

객관적 기후예측에 적합한 인공지능활용 기반기술 개발(I)

*부제: 기후예측에 적합한 인공지능활용 기반기술 적용을 위한
자료확장 및 환경구축*

**The Development of Artificial Intelligence Base
Technologies for Objective Climate Predictions (I)**

김원무, 박경원, 이운영, 이진영, 정유란

기후사업본부 예측기술과

2022.12.

APEC기후센터

평문 초록

계절 또는 계절내 스케일의 기후 예측을 위한 자료의 특성을 파악하고, 모델의 훈련 자료를 확장하고, 사용자가 쉽게 자료를 얻을 수 있는 입출력 시스템을 개발하고, 다양한 전처리 기법을 적용하여 딥러닝 모델의 구조를 최적화하고, 준지도학습을 통해 모델의 성능을 개선하여 객관적 기후예측에 적합한 AI 기반기술 개발하고자 합니다. 이러한 인공지능 모델의 개발 및 개선 과정에서 다양한 설명가능한 인공지능 기술을 적용하여 모델의 연산 과정에 대한 이해를 증진시키고 모델의 과적합, 불안정성을 해소하여 신뢰할 수 있는 기후예측 AI 기반기술을 개발하고자 하며 이를 통해 개선된 계절내 또는 계절예측 결과를 생산하고 궁극적으로 다양하고 손에 잡히는 활용 성과 창출을 위한 자료를 제공하고자 합니다. 1차년도에는 이를 위해 기후예측에 적합한 인공지능활용 기반기술 적용을 위한 자료확장 및 환경구축을 수행하였습니다.

자료 부족 완화를 위해 위성자료인 관측자료와 주별 예측일에 생산되는 자료들을 계절별로 기간을 확장하여 부족한 자료를 확장하고 확장된 자료를 ConvLSTM, U-Net 딥러닝 기법에 적용하여 예측 성능을 평가하였으며, 신기술 인공지능을 적용한 예측 모델의 정확도 향상을 위해 기후모델과 관측자료를 사용자가 쉽게 인공지능 기법에 적용하기 위한 입출력 시스템을 개발하였습니다.

딥러닝 기반의 후처리 기술 개발을 위해 먼저 계절내 예측 일 최고 및 최저기온과 일 적산 강수의 예측 성능, 특히 2-4주 예측 개선에 적합한 전처리 기법을 규명하였습니다. 또한 입력자료의 특성(피쳐) 선택을 통해 일 강수의 경우 피쳐 선택에 따른 훈련자료의 차원 감소가 예측성능 변화에 영향을 미치지 않는 것을 확인하였습니다. 그러나 일 최고 및 최저기온의 경우 피쳐 선택에 의한 훈련자료의 입력 변수가 기준 이하일 경우 예측성능이 저하될 수 있는 것으로 나타나 계절내 기온 훈련자료의 예측성능 개선에는 강수와 다른 딥러닝 훈련 모델 적용 등 다른 접근이 필요할 것으로 판단하였습니다.

준지도학습 적용을 위한 준비로 그래프와 이미지를 이용한 인공지능 모델을 구축하여 우리나라 여름철 평균온도 예측 성능을 살펴보았습니다. 인공지능모델 훈련의 가장 큰 걸림돌인 자료 부족 문제를 완화하고자 기상기후자료의 자료를 강화하는 여러 기법을 테스트하였는데, 일자료의 시간적 조합을 활용한 자료강화와 월자료의 공간적 재조합을 활용한 자료강화를 적용하여 비교하였고 목표 월만 사용하는 경우와 모든 월을 사용하는 경우의 결과를 비교하였습니다. 이미지모델의 경우, 월자료의 공간적 재조합을 활용한 자료강화를 수행하고 모든 월을 사용한 경우 모델이 비교적 잘 학습되어 자료 강화의 개선 효과를 확인할 수 있었습니다.

Executive Summary

The purpose of this study is to develop objective and climate prediction-tailored AI base technologies through the identification of data characteristics for seasonal to subseasonal climate predictions, the expansion of available data, the development of a data service system, and the optimization of deep learning model architectures by applying various pre-processing methods and semi-supervised learning.

We applied ConvLSTM and U-Net deep learning techniques to evaluate their prediction performance on weekly forecast days by additionally using satellite data in order to alleviate insufficient data issues. The ConvLSTM did not show optimized results while the U-Net deep learning model provided optimized learning results even with a small number of training data as a result of learning expanded data and showed significant accuracy improvements in terms of PCC and RMSE. We also developed a data service system and developed a GUI environment for utilizing user-oriented deep learning model training sets. This system will contribute to the easy production of input data that can be easily used for artificial intelligence models.

For the development of deep learning-based post-processing technology, this study was first conducted to identify a pre-processing technique suitable for improving the prediction performance of S2S predicted daily maximum and minimum air temperature and daily total precipitation, especially week 2 to 4. The S2S prediction raw data of six individual climate models were converted into training data to be input to the deep learning models, and MME-based S2S training data were additionally constructed from them. The pre-processing process of this study applies five scalers, and added to the pipeline consisting of cases that do not apply to these scalers, and a technique was also added to select the features characteristics of the training data according to the rank by calculating the correlation between the transformed training data and the labels. The learning model of this study simply applied TimeDistributed to the convolutional layer of U-Net, and named this as a reference model. Climate prediction data before correcting the spatial Pattern Correlation Coefficient (PCC) by lead-time evaluated for the results predicted by the Base model for the six individual climate models pre-processed with six techniques and the MME-based daily maximum and minimum air temperature and daily total precipitation training data, that is, as a result of comparison with the observed values, suitable pre-processing techniques improved the prediction performance of daily total precipitation and daily maximum and minimum air temperatures treated with Standard and Robust techniques. In particular, the prediction performance of daily total precipitation was improved in the entire lead-time, and the prediction performance of daily maximum and minimum air temperature was not improved in week 1, but improved in week 3 and 4, and it was confirmed that there is an effect of

improving the post-correction of the Base model for the S2S training data transformed into a suitable pre-processing technique. In addition, in the case of daily total precipitation to which the feature selection technique was applied, it was found that the decrease in the dimensional of the training data and the decrease in input variables according to feature selection did not affect the change in prediction performance. However, in the case of the daily maximum and minimum air temperature, it was confirmed that the prediction performance could decrease if the input variables of the training data by feature selection were below the reference(for example, selection of seven variables in training data), and the feature selection method was still effective in improving the prediction performance of week 1 of the daily maximum and minimum air temperature. As it does not appear to contribute, it is thought that a different approach, such as applying a deep learning training model different from precipitation prediction, is needed to improve the prediction performance of the S2S air temperature training data. In the next study, we will explore the reason why the base model's week 1 improvement is insignificant, optimize the hyper-parameters and structures of the base model, and add various training models to find a deep learning model that can improve the predictive performance of the S2S.

Graph- and image-based Artificial Intelligence models were developed and supervised learning results were examined to construct the research environment for the application of semi-supervised learning. In order to alleviate the lack of training data issue, data augmentation methods were tested: the construction of semi-monthly data from daily data, and the cutmix of monthly data. The use of only the target month (Target Month Only) was compared to the month-agnostic approach (Month-Agnostic) using all months. Both node classification and graph classification models were developed as graph-based AI models, and prediction performance of summertime monthly mean temperature were examined. The effect of data augmentation was not obvious since both models were not trained well. Results of the test set for July (LT2) with cutmix data augmentation and the Month-Agnostic approach outperformed other cases for the node classification model. The Month-Agnostic approach performed better for June (LT1) ($HSS > 0.35$ for all folds) and the performance was improved with data augmentation for July (LT2) for the graph classification model. 3-dimensional convolution neural networks model was developed as a image-based AI model. The model with the cutmix data augmentation and the Month-Agnostic approach was better trained and outperformed other cases ($Accuracy > 0.6$ for all months and folds). The prediction performance for the test set was improved for June (LT1) and July (LT2). Spatially significant regions were derived based on the Class Activation Maps of the last convolution layer of the model. Cases with newly activated areas where their climatological significance was previously analyzed were observed. Supervised learning results were examined and the effect of data augmentation methods was investigated in this first year study.

국문 요약

계절내 또는 계절 스케일의 기후 예측을 위한 자료의 특성을 파악하고, 훈련 자료 확장 및 입출력 시스템 개발, 다양한 전처리 기법 적용을 통한 딥러닝 모델 아키텍처 최적화, 준지도 학습(참값이 존재하거나 존재하지 않는 다양한 자료를 함께 사용하여 모델의 성능을 개선시키는 방법)을 통한 모델 성능 개선 등을 통해 객관적 기후예측에 적합한(Climate Prediction-Tailored) AI 기반기술 개발하고자 한다.

자료 부족 완화를 위해 위성자료인 관측자료와 주별 예측일에 생산되는 자료들을 계절별로 기간을 확장하여 부족한 자료를 확장하고 확장된 자료를 ConvLSTM, U-Net 딥러닝 기법에 적용하여 예측 성능을 평가하였다. 예측모델과 관측자료를 이용한 학습결과 ConvLSTM에서는 여전히 학습자료에 따라 정확도의 차이를 보여주며, 최적화된 결과를 보여주지 못한다. ConvLSTM에 비해 적은 수의 훈련자료에도 최적화된 학습결과를 제공하는 U-Net 딥러닝 모델에 확장된 자료를 학습한 결과, PCC와 RMSE의 통계적 결과에서 유의미한 정확도 향상을 보였다. 신뢰성 있는 예측결과를 제공하기 위해서 다양한 민감도 분석이 필요하고, 예측자료와 관측자료를 사용자가 쉽게 딥러닝 모델 입력자료로 산출 가능한 입출력 시스템을 설계하고, 이에 기반한 각각의 모듈을 개발하여 GUI 환경에 연동하여 사용자 중심의 딥러닝 모델 훈련셋 산출 가능한 서비스 시스템을 개발하였다. 이 시스템을 활용한다면 보다 쉽게 인공지능 기법에 활용 가능하고 정확도 높은 자료 생산에 기여할 것이다.

S2S 예측성능은 단기 및 장기 예보와 비교하여 본질적으로 예측력 원천이 부족하고 선행시간이 길어질수록 증가하는 비선형성으로 인해서 획기적인 역학 모형 성능 개선을 기대할 수 없다는 한계가 있어 통계적 보안을 통한 예측 개선 가능성을 타진할 필요가 있다. 또한 S2S 자료는 개별 기후모델의 예측 자료 생산일, 예측 길이의 다양함 등으로 인해 사용하기 어렵고 번거로운 특성이 있어서 이러한 자료에 대한 전처리 기법 및 최우수 모델 구조 선정 등 S2S 예측 자료 적용을 위한 최적 기법에 대한 연구도 부족하다. 이에 이 연구는 딥러닝 기반의 후처리 기술 개발을 위해 먼저 S2S 예측 일 최고 및 최저기온과 일 적산 강수의 예측 성능, 특히 2-4주 예측 개선에 적합한 전처리 기법을 규명하고자 하였다. 먼저, 6개(ECMWF, MSC, NCEP, KMA, UKMO, CMA) 개별 기후모델의 S2S 예측 원시자료를 딥러닝 모델에 입력하기 위한 훈련자료로 변환하고, 이들로부터 MME(multi-model ensemble) 기반 S2S 훈련자료도 추가적으로 구축하였다. 이 연구의 전처리 과정은 최고 및 최저 차이를 계산하여 입력 자료의 범위를 변형시키는 MinMax 및 MaxAbs 기법, 표준편차를 이용하는 Standard 기법 및 사분범위 등 분위수를 지정하여 변형하는 Robust와 Quantile 기법과 이런 기법이 적용되지 않는 경우로 구성된 파이프라인을 구축하였는데, 변형된 훈련자료와 예측 변수와의 상관관계를 계산하여 순위에 따라 훈련자료의 특성을 선택하는 기법(feature selection)을 추가하였다. 이 연구의 학습 모델은 U-Net에 간단하게 TimeDistributed를 합성곱 층에 적용하고 이것을 기준 모델(Base model)로 명명하였다. 6개 기법으로 전처리 된 6개 개별 기후모델 및 MME 기반 일 최고 및 최저기온과 일 강수 훈련자료를 Base model에서 예측한 결과에 대해 평가된 선행예측시간별(lead time) 공간상관계수(Spatial Pattern Correlation Coefficient)를 보정 전 기후 예측 자료, 즉 관측값

(ERA5)과 비교한 결과, 적합한 전처리 기법은 표준편차(Standard) 및 사분범위(Robust) 기법으로 처리된 일 강수와 일 최고 및 최저기온의 예측성능이 개선되었다. 특히 일 강수의 예측성능은 전체 lead-time에서 개선되었으며, 일 최고 및 최저기온의 예측성능은 1주 개선이 미흡했지만 3-4주에서 개선되어 적합한 전처리 기법으로 변형된 S2S 훈련자료에 대한 Base model의 후보정 개선 효과가 있음을 확인하였다. 또한, 피쳐 선택 기법이 적용된 일 강수의 경우 피쳐 선택에 따른 훈련자료의 차원 감소(예, 입력 변수 감소)가 예측성능 변화에 영향을 미치지 않는 것으로 나타났다. 그러나 일 최고 및 최저기온의 경우 피쳐 선택에 의한 훈련자료의 입력 변수가 기준 이하(예, 입력 변수 7개 선택)일 경우 예측성능이 저하될 수 있는 것으로 확인되었으며, 여전히 피쳐 선택 기법이 일 최고 및 최저기온의 1주 예측성능 개선에 기여하지 않는 것으로 나타나서, S2S 기온 훈련자료의 예측성능 개선에는 강수와 다른 딥러닝 훈련 모델 적용 등 다른 접근이 필요할 것으로 판단된다. 후속 연구에서는 Base model의 1주 개선이 미미한 이유를 탐색하고, Base model의 하이퍼파라미터 및 구조 등을 최적화하며 다양한 훈련 모델을 추가하여 S2S 예측 변수의 예측성능을 개선할 수 있는 딥러닝 모델을 찾고자 한다.

준지도학습 적용 환경구축을 위해 그래프기반 모델과 이미지기반 모델을 구축하고 지도학습을 수행하였다. 인공지능모델 훈련의 가장 큰 걸림돌인 자료 부족 문제를 완화하고자 기상기후자료의 자료강화를 테스트하였는데, 일자료의 시간적 조합을 활용한 자료강화와 월자료의 공간적 재조합을 활용한 자료강화를 적용하여 비교하였다. 또한 목표 월만 사용하는 경우와 모든 월을 사용하는 경우의 결과를 비교하였다. 그래프기반 모델로는 노드 분류를 위한 그래프합성곱신경망 모델과 그래프 분류를 위한 그래프합성곱신경망 모델을 각각 구축하여 3-폴드 교차검증으로 우리나라 여름철 평균기온 예측 성능을 평가하였다. 노드 분류와 그래프 분류 모델 모두 모델의 훈련이 전체적으로 잘 이루어지지 않아 자료강화의 개선 효과는 확인할 수 없었는데, 다만 노드 분류 모델의 테스트 결과에 있어서는 7월(LT2)의 경우 목표 월만 사용하고 자료강화하지 않은 경우의 성능이 낮아 모든 월을 사용한 경우나 자료강화한 경우의 개선 효과가 나타났으며(HSS>0인 폴드 확장), 그래프분류 모델의 경우 6월(LT1)에는 모든 월을 사용한 경우의 예측 성능이 우수하였고(모든 폴드에서 HSS>0.35) 7월(LT2)의 경우는 자료강화의 개선 효과를 관찰할 수 있었다. 이미지기반 모델로는 3차원 합성곱신경망 모델을 구축하여 예측 성능을 평가하였는데, 월자료의 공간적 재조합을 활용한 자료강화를 수행하고 모든 월을 사용한 경우 다른 경우에 비해 모델이 비교적 잘 학습되었으며 검증 정확도도 높게 나타났다(모든 월, 모든 폴드에서 정확도>0.6). 테스트 결과를 보면 6월(LT1)과 7월(LT2)의 경우 자료 강화의 개선 효과를 확인할 수 있었다. 모델의 마지막 합성곱층의 활성화맵을 그려 입력자료에서 중요하게 기여한 지리적 위치를 도출하였는데, 자료강화를 통해 기후학적으로 의미있는 지역이 활성화된 사례를 확인할 수 있었다. 그래프기반 모델과 이미지기반 모델 모두 매우 단순한 구조의 모델을 구축하였으며 모델 매개변수 및 초매개변수의 최적화는 수행되지 않았으므로 후속 연구를 통해 개선될 여지가 크다. 그래프기반모델의 경우 특히 노드 간의 유사성을 실제와 가깝게 도출하여 엣지 연결에 반영할 필요가 있다. 1차년도에는 단순하게 구성한 그래프기반모델 및 이미지기반모델의 지도학습 성능을 알아보고 자료강화 등의 개선 효과를 살펴보았으며 2차년도 연구를 통해 준지도학습을 적용하여 모델을 개선하고 최적화를 수행할 것이다.

목차

평문 초록	i
Executive Summary	ii
국문 요약	iv
목차	vi
표 차례	viii
그림 차례	xi
1. 서론	1
2. 방법 및 자료	4
가. 설명가능 인공지능(XAI) 기술적용을 위한 환경구축	4
(1) 설명가능 인공지능 기술의 도입	4
(2) 기준 모델 선정 및 구현	4
나. 제한된 기후자료 확장을 위한 환경구축	7
(1) 자료 파이프라인 구축: 인공지능 훈련자료 확장	7
(2) 자료 파이프라인 구축: 훈련자료 전처리 기법 테스트	22
다. 준지도학습 적용을 위한 환경구축	34
(1) 자료 수집 및 전처리	34
(2) 기상기후자료 자료강화	35
(3) 합성곱신경망(CNN) 모델 구축 및 테스트	37
(4) 그래프신경망(GNN) 모델 구축 및 테스트	39
3. 결과	43
가. 설명가능 인공지능(XAI) 기술적용을 위한 환경구축	43
(1) 기준 모델의 예측 성능 평가	43
나. 제한된 기후자료 확장을 위한 환경구축	47
(1) 자료 파이프라인 구축: 인공지능 훈련자료 확장	47
(2) 자료 파이프라인 구축: 훈련자료 전처리 기법 테스트	52
다. 준지도학습 적용을 위한 환경구축	67
(1) 합성곱신경망(CNN) 모델 훈련 및 검증	67
(2) 합성곱신경망(CNN) 모델 지도학습 성능평가	74
(3) 노드 분류 그래프신경망(CNN) 모델 훈련 및 검증	81

(4) 노드 분류 그래프신경망(CNN) 모델 지도학습 성능평가	84
(5) 그래프 분류 그래프신경망(CNN) 모델 훈련 및 검증	86
(6) 그래프 분류 그래프신경망(CNN) 모델 지도학습 성능평가	89
4. 요약 및 결론	92
사사	95
References	96

표 차례

Table	page
1. Summary of ECMWF, GLOSE5 model.	7
2. Data description of ERA5.	12
3. Overview of the characteristics of ERA5 compared with ERA-Interim.	13
4. Test years and their tercile categories for June, July and August.	77

그림 차례

Figure	page
1. U-Net based encoder-decoder model structure with convolution block attention module and conditional dropout mechanism in the bottleneck.	4
2. Operation procedure of the depth-wise separable convolution block.	5
3. Conceptual diagram of the conditional dropout mechanism.	6
4. Model architecture for the post-processing model with 3D convolution blocks.	6
5. A diagram of this study.	7
6. ECMWF S2S Precipitation data.	8
7. GPM IMERG data.	9
8. Normalized Difference Vegetation Index data from NOAA AVHRR and VIIRS sensor.	10
9. Sea surface temperature from NOAA.	11
10. ERA5 (ECMWF reanalysis v5) precipitation data.	12
11. U-Net architecture of contracting and expanding path.	17
12. Model pattern Extraction and visualization of U-Net.	20
13. The structure of in and out data system.	22
14. The work-flow of the study.	23
15. The explanation of data pre-processing step and the transformed data at each step. ...	24
16. Overview of data transforms (source: MachineLearningMastery.com).	25
17. Explain about the difference for feature extraction (left) and feature selection (right). 27	
18. Overview of feature selection (source: MachineLearningMastery.com).	29
19. Description of the total training periods for six individual climate models and MME possible periods.	30
20. Architecture of U-Net wrapped by the TimeDistributed layer in the study.	31
21. The domain of the constructed training data of Sub-seasonal to seasonal. The black box indicates the input and output domains of the deep learning training model in this study. ...	32
22. Conceptual image of the Cutmix data augmentation technique applied in this study.	37
23. Structure of the Vanilla 3D CNN model used in this study.	38
24. Conceptual diagram of the split of training, validation, and test sets in a transductive Node Classification model.	39
25. Structure of the GNN model for Node classification used in this study.	41

26. Conceptual diagram of the split of training, validation, and test sets in a Graph Classification model.	41
27. Structure of the GNN model for Graph classification used in this study.	42
28. Weights of the 3D convolution kernels of the input layer with the dimension of 2(latitude) x 3(longitude) x 7(variable) x 8(number of kernels).	44
29. Correlation coefficients between kernel structures.	44
30. Visualization of activated values of the first 3D convolution layer.	45
31. Visualization of the connection strength of the last hidden layer and the output layer.	46
32. Comparison of the result of ConvLSTM with ECMWF model at daily precipitation.	47
33. Comparison of performance of pattern correlation and RMSE of precipitation for 3 week lead-time.	48
34. Comparison of the result of U-Net with ECMWF model at daily precipitation.	48
35. The Connectivity from various module for in and out data system.	49
36. GUI of in and out data system.	50
37. Visualization of verification for training data from in and out data system.	51
38. Sensitivity analysis of feature selection from in and out data system.	51
39. Comparison of histograms of transformed training data for Sub-seasonal to seasonal (S2S) by minmax scaler and maxabs scaler: MinMax and MaxAbs scalers are the algorithm of change scale of data.	52
40. Comparison of histograms of transformed training data for S2S by standard scaler: Standard scaler is the standard deviation algorithm of change scale of data.	52
41. Comparison of histograms of transformed training data for S2S by robust scaler and quantile transformer: Robust scaler is the algorithm of change scale of data, and quantile transformer is the algorithm of change distribution of data.	53
42. Comparison of the learning performance evaluation index of the base model for the S2S training data converted by each scaler algorithm over the training period. A) is the daily minimum air temperature and B) is the daily total precipitation of ECMWF S2S training data transformed from six scalers.	54
43. Comparison of prediction performance evaluation (spatial Pattern Correlation Coefficient, PCC) of base model for the daily maximum air temperature S2S training data transformed by each scaler algorithm during test period of each individual climate model and MME. Black color presents PCC of raw S2S data, and gray represents Unscaled, orange color is MaxAbs	

scaler, blue color is Robust scaler, green presents Standard scaler, purple color is MinMax scaler, sky-blue presents Quantile transformer result.	55
44. Comparison of PCC about prediction performance evaluation of base model for the daily minimum air temperature S2S training data transformed by each scaler algorithm during test period of each individual climate model and MME.	55
45. Comparison of PCC for prediction performance evaluation of base model for the daily total precipitation S2S training data transformed by each scaler algorithm during test period of each individual climate model and MME.	56
46. Comparison of skill scores for six individual climate models and MME daily maximum air temperature S2S training data transformed into standard scaler.	57
47. Comparison of skill scores for six individual climate models and MME daily minimum air temperature S2S training data transformed into standard scaler.	57
48. Comparison of skill scores for six individual climate models and MME daily total precipitation S2S training data transformed into standard scaler.	58
49. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of MME.	59
50. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of ECMWF.	59
51. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of ECCO.	59
52. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of NCEP.	60
53. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of UKMO.	61
54. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of GloSea5.	61
55. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of CMA.	62
56. Skill score evaluation according to four groups for major variables for 6 individual climate models and MME daily total precipitation prediction.	63
57. Skill score evaluation according to four groups for major variables for 6 individual climate models and MME daily maximum air temperature prediction.	64

58. Skill score evaluation according to four groups for major variables for 6 individual climate models and MME daily minimum air temperature prediction.	64
59. Comparison of Class Activation Maps according to four groups for major variables for MME daily maximum air temperature prediction.	65
60. Comparison of differences in prediction performance for MMEs with different training periods. Left represents the PCC of MME based on the daily total precipitation, center explains the PCC of MME based on the daily maximum air temperature, and right presents the PCC for MME based on the daily minimum air temperature.	66
61. Comparison of validation loss (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.	67
62. Comparison of train loss (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.	68
63. Comparison of validation accuracy (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.	68
64. Comparison of train accuracy (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.	69
65. Comparison of validation loss (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.	70
66. Comparison of train loss (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.	70
67. Comparison of validation accuracy (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.	71
68. Comparison of train accuracy (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.	71
69. Comparison of validation accuracy (Month-Agnostic) between data pre-processing and augmentation methods.	72
70. Comparison of validation loss (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.	72
71. Comparison of train loss (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.	73
72. Comparison of validation accuracy (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.	73
73. Comparison of train accuracy (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.	73

Augmentation.	74
74. Proportion Correct (PC) between data pre-processing and augmentation methods.	75
75. Heidke Skill Score (HSS) between data pre-processing and augmentation methods.	75
76. Area Under the ROC Curve (AUC) between data pre-processing and augmentation methods.	76
77. Ranked Probability Skill Score (RPSS) between data pre-processing and augmentation methods.	76
78. Class activation map before (left) and after (right) the data augmentation.	77
79. Class activation map before (left) and after (right) the data augmentation.	78
80. Validation accuracy of LT1-3 forecast of monthly mean temperature from January to December.	78
81. Test accuracy of LT1-3 forecast of monthly mean temperature from January to December.	79
82. Test Heidke Skill Score (HSS) of LT1-3 forecast of monthly mean temperature from January to December.	79
83. Test Heidke Skill Score (HSS) of LT1 forecast of monthly mean temperature.	80
84. Test Heidke Skill Score (HSS) of LT2 forecast of monthly mean temperature.	81
85. Test Heidke Skill Score (HSS) of LT3 forecast of monthly mean temperature.	81
86. Comparison of train loss between the use of months and data augmentation methods. ..	82
87. Comparison of validation loss between the use of months and data augmentation methods.	
83	
88. Comparison of train accuracy between the use of months and data augmentation methods.	83
89. Comparison of validation accuracy between the use of months and data augmentation methods.	84
90. Comparison of validation accuracy between the use of months and data augmentation methods.	84
91. Proportion Correct (PC) between the use of months and data augmentation methods. ...	85
92. Heidke Skill Score (HSS) between the use of months and data augmentation methods. ..	85
93. Area Under the ROC Curve (AUC) between the use of months and data augmentation methods.	86
94. Ranked Probability Skill Score (RPSS) between the use of months and data augmentation	

methods.	86
95. Comparison of train loss between the use of months and data augmentation methods. ..	87
96. Comparison of validation loss between the use of months and data augmentation methods.	87
97. Comparison of train accuracy between the use of months and data augmentation methods.	88
98. Comparison of validation accuracy between the use of months and data augmentation methods.	88
99. Comparison of validation accuracy between the use of months and data augmentation methods.	89
100. Proportion Correct (PC) between the use of months and data augmentation methods. ..	89
101. Heidke Skill Score (HSS) between the use of months and data augmentation methods. ..	90
102. Area Under the ROC Curve (AUC) between the use of months and data augmentation methods.	90
103. Ranked Probability Skill Score (RPSS) between the use of months and data augmentation methods.	91
104. Comparison of Heidke Skill Score (HSS) between various AI models.	91

1. 서론

딥러닝을 포함한 인공지능(Artificial Intelligence; AI) 기술의 빠른 성장, 고성능 연산(High Performance Computing; HPC) 자원의 발전과 다양한 분야에 대한 성공적인 적용을 바탕으로 기상 및 기후 예측을 위한 AI 활용이 확대되고 있다. 기상 및 기후 예측의 다양한 적용 범위 및 시간 스케일에 대해 AI 기술이 적용되고 있는데 특히 중단기 예측을 위해 수치기상예측(Numerical Weather Prediction; NWP) 모델의 후처리를 포함한 여러 구성요소를 딥러닝으로 치환 또는 강화하거나 NWP의 전체 과정을 딥러닝으로 대체할 수 있는지 시험하는 연구 등이 활발히 수행되고 있으며(Schultz et al., 2021, Weyn et al., 2019), 계절내 및 계절 예측의 경우에도 MJO나 El Nino 등 기후지수를 예측하는 연구 등(e.g., Nooteboom et al., 2018; Ham et al., 2019; Kim et al., 2021)이 수행되고 있다.

기상 및 기후예측을 위한 AI 기술은 적용 범위에 따라 역학 모델의 전체 과정을 대체하는 Hard AI, 예측 개선을 위한 도구로 활용하는 Medium AI, 연산의 효율성을 개선하기 위해 커널과 같은 역학 모델의 일부를 모사하는 Soft AI로 나누어 생각할 수 있으며(Chantry et al., 2021), 관심을 가지는 시간 스케일에 따라 목표 변수에 영향을 미치는 물리 과정과 공간 스케일이 다르게 나타나므로 다양하게 적용될 수 있다. 기상 예측과 비교하여, 계절내/계절예측 또는 기후 역학 모델(General Circulation Models; GCMs) 결과 개선을 위해 AI 기술을 적용한 연구는 아직 많지 않으며 특히 우리나라 및 동아시아 장기 기후예측을 위한 딥러닝 모델의 개발이나 계절내 /계절예측 역학 모델 결과의 후처리 등의 연구는 미흡한 실정이다.

최근 이상기후에 따른 단기예보와 중장기 예보를 잇는 3~4주의 계절 내 예측자료가 한국 뿐만 아니라 아시아 태평양 지역에서도 다양한 기후 예·경보와 응용분야 활용에 갈수록 중요성이 증대되고 있다. 이에 따라 세계적으로 S2S (Subseasonal to Seasonal) 연구가 활발히 진행되고 있으며, 기상청은 최근 GloSea6 (Global Seasonal Forecasting Model Version 6) 모델을 개발하고 자료를 제공하고 있다. 하지만 모델에 기반한 예측자료는 다양한 변수로 계절내 예측에 어려움이 많은 것이 사실이다. 단기 예보 혹은 계절 기후 규모에서 예측성 향상을 위한 기후 예측 모델의 딥러닝 활용, 혹은 기후모델의 매개변수 모수화를 위한 딥러닝 적용 등의 연구는 최근 증가하는 반면, 계절내-계절(Sub-seasonal to seasonal, 이하 S2S) 시간 규모에 대한 연구는 상대적으로 부족한데 그것은 S2S 시간 규모의 본질적인 예측력 원천 부족과 선행시간이 길어질수록 증가하는 비선형성으로 인해 획기적인 기후 역학 모델의 예측성 개선을 기대할 수 없기 때문이다. 이 밖에도 계절내 또는 계절 예측을 위한 AI 활용에는 여러 어려움이 존재하는데 특히 1) 긴 시간 스케일로 인해 짧은 관측 기간 안에서 활용 가능한 훈련 자료가 부족하고 2) 초기 조건과 경계 조건이 모두 중요해(Reichler and Roads, 2003.) 고려해야 하는 변수의 종류와 공간 스케일이 방대하므로 자료 수집 및 처리의 어려움이 있으며 3) 관련 연구가 아직 많지 않아 최적 기술에 대한 합의가 이루어져 있지 않다.

신기술을 적용한 기법은 기존의 다양한 방법에서 발생할 수 있는 오차를 최대한 줄이기 위한 유용한 기법이지만 기후모델 자료의 정확도 높은 자료 생산에 한계가 있어 이를 극복하기 위하여 시계열 예측에 유용한 순환신경망(Recurrent Neural Network, RNN) 중에서 LSTM 기법과 공간에서 적절한 특성을 찾고 시계열 자료에 이용될 수 있는 ConvLSTM을 계절내 예측

자료 학습에 이용하고 있으나 여전히 부족한 훈련자료는 신뢰성 있는 예측자료를 생산하는데 어려움을 갖게 한다. 이 연구에서는 부족한 자료를 위성자료를 활용하여 훈련자료를 확장하고 확장된 자료들을 쉽게 딥러닝 모델에 활용할 수 있는 입출력 시스템을 개발하여 향후 민감도 분석에 활용하고자 한다.

통계적 보정, 예를 들면 기계학습 및 딥러닝 후보정(post-processing) 적용을 통해 개선 효과를 올리고자 하는 연구가 증가하고 있다. 예측된 일기 예보에 신경망(neural network) 기반의 후보정 적용을 통해 향상된 예보성능을 보여주었는데 날씨 및 기후모델에서 강수량을 개선하기 위해 이 기법을 전지구 영역으로 확장할 수 있다고 하였다(Bihlo and Popovych, 2022). 그러나 여전히 S2S 예측성 향상을 위한 기계학습 및 딥러닝 활용 연구는 상대적으로 적다. 이것은 기계학습 및 딥러닝 활용을 위한 훈련자료의 부족과 자료 전처리 기법, 모델 구조 등 기후 예측 적용을 위한 최적 기술에 대한 합의 부족 등의 어려움 때문이다. S2S 예측 기후모델 자료는 개별모델의 예측 자료 생산일(forecast frequency), 예측 길이(forecast range)의 다양함 등으로 인해 사용하기 어렵고 번거로운 특성이 있다. 또한 개별 모델의 출력 기후 변수의 수, 예를 들면 GloSea5의 경우 13개 기후 변수가 제공되며 ECMWF, ECCO, UKMO, CMA 및 NCEP 등의 출력 기후 변수는 16개 이상이다. 또한 최고 및 최저기온은 6시간 단위로 제공되는 반면, 그 외 기후 변수는 일별 단위로 파일 포맷 및 배열 변형 등 처리하고 바로 이용하기 어렵다는 문제들을 가지고 있다. 또한 기계학습 및 딥러닝 적용 시 S2S의 예측 개선 효과는 S2S 훈련자료의 전처리 기법, 예를 들면 어떤 변환기로 원시자료를 어떻게 변형시키는지 혹은 예측 목표 변수와 무관하거나 중복된 입력 변수 혹은 광범위한 정보로 구성된 원시자료에서 예측에 영향을 미치는 특성(예, 기후 변수)을 어떻게 조합 및 선택하느냐에 따라 학습 알고리즘(예, 신경망 모델)을 산만하게 하거나 잘못 이끌어서 예측성능을 저하시키는 등 예측결과가 크게 달라질 수 있는데 이처럼 기계학습 및 딥러닝 모델에 입력하기 위한 S2S 훈련자료 생산에 적합한 전처리 기법 등에 관련한 기존 연구들이 부족하다. 강수 예측에 대부분 0과 1 혹은 -1에서 1 사이의 값으로 자료의 범위를 변형시키는 정규화 기법을 활용하거나(Gerritsma, 2020; Ha et al., 2016; Singh et al., 2021) 기온 혹은 해수면 온도(sea surface temperature) 예측, 혹은 미세먼지 예측에도 정규화 기법을 활용하였다(Son et al., 2021; Xiao et al., 2019; Zhang et al., 2020). Hong et al.(2021)은 지상관측자료 기반의 기후변수 예측을 위해 각 기후변수의 평균을 빼고 표준편차로 나누는 표준화 기법을 활용하였다.

따라서 이 연구에서는 S2S 예측 원시자료 처리에 적합한 기법이 무엇인지 알아보고 변형된 훈련자료에 대한 예측성능 민감도 분석을 통해 S2S 훈련자료에 적절한 전처리 기법을 알아보고자 한다. 자료 전처리는 여러 단계로 나눌 수 있는데, 먼저 S2S 원시자료에 다양한 변환기 적용을 통해 다르게 변환된 S2S 훈련자료에 따른 예측성능을 비교한다. 두 번째는 S2S 원시자료에서 모든 특성, 즉 모든 기후 변수를 사용하지 않고 차원 감소 기법(예, 피쳐 선택)을 통해 목표 기후 변수 예측에 영향을 미칠 수 있는 변수 선택을 통해 변수, 즉 특성 수를 줄이고 우수하다고 판단된 변환기 적용을 통해 그 예측성능 변화를 평가하고자 한다. 그래서 이 연구는 딥러닝 기반의 후처리 기술 개발을 위한 첫 단계로 S2S 예측 일 최고 및 최저기온, 강수의 예측성, 특히 2-4주 예측성 개선을 위한 S2S 훈련자료에 적합한 전처리 기법을 규명하고자 하였다.

계절내 또는 계절 스케일의 기후 예측을 위한 자료의 특성을 파악하고, 훈련 자료 확장 및 입출력 시스템 개발, 다양한 전처리기법 적용을 통한 딥러닝 모델 아키텍처 최적화, 준지도 학습(참값이 존재하거나 존재하지 않는 다양한 자료를 함께 사용하여 모델의 성능을 개선시키는 방법)을 통한 모델 성능 개선 등을 통해 앞서 언급한 여러 어려움을 극복하여 객관적 기후 예측에 적합한 인공지능활용 기반기술을 개발할 필요가 있으므로 APCC의 6단계('22-' 24년) 연구를 통해 기후예측에 적합한(Climature Prediction-Tailored) AI 기반기술 개발하고자 한다.

복잡한 인공신경망의 계산 과정은 인간이 평가하기 힘든 측면이 있어, 인공신경망 기반의 인공지능 시스템의 예측 결과를 온전히 신뢰하기 어려울 뿐만 아니라, 인공신경망의 아키텍처가 적절히 구성 및 적용되었는지 평가하지 못한 채 소수의 결과론적 평가지표를 기준으로 모델의 타당성을 평가하고 있는 실정이다. 과거 AI 기술은 블랙박스로 여겨져 예측성이 우수한 경우에도 활용의 장벽이 존재해 왔는데, 근래에는 점차 AI 기술의 신뢰성(Trustworthiness)과 해석 가능성(Interpretability)이 성공적인 AI 기술을 위한 필수적인 요소로 요구되고 있다. 이 연구에서는 AI 모델의 개발 및 개선 과정에서 다양한 설명가능한 인공지능(eXplainable AI; XAI) 기술을 적용하여 모델의 연산 과정에 대한 이해를 증진하고 모델의 과적합, 불안정성을 해소하여 신뢰할 수 있는(Trustworthy) 기후예측 AI 기반기술을 개발하고자 한다. 설명가능한 인공지능은 판단에 대한 이유를 사람이 이해할 수 있는 방식으로 제시하는 인공지능으로, 기후예측을 위한 인공지능 알고리즘의 연산 과정과 내부 구조를 들여다봄으로써 아키텍처 및 초매개변수의 변화에 대한 알고리즘의 반응을 살피고, 이를 토대로 최적의 기후예측정보 생산을 위한 인공지능의 개선 방향을 설정하여 인공지능 기법을 해석 가능한 영역으로 확대할 수 있다.

또한 이 연구를 통해 개선된 계절내 또는 계절예측 결과를 생산하여 궁극적으로 다양하고 손에 잡히는(Tangible) 활용 성과 창출을 위한 자료를 제공하고자 한다. 예를 들어, 계절내 시간 스케일은 농업 분야에서 쓰이는 시간 스케일(순)과 유사하여 개선된 2-4주 예측 정보를 농업 정보로 가공하여 활용할 수 있다. 지난 4·5단계 과제를 통해 수행한 인공지능 적용 연구를 통해 축적한 경험과 노하우를 바탕으로 기후예측분야 인공지능기술 적용의 특성 및 한계를 파악하고 극복하여 기존의 여러 기후예측 방법론을 보완하고 성능을 개선하기 위한 인공지능 활용 기반기술을 개발하고자 하며 APCC는 국내를 포함한 아태지역에 대한 최적의 기후예측정보를 생산하여 경제적 손실을 경감하고 인명과 재산을 보호할 미션을 지니고 있으므로 이 연구를 통해 기후예측분야에 적합한 인공지능활용 기반기술을 개발하여 이에 부합하고자 한다. 6단계 연구의 목표는 기후예측에 적합한 인공지능활용 기반기술 개발로, 이를 통해 계절내 또는 계절 스케일의 기후 예측을 위한 자료 및 역학 모델의 특성을 파악하고, 훈련 자료 확장 및 입출력 시스템 개발, 다양한 전처리기법 적용 및 초매개변수, 다양한 알고리즘 조정 등을 통한 딥러닝 모델 아키텍처 최적화, 준지도학습을 통한 모델 성능 개선 등을 통해 앞서 언급한 여러 어려움을 극복하여 객관적 기후예측에 적합한 인공지능활용 기반기술을 개발하고자 한다. 1차년도('22년)의 목표는 기후예측에 적합한 인공지능활용 기반기술 적용을 위한 자료확장 및 환경구축으로 제한된 기후자료 확장을 위한 환경구축, 준지도학습 적용을 위한 환경구축, 설명가능 인공지능 기술적용을 위한 환경구축을 수행하였다.

2. 방법 및 자료

가. 설명가능 인공지능(XAI) 기술적용을 위한 환경구축

(1) 설명가능 인공지능 기술의 도입

이 연구에서는 설명가능한 인공지능 기술(XAI; eXplainable Artificial Intelligence)을 도입하여 인공지능 시스템의 퍼포먼스 뿐만 아니라, 인공신경망 내부 파라미터의 분배 및 자료의 처리 과정을 살펴보았다. 이를 통하여 인공신경망 모델의 구조적 적합성 여부와 인공지능의 결과 도출 과정에서 기후 자료가 물리/역학적으로 의미있는 방향으로 처리되고 있는지 확인할 수 있다. 인공신경망 내부 및 자료 처리 과정을 평가함으로써, 기후 자료 처리 및 기후 예측에 보다 적합하고 예측성이 향상된 인공신경망 기반 인공지능 기후 예측 후처리 시스템을 개발하고자 한다. 이를 위해서 설명가능한 인공지능(XAI)를 통한 계절 예측에 적합한 후처리 시스템 개발을 위한 첫 단계로서 기존 인공지능 모델을 구현하고 설명가능한 평가 환경을 구축하였다.

(2) 기준 모델 선정 및 구현

기존 인공신경망 기반 계절 예측 후처리 모델 선정과 구현을 위하여 5단계 과제를 통해 개발한 합성곱 블록 어텐션 모듈이 탑재되고 보틀넥에 조건부 드롭-아웃 메커니즘을 장착한 U-Net 기반의 인코더 디코더 모델을 포함한 다양한 종류의 아키텍처를 실험하였다.

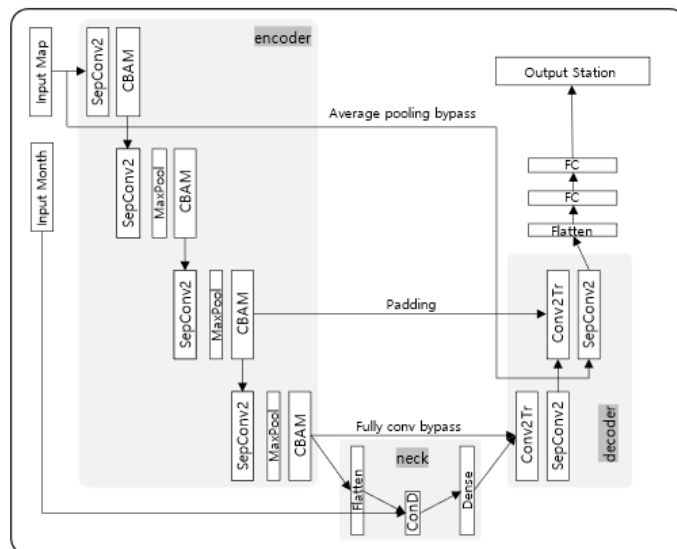


Figure 1. U-Net based encoder-decoder model structure with convolution block attention module and conditional dropout mechanism in the bottleneck.

기존 모델에서는 기후 정보를 처리하기 위해 매우 다양한 장치가 사용되었는데, 연산량을 줄이고 입력값의 마지막 차원이 일반적인 이미지 처리에서 사용하는 색상 채널이 아닌, 상호 복잡한 관련성을 가지고 있는 변수임을 고려하여 depth-wise separable 합성곱 블록을 통해 기본적인 인코딩의 뼈대를 구성하였다. 배치 정규화를 통해 활성화값의 공변적 이동을 최소화하고 Scaled Exponential Linear Unit (SELU)를 이용하여 비선형 활성화를 시키는 연산을 2회 반복하여 하나의 블록을 구성하였다. 이후 채널과 공간 어텐션을 수행하는 합성곱 블록 어텐션 모듈을 이용하여 인공지능망이 집중해야 하는 정보를 추려낼 수 있도록 조치하였다.

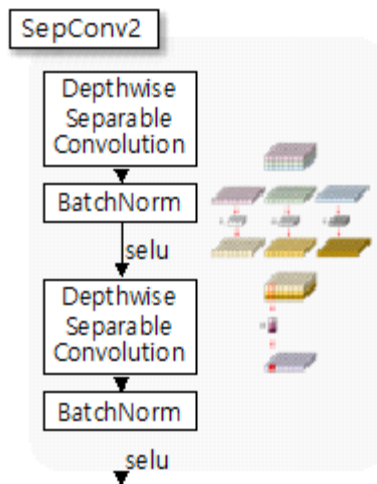


Figure 2. Operation procedure of the depth-wise separable convolution block.

