

# 시계열 데이터 기반 온실 환경 예측을 위한 딥러닝 모델의 성능 비교 및 분석

김재영, 김진홍, 최윤원, 김광주

한국전자통신연구원

jaeyoung@etri.re.kr, jinhong@etri.re.kr, yunwon.choi@etri.re.kr, kwangju@etri.re.kr

## A Study on the Prediction of Greenhouse Environment Based on Time Series Data Performance Comparison and Analysis of Deep Learning Models

Jaeyoung Kim, Jinhong Kim, Yunwon Choi, Kim Kwang-Ju

Electronics and Telecommunications Research Institute

### 요약

본 연구에서는 스마트 온실의 환경 예측을 위해 세 가지 대표적인 딥러닝 모델인 1D 합성곱 신경망(1D Convolutional Neural Network), 장단기 메모리(Long-Short Term Memory, LSTM), 심층 신경망(Deep Neural Network) 모델을 활용해서 성능을 평가하였다. 1분 간격으로 9개월 동안 수집된 데이터를 사용하여 5분, 30분, 60분의 세 가지 시간 간격에서 온도, 습도, CO<sub>2</sub>의 예측 성능을 비교하였다. 전반적으로 시간 간격이 커질수록 예측치의 오류도 커졌으며 온실 환경들에서는 최대 30분 이내의 예측치를 스마트팜 시스템의 복합환경제어에 적용할 수 있고 60분 후의 예측치들은 전반적인 추세를 예측하는 데 활용할 수 있을 것으로 분석하였고, 이를 통해 스마트 온실의 선제적 예측 제어를 통해 더욱 정밀한 환경제어를 지원할 수 있다.

### I. 서론

온실의 복합환경(온도, 습도, CO<sub>2</sub> 등)을 정밀하게 제어하기 위해서 과거의 시계열 데이터를 분석하여 그 예측치를 기반으로 정밀제어를 수행할 수 있게 된다면 스마트팜 시스템의 온실 환경 제어를 매우 효과적으로 수행할 수 있을 것으로 판단된다.

시계열 딥러닝 모델은 연속된 데이터를 분석하고 정밀한 예측을 제공하는 데 있어 매우 유용한 도구로 평가된다. 이 모델들은 과거의 복합환경 데이터를 학습하여 전통적인 방법으로는 발견하기 어려운 패턴과 추세를 식별할 수 있고 장단기 메모리 네트워크(LSTM)는 시계열 분석과 예측에 널리 사용되며, 복합환경 예측에서는 온도, 습도, CO<sub>2</sub>와 같은 변수를 예측하는 데 효과적이다. 예를 들어, 온실 내부 온도와 습도 등의 예측에 순환 신경망(RNN)의 한 유형인 LSTM 모델을 적용한 연구 결과도 있고 [2] 이 모델들의 성능은 평균 절대 오차(MAE), 평균 제곱근 오차(RMSE), 평균 제곱 오차(MSE), 결정 계수(R<sup>2</sup>)와 같은 다양한 지표로 평가되었다 [3]. 또한, 심층 신경망(DNN)은 입력과 출력 간의 복잡한 비선형 관계를 분석할 수 있기 때문에 수치 예측에 자주 사용된다.[4,5]

본 연구는 DNN, 1D CNN, LSTM 모델을 단독으로 또는 조합하여 5분, 30분, 60분 후의 온실 내부의 온도, 습도, CO<sub>2</sub> 농도를 모델 간의 예측 성능을 비교 분석하였다. 1분 단위로 수집된 환경 데이터를 바탕으로 모델을 훈련시켜, 단기 변동성과 장기 추세를 모두 반영하도록 복합환경 예측 시점을 3개의 시간 단위로 정의하였다. 개별 모델들의 성능은 MAE, RMSE, MSE, R<sup>2</sup> 지표들을 사용하여 각 모델들을 평가하였다. 이 연구는 스마트팜 시스템에 최신 인공지능 기술을 적용하여 온실 내부의 복합환경을 정밀하게 제어할 수 있도록 지원함으로써 농가의 작물 수확량과 품질을 향상시킴으로써 향후 스마트 농업의 발전에 기여할 것으로 예측된다.

### II. 본론

본 연구에 사용된 데이터셋은 경상북도 포항시에 위치한 해맞이농장에서 운영하는 아치형 딸기 비닐하우스에서 2023년 09월부터 2024년 6월까지 수집된 온실 내부의 복합환경 데이터를 사용하였다.



