

1. Literature Review

Flash floods are one of the most frequently occurring and dangerous natural disasters. Study of the phenomena has received widespread scholarly attention due to the loss of life and financial and material damage caused. Research studies typically consist of developing static flash flood susceptibility maps based on artificial intelligence models. However, there is no methodology in the literature that extends this approach to dynamic use cases. A review of research in flash flood susceptibility maps, artificial intelligence, and traffic simulation is conducted and used to construct the dynamic framework presented in section two.

Flash Flood Susceptibility Maps

A flash flood susceptibility map is a data visualization medium that improves the effectiveness of watershed management (Janizadeh et al., 2019). This is accomplished by generating a heat map of flash flood risk based on historical events and geospatial features within some area of interest. Methodological specifics vary, but the general framework is consistent.

The following steps outline the general approach undertaken in the literature to create flash flood susceptibility maps. First, a flash flood inventory is created that consists of date, time, and location information of previous flash floods. A similar dataset is compiled for non-flash flood events to train a classification model. Second, geospatial data that influences flash flood events is gathered. Commonly used features include elevation and elevation-derived products such as slope, aspect, and curvature. These features are 'static' because they do not change over time. Alternatively, rainfall data is procured to capture the dynamic component of flash flood modelling. Lastly, statistical modeling, artificial intelligence approaches, or some hybrid ensemble are used to determine the relationship between model inputs and flash flood

1

susceptibility risk (Lopez and Rodriguez, 2020; Nguyen et al., 2020; Janizadeh et al., 2019; Bui et al., 2020; Bui et al., 2019; Ngo et al., 2018; Costache and Bui, 2019). The following section provides a brief overview of the literature regarding statistical and artificial intelligence modeling approaches used to generate flash flood susceptibility maps.

Machine Learning and Deep Learning Methods

Research efforts have been made to map the flash floods events and identify the regions prone to this natural hazard event. Saharia et al. (2016) relied on the flood events archived data for the last 78 years to develop a flood severity model based on different geomorphological and climatological variables. The proposed model was then used to identify various flood hotspots and conduct seasonality-based analysis. Different machine learning models were tested to identify suitable models for flash flood susceptibility mapping by Janizadeh et al. (2019). Alternating decision tree (ADT), functional tree (FT), kernel logistic regression (KLR), multilayer perceptron (MLP), and quadratic discriminant analysis (QDA) algorithms were developed using historical data to develop and improve the existing flood mapping frameworks to update the current relief and rescue protocols. An improved real-time warning system was proposed by Acosta-Coli et al. (2018) to improve the existing warning methods for the pluvial flash floods triggered by heavy rainfall. A web-based application was implemented by the research team to elicit a timely response from the people during a potentially hazardous flash flood event.

Due to advancements made in the field of data analysis and machine learning in recent years, it has become easier to implement suitable machine learning algorithms and therefore develop more robust flood forecasting model architectures. Precipitation data from a dual-

polarization S-band Doppler weather radar (SPOL) was used as an input for the regression model by Lopez and Rodriguez (2020) to calculate timely warnings for various watersheds in the city of Sao Paulo, Brazil. A novel support vector regression (SVR) based model was proposed by Wu et al. (2019) to improve the lead times in the mountainous regions of China and develop better flash flood warning and response systems. Flash floods datasets from 1984 to 2012 were used to identify important variables needed for the model architecture to predict peak values for flood forecasting. A flash flood warning system using short message service (SMS) was also proposed by Castro et al. (2013) using the regression model based on water level and speed variables. This SMS based warning system is activated when the water level exceeds the safety threshold value so that the people living in the flood-prone areas can be made aware of the impending danger on a timely basis. This tool can also be used by the local authorities to prepare better flash flood management protocols and operations. Satellite-derived datasets were used by Chiang et al. (2007) along with the rain gauge measurements to devise a neural network-based hydrologic model that used the merged time-series dataset to analyze both intensity and pattern of the precipitation events. This model can be used to process the satellite-derived datasets for areas where gauge data points are scarce and develop respective flood forecasting models.

Compared to other natural disasters, flash flood events are relatively hard to forecast which makes it very important to inspect and update the current forecasting approaches which can then be used by the concerned authorities to develop better flash flood management policies. Therefore, machine learning and deep learning provide a powerful tool to identify the complex relationships between flood influencing factors and flash flood risk quantification. The result of these classifications can be used to identify and remove flood-affected road segments from a transportation network for re-routing simulation as presented in the following section.

Traffic Simulation

A traffic simulation is used to demonstrate how closing flood affected road segments would impact traffic flow. Simulating traffic involves creating a network that includes an origin-destination matrix, vehicle speeds, vehicle types, speed limits, traffic lights, and the physical roads themselves. The Simulation of Urban Mobility software (SUMO) integrated with deep learning models has been used to predict traffic congestion and flow from Hurricane Harvey in 2017 (Fan et al., 2020). In the simulation, the focus was major highways (50 km/hr and over). Corns et al. (2019) used the SUMO software to model optimal re-routing protocols for flood affected road segments near Valley Park, Missouri. SUMO's capacity to integrate Python functions results in greater customization that improves the quality of the simulations. Therefore, SUMO is an ideal software to capture traffic dynamics in a flash flood context. The integration of Python with SUMO will be explained in the Methodology section.

Summary

A review of the literature revealed that flash flood susceptibility maps have begun receiving greater scholastic attention. The first step in creating these maps is developing a historic flash flood database. Some of the most commonly gathered data include elevation, elevation-derived products, and rainfall. NOAA maintains a database with specific information on various natural disasters including flash floods. However, flash flood affected road segment information must be parsed from narrative episodes provided by witnesses. This results in a time-consuming procedure that inhibits the development of a database that can be used for planning purposes. Therefore, there is an opportunity to build a database that sets a standard for future data collection efforts that can then be used as an input for state-of-the-art classification techniques.

Several of the flash flood susceptibility maps present in the literature were generated using sophisticated modelling techniques such as machine learning and deep learning. As the quantity and quality of data continues to increase, so too will the application of these methods. One shortcoming of these models is the lack of predictive capabilities provided in practical applications. A model capable of providing dynamic results as rainfall is reported will provide considerable utility to local decision makers. Once the model has identified high flash flood risk road segments, those roads can then be proactively removed from the road network. To model this decision, a simulation can be constructed that captures subsequent traffic behavior. The literature is full of studies addressing each of these problems independently, but no comprehensive methodology has been presented. The methodology provided in the next section addresses this gap.

2. Methodology

The objective of this study is to provide emergency management professionals with a tool that quantifies flash flood risk at discrete locations of interest. The methodology presented in this section follows the sequence illustrated in Exhibit 2-1. Step 1 consists of developing a historic flash flood database using temporal and geospatial information of previous flash flood events. Step 2 involves the collection of explanatory data for the flash flood database developed. Step 3 extracts and converts the procured explanatory data into a tabular format for processing. Step 4 applies machine learning and deep learning techniques to predict flash flood risk of locations of interest. Lastly, Step 5 provides a traffic re-routing simulation based on the roads identified by the prediction model. The remainder of this section provides a litany to reproduce the results presented.
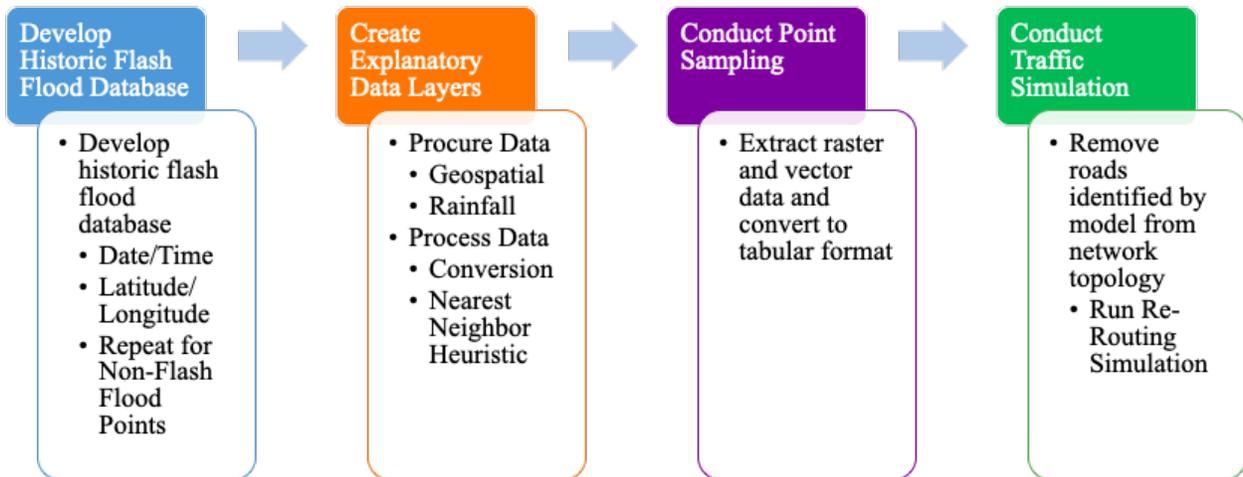
Exhibit 2-1. Model Framework

Develop Historic Flash Flood Database

Development of a historic flash flood database consists of gathering information related to previous flash flood events. The National Oceanic and Atmospheric Administration (NOAA) maintains the Storm Events Database that provides this information (NOAA, 2020). Exhibit 2-2 provides an example of the user interface provided. The tool provides the user with the ability to procure unique event type information at the county level resolution over a specified period.
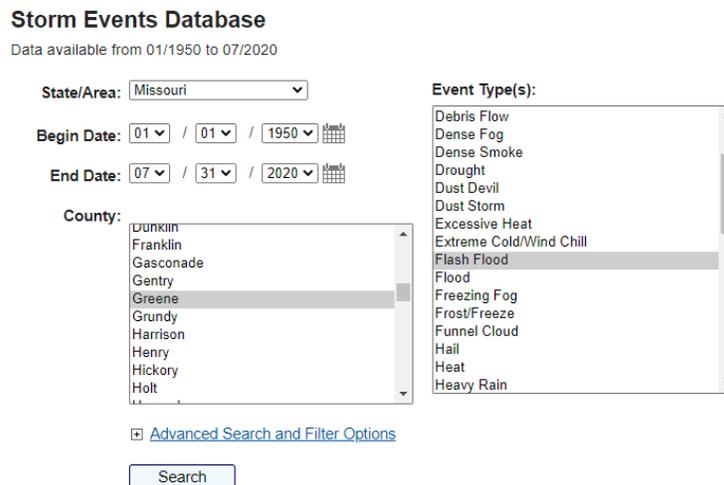
Exhibit 2-2. Storm Events Database User Interface

Once this information is specified, the user can then download an Excel or comma separated value file that includes temporal and geospatial information. However, upon further inspection it is clear that the information provided is not consistent. Each data entry consists of unique latitude and longitudinal coordinates in addition to narrative episodes provided by the spotter responsible for the entry. Exhibit 2-3 provides an example of database entries for two events. Information provided pertaining to affected roads is highlighted in yellow. Exhibit 2-4 provides a visual reference between the stated location of the event and the actual location of the

affected road given the information provided in the narrative. Due to this geographical

discrepancy, each data entry must be investigated to determine the actual location of the flash

flood affected road segments.

