

기상 데이터를 활용한 CNN-LSTM 기반 유량 예측 모델 연구

김대호^{1,2}, 김세일¹, 고진환^{2*}

¹국방기술진흥연구소, ²경상국립대학교 전자공학과

A Study on a Flow Rate Prediction CNN-LSTM Model Based on Deep Learning using Meteorological Data

Dae Ho Kim^{1,2}, Seil Kim¹, Jinhwan Koh^{2*}

¹Korea Research Institute for defense Technology planning and advancement

²Department of Electronic Engineering, GyeongSang National University

요약 기후변화, 폭우 등 이상기후가 빈번히 발생하고 홍수, 가뭄 등 자연재해는 주변 지역에 심각한 영향을 미친다. 홍수로부터의 재해를 방지하기 위해 유량을 예측하는 것은 중요하다. 하지만, 유량 예측을 위해 강우의 패턴, 지속시간, 하천의 형상, 기온 등 예측하기 어려운 다양한 변수와 공간적 상관관계들에 의해 복합적으로 영향을 받기 때문에 개념적으로 표현하기가 매우 어려우며, 본 연구에서는 순환신경망 계열의 모델을 활용하였다. 기존 연구에서는 특정 신경망 모델의 선택, 매개변수 설정을 통한 예측의 정확도를 높이는 연구가 주를 이루고 있었지만, 학습데이터의 보완 사례는 미흡한 것으로 식별되었다. 따라서, 본 연구에서는 환경부에서 제공하는 유량 자료와 기상청에서 제공하는 기상 데이터를 결합하여 예측 모델의 성능을 비교·분석하였다. 데이터는 강원도 인제군의 기상 관측소 측정 기상데이터와 근접 지역의 리빙스톤교에서 측정되는 유량 데이터이며, 결측 데이터의 처리, 정규화 등의 절차를 거쳐 학습데이터를 구성하였다. 본 연구에서 수행된 모델은 LSTM(Long Short Term Memory), GRU(Gated Recurrent Unit), CNN-LSTM(Convolutional Neural Network-Long Short Term Memory)이며, 각 모델의 은닉층의 크기, 타임스텝, 필터의 크기 등의 파라미터를 조정하여 총 20개의 딥러닝 모델로 예측 성능을 확인하였으며, 이 중 CNN-LSTM이 RMSE는 15.8402, R2는 0.9780으로 예측 정확도 등에서 가장 우수하였다.

Abstract Abnormal weather such as climate change and torrential rains frequently occur, and natural disasters such as floods and droughts have a serious impact on surrounding areas. It is very important to analyze and predict river flow in order to prevent damage from floods. It is very difficult to conceptually express flow-rate data as they are nonlinear and complexly influenced by variables that are difficult to predict, such as rainfall patterns, rainfall duration, soil quality of the surrounding area, river shape, and temperature. Artificial neural networks have shown success in various fields with such nonlinear change characteristics. Studies tend to use models with only the river water level or flow rate, precipitation, and measurement data. Research on integrated flow-rate prediction in combination with other environmental factors is incomplete. Therefore, in this study, the performance of a flow-prediction model was compared and analyzed by mixing the flow data provided by the Ministry of Environment and the meteorological data provided by the Korea Meteorological Administration. The data used in this study were meteorological data measured at a weather station provided by the Korea Meteorological Administration in Inje-gun, Gangwon-do, and flow data measured at Livingston Bridge provided by the Ministry of Environment. The models configured for the comparative analysis of performance were LSTM, GRU, and CNN-LSTM. The prediction performance was checked according to a total of 20 cases per hyper-parameter for optimal learning. The CNN-LSTM model had the highest learning speed and prediction accuracy.

Keywords : Water Flow Rate Prediction, Deep Learning, Data Combination, CNN-LSTM, LSTM, GRU

*Corresponding Author : Jinhwan Koh(Gyeongsang National University)

email: jikoh@gnu.ac.kr

Received July 17, 2023

Accepted August 10, 2023

Revised August 8, 2023

Published August 31, 2023

1. 서론

인류는 자연을 통해 필요한 것을 얻어가면서 살고 있지만, 지구 온난화로 인해 기온이 상승하는 등의 기후 변동성 및 국지성 호우로 인한 홍수 발생 빈도가 증가하여 인명, 재산피해 발생 빈도가 높아지고 있다[1].

홍수와 같은 재해를 방지하기 위해서는 정확도 높은 홍수 예보를 사전에 필요한 조치를 수행할 수 있도록 하천 유량의 분석 예측 기술은 매우 중요하다[2]. 하지만, 유량 예측은 강우 패턴, 지속시간, 주변 지역의 토질, 하천 형상, 기온 등 예측이 어려운 다양한 변수와 공간적 상관관계에 의해 복합적으로 영향을 받기 때문에 개념적으로 표현하기가 매우 어렵다. 딥러닝은 이러한 비선형 특성을 가지는 다양한 분야에서 성과를 나타내고 있으며, 본 연구에서도 순환신경망 계열 모델을 적용하여 유량을 예측하는 연구를 수행하였다.

