A Study on Estimation of Flood using Machine Learning
- Case of Samcheok Osipcheon Basin -

Presenter : Ph.D Student[1)] | An, Sung Wook

Author : Lee, Byung Hyun[2)] | Kim, Byung Sik[3)]

1). Ph.D Student, Department of Urban Environmental and Disaster Management,
Kangwon National University (E-mail:aso750@kangwon.ac.kr)
2). Research Professor, AI for Climate & Disaster Management Center, Kangwon National University
(E-mail : hydrobh2@kangwon.ac.kr)
3). Corresponding Author : Professor, Department of Artificial Intelligence & Software/Graduate School of Disaster Prevention, Kangwon National University
(E-mail: hydrokbs@kangwon.ac.kr)

AI for Climate & Disaster Management Center

GEWEX

KNU KANGWON NATIONAL UNIVERSITY

CONTENTS

1 Introduction

1.1 Research Background

1.2 Research Trends

2 Theory and methodology

2.1 Decision Tree Models

2.2 LSTM(Long Short-Term Memory)

2.3 Flowchart

3 Application and results

3.1 Current status of the study area

3.2 Building Inputs and parameter settings

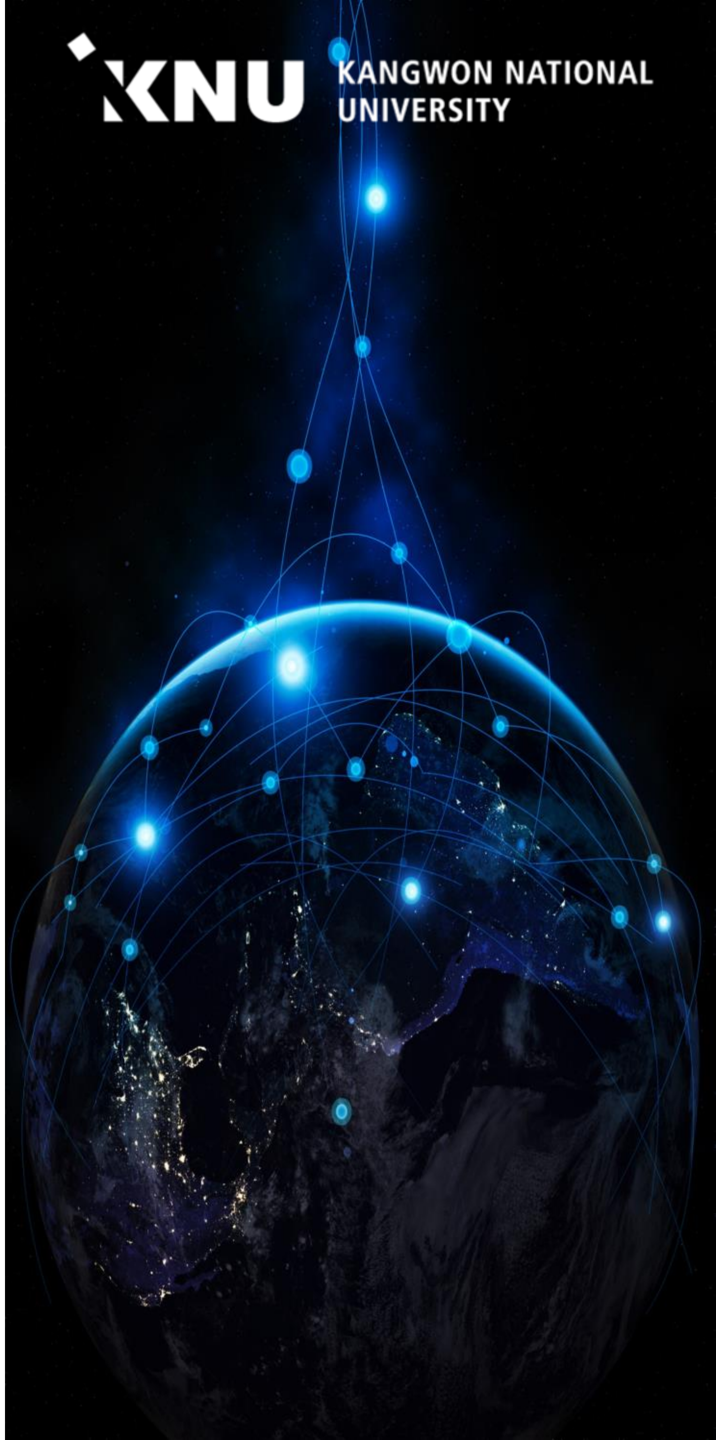
3.3 Model Training and simulation

4 Conclusion

1. Introduction

1.1 Research Background

1.2 Research Trends



1.1 Research Background

- Climate change is increasing the frequency and intensity of heavy rainfall events, causing flood damage to rise every year
- Need for real-time situation analysis system to reduce damage

Damages from flooding in :

2021
Busan caused by Typhoon Omais

Date	2021. 08. 24
Region	Busan
Damage	6 people isolated due to flooding in the main building, swimming pool building, and roadway

2022
Seoul caused by heavy rainfall

Date	2022. 08. 09
Region	Seoul, kangnam
Damage	26 fatalities and 2,280 displaced people

2023
Osong(Cheonan) by heavy rainfall

Date	2023. 07. 15
Region	Osong
Damage	23 people killed as buildings, roads, and underpasses flooded

- The scale of urban flooding is growing
- Need to establish a real-time monitoring and situation analysis system
- The physical model used tends to take a long time to analysis and overestimates the situation

---

- AI is relatively simple to build input data and can utilize various input data such as numbers, images, and voice
- Various sensor data can be linked for damage monitoring and results can be derived within a short period of time

<South Korea's Ministry of Environment Press (2020)>

Ministry of Environment pushes for flood forecasting using artificial intelligence

- Using artificial intelligence (AI) for faster and more accurate flood forecasting
- Utilizing advanced technologies such as rain radar and satellites to respond to flash floods
- Establishing scientific flood countermeasures by producing and distributing flood risk maps

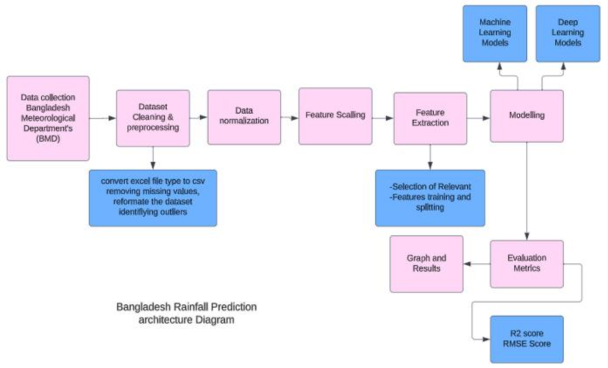
The Ministry of Environment will introduce a flood forecasting system utilizing artificial intelligence (AI) by 2025 to respond to heavy rains caused by climate change

Flood information collection sensors will be installed in 100 local rivers nationwide that are vulnerable to flooding, and a flood forecasting platform using AI will be built using information from each sensor.

