

딥러닝 기반 LSTM 모형을 이용한 저수지 수위자료 이상치 탐지

Anomaly Detection in Reservoir Water Level Data Using the LSTM Model Based on Deep Learning

양미혜* · 남원호** · 김한중*** · 김태곤**** · 신안국***** · 강문성*****

Yang, Mi-Hye* · Nam, Won-Ho** · Kim, Han-Joong*** · Kim, Taegon**** · Shin, An-Kook***** ,
and Kang, Mun-Sung*****

Abstract

Weather and hydrological phenomena have been changing due to climate change as evidenced by localized torrential rainfall and precipitation falling by more than 30% compared to the annual average. From 2013 to 2017 the ninety-nine reservoirs reached a water storage rate of 0%, making a secure stable water supply for agriculture uncertain. There is an increased need for information regarding agricultural water management to respond to the changes in the agricultural environment and climate. Therefore, automatic water level measurement facilities have been introduced to determine the real-time reservoir storage capacity and agricultural water supply. According to the Ministry of Agriculture, Food and Rural Affairs' guidelines for the installation and operation of water level measurement instruments, automatic water level facilities are currently installed at 1,734 reservoirs and 1,880 irrigation canals, with water level data generated at 10-minute intervals. The official recognition of hydrological water level data for agricultural reservoirs increased from six in 2016 to forty-nine in 2019. Anomaly detection algorithm methods for data regarding the agricultural reservoir level as well as quality control measures based on agricultural reservoir characteristics are required to minimize data quality degradation and generate reliable hydrological data over time. Though it was practically impossible to analyze the correlation between the water level or run-off and influential factors such as weather and terrain, recently a non-linear hydrological analysis has been possible using models such as Artificial Neural Networks (ANNs). This study aims to present an anomaly detection algorithm for reservoir level data using deep learning based LSTM (Long Short-Term Memory) models, in combination with other neural networks for managing quantitative information of agricultural water supply.

Key words : Agricultural Water, Water Level, Quality Control, Deep Learning, Anomaly Detection

요 지

최근 국지성 가뭄 발생 및 집중 호우, 평년 대비 강수량이 30% 이상 감소하는 등 기상 및 수문 현상이 변화하고 있다. 논벼 작물재배를 위한 용수를 공급하는 농업용 저수지의 경우 이상기후 발생으로 2013년부터 2017년까지 저수율이 0%에 도달한 저수지가 99개에 이르러 농업용수의 안정적 확보가 불확실해지고 있다. 농업환경의 변화와 기후변화에 대응하기 위해 농업용수 관리 정보화 및 과학화의 필요성이 증대되고 있으며, 실시간으로 저수지 저수량과 농업용수 공급량을 파악하기 위해 자동 수위계측시설이 도입되었다. 농림축산식품부의 저수지 자동수위계측기 설치 및 운영지침에 따라 현재 농어촌공사 관리 저수지 1,734개소 및 수로부 1,880개소에 자동수위계가 설치되어 있으며, 저수지와 수로에서 10분 간격으로 수위자료가 생성되고 있다. 농업용 저수지 수문자료의 공인지점은 2016년 6개소에서 2019년 49개소로 증가하고 있으며, 데이터 품질 저하의 최소화 및 신뢰성 있는 수문자료 생성의 필요성이 증가함에 따라 농업용 저수지의 특성을 반영한 저수지 수위 오결측 데이터 보정 방안 및 수문 자료 품질관리 방안이 요구된다. 최근 인공신경망(Artificial Neural Network, ANN) 등의 모형을 이용하여 비선형적인 수문해석이 가능해짐에 따라 농업용 저수지의 수위 변화 및 강우-유출 현상을 기상, 지형 등 영향 인자와 수위(또는 유출)와의 상관관계로부터 해석이 가능하다. 본 연구에서는 농업용수의 정량적 정보 관리를 위하여 시계열 자료의 학습에 효과적인 모형으로 다른 신경망과의 결합 등 다양한 분야에서 이용되고 있는 딥러닝(Deep Learning) 모형 중 하나인 LSTM (Long Short-Term Memory) 모형을 이용하여, 저수지 수위 자료의 오·결측 자료에 대한 이상 탐지 알고리즘을 제시하고자 한다.

핵심용어 : 농업용수, 수위 자료, 품질관리, 딥러닝, 이상 탐지

*정회원, 한경대학교 지역자원시스템공학과 석사과정(E-mail: mihye.yang@hknu.ac.kr)

Member, Master's Course, Department of Bioresources and Rural Systems Engineering, Hankyong University

**교신저자, 정회원, 한경대학교 사회안전시스템공학부 부교수(Tel: +82-31-670-5137, Fax: +82-31-670-5139, E-mail: wonho.nam@hknu.ac.kr)

Corresponding Author, Member, Associate Professor, School of Social Safety and Systems Engineering, Institute of Agricultural Environmental Science, National Agricultural Water Research Center, Hankyong National University

***한경대학교 사회안전시스템공학부 교수(E-mail: hanjoong@hknu.ac.kr)

Professor, School of Social Safety and Systems Engineering, Hankyong National University

****미네소타대학교 바이오시스템공학과 연구원(E-mail: taegon.k@gmail.com)

Research Associate, Department of Bioproducts and Biosystems Engineering, University of Minnesota

*****한국농어촌공사 농어촌연구원 주임연구원(E-mail: 2070161@ekr.or.kr)

Associate Researcher, Rural Research Institute, Korea Rural Community Corporation

*****한국농어촌공사 농어촌연구원 수석연구원(E-mail: msgang@ekr.or.kr)

Principal Researcher, Rural Research Institute, Korea Rural Community Corporation

1. 서론

농촌용수는 논·밭 용수, 생활, 공업, 환경유지 등 다기능을 하는 용수로 농업용수와 지역용수로 구분되며, 이 중 농업용수는 관개용수, 수로유지용수, 영농용수 등 농업생산 기능을 하는 용수이다. 농촌용수 관리는 한국농어촌공사와 지방자치단체로 이원화되어 있으며, 대다수가 농업용수 공급의 단일목적으로 이용되고 있다(Nam and Choi, 2013).

최근 국지성 가뭄 발생 및 집중 호우, 평년 대비 강수량이 30% 이상 감소하는 등 기상 및 수문 현상이 변화하고 있다. 논벼 작물재배를 위한 용수를 공급하는 농업용 저수지의 경우 이상기후 발생으로 2013년부터 2017년까지 저수율이 0%에 도달한 저수지가 99개에 이르러 농업용수의 안정적 확보가 불확실해지고 있다(MAFRA, 2014). 이상 수문 현상 및 농업환경의 변화에 대응하기 위해 농업용수 관리 과학화 및 정보화의 필요성이 증가하고 있다. 한국농어촌공사에서는 농림축산식품부의 저수지 자동수위측정기 설치 및 운영 지침에 따라 실시간으로 저수지 저수량과 농업용수 공급량을 파악하기 위해 자동 수위계측시설을 도입하였다(Hong et al., 2014; Nam et al., 2016). 현재 10만 톤 이상의 저수용량을 가진 공사 관리 저수지 1,734개소, 수로부 1,880개소에 자동수위계가 설치되어 있으며, 10분 단위로 수위 자료를 생성하고 있다(Nam et al., 2013; Shin et al., 2020). 수위계측 센서는 압력측정식 센서와 초음파측정식 센서가 있으며, 압력측정식 센서는 토사물의 유입, 초음파측정식은 온도, 습도, 파랑 등에 의해 계측 값이 영향을 받을 수 있다. 또한 외부 환경뿐 아니라 센서 자체의 결함 및 통신 장비 이상으로 계측 값이 영향을 받을 수 있다. 농업용 저수지 수문자료의 공인지점은 2016년 6개소에서 2019년 49개소로 증가하고 있으며, 데이터 품질 저하의 최소화 및 신뢰성 있는 수문자료 생성의 필요성이 증가함에 따라 농업용 저수지의 특성을 반영한 저수지 수위 오결측 데이터 보정 방안 및 수문 자료 품질관리 방안이 요구된다.