최근 인공신경망(ANN : Artificial Neural Network)을 통한 머신러닝은 수자원 분야에 널리 사용되고 있으며 기존의 순환신경망 계열 모델을 활용한 수자원 분야의 딥러닝 연구를 살펴보면 2011년, Sheong and Lewis.은 딥러닝 기법을 이용한 진동수조(OWC : Oscillating Water Column)의 수위를 예측 연구를 수행하였다[3]. 2017년, Tran et al.은 미국 텍사스 트리니티 강을 사례로 단일 수위 데이터 기반 표준 RNN(Recurrent Neural Network), LSTM(Long Short-Term Memory) 등의 딥러닝 알고리즘을 사용하여 침수수위를 예측하는 연구를 수행하였다[4]. 2018년, Yoon et al.은 관측된 강우량과 유량자료를 이용한 인공신경망의 신뢰성을 평가하였다[5]. 또한, Liang et al.은 LSTM 모델에 수위, 5개의 유량, 강수량을 입력 데이터로 Dongting Lake의 수위를 예측하여 서포트 벡터 머신보다 예측성이 우수함을 보였다[6]. 2019년, Le et al.은 LSTM 모형과 상류 유량 자료를 이용해 베트남 하류 지점을 유량을 예측하는 연구를 수행하였다[7]. 2020년, Lee et al.은 남강댐 유역에 위치한 관측자료 시간 간격에 따른 예측성 연구를 수행하였다[8]. 또한, Chum et al.은 CMIP5(Coupled Model Intercomparison Project Phase 5)와 LSTM 모형을 연계하여 유량 예측을 위한 최적 학습 기간을 제시하였다[9]. Kim et al.은 LSTM 모형을 이용하여 월 유량을 예측하고, SWMM(Storm Water Management Model) 모형의 모의 결과와 비교한 후에 로지스틱 회귀와 연계하여 도시 침수 범위를 예측하는 연구를 수행하였다[10]. Park et al.은 LSTM 모형을 이용한 제주도 지

역내 11개 지하수위 관측 수위자료를 대상으로 지하수위를 예측하는 모형을 제시하였다[11]. 2021년, Jung et al.은 섬진강 구례교 지점 하천 홍수위를 1, 3, 6시간 선형하여 SVM, MLP, LSTM 기반 성능분석 연구를 수행하였다[12]. 2022년, Cho et al.은 경기도 여주보 상류와 하류 데이터를 활용하여 LSTM, GRU(Gated Recurrent Unit), LSTM-GRU 3가지 모델을 활용하여 성능 비교 실험을 수행하였다[13].

수해 예방을 위한 딥러닝 기반 예측 모델 관련 연구동향 조사 결과, 수위 등의 시계열 자료를 인공신경망 모형을 활용하여 학습하고 그 결과를 예측하는 연구가 다양한 연구가 진행됨을 알 수 있다. 하지만, 특정 신경망 모델의 선택, 매개변수의 설정을 통한 예측의 정확도를 높이는 연구가 주를 이루고 있었고 학습 데이터의 확장 사례는 부족한 것으로 식별되었다.

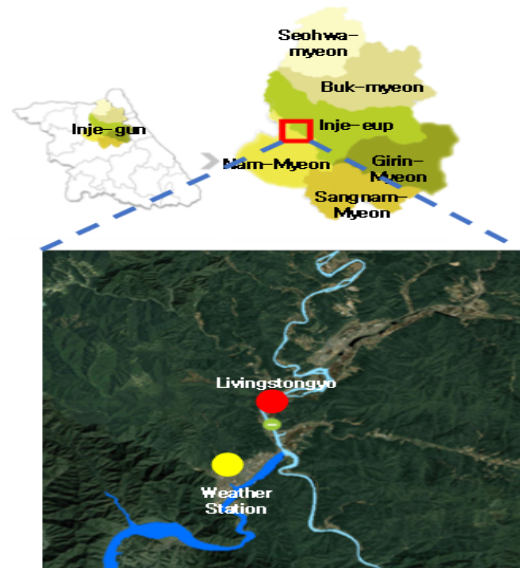


Fig. 1. LvingStongyo and Inje-Eup Weather Station

따라서 본 연구에서는 환경부에서 제공하는 유량 자료와 기상청에서 제공하는 기상데이터를 혼합하여 유량 예측 모델의 성능을 비교·분석하였다. 활용한 데이터는 강원도 인제군 기상청 제공 기상 관측소에서 측정되는 기상 데이터와 환경부 제공 리빙스톤교 유량 데이터이다. 기상 데이터를 이용한 유량 예측 모델을 설계하기 위해 측정 데이터 지리적 거리가 멀지 않은 곳을 선정했고, Fig. 1과 같이 인제군 기상 관측소와 직선거리 약 3km에 위치한 리빙스톤교를 대상으로 수집된 데이터를 혼합하여 학습 데이터로 활용하였다[14,15].

과적합 방지, 오류 방지 등 원활한 학습을 위한 결측 데이터 처리, 정규화 작업 등의 데이터를 전처리를 수행하였다. 이후 타임스텝, 은닉층의 크기, 컨볼루션 필터 등을 조정하여 LSTM, GRU, CNN-LSTM 모델 대상 파라미터를 조정하여 총 20개의 Case로 구성하였으며, 각 모델의 성능 비교를 위해 학습 속도, RMSE, R2를 통한 예측 오차를 확인하였다. 최종적으로 CNN-LSTM이 학습 속도, 예측 성능 측면에서 LSTM, GRU 보다 성능이 우수함을 확인하였다.