1.2 Research Trends

Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh
| 2023, Adel Rajab et, al., WATER

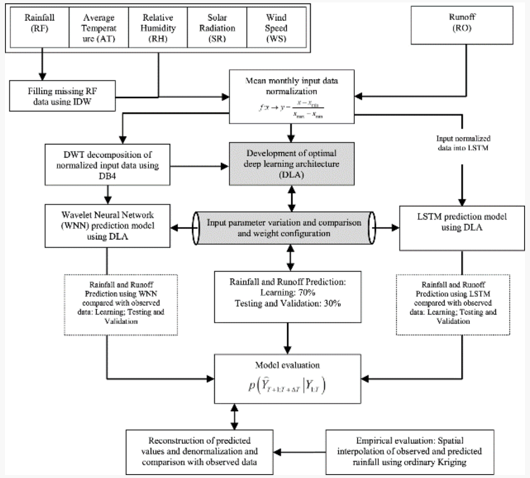
- To predict floods in Bangladesh, AI models were trained on 16 different weather data sets, including daily temperatures
- Out of 11 AI models, LSTM (Long Short-Term Memory) performed the best



Block diagram describing the proposed system

Rainfall and runoff time-series trend analysis using LSTM recurrent neural network and wavelet neural network with satellite-based meteorological data: case study of Nzoia hydrologic basin
| 2022, Yashon O. Ouma et, al., Complex & Intelligent Systems

- Collected meteorological data consisting of precipitation, average temperature, relative humidity, wind speed, and solar radiation to simulate runoff in the Nzoia River basin in Kenya
- Trained LSTM and wavelet neural networks to simulate runoff, and found that precipitation was the most important factor in the results



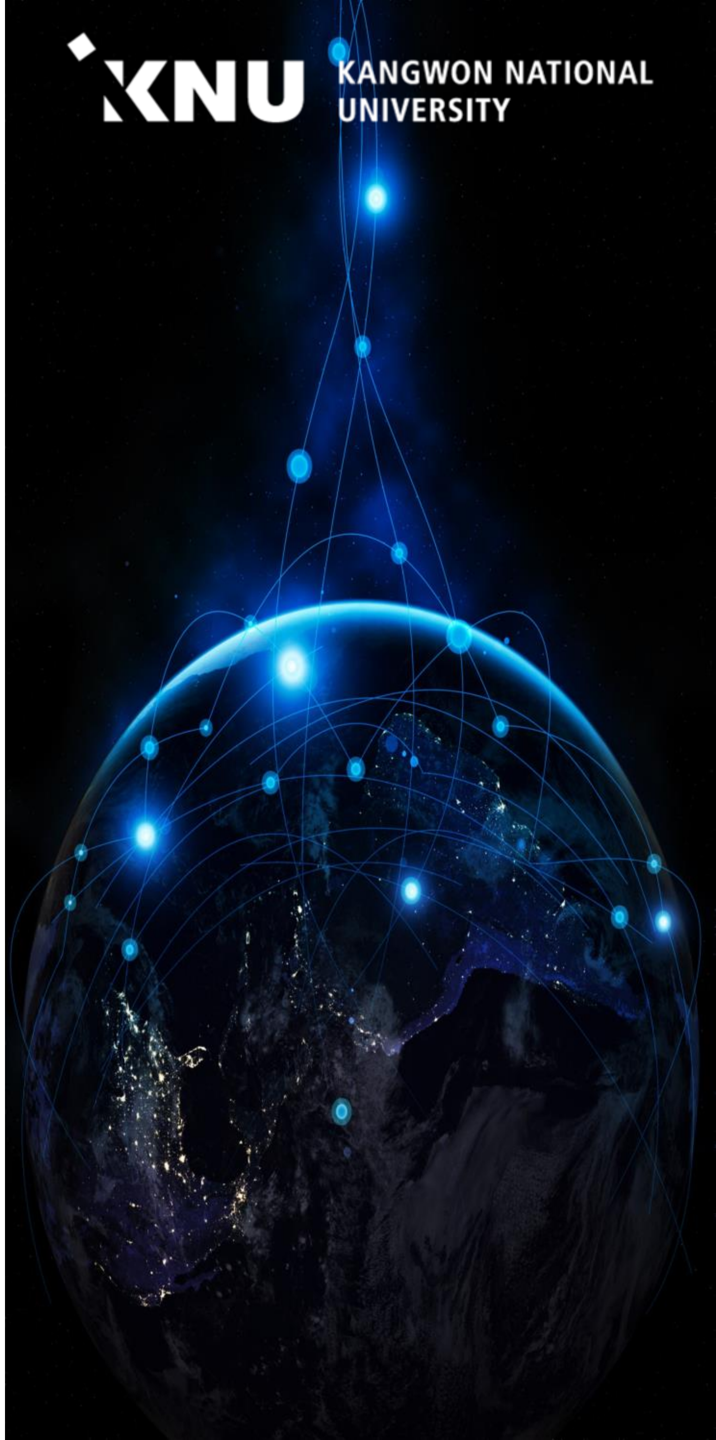
Processing Flowchart for Rainfall and Runoff Prediction Using LSTM and WNN

2. Theory and methodology

2.1 Decision Tree Models

2.2 LSTM(Long Short-Term Memory)

2.3 Flowchart

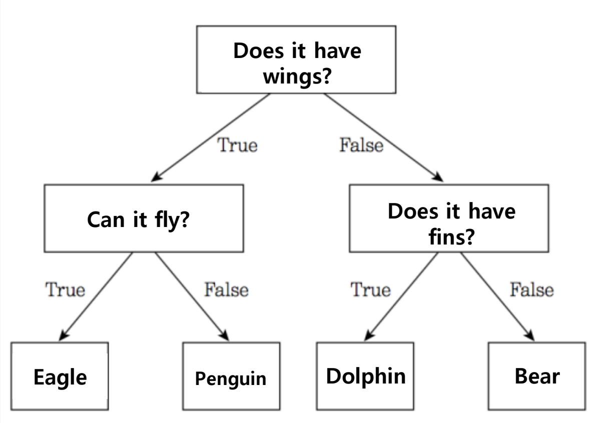


2.1 Decision Tree Models

- A decision tree-based model is a machine learning algorithm used to classify data or predict values, asking questions about the nature of the data at each "node" and iterating until it reaches the final "leaf" node, providing a prediction

RF(Radom Forest)

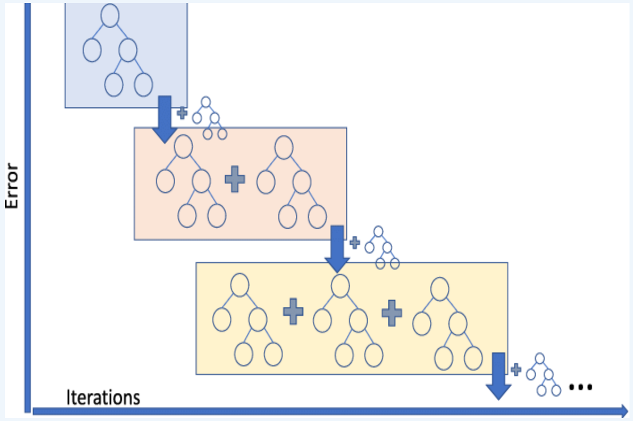
Combines results from multiple decision trees to create more accurate and robust models, with each tree trained independently on random samples of data



https://tensorflow.blog/

GBM(Gradient Boosting Machine)