이와 같은 단계를 거쳐 고차원의 자료를 잠재 차원으로 정제하는 인코딩 과정에서 단계별 정보를 기억하여 디코딩 모듈로 직접 전달함으로써, 모델의 학습 안정성을 향상했다. 특히, 일반적으로 완전연결층으로 구성되는 보틀넥에서는 기후 정보의 계절성에 관한 정보를 처리하기 위해 조건부 드롭-아웃이라는 메커니즘을 새로 개발하여 적용하였다. 조건부 드롭-아웃 메커니즘은 입력 자료의 월 정보를 받아들여 해당 월에 대응하는 보틀넥 뉴런을 제어해줌으로써 네트워크가 전체적인 특성은 공유하되, 계절 특성도 별도로 고려할 수 있도록 구성하였다. 이를 위하여 인접 월끼리는 드롭-아웃 뉴런이 일부 겹치도록 개발하였다.

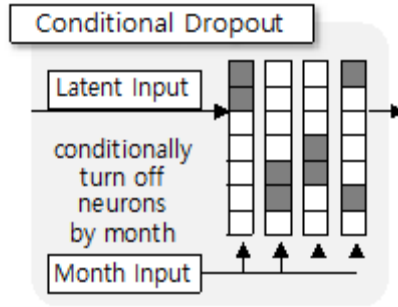


Figure 3. Conceptual diagram of the conditional dropout mechanism.

이 밖에도 바닐라 네트워크 아키텍처 혹은 단순한 합성곱 기반 신경망 아키텍처부터 복잡한 ResNet 계열의 다양한 아키텍처를 실험하였으나, 학습 자료가 부족한 상황에서 안정적인 모델 성능 평가를 얻기에는 좀 더 엄밀한 실험 설계가 필요할 것으로 판단되었다. 최종적으로는 3차원 합성곱을 이용하여 최초 다변수 계절 예측장을 처리한 후 이후 2차원 합성곱을 통해 정보를 인코딩하는 방식을 택하여 사전 학습의 학습 자료에 대한 cosine similarity 0.9, 검증 자료에 대한 cosine similarity 0.7 수준의 성능을 이루었다. 이 모델은 각 합성곱 단계에서 skip connection을 통한 bypass를 구성하여 모델의 학습 안정성을 향상한 것과 이전 복잡한 모델에 비해 학습할 파라미터의 수를 대폭 감소시킨 것이 특징이다.

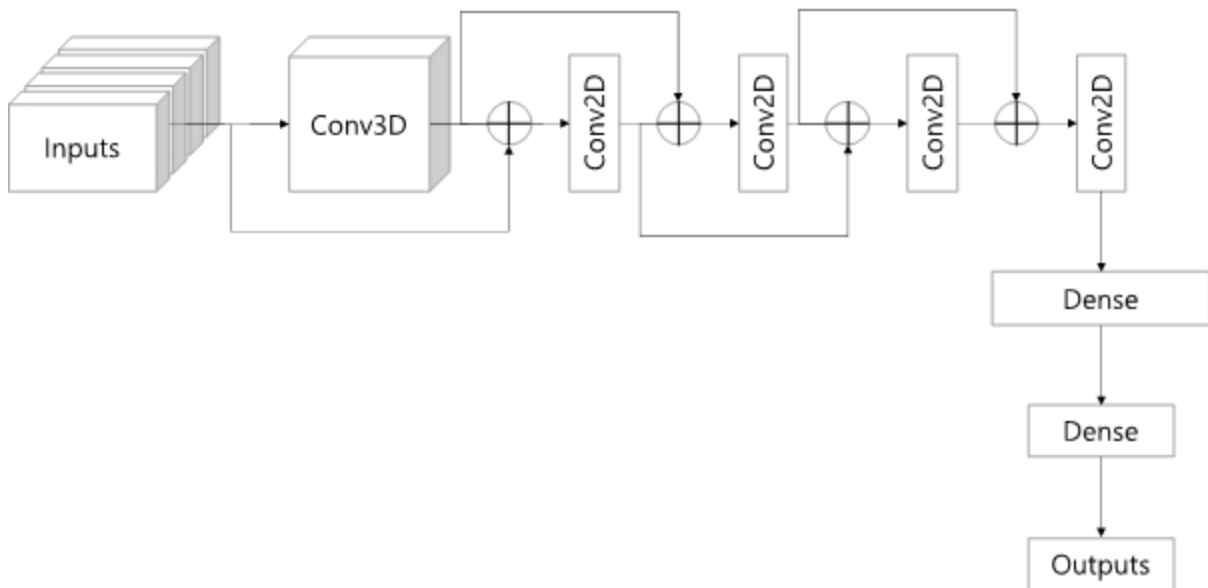


Figure 4. Model architecture for the post-processing model with 3D convolution blocks.

나. 제한된 기후자료 확장을 위한 환경구축

(1) 자료 파이프라인 구축: 인공지능 훈련자료 확장

(가) 계절내 개별 기후모델자료

자료는 부족한 자료를 계절내 예측자료와 위성자료를 이용하여 변수를 확장하고, 기간을 주별에서 계절별로 확장하고, 확장된 자료를 전처리하여 입출력 시스템의 기반 자료로 활용한다(Figure 5). 계절내 기후 예측자료는 WMO (World Meteorological Organization)의 계절내 예측 프로젝트 기반하여 다양한 기후모델 자료(ECMWF, GloSea5)를 제공하고 있다. ECMWF자료는 0-46일의 적분기간과 섭동 초기조건 앙상블을 가지고 있으며 1998년부터 2017년까지 자료를 구축하였다. 변수는 MSLP-OLR-PREC-Q700-Q850-SST-T2M-TMAX-TMIN-U50-U200-U850- V200 -W850-Z200- Z500의 자료를 딥러닝 입력변수에 맞게 전처리하였다(Figure 6). 한국 기상청에서 운용하고 있는 GloSea5 (Global Seasonal Forecasting System Version 5; MacLachlan et al., 2015; Scaife et al., 2014)는 Met Office 기후모델인 HadGEM3의 패밀리 형태의 모델로 고해상도 앙상블 모델로서 대기-지표-해양-해빙 결합모델로 대기모형은 Unified Model(UM)을 기본모델로 지면모델은 UK Land Environment Simulator (JULES; Best et al., 2011), 해양은 Nucleus for European Modeling of the Ocean (NEMO; Madec, 2008), 해빙모델은 The Los Alamos Sea Ice Mode (CICE; Hunke and Lipscomb, 2010)을 모델로 구성되어 있다. GloSea5는 N216공간해상도로 대략 위도 0.8° , 경도 0.5° 의 해상도를 가지고 있으며, 1991년부터 2010년까지 자료를 전처리 하였다(Table 1). 현재는 GloSea6 모델이 개발되어 자료를 제공중이다.



Figure 5. A diagram of this study.

Table 1. Summary of ECMWF, GLOSE5 model.

	ECMWF	GloSea5
Model version	CY41R2	GloSea5-GC2
Implement data in S2S	08/03/2016	01//11/2016
Time range	d 0-46	d 0-60
Resolution	Tco639 L91 (about 16 km) up to day 15 and Tco319 (about 32 km) after day 15 day 15	N216 L85(0.83° × 0.56°)
Ens. Size	50+1	4
Frequency	2/week	daily
Re-forecasts	on the fly	on the fly
Rfc period	past 20 years	1991-2010
Rfc frequency	2/week	4/month
Rfc size	10+1	3
Ocean resolution	1°	0.25° L75
Active Sea Ice	no	yes

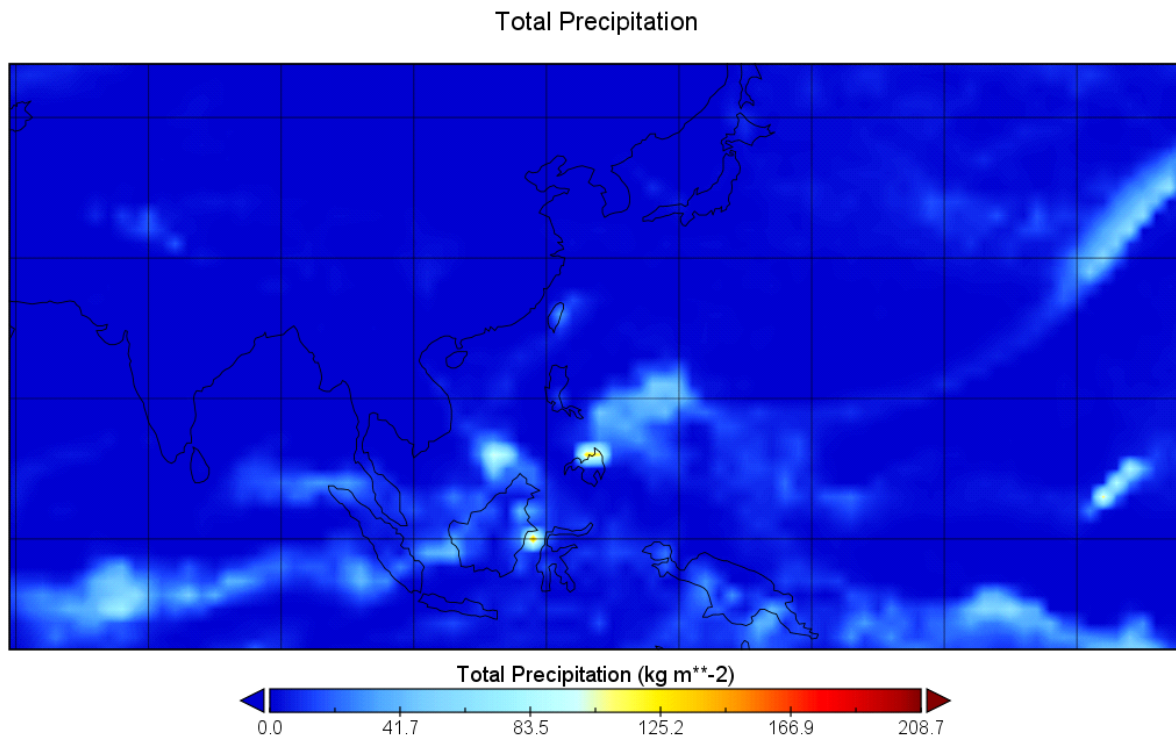


Figure 6. ECMWF S2S Precipitation data.

(나) 관측자료

관측자료는 위성자료를 기반으로 강수량, NDVI, SST을 데이터베이스화하고 자료를 수집하였다. 강수량 자료는 GPM IMERG(Integrated Multi-satellitE Retrievals for GPM)를 수집하였다 (Figure 7). GPM 자료는 Level 3 자료로 레이턴시는 전 지구 관측 후에 3.5개월 후에 자료가

업데이트 된다. 준실시간 자료를 이용하기 위해서는 GPM Early 자료를 이용하면 레이턴시 대략 6시간에 자료를 받을 수 있다. GPM IMERG자료는 2000년 6월부터 자료를 수집하였으며, HDF5를 NetCDF포맷으로 변환하여 FTP 사이트에서 자료를 수집하였다. GPM IMERG자료는 시간해상도 30분, 공간해상도 10km의 자료를 계절내 예측자료와 같은 daily, 1.5° 로 upscaling하여 자료를 데이터베이스화 하였다.

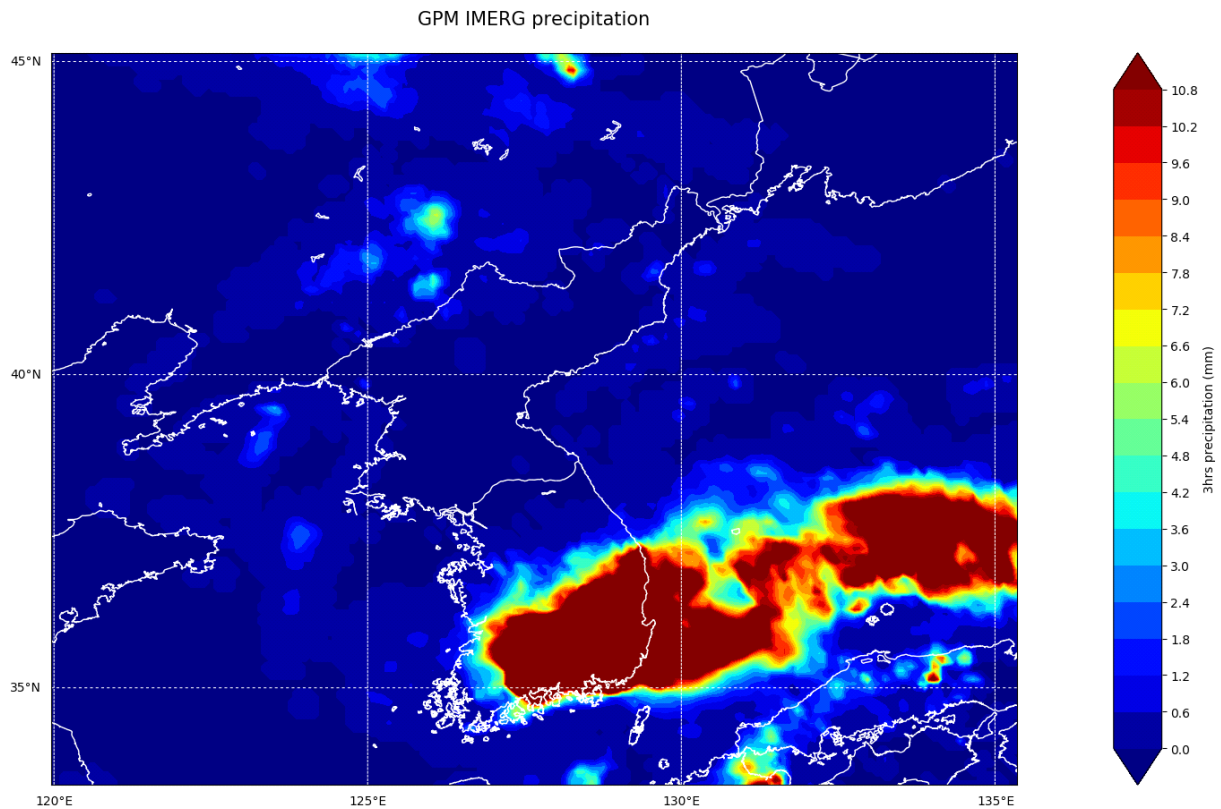


Figure 7. GPM IMERG data.

NDVI (Normalized Difference Vegetation Index)는 미국 NOAA의 CDR (Climate Data Records)프로그램 일환으로 자료를 제공하고 있으며, 1981년부터 자료를 제공하고 있다(Figure 8). 자료는 NOAA 극궤도 위성 AVHRR (Advanced very High Resolution Radiometer), VIIRS (the Visible Infrared Imaging Radiometer Suite) 센서 기반으로 만들어진 자료이며 공간해상도는 0.05° 의 해상도로 일자료를 1.5° 일자료로 변환하였다.

NOAA Climate Data Record of Normalized Difference Vegetation Index

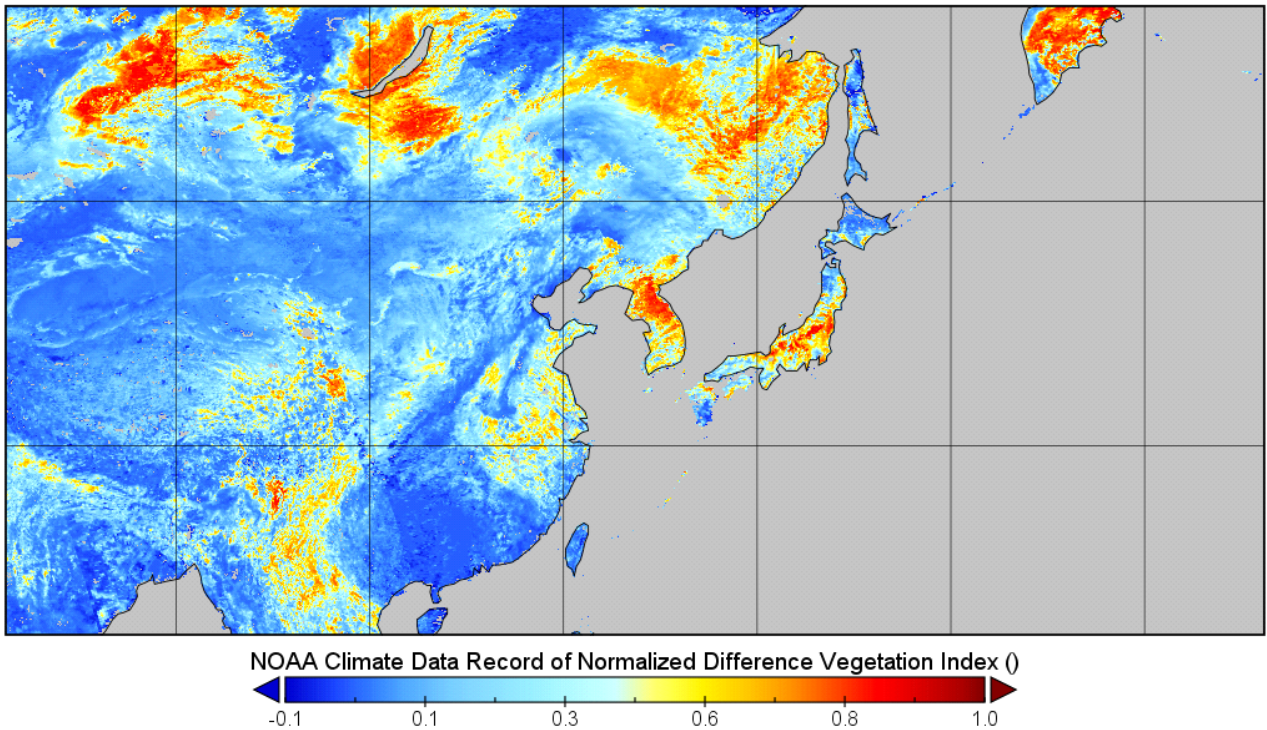


Figure 8. Normalized Difference Vegetation Index data from NOAA AVHRR and VIIRS sensor.

해수면 온도는 NOAA OI SST 자료를 이용하였다. 해수면 온도는 1981부터 일자료를 수집하였다. 공간해상도는 $0.25^{\circ} \times 0.25^{\circ}$ 의 해상도로 전지구(1140×720)를 관측한 자료이며, 자료 포맷은 NetCDF4 자료로 sea ice 자료를 결합한 자료이다(Reynolds et al., 2007). 사용할 수 있는 변수는 Sea Ice Concentration, Sea Surface Temperature, Sea Surface Temperature Anomalies 들이고, 여기서 Sea Surface Temperature 자료를 전처리하여 사용하였다(Figure 9).

Daily Sea Surface Temperature

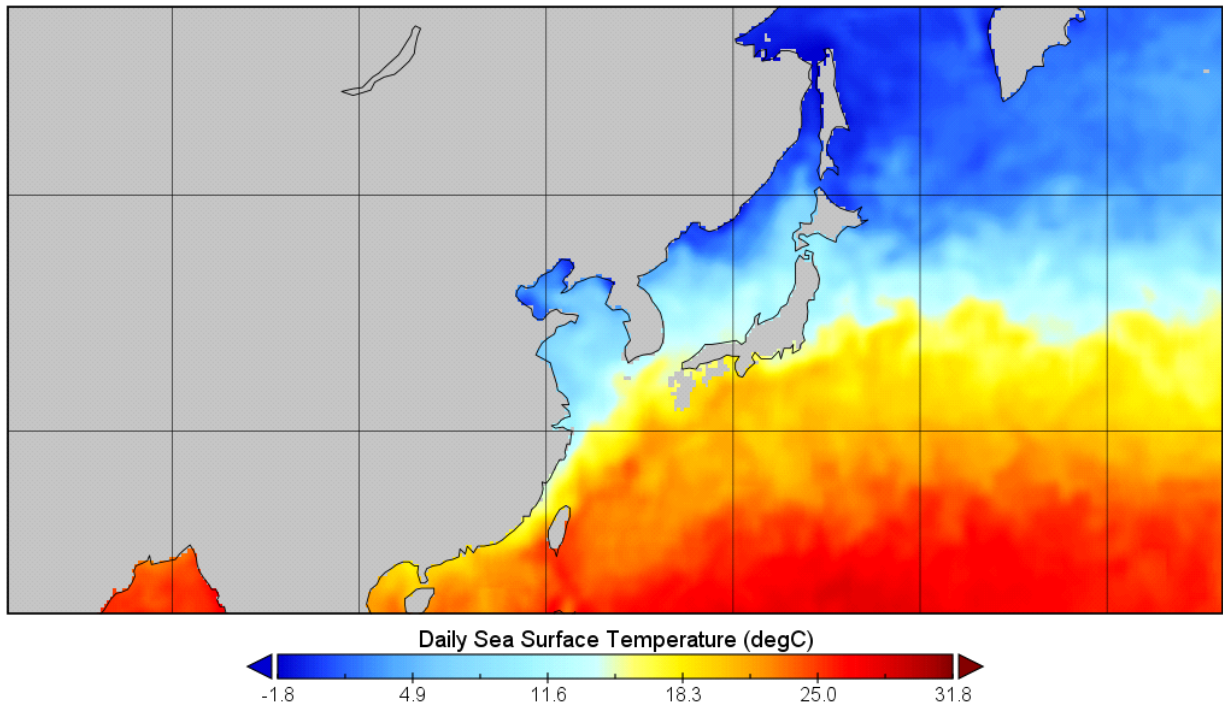


Figure 9. Sea surface temperature from NOAA.

(다) ERA5

ERA5는 ECMWF(European Center for Medium-Range Weather Forecasts)에서 5번째로 제공하는 재분석자료로 1950년부터 최근까지 자료를 제공하고 있으며, 이전 ERA-Interim을 대체한다. ERA5 재분석자료는 전지구적으로 제공받고 있는 관측자료와 모델자료를 결합하여 자료를 C3S(Copernicus Climate Change Service)에서 산출한다. ERA5는 대기, 해양, 기후 다양한 변수에 대해서 시간 간격 데이터로 제공하고 있으며, 전지구 30km 해상도와 137 연직 프로파일 해상도로 지표면에서 대기 80km까지 데이터를 제공한다. 제공되는 자료는 1979년 1월부터 2019년 12월까지 자료를 수집하였고, 자료의 포맷은 netCDF이다. 시간 해상도는 hourly 자료에서 daily 자료로, 공간적 해상도는 30km의 해상도의 자료를 계절내 예측자료와 맞게 1.5°로 변환한 자료를 이용하였다(Figure 10).

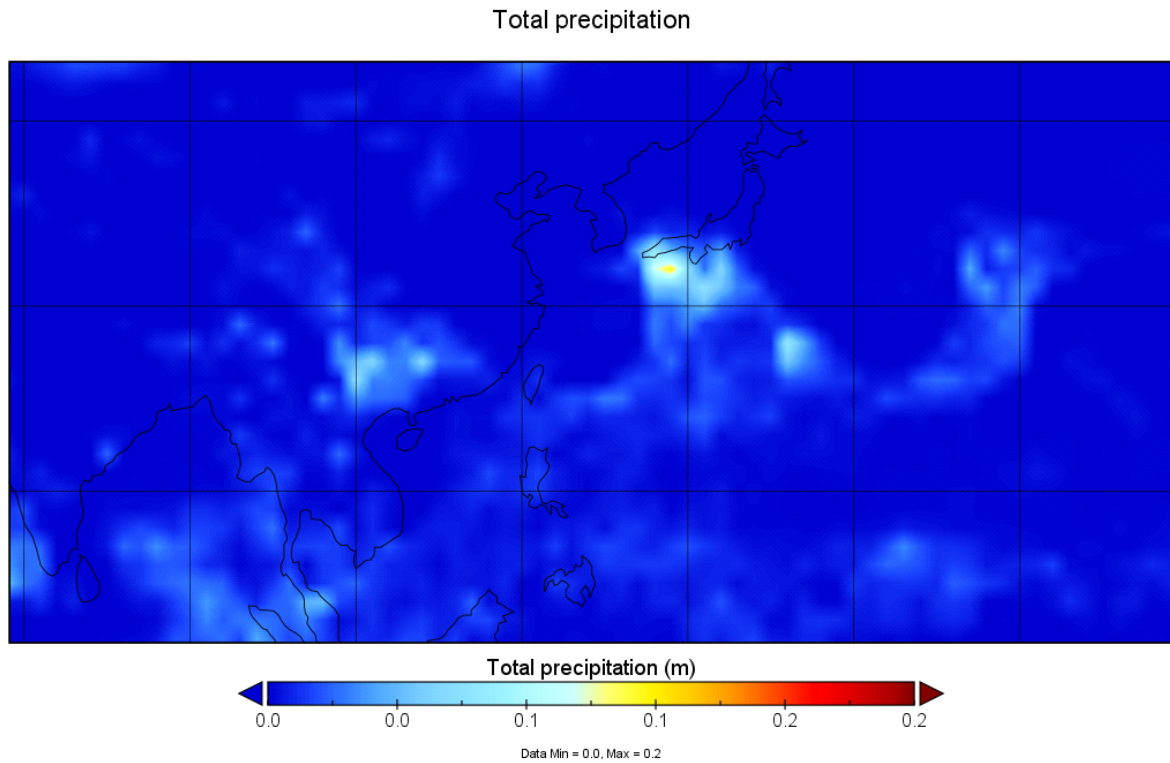


Figure 10. ERA5 (ECMWF reanalysis v5) precipitation data.

ERA5는 ECMWF, GloSea5 개별모델과 자료 기간에 맞추어 위도, 경도 배열 (61, 49)의 형태로 ECMWF 리드타임에 맞춰 최대 60일 자료를 구축하였고, 필요에 따라 GloSea5 모델의 리드타임에 자료를 수정할 수 있도록 전처리 하였다(Table 2). 타겟 변수는 강수량, 최대 및 최소 온도값을 전처리 하였다.

Table 2. Data description of ERA5.

Data Description	
Data type	Gridded
Projection	Regular latitude-longitude grid
Horizontal Coverage	Global
Horizontal resolution	Reanalysis: $0.25^{\circ} \times 0.25^{\circ}$ (atmosphere), $0.5^{\circ} \times 0.5^{\circ}$ (ocean waves) Mean, spread and members: $0.5^{\circ} \times 0.5^{\circ}$ (atmosphere), $1^{\circ} \times 1^{\circ}$ (ocean waves)
Temporal coverage	1959 to present
Temporal resolution	Hourly
File format	GRIB
Update frequency	Daily

ERA5는 이전 과제에서 이용하였던 ERA-Interim 자료와 차이가 있다. 공간해상도는 79km, 31km global로 거의 2배 좋아졌으며, 시간해상도도 6시간 1시간으로 향상되었다. 입력자료인 위성자료도 RTTOV-7에서 RTTOV-11로 보다 신뢰성 있는 자료가 입력자료로 대체되었다 (Table 3).

Table 3. Overview of the characteristics of ERA5 compared with ERA-Interim.

	ERA-Interim	ERA5	ERA5-Land
Period covered	1979-present	1950-present	1950-present
Production Period	Aug. 2006-31st Aug. 2019	Jan 2016-end 2017, then continued in near real-time	May 2018-July 2019
Model version	IFS Cycle 31r2	IFS Cycle 41r2	IFS Cycle 45r1
Assimilation system	IFS Cycle 31r2 4D-Var	IFS Cycle 41r2 4D-Var	n/a
Spatial resolution	79km globally, 60 levels to 0.1hpa	31km globally, 62km for the Ensemble of Data Assimilation(EDA), 137 levels to 0.01hpa	9km globally, 62km from ERA5 EDA equivalent surface fields
Output frequency (temporal resolution)	6-hourly analysis fields Forecast fields on surface and pressure levels 3-hourly up to 24hours, with reduce frequency up to 10days	Hourly analysis fields, 3-hourly for the Ensemble of Data Assimilations (EDA) Hourly forecast fields, 3-hourly for the Ensemble of Data Assimilations (EDA), up to 18 hours, with reduced frequency up to 10 days (not in initial release)	Hourly fields for all variables
Uncertainty estimates	None	From a 10-member Ensemble of Data Assimilations (EDA) at 63 km resolution	From ERA5-equivalent variables 10-member Ensemble of Data Assimilations (EDA) at 63 km resolution.
Model input	As in operations (inconsistent SST)	Appropriate for climate (e.g. CMIP5)	Atmospheric forcing of ERA5 with

		greenhouse gases, volcanic eruptions, SST and sea-ice cover)	additional lapse-rate correction
Input observations	As in ERA-40 and from Global Telecommunication System	In addition, various newly reprocessed datasets and recent instruments that could not be ingested in ERA-Interim	Observations indirectly influence the simulation through the atmospheric forcing from ERA5. This forcing drives the ERA5-Land single simulation and it has been obtained by assimilating observations through a 4D-VAR data assimilation system and a Simplified Extended Kalman Filter
Variational bias scheme	Satellite radiances	Also ozone, aircraft and surface pressure data	n/a
Satellite data	RTTOV-7, clear-sky, 1D-VAR rainy radiances	RTTOV-11, all-sky for various components	n/a
New parameters	ERA-Interim contains about 100 parameters on surface and single level alone, plus parameters on other level types.	ERA5 contains over 240 parameters on surface and single level alone, plus parameters on other level types. For specific parameters please compare the technical documentation linked below.	ERA5-Land contains finer-scale information of land surface component than ERA5, plus specific evaporation parameters. For specific parameters please compare the technical documentation linked below.
Parameters		A few parameters	Most of the

removed		present in ERA-Interim are not available in ERA5. For availability of specific parameters please see the technical documentation linked below.	atmospheric parameters and all the wave parameters present in ERA-5 are not available in ERA5-Land. For availability of specific parameters please see the technical documentation linked below.
Handling of accumulated parameters	Accumulated from the beginning of the forecast	Accumulated from previous post-processing	Accumulated from the beginning of the forecast

(라) 딥러닝 모델

① ConvLSTM

딥러닝 모델 중에서 CNN과 LSTM의 장점을 이용하여 격자 기반의 계절내 예측자료와 관측자료를 바탕으로 예측자료를 평가하기 위하여 ConvLSTM (Xingian Shi et al., 2015)을 이용하였다. RNN은 연속되는 계절내 예측자료 값을 시계열로 출력되는 방법으로 RNN 모델은 계절내 예측자료처럼 3-4주에 발생하는 예측값을 잘 예측하지 못하는 문제점이 있다. 이러한 문제점은 특정 시점의 오차의 출력 기울기는 출력층에서 역전파해 온 출력값의 기울기와 이후 시점에서 거슬러 온 출력 기울기 합이 되는데, 몇 개의 레이어를 걸치면서 오차를 전파하면서 기울기가 소실 또는 발산하는 문제가 있어, 특정 시점의 정보가 일정 시간이 지난 시점에까지 가중치가 전달되지 않는데 이러한 문제점을 해결하기 위해 LSTM (Long Short-Term Memory) 알고리즘이 제시되었다.

동아시아와 같은 지역에 모델 내 예측 및 관측 격자 기반의 3-4주 정확도를 평가하기 위하여 2차원 벡터를 기반으로 한 ConvLSTM을 선택하여 연구에 적용하였다. 2차원 배열을 가진 ECMWF, GloSea5의 개별모델과 관측자료를 개별 모델의 3주차, 4주차 각각의 예측자료와 예측일로부터 7일 전 관측자료를 한 개의 훈련셋으로 구성하여 타겟변수 강수량에 맞춰 ERA5 자료를 각각 3-4주차로 데이터셋을 구성하여 학습하고 결과를 분석하였다.

$$i_t = \sigma(W_{\xi}^* X_t + W_{hi}^* H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \quad [1]$$

$$f_t = \sigma(W_{xf}^* X_t + W_{hf}^* H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \quad [2]$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad [3]$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \quad [4]$$

$$H_t = o_t \odot \tanh(c_t) \quad [5]$$

ConvLSTM은 계절내 예측자료 x_t 가 2차원의 ECMWF, GloSea5과 GPM의 강수량자료를 입력 자료로 ConvLSTM의 입력, 망각, 출력 게이트 i, f, o 와 셀 입력 X , 셀 출력 C , 셀 상태 H 는 모두 3차원 텐서로 리드타임 3-4주 3차원 벡터로 예측값과 관측값을 입력으로 강수량이 산출하도록 설계하였다.

ConvLSTM은 개별모델과 관측자료는 각각 강수량을 Conv2D 필터를 통해서 피처를 뽑아내고 다시 MaxPooling을 통해서 차원을 줄이고, MaxPooling후에 다시 Conv2D를 실행하고 다시 Maxpooling을 실시하여 출력은 ERA5과 3-4주 각각의 3주차 7일자료, 4주차 7일자료를 산출하도록 ConvLSTM이 강수량을 예측 할수 있도록 Conv3D 함수로 출력이 되도록 ConvLSTM아키텍처를 설계하였다.

② U-Net

U-Net 모델의 원래 목적은 이미지 분할을 목적으로 제안된 Fully convolutional network기반의 모델이며, Figure 11과 같이 인코딩 과정과 디코딩으로 나뉘는데, 인코딩은 downsampling 과정에서 원 데이터의 특성을 추출하는 과정을 거치게 되고 디코딩 과정은 저차원에서 고차원으로 가는 과정에서 고차원의 특성을 복원해가는 과정에서 학습하게 된다(Olaf Ronneberger et al., 2015).

3×3 Convolution Layer + ReLu + BatchNorm (No Padding, Stride 1)

3×3 Convolution Layer + ReLu + BatchNorm (No Padding, Stride 1)

2×2 Max-polling Layer (Stride 2)

인코딩 과정에서는 U-Net 입력자료에서 행렬의 주변 픽셀을 고려하여 범위를 넓혀가며 feature map을 추출하고 3×3 합성곱을 수행할때마다 채널정보를 2배씩 증가시키면서 학습하는데 1->64->128처럼 레이어가 증가한다.

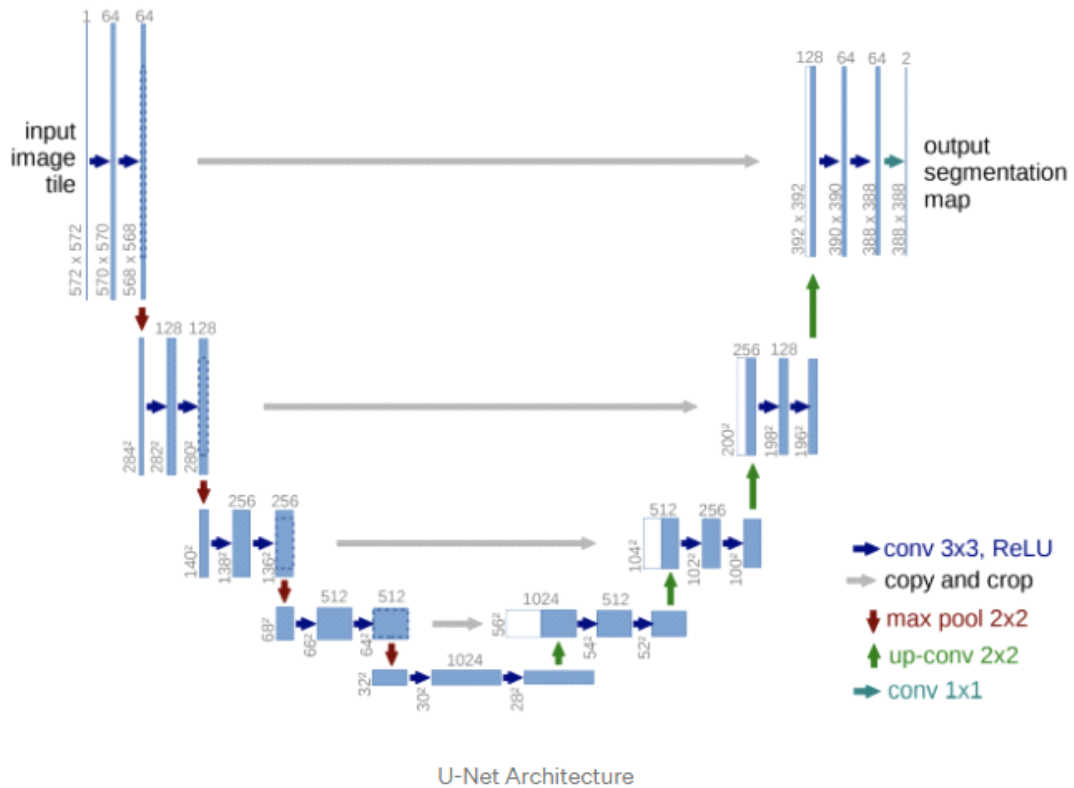


Figure 11. U-Net architecture of contracting and expanding path.

3×3 Convolution Layer + ReLu + BatchNorm (No Padding, Stride 1)

3×3 Convolution Layer + ReLu + BatchNorm (No Padding, Stride 1)

Dropout Layer

전환구간에서는 줄어든 합성곱에서 upsampling으로 가는 전 단계에서 모델을 일반화하게 한다.

2×2 Deconvolution layer (Stride 2)

3×3 Convolution Layer + ReLu + BatchNorm (No Padding, Stride 1)

3×3 Convolution Layer + ReLu + BatchNorm (No Padding, Stride 1)

디코딩과정에서는 upsampling하면서 인코딩에서 산출된 feature map의 특성을 연결시켜서 최종적으로 1×1 Convolution Layer 필터를 통과하면서 타겟변수(강수량, 최고/최저 온도)에 맞춰서 3차원 벡터가 생성됩니다. 이러한 설계에 맞춰서 ECMWF, 강수량 값과 GPM 관측자료 강수량 값을 3주차 7일 자료만 예측자료에서 얻고 3개월 치 238 sample 수를 곱하면 1,666개의

데이터가 된다. 따라서 U-Net 입력자료는 (1666, 48, 48, 2)가 되고, 1,666개 중에서 (2001-2012년)인 1,176개의 학습데이터로 이용하고 5년치(2013-2017) 490개를 테스트 데이터로 사용하였다. 강수량 값을 최대 최소값을 정규화하기 위한 MinMaxScaler를 실행하고 학습에 필요한 conv2D, dropout, maxpooling 함수로 인코딩, 디코딩을 학습하고 최종적으로 Conv2D 필터에서 예측자료를 얻게 된다.

```
inputs = Input((48,48,2))
c1 = Conv2D(16, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (s)
c1 = Dropout(0.1) (c1)
c1 = Conv2D(16, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c1)
p1 = MaxPooling2D((2, 2)) (c1)
c2 = Conv2D(32, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (p1)
c2 = Dropout(0.1) (c2)
c2 = Conv2D(32, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c2)
p2 = MaxPooling2D((2, 2)) (c2)
c3 = Conv2D(64, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (p2)
c3 = Dropout(0.2) (c3)
c3 = Conv2D(64, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c3)
p3 = MaxPooling2D((2, 2)) (c3)
c4 = Conv2D(128, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (p3)
c4 = Dropout(0.2) (c4)
c4 = Conv2D(128, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c4)
p4 = MaxPooling2D(pool_size=(2, 2)) (c4)
c5 = Conv2D(256, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (p4)
c5 = Dropout(0.3) (c5)
c5 = Conv2D(256, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c5)
u6 = Conv2DTranspose(128, (2, 2), strides=(2, 2), padding='same') (c5)
u6 = concatenate([u6, c4])
c6 = Conv2D(128, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (u6)
c6 = Dropout(0.2) (c6)
c6 = Conv2D(128, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c6)
u7 = Conv2DTranspose(64, (2, 2), strides=(2, 2), padding='same') (c6)
u7 = concatenate([u7, c3])
```

```

c7 = Conv2D(64, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (u7)
c7 = Dropout(0.2) (c7)
c7 = Conv2D(64, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c7)
u8 = Conv2DTranspose(32, (2, 2), strides=(2, 2), padding='same') (c7)
u8 = concatenate([u8, c2])
c8 = Conv2D(32, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (u8)
c8 = Dropout(0.1) (c8)
c8 = Conv2D(32, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c8)
u9 = Conv2DTranspose(16, (2, 2), strides=(2, 2), padding='same') (c8)
u9 = concatenate([u9, c1], axis=3)
c9 = Conv2D(16, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (u9)
c9 = Dropout(0.1) (c9)
c9 = Conv2D(16, (3, 3), activation='elu', kernel_initializer='he_normal', padding='same') (c9)
outputs = Conv2D(1, (1, 1), ) (c9)
model = Model(inputs=[inputs], outputs=[outputs])

```

U-Net 모델의 연결된 구조를 보면 다음과 같다(Figure 12). U-Net 모델은 입력자료의 구성이 아시아 태평양 지역의 48×48 의 배열과 ECMWF, GPM 강수량 자료만을 이용하여 학습하였다. optimizer 함수는 'adam' 으로 손실함수는 mse (mean square error) 함수로 선택하여 학습을 진행하였다. Batch normalization을 적용하지 않았으며, padding='same', kernel_initializer= ' he_normal' , activation='elu' , dropout의 파라미터를 이용하여 학습하였다. Epochs는 EarlyStopping을 적용하여 정확도가 나오면 멈출수 있도록 셋팅하였으며, 정확도 개선을 위하여 epochs와 batch size를 다양하게 값을 바꿔가며 학습을 진행하였다(Figure 12).

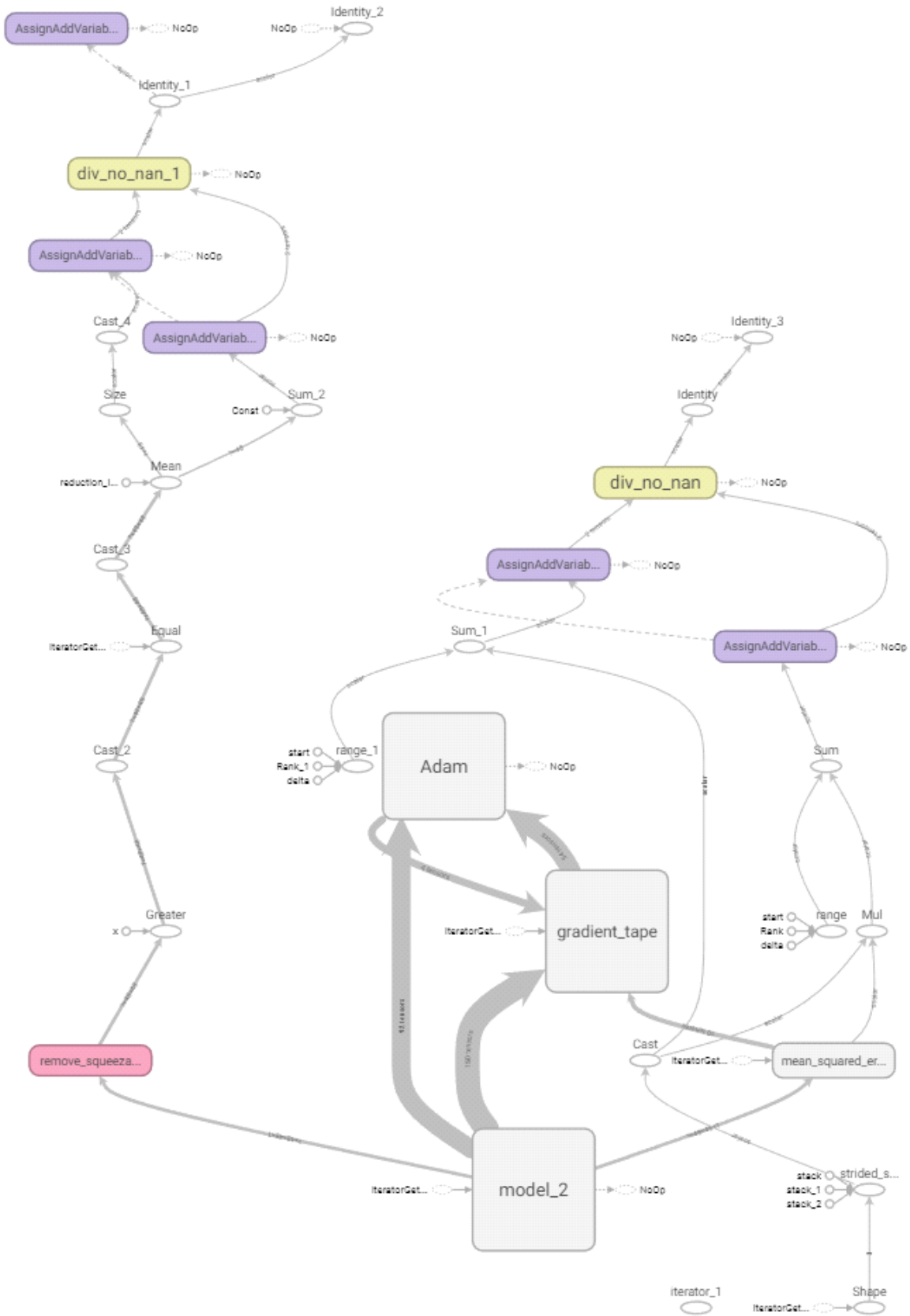


Figure 12. Model pattern Extraction and visualization of U-Net.

(마) 분석 및 검증

U-Net과 ConvLSTM의 입력자료는 3개월의 기간 확장과 계절 내 예측자료 및 관측자료의 대용량의 자료를 합성곱으로 계산하게 되면 학습을 위한 계산속도의 저하로 이어진다. U-Net, ConvLSTM의 딥러닝 모델의 계산시간을 효과적으로 단축하기 위해서는 GPU의 계산자원이 필요하다. Nvidia GTX3090을 기반으로 한 PC와 구글 클라우드에서 무료로 가상서버를 지원하기 때문에 가상서버를 이용하여 학습하였다.

예측의 정확도를 향상하기 위해서 입력자료를 다양하게 구성하였다. 예측자료의 4주간 시계열 자료와 관측자료를 예측일로부터 과거 4주간 자료를 매치시키고 입력자료를 구성하고 또한 예측자료의 3, 4주차의 자료와 관측자료의 예측일로부터 과거 1주차의 자료를 가지고 ConvLSTM과 U-Net의 훈련셋 자료를 구성하여 정확도를 평가하였다.

