Summer 2022

Development of flood prediction models using machine learning techniques

Bhanu Kanwar

Department: Engineering Management and Systems Engineering

DEVELOPMENT OF FLOOD PREDICTION MODELS USING MACHINE LEARNING TECHNIQUES

by

BHANU PARTAP SINGH KANWAR

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ENGINEERING MANAGEMENT

2022

Approved by:

Steven Corns, Advisor
Suzanna Long
Stephen Raper
Benjamin Kwasa
Venkata Nadendla

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following three articles, formatted in the style used by the Missouri University of Science and Technology:

Paper I, found on pages 2–24, has been published in the proceedings of American Society for Engineering Management 2021 International Annual Conference, in October 2021.

Paper II, found on pages 25–58, is intended for submission to *Journal of Flood Risk Management*.

Paper III, found on pages 59–84, is intended for submission to *Water* journal.

ABSTRACT

Flooding and flash flooding events damage infrastructure elements and pose a significant threat to the safety of the people residing in susceptible regions. There are some methods that government authorities rely on to assist in predicting these events in advance to provide warning, but such methodologies have not kept pace with modern machine learning. To leverage these algorithms, new models must be developed to efficiently capture the relationships among the variables that influence these events in a given region. These models can be used by emergency management personnel to develop more robust flood management plans for susceptible areas. The research investigates machine learning techniques to analyze the relationships between multiple variables influencing flood activities in Missouri. The first research contribution utilizes a deep learning algorithm to improve the accuracy and timelessness of flash flood predictions in Greene County, Missouri. In addition, a risk analysis study is conducted to advise the existing flash flood management strategies for the region. The second contribution presents a comparative analysis of different machine learning techniques to develop a classification model and predict the likelihood of flash flooding in Missouri. The third contribution introduces an ensemble of Long Short-Term Memory (LSTM) deep learning models used in conjunction with clustering to create virtual gauges and predict river water levels at unmonitored locations. The LSTM models predict river water levels 4 hours in advance. These outputs empower emergency management decision makers with an advanced warning to better implement flood management plans in regions of Missouri not served with river gauge monitoring.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere and intense sense of gratitude to my advisor Dr. Steven Corns for his judicious guidance, constant encouragement, inspiration, and facilitation during my Ph.D. studies. His experience and mentorship have been critical in motivating me to work diligently on my research project. I am thankful to him for providing timely feedback and inputs related to my work and for the preparation of this manuscript. I am also grateful to both Dr. Suzanna Long and Dr. Thomas Shoberg for providing me with key insights and concrete suggestions during the tenure of this work.

I would also like to express gratitude to my other committee members Dr. Kwasa, Dr. Raper and Dr. Nadendla for sharing research-related suggestions and feedback with me which has helped me in addressing key topics related to my research work. I am also grateful to the faculty and staff of the Department of Engineering Management and Systems Engineering for their help and encouragement during the course of my studies. I would like to thank the MidAmerica Transportation Center (MATC) and the Missouri Department of Transportation (MoDOT) for providing funds for the research projects over the last few years (grant numbers TR202023 and TR202111).

I express a deep sense of gratitude to my parents and brother for providing motivation, encouragement, everlasting affection, and moral support during the course of my Ph.D. studies. Huge thanks to my closest friends who have always believed in me and have rooted for me as I was pursuing my academic endeavors.

TABLE OF CONTENTS

Page

APPENDICES

LIST OF ILLUSTRATIONS

PAPER III

LIST OF TABLES

NOMENCLATURE

Symbol	Description
SVMs	Support Vector Machines
LSTM	Long Short-Term Memory
USGS	United States Geological Survey
NOAA	National Oceanic and Atmospheric Administration
NWS	National Weather Service
MoDOT	Missouri Department of Transportation
SEMA	State Emergency Management Agency
EFFS	European Flood Forecasting System
MLP	Multilayer Perceptron
ADT	Alternating Decision Tree
QDA	Quadratic Discriminant Analysis
PSO	Particle Swarm Optimization
HEC-FDA	Flood Damage Reduction Analysis
USACE	U.S. Army Corps of Engineers
SRMT	State Risk Management Team
FEMA	Federal Emergency Management Agency
NFHL	National Flood Hazard Layer
FRM	Flood Risk Maps
FRR	Flood Risk Reports
FRD	Flood Risk Database

ANN	Artificial Neural Network
SCIRC	Supply Chain Infrastructure Restoration Calculator
EIA	U.S. Energy Information Administration
NWSO	National Weather Service Office
LIDAR	Light Detection and Ranging
CSV	Comma-Separated Value
NDVI	Normalized Difference Vegetation Index
HEC-RAS	River Analysis System
SPOL	S-band Doppler Weather Radar
SMS	Short Message Service
DL-ANN	Deep Learning-Based Artificial Neural Networks
ROC	Receiver Operating Characteristic
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
PPV	Positive Prediction Value
TPR	True Positive Rate
FPR	False Positive Rate
AUROC	Area Under the ROC
NTD	National Transportation Dataset
API	Application Program Interface
CNN	Convolutional Neural Network

RNN	Recurrent Neural Network
1D	One-dimensional
2D	Two-dimensional
SAC-SMA	Sacramento Soil Moisture Accounting
HRU	Hydrological Response Unit
DEM	Digital Elevation Model
CC	Close-Close
CF	Close-Far
FC	Far-Close
FF	Far-Far

1. INTRODUCTION

This dissertation presents machine learning-based models that can be used to analyze the relationships between the geospatial and precipitation variables to provide predictive capabilities to stakeholders to enable better flash flood management plans. The three research contributions are:

Publication 1: A deep learning neural network using historical rainfall information and Geographic Information System (GIS) data to develop a classification model that predicts flash flood events in locations in Greene County, Missouri. The model outputs are applied to a resource calculator to determine the amount of money that would be required to restore the damaged road segments in the areas affected by flash floods.

Publication 2: A comparative analysis of logistic regression, support vector machines (SVMs), and deep learning neural network models to determine which model is most successful in classifying the occurrence of a flash flooding event in Greene County, Missouri. This research found that the deep learning algorithm was more accurate than the other models, making it a better choice to prepare advanced flash flood warnings and close roadways for public safety in flash flood-prone locations.

Publication 3: An ensemble of LSTM deep learning models is trained on the daily rainfall and river water level values for different clusters of gauges to predict water levels at unmonitored river locations associated with catchment areas of Missouri. A total of 30 different LSTM models are implemented using an ensemble learning approach to capture the intricate relationships between the time series-based input features and generate accurate multi-step predictions for the unmonitored locations.

PAPER

I. DEEP LEARNING-BASED DISASTER MANAGEMENT PLANNING AND RISK ANALYSIS OF FLASH FLOOD-PRONE REGIONS

Bhanu Kanwar[1] and Steven Corns[1]

[1]Department of Engineering Management and Systems Engineering,
Missouri University of Science and Technology, Rolla, MO 65409

ABSTRACT

An improved ability to predict flood events reduces risk to life and property. This research focuses on the use of deep learning algorithms to increase the accuracy and timeliness of flash flood predictions. Historical rainfall and Geographic Information System (GIS) data are used as inputs to a set of deep learning models. These models are then trained using historic flash flood event data to capture relationships between the weather and geographic data. Greene County, Missouri is used for this study as it encounters several weather events that have at times led to flash flood events. A risk analysis study is performed using this data to advance the current flash flood management strategies for the region. The data-driven approach is applied to publicly available data sourced from the United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), and National Weather Service (NWS).

Keywords: Flash Floods, Deep Learning, Neural Networks, Flood Risk, Missouri.

1. BACKGROUND INFORMATION

1.1. INTRODUCTION AND LITERATURE REVIEW

Flash flooding events have been responsible for significant losses over the last few decades. In addition to economic losses, these events have also been responsible for damage to infrastructure, traffic disruptions, and several fatalities in the affected areas (Ashley & Ashley, 2017). To efficiently deal with the damage caused by flash floods, it is vital to locate flood-prone locations and determine the likelihood and consequences of these events so that flood risk management protocols can be determined to ensure public safety. In addition, locating these locations will assist the city planners and local agencies such as the Missouri Department of Transportation (MoDOT) and Missouri State Emergency Management Agency (SEMA) in restoring the infrastructure elements damaged due to flash flooding, improving post-disaster relief operations for the well-being of the affected residents.

Research has been conducted over the last few years to identify the flood-prone locations and gauge the economic losses incurred by the local stakeholders due to flooding activity. Hydrological models based on LISFLOOD software are used to develop a European Flood Forecasting System (EFFS) and predict values of vital flood parameters such as water depth and inundation extent (Roo et al, 2010). To improve the existing flood warning protocols, hydrologic one-dimensional and two-dimensional flow models are proposed which can estimate both flood travel time and inundated areas in a risk-prone region (Ghimire et al, 2020). Various researchers are also relying on computational intelligence methods to locate flood-prone locations and improve current

flood warning and risk assessment protocols. A multilayer perceptron (MLP) is used to identify the time at which a swollen river might overflow its banks and cause flooding in the surrounding areas (S & S, 2020). Machine learning algorithms such as Alternating decision tree (ADT), functional tree (FT), kernel logistic regression (KLR), multilayer perceptron (MLP), and quadratic discriminant analysis (QDA) are implemented to efficiently map the locations prone to flash flooding events and update corresponding watershed management protocols (Janizadeh et al, 2019). Particle swarm optimization (PSO) based models are developed to predict flash flood-prone locations and update flood susceptibility maps in Northwest Vietnam (Bui et al, 2019). Other research teams have taken research endeavors to calculate the economic losses associated with both seasonal and untimely floods. Direct economic damages incurred due to flooding are assessed by a team of German researchers to find areas of improvement that could improve the critical processes of flood risk analysis and management (Merz et al, 2010). A separate framework is proposed by a Danish research team to efficiently conduct flood hazard and vulnerability assessment tasks and develop risk models for flood events impacting the Danish cities of Odense and Aarhus (Olsen et al, 2015). Decision tree-based flood loss models have been deployed by scientists to perform flood risk assessment studies of residential buildings and their contents in Vietnam (Chinh et al, 2017). A grid-based Geographical Information System (GIS) approach is proposed by another set of researchers to study risks posed by different waterlogging scenarios in an urban setting for distinct flood return periods (Yin et al, 2011). Information obtained from employing this GIS approach is used to calculate the depth and extent of waterlogging which might pose a threat to inundated buildings and houses. US Army

Corps of Engineers Flood Damage Reduction Analysis (HEC-FDA) software tool also provides the ability to perform economic analysis exercises and examine risk management plans related to a flood activity in a sensitive area (HEC, 2021). Frequent flood risk studies are also conducted by the U.S. Army Corps of Engineers (USACE) Silver Jacket Program in conjunction with the State Risk Management Team (SRMT) in the areas susceptible to flash floods in the state of Missouri (Silvers Jackets Website, 2021). These government agencies frequently collaborate with the State Emergency Management Agency (SEMA) to check risks posed by all flood events and develop relevant mitigation plans (SEMA, 2021; Silvers Jackets Website, 2021). Regular flood management exercises are also undertaken by these bodies to safeguard the lives and property of the local public from unexpected flooding scenarios. Silver Jackets Program also provides technical assistance for different Federal Emergency Management Agency (FEMA) Risk MAP products to effectively communicate risks emanating from all flooding events to the local public. This plethora of resourceful information can easily be obtained from relevant agencies to be used for the benefit of the local stakeholders. Risk MAP Products such as flood insurance rate maps, National Flood Hazard Layer (NFHL) databases, flood risk maps (FRM), flood risk reports (FRR), flood risk database (FRD), etc. are regularly hosted by FEMA on its web portal for use by the interested parties (FEMA.gov., 2021). The Missouri Department of Transportation (MoDOT) regularly highlights the state's flooded road segments and locations on the Traveler Information Map (Missouri Department of Transportation, 2021). Department of Transportation also shares such critical information with the people through social media applications on regular basis (Twitter, 2021).

This project focuses on the implementation of a deep learning-based artificial neural network (ANN) to locate potential flash-flood-prone locations and use a resource calculator to estimate the amount of money needed to restore flash flood-hit damaged road segments. Research suggestions can then be used to enhance flash flood risk management protocols for sensitive areas. The data required for this project is obtained from the United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), and National Weather Service (NWS).

1.2. CLASSIFICATION MODELS

The three different types of classification models implemented in this research are a deep learning-based artificial neural network model, a logistic regression model, and a support vector machine model. Deep learning-based artificial neural networks are built of layers containing mathematical functions called neurons which deduce the trends and patterns hidden in the input dataset (smartboost, 2020). It contains multiple hidden layers which receive input data from the initial layer and apply computations to it before further passing the information to the network's output layer. This output layer then processes the received information and produces a Boolean value for the classified data label. A binomial logistic regression-based model uses a logit-function to evaluate the relationship between all variables and generates probability-based output values to differentiate between flooded and non-flooded locations (scikit, 2021). Support Vector Machine (SVM) classifier allocates labeled data points on either side of a decision boundary known as hyperplane to classify data points belonging to either class label (Bambrick, 2016). All three classification algorithms assign a probability value to a location data

point belonging to either category (i.e., flooded, or non-flooded location). The algorithms assign class label 0 to all location data points with a probability value lower than 0.5 whereas the data points with a probability value of more than 0.5 are labeled with a value of 1.

1.3. SUPPLY CHAIN INFRASTRUCTURE RESTORATION CALCULATOR (SCIRC)

The Supply Chain Infrastructure Restoration Calculator (SCIRC) software tool developed by a team of researchers at Missouri S&T, Rolla, and United States Geological Survey (USGS) can be used to estimate both direct costs and resources needed to restore damaged infrastructure elements (Ojha et al, 2019). This software tool provides an approximate value of the monetary resources based on the historic reconstruction costs needed to repair damaged infrastructure elements like interstates, arterial roads, etc. The SCIRC can be used to calculate the total amount of resources required by the city planners and contractors to reconstruct an infrastructure element damaged due to a natural disaster such as a flash flood. The different types of resources needed for reconstruction purposes are power, fuel, potable water, storage area, man-hours, gray water, food, materials, etc. as shown in Figure 1. The SCIRC is populated using a bottom-up cost estimation technique and its output can be combined with event probability values to perform the approximate comparisons of risk values of different infrastructure elements. This information can then be used by the personnel in charge of post-disaster relief operations to develop and implement necessary flash flood management protocols if needed.

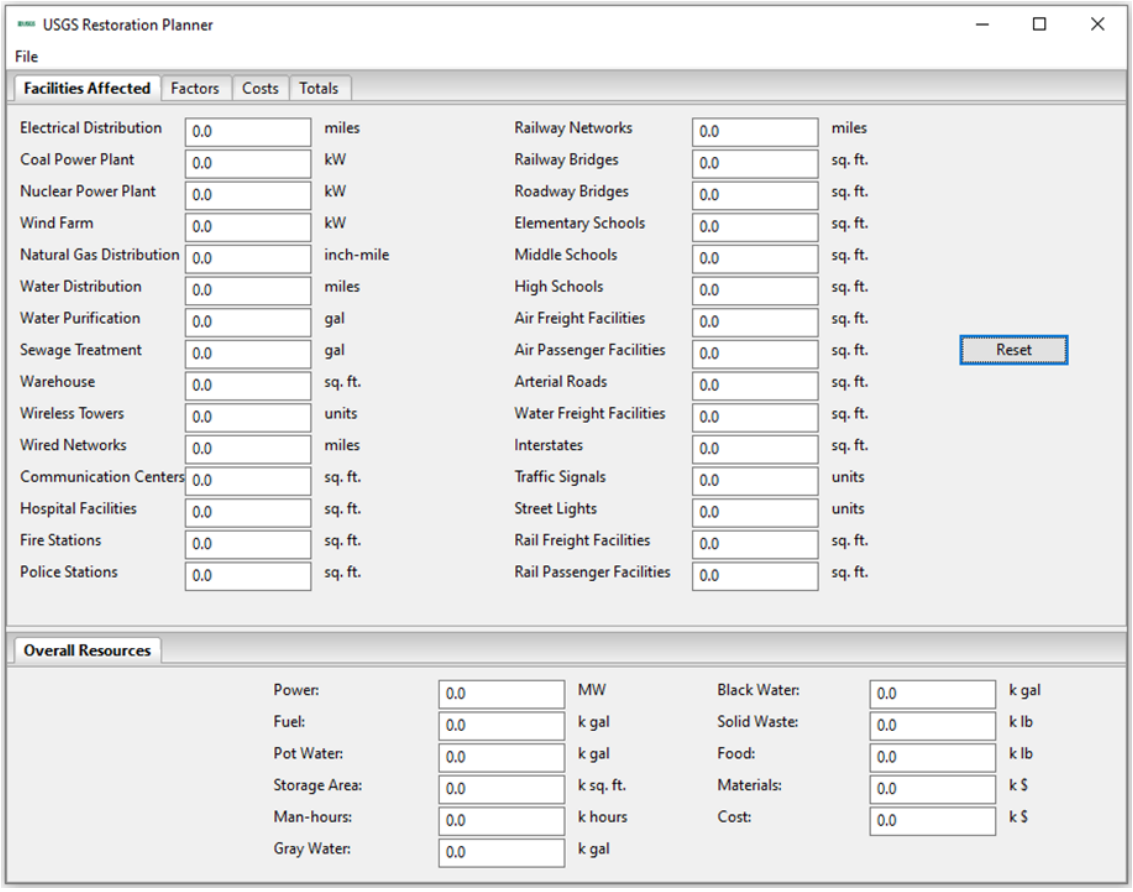

Figure 1. User interface of supply chain infrastructure restoration calculator (SCIRC).

The resource calculator constitutes five different components which are mentioned below.

1. Facilities Affected: The tab for the ‘Facilities Affected’ component consists of a list of 30 different infrastructure elements which can be selected by the user on a case-by-case basis for restoration-related calculations.
2. Factors: This component tab constitutes information related to the amount of each resource (power, water, fuel, man-hours, etc.) required to restore per unit of a chosen infrastructure element. The user can modify and update the values in this tab as per its requirements.

3. Costs: The per-unit costs information of each resource is pre-filled in the designated fields of this component tab.
4. Totals: The amount of each resource needed for the restoration of an infrastructure element is categorized under this component tab. The sum amount of all costs of resources needed for restoration is also displayed under this feature tab.
5. Overall Resources: This component tab includes data related to all the resources needed to restore all the infrastructure elements selected by the user.

The software tool can be used to calculate the amount of resources and costs involved while restoring a set of thirty different infrastructure elements selected from a variety of sectors such as transportation, education, healthcare, emergency services, city utilities, communications, traffic control, etc. Each infrastructure element requires different amounts of resources for its post-disaster restoration operations. The construction steps and processes for different infrastructure elements are reviewed to develop estimates for the resources required for their restoration. The data values for resources such as potable water, gray water, food, power, etc. are established from the number of man-hours needed for restoring respective infrastructure elements. RSMeans, a square foot estimator tool, is also used to estimate the cost of materials plus man-hours required to re-construct hospitals, schools, fire stations, police stations, and warehouses in disaster-struck areas (RSMeans, 2021). The cost of materials required to restore the remaining infrastructure elements is obtained after evaluating their respective reconstruction procedures. The cost of resources such as fuel, electricity, etc. is gathered from reviewing the similar literature work published by several government agencies and

university researchers (Environmental Protection Agency, 2021; Jiang, 2011; Michigan Water Environment Association, 2021; Ohio Environmental Protection Agency, 2021; Slate, 2021; U.S. Energy Information Administration (EIA), 2021). This cost information is only applicable to the restoration tasks for the infrastructure elements located in the Midwestern United States. The costs related information varies from region to region and will have to be updated by the user while using this tool in the other parts of the country.

A set of different linear equations are developed to calculate the total amount of resources and costs involved in reconstructing a damaged infrastructure element (Ojha et al, 2020). The total amount of resources needed to restore certain units of damaged infrastructure element is calculated by the product of 'x' units of the element with the amount of resources needed to restore per unit of the infrastructure element as shown in Equation (1) below.

$$T_{ij} = x * R_{ij} \quad (1)$$

where 'T_{ij}' is the total amount of a resource 'j' needed to restore 'x' units of a selected infrastructure element 'i' and 'R_{ij}' represents the amount of a resource 'j' needed to restore a unit of infrastructure element 'i'.

The total cost of restoring 'x' units of a selected infrastructure element 'i' is denoted by 'TCi' and is calculated from Equation (2) shown below.

$$TC_i = \sum_{j=1}^{10} (T_{ij} * C_j) \quad (2)$$

where 'C_j' represents the cost of a unit of resource 'j'.

Equation (3) is also developed to calculate the overall amount of resources required to restore multiple damaged infrastructure elements as shown below.

$$OR_j = \sum_{i=1}^{30} T_{ij} \;\; \forall j = 1, 2, 3, \ldots, 10 \quad (3)$$

where 'OR_j' is the total amount of resource 'j' required to restore multiple damaged infrastructure elements.

