MID-AMERICA TRANSPORTATION CENTER

Report # MATC-MS&T: 138-3

Final Report

WBS: 25-1121-0005-138-3

Deep Learning for Unmonitored Water Level Prediction and Risk Assessment

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

Dr. Suzanna Long, PhD
Professor and Department Chair

Samuel Vanfossan
Graduate Research Assistant

Dr. Jacob Hale, PhD
Graduate Research Assistant & Post Doctoral Researcher

Bhanu Kanwar
Graduate Research Assistant

2023

A Cooperative Research Project sponsored by U.S. Department of Transportation- Office of the Assistant Secretary for Research and Technology

MATC

Deep Learning for Unmonitored Water Level Prediction and Risk Assessment

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

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

Samuel Vanfossan
Graduate Research Assistant
Engineering Management and Systems
Engineering Department
Missouri University of Science and Technology

Dr. Jacob Hale, PhD
Graduate Research Assistant & Post Doctoral Researcher
Engineering Management and Systems
Engineering Department
Missouri University of Science and Technology

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

A Report on Research Sponsored by

Mid-America Transportation Center

University of Nebraska–Lincoln

March 2023

Technical Report Documentation Page

1. Report No. 25-1121-0005-138-3	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Deep Learning for Unmonitored Water Level Prediction and Risk Assessment		5. Report Date 30 March 2023
		6. Performing Organization Code
7. Author(s) Steven M. Corns, PhD ORCID: 0000-0002-3685-2892; Suzanna K. Long, PhD, PEM, F.ASEM, ORCID: 0000-0001-6589-5528; Jacob Hale; Bhanu Kanwar; Samuel Vanfossan		8. Performing Organization Report No. 25-1121-0005-138-3
9. Performing Organization Name and Address 223 Engineering Management 600 W. 14th St Rolla, MO 68409-0370		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 69A3551747107
12. Sponsoring Agency Name and Address Mid-America Transportation Center Prem S. Paul Research Center at Whittier School 2200 Vine St. Lincoln, NE 68583-0851 Office of the Assistant Secretary for Research and Technology 1200 New Jersey Ave., SE Washington, D.C. 20590		13. Type of Report and Period Covered January 2021-June 2023
		14. Sponsoring Agency Code MATC TRB RiP No. 91994-88
15. Supplementary Notes		

16. Abstract
In this research, deep learning models for the prediction of river gauge heights were developed using publicly available data for the unmonitored locations in Missouri. The geospatial and rainfall data for 20 different catchment areas of Missouri is used in tandem with the clustering and ensemble deep learning approaches to develop a high-performance deep learning model that efficiently captures the interdependencies between the time-series input data values. The models can accurately predict river water level values up to four hours ahead with a correlation of greater than 0.82 with most results having a correlation greater than 0.9. Using a data-based approach to develop a deep learning neural networks-based framework can assist the first responders in issuing timely and localized flood warnings for the safety of the general public. This methodology is applied to publicly available datasets obtained from The United States Geological Survey (USGS) and National Weather Service (NWS). The research project is funded by the Missouri Department of Transportation (MoDOT) and Mid-America Transportation Center (MATC).

17. Key Words Deep learning; flood management; evacuation routing and planning	18. Distribution Statement

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 34	22. Price

Table of Contents

List of Figures

List of Abbreviations (optional)

Artificial Neural Network (ANN)

Close-Close (CC)

Close-Far (CF)

Convolutional Neural Networks (CNN)

Deep Belief Networks (DBN)

Digital Elevation Model (DEM)

Far-Close (FC)

Far-Far (FF)

Genetic Algorithm (GA)

Geographic Information Systems (GIS)

Hydrological Response Unit (HRU)

Long Short Term Memory (LSTM)

Mid-America Transportation Center (MATC)

Missouri Department of Transportation (MoDOT)

Multiple Layer Perceptron (MLP)

National Weather Service (NWS)

One-dimensional (1D)

Recurrent Neural Networks (RNN)

River Analysis System (HEC-RAS)

Sacramento Soil Moisture Accounting (SAC-SMA)

Two-dimensional (2D)

United States Geological Survey (USGS)

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Abstract

In this research, deep learning models for the prediction of river gauge heights were developed using publicly available data for the unmonitored locations in Missouri. The geospatial and rainfall data for 20 different catchment areas in Missouri was used in tandem with the clustering and ensemble deep learning approaches to develop a high-performance deep learning model that efficiently captured the interdependencies between the time-series input data values. The models accurately predicted river water level values up to four hours ahead with a correlation of greater than 0.82 with most results having a correlation greater than 0.9. Using a data-based approach to develop a deep learning neural networks-based framework can assist the first responders in issuing timely and localized flood warnings for the safety of the general public. This methodology was applied to publicly available datasets obtained from the United States Geological Survey (USGS) and the National Weather Service (NWS). The research project was funded by the Missouri Department of Transportation (MoDOT) and Mid-America Transportation Center (MATC).

Executive Summary

This research project used geospatial, river water level, and rainfall feature datasets as inputs to an ensemble of deep learning neural network models to predict river water levels at unmonitored sites. Multiple gauges from 20 different catchment locations were grouped into four clusters to train an ensemble of Long Short Term Memory (LSTM) deep learning models. These four clusters represented the distance of the unmonitored site from existing gauges, labeled 'Close-Close' (CC), 'Close-Far' (CF), 'Far-Close' (FC), and 'Far-Far' (FF) respectively. Geospatial data and sequential time series data values representing rainfall and existing river gauge data were used as inputs to the models to make multistep predictions at unmonitored locations as shown in Figure E.1.

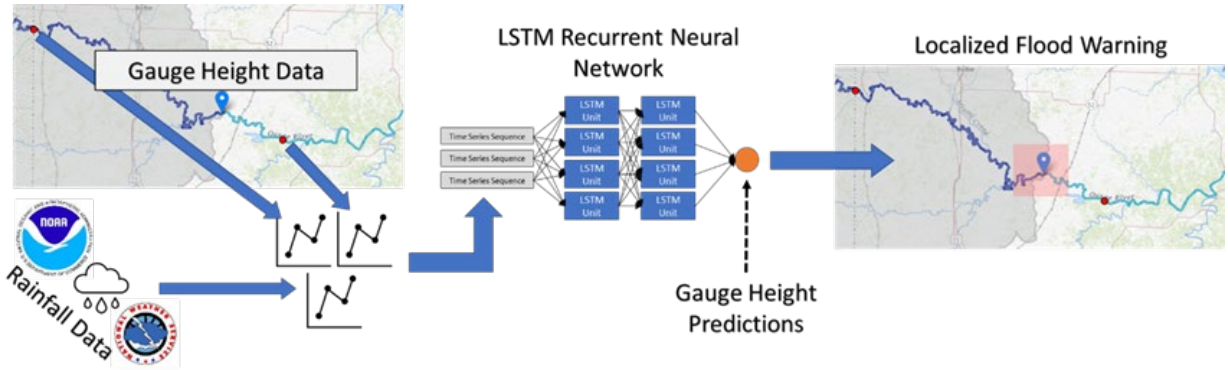

Figure E.1 Modeling Approach Overview

The use of ensemble learning methods to train multiple deep learning models increased the forecasting performance of the models as compared to a single deep learning model. Also, the application of the cluster-based deep learning ensemble model training methodology showcased a high correlation between the predicted model values and the true gauge height

values. The correlation coefficient values for the unmonitored sites in the four clusters are 0.9948, 0.9441, 0.9208, 0.8351 for ‘CC’, ‘CF’, ‘FC’, and ‘FF’ respectively. These accuracies are similar to current gauge height measurements, with sufficient detail to inform decision-makers of potentially hazardous flooding conditions. Unlike other modeling approaches, the models proposed in this study predicted correlation while making predictions at future timesteps. The deep learning-based model framework developed in this project is a novel methodology that did not exist before this project and provides significant assistance to the first responders in preparing for flooding events at an unmonitored site.

