시계열 분석 딥러닝 알고리즘을 적용한 낙동강 하굿둑 염분 예측 Prediction of Salinity of Nakdong River Estuary Using Deep Learning Algorithm (LSTM) for Time Series Analysis

우정운* · 김연중** · 윤종성*** Joung Woon Woo*, Yeon Joong Kim** and Jong Sung Yoon***

요 지: 낙동강 하굿둑은 올해 2022년 해수 유입기간을 매월 대조기마다로 확대, 하굿둑 상류 15 km 이내로 기수역 조성을 목표로 운영되고 있다. 목표 기수역 조성구간 및 염수피해 방지를 위한 신속한 의사결정을 위해 본 연구에서는 딥러닝 알고리즘 Long Short-Term Memory(LSTM)을 적용하여 낙동대교(하굿둑 상류 약 5 km)지점의 염분 예측을 수행하였다. 창녕·함안보 방류량 등 낙동강 하구역의 시·공간적 특성을 반영하기 위한 입력데이터를 구축하였으며, Sequence length에 따른 정도 변화를 통해 낙동강 하구역의 수리학적 특성을 고려한 최적모델을 구축하였다. 예측 정확도는 결정계수(R-squred)와 RMSE(root mean square error) 이용하여 통계분석을 실시하였으며. Sequence length가 12일 때 R-squred 0.997, RMSE 0.122로 가장 정도가 높았으며, 선행 예측시간은 12시간 간격까지 R-squred 0.93 이상으로 높은 정도를 보였다.

핵심용어 : 염분예측, 딥러닝, LSTM, 낙동강, 시계열 예측

Abstract : Nakdong river estuary is being operated with the goal of expanding the period of seawater inflow from this year to 2022 every month and creating a brackish water area within 15 km of the upstream of the river bank. In this study, the deep learning algorithm Long Short-Term Memory (LSTM) was applied to predict the salinity of the Nakdong Bridge (about 5 km upstream of the river bank) for the purpose of rapid decision making for the target brackish water zone and prevention of salt water damage. Input data were constructed to reflect the temporal and spatial characteristics of the Nakdong River estuary, such as the amount of discharge from Changnyeong and Hamanbo, and an optimal model was constructed in consideration of the hydraulic characteristics of the Nakdong River Estuary by changing the degree according to the sequence length. For prediction accuracy, statistical analysis was performed using the coefficient of determination (R-squred) and RMSE (root mean square error). When the sequence length was 12, the R-squred 0.997 and RMSE 0.122 were the highest, and the prior prediction time showed a high degree of R-squred 0.93 or more until the 12-hour interval.

Keywords: salinity prediction, deep learning, LSTM (long short-term memory), Nakdong river, time series prediction

1. 서 론

낙동강 하굿둑은 1987년 준공되어 조수차에 의한 해수와 담수의 순환을 인위적으로 차단하여 낙동강 하구 지역과 부산시민의 식수를 비롯한 생활용수, 농업용수, 인근 지역 공단의 공업용수 공급을 가능하게 하였다(Kim, 2020). 그러나 하굿둑 건설 이후 철새 개체 감소 및 출현 어종의 단순화와 같은 하구 생태계와 수질 악화 문제들이 꾸준히 제기되고 있다.이에 정부는 2017년 국정운영 5개년 계획·국정과제(환경부)로 '낙동강 하구 환경관리를 위한 실무협의회'를 구성하여 2017~2021년 낙동강 하굿둑 시범운영을 통한 실증실험 수행

하였다. 특히 2021년 4차례 시범운영을 통해 계절별 수문운영 방식과 수문 개도율 및 개도방식에 따른 해수유입량을 실측하였다. 이러한 결과를 바탕으로 다양한 상황에 따른 의사결정 및 모니터링 방식을 체계화하였다. 이후 2022년 2월 기수역 조성이 공식화됨에 따라 해수 유입 기간을 매월 대조기로 확대하였으며, 이로 인한 염수 피해 방지 및 기수역 조성의사결정을 하기 위하여 실시간 계측장비와 수치해석을 통한정보를 활용하고 있다(Jeong et al., 2022). 현재 기수역 조성목표 구간은 대저수문 이전의 하굿둑 상류 15 km이며, 염분이 하굿둑 상류 10~12 km 도달 시 해수유입을 중단하여 염수피해를 방지하고 있다. 이에 염수피해 방지를 위한 신속한

