

**객관적 기후예측에 적합한 인공지능 활용 기반기술
개발(III)**

부제: 기후예측에 적합한 인공지능모델 최적화 및 적용성 평가

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

김미애, 박경원, 이성규, 이윤영, 이진영, 정유란

기후사업본부 예측기술과

2024.12.

아시아·태평양경제협력체 기후센터

Executive Summary

In recent years, artificial intelligence (AI)-based medium-range weather prediction models (within 15 days) have been actively developed by leading companies and research institutes, including NVIDIA, Huawei, Google DeepMind, and the European Centre for Medium-Range Weather Forecasts (ECMWF). Notable examples include Nvidia's Spherical Fourier Neural Operators (SFNO)-based FourCastNet, Huawei's Swin-Transformer-based Pangu-Weather, ECMWF's Graph Neural Network (GNN)-based Artificial Intelligence/Integrated Forecasting System (AIFS), Fudan University's U-Transformer-based FuXi model and FuXi-S2S for subseasonal-to-seasonal (S2S) forecasting, and Google DeepMind's GNN-based GraphCast and diffusion model-based GenCast. These models are predominantly focused on medium-range weather forecasting and have not been widely extended to S2S forecasting. Furthermore, the training of AI models requires large datasets, but the ECMWF ERA-5 dataset—the primary source of climate data for AI training—spans only approximately 75 years. This limitation significantly restricts the amount of data available for training, validation, and testing.

Due to the climate crisis, the importance of high-accuracy subseasonal prediction data in the field of applied climatology is increasing. Such subseasonal prediction data face challenges in forecasting based solely on numerical models, as they rely on physical interactions among various atmospheric and oceanic processes. To overcome these difficulties, efforts have been made to enhance accuracy by expanding datasets and assigning greater weights to important patterns and features in input data. This study utilized the Attention U-Net model and subseasonal prediction models, as well as observational data-based techniques such as Filter, Wrapper, and Embedded methods, to identify the characteristics of variables. Sensitivity analysis was conducted based on various input-output systems. It was confirmed that selecting specific combinations of variables from model/observational data improved accuracy compared to using all variables for subseasonal predictions. The GUI-based input-output system facilitated the generation of input data and sensitivity analysis, contributing to the development of optimized artificial intelligence techniques. This research can serve as foundational data for improving subseasonal prediction accuracy and highlights the need for further advancements in deep learning models in the future.

Recent advances in neural network models and deep learning technologies have gained significant attention across various industries and research fields, including their growing potential in climate prediction. In the context of Sub-seasonal to Seasonal (S2S) forecasting, there is a strong emphasis on the importance of data preprocessing and improving model performance. This study focuses on enhancing the prediction accuracy of daily maximum temperatures (TMAX) in S2S forecasting using ensemble methods such as voting, bagging, boosting, and stacking. The ensemble framework was built using CNN, CNN-LSTM, U-Net, Attention U-Net, and Residual U-Net as base models, and the characteristics and performance

of each ensemble approach were systematically analyzed. Results showed that, for ECMWF-S2S TMAX forecasts, all ensemble methods improved the Anomaly Correlation Coefficient (ACC), with bagging achieving the best overall performance. For KMA (GloSea5)-S2S TMAX forecasts, stacking and boosting demonstrated consistent performance improvements across the entire lead time range (1 to 4 weeks), whereas voting and bagging were effective primarily beyond 4 weeks (20 days). Spatial distribution analyses revealed that applying ensemble techniques to ECMWF forecasts enhanced ACC across all lead times (1 to 4 weeks). Notably, the ACC spatial distribution for bagging closely aligned with the pre-ensemble distribution, showing minimal differences. In evaluations targeting specific areas of interest (e.g., four grid points over South Korea), performance improvements were more pronounced for KMA forecasts compared to ECMWF. This highlights that the effectiveness of ensemble techniques can vary depending on the training data and regional characteristics. This study underscores the value of ensemble methods in improving the stability and accuracy of S2S forecasts. The findings are expected to contribute to enhancing the precision and practicality of S2S daily maximum temperature predictions, particularly for 3–6 week lead times, and to improving the reliability of extreme heat forecasts.