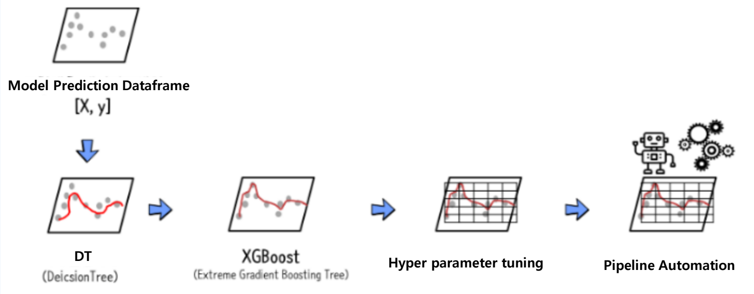
Minimizes error by sequentially training multiple weak prediction models (often decision trees), weighting the errors of the previous models to minimize error



https://medium.com/analytics-vidhya/what-is-gradient-boosting-how-is-it-different-from-ada-boost-2d5ff5767cb2

XGBoost(eXtreme Gradient Boost)

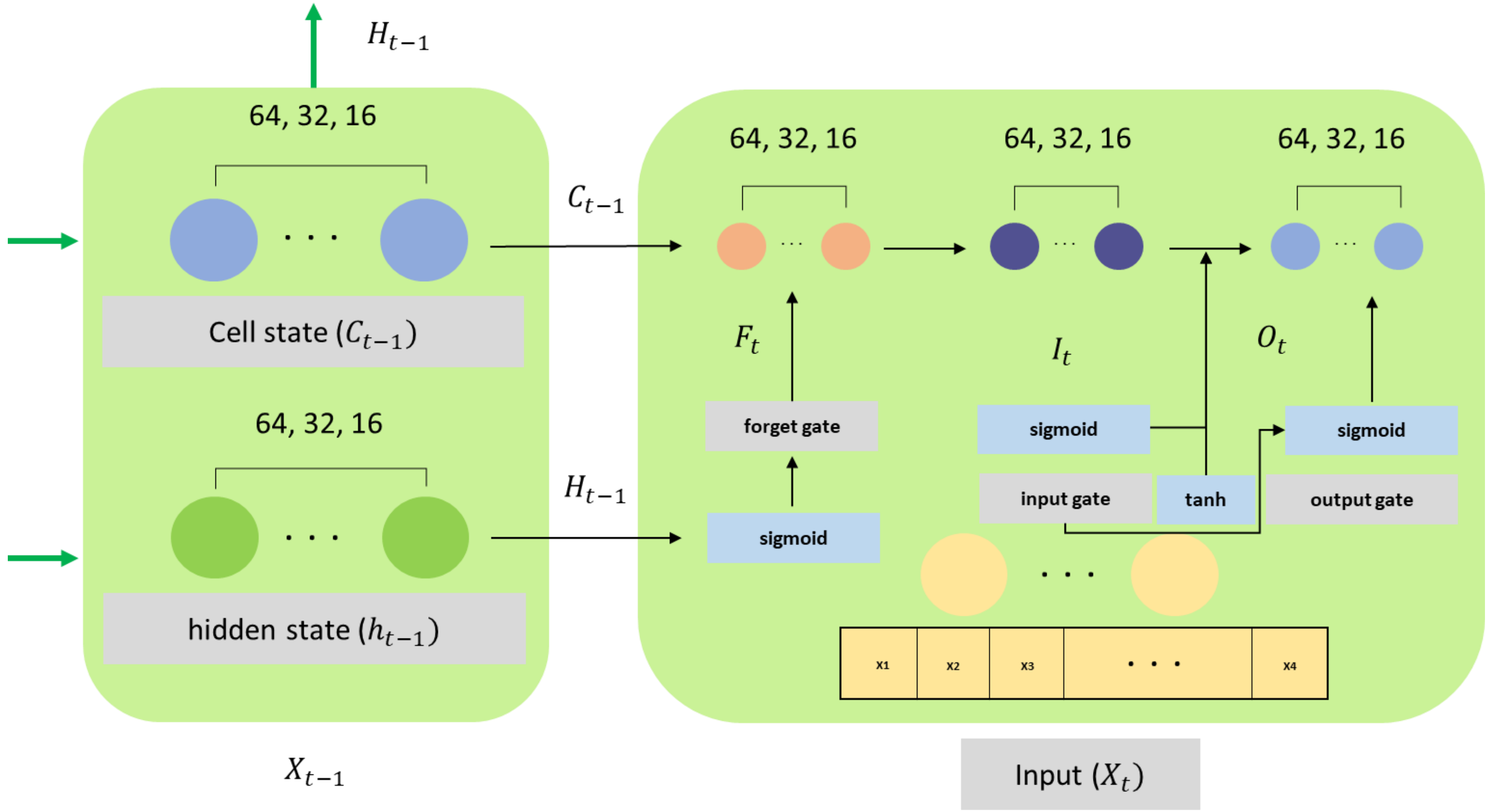
Designed to process large datasets quickly and efficiently, with features like parallel processing, tree pruning, automatic handling of missing values, and normalization to avoid overfitting