EVENT ID	BEGIN DATE	LATITUDE	LONGITUDE	EVENT NARRATIVE	EPISODE NARRATIVE
5435706	1/12/05	37.2	-93.3		A slow-moving storm system caused heavy rain to occur across much of southern and central Missouri. This event followed quickly on the heels of a previous flood event that occurred from the 4th through the 6th of January, therefore soils were nearly saturated at the onset of the event. The lack of January vegetation also contributed to increased runoff and flooding. In Greene County, the primary areas that flooded were low water crossings and low-lying areas. Other specific locations that were affected by flooding include, areas along Ward Branch Creek in Springfield, a section of Highway CC two miles west of Fair Grove, the intersection of Farm Road 235 and Highway E, the intersection of Farm Roads 231 and 2.
5470179	8/21/05	37.2	-93.3	A slow-moving thunderstorm with heavy rain caused the South Dry Sac Creek to rise out of its banks and flood a section of East Valley Water Mill Road in northeast Springfield. Several other sections of Springfield experienced flash flooding, which prompted numerous water rescues by Springfield Fire and Police Departments. One house experienced six inches of flowing water.	

**Exhibit 2-3.** Storm Events Database Sample
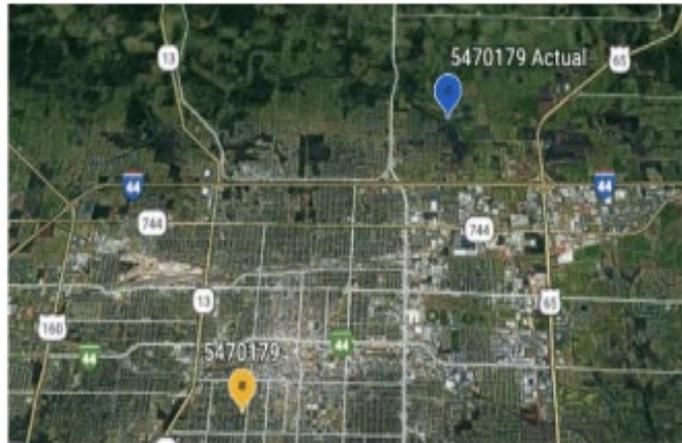
Exhibit 2-4. Geographical Discrepancy Example

8

After this procedure has been conducted for each of the entries, the resulting database can then be uploaded to a geographic information system (GIS) software for visualization. Exhibit 2-5 provides the output of this procedure (generated in ArcGIS Pro) for an area of interest that will be discussed further in the next section. The red points represent the locations determined by the narrative episode investigation and are superimposed on a road network procured from the National Transportation Dataset (NTD) (USGS, 2020a). For model training purposes, a dataset consisting of non-flash flood locations must also be developed. The results of this procedure are illustrated in Exhibit 2-6 and a represented by green squares.
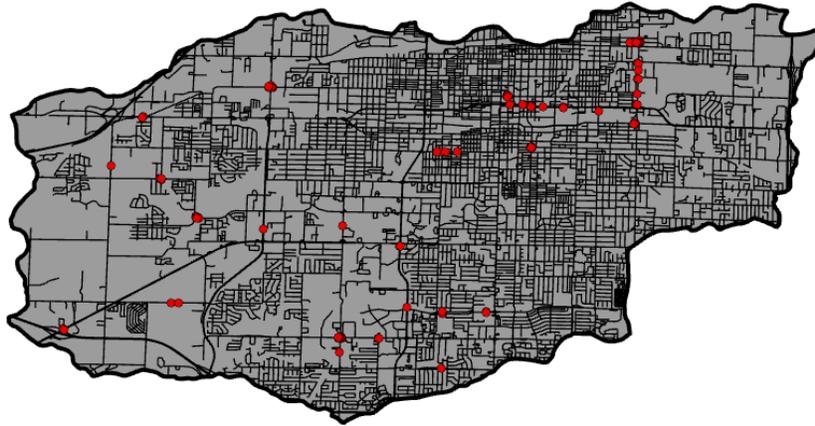

Exhibit 2-5. Flash Flood Locations
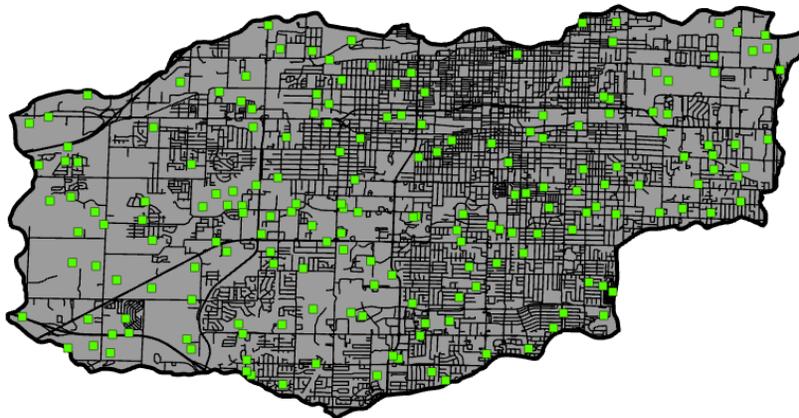
**Exhibit 2-6.** Non-flash Flood Locations

After the flash flood database is constructed, the next step is gathering flash flood explanatory data. The next section provides guidance on procuring and processing that data.

Create Explanatory Data Layers

Based on the literature, four geospatial features were identified as flash flood influencing factors: elevation, slope, aspect, and curvature. Elevation can be procured from the United States Geological Survey's (USGS) National Map (USGS, 2020b). For the area of interest presented in the previous section, elevation products are available in light detection and ranging (Lidar) format. To generate the elevation profile, 135 files were downloaded as illustrated by Exhibit 2-7. Each square represented in the 15x9 grid corresponds to a unique lidar file. This gridded visualization of the lidar data is commonly referred to as an LAS dataset. Using the LAS Dataset to Raster tool and clipping procedures, the gridded representation can then be converted into a raster elevation layer as seen in Exhibit 2-8.
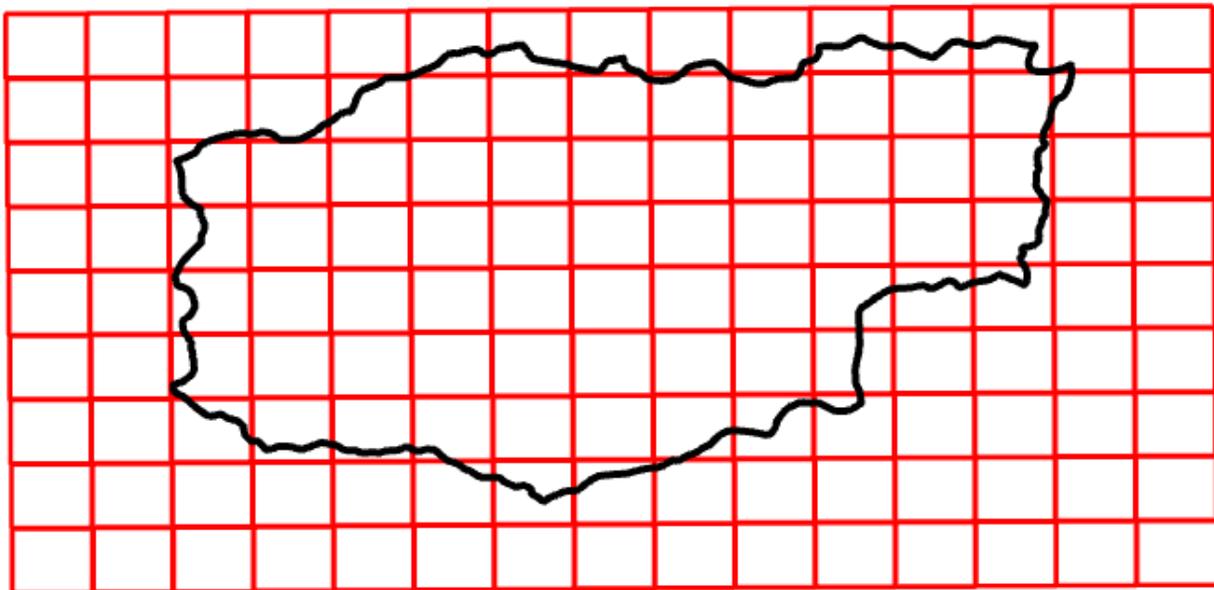

Exhibit 2-7. Lidar File Representation
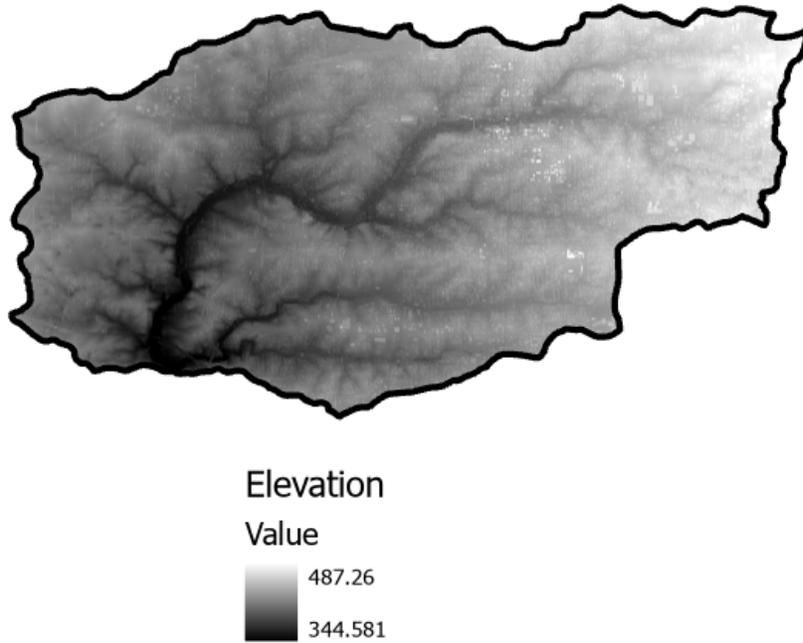
Exhibit 2-8. AOI Elevation Profile