The calculator is easy to use, and the user must enter the units of the selected infrastructure element(s) mentioned in the 'Facilities Affected' component tab to find the number of resources and the total amount of money needed for its restoration.

2. METHODOLOGY

2.1. STUDY AREA

A goal of this research project was to identify potential flash flood-prone locations in and around the city of Springfield located in Greene County, Missouri. The geospatial and weather data for this area is studied as this region of Missouri often encounters various inclement weather events over a whole year (Clayton News, 2021). According to the data collected by the local Springfield National Weather Service Office (NWSO), frequent episodes of heavy rainfall accompanied with unusual tornado activities often results in flash flood episodes in the surrounding area (US Department of Commerce, N.O.A.A., 2015). The geospatial and weather data for this area is collected and investigated to develop a deep learning-based artificial neural network (ANN) classifier that can assist in accurately locating the flash flood-prone locations. Once such locations have been studied, the regional authorities in charge of flood management responses can further use this information to perform suitable risk values-related studies and improve its disaster handling protocols.

2.2. DATA

To identify flash flood-prone locations in the Springfield area of Greene County, Missouri a deep learning-based artificial neural network (ANN) classifier is developed using the publicly available datasets. To begin with, National Oceanic and Atmospheric Administration's (NOAA) Storms Events Database is explored to prepare a custom flash flood database containing information related to flash flood events at susceptible locations. The information acquired from this database is utilized to highlight the areas of interest in the selected watershed for subsequent analysis. Light Detection and Ranging (LIDAR) based geospatial data files gathered from the United States Geological Survey (USGS) are processed using ArcGIS Pro, a geospatial analysis software, to extract raster elevation, slope, aspect, and curvature layers for the test area. Precipitation data collected from the National Weather Service (NWS) is also processed using ArcGIS Pro to collect rainfall information from January 2005 till December 2019. Eventually, the collective location data information embedded in these geospatial and rainfall layers is stored in a tabular comma-separated value (CSV) data format for further application in suitable machine learning operations.

Different machine learning models such as the logistic regression model and support vector machine (SVM) model, and another deep learning-based artificial neural network (ANN) model are implemented to select a suitable classification model for accurately classifying the flash flood-prone locations in the study area. The complete labeled location data set required for these classification models consists of 350 different data points. The target variable embedded in this dataset is assigned two distinct class label values of 0 and 1 to distinguish between both non-flash flood and flash flood

locations, respectively. So, location data points assigned with class label values of 0 and 1 signify non-flash flood and flash flood-prone locations, respectively. Out of these 350 data points, 185 points fall under the category of non-flash flood-prone locations (i.e., with class label 0) whereas the remaining 165 points signify flash flood-prone locations (i.e., with class label 1) as shown in Figure 2.

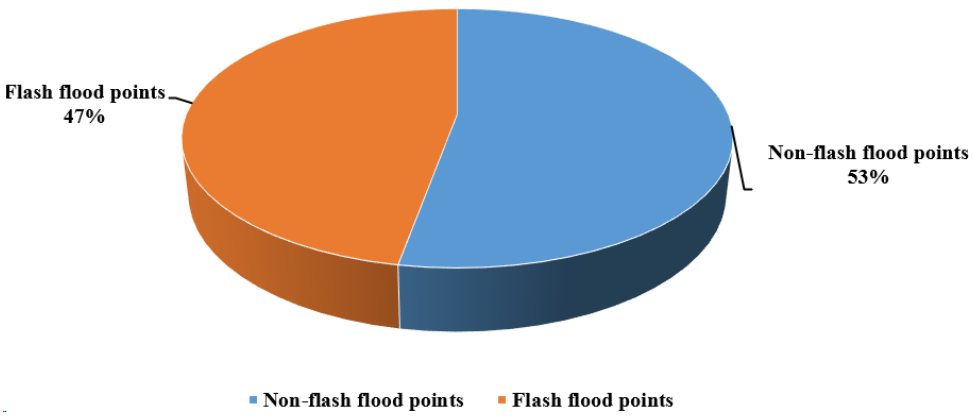

Figure 2. Distribution of location data points.

The dataset constitutes values for nine different independent variables or features such as ‘Rainfall day of Inches’, ‘Rainfall 1 day before’, ‘Rainfall 2 days before’, ‘Rainfall 3 days before’, ‘Summation of antecedent rainfall’, ‘Curvature’, ‘Aspect’, ‘Slope’ and ‘Elevation’. The dependent variable is titled ‘Class label’ and comprises binary values, 0 and 1 to signify both non-flash flood and flash flood locations. Figure 3 displays the architecture of the models implemented to classify the flash flood-prone locations. The whole dataset is divided into two sets for the tasks of training and testing the models. 262 data points (75% of the whole dataset) are allotted to the training dataset

and the remaining 88 points (25% of the whole dataset) are allocated to the testing dataset.

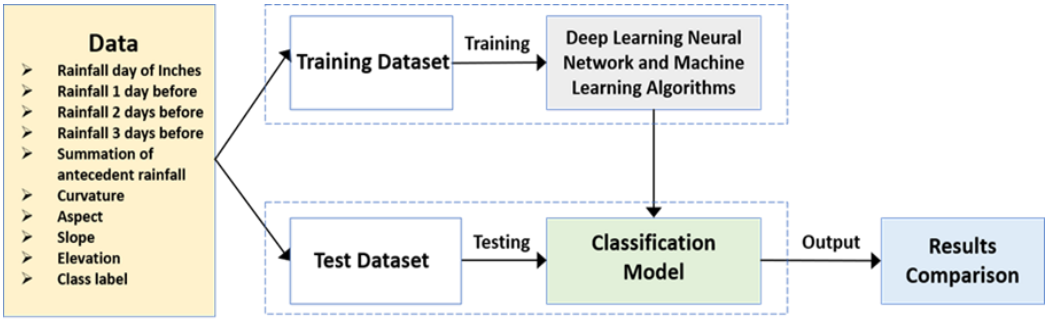

Figure 3. Location classification model architecture.

2.3. DEEP LEARNING-BASED CLASSIFICATION MODEL

The deep learning neural network uses 9 different variables as input parameters, an ‘Adam’ learning rate optimizer, and a ‘sigmoid’ activation function for the classification task. A dropout value of 0.2 is applied within the layers of the neural network to prevent overfitting (IBM, 2021). Both hyperparameters, ‘epochs’ and ‘batch size’ are assigned the values of 1000 and 10 respectively. The output ‘Dense’ layer of the network utilizes a sigmoid function to assign a value ranging from 0 to 1 to the single location point.

2.4. SUPPLY CHAIN INFRASTRUCTURE RESTORATION CALCULATOR (SCIRC) AND RISK VALUE

The classifier’s output indicates flash flooding probability values in areas of Greene County, MO. A risk value associated with damaged infrastructure can then be calculated by multiplying the respective probability values with the infrastructure’s

restoration cost. This restoration cost corresponds to the amount (in $) needed to re-construct the infrastructure elements damaged due to a flash flood. The calculator's functionality along with the event probability value of a susceptible location can be used to get higher-level estimates of the risk value associated with different infrastructure elements threatened by flash floods.

The risk value i.e., the estimated cost associated with this road section in case of a flash flood event can be calculated using Equation (4) shown below.

$$Risk\ Value = Probability\ Value * Impact\ Value \quad (4)$$

where 'Probability Value' is the likelihood of occurrence of an event and 'Impact Value' is the amount of money required in a risk event.

3. RESULTS AND DISCUSSION

3.1. DEEP LEARNING-BASED CLASSIFICATION MODEL

Out of all three classification models, the deep learning-based artificial neural network classifier achieves the highest classification accuracy of 85.23% when applied on the 88 data points-based testing feature dataset. The deep learning model described correctly classifies 75 out of 88 locations with an error rate of 14.77%. 40 locations out of a total of 75 locations are correctly classified as belonging to class label 0 and the remaining 35 locations are correctly labeled with a class value of 1. Both logistic regression and support vector machine (SVM) based classifiers deliver an accuracy score of 75% and 62.5% respectively. The parameters of the support vector machine (SVM) are

further tuned to improve its accuracy score to 80.68% which is still lower than the neural network's score as shown in Figure 4.

The neural network predicts areas with a 50% or greater probability of experiencing flash flooding using a 0.5 threshold function. By varying the threshold value from 0.5 to 0.9 (in increments of 0.1) a better evaluation of the probability of flash flooding can be found, corresponding to a 50% to 90% probability. This technique was applied to the locations previously identified in Greene County, Missouri with random rainfall amounts. This created a novel dataset for 220 values to be used as inputs to the classifier. When used to bin values in this dataset, the model identified 78 events with greater than a 50% probability of flash flooding. The total for each probability were 22 meeting only the 50% threshold, 17 meeting the 60% threshold, 14 meeting the 70% threshold, 13 meeting the 80% threshold, and 12 with 90% or greater probability.

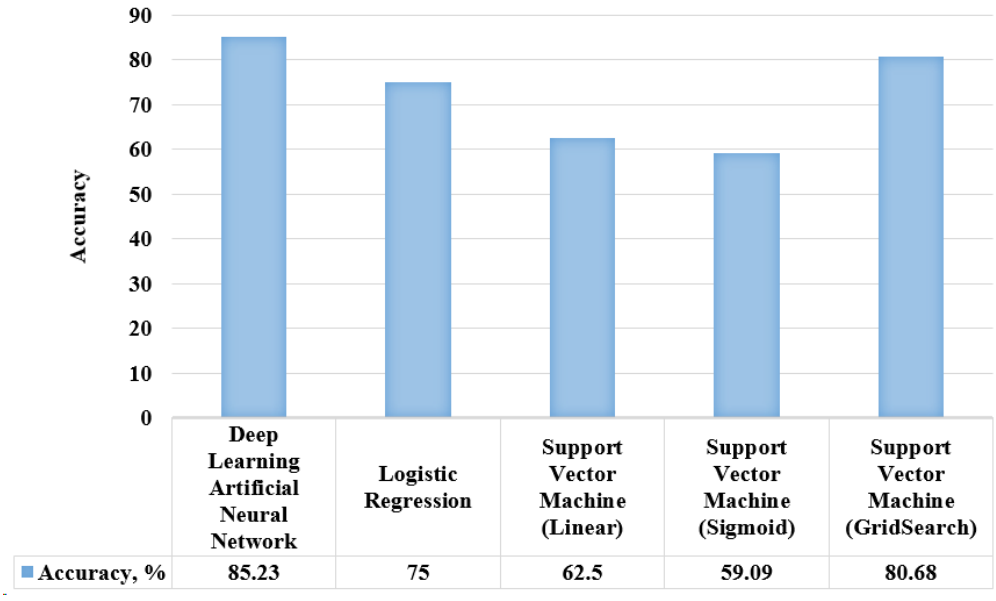

Figure 4. Comparison of accuracy scores.

3.2. RISK ANALYSIS

The deep learning-based neural network model identifies the section of West Farm Road 146 running parallel to the Wilsons Creek as one of the 12 flash flood-prone locations in the Springfield area which have a more than 90% probability value of encountering any flash flood event. The model assigns a very high flash flood probability value of 99.87% to this location which makes it a highly vulnerable spot for flash floods. Wilsons Creek is located in southwest Greene County and it frequently overflows its banks after heavy rainfall resulting in the inundation of the West Farm Road 146 (KY3, 2019). As a result, this arterial road is often closed for use by daily commuters to ensure their safety and wellbeing. In case this road is damaged due to flash floods, the local government authorities in charge of post-disaster restoration tasks will have to calculate both amounts of resources and money needed for its repair and restoration.

Using the SCIRC, the total cost to repair this road can be calculated and is equal to $8.05 per sq. ft. The cost incurred to repair per sq. ft. of this damaged road can also be called ‘Impact’ which is the amount of money spent whenever an identifiable risk event occurs (Spacey, 2017). Using Equation (4), the risk value for the West Farm Road 146 is calculated to be $8.04 per sq. ft. The deep learning-based model also identifies another section of West Sunset Street located parallel to Sunset South Creek as the location with a very high probability value of experiencing flash flood events. The model assigns a very high flash flood probability value of 99.30% to this new location. After a flash flood event, the cost to repair the damaged section of this road can also be acquired from SCIRC which is equal to $14.43 per sq. ft. Similarly, Equation (4) can be used again to calculate its risk value which is equal to $14.33 per sq. ft.

The risk value numbers can be used by the responsible personnel to prepare estimates for the budget needed to restore the damaged road sections in case of a flash flood event. The authorities can compare the risk values for different roads to prioritize spending money on the repair work as per their budget and other requirements. It can be observed that the risk value for the West Sunset Street i.e., $14.33 per sq. ft. is more than the similar value for West Farm Road 146 i.e., $8.04 per sq. ft. Since, West Sunset Street experiences more traffic volume than West Farm Road 146, the restoration crew can prioritize its re-construction over the West Farm Road 146 for the benefit of the commuters (Traffic Volume Maps, 2021). Also, a large number of heavy vehicles such as single-unit trucks and tractor-trailer combination trucks, etc. ply on West Sunset Street as compared to the West Farm Road 146 which makes its post-flood repair and restoration a priority for the responsible parties (Traffic Volume Maps, 2021). Any delays to its restoration tasks can increase the transportation costs of the goods and items transported by these heavy vehicles from one place to another. Moreover, the presence of many residential buildings constructed along the banks of Sunset South Creek on West Sunset Street makes the task of reconstructing its damaged sections more urgent to ensure the well-being of the area's residents (Google, 2021).

A similar comparison of risk values and traffic volume data for a different set of roads can assist in identifying locations in need of urgent attention after a natural hazard such as a flash flood. Prior knowledge of these estimates will help local government agencies to calculate the suitable amount of resources needed to counter the risks posed by the flash floods to the vulnerable sections of important roads. The personnel at the Missouri Department of Transportation (MoDOT) can also use this information to

develop more robust risk management plans for the safety of the commuters relying on these roads for their travels. Once such plans have been developed, adequate responses can also be taken by the Missouri State Emergency Management Agency (SEMA) to mitigate the risks posed by roads damaged because of a flash flood event. The capabilities of both deep learning-based neural network classification model and SCIRC can be used to identify the flash flood risk-prone locations and prepare adequate contingency reserves needed to repair the damaged road sections in these locations.

4. CONCLUSION AND FUTURE WORK

Different machine learning algorithms can be executed to scrutinize vast amounts of weather data generated continuously and locate areas that can witness abrupt flash flood activities. The proposed deep learning-based neural network classifier with an accuracy score of 85.23% can be adopted to identify such locations in a selected watershed basin. This classifier relies on high-quality publicly available data to efficiently diagnose and group data points into two distinct categories i.e., non-flash flood-prone and flash flood-prone locations. The architecture of the deep learning-based neural network classifier can be updated to conveniently incorporate different sets of geospatial and weather features as inputs to the model. Also, the performance of a deep learning-based classifier is not affected when implemented on a huge volume of data points which makes it suitable for application in a variety of real-world scenarios. The classifier consumes less time to analyze the data features as compared to the traditional

machine learning models which is advantageous in the time-sensitive task of identifying dangerous flash flood-prone locations.

After identifying flash flood hotspots, the probability values assigned to a susceptible road can be combined with its total restoration costs obtained from SCIRC to find an overall risk value for that location. After analyzing the risk values for different roads threatened by flash floods, the city planners can set aside sufficient resources needed for the post-disaster restoration tasks. The policymakers can also use these values to estimate the monetary impact of flash floods-induced risks on the post-disaster relief management plans. The availability of advanced risk values-related information will also assist these authorities in making apt decisions related to post-disaster resource allocation tasks.

The deep learning-based classification model can be implemented on datasets containing information related to other features such as normalized difference vegetation index (NDVI), etc. to identify flash flood-prone locations in different areas. Additional information such as building data, population, etc. can also be incorporated in the flash flood-related economic and risk studies in future work.

REFERENCES

Ashley, S. T., & Ashley, W. S. (2008, March 1). Flood Fatalities in the United States. AMETSOC. https://journals.ametsoc.org/view/journals/apme/47/3/2007jamc1611.1.xml.

Basic Waste Water Treatment Costs. Michigan Water Environment Association. (Last Accessed July 29, 2021). https://www.mi-wea.org/docs/The%20cost%20of%20Biosolids.pdf.

Bambrick, A. N. (2016, July). Support Vector Machines: A Simple Explanation. KDnuggets. https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html.

Bui, D. T., Ngo, P.-T. T., Pham, T. D., Jaafari, A., Minh, N. Q., Hoa, P. V., & Samui, P. (2019). A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *CATENA*, *179*, 184–196. https://doi.org/10.1016/j.catena.2019.04.009

Chinh, D., Dung, N., Gain, A., & Kreibich, H. (2017). Flood Loss Models and Risk Analysis for Private Households in Can Tho City, Vietnam. *Water*, *9*(5), 313. https://doi.org/10.3390/w9050313

County with the most severe weather in every state. Clayton News. (2021, March 31). https://www.news-daily.com/multimedia/photos/county-with-the-most-severe-weather-in-every-state/collection_da46773c-de17-592c-889f-d2682b9da532.html#26.

Current Flood Information. Missouri Department of Transportation. (Last Accessed June 4, 2021). https://www.modot.org/current-flood-information.

Deep Learning vs Neural Network: What's the Difference? smartboost. (2020, November 19). https://smartboost.com/blog/deep-learning-vs-neural-network/.

Diesel fuel Pump Components history - U.S. Energy Information Administration (EIA). U.S. Energy Information Administration (EIA). (Last Accessed July 29, 2021). https://www.eia.gov/petroleum/gasdiesel/dieselpump_hist.php.

Environmental Protection Agency. (Last Accessed July 29, 2021). Understanding Your Water Bill. EPA. https://www.epa.gov/watersense/understanding-your-water-bill.

FDA. HEC. (Last Accessed June 4, 2021). https://www.hec.usace.army.mil/software/hec-fda/.

FIA. HEC. (Last Accessed June 4, 2021). https://www.hec.usace.army.mil/software/hec-fia/.

Flooding. SEMA. (Last Accessed July 28, 2021). *https://sema.dps.mo.gov/plan_and_prepare/flooding.php.*

Floodplain Management in Missouri Quick Guide. (Last Accessed June 4, 2021). *https://sema.dps.mo.gov/programs/floodplain/documents/quick-guide.pdf.*

Ghimire, E., Sharma, S., & Lamichhane, N. (2020). Evaluation of one-dimensional and two-dimensional HEC-RAS models to predict flood travel time and inundation area for flood warning system. *ISH Journal of Hydraulic Engineering*, 1–17. https://doi.org/10.1080/09715010.2020.1824621

Google. (Last Accessed June 4, 2021). Google Maps. https://www.google.com/maps/place/37%C2%B009'59.0%22N+93%C2%B017'50.7%22W/@37.1651464,-93.2990385,1013m/data=!3m1!1e3!4m5!3m4!1s0x0:0x0!8m2!3d37.1663864!4d-93.2974081.

IBM. (Last Accessed November 1, 2021). *What is Overfitting?* Overfitting. Retrieved November 1, 2021, from https://www.ibm.com/cloud/learn/overfitting.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, B. T., & Lee, S. (2019). Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. *Sustainability*, *11*(19), 5426. https://doi.org/10.3390/su11195426

Jiang, J. (2011, October 28). The price of electricity in your state. NPR. https://www.npr.org/sections/money/2011/10/ 27/141766341/the-price-of-electricity-in-your-state.

Merz, B., Hall, J., Disse, M., & Schumann, A. (2010). Fluvial flood risk management in a changing world. Natural Hazards and Earth System Sciences, 10(3), 509–527. https://doi.org/10.5194/nhess-10-509-2010

Ojha, A., Kanwar, B., Long, S. K., Shoberg, T. G., & Corns, S. (2019, June 24). Supply Chain Infrastructure Restoration Calculator Software Tool-Developer Guide and User Manual. Open-File Report. https://pubs.er.usgs.gov/publication/ofr20191061.

Ojha, A., Long, S., Shoberg, T., & Corns, S. (2020). Bottom-up resource and cost estimation for restoration of supply chain interdependent critical infrastructure. Engineering Management Journal, 1–11. https://doi.org/10.1080/10429247.2020.1800387

Olsen, A., Zhou, Q., Linde, J., & Arnbjerg-Nielsen, K. (2015). Comparing Methods of Calculating Expected Annual Damage in Urban Pluvial Flood Risk Assessments. Water, 7(12), 255–270. https://doi.org/10.3390/w7010255

Risk MAP Products. FEMA.gov. (Last Accessed June 4, 2021). https://www.fema.gov/flood-maps/tools-resources/risk-map/products.

Roo, A. P. J. de, Gouweleeuw, B., Thielen, J., Bartholmes, J., Baongioannini-Cerlini, P., Todini, E., Bates, P. D., Horritt, M., Hunter, N., Beven, K., Pappenberger, F., Heise, E., Rivin, G., Hils, M., Hollingsworth, A., Holst, B., Kwadijk, J., Reggiani, P., Dijk, M. V., … Sprokkereef, E. (2010). Development of a European flood forecasting system. *International Journal of River Basin Management*, 49–59. https://doi.org/https://doi.org/10.1080/15715124.2003.9635192

RSMeans Data. RSMeans. (Last Accessed July 29, 2021). https://www.rsmeans.com/estimating-square-foot-cost.

