



Comparative analysis of wavelet transform and machine learning approaches for noise reduction in water level data

Hwang, Yukwan^a · Lim, Kyoung Jae^b · Kim, Jonggun^c · Shin, Minhwan^d · Park, Youn Shik^e · Shin, Yongchul^f · Ji, Bongjun^g

^aGraduate Student, Regional Infrastructure Engineering, Kangwon National University, Chuncheon, Korea

^bProfessor, Department of Regional Infrastructure Engineering, Kangwon National University, Chuncheon, Korea

^cAssistant Professor, Department of Regional Infrastructure Engineering, Kangwon National University, Chuncheon, Korea

^dCEO, EM Research Institute, Chuncheon, Korea

^eProfessor, Department of Regional Construction Engineering, Kongju National University, Yesan, Korea

^fAssociate Professor, Department of Agricultural Civil Engineering, Kyungpook National University, Daegu, Korea

^gAssistant Professor, Department of Regional Infrastructure Engineering, Kangwon National University, Chuncheon, Korea

Paper number: 24-008

Received: 8 February 2024; Revised: 6 March 2024; Accepted: 7 March 2024

Abstract

In the context of the fourth industrial revolution, data-driven decision-making has increasingly become pivotal. However, the integrity of data analysis is compromised if data quality is not adequately ensured, potentially leading to biased interpretations. This is particularly critical for water level data, essential for water resource management, which often encounters quality issues such as missing values, spikes, and noise. This study addresses the challenge of noise-induced data quality deterioration, which complicates trend analysis and may produce anomalous outliers. To mitigate this issue, we propose a noise removal strategy employing Wavelet Transform, a technique renowned for its efficacy in signal processing and noise elimination. The advantage of Wavelet Transform lies in its operational efficiency - it reduces both time and costs as it obviates the need for acquiring the true values of collected data. This study conducted a comparative performance evaluation between our Wavelet Transform-based approach and the Denoising Autoencoder, a prominent machine learning method for noise reduction. The findings demonstrate that the Coiflets wavelet function outperforms the Denoising Autoencoder across various metrics, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE). The superiority of the Coiflets function suggests that selecting an appropriate wavelet function tailored to the specific application environment can effectively address data quality issues caused by noise. This study underscores the potential of Wavelet Transform as a robust tool for enhancing the quality of water level data, thereby contributing to the reliability of water resource management decisions.

Keywords: Water level data, Denoising Autoencoder, Wavelet Transform, Coiflets function

웨이블릿 변환과 기계 학습 접근법을 이용한 수위 데이터의 노이즈 제거 비교 분석

황유관^a · 임경재^b · 김종건^c · 신민환^d · 박윤식^e · 신용철^f · 지봉준^g

^a강원대학교 지역건설공학과 석사과정, ^b강원대학교 지역건설공학과 교수, ^c강원대학교 지역건설공학과 조교수

^d이연연구소 대표, ^e공주대학교 지역건설공학과 교수, ^f경북대학교 농업토목공학전공 부교수, ^g강원대학교 지역건설공학과 조교수

요지

4차 산업혁명 시대에 접어들어 데이터 기반의 의사결정이 보편화되고 있다. 하지만 데이터 품질이 확보되지 않은 채 수행되는 데이터 분석은 왜곡된 결과를 낳을 가능성이 존재한다. 수자원 관리의 기초가 되는 수위 데이터도 마찬가지로 결측, 스파이크, 잡음 등 다양한 품질 문제를 가진다. 본 연구에서는 잡음으로 인해 발생하는 데이터 품질 문제를 해결하고자 하였다. 잡음은 데이터의 트렌드 분석을 어렵게 하고 비정상적인 이상치를 생성할 가능성이 있다. 본 연구는 이러한 문제를 해결하기 위해 Wavelet Transform을 이용한 잡음 제거 접근 방안을 제안한다. Wavelet Transform은 신호처리에 주로 사용되는 방법으로 잡음 제거에 효과적인 것으로 알려져 있으며 수집된 데이터의 정답 데이터(True value) 수집을 요구하지 않으므로 시간과 비용을 줄일 수 있다는 점에서 적용이 용이한 편이다. 본 연구는 Wavelet Transform의 성능 평가를 위해 대표적인 머신러닝 기반 잡음 제거 방법인 Denoising Autoencoder와 성능 비교를 수행하였다. 그 결과 Wavelet Transform 중 Coiflets 함수는, Denoising Autoencoder에 비해 Mean Absolute Error, Mean Absolute Percentage Error, Mean Squared Error 등 모든 측면에서 우수한 성능을 보이는 것으로 나타났다. 이러한 결과는 환경에 맞는 적절한 웨이블릿 함수의 선택을 통한 잡음 문제를 효과적으로 해결할 수 있음을 시사한다. 본 연구는 수위 데이터의 품질을 향상시켜 수자원 관리 결정의 신뢰성에 기여하는 강력한 도구로서 Wavelet Transform의 잠재력을 확인한 의의가 있다.

핵심용어: 수위 데이터, 잡음제거 오토인코더, 웨이블릿 변환, 코이플릿 함수

*Corresponding Author. Tel: +82-33-250-6464

E-mail: bjjj@kangwon.ac.kr (Ji, Bongjun)

1. 서론

지구 온난화 등으로 이상 기후 현상이 빈번히 발생하고 있다. 특히 우리나라는 태풍과 집중호우 발생으로 인한 홍수, 범람 피해의 발생빈도와 피해 규모가 커지고 있는 실정이다 (Kang *et al.*, 2007). 최근 10년(2011~2020년)간 호우와 태풍으로 인한 피해액이 전체 자연재해 피해 금액의 94%를 차지하며, 인명피해는 전체의 63%를 차지할 정도로 비중이 높은 것으로 나타났다. 특히, 이 기간 동안 홍수로 인한 피해가 발생한 하천의 개수와 그 횟수는 2020년에만, 전체의 64%, 52%로 나타났다(Lee *et al.*, 2022). 이처럼 강우로 인해 발생할 수 있는 피해를 예방하고 극한 상황을 가정하기 위해 다양한 모델링 기반의 접근방법이 사용되고 있다. GIS (Geographic Information System)(Escobar *et al.*, 2008) 모델링 기법을 이용하여 강우 발생 시 변화와 피해 규모 등을 예측해 왔다. 하지만 이러한 해석 모델링 기반의 접근방식은 몇 가지 한계점이 존재한다. 대표적인 한계점으로는 장기적 기후변화를 고려하지 못하고 짧은 기간의 기후변화만 예측하여 고려할 수 있다는 점이다(Reilly and David, 2000). 두 번째로는 시간에 따른 지형지물의 변화를 모델링에 반영하기 어렵다는 것이다 (Alan, 2010). 즉, 시간이 지남에 따라 도로, 건물, 자연 지형 등 지형지물이 변화하는데 GIS 기반의 접근방법은 이를 모델에 반영하지 않으면 모델의 정확성이 떨어지게 되어 예측하고자 하는 혹은 도출하고자 하는 결과의 정확성을 보장하기 힘들다. 세 번째는 모델링함에 있어 시간과 비용이 많이 소요된다는 점이다(Öberg *et al.*, 2020). 앞서 언급한 한계점들을 극복하고 현실성 높은 모델을 구현하기 위해서는 많은 파라미터들의 값을 산출하고 반영해야 하며 이를 위해 필요한 시간과 비용이 상당하다. 특히 기존에 사용하던 GIS 모델의 성능을 유지하거나 개선하기 위해 모델을 업데이트 하는 것도 많은 투자를 필요로 한다. 기존의 GIS를 이용한 해석 모델링에 대한 한계는 시간이 흘러감에 생기는 변화에 대한 고려가 어려워 유사한 데이터나 대체 값을 모색하는 데에 많은 시간과 노력이 필요하다. 이러한 해석 모델링 기반의 한계점에 대응하기 위해 머신러닝 기반의 모델들을 도입하여 새로운 해결책을 모색했다. 머신러닝에는 패턴인식, 생체인식 등 구조적인 위험을 최소화하는 목적으로 하는 Support Vector Machine (SVM)(Ko, 2005)과 대규모 데이터베이스의 특징을 발견하고 패턴을 추출하는 의사결정나무(Decision Tree)(Anthony *et al.*, 2004) 등 종류가 다양하다. 머신러닝은 데이터로부터 패턴을 학습하여 데이터를 분류, 해석하여 향후의 데이터값을 예측할 수 있다(Cho *et al.*, 2022). 머신러닝은 시계열 데이

