

# LSTM 기반 딥러닝 기법을 이용한 섬진강 구례교 지점의 홍수위 예측

## Flood Stage Forecasting at the Gurye-Gyo Station in Sumjin River Using LSTM-Based Deep Learning Models

정재원\* · 모혜림\*\* · 이준형\*\*\* · 유영훈\*\*\*\* · 김형수\*\*\*\*\*

Jung, Jaewon\*, Mo, Hyelim\*\*, Lee, Junhyeong\*\*\*, Yoo, Younghoon\*\*\*\*, and Kim, Hung Soo\*\*\*\*\*

### Abstract

Instances of flood damage caused by extreme storm rainfall due to climate change and variability have been showing an increasing trend. Particularly, a flood forecasting and warning system has been recognized as an important nonstructural measure for flood damage reduction, including loss of life. Flood forecasting and warning have been performed by the forecasts of flood discharge and flood stage using the physically based rainfall-runoff models. However, recently, studies involving the application of a machine learning-based flood forecasting models, which addresses the limitations of extant physically based flood stage forecasting models, have been performed. We may require various case studies to determine more accurate methods. Therefore, this study performed the real-time forecasting of the river water level or stage at the Gurye station of the Sumjin river with lead times of 1, 3, and 6 h by applying a long short-term memory (LSTM)-based deep learning model. In addition, the applicability of the LSTM model was evaluated by comparing the results with those from widely used models based on support vector machine and multilayer perceptron. Consequently, we noted that the LSTM model exhibited a relatively better forecasting performance. Therefore, the applicability of the LSTM model should be extensively studied for flood forecasting applications.

**Key words :** Machine Learning, Deep Learning, LSTM, Flood Stage Forecasting

### 요 지

기후변화와 기후변동으로 인한 극한 호우의 발생으로 홍수 피해는 증가하는 추세에 있다. 특히, 인명 손실을 포함한 하천홍수 피해 저감을 위해 비구조적 방법인 홍수예경보시스템은 매우 중요하게 인식되고 있으며, 홍수예경보는 물리적 모형인 강우-유출 모형에 의한 홍수량 및 홍수위 예측을 통해 이루어져 왔다. 그러나 최근에는 기존 물리적 기반의 홍수위 예측 모형의 단점을 보완한 머신러닝 기반의 홍수위 예측 모형을 적용한 연구들이 진행되고 있으나 다양한 사례에 대한 적용 연구가 지속적으로 필요한 실정이다. 따라서 본 연구에서는 Long Short-Term Memory (LSTM) 기반의 딥러닝 기법을 적용하여 섬진강 수계 구례 지점의 하천 수위를 1, 3, 6시간 선행 예측하고, 선행 예측 시간에 따른 예측 정확도를 비교 분석하고자 한다. 또한 기존의 서포트벡터머신(SVM) 모형, 다층퍼셉트론(MLP) 모형을 적용한 결과와의 비교 분석을 통해 LSTM 모형의 적용성을 평가하고자 하였다. 평가 결과, LSTM 모형이 다른 모형 보다 우수한 예측 성능을 보이고 있음을 알 수 있었으며, 향후 홍수 예측을 위해 LSTM 모형의 적용성을 보다 심층적으로 연구할 필요가 있다.

**핵심용어 :** 머신러닝, 딥러닝, LSTM, 홍수위 예측

\*정회원, 인하대학교 수자원시스템연구소 선임연구원(E-mail: jungjw89@gmail.com)

Member, Senior Researcher, Institute of Water Resources System, Inha University

\*\*승실대학교 생명정보학과 석사과정(E-mail: mohyelim7@gmail.com)

Master's Course, Department of Bioinformatics and Life Science, Soongsil University

\*\*\*정회원, 인하대학교 스마트시티공학과 박사과정(E-mail: lee\_junhyeong@naver.com)

Member, Ph.D. Candidate, Department of Civil Engineering, Inha University

\*\*\*\*정회원, 인하대학교 스마트시티공학과 박사과정(E-mail: dudgns5971@naver.com)

Member, Ph.D. Candidate, Department of Civil Engineering, Inha University

\*\*\*\*\*교신저자, 정회원, 인하대학교 사회인프라공학과 교수(Tel: +82-32-872-8729, Fax: +82-32-876-9787, E-mail: sookim@inha.ac.kr)

Corresponding Author, Member, Professor, Department of Civil Engineering, Inha University

## 1. 서론

최근 이상기후로 발생에 따라 집중호우의 발생 빈도 및 강도가 증가하였고 이로 인한 홍수 피해가 매년 증가하고 있다. 행정안전부의 통계자료에 의하면 최근 10년간(2009-2018) 호우로 인한 피해액은 연간 약 1천 5백억 원으로 태풍, 대설, 지진, 풍랑, 강풍 등의 원인들 중 가장 큰 피해액이 발생한 것으로 나타났다(MOIS, 2019). 또한 지난 2020년 한반도 전역에 발생한 집중호우로 인해 31명이 사망하고 약 7,000명의 이재민이 발생하였으며, 3만 건 이상의 시설과 2만 6천여 헥타르의 농경지가 피해를 입기도 하였다. 특히 섬진강 일부 지역에서는 급격히 증가하는 하천 수위를 예측하지 못하여 인명피해를 비롯한 극심한 피해가 발생하였다. 또한 이러한 하천 범람은 침수피해 뿐 아니라 주요 기반 시설물의 손상으로 인해 교통 및 통신 두절 등 주변 지역으로의 2차 피해를 연쇄적으로 유발한다. 따라서 이러한 피해를 줄이기 위한 연구가 시급한 실정이다(Park and Kim, 2020; Jeong et al., 2019).