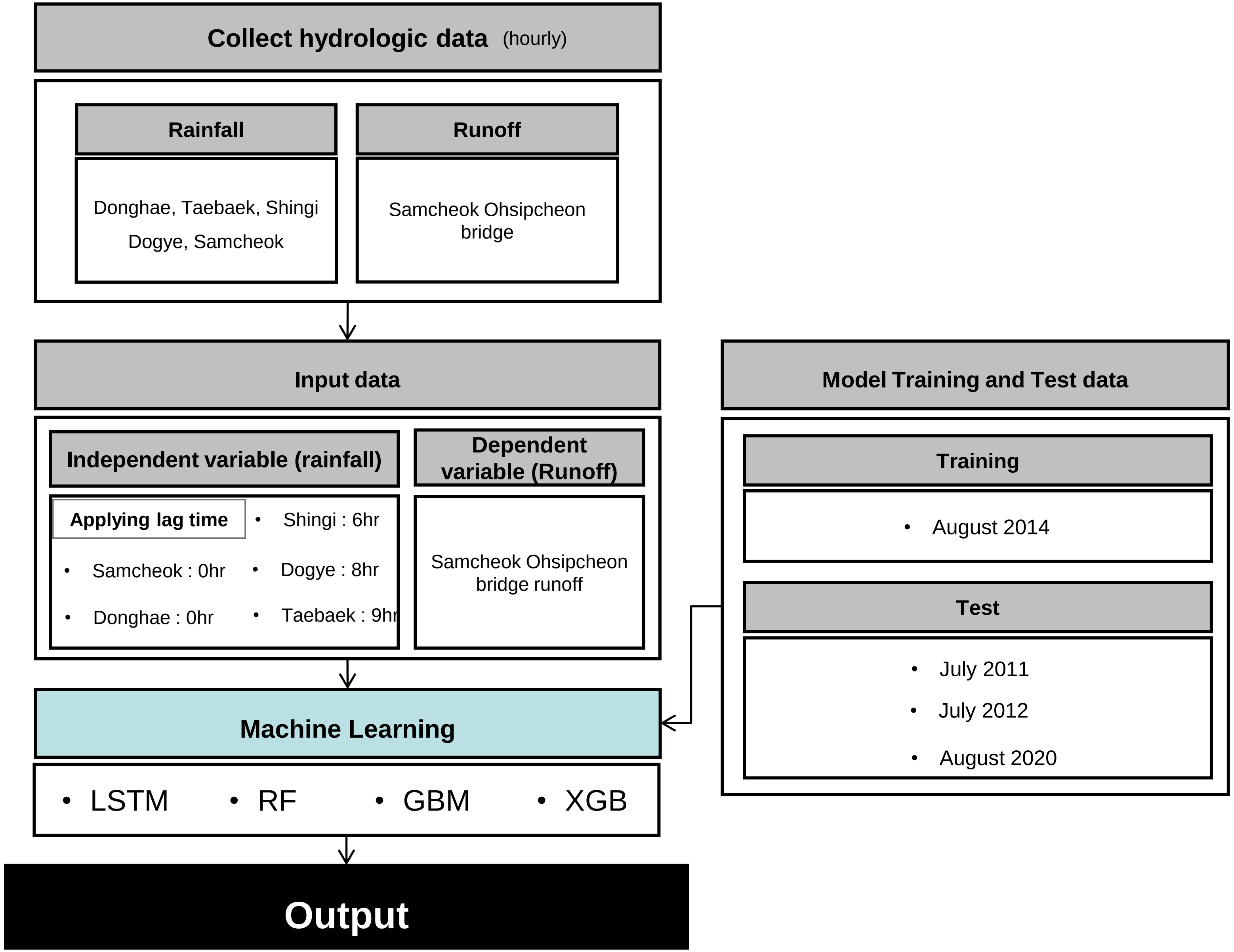
https://statkclee.github.io/model/model-python-xgboost-hyper.html

2.2 LSTM(Long Short-Term Memory)

- Long Short-Term Memory (LSTM) is a model that improves on the long-term dependency problem of RNNs, calculating how much of the past is likely to be forgotten or remembered based on information at the current time, and performs better on longer sequences than RNNs