터를 입력하고 처리하는 과정에 어려움이 있었고 이를 개선하기 위해 머신러닝 기법을 접목하여 연구되고 있다. 다양한 머신러닝 기반의 모델 중 기후변화 예측의 어려움과 지형지물의 변화 등, 시간에 따른 변화에 대한 문제에 뛰어난 효과가 있는 RNN (Recurrent Neural Network)의 한 종류인 LSTM (Long Short-Term Memory) Networks (Yu *et al.*, 2019)란 신경망을 사용한 방법이 있다. LSTM 신경망은 RNN의 한계인 장기 의존성 문제를 극복하기 위해 은닉층에 셀 상태(cell state)구조를 추가한 신경망이다. 이 셀 상태는 정보를 장기적으로 기억하여 과거 정보와 현재 정보에 대한 업데이트를 수행하면서 향후 시간의 흐름에 따른 기후 및 지형 등의 변화를 예측할 수 있을 뿐만 아니라 데이터의 과적합(Overfitting) 문제 또한 완화한다(Kim *et al.*, 2020). LSTM 신경망을 활용하여 예측할 수 있음에 따라 시간과 비용 역시 감소하게 되었다. 머신러닝 기반의 접근방법은 데이터를 수집한 후 수집된 데이터에서 특성을 추출하여 학습하여 결과를 도출해 낸다. 이러한 머신러닝 기반 방법을 통해 확보한 데이터의 품질은 정확성과 신뢰성이 있는 데이터라고 가정된 상태에서 특성을 추출한다. 그러나 우리가 확보할 수 있는 데이터들의 품질은 항상 높은 품질이라고 확정할 수 없다는 또 다른 문제가 있다(Kim, 2019). 그렇기 때문에 확보한 데이터에 대해서 데이터 품질 확보가 중요하게 되었고 여러 분야에서 품질 확보를 위해 다양한 방법들이 사용된다(Choi *et al.*, 2023). 본 논문에서는 수위 데이터에서 미세하게 발생하는 잡음(Noise)에 대한 데이터 품질 향상을 목적으로 Wavelet Transform이 적용 가능할 것이라는 가설을 설정하고 검증하였다. 데이터상에서의 잡음은 정확한 패턴을 파악하기 어려우며 분석 결과의 왜곡을 일으킬 수 있기에 이 잡음의 제거는 중요하다. 본래 Wavelet Transform은 신호 및 영상의 잡음 처리에 널리 사용되고 있으며 이러한 장점들로 금융 및 교통 등 다양한 분야에서도 적용 사례가 존재한다(Jo, 2009). 수위 데이터에서의 Wavelet Transform을 적용한 잡음제거의 성능을 확인하기 위한 기준을 선정하기 위해 Denoising Autoencoder 모델과 성과지표인 MAE (Mean Absolute Error), MSE (Mean Squared Error), MAPE (Mean Absolute Percentage Error)를 비교하여 성능 결과를 확인하고자 한다.

2. 재료 및 방법

2.1 데이터 수집

연구를 위해 수집된 수위 데이터는 수로에 레이저 센서를

설치하여 수집한 데이터이다. 수로는 강원 특별자치도 원주시 흥업면 흥업리에 위치하여 있으며 데이터 수집 기간은 2022년 11월 16일 16시부터 2023년 4월 11일 16시까지, 매 10분 간격으로 측정하였다. 데이터는 0.001 m의 해상도를 가지며 총 20,585개의 수집된 데이터에서 최저수위는 0.038 m, 최고 수위는 0.325 m로 나타났음을 확인할 수 있다(Table 1).

데이터 확인 결과, 데이터 품질에 잡음 문제가 있음을 확인할 수 있었다. 수집된 데이터 전반에 걸쳐 잡음으로 추정되는 임의의 변동성이 데이터상에서 관측되었다. 잡음은 여러 도메인에서 발생하는 주기성이 없는 무작위한 신호이다(Kuo, 2018). 수위 데이터에서는 수로의 물이 흘러가거나 이물질로 인해 발생하는 미세한 높낮이의 차이를 레이저 센서가 측정함에 따라 잡음이 발생한 것으로 추정된다. 이러한 잡음을 처리

하지 않고 수위에 대해 분석할 경우, 수위의 변동성을 과대하게 산출할 가능성이 있다.

2.2 데이터 전처리

수집된 수위 데이터에는 부분적으로 결측 데이터가 존재하였다(Fig. 1(a)). 결측 데이터는 데이터 분석의 왜곡을 야기할 수 있는 문제 중 하나이다(Graham, 2009). 특히 결측 데이터는 머신러닝 모델의 학습을 불가능하게 한다. 따라서 결측 데이터를 전처리하여 사용하였다. 결측 데이터의 전처리를 위해 선형보간법(Linear Interpolation)(Davis, 1975)을 사용하였는데 선형보간법은 결측 데이터를 사이에 둔 결측되지 않은 두 개의 데이터 포인트 사이에 직선을 그리고, 그 직선으로 결측치를 대체한다. 이를 수식으로 나타내면 Eq. (1)과 같으며 본 연구에서는 선형보간을 진행하여 결측 데이터를 대체하였다(Fig. 1(b)). 결측 데이터를 선형보간법으로 대체한 후 총 21,026개의 데이터를 실험에 활용하였다.

Table 1. Statistical characteristics of collected data

	Number of data	Mean	Median	Standard Deviation	Max	Min
Water Level (m)	20,585	0.128	0.124	0.025	0.325	0.038

$$y = y_1 + \frac{(x - x_1) * (y_2 - y_1)}{x_2 - x_1} \quad (1)$$

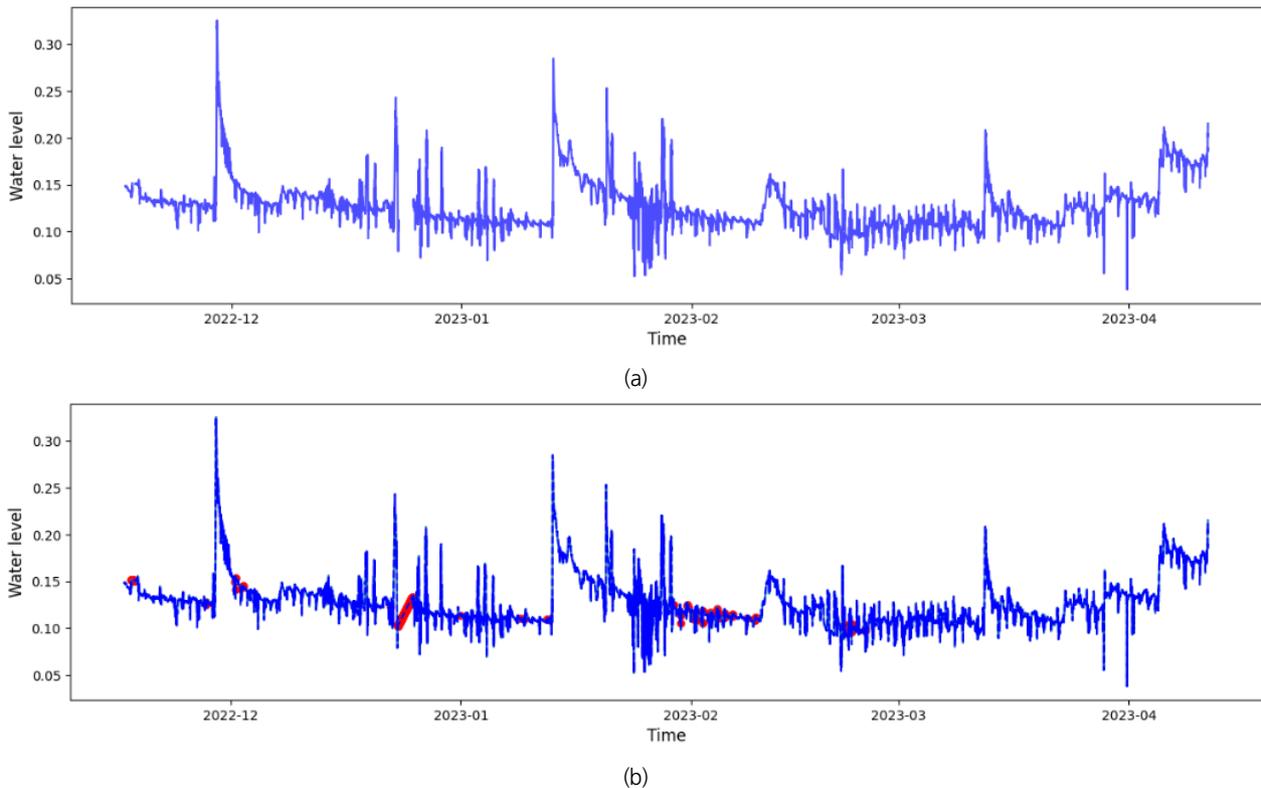


Fig. 1. Collected water level data ((a): raw data, (b): linear interpolation is applied)

3. 사용모델 및 평가방법 소개

본 연구에서 잡음 제거를 위해 사용한 방법 중 하나는 Wavelet Transform이다. 수위 데이터에 Wavelet Transform을 적용, 평가하고 최적 성능의 Wavelet 함수를 선정하기 위해 정답 데이터를 생성한 후 다양한 Wavelet 함수들의 잡음 제거 결과와 비교하여 그 성능을 평가하는 방법을 제안한다. Wavelet Transform 외에도 머신러닝 기법 중 하나인 Denoising Autoencoder 모델을 사용하였으며 해당 방법에 대해 본 장에서 서술한다. 또한 객관적인 성능 평가를 위해 널리 사용되는 성능 평가 지표에 대해서도 소개한다.

3.1 정답 데이터 생성

우선 두 모델의 잡음 제거 성능을 평가하기 위해 정답 데이터(True data)를 생성하였다. Fig. 2는 수위/수질 데이터 수집 업체의 데이터 수집 전문가가 실제 계측을 통해 수집한 값을 이용하여 잡음을 제거한 데이터이다. 해당 결과를 정답 데이터로 가정하여 Wavelet Transform과 Denoising Autoencoder를 통해 생성된 잡음 제거 성능을 평가하였다.