Chapter 1 Literature Review

Flooding events cause economic and personal damages to people living in the flood-prone areas of Missouri. Multiple studies have been conducted to calculate the penetration of floodwaters in proximity to gauged locations using mathematical and machine learning-based models. However, there is a lack of methodology in the literature that addresses the issue of calculating similar water levels at unmonitored locations. Various research studies in the fields of river gauge level monitoring, flood predictions, and neural networks-based prediction models were reviewed to develop a methodology to predict flash flood activity at unmonitored locations.

1.1 Flood Susceptibility Mapping and Machine Learning

Various modeling techniques have been applied by researchers to develop flood susceptibility maps using computational intelligence tools such as neural networks, genetic algorithms (GA), etc. The flow of the river Nile was forecasted by using an artificial neural network (ANN) based flood prediction model in flood-sensitive areas of Sudan (Elsafi, 2014). A hybrid autoencoder-multilayer perceptron model was developed for flood susceptibility mapping of the flood-prone areas in Iran and India (Ahmadlou et al., 2020). This hybrid model relied on multiple variables to deliver better results than a traditional multiple layer perceptron (MLP)-based model. A Long Short Term Memory (LSTM)-based neural network model was trained to calculate the river water levels in the Russian River Basin, California, USA (Han et al., 2021). The outputs of this hourly runoff LSTM forecasting model can be used to make short-term flood predictions for the selected area. Flood susceptibility maps were also developed using the convolutional neural networks (CNN) and recurrent neural networks (RNN) on historical flood information and various geospatial features (Panahi et al., 2021). Genetic Algorithm (GA) based optimization techniques have been adopted to optimize deep belief networks (DBN) and develop

novel methodology to predict flash flood susceptibility in the flash flood-prone regions of Iran (Shahabi et al., 2021). One-dimensional (1D) and two-dimensional (2D) hydrologic flow computations have been conducted to calculate flood travel time and inundated areas in the flood-prone regions of Ohio, USA (Ghimire et al., 2020). Researchers have also used diverse sets of variables such as groundwater levels, depth, average wind speed, tides, etc., to apply neural networks-based regression models to predict flooding events in Mohawk River, New York (Tsakiri et al., 2018). The results from a Geographic Information Systems (GIS) simulator were analyzed to develop an efficient rainfall-runoff model and predict the rise in flood water levels more accurately (Chiari et al., 2000). Hydrologic simulation software such as the River Analysis System (HEC-RAS) is often implemented by authorities in charge of implementing flood disaster mitigation plans to develop flood inundation maps of the risk-prone regions of the United States (US Army Corps of Engineers, 2021).

1.2 Ungauged Basins

Even though there is a lack of research efforts in the field of flood forecasting for the areas without any river gauge installations, a few researchers have developed methodologies to address this issue. A regional flood frequency analysis study was conducted using the method of L-moments and index flood to predict flood quantiles for a mid-Norway region with 26 unregulated catchments (Hailegeorgis & Alfredsen, 2017). Machine learning techniques such as k-fold validation were used to train the Long Short Term Memory (LSTM) networks instead of relying on the Sacramento Soil Moisture Accounting (SAC-SMA) model to accurately capture catchment-level rainfall-runoff behaviors (Kratzert et al., 2019). Hydrologic similarities between basins were studied to estimate streamflow values and develop a rainfall-runoff model for the ungauged Karkheh River Basin, Iran (Choubin et al, 2019). Geospatial features such as

Hydrological Response Unit (HRU) images were analyzed in QGIS software to study similarities between 33 catchments in the Western Black Sea Region of Turkey (Aytaç, 2020). Classification techniques such as fuzzy c-means and k-Nearest Neighbour were relied on to develop machine learning models to classify drainage basins and predict streamflow values in ungauged basins (Papageorgaki & Nalbantis, 2016).

Most of the flash flood susceptibility mapping and warning systems methods presented in the literature were developed for areas that have river water level prediction infrastructure installed at appropriate sites. However, it is equally important to predict flash flood events for unmonitored locations in order to develop robust warning systems for the benefit of the residents of these areas. With the availability of huge volumes of geospatial and precipitation datasets, the predictive capabilities of deep learning neural network models can be harnessed to make high-quality river water level predictions for such locations. In this study, an ensemble of multiple deep learning models was implemented to capture relationships between different variables such as gauge height, rainfall, etc., and develop virtual gauges for unmonitored locations in the flash flood-prone catchments of Missouri. The authorities responsible for flood prediction and management tasks can rely on the output of these virtual gauges to take necessary precautions for the safety of the general public.

Chapter 2 Methodology

The objective of this research project was to develop virtual monitors for unmonitored locations in flood-prone catchment areas of Missouri. Figure 2.1 illustrates the methodology used to collect data and develop deep learning-based models to predict water levels at locations without any water level monitoring infrastructure. The first step of the methodology involved developing the catchment database using ArcGIS Pro and datasets from the United States Geological Survey (USGS). In the second step, the flowline distances between different sets of gauges were calculated to identify different groups of gauges needed to develop deep learning-based prediction models. The water level or stage values for gauges in these groupings were obtained from the USGS's data archive. Also, the daily rainfall observation values for the catchments were gathered from the National Weather Service (NWS) archive. Finally, the gauge and rainfall values were used as inputs to the deep learning models that can assist in the task of predicting water levels at unmonitored locations.

Figure 2.1 Model Framework

2.1 Develop Catchment Database

The geospatial data files which contain topographic maps and geographic information system (GIS) datasets for different parts of Missouri were downloaded from the USGS's National Map data repository (The National Map, 2022). The one-meter Digital Elevation Model (DEM) files for Missouri were downloaded using National Map's data download application

(USGS National Map, 2022). Even though the one-meter DEM coverage is limited in some areas of the state as shown in Figure 2.2, these high-quality data files obtained from the light detection and ranging (Lidar) source are suitable for capturing the geospatial features of a location efficiently.

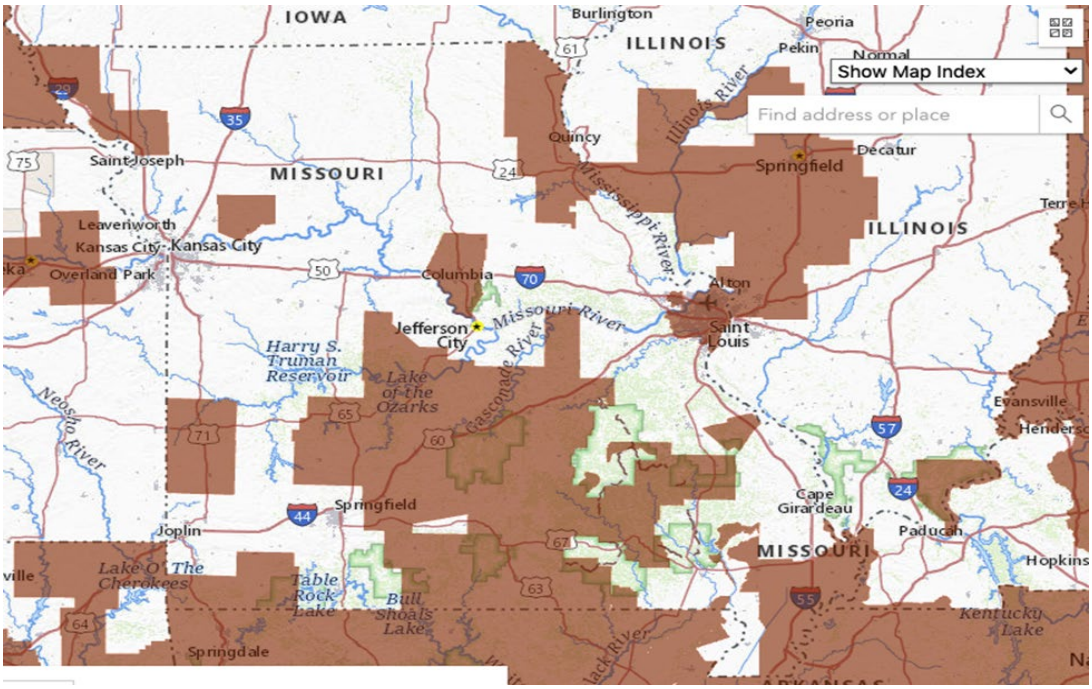

Figure 2.2 One-meter DEM Coverage for Missouri

The goal of the study was to capture the behavior of the gauges in the state's catchment areas to develop deep learning-based virtual gauges to monitor water levels at unmonitored locations. Once developed, these virtual gauges can provide more accurate local flood predictions for the area. The gauge information from the USGS National Water Dashboard's database was also used to identify 50 different catchments in the state where gauges have been installed by the authorities to monitor river and stream water levels (USGS, 2022). Then, the DEM files for these 50 different gauged catchment areas were uploaded to the geospatial

information system (GIS) software ArcGIS Pro for visualization. These files were further processed using the software's inbuilt functions to extract values of various geospatial features such as elevation, slope, area, perimeter, etc. for the catchments. The purple polygons in Figure 2.3 highlight the locations of such gauged catchments in Missouri.

