

## SimVP 기법과 레이더 강수 데이터를 이용한 AI 초단기 강수 예측 모델 개발

표성훈<sup>1,2</sup>, 차지은<sup>1,2</sup>, 백유현<sup>1,2</sup>, 박기준<sup>1,2</sup>, 이해숙<sup>1,2</sup>, 고지훈<sup>3</sup>, 강신환<sup>3</sup>

<sup>1</sup>국립기상과학원 인공지능기상연구과  
<sup>2</sup>기상청 인공지능기상기술연구회  
<sup>3</sup>한국과학기술원(KAIST) 김재철AI대학원

기상 예측 분야에서 초단기 강수 예측 기술은 6시간 이내에 발생하는 다양한 유형의 예측하기 어려운 강수로부터 국민의 재산과 안전을 보호하는 데 중요한 역할을 한다. 최근 전 세계적으로 기존의 수치 모델을 이용한 기상 예측 기술을 넘어 AI를 활용한 기상 예측 기술이 활발하게 개발되고 있으며, 그 성능 또한 수치 모델 및 레이더 실황과 유사한 수준에 도달하고 있다.

해당 개발에 사용한 SimVP(Simpler yet Better Video Prediction) 기법은 convolution 연산만을 활용하여 영상 예측면에서 vision transformer 모델보다 더 우수한 성능을 낸다고 알려져있으며 본 연구는 SimVP와 생성적 적대 신경망(Generative Adversarial Networks; GAN)을 결합하였다. 학습된 SimVP가 시간대별 예측 이미지를 생성한 뒤 GAN의 판별자가 예측 이미지와 실제 이미지를 비교하여 판단하는 방식으로 강수 구조의 움직임을 실제와 유사하게 예측하는 초단기 강수 예측 모델을 개발한 내용을 다루고 있다. 본 모델 개발에는 2018, 2019, 2021, 2022년의 강수량 레이더 데이터와 AWS 데이터를 학습에 사용하였고, 2020년 데이터를 테스트에 사용하여 학습이 완료된 모델에 현재시점에서 -6시간 데이터를 입력하면 +6시간 예측을 10분 간격으로 예측하는 모델을 개발하였다. 개발된 모델은 강수량의 구간별 강수 확률을 예측하여 최종적으로 구간화된 강수강도, 강수구조, 강수유무 등을 예측한다. 최근 개선된 모델은 2022년에 개발된 이전 모델과 비교하여 FSS= 40.7%, F1= 28.7%, CSI= 32.4%의 성능 향상을 이루었으며, SAL (struture-amplitude-location), PSD(Power spectral density)등 시공간 정확도 분석을 진행하였을 때 높은 예측 성능을 보였다. 추가로 모델에 AWS 실시간 데이터로 Fine-tuning 을 진행하여 AWS 지점에 대한 예측도 높은 수준의 정확도로 예측이 가능해졌다.

이를 통해 6시간 동안 10분 간격으로 한반도 레이더 영상의 예측 이미지 및 강수량 데이터와 AWS에 대한 예측자료를 제공할 수 있으며 강수 영역, 강수 유무, 강수 강도에 대한 정보를 예보관이 초단기 강수 예측에 활용할 수 있다. 이는 기존 초단기 강수 예측보다 더욱 신속하고 정확한 예측이 가능하여 기후변화로 인해 점차 증가하는 극한기상현상에 효과적으로 대응이 가능 할 것이다.

Key words: RADAR, AI, SimVP, Generative Adversarial Networks, Precipitation

※ 이 연구는 기상청 국립기상과학원 「AI 기상예측기술개발」 (KMA2021-00121)의 지원으로 수행되었습니다.