Slope, aspect, and curvature layers can be quickly generated using the correspondingly named tools in a GIS software. Each layer requires the elevation layer presented in Exhibit 2-8 as an input. Exhibits 2-9 through 2-11 represent the outputs of this operation.
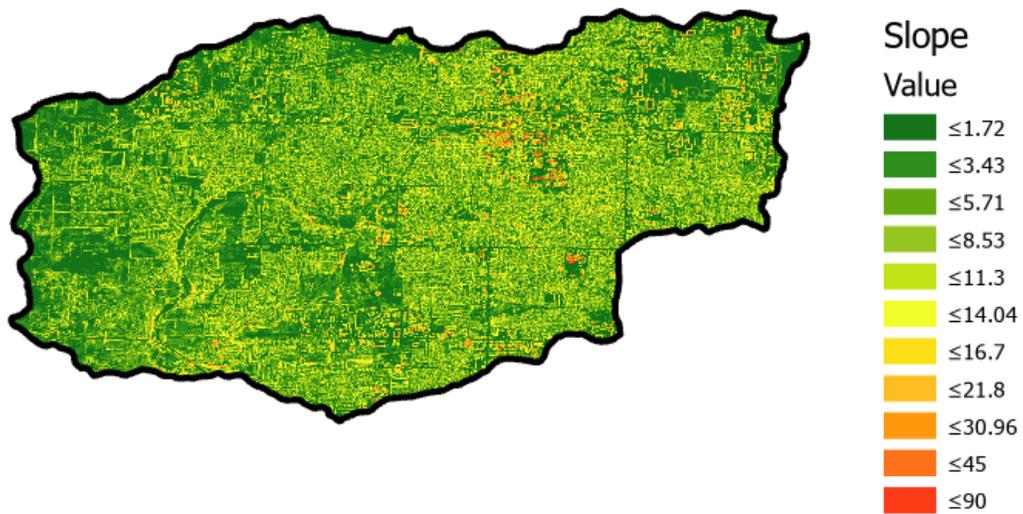
Exhibit 2-9. AOI Slope Profile
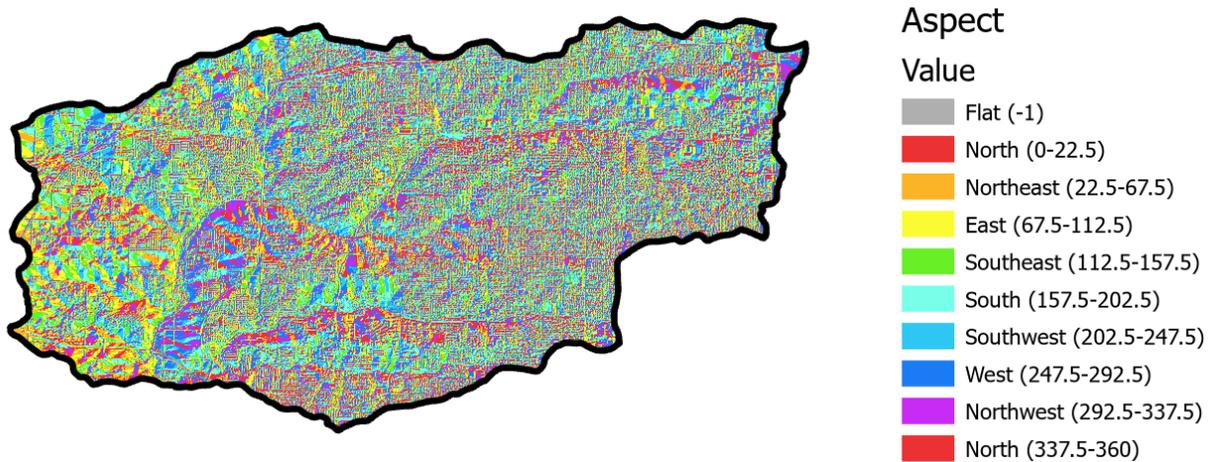
Exhibit 2-10. AOI Aspect Profile
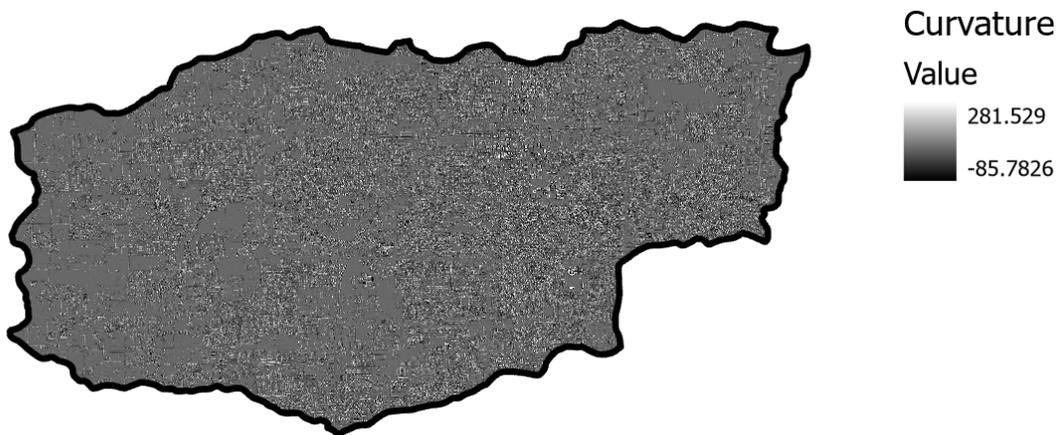
Exhibit 2-11. AOI Curvature Profile

Elevation, slope, aspect, and curvature each represent static explanatory layers that do not change significantly over time. Alternatively, rainfall provides a dynamic input that varies. Each of the entries in the flash flood database is associated with a specific date. Therefore, rainfall data can be procured from the National Weather Service (NWS) Archive. Rainfall data from January 1, 2005-June 27, 2017 is available in point format (NWS, 2020a). A nearest neighbor heuristic is

used to apply a rainfall amount to flash flood and non-flash flood locations. From June 28, 2017-Present, rainfall data is available in a raster format (NWS, 2020b). In these instances, rainfall amount is assigned based on the underlying pixel value. Flash floods often occur quickly following a rainfall event. However, soil moisture before the causative rainfall event plays a significant role in the flash flood event. Therefore, rainfall data for each of the three days prior to the flash flood event is also captured to serve as a proxy for soil moisture. Once the rainfall amounts have been assigned, explanatory data layer values can be extracted using point sampling as described in the following section.

Conduct Point Sampling

After the flash flood database has been compiled and explanatory data layers have been generated, point sampling is the final step prior to classification models being applied. Point sampling consists of extracting data layer values at discrete locations. In this context, explanatory raster layer values for each of the discrete flash flood and non-flash flood locations are extracted and compiled in a tabular format. Using the ArcGIS Extract Multi Values to Points tool, this can be done quickly. Exhibit 2-12 is the final visual output. Exhibit 2-13 is a sample of the corresponding attribute table for the flash flood test set.
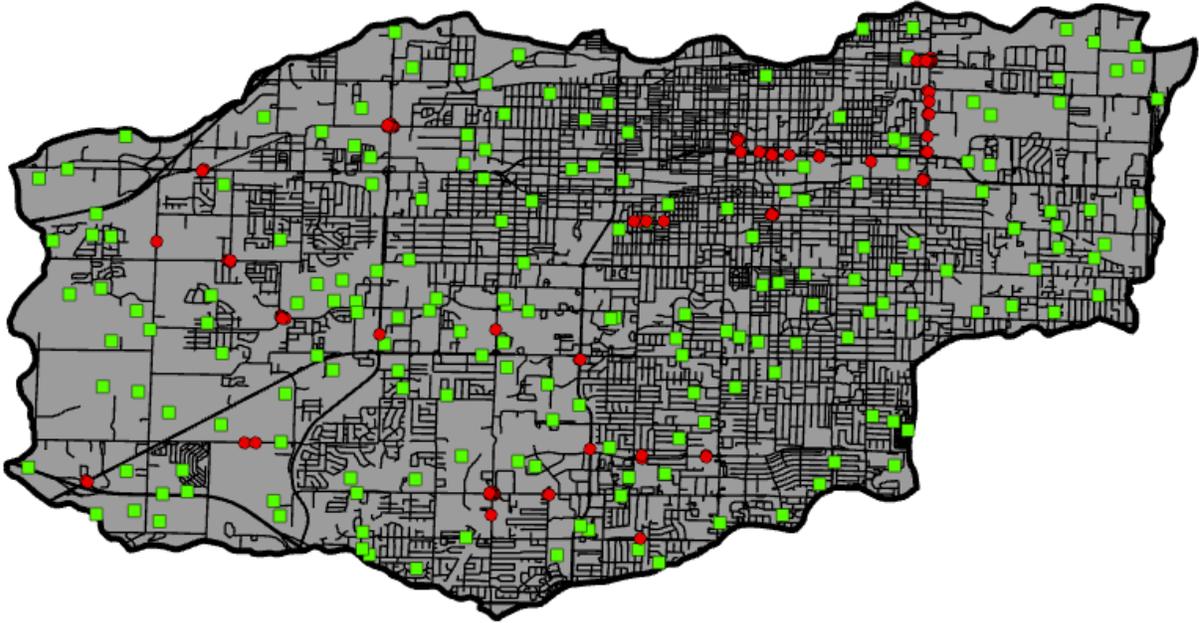
Exhibit 2-12. Final Visual Output

Each data point in both the flash flood and non-flash flood datasets has a similar entry. ID is a unique identification number that denotes that point's position in the dataset. Name refers to the original Storm Event's Database identification number. Latitude and longitude are the coordinates for the entry. Date provides the start of the event. Each of the rainfall points corresponds to the observed rainfall at the entry. Summation of antecedent rainfall is the result of adding up all prior rainfall amounts. Curvature, aspect, slope, and elevation are the result of point sampling extractions explained earlier in this section. With the flash flood database now in tabular format it can be routinely used in an array of machine learning and deep learning classification tools. Model specifics are presented in the following section.

14

ID	1
Name	22814a
Latitude	37.15
Longitude	-93.31
Date	2007-03-20
Rainfall Day of (in)	0.39
Rainfall 1 Day Before (in)	0.00
Rainfall 2 Days Before (in)	0.01
Rainfall 3 Days Before (in)	0.00
Summation of Antecedent Rainfall (in)	0.01
Curvature (dimensionless)	0.33
Aspect (degrees)	227.36
Slope (degrees)	1.05
Elevation (m)	383.49

Exhibit 2-13. Point Sampling Output Example

Machine Learning and Deep Learning Models

In order to identify points (with latitude and longitude coordinates) susceptible to flash flooding, different classification-based machine learning, and deep learning techniques were implemented on the output dataset obtained from the point sampling operation. The dataset consists of two classes, 0 and 1 which represent non-flash flood and flash flood labels, respectively. Artificial neural network (ANN), logistic regression, and support vector machine (SVM) algorithms were used to develop classification models and classify the binary labels into two different categories. The dataset contains a total of 350 points/observations with respective values for variables or features mentioned in Exhibit 2-13. Out of these 350 points, 185 points belong to class label 0 i.e. non-flash flood points or locations whereas 165 points belong to label 1 i.e. flash flood points or locations as shown in Exhibit 2-14.
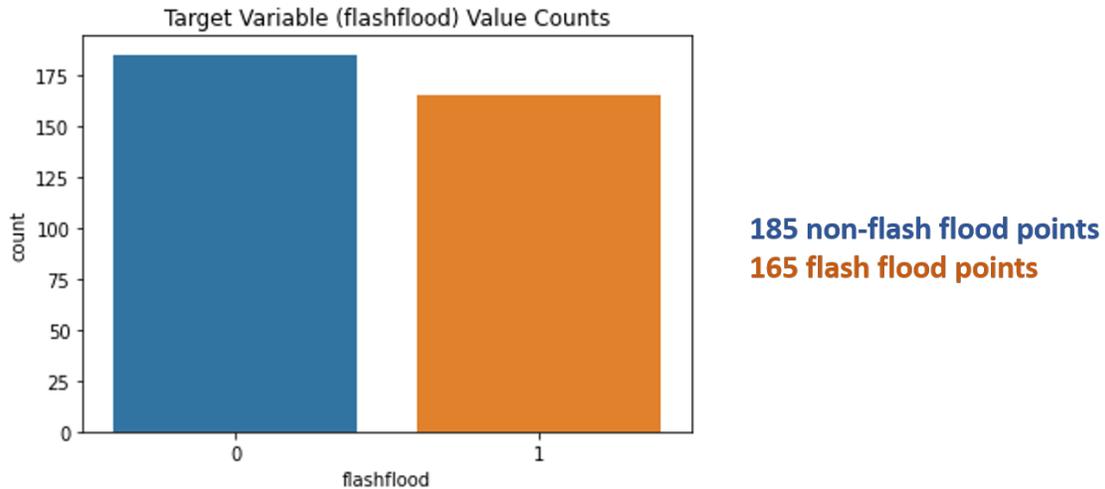
**Exhibit 2-14.** Flash Flood Data Point Distribution