2.3 Flowchart

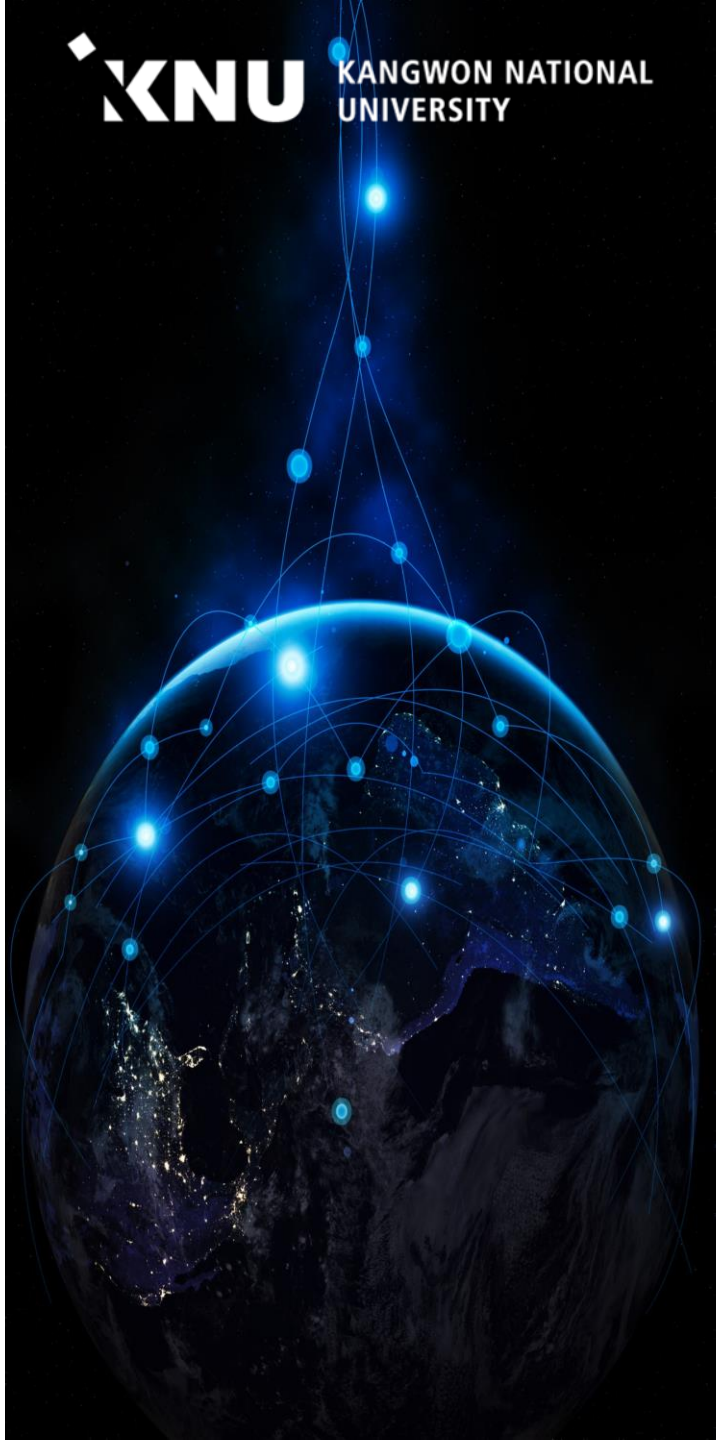


3. Application and results

3.1 Current status of the study area

3.2 Building Inputs and parameter settings

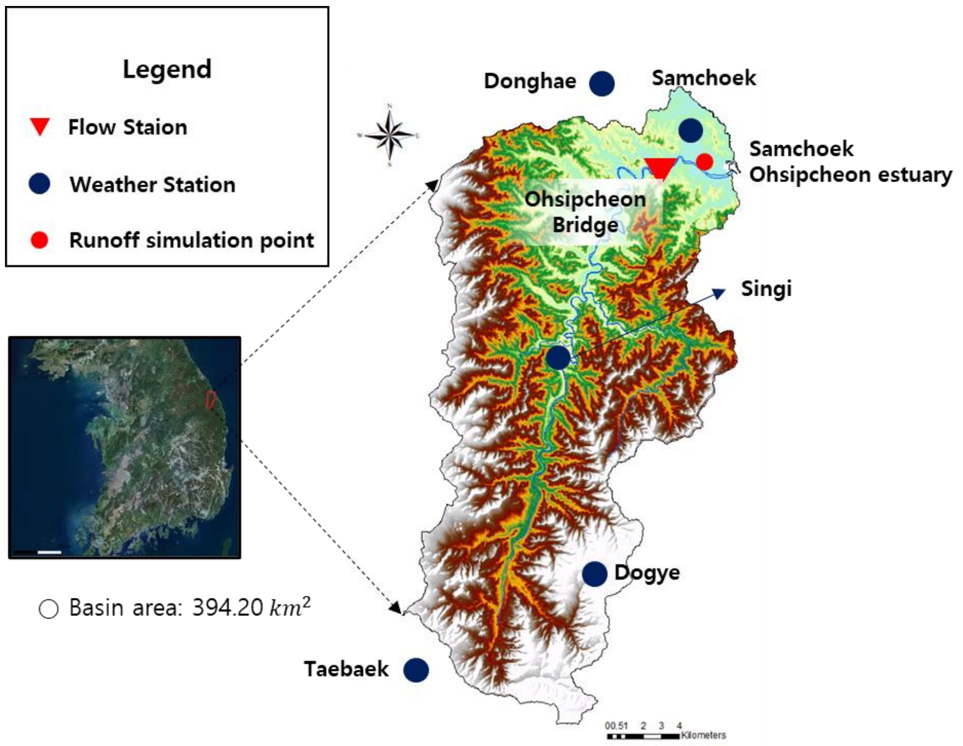
3.3 Model Training and simulation



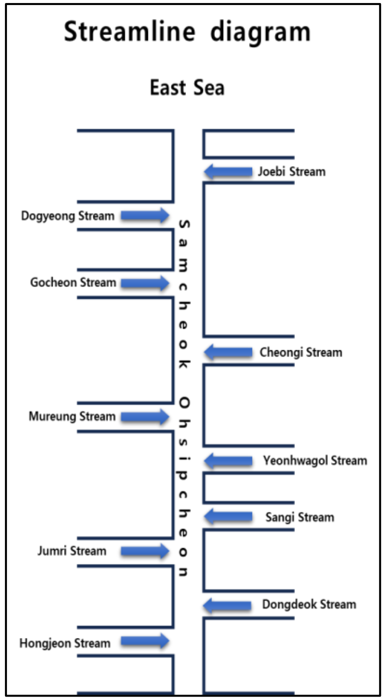
3.1 Current status of the study area

- The Osipcheon watershed in Samcheok-si, Gangwon-do was selected as the target watershed in this study
- There are five weather stations (Donghae, Taebaek, Shingi, Samcheok, and Dogye) and one flow station in the vicinity of Samcheok Osipcheon

Status of rainfall-runoff simulation points and stations



○ Basin area: 394.20 km^2



Weather and flow station status

Station name	Observation start date	X (N)	Y (E)	Note
Donghae	1992-05-01	447418	387813	Weather stations
Taebaek	1985-08-01	409787	376664	
Shingi	2002-12-24	429528	384830	
Samcheok	2003-06-16	441157	391301	
Dogye	2010-07-08	415918	386018	
Samcheok (Ohsipcheon bridge)	2006 1992-05-01	537671	214614	Flow stations

Watershed Area (km^2)	Total Stream Length (km^2)	Planning frequency (years)	Planned flood volume (m^3/s)
394.20	58.10	100	2,760

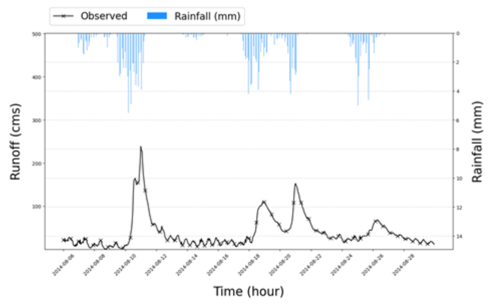
3.2 Building Inputs and parameter settings

- Using August 2014, the month with the fewest missing data, as training material
- This study analyzes the results of runoff simulations for 2011, 2012, and 2020, when Samcheok City was affected by typhoons and heavy rains using artificial intelligence models
- Trained 3 decision tree-based models (XGB, GBM, RF) and LSTM model

Samcheok City Statistical Yearbook

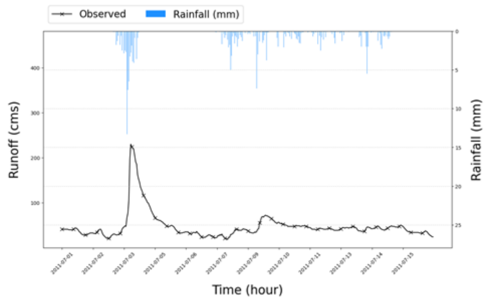
Year	Victims(people)	Damage amount(Dollar)					
		total	Buildings	Ships	Farmland	Public Utilities	Etc.
July 2011	-	4,998,750	-	72,755	-	2,867,286	2,058,708
July 2012	-	1,924,202	2,161	2,370	1,756	1,913,530.26	4,384
August 2014	12	1,108,375	54,034	-	-	180,986	873,354
August 2020	106	10,537,554	1,510,201	36,422	70,260	10,266,507	13,343

Training

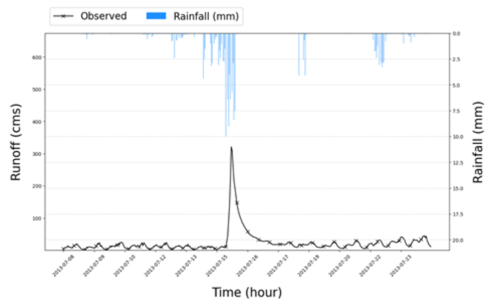


August 2014

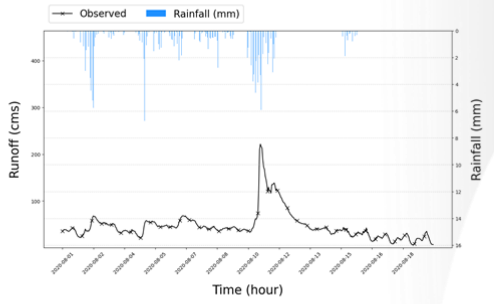
Test



July 2011



July 2012



August 2020

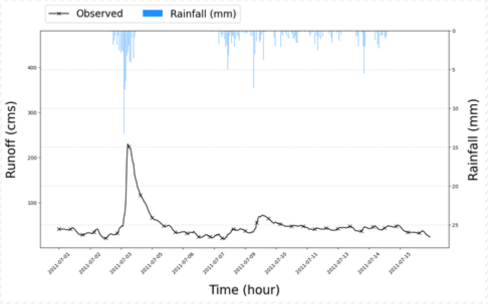
XGB
GBM
RF

LSTM

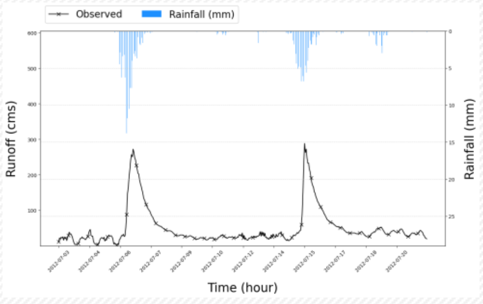
3.2 Building Inputs and parameter settings