2. 딥러닝 모델

2.1 순환신경망(RNN: Recurrent Neural Network)

순차 데이터는 '어떤 순서를 가진 데이터로써 순서가 변경될 경우 데이터 특성을 잃어버리는 데이터를 말한다. 순서를 가지는 문장 등의 자연어, 주가, 음성, DNA 염기서열 등의 연속성을 가진 시계열(Time Series Data)이 모두 순차 데이터에 해당된다. 순환신경망은 인공신경망 알고리즘의 일종으로 순차 데이터를 분석하기 위해 과거 정보를 기억하는 기능을 가지고 있다. 과거 정보를 기억하는 대표적 방법은 이전에 사용한 데이터를 재사용하는 것이며, 신경망 층에서 순환이 필요하다. 이러한 알고리즘을 순환신경망(Recurrent Neural Network, RNN)이라 한다.

순환신경망은 아래 Fig. 2와 같이 시간마다 가중치로 겹겹이 쌓인 구조이며, 모든 과거 정보를 참고하는 특성으로 입력 순차열이 길 경우, 앞에 있는 시간이 뒤에 있는 시간으로 충분히 전달되지 않는 장기 의존성(Long Term Dependencies)의 문제가 발생할 수 있다.

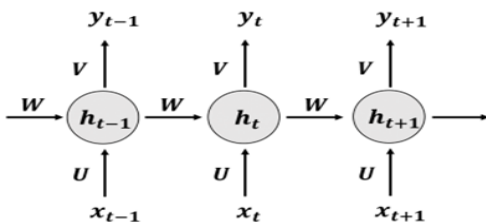


Fig. 2. RNN Structure

2.2 LSTM(Long Short-Term Memory)

RNN 단점인 장기 의존성 문제를 해소하기 위해 Hochreiter & Schmidhuber가 1997년에 LSTM를 제

시하였다[16]. LSTM은 Fig. 3과 같이 RNN과 같은 체인 구조를 가지고 있지만, 하나의 셀 안에 여러 개의 게이트를 가진다.

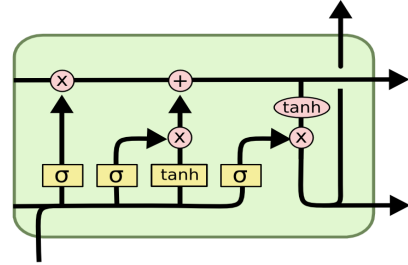


Fig. 3. LSTM Structure

LSTM의 핵심은 셀 상태(cell state)를 통해 이전 정보를 기억할 것인지 망각할 것인지 결정하는 것이며, 저장된 정보에 현재의 정보를 업데이트한 후에 셀 상태에 따라 활성화 함수에 의해 출력을 결정한다.

2.3 GRU(Gated Recurrent Unit)

GRU는 LSTM에서 영감을 받아 더 간략한 구조를 가지는 RNN 프레임워크 일종이다. GRU는 2014년 최초로 제안한 방법이며, Fig. 4와 같이 LSTM과 유사하게 게이트 개념을 사용하지만, 조금 더 간소화하였다[17]. GRU는 LSTM의 셀 상태가 존재하지 않고, 은닉층만을 가지며, 은닉층 값이 LSTM 셀 상태 역할을 대신한다.

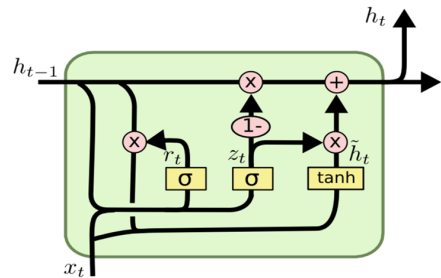


Fig. 4. GRU Structure

2.4 CNN-LSTM

CNN-LSTM은 컨볼루션 계층에서 합성곱 연산을 하는 필터가 시퀀스 위를 슬라이딩하여 인접 데이터에 대한 공간적 특성을 추출하고, 시간 정보 파악에 최적화된 알고리즘인 LSTM을 지나며 시간적 특성을 반영하는 것이 주요 특징이다[18]. 연구에 활용한 CNN-LSTM 구조는 Fig. 5와 같이 컨볼루션 계층, 풀링 계층, LSTM 층으로 구성되며, 입력 데이터에 대한 합성곱 연산 후 활성화

화 함수(본 연구에서는 ReLU 함수 적용)를 통해 특징을 추출한다. 이후 풀링 계층에서 검출된 특징 데이터를 크기를 줄여주는 작업을 수행하고, LSTM 계층을 통해 시계열 예측을 수행한다.

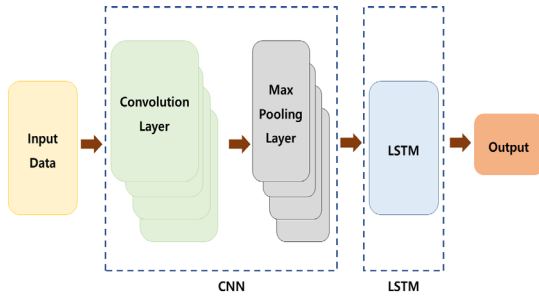


Fig. 5. CNN-LSTM Structure

3. 성능평가

3.1 데이터 수집 및 전처리

본 연구에 사용된 데이터는 기상청에서 제공하는 강원도 인제군 기상 관측소에서 측정한 종관기상관측(ASOS: Automated Synoptic Observing System) 자료 데이터(온도, 습도, 강수량, 전운량 등 8개 요소)와 환경부 한강홍수통제소(HRFCO: Han River Flood Control Office)에서 제공하는 근접 지역에 위치한 인제군 리빙스틴교 유량 계측 데이터를 각각 수집 및 시간 단위의 시계열 데이터로 혼합하였다. 데이터는 2019년 1월 1일 00시부터 2023년 3월 12일 23시까지 총 36,768시간으로 구성하였다.