^{*}인제대학교 건설기술연구소 전임연구원(Associate Researcher, Department of Civil and Environmental Engineering (Institute of Construction Technology Center), Inje University)

^{**}인제대학교 건설환경공학부 연구교수(Research Doctor, Department of Civil and Environmental Engineering, Inje University)

^{***}인제대학교 건설환경공학부 교수(Corresponding author: Jong Sung Yoon, Professor, Department of Civil and Environmental Engineering, Inje University, 197, Inje-ro, Gimhae-si, Gyeongsangnam-do 50834, Korea, Tel: +82-55-320-3694, civyunjs@inje.ac.kr)

의사결정을 위해서는 염수의 도달거리 및 도달시간에 대한 빠른 예측이 필요하다.

Blumberg and Mellor(1987)은 유체역학 방정식과 수송 방정식의 결합을 통해 하구에서의 밀도류 2차원 해수유동 모델을 개발하였고, Parsa et al.(2007)은 다양한 수치해석 모델들을 Bahmanshir Estuary에 적용하여 염분 예측 정도를 비교하였다. 이후 EFDC(Environmental Fluid Dynamics Code) 등다양한 수치해석 프로그램을 활용한 하구역에서의 염수거동에 대한 연구가 활발히 이루어지고 있다. Kim(2020)은 EFDC를 이용한 낙동강 하굿둑 개도율에 따른 염수 소상거리 및 창녕·함안보 방류량에 따른 도달시간 및 염분 플러싱에 관한 연구를 진행하였으며, Han et al.(2011)은 ECOMSEM을 활용하여 낙동강 하구둑 수문 완전 개방시 유입 유량과 조위에 따른 염분 침투 현상에 대한 민감도 분석을 실시하였다. Ahn et al.(2020)은 섬진강 하구역(약 78 km)의 광범위한 영역의대한 수치해석을 실시하였으며, 하천 유량과 조석에 따른 염수 침투 거리 분석을 실시하였다.

물리 모델에 기반한 수질예측의 경우 정밀하고 광범위한 예측은 가능하지만, 2차원 및 3차원 흐름을 고려해야 하기 때문에 계산이 복잡하고 계산시간이 긴 단점이 있다. 최근 국내·외에서 AI(Artificial Intelligence)를 통한 수질예측 연구가활발히 진행되고 있는데, 이는 물리모델에 비해 AI모델의 계산구조가 간단하고 방대한 데이터를 효과적으로 처리할 수 있기 때문이다(Melesse et al., 2020; Yang et al., 2020; Kim and Lee, 2022).

Chen et al.(2017)은 인공신경망(Artificial Neural Network, ANN) 모델과 3차원 수치해석 모델 비교 연구를 수행하였는데, 단기간 염분 예측 비교 결과 인공신경망 모델이 3차원 수치해석 모델보다 좋은 정도를 보이고 있다. Barzegar et al. (2020)은 DO와 클로로필a(Chl-a)의 두 가지 수질 변수 예측을 위해 CNN-LSTM 결합모델과 벡터자기회귀 모델(Vector Auto Regressive Model, VAR) 등 다양한 딥러닝 알고리즘모델을 구축하고 모델 예측정도를 비교하는 연구를 수행하였다. 시계열 데이터 분석에 있어서 LSTM(Long Short-term Memory)의 예측 정도가 가장 우수한 것으로 나타났다. 국내의 경우 ANN 모델을 이용한 청평댐 하류의 수질인자 예측을

통해 8개 수질인자 중 7개 인자에 대한 유의미한 예측 성과를 보았으며(Seo and Yun, 2016), 머신러닝(Machine Learning) 기법을 활용하여 낙동강 중류 지역의 수질인자 및 유랑을 입력데이터로 사용한 클로로필a(Chl-a) 예측 모델개발에 대한 연구를 수행하였다(Lee et al., 2020). 해양수질자동측정망 데이터를 사용하여 hidden node, epch, 예측시간에 따른 재현 정확도 비교를 통한 최적화 모델 개발을 수행하였다(Park and Kim, 2021). 하지만 머신러닝 및 딥러닝을 통한 하구에서의 염분예측 연구는 부족한 실정이다.