정확도를 평가하기 위하여 통계적으로 가장 많이 이용하는 RMSE와 PCC를 사용하여 분석하고 비교하였다.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (X_t - \hat{X}_t)^2}{n}} \quad [6]$$

X_t 는 ConvLSTM과 U-Net으로 학습하여 예측된 값과 \hat{X}_t 는 ERA5을 이용하여 RMSE값과 산출하고 계절내 예측모델과 ERA5의 공간패턴을 검증하는 PCC값을 이용하여 정확도 비교 평가를 실시하였다.

$$PCC(t) = \frac{\frac{1}{n} \sum_{y=1}^n (S'_{yt} - \bar{S}_t)(O_{yt} - \bar{O}_t)}{\sqrt{\frac{1}{n} \sum_{y=1}^n (S'_{yt} - \bar{S}_t)^2} \sqrt{\frac{1}{n} \sum_{y=1}^n (O_{yt} - \bar{O}_t)^2}} \quad [7]$$

PCC는 위의 식으로 계산할 수 있으며 S와 O는 각각 U-Net과 ConvLSTM의 학습하여 예측 값과 ERA5의 타겟변수 값을 의미한다.

(바) 입출력 시스템 설계

자료확장과 입력자료를 바탕으로 정확도를 개선하기 위해서 다양한 입력자료와 feature를 선택할 수 있는 입출력 시스템을 개발하여 민감도 분석을 실시하고자 한다. 입출력 시스템은 feature selection 하는 부분과 사용자 중심의 조건을 맞출 수 있는 전처리 부분으로 구성하여 다양한 딥러닝 모델 입력자료를 산출하게끔 설계하였다. 각각의 모듈은 독립적으로 연결이 될 수 있도록 설정하였으며, 이러한 모듈을 사용자가 객체지향적으로 이용할 수 있도록 최종적으로 GUI 환경과 연결하여 GUI 기반 입출력 시스템이 될 수 있도록 설계하였다(Figure 13).

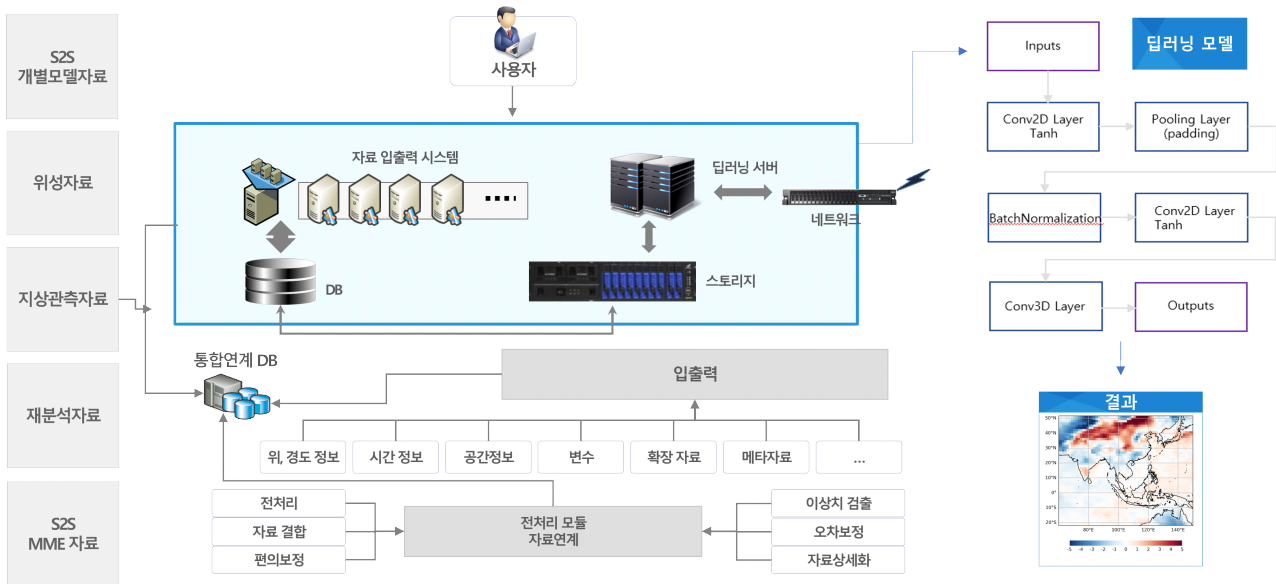


Figure 13. The structure of in and out data system.

(2) 자료 파이프라인 구축: 훈련자료 전처리 기법 테스트

(가) 실험 개요

Figure 14은 이 연구에 대한 실험 개요도이다. 6개 개별 기후모델 및 MME (Multi-Model Ensemble, 이하 MME) 입력 자료에 6개 전처리 및 차원 감소 기법이 적용 및 변환된 훈련자료를 기준 딥러닝 훈련 모델(그림에서 Base model)에 의해 훈련 후 예측성능을 평가하는 전체 과정을 설명한 것이다. Figure 14에서 첫 번째 박스 단계는 6개 개별 기후모델 및 MME 원시 입력 자료를 설명한다. 이 연구의 입력 자료는 6개 개별 기후모델로, European Centre for Medium-Range Weather Forecasts (이하 ECMWF)의 S2S 예측 자료(Vitart and Robertson, 2018; de Andrade et al., 2019; ECMWF, 2020)와 날씨 예보 현업 연구 대상 모델인 Global Seasonal Forecast System의 버전 5(이하 GloSea5)의 S2S (Scaife et al., 2014; MacLachlan et al., 2015) 뿐만 아니라, Meteorological Service of Canada(이하 ECCC), NOAA의 National Centers for Environmental Prediction (이하 NCEP), China Meteorological Administration (이하 CMA) 및 영

국(United Kingdom)의 Met Office (이하 UKMO) 등 6개 개별 기후모델을 이용하였다. 이 6개 개별 기후모델의 원시자료를 모두 결합한 것을 MME라고 명명하였다. 두 번째 박스에서는 입력 자료의 특성(예, 변수 등)의 크기 혹은 분포 등을 계산 및 변형하는 과정으로 원시자료를 딥러닝 훈련 모델에서 인식하기 용이한 훈련자료로 변환하는 단계를 설명한다. 세 번째 박스는 이 연구에서는 기본 U-net을 간략하게 수정한 Base model로부터 변환된 자료를 훈련하는 과정을 설명한다.

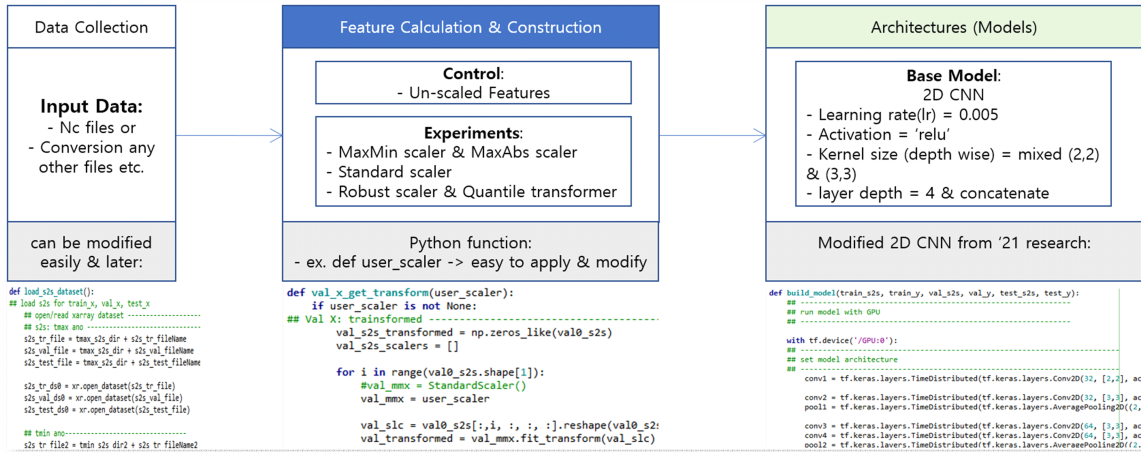


Figure 14. The work-flow of the study.

(나) 전처리 기법

전처리(preprocessing)는 흔히 데이터 전처리(data preprocessing)이다. 데이터를 모델이 이해할 수 있는 형태로 변환하거나 자료의 품질을 올리는 일련의 과정을 일컫는데, feature tuning이라고도 한다. 전처리는 여러 단계로 이루어지는데, 단계별로 처리된 데이터는 3가지 형태로 명명할 수 있다(Figure 15). 첫 번째가 원시자료(raw data)로 초기에 수집된 원본 데이터로 분석이나, 기계학습 혹은 딥러닝 학습 용도로 전혀 전처리가 되지 않은 데이터로 데이터 저장소 등에 저장된 데이터가 이에 해당된다. 두 번째는 정제자료(prepared data)는 Data engineering, 즉 전처리 기법에 따라서 학습을 위한 데이터로만 추출한 데이터(예, subset)를 의미한다. 예를 들면, 앞서 언급한 바와 같이 S2S 예측 자료는 다루기 어려운데, 이것을 (시간, 위도, 경도, 변수) 등으로 바꾸고, 모형 자료라서 -값이나 0이 없겠지만 nan 데이터가 있는지도 확인해서 준비된 데이터를 훈련자료라고 할 수 있다. 이 연구에서는 원시자료가 정제자료까지의 단계를 말한다. 세 번째가 훈련 자료(예, engineered feature 혹은 engineered data)이다. 정제된 데이터가 딥러닝 학습에 적절한 형태로 재가공 되어야 하는데, 이를 Feature Engineering 이라고 한다. 예를 들어, 숫자와 같은 값을 0에서 1 사이로 변환시키거나 범주형 자료로 예를 들어 남자 혹은 여자를 0 혹은 1과 같은 값으로 변환시키고, 전체 데이터를 학습(training) 및 평가용(validation or test)으로 7:3 분할하여(split) 저장하는 등의 과정이 이에 해당된다.

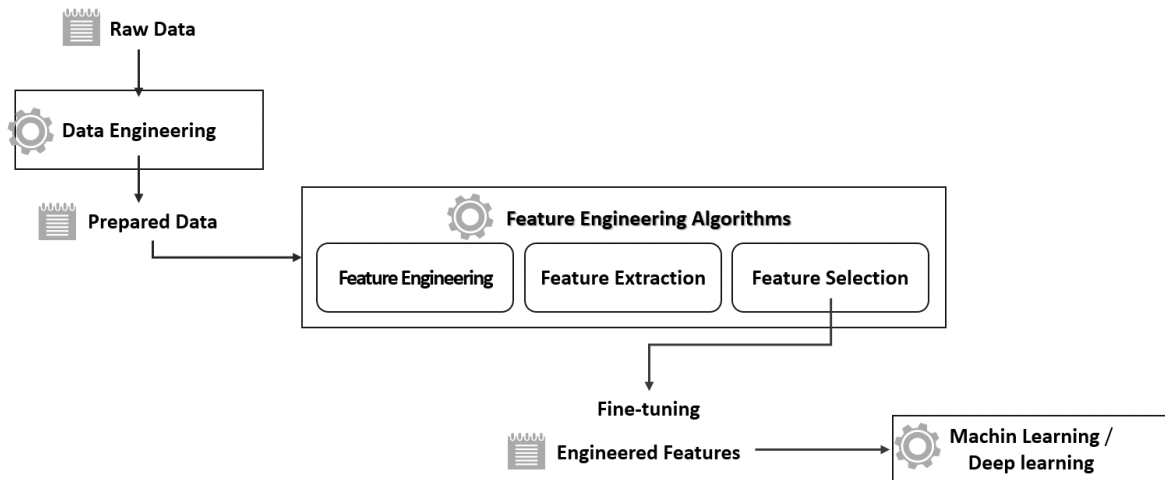


Figure 15. The explanation of data pre-processing step and the transformed data at each step.

전처리 종류는 데이터 과학에서 여러 단계로 분류되는데, 이 연구에서는 크게 데이터 클리닝(data cleaning), 데이터 통합(data integration), 데이터 변환(transformation), 데이터 축소(reduction)로 크게 3가지 종류를 기준으로 삼았다. 이 중 이 연구에서는 데이터 변환 단계를 전처리(이하 Scaler)라고 명명하였다. 또한 어떤 책에서는 scaler를 데이터 변환기(transformer)라고 하는 예도 있다.

다양한 Scaler를 적용한 예측 성능 평가는 숫자형 데이터를 모두 학습시키면 성능이 좋은 모델이 될 것인가? 라는 가설에서 이 연구는 시작되었다. 일반적인 기계학습 혹은 딥러닝 알고리즘들은 매우 다양한 범위의 숫자형 데이터를 학습시키면 제대로 성능을 보여주지 못하는 경우가 많다. 예를 들어, 특정 데이터의 범위가 -500에서 39,320이라면 매우 다양한 데이터가 존재하게 된다. 즉 데이터의 범위가 매우 크게 된다. 이러한 상태에서는 제대로 학습할 수 없어서 이것을 방지하기 위해서 숫자형 데이터의 범위를 줄여주는 방법을 사용하게 된다. 이것을 데이터 이산화(discretize), 표준화(standardize), 정규화(normalize) 등이라고도 한다(Figure 16). 정규화 기법에는 0과 1 혹은 -1에서 1 사이로 데이터의 범위를 변형하는 MinMax 와 MaxAbs 기법이 있다. 또한 표준화에는 평균을 빼고 표준편차로 나누는 Standard 기법이 있다. 데이터 변형에서 먼저 데이터의 유형이 수치 혹은 범주형이냐에 따라 달라진다. 이 연구의 데이터 유형은 수치형으로 세 가지 기법 중에서 규모 변형 및 분포 변형을 선택하였다. 규모 변형에는 정규화와 표준화 중 표준화를 선택하고 사분범위도 선택하였다. 분포 변형에서는 역시 분위로 결정하는 알고리즘을 선택하였다.

이 연구에서는 Scaler를 적용하지 않은 것을 대조군으로 Unscaled라고 명명하고, 최고 및 최저 차이를 계산하는 MaxMin Scaler 및 MaxAbs Scaler, 표준편차를 계산하는 Standard Scaler, 사분 범위를 계산하는 Robust Scaler와 Scaler는 아니지만 Quantile transformer 기법 등 총 6개 기법으로 입력 자료를 표준화하여 이에 따른 예측성능 변화를 평가하였으며, 입력 변수의 차원 감소를 위해 Feature Selection 기법을 통해 목표 예측 변수에 영향을 주는 주요 변수의 변화

에 따른 예측성능 변화를 평가하였다. 이 일련의 과정은 파이썬 모듈로 작성되었다.

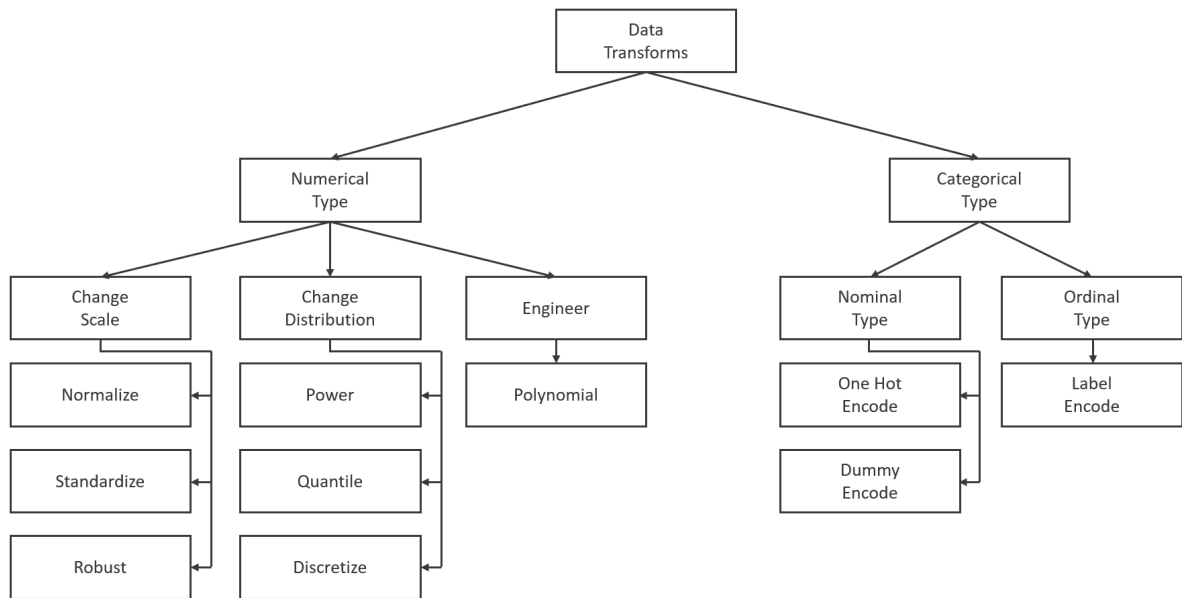


Figure 16. Overview of data transforms (source: MachineLearningMastery.com).

① 최고 및 최저 차이 알고리즘

MinMax scaler는 Eq. [8]로 표현되며, 정규화 기법 중에서 가장 많이 사용되는 Scaler로 최솟값은 0으로, 최댓값은 1로 변환, 나머지는 0과 1 사이의 값에서 변환된다. MaxAbs scaler는 특성(예, feature)를 maximum absolute 값으로 변환한다(Eq. [9]).

$$\frac{X_i - \min(x)}{\max(x) - \min(x)} \quad [8]$$

$$\frac{x}{\max(|x|)} \quad [9]$$

② 표준편차 알고리즘

Standard Scaler는 평균을 제거하고 분산에 맞게 변환된다(예, 표준정규분포, Eq. [10]).

$$\frac{x - \mu}{\sigma} \quad [10]$$

여기서, μ (뮤)는 훈련자료의 평균값이고, σ (시그마)는 훈련자료 표준편차이다.

③ 사분범위 알고리즘

Robust Scaler는 사분범위(IQR, interquartile range)로 변환하는데, IQR은 1사 분위(25%)와 3사 분위(75%)값 사이를 말하며, Eq. [11]에 의해 계산된다. Quantile transformer는 지정된 분위수에 맞게 변환되는데, 일반적으로 가우시안 분포와 유사하게 변환된다.

$$\frac{x_i - x_{med}}{x_{75} - x_{25}} \quad [11]$$

(다) 차원 감소 적용

① Feature Selection 기법

딥러닝 혹은 기계학습 훈련 모델에 제공하는 입력 변수를 특성(feature)이라고 하는데, 데이터 집합(set)에는 어떤 모델을 훈련 및 예측하는데 전혀 관련이 없는 잉여 데이터 혹은 잡음(noise) 등을 포함하고 있을 수 있다. 따라서, 최적의 모델을 훈련 및 예측하기 위한 필수 특성만 사용하고 있는지 확인할 필요가 있다. 왜냐면 특성이 너무 많으면 훈련 모델이 중요하지 않은 패턴을 갈무리하게 되므로 잡음으로부터 학습할 수도 있기 때문이다. 그렇게 되면, 즉, 데이터가 많으면 훈련 프로세스 및 학습 모델이 느려질 수 있고, 이렇게 관련 없는 데이터(셋)로부터 학습 알고리즘이 산만하게 되거나 학습이 되지 않아서 모델의 예측성이 부정확, 즉 예측 성능이 저하될 수 있다.

데이터 집합에서 관련 없는 데이터를 어떻게 처리하느냐가 중요하며 이런 기술을 크게 피쳐 엔지니어링(특성 공학) 기술이라고 한다. 피쳐 엔지니어링(feature engineering)은 모델의 성능을 높이고 경쟁(competition)에서 좋은 성적을 얻기 위해서는 반드시 필요한 기술 중 하나이다. 이와 같이 관련 없는 데이터를 걸러내는 기법 중에 Feature Selection(특성 선택) 혹은 Feature Extraction(특성 추출)이 있는데, 변수를 선택하여 사용한다는 점에서 비슷해 보이지만 이 둘은 다른 기법이다(Figure 17). 특성 추출이 새로운 차원의 공간으로 데이터를 투영하는데, 변수들이 가지고 있던 고유의 성질을 잃어버릴 수 있다는 단점이 있는 동시에 변수들을 조합하여 새로운 잠재 변수들을 통해서 데이터를 잘 설명할 수 있는 표현형을 만들 수 있다는 장점이 있다. 반면, 특성 선택은 사용자가 더 해석하기 쉽게 모델을 단순화할 수 있고 훈련 시간을 단축할 수 있으며 과적합(over-fitting)을 줄여 좀 더 일반화(generalization)한다는 장점들이 있다. 요약하면, 특성 추출이 데이터를 어떻게 유용하게 만들 것인가의 문제라면 특성 선택은 좀 더 구체적으로 데이터에서 유용한 피쳐를 어떻게 선택할 것인가에 더 초점이 맞춰져 있다.

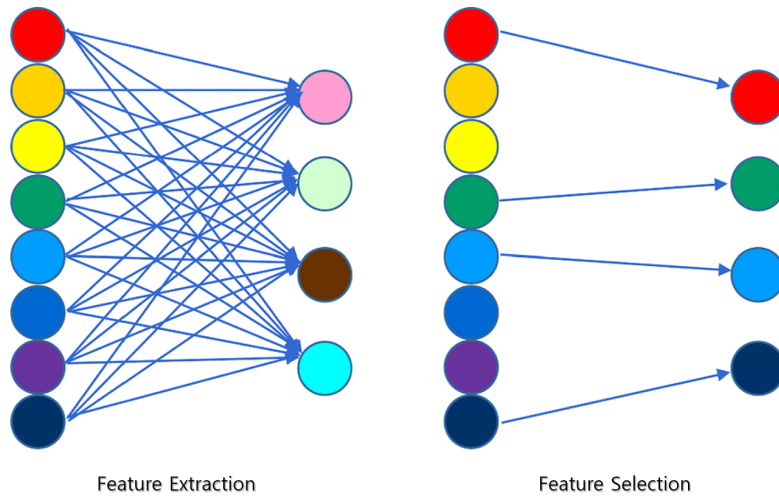


Figure 17. Explain about the difference for feature extraction (left) and feature selection (right).

이 연구에서는 특성 선택 기법으로 S2S 새 훈련자료를 구성하고 이에 따른 예측 성능 변화를 평가하였다. 특성 선택(feature selection)에서 피처를 선택하는 것은 어떻게 보면 매우 직관적인 아이디어이다. 즉, 원시자료에서의 피처를 모두 사용하기에는 computing power와 memory 측면에서 매우 비효율적이기에 일부 필요한 피처만 선택한다는 아이디어에서 출발한 것이다. 간단히 말하면, 모델을 구성하기 위한 중요 피처(변수)를 선택하는 과정으로, 가장 밀접하게 연관되어있는 특성만을 사용하게 됨으로써 모델 성능의 향상을 기대할 수 있으며, 또한 처리할 데이터의 양이 줄어들므로 처리 속도 또한 빨라지게 된다.

특성 선택 알고리즘에는 크게 세 가지가 있다. 이 알고리즘 중에서 어떤 특성 선택 알고리즘을 사용할지는 우리가 사용하는 입력 및 출력 자료의 유형에 달려 있다. 즉, 숫자형 혹은 범주형이냐에 따라 특성 선택 알고리즘이 달라진다. Figure 18와 같이 특성 선택 알고리즘은 filter methods, wrapper methods 및 Embedded Methods로 분류된다.

Filter Methods는 관련성을 찾는 방법으로, 통계적 측정 방법을 사용해서 피처들의 상관관계를 알아내는 방법이다. 하지만, 피처 간의 상관계수가 반드시 모델에 적합한 피처라고는 할 수 없고, 데이터 집합(셋)의 조정이 정확하지 않다는 단점이 있다. 대신 계산 속도가 빠르고 피처 간 상관관계를 알아내는 데 적합하기 때문에 Wrapper method를 사용하기 전 전처리하는 것에 사용할 수도 있다. Filter Feature Selection 방법은 모델에 적용하기 전에 피처들을 걸러내는 방법으로 전처리 과정이라 볼 수 있다. 통계적인 계산에 의한 영향력을 바탕으로 각 Feature 들을 랭킹화한 뒤 랭킹에 따라 데이터 집합에서 그대로 남아있거나 제거되게 된다. 피어슨 상관분석, 카이제곱검정 등을 사용하여 상관관계를 파악한다. 특히, filter methods는 전처리 단계에서 통계적 기법을 통해서 주요 변수를 선택하는 방법으로 다소 용이한 방법이지만 필수적이면서 중요한 방법이다.

이 세 가지 방법 중 어떤 것을 사용하든 주의점이 있는데, 훈련자료에서 피처를 고른다면, 훈련자료에 과적합 될 수 있다. 그러므로 훈련자료, 테스트 자료를 제외한 데이터에서 선택하는 것이 중요하다. 모든 데이터에서 feature selection을 진행하면, 교차 검증에서 똑같이 선택된 feature를 사용하게 되므로 결과가 편향될 수 있기 때문이다.

또한, 어떤 특성 선택이 이 연구에서 채택한 모델에 적합한지는 사용할 자료 형태, 즉 입력 및 출력 변수의 유형이 무엇이냐에 따라 다르다. 간단히 말하면, 변수의 유형은 두 가지로 나뉜다. 숫자(수치형, numeric) 변수 혹은 범주형(catagory) 변수인데, 이 연구에서 이용할 S2S 예측 자료는 숫자형이기 때문에 상관관계를 계산하는 filter method 알고리즘을 사용하였으며, 이 모든 과정은 파이썬 모듈로 작성되었다.

② 주요 특성 선택 기반 훈련 자료 구성

Figure 18은 특성 선택 알고리즘을 설명한 것이다. 특성 선택 알고리즘은 먼저 지도학습과 비지도학습으로 나뉜다. 지도학습은 다시 자동으로 피처를 선택하는 방법 혹은 best model에서 선택된 것을 뽑는 방법 및 점수화해서 순위로부터 피처를 선택하는 방법 등이 있다. 이때 점수화하여 순위를 결정하는 알고리즘으로는 통계적 기법에서 상관관계 계산 등으로 점수화하는 방법 등이 있다. 통계 기법은 다시 입력 및 출력 데이터의 유형에 따라 달라진다.

특성 선택 및 특성 추출은 차원 감소 알고리즘과 연관된다(Figure 18). 데이터 집합(셋)에 대한 입력 특성의 수는 데이터의 차원으로 간주할 수 있기 때문이다. 데이터의 차원 감소의 대표적인 접근 기법은 행렬 분석 기술(matrix factorization technique)인데, 이 행렬 분석 기법의 대표적인 방법이 또한 주성분 분석 (PCA: Principal Component Analysis)이다. 이것은 특성 선택의 대안으로 원시자료의 가장 중요한 속성을 여전히 유지하는 저차원 공간에 데이터를 투영하는 것이지만 특성 선택과 다른 점은 원시자료와 직접적인 관련이 없는 데이터가 투영될 때, 예를 들면 종속성, 상관이 있는 변수들이 제거될 때 해석하기 어렵다는 단점이 있다.

이 연구에서는 파이썬에서 제공하는 SelectKBest module로부터 상관관계를 계산하여 가장 점수가 높은 변수에 대해 순위를 결정하였다. 구축된 S2S 훈련자료는 대체로 16개 기후 변수로 구성되었는데, 이 중 예측을 위한 목표 변수와의 상관 순위를 8위까지로 설정해서 8개 주요 변수로 이루어진 새로운 훈련 데이터 집합을 기준 훈련 자료(예, 변수 8개 사용 그룹)로 만들고, 이를 기준으로 랜덤으로 변수를 +/- 적용하여 3개의 그룹(예, 변수 10개 이상 사용 그룹, 변수 9개 사용 그룹, 변수 7개 이하 사용 그룹)을 더 추가로 구성해서 변수 조합에 따른 예측성능 변화를 평가하였다.

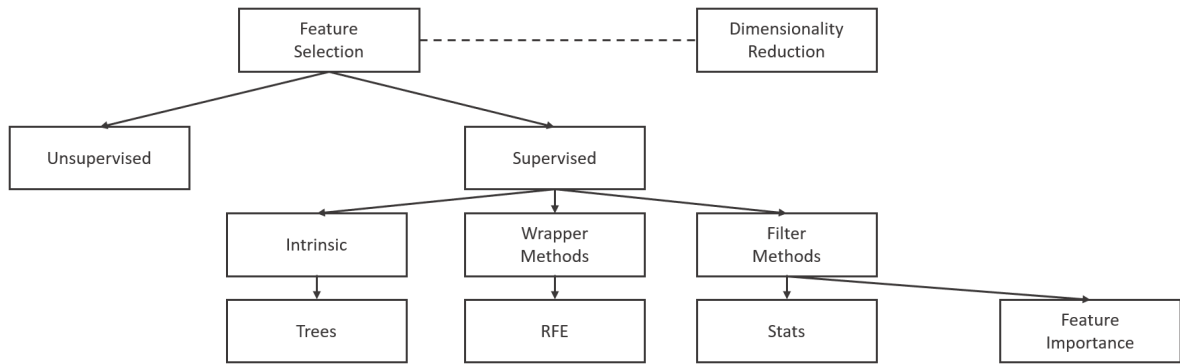


Figure 18. Overview of feature selection (source: MachineLearningMastery.com).

(라) MME 방법

① S2S 예측 자료

이 연구에서는 예측성이 가장 우수하다고 평가되고 있는 European Centre for Medium-Range Weather Forecasts(이하 ECMWF)의 S2S 예측 자료(Vitart and Robertson, 2018; de Andrade et al., 2019; ECMWF, 2020)와 날씨 예보 현업 연구 대상 모델인 Global Seasonal Forecast System의 버전 5(이하 GloSea5)의 S2S(Scaife et al., 2014; MacLachlan et al., 2015)뿐만 아니라, Meteorological Service of Canada(이하 ECCC), NOAA의 National Centers for Environmental Prediction(이하 NCEP), China Meteorological Administration(이하 CMA) 및 영국(United Kingdom)의 Met Office(이하 UKMO) 등 6개 개별 기후모델을 이용하였다.

개별 기후모델의 기후 변수는 TMAX 및 TMIN(Maximum and minimum air temperature at 2m), T2M(2 meter air temperature), MSLP(Mean Sea Level Pressure), PREC(Precipitation), SST(Sea surface temperature), OLR(Top net thermal radiation), U50, U200, U850, V200, V850(U and V component of wind at 50, 200 and 850hPa), W500(Vertical velocity at 500hPa), Q850(Specific humidity at 850hPa), Z200 및 Z500(Geopotential height at 200 and 500hPa) 등 약 16개이다. 이 중 UKMO는 MSLP을, KMA는 SST, OLR, Q700, Q850을 이용할 수 없다. S2S 예측 기후 변수의 시간 해상도는 TMAX 및 TMIN(6시간 간격의 시간 자료)을 제외하고 일별이며(daily), 공간 해상도는 1.5도(적도 부근에서 약 100km)의 전지구 규모로 위도 121개 및 경도 240의 격자 배열을 갖는다. 또한 앙상블 수, 과거 재현 예측(이하 Hindcast) 자료의 생산 방법, 선행예측시간(lead-time), Hindcast의 기간, 예측 자료의 생산 주기 등이 모두 다르다.

S2S 예측 자료의 앙상블은 기후모델의 기후 변수가 증가하면 (그) 앙상블도 증가되는데, 이것은 합성곱신경망(Convolutional Neural Network, 이하 CNN) 네트워크의 속도 및 컴퓨팅 계산의 문제가 될 수 있어 개별 기후모델의 기후 변수의 앙상블은 control과 perturbation에 상관 없이 평균하였다. 예를 들면, ECMWF 기후모델의 경우 control에 대한 17개 기후 변수와 perturbed에 대한 17개 기후 변수는 Hindcast에 대해 11개 앙상블로 총 187개(11 ensembles x 17 variables)로 구성될 뿐만 아니라, 이것이 전체 도메인 격자(121 x 240)뿐만 아니라 기간 및 주기(20년 x 46일 x 52주)로 구성되기 때문에, read/open 및 속도에 어려움이 발생하였다. 따라

서 입출력 및 속도 문제 해결을 위해 (control 및 perturbation에 상관없이) 17개 변수별로 해당 앙상블 11개를 평균하였다. 마찬가지로 GloSea5, UKMO, CMA, NCEP, ECCC 기후 모델도 각각의 기후 변수별로 앙상블을 평균하여 처리하였다.

② MME 자료

이 연구에서 6개 개별 기후모델의 S2S 예측 자료를 모두 결합하여 MME를 생산하였다. 6개 개별 기후모델의 Hindcast 기간 전체가 포함되는 범위는 1999년에서 2010년인데, 12년에 불과하므로 이 기간을 MME 가능 기간으로 설정하기에는 너무 부족하였다. 6개 개별 기후모델의 Hindcast의 모든 기간이 완벽하게 포함되지 않더라도 Hindcast 기간을 최대 20년에 맞추기 위해, NCEP 및 GloSea5 기후모델의 앞뒤로 4년을 포함하지 않더라도 4개 개별 기후모델이 모두 배치될 수 있는 기간으로 1995년에서 2014년을 선정하였다. MME 시작일은 ECMWF와 ECCC의 시작일 중 목요일을 기준으로 설정하였다. 예를 들어, 2018년 1월 4일은 목요일로 ECMWF와 ECCC의 시작일인데, 시작일이 매일, 즉 일별인 CMA와 NCEP은 1월 4일과 결합이 될 수 있지만 GloSea5와 UKMO는 1월 1일이 시작일로 앞에 3일 패스하고 1월 4일과 결합하게 된다. 선행 예측시간(lead-time)도 각각 다른데, MME의 선행예측시간은 30일로 설정하였다. 또한, S2S 예측 자료들은 학습에 활용하기 쉽게 다차원 배열(예, 시간, 선행예측시간, 위도, 경도, 변수로 구성된 5차원 배열) 데이터 구조로 구성하였다.

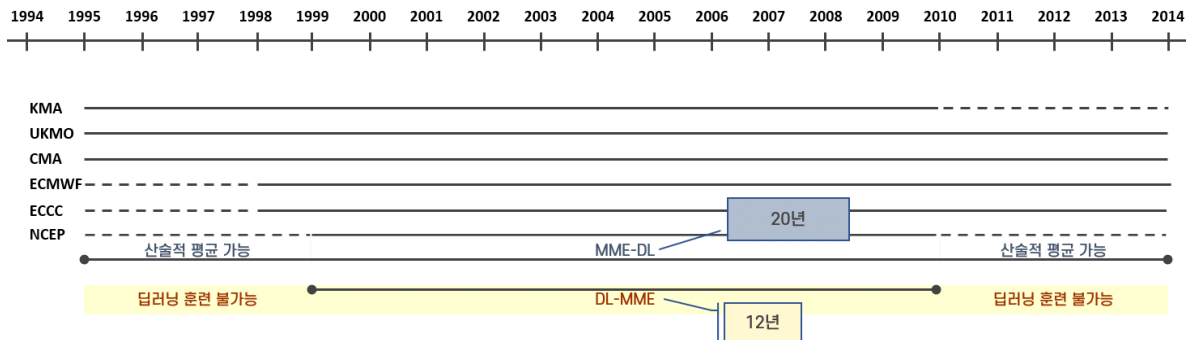


Figure 19. Description of the total training periods for six individual climate models and MME possible periods.

(마) 적용 모델

합성곱 장단기기억(Convolutional Long-Short Term Memory, 이하 ConvLSTM)과 U-Net을 혼합한 모델 연구 등이 많이 있는데(Agrawal et al., 2019; Ayzel et al., 2020; 신흥준 등, 2021), 이 연구에서는 선례 연구에서 작성된 기본 U-Net에 간단한 방법으로 U-Net을 수정하여 S2S 예측 최고 및 최저기온, 강수의 예측성 개선을 시도했다. U-Net은 이미지의 전반적인 특성 정보를 얻기 위한 수축 형태(Contracting Path)의 네트워크와 정확한 지역화를 위한 팽창 형태(Expanding Path)의 네트워크를 U자형의 대칭 형태로 갖는데(Ronneberger et al., 2015; Weyn

et al., 2021), 이 연구에서는 TimeDistributed 레이어를 U-Net의 Convolution 레이어에 추가하여 시계열 효과를 주고자 했다(Figure 20). TimeDistributed는 차원 변경 없이 간단하게 2D convolution을 사용할 수 있기 때문에 5차원 배열의 훈련자료가 바로 입력된다.

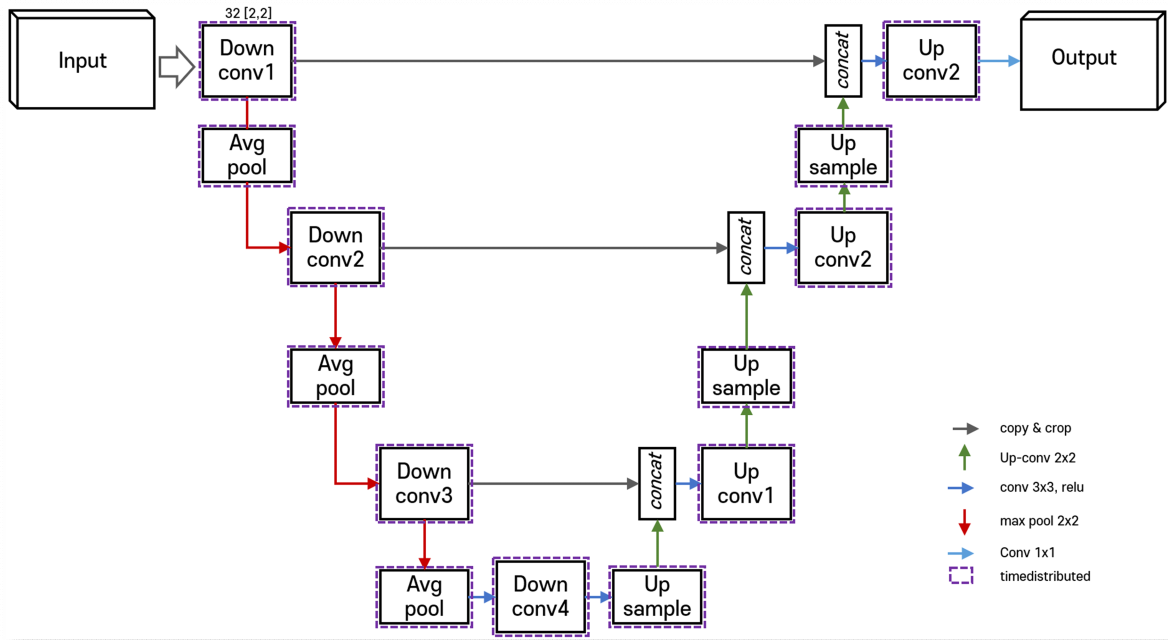


Figure 20. Architecture of U-Net wrapped by the TimeDistributed layer in the study.

(바) 훈련 환경 및 도메인 설정

훈련 기간과 테스트 기간은 개별 기후모델의 Hindcast 기간 및 MME 기간에 따라 각각 달라지는데 Hindcast 기간 전체에 대해 80%는 훈련 기간, 20%는 테스트 기간으로 계산된다. 예를 들면, MME의 Hindcast 기간이 1995년부터 2014년이므로 1995년부터 2010년까지 16년(80%)은 훈련 기간으로, 2011년부터 2014년까지 4년(20%)은 테스트 기간이 된다.

CNN 네트워크 규모 등을 고려할 때 교차 검증(k-fold cross validation)은 시간이 많이 걸리는데, 이 연구에서는 이런 점을 고려하여 3-fold 교차 검증을 수행했으며, 훈련자료 구축은 전지구 규모이지만, 이번 연구에서 속도 및 컴퓨팅 계산을 고려하여 딥러닝 훈련 모델에 입력될 때의 입력 및 출력 도메인은 경도 52에서 160, 위도 -50에서 58의 1.5도로 y축(경도)으로 72개, x축(위도)으로 72의 격자를 가진 동아시아 지역으로 설정하였다.

딥러닝 훈련 모델은 케라스(Keras) 및 텐서플로우(TensorFlow) 프레임워크를 사용하였으며, CNN의 최적화를 위한 손실함수(loss function)는 평균절대오차(Mean Square Error, MSE)로 계산되었으며, 측정함수(metrics)로는 일 최고 및 최저기온과 강수에 대한 오차와 정확도 평가를 위해 코사인 유사도 및 평균평방근오차(Root Mean Square Error, RMSE)를 계산하여 학습하였다. 예를 들면, 예측 일 강수량이 생성되면 예측 시점의 관측 강수량과의 편차를 나타내는

손실함수가 계산된다. 일 최고 및 최저기온 역시 동일한 과정으로 예측된다. 손실함수를 최소화하면서 적용 모델의 매개변수 갱신을 위한 optimizer는 최근 가장 성능이 좋다고 알려진 Adam optimizer를 사용하였으며, 학습율(learning rate)은 0.005로 설정한다. 반복 학습 횟수(epoch)는 50, 1회 학습, 즉 1 epoch일 때 크기(batch size)는 80으로 설정했다.

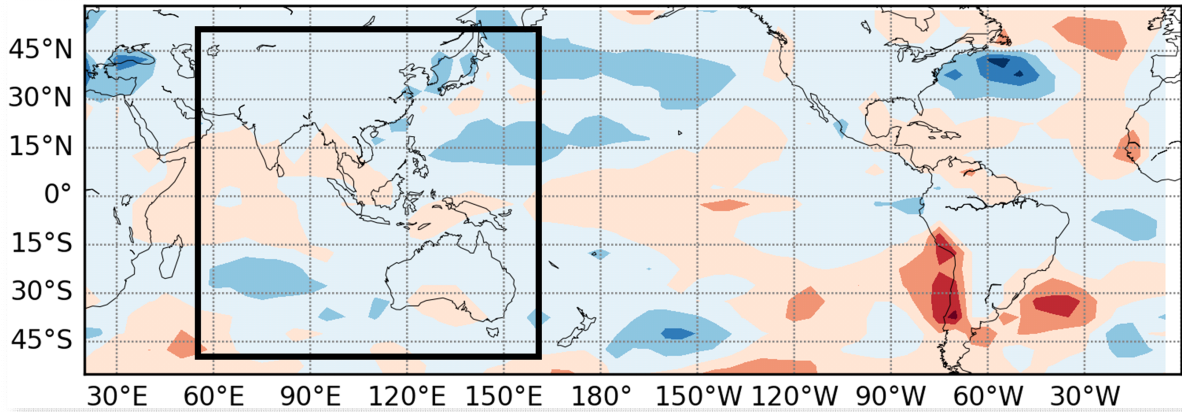


Figure 21. The domain of the constructed training data of Sub-seasonal to seasonal. The black box indicates the input and output domains of the deep learning training model in this study.

(사) 평가 방법

① 평가 지수

훈련 모델에서 예측된 개별 기후모델 및 MME 기반 일 강수의 공간적 예측 정확성 및 변동성을 평가를 위해 이 연구에서는 선행예측시간에 대한 편차공간상관계수(anomaly spatial Pattern Correlation Coefficient, 이하 PCC)를 계산하였다(Eq. [12]).

$$PCC_i = \frac{\sum_{i=1}^N \Delta x_{i,j} \Delta y_{i,j}}{\sqrt{\sum_{i=1}^N \Delta x_{i,j}^2} \sqrt{\sum_{i=1}^N \Delta y_{i,j}^2}} \quad [12]$$

여기서, $\Delta x_{i,j}$ 는 격자 i 에서 i 일째의 훈련 전 혹은 학습 후 각각의 개별 기후모델 및 MME 기반 예측 기후자료의 아노말리이며, $\Delta y_{i,j}$ 는 i 는 임의의 격자이고, M 은 전체 격자수이다. j 는 선행예측시간(lead-time), N 은 예측 빈도 전체 개수를 각각 나타낸다.

개선도(improvement, 혹은 skill score)는 이 연구에서 활용한 각각의 평가 지수에 대한 상

대적 변화를 쉽게 이해하고 설명하기 위해 계산하였다(Eq. [13]). 예를 들면, 딥러닝 훈련 모델에서 학습하지 않은 원래 S2S의 예측성능 대비, 훈련 모델에서 학습하여 보정된 S2S의 예측성능을 평가한 지수(예, PCC 및 RMSE 등)의 변화량을 계산한 것이다.

$$\text{Improvement} = \frac{PCC_{DL} - PCC_{None}}{PCC_{None}} \quad [13]$$

여기서, PCC_{DL} 는 딥러닝 훈련 모델에 의해 예측된 목표 기후 변수(최고 및 최저기온, 강수) 예측 결과를 평가한 지수(예, 훈련 후 평가된 PCC)이고, PCC_{None} 은 입력 전, 즉 훈련 모델에서 학습되지 않은 S2S 예측 강수의 원시 훈련자료에 대해 평가한 지수(예, 훈련 전 평가된 PCC)이다.

② 가중치 시각화

모델이 예측한 결과에 대한 이유를 이해하는 것은 매우 중요하며, 모델이 올바른 기능을 기반으로 결정을 내리도록 진단하는 일이 필요하다. 예를 들면, 이미지의 손상된 부분을 잘 예측하는 CNN을 구축했다고 했을 때, 예측과 상관관계가 있는 정보가 아닌 손상된 표면을 기반으로 예측했는지 확신할 수 없다. 그래서 제안된 유용한 방법이 saliency map(돌출 맵)이다. 모델에서 예측된 이미지에서 예측에 가장 잘 기여한 픽셀을 시각화한 것으로 (Simonyan et al., 2014; Chowdhury et al., 2020), 전체 이미지 픽셀값을 미세하게 조정하여 예측된 클래스(class)의 변화를 계산함으로써 모델의 궁극적인 예측에 대한 각 픽셀의 상대적 중요도를 측정한다. 이 연구에서 saliency map을 계산하기 위해 케라스(Keras)와 텐서플로우(TensorFlow)를 사용하였으며, 최근에는 keras-vis의 Saliency 라이브러리(명령어)를 바로 적용해서 saliency map을 계산할 수 있다.