국내 수문자료 품질관리를 위해 다양한 연구들이 진행되어 왔으며, Choi and Kim (2009)은 수문자료 관측현황 및 수문자료관리시스템을 조사하고 분석하여 수문자료 관리 개선방안을 제시하였다. MOLIT (2010)에서는 신뢰성 있는 수문자료 생산을 위해 수문조사시설의 설치환경, 유지, 관리, 수문자료 품질 관리, 수문조사 방법 및 기준 등 표준화 방안을 제시하고, 관련 기준 및 매뉴얼 작성 방안 등을 제시한 바 있다. 또한 수문자료 품질관리 업무 가이드라인을 작성하여 체계적인 수문자료 품질관리 절차와 방법을 제시하였다(MOLIT, 2011). Kang et al. (2010)은 노후화된 농업용수로 시스템의 누수 및 관리손실량을 저감하기 위해 계측기 배치 위치, 계측자료 해석 방안, 자료 품질보증 등 효율적인 농업용수로 시스템 관리 방안을 제시하였다. Oh et al. (2014), Oh et al. (2015)은 관측기기 종류 및 특성, 자연현상에 의한

장애요인 등 오류 발생의 원인을 조사하고 품질관리 기법 및 수문검사 조사 등을 실시하여 수문자료 품질 개선 방안을 제시하였다. K-water (2017)는 LSTM 알고리즘을 도입하여 시뮬레이션 수행 후 데이터에 적합한 모델을 구축하여 오·결측 데이터 관리 업무 자동화 프로그램을 개발하였다. KRC (2018)는 저수지 및 수로부 수위 계측자료에서 발생하는 이상치를 보정하기 위해 수위변화자료의 임계값을 설정하여 1차 필터링 후, 저수지별 통계값을 이용하여 2차 필터링하는 이상치 보정 모듈을 개발하였다. K-water (2019)는 오·결측 기법 및 매개변수 최적화 방안 도출, DB 내 시스템 품질관리 자동화를 위해 품질관리 적용 대상 및 주기에 따른 알고리즘 구축에 대한 표준화를 마련하였다.

수위자료 모의 및 예측을 위해 강우-유출 현상 등 수문해석이 이루어져야 한다. 수문해석을 위해 다양한 개념적, 물리적 모형이 개발되었으나 필요한 모든 변수의 반영이 어려운 실정이다. 또한 모형이 정교할수록 매개변수가 증가하게 되고, 검증 및 보증되지 않은 매개변수는 결과에 불확실성을 초래할 수 있다(Jung et al., 2018). 농업용 저수지 수위 변화 및 기상 등 영향 인자와의 상관관계 분석은 실무적으로 불가능하였으나, 최근 인공신경망(Artificial Neural Network, ANN) 등과 같이 black-box 형태의 모형을 이용하여 비선형적인 수문해석이 가능해졌다(Yang et al., 2019).

본 연구에서는 딥러닝(Deep Learning) 기반의 LSTM 모형을 활용하여 강우를 고려한 저수지 수위 자료 예측을 수행하였다. 예측한 수위 자료를 바탕으로 농업용 저수지 원시 수위 자료를 검토하여 수위계측센서에서 발생할 수 있는 이상치 탐지 방안을 제시하고자 한다.

2. 연구방법

2.1 LSTM 모형

Recurrent Neural Network (RNN)는 Deep Neural Network (DNN)의 일종으로 시계열 데이터처리에 유용한 모형으로, 과거의 정보 및 시간의 흐름을 고려한 인식, 예측 등의 문제 해결에 활용 가능하다(Brownlee, 2017; Lee and Jun, 2018). 본 연구에서는 저수지 수위 자료와 인근 관측소의 강수량 자료를 활용하여 저수지 수위 자료를 예측하고 이를 저수지 수위 이상치 탐지에 활용하고자 한다. 따라서, 입력 변수와 시간의 흐름을 고려하여 저수지 수위 자료를 예측해야 하므로 RNN 구조를 활용하였다.

RNN은 이전 셀의 출력 값이 다시 입력 값으로 사용되는 병렬 체인 구조를 가지고 있다(Jang et al., 2019). 체인이 길어지면 기울기를 전파해야 하는 레이어 수가 늘어남에 따라 활성화 함수를 통과하는 횟수도 증가하여 기울기가 작아지는 기울기 소멸 문제, 과거의 학습 결과의 영향력이 감소하는 장기 의존성 등의 단점이 있다(Kim et al., 2016; Shin et al., 2017). 이러한 단점을 보완하기 위해 Hochreiter and Schmidhuber (1997)에

의해 LSTM이 제안되었다. LSTM은 RNN의 Hidden Node에 Cell State를 추가한 구조로, Cell State를 통해 이전의 가중치를 기억, 제거 여부를 결정하여 장시간, 대량의 데이터를 이용하여 학습할 경우 발생하는 기울기 소실 문제를 해결 하였다 (Brownlee, 2018; Jeon et al., 2019; Ryu and Han, 2019).

LSTM 모형 구동을 위해 입력 시퀀스의 각 시간 스텝마다 다음 시간 스텝 값을 예측하기 위해 Sequence-to-sequence Regression LSTM 네트워크를 통해 학습을 진행하였다. 네트워크 구조는 Fig. 1의 LSTM Architecture와 같이 구성하였다. 학습에 사용되는 최적화 알고리즘은 ADAM (Adaptive Moment Estivation) 알고리즘을 사용하였다. 기울기가 발산하는 것을 방지하기 위해 Gradient Threshold는 1, Initial Learn Late는 기본 설정 값인 0.005로 설정하였으며, Learn Rate Deep Period와 Learn Rate Drop Factor를 각각 기본 값인 125와 0.2로 지정하여 125 Epoch가 지나면 0.2를 곱하여 학습률을 떨어뜨리도록 하였다. 학습 횟수(Max Epoch)는 1,000회이며, Hidden Layer와 Hidden Unit 구성을 위해 Hidden Layer를 1, 2, 3개, Hidden Unit을 100, 200, 300개로 나누어 Test Dataset과 LSTM 결과 값의 RMSE (Root Mean

Square Error)를 산정하였다.