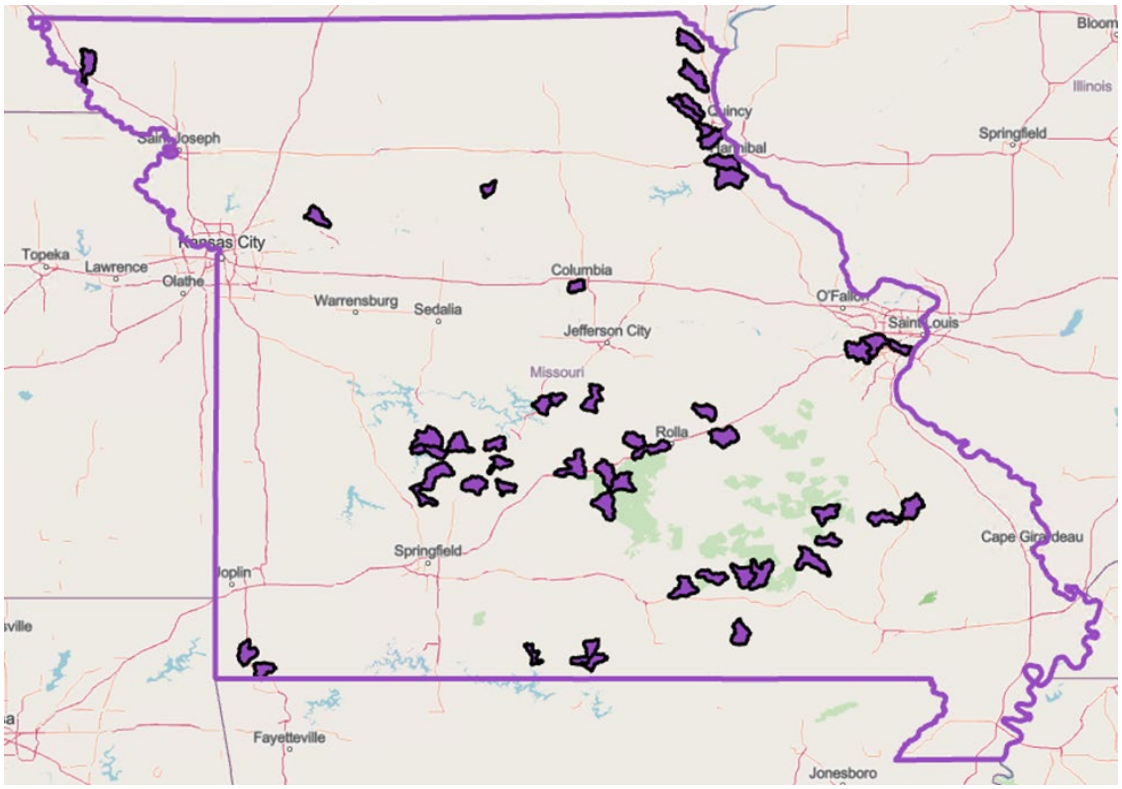

Figure 2.3 Gauged Catchments in Missouri

2.2 Calculate Flowline Distances

The gauges used in this study were divided into three different groups: upstream gauges, gauges of interest, and downstream gauges. The gauge of interest is located between the upstream and downstream gauges as shown in Figure 2.4. The development of a virtual monitor for this gauge of interest will provide a focused flood warning for a specified region without

spending any resources to install and maintain an actual gauge at this location. The distance value between the gauges interacts differently with the other model inputs, so it is necessary to determine the stream distances between the gauges.

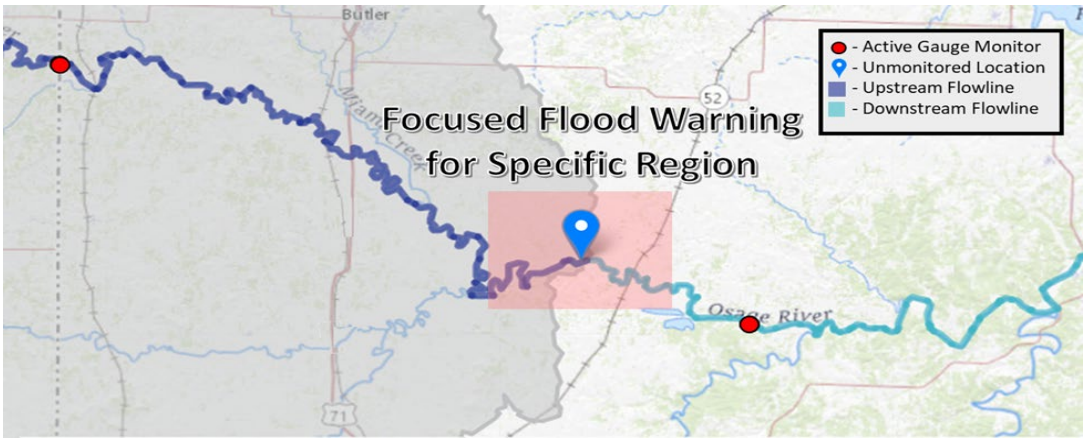

Figure 2.4 Gauge Grouping

Flowlines were drawn in ArcGIS Pro to calculate the polyline distance between an upstream gauge and a downstream gauge in a gauged catchment. Figure 2.5 shows a flowline drawn between two actively monitored gauges on Osage River near Bagnell, MO. Similar lines were drawn between all monitored gauges located in the catchments to extract respective polyline distances between them.

Figure 2.5 Measuring Polyline Distance between Two Monitored Gauges

The number of catchments needed for the predictive analytics tasks was reduced from 50 to 21 in this study due to the lack of availability of gauge height and rainfall data for the remaining 29 catchments. A total of 42 different flowlines were drawn between the upstream, gauge of interest, and downstream gauges, which constitute 21 complete chains of gauges in the catchment database that originally comprised 50 gauged catchments. Later, information for one gauge of interest had to be removed from the database due to the limited availability of the gauge readings for its upstream gauge located on St. Francis River. It is also worth noting that some gauges are located outside the state of Missouri as shown in Figure 2.6.