S, V. B., & S, S. (2020). Flood Prediction System using Multilayer Perceptron Classifier and Neural Networks. https://www.irjet.net/archives/V7/i5/IRJET-V7I51189.pdf.

sklearn.linear_model.LogisticRegression¶. scikit. (Last Accessed June 4, 2021). https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Slate. (2013, July 12). You are paying 300 times more for bottled water than tap water. Slate Magazine. http://www.slate.com/blogs/business_insider/2013/07/12/cost_of_bottled_water_vs_tap_water_the_difference_will_shock_you.html.

Solid waste disposal fees. Ohio Environmental Protection Agency. (Last Accessed July 29, 2021). https://ohioepa.custhelp.com/app/answers/detail/a_id/2270/~/solid-waste-disposal-fees.

Spacey, J. (2017, April 12). What is a Risk Value? Simplicable. https://simplicable.com/new/risk-value.

State Teams. Silver Jackets Website. (Last Accessed June 4, 2021). https://silverjackets.nfrmp.us/State-Teams/Missouri.

Traffic Volume Maps: Missouri Department of Transportation. Traffic Volume Maps. (Last Accessed June 4, 2021). https://www.modot.org/traffic-volume-maps.

Twitter. (Last Accessed June 4, 2021). *Twitter*. MoDOT (@MoDOT) / Twitter. https://twitter.com/MoDOT?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor.

US Department of Commerce, N.O.A.A. (2015, March 27). *SEVERE WEATHER CLIMATOLOGY FOR THE SPRINGFIELD, MO. NWSO COUNTY WARNING AREA*. National Weather Service. https://www.weather.gov/sgf/severe_climatology_paper.

VIDEO: Wilson's Creek spills over its banks, leading to flooding in southwest Greene County. KY3. (2019, May 23). https://www.ky3.com/content/news/VIDEO-Wilsons-Creek-spills-over-its-banks-leading-to-flooding-in-southwest-Greene-County-510324601.html?jwsource=cl.

Yin, Z., Yin, J., Xu, S., & Wen, J. (2011). Community-based scenario modelling and disaster risk assessment of urban rainstorm waterlogging. *Journal of Geographical Sciences*, *21*(2), 274–284. https://doi.org/10.1007/s11442-011-0844-7

II. DEEP NEURAL NETWORK CLASSIFIER FOR FLASH FLOOD SUSCEPTIBILITY

Bhanu Kanwar[1], Jake Hale[1], Steven Corns[1] and Suzanna Long[1]

[1]Department of Engineering Management and Systems Engineering,
Missouri University of Science and Technology, Rolla, MO 65409

ABSTRACT

Compared to normal flood events, flash flooding events are challenging to predict due to their localization and the speed at which they occur. To assess flash flood risks this research compares three machine learning techniques used to predict the likelihood of flooding in Greene County, Missouri: logistic regression, support vector machines, and a deep learning neural network. Publicly available geospatial information and precipitation data associated with the area are used as inputs to these algorithms to classify the likelihood that a flash flooding event had occurred. These predictions are validated against historic flash flood occurrences documented by the National Oceanic and Atmospheric Administration. When compared to a logistic regression model and a support vector machine, the deep learning neural network model had a higher classification accuracy of 85.23%. Adjusted into risk levels, this provides advance warning of high-risk flash flood events to allow emergency managers the ability to issue warnings and close susceptible roadways. These higher quality predictions can reduce danger to the public and assist key stakeholders in developing robust flood management responses.

Keywords: Flash Floods, Deep Learning, Neural Networks, Geospatial, Precipitation.

1. INTRODUCTION AND LITERATURE REVIEW

Natural disasters like flash floods are known to cause both economic and personal damages to the inhabitants of the areas located in both urban and rural watersheds. Flash floods have also been culpable for widespread fatalities across the country (Ashley et al, 2008). Most of these fatalities happen when commuters are stuck inside their stalled vehicles on a submerged road segment (SEMA, 2021). Missouri lost 11 people on a single unfortunate night due to sudden heavy downpour and flash flood (SEMA, 2021). In 2016, 7 people lost their lives due to a flash flooding activity caused by sudden rains in Pulaski County, Missouri (St. Louis Post Dispatch, 2016). In recent decades, climate change has led to a surge in precipitation events which has worsened the situation in the country (Davenport et al, 2021). The ever-increasing cost of floods has motivated researchers to develop and test various methods which can assist in the urgent task of accurately predicting the areas threatened by flash flooding.

Research activities have been conducted in both public and private sectors to predict flood events in susceptible regions. Simulation software like the River Analysis System (HEC-RAS) developed by the US Army Corps of Engineers has been used by flood disaster management teams to predict flooding events (US Army Corps of Engineers, 2021). In addition to HEC-RAS, hydrodynamic models like LISFLOOD-FP and TELEMAC-2D have also been applied to calculate the extent of inundation in a floodplain using parameters like channel friction, etc. and predict river flooding in a

potentially vulnerable location (Horritt et al, 2002). LISFLOOD based hydrological models have also been used by researchers to develop a European Flood Forecasting System (EFFS) which can predict feature values such as water depth and inundation extent (Roo et al, 2010). Hydrologic one-dimensional (1D) and two-dimensional (2D) flow computations are applied to estimate the flood travel time and its inundation area and improve the respective flood warning systems in the risk zones (Ghimire et al, 2020).

Various mathematical models have also been developed to identify key parameters which aid in forecasting the flood hazard efficiently. Bayesian methods such as ensemble Bayesian forecasting system, Bayesian multi-model combination, and others have been shown to be effective in predicting the water level stages of a swollen river after incorporating river parameters such as discharge and runoff in the flood prediction models (Han et al, 2017). Both geomorphological and climatological features have been studied to develop a flood severity model and conduct a seasonality-based analysis to identify risk-prone areas (Saharia et al, 2017). A multilayer perceptron (MLP) model was developed to determine the time at which a river might overflow its banks at certain locations during a flood activity (Baalaji et al, 2020). Machine learning techniques such as alternating decision trees (ADT), functional trees (FT), kernel logistic regression (KLR), multilayer perceptrons, and quadratic discriminant analysis (QDA) algorithms have been implemented on historical databases to develop a model for efficient flash flood susceptibility mapping needed to enhance area's watershed management approach (Janizadeh et al, 2019). Computational intelligence methods like particle swarm optimization (PSO) have also been used to predict the flash flood locales and develop respective flash flood susceptibility maps of Northwest Vietnam (Bui et al, 2019). Dual-

polarization S-band doppler weather radar (SPOL) based precipitation data was used as an input to a model to relay timely flash flood emergency warnings for risk-prone watersheds in Sao Paulo, Brazil (López et al, 2020). A linear regression model-based flash flood warning system was implemented to send short message service (SMS) messages and alert inhabitants of a flood-prone area before the onset of flash flooding in the Philippines (Castro et al, 2013). Taiwanese researchers have relied on large volumes of satellite-derived datasets and rain gauge data points to develop a neural network-based model which can analyze precipitation events resulting in flash floods (Chiang et al, 2007).

Most flash flood-related research focuses on the task of predicting such events in a selected vulnerable region. However, such studies have rarely addressed the need to accurately identify locations that may be affected by flash floods. Failure to identify flash flood-prone traffic areas such as roads can lead to personal losses incurred by the commuters who use them for their commutes. The availability of a large quantity of data accompanied by capable computer hardware has made it possible to implement machine learning techniques and identify patterns to develop better flood identification models for such locations. Since a flash flood activity is both immediate and dangerous, it is imperative to develop models that can correctly identify potential flash flood locations in a susceptible region. The implementation of an efficient machine learning-based classification model will provide the ability for the stakeholders to analyze the complex datasets and prepare suitable flood management responses. This project combines geospatially analyzed flash flood data with a deep learning-based neural network model to address the dynamic nature of volatile flash flooding events and locate potential flood-

prone locations in the region of Greene County, Missouri. The model relies on a data-driven approach to accurately classify and identify different flash flood-prone road segments in this region.

2. MODEL OVERVIEW

The goal of this research is to classify the likelihood of flash flooding at locations that are historically susceptible to flash flooding events. This study is focused on locations in Greene County, Missouri, USA. Three different machine learning models were used to predict flash flood events: deep learning-based artificial neural networks (DL-ANN), logistic regression, and support vector machine (SVM). These algorithms use supervised learning, where input data is used to train the algorithms to give a tagged response associated with that input data. Since the target variable values are already embedded in the labeled dataset used as input to the model, the developed classification model can be conveniently tuned to improve its performance.

With the advent of modern technology, the process of collecting weather, geospatial and meteorological data has become relatively streamlined. The accelerated advancement in the field of machine learning in conjunction with the availability of efficient computer hardware has led to an increase in the application of machine learning techniques in various research sectors. This research project also depends on such techniques to address the real-world challenge of determining the likelihood of a flash flood at the locations considered in this study.

2.1. DEEP LEARNING ARTIFICIAL NEURAL NETWORK

Deep learning artificial neural networks (DL-ANNs) are an elaboration on artificial neural networks (ANNs); mathematical models intended to replicate the architecture of the human brain. They can be used to detect trends and patterns present within the input data by approximating the complex functions representing these patterns. Each ANN consists of an input layer, an output layer, and hidden layers populated with mathematical functions referred to as neurons. These neurons mirror biological neurons and gather the information presented by the data to produce output results to be used by another set of neurons in the next connected layer. Once the necessary information is shared between the input and the hidden layers, the neurons apply a specified activation function to the received input before further feeding the resultant to the final output layer to produce model output. A deep learning artificial neural network is an artificial neural network with several hidden layers (usually three or more).

For the binary classification tasks, the output layer of the DL-ANN consists of a sigmoid activation function which generates an output value in the range [0,1] as shown in Equation (1).

$$S(x) = \frac{1}{1+e^{-x}} \tag{1}$$

Sigmoid function-based classification models use a default probability threshold value of 50% or 0.5 to distinguish between different probability values and identify if an event has occurred. For this research, if the probability values are lower than the threshold value of 50% then flash flooding is not predicted to happen. If the value is equal to or greater than the 50% threshold value, then a flash flooding event is predicted to happen.

2.2. LOGISTIC REGRESSION

Logistic regression is a conventional statistical approach to create a model for binary classification problems. The logistic regression model evaluates the relationship between dependent and independent variables in a dataset and uses a sigmoid function to generate a probability-based output value between 0 and 1 to allocate binary class labels to different location data points (Kambria, 2019). As with the DL-ANN, returned values of less than 0.5 indicate no flash flooding is predicted at the location under evaluation and values of 0.5 or greater indicate a flash flood will occur at that location.

2.3. SUPPORT VECTOR MACHINE

The third machine learning method examined is a support vector machine (SVM). SVMs have two major components: the kernel and a hyperplane. The kernel is a mathematical function that uses the input data and transforms it into the required output format for specified applications (scikit, 2021). Some examples of kernels are the linear kernel, nonlinear kernel, sigmoid kernel, and polynomial kernel, etc. The hyperplane represents a decision boundary that divides the differently labeled data points (yes or no as to whether a flash flood is predicted), with the points bordering the decision boundary are known as support vectors. Support vector machine classifiers identify the best decision boundary, or hyperplane, to correctly distinguish between labeled points in a dataset. The distance between the nearest data point and the hyperplane is called margin (MathWorks, 2021). The model identifies the hyperplane such that the margin value of a data point belonging to a particular class is as high as possible. To deliver high

classification accuracy, all data points associated with a given class should also be situated on the correct side of this hyperplane.

To further improve a SVM's classification performance, the Grid Search approach is implemented to tune the settings of the algorithm, called hyperparameters. These hyperparameters include misclassification cost, gamma, and the kernel used. Misclassification cost represents an acceptable amount of misclassification error for the classification task and the gamma value highlights the influence of the data points located on either side of the hyperplane. In the Grid Search approach, the model tested different combinations of hyperparameter values chosen through trial and error to identify a set of values that can improve its classification accuracy score. The dataset is then re-fitted on the updated model for training purposes. Once the model is re-trained, it is run on the testing dataset to classify data points between the two classes.

2.4. DATA SOURCES

The data used to determine flash flooding in susceptible areas of Greene County, Missouri is publicly available, open-source data acquired from the United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), and National Weather Service (NWS). Information related to previous flash flood events in the county are gathered from the 'Storm Events Database' to prepare a database of the flash flood-prone locations in the region. This database is maintained by NOAA and contains information on the time and location of severe weather events across the country (NOAA, 2021). The 'Storm Events Database' contains exhaustive information related to flash floods that have occurred in the area of interest dating back to 1950. NOAA's

National Weather Service (NWS) maintains and develops this database using comprehensive inputs from trained storm spotters and emergency response personnel to collect the flash flood event details at an affected location. It is the responsibility of the spotter to gather information on the areas and road segments affected by a flash flood and store the necessary details of the event in the database. The details of the flash flood events in the test region are collected from this database to identify the affected roads between 2005 and 2019. Elevation data, rainfall data, and soil moisture are used as inputs for all the models. Elevation data is available from the USGS's National Map to develop the raster elevation layer of the area of interest (USGS, 2021). Rainfall data is collected from the NWS for the period matching the 'Storm Events Database' (NOAA, 2021). Soil moisture is accounted for by using rainfall amounts for the three days preceding the time frame being evaluated using data from NWS. To develop a machine learning model capable of predicting these events, the data must be divided into a training set that the algorithms will use to determine relationships and a testing set that will be used to validate the accuracy of the model.

2.5. MODEL EVALUATION METRICS

The classification accuracy of the models is validated using three performance metrics (Accuracy, Precision, and Area Under the Receiver Operating Characteristic Curve) and displayed in a confusion matrix and receiver operating characteristic (ROC) curve graph. Different sets of confusion matrices are generated to represent the number of actual and predicted flash flood locations in the test region. The four main components of a confusion matrix are True Positive (TP), True Negative (TN), False Positive (FP), and

False Negative (FN). True positive represents a predicted true value that is actually true and false positive is a predicted true value that is actually false. True negative is the predicted false value that is actually false and false negative is the predicted false value that is actually true. The accuracy score of a model is calculated as the sum of both True Positive and True Negative data values divided by the total number of values used as input to the model as shown in Equation (2). The accuracy scores of all the models are compared against each other to choose the most accurate flash flood classification model.

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{(Total\ number\ of\ values\ or\ observarions)} * 100 \quad (2)$$

Positive Prediction Value (PPV), or Precision, is a measure of how well the model predicted positive results. It is calculated as the number of true positive data values divided by the sum of true positive and false positive values as shown in Equation (3).

$$Positive\ Prediction\ Value = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

True Positive Rate (TPR), or Recall, determines how well the model correctly identifies true positives. It is defined as the number of true positive values divided by the sum of true positive and false negative data values as shown in Equation (4).

$$True\ Positive\ Rate = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

False Positive Rate (FPR) determines the ratio of negative cases that are incorrectly labeled as positive by the classification model. It is defined as the number of false positive values divided by the sum of false positive and true negative data values as shown in Equation (5).

$$False\ Positive\ Rate = \frac{False\ Positive}{False\ Positive + True\ Negative} \quad (5)$$

The receiver operating characteristic (ROC) curve and the area under the ROC (AUROC) values are calculated to analyze the classification performances of the machine learning models (Google, 2021). The ROC curve is a plot between the TPR and the FPR that visualizes a model's performance for different classification threshold values. AUROC is the area below the ROC curve of a classifier with values in the range from 0 to 1. An AUROC value of 0.5 means that the model is not able to distinguish between different classes, whereas a score between the values of 0.7 and 0.8 signifies a model with an acceptable classification performance. The AUROC values lying between the ranges of 0.8 to 0.9 and 0.9 to 1.0 denote that the models have excellent and outstanding classification capabilities, respectively (Mandrekar, 2015).

3. METHODOLOGY

3.1. STUDY AREA

The area of interest for the classification models developed in this study is Greene County, Missouri, and in particular the locations with commonly used roads. Greene County is in the southwestern part of Missouri and is one of the most populous counties in Missouri, containing the Springfield metropolitan area (US Census Bureau, 2021). Greene county also experiences inclement weather conditions throughout the year (Clayton News, 2021). As per the local Springfield National Weather Service Office (NWSO), the county experiences an average of 39 heavy rain events each year that can lead to flash flood events, especially in and around the areas of Springfield (NOAA, 2015). The Springfield metropolitan area consists of busy highways such as Interstate 44,

U.S. Route 60, U.S. Route 65, and U.S. Route 165, etc. which pass through the county (MoDOT, 2021). The area's creeks and ravines can experience flash flooding, with damage to road segments resulting in both economic and personal losses to its residents (Louzader, 2020). Local law enforcement authorities must sometimes conduct rescue operations for the commuters stuck in their vehicles on flash flood-hit roads (KY3, 2021). Hence, it is crucial to investigate both geospatial and rainfall information of the region to develop a dependable model which can identify roads segments susceptible to flash flooding activities to anticipate and minimize the impact of these events. Identifying these locations in the region will assist the city planners and government agencies such as the Missouri Department of Transportation (MoDOT) and the Missouri State Emergency Management Agency (SEMA) in taking necessary precautionary measures in advance.

3.2. DATA COLLECTION

To prepare the dataset for machine learning models, information on historic flash flood events is collected from NOAA's database. The events can be located by searching the database to determine when they occurred, but all information related to location is entered as text that must be analyzed to determine where the event took place. All the events used for training the machine learning algorithms had to be manually analyzed to determine these locations. Once the time and location of these events are known, information on the geographical features of the area that can affect whether a flash flood occurs is determined. Both the time and geographical location data are used to gather rainfall data during and three days prior to the possible flash flood event to account for

soil saturation and the amount of rainfall for the event. This information is used as the input to the models for identifying the flash flood-prone locations.

Pre-processing is performed on the data to gather and convert it into a suitable format to be used as inputs for the different classification models. A database containing historic flash flood events information is built to create explanatory data layers based on four different geospatial variables: elevation, slope, aspect, and curvature. The entries in this database also contain the latitude and longitude information of the locations of interest to accurately visualize them in ArcGIS Pro. The location information from this database is combined with the road network obtained from the National Transportation Dataset (NTD) to highlight road segments that may be affected by flash floods (USGS, 2021).

A labeled location data set of 350 data points is built from the historic flood event database depicting a set of locations in Greene County using a point sampling operation in ArcGIS Pro. These location points are labeled as either locations where flash flooding has occurred or locations where it has not occurred. Out of a total of 350 data points, 165 data points correspond to recorded flash flood events and 185 are categorized as non-flash flood events as shown in Figure 1. The flash flood locations are represented with red dots in Figure 2 and the non-flash flood locations are represented with green dots in Figure 3. The elevation data for the test region are gathered by the USGS and stored in Light Detection and Ranging (LIDAR) format. The lidar-based elevation data for this region is analyzed using ArcGIS Pro (esri, 2021) to generate the raster elevation layer of the selected area using ArcGIS Pro's in-built raster tool and clipping procedures as displayed in Figure 4.

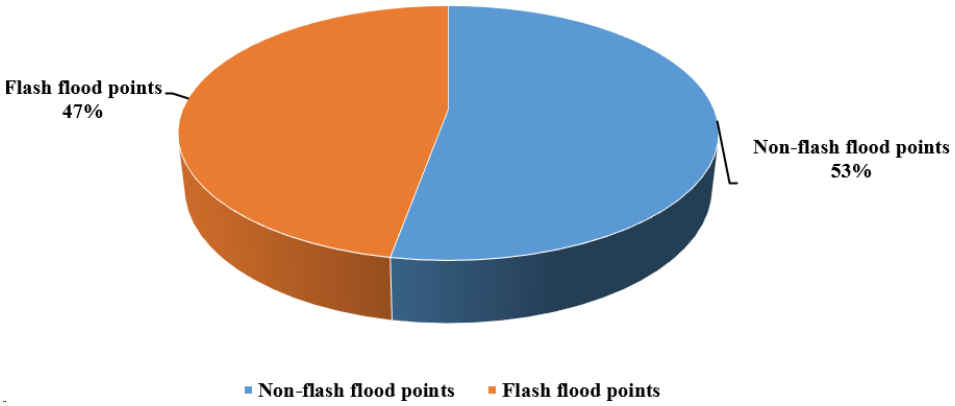

Figure 1. Datapoints distribution in dataset.

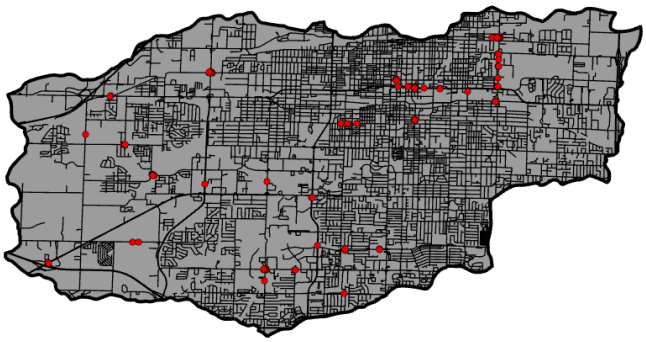

Figure 2. Flash flood locations.