실시간 수위 예측을 통한 예경보 시스템은 극한 호우 발생시 실제 차량 및 보행자의 통행을 제한함으로써 인명 및 재산 피해를 막을 수 있는 가장 실효성 있는 비구조적 대책이라 할 수 있다(Jung et al., 2018; Liu et al., 2018). 국내의 경우 「하천법시행규칙 제 23조 제 2항」에 의하여 한국 홍수통제소에서 주요 지점 65개소에 대해 홍수예경보를 발령하고 있으며, 이때 Navier-Stokes 방정식을 비롯한 미분방정식을 해석하는 수치모형인 저류함수모형을 통해 하천 수위를 예측하고 있다(Yoo et al., 2019; Lee, 2021). 하지만 저류함수모형의 경우 수문 및 지형 자료가 충분하지 않아 국내 유역의 특성을 고려한 객관적이고 합리적인 매개변수 추정방법이 제시되어 있지 않다는 한계가 지적되었다(Song et al., 2006; Kim et al., 2008; Choi et al., 2010; Park et al., 2021; Lee, 2021). 그 간의 이러한 문제점을 해결하고 유역 특성에 맞는 최적의 매개변수를 산정하기 위한 많은 노력의 결과로, 예경보 지점 65개소에 대해서는 홍수 예측의 믿을 만한 결과를 얻을 수 있게 되었다. 하지만 이는 대유역 중심의 하천 홍수를 대상으로 개발된 것이기 때문에 대부분의 예경보 지점이 하천 본류에 위치하고 있어 그 외 지점들에 대한 실시간 예측은 어려운 실정이다.

이를 해결하기 위한 수위 예측 연구 사례를 보면, 크게 개념적 모형, 물리적 기반 모형, 자료 기반 모형을 적용한 연구로 나눌 수 있다(Mosavi et al., 2018; Zounemat-Kermani et al., 2020). 앞의 두 모형을 적용하는 경우에는 정확한 예측을 위해 필요로 하는 기상학적, 지형학적, 지질학적 매개변수가 많아 가용한 정보가 없을 시 적용하기 어렵다는 단점이 있다. 또한 물리적 기반 모형의 경우에는 구축 및 모의에 상대적으로 많은 시간이 소요되기 때문에 후속조치를 위한 선행시간 확보가 필수인 홍수 예보에서 그 활용이

제한적일 수 있다(Jung et al., 2018). 반면 자료 기반 모형의 경우 입출력간의 관계가 블랙박스 처리되어 유역 및 하도에서 발생하는 물리적인 수리수문학적 특성을 반영할 수 없다는 단점이 있으나, 비교적 적은 양의 변수를 통해서도 예측이 가능하다는 장점을 가진다. 또한 특히 하천 수위와 같이 비선형적 특성을 가지는 자료를 예측하는데 있어서 우수한 예측 성능을 발휘하는 것으로 나타나 최근 컴퓨팅 성능 및 기술 발전과 더불어 많은 연구에 적용되고 있다(Maier and Dandy, 2000).

실제 이러한 자료기반 모형, 소위 머신러닝 모형을 적용한 연구사례를 보면 주로 인공신경망 모형(Artificial Neural Network, ANN), 뉴로-퍼지 추론 시스템(Adaptive Network-based Fuzzy Inference System, ANFIS), 서포트 벡터 머신(Support Vector Machine, SVM), 그라디언트 부스트(Gradient Boost, GB), 다층 퍼셉트론(Multilayer Perceptron, MLP) 등을 이용한 수위예측 연구가 수행되어왔다(Thirumalaiah and Deo, 1998; Suykens and Vandewalle, 1999; Supharatid, 2003; Chiang et al., 2010; Chen et al., 2012; Hidayat et al., 2014; Dineva et al., 2014; Gizaw and Gan, 2016; Kim et al., 2016; Kim and Tachikawa, 2017; Mosavi and Edalatifar, 2018; Taherei Ghazvinei et al., 2018; Yoo et al., 2019). 또한 근래에 들어서는 심층신경망 구조를 갖는 딥러닝 기반의 순환신경망 모형(Recurrent Neural Network, RNN)과 비선형 자기회귀 신경망 모형(Nonlinear autoregressive network with exogenous inputs, NARX)와 같이 발전된 형태의 머신러닝 모형을 적용한 연구들이 수행되고 있다(Lee and Tuan Resdi, 2016; Yoo et al., 2019; Lee et al., 2020). Yoo et al. (2019)은 기존 ANN 모형을 활용하여 수위예측을 한 연구들이 시간적 매개변수를 고려하지 못하였다는 한계를 지적하였고, 이를 보완할 수 있는 NARX 모형을 적용하여 한강대교의 수위를 예측하였다. 하지만 NARX모형의 경우 입력자료의 길이가 길 때 발생하는 기울기 소실(Vanishing gradient) 문제를 한계로 지적하였고, 이를 해결하기 위해 개발된 Long-Short Term Method (LSTM) 모형을 적용할 것을 제안하였다. LSTM 모형을 적용한 수위 예측 연구의 경우는 최근 몇 년 동안 수행되었으나 다양한 사례에 대한 적용 연구가 지속적으로 필요한 실정이다(Jung et al., 2018; Lee et al., 2018; Yoo et al., 2019; Song et al., 2020; Park and Kim, 2020).

따라서 본 연구에서는 LSTM 모형을 적용하여 섬진강 수계 구례 지점의 하천 수위를 1, 3, 6시간 선행 예측하고, 선행 예측 시간에 따른 예측 정확도의 비교 분석하고자 한다. 또한 기존의 SVM 모형, MLP 모형을 적용한 결과와의 비교 분석을 통해 LSTM 모형의 적용성을 평가하고자 한다.

## 2. 적용 모형

### 2.1 서포트 벡터 머신

서포트 벡터 머신(SVM)은 Vapnik (1995)에 의해 제시된 방법이며, 선형으로 서로 다른 클래스의 벡터들을 그 사이의 거리에 대한 최대의 마진(margin)으로 분류할 수 있는 서포트 벡터들로 이루어진 초평면을 찾는 방법이다(Lee et al., 2016; Choi and Moon, 2013). 선형으로 분류 불가능한 데이터에 대해서는 커널함수를 이용하여 평면에서 고차원 공간에 사상시킨 뒤 분류하는데, 대표적인 커널함수의 종류로는 polynomial, sigmoid, Radial Basis Function (RBF) 등이 있다(Park, 2006). SVM에  $\epsilon$ -무감도 손실함수( $\epsilon$ -insensitive loss function)를 도입하여 회귀(regression) 분석에 사용할 수 있도록 확장한 방법을 서포트 벡터 회귀(Support Vector Regression, SVR)라 한다(Kim et al., 2012). 즉, SVM은 분류 문제에서 데이터를 '+1' 클래스와 '-1' 클래스로 구분하는 데 사용되지만, SVR은 임의의 실수값을 예측할 수 있도록 SVM을 일반화한 방법이라 할 수 있다(Lee, 2021). SVR 모형 구조의 모식도는 다음 Fig. 1과 같다.