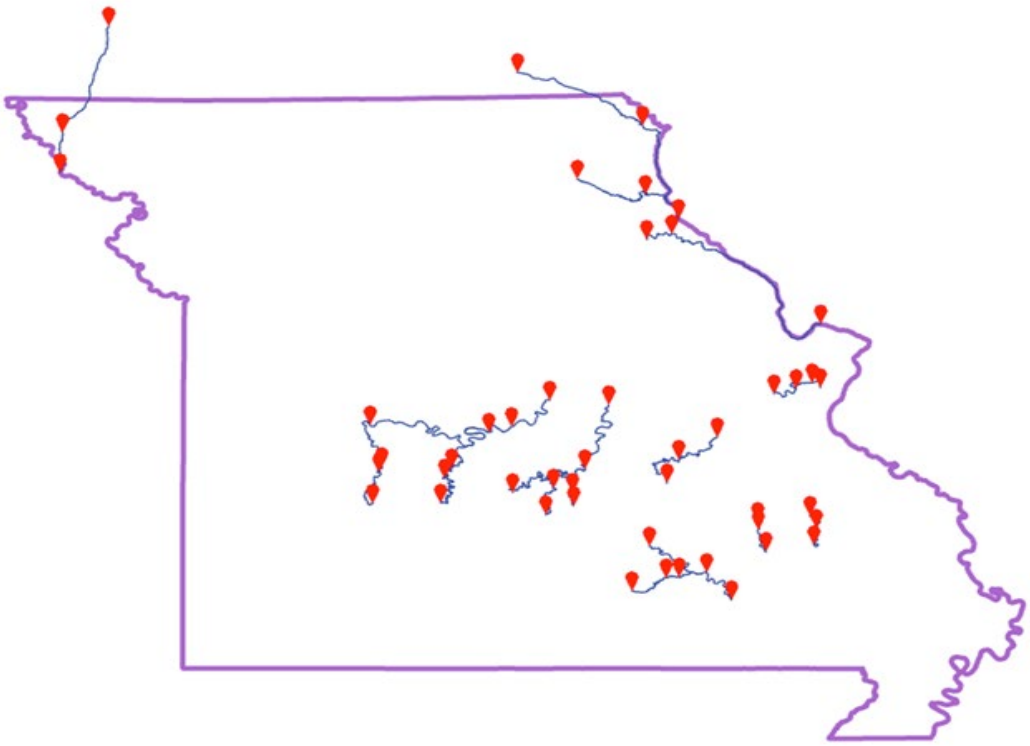

Figure 2.6 River Flowlines Between Monitored Gauges

The geodesic distances between these gauges were compiled in a tabular format using ArcGIS Pro and are presented in Figure 2.7. These upstream and downstream distances were measured in kilometers and were calculated using ArcGIS Pro's 'Calculate Geometry' tool. The geodesic distance represents a more accurate distance between two points on the earth's curved surface as the elevation difference between them is reflected in the calculations (ESRI, 2022). This proximity information enabled the development of specialized deep learning-based prediction models, trained to make predictions at differing distances from sponsoring input gauges.

ID	Catchment	Upstream [km]	Downstream [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
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

Figure 2.7 Flowline Geodesic Distances

2.3 Download Gauge and Rainfall Data

The river gauge readings for the upstream and downstream gauges were procured from the USGS National Water Information System after conducting the relational mapping operation between these waterway gauges by calculating the geodesic distances between them (USGS National Water Information System, 2022). The historic readings for both the upstream and downstream gauges were downloaded for a date range lying between September 1, 2016 and December 30, 2021. These data values were then resampled at 30-minute intervals using Python scripts to use them as inputs to the deep learning-based models. The historic readings for the unmonitored gauges of interest were not used as inputs to these models as this information would

not be available to the user in the eventual scenario where a deep learning model is relied on to predict gauge heights at such a location.

The National Weather Service (NWS) Archive maintains daily rainfall data values for the contiguous United States. The rainfall data from September 1, 2016 - December 30, 2021 was obtained from the National Weather Service (NWS) archive for the catchments. The daily rainfall data from September 1, 2016 - June 27, 2017 was available in the point format (NWS, 2022). The rainfall data was also available in the raster format from June 28, 2017 - December 30, 2021 (NWS, 2022). The daily rainfall observation values were procured using the 'Clip' tool in ArcGIS Pro and were divided by 48 to obtain values at a time interval of 30 minutes. The complete dataset containing time series-based upstream and downstream gauge values and rainfall values was used as an input to the deep learning neural networks.

2.4 Implement Deep Learning Models

In order to predict the gauge readings at the central gauge of interest, the readings from the upstream and downstream gauges were used as inputs to an ensemble of multiple Long Short Term Memory (LSTM) neural network-based models. Apart from using these two model input features, the average localized rainfall values within the catchment were also used to make such predictions. The predicted gauge height values were then compared with the true time series labels at the gauge of interest to evaluate the performance of the deep learning models. Eventually, the learning outcome of these prediction models can be applied to when predicting gauge height values at unmonitored locations. The dataset contained a total of 118,021 observations for all three features: upstream gauge height, downstream gauge height, and catchment rainfall. These time series-based data observations represented feature values at a time interval of thirty minutes. Different data processing techniques such as data cleaning, exploratory

data analysis, data normalization, etc., were conducted on this multivariate dataset to prepare a suitable format for an ensemble of multiple LSTM-based deep learning neural networks. Figure 2.8 shows the framework of the LSTM-based deep learning neural networks implemented on the multivariate time series dataset to predict the gauge heights at the gauge of interest.

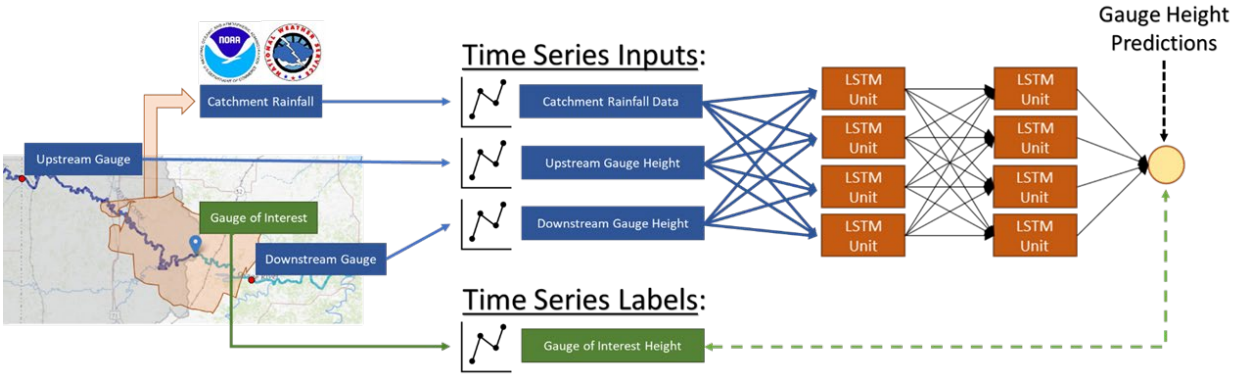

Figure 2.8 LSTM Deep Learning Neural Networks Framework for Gauge Height Predictions

The entire dataset was divided into two sets of training and testing datasets for analysis. 80% of the whole dataset constituted the training dataset and the remaining 20% of it was used as a testing dataset to evaluate the model performance. Also, 15% of the training dataset was allotted to a validation dataset to tune the parameters of the deep learning models and improve their prediction accuracy.

2.5 Gauge Grouping Clusters

Figure 2.9 details how each of the 20-gauge groupings in Figure 2.6 were partitioned into four different grouping clusters. These clusters were delineated by the distance between individual gauges within each gauge grouping. Specifically, for each gauge grouping, two distance values were determined: the distance between the gauge of interest and the upstream gauge and between the gauge of interest and the downstream gauge. The median exhibited value

for each of these distances was then determined with respect to all 20-gauge groupings. For each gauge grouping, their respective distance values were compared to these median values to assign a fuzzy tag describing the general proximity of the individual gauges in the grouping. For instance, a gauge grouping that maintained a distance less than the 20-grouping median between the gauge of interest and the upstream gauge was assigned a 'Close' tag as the relationship. If the same grouping's distance between the gauge of interest and the downstream gauge was greater than the 20-group median, a 'Far' tag would be assigned as the relationship. A labeling mechanism describes these tags in the leftmost graph of Figure 2.9: 'Close-Close' (CC), 'Close-Far' (CF), 'Far-Close' (FC), and 'Far-Far' (FF). Therein, the first tag describes the gauge of interest-to-upstream gauge distance with the second tag describing the gauge of interest-to-downstream gauge distance. Data from gauge groupings within each cluster was then used to train respective models to predict the behavior of virtual gauges that would fall into the same cluster.