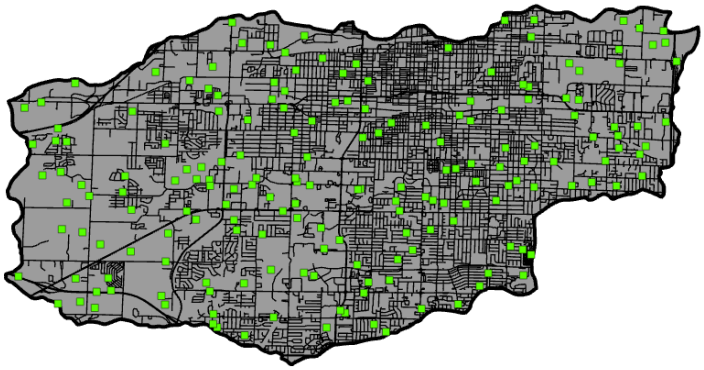

Figure 3. Non-flash flood locations.

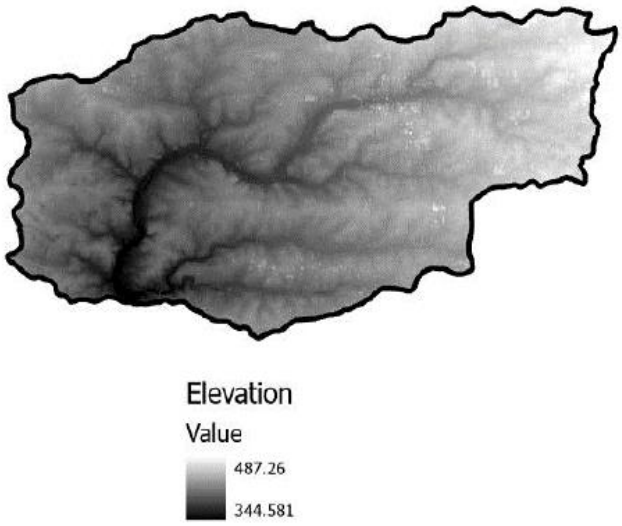

Figure 4. Elevation profile.

ArcGIS Pro's slope, aspect, and curvature tools are also used to obtain the remaining slope as shown in Figure 5, aspect as shown in Figure 6, and curvature as shown in Figure 7 layers for the region.

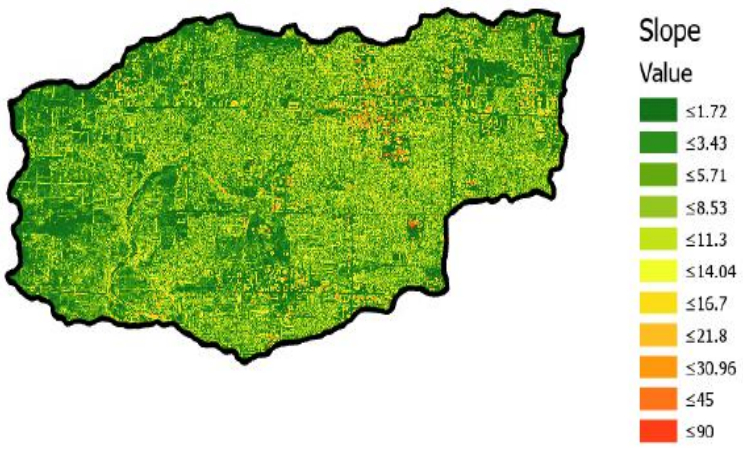

Figure 5. Slope profile.

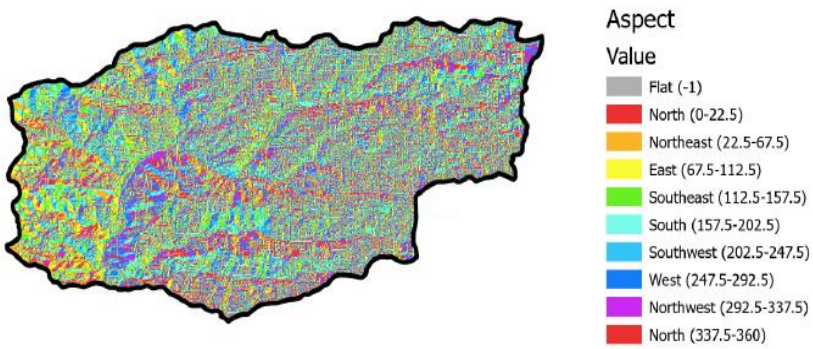

Figure 6. Aspect profile.

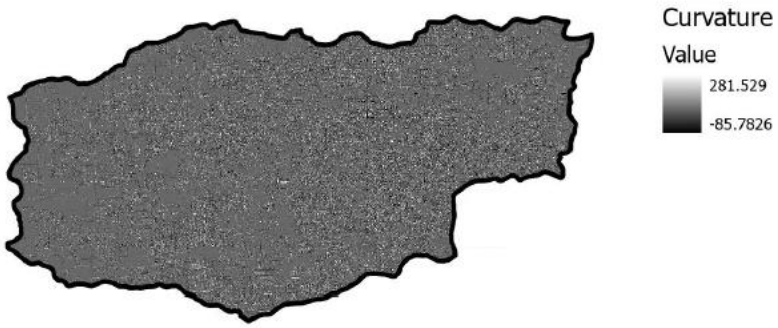

Figure 7. Curvature profile.

Rainfall data from NOAA does not map directly to the locations in the geographically data set. To compensate for this, a nearest neighbor algorithm is used in ArcGIS Pro to associate the historical rainfall information to the location data so that the input information can be properly aligned (esri, 2021). This was done for all data gathered from January 2005 to June 2019 and associated with the explanatory layers.

The collected dataset is subjected to data cleaning, exploratory analysis, and data normalization to put it into a common format for analysis. This also assists in getting a basic understanding of the multi-variate dataset under examination. The nine different independent variables in the dataset are ‘Rainfall day of Inches’, ‘Rainfall 1 day before’,

'Rainfall 2 days before', 'Rainfall 3 days before', 'Summation of antecedent rainfall', 'Curvature', 'Aspect', 'Slope' and 'Elevation' as illustrated in Figure 8. The dataset's dependent variable is a binary titled 'Class label' indicating whether a flash flood event occurred. The values for all locations are compiled and saved in a tabular comma-separated value (CSV) file format.

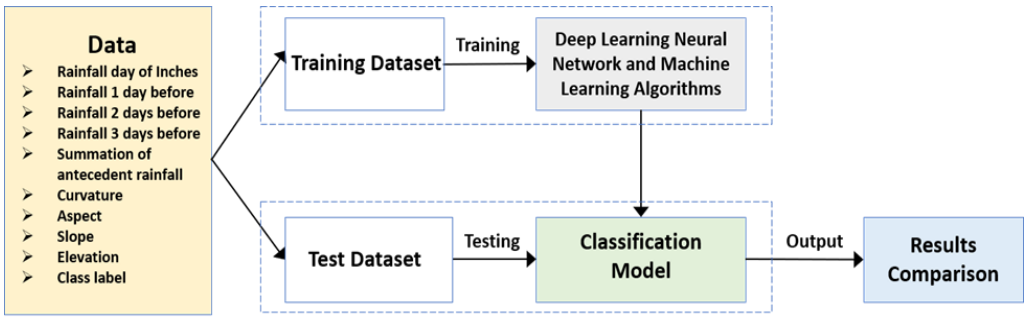

Figure 8. Classification model architecture.

The whole database is divided into two sets of training and testing datasets for further analysis. The training set consists of 75% of the total dataset whereas the remaining 25% is assigned to the testing set. Specifically, the training set is made of 262 data points or observations whereas the testing set comprises 88 points. Apart from the labeled variables information, the testing set is also tagged with the respective truth values (class label) required to investigate the accuracy of the classification models.

3.3. CLASSIFICATION MODELS

The data consists of nine different geospatial and precipitation variables which are used as inputs to all the models. The deep learning neural network model's architecture

consists of a ‘Sequential’ class that utilizes Keras’ sequential application program interface (API) to incrementally add four different hidden layers to it (Keras, 2021). The ‘Dense’ fully connected input layer uses a one-dimensional array of nine input elements to the model with a default rectified linear unit or ‘ReLU’ activation function. The ‘ReLU’ activation function applies a non-linear transformation to the input data to make it linearly separable for the binary classification task of classifying locations (Keras, 2021). The fully connected hidden layers consist of the ‘Sigmoid’ activation function applied to its input values. The final output ‘Dense’ layer of the model uses the sigmoid function to assign a value ranging from 0 to 1 to the single location data point.

A dropout value of 0.2 is also applied between the hidden layers of the neural network. Using a dropout value of 0.2, 20% of a layer’s output neurons are randomly dropped to prevent overfitting of the model during its training phase. An overfitted model performs well on the training data but can produce a high-test error when evaluating an unseen testing dataset (IBM, 2021). ‘Binary Cross-entropy’ loss function and ‘Adam’ learning rate optimizer are used for compiling the deep learning model after defining its architecture. The purpose of using the binary cross-entropy loss function is to find the error in the model’s binary classification learning process (TensorFlow, 2021). The model is trained on the training dataset using the ‘epochs’ and ‘batch size’ values of 1000 and 10, respectively. After training the model, the testing dataset is fitted to it to obtain the class labels and the probability values of each test data point. These output class label values assist in distinguishing between different location test data points belonging to either the flash flood or non-flash flood category.

All models are initially trained on the same training dataset to capture the relationships between its different variables. Once trained, the models are then used to evaluate the testing dataset to evaluate the overall accuracy of the methods. In the dataset, a one denotes a flash flood occurred and a zero denotes a flash flood did not occur. For the predicted values, a one denotes a predicted flash flood and a zero denotes a flash flood was not predicted. When the testing data set was generated, 46 of the data points represented locations where there was flash flooding and 42 of the data points represented locations where there was no flash flooding.

4. RESULTS

4.1. DEEP LEARNING ARTIFICIAL NEURAL NETWORK

The classification performance of the deep learning neural network model can be visualized using a confusion matrix. The parameters of the matrix are used to calculate performance metrics such as accuracy and precision. The deep learning neural network model achieves a classification accuracy of 85.23% and a precision of 94.59% when tested on the testing set with feature information for 88 locations. It correctly classifies a total of 75 out of 88 locations with an error rate of 14.77%. The model also correctly identifies 35 potential hazardous locations with class label 1 and 40 locations are correctly depicted by class label 0 as shown in Figure 9. As per the model, 13 locations are incorrectly predicted with 11 false negatives and 2 false positives.

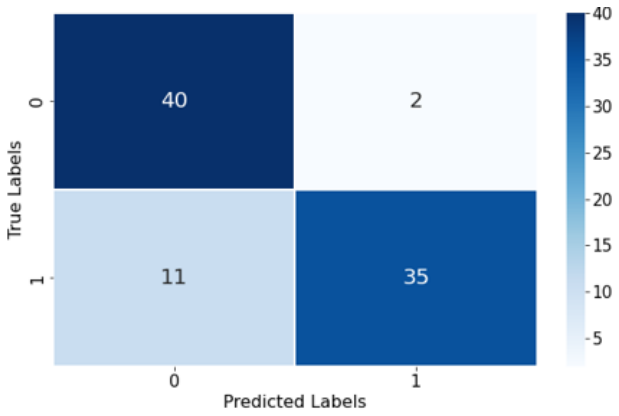

Figure 9. Deep learning artificial neural network confusion matrix.

4.2. LOGISTIC REGRESSION

The logistic regression classifier did not perform as well as the DL-ANN classification model and produced an accuracy score of 75% and a precision of 90%. The regression classifier only correctly classified 27 locations where flash flooding occurred, and 39 locations are classified as non-flash flood locations as shown in Figure 10. As per the model, 22 locations are incorrectly predicted with 19 false negatives and 3 false positives.

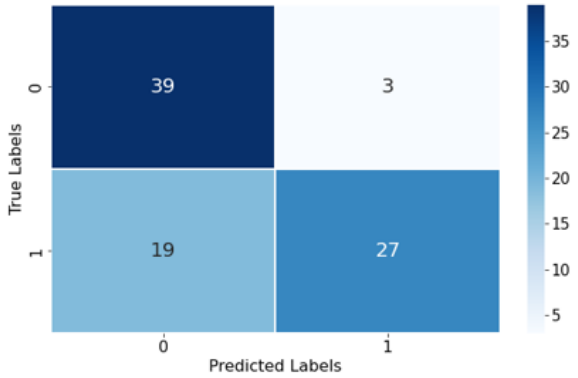

Figure 10. Logistic regression confusion matrix.

4.3. SUPPORT VECTOR MACHINE

Initially, the support vector machine model performed worse than the other classifiers, achieving an accuracy score of 59.09% and a precision of 65.62%. When Grid Search was implemented to tune the hyperparameters the results improved, with an accuracy score of 80.68% and a precision of 85.30%. The model correctly classified 35 locations where flash flooding occurred, and 36 locations classified as non-flash flood as shown in Figure 11. The error rate also dropped to 19.32%, which is less than the error rate of the logistic regression model but more than the DL-ANN's error rate.

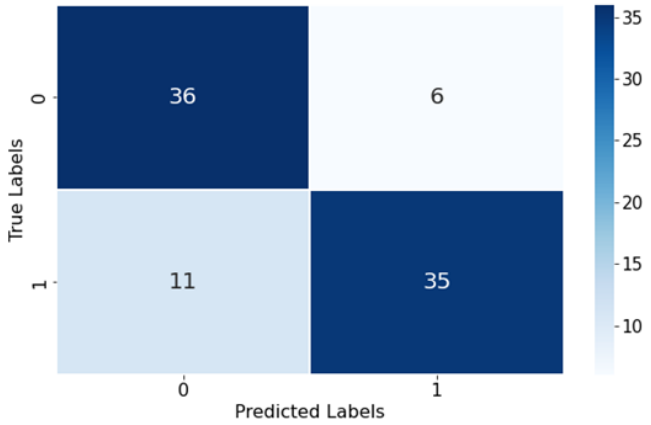

Figure 11. Support vector machine with grid search confusion matrix.

5. DISCUSSION

The accuracy, precision and AUROC values of all models are compared to identify the best performing model. When considering accuracy as shown in Figure 12, the support vector machine classifier without any hyperparameters tuning is the worst-

performing model, followed by the regression model. When Grid Search was added to the SVM model it performed better than the regression model, but still not as well as the DL-ANN.

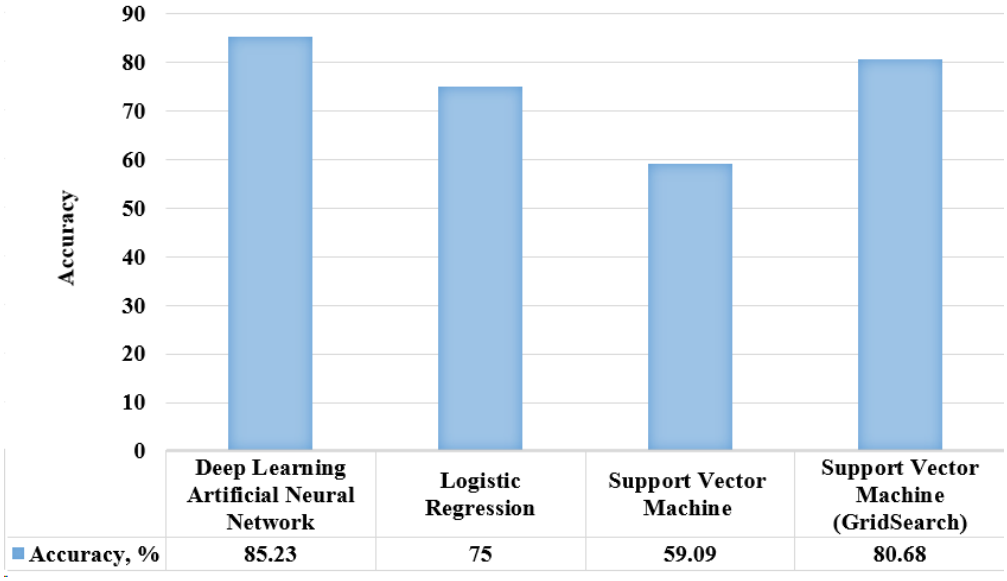

Figure 12. Classification accuracy scores comparison.

Precision is a measure of the classification model's ability to predict positive results. Figure 13 shows that the deep learning-based model achieves the highest precision score of 94.59% out of all classification models. It means that as compared to other models, when the DL-ANN predicts flash floods, it is correct 94.59% of the time. As a result, the DL-ANN model is more reliable than the other models when classifying a given event as a flash flood event.

The deep learning-based model achieves the highest accuracy score of 85.23%. It also had the highest AUROC value of 0.8727. Figure 14 illustrates the AUROC for the deep learning model which is the area under the yellow-colored ROC curve.

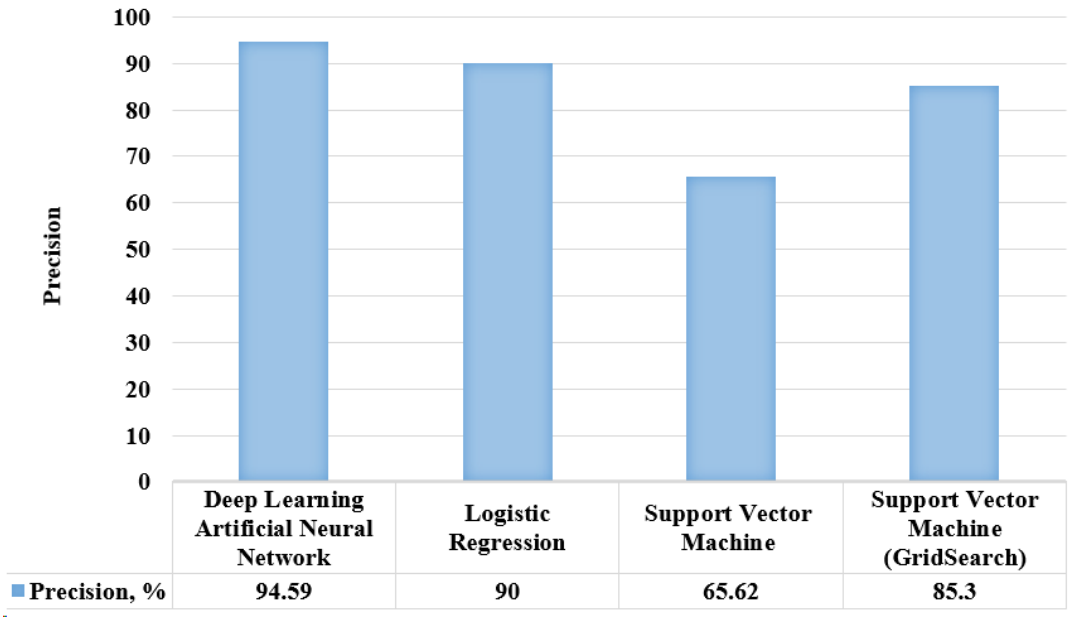

Figure 13. Precision scores comparison.

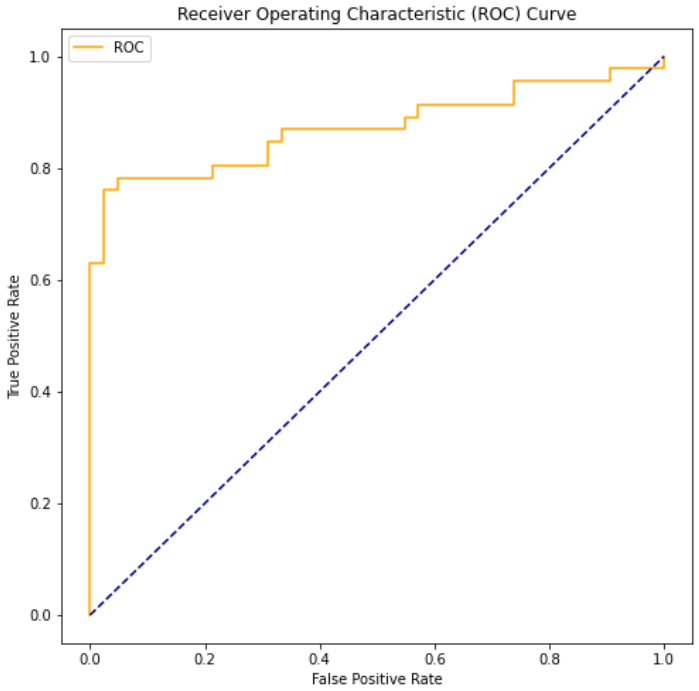

Figure 14. Deep learning artificial neural network ROC curve.

The logistic regression classification model performed worse than the DL-ANN, with an AUROC score of 0.7893 as shown in Figure 15. This means that as compared to the logistic regression model, the deep learning-based classifier produces a greater number of true positive values and a lower number of false positives. For reference, a dashed line indicating an AUROC value of 0.5 has been included. An AUROC value of 0.5 means that a classifier is not able to distinguish between the data values belonging to either a positive or negative class.