2.2 대상 저수지

농업용수 공급을 위한 저수지는 총 17,240개소이며, 한국농어촌공사에서 3,411개소, 시군에서 13,829개소를 관리하는 이원적 체계로 구성되어 있다(Kim et al., 2017). 한국농어촌공사에서는 농업용수 물관리 정보화를 위하여 자동수위계측시설을 도입하여 현재 저수지 1,734개소, 수로부 1,880개소에 설치하였다. 자동수위계측시설을 통해 생산된 수위 자료는 농촌용수종합정보시스템(Rural Agricultural Water Resource Information System, RAWRIS)을 통해 10분 단위 실시간 자료로 제공되고 있다. 이 자료는 실시간 계측데이터로 현장환경 및 품질관리에 의해 수정되어 수문자료 공인과정을 거쳐 제공된다. LSTM 모형을 활용하여 이상치를 탐지하고 탐지된 이상치의 검증을 위해 검증자료가 필요하므로 공인지점을 대상 저수지로 선정하였다. 본 연구에서는 대상 저수지를 저수지 공인지점 중 손항저수지, 수양저수지, 용암저수지로 선정하였다. 각 저수지에 대한 제원은 Table 1과 같다.

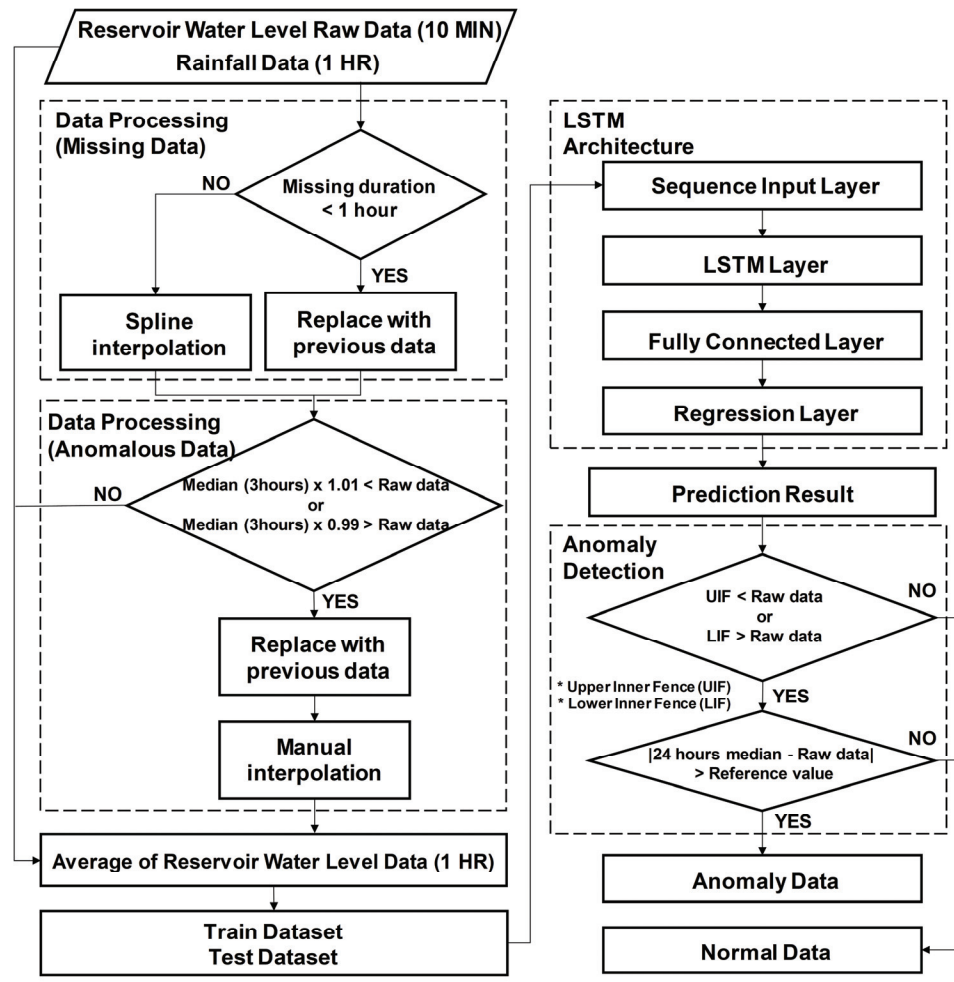


Fig. 1. Flow Chart for Anomaly Detection Algorithm of Reservoir Water Level Data

Table 1. Properties of Reservoirs

Facility Name	Sonhang	Suyang	Yongam
Place of Location	Gyeongsangnam-do	Jeollanam-do	Sejong-si
Irrigated Area (ha)	133.70	1,497.00	447.50
Total Storage Capacity (1,000m ³)	531.48	11,926.00	4,867.90
Flood Water Level (EL.m)	226.22	48.41	61.30
Dead Storage Level (EL.m)	210.00	37.66	50.16

2.3 저수지 수위 자료 및 기상 인자

농업용 저수지는 관개기인 4월부터 9월에 수문을 개방하여 용수를 공급하여 수위 자료가 감소하는 시계열적 특징을 갖고 있다. 저수지 유역의 유입량의 대부분은 강수량으로 관개기에 강우가 존재할 경우 별도의 처리 없이 그대로 경지에 공급이 가능하다. 따라서, LSTM 모형의 예측률을 높이기 위해 학습인자를 저수지 수위와 강수량으로 선정하였다.

본 연구에서는 각 대상저수지의 2014.01.01. 0:00에서 2019.10.31. 23:50까지 10분 단위 원시 수위 자료와 동일한 기간의 강수량 자료를 수집하였다. RAWRIS에서 실시간 10분 단위 계측 자료가 제공되고 있다는 점과 계측자료의 정확성의 확보를 통해 공인자료의 신뢰도를 높이고자 10분 단위 계측자료를 수집하였다.

각 저수지마다 인접한 기상청의 기상관측소를 선정하여 방재기상관측(Automatic Weather Station, AWS)의 강수량 자료를 수집하였다. 수양저수지는 장성(지점코드: 730), 오창저수지는 오창(지점코드: 693), 용암저수지는 세종연서(지점코드: 611) 관측소를 선정하였다. 기상청에서 제공하는 강수량 자료는 저수지 수위 자료의 수집 단위인 10분 단위의 강수량 자료를 제공하지 않아 시간 단위의 강수량 자료를 구축하였다.

2.4 이상치 탐지 알고리즘

저수지 수위 자료의 이상치를 탐지하기 위해 Fig. 1의 Anomaly Detection 조건을 설정하였다. Fig. 2와 같이 LSTM 데이터(i)의 전 1시간 데이터(i-7~i-1)의 사분위범위(InterQuartile Range, IQR)를 구하여 원시데이터가 IQR의 1.5배를 벗어나지 않으면 정상치로 판단하였다. 원시 데이터가 IQR의 1.5배를 벗어나는 경우, 이상치로 판단하고자 하였으나 수위 자료는 파랑에 의해 항상 일정하지 않은 특징을 갖고 있다. 따라서, 원시 데이터(i)가 IQR의 1.5배를 벗어나더라도 원시 데이터(i)와 1시간 데이터(i-7~i-1)의 중앙값과의 차이가 손항저수지의 경우 1, 수양저수지의 경우 0.1, 용암저수지의 경우 0.1보다 큰 경우 이상치로 판단하였다. 이 값은 수위가 정상적인 거동을 보일 때 평균적으로 발생하는 변화 값을 산정한 것이다.