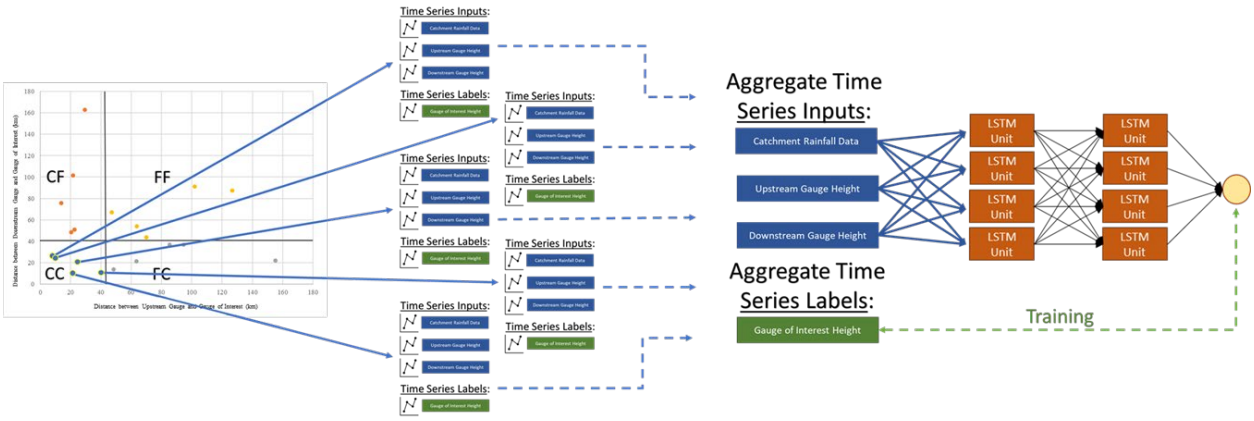

Figure 2.9 Gauge Groupings and Four Clusters

This clustering was completed to ease the prediction burden that would be incurred if a single model was used to make accurate predictions regardless of the proximity of the gauges used to provide input feature data. In this way, four specialized models were created to predict virtual gauge heights for ‘CC’, ‘CF’, ‘FC’, and ‘FF’ scenarios, respectively.

For each of the four grouping clusters with five-gauge groupings each, four-gauge groupings were used in an aggregate fashion to train an ensemble model dedicated to predictions on the grouping cluster. A final gauge grouping was kept out of training to test the generalization of the trained model for its ability to process new data as shown in Figure 2.10. The four novel gauge groupings were reserved to evaluate their respective grouping clusters. This approach simulated the application of a trained cluster ensemble to a scenario where the virtual gauge was desired while enabling the comparison of the model’s output to known true gauge-of-interest values.

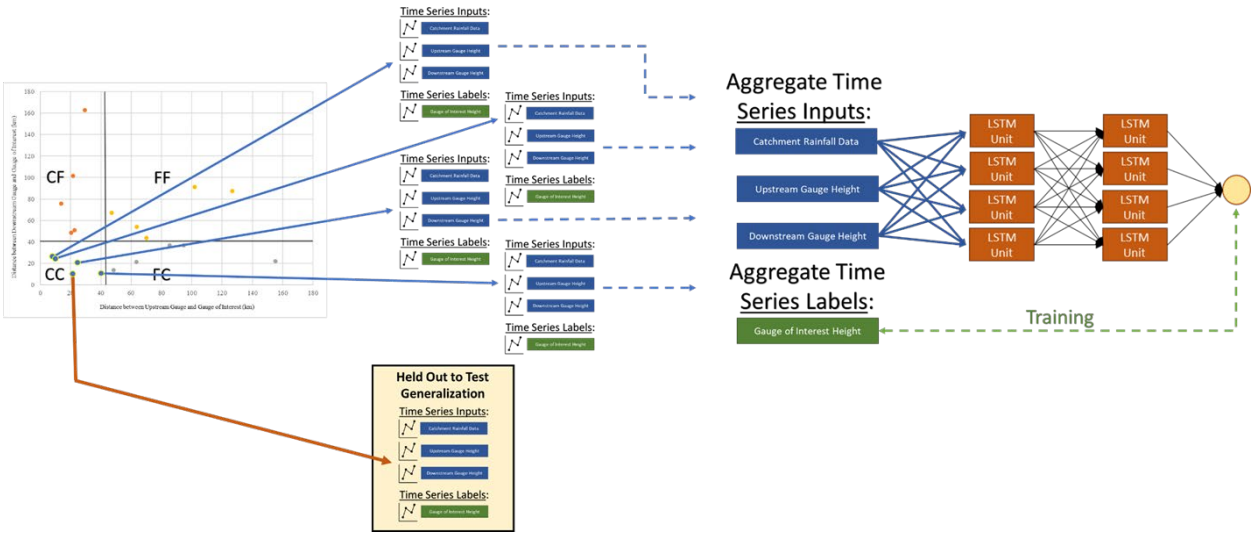

Figure 2.10 Gauge Groupings and Deep Learning Models

2.6 Long Short Term Memory (LSTM) Deep Neural Networks

A deep learning neural network is a mathematical model which consists of multiple layers with mathematical functions called neurons that process the input data values and pass the corresponding output information to the next layer in the network. The three different types of layers in a deep learning neural network are an input layer, hidden layers, and an output layer. The neurons in the input layer receive the input dataset and transfer the necessary information to the interconnected hidden layers. This information is processed within the hidden layers and its output is sent to the output layer of the network. Later, the final output layer generates an appropriate model output value for a regression task such as predicting water levels at a gauge of interest.

An LSTM-based deep learning neural network was developed to predict the gauge height values for the gauges of interest from each cluster. Long Short Term Memory (LSTM) networks are a category of recurrent neural networks that are suitable to make predictions for an input time series-based feature dataset. These neural networks capture the interdependencies present between the different time steps of a huge input dataset to generate accurate prediction results. These interdependencies are preserved in its network through a set of gating mechanisms that aid in the efficient retention and movement of critical information between different steps of a sequential time series dataset (laddad, 2019). The LSTM-based deep learning neural networks implemented in the research project used input feature values from the previous 30 timesteps (15 hours) to predict gauge height level readings which are 8 timesteps (4 hours) ahead in the future. The architecture of the LSTM implemented in this study consisted of eight different layers arranged sequentially. The time series dataset with values for different features such as upstream gauge height, downstream gauge height, and average catchment rainfall was used as input to the

initial 50-unit layer of the LSTM network as shown in Figure 2.11. These units represented the dimension of outputs and the number of parameters in the LSTM layer (Tung, 2022). A dropout regularization layer was added next to help avoid overfitting (Ampadu, 2021). An overfitted model will perform poorly on the test dataset producing high test error values and inaccurate prediction results (IBM, 2021). A dropout value of 0.2 was used to randomly drop 20% of a layer's output neurons to prevent overfitting while training the model with the input data values. This pair of a 50-unit LSTM layer and 0.2 or 20% dropout layer was repeated twice in the architecture of the LSTM deep neural network. The penultimate layer with 50 units passed its outputs to the final one-unit 'Dense' layer. This 'Dense' layer was used to extract a single-value prediction from the model for the gauge height of the gauge of interest four hours ahead. Whilst training, this prediction value was compared to the true four-hour-ahead value from the physical gauge of interest to determine error and backpropagation proceedings. When applied to novel scenarios, the output of the model can be regarded as a prediction for a virtual gauge between the upstream and downstream gauges whose time-series data was passed as the input to the deep learning model.

```
# Model Construction
model = Sequential()
# First LSTM Layer and Dropout Regularization
model.add(LSTM(units=50, return_sequences=True, input_shape=(X_trainIn.shape[1], numFeatures)))
model.add(Dropout(0.2))
# Second LSTM Layer and Dropout Regularization
model.add(LSTM(units=50, return_sequences=True))
model.add(Dropout(0.2))
# Third LSTM Layer and Dropout Regularization
model.add(LSTM(units=50, return_sequences=True))
model.add(Dropout(0.2))
# Final LSTM Layer
model.add(LSTM(units=50))
# Dense Output Layer
model.add(Dense(units=1))
```

Figure 2.11 LSTM Deep Learning Neural Network Architecture

An ensemble learning approach was implemented on the input feature dataset to train multiple LSTM-based deep learning models for all gauges in each cluster category. The prediction values from these multiple models were combined to reduce the variance of the predictions and generate better prediction results as compared to a single deep learning model (Brownlee, 2019). For each cluster, 30 identical instances of the LSTM model were trained and combined into an ensemble to make both an average prediction and a prediction interval about this average. This ensemble method was also used to mitigate the impact that initialized weights may have on the final prediction of a single model. Training an ensemble of models on the same data while providing different weight initializations reduced the overall bias these initial weights may have had on a final solution. Further, the variance between the predictions made by individual models within the ensemble describe a band about the ensemble's average prediction where the true value is expected to lie.