본 연구에서는 LSTM 모델을 사용하여 낙동강 하구역의 시·공간적 특성을 고려한 염분농도를 예측하였다. 창녕·함안보 방류량의 도달시간과 염분 농도의 상관관계를 예측모형의 Sequence length에 따라 재현 정도를 비교하였고, 예측 선행시간에 따른 재현 정확도 비교를 통한 LSTM모델의 장단기 예측성능을 평가하였다.

2. 연구 방법

2.1 Long Short-Term Memory (LSTM)

RNN(Recurrent Neural network)은 입력과 출력을 연관된 연속의 데이터 단위로 처리하는 모델이다. LSTM 모델은 순환신경망 모델에서 발생하는 기울기 소실(gradient vanishing) 문제를 특수한 게이트 순환 구조로 해결하고 장기적인 시간의 종속성(long-term dependencies)을 다루는데 효과적이다. 이에 시계열예측, 번역, 음성인식뿐만 아니라 위에 설명한 CNN-LSTM과 같이 다른 신경망과의 결합을 통해 다양한 분야에 활용되고 있다(Hochreiter and Schmidhuber, 1997; Barzegar et al., 2020).

Fig. 1에서 보듯이, LSTM에서는 기중치 반영 및 활성화 함수 변환을 통하여 입력값에서 출력값으로 변환하는 단계를 하나의 셀(cell)로 보고, 셀 내부의 상태량인 셀 스테이트(cell state)를 입력게이트(input gate), 망각게이트(forget gate), 출력게이트(output gate)를 이용하여 총 4단계의 계산과정을 통해 가중치 소실문제가 발생하지 않도록 조절한다. LSTM 첫 번째 단계에서는 특정 정보의 제거 여부를 망각게이트(f_i)를 통하여 식(1)과 같이 결정한다.

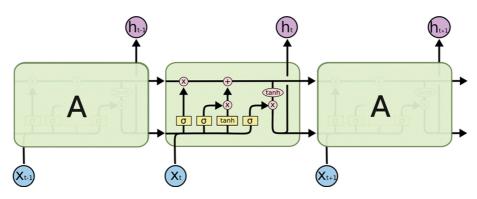


Fig. 1. Structure of LSTM.

$$f_t = \sigma(W_t[h_{t-1}, x_t] + b_t)$$
 (1) $h_t = o_t \tanh(C_t)$

여기서 f_i 는 0, 1의 값을 가지는 망각게이트 값, W_j 는 망각게이트 가중치, b_j 는 망각게이트 편향값, o는 시그모이드 함수이다.

두 번째 단계에서는 입력게이트 (i_i) 를 통하여 새로운 정보의 저장 여부를 식(2)과 같이 결정하며, i_i 는 0 또는 1의 값을 가지는 입력게이트 값, W_i 는 입력게이트 가중치, b_i 는 입력게이트 편향값이다.

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

세 번째 단계에서는 입력게이트와 출력게이트의 값을 이용하여 셀 스테이트를 식(3), 식(4)와 같이 업데이트 한다. 여기서, \tilde{C}_i 는 tanh로 구성되어 있는 셀 스테이트 중간값이며, C_i 는 업데이트된 시점 t에서의 셀 스테이트를 나타낸다.