각 수집한 기상 데이터 및 유량 데이터를 병합하여, Table 1에서 보는 바와 같이 유량, 온도, 강수량, 풍속, 습도, 증기압, 이슬점 온도, 현지 기압, 전 운량으로 총 9개의 변수로 구성되어 있다. 각 변수는 일부 결측 데이터를 포함하고 있으며, 해당 데이터는 왜곡 또는 편향된 예측 결과를 가져올 수 있으므로, 다른 값으로 대체하였다.

Table 1. Data Status with Missing Values

Feature	Unit	Source	Region	Number of Missing Value
Water Flow	m^3/s	HRFCO	Inge-Gun	147
Temperature	℃	ASOS	Inge-Gun	9
Precipitation	mm	ASOS	Inge-Gun	33,788
Windspeed	m/s	ASOS	Inge-Gun	35
Humedad	%	ASOS	Inge-Gun	8
VaporPressure	hPa	ASOS	Inge-Gun	12
Dew Point Temperature	℃	ASOS	Inge-Gun	12
Atmospheric Pressure	hPa	ASOS	Inge-Gun	12
Total Cloud	Decile	ASOS	Inge-Gun	453

이전 값과 연속성을 가질 수 있는 기온, 풍속, 습도, 증기압, 이슬점 온도, 현지 기압, 유량은 이전 값으로 대체했고, 이전 값에 크게 영향을 받지 않는 전 운량과 강수량은 0으로 대체하였다. 데이터 처리 후 데이터 정보는 아래 Fig. 6과 같이, 결측 데이터가 없음을 확인하였으며, 그림 28과 같이 2019년 1월 1일 00시부터 2023년 3월 12일 23시까지 총 36,768개의 데이터와 9개의 변수로 구성되어 있다.

데이터는 같이 각 변수 별 크기, 범위가 다르기 때문에 특정 변수에 대한 학습 편향 및 과적합이 발생할 수 있어 학습이 제대로 되지 않을 수 있다. 따라서, 모델이 효과적으로 학습할 수 있도록 모든 데이터 간 규모를 동일 선상에서 맞추기 위해 정규화(Min-Max Normalization)를 통해 모든 데이터를 0에서 1 사이의 값으로 변환시켰다.

유량예측 모델의 성능검증을 위해 LSTM, GRU, CNN-LSTM 등의 딥러닝 모델을 설계하였다. 각 딥러닝 모델의 학습/테스트 데이터 크기는 Fig. 7과 같이 전체 데이터 36,768 시계열 데이터 중에서 65%는 학습데이터로, 10%는 검증 데이터, 25%는 테스트 데이터로 분할하였다.

	WaterFlow	Temperature	Precipitation	Windspeed	Humedad	VaporPressure	DewPointTemperature	AtmosphericPressure	TotalCloud
count	36768.000000	36768.000000	36768.000000	36768.000000	36768.000000	36768.000000	36768.000000	36768.000000	36768.000000
mean	28.521398	10.642869	0.127597	1.524992	68.658671	10.868981	4.264039	993.004844	5.082789
std	110.361405	11.038592	0.987863	1.278313	21.623739	7.902168	11.876340	7.740442	4.025127

Fig. 6. Data Information After handling missing data

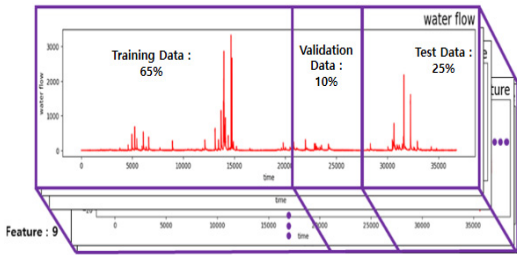


Fig. 7. Water flow prediction process

데이터 정규화를 위해 Eq. 1 식에 따라 정규화를 수행하였으며, 열값에서 최솟값을 뺀 값을 열의 최댓값에서 열의 최솟값으로 뺀 값으로 나눠주면 0과 1 사이의 정규화 된 값으로 변환시킬 수 있다[19].

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3.2 학습환경 구축

입력 데이터를 통한 유량 예측은 아래 Fig. 8와 같이 학습되며, 타임스텝(N hours)에 따라 9개의 변수 데이터를 통해 N+1 시점의 1개의(1시간) 유량을 예측하게 된다. 이를 계속 반복하여 유량 시퀀스 예측을 수행하였다.