In the second year (2023) of this project, we developed an image-based deep learning model for classifying MJO phases using semi-supervised learning (SSL) techniques. For this purpose, we first optimized a supervised learning model by testing various combinations of input variables, model architectures, and training data split methods based on different MJO indices. Then, the optimized model was trained in the SSL framework with both unlabeled and labeled data to classify MJO phases. The semi-supervised model demonstrated its effectiveness in learning critical climate patterns even with reduced labeled data, achieving comparable or superior accuracy to supervised models with all labeled data. Sensitivity analyses revealed that spatial characteristics of climate data should be properly taken into account for data augmentation in SSL. Horizontal flipping reduced accuracy while vertical flipping improved performance for MJO phase classification. Building on this, the study extended the semi-supervised approach for MJO phase classification to RMM prediction in the present year of 2024. A modified ResNet-18 model was developed to predict RMM1 and RMM2 indices for lead times of 0 to 40 days. By integrating anomaly and background climate variables (U200, U850, OLR, VP, TS, TCWV, hadvect), the model achieved skillful RMM predictions up to 22 days based on BCOR 0.5. A custom loss function incorporating MJO amplitude further reduced amplitude-related errors, emphasizing the importance of domain-specific adjustments in deep learning-based MJO research. Interpretable AI techniques highlighted the roles of background wind fields and moisture anomalies in MJO prediction, offering insights for future studies. Finally, probabilistic prediction models were developed using ensemble perturbation methods for initial conditions and model parameters, including ERA5 ensemble data, Perlin noise, and MC dropout. Results suggested that Perlin noise method showed more ensemble spreads over lead times compared to others but was

inconsistent across MJO cases, reflecting the challenge of generating ensemble spreads using deep learning models. Future research will explore advanced AI models, high-resolution data for detailed spatial pattern identification to further enhance climate predictability.

To improve the predictability of temperature probabilities on S2S timescales, we developed a deep learning-based 3-6 week temperature probability forecast model for East Asia. The model employs a U-NET architecture enhanced with attention mechanisms and eXplainable Artificial Intelligence (XAI) to build one-month temperature prediction models. This served as the basis for additional models, incorporating expanded input data from the Atmosphere-Land Surface-Ocean (ALO) system and leveraging state-of-the-art deep learning techniques. A total of 31 models were constructed, including ensemble learning models, N-step forecasting models, and retraining-based models. These models were evaluated against the ECMWF S2S model using Heidke Skill Score (HSS).

The input variables used in the model design were T2M (atmosphere) and SST (ocean) from the ERA-5 reanalysis dataset, and NDVI (land) from MODIS and AVHRR. These models are 7-day or 14-day inputs and projected 1 day into the future, and for the 3-6 week temperature probability forecasts, a rolling prediction technique was used to build a temperature probability prediction system using the ECMWF S2S model 1-2 week forecast and hindcast data as inputs. Among the input variables of the models, T2M and SST variables are different from the NDVI data, which is only available for 1982-2019, so the evaluation period of the models was divided into 2018-2019 and 2018-2022. For the evaluation period 2018-2022, NDVI data is not available for 2020, so 2019 data was used. The evaluation results show that for the 2018-2019 period, the top-performing model was DL-SE-7+SST+NDVI, followed by DL-SE+FLT2+BTLNCK-7 and NM-SE-7. ECMWF ranked 25th. For the 2018-2022 period, DL-SE-14 ranked first, DL-SE+BIGDATA+FLT2+BTLNCK-14 ranked second, and DL-SE+BIGDATA-14 ranked third, with ECMWF ranking 18th. During El Niño events, the top three models were DL-SE-7+SST+NDVI, DL-SE+FLT2+BTLNCK-7, and DL-SE+BIGDATA-14, with ECMWF ranking 7th. During El Niño events, the top three models were DL-SE-7+SST+NDVI, DL-SE+FLT2+BTLNCK-7, and DL-SE+BIGDATA-14, with ECMWF ranking seventh. Models incorporating the ALO system outperformed others during the 2018-2019 period, suggesting that land (NDVI) and ocean (SST) data positively influenced 4-6 week forecasts. However, for 2020 onwards, the limited availability of NDVI data reduced the predictive power of models utilizing this variable. Overall, SST and NDVI data were found to have a positive impact on the predictability of the models for S2S temperature forecasts.

Improvements of monthly climate predictions were tested by applying a simple data augmentation technique, Cutmix, to a simple convolution neural network (CNN) model. The contribution of the data augmentation method were assessed for the predictions of the summer/winter monthly mean temperature and 1-month Standardized Precipitation Index

(SPII). Summer/winter validation accuracy values for both variables increased with data augmentation during the training phase. Epistemic uncertainty was reduced for the test phase when examined by Monte Carlo dropout. Individual monthly predictions, however, only showed improvements for January (LT2), February (LT3) and August (LT3) for monthly mean temperature, and for July (LT2), August (LT3), and December (LT1) for SPII. Confusion matrices were reviewed; hit rate were improved most of the cases. Although the contribution of simple data augmentation techniques can be limited with overly simplified models even though the cause of poor performance is the lack of training data, some improvements were observed by using a data augmentation technique overall.