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$
(3)

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{4}$$

마지막 단계는 출력게이트와 셀 스테이트를 이용하여 출력 값을 다음 식(5), 식(6)와 같이 계산한다.

$$o_{t} = \sigma(W_{o}[h_{t-1}, x_{t}] + b_{o})$$
(5)

(6)

2.2 실험모형 구축 및 성능평가 방법

본 연구에서는 LSTM 모형을 구축하여 낙동강 하구역에서 수위 및 창녕·함안보 방류량에 따른 염분 예측을 수행하고 Sequence length과 선행시간(time lead)에 따른 예측 정도를 분석하는 것을 목적으로 한다. 각 case 조건은 Table 1과 같다. 예측 1시간에 대한 학습 Sequence length는 1시간, 3시간, 6시간, 9시간, 12시간, 24시간으로 하고, 그 중 가장 높은 정도를 가지는 Sequence lengh 12시간을 사용하여 선행시간 3시간(t+3), 6시간(t+6), 9시간(t+9), 12시간(t+12), 24시간(t+24)에 대한 재현 정도를 비교하였다. epoch 횟수는 100, optimizer는 Adam, learning rate는 0.001, Dropout 비율은 20%로 설정하였다. RMSE(Root mean square error)

Table 1. Experiment cases condition

Case 1						
Sequence length	Optimizer	Epoch	Learning rate	Dropout	Time lead	
1, 3, 6, 9, 12, 24 (hours)	Adam	100	0.001	0.2	1 hour	
Case 2						
Time lead	Optimizer	Epoch	Learning rate	Dropout	Sequence length	
3, 6, 9, 12, 24 (hours)	Adam	100	0.001	0.2	12 hour	

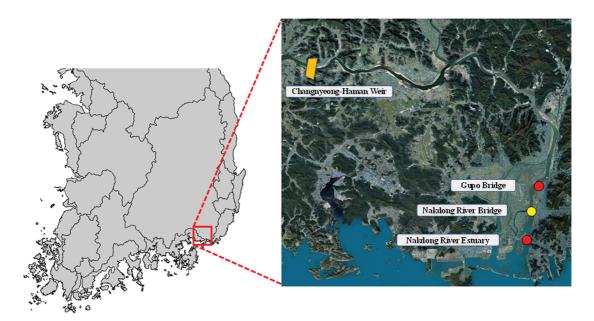


Fig. 2. Study area.

Feature		Mean	Max	Min
Nakdong River Bridge	Bottom Salinity (psu) Bottom EC (μs) Bottom Temperature (°C)	0.80 1,487.03 16.23	12.10 20,258.00 30.30	0.07 145.00 2.40
Changnyeong-Haman Weir	Discharge (m³/s) Water level (m) Rainfall (mm)	182.49 4.35 0.13	999.64 11.38 12.80	0.00 2.03 0.00
Gupo Bridge	Water level (m)	2.39	4.87	1.59
Nakdong River Estuary	Upstream Water level (m) Downstream Water level (m) Rainfall (mm)	0.85 0.13 0.13	1.76 1.47 14.8	-0.17 -1.06 0.00

Table 2. The mean, max, min of features before data pre-processing

와 R-Squared로 모델의 예측 정도를 평가하였으며 해당 식은 다음과 같다. RMSE는 0에 수렴할수록 R-Squared는 1에 수렴할수록 좋은 예측력을 가진다고 판단할 수 있다.

$$RMSE = \sqrt{(MSE)} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y})^2}$$
 (7)

$$R^{2} = 1 - \frac{\Sigma (y_{i} - \hat{y})^{2}}{\Sigma (y_{i} - \overline{y})^{2}}$$
(8)

2.3 연구 대상지 및 입력자료 구축

본 연구의 대상유역은 낙동강의 창녕·함안보에서부터 낙동 강 하굿둑까지 약 80 km 구간으로 설정하였다. 자료는 약 3 년간(2019.01.01.~2020.12.31.)의 정점별 1시간단위 데이터이 며, 염분데이터는 부산환경연구원에서 제공하는 낙동대교지 점의 저층데이터를 사용하였고 수리·수문데이터는 My Water (물정보포털)의 창녕·함안보, 구포대교, 낙동강 하굿둑지점 데 이터를 사용하였다(Fig. 2). 데이터 항목으로는 낙동대교의 염 분도, 전기전도도, 수온과 구포대교 수위, 하굿둑 내수위, 외 수위, 강우량 데이터를 사용하였으며, 창녕함안보의 총방류량, 수위와 강우량 데이터로 총 10가지 항목으로 구성되어 있다. 각 데이터의 결측 및 이상치는 선형보간하여 전처리하였다. 학 습데이터는 2년치를 사용하였으며, 1년치를 모델 검증에 사 용하였다. 자료는 염분도(PSU)와 전기전도도의 Value Range에 큰 차이가 있어 Min-Max Scaling을 통한 정규화(Normalization) 로 데이터 전처리 하였고, 정규화에 사용된 식은 다음과 같 다. 전처리 전의 각 항목의 평균, 최대, 최소는 Table 2에 제 시하였다.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{9}$$