- Collected hourly observed runoff data from the Han River Flood Control Center of the Ministry of Environment for July 2011, July 2012, August 2014, and August 2020 at Samcheok Osipcheon Bridge
- Collected hourly observed rainfall from five meteorological observation stations (Taebaek, Dogye, Shingi, Samcheok, and Donghae) from the Korea Meteorological Administration

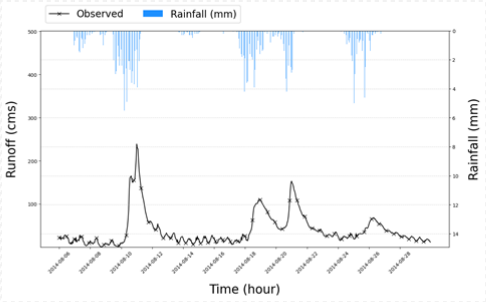
Time Series



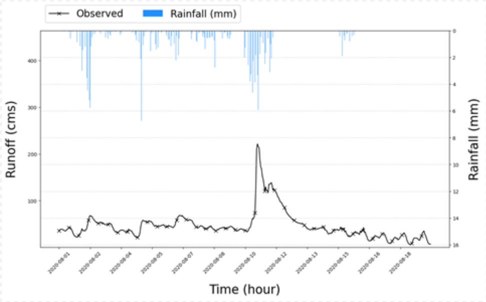
July 2011



July 2012



August 2014



August 2020

Rainfall events

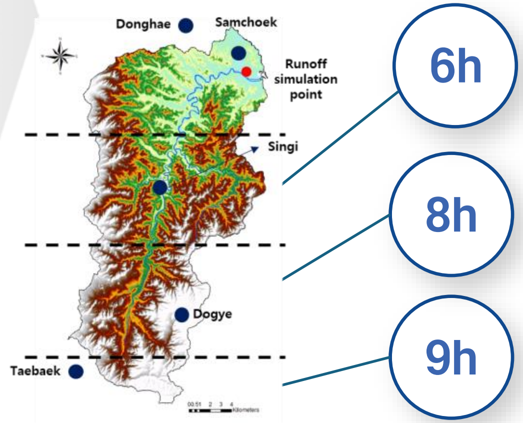
	Rainfall duration (hour)	Peak flood volume (cms)	Cumulative rainfall (mm)
July 2011	169	230.24	192.18
July 2012	154	288.23	285.4
August 2014	284	238.82	336.12
August 2020	194	221.38	200.26

Samcheok Osipcheon bridges runoff and weather station 1-hour statistics(2011 Jul, 2012 Jul, 2014 Aug, 2020Aug)

Station name	Count	mean	std	min	max	Var	Note
Donghae	2976	36.73	33.75	1.06	288.23	1139.00	Weather Station
Taebaek		0.43	1.72	0.00	31.00	2.95	
Samcheok		0.47	1.88	0.00	29.50	3.53	
Dogye		0.38	1.67	0.00	26.67	2.78	
Shingi		0.44	1.77	0.00	33.00	3.15	
Ohsipcheon bridge		0.38	1.51	0.00	27.50	2.27	Flow Station

Applying lag time

- **Applying the 9-hour arrival time** of the Samcheok-Osibcheon Estuary branch of the Samcheok-Osibcheon River Basic Plan to reflect basin characteristics
- Considering the location of the weather stations, **apply 9, 8, and 6 hours as the lag time** for Taebaek, Dogye, and Shingi stations, respectively



3.2 Building Inputs and parameter settings

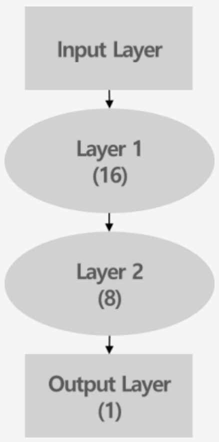
- Hyperparameterized XGB, GBM, and RF models using GridSearch
- Set the training number with the least loss to the optimal training number for the LSTM model

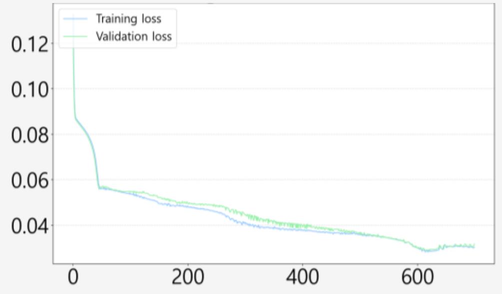
XGB	
colsample_bytree	0.7
learning_rate	0.05
max_depth	3
n_estimators	50
subsample	1

GBM	
learning_rate	0.05
max_depth	3
n_estimators	50
subsample	0.7

RF	
max_depth	5
max_features	auto
min_samples_leaf	1
min_samples_split	2
n_estimators	100

<table>
<tr><th colspan="6">LSTM</th></tr>
<tr><th colspan="2">Layer 1
(LSTM)</th><th colspan="2">Layer 2
(LSTM)</th><th colspan="2">Output Layer
(Dense)</th></tr>
<tr><td>units</td><td>16</td><td>units</td><td>8</td><td>units</td><td>1</td></tr>
<tr><td>activation</td><td>tanh</td><td>activation</td><td>tanh</td><td rowspan="2">activation</td><td rowspan="2">relu</td></tr>
<tr><td>l2 regularizer</td><td>0.0001</td><td>l2 regularizer</td><td>0.0001</td></tr>
<tr><td rowspan="2">Compile</td><td>epoch</td><td>608</td><td>loss</td><td colspan="2">MSE</td></tr>
<tr><td>optimizer</td><td>Adam</td><td>learning rate</td><td colspan="2">0.00004</td></tr>
</table>





LSTM Loss Graph

3.2 Model evaluation metrics

- Statistically compare the margin of error between ground truth and model predictions with Mean Error (ME), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)
- Evaluate performance for peak flood volumes with Normalized Peak Error (NPE)

