## 펌프장 유입량 예측을 위한 BILSTM 기반 모델 개발 보고서

**1. 서론**

도시화가 가속화됨에 따라 강우 시 하천 범람이나 배수 지연과 같은 침수 피해가 빈번히 발생하고 있습니다. 특히 펌프장은 이러한 문제를 해결하는 핵심 인프라로, 유입량을 효과적으로 제어하는 것이 매우 중요합니다. 따라서 펌프장 유입량을 사전에 예측할 수 있다면, 강우 상황에 따라 펌프 가동 시점과 강도를 조절할 수 있어 침수 피해를 최소화할 수 있습니다.

본 연구의 목적은 펌프장 유입량 데이터를 기반으로 향후 3시간의 유입량을 예측하는 모델을 개발하여, 실시간 모니터링 및 대응 체계 고도화를 위한 기초 자료를 제공하는 것입니다.

**2. 데이터 설명**

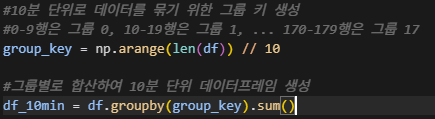
**2.1 데이터 출처 및 구성**

본 연구에서는 Train, Test 데이터를 별도로 구성하였으며, 입력 변수(독립 변수)로는 Rain, Flow, 수위(모니터링 데이터)를 사용하였고, 출력 변수(종속 변수)는 Flow 유입량 데이터입니다.

* Rain: 10분 단위로 수집된 3시간 분량의 강우량 데이터
* Flow: 1분 단위로 수집된 3시간 분량의 유입량 데이터 (종속 변수도 동일한 해상도 사용)
* 모니터링 데이터: 수위 데이터만을 사용하였으며, 총 4개의 파일을 통합(concat)하여 구성하였습니다.

**2.2 데이터 전처리**

* 시간 단위 통일: Flow 데이터는 1분 단위이므로, 이를 10분 단위로 변환하여 Rain 데이터와 시간 해상도를 일치시켰습니다.



* 음수 값 처리: Rain 데이터 내 음수 값은 모두 절대 값으로 변환하였습니다.



* 행 수 정규화: 시계열 예측을 위한 고정 입력 구조를 구성하기 위해, 모든 데이터셋을 18행 단위로 통일하였습니다.



* 수위 데이터 통합: 수위 데이터는 4개 파일을 concat 함수를 이용해 하나의 통합 데이터셋으로 구성하였고, 앞서 언급한 기준에 맞춰 동일한 길이로 정규화하였습니다.





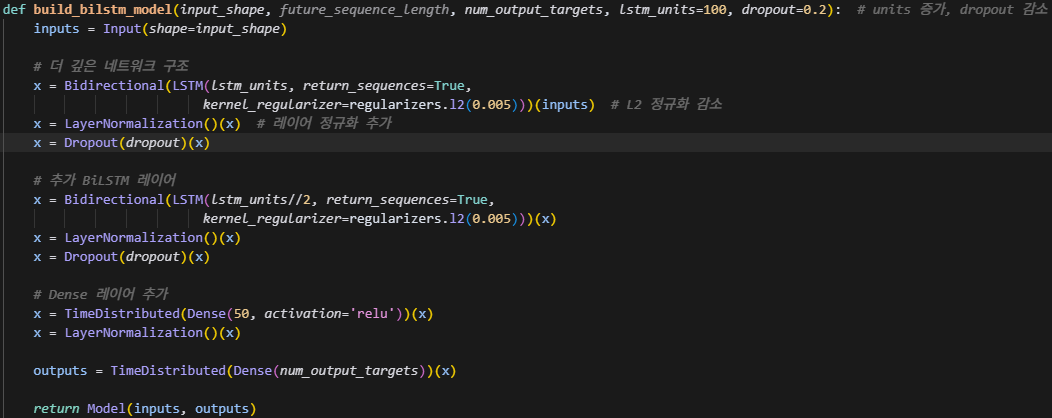
**2.3 모델 선정 및 비교**

예측 성능을 비교하기 위해 다음과 같은 시계열 기반 딥러닝 모델들을 실험적으로 적용하였습니다:

* LSTM (Long Short-Term Memory)
* TCN (Temporal Convolutional Network)
* Transformer
* BILSTM (Bidirectional LSTM)

비교 결과, BILSTM이 가장 우수한 예측 성능을 보였습니다. BILSTM은 시계열 데이터의 과거와 미래 정보를 모두 반영할 수 있어, 일반 LSTM보다 정밀한 예측이 가능하다는 장점이 있습니다. 따라서 본 연구에서는 최종 예측 모델로 BILSTM을 채택하였습니다.