3. 결과 및 고찰

3.1 Sequence length에 따른 염분 농도 예측평가

Sequence length에 따른 모델의 예측 정확도는 Table 3

Table 3. Model Performance Comparison about Sequence Length

Case 1	RMSE	R-Squared
Case1-SL-1	0.140	0.996
Case1-SL-3	0.132	0.997
Case1-SL-6	0.150	0.995
Case1-SL-9	0.135	0.996
Case1-SL-12	0.122	0.997
Case1-SL-24	0.143	0.996

에 나타내었으며, 그에 따른 검증데이터 1년(2020년도)에 대한 염분농도를 Fig. 3에 나타내었다. Sequence length에 따른 모델 예측 정도의 차이는 거의 없는 것으로 나타났다. 실험케이스의 정도는 R-Squared 값이 0.995~0.997, RMSE 값은 0.122~0.150(psu)로 상당히 높은 정도를 보이는 것으로 나타났다. Case1-SL-12시간일 때 R-Squared 값이 0.997, RMSE 값이 0.122 psu로 가장 예측정도가 좋았으며, Case1-SL-24일 때 예측 정도가 가장 낮았다. 이는 낙동대교에서의 염분농도가 창녕·함안보 방류에 의한 도달시간(약12시간)과 하굿둑 수문운영에 따른 염분 침투시간(약12시간)의 영향을 받는 것으로 보이며, 이는 낙동강 하구역에서의 수리학적 특성을 잘 나타내는 것으로 판단된다(Kim, 2020).

3.2 선행시간(time lead) 따른 염분 농도 예측평가

앞선 실험 모델 중 가장 높은 정도를 보이는 Case1-SL-12에 대하여 선행시간(time lead)에 따른 염분예측 정도를 Table 4에 정리하였으며, 검증기간에 따른 염분농도 비교를 Fig. 4에 나타내었다. 선행시간의 간격이 길어질수록 정도가 떨어지는 것으로 나타났다. 각 Case별 정도는 3~24시간순으로 R-Squared 값이 [0.98, 0.96, 0.95, 0.93, 0.74]으로 나타났으며, RMSE 값은 [0.296, 0.481, 0.486, 0.592, 1.012]로 나타났다. 선행시간이 24시간일 때 R-Squared 값이 0.74, RMSE 값이 1.012 psu로 가장 정확도가 낮았으며, 선행시간 24시간 이후 급격하게 정도가 떨어지는 것으로 보인다. 이는 입력데이터로 사용되는 과거 Sequence에 대한길이가 예측 선행시간보다 짧아 정도가 떨어진 것으로 판