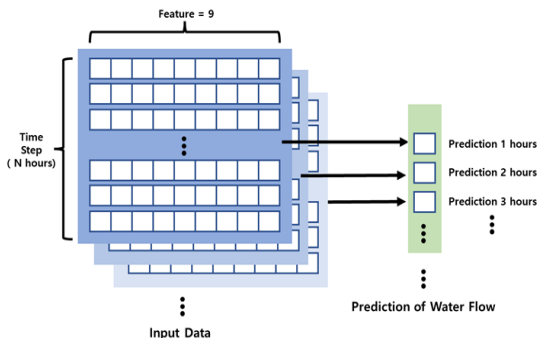


Fig. 8. Water flow prediction process

모델의 컴파일과 학습 훈련을 위해 적용된 파라미터들은 Table 2와 같으며, 타임 스텝과 은닉층, CNN-LSTM 모델의 경우에는 컨볼루션 필터의 크기를 가변 값으로 설정하여 구성하였다.

예측 정확도 정량화 및 모델별 비교·분석을 위해 Eq. 2 RMSE(Root Mean Squared Error)와 Eq. 3 R2(R-Squared)를 측정하여 성능평가를 수행하였다. RMSE는 값이 작을수록, R2는 1에 가까울수록 예측 성능이 우수하다.

Table 2. Parameter setting value for learning

Division		Setting Value
Variable Value	Time Step	24, 48
	Hidden Layer	180-30, 120-30, 90-30
	Convolution Filter	4, 8 (CNN-LSTM)
Fixed Value	Number of Feature	9
	Batch Size	16
	Activation Function	ReLU
	Stride	1(CNN-LSTM)
	Pooling	Max Pooling, Size : 2 (CNN-LSTM)
	Learning rate	0.001
	Epoch	200
	Early Stopping	20
	Loss Function	MSE
	Optimizer	Adam
Performance Evaluation	RMSE, R2	

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{pred} - y_{actual})^2} \quad (2)$$

$$R^2_{score} = \frac{\sum_i^n (y_{pred} - y_{mean})^2}{\sum_i^n (y_{actual} - y_{mean})^2} \quad (3)$$

연구에서 활용한 PC 환경은 CPU는 i5-10400, GPU는 RTX 2060, RAM은 16GB로 구성된다. 또한, 텐서플로우(Tensorflow)에서 GPU 병렬 연산을 위해 CUDA (Computed Unified Device Architecture)와 딥러닝 라이브러리인 cuDNN(CUDA Deep Neural Network)를 구축하였다.

3.3 모델 성능 분석

연구에서 모델의 성능 분석을 위해 활용한 딥러닝 네트워크는 Table 3과 같이 LSTM, GRU, CNN-LSTM로 분류할 수 있으며 타임 스텝은 24시간(1일), 48시간(2일), 은닉층은 90-30, 120-30, 180-30으로 설정했다. CNN-LSTM의 경우 CNN 네트워크가 추가되기 때문에 컨볼루션 필터 크기를 4, 8, 스트라이드는 모두 1로 적용하였다. 이를 통해 총 20가지 Case로 구성하였다. 각 Case 별 학습하여야 할 파라미터 수는 Table 3과 같이 CNN-LSTM, GRU, LSTM 순으로 적었고, 은닉층 크기 값이 커지면 학습 파라미터가 많아진다고 해석할 수 있다.

Table 3. Case according to deep learning network

Division		Variable Value				Total Parameter
		Time Step (hour)	Hidden Layer	Convolution		
				Stride	Filter	
LSTM	Case 1	24	90-30	-	-	50,551
	Case 2	24	120-30	-	-	80,551
	Case 3	24	180-30	-	-	162,151
	Case 4	48	90-30	-	-	50,551
	Case 5	48	120-30	-	-	80,551
	Case 6	48	180-30	-	-	162,151
GRU	Case 7	24	90-30	-	-	38,281
	Case 8	24	120-30	-	-	60,871
	Case 9	24	180-30	-	-	122,251
	Case 10	48	90-30	-	-	38,281
	Case 11	48	120-30	-	-	60,871
	Case 12	48	180-30	-	-	122,251
CNN-LSTM	Case 13	24	30	1	4	4,271
	Case 14	24	60	1	4	15,701
	Case 15	48	30	1	4	4,271
	Case 16	48	60	1	4	15,701
	Case 17	24	30	1	8	4,791
	Case 18	24	60	1	8	16,701
	Case 19	48	30	1	8	4,791
	Case 20	48	60	1	8	16,701

Table 4. Prediction performance

Division		Training Time (sec/epoch)	RMSE (m^3/s)	R2	Average RMSE
LSTM	Case 1	13	25.1574	0.9445	26.769
	Case 2	13	26.7029	0.9375	
	Case 3	14	32.0185	0.9101	
	Case 4	17	27.3736	0.9344	
	Case 5	18	24.2473	0.9486	
	Case 6	19	25.1143	0.9448	
GRU	Case 7	11	39.3885	0.8639	38.399
	Case 8	11	44.9875	0.8225	
	Case 9	11	36.4848	0.8832	
	Case 10	15	35.0375	0.8926	
	Case 11	13	37.5970	0.8763	
	Case 12	13	36.8966	0.8809	
CNN-LSTM	Case 13	9	18.7271	0.9693	20.593
	Case 14	9	15.8402	0.9780	
	Case 15	10	21.9961	0.9576	
	Case 16	9	21.2203	0.9606	
	Case 17	10	20.6231	0.9627	
	Case 18	8	18.3206	0.9705	
	Case 19	7	20.1482	0.9645	
	Case 20	9	27.8702	0.9320	