### 2.2 다층 퍼셉트론

ANN은 시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)가 학습을 통해 시냅스의 결합 세기를 변화시키며 문제를 해결하는 모형을 말한다(Kang and Lee, 2011). ANN 모형에는 여러 구조가 있으며 퍼셉트론은 입력층, 출력층, 가중치 및 임계치 벡터로 구성된 구조를 말한다(Hong and Ahn, 2021). 그 중에서 다층 퍼셉트론(MLP)은 입력층과 출력층 사이에 하나 이상의 은닉층이 존재하는 ANN 모형으로, 기존의 ANN이 비선형 문제는 해결 못하는 한계를 해결하기 위해 개발되었다. MLP의 입력층은 다수의 입력 데이터를 받고, 출력층은 데이터의 출력을 담당한다. 은닉층은 입력층과 출력층 사이에서 두 층을 연결시킨다. 은닉층의 층수와 각 층의 노드 개수를 설정하여 모델을 구성한다(Lee et al., 2013). MLP 모형 구조의 모식도는 다음 Fig. 2와 같다.

### 2.3 Long Short-Term Memory (LSTM)

LSTM은 시계열 데이터에서 직접 학습하는 RNN의 일종이다(Han et al., 2021). 일반적으로 RNN은 기존 알고리즘과 다르게 이전 단계의 정보를 저장하고 다음 단계에서 이를 입력으로 사용함으로써 긴 시퀀스에 대해 예측이 가능하기 때문에 연속적인 데이터의 분석에 주로 사용된다. RNN에서 발생하는 오차경사의 기울기 소실 및 최적화 오류를 해결하고 입력 자료에 대한 정보를 더욱 장기적으로 기억하기 위해 은닉층에 셀 상태(cell state)구조를 추가한 모형이다(Jung et al., 2018).

다음 Fig. 3은 LSTM 구조의 모식도이다. LSTM에서는 입력게이트(Input gate,  $i_t$ ), 망각게이트(forget gate,  $f_t$ ), 출력

게이트(output gate,  $o_t$ )가 있어 전 단계의 정보 중에서 버릴 것과 업데이트 할 것을 결정하여 다음 단계로 전달한다.

먼저 망각게이트( $f_t$ )는 어떤 정보를 버릴 것인가에 대한 계산을 하며, 이전 단계의  $h_{t-1}$ 과 현재 단계의  $x_t$ 를 시그모이드 함수에 적용하여 0과 1사이의 값을 산정한다. 이를 현재 상태와 곱해지게 되며, 이 과정에서 정보를 사용할지 제거할지를 선택한다.

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (1)$$

여기서  $\sigma$ 는 활성화함수,  $W_f$ 는 망각게이트의 가중치,  $b_f$ 는 편향을 나타낸다.

다음으로 입력게이트( $i_t$ )라 불리는 시그모이드 함수가

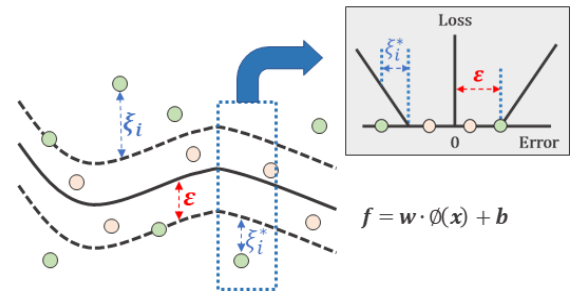


Fig. 1. SVR Structure

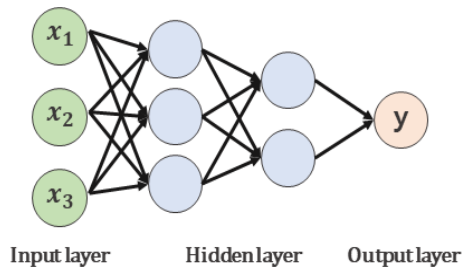


Fig. 2. MLP Structure

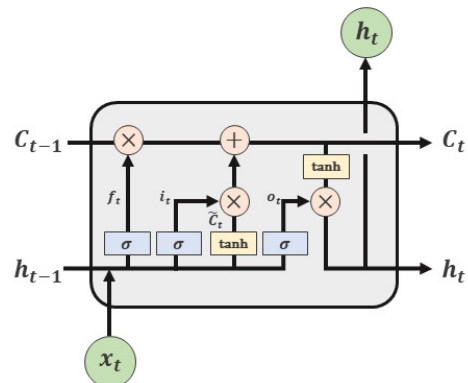


Fig. 3. LSTM Structure

어떤 데이터를 업데이트 할지를 결정한다. 하이퍼탄젠트 함수(tanh)를 통해 새로운 후보 값들인  $\tilde{C}_t$  벡터를 만들게 되고,  $i_t$  값과 조합하여 셀 상태에 더한다.

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C) \quad (3)$$

Eqs. (1)~(3)의 결과들을 이용하여, 과거 상태인  $C_{t-1}$  을 업데이트해서 새로운 상태  $C_t$  를 만든다. Eq. (4)를 이용하여 현재 단계에 대한 정보 상태를 업데이트 할 수 있다.

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

마지막으로 출력게이트( $o_t$ )를 통해 최종적인 결과가 도출되며, 셀 상태의 어느 부분을 도출할지 결정하는 단계로 볼 수 있다. 이를 위해 Eq. (5)를 이용하고, Eq. (6)을 통해 현재 셀 상태에 대한 업데이트를 수행한다.