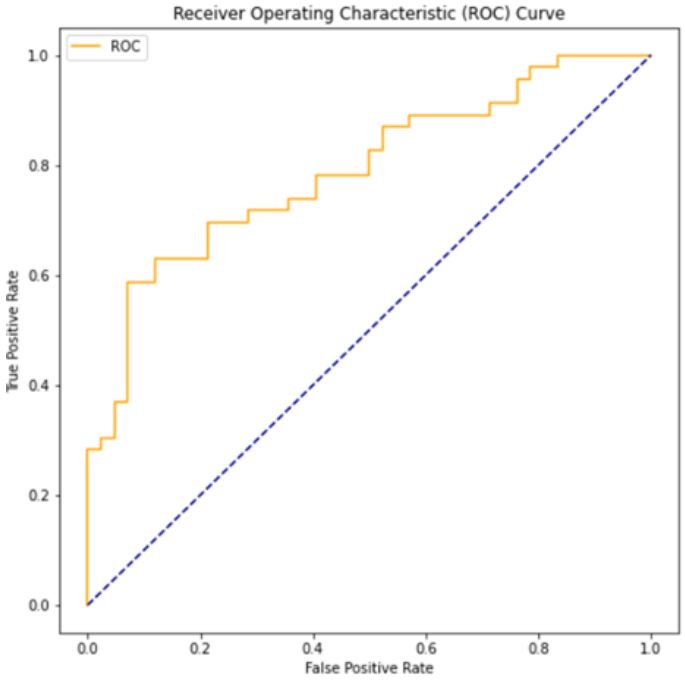

Figure 15. Logistic regression ROC curve.

The support vector machine achieved the lowest AUROC value (0.5973), but this improved to 0.8090 after the implementation of Grid Search as shown in Figure 16. Both the SVM with Grid Search and the DL-ANN models had eleven false negative

predictions, leading to a lower true positive rate. The DL-ANN model generates only 2 false positive values while the SVM with Grid Search reported 6 false positives as shown in Figures 9 and Figures 11, respectively. DL-ANN model's lower number of false positives means that the stakeholders will have to deal with a fewer number of false flood alarms. Instead, they can focus on the real scenarios when there are actual floods in the region. Figure 17 shows that the DL-ANN model has the highest AUROC value as compared to the other models which makes it a more reliable model for the task of identifying locations with flash flood.

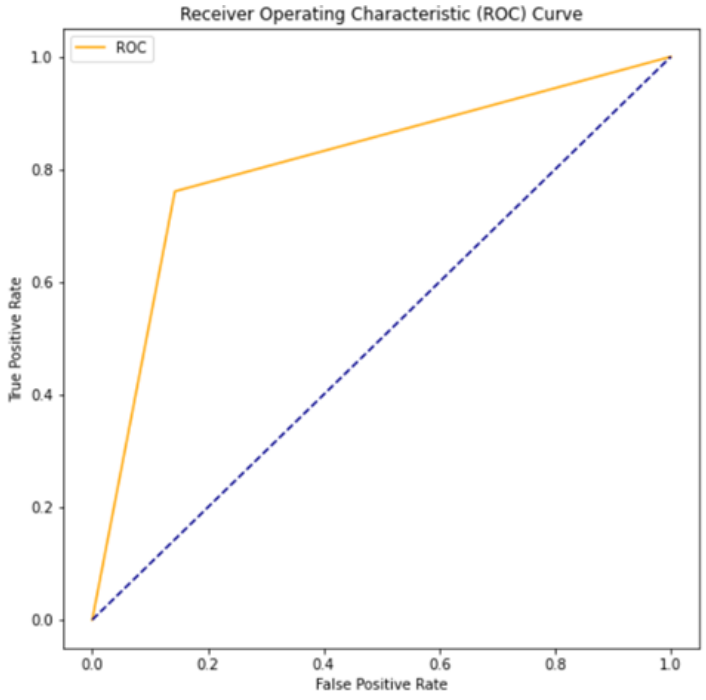

Figure 16. Support vector machine with grid search ROC curve.

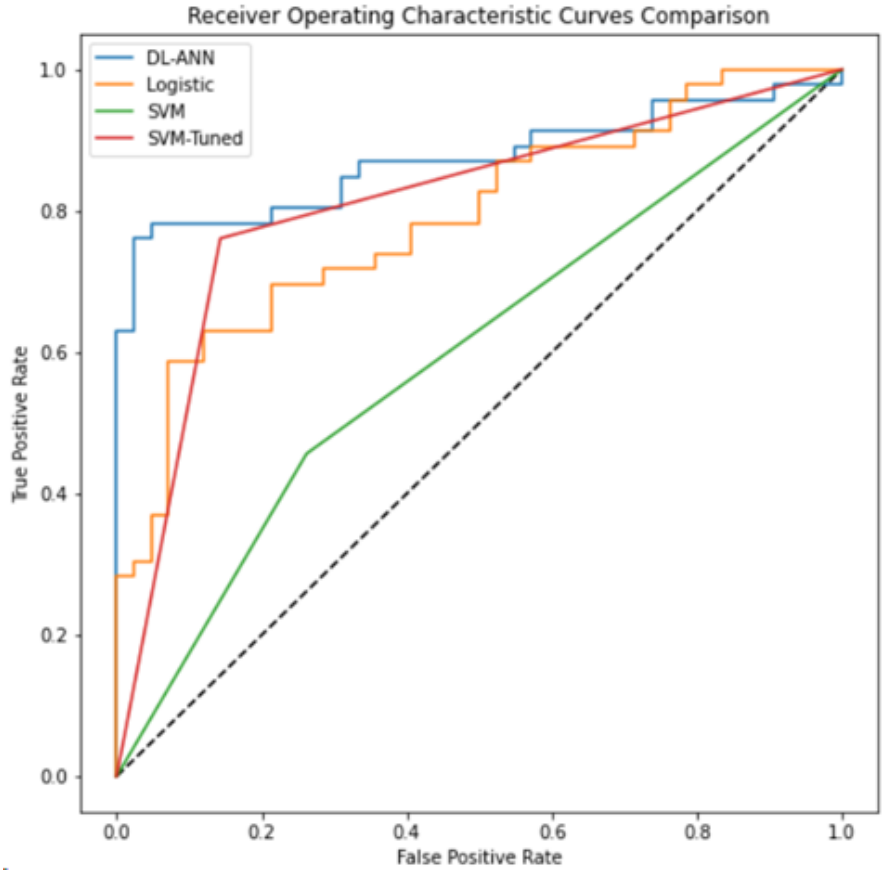

Figure 17. ROC curves comparison.

Based on a comparison of the results for accuracy, precision and AUROC values, the DL-ANN model emerged as a best model of those considered for classifying flash flood locations, with accuracy, precision and AUROC values of 85.23%, 94.59% and 0.8727, respectively as shown in Table 1. While there were more false positives in the SVM with Grid Search model than in the DL-ANN, the number of false negatives was the same (11), meriting some consideration for further analysis. The deep learning-based classifier accurately identified 75 locations out of a total of 88 locations at risk of experiencing flash floods in the Springfield region of Greene County.

Table 1. Comparison of performance metrics of classification models.

Classification Models	**Deep Learning Artificial Neural Network**	**Logistic Regression**	**Support Vector Machine**	**Support Vector Machine with Grid Search**
Accuracy Score	85.23%	75%	59.09%	80.68%
Precision Score	94.59%	90%	65.62%	85.30%
AUROC Value	0.8727	0.7893	0.5973	0.8090

As an example, the deep learning-based classification model identifies Chestnut Expressway as one of the locations affected by flash floods. Chestnut Expressway often gets flooded after periods of heavy rainfall, resulting in stalled vehicles in floodwaters (KY3, 2020). After a heavy downpour of rain, the low-lying segments of this road must be regularly closed for traffic activity due to the occurrence of flash floods in the area (Randall, 2021; Strohl, 2019). In 2019, a section of Chestnut Expressway was closed for traffic as it was inundated with 4 feet of floodwater (Simmons, 2019). This location was correctly predicted by the deep learning-based artificial neural network model. Also correctly identified by the deep learning-based classification model, the section of West Farm Road 146 running parallel to Wilsons Creek, which is also prone to flash flood activity. This road segment was overtopped when the adjoining Wilsons Creek overflowed its banks after a period of rainfall in the region (KY3, 2019). The floodwaters effecting Wilsons Creek have also impacted the region's various other roads in the past (Simmons, 2019).

6. CONCLUSIONS

Three different models are implemented on the database generated from the publicly available geospatial and rainfall information to determine which best predicts flash flooding events. The deep learning neural network model had the best performance identifying locations such as road segments that were affected due to flash floods. The deep learning-based classification model had the highest accuracy, precision and AUROC values. The deep learning model also takes less time to analyze large datasets as compared to logistic regression and SVM models, making it suitable for the time-sensitive task of forecasting flash flood events in susceptible regions. Its architecture can also be updated to incorporate more complex datasets with a variety of weather and geospatial variables if additional data becomes available. Using the model's output, the region's different infrastructure elements such as roads can be monitored by the authorities to warn commuters of any anticipated flash flood activity in the area. The timely issuance of road closure warnings in such locations will also help the commuters to prepare for any traffic delays and detours well in advance of their commute. The output of the model can also assist in preparing advanced flash flood-related information needed to safeguard the lives and property of the people living in these flood-prone areas. The DL-ANN classifier captures the intricate relationships between data variables to demonstrate an efficient approach that can be used by personnel in charge of disaster management to capably deal with the problems posed by flash flooding. Those in charge of taking key decisions in such scenarios can use the model output to augment their flood warning responses and prevent loss of both life and property.

7. FUTURE WORK

The DL-ANN model proposed in this research can be deployed by people overseeing the flash flood prediction and management responses in flood-sensitive locales. Notwithstanding the neural network's effectiveness, it can be tested on datasets containing miscellaneous features like normalized difference vegetation index (NDVI), etc. in future research undertakings. The scalability of this model can further be examined by employing it in different flash flood-prone drainage basins. An economic study can also be conducted to understand the direct and indirect economic losses endured by people and areas affected by flash flood accidents.

Rather than a set yes/no result, probabilities could be reported as output from some of these algorithms to determine the likelihood of flash flooding. While the exact values would not be overly useful, binning the probabilities and using this as part of a risk analysis could draw attention to some of the events that would otherwise be misclassified. In addition, many deep learning methods exist, so a more thorough analysis of the different algorithms could produce improved results. Given the time sensitive nature of the data used to capture soil moisture and rainfall data, time series analysis methods may prove beneficial. This could include algorithms incorporating memory structures such as Long Short-Term Memory machines or deep learning neural networks used to drive a finite state machine. Any reduction in error realized by these methods would be a benefit to public safety.

REFERENCES

Ashley, S. T., & Ashley, W. S. (2008, March 01). Flood Fatalities in the United States. Retrieved from https://journals.ametsoc.org/view/journals/apme/47/3/2007jamc1611.1.xml

Baalaji S, V., & S, S. (2020). Flood Prediction System using Multilayer Perceptron Classifier and Neural Networks. International Research Journal of Engineering and Technology (IRJET), 07(05). https://www.irjet.net/archives/V7/i5/IRJET-V7I51189.pdf

Bogan, J. (2016, January 20). *7 killed in Missouri flooding last month died just minutes and a few miles apart*. STLtoday.com. Retrieved November 1, 2021, from https://www.stltoday.com/news/local/metro/7-killed-in-missouri-flooding-last-month-died-just-minutes-and-a-few-miles-apart/article_71f566f6-2844-5884-bf1e-09533ab7508f.html.

Bui, D. T., Ngo, P. T., Pham, T. D., Jaafari, A., Minh, N. Q., Hoa, P. V., & Samui, P. (2019). A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. Catena, 179, 184-196. doi:10.1016/j.catena.2019.04.009

Castro, Joel T. de, Gabriel, Salistre M., Byun, Young-Cheol, & Gerardo, B. (2013). Flash Flood Prediction Model Based on Multiple Regression Analysis for Decision Support System. Retrieved from https://www.researchgate.net/publication/289269355_Flash_Flood_Prediction_Model_based_on_Multiple_Regression_Analysis_for_Decision_Support_System

Chiang, Y., Hsu, K., Chang, F., Hong, Y., & Sorooshian, S. (2007, April 20). Merging multiple precipitation sources for flash flood forecasting. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022169407002090

Clayton News. (2021, March 31). County with the most severe weather in every state. Retrieved November 1, 2021, from https://www.news-daily.com/multimedia/photos/county-with-the-most-severe-weather-in-every-state/collection_da46773c-de17-592c-889f-d2682b9da532.html#26.

Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. Proceedings of the National Academy of Sciences, 118(4). doi:10.1073/pnas.2017524118

esri. (Last Accessed November 1, 2021). *How average nearest neighbor works*. How Average Nearest Neighbor works-ArcGIS Pro | Documentation. Retrieved November 1, 2021, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-average-nearest-neighbor-distance-spatial-st.htm.

esri. (Last Accessed November 1, 2021). *What is a las dataset?* What is a LAS dataset?-ArcGIS Pro | Documentation. Retrieved November 1, 2021, from https://pro.arcgis.com/en/pro-app/latest/help/data/las-dataset/what-is-a-las-dataset-.htm.

Ghimire, E., Sharma, S., & Lamichhane, N. (2020). Evaluation of one-dimensional and two-dimensional HEC-RAS models to predict flood travel time and inundation area for flood warning system. ISH Journal of Hydraulic Engineering, 1-17. doi:10.1080/09715010.2020.1824621

Google. (Last Accessed November 7, 2021). Classification: Roc curve and AUC | machine learning crash course. Google. Retrieved November 7, 2021, from https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc.

Han, S., & Coulibaly, P. (2017). Bayesian flood forecasting methods: A review. Journal of Hydrology, 551, 340-351. doi:10.1016/j.jhydrol.2017.06.004

Horritt, M., & Bates, P. (2002). Evaluation of 1D and 2D numerical models for predicting river flood inundation. Journal of Hydrology, 268(1-4), 87-99. doi:10.1016/s0022-1694(02)00121-x

IBM. (Last Accessed November 1, 2021). *What is Overfitting?* Overfitting. Retrieved November 1, 2021, from https://www.ibm.com/cloud/learn/overfitting.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E., Lee, S. (2019). Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. Sustainability, 11(19), 5426. doi:10.3390/su11195426

Kambria. (2019, August 27). Logistic regression for machine learning and classification. Retrieved November 1, 2021, from https://kambria.io/blog/logistic-regression-for-machine-learning/.

Keras. (Last Accessed November 1, 2021). Keras documentation: Layer Activation functions. Retrieved November 1, 2021, from https://keras.io/api/layers/activations/.

Keras. (Last Accessed November 1, 2021). Keras Documentation: The Functional API. Retrieved November 1, 2021, from https://keras.io/guides/functional_api/#first-example-a-densely-connected-network.

KY3. (2019, May 23). Video: Wilson's Creek spills over its banks, leading to flooding in southwest Greene County. https://www.ky3.com. Retrieved November 7, 2021, from https://www.ky3.com/content/news/VIDEO-Wilsons-Creek-spills-over-its-banks-leading-to-flooding-in-southwest-Greene-County-510324601.html?jwsource=cl.

KY3. (2020, May 22). Pictures: Flooding reported across Springfield and Greene County, MO. https://www.ky3.com. Retrieved November 7, 2021, from https://www.ky3.com/content/news/Flooding-reported-across-Springfield-and-Greene-County-Missouri-570689491.html.

KY3. (2021, July 17). *First alert weather: Flash flooding impacts some Springfield commutes, leads to water rescues*. https://www.ky3.com. Retrieved November 1, 2021, from https://www.ky3.com/2021/07/17/first-alert-weather-flash-flooding-impacts-some-springfield-commutes-flash-flood-warning-remains-effect/.

Laio, F., Porporato, A., Revelli, R., & Ridolfi, L. (2003). A comparison of nonlinear flood forecasting methods. Water Resources Research, 39(5). doi:10.1029/2002wr001551

López, A. S., & Rodriguez, C. A. (2020). Flash Flood Forecasting in São Paulo Using a Binary Logistic Regression Model. Atmosphere, 11(5), 473. doi:10.3390/atmos11050473

Louzader, D. (2020, May 22). Flash Flooding In Springfield Metro Area. Retrieved from https://www.ktts.com/2020/05/22/flash-flooding-in-springfield-metro-area/

Mandrekar, J. N. (2015, November 20). Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology. Retrieved November 7, 2021, from https://www.sciencedirect.com/science/article/pii/S1556086415306043#bib5.

MathWorks. (Last Accessed November 1, 2021). Support Vector Machine (SVM). Retrieved November 1, 2021, from https://www.mathworks.com/discovery/support-vector-machine.html.

MoDOT. (Last Accessed November 1, 2021). Retrieved November 1, 2021, from http://traveler.modot.org/map/?district=SOUTHWEST.

NOAA. (2015, March 27). *Severe weather climatology for the Springfield, MO. NWSO County Warning Area*. National Weather Service. Retrieved November 1, 2021, from https://www.weather.gov/sgf/severe_climatology_paper.

NOAA. (Last Accessed November 1, 2021). Storm Events Database. Retrieved November 1, 2021, from https://www.ncdc.noaa.gov/stormevents/.

Randall, C. (2021, July 18). Flooding causes numerous water rescues in Springfield. OzarksFirst.com. Retrieved November 8, 2021, from https://www.ozarksfirst.com/local-news/local-news-local-news/flooding-causes-numerous-water-rescues-in-springfield/.

Roo, A. P. J. de, Gouweleeuw, B., Thielen, J., Bartholmes, J., Baongioannini-Cerlini, P., Todini, E., Bates, P. D., Horritt, M., Hunter, N., Beven, K., Pappenberger, F., Heise, E., Rivin, G., Hils, M., Hollingsworth, A., Holst, B., Kwadijk, J., Reggiani, P., Dijk, M. V., Sprokkereef, E. (2010). Development of a European flood forecasting system. International Journal of River Basin Management, 49–59. https://doi.org/https://doi.org/10.1080/15715124.2003.9635192

Saharia, M., Kirstetter, P., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping Flash Flood Severity in the United States. Journal of Hydrometeorology, 18(2), 397-411. doi:10.1175/jhm-d-16-0082.1

scikit. (Last Accessed November 1, 2021). 1.4. Support Vector Machines. Retrieved November 1, 2021, from https://scikit-learn.org/stable/modules/svm.html.

scikit. (Last Accessed November 1, 2021). Sklearn.model_selection.GRIDSEARCHCV. Retrieved November 1, 2021, from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

SEMA. (Last Accessed November 1, 2021). *Flooding*. Retrieved November 1, 2021, from https://sema.dps.mo.gov/plan_and_prepare/flooding.php.

Simmons, L. (2019, May 23). Flash flooding closes roads around Springfield Airport, Willard, MO. https://www.ky3.com. Retrieved November 7, 2021, from https://www.ky3.com/content/news/Flash-flooding-near-Springfield-airport-Willard-510320531.html.

Statista. (2021, October 2). Economic damage caused by floods in the U.S. 2019. Retrieved November 1, 2021, from https://www.statista.com/statistics/237420/economic-damage-caused-by-floods-and-flash-floods-in-the-us/.

Strohl, B. (2019, May 23). Chestnut Expressway, major roads near Springfield-Branson Airport closed from flooding. KTTS. Retrieved November 7, 2021, from https://www.ktts.com/2019/05/23/chestnut-expressway-major-roads-near-springfield-branson-airport-closed-from-flooding/.

TensorFlow. (Last Accessed November 1, 2021). Tf.keras.model ： tensorflow core v2.6.0. Retrieved November 1, 2021, from https://www.tensorflow.org/api_docs/python/tf/keras/Model?hl=ru.

US Army Corps of Engineers. (Last Accessed November 1, 2021). HEC-RAS. Retrieved November 1, 2021, from https://www.hec.usace.army.mil/software/hec-ras/.

US Census Bureau. (Last Accessed November 1, 2021). Greene County Quickfacts from the US Census Bureau. Retrieved November 1, 2021, from https://web.archive.org/web/20110607041738/http://quickfacts.census.gov/qfd/states/29/29077.html.

USGS. (2021, July 16). U.S. Geological Survey, National Geospatial Technical Operations Center. Retrieved November 1, 2021, from https://www.sciencebase.gov/catalog/item/5a61c92fe4b06e28e9c3bd7b.

USGS. (Last Accessed November 1, 2021). *National Geospatial Program*. The National Map. Retrieved November 1, 2021, from https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map.

USGS. (Last Accessed November 1, 2021). National hydrography. Retrieved November 1, 2021, from https://nhd.usgs.gov/useGuide/Robohelpfiles/NHD_User_Guide/Feature_Catalog/Hydrography_Dataset/NHDFlowline/NHDFlowline.htm.