Exhibit 2-15 shows the artificial neuron network (ANN) developed to receive the output data from point sampling as input and perform the binary classification task to identify potential flash flood locations. The different layers of a neural network contain neurons that represent mathematical functions required to accept input data and generate output values for the succeeding layer. The neurons in the input layer receive the input data's variables and pass it on to the hidden layer which applies the activation function to the received data before further passing its respective output values to the output layer. The output from the hidden layer is processed by the neurons in the output layer before generating a Boolean value for the classified label. For the binary classification task, the sigmoid activation function is utilized to generate the required output values in the range [0,1]. The output values generated by this sigmoid function in the final output layer of the neural network correspond to probability values of a given observation belonging to class 1 or positive class representing flash flood-prone locations. By default, any output probability value below the threshold value of 0.5 is labeled as class 0 (non-flash flood-prone locations) and a value above 0.5 falls under class 1 (flash flood-prone locations). Apart from this default probability

threshold value of 0.5, different threshold values of 0.6, 0.7, 0.8, and 0.9 are also used to test the models and identify respective flash flood-prone locations belonging to the positive class 1.
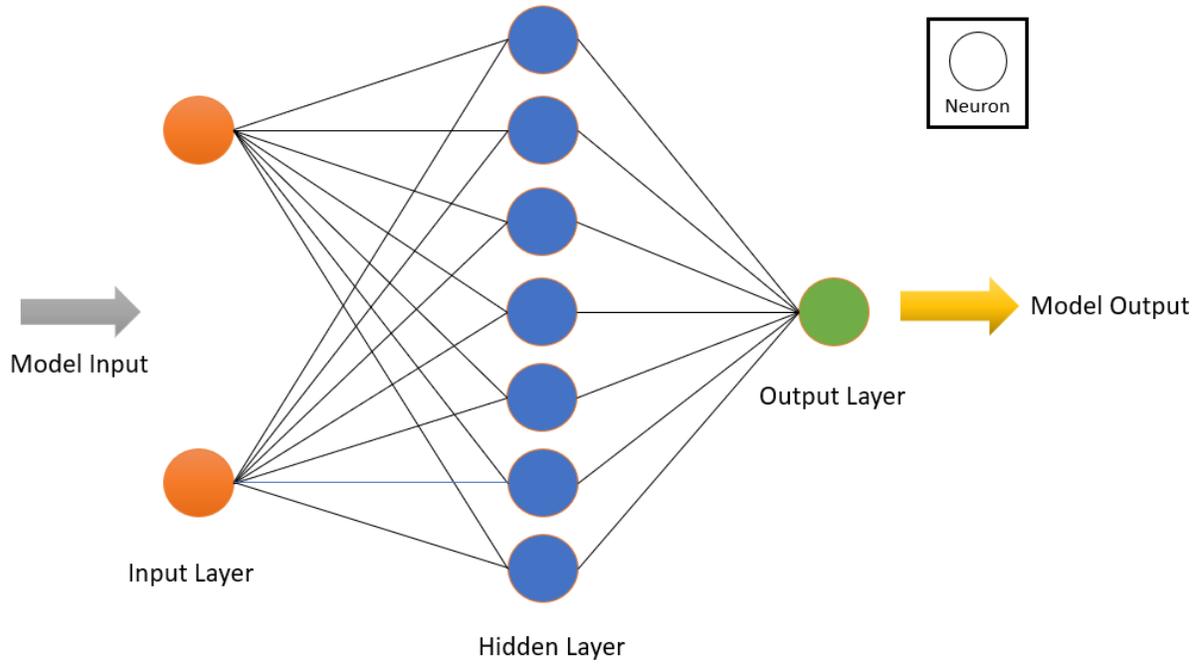
**Exhibit 2-15.** Artificial Neural Network

A model based on the logistic regression model is also developed to analyze the variables from the 350 input data points and then generate the binary output variable based on the logistic function as displayed in Exhibit 2-16. The accuracy of this model is then compared with the performance of the above-mentioned neural network model to conduct further analysis.
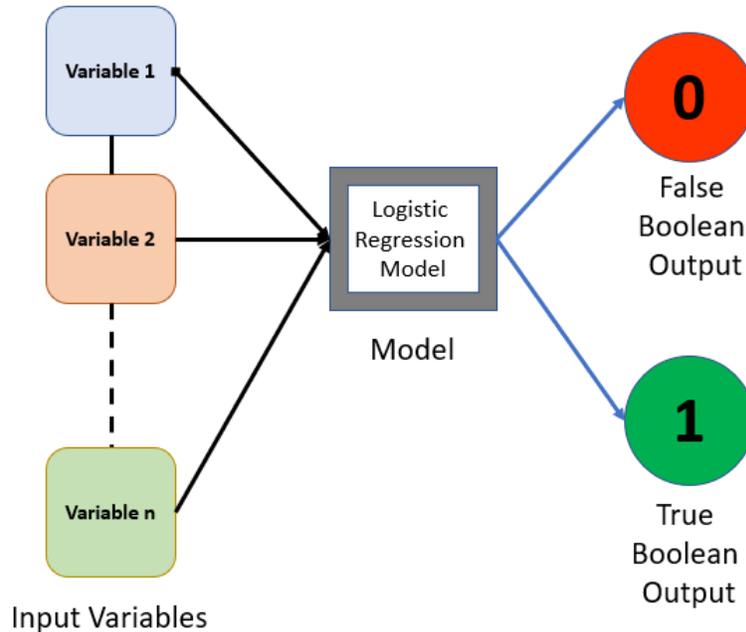
Exhibit 2-16. Logistic Regression

A third machine learning model, support vector machine (SVM) is also proposed and implemented for the classification task as exhibited in Exhibit 2-17. A support vector machine is a supervised learning algorithm that receives the labeled input data to further partition it into two different classes, class 0 and class 1. We can plot the available observations in an n-dimensional space (where 'n' represents the number of features or variables) and then identify the suitable hyperplane to distinguish between the distinct classes. In binary classification tasks, the hyperplane is a decision boundary that separates the two classes, and the support vectors are the data points adjacent to it. For the project, the input data is classified using support vector machine algorithms based on both linear and sigmoid kernels, respectively. A 'GridSearch' approach was also implemented to identify the optimum values of the support vector machine hyperparameters. Another support vector algorithm was then executed using the hyperparameter values obtained from the 'GridSearch' process. The performance metrics of both the support vector machine

models are analyzed and compared with the similar metrics from both artificial neural network and logistic regression models to choose the best classification model.
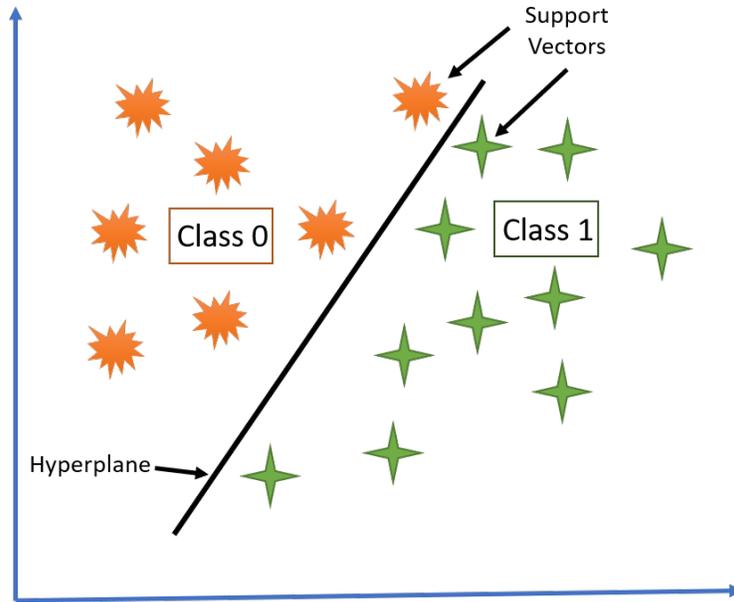
Exhibit 2-17. Support Vector Machine

The dataset derived from the point sampling process is inspected using Python 3.6 programming language before conducting data analysis operations (data processing, data cleaning, and exploratory data analysis, etc.) to identify relevant variables for our models. The processed 350 data points are divided between two different datasets, the training set, and the testing set. 75% of the total dataset i.e. 262 observations are allocated to the training set and the remaining 25% i.e. 88 observations are allotted to the other testing dataset. The training dataset consists of labeled variables information for 262 observations needed for training the models while the testing dataset is made up of variable information for the remaining 88 observations along with the respective truth values needed to test the accuracy of the classification models. Both training and testing datasets are then converted into suitable data formats to be used for analysis using artificial

neural network, logistic regression, and support vector machines as shown in Exhibit 2-18. In the next section, roads identified as flood affected by the best model are then removed from the road network in a simulation setting.
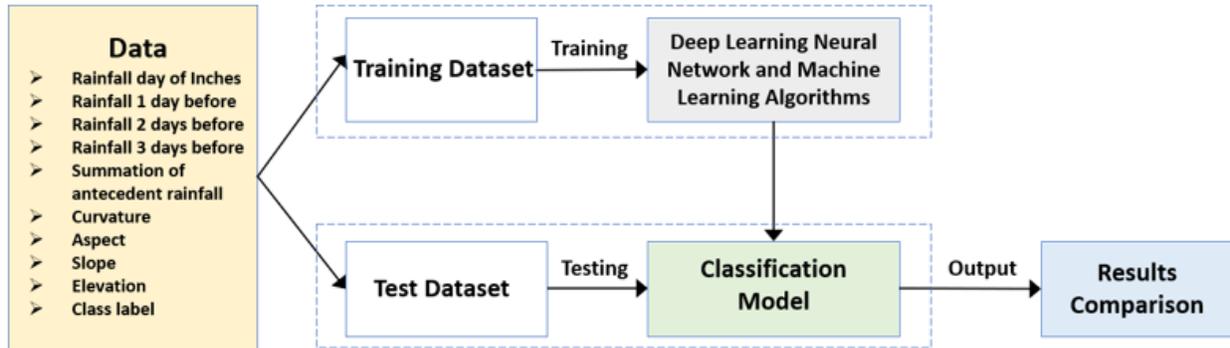