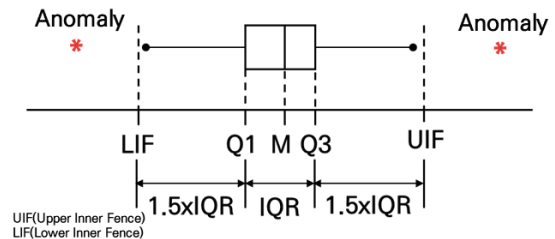


Fig. 2. Definition of Interquartile Range

3. 결과 및 고찰

3.1 LSTM 모형 적용을 위한 데이터 전처리

LSTM 모형을 통해 저수지 수위 자료, 강수량 자료, 수로부 수위 자료의 학습을 통한 저수지 수위 자료의 예측을 위해 보정이 이루어지지 않은 원시 수위 자료의 전처리 과정이 필요하다. 수위 자료의 전처리는 오측과 결측으로 구분하여 수행하였다. 결측의 경우, 결측지속기간이 1시간 이하인 경우는 이전 데이터로 대체하였으며, 결측지속기간이 1시간 초과인 경우 스플라인(Spline) 보간을 하였다. 오측의 경우, 오측자료의 직전 3시간 중앙값을 기준으로 $\pm 1\%$ 보다 크거나 작은 경우 이전 데이터로 대체하였으며, 이후 보정되지 않은 값에 대해서는 수동보정을 하였다. 전처리된 데이터는 강수량 자료와 시간단위를 맞추기 위해 해당 시간대의 강수량을 10분 단위로 동일하게 작성하여 학습 데이터를 구축하였다. Figs. 3-5는 대상 저수지별 구축된 학습 및 테스트 데이터를 나타내었으며, (a)는 원시 학습 데이터, (b)는 전처리된 학습 데이터, (c)는 원시 테스트 데이터, (d)는 전처리된 테스트 데이터이다.

전처리된 10분 단위 수위 자료를 LSTM 모형에 적용하기 전 정확도를 확인하기 위해 해당 저수지 공인 수위 자료와 비교하였다. 공인 자료는 매년 수문조사연보를 통해 제공되고 있으며, 본 연구의 대상 저수지인 손항저수지, 수양저수지, 용암저수지 모두 공인 수위 자료가 2018년부터 제공되어 테스트 데이터셋인 2019년 수위 자료로 정확도를 확인하였다. 공인 수위 자료는 일 단위로 제공되어 매일 0시의 전처리 자료와 비교하였을 때 대상 저수지 모두 R² 값이 0.99로 나타났고, RMSE 값은 손항저수지 0.216, 수양저수지 0.107, 용암저수지 0.399로 나타나 LSTM 모형에 적용하기에 적절하다고 판단하여 학습 및 테스트 데이터셋을 구성하였다.