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

### 3. 모형 적용 및 분석

#### 3.1 대상 유역 및 자료 수집

대상 유역은 지난해 큰 피해를 입은 섬진강 유역으로, 중류에 위치한 구례군(구례교) 지점의 수위를 예측하였다. 섬진강 유역은 국내 5대강 유역으로 전라북도과 전라남도, 경상남도의 세 개도에 걸쳐있고 유역면적으로는 네 번째 큰 유역이지만, 다른 하천에 비하여 산업화 과정에서 상대적으로 소외되어왔다(Jung et al., 2020). 또한 상대적으로 치수 능력이 부족하여 홍수피해에 대해 가장 취약할 것으로 평가된 바 있다(Lee et al., 2011). 구례군(구례교) 수위 관측 지점은 섬진강의 중류부에 위치하여 조위의 영향을 받지 않으며, 상류에 구례군(송정리) 수위 관측 지점이 위치하고 있어 상류 수위자료를 예측에 사용할 수 있다. 또한 인근 강우 관측소 중 관측 기간이 15년 이상인 남원 지점의 강수량 자료를 예측에 활용할 수 있기 때문에 예측 대상 지점으로 선정하였다(Fig. 4).

구례군(구례교)과 구례군(송정리) 지점에 대해 2006년부터 2020년까지의 총 15년간의 10분 단위 수위자료를 수집하였고, 동일한 기간의 남원 지점의 강수량 자료를 수집하였다. 기간 내의 홍수기 자료인 6월에서 9월 자료를 분석에 사용하였으며, 모형의 학습을 위한 훈련자료로는 2006년부터 2017년까지 12년간의 자료를 사용하였으며, 2018년부터 2020년까지의 3년 기간을 검증 및 시험자료로 사용하여 구례군(구

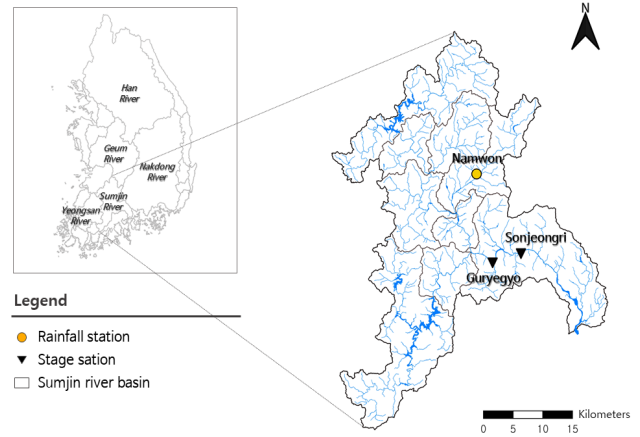


Fig. 4. Location Map of the Target Basin and Stations

례교)의 실제 수위 관측과 비교하여 모형의 적용성을 평가하였다.

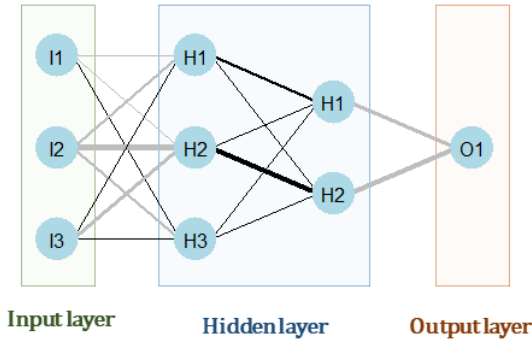
#### 3.2 모형 구축

##### 3.2.1 SVR 모형 구축

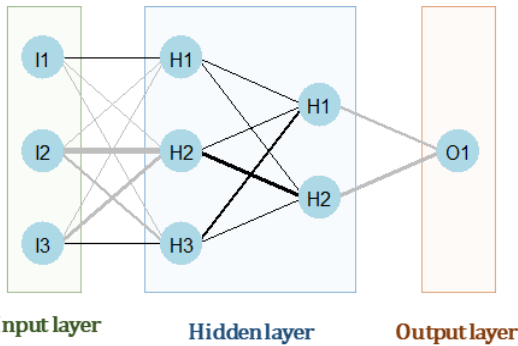
본 연구에서는 구례군(구례교)의 수위를 선행시간 1, 3, 6시간별로 예측하고 예측정확도를 비교 분석하는 것을 목적으로 한다. 변수로 사용되는 시계열 자료는 정규분포를 따르는 것으로 가정하여 평균을 빼고 표준편차로 나누는 standard scaler 변환을 통해 정규화하여 사용하였다. SVR 모형 구축 시 커널함수(Kernel function)로는 선행 연구사례에서 성능이 검증된 Radial basis function을 적용하였다(Tay and Cao, 2001; Park and Kim, 2007; Kim and Ahn, 2010; Lesmeister, 2017; Choi et al., 2019). 모형의 하이퍼파라미터인 C값과  $\epsilon$  (epsilon)값은 랜덤하게 선택하여 각각 최적의 결과값을 가지는 0.01과 0.2로 설정하였고,  $\gamma$  (gamma)값은 기본값인 1로 선택하였다.

##### 3.2.2 MLP 모형 구축

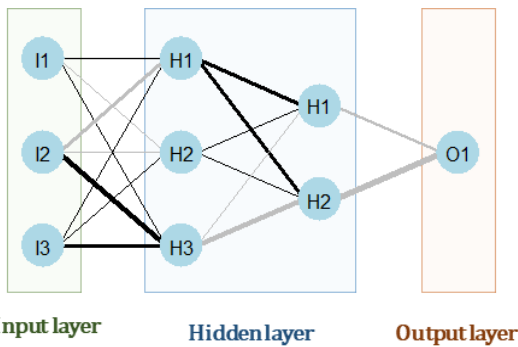
MLP 모형은 입력층과 출력층 사이에 하나 이상의 중간층인 은닉층(hidden layer)이 존재하는 신경망으로 본 연구에서는 2개의 은닉층으로 다음 Fig. 5와 같이 첫 번째 은닉층에 3개 노드, 두 번째 은닉층에 2개 노드로 구성하였다. 이때 은닉층의 노드(node) 수는 많아질수록 출력시 일반화가 어렵고, 적으면 복잡한 의사결정 경계를 만들기 어렵기 때문에 적절한 노드 개수의 선정이 중요하다. 본 연구에서는 시행착오법(trial and error)을 통해 최적의 개수를 선정하였다. 각 시계열 자료는 min-max scaler 변환을 통해 0~1값으로 정규화하여 사용하였다. 학습시 역전파 알고리즘(Back Propagation Algorithm)을 통해 오차를 최소화하였고, 학습횟수는 기본값인 100회로 학습하였다.