To explore AI applications in seasonal climate prediction, this study aims to develop deep learning models for post-processing the Multi-Model Ensemble (MME) forecast of 3-month average precipitation over East Asia. Three different post-processing approaches are applied: ensemble SubSampling (SS), DeBiasing (DB), and Transfer-Learning based Diagnosing (TLD). High-performing architectures and hyperparameters are identified through extensive sensitivity testing. The conventional Quantile Mapping (QM) bias-correction method is used as the baseline for evaluating AI model performance. Among the 12 test seasons in 2023, the DB approach was identified as the best-performing model for 7 seasons, followed by SS for 3 seasons, and TLD and QM each for 1 season. This highlights the effectiveness of AI techniques in seasonal forecast post-processing. However, the significant variability in performance across different post-processing approaches, model architectures, and seasons underscores the need for the MME method to mitigate the uncertainties of individual AI-based post-processing models. Therefore, this study develops codes for calculating both deterministic and probabilistic MMEs of post-processing models and historical performance. It finds that a simple composite mean of 10 post-processing models generally outperforms individual models, with a 40% improvement in ACC compared to the operational APCC-MME. Lastly, an effective visualization layout is designed, potentially introducing a new prototype for an AI-aided APCC-MME.

국문 요약

최근 엔비디아, 화웨이, 구글 딥마인드, 유럽중기예보센터(ECMWF, European Centre for Medium-Range Weather Forecasts) 등 세계적인 기업과 연구소를 중심으로 인공지능 기반 중기 날씨 예측 모델(15일 이내)이 활발히 개발되고 있다. 주요 사례로는 엔비디아의 SFNO (Spherical Fourier Neural Operators) 기반 FourCastNet, 화웨이의 Swin-Transformer 기반 Pangu-Weather, ECMWF의 GNN (Graph Neural Network) 기반 AIFS, 중국 푸단 대학의 U-Transformer 기반 FuXi 모델 및 S2S 예측용 FuXi-S2S, 그리고 구글 딥마인드의 GNN 기반 GraphCast와 디퓨전 모델 기반 GenCsat 모델이 있다. 대부분이 15일 이내 중기 날씨 예측에 국한되어 있으며 계절 내 및 계절 예측으로의 확장은 미흡한 상황이다. 또한 인공지능 학습에는 대량의 자료가 필요하지만, 주요 기후 자료로 사용되는 ECMWF ERA-5 자료는 75년 정도의 자료만 이용 가능하다. 이를 학습, 검증, 테스트용으로 데이터셋을 구축하면 실제 학습에 사용되는 자료는 부족하다는 한계가 있다.

기후위기로 인하여 응용기후분야에 정확도 높은 계절 내 예측자료의 중요성이 커지고 있다. 이러한 계절 내 예측자료는 물리적 기반으로 다양한 대기 및 해양의 상호작용을 수치모델만으로 예측하기 어려운 문제에 직면하고 있다. 이러한 어려움을 극복하기 위하여 자료 확장과 입력자료에서 보다 중요한 패턴과 특징에 가중치를 줘서 정확도를 개선하고자 하는 Attention U-Net 모델과 계절 내 예측 모델 뿐만 아니라 관측자료를 기반으로 Filter, Wrapper, Embedded 등의 기법을 이용하여 변수가 가지고 있는 특징들을 찾아내어 다양한 입출력 시스템에 기반한 민감도 분석을 실시하였다. 모델/관측자료의 변수의 선택으로 일부 변수조합으로 계절내 예측보다 정확도가 향상됨을 확인할수 있었다. GUI 기반 입출력 시스템은 입력자료 생성과 민감도 분석을 용이하게 하여 최적화된 인공지능 기법 개발에 기여하였다. 본 연구는 계절 내 예측 정확도 향상을 위한 기초 자료로 활용될 수 있으며, 향후 딥러닝 모델의 추가적인 개선이 필요하다