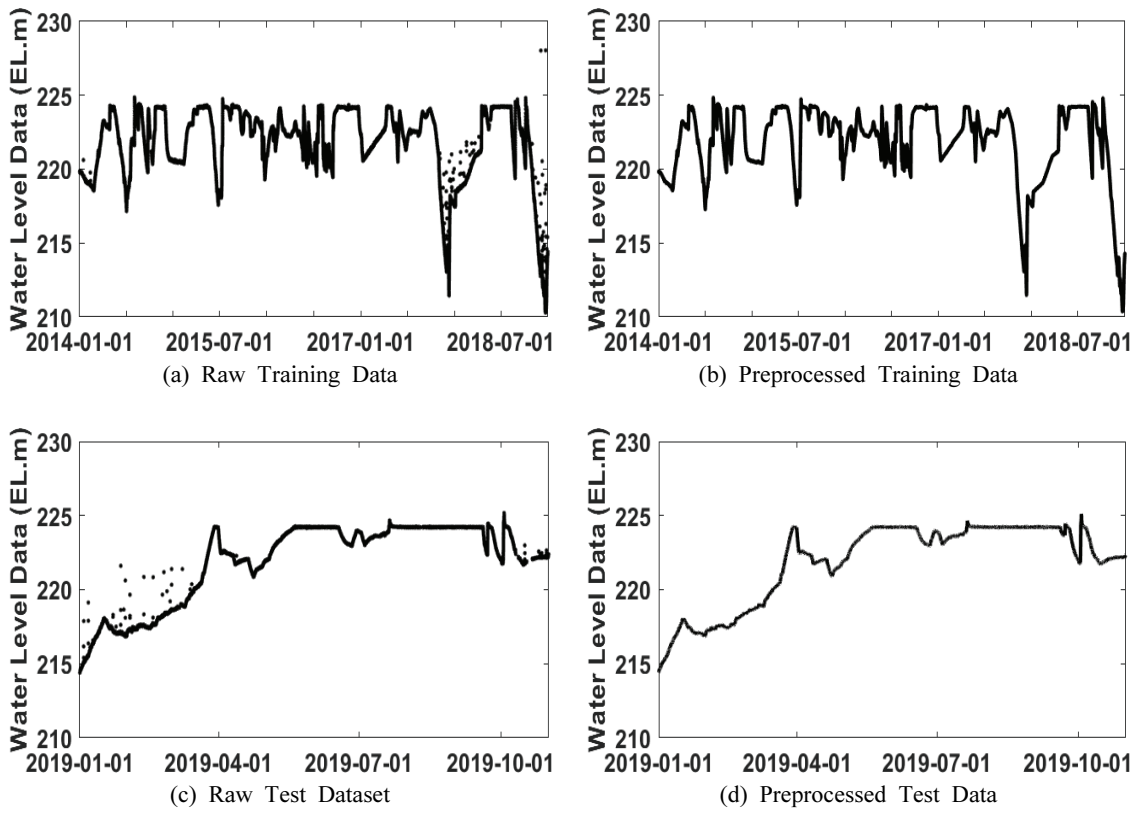


Fig. 3. Training Data Set and Test Data Set of Sonhang Reservoir

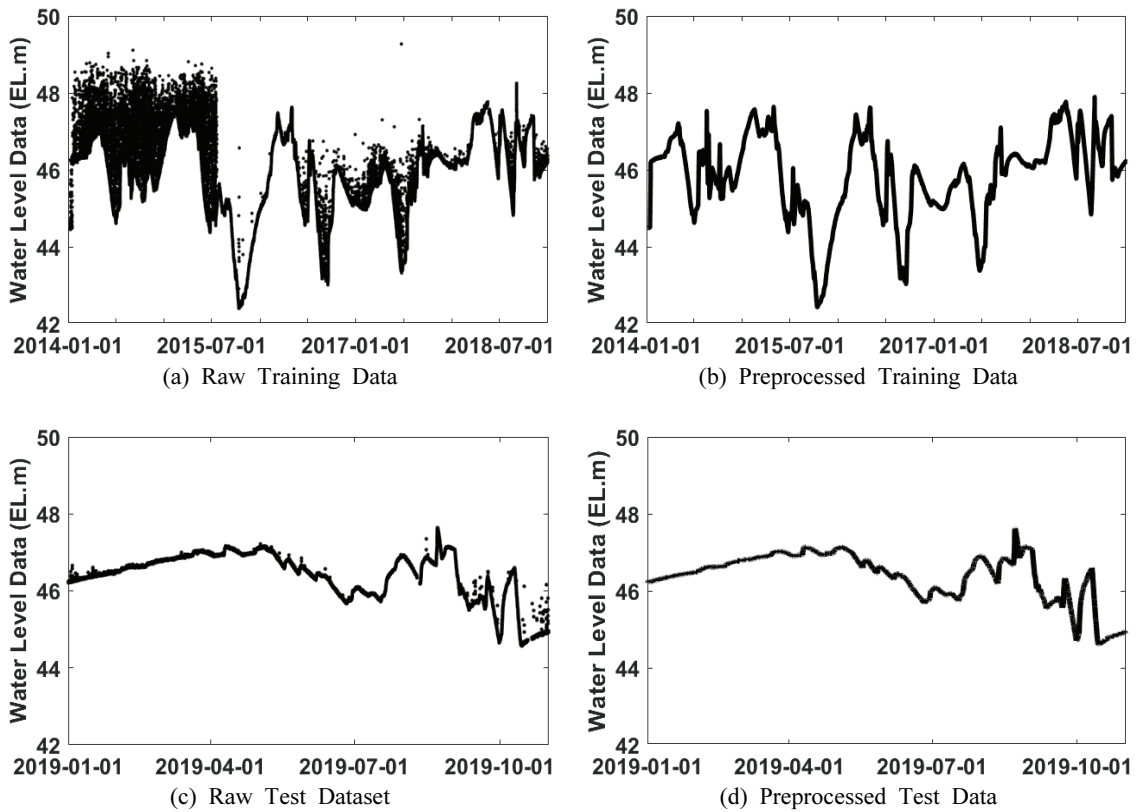


Fig. 4. Training Data Set and Test Data Set of Suyang Reservoir

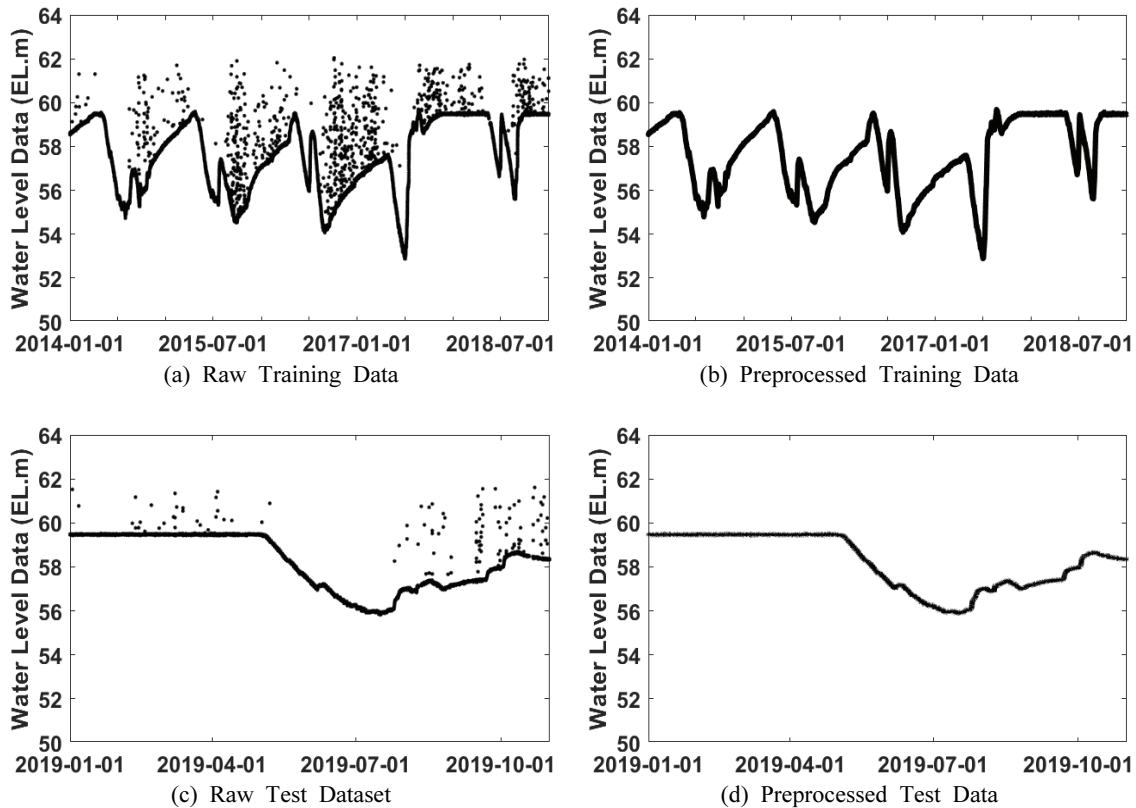


Fig. 5. Training Data Set and Test Data Set of Yongam Reservoir

3.2 LSTM 모형 적용

저수지 수위 자료 예측에 가장 적합한 LSTM 모형의 구축을 위해 Hidden Layer와 Hidden Unit을 조정하여 오차율을 통해 수위예측의 정확성을 확인하였다. 본 연구의 대상 저수지 중 용암저수지 데이터를 적용하였으며, 오차율 확인을 위해 RMSE를 사용하였다. 테스트 데이터셋과 LSTM 결과의 RMSE 값을 Table 2와 같이 비교하였다. RMSE 값이 0.034로 가장 낮았던 Hidden Layer 1개, Hidden Unit 200개의 구성으로 구축하였다.

Table 2. Comparison of RMSE on Test Datasets and Authorized Data According to Changes in the Number of Hidden Layers and Hidden Units

Hidden Unit	Hidden Layer		
	1	2	3
100	0.0743	0.0575	0.0598
200	0.0341	0.0558	0.0601
300	0.0521	0.1287	0.0503

Figs. 6, 8, 10의 (a)는 각 저수지별 LSTM 모형을 적용하여 2019년 예측 수위와 공인 자료를 나타낸 그래프이며, (b)는 공인 자료와 LSTM 자료의 산점도이다. Figs. 7, 9, 11은 각각 손항, 수양, 용암저수지의 LSTM 데이터와 공인데이터

의 분포를 나타낸 그래프이다.

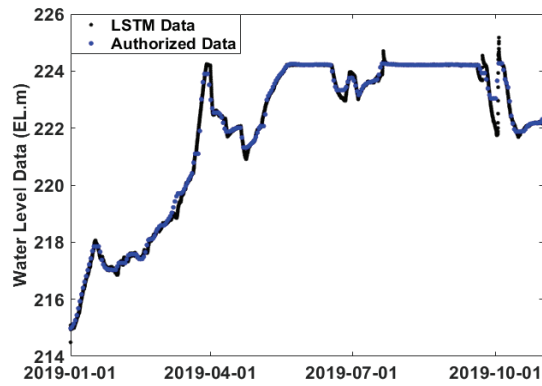
손항저수지는 4월과 9월, 10월의 수위가 급변하는 구간을 제외하고 공인 자료와 비슷한 결과를 보였다. 수양저수지는 8월, 9월, 10월 일부 자료에서 분포가 차이가 발생하였다. 손항저수지의 경우와 같이 수위가 급변하는 형상으로 예측되었다. 분포도는 분포가 대각선에 가까울수록 LSTM 결과 정확도가 높으며, 용암저수지는 공인자료와 비교하였을 때, 대부분 정확한 결과가 산정되었다. Figs. 6, 8, 10의 (b)를 확인하였을 때, 대상 저수지 모두 점들이 앞서 수위가 급변하는 구간을 제외하고 유사한 결과를 도출하였다.