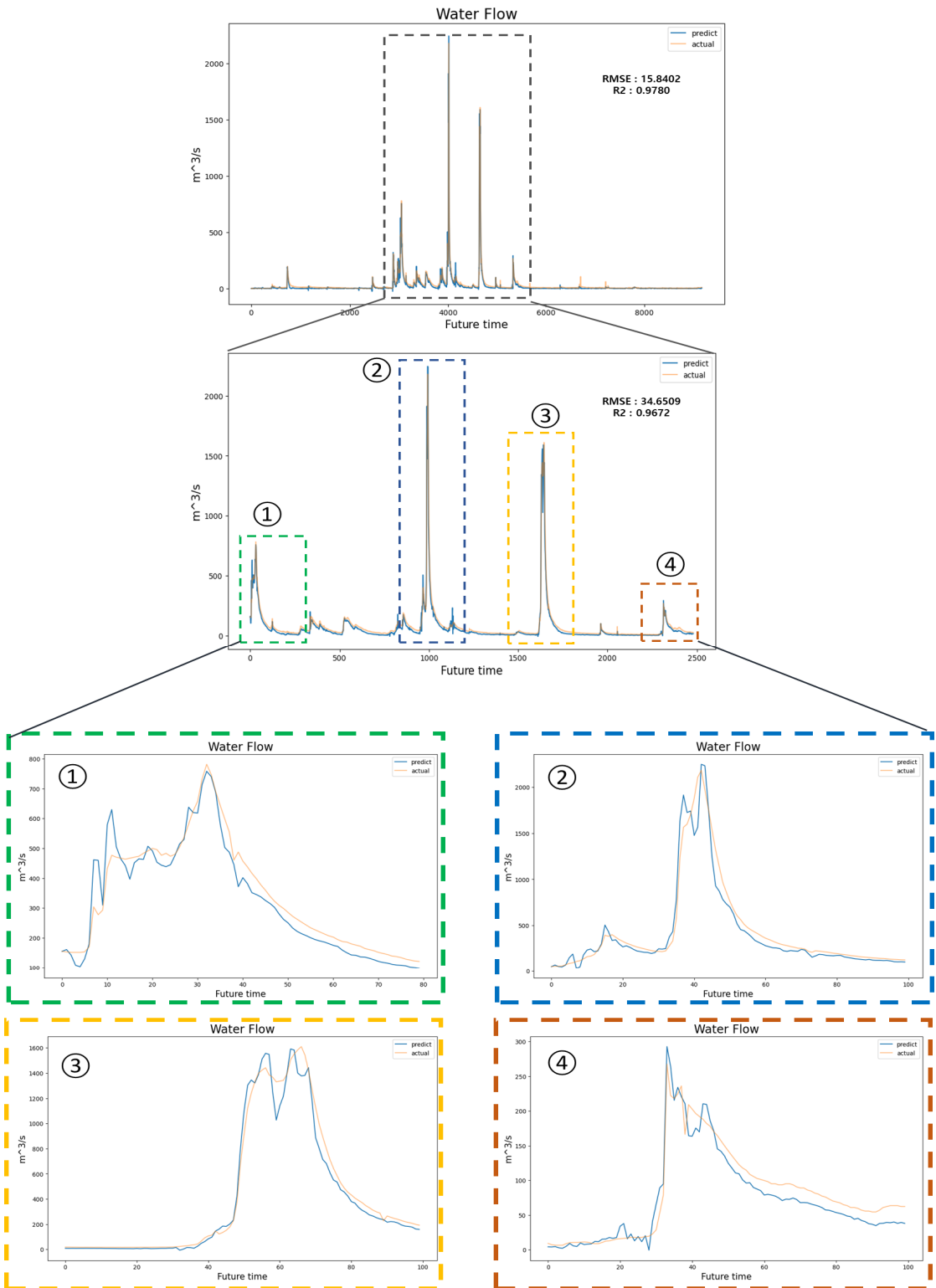


Fig. 9. Water flow forecast visualization(CNN-LSTM)

20가지 Case에 따른 성능 분석을 Table 4과 같이 정리될 수 있으며, 첫 번째는 LSTM 모델은 Case 5가 RMSE가 24.2473, R2가 0.9486으로 가장 높았으며, LSTM 모델들의 평균 RMSE는 26.769로 산출되었다. 두 번째, GRU 모델은 Case 10이 RMSE가 35.0375, R2는 0.8926으로 성능이 가장 높았으며, 모델들의 평균 RMSE는 38.399로 산출되었다. 마지막 CNN-LSTM은 Case 14가 RMSE는 15.8402, R2는 0.9780으로 가장 높았고, CNN-LSTM 모델 평균 RMSE는 20.593로 산출되었다.

성능 지표(RMSE, R2)에 따라 CNN-LSTM, LSTM, GRU 순으로 예측 정확도가 높음을 확인할 수 있다. 학습 속도는 1 에폭 당 소비되는 시간(초) 기준으로 CNN-LSTM, GRU, LSTM 순으로 학습 속도가 빨랐다. 결론적으로 CNN-LSTM이 LSTM, GRU 대비 유량 예측에 가장 적합한 모델이라고 해석할 수 있다.