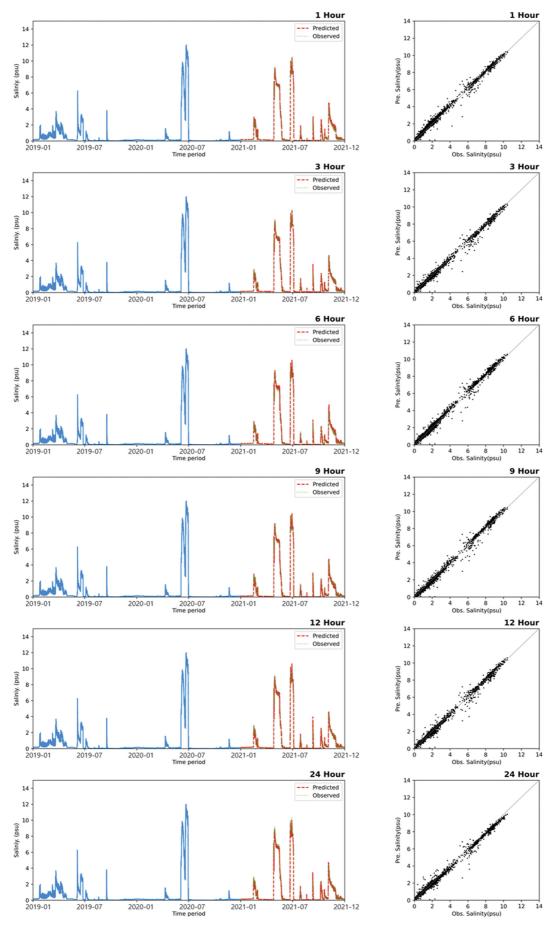


Fig. 3. Predicted and observed values of Case 1 (Sequence length).

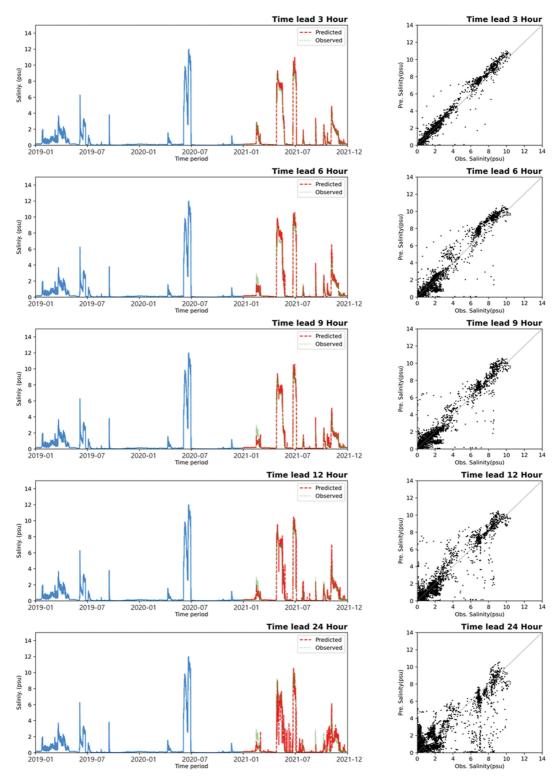


Fig. 4. Predicted and observed values of Case 2 (Time lead).

Table 4. Model Performance Comparison about Time Lead

Table Williams Companies accur into Zeus					
RMSE	R-Squared				
0.296	0.98				
0.481	0.96				
0.486	0.95				
0.592	0.93				
1.012	0.71				
	RMSE 0.296 0.481 0.486 0.592				

단되며, 이는 입력데이터의 Sequence length 및 입력데이터 다양하를 통한 연구가 필요할 것으로 보인다. 단기간 예측에 대한 정도는 R-Squared 0.9 이상으로 뛰어난 것으로 판단되며, 이는 낙동대교 염분농도 예측을 통해 낙동강 하 굿둑의 수문운영 의사결정에 신속하게 적용 가능할 것으로 판단된다.

4. 결 론

본 연구에서는 LSTM 모델을 활용하여 낙동강 하구역의 낙동대교지점 염분도 예측을 수행하였다. 낙동강 하구역의 수리·수문학적 특성을 고려하기 위해 Sequence length에 따른 예측 정도를 비교하였으며, 모델 결과 Case1-SL-12시간일 때가장 높은 정도를 보이는 것으로 나타났다. 이는 낙동대교에서의 염분농도와 창녕·함안보 유량 도달시간 및 하굿둑 염분 첨투시간과의 상관관계를 모델이 잘 재현하는 것으로 판단된다(Kim, 2020). 또한, 선행시간(time lead)에 따른 예측정도는 12시간 간격까지는 R-Squared 값이 0.93으로 높은 정도를 보이고 있으며, 24시간에서의 예측 정도는 0.71로 급격하게 떨어지는 것으로 나타났다. 이는 LSTM 모형이 12시간 이하의 단기 예측에는 높은 정도를 보이나, 24시간 이상의 장기예측에는 정도가 떨어지는 것으로 보인다.