최근 신경망 모델의 발전과 딥러닝 기술의 응용은 다양한 산업과 연구 분야에서 주목받고 있으며, 기후 예측에서도 활용 가능성이 증가하고 있다. 특히, S2S(Sub-seasonal to Seasonal) 예측에서는 데이터 전처리 및 모델 성능 개선의 필요성이 강조되고 있다. 본 연구는 앙상블 기법(보팅, 배깅, 부스팅, 스택킹)을 활용하여 S2S 일 최고기온의 예측 성능을 개선하는 데 초점을 맞췄다. 선정된 각각의 앙상블 기법의 특성과 성능을 분석하였고 선례 연구(6단계 1-2차년)에서 구축한 신경망 모델들, CNN, CNN-LSTM, U-Net, Attention U-Net, Residual U-Net을 앙상블 멤버 모델로 구성하였고 표준화를 통해 훈련자료를 전처리하고 각각의 앙상블 기법에서 훈련 및 예측 평가하였다. 연구 결과, ECMWF-S2S TMAX 예측에서는 모든 앙상블 기법의 ACC가 향상되었으며, 특히 배깅이 가장 우수한 성능을 보였다. KMA(GloSea5)-S2S TMAX에서는 스택킹과 부스팅이 선행예측 시간 규모 전체의(1주에서 4주) 예측 성능을 개선하였으나, 보팅과 배깅은 4주(20일) 이후에서만 효과적이었다. 공간 분포 평가에서도 ECMWF에서 앙상블 기법 적용 후 ACC는 선행예측 1주에서 4주 모두 개선되었고, 특히 배깅의 ACC 공간 분포는 훈련 전 ACC 공간 분포와의 차이가 대부분 0으로 관찰되었다. 관심 즉 특정 지역(예, 남한 격자 4지점)에 대한 평가에서 ECMWF보다 KMA에서 성능 향상을 보였다. 이는 앙상블 기법이 훈련 자료

및 지역 특성에 따라 다르게 작용할 수 있음을 시사한다. 본 연구는 S2S 예측의 안정성과 성능을 개선하기 위해 앙상블 기법의 유용성을 입증하였으며, 이를 통해 S2S의 일 최고기온의 3-6주 예측의 정확도 향상과 실용성(예, 이상고온 예측 성능 개선 등)을 높이는 데 기여할 수 있을 것으로 기대한다.

2차년도(2023)에 준지도학습 기법을 활용하여 MJO 위상을 분류하는 이미지 기반 인공지능 모델이 개발되었다. 이를 위해 먼저 다양한 MJO 지수에 따른 입력 변수 조합, 모델 구조, 훈련 데이터 분리 방법 등을 테스트해 지도학습 기반의 최적 MJO 위상 분류 모델을 구축했다. 준지도학습 기반 모델은 라벨링된 데이터가 적은 상황에서도 중요한 기후 패턴을 효과적으로 학습하며, 지도학습 모델과 유사하거나 더 높은 정확도를 보이는 사례를 확인했다. 자료증강 기법이 기후 데이터의 공간적 특성에 맞게 적용될 때 모델 성능을 향상시키는 것을 확인했다. MJO 위상 분류에서 수평 반전은 정확도를 낮추지만, 수직 반전은 성능을 향상시켰다. 3차년도(2024)에는 준지도학습을 확장하여 선행시간별 RMM1과 RMM2 지수를 예측하는 인공지능 모델을 개발했다. ResNet-18을 수정한 모델에 U200, U850, OLR, VP, TS, TCWV, hadvect 기후 변수의 편차와 배경장을 함께 활용하여, BCOR 0.5 기준 최대 약 22일까지 예측 성능을 달성했다. MJO 진폭을 손실함수에 포함해 진폭 관련 오류를 개선하여 인공지능 기반 MJO 연구에서 도메인에 특화된 모델 조정의 중요성을 확인하였다. 또한 해석가능한 인공지능 기법을 통해 바람장의 배경 상태와 수증기 편차가 MJO 예측에 중요한 영향을 미친다는 점을 확인했다. 나아가 RMM 단정예측 인공지능 모델을 기반으로 초기장 섭동과 모델 파라미터 섭동을 활용해 확률예측 모델을 개발했다. Perlin noise를 이용한 초기장 섭동과 MC dropout을 적용한 모델 섭동을 통한 앙상블 분산을 분석한 결과, Perlin noise이 선행시간이 길어질수록 분산을 가장 크게 증가시켰으나, 일부 사례에서는 분산 증가가 제한적이었다. 이는 인공지능 모델이 평균에 수렴하는 특성으로 앙상블 확산을 산출하는 데 어려움이 있을 수 있음을 시사한다. 향후 연구에서는 최신의 인공지능 모델에 배경장과 편차 정보를 통합 활용하고 고해상도 데이터를 활용해 기후의 공간적 패턴과 특성을 더욱 정밀하게 확인하고 예측력을 향상시키는 방안을 모색할 예정이다.