3.2 Wavelet transform

Wavelet Transform의 용어는 1984년 Morlet (Morlet and Grossmann, 1984)에 의해 처음으로 개발되었다. 이는 시간 및 주파수 도메인에서 신호의 특징을 동시에 분석하는 방법이다(Chun-Lin, 2010). 이는 푸리에 변환(Fourier Transform) (Fourier, 1822)과 달리 시간-주파수 해상도를 가짐으로써 신호의 특징을 더 상세하게 분석할 수 있도록 도와준다. 푸리에 변환은 시간 영역의 신호를 주파수 영역으로 변환하여 분석하기에 시간정보와 주파수 정보 간의 Trade-off가 존재하며, 시간과 주파수를 동시에 정확히 파악하는 데에 어려움이 있다. 그러나 Wavelet Transform은 시간과 주파수에 대한 정보를 단계별로 계산하면서 두 정보를 동시에 분석할 수 있어 효율적

이며 유연한 분석이 가능하다(Kwon and Lee, 2023). Wavelet Transform은 Wavelet 함수라고 불리는 특정한 함수를 사용하는 데 이 함수들은 정해진 시간 내에서 증가와 감소를 반복하고 진동하며 평균이 0인 특징을 가진다. Wavelet 함수들은 불규칙적이고 비대칭적이며 0에서 시작하여 진폭이 커지다가 작아지며 다시 0으로 수렴하는 조건이 충족하면 되기에 종류는 무한하다(Lee and Yamamoto, 1994). Wavelet Transform은 Eq.(2)를 사용하여 Wavelet 함수의 주파수를 조정하고 시간 축을 따라 이동시키며 신호를 시간에 따라 분해하는 방식으로 원본 신호와의 상관관계를 측정한다.

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \frac{t-b}{a} dt \quad (2)$$

여기서 a 는 Wavelet의 스케일링(Scaling) 인자(주파수 조정), b 는 위치변화(시간 이동) 인자, ψ 는 모(Mother) Wavelet으로 수많은 Wavelet 함수 중 선택한 함수의 쉐레복소수를 의미한다(Sifuzzaman *et al.*, 2009). Wavelet Transform은 주파수의 조정을 통해 얻은 상관관계를 분석하여 신호 및 이미지의 잡음 처리에 주로 사용된다. Wavelet Transform에 적용되는 Wavelet 함수들은 각기 다른 특징을 가지므로 목적과 데이터의 특성에 따라 선택해야 한다. Wavelet 함수에는 이산(Discrete) Wavelet과 연속(Continuous) Wavelet으로 나뉜다(Heiland Walnut, 1989). 이산 Wavelet은 Python에서 적용할 수 있는 다양한 매개변수의 Wavelet 함수를 적용했으나 연속 Wavelet은 매개변수의 개념이 존재하지 않는다. 수위 데이터에서의 최적 Wavelet 함수와 매개변수는 정해져 있지 않으므로 본 연구에서는 다양한 종류의 Wavelet 함수를 적용하고 적절한 함수 및 매개변수를 정의하고자 하였다. 본 연구에서는 연속 Wavelet 2가지(Morlets, Mexican hat)와 이산 Wavelet 4가지(Biorthogonal, Coiflets, Daubechies, Symlets)를 적용하였다. Morelets은 주파수 영역에서의 해석에 강점이 있으

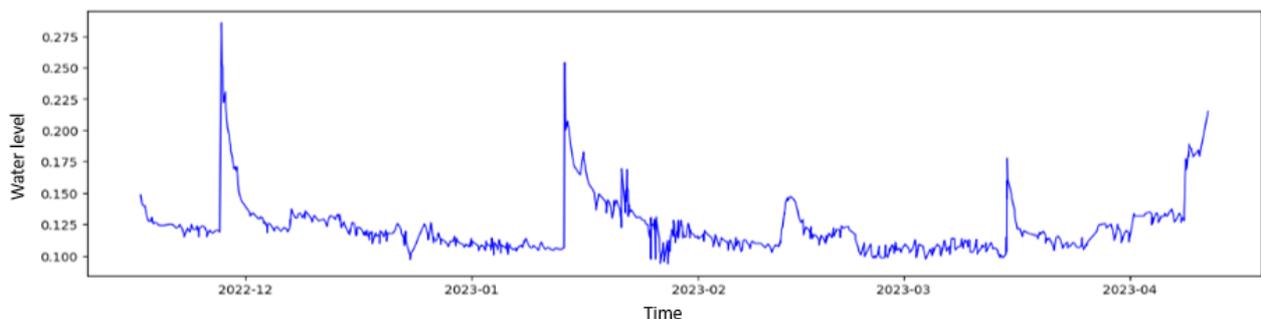


Fig. 2. Selected true data

며(De Moortel *et al.*, 2004), Mexican hat은 Morlets과 반대로 시계열 영역에서의 해석에 강점이 있다(Baliunas *et al.*, 1997). Biorthogonal은 신호 및 이미지 처리 분야에 강점이 있으며(Sridhar *et al.*, 2013), Coiflets은 부드러운 곡선이나 연속적인 신호를 표현하는 데에 강점이 있다(Oh and Seo, 2011). Daubechies은 신호가 불연속성, 자기 유사성 등을 가지고 있는 경우에 강점이 있으며(Chan and Shen, 2005), Symlets은 Daubechies 함수와 유사하지만 대칭성을 가져 음향 인식 분야에 강점이 있다(Głowacz, 2014). Python에서 다음의 6가지

Wavelet 함수들에 적용할 수 있는 매개변수에 변화를 주어 실험을 진행하였다. Fig. 3는 본 연구에서 사용한 6가지 Wavelet 함수들의 기본 그래프이다.

3.3 Denoising autoencoder

Denoising Autoencoder는 Autoencoder의 특수한 형태이다. AutoEncoder (Hinton and Salakhutdinov, 2006)는 라벨을 필요로 하지 않는 비지도 학습 방법 중 하나로 입력 데이터를 압축시킨 후 압축된 데이터를 다시 본래의 입력 형태로 복

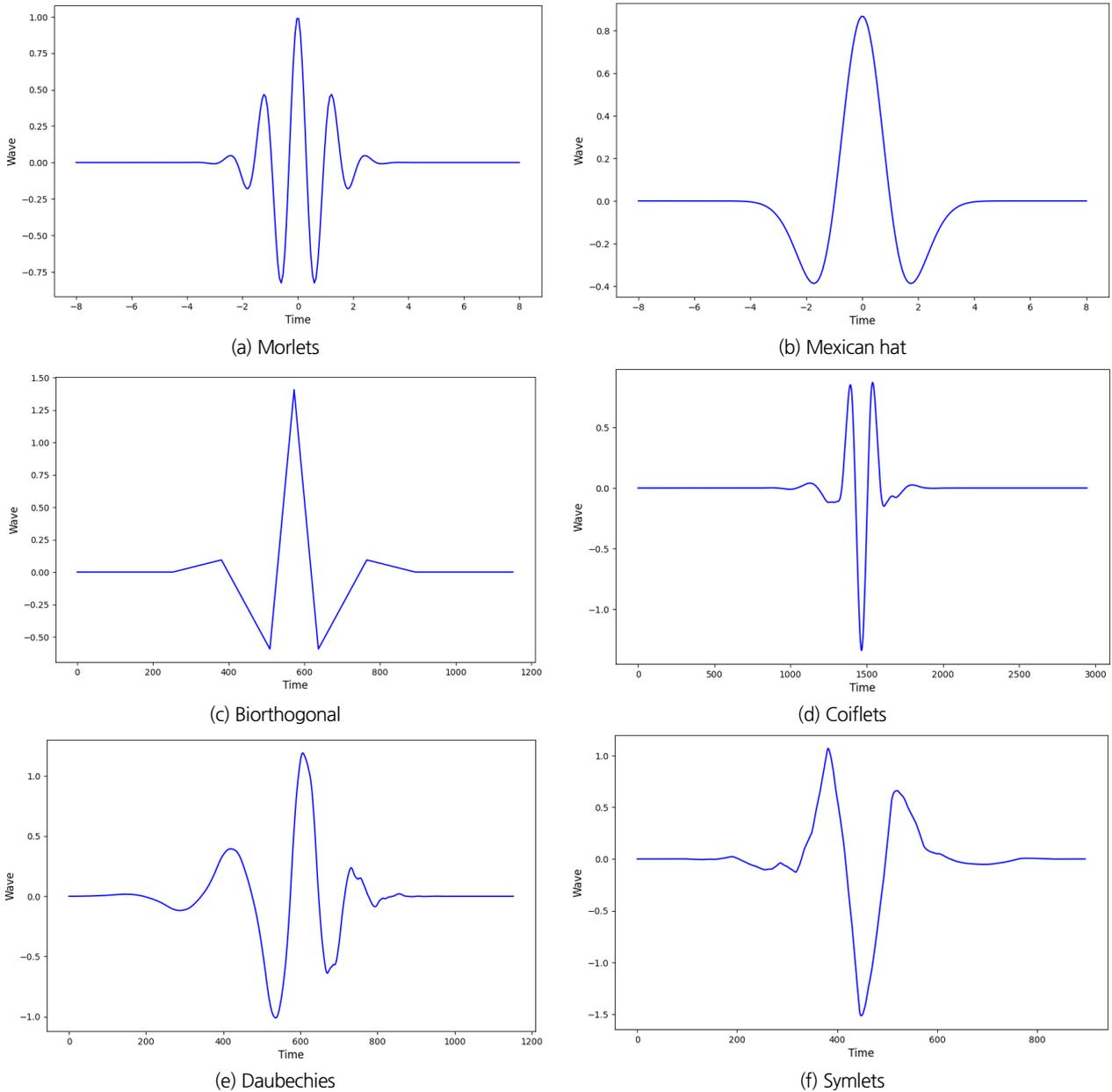


Fig. 3. Examples of wavelet function

원시키는 일종의 인공신경망이다. 이때, 데이터를 압축하는 부분을 인코더(Encoder), 다시 복원하는 부분을 디코더(Decoder)라고 부른다(Fig. 4). 인코딩 과정에서 추출한 의미 있는 데이터를 잠재 벡터(Latent Vector) 혹은 잠재 표현(Latent Representation)이라고 부르며 디코딩은 잠재 벡터를 입력 데이터와 같은 차원으로 재구성하는 역할을 한다(Umberto, 2022).