III. ENSEMBLE DEEP LEARNING FOR PREDICTING GAUGE HEIGHT AT UNMONITORED RIVER LOCATIONS

Bhanu Kanwar[1], Samuel Vanfossan[1], Jake Hale[1], Steven Corns[1] and Suzanna Long[1]

[1]Department of Engineering Management and Systems Engineering,
Missouri University of Science and Technology, Rolla, MO 65409

ABSTRACT

Flooding is a common occurrence in Missouri and poses a serious threat in susceptible areas of the state. To anticipate flooding, gauges have been installed along rivers to monitor the water level, but there are many catchment areas in Missouri that are not serviced by these gauges, leaving a large portion of the state that is susceptible to flooding unmonitored. This research presents an ensemble of Long Short-Term Memory based deep learning models coupled with a clustering algorithm to predict river water levels at these unmonitored locations. The implementation of ensemble learning with deep learning models improves the forecasting performance of the models compared to a standalone deep learning prediction model. The proposed models predict the river water levels 4 hours ahead in the future with locations having a correlation greater than 0.9. This provides emergency management personnel with a manifold increase in flood information, allowing greater accuracy in flood alerts so that safety measures can be implemented efficiently and effectively to improve public safety.

Keywords: Flooding, LSTM, Ensemble Learning, Deep Learning.

1. INTRODUCTION AND LITERATURE REVIEW

This research presents algorithms and a methodology to predict water levels in rivers using virtual gauges to provide gauge information at unmonitored locations. This addresses a lack of reliable river water level information at such locations to determine the timing and extent of flood events in these areas.

Several research studies have been conducted in the field of forecasting flood events at susceptible locations. Many of these studies focus on the task of predicting river water levels at locations where gauges have been installed to monitor any variation in the water levels. A review of the literature related to research areas such as river water level monitoring, flood predictions, and machine learning-based prediction modeling is given here. It should be noted that there is a lack of available methodologies in the reviewed literature that utilizes deep learning models to predict river water levels and forecast flooding events at unmonitored locations.

Computational intelligence models such as neural networks and genetic algorithms have been implemented by researchers to study the susceptibility of floods in different flood-prone areas. An artificial neural network (ANN) based model was developed to create a flood water level prediction model for the River Nile in Sudan (Elsafi, 2014). A hybrid autoencoder-multiplayer perceptron model was implemented on a multivariate dataset to perform flood susceptibility mapping in the regions of Iran and India (Ahmadlou et al., 2020). A Long Short-Term Memory (LSTM) based neural network has been used to generate hourly runoff water level values and forecast floods in the Russian river basin, California, USA (Han et al., 2021). Convolutional neural

networks (CNN) have also been used in tandem with recurrent neural networks (RNN) to develop flood susceptibility maps based on historical flood and geospatial features information of Northern Iran (Panahi et al., 2021). Genetic algorithms have been used to optimize the deep belief networks and forecast flash floods in the flood-sensitive regions of Iran (Shahabi et al., 2021). One-dimensional (1D) and two-dimensional (2D) hydrologic flow computational models were implemented by a team of researchers to estimate the flood travel time and affected areas in Ohio, USA (Ghimire et al., 2020). Neural networks-based regression models have also been tested on multivariate datasets with features such as groundwater levels, depth, average wind speed, tides, etc. to forecast flood activities in the Mohawk River, New York (Tsakiri et al., 2018). A Geographic Information Systems (GIS) simulator was also developed to predict both rainfall-runoff values and seasonal flooding activity (Chiari et al., 2000). Other tools are available through government agencies, such as River Analysis System (HEC-RAS) to develop inundation maps of areas in the United States vulnerable to floods (US Army Corps of Engineers, 2021).

Research has been conducted to develop flood forecasting models for areas that lack adequate gauge networks to monitor water levels of rivers and streams. Machine learning cross-validation techniques such as k-fold cross-validation has been used to improve a neural network model that accurately captures the rainfall-runoff patterns at catchment levels in California which was then compared to the Sacramento Soil Moisture Accounting (SAC-SMA) model used for a similar task (Kratzert et al., 2019). GIS software tools such as QGIS have been used to study the Hydrological Response Unit (HRU) images of the Western Black Sea Region of Turkey and identify similarities

between 33 different catchments in the area (Aytaç, 2020). In Iran, hydrologic similarities between different watershed basins were analyzed to develop a rainfall-runoff model that forecasts streamflow values for the ungauged Karkheh River Basin (Choubin et al, 2019). Fuzzy C-Means and k-Nearest Neighbor-based machine learning classification models were used to classify the drainage basins and estimate the future streamflow values in the ungauged basins (Papageorgaki & Nalbantis, 2016).

River water level prediction relies on sequential data as well as spatial data values. LSTMs are suitable for capturing interdependencies between the sequential input time series-based dataset (Laddad, 2019). Also, the ensemble learning approach assists in minimizing the variance of the combined model outputs to generate better results as compared to a standalone LSTM-based model (Brownlee, 2019). An analysis of several research publications shows that most of the flood forecasting-related research work has been conducted for the areas where gauges have been installed to collect river water level height values. However, researchers have also started to take an interest in developing flood forecasting models for areas with inadequate gauge networks. To develop an accurate model a large amount of data must be gathered and analyzed.

2. METHODOLOGY

There are several factors leading to changes in water level at unmonitored locations that must be considered. These include rainfall, soil moisture, and the upstream/downstream gauge height levels. The effect of upstream and downstream gauges may vary depending on distance from the prediction site and must be considered

when creating predictive models. These data form a multivariate dataset that is then used as an input to an ensemble of LSTM deep learning models to predict river water levels as shown in Figure 1.

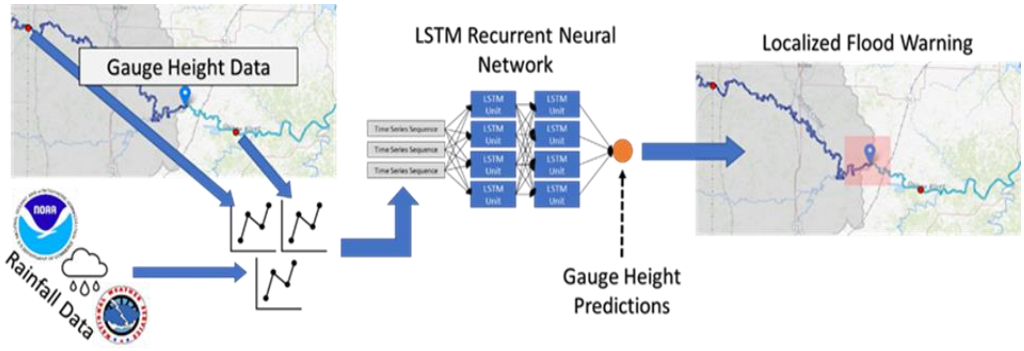

Figure 1. Model framework.

A total of 20 different locations are selected to analyze its data variables and develop the prediction model for this project. Each location consists of a gauge group that includes the virtual gauge or gauge of interest for that location, an existing gauge upstream of the location, and an existing gauge downstream of the location. These gauge groups are assigned to different clusters to train four specialized models that can predict virtual gauge heights for distinct clustering scenarios.

2.1. MODEL WORKFLOW

Figure 2 displays the four steps that make up the workflow of the model developed for this project. The first step gathers available data describing the catchment under consideration. The second establishes the distance between the virtual gauge and the upstream/downstream gauges. The third step is to gather the gauge and rainfall data,

and the fourth step uses the provided data to cluster the gauge groups and analyze the data to create the predictive model.

Figure 2. Model workflow.

2.1.1. Develop Catchment Database. The USGS National Map's data download application is used to download the 1-meter Digital Elevation Model (DEM) files that contain the topographic and geographic information for different parts of Missouri (The National Map, 2022, USGS National Map, 2022). The downloaded 1-meter DEM files are generated from the geospatial information collected using the light detection and ranging (Lidar) data source and provide high-resolution geographic information for a selected area. Unfortunately, the 1-meter DEM coverage is not currently available for all areas of Missouri as shown in Figure 3.

In addition to the 1-meter DEM data, candidate areas must also have gauge information available upstream and downstream of the desired virtual gauge. The gauge information from the USGS National Water Dashboard's database is analyzed to determine the location of gauges installed to collect river water levels information (USGS, 2022). Catchment areas with 1-meter DEM coverage and where the gauges have been installed for data collection purposes are selected for further study. The 1-meter DEM files for the selected catchment areas are uploaded on the ArcGIS Pro software for visualization purposes.

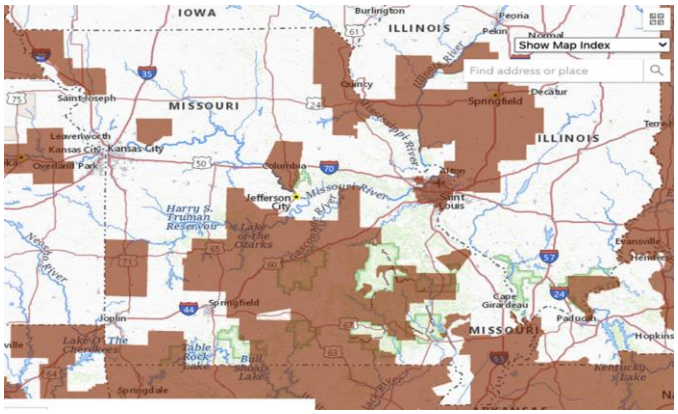

Figure 3. 1-meter DEM coverage for Missouri.

ArcGIS Pro is used to process the DEM files and extract the information related to the gauged catchment's geospatial variables such as elevation, slope, area, perimeter, etc. The locations of the gauged catchments are represented by purple-colored polygons as shown in Figure 4. After reviewing the gauge height and rainfall data for the 50 catchments with 1-meter DEM data, 21 catchments are selected for analysis as the gauge height and rainfall data are not available for the other 29 catchments. Another location was rejected due to limited availability of data on an upstream gauge located on the St. Francis River.

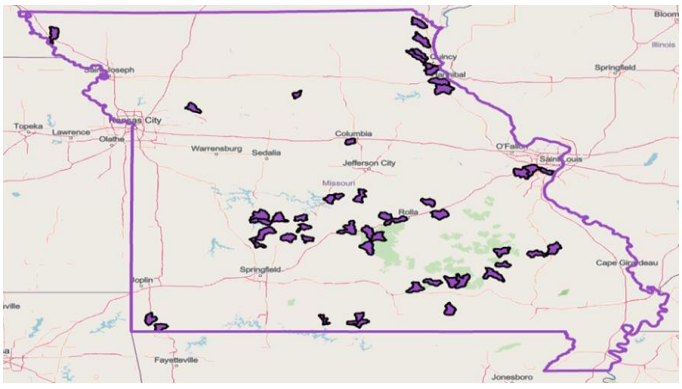

Figure 4. Gauged catchments in Missouri.

2.1.2. Calculate Flowline Distances. The flowline distances between the actively monitored upstream and downstream gauges located in the selected catchments are drawn in the ArcGIS Pro software to calculate the polyline distances between the gauges. A total of 40 flow lines are considered to calculate the polyline distances from the virtual gauge to the upstream and downstream gauges in the selected catchments. ArcGIS Pro is also used to calculate and compile the geodesic distances between the gauges in a tabular format as shown in Table 1. The geodesic distances between the gauges are calculated in the kilometer units format using ArcGIS Pro's 'Calculate Geometry' tool. The purpose of using the geodesic distances is to capture the curvature along the surface of the earth.

Table 1. Distances between virtual gauge and the upstream/downstream gauges.

ID	Catchment	Upstream Distance (km)	Downstream Distance (km)
1	BagnellDam	155.28	21.95
2	BayCreek_JacksFork	39.79	10.54
3	BrushCreek_FoxRiver	101.98	91.38
4	CraneCreek_PommedeTerreRiver	48.27	13.66
5	DemocratRidge_BigPiney	20.51	48.58
6	DuncanCreek_GasconadeRiver	126.75	87.49
7	GumCreek_OsageRiver	22.58	50.85
8	HamiltonCreek_MeramecRiver	24.86	20.97
9	JacksCreek_NianguaRiver	21.50	101.45
10	MillCreek_PommedeTerreRiver	13.89	75.68

Table 1. Distances between virtual gauge and the upstream/downstream gauges. (Continued)

11	OutletEastFork_BlackRiver	8.77	25.66
12	OutletJacksFork	10.52	23.91
13	PruettCreek_MeramecRiver	63.69	53.97
14	RockyCreek_CurrentRiver	69.99	43.86
15	SmithBranch_Roubidoux	47.09	66.95
16	SouthFabiusRiver_OutletFabiusRiver	85.42	36.92
17	SweetHollowCreek_NianguaRiver	63.46	21.23
18	TarkioRiver	94.56	36.91
19	TurkeyCreek_SaltRiver	29.49	162.86
20	ValleyPark	21.56	9.79

2.1.3. Download Gauge and Rainfall Data. The river water level values for the upstream and downstream gauges are gathered from the USGS National Water Information System's database (USGS National Water Information System, 2022). The data values for the upstream and downstream gauges are downloaded from September 1, 2016 till December 30, 2021. The data is in time-series format and is resampled at 30-minute time intervals. The daily rainfall data values for the catchments were downloaded from the National Weather Service (NWS) data archives for the period from September 1, 2016 to December 30, 2021. The rainfall data from September 1, 2016 - June 27, 2017 is available in point format while the data from June 28, 2017 - December 30, 2021 is available in the raster format (NWS, 2022). The rainfall data values are processed using

the ‘Clip’ tool in ArcGIS Pro to put them into a common format and resampled at 30-minute intervals. These data are then combined to generate a single multivariate dataset that is used as input to the deep learning-based models.

The data values are subjected to data processing techniques such as data cleaning, exploratory data analysis, data normalization, etc., to prepare the correctly formatted input values for the models. The dataset is divided into two sets for training and testing the models. 65% of the dataset is used for training the models, 15% is used for validation, and the remaining 20% is used for testing purposes. During the model testing phase, the testing dataset is used to evaluate the performance of the model. The predictions generated by the deep learning models are compared against the actual water level values at the virtual gauge to evaluate their accuracy.

2.1.4. Implement Deep Learning Models. The deep learning models are trained on the input dataset of Section 2.1.3, which consists of 118,021 observations. To increase the accuracy of the deep learning models, preprocessing of the virtual gauges by proximity to the upstream and downstream gauges was performed by clustering the gauge groups (Section 2.2). This resulted in four clusters, each consisting of 5 different gauge groups. Of the five gauge groups, four are used to train the ensemble deep learning models and make predictions about the group cluster. The fifth gauge group is reserved to test the generalizability of the model when applied to an unseen dataset.

An ensemble of 30 LSTM-based models takes in the input dataset to predict river levels at unmonitored central gauges of interest. A set of 30 LSTM models are trained for each cluster to generate average prediction values and prediction intervals for these values. The prediction interval for the ensemble’s average prediction can be derived from

the variance between the predicted values of the models within the ensemble. The developed models can be applied to a new dataset to predict gauge height values for novel gauge groups. Thus, the models from each cluster can better predict results for a scenario that would fit in a particular cluster.

2.2. GAUGE GROUP CLUSTERS

The 20-gauge groups selected earlier are clustered into four different groups based on the distance between the individual gauges lying within each gauge group. A gauge group consists of three different types of gauges such as an upstream gauge, a downstream gauge, and a virtual gauge or gauge of interest. Two different types of distances are determined for each gauge group, the distance between the virtual gauge and the upstream gauge and the distance between the virtual gauge and the downstream gauge. The median value for these distances is determined for all 20-gauge groups. The distance values for each gauge group are compared against the median values to allocate a tag specifying the general proximity of the individual gauges in a group. A ‘Close’ tag is assigned to a gauge group where the distance between the virtual gauge and the upstream gauge is less than the median value and a ‘Far’ tag is assigned if the distance is greater than the median value. This comparison is repeated for the distance between the virtual gauge and the downstream gauge to generate four clusters of gauge groups: ‘Close-Close’ (CC), ‘Close-Far’ (CF), ‘Far-Close’ (FC), and ‘Far-Far’ (FF) tags to each of the 20-gauge groups. The four different types of tags assigned to all gauge groups are shown in the graph in Figure 5. The feature values for gauge groups within each of the

four clusters are used to develop models that can predict the water levels at virtual gauges that lie in the same cluster.

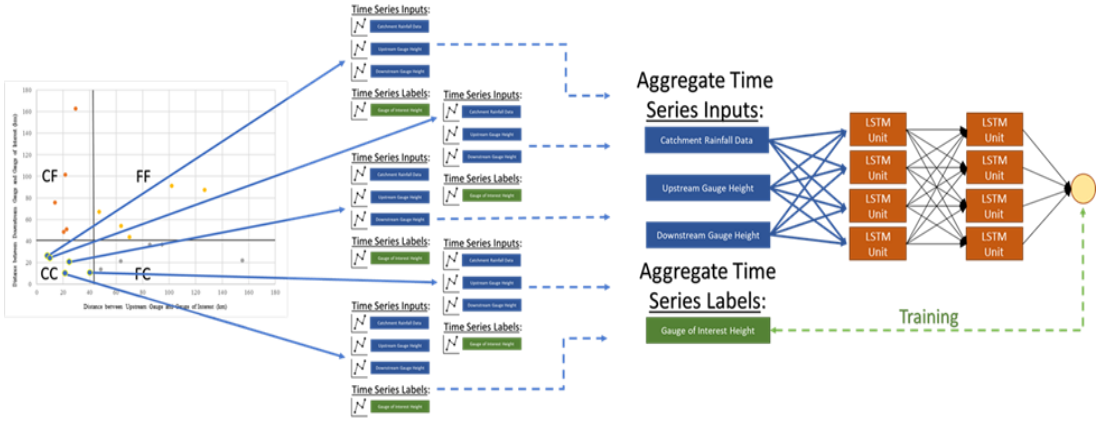

Figure 5. Gauge groups and LSTM deep learning models.

2.3. LSTM DEEP LEARNING MODEL ARCHITECTURE

The time series-based input dataset consists of values that are resampled at 30-minute time intervals. The data values from the previous 30 timesteps (or 15 hours) are used as inputs to the LSTM-based models to predict gauge height values that are 8 timesteps (or 4 hours) in the future. The architecture of the LSTM models consists of 8 different sequentially arranged layers. The input data values from the dataset are passed into the initial 50-unit layer of the network. The units of the input layer specify the dimension of outputs and the number of parameters in the LSTM layer (Tung, 2022). A dropout layer is also added to the architecture to randomly drop 20% of a layer's output neurons and prevent overfitting. An overfitted model produces high testing errors when implemented on the test dataset (Ampadu, 2021; IBM, 2021). Finally, a 1-unit 'Dense' layer is used to receive information from the preceding layers and predict a single-value

gauge height for the virtual gauge. When applied to novel scenarios, the outputs of the LSTM deep learning model predict the value for a virtual gauge lying between an upstream and downstream gauge.

3. RESULTS AND DISCUSSION

The ensemble of LSTM deep learning models is implemented to predict gauge height values using virtual gauges for unmonitored locations. The model outputs are compared against the actual readings for the gauge group clusters to calculate the performance metrics. The relationship between the predicted and true gauge height values for the CC gauge group is shown in Figure 6.

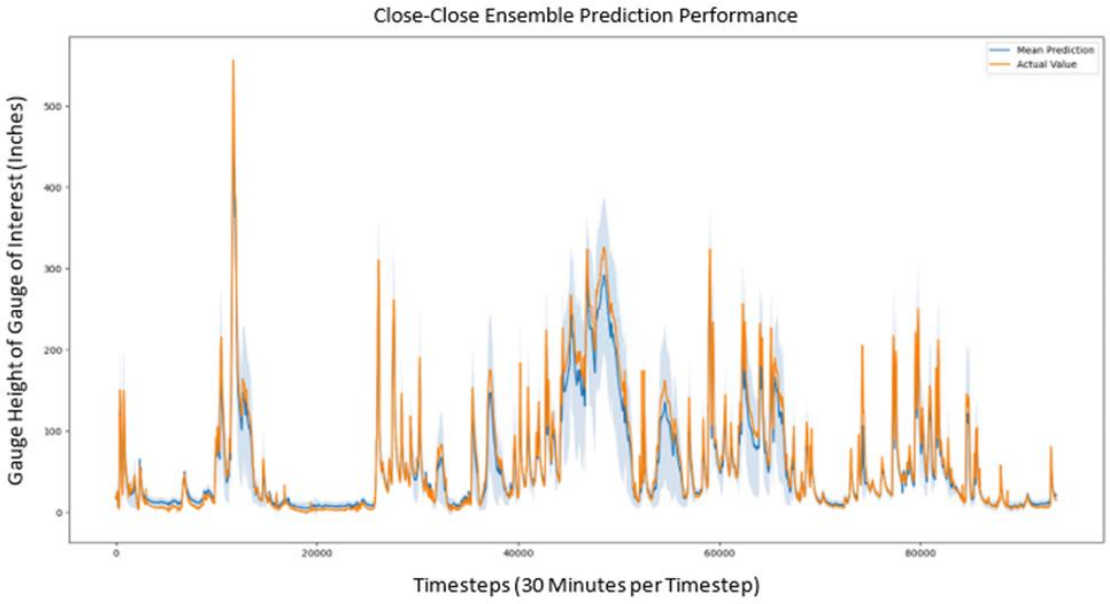

Figure 6. Comparison of predicted and true gauge height values for CC gauge group.