계절 내 및 계절 시간 규모의 기온확률 예측성을 높이기 위해 딥러닝 기반의 동아시아지역 3-6주 기온 확률예측 모델을 개발하고자 하였다. U-NET 아키텍처 기반으로 어텐션 메커니즘, 설명가능한 인공지능(XAI, eXplainable Artificial Intelligence) 등을 활용하여 1개월 기온 예측 모델을 구축하고, 이 모델을 기초로 Atmosphere-Land Surface-Ocean (ALO) 시스템을 기초로 입력자료 확장을 고려한 모델, 최신 딥러닝 아키텍처를 이용한 기온 예측 모델, 앙상블 학습 기법을 이용한 기온 예측 모델, N 스텝 예측 방법을 이용한 기온 예측 모델, 재학습 방법을 이용한 기온 예측 모델 등 동아시아지역 3-6 기온확률 예측을 위해 총 31개 모델을 구축하고 ECMWF S2S 모델과 HSS 방법으로 비교 평가하였다. 모델 설계에 사용한 입력 변수는 ERA-5 재분석자료의 T2M(대기)과 SST 자료(해양), MODIS와 AVHRR 자료로 구축한 NDVI 자료(지면)를 이용하였다. 이 모델들은 14일 입력자료로 미래 1일을 맞추는 모델로, 3-6주 기온확률 예측을 위해 순환 예측(rolling prediction) 기법을 이용하여 ECMWF S2S 모델의 forecast와 hindcast 1-2주 예측자료를 입력자료로 하는 기온확률 예측 체계를 구축하였다. 모델의 입력 변수 중 NDVI 자료는 T2M과 SST 변수는 달리 1982~2019년 자료만 구축된 데이터셋을 이용하였기 때문에 모델의 평가 기간을 2018~2019년과 2018~2022년으로 나누어 수행하였다. 평가기간 2018~2022년은 NDVI 자료는 2020년 이후 자료는 없기 때문에 2019년 자료를 이용하였다. 평가

결과, 평가기간 2018~2019년은 1위 DL-SE-7+SST+NDVI, 공동 2위 DL-SE+FLT2+BTLNCK-7, NM-SE-7, EM-NP-14, nP-DL-SE+BIGDAA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI, RL-DL-SE-14+SST+NDVI였으며, ECMWF는 25위였다. 평가기간 2018~2022년은 전체 평가 기간에서 총 32개 모델(ECMWF 포함) 중 1위 DL-SE-14, 2위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 3위 DL-SE+BIGDATA-14였으며, ECMWF는 18위였다. 엘니뇨 발생 시기는 공동 1위 DL-SE-7+SST+NDVI, DL-SE+FLT2+BTLNCK-7. 3위 DL-SE+BIGDATA-14이며 ECMWF는 공동 7위였다. 2018~2019년은 ALO 시스템을 기초로 입력자료 확장을 고려한 모델들이 주로 좋은 예측성을 보였다. 이는 지면자료(NDVI)와 해양자료(SST) 자료가 4~6주 예측에 긍정적인 영향을 준 것으로 판단된다. 반면 2018~2022년은 NDVI 자료 구축 기간이 2018~2019년으로 2020년부터는 2019년 자료를 이용하였기 때문에 T2M 자료만 이용한 모델에 비해 상대적으로 예측성이 떨어지는 경향을 보였다. ALO 시스템 관점에서 계절 내 및 계절 기온 예측에서 SST와 NDVI 자료가 모델의 예측성에 긍정적인 영향을 미치는 것으로 나타났다.