Autoencoder는 입력된 데이터에 포함되지 않은 이상 특성이 발견되는 경우 재구성 성능이 떨어진다. 이러한 성능저하는 주로 이상치를 탐지하는 데에 활용된다(Chen *et al.*, 2018). 그러나 이상치 탐지를 목적으로 하는 것뿐만 아니라, 데이터의 특성을 고려하여 인코더와 디코더를 조정하는 등의 변형이 적용된 Autoencoder들이 있다(Baldi, 2012). 이 중에서 잡음 제거에 특화된 Autoencoder로 Denoising Autoencoder가 존재한다.

Denoising AutoEncoder는 Fig. 5와 같이 기존 AutoEncoder

와 유사한 모델 구조를 가진다. 주요 차이점은 입력 데이터에 잡음(Noise)을 추가하되 최종적으로 산출된 출력값은 잡음을 추가하기 전 입력값을 예측하도록 학습한다는 점이다(Creswell and Bharath, 2018). 즉 Denoising autoencoder는 인코딩과 디코딩을 거치며 잡음을 제거하는 역할을 수행한다.

Denoising Autoencoder 또한 Neural Network의 한 종류이기 때문에 레이어의 개수, 노드의 개수, 학습률 등과 같은 하이퍼 파라미터의 값에 성능이 크게 영향받는다. 따라서 본 연구에서는 Denoising Autoencoder의 최적 하이퍼파라미터 선정을 위해 인코더 및 디코더 구조에 변화를 주었고, 각 변화에 따른 성능을 비교했다. 공통적으로 인코더는 각 레이어를 거치며 노드 수가 절반으로 줄어드는 구조로 구성하였으며 디코더는 각 레이어를 거치며 두 배로 노드 수를 증가시켜 입력 데이터의 차원까지 증가하도록 구성하였다. 추가로 연구에 사용되는 입력 데이터는 잡음을 가지고 있는 상태이기에 잡음을 추가하는 과정은 생략하였다.

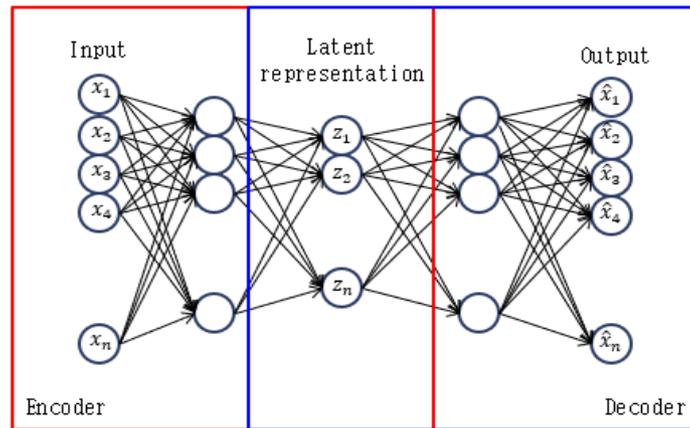


Fig. 4. Autoencoder neural network

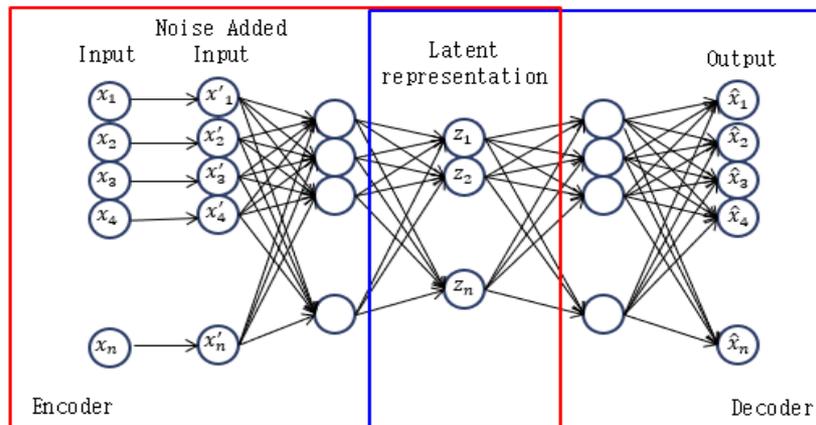


Fig. 5. Denoising autoencoder neural network

3.4 성능평가 지표

본 연구에서는 수위 데이터의 잡음 제거 성능을 정량적으로 평가하기 위하여 회귀(Regression) 모델의 성능평가에 사용되는 대표적인 지표들을 도입하였다. 크게 Wavelet Transform과 Denoising Autoencoder의 성능을 평가하였으며 더 자세히는 Wavelet Transform의 경우 6가지 Wavelet 함수와 매개변수의 값 조정에 따른 성능 변화를 측정하였고 Denoising Autoencoder의 경우 레이어의 수에 따라 4가지 조건을 실험하였다. 성능 평가 지표는 평균 절대 오차(Mean Absolute Error, MAE), 평균 제곱 오차(Mean Squared Error, MSE), 평균 절대 백분율 오차(Mean Absolute Percentage Error, MAPE)이며 Eqs. (3)~(5)이 각각을 나타낸다.

$$MAE = \frac{1}{n} \sum_{t=1}^n |a_t - p_t| \quad (3)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (a_t - p_t)^2 \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|a_t - p_t|}{y_t} * 100\% \quad (5)$$

여기서, n 은 데이터의 수, a_t 는 t 시점의 정답 데이터, p_t 는 t 시점의 잡음 제거 데이터, y_t 는 t 시점의 원본 데이터이다. MAE와 MSE는 전체 정확도를 나타내는 척도이며, MAE는 모든 오차에 동일한 중요성이 부여되지만, MSE는 큰 오차에 추가 가중치가 부여된다는 차이가 있다. MAE와 MSE는 0에 가까울수록 높은 예측 결과를 나타낸다. MAPE는 MAE에 해당하는 상대적인 척도로 예측의 정량적인 방법에서 일반적으로 사용되는 정확도 척도이다(Eva and Oskar, 2012). Table 2와 같이 MAPE의 값이 10% 미만이면 우수, 10~20% 사이는 좋음, 20~50% 사이는 허용 가능, 마지막으로 5% 이상은 부정확함을 나타낸다(Lewis, 1982).

Table 2. MAPE performance evaluation accuracy

MAPE Result	Performance
0-10%	Excellent accurate forecasting
10-20%	Good forecasting
20-50%	Acceptable forecasting
50-100%	Inaccurate forecasting

4. 연구 결과

4장에서는 Wavelet Transform과 Denoising Autoencoder를 적용하여 원본 데이터의 잡음을 제거한 결과 데이터를 시각화하고, 생성한 정답 데이터와 비교하여 평균 절오차(MAE), 평균 제곱 오차(MSE), 평균 절대 백분율 오차(MAPE)를 통해 성능평가를 진행하였다.

4.1 잡음 제거 결과

실험은 Python 3을 사용하여 수행하였으며 Wavelet Transform 구현을 위해서는 PyWavelets를, Denoising Autoencoder 구현을 위해서는 PyTorch 라이브러리를 사용하였다. 다음의 Figs. 6~10은 Wavelet Transform의 Wavelet 함수에 따른 매개변수별 결과 그래프이다. 이때 선택한 매개변수는 1, 3, 5로 공통되게 지정했으며 몇몇 함수에서는 Python에서 적용이 불가능하여 다음과 같이 지정하였다.

Fig. 6은 Biorthogonal wavelet 함수를 적용한 결과이다. Python 내에서 Biorthogonal 함수는 bior1과 같이 정수로 되어있지 않고 소수점으로 적용할 수 있다. 이에 각 매개변수 1, 2, 3에서 첫 번째 소수점인 Biorthogonal1.1, Biorthogonal2.2, Biorthogonal3.1으로 지정하였다.

Fig. 7은 Coiflets wavelet 함수를 적용한 결과이다. Python 내에서 Coiflets 함수는 매개변수 1, 3, 5를 모두 적용할 수 있으므로 Coiflets 1, Coiflets 3, Coiflets5로 지정하였다.

Fig. 8은 Wavelet 함수로 Daubechies wavelet 함수를 적용한 결과이다. Python 내에서 Daubechies 함수는 매개변수 1, 3, 5를 모두 적용할 수 있으므로 Daubechies1, Daubechies3, Daubechies5로 지정하였다.

Fig. 9는 Wavelet 함수로 Symlets wavelet 함수를 적용한 결과이다. Python 내에서 Symlets 함수는 매개변수 1을 적용할 수 없고 매개변수 2부터 적용이 가능하다. 이에 매개변수 2, 3, 5를 적용하여 symlets2, symlets3, symlets5로 지정하였다.

Fig. 10은 Wavelet 함수로 Mexican hat 함수와 Morlets 함수를 적용한 결과이다. Python 내에서 연속 Wavelet 함수는 적용할 수 있는 매개변수가 없다. 이에 연속 Wavelet인 Mexican hat과 Morlets는 여러 매개변수를 지정하지 못하고 함수 자체로 지정하였다.

이산 Wavelet의 모든 함수의 매개변수의 값을 증가시키에 따라 세부적인 특징의 분석을 통해 도출하고자 하는 정답 데이터와 유사한 결과를 얻을 수 있음을 확인했다.