Chapter 3 Results and Discussion

An ensemble of LSTM-based deep learning neural networks is trained on the time series dataset containing gauge height and catchment rainfall values to predict gauge height readings at virtual gauges located in flood-prone regions. The model outputs obtained after analyzing the aggregate input feature dataset are compared against the true gauge of interest values to evaluate the capabilities of the prediction models. The deep learning-based models assist in efficiently capturing the relationships between the input data feature points. Also, the application of the ensemble learning approach to train such models to make high-quality predictions with low generalization errors. The predicted values are compared against the true gauge readings for all the 'Close-Close' (CC), 'Close-Far' (CF), 'Far-Close' (FC), and 'Far-Far' (FF) gauge groupings to evaluate the generalization capabilities of respective gauge grouping ensembles.

Figure 3.1 details the relationship between model predictions and true gauge height values for the novel 'Close-Close' (CC) gauge grouping used to test the generalization of the trained 'Close-Close' (CC) ensemble. The predicted values accurately trace the true values, capturing the intricacies and progressions of the gauge of interest's rise and fall in water levels. Additionally, a 95% prediction interval about the 'Close-Close' (CC) ensemble's mean predicted value is shown in Figure 3.1 as light blue. These prediction interval bounds are specified by the values that lie at a distance of '1.96σ' above and below the mean prediction values obtained after training an ensemble of LSTM deep learning models. Ideally, the predicted values should lie between the prediction interval to highlight more accurate forecast results from a prediction model. Here, 'σ' represents the standard deviation of the predictions calculated by the 30 LSTM-based deep learning models comprising the 'Close-Close' (CC) ensemble. Also, these prediction-

interval bounds fluctuate in scenarios where the ensemble models show less variability and display a wider range and more nonconformity between the model predictions exists.

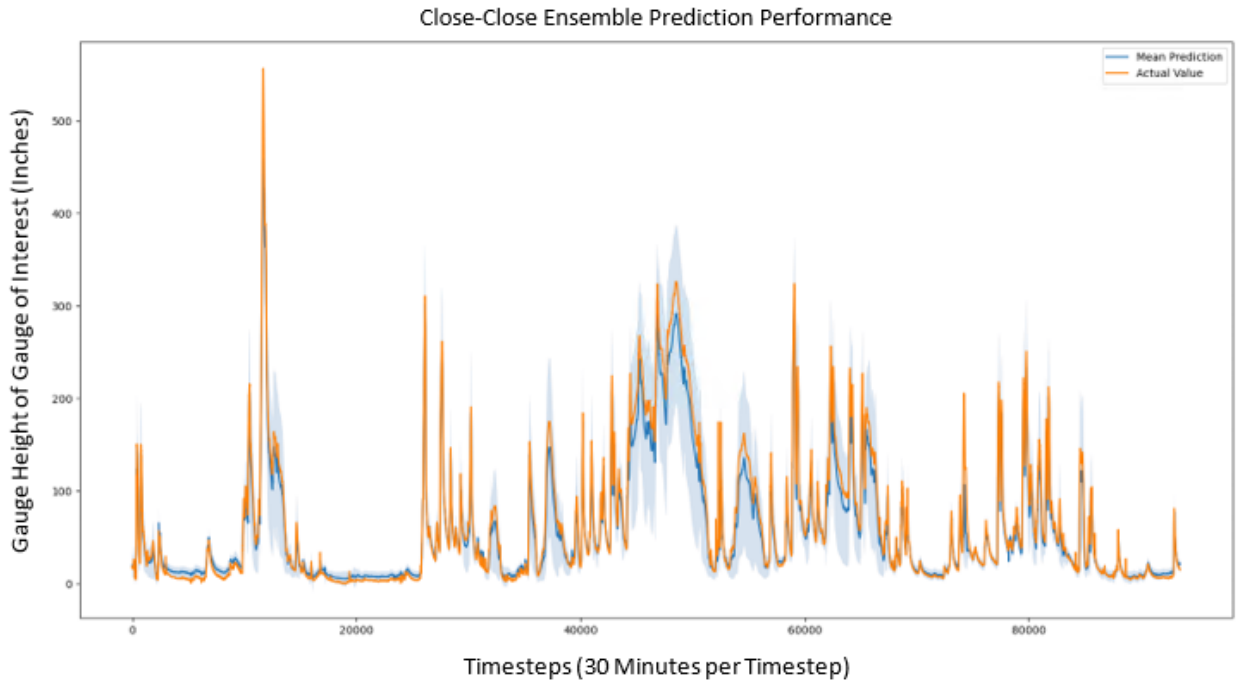

Figure 3.1 Comparison of Predicted and True Gauge Height Values for 'Close-Close' (CC) Gauge Grouping

Summary statistics results for the 'Close-Close' (CC) ensemble's prediction performance metrics are shown in Figure 3.2. These metrics are obtained from the application of the trained 'Close-Close' (CC) ensemble to a novel gauge grouping scenario that fits the 'Close-Close' (CC) cluster classification. The mean and median absolute deviation values between the true and predicted values are 8.683 and 5.119 inches, respectively which means that the spectrum of the gauge grouping's gauge of interest is 556.44 inches. Alternatively, it means that the difference between the maximum and minimum gauge height values exhibited by the gauge of interest is 556.44 inches. This spectrum also implies that the mean and median absolute deviation relate to

a 1.56% and 0.92% deviation from the true value when compared to the exhibited spectrum. The width of the generated 95% prediction interval is quite wide at 55.579 inches, on average. This width does not detract from the predictive capability of this ensemble as the average model prediction exhibits a 0.9948 correlation coefficient to the true gauge of interest gauge height. As the gauge of interest displays a 0.8784 correlation coefficient, on average, to upstream and downstream gauge heights, an increase of 0.1155 is attributable to the ensemble model. This relates to a 13.133% increase in the correlation coefficient performance beyond the information available from upstream and downstream gauges directly.

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%
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 3.2 ‘Close-Close’ (CC) Ensemble Summary Statistics

The relationship between the predicted and true gauge height values for the ‘Close-Far’ (CF) gauge grouping is shown in Figure 3.3. The predicted values trace the true values to some extent and still manage to capture the gauge of interest’s behavior and its underlying patterns. A

95% prediction interval about the ‘Close-Far’ (CF) ensemble’s mean predicted value is shown in light blue. While the prediction intervals generated by the ‘Close-Far’ (CF) ensemble also fluctuate to match individual model variation, this implementation of the ‘Close-Far’ (CF) ensemble does not exhibit as much variability as was demonstrated in the ‘Close-Close’ (CC) ensemble. Even though there is generally some deviation between the ensemble predicted values and their true-valued counterparts, the characteristic mirroring between the two sequences is still useful.