Exhibit 2-18. Deep Learning Neural Network and Machine Learning Algorithms Framework

Traffic Simulation

The simulation used to reroute traffic based on the best classification model is used as an input for SUMO. After the neural network gave an output of 1 (flooded) or 0 (not flooded) a network was formed for the values given a "1". The networks were exported from Open Street Maps (Exhibit 2-19) and converted to an xml document. The xml network was then edited using the text editor NetEdit. This application allows the user to edit the converted networks, for this report it was used to delete the roads that were predicted to flood (Exhibit 2-20). Once the networks had the appropriate roads removed a Python file was generated and integrated to assign "random trips" for the cars in the simulation. This was considered to be the standard flow of traffic. One the Python file is integrated a SUMO simulation was generated and data was extracted to predict the travel time. The travel time was then compared to the original travel time

i.e. if the roads were not flooded (Exhibit 2-21). In the next section a case study is presented that demonstrates methodological effectiveness.
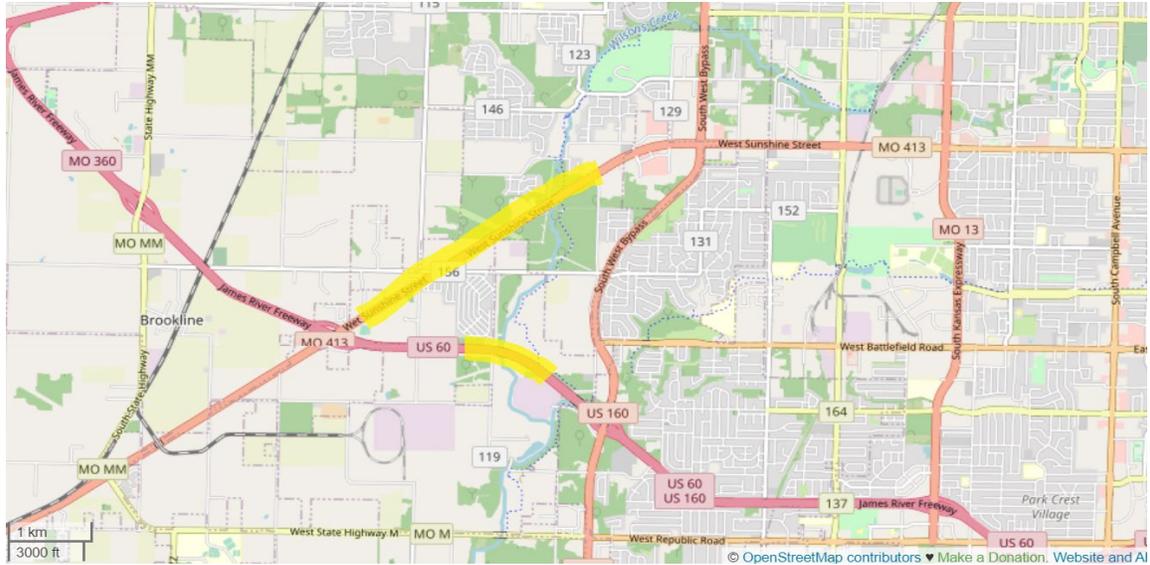
Exhibit 2-19. Open Street Map Example with Roads Highlighted
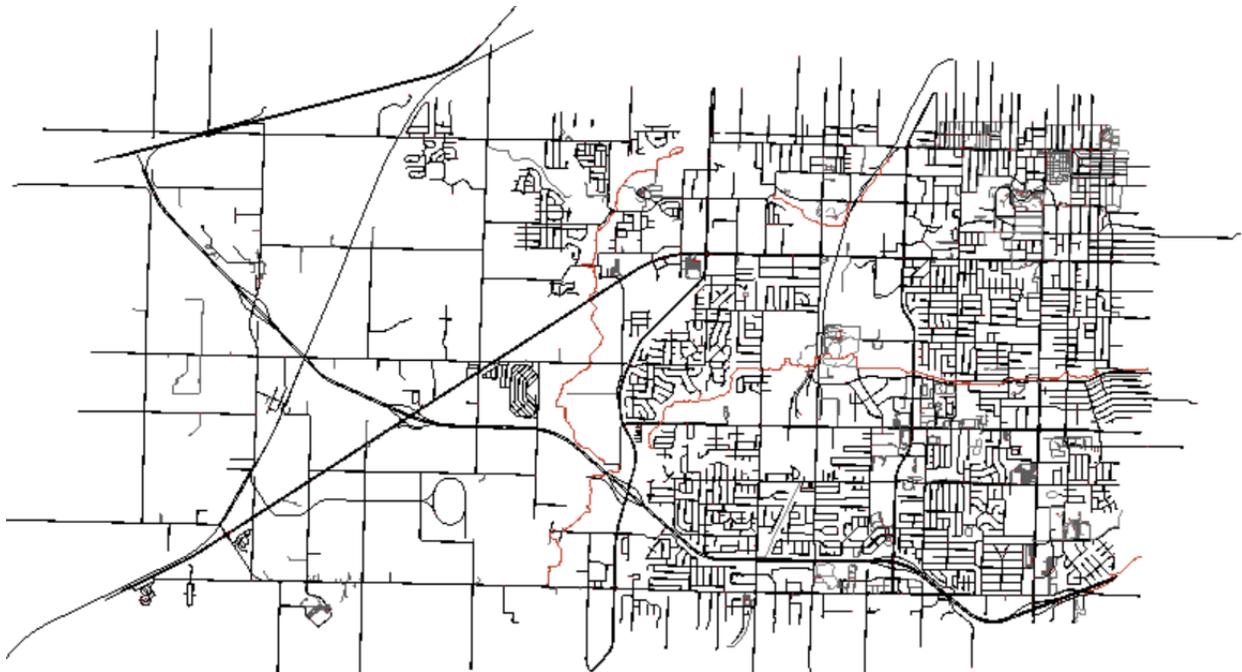
**Exhibit 2-20.** Original Road Network with Flowlines in Red
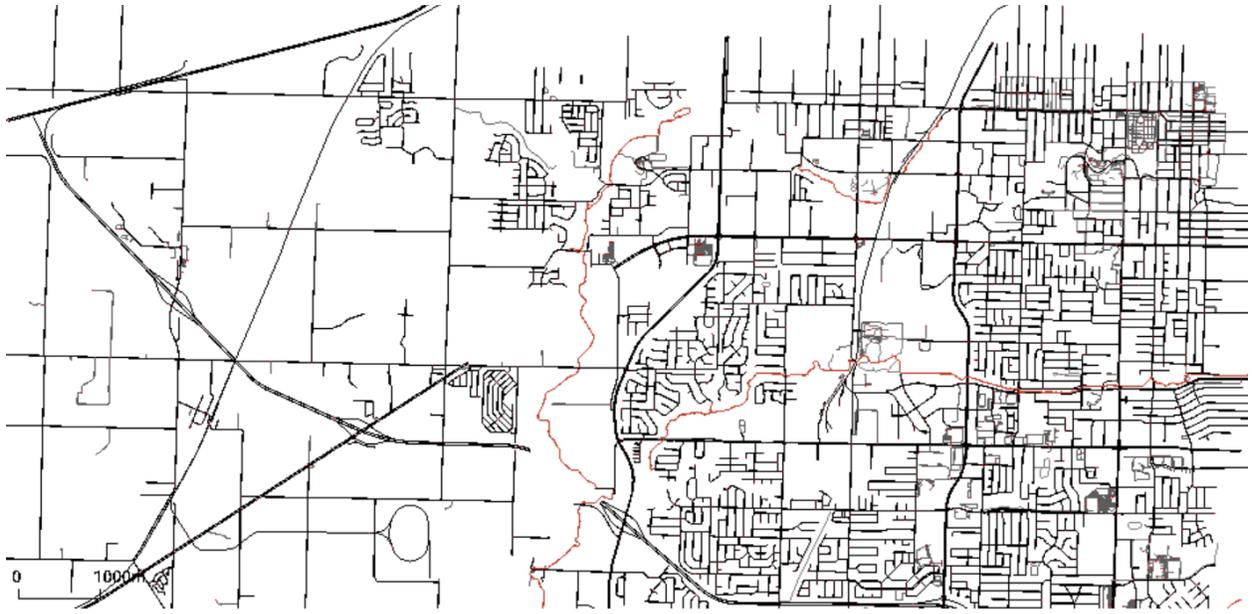
Exhibit 2-21. Road Network with Roads Removed

3. Results and Discussion

In consultation with key stakeholders, several counties in Missouri were identified as potential test sites. Ultimately, Greene County was chosen to demonstrate methodological efficacy given data availability and flash flood frequency.

Study Area

Greene County is located in the southwestern part of Missouri as illustrated by Exhibit 3-1 (Benbennick, 2006). The county includes the Springfield Metropolitan Area with an estimated population of almost 170,000 residents (Census Bureau, 2020).
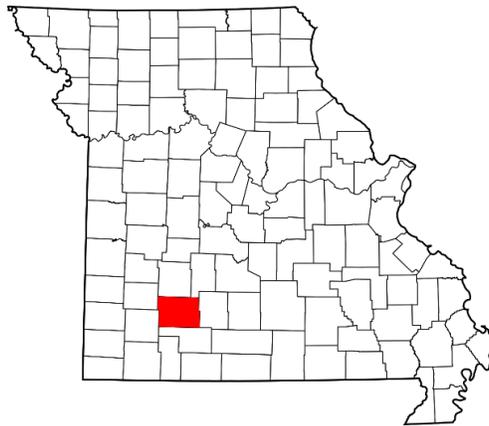
Exhibit 3-1. Greene County, Missouri

A population center of this size generally results in high traffic flow and significant economic operations. Therefore, flash floods pose a significant risk to motorists and the region's economic livelihood. While rainfall is the most common flash flood inducing factor, it is generally dependent on watershed characteristics. Exhibit 3-2 presents the flash flood locations from the Storm Events Database for Greene County with respect to 12-digit hydrologic unit code boundaries (subwatershed) (USGS, 2020c). The purple line corresponds to the county outline

and the green lines delineate the subwatershed boundaries. Discrete flash flood locations are represented as a heat map to demonstrate the high frequency of flash floods within a specific subwatershed.
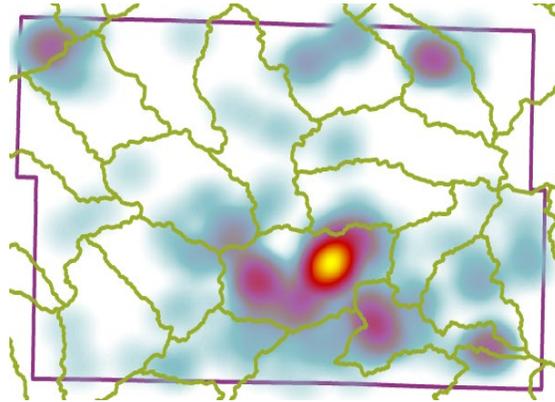
Exhibit 3-2. Greene County Flash Flood Hotspot