Denoising Autoencoder의 경우, 입력 데이터의 노드 수의 변화(Condition)에 변화를 주어 성능을 비교하였다. 첫 번째

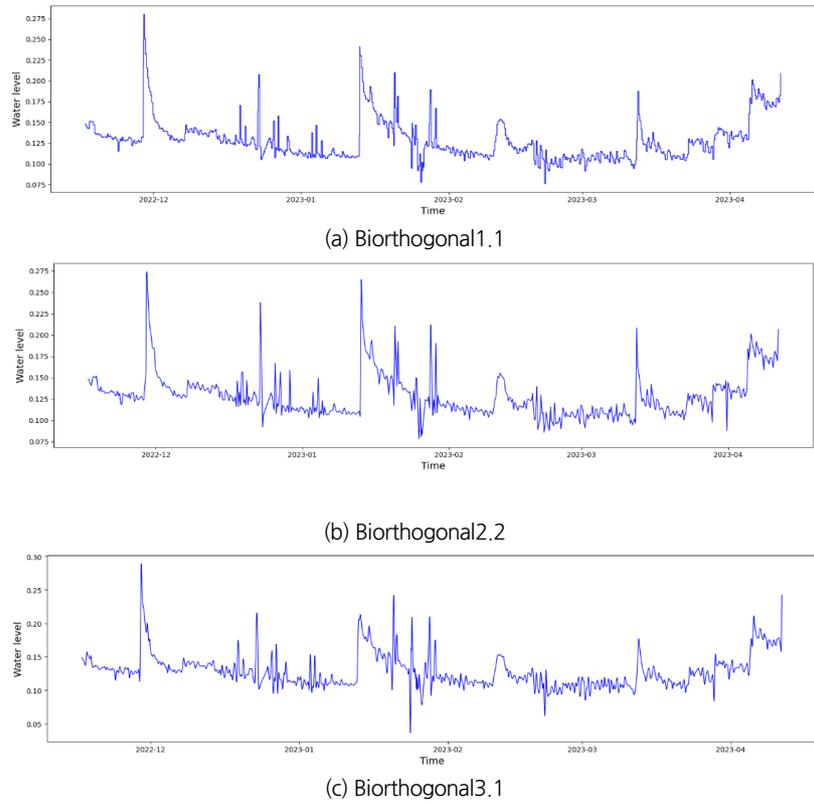


Fig. 6. Result of Biorthogonal wavelet function

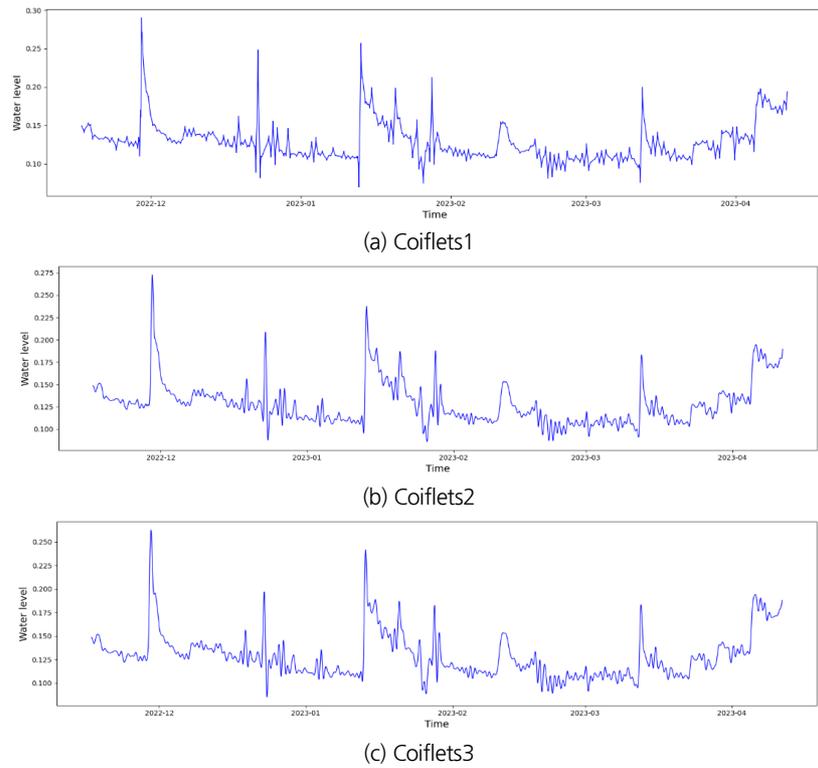


Fig. 7. Result of Coiflets wavelet function

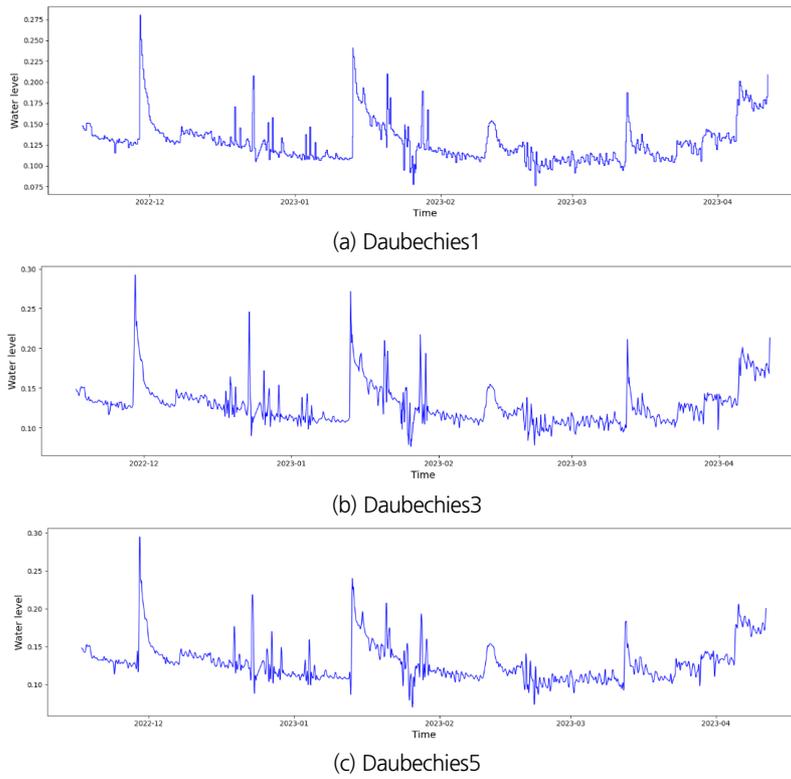


Fig. 8. Result of Daubechies wavelet function

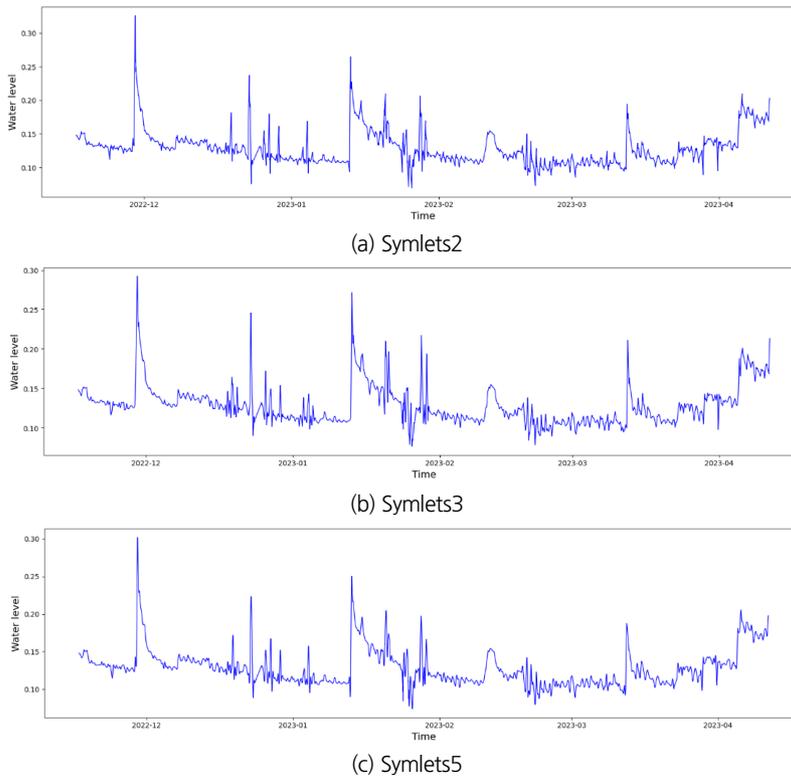
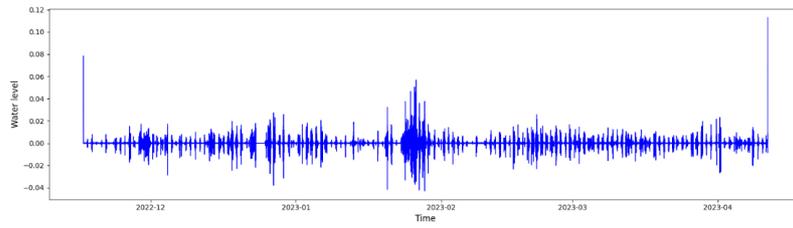
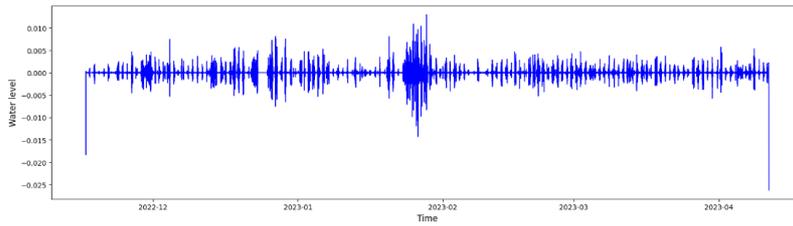