LSTM 모형 적용 결과를 공인 자료와 비교하였을 때 손항저수지의 평균 RMSE는 0.216, 수양저수지의 평균 RMSE는 0.107, 용암저수지의 평균 RMSE는 0.091이며, 대상 저수지 모두 R^2 값이 0.99로 나타났다.

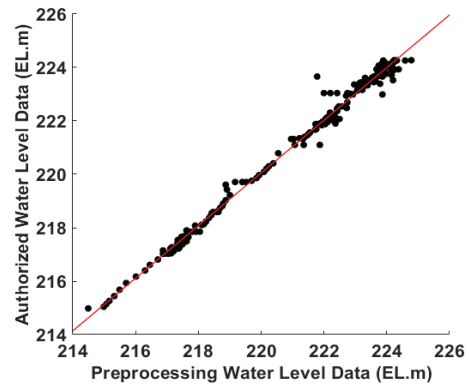
3.3 수위 자료 이상치 탐지

Figs. 12, 13, 14는 각 대상 저수지의 이상치 탐지 결과를 도시하였다. (a)는 오측 탐지를 확인하기 위한 그래프이다. (b)는 결측 탐지를 확인하기 위한 그래프이며, (b)에서 0은 결측을 의미한다.

Fig. 12의 (a)에서 손항저수지 수위자료 중 1-3월과 10월의 오측 데이터와 9월 일부구간이 이상치로 탐지되었다. 물결을 고려하여 이상치 탐지 조건을 설정하였으나 9월의 수위



(a) Verification of LSTM Data



(b) Scatterplot of Authorized Data and LSTM Data

Fig. 6. Verification of Sonhang Reservoir LSTM Data

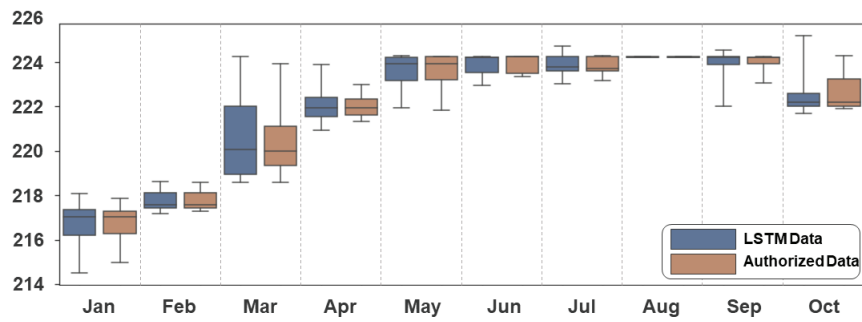
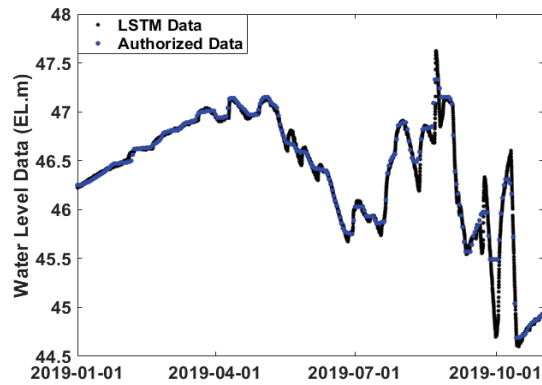
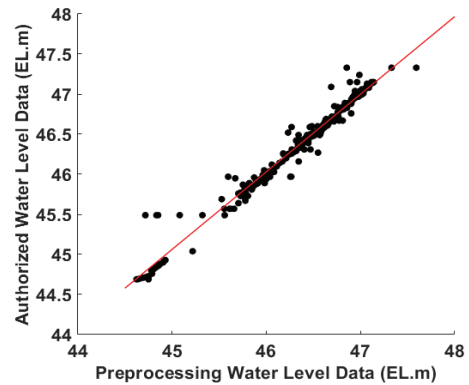


Fig. 7. Comparison of Monthly Distribution of Sonhang Reservoir LSTM Data and Authorized Data



(a) Verification of LSTM Data



(b) Scatterplot of Authorized Data and LSTM Data

Fig. 8. Verification of Suyang Reservoir LSTM Data

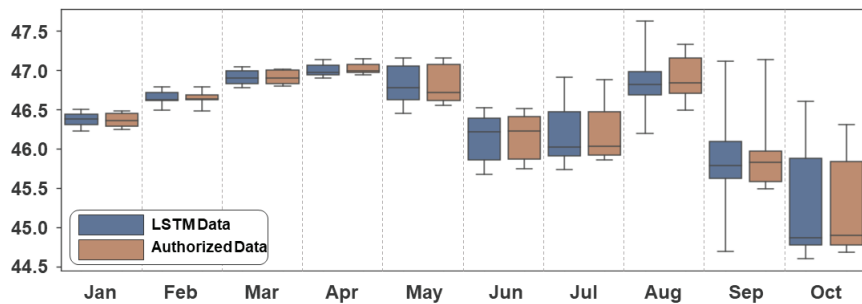
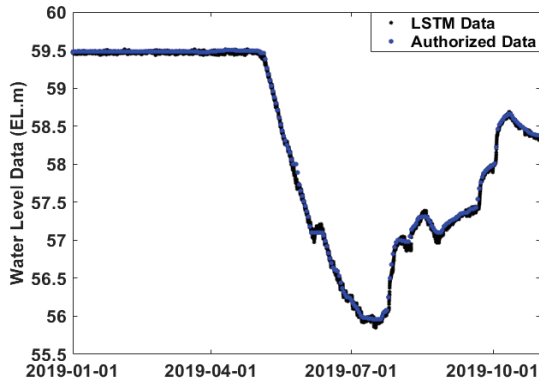
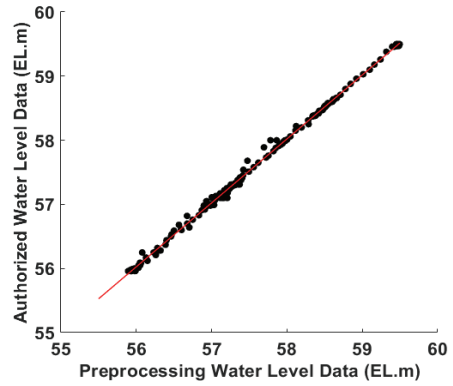


Fig. 9. Comparison of Monthly Distribution of Suyang Reservoir LSTM Data and Authorized Data