ME

$$ME = \frac{1}{n}\sum_{i}^{n}(Q_{oi} - Q_{si})$$

RMSE

$$RMSE = \sqrt{\frac{\sum_{i}^{n}(Q_{oi} - Q_{si})^2}{n}}$$

MAE

$$MAE = \frac{\sum_{i}^{n}|Q_{oi} - Q_{si}|}{n}$$

NPE

$$NPE = \frac{|Q_{si}^{max} - Q_{oi}^{max}|}{Q_{o}^{max}}$$

Q_{oi} = Observed runoff

Q_{si} = Simulated runoff

n = Number of data

Q_{oi}^{max} = Maximum value of observed outflow

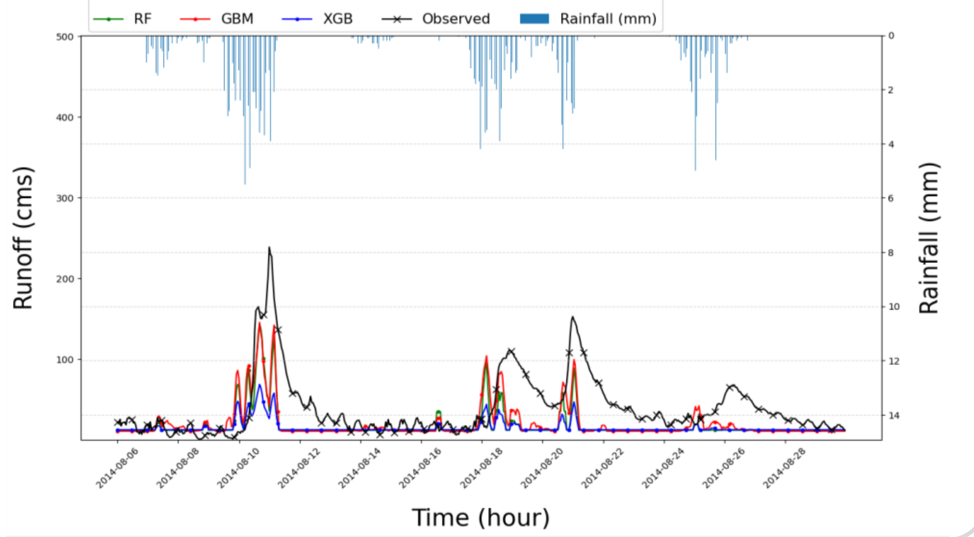
Q_{si}^{max} = Maximum value of simulated outflow

3.3 Model Training and simulation(Training)

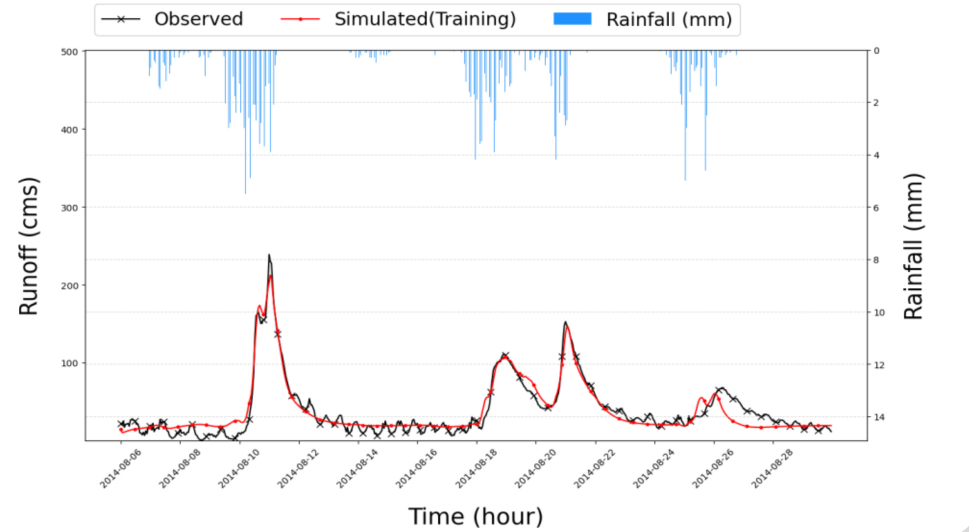
- **Since the DT model is not a model that can consider time series characteristics**, a **moving average was applied** to the simulation results with a window size of 9 hours for the Samcheok Osipcheon River Maintenance Basic Plan
- GBM has the lowest NPE among the three DT models
- The training results of the LSTM model showed ME, MAE, and RMSE of 0.9, 7.6, and 10.2, respectively, and NPE, the error for the peak flood volume, was 0.1

	Training			
	ME	MAE	RMSE	NPE
XGBoost	26.5	26.5	41.5	0.7
GBM	18.7	25.1	36.5	0.3
Random Forest	21.1	25.5	38.0	0.4
LSTM	0.9	7.6	10.2	0.1

DT (RF, XGB, GBM)

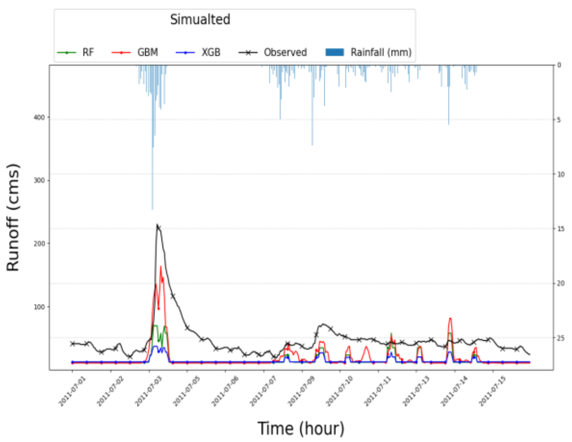


LSTM

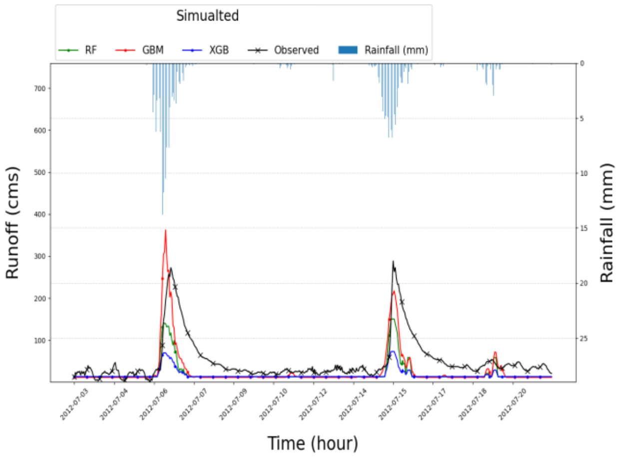


3.3 Model Training and simulation(Test)

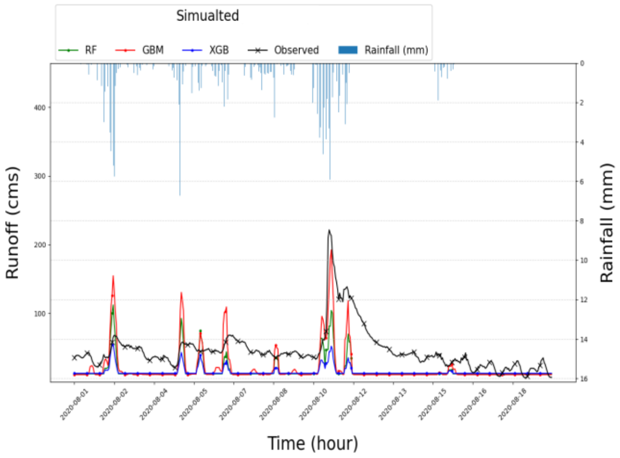
DT (RF, XGB, GBM)