Fig. 9. Result of symlets wavelet function

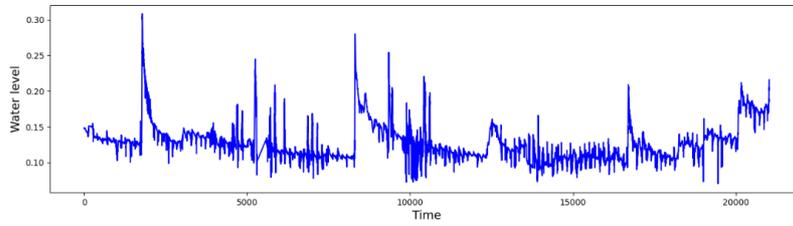


(a) Mexican hat

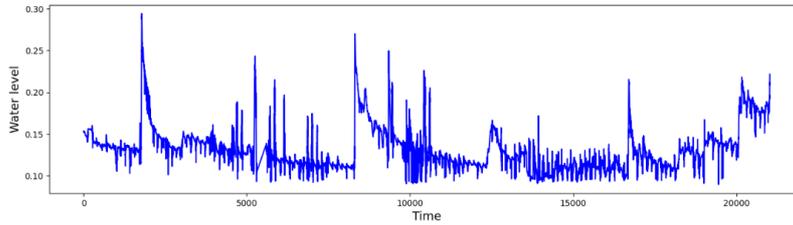


(b) Morlets

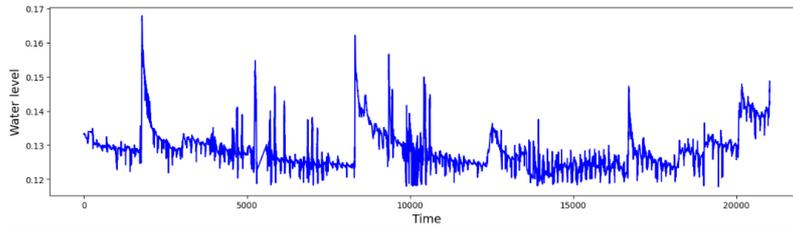
Fig. 10. Result of Mexican hat, Morlets wavelet function



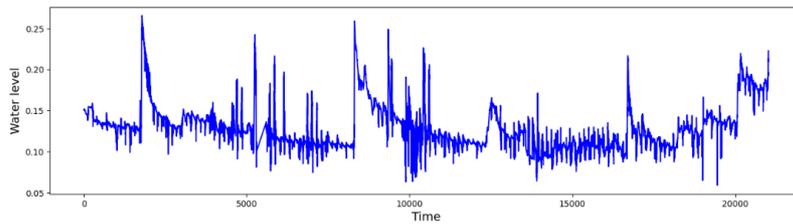
(a) Condition1



(b) Condition2



(c) Condition3



(d) Condition4

Fig. 11. Result of each condition

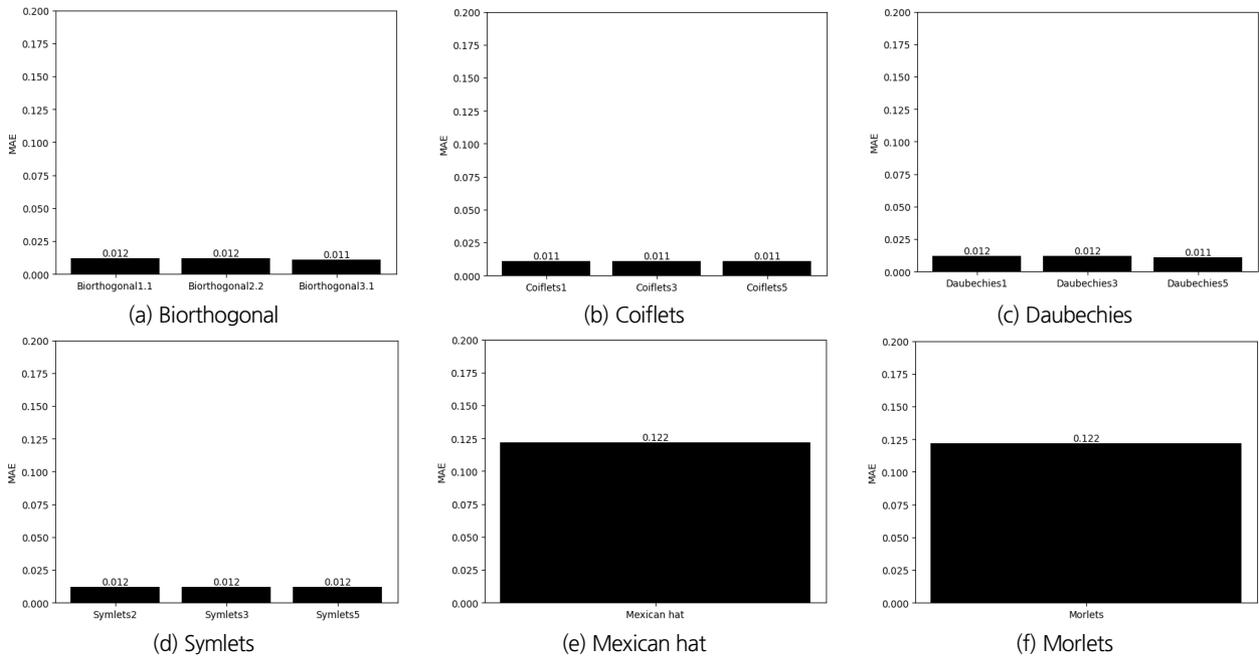


Fig. 12. MAE of the each wavelet function

레이어의 노드 수를 256개, 128개, 64개, 32개로 변경하며 실험을 수행하였으며 각각의 조건을 Condition1, 2, 3, 4로 지정하였다. 다음의 Fig. 11은 각 Condition에 따른 결과 그래프이다.

4.2 잡음 제거 결과에 대한 성능평가

4.2.1 MAE를 이용한 모델의 성능평가

Fig. 12는 각 Wavelet 함수를 이용하여 MAE를 산출한 결과를 나타낸 것으로, 이산 Wavelet에서는 약 0.011의 값이 나타났다. 연속 Wavelet에서는 0.122로 이산 Wavelet 함수들보다 높은 MAE가 나타났음을 확인할 수 있다. 그럼에도 최대 0.122에서 최소 0.011의 MAE 값이 나온 것으로 보아 어떠한 Wavelet 함수를 사용하더라도 준수한 MAE 성능평가를 보인다. 이러한 결과는 수위 데이터에서 이산 Wavelet이 연속 Wavelet보다 더 높은 성능을 보였다.

Fig. 13는 Denoising Autoencoder의 각 Condition별 MAE를 산출한 결과를 나타낸 것으로, 평균적으로 0.015의 MAE의 결과값이 나타났다. 이러한 결과는 모든 Condition에서 준수한 MAE 성능평가를 보였다.

두 모델의 MAE를 비교하였을 때, Wavelet Transform은 평균이 약 0.012, 비교 대상인 Denoising Autoencoder는 평균이 약 0.015로 Wavelet Transform이 Denoising Autoencoder보다 근소하지만, 더 높은 성능의 결과를 보임을 확인할 수 있었다.

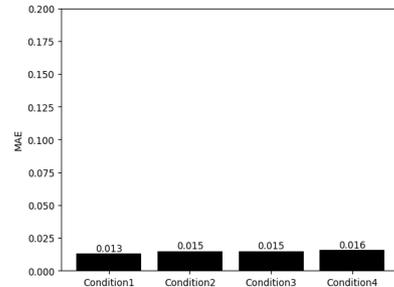


Fig. 13. MAE of the each condition

4.2.2 MSE를 이용한 모델의 성능평가

Fig. 14은 각 Wavelet 함수를 이용하여 MSE를 산출한 결과를 나타낸 것으로, 이산 Wavelet에서는 모두 0의 MSE가, 연속 Wavelet에서는 0.015의 MSE가 나타났음을 확인할 수 있었다. MSE의 수치적으로는 어떠한 Wavelet 함수를 사용하더라도 우수한 결과를 보여주며, MAE와 마찬가지로 이산 Wavelet이 연속 Wavelet보다 더 높은 성능을 보였다.

Fig. 15은 Denoising Autoencoder의 각 Condition별 MSE를 산출한 결과를 나타낸 것으로, 모든 Condition에서 0.000의 MSE가 나타났음을 확인할 수 있었다. 이는 MAE와 마찬가지로 모든 Condition에서 준수한 성능을 보였다.

Wavelet Transform과 Denoising Autoencoder에서의 MSE 결과는 연속 Wavelet을 제외하고 0.000의 값으로 높은 성능의 결과를 보임을 확인할 수 있었다.