BILSTM



2.3-1 모델 레이어 구조

•첫 번째 BiLSTM 레이어:



양방향 LSTM 사용,L2정규화(0.005)로 과적합 방지

•정규화 및 드롭아웃:



•두 번째 BiLSTM 레이어:



•Dense 레이어:



각 시퀸스 스텝마다 50개의 뉴런을 가진 Dense 레이어

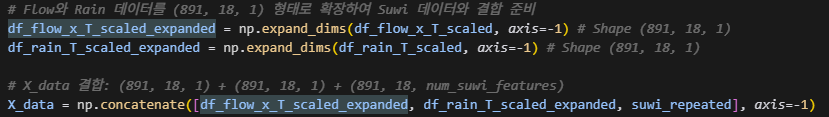
ReLU활성화 함수 사용

•출력 레이어

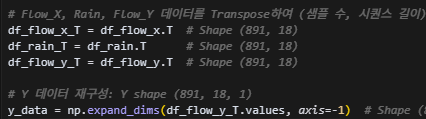


**2.4 학습 방식**

* 수위 데이터는 항상 입력 변수에 고정적으로 포함하였습니다.



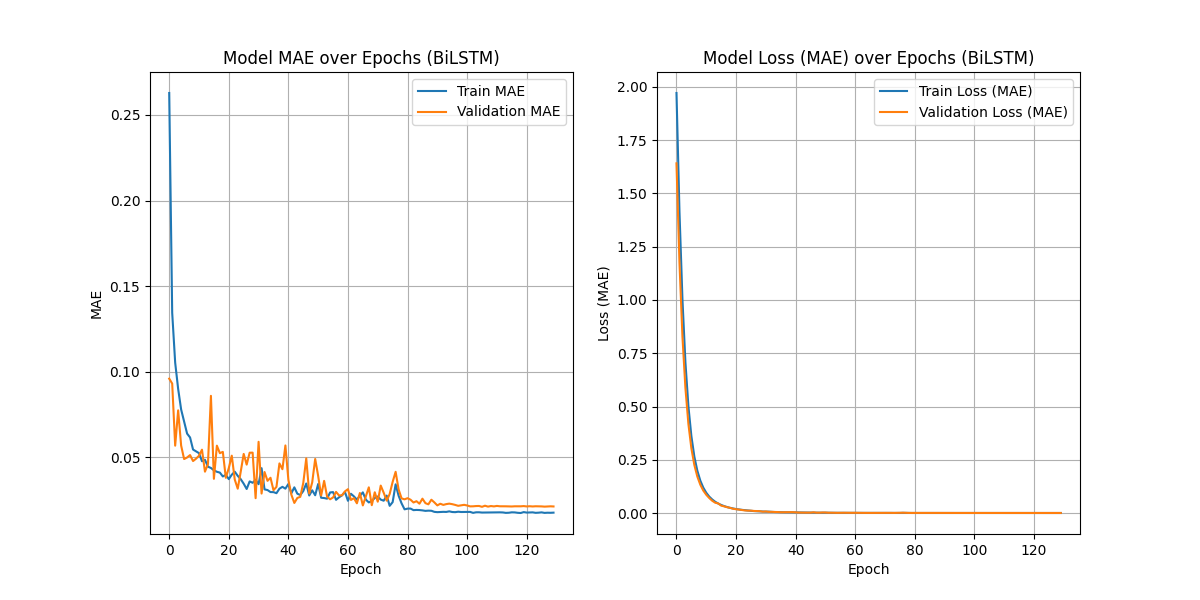
* Rain, Flow 데이터는 각각의 열이 시간적으로 일대일로 대응되도록 구성하고, 종속 변수인 유입량 데이터와도 정확히 매칭시켜 모델에 입력하였습니다.

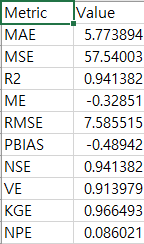


* Train 데이터는 총 891개의 시퀀스(열)를 학습에 사용하였으며, Test 데이터는 9개의 시퀀스를 대상으로 성능을 검증하였습니다.