(a) Verification of LSTM Data



(b) Scatterplot of Authorized Data and LSTM Data

Fig. 10. Verification of Yongam Reservoir LSTM Data

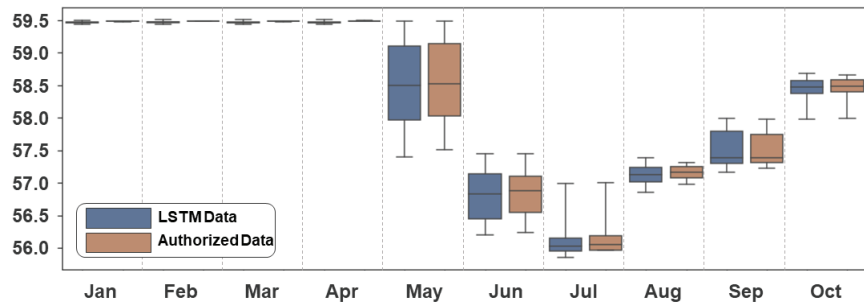
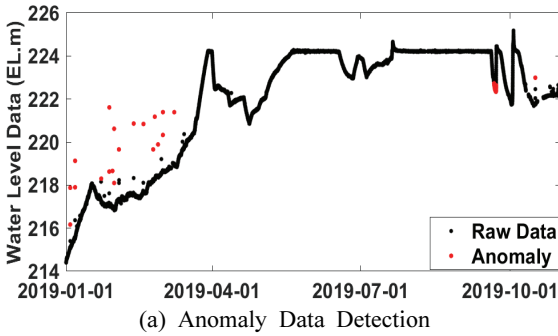
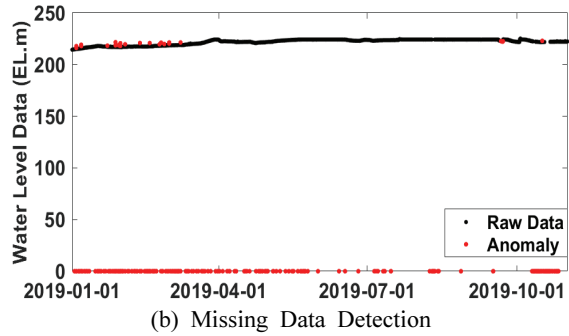


Fig. 11. Comparison of Monthly Distribution of Yongam Reservoir LSTM Data and Authorized Data

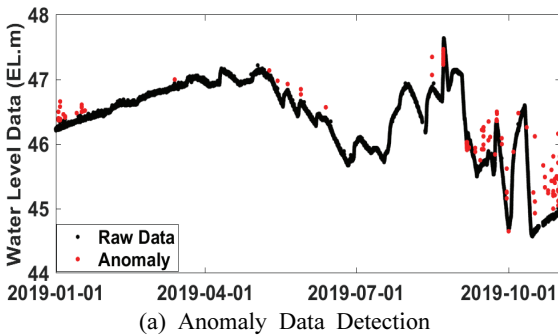


(a) Anomaly Data Detection

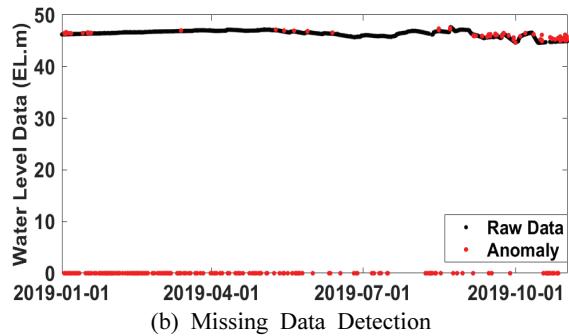


(b) Missing Data Detection

Fig. 12. Anomaly Detection of Sonhang Reservoir Water Level Data



(a) Anomaly Data Detection



(b) Missing Data Detection

Fig. 13. Anomaly Detection of Suyang Reservoir Water Level Data

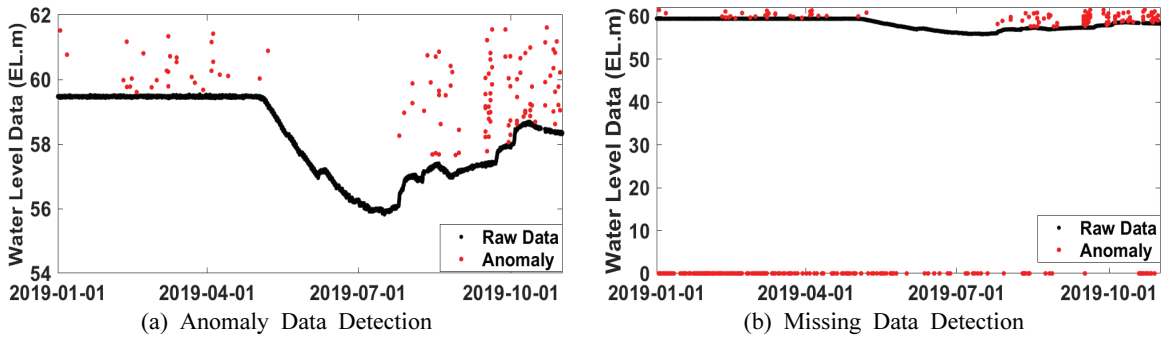


Fig. 14. Anomaly Detection of Yongam Reservoir Water Level Data

변동이 급격하여 탐지된 것으로 추정된다. Fig. 13의 (a)에서 수양저수지 수위자료 오측 데이터는 모두 탐지된 것을 확인하였다. 8월 일부구간은 손항저수지의 경우와 같이 수위가 급변하여 이상치로 탐지된 것으로 판단된다. Fig. 14의 (a)에서 오측 데이터가 모두 탐지되었음을 확인하였다. 또한 Figs. 12, 13, 14의 (b)에서 결측 데이터가 탐지된 것을 확인하였다.

본 연구에서 제시한 알고리즘을 적용하여 수위 자료의 이상치 탐지하였을 때, 원시 데이터의 이상치 비율에 관계없이 결측 데이터는 탐지가 가능하였다. 오측 데이터의 경우 탐지가 가능하였고, 수위 급변 구간의 경우 일부 이상치로 탐지 가능하였다.

4. 결론

본 연구에서는 딥러닝 기반 LSTM 모형을 활용하여 저수지 수위 자료 이상치 탐지 방안을 제안하였다. LSTM 모형 학습을 위해 저수지 수위 자료, 강수량 자료를 구축하였으며, 강수량 자료와 시간 단위를 동일하게 학습자료를 구축하였다. 학습을 위해 저수지 수위 자료를 결측과 오측으로 구분하여 전처리하였으며, 학습 데이터셋과 테스트 데이터셋으로 구분하여 사용하였다. 학습 모형 구축을 위해 Hidden Layer와 Hidden Unit을 조정하여 오차율이 가장 낮았던 Hidden Layer 1개, Hidden Unit 200개의 구성으로 구축하여 학습을 진행하였다. LSTM 모형을 통해 수위자료를 예측한 결과, 손항저수지와 수양저수지 일부 구간에서 수위가 급변하는 형태로 예측되었으나 대부분의 경우 높은 예측률을 보였다.