The predicted (virtual) gauge height values are in blue and the actual gauge height values are in orange as seen in Figure 6. The 95% prediction interval for the outputs of the trained CC ensemble is displayed in a light-blue colored shaded area. The trained CC ensemble model is applied to validation gauge group data for the CC cluster classification scenario. The performance metrics of the model are displayed in Table 2. The mean absolute deviation and median absolute deviation values for the true and predicted gauge height values are 8.683 and 5.119 inches, respectively. Both mean absolute deviation and median absolute deviation values exhibit a 1.56% and 0.92% deviation from the true values, respectively, when compared to the virtual gauge range. The ensemble model resulted in a correlation coefficient between the predicted values and the true values of 0.9948. The correlation coefficient value between the gauge height values obtained from the virtual gauge and upstream and downstream gauges is 0.8794. The implementation of the ensemble learning approach improves the correlation coefficient value by 0.1155 while increasing prediction accuracy. This signifies an improvement of 13.133% in the correlation coefficient values beyond the information available from upstream and downstream gauges directly.

Table 2. CC ensemble summary statistics.

Close-Close Gauge Grouping		
Gauge of Interest Spectrum	556.44 inches	
Performance Metric	Inches	Percent of Spectrum
Mean Absolute Deviation	8.683	1.56%
Median Absolute Deviation	5.119	0.92%

Table 2. CC ensemble summary statistics. (Continued)

Average 95% Prediction Interval Width	55.579	9.99%
Model Prediction to True Gauge Height Correlation Coefficient	0.9948	
Upstream-Downstream Gauge Height to True Gauge Height Correlation Coefficient	0.8794	
Correlation Coefficient Improvement Attributable to Model	0.1155	
Percent Correlation Coefficient Improvement Attributable to Model	13.133%	

Figure 7 displays the relationship between the predicted and true gauge height values for the CF gauge group. The blue-colored curve for the predicted values traces the orange-colored curve for the true values and manages to capture the underlying patterns of the water level readings at the virtual gauge. The CF ensemble does not demonstrate as much variability as the CC ensemble approach but can still display a high level of accuracy. The mean absolute deviation and median absolute deviation values for the predicted and true values for the CF ensemble are 4.624 and 3.679 inches, respectively as shown in Table 3. These values demonstrate a 1.49% and 1.19% deviation from the true gauge height values with respect to the virtual gauge range of 309.36 inches for the CF ensemble. The predictive performance of this ensemble is lower than the CC cluster in terms of absolute deviation values, the percentage of the virtual gauge's range covered by the deviation metrics is similar.

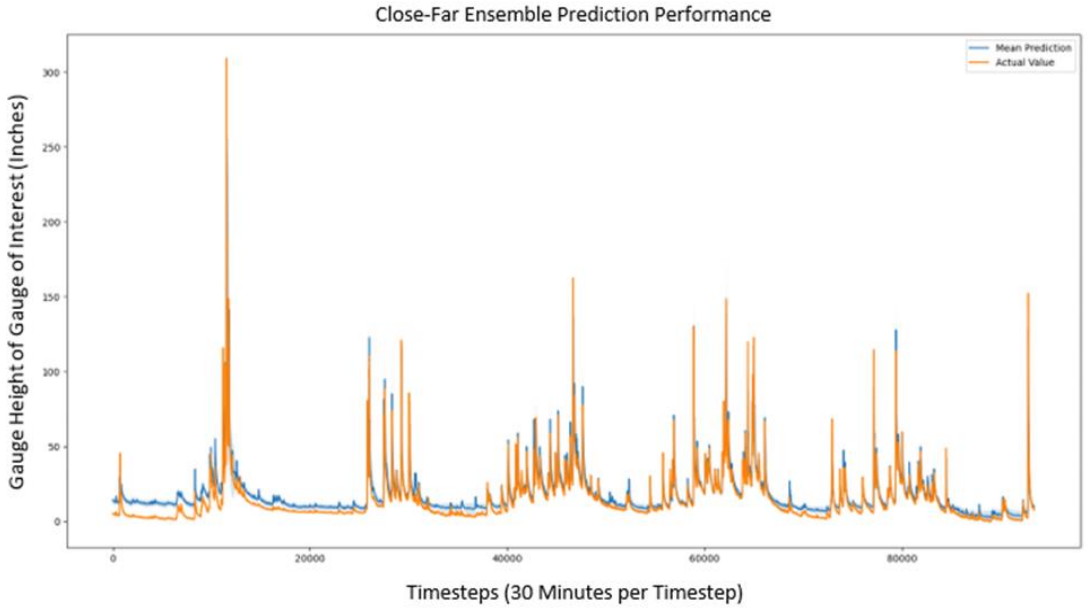

Figure 7. Comparison of predicted and true gauge height values for CF gauge group.

The width of the average 95% prediction interval is 6.704 which represents the 2.17% width of the virtual gauge range value of 309.36 inches as shown in Table 3. In this case, the implementation of the ensemble learning approach improves the correlation coefficient value between the predicted and true gauge height values by 7.694%.

Table 3. CF ensemble summary statistics.

Close-Far Gauge Grouping		
Gauge of Interest Spectrum	309.36 inches	
Performance Metric	Inches	Percent of Spectrum
Mean Absolute Deviation	4.624	1.49%
Median Absolute Deviation	3.679	1.19%
Average 95% Prediction Interval Width	6.704	2.17%

Table 3. CF ensemble summary statistics. (Continued)

Model Prediction to True Gauge Height Correlation Coefficient	0.9441
Upstream-Downstream Gauge Height to True Gauge Height Correlation Coefficient	0.8766
Correlation Coefficient Improvement Attributable to Model	0.0674
Percent Correlation Coefficient Improvement Attributable to Model	7.694%

The relationship between the predicted and the true gauge height values for the FC gauge group used to test the generalizability of the trained FC ensemble is shown in Figure 8. The FC ensemble predicted values largely trace the true values and succeed in capturing the behavior of the virtual gauge for the ensemble as seen in Figure 8. The summary statistics values of different performance metrics of the FC ensemble are shown in Table 4. The metrics are generated after the implementation of the trained FC ensemble to the gauge group scenario that fits the FC cluster classification. Both mean and median absolute deviation values represent the 1.74% and 1.35% deviation from the true gauge height values respectively when compared against the virtual gauge spectrum value of 130.92 inches. The average 95% prediction interval width is 3.804 inches which is 2.91% of the virtual gauge spectrum. The ensemble modeling approach also improves the correlation between the model predictions and the true gauge height values by 3.605%.

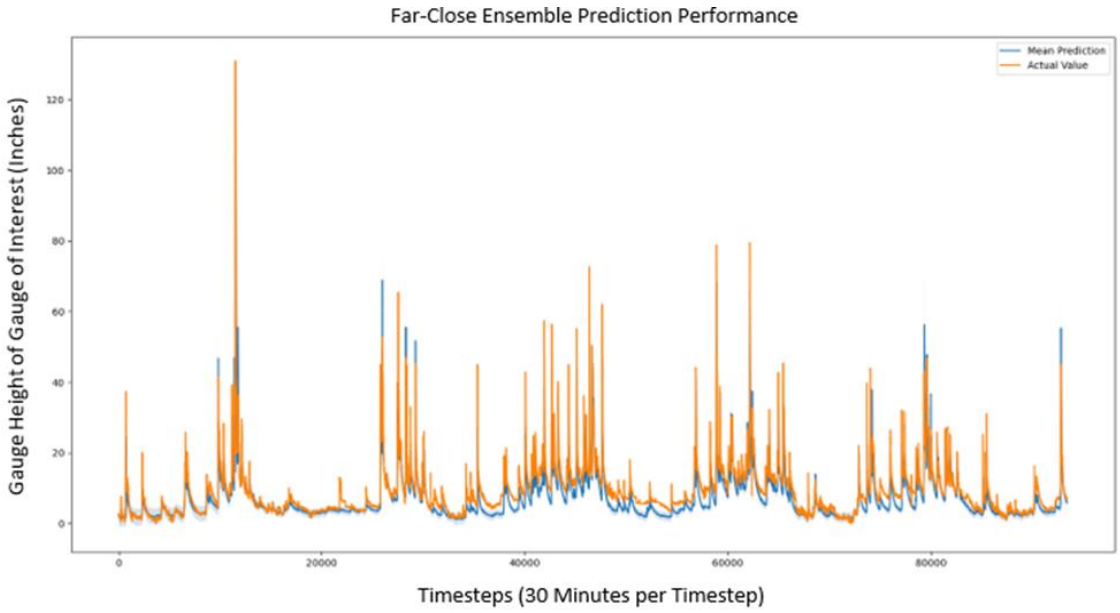

Figure 8. Comparison of predicted and true gauge height values for FC gauge group.

Table 4. FC ensemble summary statistics.

Far-Close Gauge Grouping		
Gauge of Interest Spectrum	130.92 inches	
Performance Metric	Inches	Percent of Spectrum
Mean Absolute Deviation	2.278	1.74%
Median Absolute Deviation	1.766	1.35%
Average 95% Prediction Interval Width	3.804	2.91%
Model Prediction to True Gauge Height Correlation Coefficient	0.9208	
Upstream-Downstream Gauge Height to True Gauge Height Correlation Coefficient	0.8888	
Correlation Coefficient Improvement Attributable to Model	0.0320	

Table 4. FC ensemble summary statistics. (Continued)

Percent Correlation Coefficient Improvement Attributable to Model	3.605%

Figure 9 shows the relationship between the predicted gauge height values and the true values for the FF gauge group. The FF ensemble predictions also trace the true gauge height values, highlighting the ability of the ensemble to efficiently generalize to an FF gauge group scenario.

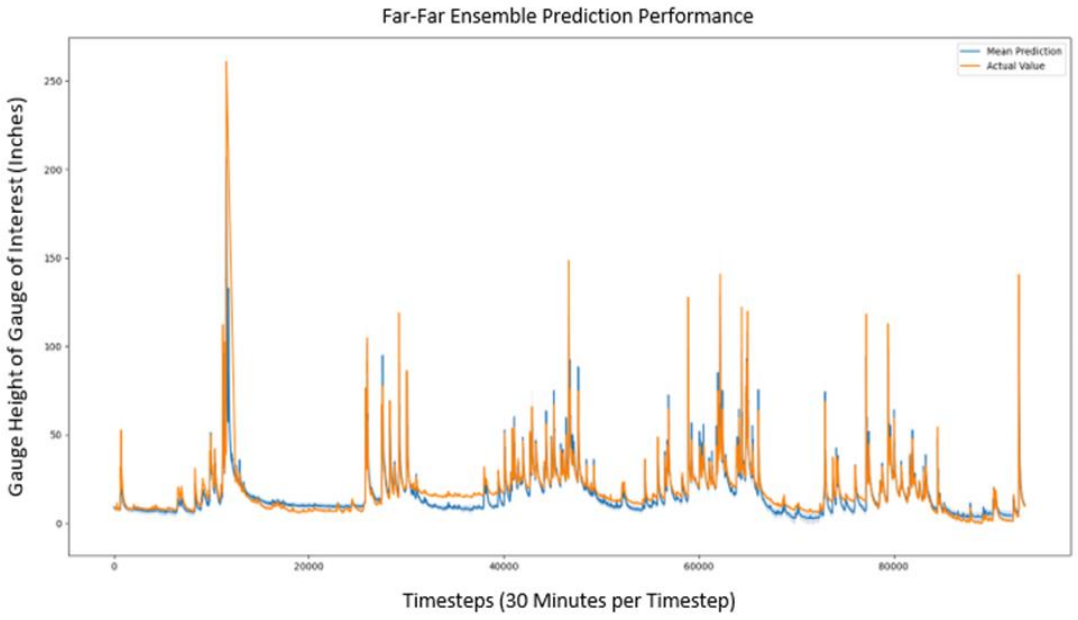

Figure 9. Comparison of predicted and true gauge height values for FF gauge group.

The. virtual gauge range for the FF group is 260.88 inches as seen in Table 5. The 6.337 inches wide average 95% prediction interval covers 2.43% of the total virtual gauge range. Both mean and median absolute deviation values for the gauge group are

3.994 and 2.738 respectively. The use of the ensemble learning approach for the gauge group improves the correlation between the predicted values and true gauge height values by 8.661% which is a statistically significant improvement from the upstream and downstream gauges directly.

Table 5. FF ensemble summary statistics.

Far-Far Gauge Grouping		
Gauge of Interest Spectrum	260.88 inches	
Performance Metric	Inches	Percent of Spectrum
Mean Absolute Deviation	3.944	1.51%
Median Absolute Deviation	2.738	1.05%
Average 95% Prediction Interval Width	6.337	2.43%
Model Prediction to True Gauge Height Correlation Coefficient	0.8351	
Upstream-Downstream Gauge Height to True Gauge Height Correlation Coefficient	0.7685	
Correlation Coefficient Improvement Attributable to Model	0.0666	
Percent Correlation Coefficient Improvement Attributable to Model	8.661%	

The four different ensemble models in this research are used to train the associated gauge groups clusters. The ensemble model has a mean absolute deviation of

4.882 inches across the four different clustering scenarios and generates accurate gauge height predictions for virtual gauges for the different clusters. The median absolute deviation of the gauge groups is 3.326 inches. The mean absolute deviation value for all four gauge groups is greater than the median absolute deviation. The ensemble models also achieve higher correlation coefficients between the predicted and true gauge height values. The correlation coefficient values for the unmonitored sites in the four clusters are 0.9948, 0.9441, 0.9208, and 0.8351 for CC, CF, FC, and FF respectively. The models produce more accurate prediction results beyond what is available through a simple correlation modeling approach. The models developed in this research predict correlation while making predictions at future timestep values. The ensemble learning-based models predict gauge height values that are 8 timesteps (or 4 hours) into the future while achieving better correlation coefficient values than traditional deep learning approaches.

4. CONCLUSIONS

Clustering and ensemble deep learning-based models are implemented in this project to create virtual gauges and predict water levels at unmonitored locations in Missouri. Using publicly available data, a total of 20 different virtual gauge locations were analyzed to predict water levels. The use of an ensemble of LSTM models gives a higher accuracy than using individual LSTM models. The dataset is efficiently processed by the LSTM-based neural networks in the ensemble to capture relationships between the features and produce generalized prediction models. Dividing the gauge groups into four

clusters characterized by the flowline distances allows the algorithms to take advantage of similarities between gauge groups to improve the predictions even more.

The accuracy of the prediction at most of the virtual gauge sites is about 5%, which is comparable to the physical gauges used to monitor the water level. The exception is the FF gauge group, but the small error associated with it makes the information from the virtual gauges very valuable for emergency management personnel to better understand flooding situations and plan responses. The FF gauge group has the lowest accuracy, but with an improved correlation coefficient of 8.661% attributable to the implemented methodology, it still provides enough information to determine if there is a probability of a road being overtopped or flood risk. As more gauges are installed these virtual gauges will only increase in accuracy. The data provided by these gauges allows state and federal emergency management personnel to send observers or inspection teams to determine the danger to the traveling public. By providing information in 30-minute intervals with predictions four hours in advance emergency management personnel will have more options on how to address flooding. Teams can be sent out to investigate before a flood event occurs, and roads can be closed before sections are overtopped, removing the temptation of drivers to proceed across a flooded road. Advance flood warnings can also be issued by the local government authorities to ensure the safety of the people. The local authorities can rely on the multistep prediction values to take necessary precautions to avert any dangers posed by flood events in a susceptible region.

5. FUTURE WORK

The developed models can be used in scenarios when the upstream and downstream gauge data and rainfall data are available for the catchment area where the virtual gauge is planned for implementation. The models will have to be modified accordingly if different subsets of inputs are to be used for its development. Additional gauge groups and data variables can be incorporated into inputs to implement models that can capture a wide range of scenarios and deliver high-quality results. If possible, granular rainfall data can be used instead of the resampled daily rainfall values to test a new deep learning algorithm and compare its performance with the model proposed in this research project.

REFERENCES

Ahmadlou, M., Al-Fugara, A., Al-Shabeeb, A. R., Arora, A., Al-Adamat, R., Pham, Q. B., Al-Ansari, N., Linh, N. T., & Sajedi, H. (2020). Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder Neural Networks. Journal of Flood Risk Management, 14(1). https://doi.org/10.1111/jfr3.12683

Ampadu, H. (2021, May 10). Dropout in deep learning. AI Pool. Retrieved February 22, 2022, from https://ai-pool.com/a/s/dropout-in-deep-learning

Aytaç, E. (2020). Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based K-means clustering, a demonstration on western Black Sea region of Turkey. International Soil and Water Conservation Research, 8(3), 321–331. https://doi.org/10.1016/j.iswcr.2020.05.002

Brownlee, J. (2019, August 6). Ensemble learning methods for deep learning neural networks. Machine Learning Mastery. Retrieved February 22, 2022, from https://machinelearningmastery.com/ensemble-methods-for-deep-learning-neural-networks/

Chiari, F., Delhom, M., Santucci, J. F., & Filippi, J. B. (2000). Prediction of the hydrologic behavior of a watershed using artificial neural networks and geographic information systems. SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. 'Cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions' (Cat. No.00CH37166). https://doi.org/10.1109/icsmc.2000.885021

Choubin, B., Solaimani, K., Rezanezhad, F., Habibnejad Roshan, M., Malekian, A., & Shamshirband, S. (2019). Streamflow regionalization using a similarity approach in ungauged basins: Application of the Geo-environmental signatures in the Karkheh River Basin, Iran. CATENA, 182, 104128. https://doi.org/10.1016/j.catena.2019.104128

Elsafi, S. H. (2014). Artificial Neural Networks (Anns) for flood forecasting at Dongola station in the River Nile, Sudan. Alexandria Engineering Journal, 53(3), 655–662. https://doi.org/10.1016/j.aej.2014.06.010

esri. (Last Accessed February 19, 2022). Geodesic versus planar distance. Geodesic versus planar distance-ArcGIS Pro | Documentation. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/geodesic-versus-planar-distance.htm

Ghimire, E., Sharma, S., & Lamichhane, N. (2020). Evaluation of one-dimensional and two-dimensional HEC-ras models to predict flood travel time and inundation area for flood warning system. ISH Journal of Hydraulic Engineering, 28(1), 110–126. https://doi.org/10.1080/09715010.2020.1824621

Hailegeorgis, T. T., & Alfredsen, K. (2017). Regional Flood Frequency Analysis and prediction in ungauged basins including estimation of major uncertainties for mid-norway. Journal of Hydrology: Regional Studies, 9, 104–126. https://doi.org/10.1016/j.ejrh.2016.11.004

Han, H., Choi, C., Jung, J., & Kim, H. S. (2021). Deep learning with long short-term memory based sequence-to-sequence model for rainfall-runoff simulation. Water, 13(4), 437. https://doi.org/10.3390/w13040437

IBM. (Last Accessed November 1, 2021). What is Overfitting? Overfitting. Retrieved November 1, 2021, from https://www.ibm.com/cloud/learn/overfitting.