향후 낙동강 기수역 조성에 따른 염수피해 방지를 위한 신속한 수문운영 의사결정을 위해 낙동강 하구역의 여러 염분 관측지점 데이터를 학습하고 낙동강 하구만의 수리·수문학적 특성을 고려한 염분예측 모델 연구가 필요하다고 판단된다. 그리고 장기예측 정도 향상을 위해 학습데이터간의 상관관계 및 EDA(Exploratory Data Analysis)를 통한 데이터 전처리연구가 꼭 필요할 것으로 보인다.

감사의 글

본 연구는 2020년 한국연구재단의 이공분야기초연구사업 (NRF-2020R1I1A3A0403784313)의 재원으로 수행된 연구결과 중 일부임을 밝히며, 연구비 지원에 감사드립니다.

References

- Ahn, J.M., Lee, K. and Lyu, S. (2020). Effect of changes in watershed runoff characteristics on salinity intrusion in estuary using EFDC. KSCE J. Civ. Eng., 24(1), 87-98.
- Barzegar, R., Aalami, M.T. and Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. Stochastic Environmental Research and Risk Assessment, 34(2), 415-433.
- Blumberg, A.F. and Mellor, G.L. (1987). A description of a threedimensional coastal ocean circulation model. Three-dimensional Coastal Ocean Models, 4, 1-16.
- Chen, W., Liu, W., Huang, W. and Liu, H. (2017). Prediction of salinity variations in a tidal estuary using artificial neural net-

- work and three-dimensional hydrodynamic models. Comput. Water Energy Environ. Eng., 06, 107-128.
- Han, C.-S., Park, S.-K., Jung, S.-W. and Roh, T.-Y. (2011). The study of salinity distribution at nakdong river estuary. J. Korean Soc. Coast. Ocean Eng., 23(1), 101-108 (in Korean).
- Hochreiter, S. and chmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
- Jeong, S., Lee, S., Hur, Y.T., Kim, Y. and Kim, H.Y. (2022). Development of seawater inflow equations considering density difference between seawater and freshwater at the Nakdong River estuary. J. Korea Water Resources Association, 55(5), 383-392 (in Korean).
- Kim, T. (2020). Study on th Behavior Analysis of Salt Water According to the Operation of Nakdong River Estuary Barrage. Inje University.
- Kim, T. and Lee, W.-D. (2022). Review on applications of machine learning in coastal and ocean engineering. J. Ocean Eng. Technol., 36(3), 194-210.
- Lee, S.-M., Park, K.-D. and Kim, I.-K. (2020). Comparison of machine learning algorithms for Chl-a prediction in the middle of Nakdong River (focusing on water quality and quantity factors). J. Korean Soc. Water Wastewater, 34, 277-288 (in Korean).
- Melesse, A.M., Khosravi, K., Tiefenbacher, J.P., Heddam, S., Kim, S., Mosavi, A. and Pham, B.T. (2020). River water salinity prediction using hybrid machine learning models. Water, 12, 2951.
- Park, S. and Kim, K. (2021). Prediction of DO concentration in nakdong river estuary through case study based on long short term memory model. J. Korean Soc. Coast. Ocean Eng., 33(6), 238-245 (in Korean).
- Parsa, J., Etemad-Shahidi, A. and Hosseiny, S. (2007). Evaluation of computer and empirical models for prediction of salinity intrusion in the bahmanshir estuary. J. Coast. Res., 50(SI), 658-662.
- Seo, I.W. and Yun, S.H. (2016). Forcasting water quality by ANN model at the downstream of cheongpyeong dam. KSCE 2016 CONVENTION PROGRAM, 41-42.
- Yang, H., Lee, K., Choo, Y. and Kim, K. (2020). Underwater acoustic research trends with machine learning: Ocean parameter inversion applications. J. Ocean Eng. Technol., 34(5), 371-376.

Received 11 July, 2022

1st Revised 16 August, 2022

2nd Revised 24 August, 2022

Accepted 24 August, 2022