다. 준지도학습 적용을 위한 환경구축

(1) 자료 수집 및 전처리

(가) 일별·월별 관측 및 재분석 자료 수집

우리나라 계절예측을 위한 인공지능모델의 입력 변수로 강수, 해수면온도, 하층기온, 지위 고도, 해면기압의 다섯 개의 변수를 이용하였다. 일별 강수는 CPC Global Precipitation 자료를 이용하였으며(ftp://ftp.cdc.noaa.gov/Datasets/cpc_global_precip에서 다운로드, 육상 강수만 포함) 월별 강수는 CMAP 강수 자료를 이용하였다(<ftp://ftp.cpc.ncep.noaa.gov/precip/cmap/>). 해수면온도는 NOAA OISST V2 High Resolution Dataset (<ftp://ftp.cdc.noaa.gov/Datasets/noaa.oisst.v2.highres>)을 이용하였으며 하층기온(850hPa), 지위고도(500hPa), 해면기압 일별 자료는 NCEP/DOE Reanalysis II (NCEP2) 자료를 이용하였다(<ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis2.derived/> 및 <ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis2.dailyavgs/>).

(나) 관측 자료 부재한 시기를 위한 예측 자료 수집

관측 자료가 부재한 기간에 대해서는 역학 기후모델의 예측자료를 활용하였는데, 관측 및 재분석 자료와 동일하게 강수, 해수면온도, 하층기온(850hPa), 지위고도(500hPa), 해면기압 다섯 개의 변수에 대해 APCC MME 자료를 수집하여 이용하였다.

(다) 공간 및 시계열 전처리

① 공간 전처리

입력 자료를 이미지기반 인공지능모델과 그래프기반 인공지능모델에 활용하기 위해 두 가지 공간해상도로 전처리하였다. 이미지기반 모델의 입력으로 사용하기 위해서 모든 자료를 좌상단(upper-left) 격자의 중앙점이 위경도 (90, 0)에서 시작하는 $2.5^{\circ} \times 2.5^{\circ}$ 해상도로 리샘플(resample, 원자료의 공간해상도에 따라 해수면온도와 일별 강수는 average, 월별 강수는 bilinear 내삽)하고 공통 영역인 72×144 격자로 잘라내었다. 그래프기반 모델의 입력으로 사용하기 위해서는 모든 자료를 좌상단(upper-left) 격자의 좌상단 모서리 점이 위경도 (90, 0)에서 시작하는 $15^{\circ} \times 15^{\circ}$ 해상도로 리샘플(resample, 모든 변수 average 내삽)하고 공통 영역인 12×24 격자로 잘라내었다.

② 시계열 전처리

모든 자료를 각 격자에 대하여 표준화 및 정규화 시계열 전처리하였다. 월별 자료는 월별, 격자별로 표준화 및 정규화하였으며 일별 자료는 자료강화 적용 시에 일별 자료 그대로 사용하지 않고 상반기, 하반기, 전월을 사용하므로 각각에 대해 기간별, 격자별로 표준화 및 정규화

하였다. 예측인자 값들은 월 간에(Month by Month) 서로 독립적이라 가정하였으며 전처리를 통해 각 월이나 기간에 대한 예측인자의 공간 분포를 연도 연간(intra-annual) 혹은 계절내(intra-seasonal) 예측인자 값들의 상대적인 차이는 고려하지 않도록 월별 또는 기간별로 전처리하였다. 또한 기후분석을 위해 평년 기후 값을 뺀 편차(Anomaly)를 보는 것처럼 위치별로 예측인자의 전체 기간에 대한 상대적인 값을 구하고 그 공간 분포를 보기 위해 격자별로 전처리하였는데, 후속 연구를 통해 격자별로 전처리하는 경우와 전체 영역의 모든 격자에 대해 전처리하는 경우의 결과를 비교할 예정이다.

표준화는 기준기간(1991-2020)에 대해서 평균과 표준편차를 도출하고 이를 이용하여 값을 산정하였다. 기준기간 외의 자료가 기준기간 자료 범위 밖의 값을 가지는 경우 이상치를 가질 수 있으며 기준기간 내의 자료일지라도 빈도가 낮은 극한값을 가지는 경우 전처리 후의 값이 매우 크거나 작아질 수 있으므로 [-3.09, 3.09]의 제한 범위를 두었다.

$$\text{Standardized Value} = \frac{x - \text{Mean}}{\text{Stdev}} \quad [14]$$

정규화는 기준기간(1991-2020)에 대해서 최댓값과 최솟값을 도출하고 이를 이용하여 값을 산정하였다. 자료의 제한 범위를 주지 않았을 때 자료 검증 과정에서 일부 격자에서 이상치가 발생하는 것을 관찰하였으므로(2021년 1월 CPC Global 강수 자료의 경우 아프리카 대륙 사막 지역에서 정규화한 값이 6842.8에 이르는 이상치 발생함 - 강수가 거의 없는 지역이나 소량의 강수가 발생했던 경우가 있는데 이 이상의 강수가 2021년에 발생하여 정규화한 값이 매우 커짐) 전처리 후의 자료에 대해 [0, 1]의 제한 범위를 두었다.

$$\text{Normalized Value} = \frac{x - \text{Minimum}}{\text{Maximum} - \text{Minimum}} \quad [15]$$

(2) 기상기후자료 자료강화

일자료 시간적 조합(semi-monthly)을 활용한 자료강화와 월자료 공간적 재조합(Cutmix 응용)을 활용한 자료강화 두 가지 기법을 적용하여 예측모델 학습개선 여부를 평가할 수 있도록 하였다.

(가) 일자료 시간적 조합(semi-monthly)을 활용한 자료강화

월별 자료는 해당 기간의 일별 변동성과 그 영향을 반영하기 어려우므로 이를 일부나마 반영하고자 일별 자료를 전월 상반기(1-15일), 하반기(16-말일)에 대해 합산(강수) 또는 평균하

여 활용하였다. 입력자료의 7개월 기간(6개월 선행-동시)에 대하여 관측자료가 존재하는 기간은 전월, 상반기, 하반기 중 선택할 수 있으며 관측자료가 존재하지 않는 기간은 APCC MME 월별 자료만을 사용하므로 전월 자료만을 사용하였다.

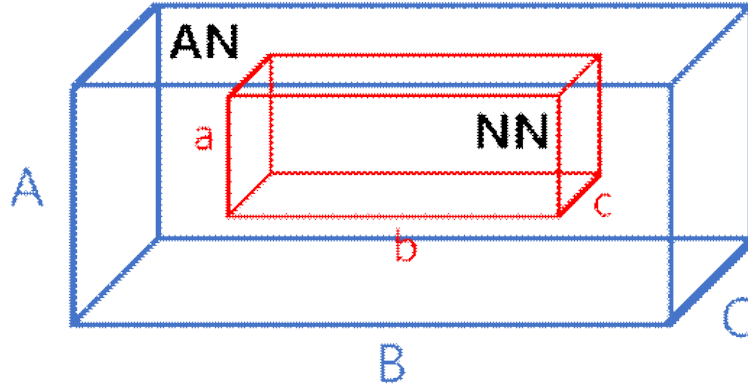
예측 선행시간에 따라 관측자료가 존재하는 기간이 달라지므로, 아래와 같이 관측과 예측의 기간이 결정되었다:

- 1개월 선행예측(LT1)의 경우 lag [6, 5, 4, 3, 2]에 대해 관측자료를 이용하고 lag [1, 0]에 대해 MME 예측자료를 이용
- 2개월 선행예측(LT2)의 경우 lag [6, 5, 4, 3]에 대해 관측자료를 이용하고 lag [2, 1, 0]에 대해 MME 예측자료를 이용
- 3개월 선행예측(LT3)의 경우 lag [6, 5, 4]에 대해 관측자료를 이용하고 lag [3, 2, 1, 0]에 대해 MME 예측자료를 이용

(나) 월자료 공간적 재조합(Cutmix 응용)을 활용한 자료강화

Cutmix 방법은 (세로, 가로, 채널)의 3차원 입력자료를 채널에 대해서는 고정하고 세로 x 가로에 대해 정사각 형태의 2차원 영역을 치환하며, 면적비로 각 카테고리의 가중치를 곱하여 레이블을 산정한다(Yun et al. 2019). 해당 방법을 이 연구의 4D 자료에 적용하되, 변수에 대해서는 고정하고 (위도, 경도, 기간)의 3D 부피에 대해 치환하여 부피 비로 각 카테고리의 가중치를 곱하여 레이블을 산정하도록 하였다. 여기서 기간은 모델의 입력자료로 활용하는 예측인자의 선행기간 윈도우를 의미한다. 위도와 경도에 대해서는 전체 범위의 1/8 이상, 7/8 이하로 치환하도록 하고 기간에 대해서는 1개월 이상, 6개월 이하로 치환하도록 하였다.

기존 논문에서는 입력자료의 수가 많으므로 입력자료의 일부에 대해 cutmix를 적용하여 자료강화 하였으나, 이 연구에서는 입력자료의 수가 적으므로, 교차검증 시 각 폴드에 대해 훈련연도의 모든 조합에 대해 Cutmix를 적용하였다.



$$AN: NN: BN = \frac{ABC - abc}{ABC}; \frac{abc}{ABC}; 0$$

Figure 22. Conceptual image of the Cutmix data augmentation technique applied in this study.

(3) 합성곱신경망(CNN) 모델 구축 및 테스트

(가) 합성곱신경망(CNN) 모델의 구성

모델은 두 개의 합성곱신경망 블록으로 이루어진 특성 추출(Feature extraction) 네트워크와 완전연결층 및 Softmax 함수의 분류 네트워크로 이루어져 있으며 기본적인 정보는 다음과 같다.

- 입력자료 차원: (위도, 경도, 기간, 변수)의 (72, 144, 7, 5)
- 목표로 하는 분류 클래스: 우리나라 평균기온 삼분위 예측 카테고리 3개
- 모델의 최적화 함수(optimizer): adam
- 모델의 손실 함수(loss): categorical cross entropy
- 학습율: epoch 5까지 증가하였다가 지수적으로 감소하도록 설정
- Batch 크기: 32

특성 추출 네트워크의 첫 번째 합성곱 블록의 경우 컨볼루션 층은 128개의 (3, 3, 3) 커널 (필터)을 사용하며 활성화 함수는 ReLU를 적용하였다. 커널의 개수는 비교적 크게 설정하였는데 입력자료에서 우리나라 기후예측을 위한 많은 인자가 존재한다고 가정하고 다수의 특성맵(feature map)을 추출하고자 하였기 때문이다. 3D 맥스 풀링의 크기는 (2, 2)로 설정하였다. 두 번째 합성곱 블록의 경우 컨볼루션 층은 256개의 (3, 3, 3) 커널을 사용하며 활성화 함수는 역시 ReLU를 적용하였고 3D 맥스 풀링의 크기는 (2, 2)로 설정하였다.

두 차례의 맥스 풀링으로 “기간” 부분이 납작해졌으므로 Squeeze 함수를 적용하여 3차원으로 압축하여 Global 맥스 풀링을 적용하였으며 512개 노드를 지닌 완전연결층을 연결하고 Softmax 함수를 통해 AN, NN, BN 카테고리의 확률을 출력하도록 하였다.

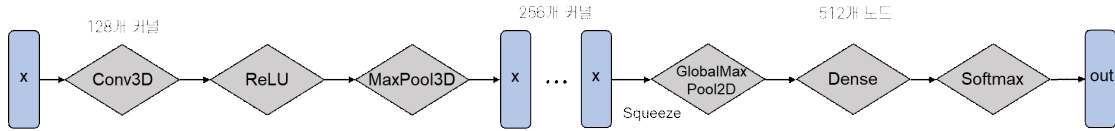


Figure 23. Structure of the Vanilla 3D CNN model used in this study.

(나) 합성곱신경망(CNN) 모델 테스트 설계

여름철(6월 LT1, 7월 LT2, 8월 LT3) 평균기온 예측 성능 개선이 목표인데 월별, 선행시간 별로 모델을 따로 구성하는 경우 가용 자료의 수가 매우 적다. 따라서 모델 훈련 시 모든 월의 자료를 다 사용하는 경우(Month-Agnostic)를 제안하고 목표 월만 사용하는 경우(Target Month Only)와 비교하였다. Month-Agnostic의 경우 월에 상관없이 예측인자의 유사한 공간 분포가 있으면 유사한 결과가 나타난다는 가정에 기반하여 구상하였다.

자료강화한 자료를 포함하는 경우 전체 자료를 훈련과 검증/테스트 자료로 무작위로 나누면 같은 시점(월)에 대한 강화자료가 훈련과 검증/테스트 양쪽에 모두 포함되어 오염이 일어날 수 있으므로, 자료 분리 시 연도를 나누도록 하였다. 전체 기간에서 모델의 성능을 최종 평가할 테스트셋을 제외하고 나머지 기간을 3개의 폴드로 나누어 교차검증을 수행하였다. 5개의 폴드도 고려하였으나, 이 경우 검증 연도 수가 일정하지 않고(4년 또는 5년) 폴드별 모델 결과의 차이가 너무 많이 나타나 3개의 폴드를 사용하기로 결정하였다.

만들어진 폴드별 모델은 월에 상관없는 선행시간별 모델이 될 것이나 6월(LT1), 7월(LT2), 8월(LT3)의 테스트 결과를 최종 평가할 것을 염두에 두고 카테고리별 예측 성능을 고루 살펴 보기 위해 테스트 연도와 훈련/검증연도를 관심 월에 대해 아래와 같이 선정하여 사용하였다.

- 테스트 연도에는 AN, NN, BN인 해가 각각 2년 이상씩 포함되도록 함
- 각각의 폴드도 검증연도가 각 카테고리를 2년 이상씩 포함하도록 선정함
- 만일 테스트 연도 제외한 남은 기간에 모든 폴드에 2년씩 할당할 만큼 해당 카테고리를 참값으로 하는 연도가 존재하지 않는다면 $\min(\text{floor}(\text{해당 카테고리의 전체 빈도} / 3), 2)$ 이상 포함하도록 선정함

(4) 그래프신경망(GNN) 모델 구축 및 테스트

(가) 노드 분류를 위한 그래프신경망(GNN) 모델의 구성

노드 분류 모델은 모든 노드(Node)가 포함된 하나의 그래프를 이용하여 구축할 수 있는데, 이 그래프는 훈련, 검증, 테스트 노드를 모두 포함하며(준지도학습) 노드의 라벨 여부로 훈련, 검증, 테스트셋으로 자료를 나눈다.

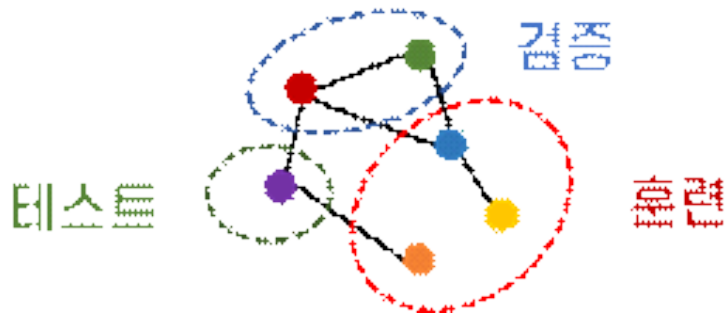


Figure 24. Conceptual diagram of the split of training, validation, and test sets in a transductive Node Classification model.

모델은 세 개의 그래프합성곱신경망 블록으로 이루어진 특성 추출(Feature extraction) 네트워크와 완전연결층 및 Softmax 함수의 분류 네트워크로 이루어져 있으며 기본적인 정보는 다음과 같다.

- 각 월이 노드를 구성(ex. 2017년 7월)
- 노드 특성(Node features): (위도, 경도, 기간, 변수)의 1D 벡터로 구성된 입력자료 값으로 $15^\circ \times 15^\circ$ 해상도의 자료를 사용했으므로 벡터의 길이는 $12 \times 24 \times 7 \times 5 = 10,080$ 이 됨
- 노드 라벨(Node label): 우리나라 평균기온 삼분위 예측 카테고리 3개
- 엣지(Edge) 및 가중치: 노드 사이의 연결 및 연결 강도를 나타냄. 엣지 연결 및 가중치 초깃값으로 각 노드 쌍에 대해 노드 특성 벡터 사이의 코사인 유사도를 구하고 그 값이 0.1 초과인 경우(대략 상위 25% 정도에 해당됨) 엣지를 연결하여 그 코사인 유사도를 가중치 초깃값으로 설정함. 엣지의 가중치를 학습 가능한 매개변수로 지정하여, 모델의 학습 과정에서 함께 학습될 수 있도록 함
- 모델의 최적화 함수(optimizer): adam
- 모델의 손실 함수(loss): categorical cross entropy

- 학습율: 지수적으로 감소하도록 설정(lr_lambda = 0.95 ** epoch)
- Batch 크기: 32

노드 임베딩의 인코더로 GCN (Graph Convolutional Network) 모델을 사용하였다. GCN은 가장 간단한 딥인코더의 하나로 이웃의 메시지를 평균하고 자기 자신 노드의 메시지를 결합하여 신경망을 적용한다. k 번째 층(layer k)에서 노드 v 의 임베딩(hidden representation)을 $h_v^{(k)}$ 라고 하면, $h_v^{(0)}$ 는 초기 0번째 층의 임베딩으로 처음 입력된 노드 특성과 같으며, N_v 는 노드 v 의 이웃 노드의 집합, $|N(v)|$ 는 노드 v 의 degree (이웃 노드의 개수), K 는 층의 개수, σ 는 활성화 함수이고 z_v 는 K 개의 층을 지난 후 이웃 정보의 결합을 통해 산정된 최종 임베딩이 된다. 이때, W_k 와 B_k 는 학습 가능한 매개변수로 각각 이웃 결합(neighborhood aggregation)을 위한 가중치 행렬과 자기 노드의 은닉 벡터를 변환(self transformation)하기 위한 가중치 행렬이며 모든 노드에 대해 공통으로 적용된다.

$$h_v^{(0)} = X_v \quad [16]$$

$$h_v^{(k+1)} = \sigma \left(W_k \sum_{u \in N(v)} \frac{h_u^{(k)}}{|N(v)|} + B_k h_v^{(k)} \right), \forall k \in \{0, \dots, K-1\} \quad [17]$$

$$z_v = h_v^{(K)} \quad [18]$$

분류를 위한 지도학습에서 손실함수로 교차 엔트로피를 사용하는데 y 는 분류 카테고리 라벨(참값), z_v^T 는 인코더를 통과한 노드 임베딩이며 θ 는 분류를 위한 가중치이다.

$$L = \sum_{v \in V} y_v \log(\sigma(Z_v^T \theta)) + (1 - y_v) \log(1 - \sigma(Z_v^T \theta)) \quad [19]$$

첫 번째 GCN 블록은 노드 임베딩과 Batch Normalization, ReLU를 이용한 활성화와 자료의 30% Dropout 과정을 거친다. 노드 임베딩의 차원은 256으로 지정하였다(노드별 임베딩 벡터의 크기). 두 번째 GCN 블록도 동일한 형태를 가지며, 세 번째 GCN 블록은 노드 임베딩을 수행한 후 Softmax 함수를 이용하여 AN, NN, BN의 삼분위 카테고리 각각에 대한 확률을 출력한다.

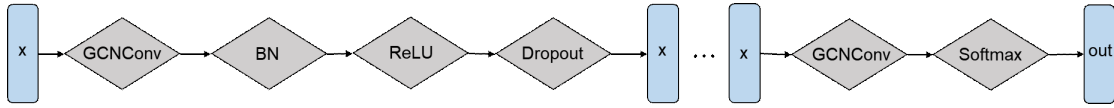


Figure 25. Structure of the GNN model for Node classification used in this study.

(나) 그래프 분류를 위한 그래프신경망(GNN) 모델의 구성

그래프 분류 모델은 각 월이 그래프를 구성(ex. 2017년 7월)하도록 전체 사용 자료의 개수만큼 그래프를 구축하였다. 그래프의 라벨 활용 여부로 훈련, 검증, 테스트셋으로 자료를 나누었으며 그래프 라벨(graph label)이 우리나라 평균기온의 삼분위 카테고리라 된다.

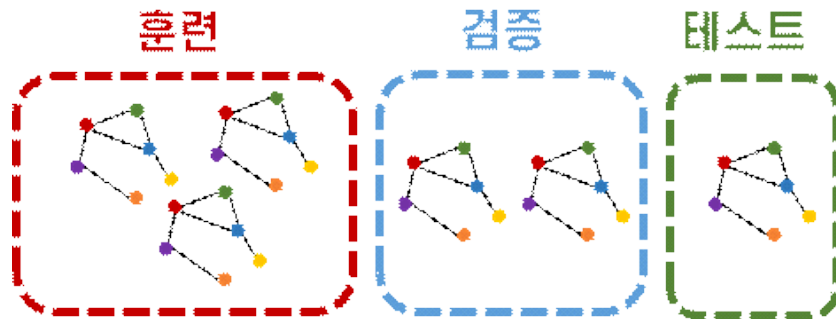


Figure 26. Conceptual diagram of the split of training, validation, and test sets in a Graph Classification model.

그래프 안의 각 노드는 공간 격자(위도, 경도)로 이루어지며 노드 특성은 (기간, 변수)의 1D 벡터로 구성된 입력자료 값으로 이 벡터의 길이는 $7 \times 5 = 35$ 가 된다. 엣지(edge)는 각 노드 쌍에 대해 노드 특성들 사이의 코사인 유사도를 구하고 그 값이 0.1 초과인 경우 또는 인접한 그리드인 경우에 대해 연결하며 가중치는 사용하지 않았다. 그 밖의 정보는 다음과 같다.

- 모델의 최적화 함수(optimizer): adam
- 모델의 손실 함수(loss): categorical cross entropy
- 학습율: 지수적으로 감소하도록 설정($lr_lambda = 0.95 ** epoch$)
- Batch 크기: 32

모델의 형태는 앞서 설명한 노드 분류 모델의 세 GCN 블록을 그대로 노드 임베딩을 위해 이용하고 Global mean pooling을 통해 이를 평균하여 취합(Aggregation)한다. 이후 노드 256개의 완전연결층을 연결하고 Softmax 함수를 통해 AN, NN, BN의 삼분위 카테고리 각각에 대한

확률을 출력한다.

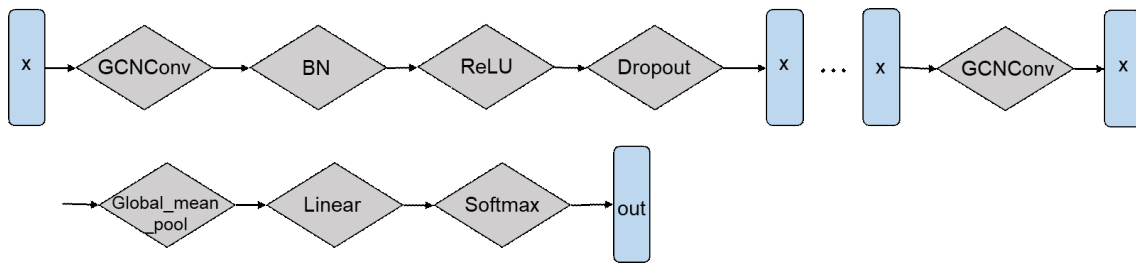


Figure 27. Structure of the GNN model for Graph classification used in this study.

(다) 그래프신경망(GNN) 모델 테스트 설계

이미지기반의 합성곱신경망모델과 마찬가지로 여름철(6월 LT1, 7월 LT2, 8월 LT3) 평균기온 예측 성능 개선이 목표이다. 이 경우에도 월별, 선행시간별로 따로 모델을 구성하는 경우 가용 자료의 수가 매우 적으므로 모델 훈련 시 목표 월만 사용하는 경우(Target Month Only)와 모든 월의 자료를 다 사용하는 경우(Month-Agnostic)를 비교하였다. 합성곱신경망모델 테스트의 경우와 마찬가지로 3-fold 교차검증을 수행하였다.

3. 결과

가. 설명가능 인공지능(XAI) 기술적용을 위한 환경구축

(1) 기준 모델의 예측 성능 평가

인공신경망 기반 계절 예측 후처리 모델의 성능을 평가하기 위하여 사전 학습과 미세 조정 학습의 두 가지 단계를 통해 모델을 훈련하고 학습 과정을 평가하였다. 구체적으로, 사전 학습을 위해서는 비교적 장기 기후 정보가 제공되는 NCEP1 재분석 자료를 이용하였고, 미세 조정학습을 위해서는 APCC MME의 2018년 버전 과거 재현 자료와 최신 과거 재현/예측 자료 및 관측소 관측 자료를 사용하였다. 일부 예측 안정성 향상을 위한 실험에서는 두 자료를 교차 혼합하여 사전 학습을 수행하기도 하였다. 이를 통해 사전 학습의 학습 자료 cosine similarity 0.9 와 검증 자료 cosine similarity 0.7 수준을 달성한 3-차원 합성곱 기반 모델로 미세 조정학습을 수행 중에 있다.

합성곱 커널의 가시화를 통해서 인공신경망 자체가 어떻게 구성되어 있는지 확인하고 이를 기반으로 모델을 개선하고자, 3-차원 합성곱을 통해 다변수 입력을 처리하는 층의 커널 모습을 확인해 보았다. 각 열이 하나의 커널을 나타내고 행은 개별 커널 내의 변수를 나타내며, 입력 자료는 변수별로 정규화가 완료된 상태이다. 전체적으로 강수에서 뚜렷한 패턴이 많이 잡히고, 이와 함께, 하층 기온과 동서류에서 상위 패턴이 나타나는 것을 확인할 수 있다. 하층 남북류의 경우 동서류와 함께 동작하는 경우와 그렇지 않은 경우로 나뉘는 것을 볼 수 있습니다. 커널 구조간 상관도를 살펴보았을 때, 변수 간 분포가 비슷한 구조를 띠는 경우(예를 들어, 2-5 혹은 6-8 커널)가 종종 관찰된다. 커널 크기 조정을 통해 기후 변수 간의 물리/역학적 관계를 검토하거나, 커널 간 상관도를 하이퍼파라미터 조정에 사용할 수도 있을 것으로 판단된다. 즉, 현재 $2.5^\circ \times 2.5^\circ$ 크기의 격자점에 대해 2×3 (위도 \times 경도)를 사용하였으나, 이는 중 위도에서 일반적인 중관 규모의 기상 현상을 표현하기에는 공간적인 범위가 협소하다. Dilation 기법 등을 통하여 커널의 크기를 확대하는 방법을 통해 커널의 공간 범위를 확대할 수 있을 것으로 보인다.

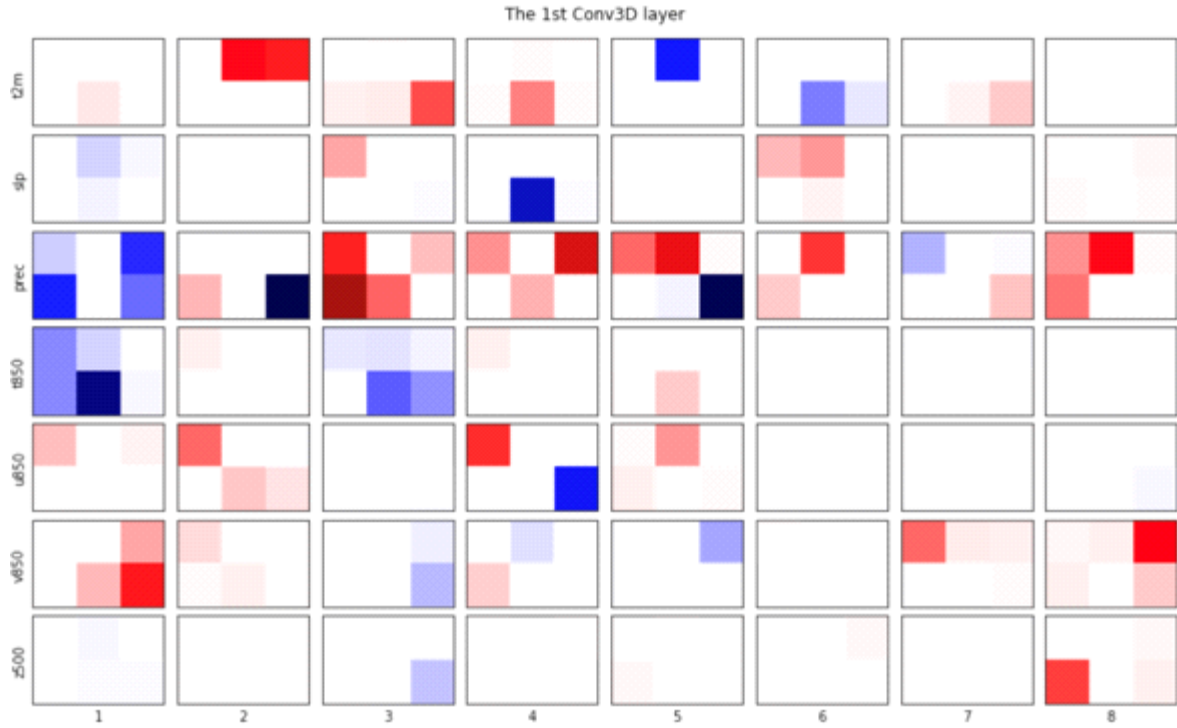


Figure 28. Weights of the 3D convolution kernels of the input layer with the dimension of 2(latitude) x 3(longitude) x 7(variable) x 8(number of kernels).

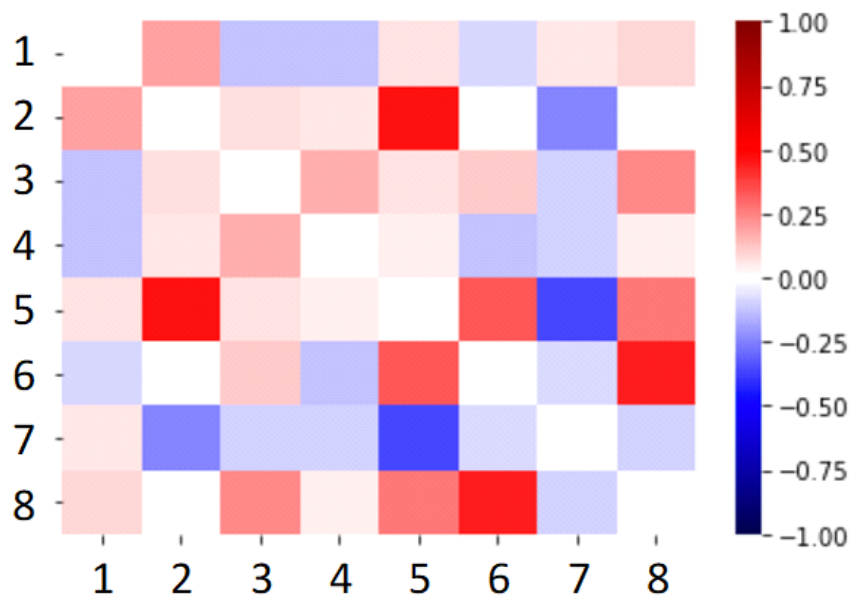


Figure 29. Correlation coefficients between kernel structures.

각 층의 활성화 정도와 각 층을 통과하는 정보량을 이용하여 인공지능망의 복잡도를 일부 평가해볼 수 있다. 앞서 살펴본 활성화맵과 비슷한 모습을 보이고는 있는 아래 그림은 실제 인공

신경망 안에서 첫 번째 3-차원 합성곱 층을 통해 활성화된 평균 정보를 시각화한 것이다. 즉, 커널을 통과한 정보량이라 볼 수 있다. 대부분의 경우, 기후적 측면에서 유의한 정보가 추출되어야 할 지역에서 유의미한 시그널들이 관찰되는 것을 볼 수 있다. 전체 채널을 평균해보았을 때 적도 태평양 지역에서 정보량이 중요한 것으로 보인다. 하지만 한편으로는 일부 커널의 통과 값이 기후적으로 의미 있는 패턴을 보이지 못하는 경우, 추출해야 하는 정보와 이를 추정할 수 있는 자료의 양에 비해 지나치게 복잡한 모델 파라미터를 사용했을 가능성이 있다. 이를 이용하여 물리적 관점에서 하이퍼파라미터 튜닝을 시도할 수 있을 것이다.

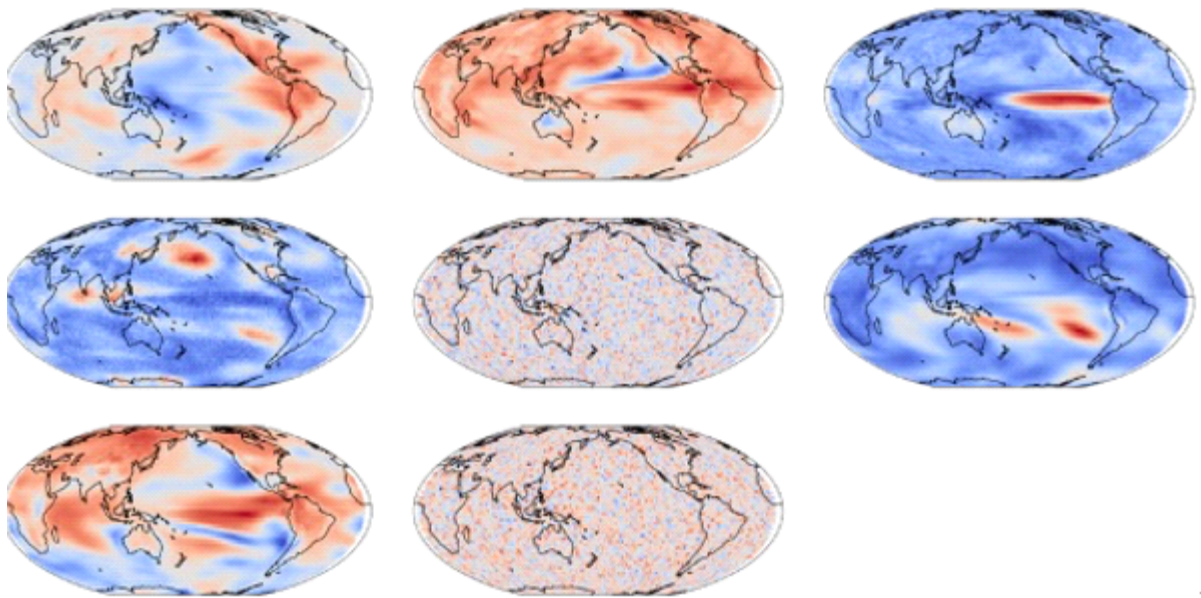


Figure 30. Visualization of activated values of the first 3D convolution layer.

정보를 입력받는 초기 층뿐만 아니라, 마지막 층에 대한 분석도 인공지능망 모델 개선에 흥미로운 정보를 줄 수 있다. 다음 그림은 관측재분석 자료를 이용해 사전 전이 학습을 수행한 인공지능망 모델의 맨 마지막 완전연결층의 연결 강도와 맨 마지막 완전연결층을 제외한 인공지능망 모델은 고정하고 실제 자료(예를 들어 MME-관측소 자료의 조합)를 이용하여 추가 미세 조정 학습을 수행한 후 변화한 완전연결층의 연결 강도 변화를 보여주고 있다.

이 과제에서는 적은 수의 자료를 이용하여 복잡한 기후 현상을 적합하기 위하여 전이 학습이라는 기법을 사용하는데, 왼쪽 그림의 경우 전이 학습을 통해 적합한 모델의 파라미터가 실제 자료를 처리하는 데 그대로 사용해도 무리가 없는 안정적인 피팅을 보여준다. 반면 오른쪽 그림의 경우 사전 학습 자료를 통해 학습한 패턴과 실제 자료가 맺는 현상 사이에 큰 차이가 존재하며, 최종 미세 학습 과정에서 인공지능망 모델이 크게 드리프트하는 모습을 보여주고 있다. 또한, 일부 뉴런에서는 전체 자료를 투입하더라도 뚜렷한 활성화가 관찰되지 않는 경우도 있는 것으로 나타나고 있다. 이러한 특성을 분석함으로써, 과대/과소 적합 없이 자료를 설명하기 위한 적절한 수의 뉴런(하이퍼파라미터)으로 구성하였는지, 사전 학습을 통해 형성된

네트워크가 실제 자료에서 안정적으로 동작하는지 등을 파악할 수 있고, 이를 인공신경망 모델을 구성하는 데 활용할 수 있을 것으로 보인다.

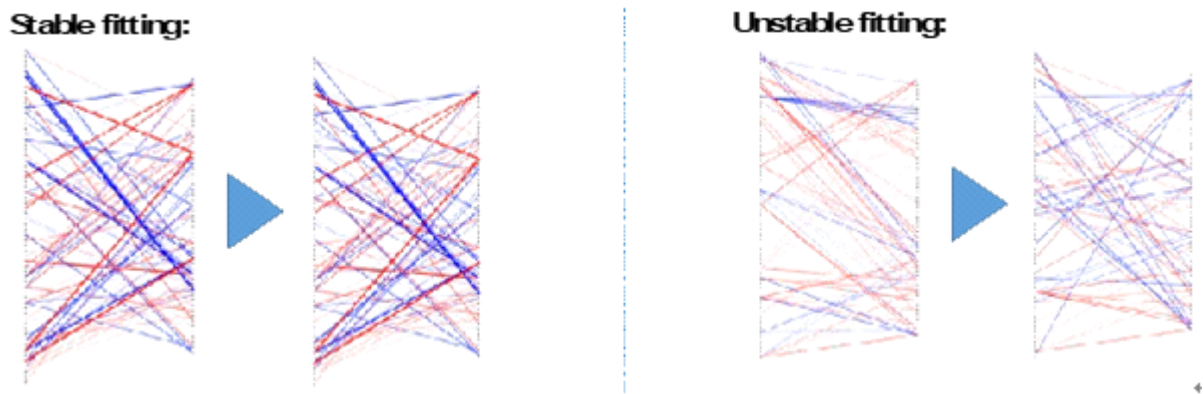


Figure 31. Visualization of the connection strength of the last hidden layer and the output layer.

나. 제한된 기후자료 확장을 위한 환경구축

(1) 자료 파이프라인 구축: 인공지능 훈련자료 확장

(가) 개별모델 딥러닝 모델 학습결과

자료확장을 위하여 ECMWF와 GPM 강수량자료를 ConvLSTM 5D 입력자료(17년, 28일, 위도, 경도, 변수)로 구성하고, 이를 3개월로 확장(238, 28일, 위도, 경도, 변수)확장하고 훈련셋을 구성하였다. ConvLSTM2D함수를 이용하여 모델을 구성하였고, 커널은 (3, 3), maxpooling은 (2, 2), filter는 16로 두고 activation은 relu 함수를 사용하여 모델을 구성하였다. Optimizer는 아담함수를 loss는 mse 함수를 이용하고 batch_normalization을 적용하고 최종적으로 fully-connected layer를 격자형태로 자료를 산출할수 있도록 Conv3D함수를 이용하여, ERA5자료에 데이터가 fitting되도록 ConvLSTM모델을 구성하고 학습을 진행하였다. ConvLSTM의 모델이 과거 학습데이터의 패턴이 비슷한 경우에는 잘 학습되지만 그렇지 못한 경우에는 부족한 데이터로 인하여 학습이 거의 되지 못하고, 시계열로 초기 feature 가중치가 3-4주까지 전달되는데 어려움을 가진다(Figure 32).

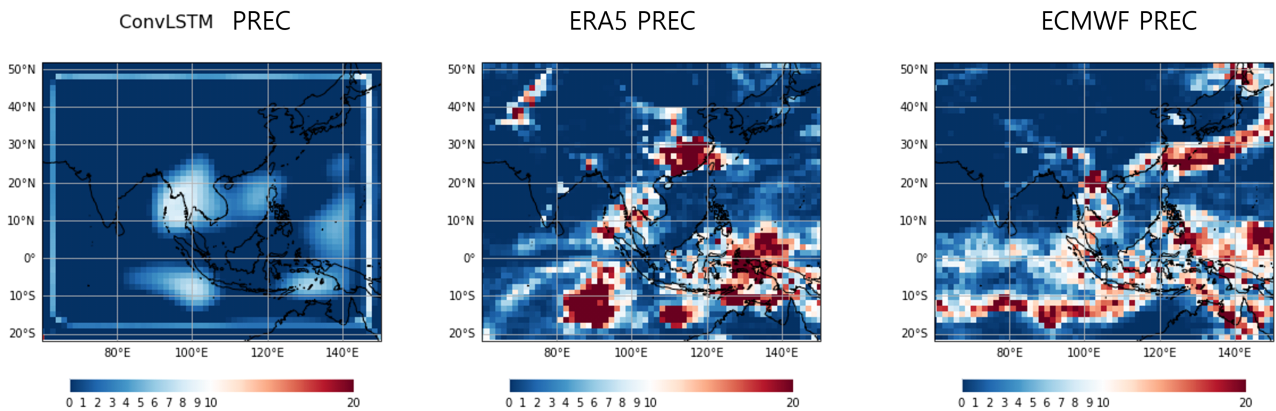


Figure 32. Comparison of the result of ConvLSTM with ECMWF model at daily precipitation.

이에 반해 적은 수의 자료에도 학습 능력이 높은 U-Net 딥러닝 모델에 ECMWF, GPM 강수량 자료를 3주차 7일 치 자료와 예측일에서 과거 7일 치 자료를 뽑아내서 3개월 238 샘플갯수를 곱하여 1,666개의 입력자료를 (1666, 48, 48, 2)으로 만들어 12년 치(2001-2012)인 1,176개를 학습데이터로 이용하고 5년 치(2013-2017) 490개를 테스트 데이터로 활용하여 학습결과 PCC가 ECMWF 0.26보다 높은 0.39이고, RMSE는 9.6, 11.3으로 작아지긴 했지만 크게 개선되는 경향은 보여주지 못하였다(Figure 33). 자료를 확장하여 U-Net의 입력자료로 활용했지만 여전히 1,176개의 자료 수는 많지 않은 양의 자료이며, RMSE는 크게 개선되지 않는 결과로 보아 과적합일 경우가 있어 민감도 분석을 통해 다양한 방법으로 feature를 선택하여 결과를 분석할 필요성이 있다(Figure 34).

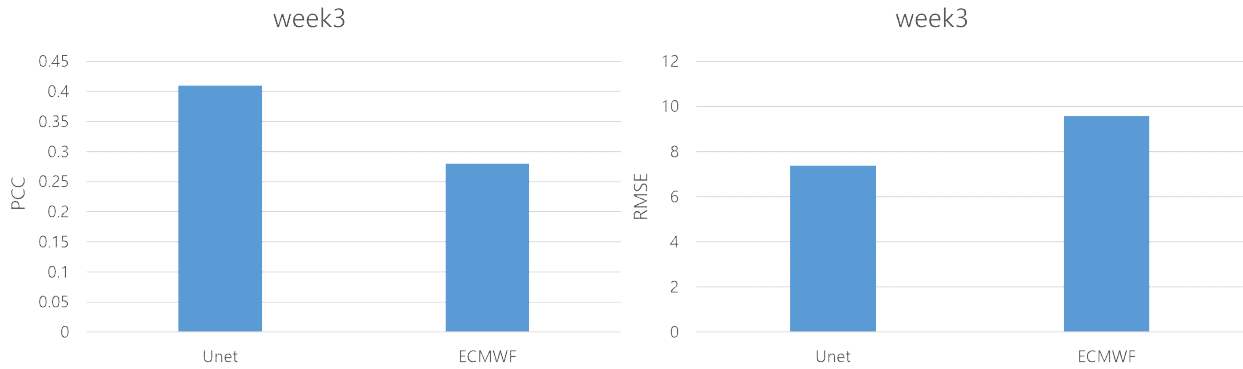


Figure 33. Comparison of performance of pattern correlation and RMSE of precipitation for 3 week lead-time.

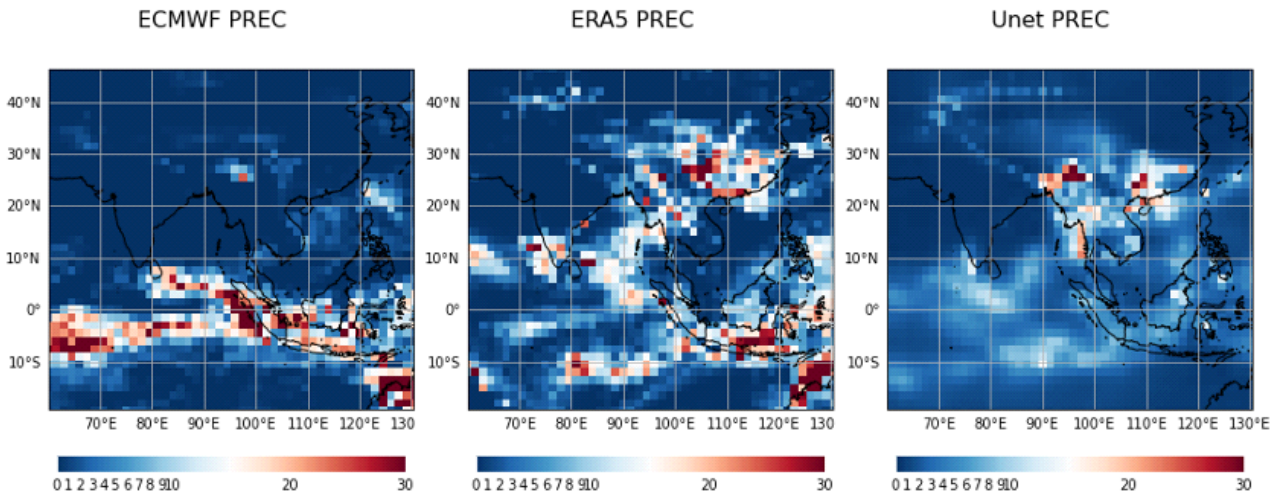


Figure 34. Comparison of the result of U-Net with ECMWF model at daily precipitation.