(a) 1-hour lead time



(b) 3-hour lead time



(c) 6-hour lead time

Fig. 5. Visualization of MLP Models

### 3.2.3 LSTM 모형 구축

LSTM 모형의 학습과 예측을 위해서 상류지점인 구례군(송정리), 구례군(구례교)의 10분 단위 수위자료와 선행 5일 강수량 자료를 사용하였다. 이때 1, 3, 6시간 선행 예측을 위해서 자료는 각각 1, 3, 6시간 전의 자료를 학습에 사용하였고, 각 시계열 자료는 min-max scaler을 통해 정규화하였다. 학습시 과적합을 방지하기 위하여 손실함수(loss function)에서 loss의 변화가 더 이상 없거나 오차가 증가할 때 학습을 멈추는 학습 조기 종료(early stopping) 기법을 적용하였다. 또한 오버피팅과 오버피팅 및 언더피팅을 방지하기 위하여 하이퍼 파라미터를 선택적으로 조절하였다. 학습속도

Table 1. Comparisons of ML Models' Performance

| Lead time | Model | Model performance |                |        |
|-----------|-------|-------------------|----------------|--------|
|           |       | NSE               | R <sup>2</sup> | RMSE   |
| 1-hour    | SVR   | 0.6754            | 0.7019         | 0.0216 |
|           | MLP   | 0.9591            | 0.9609         | 0.0315 |
|           | LSTM  | 0.9796            | 0.9815         | 0.0291 |
| 3-hour    | SVR   | 0.6558            | 0.6869         | 0.0373 |
|           | MLP   | 0.9178            | 0.9249         | 0.0657 |
|           | LSTM  | 0.7810            | 0.8042         | 0.0612 |
| 6-hour    | SVR   | 0.6018            | 0.6504         | 0.1056 |
|           | MLP   | 0.8530            | 0.8655         | 0.0856 |
|           | LSTM  | 0.6872            | 0.7862         | 0.2121 |

(learning rate)는 0.00001, 학습횟수(epoch)는 10회로 구성하였으며, 최적화 알고리즘으로는 'Adam'을 사용하고 loss function으로는 평균 제곱 오차(Mean Squared Error)를 선택하였다.

## 4. 모의 결과 및 분석

예측된 수위의 정확도를 정량화하기 위해서는 Nash-Sutcliffe Efficiency (NSE), R<sup>2</sup>, Root Mean Square Error (RMSE)의 세 개의 평가지표를 사용하였고, 이를 통해 1, 3, 6시간 선행 예측 결과를 비교·분석하였다.

$$NSE = 1 - \frac{\sum_{t=1}^N (O_t - P_t)^2}{\sum_{t=1}^N (O_t - \bar{O})^2} \quad (7)$$

$$R^2 = \frac{\sum_{t=1}^N (O_t - \bar{O})(P_t - \bar{P})}{\sqrt{\sum_{t=1}^N (O_t - \bar{O})^2 (P_t - \bar{P})^2}} \quad (8)$$

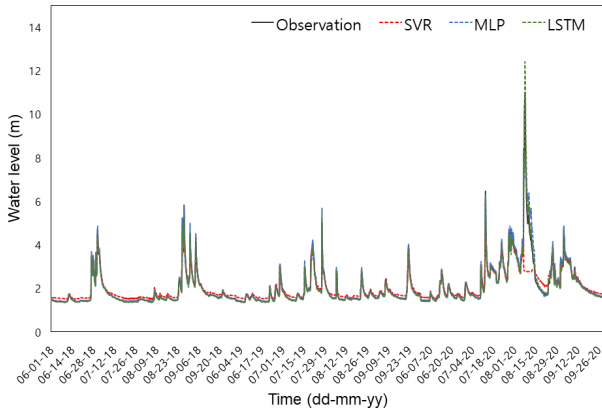
$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (O_t - P_t)^2} \quad (9)$$

여기서,  $O_t$ 와  $P_t$ 는 시간  $t$ 에서의 관측 수위와 예측 수위,  $\bar{O}$ 와  $\bar{P}$ 는 관측 수위와 예측 수위의 평균값,  $N$ 은 자료의 개수를 의미한다.

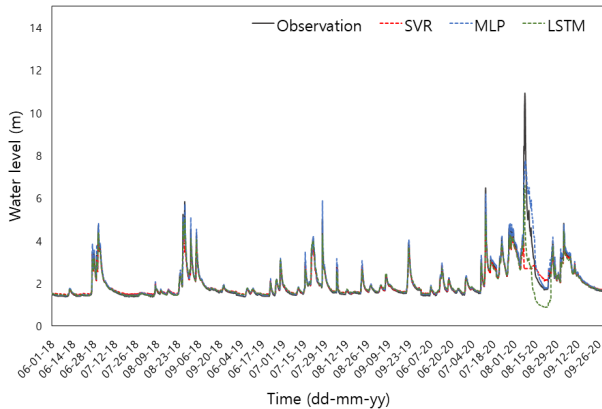
각 모형의 예측 결과와 관측 자료를 비교하여 안정적인 평가지표를 예측모형별, 선행시간별로 정리하였다(Table 1). 이를 통해 모형의 예측 성능을 비교한 결과, 모형간 비교시에는 SVR 모형을 통한 예측 결과의 NSE, R<sup>2</sup> 값이 상대적으로 작고 RMSE 값이 크게 나타났다. MLP와 LSTM

모형의 경우 상대적으로 우수한 예측 성능을 보였으며, 1시간 선행 예측시에는 LSTM 모형이 더 높은 성능을 보였다. 하지만 선행시간별 예측성능의 비교시에는 LSTM 모형의 예측 성능이 큰 폭으로 감소함을 보였다.