마지막으로 연구에서 적용한 모델 중 가장 높은 정확도를 보여준 CNN-LSTM 모델(Case 14)을 대상으로 테스트 데이터에 대한 예측 결과를 시각화하면 Fig. 9과 같이 나타낼 수 있다. 테스트 데이터(전체 데이터의 25%) 1년(약 9,000시간) 중 유량이 급격히 증가하는 시점이 많아 홍수 등의 재해 가능성이 많은 6~9월 (2,500시간) 기간의 RMSE는 34.6509, R2는 0.9672로 높은 성능을 유지함을 확인하였다.

또한, 해당 기간 내에서 유량이 급격히 증가하는 4개의 기간(각 100시간)에 대해서도 실제 값(살색)과 유사한 딥러닝 예측 결과(파란색)를 도출할 수 있었다.

4. 결론

본 논문에서 복잡한 환경조건에 영향을 받고, 불확실성이 높은 유량 데이터에 대해 예측 모형을 개발하였으며, 딥러닝 학습기법 중에서 시계열 자료에 대한 학습에 특화된 인공신경망인 순환신경망 계열 기법을 이용하였다. 기존 연구에서 하천, 댐 등의 유량을 예측하기 위해 딥러닝 적용하기 위한 다양한 노력을 기울이고 있다. 하지만, 물 관련 데이터(유량, 강수량, 수위)는 해당 장소의 계측 자료만 제공하지만, 유량 데이터 특성상 주변 환경 인자(기상 등)에도 영향을 받기 때문에 환경 데이터를 결합하는 연구를 수행하였다. 하지만, 동일 지역 내 측정된 유량 및 기상 데이터 확보가 제한되어 유량 측정 지점과 인접지역에서 측정하는 기상 데이터를 혼합하여 데이

터를 생성하였다.

연구에 활용한 데이터는 강원도 인제군에 위치하는 기상 관측소에서 측정되는 기상 데이터와 근접 위치 리빙 스텐교에서 측정되는 유량 데이터를 2019년 1월 1일 0시부터 2023년 3월 12일 23시까지 1시간 단위로 데이터 수집하였고 수집된 데이터의 결측 데이터를 각 변수 특성에 맞추어 값을 대체하였으며 모든 값을 0에서 1 사이의 값으로 정규화하였다.

성능 비교분석을 위한 구성한 모델은 LSTM, GRU, CNN-LSTM이며, 최적 학습을 위해 하이퍼 파라미터 별 총 20개의 Case에 따른 각각의 예측 성능을 확인할 결과, 학습 속도, 예측 정확도 등의 측면에서 CNN-LSTM 모델이 가장 우수하다고 식별되었고, 테스트 데이터 전체 중 유량이 급격히 증가하는 시점에서도 높은 예측 성능을 유지하였다. 이는 과거 유입 자료 및 강수량 자료만을 이용한 예측 모델은 그 밖의 다른 영향 요소의 변동에 따라 과소 추정될 가능성이 있어 기상 데이터 혼합을 통한 입력 자료의 확대를 통해 예측 성능을 높일 수 있을 것으로 해석할 수 있다.

본 연구에서 활용한 기상 결합 데이터는 실제 기상 데이터 측정 장소와 유량 데이터 측정 장소가 동일 위치가 아니므로 거리(3km)에 따른 일부 환경 오차가 있을 것으로 판단된다. 그러므로, 댐 운영에 필요한 유량, 수위 데이터 외 기상 데이터 측정을 위한 각 하천 지점에 기상 센서를 구축할 시 거리 오차를 줄일 수 있을 것으로 판단된다. 이를 통해 동일 위치에서의 다양한 환경 수집 데이터를 확보할 수 있게 되고 유량 예측 오차를 줄일 수 있을 것으로 기대된다. 또한, 제방의 지각 상태 등 추가적인 고려 요소에 대해서도 고려된다면 예측 성능을 높일 수 있을 것으로 판단된다.

본 논문에서 제안하는 기상 데이터를 활용한 유량 예측 연구 수행은 향후 홍수 등의 수해를 조기에 예측할 수 있는 기반 기술로써 하천 제방 유실 등의 물적 피해, 인명피해를 최소화할 수 있는데 큰 기여가 될 것으로 기대된다.

References

- [1] Ministry of the Interior and Safety, "2021 Disaster Yearbook, Korea", pp.12-20, Available From: <http://www.mois.go.kr> (accessed May. 20, 2023)
- [2] H. R. Kim, C. H. Kim, "2013 Han River Flood Control Center Flood Forecasting Status", Korea Water Resources