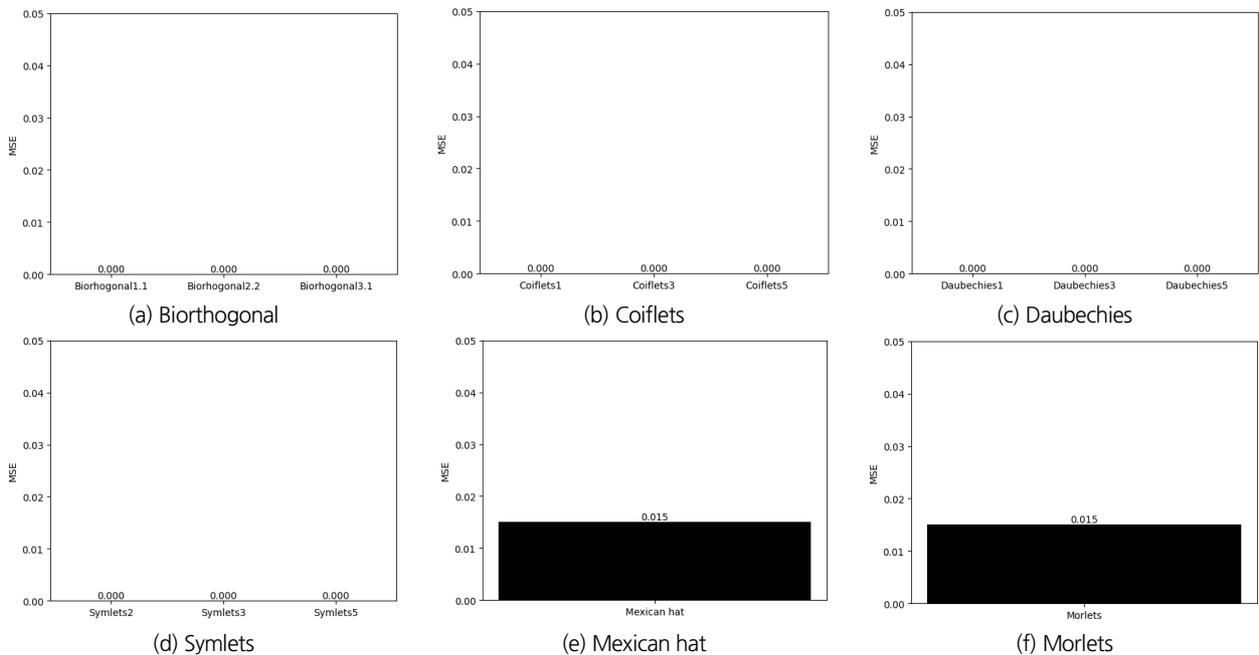


Fig. 14. MSE of the each wavelet function

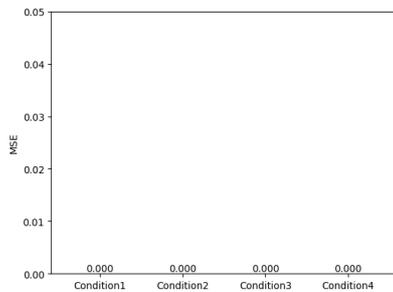


Fig. 15. MSE of the each condition

4.2.3 MAPE를 이용한 모델의 성능평가

Fig. 16은 각 Wavelets 함수를 이용하여 MAPE를 산출한 결과를 나타낸 것으로, 이전의 성능평가들과 유사한 결과가 나타났다. 이산 Wavelet에서는 8~9%대의 MAPE 값이, 연속 Wavelet에서는 100%의 MAPE가 나타났다.

Fig. 17는 Denoising Autoencoder의 각 Condition별 MAPE를 산출한 결과를 나타낸 것으로, Condition1은 10% 내의 우수한 성능평가를, 그 외 Condition에서는 12%~13%의 준수한 성능평가를 확인할 수 있었다.

4.3 모델별 결과 비교분석

Wavelet Transform의 경우, Wavelet 함수의 매개변수가 높을수록 미세한 세부 특성을 감지 및 분석이 가능하다. 하지만 단순히 높기만 해서 항상 우수한 결과를 가져오는 것이 아

니라 잡음이 함께 증가할 가능성이 있다. 그렇기 때문에 최적의 잡음 제거를 위해 최적의 매개변수를 찾아 적용해야 한다 (Jonathan and Sofia, 2010). 모든 Wavelet 함수에서 MAE, MSE는 준수한 결과를 산출했다. 그러나 연속 Wavelet의 MAPE는 99.9%와 100%로 Table 2를 통해 부적합함을 확인했다. 이로 인해 수위 데이터에서는 연속 Wavelet보다 이산 Wavelet을 적용하는 것이 효과적임을 확인할 수 있다. 적용할 수 있는 이산 Wavelet 중 Coiflets의 높은 매개변수에서 0.011, 0, 8.506%의 MAE, MSE, MAPE값을 얻었다. MAE, MSE는 0에 가까운 수치로 적용할 수 있음을 보였다. Denoising Autoencoder의 경우, 입력 데이터의 노드 수가 많을수록 세밀한 추출을 통해 의미 있는 잠재 벡터를 얻게 되어 보다 높은 성능을 보여준다. 그러나 노드 수가 과하거나 부족하면 과적합, 과소적합이 일어나 성능을 저하시킬 수 있다. 비교한 Condition 중 입력 데이터의 노드 수가 256개로 가장 많았던 Condition1에서 0.013, 0.000, 9.932%의 MAE, MSE, MAPE를 얻었다. Table 2를 통해 Baseline으로 지정한 Denoising Autoencoder는 Condition1에서는 ‘우수’, 그 외의 Condition에서는 ‘우수’에 가까운 ‘ 좋음’의 잡음 제거 예측률을 보여주었지만, Wavelet Transform은 연속 Wavelet에서는 ‘부정확함’, 이산 Wavelet에서는 ‘우수’ 잡음 제거 예측률을 보였다. Table 3은 각 Wavelet 함수와 Condition에서 나타난 성능평가 중 가장 우수한 결과를 정리하여 보여준다.

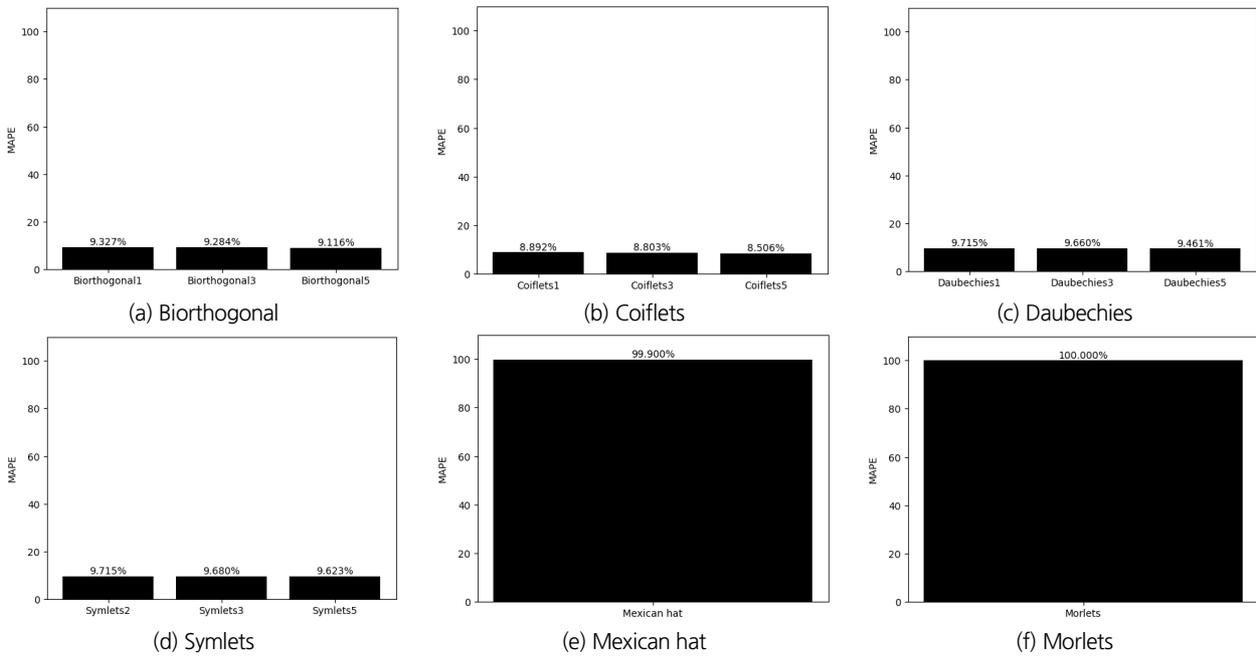


Fig. 16. MAPE of the each wavelet function

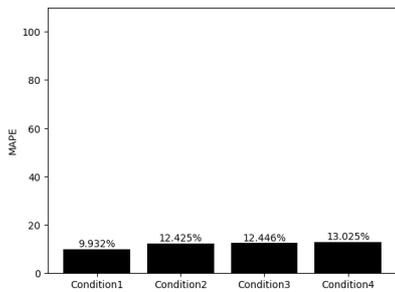


Fig. 17. MAPE of the each condition

Table 3. Highest performance results

Model	Result	MAE (Mean Absolute Error)	MSE (Mean Squared Error)	MAPE (Mean Absolute Percentage Error)
Wavelet	Biorthogonal	0.011	0.000	9.116%
	Coiflets	0.011	0.000	8.506%
	Daubechies	0.011	0.000	9.461%
	Symlets	0.012	0.000	9.623%
	Mexican hat	0.122	0.015	99.9%
	Morlets	0.122	0.015	100.0%
Denoising Auto-encoder	Condition 1	0.013	0.000	9.932%
	Condition 2	0.015	0.000	12.425%
	Condition 3	0.015	0.000	12.446%
	Condition 4	0.016	0.000	13.025%