그림 1. 해맞이농장 온실운영시스템 및 센서류

모델들을 학습하기 위해서 7개의 내부 환경 센서(온도(좌, 우), 습도(좌, 우), CO<sub>2</sub>(좌, 우), 일사량)를 통해 데이터를 수집하였다. 수집된 원시 데이터셋에서 결측치는 0.1% 정도로 매우 작기 때문에 결측치를 전체의 평균값으로 처리하였다. 이상치의 경우는 학습에 영향을 주지 않아서 특별히 전처리는 수행하지 않았으며 전처리 후 데이터셋은 각각 6:2:2의 비율로 학습, 검증 및 테스트 데이터셋으로 나누어 수행하였다.

세 개의 모델(DNN, LSTM, 1D CNN)을 3개의 시간 간격(5분, 30분, 60분)에 대해서 학습 및 평가를 수행하였다. 전반적인 모델들의 성능 평가를 위해 MAE, RMSE, MSE, R<sup>2</sup> 를 지표로 활용하였으며 MAE는 오차를 절대값으로 처리하여 각 오차를 동일하게 취급하는 반면, RMSE와 MSE는 오차를 제곱하여 큰 오차에 더 큰 페널티를 부여하게 되어 RMSE/MSE가 이상치에 더 민감하게 반응하게 된다. 또한, 높은 R<sup>2</sup> 값은 모델이 효과적으로 패턴을 학습했음을 나타내며, 낮은 R<sup>2</sup> 값은 모델이 패턴을 잘 포착하지 못했음을 나타낸다.

모든 성능 지표를 살펴보면 전반적으로 시간 간격이 커질수록 예측치의 오류도 커지고 있으며 온도 측면에서는 최대 30분 이내의 예측치를 실제 스마트팜 시스템의 복합환경제어에 적용할 수 있고 60분 후의 예측치들은 전반적인 추세를 예측하는 데 활용할 수 있을 것으로 판단된다.

표 1. 1D CNN 모델의 시간 간격별 성능 비교

지표	5분			30분			60분		
	온도	습도	CO <sub>2</sub>	온도	습도	CO <sub>2</sub>	온도	습도	CO <sub>2</sub>
MAE	0.38	1.38	4.47	1.16	2.86	8.85	2.17	3.82	11.84
RMSE	0.52	2.00	5.85	1.45	3.79	11.08	2.72	5.13	14.93
MSE	0.27	3.99	34.26	2.12	14.39	122.80	7.39	26.27	222.90
R <sup>2</sup>	0.99	0.97	0.95	0.94	0.91	0.81	0.78	0.83	0.66

표 2. LSTM 모델의 시간 간격별 성능 비교

지표	5분			30분			60분		
	온도	습도	CO <sub>2</sub>	온도	습도	CO <sub>2</sub>	온도	습도	CO <sub>2</sub>
MAE	0.25	1.18	3.86	0.95	2.42	7.88	1.54	3.20	12.55
RMSE	0.35	1.79	5.12	1.25	3.43	10.18	2.12	4.43	15.28
MSE	0.12	3.21	26.26	1.56	11.76	108.73	4.51	19.64	233.30
R <sup>2</sup>	0.99	0.98	0.96	0.95	0.92	0.84	0.87	0.87	0.64

표 3. DNN 모델의 시간 간격별 성능 비교

지표	5분			30분			60분		
	온도	습도	CO <sub>2</sub>	온도	습도	CO <sub>2</sub>	온도	습도	CO <sub>2</sub>
MAE	0.26	1.42	4.62	0.76	2.60	11.43	1.63	3.26	17.59
RMSE	0.40	2.12	6.69	1.05	3.48	15.68	2.02	4.51	21.18
MSE	0.16	4.52	44.82	1.11	12.11	245.75	4.09	20.37	448.88
R <sup>2</sup>	0.99	0.97	0.93	0.97	0.92	0.62	0.88	0.87	0.30