(나) 입출력 시스템 개발

입출력 시스템의 개별 모듈은 사용자 중심으로 다양한 조건별로 데이터를 읽거나 전처리 하는 각각의 모듈로 구성된다. 위, 경도별로 자료를 Clipping 할 수 있는 모듈과 계절 내 예측 모델(ECMWF, GloSea5)을 선택하고, 관측자료인 GPM 강수량, NDVI, SST 자료를 선택하고 각각의 자료 품질관리 할 수 있는 모듈과 모델별 feature를 선택하는 모듈, 자료를 NetCDF로 산출할 수 있는 모듈과 딥러닝 모델(ConvLSTM, U-Net, CNN)의 입력자료로 만들어 줄 수 있는 모듈, 최종적으로 산출된 자료를 검증할 수 있는 가시화 모듈을 개발하였다(Figure 35).



Figure 35. The Connectivity from various module for in and out data system.

입출력 시스템 모듈은 Clipping, 기간 선정, 역학 모델 선택, 관측자료 선택, 최종자료 산출 가시화와 같은 모듈을 GUI 환경과 연계되어 객체지향적으로 수행되는 시스템이다. GUI 환경은 상태 바에서 “Processing” 으로 시작해서 모든 모듈이 돌아가서 자료를 산출하면 “End” 상태로 표시된다. 타겟변수는 강수, 최고/최저 온도를 선택 가능하며, 기간은 1991년부터 2010년까지 20년간 자료를 선정할 수 있도록 표시하였고, 더불어 위, 경도를 입력하면 입력된 자료에 맞게 입력자료가 산출되도록 GUI를 구성하였다. 역학 모델은 GloSea6, ECMWF를 선택 가능하면 관측자료는 GPM, NDVI, SST 등의 위성자료 선택 가능하다. Feature 변수를 선택하고 학습할 딥러닝 모델을 선택하여 GenerateData 버튼을 클릭하면 최종 학습할 자료가 지정된 디렉토리에 생성된다(Figure 36). Visualization을 클릭하면 최종 학습자료가 잘 산출되었는지 자료 가시화를 통해서 확인할 수 있다(Figure 37).

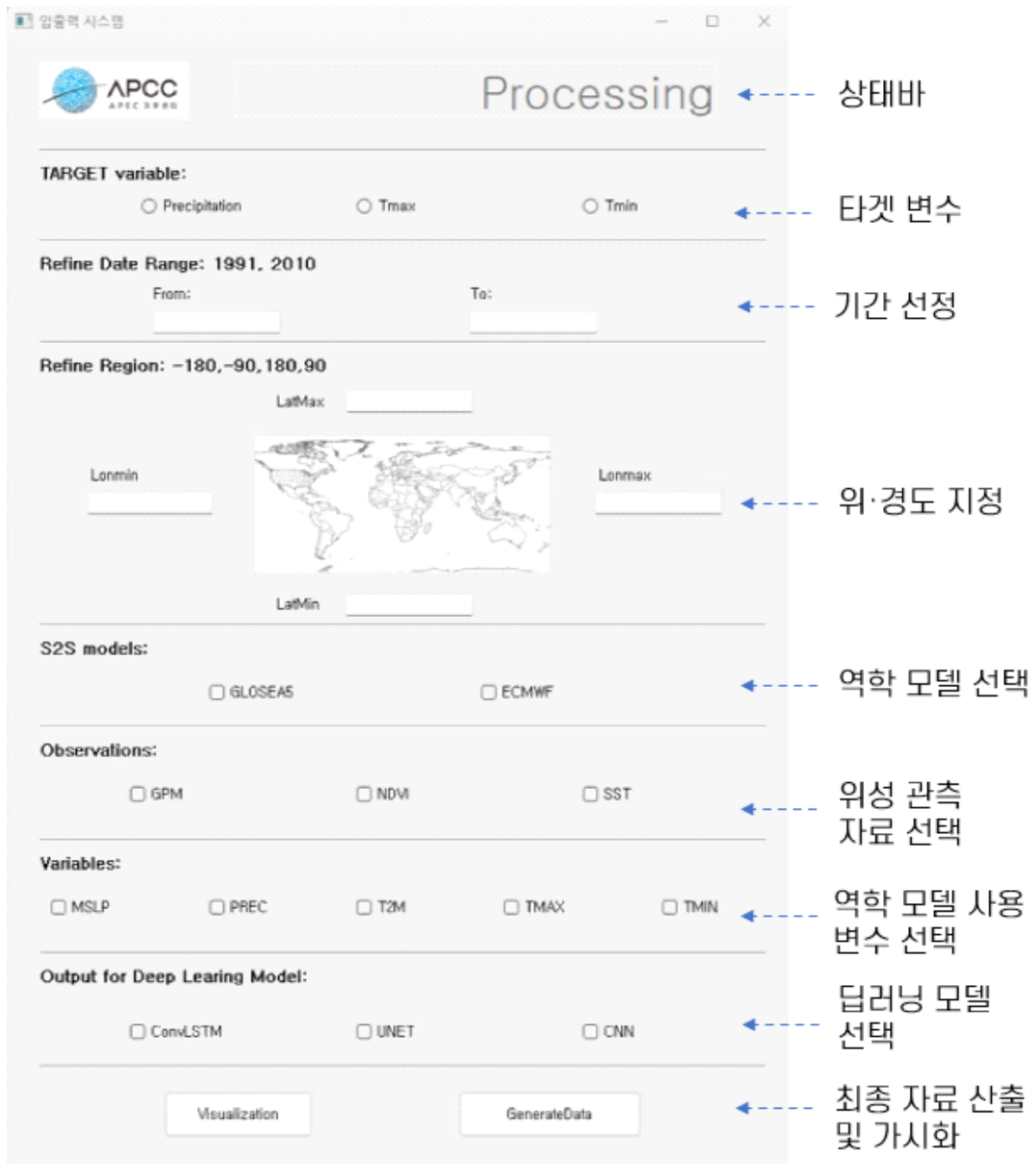


Figure 36. GUI of in and out data system.

ConvLSTM PREC

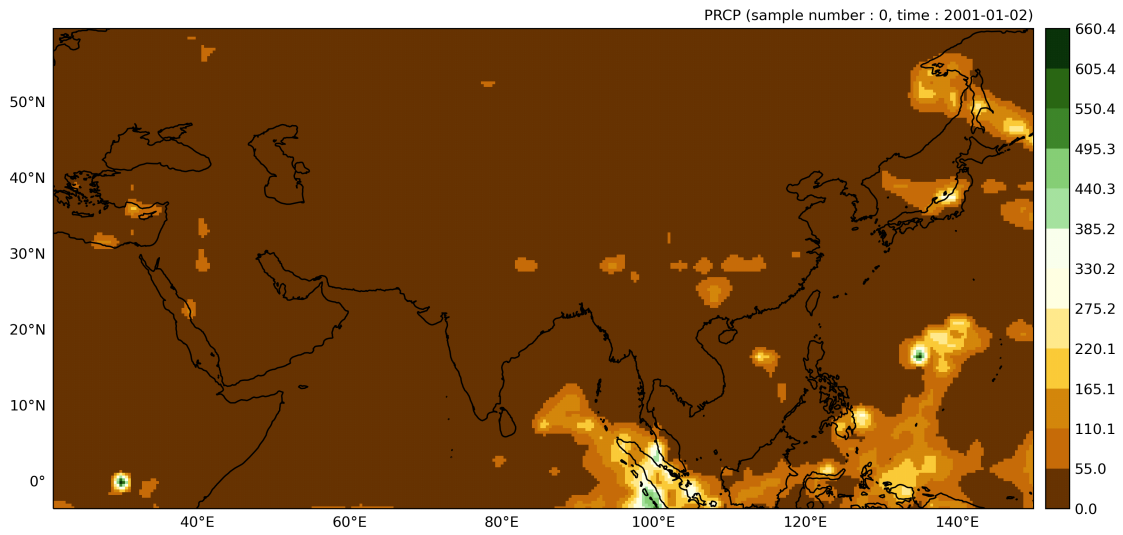


Figure 37. Visualization of verification for training data from in and out data system.

최종적으로 개발된 입출력 시스템은 사용자 중심으로 다양한 예측자료와 관측자료를 선택하고 필요시에 feature를 선택하면서 딥러닝 모델을 고정하고 민감도 분석을 통하여 입력자료에 따른 정확도 높은 딥러닝 모델을 산출하고자 한다(Figure 38).

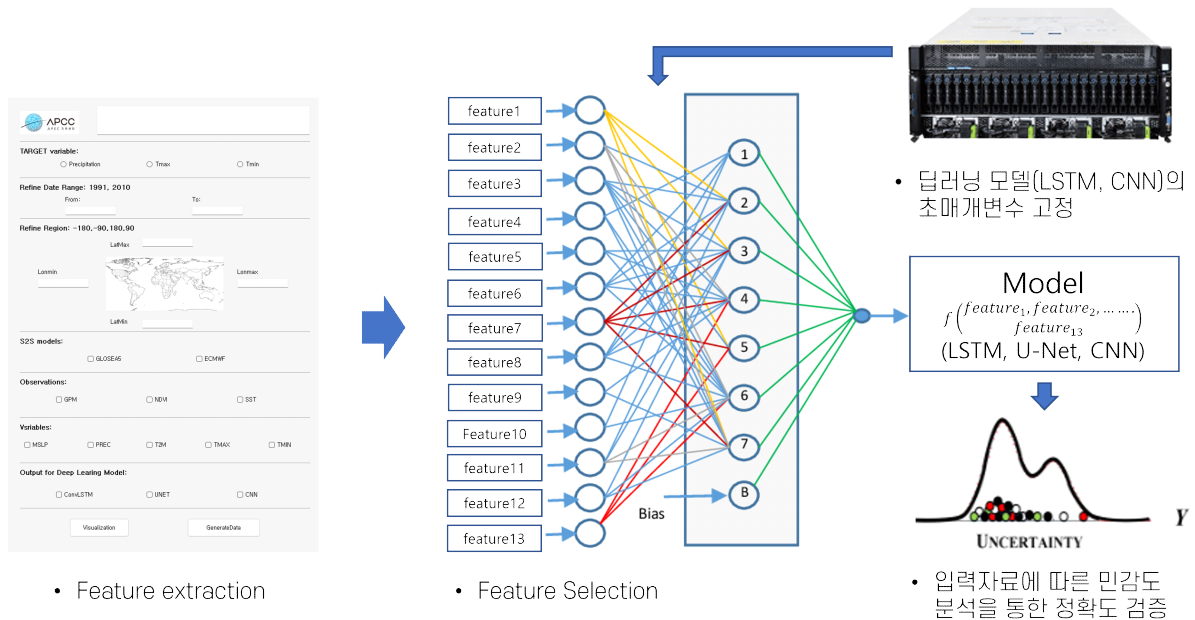


Figure 38. Sensitivity analysis of feature selection from in and out data system.

(2) 자료 파이프라인 구축: 훈련자료 전처리 기법 테스트

(가) 전처리 기법에 따른 예측 성능 평가

① 전처리 기법에 따른 변형

Figure 39는 최고 및 최저 차이 알고리즘으로 변환된 계절내-계절(Sub-seasonal to seasonal, 이하 S2S) 훈련자료이다. MaxAbs Scaler는 -1에서 1 사이의 값으로 변환되는데, 이 연구의 S2S 훈련자료도 -0.5에서 0, 그리고 0.5 분포로 변환되었다. MinMax Scaler는 0에서 1 사이의 값을 갖도록 변환되는데, 역시 S2S 훈련자료가 0에서 1 사이에서 변환되었음을 확인할 수 있다. 그러나 Figure 39의 MinMax 결과에서 0과 1이 전혀 없는 것으로 나타났는데, 이것은 개별 기후모델의 S2S 원시자료에서 최저 혹은 최고값이 없을 뿐만 아니라 관측값(ERA5)에도 없다는 의미로 모델값이기 때문으로 생각한다.

Figure 40은 표준편차 알고리즘으로 변환 S2S 훈련자료이다. Standard Scaler는 평균을 제거하고 정규 분포로 변환되는데, 히스토그램을 비교해보면, 변환된 S2S 훈련자료와 참값 레이블의 분포에서 차이가 발생하였다.

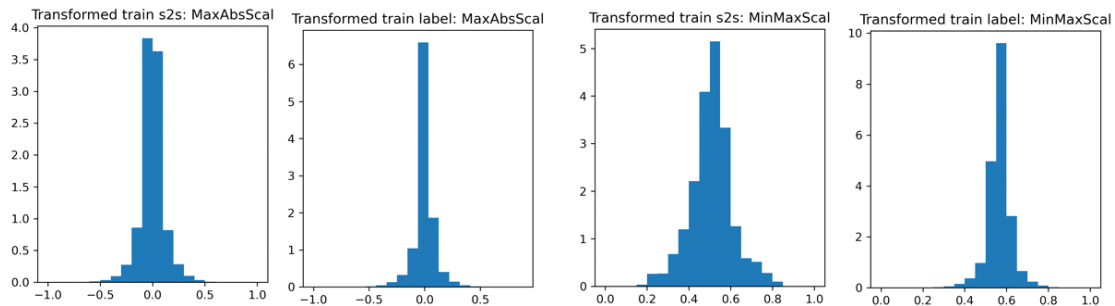


Figure 39. Comparison of histograms of transformed training data for Sub-seasonal to seasonal (S2S) by minmax scaler and maxabs scaler: MinMax and MaxAbs scalers are the algorithm of change scale of data.

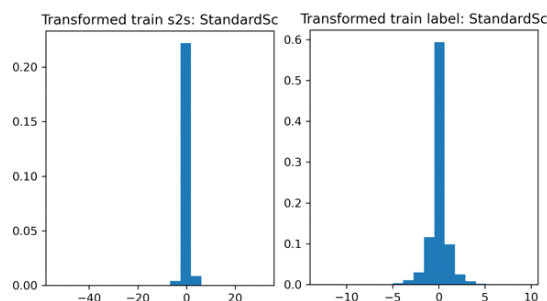


Figure 40. Comparison of histograms of transformed training data for S2S by standard scaler: Standard scaler is the standard deviation algorithm of change scale of data.

Figure 41은 사분범위 알고리즘으로 변환된 S2S 훈련자료이다. Robust Scaler는 사분범위 (IQR, interquartile range)에 의해 변환되는데, Figure 41과 같이 S2S 훈련자료의 IQR과 참값 레이블의 사분범위가 다름을 확인할 수 있다. Quantile Transformer도 분위수에 의해 변환되는데 가우시안 분포와 유사하게 변환되었다.

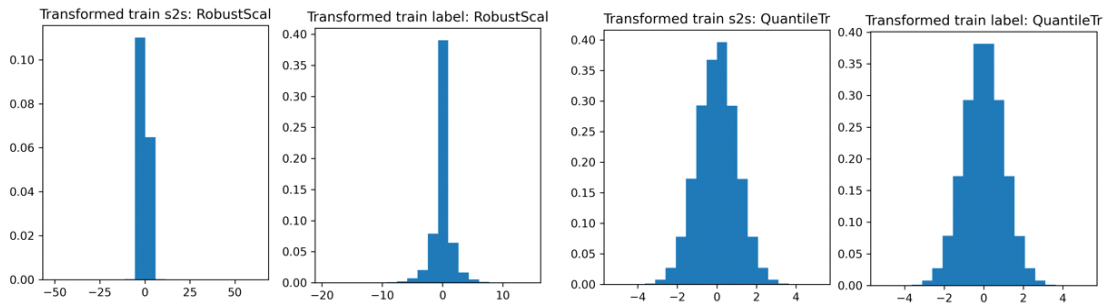


Figure 41. Comparison of histograms of transformed training data for S2S by robust scaler and quantile transformer: Robust scaler is the algorithm of change scale of data, and quantile transformer is the algorithm of change distribution of data.

② 전처리 기법 파이프라인 구축

Figure 42는 이 연구에서 Scaler를 적용하지 않은 Unscaled와 5개 Scalers로 변환된 S2S 훈련자료에 대한 훈련 모델의 학습 성능 결과이다. 평가 지수는 코사인 유사도, 평균평방근오차 (root mean square error, 이하 RMSE)와 평균절대오차(mean absolute error, 이하 MAE)를 계산하여 metric으로 구성하였다. Base model에서 여섯 개 개별 기후모델과 MME에 대한 각각의 전체 훈련 기간에 대한 학습 성능에 대한 오차 스프레드를 비교하였다. 훈련 기간은 개별 기후 모델의 20년 기간에 대해 80%가 이용된 것이다.

평가 지수 중 코사인 유사도(cosine similarity)는 재현 혹은 유사 정도(예, 얼마나 잘 모사할 수 있는가)를 계산하는 척도(measure)인데, MinMax Scaler로부터 변환된 S2S 훈련자료의 코사인 유사도가 거의 1(예, 0.999)로 나타났다. 이것은 MinMax에 의해 변환된 S2S 훈련자료 (Figure 39)에서 0과 1이 없는데에 기인된 것으로 판단된다. 추후 MinMax Scaler로 변환된 S2S 훈련자료에서 평균을 이동(shift)하여 다시 Base mode로부터 훈련한 후 비교하는 것이 필요할 것으로 판단된다.

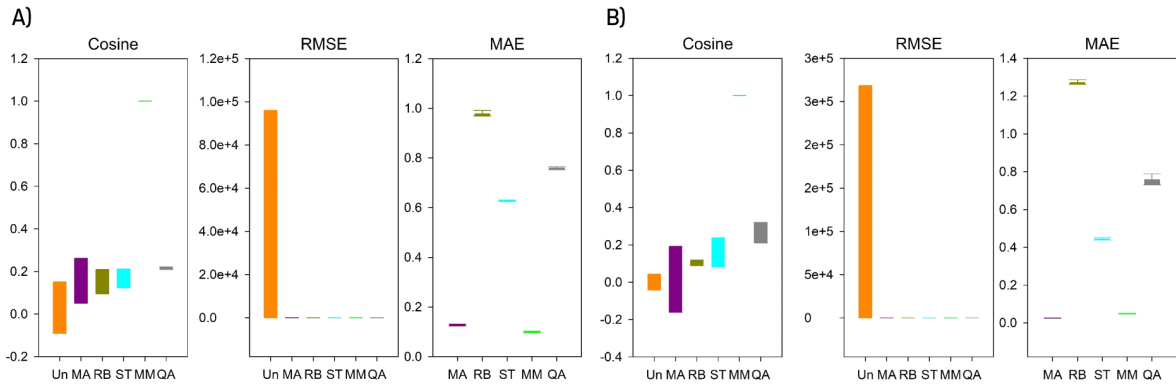


Figure 42. Comparison of the learning performance evaluation index of the base model for the S2S training data converted by each scaler algorithm over the training period. A) is the daily minimum air temperature and B) is the daily total precipitation of ECMWF S2S training data transformed from six scalers.

③ 전처리 기법에 예측 성능 평가

Scaler별 학습 성능을 상세히 비교하기 위해 선행예측시간(lead-time)에 따른 공간상관계수 (spatial Pattern Correlation Coefficient, 이하 PCC)를 계산하였다(Figure 43에서 Figure 45). 레전드에서 검정색 및 S2S는 각각 다른 훈련 기간 동안 학습된 Base model로부터 목표 예측 변수(예, 일 최고 및 최저기온과 일 적산 강수)에 대해 딥러닝 적용하지 않은 원시자료와 참값으로 입력한 재분석 자료인 관측(이하 ERA5)와의 PCC이며, 나머지 색상(회색: Unscaled, 오렌지색: MaxAbs scaler, 파란색: Robust scaler, 초록색: Standard scaler, 보라색: MinMax scaler, 하늘색: Quantile transformer)과 각각의 도형(동그라미: ECMWF, 네모: GloSea5, 위-세모: ECCO, 아래-세모: UKMO, 마름모: NCEP, 가위: CMA, 별: MME)은 개별 기후모델과 MME 원시자료에 Scaler를 적용하지 않은 것과 다섯 개 Scaler를 적용한 후 Base model에서 훈련된 결과를 ERA5에 대해 PCC를 계산하였다.

Figure 43에서 Figure 44는 일 최고 및 최저기온에 scaler를 적용하지 않은 것과 다섯 개 scaler 알고리즘에 의해 변환된 여섯 개의 개별 기후모델 및 MME의 S2S 훈련자료를 Base model에서 훈련하고, 각각의 개별 기후모델의 테스트 기간에 대해 예측된 결과를 ERA5의 일 최고 및 최저기온과의 PCC를 계산한 산포도이다. 기온의 경우, 변형되지 않은 원시자료의 PCC(검정색, 전체 평균 PCC: 0.8) 자체가 매우 높아서 전반적으로 Base model의 예측성능 개선이 뚜렷하게 나타나지 않았다. 즉, scaler가 적용되지 않은 Unscaled(회색)와 MinMax scaler(보라색)에 의해 변형된 S2S 훈련자료의 PCC를 제외하고 대부분의 scaler 알고리즘의 예측성능 차이는 뚜렷하지 않았다. 그러나 여섯 개 개별 기후모델 및 MME 원시자료의 PCC(검정색)와 비교하여 기온에 대한 Base model 후보정 효과는 선행예측시간 3주 이후에 모두 나타나는 것을 확인할 수 있었다.

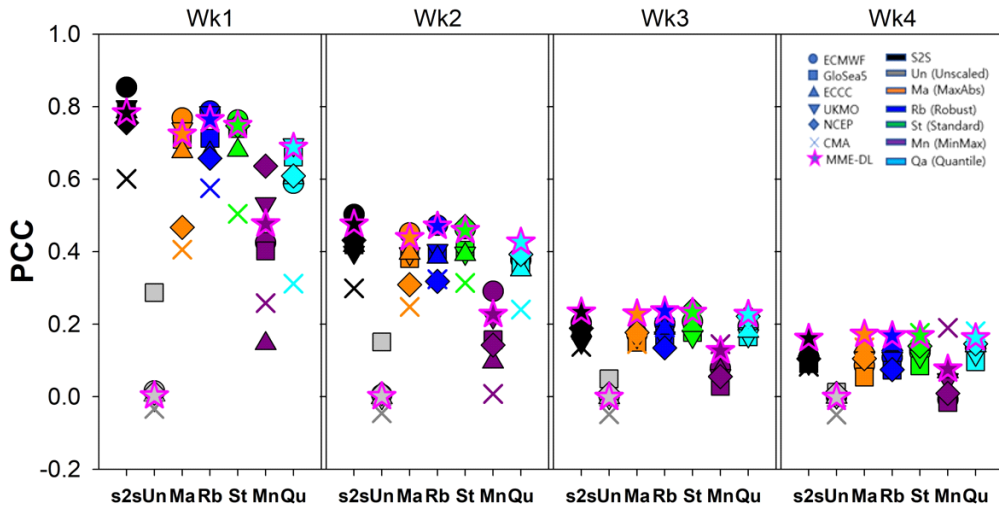


Figure 43. Comparison of prediction performance evaluation (spatial Pattern Correlation Coefficient, PCC) of base model for the daily maximum air temperature S2S training data transformed by each scaler algorithm during test period of each individual climate model and MME. Black color presents PCC of raw S2S data, and gray represents Unscaled, orange color is MaxAbs scaler, blue color is Robust scaler, green presents Standard scaler, purple color is MinMax scaler, sky-blue presents Quantile transformer result.

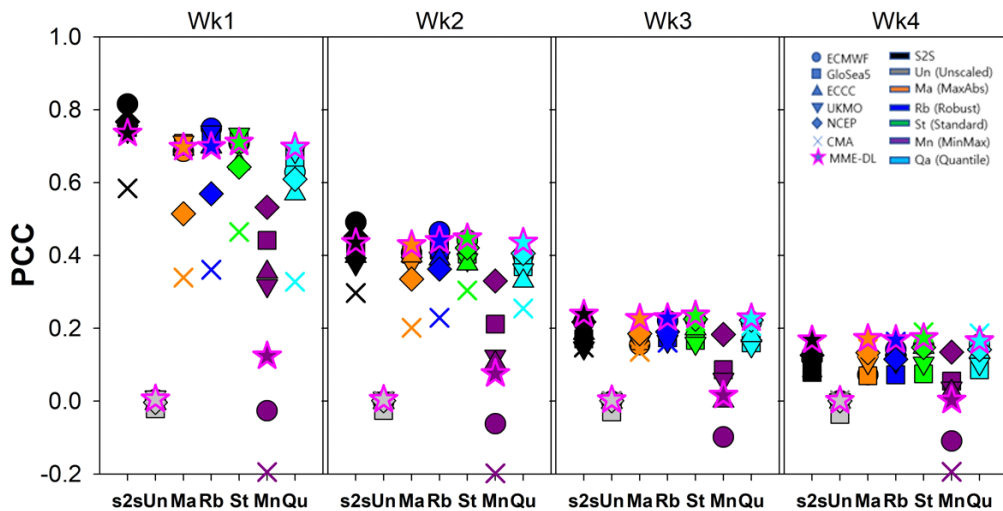


Figure 44. Comparison of PCC about prediction performance evaluation of base model for the daily minimum air temperature S2S training data transformed by each scaler algorithm during test period of each individual climate model and MME.

강수의 경우 역시 scaler를 적용하지 않은 것과 다섯 개 scaler 알고리즘에 의해 변환된 여섯 개의 개별 기후모델 및 MME의 S2S 훈련자료를 Base model에서 훈련하고, 각각의 개별

기후모델의 테스트 기간에 대해 예측된 결과를 ERA5의 일 적산 강수와의 PCC를 계산한 산포도이다(Figure 45). Base model의 강수 예측성능은 기온과 다르게 선행예측시간 1주에서부터 확실하게 나타났는데, 원시자료(검정색)의 PCC 평균 0.3과 비교했을 때 0.1 상승으로 확실하게 나타났다. 또한 Base model의 후보정 효과는 선행예측시간 4주까지도 유지되는 것을 확인할 수 있었다. 가장 뚜렷한 개선 효과를 보인 scaler도 확실하게 나타났는데, Robust Scaler(파란색)와 Standard Scaler(초록색)로 나타났다.

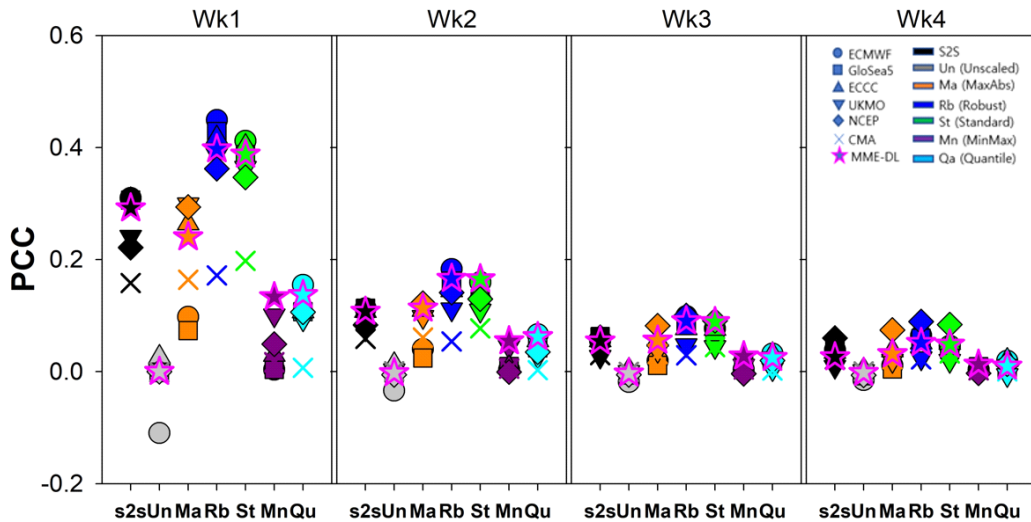


Figure 45. Comparison of PCC for prediction performance evaluation of base model for the daily total precipitation S2S training data transformed by each scaler algorithm during test period of each individual climate model and MME.

이와 같이 scaler 적용에 따른 목표 기후변수(예, 일 최고 및 최저기온, 강수)에 대한 Base model의 예측성능의 변화를 비교한 결과, Robust 및 Standard Scaler에 의해 변형된 S2S 훈련 자료로 확인되었다.

앞에서는 scaler에 대한 예측성능 변화를 평가하였다면 Figure 46에서 Figure 48은 효과가 우수하다고 판단된 scaler에 대한 모델별 예측성능을 비교한 것이다. 즉, 각각의 여섯 개 개별 기후모델 및 MME의 일 최고 및 최저기온과 일 적산 강수에 Standard Scaler를 적용하여 변형된 각각의 S2S 훈련자료에 대한 Base model의 예측 성능 지수로 계산된 PCC의 변화를 평가하기 위해 도입한 평가 지수의 상대적 변화를 계산하고 비교하였다(Figure 46에서 Figure 48).

다양한 scaler 적용에 따른 결과와 유사하게 일 최고 및 최저기온의 경우 여섯 개 개별기후모델뿐만 아니라 MME에서 선행예측시간 1-2주의 후보정 효과는 나타나지 않았다. 그러나 3-4주에서는 나타났으며 예측성이 낮다고 알려진 개별 기후모델의 후보정 효과가 큰 것으로 나타났다. 예를 들면, 예측성이 낮다고 알려진 NCEP 이나 CMA의 skill score 변화가 큰 것은 NCEP 및 CMA 원시자료의 PCC가 매우 낮는데 Base model에 의한 후보정으로 PCC가 높아졌

기 때문이다. 다만 아쉬운 것은 일 최고 및 최저기온 MME의 예측 성능 개선이 3-4주에서만 나타났다는 것이지만 이 연구의 목표가 1-2주 보다 3-4주 예측 성능을 개선하는 것이므로 연구 목표에는 도달했다고 판단된다.

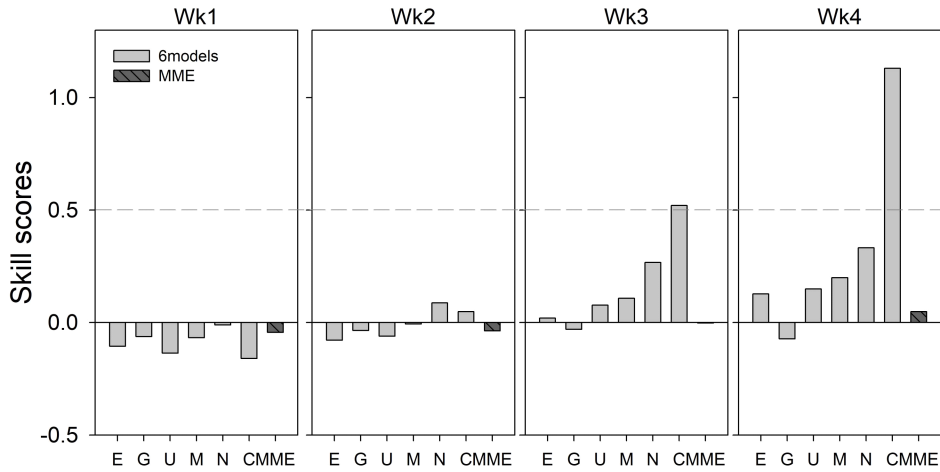


Figure 46. Comparison of skill scores for six individual climate models and MME daily maximum air temperature S2S training data transformed into standard scaler.

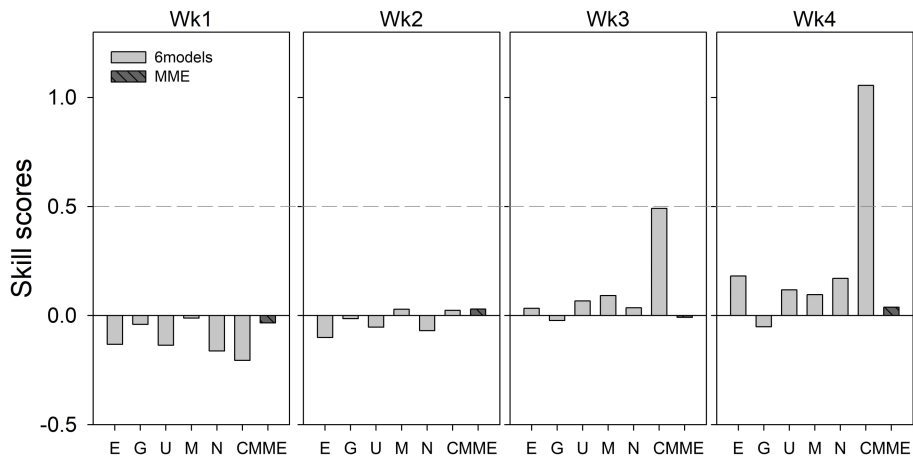


Figure 47. Comparison of skill scores for six individual climate models and MME daily minimum air temperature S2S training data transformed into standard scaler.

강수의 PCC 변화는 기온과 다르게 나타났다. 앞의 결과와 같이 강수는 여섯 개 개별 기후모델 및 MME 전체 선행예측시간에서 Base model의 후보정 효과가 모두 나타났다. 특히 사항은 ECCC가 선행예측시간 4주에서 매우 큰 PCC 변화를 보였는데, 이것은 ECCC 원시자료의

PCC가 매우 낮았는데 Base model에 의해 4주 후보정 효과가 극대화되었기 때문으로 판단된다. 또한 기온보다 강수의 Base model 후보정 효과는 MME에서 매우 높게 나타나서 기후 변수에서도 차이가 발생하였다.

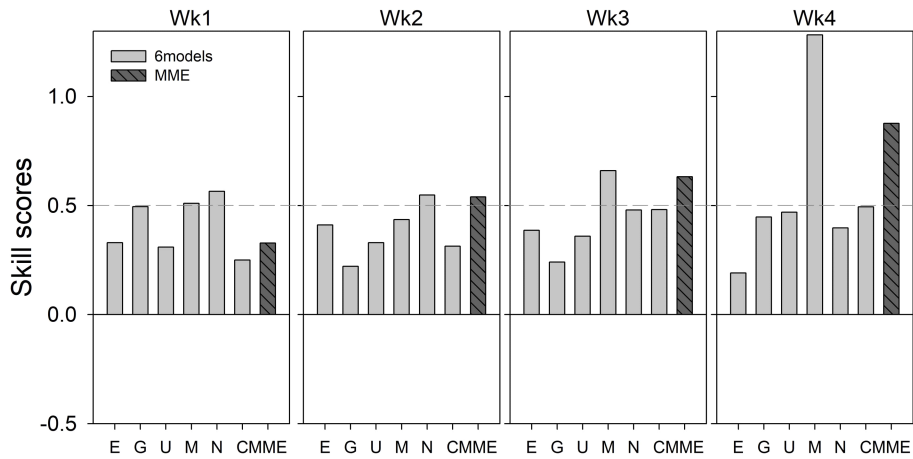


Figure 48. Comparison of skill scores for six individual climate models and MME daily total precipitation S2S training data transformed into standard scaler.

(나) 피쳐 선정에 대한 예측 성능 평가

① 주요 변수 선택 및 새 훈련자료 구성

파이썬에서 제공하는 SelectKBest module로부터 상관관계를 계산하여 가장 점수가 높은 변수에 대해 순위를 결정하여 목표 기후 변수를 예측하는데 영향을 미치는 주요 변수를 선택하였다. Figure 49은 Standard Scaler에 의해 변형된 MME 기반 일 최고 및 최저기온과 강수의 S2S 훈련자료로써 이 S2S 훈련자료에서 예측에 영향을 미친 주요 변수를 선택하였다. 일 최고 및 최저기온에 영향을 미치는 주요 변수로는 S2S 최고 및 최저기온과 평균기온을 제외하고 해면기압과 200 및 500기압에서의 지위고도, 해수면 온도 및 장파복사 플럭스, 850기압에서의 습도가 기온 예측에 영향을 미칠 수 있는 주요 변수로 선택되었다.

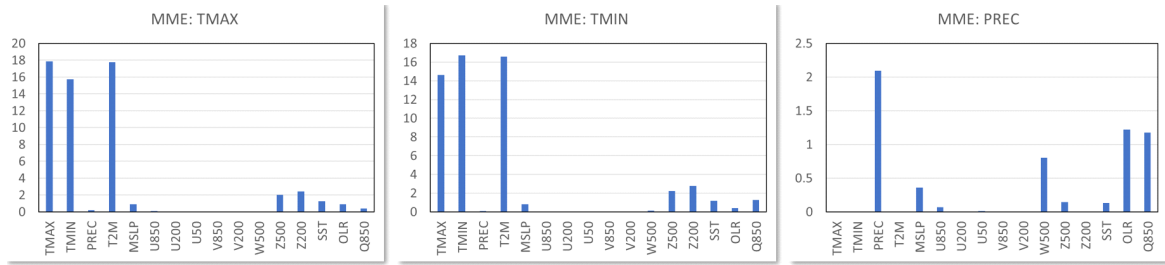


Figure 49. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of MME.

MME 강수에 영향을 미칠 수 있는 주요 변수로는 기온과 같이 S2S 강수를 제외하고, 해면 기압과 850기압에서의 동쪽 방향의 바람, 500기압에서의 지위고도와 수직 상승으로 나타났고, 기온과 같이 해수면 온도와 장파복사 플럭스 및 850기압에서의 습도가 주요 변수로 나타났다.

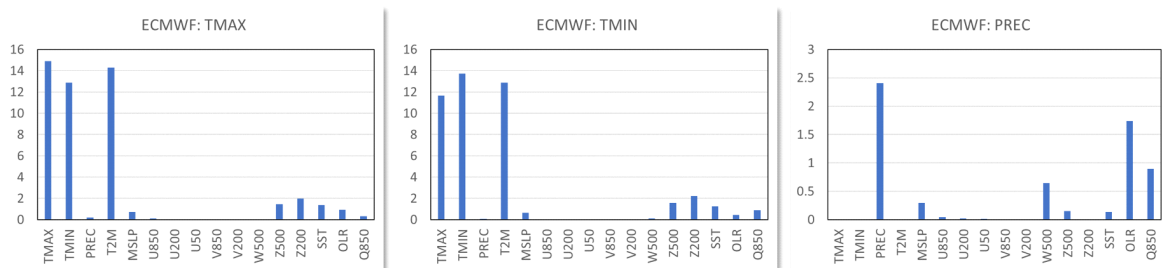


Figure 50. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of ECMWF.

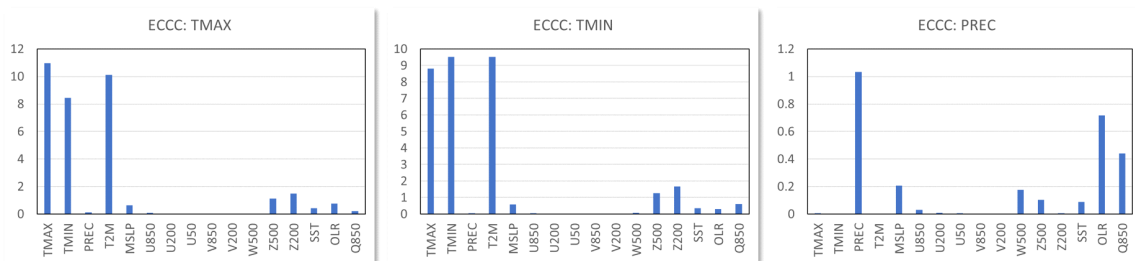


Figure 51. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of ECCC.

Figure 50에서 Figure 51은 예측 빈도가 목요일로 같은 개별 기후모델 ECMWF와 ECCC의 일 최고 및 최저기온, 강수 원시자료를 Standard Scaler에 의해 변형시켰다. 변형된 ECMWF와 ECCC의 일 최고 및 최저기온과 강수 S2S 훈련자료에서 이들의 일 최고 및 최저기온, 강수 예

측에 영향을 미친 주요 변수를 선택하였다. 일 최고 및 최저기온에 영향을 미치는 주요 변수로는 MME와 같이 S2S 최고 및 최저기온과 평균기온을 제외하고 해면기압과 200 및 500기압에서의 지위고도, 해수면 온도 및 장파복사 플럭스, 850기압에서의 습도가 기온 예측에 영향을 미칠 수 있는 주요 변수로 선택되었다. ECMWF와 ECCC 강수에 영향을 미칠 수 있는 주요 변수로는 역시 기온과 같이 S2S 강수를 제외하고, 해면기압과 850기압에서의 동쪽 방향의 바람, 500기압에서의 지위고도와 수직 상승으로 나타났고, 기온과 같이 해수면 온도와 장파복사 플럭스 및 850기압에서의 습도가 주요 변수로 나타났는데, MME 강수와 다르게 ECMWF와 ECCC 강수 예측에서는 장파복사 플럭스가 다른 주요 변수에 비교했을 때 높은 영향력을 가질 수 있는 것으로 나타났다. 반면 예측 빈도가 일별인 NCEP은 MME의 일 최고 및 최저기온, 강수에 대한 주요 변수와 매우 흡사하게 나타났다(Figure 52).

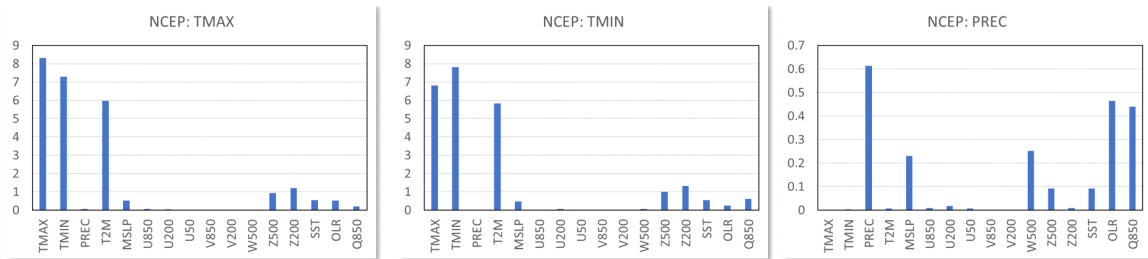


Figure 52. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of NCEP.

Figure 53은 UKMO의 일 최고 및 최저기온, 강수에 영향을 미칠 수 있는 주요 변수를 나타낸 것이다. UKMO는 다른 다섯 개 개별 기후모델과 달리 해면기압이 없으며 UKMO 원시자료를 Standard Scaler에 의해 변형시켰다. 변형된 UKMO의 일 최고 및 최저기온과 강수 S2S 훈련자료에서 이들의 일 최고 및 최저기온, 강수 예측에 영향을 미친 주요 변수를 선택하기 위해 상관관계를 계산한 결과, 일 최고 및 최저기온에 영향을 미치는 주요 변수로는 MME 및 ECMWF와 ECCC 기후모델과 같이 S2S 최고 및 최저기온과 평균기온을 제외하고 200 및 500기압에서의 지위고도, 해수면 온도 및 장파복사 플럭스, 850기압에서의 습도가 기온 예측에 영향을 미칠 수 있는 주요 변수로 선택되었다. UKMO 강수에 영향을 미칠 수 있는 주요 변수는 해면기압이 없고, 850기압에서 동쪽 방향의 바람, 500기압에서의 지위고도와 수직 상승으로 나타났고, 기온과 같이 해수면 온도와 장파복사 플럭스 및 850기압에서의 습도가 주요 변수로 나타났다. 역시 ECMWF와 ECCC 강수 예측과 같이 장파복사 플럭스가 다른 주요 변수와 비교했을 때 높은 영향력을 가질 수 있는 것으로 나타났다.

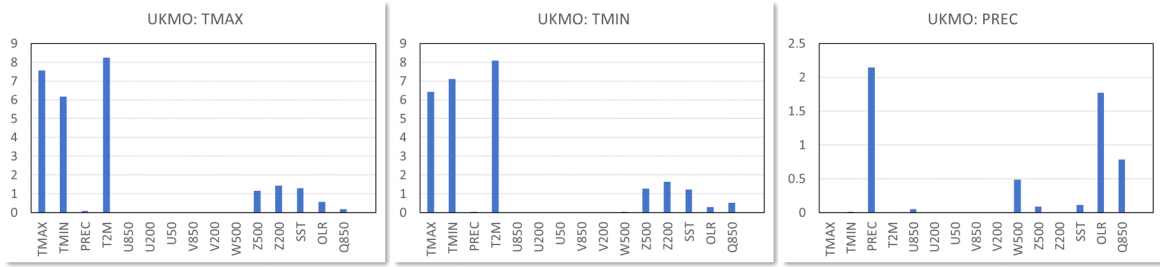


Figure 53. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of UKMO.