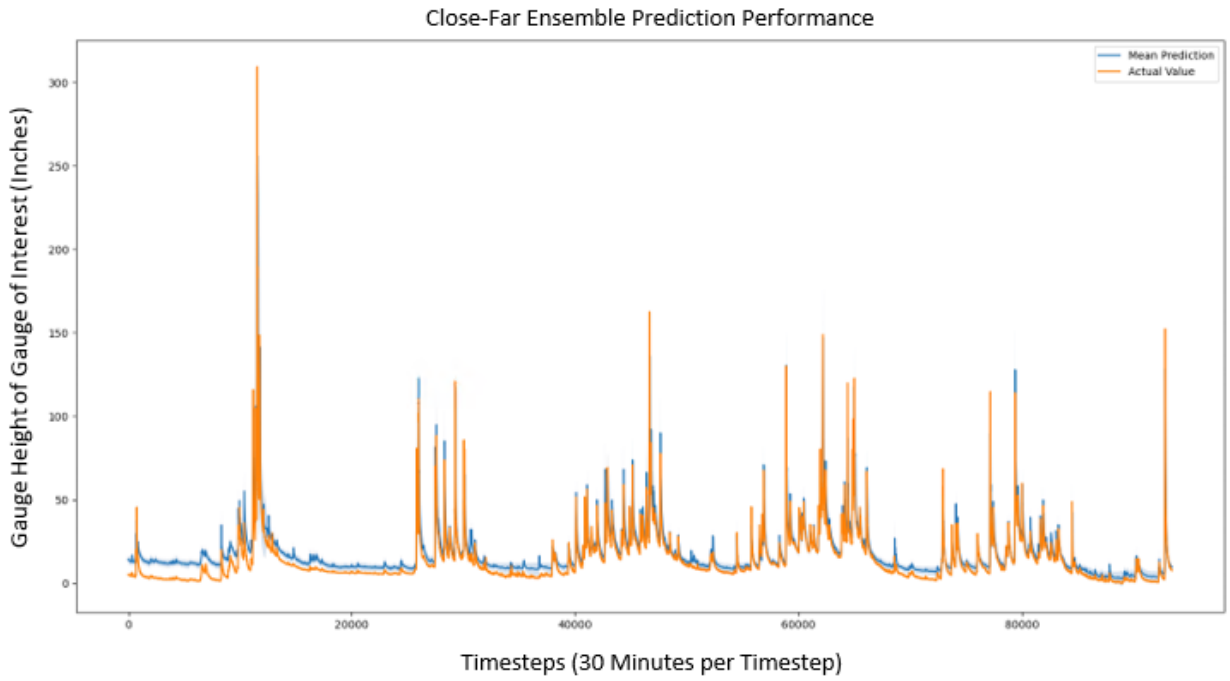

Figure 3.3 Comparison of Predicted and True Gauge Height Values for ‘Close-Far’ (CF) Gauge Grouping

The mean and median absolute deviation between the true and predicted values for the ‘Close-Far’ (CF) ensemble are 4.624 and 3.679 inches, respectively as shown in Figure 3.4. Here, the spectrum of the gauge grouping’s gauge of interest is 309.36 inches. This spectrum

implies that the mean and median absolute deviation relate to a 1.49% and 1.19% deviation from the true value when compared to the exhibited spectrum. This also means that a direct comparison of the absolute deviation values in inches is not a fair comparison. The 'Close-Far' (CF) ensemble's prediction performance is much lower in terms of explicit deviation but the percentage of the gauge-of-interest's spectrum these deviations comprise is relatively similar. The width of the 95% prediction interval is narrow at 6.704 inches or 2.17% of the gauge of interest's spectrum on average. The correlation coefficient value between the ensemble's predictions and actual values of the gauge is 0.9441 which is an improvement of 7.694% beyond the information discernable from the upstream and downstream gauges directly.

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%
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%	

Figure 3.4 'Close-Far' (CF) Ensemble Summary Statistics

Figure 3.5 shows the relationship between model predictions and true gauge height values for the novel 'Far-Close' (FC) gauge grouping used to test the generalization of the

trained ‘Far-Close’ (FC) ensemble. The predicted values trace the true values to some extent and still succeed in capturing the gauge-of-interest’s behavior and patterns, enabling utility from this ensemble of models. A 95% prediction interval about the ‘Far-Close’ (FC) ensemble’s mean predicted value is displayed in light blue.

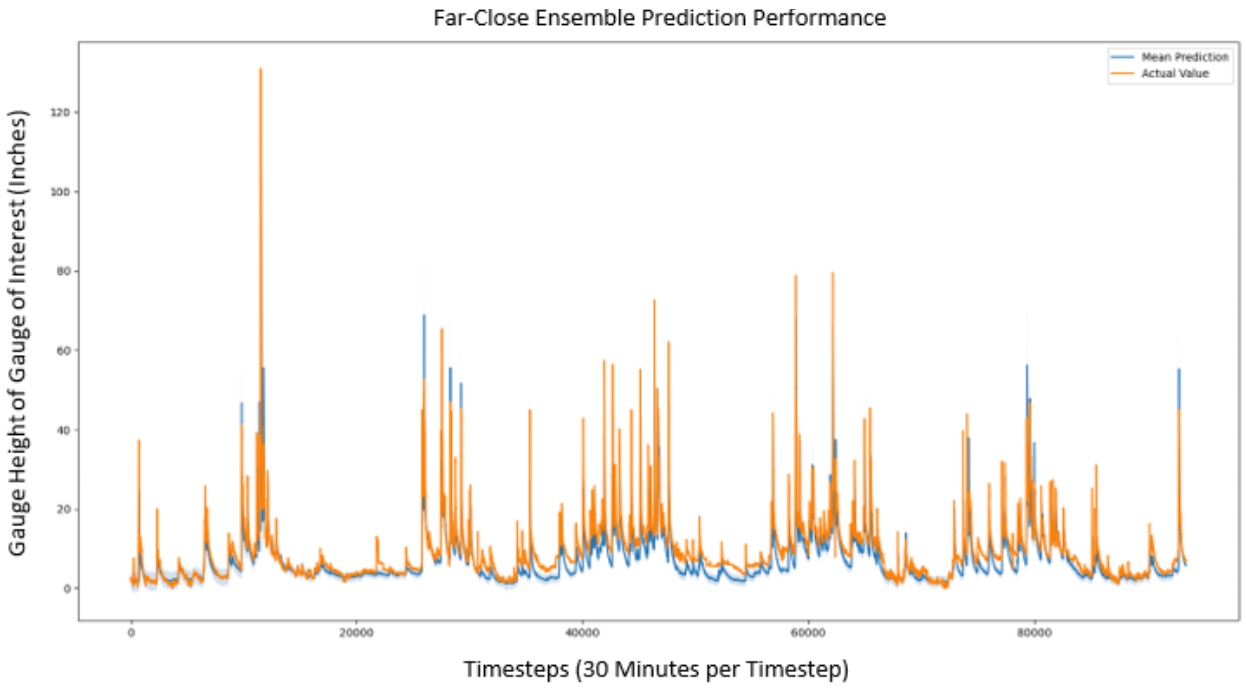

Figure 3.5 Comparison of Predicted and True Gauge Height Values for ‘Far-Close’ (FC) Gauge Grouping

The summary statistics of performance metrics for the ‘Far-Close’ (FC) ensemble shown in Figure 3.6 are derived from the application of the trained ‘Far-Close’ (FC) ensemble to a novel gauge grouping scenario that fits the ‘Far-Close’ (FC) cluster classification. The mean and median absolute deviation parameter values between the true and predicted gauge height values are shown to be 2.278 and 1.766 inches, respectively. The spectrum of the gauge grouping’s gauge-of-interest equals 130.92 inches. This spectrum implies a mean and median absolute

deviation of 1.74% and 1.35%, respectively, of the true gauge height value to the spectrum. The width of the 95% prediction interval is 3.804 inches or 2.91% of the gauge-of-interest's spectrum, on average. The ensemble's predictions demonstrate a 0.9208 correlation coefficient to the actual values of the gauge which results in an improvement of 3.605% beyond the information discernable from the upstream and downstream gauges directly.

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	
Percent Correlation Coefficient Improvement Attributable to Model:	3.605%	

Figure 3.6 'Far-Close' (FC) Ensemble Summary Statistics

In the case of the 'Far-Far' (FF) gauge grouping, the predicted values trace the true values to some extent and accurately capture the gauge-of-interest's behavior and patterns as shown in Figure 3.7. A 95% prediction interval about the 'Far-Far' (FF) ensemble's mean predicted value is shown in light blue. With some exceptions, the 'Far-Far' (FF) ensemble predictions closely trace the true gauge height values, suggesting an ability to competently generalize to a novel 'Far-Far' (FF) gauge grouping scenario.