July 2011

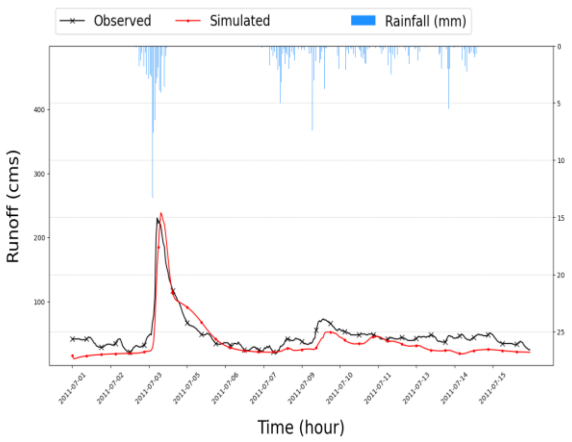


July 2012

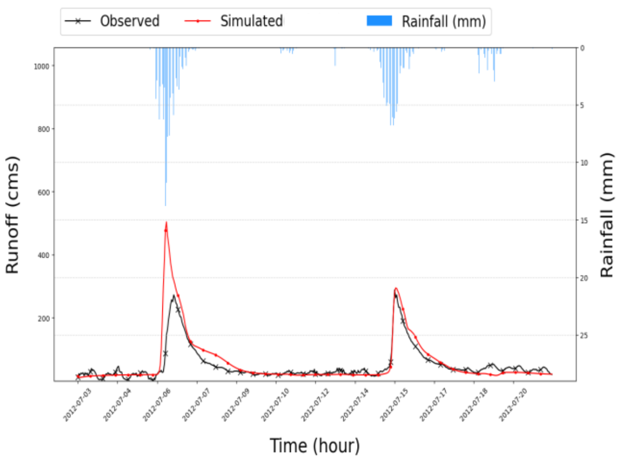


August 2020

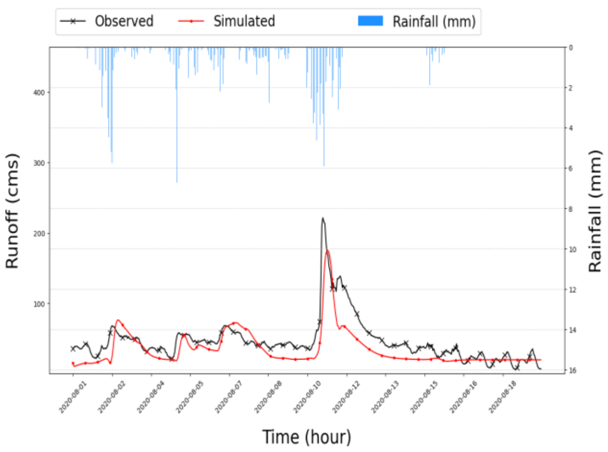
LSTM



July 2011



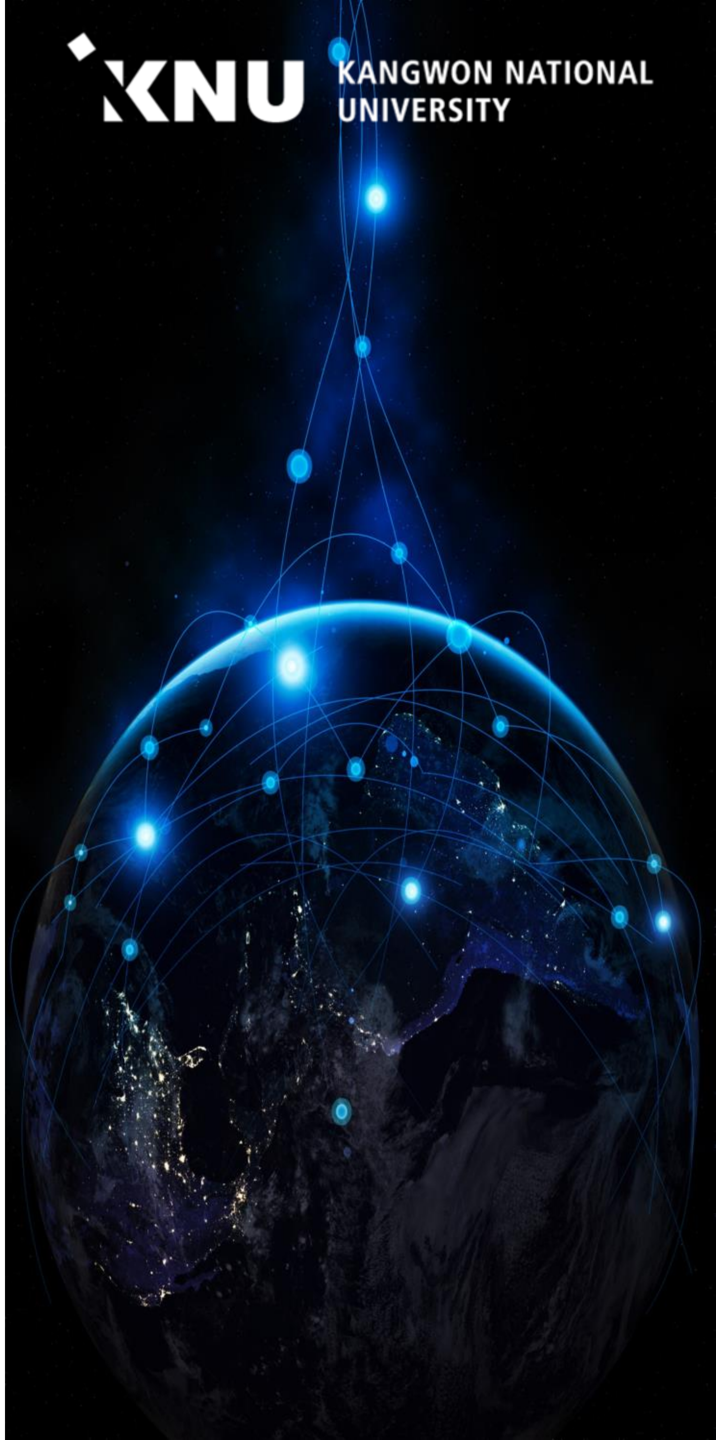
July 2012



August 2020

	Test July 2011			
	ME	MAE	RMSE	NPE
XGBoost	33.3	33.3	43.3	0.8
GBM	28.1	29.3	35.1	0.2
RF	31.9	32.2	40.8	0.7
LSTM	10.4	14.1	17.3	0.03
	Test July 2012			
	ME	MAE	RMSE	NPE
XGBoost	33.4	34.2	56.1	0.7
GBM	23.5	30.2	43.2	0.2
RF	29.4	31.3	47.2	0.4
LSTM	-7.6	18.0	42.6	0.7
	Test August 2020			
	ME	MAE	RMSE	NPE
XGBoost	31.5	31.7	42.3	0.7
GBM	25.5	30.2	37.7	0.1
RF	29.1	30.8	39.9	0.4
LSTM	11.7	15.7	21.4	0.2

4. Conclusion



4 Conclusion

- To reflect the characteristics of the watershed, 9, 8, and 6 hours were applied as **lag time** for Taebaek, Dogye, and Shingi stations, respectively
- **August 2014 was used as training data** for XGB, GBM, RF, and LSTM models, and **July 2011, July 2012, and August 2020 were used as tests**
- Model training results show that the decision tree-based model simulates runoff based on rainfall events more closely than the observed runoff. overall, the LSTM model simulates the peak flood volume and flood reach better than the decision tree-based model
- The results of this study confirm the performance of the LSTM model in simulating runoff using 1-hour observed rainfall, and it is expected that it can be used for ungauged watersheds and real-time flood forecasting with future plans to utilize multiple meteorological observables