This evidence justifies the selection of this subwatershed as a test bed for the methodology presented. Exhibit 3-3 was created by following the steps presented in Section 2 for developing a historic flash flood database, creating explanatory data layers, and conducting point sampling for discrete locations. Explanatory data is then extracted in tabular format and used as an input in machine learning and deep learning models. The results of those procedures are presented in the next section.
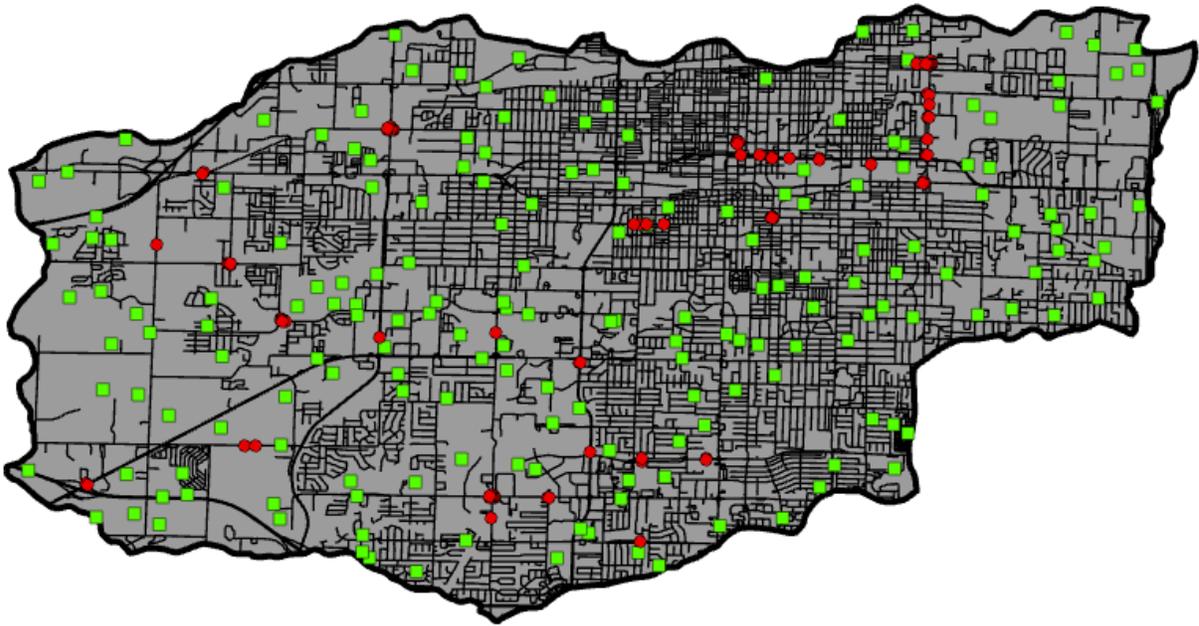
Exhibit 3-3. Final Visual Output

Machine Learning and Deep Learning Model Results

Three different types of artificial intelligence models, artificial neural network (ANN), logistic

regression, and support vector machine (SVM) are implemented to conduct the classification

task and identify the locations prone to flash flooding events. The artificial neural network

consists of 9 different input parameters or variables, the 'Adam' learning rate optimizer, and the

'sigmoid' activation function. The hyperparameters, 'epochs' and 'batch size' are assigned

values of 1000 and 10 respectively. The artificial neural network is run on the test data

containing feature information for 88 locations and delivers a classification accuracy of 85.23%.

The model correctly classified a total of 75 out of 88 locations and has an error rate of 14.77%.

The confusion matrix is shown in Exhibit 3-4 and displays the model performance of this

'sigmoid' based artificial neural network. Out of 88 total observations, the model identified 37

locations that are prone to flash floods and assigned them class labels equal to one. The

remaining 51 observations were classified as non-flash flood-prone locations by the classifier. A total of 46 and 42 observations are labeled as true flash flood and non-flash flood locations respectively in the testing dataset.
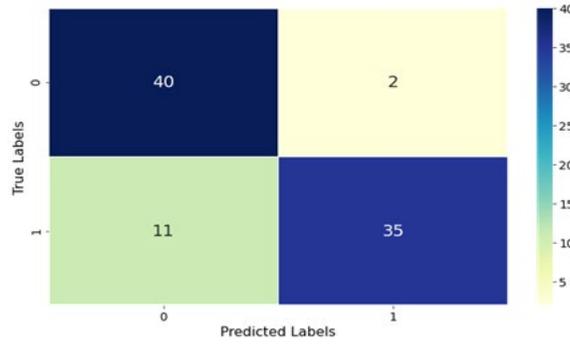
Exhibit 3-4. Artificial Neural Network Confusion Matrix

A second logistic model is implemented using the same dataset with a resulting classification accuracy score of 75% which is less than the artificial neural network's similar performance metric. Also, the error rate of this regression classifier has increased to 25%. Exhibit 3-5 shows that as compared to the true positive (label 1) and true negative (label 0) class values of 46 and 42 respectively, the logistic regression model classified 30 locations that are susceptible to flash flooding whereas 58 locations were predicted to have no flash flooding events.
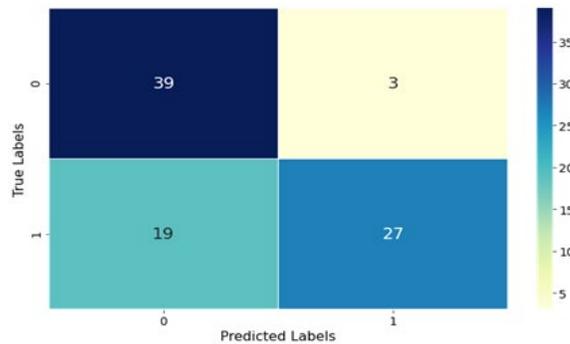
Exhibit 3-5. Logistic Regression Confusion Matrix

The support vector machine (SVM) classifiers based on both 'linear' and 'sigmoid' kernels performed worse than the other two classifiers and produced accuracy scores of 62.50% and 59.09% respectively. Its hyperparameters (Misclassification cost C, Gamma and kernel) were then tuned using the GridSearch approach and the accuracy score of the model improved to 80.68% as shown by the confusion matrix in Exhibit 3-6. Here, the error rate of 19.32% is less than the error rate of the logistic regression model but more than the neural network's error rate.
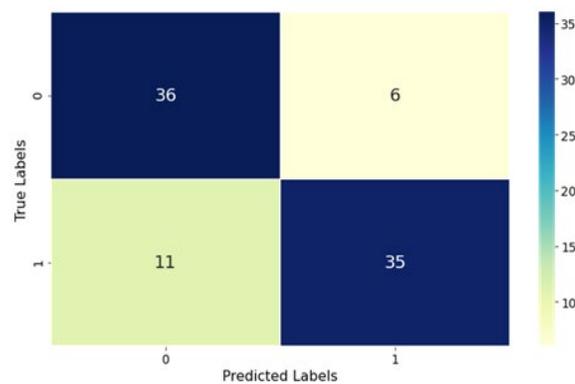
Exhibit 3-6. GridSearch Support Vector Machine Confusion Matrix

Finally, the artificial neural network-based classifier with the highest accuracy score of 85.23% is chosen to test the dataset containing randomly determined rainfall amounts for the selected test region of Greene County, Missouri.

Case Study

Given the accuracy of the classification model presented, it is appropriate to apply the model to locations that are of interest to local stakeholders. One such set of locations would be the intersection of flowlines, extracted from the National Hydrography Dataset (NHD), and the road network. Exhibit 3-7 provides a visualization of these locations (220 points) superimposed on the elevation profile for the AOI.
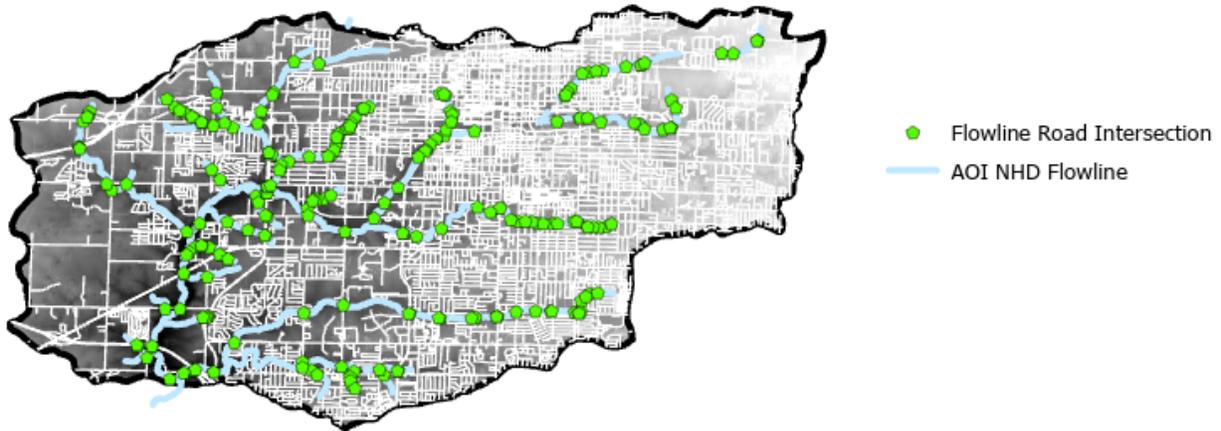
Exhibit 3-7. Case Study Visual

Using the methodology presented, the explanatory data layers and subsequent point sampling can be quickly compiled. However, this case study is intended to simulate future events. Therefore, the rainfall amount is randomly generated and assigned to each of the points. Classification results for each of the points are presented in the following section.

Case Study Classification Results

A test dataset with random rainfall amounts for 220 different locations in Greene County is fed into the artificial neural network model with the highest accuracy score of 85.23% to identify the locations susceptible to flash flood events. Using the 'sigmoid' activation function based binary classifier, a set of locations is identified for different probability threshold values of 0.5 (50%), 0.6 (60%), 0.7 (70%), 0.8 (80%) and 0.9 (90%). The model automatically assigns the class label value of 1 to all the observations above these threshold values and classifies them as potential flash flood-prone locations. This test dataset is analyzed and processed before passing it in the selected artificial neural network. Based on different data processing operations, the number of relevant variables or features in this test dataset is reduced to 9 which is the same number of variables used to test and compare different classification models previously. After implementing

the neural network on this new dataset, the classifier identified 22, 17, 14, 13, and 12 flash flood-prone locations for the different probability threshold values of 50%, 60%, 70%, 80%, and 90%, respectively. Model classification results are visually presented in Exhibit 3-8. Additionally, a histogram of classifications is also provided in Exhibit 3-9. An example simulation is provided for a set of road segments that possess flash flood risk probabilities greater than 90%.