Kratzert, F., Klotz, D., herrnegger, mathew, Sampson, A. K., Hochreiter, S., & Nearing, G. (2019). Towards improved predictions in ungauged basins: Exploiting the power of Machine Learning. https://doi.org/10.31223/osf.io/4rysp

laddad, A. (2019, March 25). Basic understanding of LSTM. Medium. Retrieved February 23, 2022, from https://blog.goodaudience.com/basic-understanding-of-lstm-539f3b013f1e

National Hydrography. Access National Hydrography Products | U.S. Geological Survey. (2022). Retrieved February 18, 2022, from https://www.usgs.gov/national-hydrography/access-national-hydrography-products

NWS. (Last Accessed February 21, 2022). NCEP Stage IV Daily Accumulations (06/28/2017 – 12/30/2021). https://water.weather.gov/precip/downloads/

NWS. (Last Accessed February 21, 2022). Stage III Daily Accumulations (09/01/2016 – 06/27/2017). https://water.weather.gov/precip/archive/

Panahi, M., Jaafari, A., Shirzadi, A., Shahabi, H., Rahmati, O., Omidvar, E., Lee, S., & Bui, D. T. (2021). Deep Learning Neural Networks for spatially explicit prediction of flash flood probability. Geoscience Frontiers, 12(3), 101076. https://doi.org/10.1016/j.gsf.2020.09.007

Papageorgaki, I., & Nalbantis, I. (2016). Classification of drainage basins based on readily available information. Water Resources Management, 30(15), 5559–5574. https://doi.org/10.1007/s11269-016-1410-y

Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B. T., Mansouripour, F., Geertsema, M., Clague, J. J., & Bui, D. T. (2021). Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. Geoscience Frontiers, 12(3), 101100. https://doi.org/10.1016/j.gsf.2020.10.007

The National Map. (Last Accessed February 15, 2022). The National Map - Data Delivery. The National Map - Data Delivery | U.S. Geological Survey. https://www.usgs.gov/the-national-map-data-delivery

Tsakiri, K., Marsellos, A., & Kapetanakis, S. (2018). Artificial neural network and multiple linear regression for flood prediction in Mohawk River, New York. Water, 10(9), 1158. https://doi.org/10.3390/w10091158

Tung, L. K. (Last Accessed February 22, 2022). Units in LSTM. Retrieved February 22, 2022, from https://tung2389.github.io/coding-note/unitslstm

US Army Corps of Engineers. (Last Accessed November 1, 2021). HEC-RAS. Retrieved November 1, 2021, from https://www.hec.usace.army.mil/software/hec-ras/

USGS National Map. TNM Download V2. (Last Accessed February 21, 2022). https://apps.nationalmap.gov/downloader/#/

USGS National Water Information System. (Last Accessed February 21, 2022). USGS Water Data for the Nation. https://waterdata.usgs.gov/nwis

USGS. (Last Accessed February 17, 2022). National Water Dashboard. USGS. https://dashboard.waterdata.usgs.gov/app/nwd/?region=lower48&aoi=default

SECTION

2. CONCLUSIONS AND RECOMMENDATIONS

This dissertation examines the use of machine learning models to assist first responders and emergency management personnel in preparing robust flood and flash flood management plans before these events occur. The models developed use time series-based datasets that contain feature values for distinct sets of geospatial and precipitation variables. The deep learning models capture the temporal dependencies between the sequential feature values to make predictions for future timesteps. The LSTM models in paper 3 utilize the temporal input feature values to accurately predict the river water levels at unmonitored locations in flood-prone catchment areas. Similarly, the deep learning models developed for papers 1 and 2 can assist stakeholders in classifying locations in Missouri where flash flooding is likely to occur.

The first research contribution proposed a deep learning-based neural network to predict flash flood locations, demonstrating it has a classification accuracy of 85.23. The model accurately distinguishes between two sets of flash flood and non-flash flood events in Greene County, Missouri. The output values of the model represent the flash flood probability values for a road segment which, when combined with its restoration costs signifies a risk value for the location. City planners can use these values to prioritize the allocation of resources needed to restore roads damaged by flash flood activities. The risk information can also be used to determine the required contingency reserves to restore damaged critical infrastructure elements.

A deep learning-based neural network is compared to other machine learning methods in the second research contribution to identify which is best for predicting flash flooding events in Missouri. The deep learning neural network not only has a high classification accuracy, but also a higher precision and AUROC score, highlighting its superior performance in identifying flash flood events. The model's architecture consists of dropout layers to prevent overfitting and, as a result, it can generalize well to the unseen binary test dataset. The information gathered from implementing the model for a flash flood-prone area can aid stakeholders in developing better flood management protocols for the safety of the area's residents. Because it is generalizable it can be implemented in many areas with sufficient retraining.

The prediction of flood events at unmonitored locations is addressed in the third research contribution to this dissertation. Publicly available datasets are used as inputs to an ensemble of LSTM models to ultimately predict river water levels at unmonitored locations that may experience flooding. The highly accurate multistep future values are obtained by applying clustering as a preprocessor of the data and then using ensemble learning to determine gauge height. Emergency managers can use the results to take precautionary steps needed in the event of flooding without installing physical gauges for monitoring water levels at unmonitored locations. This approach also provides necessary data with negligible resources to inform decision making that ensures the safety of the people residing in these locations.

Future research work can include the task of further testing the performances of the developed models by incorporating more varied features such as normalized difference vegetation index (NDVI), granular rainfall data, etc. into the input datasets.

The efficacy of the models can also be tested by implementing them in diverse sets of locations all over the country. However, such research exercises might require modifying the model architecture to match the new scenarios. Researchers can also develop programming tools such as widgets and scripts that can simplify the time-intensive task of collecting crucial input datasets required for developing the models. Since floods and flash floods typically cause economic losses, the key stakeholders can also sponsor research studies that investigate both direct and indirect costs associated with these events.

APPENDIX A.

COMPARISON OF PREDICTED AND TRUE GAUGE HEIGHT VALUES FOR CC GAUGE GROUP

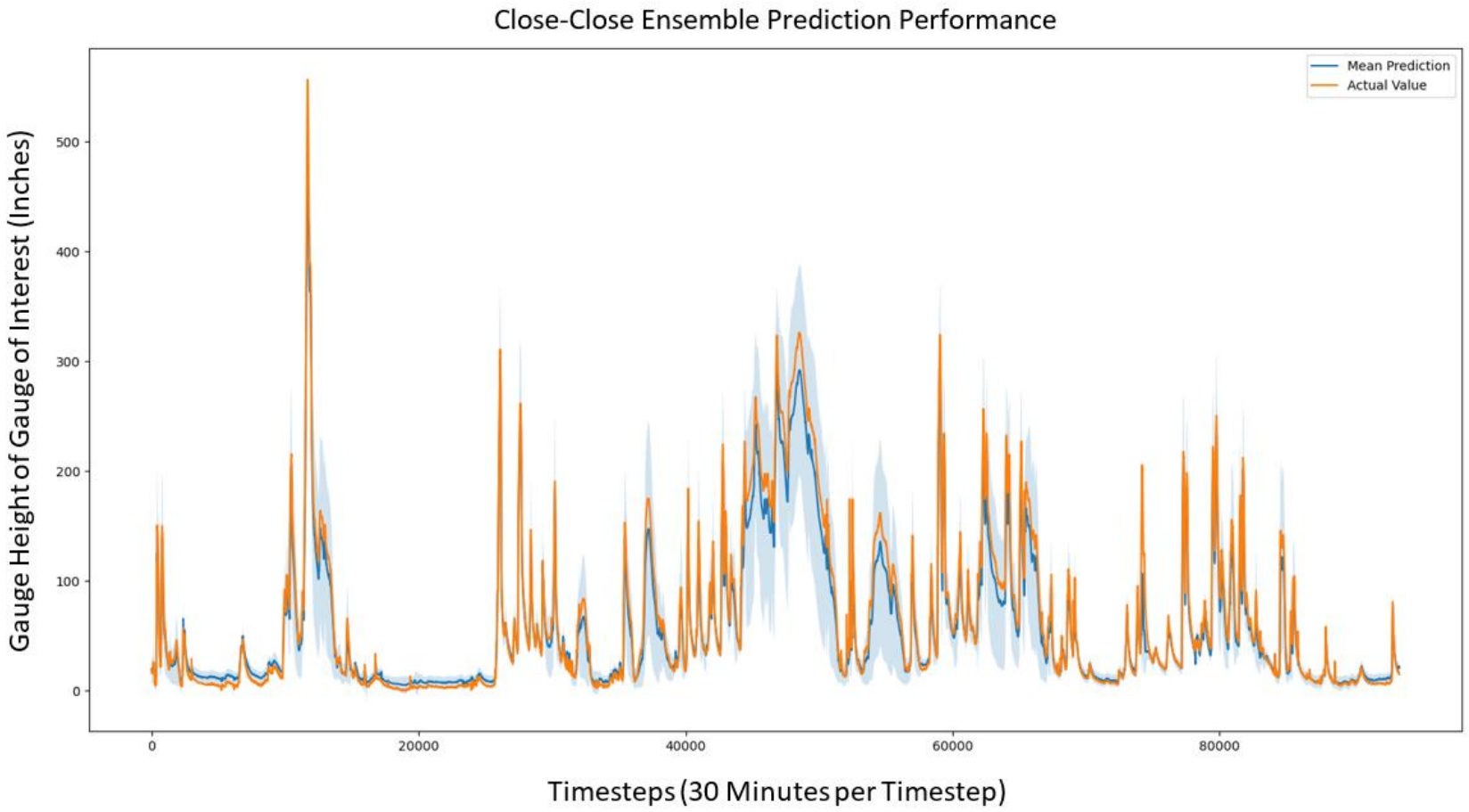

Figure A. Comparison of predicted and true gauge height values for CC gauge group.

APPENDIX B.

COMPARISON OF PREDICTED AND TRUE GAUGE HEIGHT VALUES FOR CF GAUGE GROUP

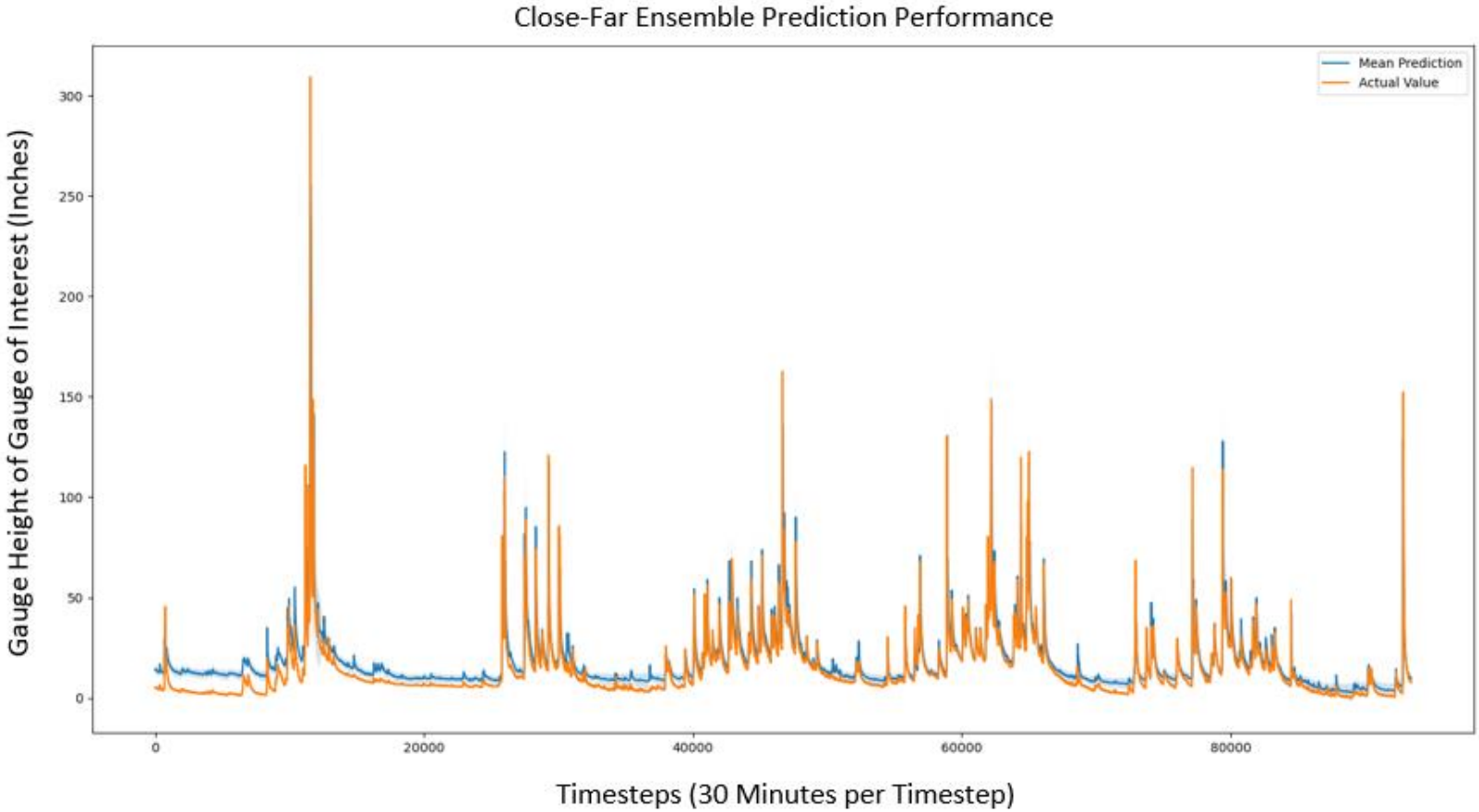

Figure B. Comparison of predicted and true gauge height values for CF gauge group.

APPENDIX C.

COMPARISON OF PREDICTED AND TRUE GAUGE HEIGHT VALUES FOR FC GAUGE GROUP

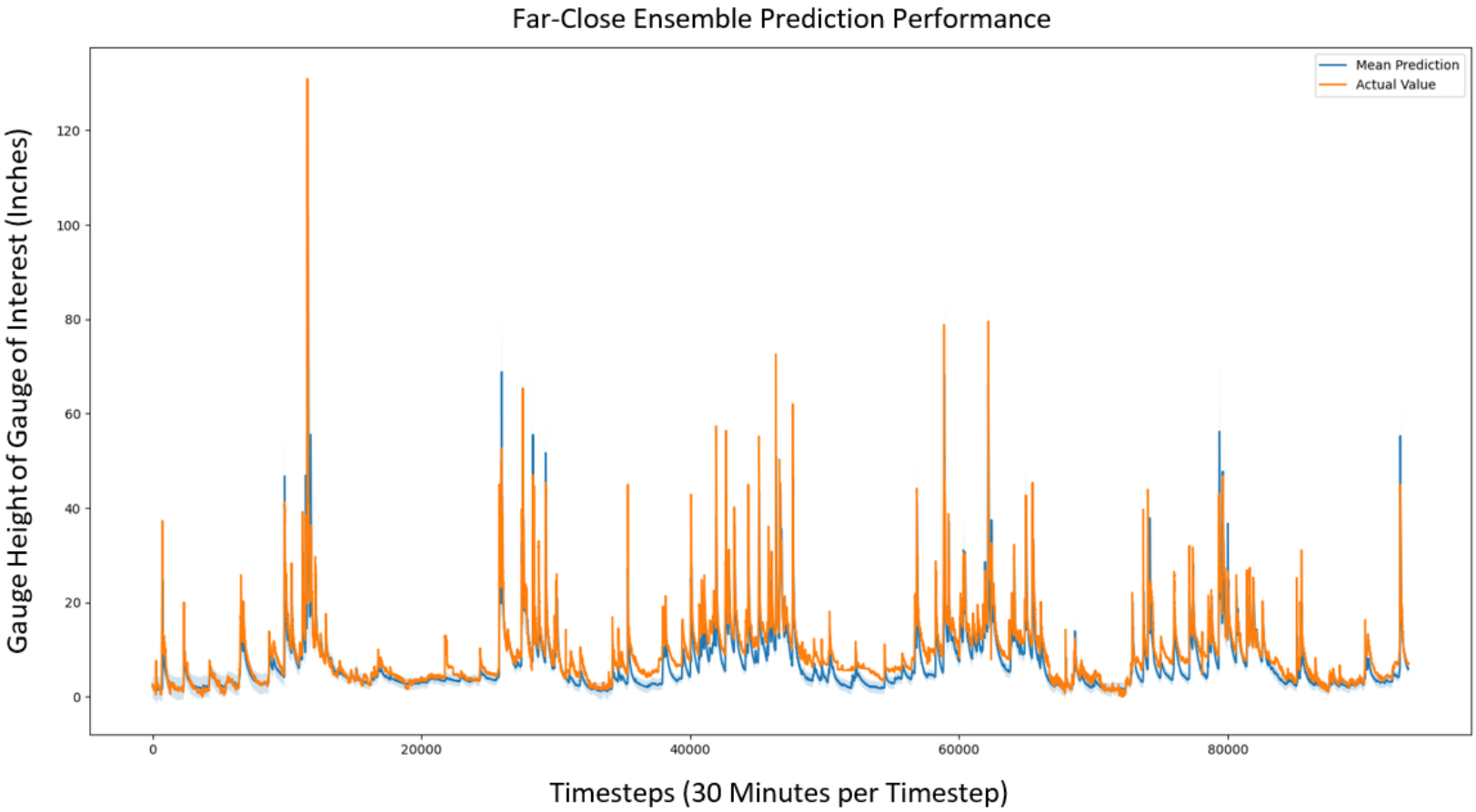

Figure C. Comparison of predicted and true gauge height values for FC gauge group.

APPENDIX D.

COMPARISON OF PREDICTED AND TRUE GAUGE HEIGHT VALUES FOR FF GAUGE GROUP

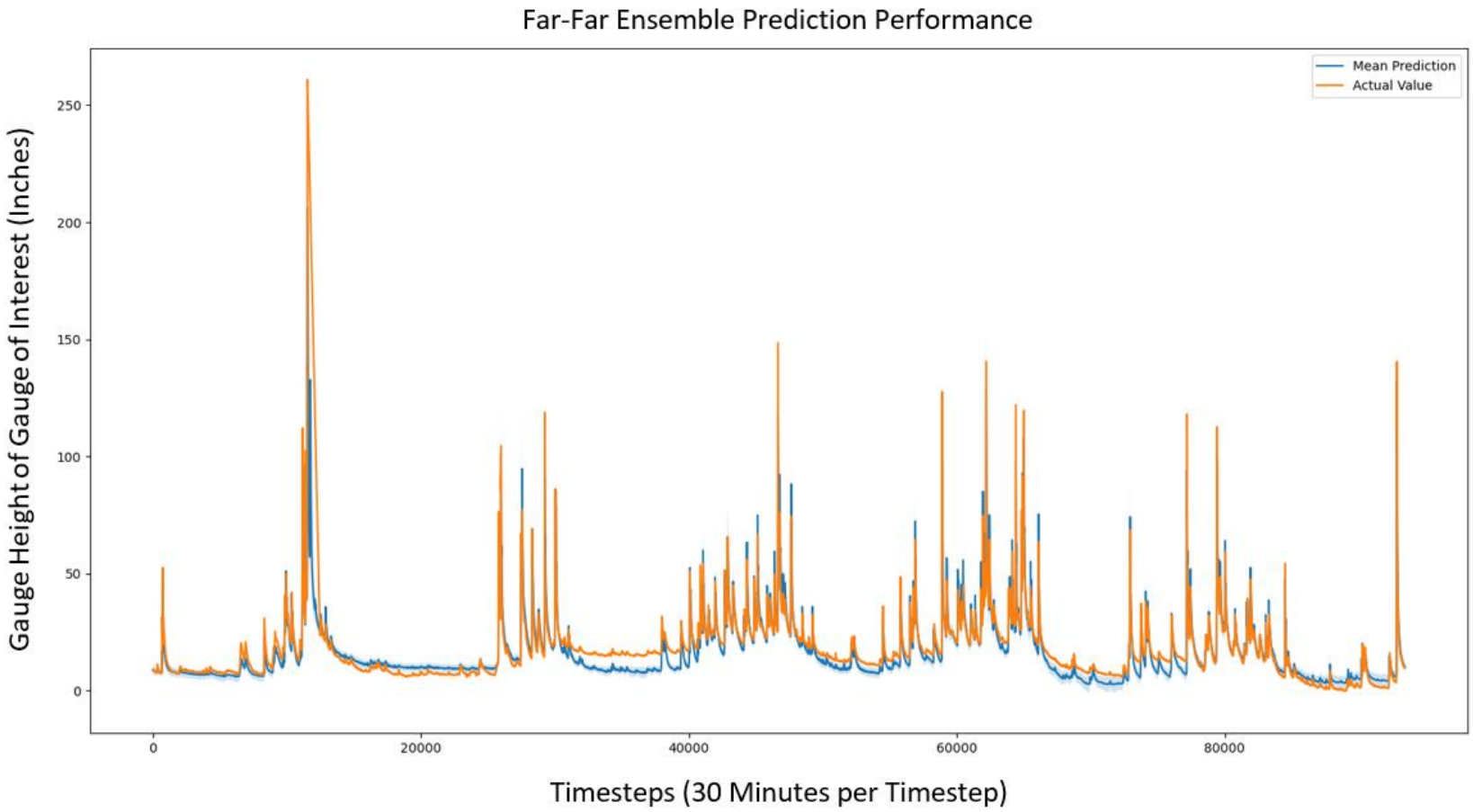

Figure D. Comparison of predicted and true gauge height values for FF gauge group.

REFERENCES

Teton County. (Last Accessed April 20, 2022). Impacts of a flash flood. Impacts of a Flash Flood | Teton County, WY. https://www.tetoncountywy.gov/412/Impacts-of-a-Flash-Flood

VITA

Bhanu Partap Singh Kanwar was born in Ludhiana, India. He finished his schooling from Ludhiana and received a degree of Bachelor of Technology in Mechanical Engineering from Punjab Technical University in May 2008. He also received a master's degree in Mechanical Engineering from Missouri University of Science and Technology (Missouri S&T), Missouri in May 2015. He began his Ph.D. program in Engineering Management at Missouri S&T in August 2016. During his Ph.D. program, he worked on projects in the fields of computational intelligence, machine learning, deep learning, and data analytics to solve real-world problems. He was a member of The American Association of Geographers (AAG), The American Society for Engineering Management (ASEM), and The Institute of Industrial and Systems Engineers (IISE). Bhanu held the position of Graduate Research Assistant under Dr. Steven Corns where he developed machine learning-based flood prediction models using real-world datasets. He was also an instructor for graduate-level courses at Missouri S&T. Bhanu received his Ph.D. in Engineering Management from Missouri S&T in July 2022. His research interests included data science and machine learning.