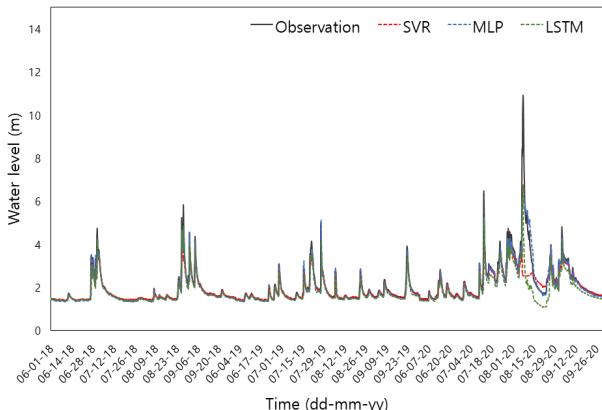
다음 Fig. 6에서는 모형별 예측 결과와 관측 수위 자료를 시계열 그래프로 선행 예측시간별로 도시하였다. 전반적인 수위 변화를 잘 예측하는 것으로 나타났으나, 2020년도 집중



(a) 1-hour lead time



(b) 3-hour lead time



(c) 6-hour lead time

Fig. 6. Water Level Prediction Results of ML Models

호우시 발생한 침투수위의 경우 LSTM모형을 제외한 SVR과 MLP 두 모형은 과소 추정되는 것으로 나타났다. 또한 선행예측시간이 길어질수록 관측값과 예측값의 오차가 크게 발생하고 있는 것으로 나타났다.

Fig. 7은 예측값(prediction)과 관측값(observation) 간의 산점도를 통해 각 모형별, 선행시간별 예측 결과를 도시하였다. SVR과 MLP 모형의 경우 고수위의 경우 과소 추정되는 양상을 뚜렷하게 보였으며, LSTM의 경우에는 1시간 선행 예측 시에는 비교적 고수위를 잘 예측하나 3시간, 6시간 선행 예측시에는 다른 두 모형과 비슷한 결과를 보였다.

## 5. 결론

본 연구에서는 홍수로 인한 피해를 저감하기 위한 방안으로 섬진강 구례 지점에 대해 1, 3, 6시간 선행하여 하천 수위를 예측하는 연구를 수행하였다. 예측을 위한 모형으로는 머신러닝 3개 모형, SVR, MLP, LSTM 모형을 각각 적용하였으며 선행 예측 결과의 비교를 통해 모형 성능을 비교·평가하였다. 연구 결과에서는 일반적으로 3개 모형 결과 모두 1시간 선행 예측 성능이 가장 높게 평가되었으며, 선행시간이 길어질수록 오차가 커지고 예측 정확도가 낮아지는 것으로 나타났다. 2020년도 발생한 극한 사상에 대한 예측 결과에서는 SVR과 MLP 모형의 경우 다소 큰 오차가 발생하였고 실제 관측수위보다 과소 추정된 것으로 나타났다. LSTM 모형의 예측 결과에서는 1시간 예측 시에는 침투값을 비교적 잘 반영하였으나 3, 6시간 선행예측 결과는 동일하게 과소 추정되었다. 3, 6시간 선행 예측시에는 MLP 모형을 이용하는 것이 적합할 것으로 판단되나, 종합적으로 LSTM 모형을 이용한 1시간 선행 예측 결과의 예측 성능이 가장 우수한 것으로 나타났다(NSE: 0.98, R2: 0.98, RMSE: 0.22 m).

본 연구를 통해 연속된 시계열 예측을 위해 기율기 소실 문제를 개선한 모형인 LSTM 모형의 적용성을 확인하였다. 이를 통해 현재 홍수예경보 시스템이 가지는 공간적 적용에 있어서의 한계와 기존 물리적 기반 모형이 지니는 자료 수집 및 모형 구축 과정에서의 한계를 보완할 수 있을 것이라 기대된다. 하지만 선행시간 1시간 이상의 예측에 대해서는 보다 정확한 예측을 위해 지속적인 연구가 필요할 것으로 판단된다. 또한 2020년도 발생한 홍수시의 침투 수위의 경우에 분석기간(2006.06~2020.09) 내 최대 수위로, 예측 결과와 관측값의 오차가 가장 많이 발생하였다. 이는 입력 자료를 통해 학습하고 이를 기반으로 예측하는 머신러닝 모형의 한계로 생각된다. 따라서 실제 집중호우 발생시 실시간 수위 예측을 위해서는 이와 같은 한계를 고려하고 기존 모형의 결과와 상호 보완할 수 있는 방안을 찾는다면 보다 정확한 예측결과를 도출 할 수 있을 것으로 기대된다.