5. 결론

본 논문은 다양한 데이터 기반 의사결정에 활용될 수 있는 수위 데이터의 데이터 품질을 향상시키기 위한 방안으로 수위 데이터에서의 잡음(Noise)을 제거하는 것을 목표로 하였다. 구체적으로는 Wavelet Transform의 적용 가능성을 평가하고자 하였다. 실제 수집한 데이터는 잡음이 다수 분포하고 있어 데이터 해석에 왜곡을 낳을 수 있었다. 이에 수집한 데이터를 데이터 수집 전문가가 수정하여 정답 데이터를 생성하였다. 정답 데이터는 Wavelet Transform을 적용하여 잡음을 제거한 결과의 평가에 활용되었다. Wavelet Transform의 객관적인 성능을 평가하기 위해 Denoising Autoencoder를 비교군으로 삼았고 MAE, MSE, MAPE를 통해 잡음 제거 성능을 평가하였다. Denoising Autoencoder는 입력 데이터의 노드 수(Condition)와 추가 잡음(Noise Factor)에 변화를 주었다. 입력 데이터의 노드 수가 많은 Condition1에서 MAE는 0.013, MSE는 0.000, MAPE는 9.932%로 가장 우수한 수준의 성능 평가 결과를 보였다. Wavelet Transform은 연속 Wavelet에서는 적용이 부적합함을 보였으며 이산 Wavelet에서는 Coiflets 함수의 매개변수가 높을 때, MAE는 0.011, MSE는 0, MAPE는 8.506%로 가장 우수한 수준의 성능평가 결과를 보였다. Wavelet Transform과 Denoising Autoencoder의 성능평가 비교를 통해 신호 및 이미지뿐만 아니라 수위 데이터에서도 이산 Wavelet을 사용한 Wavelet Transform은 잡음 제

거가 가능함을 확인할 수 있었다. Wavelet Transform은 연산 속도가 빠르고 많은 메모리를 요구하지 않으며 효과적으로 잡음을 제거할 수 있어, 기존 수위 관리 시스템에 적용이 용이하다. Wavelet transform이 효과적이고 효율적으로 수위 데이터의 잡음을 제거함에도 불구하고 수위 데이터의 특성에 따라 초기에 Wavelet 함수를 지정해야 한다는 불편함도 일부 존재한다. 추후 최적의 Wavelet 함수 선정에 대한 가이드라인이 연구된다면 시행착오를 줄일 수 있을 것으로 판단된다.

감사의 글

본 결과물은 농림식품기술기획평가원의 농업용수 물순환 체측 테스트베드 운영 및 물관리 시스템 개발 사업(322081-3)과 2023년도 강원대학교 대학회계 학술연구조성비로 연구가 진행되었습니다.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Alan, T.M. (2010). "Advances in location modeling: GIS linkages and contributions." *Journal of Geographical Systems*, Vol. 12, pp. 335-354.
- Anthony, J.M., Robert, N.F., Yang, L., Nathaniel, A.W., and Steven, D.B. (2004). "An introduction to decision tree modeling." *Journal of Chemometrics*, Wiley, Vol. 1, No.18, pp. 275-285.
- Baldi, P. (2012). "Autoencoders, unsupervised learning, and deep architectures." *In Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, PMLR, Scotland, pp. 37-49.
- Baliunas, S., Frick, P., Sokoloff, D., and Soon, W. (1997). "Time scales and trends in the central England temperature data (1659-1990): A wavelet analysis." *Geophysical Research Letters*, Vol. 24, No. 11, pp. 1351-1354.
- Chan, T.F., and Shen, J. (2005). *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Society for Industrial and Applied Mathematics, SIAM, Philadelphia, PA, U.S.
- Chen, Z., Yeo, C.K., Lee, B.S., and Lau, C.T. (2018). "Autoencoder-based network anomaly detection." *Wireless Telecommunications Symposium*, IEEE, Phoenix, AZ, U.S., pp. 1-5.
- Cho, J.H., Choi, J.H., Kim, Y.S., Yeon, J.H., Lee, S.J., and Ji, B.J. (2022). "A study on the prediction of slope collapse based on LSTM Autoencoder." *Gangwon Agricultural Life and Environment Research*, Vol. 34, No. 3, pp. 345-353.
- Choi, J., Lim, K.J., and Ji, B. (2023). "Robust imputation method with context-aware voting ensemble model for management of water-quality data." *Water Research*, Vol. 243, 120369.
- Chun-Lin, L. (2010). *A tutorial of the wavelet transform*. NTUEE, Taiwan.
- Creswell, A., and Bharath, A.A. (2018). "Denoising adversarial autoencoders." *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 4, pp. 968-984.
- Davis, P.J. (1975). *Interpolation and approximation*. Courier Corporation, MA, U.S.
- De Moortel, I., Munday, S.A., and Hood, A.W. (2004). "Wavelet analysis: The effect of varying basic wavelet parameters." *Solar Physics*, Vol. 222, No. 2, pp. 203-228.
- Escobar, F., Hunter, G., Bishop, I., and Zerger, A. (2008). Introduction to GIS, accessed 13 December 2023, <<http://www.sli.unimelb.edu.au/gisweb/>>.
- Eva, O., and Oskar, O. (2012). "Forecasting using simple exponential smoothing method." *Acta Electrotechnica et Informatica*, ISSN, Vol. 12, No. 3, pp. 62-66.
- Fourier, J.B.J. (1822). *Théorie analytique de la chaleur*. Firmin Didot, Paris.
- Głowacz, A. (2014). "Diagnostics of direct current machine based on analysis of acoustic signals with the use of symlet wavelet transform and modified classifier based on words." *Eksploatacja i Niezawodność*, Vol. 16, No. 4, pp. 554-558.
- Graham, J.W. (2009). "Missing data analysis: Making it work in the real world." *Annual Review of Psychology*, Vol. 60, pp. 549-576.
- Heil, C.E., and Walnut, D.F. (1989). "Continuous and discrete wavelet transforms." *SIAM Review*, Vol. 31, No. 4, pp. 628-666.
- Hinton, G.E., and Salakhutdinov, R.R. (2006). "Reducing the dimensionality of data with neural networks." *Science*, Vol. 313, No. 5786, pp. 504-507.
- Jo, H.Y. (2009). "A study of construction index fund using wavelet analysis." *The Journal of Information Systems*, Vol. 18, No. 3, pp. 351-373.
- Jonathan, M.L., and Sofia, C.O. (2010). "On the analytic wavelet transform." *IEEE Transactions on Information Theory*, IEEE, Vol. 56, No. 8, pp. 4135-4156.
- Kang, S.M., Park, M.J., Kim, S.H., and Kim, S.J. (2007). "A study on the mitigation of inundation damage using flood inundation analysis model FLUMEN." *Academic Presentation of the Korean Water Resources Association*, pp. 815-819.
- Kim, D.K. (2019). "Introduction to machine learning." *Korean Society Of Facilities Engineering*. Vol. 48, pp. 80-82.
- Kim, H.L., Han, K.Y., and Lee, J.Y. (2020). "Prediction of urban flood extent by lstm model and logistic regression." *Journal of the Korean Society of Civil Engineers*, ISSN, Vol. 40, No. 3, pp. 273-283.
- Ko, J.P. (2005). "Solving multi-class problem using support vector machines." *Journal of Information Science Society*, Vol. 32, No. 12, pp. 1260-1270.

- Kuo, H.H. (2018). *White noise distribution theory*. CRC press, Boca Raton, FL, U.S.
- Kwon, H.S., and Lee, M.H. (2023). "Stock index forecasting using combined model of wavelet transform lstm and multi-head attention." Vol. 40, No. 2, pp. 97-112.
- Lee, D.T., and Yamamoto, A. (1994). "Wavelet analysis: theory and applications." *Hewlett Packard journal*, Vol. 45, p. 44.
- Lee, M.H., Park, J.S., and Won, Y.S. (2022). "Future of flood forecasting considering flood damage phenomenon." *Journal of the Korean Meteorological Society's Academic Conference*, p. 226-226.
- Lewis, C.D. (1982). *Industrial and business forecasting methods*. Butterworths, London.
- Morlet, J., and Grossmann, A. (1984). "Decomposition of hardy function into square integrable wavelets of constant shape." *SIAM Journal on Mathematical Analysis*, Vol. 15, No. 4, pp. 723-736.
- Öberg, G., Metson, G. S., Kuwayama, Y., and A. Conrad, S. (2020). "Conventional sewer systems are too time-consuming, costly and inflexible to meet the challenges of the 21st century." *Sustainability*, Vol. 12, No. 16, 6518.
- Oh, J.h., and Seo, K.D. (2011). "Analysis of the intensity and duration of typhoons on the Korean Peninsula using wavelet analysis." *Collections of Academic Conferences of the Korean Chosun Association*, pp. 2040-2043.
- Reilly, J., and David, S. (2000). "Irreversibility, uncertainty, and learning: Portraits of adaptation to long-term climate change." *Climatic Change*, Vol. 45, No. 1, pp. 253-278.
- Sifuzzaman, M., Islam, M.R., and Ali, M.Z. (2009). "Application of wavelet transform and its advantages compared to fourier transform." *Journal of Physical Sciences*, ISSN, Vol. 13, pp. 121-134.
- Sridhar, S., Kumar, P.R., and Ramanaiah, K.V. (2013). "Daubechies, coiflet, symlet and biorthogonal wavelet transforms for effectual image compression and analysis of fidelity measures." *I-Manager's Journal on Digital Signal Processing*, Vol. 1, No. 2, pp. 1-9.
- Umberto, M. (2022). "An introduction to autoencoder." arXiv Preprint, arXiv:2201.03898.
- Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). "A review of recurrent neural networks: LSTM cells and network architectures." *Neural Computation*, Vol. 31, No. 7, pp. 1235-1270.