## VQGAN과 Transformer를 이용한 실시간 AI 초단기 정량강수 예측모델

차지은, 이해숙, 표성훈, 박기준

기상청 국립기상과학원 인공지능기상연구과

현재 기상청(KMA)은 KLAPS(Korea Local Analysis and Prediction System) MAPLE(McGill Algorithm for Precipitation Nowcasting by Lagrangian Extrapolation)과 두 모델을 합성한 BLEND 모델 등을 이용하여 초단기 강수 예측을 수행 중이다. '21년부터 국립기상과학원과 KAIST는 공동 연구를 통하여, 기존 모델 대비 고해상도(1km)의 AI 초단기 정량강수 예측모델을 개발하였다. AI모델은 수치모델 비해 예측 수행 시간이 짧으며(1분여) 강수의 비선형도 학습할 수 있다는 강점을 지닌다. 해당 연구의 AI 초단기 정량강수 예측 모델은 고해상도 자료를 토근화하여 시공간 강수 패턴을 추출하는 3D-VQGAN과 최대 6시간의 강수 예측을 수행하는 MeBT(Memory-Efficient Bidirectional Transformer) 트랜스포머 구조를 사용한다. 입력과 타겟 모두에 500m 해상도 레이더(HSR\_KMA) 반사도를 사용하였으며, 2014년부터 2023년까지의 9년간의 데이터를 학습하였다(2020년은 검증 데이터로 사용). 입력에는 10분 간격 과거 2시간까지의 값을 사용하며, 10분 간격 6시간까지의 예측을 수행한다. 개발 초기에 메모리 제약으로 인해 12개 레이어와 작은양의 파라미터를 가진 MeBT12(임시로 지칭) 모델을 개발하였다. 이후 예측성 개선을 위하여, Flash-Attention을 도입하고, 메모리 효율성을 개선을 통해 레이어와 파라미터를 증가시킨 MeBT24(임시로 지칭)를 개발하였으며, MeBT12에 비하여 우수한 강수 예측성능을 확인하였다. MeBT24의 예측성능이 3시간 이후 감소하는 현상이 나타났기에, 이를 해결하려, 짧은 예측 주기(10분 간격)로 학습을 이어가는 커리큘럼 학습 모델(MeBT24\_curr)을 추가 개발하였다. 해당 모델은 현재 90분까지의 학습만이 진행 되었으나, 110분까지의 예측에서 세 모델 중 가장 높은 성능을 보였다(90분 이후는 재귀적 예측 수행). 2020년 검증 데이터 일부에 대한 성능 평가 결과, 0.1mm/h 이하 강수에 대한 1시간 예측에서, MeBT24와 MeBT24\_curr의 CSI는 0.8이었으며, MeBT12는 0.76을 기록했다. 다섯 가지로 분류된 강수 유형(대류, 저기압, 전선, 지형, 기타)에 대한 성능 검증에서는 저기압형>전선형>지형형 순으로 높은 예측성능을 보였으며, 내륙에서 발생하는 대류성 강수에서는 비교적 낮은 예측성능을 보였다. 다만, 한반도 남서 이외의 방향에서 유입되는 강수와 축이 회전하는 강수 패턴에 대한 예측 정확도가 떨어지고, 강한 강수(10mm/h이상)를 과대 모의하는 경향을 보여 개선과정에 있다. 현재 맥락길이 연장 실험을 진행 중이며, 디코더의 최적화 등 추가 개선실험을 통하여 예측성능을 향상할 예정이다. 올해('24년) 7월부터는 MeBT24와 MeBT12 모델의 예측을 실시간으로 수행하고 기상청 내부 웹페이지에 표출하는 서비스를 제공하고 있다. 해당 모델들은 일 년여의 시범 운영과 성능 검증을 마치면, 내년 여름 대국민 서비스를 시작할 예정이다.