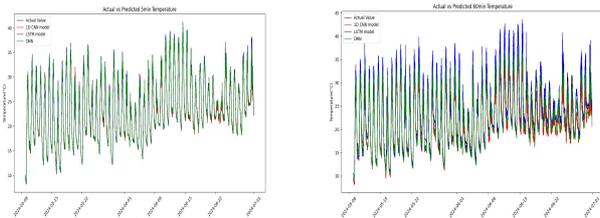


그림 1. 온도 5분 및 60분 후 모델들의 예측치 비교

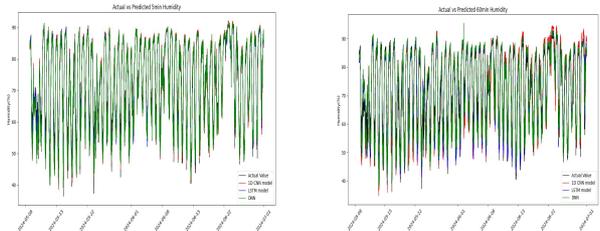


그림 2. 습도 5분 및 60분 후 모델들의 예측치 비교

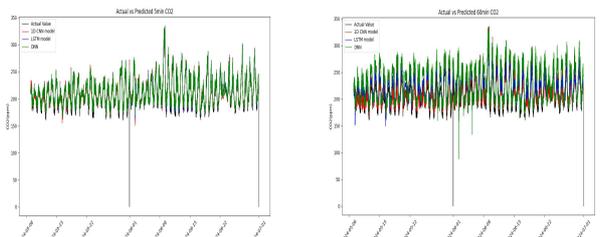


그림 3. CO<sub>2</sub> 5분 및 60분 후 모델들의 예측치 비교

온도 예측 성능에서는 3개의 모델들이 모두 만족할 수준으로 나타났고 5분 후 예측 성능에서는 표2와 같이 시계열 데이터 분석에 강한 LSTM 모델이 근소하게 우수한 성능을 보였고 표 3과 같이 1D CNN 모델이 가장 나쁜 성능을 나타냈다. 30분 후 온도 예측 성능에서는 DNN 모델과 LSTM 모델이 비슷한 성능을 보였다.

습도 예측 성능에서는 3개 모델 모두 유사한 성능을 보였고 CO<sub>2</sub> 성능 예측에서는 표 3과 같이 DNN 모델이 상당히 저조한 결과를 나타내었다. 특징적인 부분은 비선형성과 복잡한 피쳐 간의 관계를 학습하는 데 효과적으로 알려진 DNN 모델에서 60분 후의 CO<sub>2</sub> 예측 성능이 급격히 떨어지는 모습을 보였다.

### III. 결론

본 논문에서는 스마트팜 시스템이 설치된 온실 환경에서 9개월 간 온실 작물의 성장에 직접적인 영향을 미치는 온도, 습도, CO<sub>2</sub> 데이터를 수집하여 3개의 인공지능 모델과 3개의 시간 구간을 적용하여 그 예측 성능을 분석하였다. 전반적으로 시간 간격이 커질수록 예측치의 오류도 커졌으며 온실 환경들에서는 최대 30분 이내의 예측치를 스마트팜 시스템의 복합환경제어에 적용할 수 있고 60분 후의 예측치들은 전반적인 추세를 예측하는 데 활용할 수 있을 것으로 분석하였고 전체적으로, 이 연구의 결과는 정밀 온실 제어에 달려있는 기반 예측 모델을 적용할 가능성을 보여주었다.

향후 온실 외부 환경 및 구동기 데이터도 포함해서 분석하고 인공지능 모델을 결합하여 더욱 정확한 모델을 개발할 예정이다.

### ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [24ZD1150, 대경권 지역산업 기반 ICT 융합기술 고도화 지원사업(팜)]

### 참고 문헌

- [1] El Ghomari, M.Y., Tantau, H.J., Serrano, J., “Non-linear constrained MPC: realtime implementation of greenhouse air temperature control”, *Comput. Electron. Agric.* Vol. 49, pp. 345-356, Dec. 2005.
- [2] Jung. D.-H., Kim. H.S., Jhin, C., Kim. H.-J. and Park. S.H., “Time-series analysis of deep neural network models for prediction of climatic conditions inside a greenhouse”, *Comput. Electron. Agric.* Jun. 2020.
- [3] Van Klompenburg. T., Kassahun. A., Catal. C., “Crop yield prediction using machine learning: A systematic literature review”, *Comput. Electron. Agric.* Oct. 2020.
- [4] Dariouchy. A., Aassif. E., Lekouch. K. and Bouriden. L. and Maze. G., “Prediction of the intern parameters tomato greenhouse in a semi-arid area using a time-series model of Deep neural networks”, *Measurement*, vol. 42, pp. 456-463, Apr. 2009.
- [5] Escamilla-García, A., Soto-Zarazúa, G.M., Toledano-Ayala. M. and Rivas-Araiza. E. and Gastelum-Barrios. A., “Applications of Deep neural networks in greenhouse technology and overview for smart agriculture development”, *Appl Sci.* May. 2020.