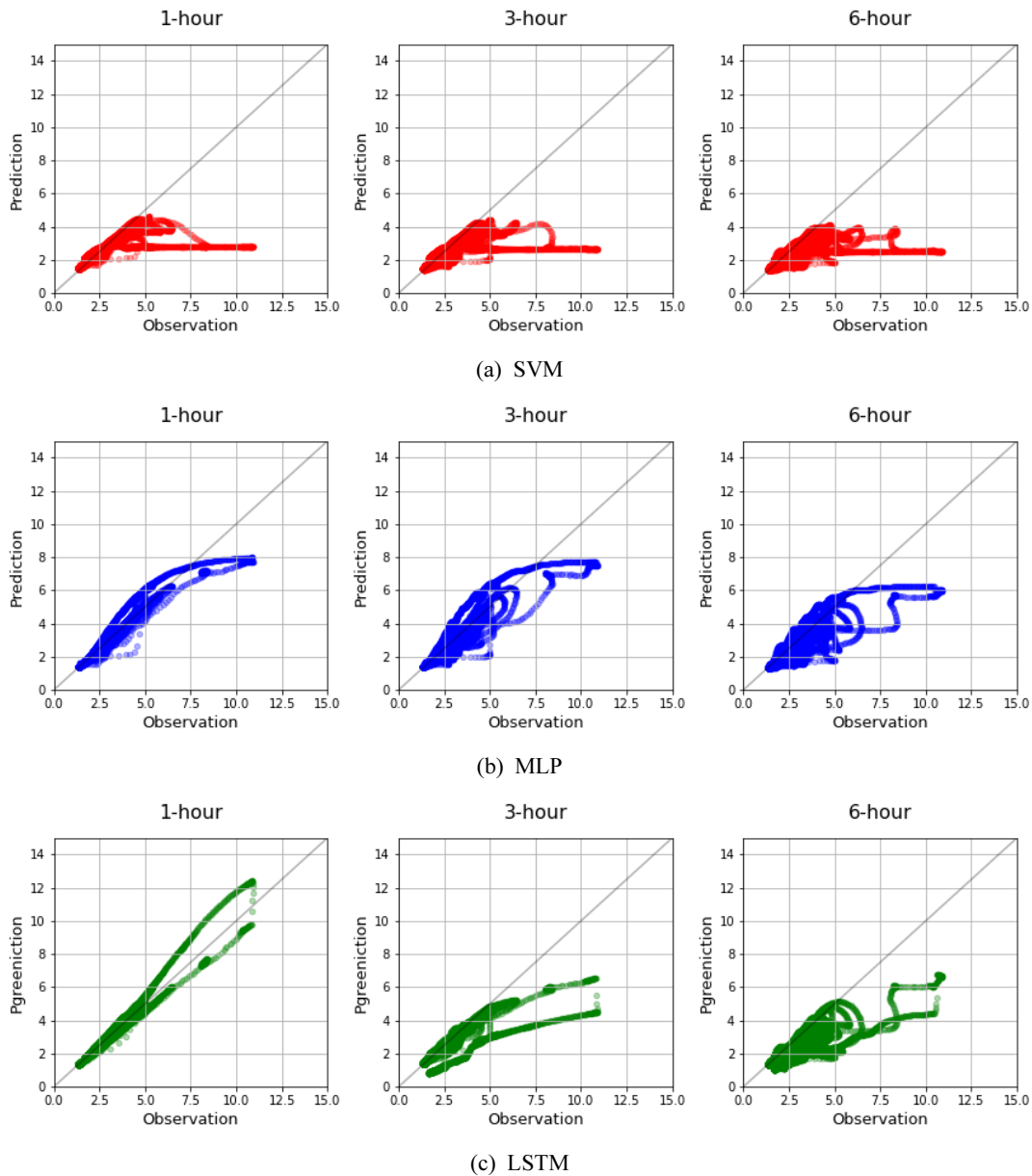


Fig. 7. Scatter Plots of Observation Versus Prediction Using ML Models

### 감사의 글

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2017R1A2B3005695).

### References

- Chen, W.B., Liu, W.C., and Hsu, M.H. (2012). Comparison of ANN Approach with 2D and 3D Hydrodynamic Models for Simulating Estuary Water Stage. *Advances in Engineering Software*, Vol. 45, No. 1, pp. 69-79.
- Chiang, Y.M., Chang, L.C., Tsai, M.J., Wang, Y.F., and Chang, F.J. (2010). Dynamic Neural Networks for Real-time Water Level Predictions of Sewerage Systems-Covering Gauged and Ungauged Sites. *Hydrology and Earth System Sciences*, Vol. 14, No. 7, pp. 1309-1319.
- Choi, C., Kim, J., Kim, J., and Kim, H.S. (2019). Development of combined heavy rain damage prediction models with machine learning. *Water*, Vol. 11, No. 12, pp. 2516.
- Choi, J.N., Ahn, W.S., Kim, H.S., and Park, M.K. (2010). Parameter Estimation of the Storage Function Model:

1. Development of the Universal Model for the Parameter Estimation. *J. Korean Soc. Hazard Mitig.*, Vol. 10, No. 6, pp. 119-130.
- Choi, K.S., and Moon, I.J. (2013). Two climate factors in May that affect Korean rainfall in September. *Acta Oceanologica Sinica*, Vol. 32, pp. 32-47.
- Dineva, A., Várkonyi-Kóczy, A.R., and Tar, J.K. (2014). Fuzzy expert system for automatic wavelet shrinkage procedure selection for noise suppression. *Proceedings of the 2014 IEEE 18th International Conference on Intelligent Engineering Systems (INES)*, Tihany, Hungary, pp. 163-168.
- Gizaw, M.S., and Gan, T.Y. (2016). Regional flood frequency analysis using support vector regression under historical and future climate. *J. Hydrol.*, Vol. 538, pp. 387-398.
- Han, H.C., Choi, C.H., Jung, J.W., and Kim, H.S. (2021). Deep Learning with Long Short Term Memory Based Sequence-to-Sequence Model for Rainfall-Runoff Simulation. *Water*, Vol. 13, No. 4, pp. 437.
- Hidayat, H., Hoitink, A.J.F., Sassi, M.G., and Torfs, P.J.J.F. (2014). Prediction of discharge in a tidal river using artificial neural networks. *Journal of Hydrologic Engineering*, Vol. 19, No. 8, pp. 04014006.
- Hong, S.K., and Ahn, J.H. (2021). Seismic Ground Response Analysis Based on Multilayer Perceptron and Convolution Neural Networks. *J. Korean Soc. Hazard Mitig.*, Vol. 21, No. 1, pp. 231-238.
- Jeong, M.G., Kim, G.T., and Gwon, H.H. (2019). Secondary damage response system for major infrastructure caused by natural disasters-Focused on flood disaster management technology. *Water for the future*, Vol. 52, No. 12, pp. 50-54.
- Jung, J.S., Lee, H.S., and Park, Y.K. (2020). Usage Improvement of Measured Flowrates for Water Management in Seomjin River Basin. *Journal of Korean Society for Environmental Technology*, Vol. 21, No. 6, pp. 480-488.
- Jung, S.H., Cho, H.S., Kim, J.Y., and Lee, G.H. (2018). Prediction of water level in a tidal river using a deep-learning based LSTM model. *Journal of Korea Water Resources Association*, Vol. 51, No. 12, pp. 1207-1216.
- Kang, B.S., and Lee, B.K. (2011). Application of Artificial Neural Network to Improve Quantitative Precipitation Forecasts of Meso-scale Numerical Weather Prediction. *Journal of Korea Water Resources Association*, Vol. 44, No. 2, pp. 97-107.
- Kim, B.J., Kwak, J.W., Lee, J.H., and Kim, H.S. (2008). Calibration and Estimation of Parameter for Storage Function Model. *Journal of Korean Society of Civil Engineers*, Vol. 28, No. 1B, pp. 21-32.
- Kim, S., and Tachikawa, Y. (2017). Real-time River-stage Prediction with Artificial Neural Network based on Only Upstream Observation Data. *Annual Journal of Hydraulic Engineering*, Vol. 62, pp. 1375-1380.
- Kim, S., Matsumi, Y., Pan, S., and Mase, H. (2016). A real-time forecast model using artificial neural network for after-runner storm surges on the Tottori Coast, Japan. *Ocean Eng.*, Vol. 122, pp. 44-53.
- Kim, S.J., Ryoo, E.C., Jung, M.K., Kim, J.K., and Ahn, H.C. (2012). Application of support vector regression for improving the performance of the emotion prediction model. *Journal of Intelligence and Information Systems*, Vol. 18, No. 3, pp. 185-202.
- Kim, S.W., and Ahn, H.C. (2010). Development of an intelligent trading system using support vector machines and genetic algorithms. *Journal of Intelligence and Information Systems*, Vol. 16, No. 1, pp. 71-92.
- Lee, G., Jung, S., and Lee, D. (2018). Comparison of physics-based and data-driven models for streamflow simulation of the Mekong river. *Journal of Korea Water Resources Association*, Vol. 51, No. 6, pp. 503-514.
- Lee, H.H., Chung, S.H., and Choi, E.J. (2016). A case study on machine learning applications and performance improvement in learning algorithm. *Journal of Digital Convergence*, Vol. 14, No. 2, pp. 245-258.
- Lee, J.H., Kim, J.S., Jang, H.W., and Lee, J.C. (2013). Drought Forecasting Using the Multi Layer Perceptron (MLP) Artificial Neural Network Model. *Journal of Korea Water Resources Association*, Vol. 46, No. 12, pp. 1249-1263.
- Lee, J.S. (2021). *Development and Application of Artificial Intelligence Based Model for Real Time Flood Forecasting*. Ph.D. dissertation, Inha University, Incheon, Korea.
- Lee, M.H., Jung, I.W., and Bae, D.H. (2011). Korean Flood Vulnerability Assessment on Climate Change. *Journal of Korea Water Resources Association*, Vol. 44, No. 8, pp. 653-666.
- Lee, W.K., and Tuan Resdi, T.A. (2016). Simultaneous hydrological prediction at multiple gauging stations using the NARX network for Kemaman catchment, Terengganu, Malaysia. *Hydrological Sciences Journal*, Vol. 61, No. 16, pp. 2930-2945.