Key words: AI 초단기 정량 강수 예측 모델, 레이더 반사도, 실시간 운영, 현업화 예정

## 인공지능을 이용한 폭염기간 보행자 규모 연직기온 예측 모델 개발

강민수<sup>1</sup>, 박문수<sup>2</sup>, 백기태<sup>1</sup>, 김석철<sup>1</sup>

<sup>1</sup>세종대학교 기후환경융합센터

<sup>2</sup>세종대학교 기후환경융합학과

기후변화에 따른 지속적인 기온상승으로 도시 열섬 현상이 심화되어 인명피해가 도시지역에서 지속적으로 발생하고 있다. 서울은 대한민국의 수도로, 고도로 밀집된 대도시로서 지리적, 토지이용 등에서 복잡한 특성을 가지고 있다. 이러한 도시지역의 폭염으로 인한 피해를 줄이기 위해서는 보행자 규모의 정확한 기온 관측, 분석 및 예측이 필요하였다. 보행자 규모 기상장 연직 기온 예측 모델은 서울에서 수행된 2건의 고해상도 도시 기상관측캠페인 데이터를 활용하여 4종류의 머신러닝 기법들을 사용하여 개발되었다. 2019년 8월 광화문 광장에서 3차원 빌딩숲 규모 기상관측캠페인이, 2023년 8월 잠실 도시 기상관측캠페인이 진행되었다. 도시 지역의 보행자의 주변 세부 환경을 반영하기 위해 열 쾌적 유형을 분류하여 정의하였다. 모델 구축을 위한 입력 자료는 기상청의 종관기상관측소와 방재기상관측소의 기상관측항목들과 두 캠페인의 고도별 기온, 순환 시간 및 열 쾌적 유형 자료들의 짝으로 구성하였다. 보행자 규모 기온 예측 결과는 관측 값과 모델 예측 값 사이의 회귀 평가 지표 분석을 통해 각 머신러닝 모델 성능에 대해 평가하였다. 최적의 성능을 나타내는 모델을 찾기 위해 모델 검증 및 적용 프로세스의 열 쾌적 유형, 4가지의 머신러닝 모델 및 4종류의 입력 자료(AWS 기온, AWS 기온 편차, ASOS 기온, ASOS 기온 편차 자료)에 따라 평균 편차, 평균 절대 오차, 평균 제곱근 오차, R2 값을 상세히 분석하였다. 최적 모델은 최소의 평균 제곱근 오차 값과 최대의 R2 값을 가지는 입력 자료와 머신러닝 모델로 선정하였다. 최적 모델은 잠실 캠페인과 광화문 광장에서 다른 기간에 적용하여 그 성능을 시험해보았다. 개발된 모델은 기상청의 기상 자료와 TCT만 있으면 보행자 규모의 연직 기온을 예측할 수 있었다. 또한 도심지역 폭염 시에 보행자 체감 기온을 제공할 수 있다는 가능성과 상당히 신뢰성 있는 성능을 보여주었다.

Key words: 폭염, 머신러닝, 보행자 체감 기온, 열섬현상, 보행자 규모 기상장 연직 기온 예측 모델

※ 이 연구는 한국연구재단의 “2021R1I1A2052562 (보호연구)”의 지원으로 수행되었습니다.

## 기계학습/딥러닝 기반의 호우, 낙뢰 가능성 예측

이창재, 임윤진, 이상진, 김병권, 인희진

기상청

호우와 낙뢰는 모두 생명과 재산에 피해를 줄 수 있는 기상현상이지만, 이를 정확히 예측하는 것은 여전히 도전적인 과제로 남아있다. 이번 연구에서는 수치모델에서 제공하는 기상 변수(바람, 기온, 기압, 습도)를 활용해 여러 기상 변수와 호우, 낙뢰 사이의 숨겨진 구조(hidden structure)를 탐색하고, 발생 조건을 분류(classification) 하는 모델을 개발하여 호우와 낙뢰의 가능성을 예측하고, 그 결과를 검증해보고자 하였다.

이를 위해 결정 트리(decision tree) 기반의 기계학습 기법과 기상 변수들의 공간적 분포를 고려할 수 있는 Convolutional Neural Network(CNN) 기법을 사용하여 각각 모델을 구축하였다. 구축된 모델의 학습은 최근 3년간 5월부터 10월까지의 ECMWF(유럽중기예측센터) 수치모델의 기상 변수와 기상청 레이더 누적강수 자료와 낙뢰 관측자료를 라벨(label)을 활용하였으며, 각 모델의 성능은 2024년 5월부터 8월까지 기간에 대해 결과들을 각각 비교해보고자 한다. 또한, 이를 수치모델 예측결과에도 적용해 예측 시간별 검증결과(Frequency Bias, CSI, POD, FAR)도 함께 도출하여 최적의 방법과 활용 가능성에 대해 발표하고자 한다.

Key words: 호우, 낙뢰, 기계학습, 딥러닝, 검증

## GOES ABI자료를 사용한 대기경계층 가강수량 산출 머신러닝 모델 개발

이윤진<sup>1</sup>, Kyle Hilburn<sup>2</sup>

<sup>1</sup>서울대학교 기초과학연구원

<sup>2</sup>Cooperative Institute for Research in the Atmosphere/Colorado State University