●	≤0.078794
■	≤0.212014
▲	≤0.441440
■	≤0.742652
⬡	≤0.999870
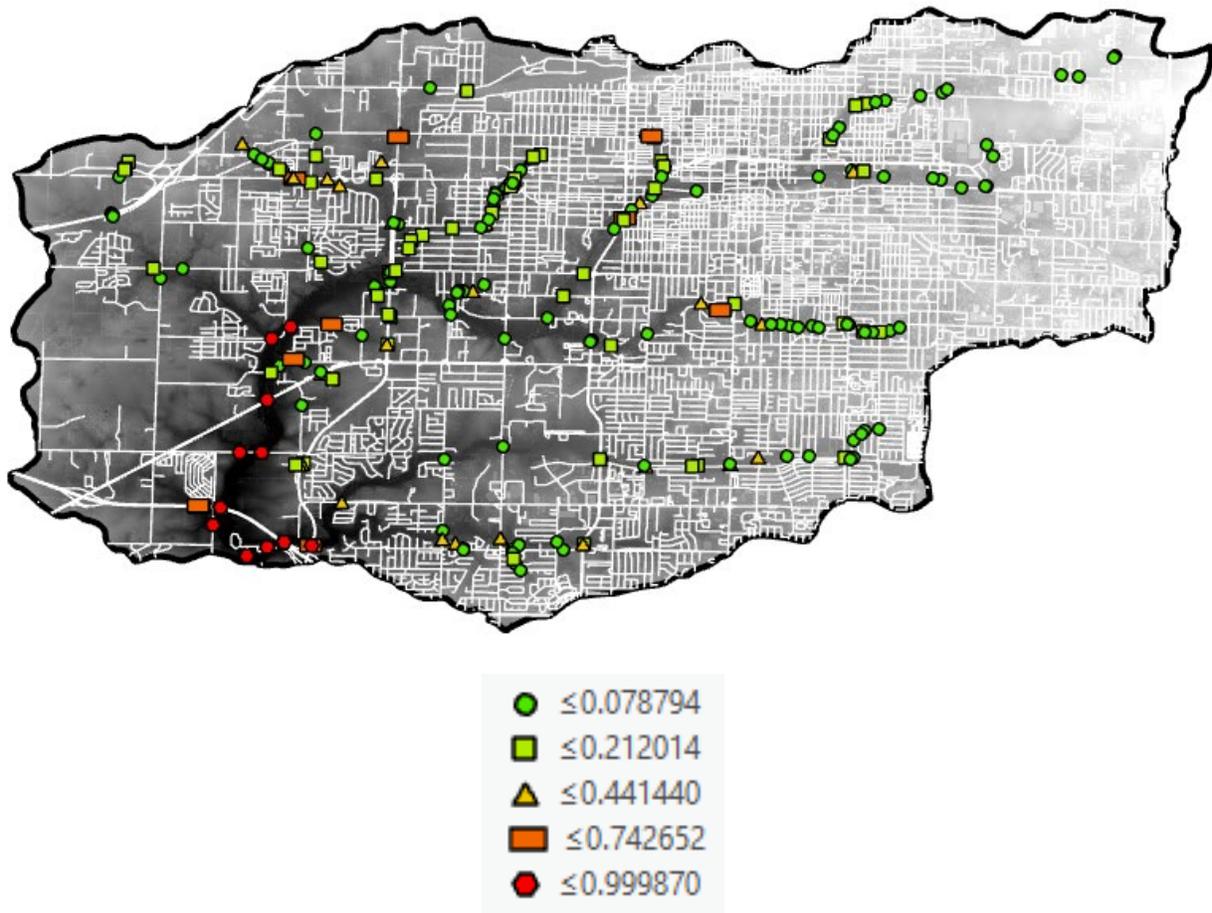
Exhibit 3-8. Classification Results
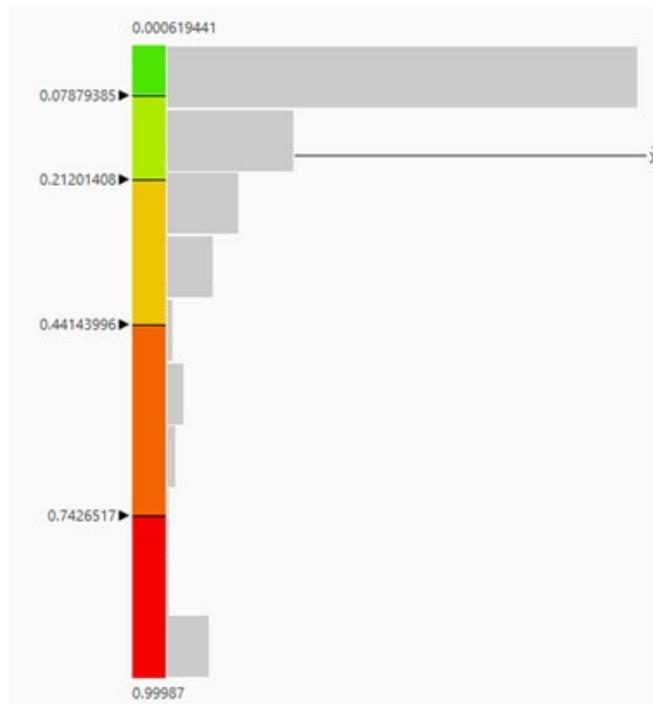
Exhibit 3-9. Histogram of Classification Scheme

Case Study Traffic Simulation Results

For this simulation part of the James River Freeway and West Sunshine Street were chosen to be removed in the simulation (highlighted in yellow in Exhibit 3-10). These roads were chosen because both points had a predictability percentage of 90% or higher.

The results of removing the two roads increased the average travel time and average wait time (Exhibit 3-11 and 3-12). The simulation where the networks were eliminated are in blue and the original network is depicted in green. The average travel time shows the average time it took a car to travel from their origin to their final destination. The average wait time shows the average amount of time a car had to wait at an intersection (due to congestion) during their travel to their destination.
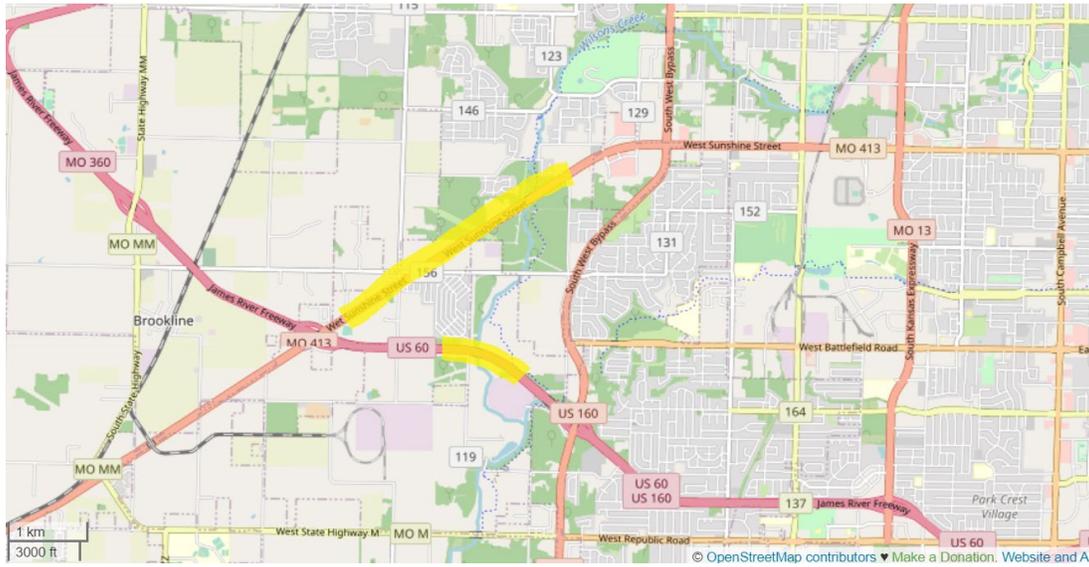
Exhibit 3-10. Roads Removed for Simulation Highlighted in Yellow
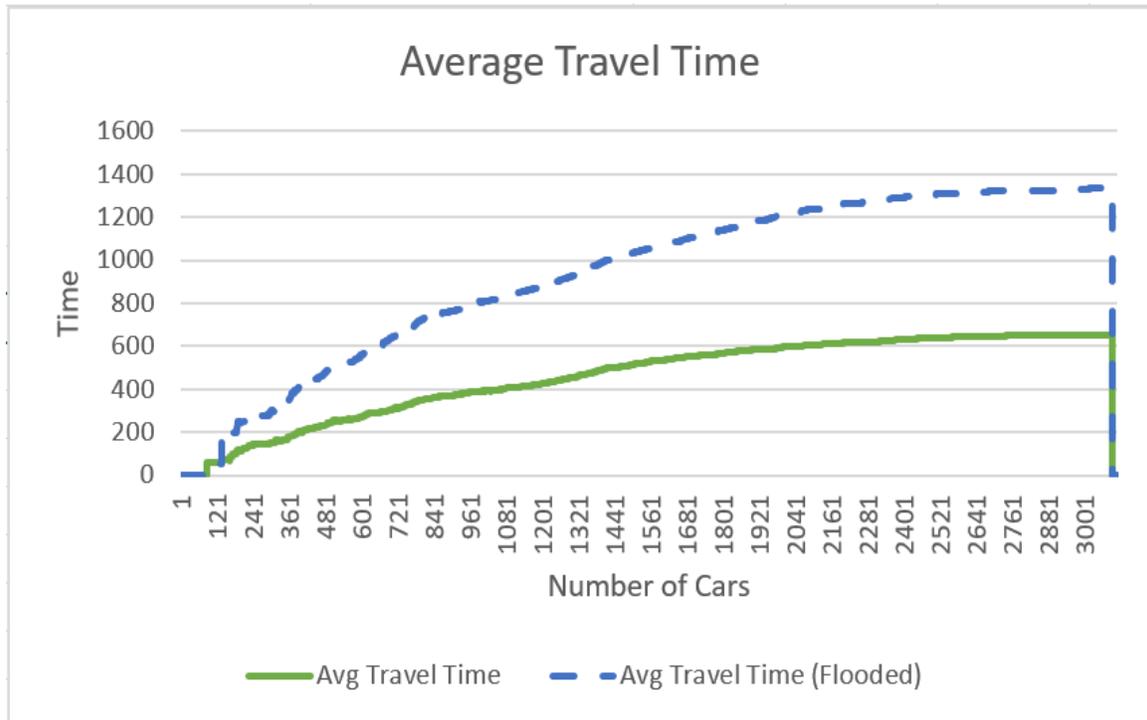
Exhibit 3-11. James River Freeway Simulation Result
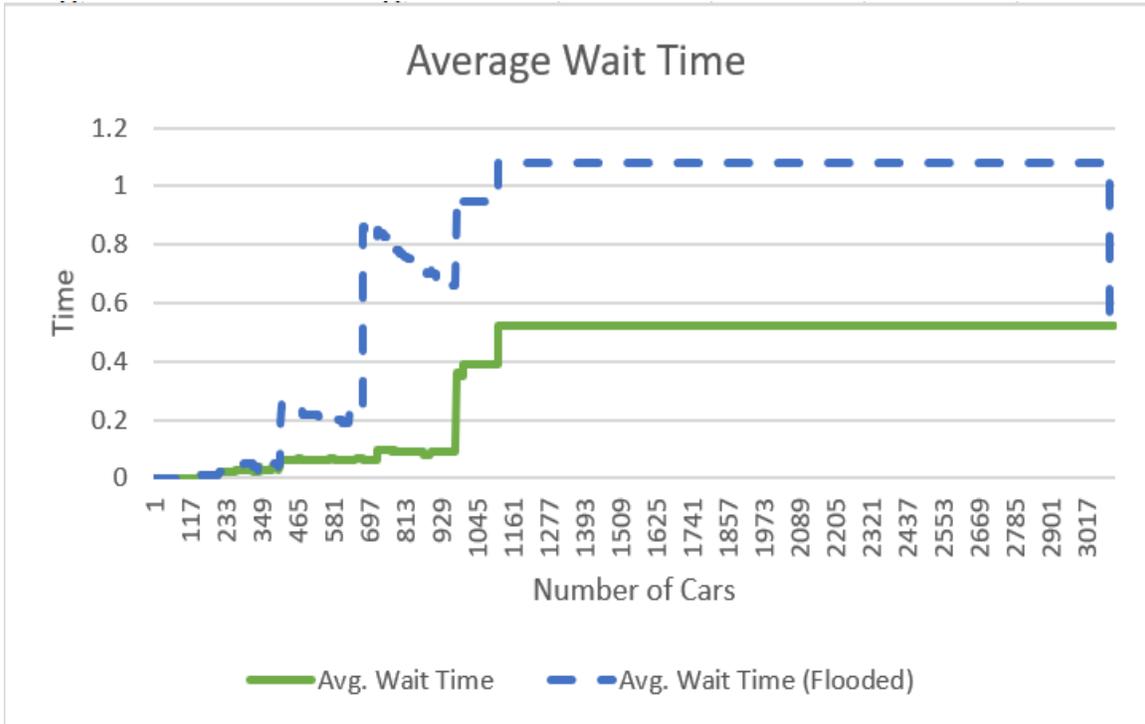
Exhibit 3-12. West Sunshine Street Simulation Result