LSTM 모형을 통해 생성된 저수지 수위 자료를 바탕으로 두 가지의 이상치 탐지 조건을 적용하여 이상치 탐지 비율을 확인하였다. LSTM 모형을 활용한 농업용 저수지 수위 자료 이상치 탐지 결과이상치 비율에 관계없이 결측은 탐지가 가능하였으며, 오측의 경우 일부 수위 급변 구간이 이상치로 탐지되었으나 대부분의 이상치가 탐지되었다.

본 연구의 결과는 저수지 수위 자료 품질관리를 위해

딥러닝 기반 LSTM 모형 및 관련 인자인 강수량 자료를 활용하였으며, 정확한 수위자료 예측 모형 구축 및 품질관리를 위해 대상 저수지 및 학습 데이터의 누적 필요할 것으로 판단된다. 향후 LSTM을 이용한 저수지 수위 예측 모형이 구축되고 수위자료의 보정 방안이 마련된다면 이상치 탐지 및 보정 과정의 자동화가 가능해져 노동력 및 시간의 절감이 가능하다고 판단된다.

감사의 글

본 연구는 행정안전부 극한재난대응기반기술개발사업의 연구비 지원 (2019-MOIS31-010)에 의해 수행되었습니다. 또한 본 연구는 농업용 저수지 및 수로부의 수문자료 품질관리 기준 정립 연구사업의 연구비지원에 의해 수행되었습니다.

References

Brownlee, J. (2017). *Long short-term memory networks with python: develop sequence prediction models with deep learning*. Retrieved from <https://machinelearningmastery.com/lstms-with-python/>

Brownlee, J. (2018). *Machine learning algorithms in python, machine learning mystery: Understand your data, create accurate models and work projects end-to-end*. Retrieved from <https://machinelearningmastery.com/machine-learning-with-python/>

Choi, H.S., and Kim, J.Y. (2009). Quality control and improvement of hydrologic data at the Geum river flood control center. *Magazine of Korea Water Resources Association*, Vol. 42, No. 2, pp. 49-54.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780.

Hong, E.M., Nam, W.H., Choi, J.Y., and Kim, J.T. (2014).

- Evaluation of water supply adequacy using real-time water level monitoring system in paddy irrigation canals. *Journal of the Korean Society of Agricultural Engineers*, Vol. 56, No. 4, pp. 1-8.
- Jang, W., Lee, Y., Lee, J., and Kim, S. (2019). RNN-LSTM based soil moisture estimation using Terra MODIS NDVI and LST. *Journal of the Korean Society of Agricultural Engineers*, Vol. 61, No. 6, pp. 123-132.
- Jeon, B.K., Lee, K.H., and Kim, E.J. (2019). Development of a prediction model of solar irradiances using LSTM for use in building predictive control. *Journal of the Korean Solar Energy Society*, Vol. 39, No. 5, pp. 41-52.
- Jung, S.H., Lee, D.E., and Lee, K.S. (2018). Prediction of river water level using deep-learning open library. *J. Korean Soc. Hazard Mitig.*, Vol. 18, No. 1, pp. 1-11.
- Kang, M.G., Jeong, H.S., and Kim, J.T. (2010). Efficient management of agricultural canal systems through quality management of water level and water quantity data. *Magazine of Korean Society of Agricultural Engineers*, Vol. 52, No. 2, pp. 87-96.
- Kim, D.S., Kang, S.M., Kim, J.T., Kim, J.D., Kim, H.H., and Jang, J.U. (2017). Development and implementation of prototype for intelligent integrated agricultural water management information system and service including reservoirs managed by city and county. *Journal of the Korean Society of Rural Planning*, Vol. 23, No. 3, pp. 163-174.
- Kim, Y.H., Hwang, Y.K., Kang, T.G., and Jung, K.M. (2016). LSTM language model based Korean sentence generation. *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 41, No. 5, pp. 592-601.
- Korea Rural Community Corporation (KRC). (2018). *Technical development for data management and utilization using automatic water level measurements*.
- Korea Water Resources Corporation (K-water). (2017). *Development of automatic detection and refining system for abnormal data*.
- Korea Water Resources Corporation (K-water). (2019). *Development of quality control algorithm for standard database of water information*.
- Lee, T.H., and Jun, M.J. (2018). Prediction of Seoul house price index using deep learning algorithms with multivariate time series data. *SH Urban Research & Insight*, Vol. 8, No. 2, pp. 39-56.
- Ministry of Agriculture, Food and Rural Affairs (MAFRA). (2014). *A study on the improvement of cost-bearing system for using and managing agricultural water and repair facilities*.
- Ministry of Land, Infrastructure and Transport (MOLIT). (2010). *A study on standardization of hydrological investigation methods and standards*.
- Ministry of Land, Infrastructure and Transport (MOLIT). (2011). *Guideline user guide for service. Establishment of national hydrological data quality management system*.
- Nam, W.H., and Choi, J.Y. (2013). Development of operation rules in agricultural reservoirs using real-time water level and irrigation vulnerability index. *Journal of the Korean Society of Agricultural Engineers*, Vol. 55, No. 6, pp. 77-85.
- Nam, W.H., Choi, J.Y., Hong, E.M., and Kim, J.T. (2013). Assessment of irrigation efficiencies using smarter water management. *Journal of the Korean Society of Agricultural Engineers*, Vol. 55, No. 4, pp. 45-53.
- Nam, W.H., Hong, E.M., and Choi, J.Y. (2016). Assessment of water delivery efficiency in irrigation canals using performance indicators. *Irrigation Science*, Vol. 34, No. 2, pp. 129-143.
- Oh, G.L., Lee, S.J., Choi, B.C., Kim, J., Kim, K.R., Choi, S.W., et al. (2015). Quality control of agro-meteorological data measured at Suwon weather station of Korea Meteorological Administration. *Korean Journal of Agricultural and Forest Meteorology*, Vol. 17, No. 1, pp. 25-34.
- Oh, S.R., Kim, J.Y., Choe, Y.J., and Choe, H.J. (2014). A study on the improvement of hydrologic data quality in water level observation sector. *Magazine of Korea Water Resources Association*, Vol. 47, No. 7, pp. 74-80.
- Ryu, B.H., and Han, C.S. (2019). LSTM based hydraulic excavator angular velocity prediction model. *Journal of Institute of Control, Robotics and Systems*, Vol. 25, No. 8, pp. 705-712.
- Shin, D.H., Choi, K.H., and Kim, C.B. (2017). Deep learning model for prediction rate improvement for stock price using RNN and LSTM. *Journal of Korean Institute of Information Technology*, Vol. 15, No. 10, pp. 9-16.
- Shin, J.H., Nam, W.H., Bang, N.K., Kim, H.J., An, H.U., Do, J.W., et al. (2020). Assessment of water distribution and irrigation efficiency in agricultural reservoirs using SWMM model. *Journal of the Korean Society of Agricultural Engineers*, Vol. 62, No. 3,

pp. 1-13.

Yang, M.H., Nam, W.H., Kim, T., Lee, K., and Kim, Y. (2019). Machine learning application for predicting the strawberry harvesting time. *Korean Journal of Agricultural Science*, Vol. 46, No. 2, pp. 381-393.

<i>Received</i>	November 30, 2020
<i>Revised</i>	November 30, 2020
<i>Accepted</i>	December 15, 2020