하층 수증기는 대류 발생에 기여하는 요소 중 하나로 특히 습윤한 기단과 건조한 기단이 만나 생성된 건조선(dryline)을 따라 발달하는 폭풍우 예측에 중요한 요소이다. 라이더나 레디오존데 등과 같은 지상 기반 관측 시스템은 하층 수증기에 대한 정확한 정보를 제공할 수 있지만, 공간적으로 제한된 정보를 제공하거나, 낮은 시간 해상도로 인해 빠르게 발달하는 건조선 추적에 어려움이 있다. 정지궤도위성은 높은 시공간 해상도의 관측자료를 제공하지만, 현재의 정지궤도위성 채널 대부분 중상층 대기의 수증기에 민감하여 하층 수증기를 직접적으로 관측할 수 있는 채널은 없다. 하지만 “깨끗한 대기창”이라 불리는 10.3 $\mu\text{m}$  채널과 “오염된 대기창”이라 불리는 12.3 $\mu\text{m}$  채널의 차이인 Split Window Difference(SWD)를 사용하여 하층 수증기량을 추정할 수 있다. 본 연구에서는 미국의 Geostationary Operational Environmental Satellites(GOES)의 Advanced Baseline Imager(ABI)자료를 입력자료로 사용한 대기경계층 가강수량 추정 머신러닝 모델을 개발하였다. U-Net 모델을 사용하여 High-Resolution Rapid Refresh(HRRR)모델 기반의 대기경계층 가강수량을 참값으로 학습시켰으며, 학습 이후 HRRR모델과 레디오존데 기반의 대기경계층 가강수량으로 검증하였다. 여러 입력자료 조합 중 ABI 채널 13, 14, 15 (10.3, 11.2, 12.3 $\mu\text{m}$ ) 밝기온도, SWD, HRRR 모델의 지표면 온도, 태양 천정 각을 입력자료로 사용하였을 경우 모델 성능이 가장 좋았으며, 추가적으로 ABI 가강수량 산출물을 입력자료로 사용했을 경우 성능이 더욱 개선되었다. 학습된 머신러닝 모델을 여러 건조선 사례에 적용했을 때 대기경계층 가강수량이 급격하게 감소하는 건조선을 잘 모의하고, 모의된 건조선을 따라 대류운들이 발달하는 것을 볼 수 있었다. 본 연구에서 개발된 머신러닝 모델은 건조선을 따라 발달하는 대류운들을 미리 예측하는데 도움될 것으로 보이며, 본 연구에서 개발된 모델이 기후가 다른 한국에서도 천리안 2A호자료를 사용하여 적용될 수 있는지 분석할 것이다.

Key words: 정지궤도위성, 하층 수증기, 머신러닝, 건조선

## Generation of Rainfall Maps from Satellite Images Using Deep Learning

Yerim Lim<sup>1</sup>, Hyun-Jin Jeong<sup>2</sup>, Yong-Jae Moon<sup>1,2</sup>,  
Yeji Choi<sup>3</sup>, Doyi Kim<sup>3</sup>, Jihye Kim<sup>3</sup>

<sup>1</sup>School of Space Research, Kyung Hee University

<sup>2</sup>Department of Astronomy and Space Science, College of Applied Science, Kyung Hee University

<sup>3</sup>Earth Intelligence, SI Analytics

In this study, we propose a rainfall maps generation model from satellite images using deep learning without Z-R relationships. The input data for the model are multi-channel satellite images, which cover visible to infrared wavelengths, and the target data (ground truth) hybrid surface precipitation (HSP) data (in unit of mm/hr) of Korea Meteorological Administration (KMA) from 2021 to 2023. Hybrid surface precipitation (HSP) consist of radar bins at the lowest altitudes that are not affected by beam blockage or ground clutter. Our deep learning model, called Pix2PixCC, is an improved version of Pix2Pix for image translation with an additional inspector network to calculate the correlation between input and target pairs. We compare our results with GEO-KOMPSAT-2A (GK-2A) rain rate (mm/hr) from May 2023 to May 2024 using the Correlation Coefficient (CC) and Root Mean Square Error (RMSE). Our model has a much better performance than GK-2A rain rate in terms of metrics. For the entire period, the CC and RMSE between our model predictions and the ground truth are 0.45 and 0.71 mm/hr, respectively, at a resolution of 8 km. For August 2023, the CC and RMSE are 0.59 and 1.12 mm/hr, respectively at the same resolution. This study is significant in that it proposes a deep learning model for directly generating rainfall maps from satellite images, which can be expanded for global regions.

Key words: Rainfall estimates, Deep learning, Weather satellites, Hybrid surface precipitation.