4. Conclusions

A model framework based on high-quality publicly available datasets is developed and implemented to identify distinct locations in Greene County, Missouri which are prone to flash flooding. A combination of flood, hydrological, geospatial, and transportation datasets is relied upon to identify the roads which might be inaccessible due to flash flood events. The proposed framework provides updated critical information to the local authorities in charge of the emergency operations so that they can implement crucial public safety measures.

The model evaluates precipitation and geospatial information for various locations in the selected test region and utilizes an accurate deep learning algorithm to assign flash flood risk probabilities for potential flash flood-prone points. In comparison to other classification models, the deep learning algorithm provides a more detailed option to analyze the interactions between different variables and pinpoint vulnerable road segments. The inclusion of diverse geospatial features in the dataset further improves the neural network's capability to make accurate predictions. Additionally, the framework's ability to suggest alternative options to reroute traffic provides city planners with additional information to advise and assist commuters. Timely warnings and information can then be disseminated through reliable public information systems to warn commuters to avoid damaged or submerged road segments.

5. Limitations and Future Work

The results presented here demonstrate the utility of the methodology provided. However, there are some model limitations that constitute areas of future work. Development of the historic flash flood database was dependent on written narrative episodes provided in the Storm Events Database. Some of the information provided was inconsistent with actual road listing information. For example, on one occasion state routes and farm roads were used interchangeably. One solution to this problem is for witnesses or local authorities to provide specific coordinate information on flood affected road segments. This action would reduce the time required to build the database and the potential for clerical errors.

Machine learning and deep learning models require copious amounts of data. Out of 350 total data points, the flash flood locations were represented by 165 entries. These methods tend to require significantly greater quantities of data, and the accuracy of the model underscores this fact. As data gathering methods are improved and more data becomes available, the accuracy of the model presented will continue to improve. The result will be a more powerful tool at the disposal of local decision makers.

Providing local decision makers with a simulation of traffic behavior is an important part of the methodology presented. However, the simulation is constrained on the basis of how many intersections can be considered. Simulation results are then presented as a series of findings instead of one singular output. This sort of software limitation could be addressed by improving the capacity of the software chosen or selecting an alternative way to conduct the simulation. Addressing each of the areas would substantively improve the quality of the results presented.

References

Acosta-Coll, Melisa, Francisco Ballester-Merelo, Marcos Martinez-Peiró, and Emiro De la
Hoz-Franco. 2018. "Real-Time Early Warning System Design for Pluvial Flash
Floods—A Review." Sensors 18 (7): 2255. https://doi.org/10.3390/s18072255.

Benbennick, David. "Commons: United States County Locator Maps."
https://commons.wikimedia.org/wiki/Commons:United_States_county_locator_maps
(accessed October 4, 2020).

Bui, Dieu Tien, Phuong Thao Thi Ngo, Tien Dat Pham, Abolfazl Jaafari, Nguyen Quang
Minh, Pham Viet Hoa, and Pijush Samui. 2019. "A Novel Hybrid Approach Based on
a Swarm Intelligence Optimized Extreme Learning Machine for Flash Flood
Susceptibility Mapping." *Catena* 179 (March): 184–96.
https://doi.org/10.1016/j.catena.2019.04.009.

Castro, Joel T. de, Gabriel M. Salistre, Young-Cheol Byun, and B. Gerardo. 2013. "Flash
Flood Prediction Model Based on Multiple Regression Analysis for Decision Support
System." Undefined. 2013. https://www.semanticscholar.org/paper/Flash-Flood-
Prediction-Model-based-on-Multiple-for-Castro-
Salistre/8366f68c5c80a6332b9cbdd8f10d6cbcce240c57.

Census Bureau. "Population and Housing Unit Estimates Tables."
https://www.census.gov/programs-surveys/popest/data/tables.2019.html
(accessed October 5, 2020).

Chiang, Yen-Ming, Kuo-Lin Hsu, Fi-John Chang, Yang Hong, and Soroosh Sorooshian. 2007. "Merging Multiple Precipitation Sources for Flash Flood Forecasting." Journal of Hydrology 340 (3): 183–196. https://doi.org/10.1016/j.jhydrol.2007.04.007.

Corns, S., Long, S., Hale, J., Gude, H., Kanwar, B. Predictive Deep Learning for Flood Evacuation Planning and Routing, 2019.
Retrieved from https://spexternal.modot.mo.gov/sites/cm/CORDT/cmr20-002.pdf

Costache, Romulus, and Dieu Tien Bui. 2019. "Spatial Prediction of Flood Potential Using New Ensembles of Bivariate Statistics and Artificial Intelligence: A Case Study at the Putna River Catchment of Romania." *Science of the Total Environment* 691: 1098–1118. https://doi.org/10.1016/j.scitotenv.2019.07.197.

Fan, C., Jiang, X., & Mostafavi, A. (2020, September 02). Adaptive Reinforcement Learning Model for Simulation of Urban Mobility during Crises. Retrieved from https://arxiv.org/abs/2009.0135

Janizadeh, Saeid, Mohammadtaghi Avand, Abolfazl Jaafari, Tran Van Phong, Mahmoud Bayat, Ebrahim Ahmadisharaf, Indra Prakash, Binh Thai Pham, and Saro Lee. 2019. "Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran." *Sustainability (Switzerland)* 11 (19): 1–19. https://doi.org/10.3390/su11195426.

López, Andrea Salomé Viteri, and Carlos Augusto Morales Rodriguez. 2020. "Flash Flood Forecasting in Sao Paulo Using a Binary Logistic Regression Model." *Atmosphere* 11 (5). https://doi.org/10.3390/ATMOS11050473.

National Centers for Environmental Information, National Oceanic and Atmospheric Administration. "Storm Events Database." https://www.ncdc.noaa.gov/stormevents/ (accessed June 15, 2020).

National Weather Service. "NCEP Stage IV Daily Accumulations (06/28/2017 – Current)." https://water.weather.gov/precip/downloads/ (accessed August 15, 2020b).

National Weather Service. "Stage III Daily Accumulations (1/01/2005 – 06/27/2017)." https://water.weather.gov/precip/archive/ (accessed August 15, 2020a).

Ngo, Phuong Thao Thi, Nhat Duc Hoang, Biswajeet Pradhan, Quang Khanh Nguyen, Xuan Truong Tran, Quang Minh Nguyen, Viet Nghia Nguyen, Pijush Samui, and Dieu Tien Bui. 2018. "A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data." *Sensors (Switzerland)* 18 (11). https://doi.org/10.3390/s18113704.

Nguyen, Viet Nghia, Peyman Yariyan, Mahdis Amiri, An Dang Tran, Tien Dat Pham, Minh Phuong Do, Phuong Thao Thi Ngo, Viet Ha Nhu, Nguyen Quoc Long, and Dieu Tien Bui. 2020. "A New Modeling Approach for Spatial Prediction of Flash Flood with Biogeography Optimized CHAID Tree Ensemble and Remote Sensing Data." *Remote Sensing* 12 (9). https://doi.org/10.3390/RS12091373.

Saharia, Manabendra, Pierre-Emmanuel Kirstetter, Humberto Vergara, Jonathan J. Gourley, Yang Hong, and Marine Giroud. 2017. "Mapping Flash Flood Severity in the United States." Journal of Hydrometeorology 18 (2): 397–411. https://doi.org/10.1175/JHM-D-16-0082.1.

Tien Bui, Dieu, Nhat Duc Hoang, Francisco Martínez-Álvarez, Phuong Thao Thi Ngo, Pham Viet Hoa, Tien Dat Pham, Pijush Samui, and Romulus Costache. 2020. "A Novel Deep Learning Neural Network Approach for Predicting Flash Flood Susceptibility: A Case Study at a High Frequency Tropical Storm Area." *Science of the Total Environment* 701: 134413. https://doi.org/10.1016/j.scitotenv.2019.134413.

United States Geological Survey. "National Hydrography Dataset." https://www.usgs.gov/core-science-systems/ngp/national-hydrography/national-hydrography-dataset?qt-science_support_page_related_con=0#qt-science_support_page_related_con (accessed September 12, 2020c).

United States Geological Survey. "National Transportation Dataset." https://catalog.data.gov/dataset/usgs-national-transportation-dataset-ntd-downloadable-data-collectionde7d2 (accessed June 20, 2020a).

United States Geological Survey. "The National Map." https://viewer.nationalmap.gov/advanced-viewer/ (accessed August 1, 2020b).

Viteri López, Andrea Salomé, and Carlos Augusto Morales Rodriguez. 2020. "Flash Flood Forecasting in São Paulo Using a Binary Logistic Regression Model." Atmosphere 11 (5): 473. https://doi.org/10.3390/atmos11050473.

Wu, Jian, Haixing Liu, Guozhen Wei, Tianyu Song, Chi Zhang, and Huicheng Zhou. 2019. "Flash Flood Forecasting Using Support Vector Regression Model in a Small Mountainous Catchment." Water 11 (7): 1327. https://doi.org/10.3390/w11071.