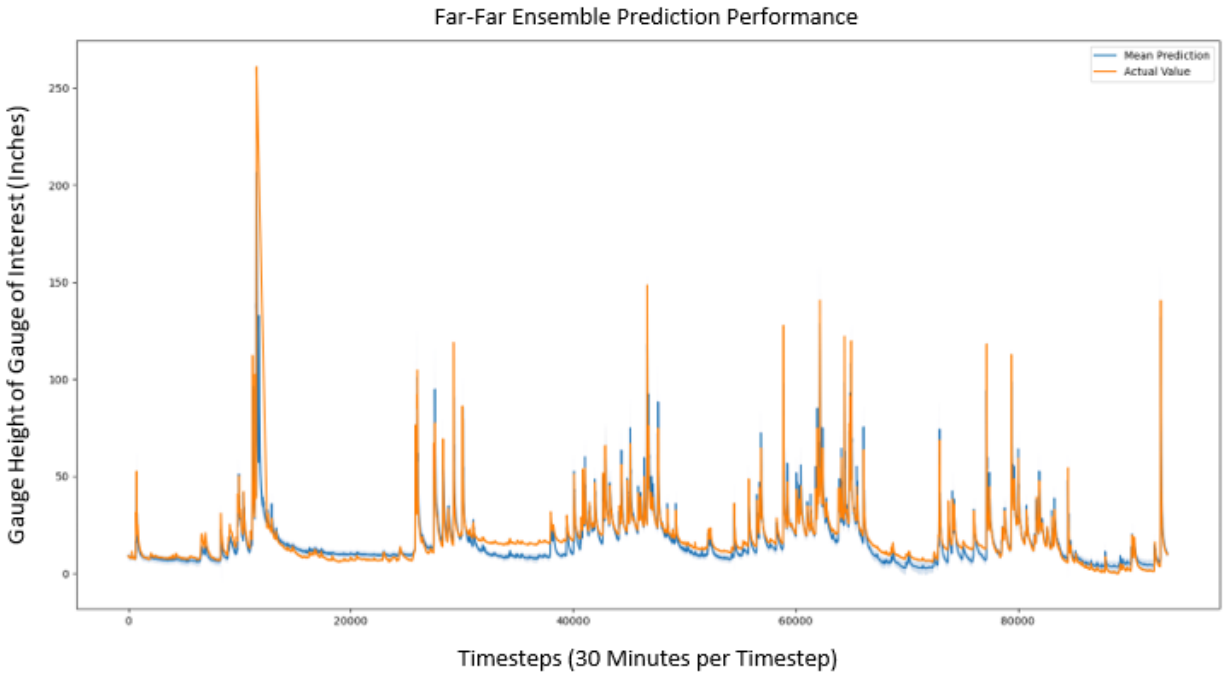

Figure 3.7 Comparison of Predicted and True Gauge Height Values for 'Far-Far' (FF) Gauge Grouping

The mean and median absolute deviation readings between the true and predicted values are shown to be 3.944 and 2.738 inches, respectively for the 'Far-Far' (FF) ensemble as illustrated in Figure 3.8. The value of the spectrum of the gauge grouping's gauge-of-interest is 260.88 inches, which implies that the mean and median absolute deviation represent a 1.51% and 1.05% deviation from the true value when compared to the spectrum exhibited. The width of the 95% prediction interval generated is 6.337 inches or 2.43% of the gauge of interest's spectrum, on average. The ensemble's predictions have a 0.8351 correlation coefficient to the actual values of the gauge which showcases an improvement of 8.661% beyond the information discernable from the upstream and downstream gauges directly.

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%	

Figure 3.8 ‘Far-Far’ (FF) Ensemble Summary Statistics

The ensemble of LSTM-based deep learning models is trained for different sets of ‘Close-Close’ (CC), ‘Close-Far’ (CF), ‘Far-Close’ (FC), and ‘Far-Far’ (FF) clusters characterized by the flowline distances between individual gauges within each gauge grouping. These ensemble models make high-quality predictions for the novel gauges-of-interest within the respective four cluster scenarios with a mean absolute deviation of 4.882 inches across the four examined instances. There is a 1.577% deviation in relation to the average spectrum across the four novel gauges-of-interest. Relatedly, the median absolute deviation across the four novel gauge groupings is 3.326 inches which is 1.127% of the average four-gauge spectrum. The mean absolute deviation for all four novel gauge groupings is greater than the median absolute deviation which represents a right-skewed distribution of the absolute deviation values. This indicates there are some instances of larger absolute deviations that draw the mean to the right of the median.

The deep learning ensemble model's correlation coefficient value between its predicted values and the true gauge values is greater than the correlation coefficient value between the upstream, downstream gauges, and the true gauge values. The model demonstrates an improvement in the prediction of gauge-of-interest information beyond what is available by considering the upstream and downstream gauges. Also, the average correlation coefficient between all gauges considered in the modeling efforts is 0.677, which shows that this approach provides predictive information beyond what is available through simple correlation modeling. Moreover, these correlation coefficients are calculated in-time, so that the noted correlation coefficients achieved by examining other gauges could be achieved for the current timestep(t=now). The ensemble models developed in this study achieve the superior correlation coefficient values while efficiently predicting the gauge height values eight timesteps, or four hours, into the future (t=now+4 hours). Here, the value of each timestep is 30 minutes. The models provide advanced knowledge of changes in the gauge height values at the gauge-of-interest that grants a head start to emergency planning and mitigation officials to notify the public of impending flooding scenarios. The accurate multi-step gauge height predictions provide sufficient time for the authorities to issue warnings for the safety of the general public.

Also, the learning outcomes of the LSTM-based deep learning neural networks can be implemented using an unseen feature dataset to make gauge height predictions for novel gauge groupings. The models developed for each cluster can be used to make predictions about a scenario that would fit into a cluster lacking a gauge-of-interest. The models associated with each cluster can be used appropriately to create virtual gauges between an upstream and downstream gauge pair that can be used to make gauge height predictions without installing and maintaining a physical gauge at unmonitored locations. In this case, the authorities do not have

to allocate resources to install gauges at such locations and can rely on deep learning models to make water level predictions.

Chapter 4 Conclusions

A computational intelligence-based methodology was implemented to develop virtual gauges and predict river water levels for unmonitored catchment locations in Missouri. A set of 20 different gauges were selected and divided into different clusters based on the distances between the gauges in each cluster. The gauge groupings from the clusters were then used in an aggregated manner to implement an ensemble of Long Short Term Memory (LSTM)-based deep learning neural networks. The deep learning neural networks were applied on a high-quality feature data set to make model predictions that were used to develop virtual gauges and predict river water levels at unmonitored sites. The multivariate feature dataset needed to develop the model framework was gathered from the data archives of reliable government agencies such as the United States Geological Survey (USGS) and the National Weather Service (NWS).

The implementation of an ensemble learning approach in this research project assisted in the task of efficiently processing and analyzing the huge volumes of data points to make accurate predictions. The LSTM's parameters were also tuned accordingly to capture the intricate relationships between distinct data variables and develop a generalized neural network that generates highly accurate model outputs. These outputs were generated four hours into the future. This gives first responders the ability to use the model outputs where real-time water level information is needed to take advance actions for the safety of the public. Also, the authorities can focus on making critical flash flood precaution-related decisions without being concerned about the allocation of resources such as time and money to install actual gauges to predict future water levels at a given site.

A few limitations about the research presented are worth mentioning. The models developed can only be employed in scenarios where upstream and downstream gauge data are

available in a timely manner, along with rainfall data within the same rainfall catchment that the virtual gauge is planned for. If modified models are created that require different subsets of these input features, the models may be applied to a larger number of locations and scenarios. For instance, if a model is trained to make predictions while requiring only a downstream gauge and rainfall data, it may be more broadly applicable. This broadened range of application may be of considerable utility, even if the general accuracy is reduced given the leaner inputs to the model.

These relatively restrictive input feature requirements also limit the number of gauge groupings that would be used to train the developed models. While this research included only 20-gauge groupings in their training and validation efforts, a less demanding input feature prerequisite would have allowed for the inclusion of many more gauge groupings in the training set. Increasing the number of gauge groupings used in training these predictive models may be beneficial as it could expose the models to scenarios and intricacies that are not captured by the current training set. While this research focused on supplying training data comprising all three of these critical hydrologic features, other, less demanding models may receive a boost from a broadened training set.

Additionally, the models developed in this research utilized average daily rainfall data from the gauge of interest catchments in crafting the rainfall input feature. To do so, the daily values were resampled into 30-minute increments, essentially describing a steady rainfall throughout the day amounting to the average observed daily total for the catchment. While this provides some information about the rainfall experienced within a catchment, this procedure introduces a lot of noise and may lead to less accurate prediction results than expected from a more granular rainfall reading. While this effect on accuracy has not been verified, it stands to

reason that the absence of correspondingly granular rainfall readings may impact the performance achieved by a model making predictions at a step of only four hours ahead.

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