Figure 54는 GloSea5의 일 최고 및 최저기온, 강수에 영향을 미칠 수 있는 주요 변수를 나타낸 것이다. GloSea5도 다른 다섯 개 개별 기후모델과 달리 해수면 온도, 장파복사 플럭스, 850기압에서의 습도 등 세 개의 변수가 없지만 GloSea5 원시자료에 Standard Scaler를 적용하여 변형시켰다. 변형된 GloSea5의 일 최고 및 최저기온에 영향을 미치는 주요 변수로는 MME 및 ECMWF와 ECCO 기후모델과 같이 S2S 최고 및 최저기온과 평균기온을 제외하고 해면기압과 200 및 500기압에서의 지위고도로 나타났다. 해수면 온도 및 장파복사 플럭스나 850기압은 없기 때문이다. GloSea5 강수에 영향을 미칠 수 있는 주요 변수는 역시 다른 개별 기후모델과 마찬가지로 해면기압과 850기압에서 동쪽 방향의 바람, 500기압에서의 지위고도와 수직 상승으로 나타났다.

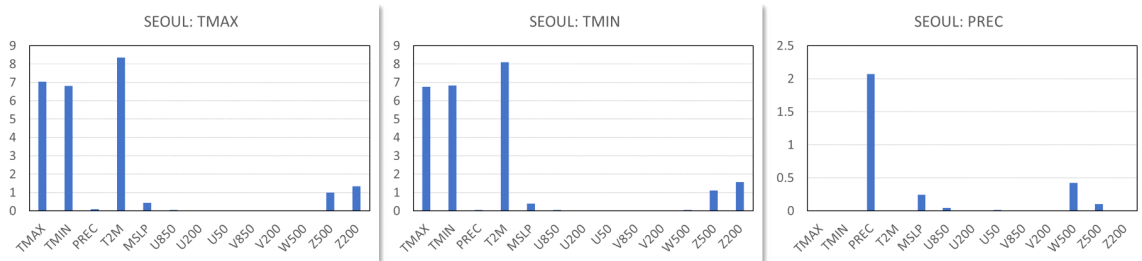


Figure 54. Explanation of the main variables extracted by feature selection algorithm for the daily maximum and minimum temperature, and precipitation of GloSea5.

Figure 55는 예측성이 낮다고 알려진 CMA의 일 최고 및 최저기온, 강수에 영향을 미칠 수 있는 주요 변수를 나타낸 것이다. CMA는 NCEP과 같이 예측 빈도가 일별이며 CMA 원시자료에 Standard Scaler를 적용하여 변형시켰다. 변형된 CMA의 일 최고 및 최저기온에 영향을 미치는 주요 변수는 앞서 언급한 개별 기후모델 및 MME와 같이 S2S 최고 및 최저기온과 평균기온을 제외하고 해면기압과 200 및 500기압에서의 지위고도, 해수면 온도, 장파복사 플럭스 및 850기압에서의 습도로 나타났다. CMA 강수에 영향을 미칠 수 있는 주요 변수는 앞서 언급한 다른 개별 기후모델 및 MME와 다소 다르게 나타났는데, 해면기압과 850기압에서 동쪽 방

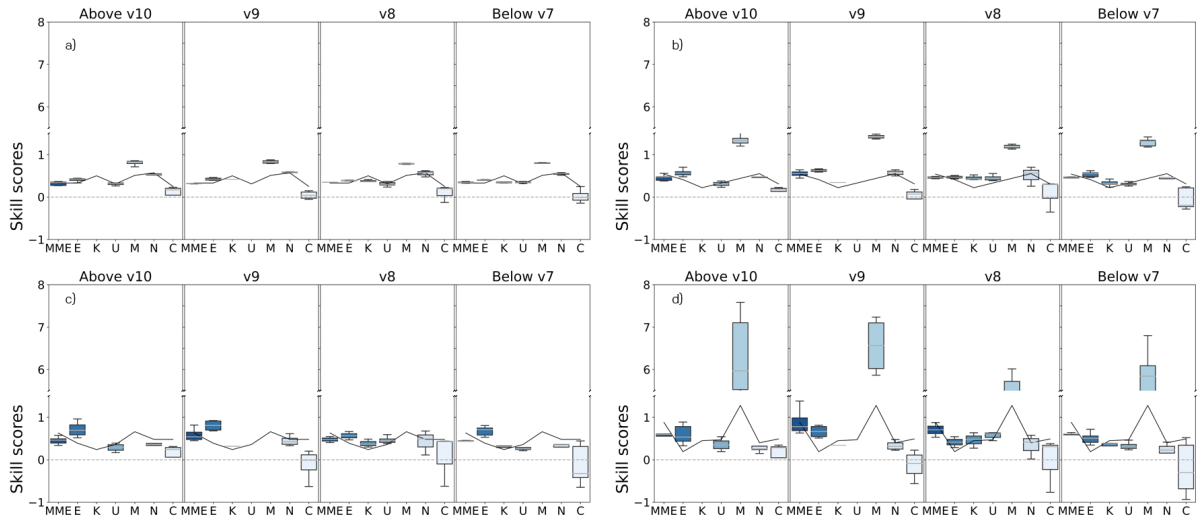


Figure 56. Skill score evaluation according to four groups for major variables for 6 individual climate models and MME daily total precipitation prediction.

Figure 56은 여섯 개 개별 기후모델 및 MME의 강수 예측에 대해 이 연구에서 설정한 4개 주요 변수 선택에 따른 그룹에 대한 PCC의 변화를 선행예측시간에 따라 평가한 것이다. 강수 예측에서 주요 변수의 변화는 예측 성능 변화에 큰 영향을 미치지 않는 것으로 나타났다. 예를 들면, 선행예측시간 1-2주에서 기준 주요 변수에 대한 훈련자료뿐만 아니라 랜덤하게 주요 변수를 +/- 한 경우에도 예측 성능은 검은 점선, 즉 원시자료의 PCC보다 항상 높게 유지되었다. 다만 ECCG는 주요 변수 선택의 변화에 대해 선행예측시간 4주에서 변화가 많아서 ECCG의 4주 예측성능의 불확실성이 나타났다.

6개 개별 기후모델 및 MME의 일 최고 및 최저기온 예측에 대해 이 연구에서 설정한 4개 주요 변수 선택에 따른 그룹에 대한 PCC의 변화를 선행예측시간에 따라 평가한 결과(Figure 57 to Figure 58), 일 최고 및 최저기온 예측에서 주요 변수의 변화는 선행예측시간 1주에서 원시자료의 PCC에 대해 Base model 후보정 기법에 의해서도 일 최고 및 최저기온의 예측 성능을 높이지 못했다. 또한, 강수 예측은 주요 변수의 변화가 예측성능 변화에 영향을 미치지 못했는데 기온의 경우는 주요 변수 선택 7개 미만에 대해서는 예측성능이 저하될 수 있는 것으로 나타났다.

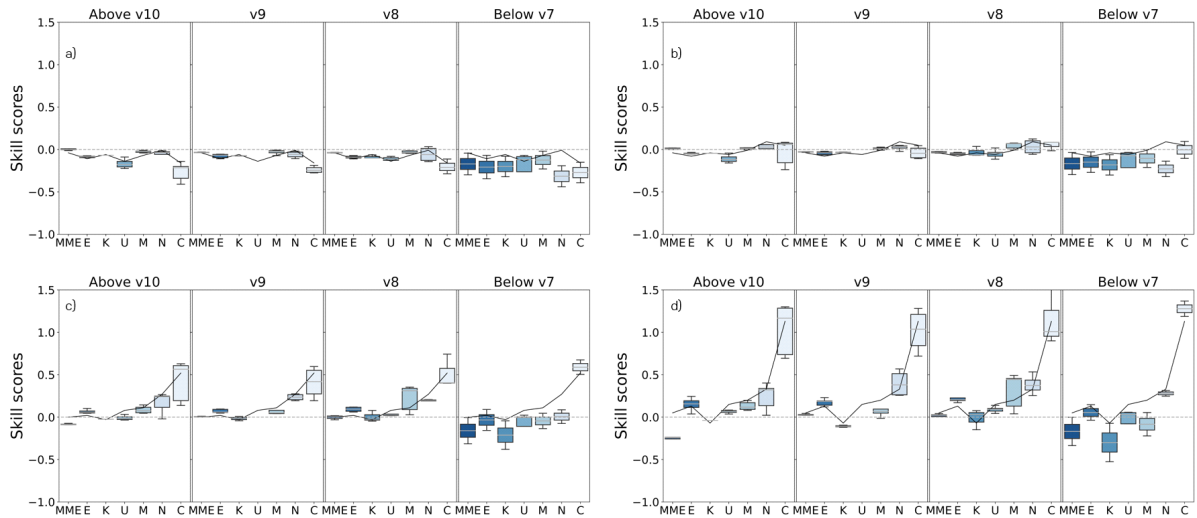


Figure 57. Skill score evaluation according to four groups for major variables for 6 individual climate models and MME daily maximum air temperature prediction.

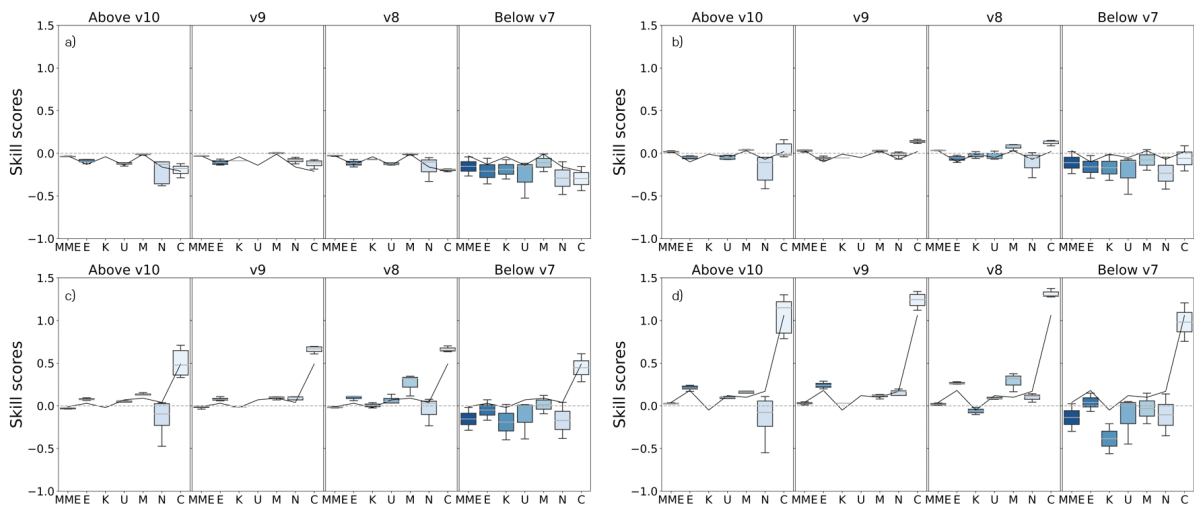


Figure 58. Skill score evaluation according to four groups for major variables for 6 individual climate models and MME daily minimum air temperature prediction.

전반적으로 특성 선택에 의한 주요 변수의 변화는 강수 예측성능 변화에는 영향을 미치지 못하는 것으로 나타났으며 기온은 주요 변수 7개 미만이 일 최고 및 최저기온의 예측성능을 저하시킬 수 있는 것으로 나타났다.

③ 주요 변수 변화에 따른 가중치 변화

Figure 59는 테스트 기간(2011-2014) 중 2013년 예측 날짜 1월 4일에서 3주로부터 7일에 대한 재분석 ERA5의 일 최고기온 편차와 Base model에 의해 예측된 일 최고기온의 Class

Activation Map(일종의 돌출 맵)을 나타낸 것인데, CAM은 특성 선택에서 주요 변수 선택에 따른 4개 그룹에 대한 CAM이다. CAM은 노란색에서 빨간색으로 갈수록 중요한 것이고 파란색으로 갈수록 예측에 미친 영향력이 작다는 것을 의미한다. 모든 변수 즉 16개 변수를 모두 입력했을 때는 노란색과 빨간색 지역이 ERA5의 일 최고기온 편차 차이가 높은 지역들과 유사해지는 것을 확인할 수 있다. 그러나 주요 변수 선택이 6개 미만에서는 노란색과 빨간색 지역이 ERA5의 편차 차이가 높은 지역과 달라지는 것을 확인할 수 있다. 다시 말해서 딥러닝 훈련을 위한 Base model이 일 최고기온의 편차 차이가 높은 지역을 보고 있다고 해석할 수 있다.

최근 CAM 혹은 Saliency map의 정성적 해석보다는 saliency map 혹은 CAM에 대한 정량적 해석을 시도한 연구들이 많아지고 있다(Alqaraawi et al., 2020; Fahim et al., 2022). 후속 연구에서 CAM 혹은 saliency map에서 영향을 미치는 픽셀이 보이는 것과 보이지 않는 것에 대한 정량적 해석 기법의 개발이 필요할 것이다.

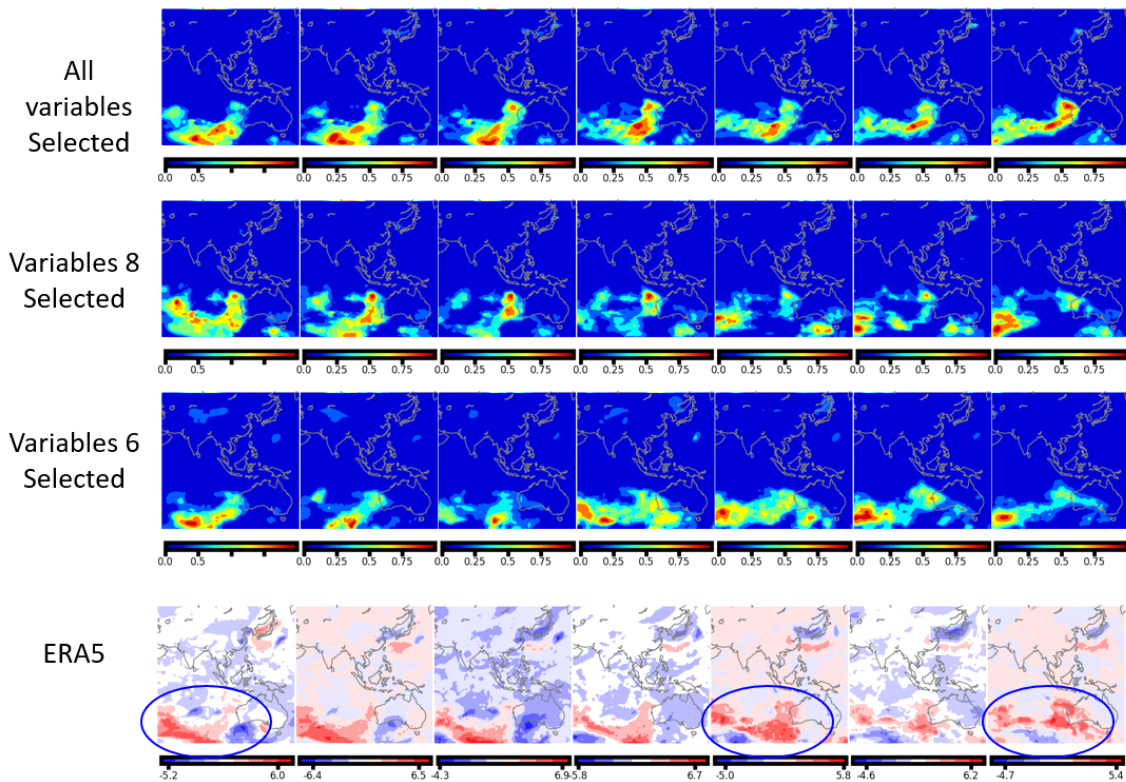


Figure 59. Comparison of Class Activation Maps according to four groups for major variables for MME daily maximum air temperature prediction.

(다) MME 기법의 훈련 기간별 성능 평가

① MME 예측 성능 평가

Figure 60은 MME 기반의 일 최고 및 최저기온과 강수에 대해 Base model의 예측성능을 PCC로 평가한 결과이다. 회색 막대그래프는 ECWFM의 S2S 원시자료, 즉 딥러닝이 적용되지

않은 자료에 대한 ERA5와의 PCC이며, 파란색 막대 그래프는 여섯 개 개별 기후모델의 S2S 원시자료를 모두 결합한 MME, 즉 MME 기반 S2S 원시자료에 대한 ERA5와의 PCC이다(MME). 오렌지색 막대 그래프는 MME 기반의 S2S 원시자료를 Standard scaler로 변형시키고 Base model에 의해 예측된 결과에 대한 ERA5와의 PCC이며(MME-DL) 노란색 막대 그래프는 여섯 개 개별 기후모델에 Standard scaler로 변형하여 Base model에서 훈련을 시켜서 예측한 결과들을 결합한 것에 대한 ERA5와의 PCC이다(DL-MME). 파란색 막대 그래프가 어떤 보정도 적용하지 않은 MME 훈련자료에 대한 PCC인데, 일 강수뿐만 아니라 일 최고 및 최저기온에서 선행예측시간 2주의 PCC는 ECMWF(회색)보다 높은 것을 확인할 수 있다. 이것은 여러 기후모델을 한 번에 고려하여 예측 신뢰성을 증가시킬 수 있다는 MME 기법의 장점 (Monhart et al., 2018; Vitart and Robertson, 2018; Weyn et al., 2021; Weigel et al., 2008)에서 비롯된 것으로 판단된다. 따라서, 딥러닝 훈련 모델 적용 전에 이미 MME 기법을 통해 S2S 예측 강수, 일 최고기온 및 최저기온의 PCC를 1차 개선시킬 수 있다는 것을 확인하였다.

또한, 오렌지색 막대 그래프를 보면 일 최고 및 최저기온에서 선행예측시간 1주를 제외하고 강수는 전체 선행예측시간 전체의 PCC를 개선시킬 수 있으며, 일 최고 및 최저기온은 선행예측시간 2주 이후의 PCC를 개선시킬 수 있다는 것을 확인하였다. 다만, 아쉬운 점은 개별 기후모델에 Base model 훈련 후 MME 한 경우의 PCC가 다소 낮은 것인데, 이 경우도 선행예측시간 1주를 제외하면 후보정하지 않은 MME 및 ECMWF보다 높기 때문에 딥러닝 훈련에 의한 후보정을 통해 S2S의 선행예측시간 2-4주 예측성능을 개선할 수 있을 것으로 판단된다.

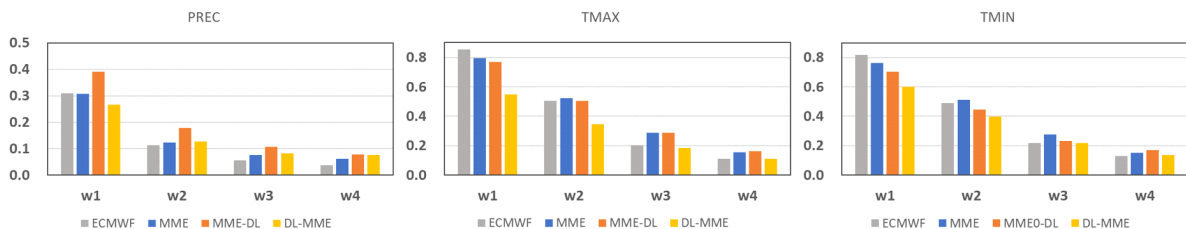


Figure 60. Comparison of differences in prediction performance for MMEs with different training periods. Left represents the PCC of MME based on the daily total precipitation, center explains the PCC of MME based on the daily maximum air temperature, and right presents the PCC for MME based on the daily minimum air temperature.

다. 준지도학습 적용을 위한 환경구축

(1) 합성곱신경망(CNN) 모델 훈련 및 검증

(가) 일자료 시간적 조합(semi-monthly)을 활용한 자료강화: 모든 월 사용(Month-Agnostic)

자료 전처리 방법으로 표준화 및 정규화를 테스트하였는데, 결과에 큰 차이가 없어 표준화 방법을 사용한 경우의 결과만을 포함하였다(Figures 61-64). 검증 손실(validation loss)을 모니터링하며 최적의 Epoch을 정하였다(그래프에 표시된 파란색/붉은색 점). Semi-monthly 방법의 경우 validation loss를 보면 자료강화하지 않은 경우(파란선) epoch이 증가해도 loss가 감소하지 않으며 자료강화한 경우(주황선) loss가 오히려 증가함을 볼 수 있다. Train loss를 살펴보면 충분히 학습되지 않은 epoch에서 validation loss 정체/증가로 인해 학습이 중단되었음을 볼 수 있다. 폴드별 모델의 성능은 검증 정확도(validation accuracy)로 평가할 수 있는데, validation accuracy와 train accuracy를 확인하면 충분히 학습되지 않은 상황에서 학습이 중단되었으므로 best epoch에서 validation accuracy와 train accuracy 모두 비교적 낮게 나타난다. 목표 월만 사용하는 경우 Semi-monthly 자료강화 방법 적용 시 지속적인 에러 발생으로 결과 비교가 불가능하여 제외하였다.

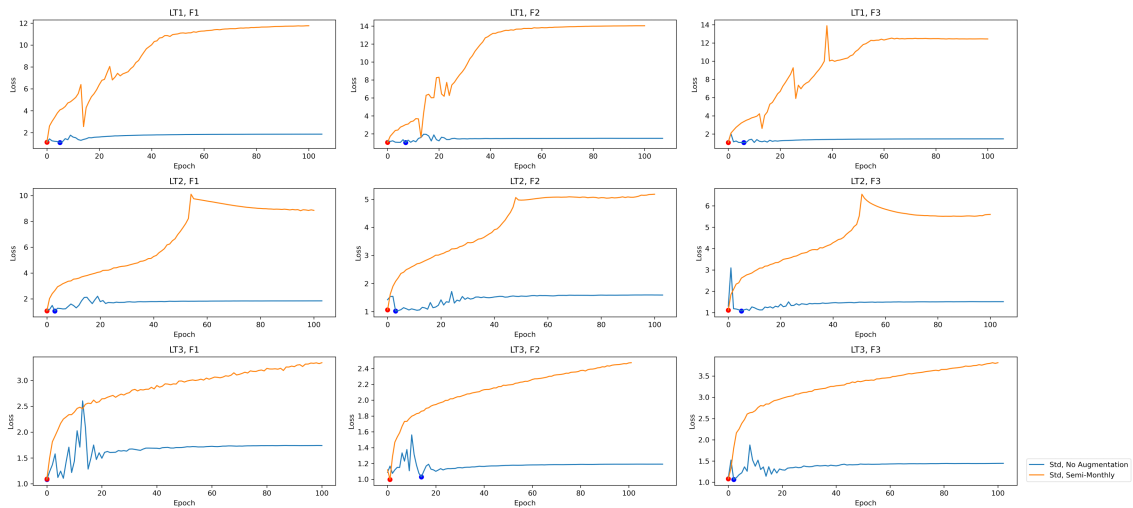


Figure 61. Comparison of validation loss (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.

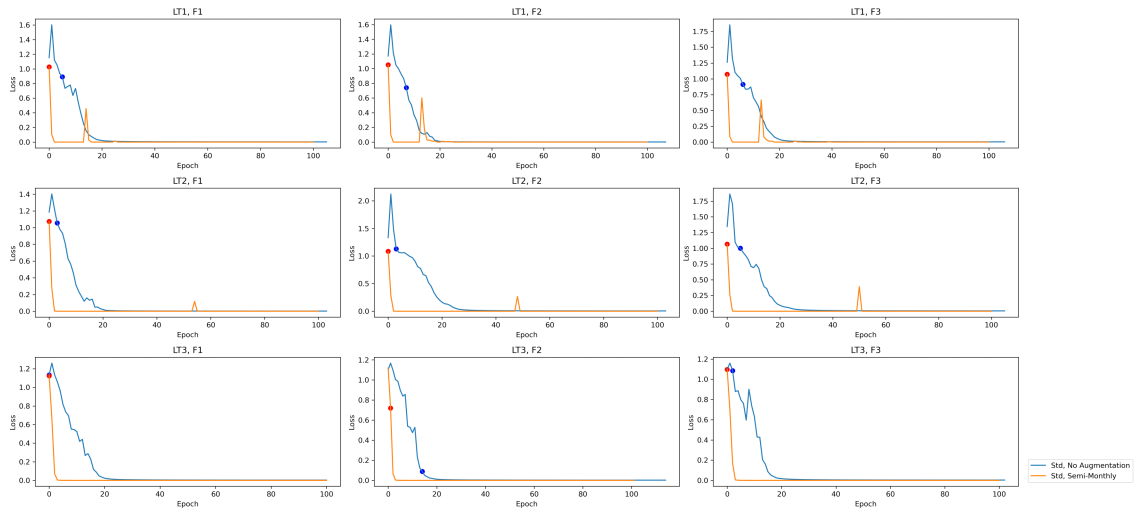


Figure 62. Comparison of train loss (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.

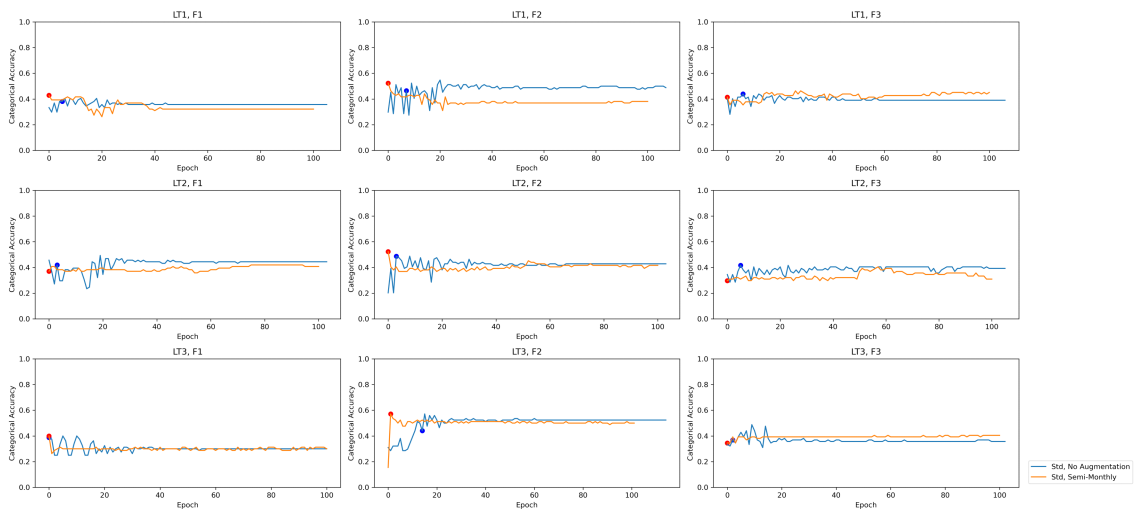


Figure 63. Comparison of validation accuracy (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.

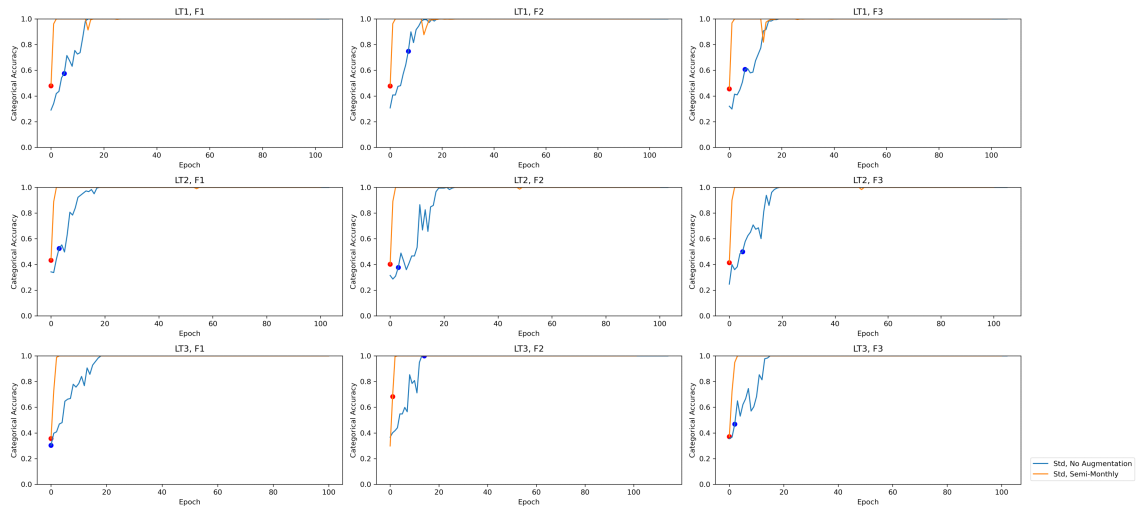


Figure 64. Comparison of train accuracy (Month-Agnostic): Augmentation (Semi-monthly) vs. No Augmentation.

(나) 월자료 공간적 재조합(Cutmix 응용)을 활용한 자료강화: 모든 월을 사용하는 경우

Cutmix 방법의 경우 validation loss를 보면 자료강화 하지 않은 경우(파란선) epoch 증가함에 따라 나중에는 loss가 증가하는데 자료강화한 경우(주황선) loss가 상대적으로 작으며 또한 epoch이 커짐에 따라 일부 경우를 제외하고 크게 증가하지 않음을 볼 수 있다(Figure 65). Training loss를 살펴보면 자료강화하지 않은 경우에는 충분히 학습이 되지 않은 epoch에서 validation loss 정체/증가로 인해 학습이 중단되었으나, 자료강화된 경우에는 학습이 중단되기 전에 loss가 비교적 많이 줄어들었음을 볼 수 있다(Figure 66). Validation accuracy와 train accuracy를 확인하면 자료강화하지 않은 경우에는 충분히 학습되지 않은 상황에서 학습이 중단되었으나 자료강화한 경우 학습이 비교적 충실히 이루어져 best epoch에서 validation accuracy도 비교적 개선되었음을 볼 수 있다(Figures 67-68).

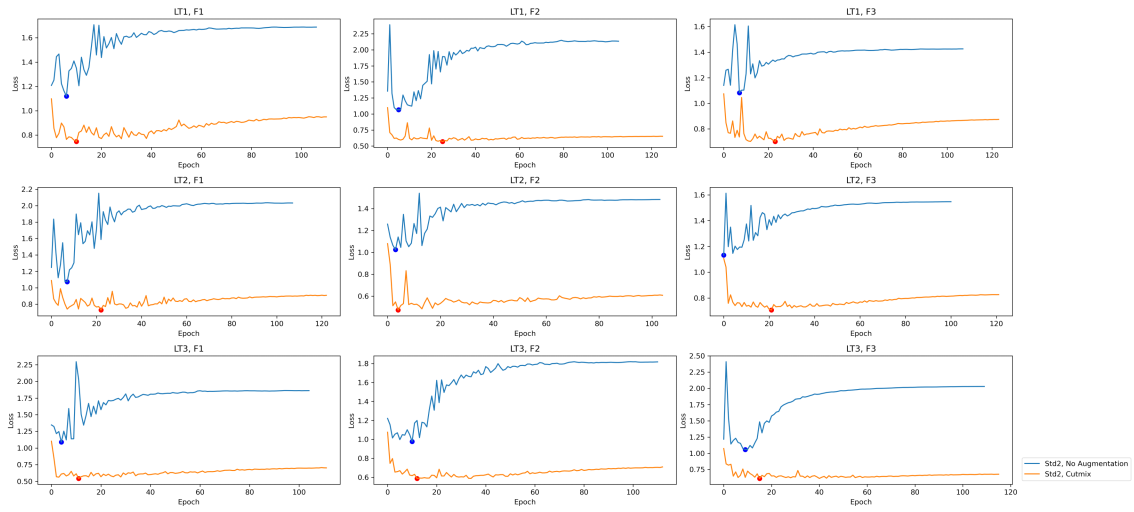


Figure 65. Comparison of validation loss (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.

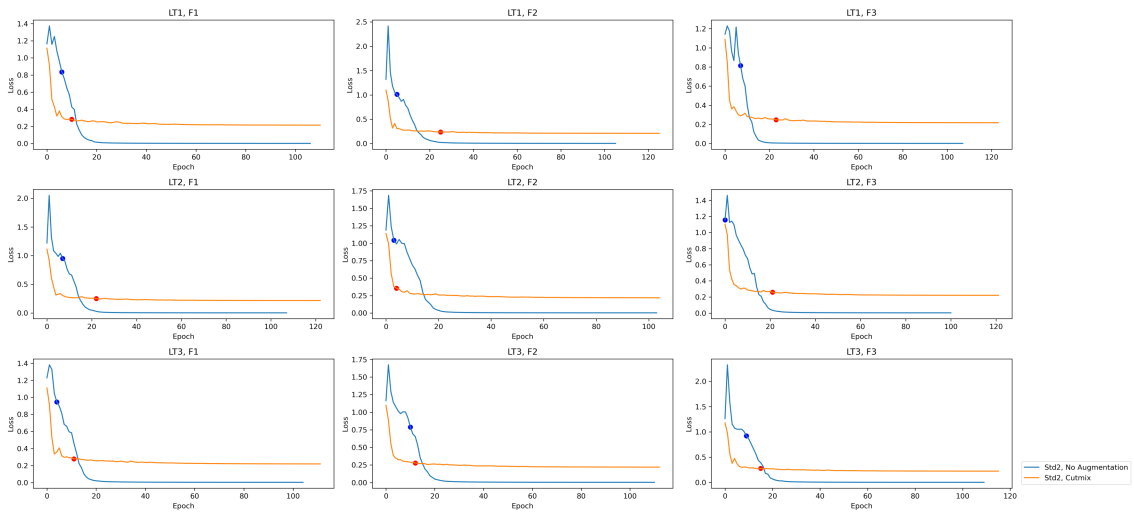


Figure 66. Comparison of train loss (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.

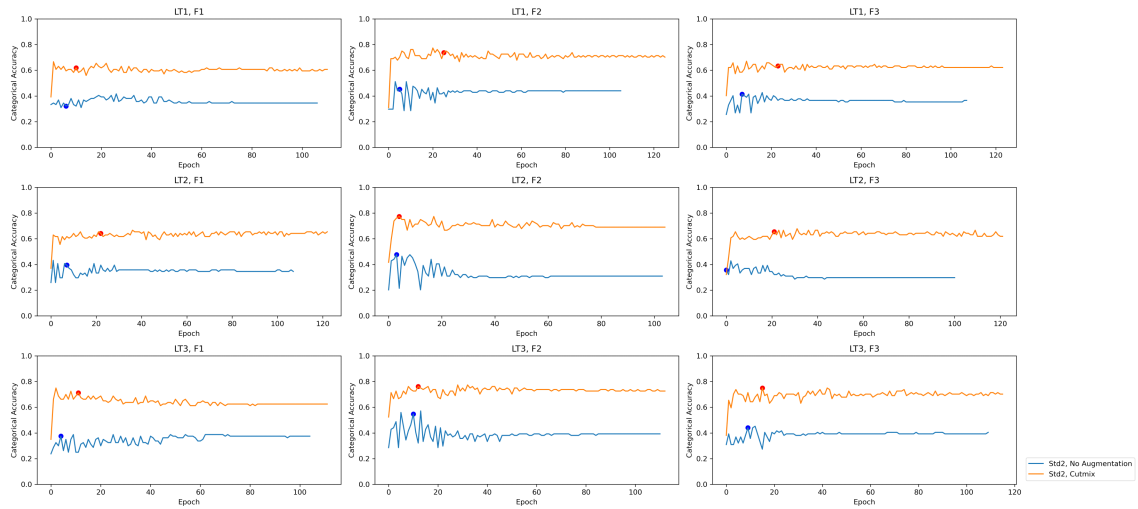


Figure 67. Comparison of validation accuracy (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.

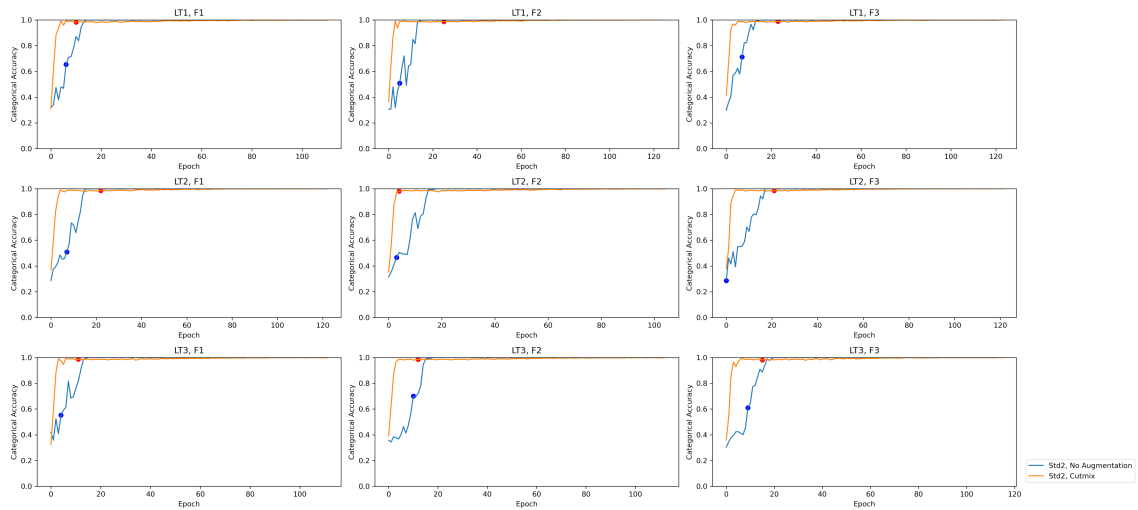


Figure 68. Comparison of train accuracy (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.

3개의 fold에 대한 validation accuracy를 box plot으로 나타내어 방법 간에 비교하면 다음 그림과 같다(Figure 69). Semi-monthly 방법의 경우 개선여부가 일관적으로 드러나지 않는데 반하여 Cutmix 방법으로 모든 월을 사용한 경우 validation accuracy는 명확하게 개선되었음을 알 수 있다. 자료 전처리 기법에 있어 표준화와 정규화 방법 간의 결과 차이는 뚜렷하지 않으나 LT3 모델 F3의 경우 학습이 제대로 이루어지지 않아 Cutmix (표준화) 방법이 더 적합한 것으로 여겨진다.

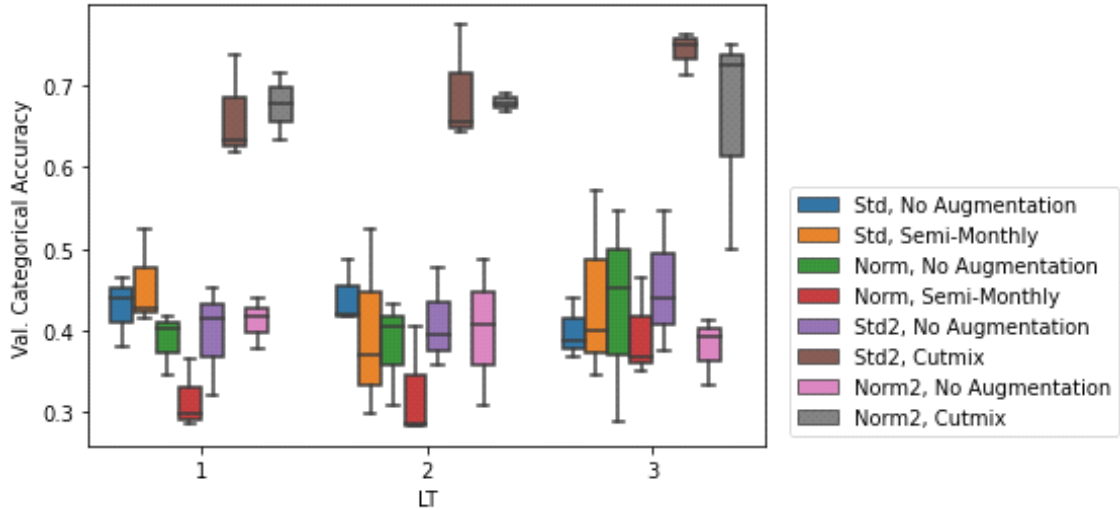


Figure 69. Comparison of validation accuracy (Month-Agnostic) between data pre-processing and augmentation methods.

(다) 월자료 공간적 재조합(Cutmix 응용)을 활용한 자료강화: 목표 월만 사용(Target Month Only)

목표 월만 사용하여 Cutmix 자료강화한 경우(주황선) 대부분의 경우 loss가 상대적으로 작지만 비슷한 경우도 존재하였다(Figure 70). Training loss를 살펴보면 자료강화하지 않은 경우와 자료강화한 경우 모두 많은 경우에 충분히 학습이 되지 않은 epoch에서 validation loss 정체/증가로 인해 학습이 중단되었음을 볼 수 있었으며(Figure 71) validation accuracy와 train accuracy에서도 자료강화하지 않은 경우와 자료강화한 경우 모두 많은 경우에 충분히 학습되지 않은 상황에서 학습이 중단되었음을 볼 수 있었다(Figures 72-73).

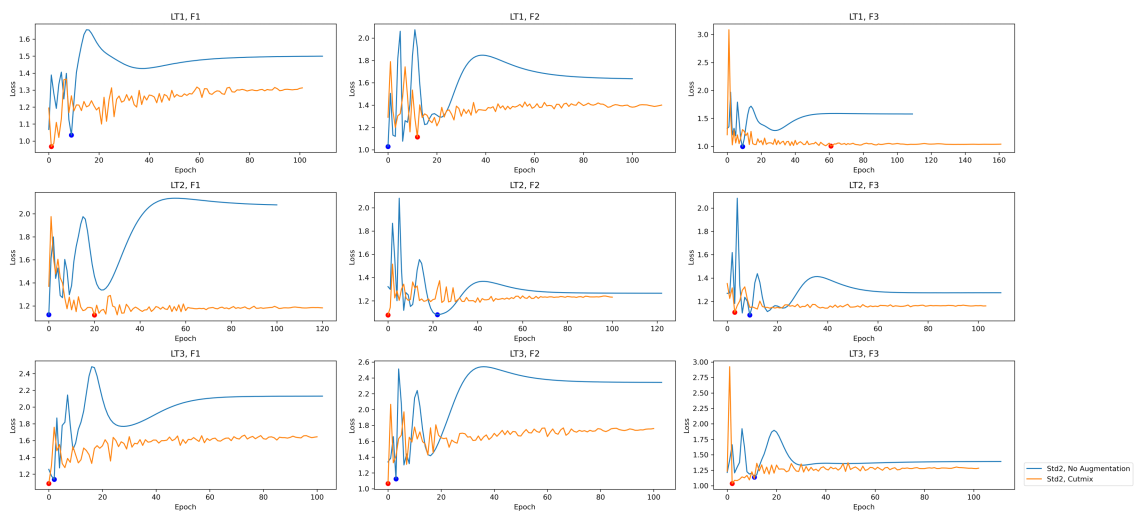


Figure 70. Comparison of validation loss (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.

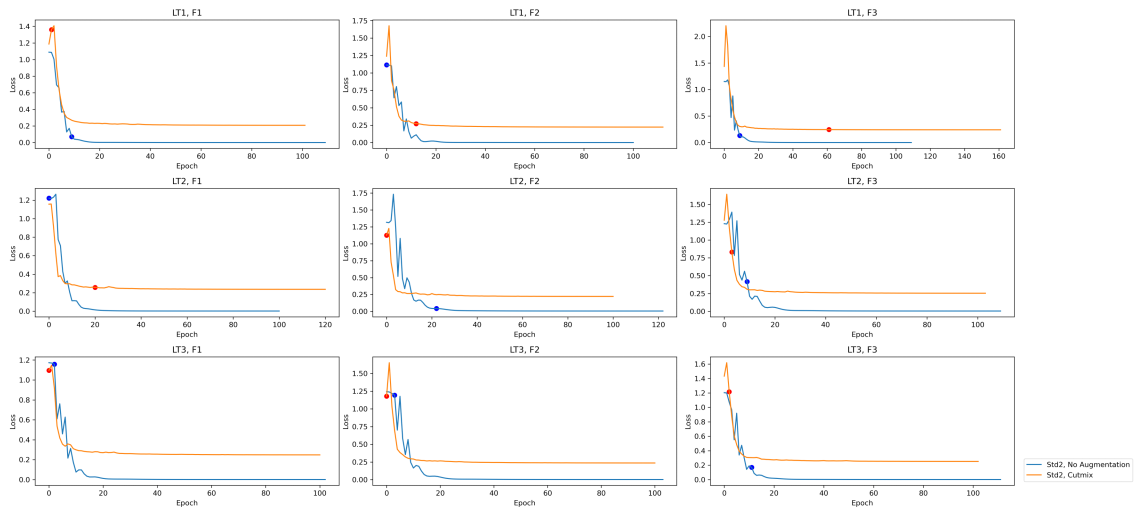


Figure 71. Comparison of train loss (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.

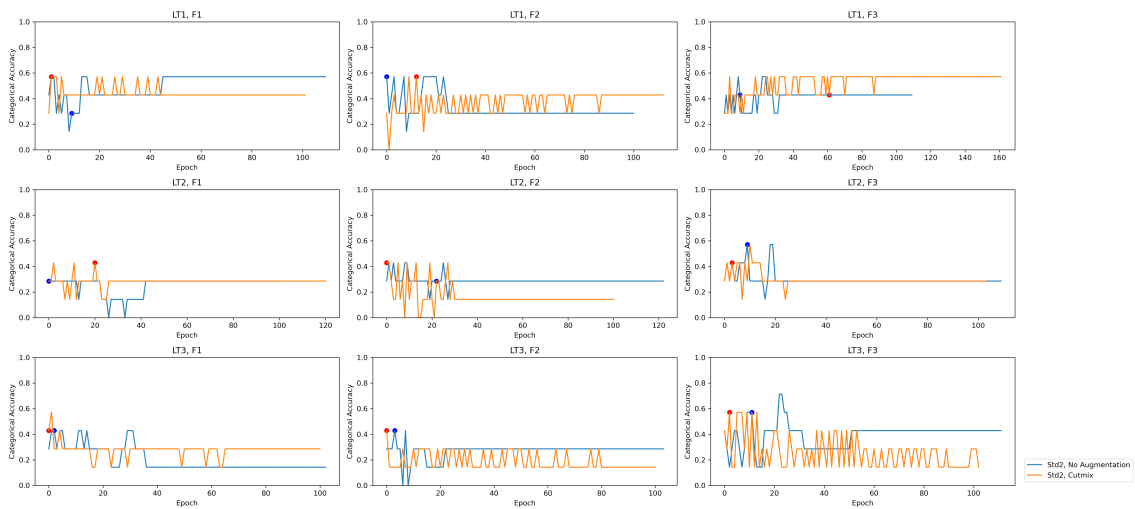


Figure 72. Comparison of validation accuracy (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.