- Association, Vol.46, NO.12, pp.83-92, 2013, Available From:
https://koreascience.kr/article/JAKO20131895817904_9.page
- [3] Sheng W, Lewis A, "Short-Term Prediction of an Artificial Neural Network In an Oscillating Water Column", International Journal of Offshore and Polar Engineers, United State, 2011, Available From:
<https://onepetro.org/IJOPE/article-abstract/35578/Short-Term-Prediction-of-an-Artificial-Neural?redirectedFrom=fulltext>
- [4] Tran. Q, S. K. Song, "Water Level Forecasting based on Deep Learning: A Use Case of Trinity River-Texas-The United States", Journal of KIIE, Vol.44, no.6, pp. 607-612, 2017, Available From:
<https://doi.org/10.5626/JOK.2017.44.6.607>
- [5] E. H. Yoon, J. B. Park, J. H. Lee, H. Y. Shin, "Reliability evaluations of time of concentration using artificial neural network model -focusing on Oncheoncheon basin", Journal of Korea Water Resources Association, Vol.51, No.1, pp.71-80, 2018,
 DOI: <https://doi.org/10.3741/JKWRA.2018.51.1.71>
- [6] C. Liang, H. Li, M. Lei, Q. Du, "Dongting Lake Water Level Forecast and Its Relationship with the Three Gorges Dam Based on a Long Short-Term Memory Network", Water 10(10):1389, Oct. 2018,
 DOI: <https://doi.org/10.3390/w10101389>
- [7] X. H. Le, H. H. Viet, G. Lee, "River streamflow prediction using a deep neural network: a case study on the Red River, Vietnam", Koren Journal Of Agricultural Science, Oct. 2019,
 DOI: <https://doi.org/10.7744/kjoas.20190068>
- [8] S. S. Lee, H. U. An, Y. T. Hur, Y. S. Kim, J. S. Byun, "Flood prediction in the Namgang Dam basin using a long short-term memory (LSTM) algorithm", Korean Journal of Agricultural Science, Vol.47, No.3, pp.471-483, 2020,
 DOI: <https://doi.org/10.7744/kjoas.20200036>
- [9] B. S. Chun, Y. C. Shin, T. H. Lee, S. W. Kim, K. J. Lim, Y. H. Jung, J. W. Do, "Estimation of Optimal Training Period for the Deep-Learning LSTM Model to Forecast CMIP5-based Streamflow", Journal of the Korean Society of Agricultural Engineers, Vol.64, No.1, pp.39-50, 2022,
 DOI: <https://doi.org/10.5389/KSAE.2022.64.1.039>
- [10] H. I. Kim, K. Y. Han, J. Y. Lee, "Prediction of Urban Flood Extent by LSTM Model and Logistic Regression", Korean Society of Civil Engineers, Vol.40, No.3, pp.273-283, 2020,
 DOI: <https://doi.org/10.12652/Ksce.2020.40.3.0273>
- [11] C. H. Park, I. M. Chung, "Evaluating the groundwater prediction using LSTM model", J.Korea Water Resour. Assoc, Vol.53, No.4, pp. 273-283, 2020,
 DOI: <https://doi.org/10.3741/JKWRA.2020.53.4.273>
- [12] J. W. Jung, H. L. Mo, J. H. Lee, Y. H. Yoo, H. S. Kim, "Flood Stage Forecasting at the Gurye-Gyo Station in Sumjin River Using LSTM-Based Deep Learning Models", Journal of The Korean Society of Hazard Mitigation, Vol.21, No.3, pp. 193-201, 2021,
 DOI: <https://doi.org/10.9798/KOSHAM.2021.21.3.193>
- [13] M. W. Cho, J. S. Park, S. J. Kim, H. K. Jung, "Design and implementation of flood management system using water level prediction model", Journal of the Korea Institute Of Information and Communication Engineering, Vol.27, No.2, pp.266-273,
 DOI: <https://doi.org/10.6109/jkiice.2023.27.2.266>
- [14] Korea Meteorological Administration, Meteorological Data Open Portal, Automated Synoptic Observing System, Available From:
<https://data.kma.go.kr/cmnn/main.do> (accessed Mar. 15, 2023)
- [15] Han River Flood Control Office, Real-time hydrological data, Available From:
<https://www.hrfco.go.kr/main.do> (accessed Mar. 15, 2023)
- [16] Hochreiter S, Schmidhuber. J, "Long Short-Term Memory", Neural Computation, Vol.9, Issue.8, pp.1735-1780, Nov. 1997,
 DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [17] J. Y. Chung, C. Gulcehre, K. H. Cho, Y. H. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling", NIPS 2014 Workshop on Deep Learning, 2014,
 DOI: <https://doi.org/10.48550/arXiv.1412.3555>
- [18] T. Y. Kim, S. B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks", Energy 2019, Vol.182, pp.72-81, 2019,
 DOI: <https://doi.org/10.1016/j.energy.2019.05.230>
- [19] J. M. Jo, "Effectiveness of Normalization Pre-Processing of Big Data to the Machine Learning Performance", The Journal of The Korea Institute of Electronic Communication Sciences, Vol.14, No.3, pp.547-552, 2019,
 DOI: <https://doi.org/10.13067/JKIECS.2019.14.3.547>

김 대 호(Dae Ho Kim)

[정회원]



- 2016년 8월 : 전북대학교 전자공학과 (공학사)
- 2017년 9월 ~ 2020년 12월 : 국방기술품질원 연구원
- 2021년 1월 ~ 현재 : 국방기술진흥연구소 연구원
- 2021년 9월 ~ 현재 : 경상대학교 전자공학과 석사과정

<관심분야>

국방기획, 정보통신, 컴퓨터공학

김 세 일(Seil Kim)

[정회원]



- 2019년 8월 : 충남대학교 전파공학과 (공학석사)
- 2019년 12월 ~ 2020년 12월 : 국방기술품질원 연구원
- 2021년 1월 ~ 현재 : 국방기술진흥연구소 연구원

<관심분야>

정보경영, 정보통신

고 진 환(Jinhwan Koh)

[정회원]



- 1999년 2월 : 시라큐스대학교 전기공학 박사 졸업
- 2003년 2월 ~ 현재 : 경상국립대학교 공과대학, 그린에너지연구원, 전자공학과 교수

<관심분야>

레이더시스템, 전자파신호처리, 마이크로프로세서