- Lee, Y.H., Kim, H.I., Han, K.Y., and Hong, W.H. (2020). Flood Evacuation Routes Based on Spatiotemporal Inundation Risk Assessment. *Water*, Vol. 12, No. 8, pp. 2271.
- Lesmeister, C. (2017). *Mastering machine learning with R, 2nd ed.* Packt Publishing Ltd., Birmingham, UK.
- Liu, C., Guo, L., Ye, L., Zhang, S., Zhao, Y., and Song, T. (2018). A review of advances in China's flash flood early-warning system. *Natural hazards*, Vol. 92, No. 2, pp. 619-634.
- Maier, H.R., and Dandy, G.C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental modelling & software*, Vol. 15, No. 1, pp. 101-124.
- Ministry of the Interior and Safety (MOIS). (2019). *Statistical Yearbook of Natural Disaster 2018 MOIS*. Seoul, Korea.
- Mosavi, A., and Edalatifar, M. (2018). A hybrid neuro-fuzzy algorithm for prediction of reference evapotranspiration. In G. Laukaitis (Ed.), *International Conference on Global Research and Education* (pp. 235-243). Springer, Cham, Switzerland.
- Mosavi, A., Ozturk, P., and Chau, K.W. (2018). Flood prediction using machine learning models: Literature review. *Water*, Vol. 10, No. 11, pp. 1536.
- Park, J.S. (2006). *An empirical comparison between support vector regression and neural networks*. Master's thesis, Dongguk University, Seoul, Korea.
- Park, M.K., Jung, J.W., Joo, H.J., Kim, Y.S., Kwak, J.W., Kim, J.W., et al. (2021). A synthetic equation for storage function model parameter estimation based on kinematic wave approximation. *Hydrological Sciences Journal*, Vol. 66, No. 3, pp. 544-554.
- Park, S.H., and Kim, H.J. (2020). Design of Artificial Intelligence Water Level Prediction System for Prediction of River Flood. *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 24, No. 2, pp. 198-203.
- Park, U.Y., and Kim, G.H. (2007). A study on predicting construction cost of apartment housing projects based on support vector regression at the early project stage. *Journal of the Architectural Institute of Korea, Structure & Construction*, Vol. 23, No. 4, pp. 165-172.
- Song, J.H., Kim, H.S., Hong, I.P., and Kim, S.W. (2006). Parameter Calibration of Storage Function Model and Flood Forecasting (1) Calibration Methods and Evaluation of Simulated Flood Hydrograph. *Journal of Korean Society of Civil Engineers*, Vol. 26, No. 1B, pp. 27-38.
- Song, T., Ding, W., Wu, J., Liu, H., Zhou, H., and Chu, J. (2020). Flash flood forecasting based on long short-term memory networks. *Water*, Vol. 12, No. 1, pp. 109.
- Supharatid, S. (2003). Application of a neural network model in establishing a stage-discharge relationship for a tidal river. *Hydrological processes*, Vol. 17, No. 15, pp. 3085-3099.
- Suykens, J.A., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.*, Vol. 9, pp. 293-300.
- Taherei Ghazvinei, P., Hassanpour Darvishi, H., Mosavi, A., Yusof, K.B.W., Alizamir, M., Shamshirband, S., and Chau, K.W. (2018). Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *Eng. Appl. Comput. Fluid Mech.*, Vol. 12, pp. 738-749.
- Tay, F.E., and Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, Vol. 29, No. 4, pp. 309-317.
- Thirumalaiah, K., and Deo, M.C. (1998). Real-Time Flood Forecasting Using Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 13, No. 2, pp. 101-111.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York, NY, USA.
- Yoo, H.J., Lee, S.O., Choi, S.H., and Park, M.H. (2019). A Study on the Data Driven Neural Network Model for the Prediction of Time Series Data: Application of Water Surface Elevation Forecasting in Hangang River Bridge. *Korean Society of Disaster & Security*, Vol. 12, No. 2, pp. 73-82.
- Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., and Hinkelmann, R. (2020). Neurocomputing in surface water hydrology and hydraulics: a review of two decades retrospective, current status and future prospects. *Journal of Hydrology*, Vol. 588, pp. 125085.

---

|          |              |
|----------|--------------|
| Received | May 28, 2021 |
| Revised  | May 31, 2021 |
| Accepted | June 7, 2021 |