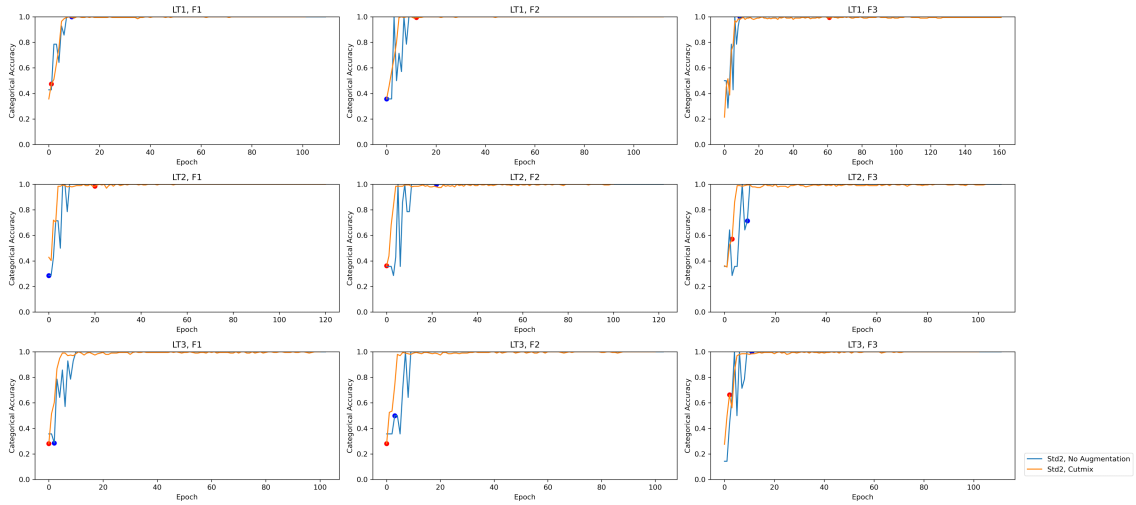


Figure 73. Comparison of train accuracy (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.

(2) 합성곱신경망(CNN) 모델 지도학습 성능평가

(가) 3D 합성곱신경망모델 최종 성능 평가

모델의 최종 검증을 위해 따로 남겨놓은 테스트셋(Table 4)에 대한 평가는 앞서 개발한 3개의 폴드에 대한 모델을 각각 이용한 결과와 그 앙상블의 성능 평가로 수행하였다(Figures 74-77). 앙상블(아래 그림에서 색깔점으로 표시)은 PC (total accuracy)와 HSS의 경우에는 투표결과(voting, 다수결)의 성능 평가로 얻었으며 AUC와 RPSS의 경우에는 3개 모델에서 각각 얻은 삼분위 확률을 각각의 카테고리에 대해 평균하여 산정하였다. 테스트셋 평가 결과, 6월(LT1)과 7월(LT2)의 경우 Cutmix (표준화) 앙상블 결과를 살펴보면 PC, HSS, AUC, RPSS에서 모두 자료강화의 개선 효과를 확인할 수 있었다. 다만 8월(LT3)의 경우는 개선되지 않았다.

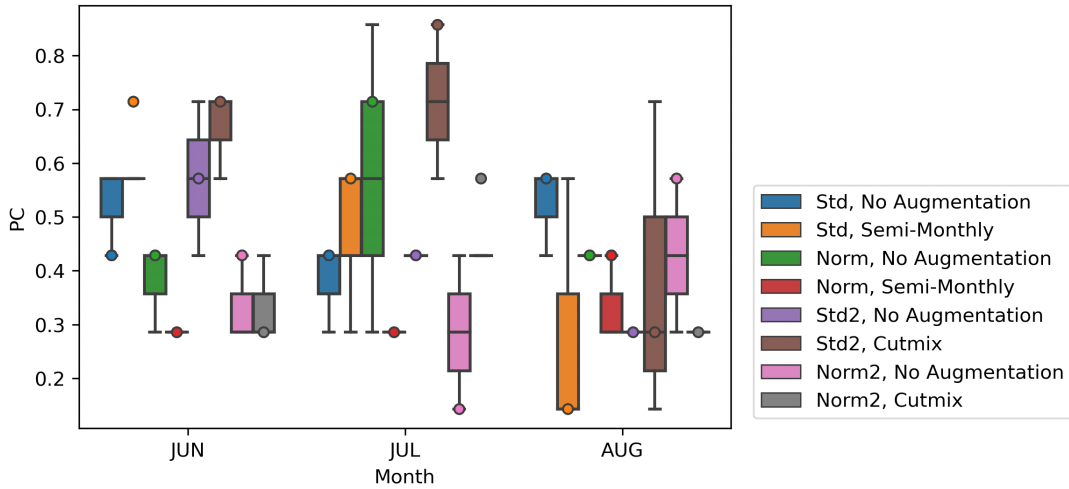


Figure 74. Proportion Correct (PC) between data pre-processing and augmentation methods.

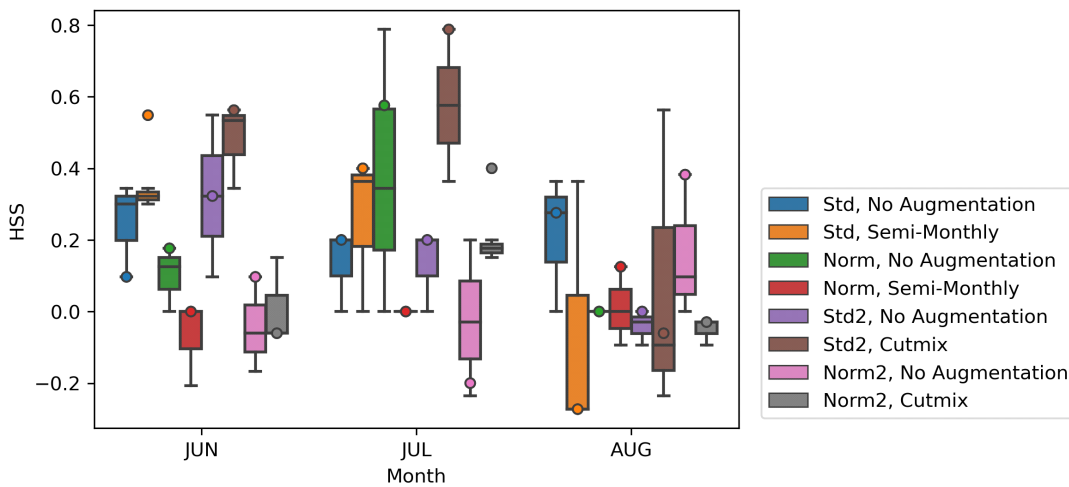


Figure 75. Heidke Skill Score (HSS) between data pre-processing and augmentation methods.

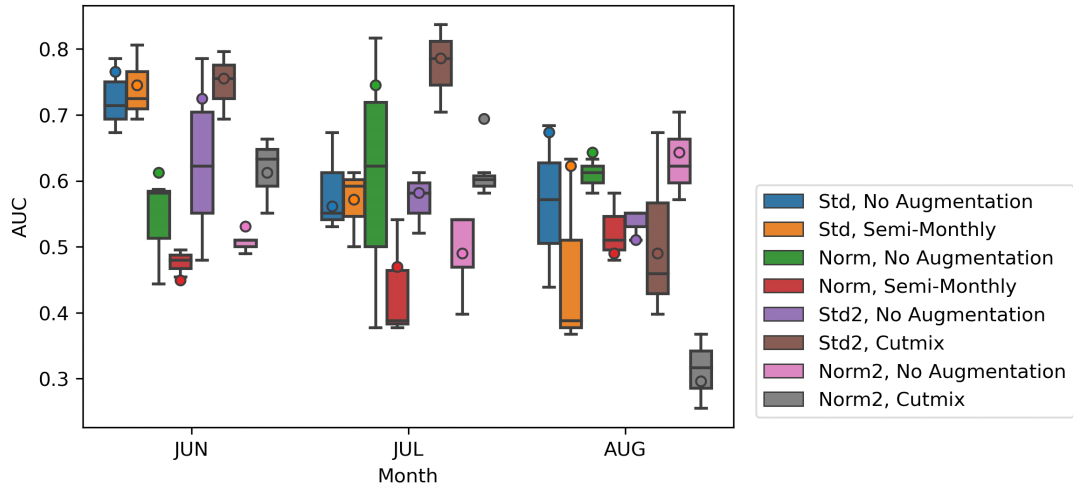


Figure 76. Area Under the ROC Curve (AUC) between data pre-processing and augmentation methods.

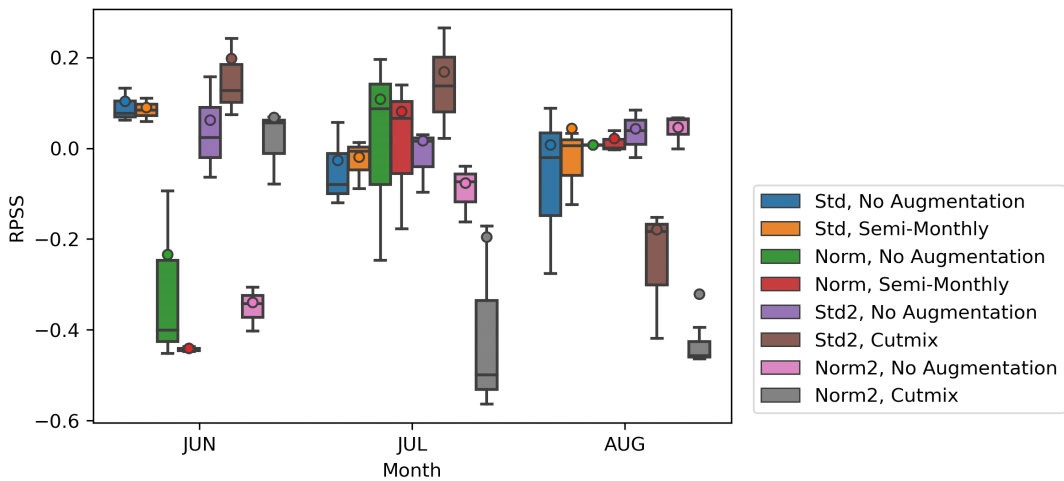


Figure 77. Ranked Probability Skill Score (RPSS) between data pre-processing and augmentation methods.

(나) 활성맵 변화 검토

Cutmix 방법에 대해 월별, 폴드별로 활성맵을 그려 입력자료에서 중요하게 기여한 지리적 위치를 도출하여 자료강화 전후를 비교하였다. 이때 활성맵은 완전연결층 직전의 공간정보를 이용하여 도출하였다. Cutmix 자료강화 통해 개선된 사례 중 삼분위 카테고리 참값이 NN이 아니고 2018년 이후인 경우의 활성화맵을 살펴보았다.

Table 4. Test years and their tercile categories for June, July and August.

MONTH	TEST YEARS	TERCILE CATEGORY
JUN	2018 1998 2000 2007 2005 2021 1995	AN BN NN NN AN AN BN
JUL	1999 2021 2020 2018 1993 1995 2005	BN AN BN AN BN NN NN
AUG	2014 2020 1999 1997 1998 2017 1995	BN AN BN NN NN NN AN

- 2018년 7월(폴드3 모델): BN로 낮게 예측 → AN로 개선

2018년 7월 기후분석토의 보고서(기후분석과 작성)의 내용을 살펴보면, 고온의 원인을 다음과 같이 기술하였다: “우리나라 대기 상하층에 정체된 고기압으로 인해 하강기류에 의한 기온 상승과 일사량 증가로 고온이 지속되면서 폭염이 발생함. 티베트 고기압과 북태평양 고기압 세력이 우리나라까지 확장, 열대 서태평양 대류활동이 강해지면서 중위도에 하강기류가 발달, 중위도 전반에 걸쳐 파동형태로 늘어난 고기압 편차 발생, 중위도 제트기류가 약화되며 대기 상층의 흐름 정체 등이 폭염 발생과 지속에 기여함.” 자료강화 후 활성화맵을 살펴보면 북태평양 지역이 활성화되었음을 볼 수 있다(Figure 78).

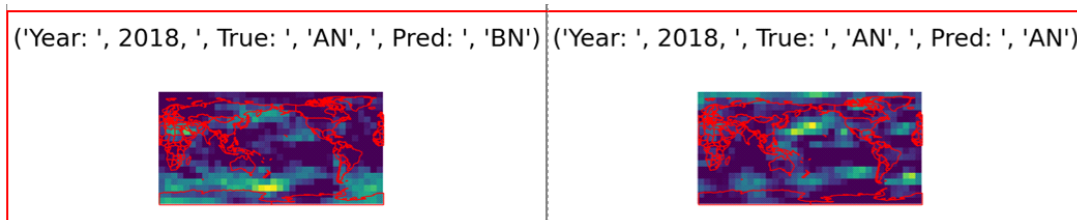


Figure 78. Class activation map before (left) and after (right) the data augmentation.

- 2021년 7월(폴드3 모델): BN로 낮게 예측 → AN로 개선

2021년 7월 기후분석토의 보고서(기후분석과 작성)의 내용을 살펴보면, 고온의 원인을 다음과 같이 기술하였다: “7월 초에는 열대 인도양부터 서태평양까지 대류활동이 활성화되면서 장마 시작과 많은 강수에 영향을 주었으며, 이후 양의 북극진동과 북대서양진동의 영향을 받아 북유럽부터 전파된 파동으로 인해 상층 고기압성 편차가 강화되며 기온이 급 상승하였음.” 자료강화 후 활성화된 맵을 보면 극지방이 활성화되었음을 볼 수 있다(Figure 79).

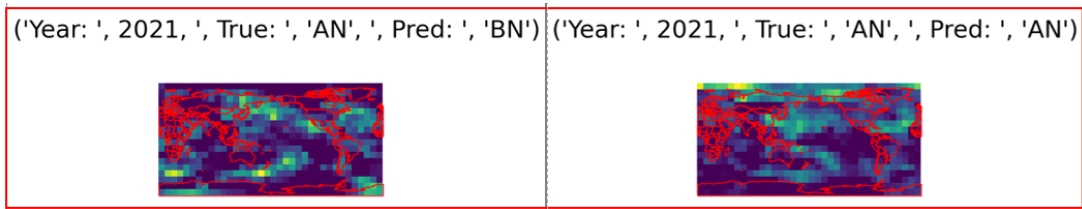


Figure 79. Class activation map before (left) and after (right) the data augmentation.

(다) 모든 월에 대한 지도학습 성능 평가

3D 합성곱 신경망 모델을 평균기온의 1-12월 LT1-3 예측에 적용하여 성능을 평가하였다. Validation Accuracy를 보면 모든 월과 선행시간에서 자료강화한 경우의 정확도가 높으며 (Figure 80) test Accuracy의 경우 대부분의 월 및 선행시간에서 자료강화한 경우의 정확도가 높거나 비슷함을 볼 수 있다(Figure 81). 8월 LT3 예측의 경우만 개선이 특히 미흡하게 나타난다. Test Heidke Skill Score (HSS)를 보면 마찬가지로 대부분의 월 및 선행시간에서 자료강화한 경우의 HSS가 높거나 비슷하게 나타난다(Figure 82).

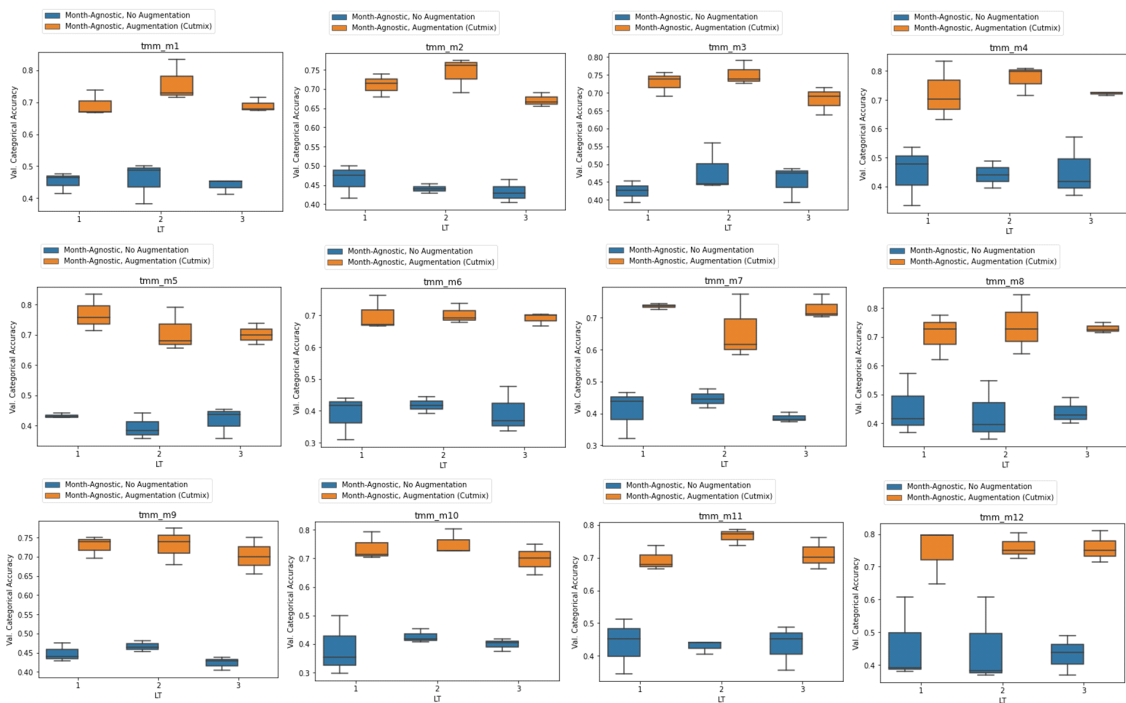


Figure 80. Validation accuracy of LT1-3 forecast of monthly mean temperature from January to December.

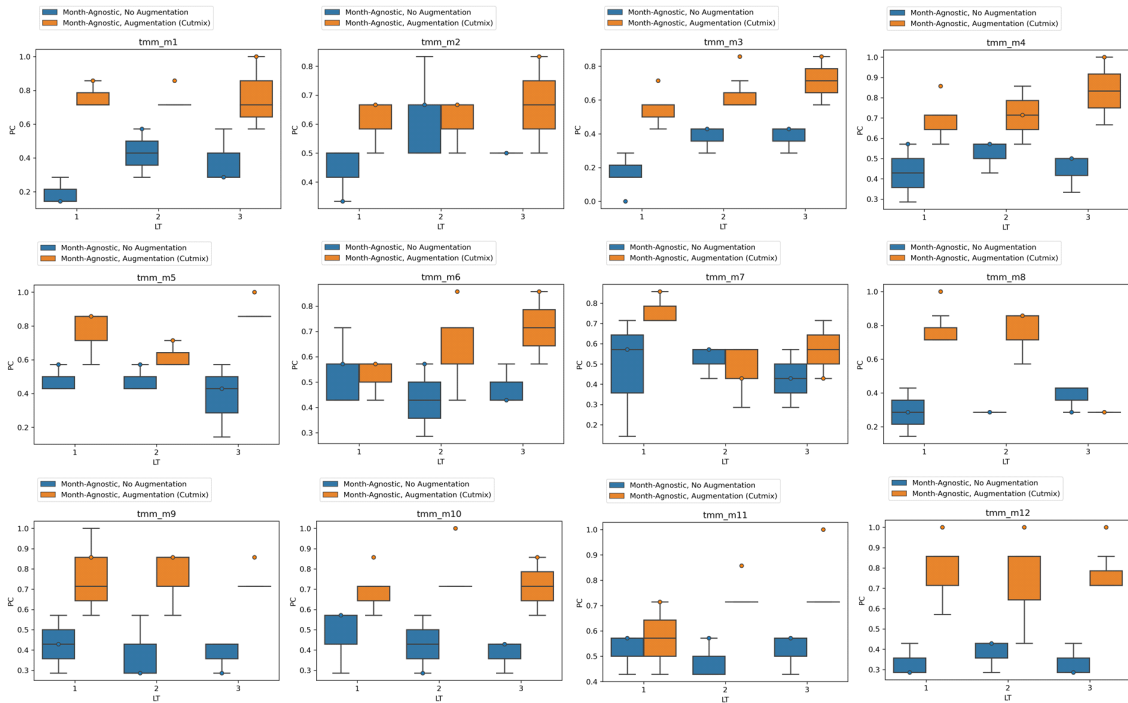


Figure 81. Test accuracy of LT1-3 forecast of monthly mean temperature from January to December.

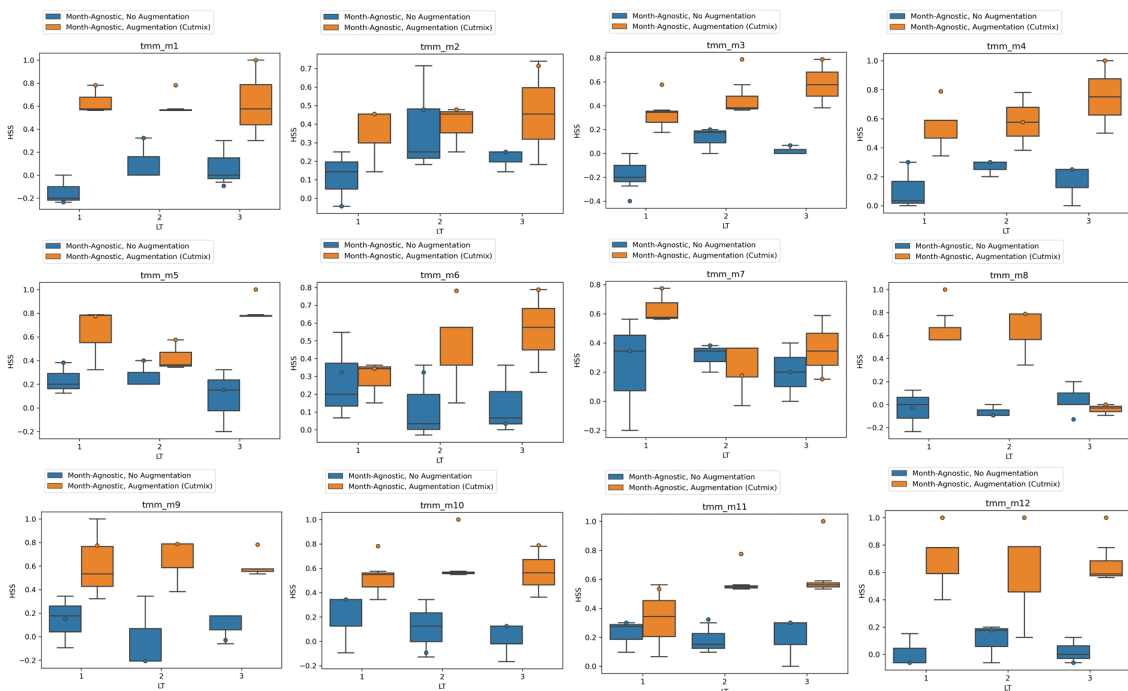


Figure 82. Test Heidke Skill Score (HSS) of LT1-3 forecast of monthly mean temperature from January to December.

(라) GP 모델과의 지도학습 성능 비교

5단계 연구에서 구축한 가우시안 프로세스(GP) 모델을 3D 합성곱신경망모델 테스트를 위해 설정한 것과 같은 3개 폴드에 대해 교차검증을 수행하고 동일한 테스트 자료에 대해 결과를 얻어 3D CNN 모델 결과와 성능 비교하였다(3-fold 결과 분포 및 앙상블, Figures 83-85). 월별로 다르게 나타나나, CNN의 성능이 전체적으로 더 우수하였다.

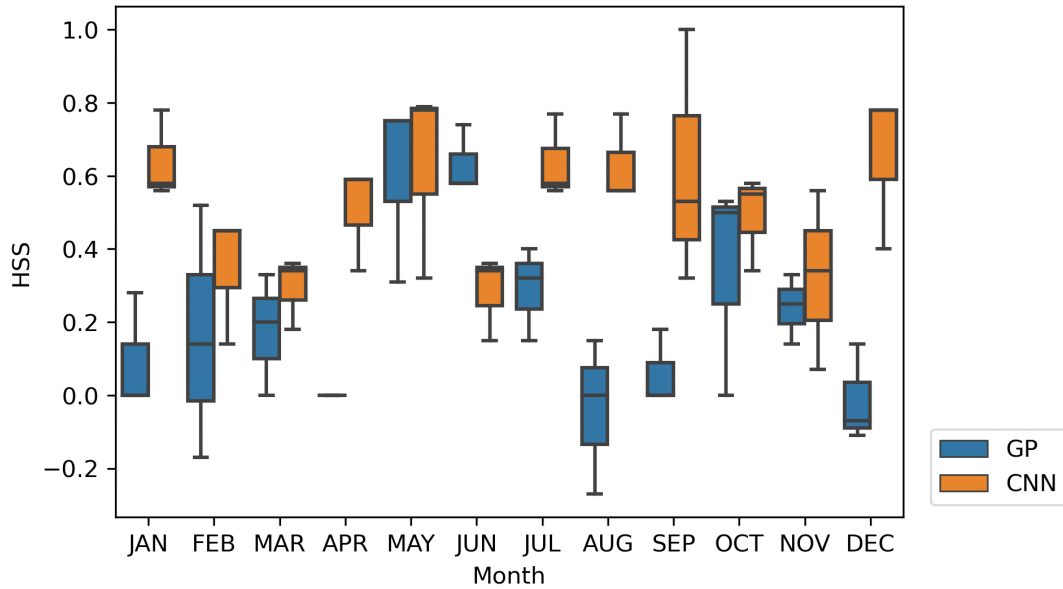


Figure 83. Test Heidke Skill Score (HSS) of LT1 forecast of monthly mean temperature.

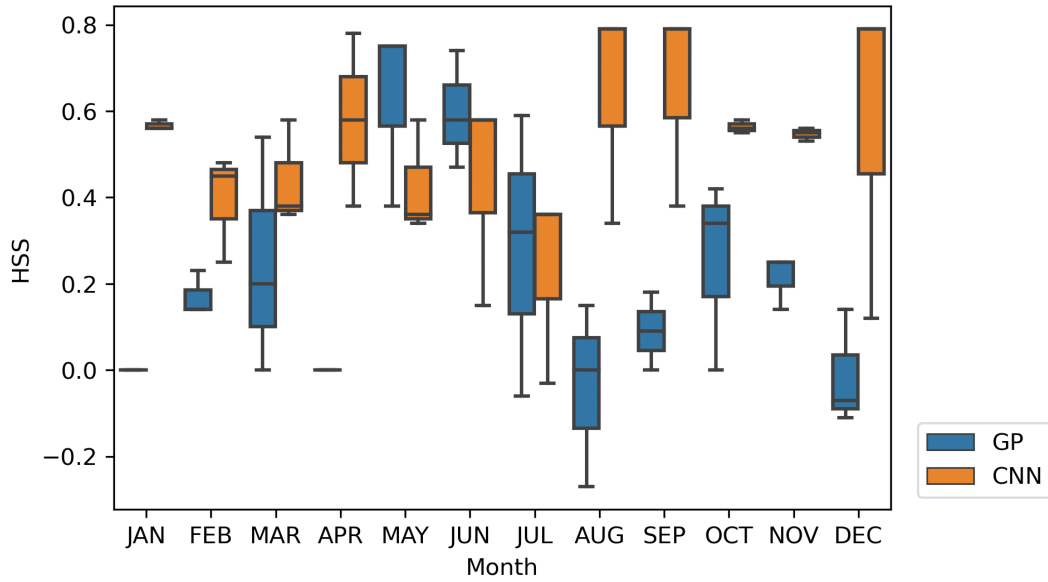


Figure 84. Test Heidke Skill Score (HSS) of LT2 forecast of monthly mean temperature.

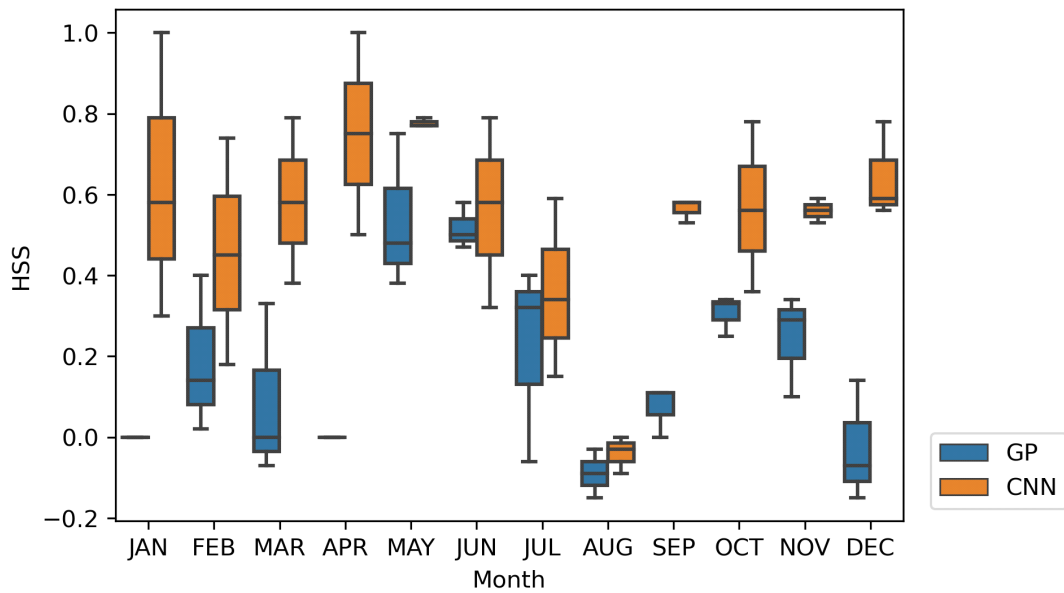


Figure 85. Test Heidke Skill Score (HSS) of LT3 forecast of monthly mean temperature.

(3) 노드 분류 그래프신경망(CNN) 모델 훈련 및 검증

Epoch 200 까지 모델을 구동한 후 validation loss가 가장 작은 epoch을 best epoch으로 선정하여 결과를 도출하였다. 목표 월만 사용한 경우(Target Month Only)와 모든 월을 사용한 경우(Month-Agnostic), 그리고 자료강화하지 않은 경우(No Augmentation)와 Cutmix 방법을 이용

하여 자료강화한 경우(Augmentation)의 결과를 비교하였으며 전처리 기법은 표준화를 활용하였다.

Train loss를 살펴보면 목표 월만 사용하고 자료강화 하지 않은 경우의 loss가 가장 낮았으며 그 다음 목표 월만 사용하고 자료강화한 경우가 낮았다(Figure 86). 모든 월을 사용한 경우는 train loss가 크게 낮아지지 않아 모델의 학습이 잘 이루어지지 않았음을 알 수 있었으며 (Figure 86) validation loss를 보면 목표 월과 폴드에 따라 개선 여부가 일관적이지 않았다 (Figure 87). Train accuracy를 살펴보면 목표 월만 사용하고 자료강화 하지 않은 경우의 accuracy가 가장 높았으며 그 다음 목표월만 사용하고 자료강화한 경우가 높았다(Figure 88). 모든 월을 사용한 경우는 train accuracy가 크게 높지 않아 모델의 학습이 잘 이루어지지 않았음을 알 수 있었다(underfitting, Figure 88). Validation accuracy를 보면 목표월과 폴드에 따라 개선 여부가 일관적이지 않으며 accuracy도 낮게 나타났다(Figure 89).

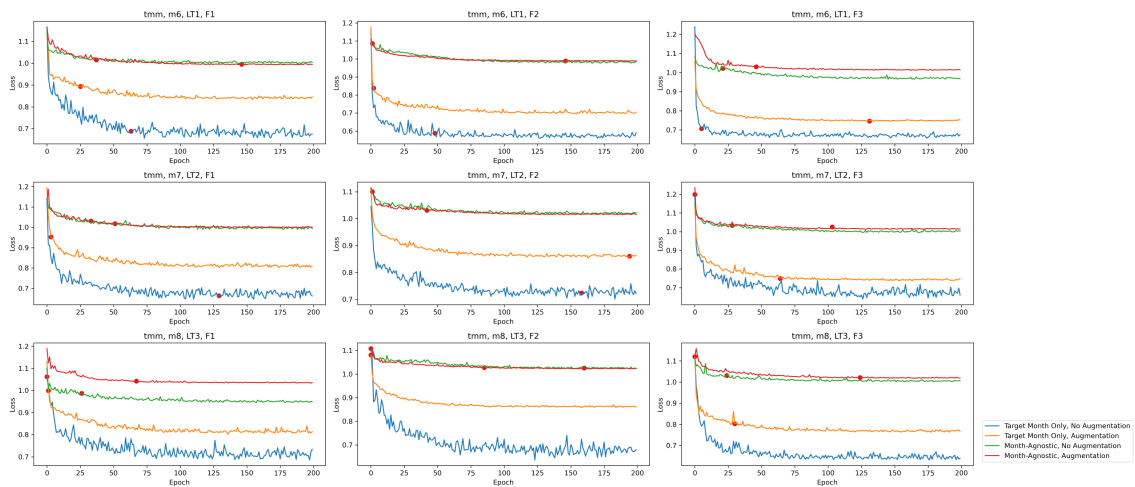


Figure 86. Comparison of train loss between the use of months and data augmentation methods.

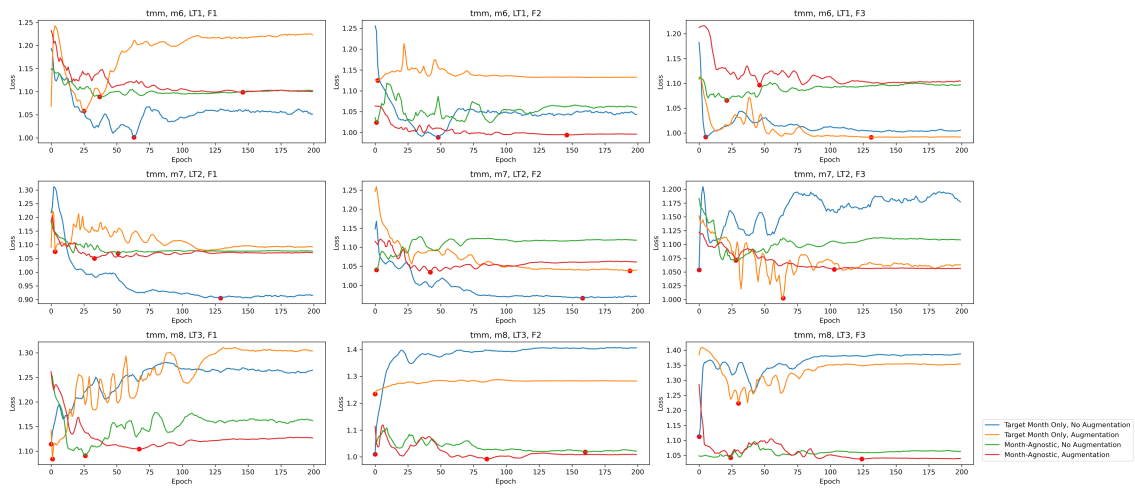


Figure 87. Comparison of validation loss between the use of months and data augmentation methods.

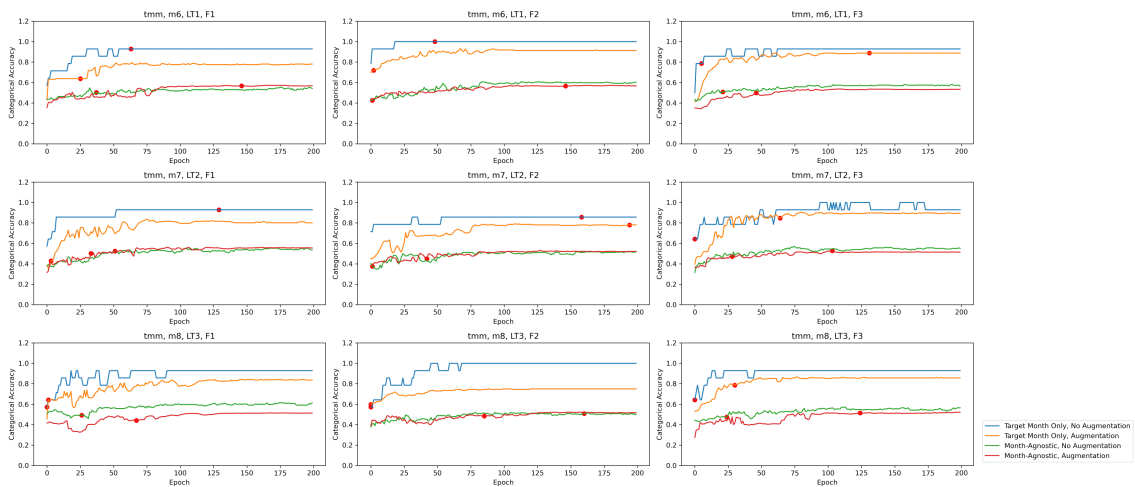


Figure 88. Comparison of train accuracy between the use of months and data augmentation methods.

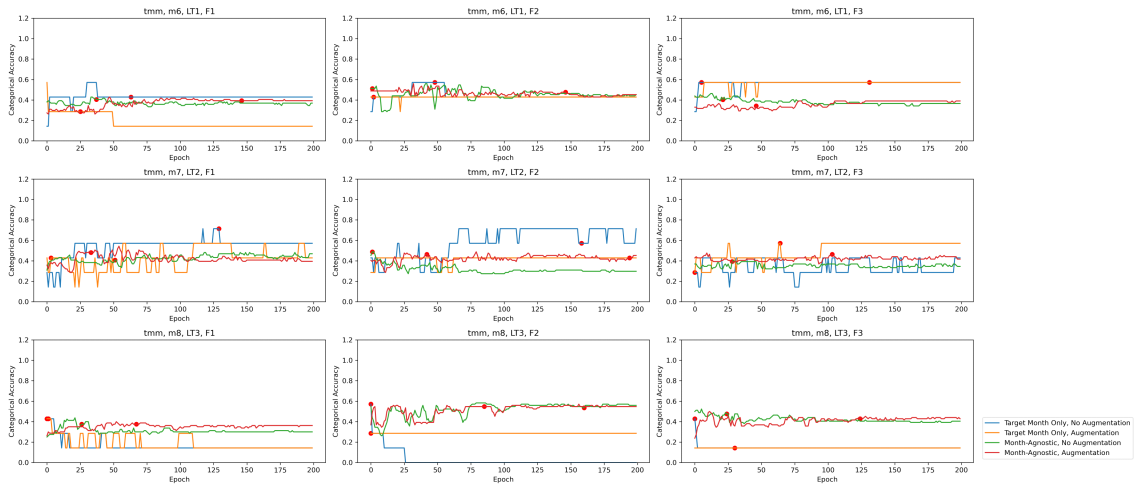


Figure 89. Comparison of validation accuracy between the use of months and data augmentation methods.

3개의 fold에 대한 validation accuracy를 box plot으로 나타내어 비교하면 다음 그림과 같다(Figure 90). 목표 월만 사용하고 자료강화 하지 않은 경우가 정확도 가장 높으며 모든 월의 사용이나 자료강화의 개선 효과가 드러나지 않았다.

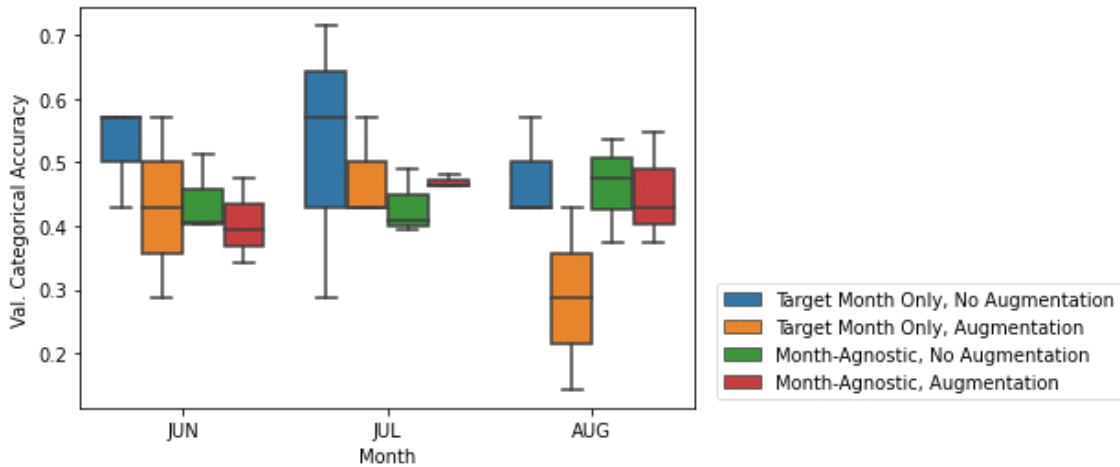


Figure 90. Comparison of validation accuracy between the use of months and data augmentation methods.

(4) 노드 분류 그래프신경망(CNN) 모델 지도학습 성능평가

이 연구에서 사용한 노드 분류 모델의 경우 훈련, 검증, 테스트 자료 모두에 대해 하나의 그래프를 구축하여 활용하므로(라벨에 대해서만 분리) 태생적으로 준지도학습이 된다. 모델의 최종 검증을 위한 테스트셋에 대한 평가는 이미지기반 모델 평가와 마찬가지로 3개의 폴드에 대한 모델을 각각 이용한 결과와 그 앙상블의 성능 평가로 수행하였다. 테스트셋 평가 결과, 7월(LT2)의 경우 목표 월만 사용하고 자료강화하지 않은 경우의 성능이 워낙 낮아 모든 월을 사

용한 경우와 자료강화한 경우의 개선 효과가 뚜렷이 나타났으나, 그 밖의 경우에는 개선의 효과가 나타나지 않았다(Figures 91-94).

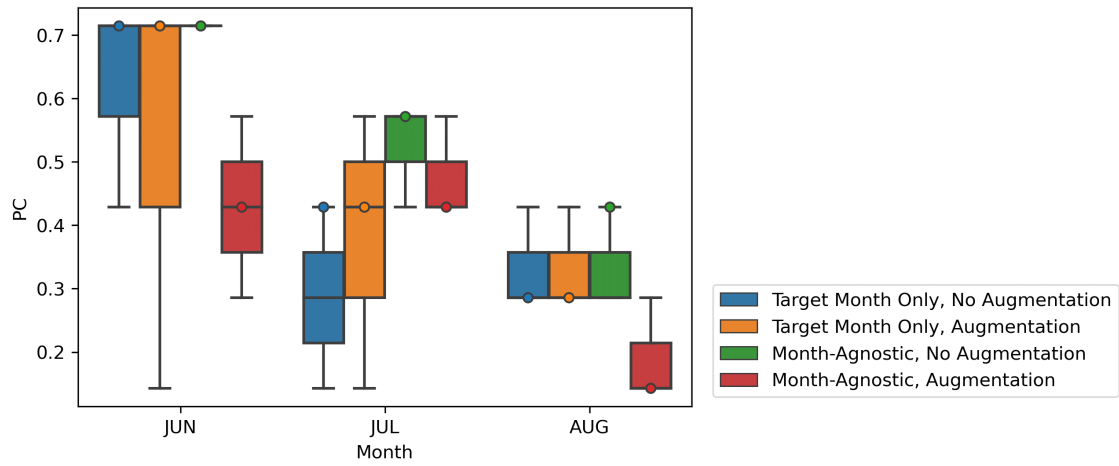


Figure 91. Proportion Correct (PC) between the use of months and data augmentation methods.

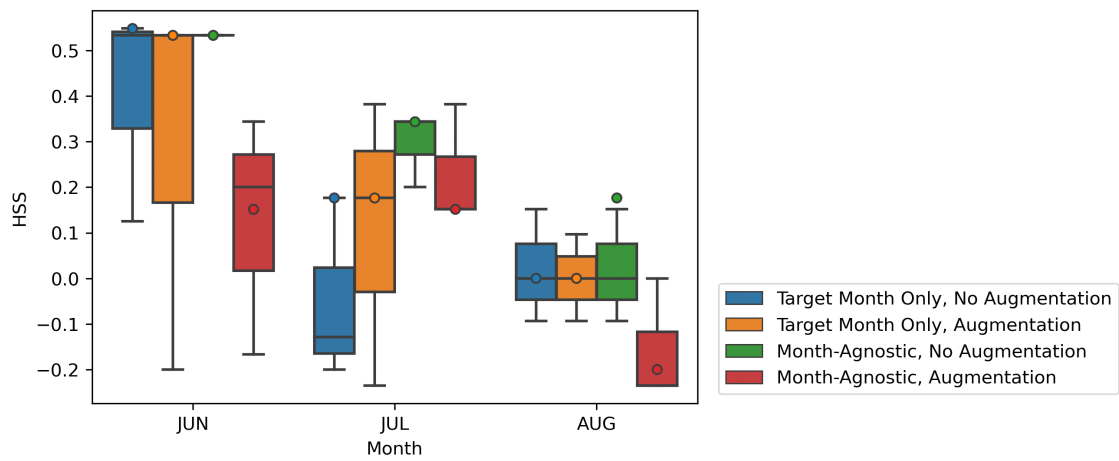


Figure 92. Heidke Skill Score (HSS) between the use of months and data augmentation methods.