**2.5 평가지표 및 해석**

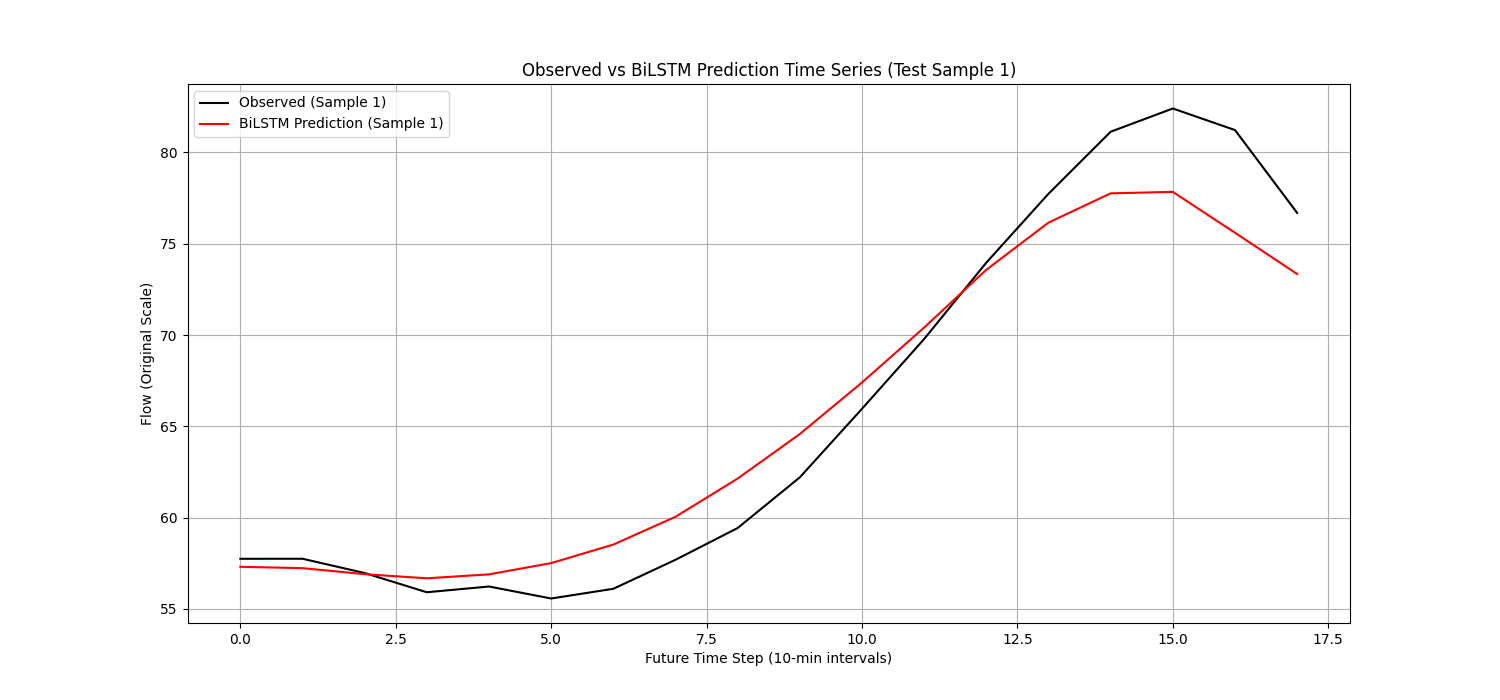
모델 성능 평가는 다음과 같은 수치적 지표를 기반으로 진행하였습니다:





* R² (결정계수), NSE (Nash-Sutcliffe Efficiency), KGE (Kling-Gupta Efficiency) 등의 지표에서 모두 0.9 이상의 높은 예측 정확도를 나타내었습니다.
* ME (Mean Error), PBIAS (Percent Bias) 값은 0에 가까워 예측값의 편향이 거의 없음을 보여주었습니다.
* 특히 테스트 데이터(9개 샘플)에 대해서도 예측 오차가 작고, 전반적으로 모든 지표에서 우수한 결과를 나타냈습니다.
* NPE 값이 0.08로 정규화된 예측오차가 작음을 나타냈습니다.

**2.6예측결과 해석**

 **전반적인 추세 예측 정확도 우수**

* 예측곡선은 전체적으로 관측곡선과 유사한 형태의 상승 곡선을 따라가고 있음
* 이는 BiLSTM 모델이 시계열의 **패턴(상승/감소 트렌드)** 을 잘 학습했음을 보여줌

** 초기 구간 예측 정확도 높음 (0~10 타임스텝)**

* 초기에는 관측값과 예측값이 거의 유사하여 예측이 정확함
* 실제 유입량의 점진적 증가를 BiLSTM이 적절히 반영함

** 중후반부의 피크값에서 오차 존재 (11~17 타임스텝)**

* 관측값은 15타임스텝 부근에서 약 83 정도의 최고값을 기록
* 반면 예측값은 최고점에 도달하기 전 **완만한 정체 후 감소**함
* 이는 모델이 **급격한 피크나 변곡점**을 정확히 포착하지 못한 것으로 보임

** 과소 예측 경향**

* 전반적으로 예측 곡선이 관측 곡선보다 낮게 위치 → **과소 추정**
* 이는 앞서 본 지표 중 **ME: -0.33, PBIAS: -0.49%** 와 일관됨
* 그러나 차이는 크지 않아, **패턴은 잘 맞추되 수치적 차이가 미세하게 존재**함

**종합적으로 볼 때**

BILSTM 모델은 펌프장 유입량 예측에서 **높은 정확도, 낮은 편향, 우수한 시계열 추적 능력**을 갖춘 효과적인 도구로 판단됩니다. 이는 향후 도시 침수 대응, 펌프장 자동 제어 시스템 등의 분야에서 **선제적 의사결정을 위한 핵심 기술**로 활용될 수 있을 것입니다.

**3. 결론**

본 연구에서는 펌프장 유입량 데이터를 활용하여 3시간 후 유입량을 예측하는 시계열 딥러닝 모델을 개발하였습니다. LSTM, TCN, Transformer 등 다양한 모델을 비교한 결과, BILSTM이 가장 높은 예측 정확도를 보였으며, 최종 모델로 선정되었습니다.

특히, 수위 데이터를 고정 입력으로 사용하고 강우량 및 유입량 데이터를 정제하여 시계열 구조로 구성한 점이 모델 성능 향상에 크게 기여하였습니다. 본 모델은 향후 도시 침수 대응 체계의 고도화, 펌프장 자동 운영 시스템 등에 실시간 제어 기반 기술로 활용될 가능성이 높습니다.