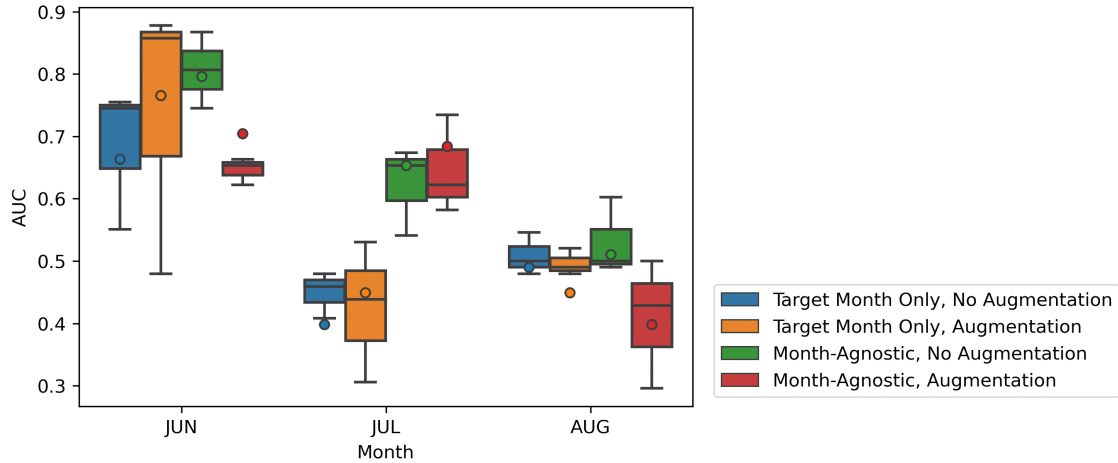


Figure 93. Area Under the ROC Curve (AUC) between the use of months and data augmentation methods.

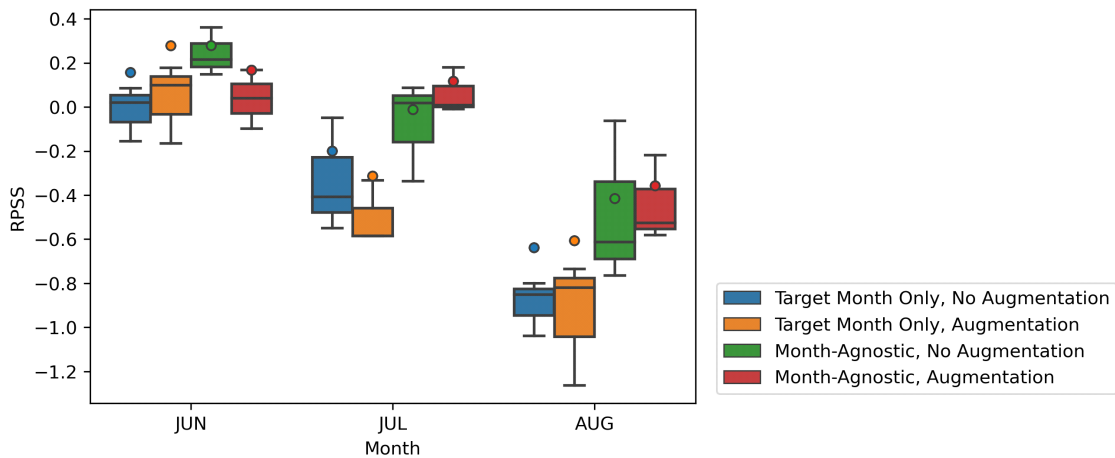


Figure 94. Ranked Probability Skill Score (RPSS) between the use of months and data augmentation methods.

(5) 그래프 분류 그래프신경망(CNN) 모델 훈련 및 검증

노드 분류 모델과 마찬가지로 Epoch 200까지 모델을 구동한 후 validation loss가 가장 작은 epoch을 best epoch으로 선정하여 결과를 도출하였다. 목표 월만 사용한 경우(Target Month Only)와 모든 월을 사용한 경우(Month-Agnostic), 그리고 자료강화하지 않은 경우(No Augmentation)와 Cutmix 방법을 이용하여 자료강화한 경우(Augmentation)의 결과를 비교하였으며 전처리 기법은 표준화를 활용하였다.

Train loss를 살펴보면 목표 월만 사용하고 자료강화 하지 않은 경우의 loss가 가장 낮았으며 그 다음 목표월만 사용하고 자료강화한 경우가 낮았다. 그런데 목표월만 사용하고 자료강화한 경우와 모든 월을 사용한 경우는 epoch을 되풀이하며 loss의 변동이 심하여 불안정한 상태

를 보였다(Figure 95). 모든 월을 사용한 경우는 train loss가 크게 낮아지지 않아 모델의 학습이 잘 이루어지지 않았음을 알 수 있었으며(Figure 95) 전체적으로 validation loss를 보면 목표 월과 폴드에 따라 개선 여부가 일관적이지 않았다(Figure 96). Train accuracy를 살펴보면 목표 월만 사용한 경우의 accuracy가 매우 높았다(Figure 97). 모든 월을 사용한 경우는 train accuracy가 크게 높지 않아 모델의 학습이 잘 이루어지지 않았음을 알 수 있었으며 (underfitting, Figure 97), validation accuracy를 보면 목표 월과 폴드에 따라 개선 여부가 일관적이지 않았다(Figure 98).

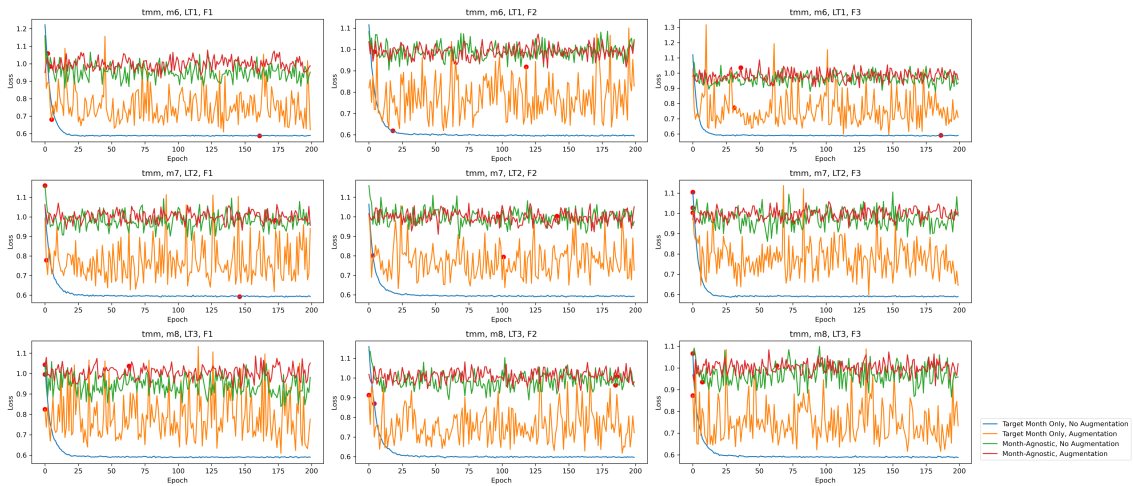


Figure 95. Comparison of train loss between the use of months and data augmentation methods.

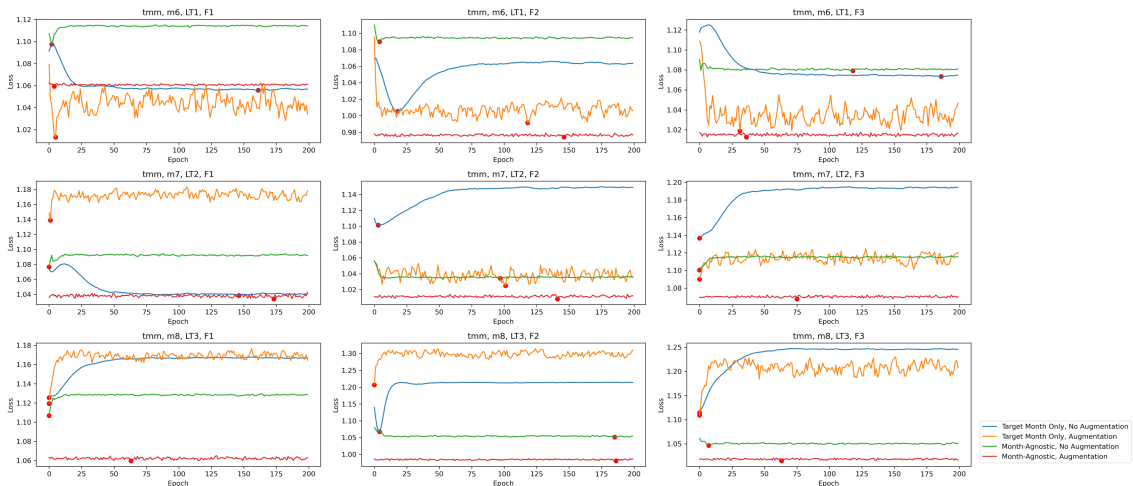


Figure 96. Comparison of validation loss between the use of months and data augmentation methods.

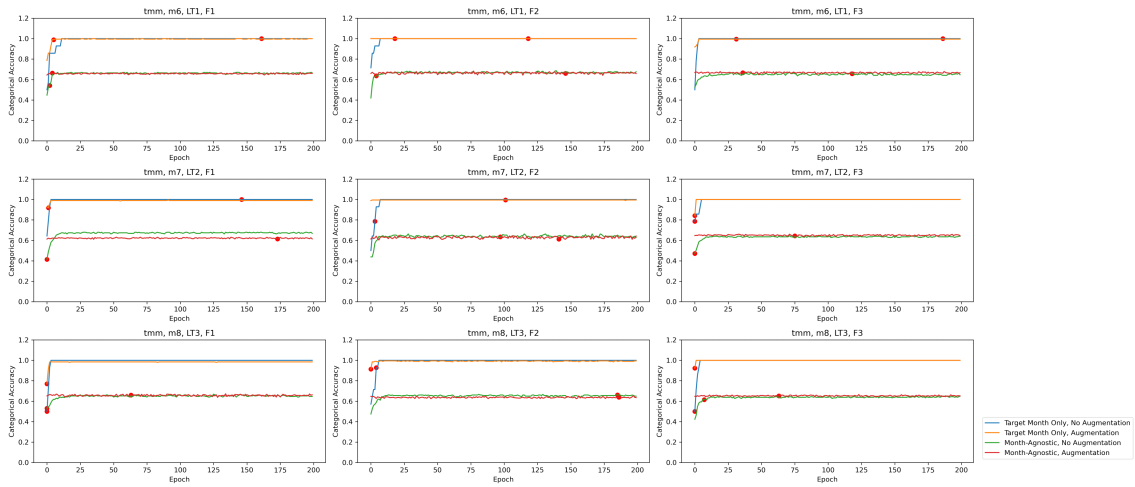


Figure 97. Comparison of train accuracy between the use of months and data augmentation methods.

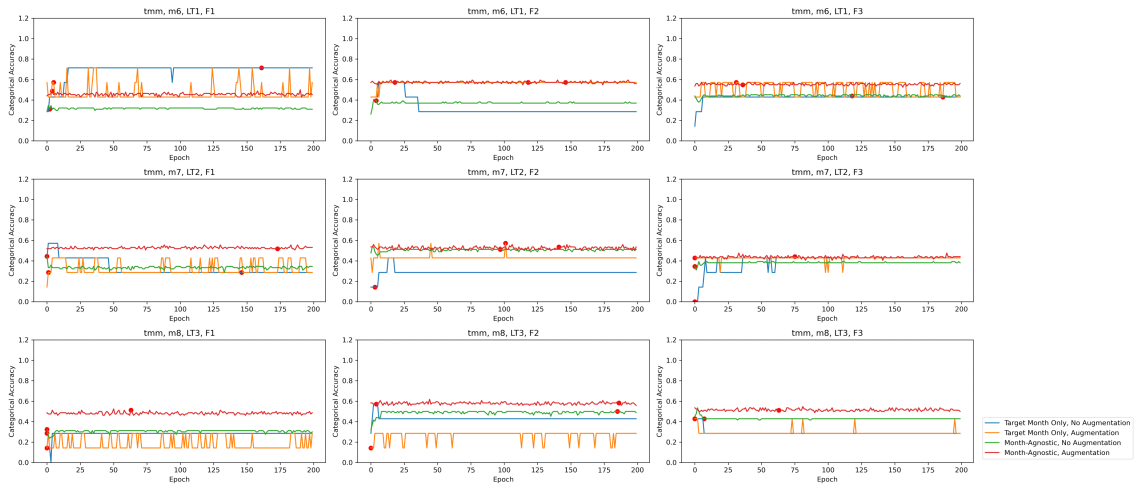


Figure 98. Comparison of validation accuracy between the use of months and data augmentation methods.

3개의 fold에 대한 validation accuracy를 box plot으로 나타내어 비교하면 다음 그림과 같다(Figure 99). 6월(LT1)의 경우 목표 월만 사용하고 자료강화 하지 않은 경우가 정확도 가장 높으나 7월(LT2)의 경우에는 매우 낮게 나타났다. 모든 월을 사용이나 자료강화한 경우 다른 경우와 비교해 정확도가 높은 것을 알 수 있다.

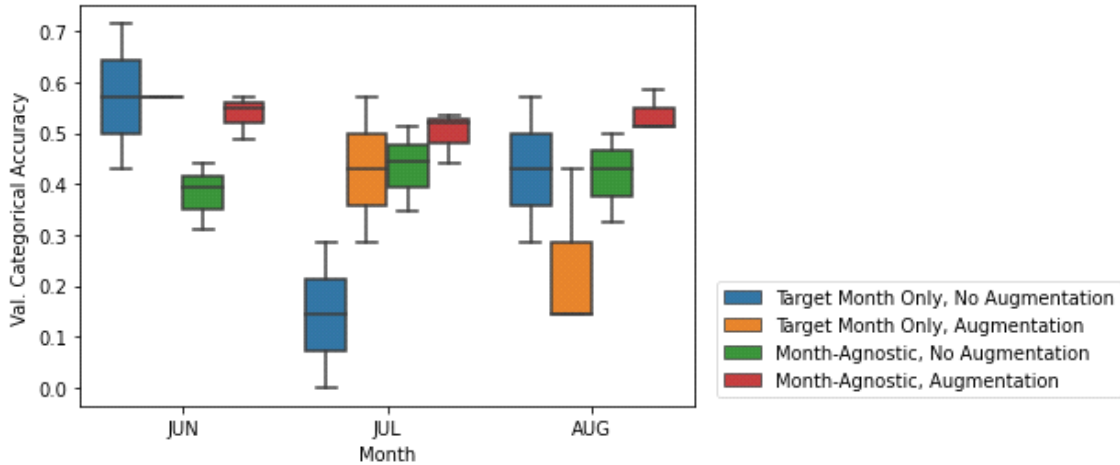


Figure 99. Comparison of validation accuracy between the use of months and data augmentation methods.

(6) 그래프 분류 그래프신경망(CNN) 모델 지도학습 성능평가

테스트셋 평가 결과, 6월(LT1)에는 목표 월만 사용한 결과보다 모든 월을 사용한 경우의 예측 성능이 우수하였으며 7월(LT2)의 경우는 목표 월, 모든 월을 사용한 경우 각각에 대해서 자료강화의 개선 효과를 관찰할 수 있었다(Figures 100-103). 그래프기반 모델의 경우 초기 엣지 연결 및 가중치 설정에 있어 노드 특성의 코사인 유사도를 활용하였는데 노드 간의 실제 유사성을 잘 반영하지 못하여 학습에 영향을 미친 것으로 보인다. 가중치는 노드 특성을 이용하여 초깃값을 설정할 수 있으나 엣지 연결 자체는 훈련 기간의 가용한 자료를 적극 활용하여 도출할 필요가 있으며 지금의 모델은 매개변수 및 초매개변수 최적화가 이루어지지 않은 상태이므로 후속 연구를 통해 개선의 여지가 있다.

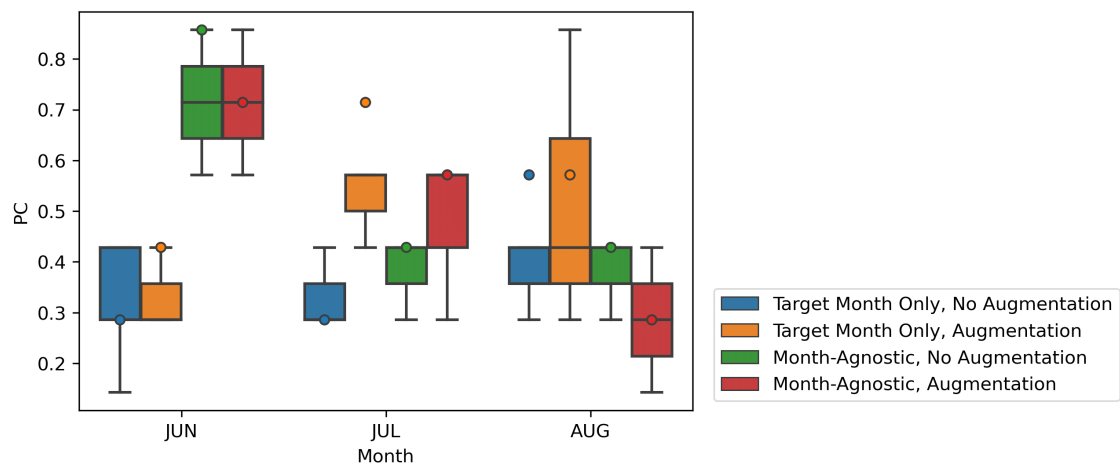


Figure 100. Proportion Correct (PC) between the use of months and data augmentation methods.

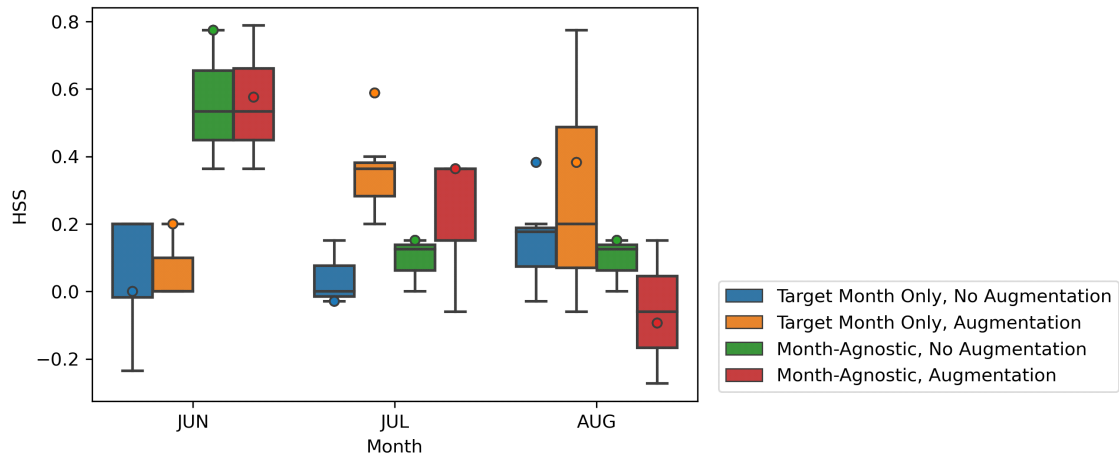


Figure 101. Heidke Skill Score (HSS) between the use of months and data augmentation methods.

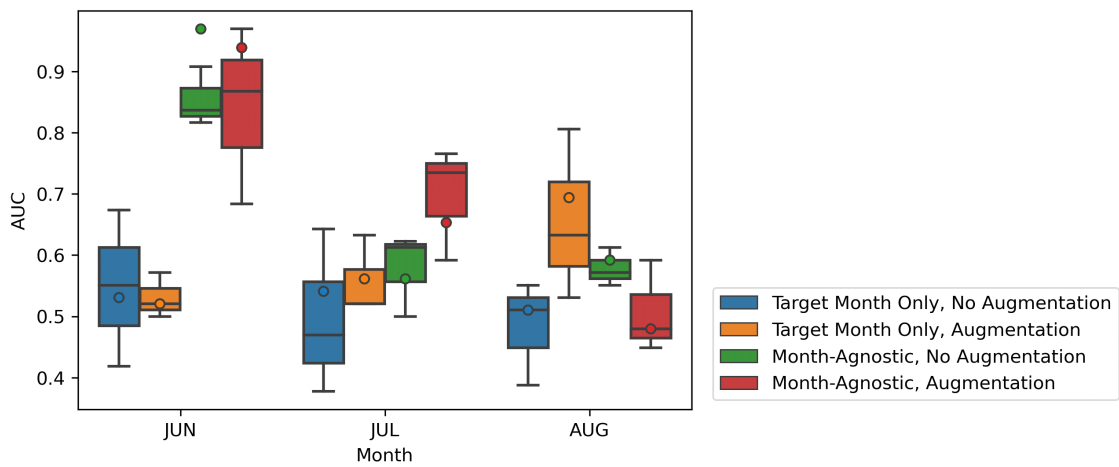


Figure 102. Area Under the ROC Curve (AUC) between the use of months and data augmentation methods.

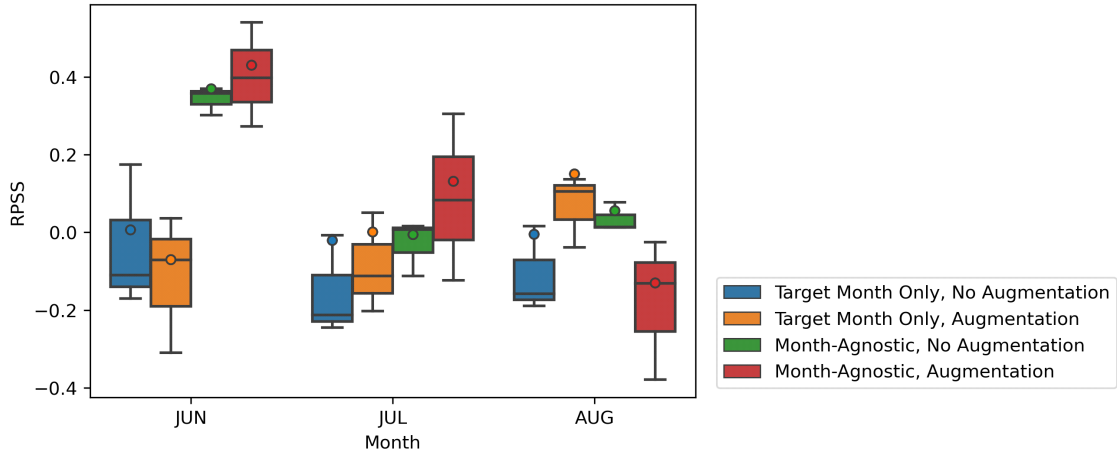


Figure 103. Ranked Probability Skill Score (RPSS) between the use of months and data augmentation methods.

GP 모델 결과와 이미지기반의 합성곱신경망모델, 그래프기반의 노드 분류 및 그래프 분류 모델의 평균기온 예측 HSS를 비교하였는데(그래프기반 모델의 경우 모든 월을 사용하고 자료 강화한 경우의 성능이 가장 높은 것은 아니었지만 편의상 이 경우의 결과를 이용함; 각 box 안에는 3개의 폴드에 대한 값), 6월(LT1) 예측의 경우 GP 모델과 GNN 그래프 분류 모델의 성능이 비교적 높았으나 7월(LT2) 예측의 경우는 CNN 모델의 성능이 가장 우수하였다. 8월의 경우에는 몇몇 폴드에 대한 경우를 제외하고는 전체적으로 성능이 좋지 않았다(Figure 104).

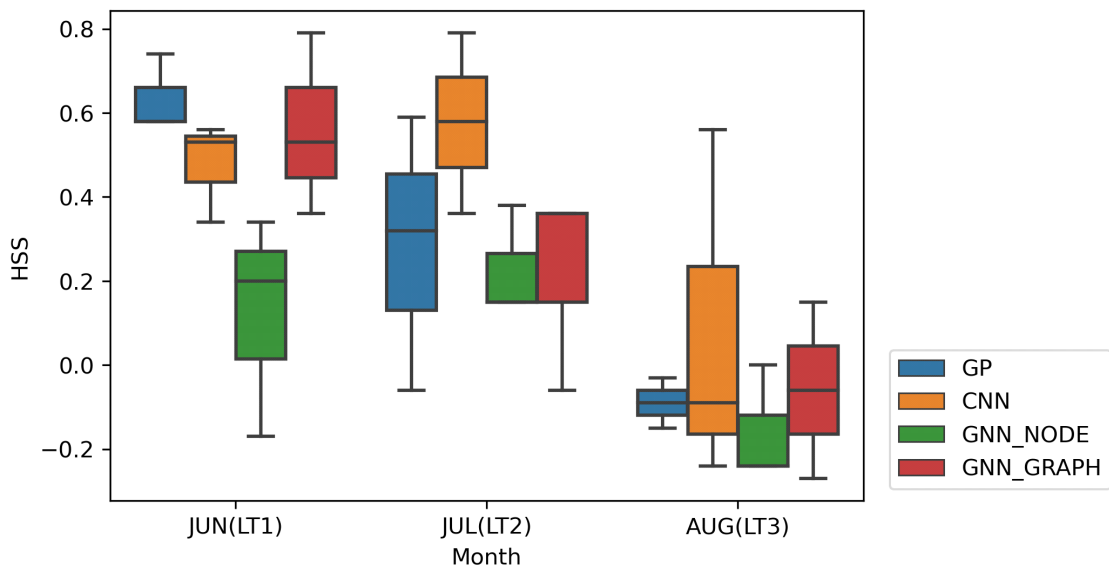


Figure 104. Comparison of Heidke Skill Score (HSS) between various AI models.

4. 요약 및 결론

동아시아 태평양 지역의 3-4주의 계절내 예측자료 생산을 위하여 자료를 확장하고 다양한 격자기반 딥러닝 프레임 ConvLSTM, U-Net 모델 구축과 정확도 높은 예측자료를 위한 민감도 분석을 위한 입출력 시스템을 개발하였다. 부족한 자료는 관측자료와 기간확장을 통하여 자료를 확장하고 확장된 자료는 ECMWF, GloSea5의 예측모델과 GPM 위성자료의 강수량 자료에 대해서 훈련셋을 구축하고 재분석자료 ERA-Interim에서 ERA5로 전처리하여 훈련셋을 구성하였다.

확장된 예측모델, 관측자료의 훈련셋은 ConvLSTM과 U-Net 아키텍처에서 다양하게 학습한 결과 ConvLSTM은 여전히 부족한 훈련셋에 기반하여 정확도 향상에 어려움이 있었으나 U-Net 모델에서는 성능이 개선되는 것을 확인하였다. 이에 보다 다양한 방법으로 딥러닝 모델 입력자료를 사용자가 쉽게 만들어서 민감도 분석을 할 수 있는 입출력 시스템을 개발하였다. 입출력 시스템은 예측모델, 관측자료, 위/경도 선택, 변수 선택을 통한 딥러닝 모델 입력자료를 산출할 수 있으며, 보다 직관적인 서비스를 위하여 GUI 기반으로 자료를 산출하고 가시화를 통하여 검증할 수 있는 시스템으로 개발되었다. 기 개발된 입출력 시스템은 자료 쉽게 훈련셋으로 변환이 가능하기 때문에 이에 기반한 보다 최적화된 인공지능 기법을 개발하는데 도움이 될 것이다.

목표 변수와 무관하거나 중복된 입력 변수 및 광범위한 정보로 구성된 훈련자료는 학습 알고리즘(예, 딥러닝 신경망)을 산만하게 하거나 잘못 이끌어서 예측성능을 저하시킬 수 있다. 바꿔 말하면, 전처리가 잘 된 훈련자료는, 컴퓨팅 계산 자원을 절감하고 훈련 모델의 성능 개선을 가능하게 한다. 이에 이 연구는 MinMax 및 MaxAbs 기법, 표준편차를 이용하는 Standard 기법 및 사분범위 등 분위수를 지정하여 변형하는 Robust와 Quantile 기법과 이런 기법이 적용되지 않는 경우 등으로 구성된 파이프라인을 구축하였다. 또한 변형된 훈련자료와 예측 목표 변수와의 상관관계를 계산하여 순위에 따라 훈련자료의 특성을 선택하는 기법(feature selection)을 추가함으로써, 전처리 파이프라인 및 피쳐 선택 기준을 통해 다루기 어려운 계절내-계절(Sub-seasonal to seasonal, 이하 S2S) 예측 원시자료를 모델의 훈련자료로 바로 전환할 수 있도록 하였다. 이 연구의 학습 모델은 U-Net에 간단하게 TimeDistributed를 합성곱 층에 적용하고 이것을 기준 모델(Base model)로 명명하였다. 또한, 여러 기후모델을 한 번에 고려하는 다중모델앙상블(multi-model ensemble, 이하 MME) 기법이 기후 예측의 신뢰성을 증가시킨다고 알려져 있는데, 6개 개별 기후모델의 S2S 훈련자료뿐만 아니라 개별 기후모델의 예측 빈도 및 예측 기간을 전체적으로 고려한 MME 기반 S2S 훈련자료도 생산하여 전처리 기법 및 Base model에서 MME 대한 예측성능을 평가하였다. 6개 기법으로 전처리 된 6개 개별 기후모델 및 MME 기반 일 최고 및 최저기온과 일 강수 S2S 훈련자료를 Base model에서 훈련한 예측값을 선행예측시간별(lead time) 공간상관계수(spatial Pattern Correlation Coefficient)를 보정 전 기후 예측 자료, 즉 관측값(ERA5)과 비교한 결과, 적합한 전처리 기법은 표준편차(Standard) 및 사분범위(Robust) 기법으로 처리된 일 강수 예측성능은 전체 lead-time 1-4주에서 모두 개선되었지만 일 최고 및 최저기온의 예측성능 개선은 3-4주에서만 나타났다. 또한, 피쳐 선택 기법이 적용된 일 강수의 경우 피쳐 선택에 따른 훈련자료의 차원 감소, 즉 입력 변수 감소가 예측

성능 변화에 영향을 미치지 않는 것으로 나타났다. 그러나 일 최고 및 최저기온의 경우 피쳐 선택에 의한 훈련자료의 입력 변수가 기준 이하, 예를 들면 입력 변수 7개 선택일 경우 예측 성능이 저하될 수 있는 것으로 확인되었다. 또한 여전히 피쳐 선택 기법이 일 최고 및 최저기온의 1주 예측성능 개선에 기여하지 않는 것으로 나타나서, S2S 기온 훈련자료의 예측성능 개선에는 강수와 다른 접근, 예를 들면, 다른 딥러닝 훈련 모델 적용 등이 필요할 것으로 판단된다. 동아시아 지역 6개 개별 기후모델 및 MME 기반 S2S 예측 훈련자료의 예측성능 향상을 위한 6개 전처리 기법 및 피쳐 선택에 따른 예측 성능 평가 결과와 이 연구를 통한 교훈을 다음과 같이 요약하였다.

- 훈련 모델에서 적절한 전처리 기법을 통해 변형된 S2S 예측 훈련자료의 예측성능이 개선된다는 본 연구의 가설은 6개 개별 기후모델 및 MME 기반 일 강수의 S2S 예측 자료의 PCC가 전체 선행예측시간에서 개선됨으로써 확인되었다. 그러나 일 최고 및 최저기온의 1주 개선은 여전히 미흡하였지만 3-4주 예측성은 개선되었다.
- 많은 기후 변수로 구성된 S2S 입력 자료는 훈련 알고리즘을 산만하게 하여 훈련 모델의 예측성능을 저하시킬 수 있으므로 목표 변수와 무관하거나 중복된 입력 변수를 제거하는 차원 감소 기법으로 S2S 훈련자료의 예측성능을 강화시킬 수 있다는 이 연구의 가설에서 일 강수 예측성능에는 피쳐 선택의 영향이 크지 않음을 확인하였다. 그러나 기온은 S2S 입력 변수 7개 선택 이하에서 예측성능이 저하될 수 있음을 확인하였다. 그러나 차원 감소에 의한 중복 정보 제거에도 기온의 1주 예측 개선은 여전히 미흡하였다. 기온의 1주 예측 개선과 관련해서는 전처리 기법 이외 다른 접근이 필요한 것으로 판단된다.
- 예측성능이 높다고 알려진 ECMWF와 동등한 예측성능을 가진 MME 기반 S2S 예측 자료를 구축하였으며, 딥러닝 후보정 통해 선행예측시간 2-4주 예측성능 개선을 확인하였다.

이 연구에서 다양한 예측 빈도 및 예측 기간 등으로 이용하기 어려운 S2S 예측 원시자료에 다양한 전처리 기법 적용이 가능하도록 파이프라인을 구축하고 S2S 예측성능 개선과 관련한 적절한 전처리 기법을 확인하였다. 또한, 여러 개 기후 예측 자료를 한 번에 고려한 동아시아 지역 MME 기반 S2S 훈련자료를 구축하여 MME 기법만으로도 S2S 예측성능 개선이 가능함을 확인하고 딥러닝 훈련을 통해 예측성능 강화도 확인하였다. 그러나 기온의 1주 예측성능이 미흡한 점 등 최적화 전처리 기법만으로는 예측성능을 높이는 데에 한계가 있다는 것을 알게 된 만큼 후속 연구를 통해 다른 접근, 예를 들면 기온에는 다른 딥러닝 훈련 모델을 적용하고 그 모델 구조 및 파라미터 조정 등의 최적화 및 민감도 분석 등을 통해 S2S 예측 목표 변수에 대한 최적의 모델 및 개선 방향 등을 제시할 수 있을 것으로 기대한다.

준지도학습 적용 환경구축을 위해 그래프기반 모델과 이미지기반 모델을 구축하고 지도학습을 수행하였다. 인공지능모델 훈련의 가장 큰 걸림돌인 자료 부족 문제를 완화하고자 기상기후자료의 자료강화를 테스트하였는데, 일자료의 시간적 조합을 활용한 자료강화와 월자료의 공간적 재조합을 활용한 자료강화를 적용하여 비교하였다. 또한 목표 월만 사용하는 경우와 모든 월을 사용하는 경우의 결과를 비교하였다.

그래프기반 모델로는 노드 분류를 위한 그래프합성곱신경망 모델과 그래프 분류를 위한 그래프합성곱신경망 모델을 각각 구축하여 3-폴드 교차검증으로 우리나라 여름철 평균기온 예측 성능을 평가하였다. 노드 분류와 그래프 분류 모델 모두 모델의 훈련이 전체적으로 잘 이루어지지 않아 자료강화의 개선 효과는 확인할 수 없었는데, 다만 노드 분류 모델의 테스트 결과에 있어서는 7월(LT2)의 경우 목표 월만 사용하고 자료강화하지 않은 경우의 성능이 낮아 모든 월을 사용한 경우나 자료강화한 경우의 개선 효과가 나타났으며(HSS>0인 폴드 확장), 그래프분류 모델의 경우 6월(LT1)에는 모든 월을 사용한 경우의 예측 성능이 우수하였고(모든 폴드에서 HSS>0.35) 7월(LT2)의 경우는 자료강화의 개선 효과를 관찰할 수 있었다. 이미지기반 모델로는 3차원 합성곱신경망 모델을 구축하여 예측 성능을 평가하였는데, 월자료의 공간적 재조합을 활용한 자료강화를 수행하고 모든 월을 사용한 경우 다른 경우에 비해 모델이 비교적 잘 학습되었으며 검증 정확도도 높게 나타났다(모든 월, 모든 폴드에서 정확도>0.6). 테스트 결과를 보면 6월(LT1)과 7월(LT2)의 경우 자료 강화의 개선 효과를 확인할 수 있었다. 모델의 마지막 합성곱층의 활성화맵을 그려 입력자료에서 중요하게 기여한 지리적 위치를 도출하였는데, 자료강화를 통해 기후학적으로 의미있는 지역이 활성화된 사례를 확인할 수 있었다.

그래프기반 모델과 이미지기반 모델 모두 매우 단순한 구조의 모델을 구축하였으며 모델 매개변수 및 초매개변수의 최적화는 수행되지 않았으므로 후속 연구를 통해 개선될 여지가 크다. 그래프기반모델의 경우 특히 노드 간의 유사성을 실제와 가깝게 도출하여 엣지 연결에 반영할 필요가 있다. 1차년도에는 단순하게 구성한 그래프기반모델 및 이미지기반모델의 지도학습 성능을 알아보고 자료강화 등의 개선 효과를 살펴보았으며 2차년도 연구를 통해 준지도학습을 적용하여 모델을 개선하고 최적화를 수행할 것이다.

사사

이 연구는 APCC의 지원을 통해 수행 되었습니다.

References

- 신홍준, 윤성심, 최재민, 2021: 시간 연속성을 고려한 딥러닝 기반 레이더 강우예측. 한국수자원학회 논문집 54, 301-309.
- Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., Hickey, J., 2019: Machine Learning for Precipitation Nowcasting from Radar Images, <https://doi.org/10.48550/arxiv.1912.12132>.
- Alqaraawi, A., Schuessler, M., Weiß, P., Costanza, E., Berthouze, N., 2020: Evaluating saliency map explanations for convolutional neural networks. *Proc. Int. Conf. on Intelligence User Interfaces*, 275-285, <https://doi.org/10.1145/3377325.3377519>.
- Ayzel, G., Scheffer, T., Heistermann, M., 2020: RainNet v1.0: a-convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, **13**, 2631-2644, <https://doi.org/10.5194/gmd-13-2631-2020>.
- Best, M. J., Pryor G., Clark D. B., Rooney D. B., Essery D. B. H., Menard C. B., Edwards J. M., Hendry J. M., Porson A., Gedney N., Mercado L. M., Sitch S., Blyth E., Boucher O., Cox P. M., Grimmond C. S. B., and Harding R. J., 2011: The Joint UK Land Environment Simulator (JULES), model description - Part 1: Energy and water fluxes. *Geoscientific Model Development*, **4**, 677-699, <https://doi.org/10.5194/gmd-4-677-2011>.
- Bihlo, A., Popovych, R.O., 2022: Physics-informed neural networks for the shallow-water equations on the sphere. *Journal of Computational Physics*, 456, 111024, <https://doi.org/10.1016/j.jcp.2022.111024>.
- Chantry, M., Christensen, H., Dueben, P., Palmer, T. 2021: Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. *Philosophical Transactions of the Royal Society A*, **379**, 20200083. <https://doi.org/10.1098/rsta.2020.0083>.
- Chowdhury, M.E.H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z. Bin, Islam, K.R., Khan, M.S., Iqbal, A., Emadi, N. Al, Reaz, M.B.I., Islam, M.T., 2020: Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access*, **8**, 132665-132676, <https://doi.org/10.1109/ACCESS.2020.3010287>.
- de Andrade, F.M., Coelho, C.A.S., Cavalcanti, I.F.A., 2019: Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dynamics*, **52**, 5451-5475, <https://doi.org/10.1007/s00382-018-4457-z>.

- ECMWF, 2020: <https://www.ecmwf.int/>
- Fahim, M.A.N.I., Saqib, N., Siam, S.K., Jung, H.Y., 2022: Rethinking Gradient Weight's Influence over Saliency Map Estimation. *Sensors*, **22**, <https://doi.org/10.3390/s22176516>.
- Gerritsma, A., 2020: Super-resolution neural networks: A strategy to downscale model output data. M.S. thesis, The Wageningen University, 35 pp.
- Ha, J.-H., Lee, Y.H., Kim, Y.-H., 2016: Forecasting the Precipitation of the Next Day Using Deep Learning. *Journal of the Korean Institute of Intelligent Systems*, **26**, 93–98, <https://doi.org/10.5391/jkiis.2016.26.2.093>.
- Ham, Y.-G., Kim, J.-H., Luo, J.-J. 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572. <https://doi.org/10.1038/s41586-019-1559-7>.
- Hong, S., Hwan Kim, J., Sung Choi, D., Baek, K., 2021: Development of Surface Weather Forecast Model by using LSTM Machine Learning Method. *Atmosphere*, **31**, 73–83, <https://doi.org/10.14191/Atmos.2021.31.1.073>.
- Hunke, E. C., and Lipscomb W. H., 2008: The Los Alamos sea ice model documentation and software user's manual, Version 4.0, LA-CC-06-012, Los Alamos National Laboratory, N.M.
- Kim, H., Ham, Y.G., Joo, Y.S., Son, S.W. 2021: Deep learning for bias correction of MJO prediction, *Nature Communications*, **12**, 3087. <https://doi.org/10.1038/s41467-021-23406-3>.
- MacLachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereday, D., Scaife, A.A., Gordon, M., Vellinga, M., Williams, A., Comer, R.E., Camp, J., Xavier, P., Madec, G., 2015: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, **141**, 1072–1084. <https://doi.org/10.1002/qj.2396>.
- Monhart, S., Spirig, C., Bhend, J., Bogner, K., Schär, C., Liniger, M.A., 2018: Skill of Subseasonal Forecasts in Europe: Effect of Bias Correction and Downscaling Using Surface Observations. *Journal of geophysical research-Atmospheres*, **123**, 7999–8016, <https://doi.org/10.1029/2017JD027923>.
- Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P., Gentine, P., 2021. Assessing the Potential of Deep Learning for Emulating Cloud Superparameterization in Climate Models With Real-Geography Boundary Conditions. *Journal of Advances in Modeling Earth Systems*, **13**, 1–26, <https://doi.org/10.1029/2020MS002385>.
- Nooteboom, P.D., Feng, Q.Y., López, C., Hernández-García, E., Dijkstra, H.A. 2018: Using

- Network Theory and Machine Learning to predict El Nino, *Earth System Dynamics*, **9**, 969–983. <https://doi.org/10.5194/esd-9-969-2018>.
- Reichler, T.J., Roads, J.O. 2003: The role of boundary and initial conditions for dynamical seasonal predictability, *Nonlinear Processes in Geophysics*, **10**, 211–232. <https://doi.org/10.5194/npg-10-211-2003>.
- Ronneberger, O., Fischer, P., Brox, T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. <https://doi.org/10.48550/arxiv.1505.04597>.
- Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D., Folland, C.K., Gordon, M., Hermanson, L., Knight, J.R., Lea, D.J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A.K., Smith, D., Vellinga, M., Wallace, E., Waters, J., Williams, A., 2014: Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, **41**, 2514–2519, <https://doi.org/10.1002/2014GL059637>.
- Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A., Stadtler, S. 2021: Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, **379**, 20200097. <https://doi.org/10.1098/rsta.2020.0097>.
- Simonyan, K., Vedaldi, A., Zisserman, A., 2014: Deep inside convolutional networks: Visualising image classification models and saliency maps. *2nd Int. Conf. Learn. Represent. ICLR 2014 - Work. Track Proc.* 1–8, <https://doi.org/10.48550/arxiv.1312.6034>.
- Singh, M., Kumar, B., Rao, S., Gill, S.S., Chattopadhyay, R., Nanjundiah, R.S., Niyogi, D., 2021: Deep learning for improved global precipitation in numerical weather prediction systems 5. <https://doi.org/10.48550/arxiv.2106.12045>.
- Son, H., Kim, S., Yeon, H., Kim, Y., Jang, Y., Kim, S.E., 2021: Visual analysis of spatiotemporal data predictions with deep learning models. *Applied Sciences*, **11**, <https://doi.org/10.3390/app11135853>.
- Vitart, F., Robertson, A.W., 2018: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, **1**, 3, <https://doi.org/10.1038/s41612-018-0013-0>.
- Weigel, A.P., Liniger, M.A., Appenzeller, C., 2008: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, **134**, 241–260, <https://doi.org/10.1002/qj.210>.
- Weyn, J.A., Durran, D.R., Caruana, R. 2019: Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data. *Journal of Advances in Modeling Earth Systems*, **11**, 2680–2693.

<https://doi.org/10.1029/2019MS001705>.

- Weyn, J.A., Durran, D.R., Caruana, R., Cresswell-Clay, N., 2021: Sub-Seasonal Forecasting With a Large Ensemble of Deep-Learning Weather Prediction Models. *Journal of Advances in Modeling Earth Systems*, **13**, <https://doi.org/10.1029/2021MS002502>.
- Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., Xu, L., Chen, Z., Gong, J., 2019: A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling & Software*, **120**, 104502, <https://doi.org/10.1016/j.envsoft.2019.104502>.
- Xingian S., Zhouong C., Hao W., Dit-Yan Y., Wai-kin W., and Wang-chun W., 2015: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, arXiv:1506.04214v2.
- Yun, S., D. Han, S. J. Oh, S. chun, J. Choe, and Y. Yoo. 2019: CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. Proc. IEEE/CVF Int. Conf. on Computer Vision. 2019, pp. 6023-6032. <https://doi.org/10.48550/arXiv.1905.04899>.
- Zhang, Z., Dong, Y., Yuan, Y., 2020: Temperature Forecasting via Convolutional Recurrent Neural Networks Based on Time-Series Data. *Complexity* 2020. <https://doi.org/10.1155/2020/3536572>.