단순한 인공지능모델에 간결한 자료강화기법을 적용하여도 월별 기후예측을 개선할 수 있을지 가능성을 타진하고자 하였다. 합성곱신경망모델(CNN)을 구축하여 여름철과 겨울철, 평균기온과 강수(SPI1) 예측에 대한 Cutmix 자료강화의 개선 성능을 평가하였다. 자료강화한 경우 훈련 과정에서의 검증 정확도는 여름철, 겨울철 그리고 평균기온 및 강수(SPI1)에 대해 모두 증가하였으며, Monte Carlo Dropout 적용하여 살펴본 결과 최종 예측 결과의 epistemic 불확실성은 감소하였다. 그러나 월별로 보았을 때 평균기온은 1월(LT2), 2월(LT3), 8월(LT3), 그리고 강수(SPI1)는 7월(LT2), 8월(LT3), 12월(LT1)의 경우에만 개선이 관찰되었다. Confusion matrix를 통해 개별 예측 결과를 전체적으로 검토하였는데, 자료강화한 경우 대부분의 경우 적중률이 증가하였음을 확인하였다. 결론적으로 평균기온이나 강수(SPI1) 예측 부진의 주요 원인이 자료 부족이더라도 지나치게 단순한 모델로는 자료강화의 효과가 제한적일 수 있으나 전반적으로 개선의 효과를 확인할 수 있었다.

인공지능 기술의 계절 예측 분야 활용성을 탐색하고자 다중모델앙상블(Multi-Model Ensemble, MME) 예측, 구체적으로 동아시아 3개월 평균 강수에 대한 후처리모델 개발 업무를 수행하였다. 이를 위해 성능이 우수할 것으로 추정되는 앙상블만을 선별하는 방식(SubSampling, SS), 예측장 오차를 제거하는 방식(DeBiasing, DB), 전이 학습 기반으로 강수를 진단(Transfer-Learning based Diagnosing, TLD) 방식으로 인공지능 모델을 구현하고, 전통적인 편차보정 방식(Quantile Mapping, QM)를 기초선으로 성능 개선 평가를 진행하였다. 테스트 기간 2023년 12 계절 전체에서 많은 후처리 모델이 현업 APCC-MME 성능을 넘어섰고, 구체적으로 DB가 7개, SS가 3개, TLD/QM이 각각 1개 시즌에서 최고 성능을 보여 인공지능 기술 적용의 타당성/우수성을 확인하였다. 하지만 후처리 방식별, 계절별, 버전별로 성능 변동성이 매우 크기 때문에, 이러한 불확실성을 극복하고자 후처리 모델에 대해서도 MME 기법을 적용해 보았다. 성능면에서 개별 후처리 모델들보다 전반적으로 우수하고, 현업 역학 모델 MME에 비해서는 약 40% ACC 개선을 확인할 수 있었다. 최종적으로 다중 후처리 모델 합산 예측장과 과거 성능 평가 정보에 대한 통합 표출 모듈을 개발함으로써 인공지능의 도움을 받는 새로운 형태의 APCC-MME 원형을 제시하였다.

목차

Executive Summary	i
국문 요약	v
1. 서론	1
2. 방법 및 자료	9
2.1. 자료 확장에 따른 모델 성능의 민감도 분석 및 입출력 시스템 설계	9
2.1.1. 자료	9
2.1.2. 방법	15
2.1.2.7 분석 및 검증	21
2.2. 전처리/초매개변수 기반 다중 인공지능 모델 도출 및 모델간 앙상블	28
2.2.1. 앙상블 기법 선정 및 특징 분석	30
2.2.2. 앙상블 기법 위한 멤버 모델 구성	34
2.2.3. 계절내 기후 예측 훈련자료	41
2.2.4. 구동 환경, 연구 지역 및 평가 방법	42
2.3. 준지도학습을 통한 MJO 지수 확률예측 인공지능모델 개발	45
2.3.1. 사용된 자료 및 전처리 방법	45
2.3.2. 준지도학습을 적용한 MJO 위상 분류 모델 개발 방법	47
2.3.3. 선행시간별 MJO RMM 지수 확률예측 인공지능 모델 개발 방법	52
2.3.4. MJO 확률예측을 위한 인공지능 모델 구축 방법	54
2.4. 3-6주 동아시아 기온확률 예측 인공지능 모델 개발	56
2.4.1. 연구자료	56
2.4.2. 3-6주 동아시아 기온확률 예측 인공지능 모델 개발	59
2.4.3. 피쳐붕괴 분석을 통한 딥러닝 모델 분석	65
2.5. 자료강화기법을 이용한 인공지능모델의 월별 기후예측 개선	68
2.5.1. 자료 수집 및 전처리	68
2.5.2. 자료강화기법 적용	68
2.5.3. 인공지능모델 및 실험 구성	69
2.6. 인공지능 기술을 활용한 다중모델 앙상블 강수 역학예측 후처리	71
2.6.1. 훈련 자료 및 전처리 방법	71
2.6.2. 후처리 대상 지역 및 성능 평가 방법	74
2.6.3. 역학 모형 후처리 접근 방식 및 딥러닝 모형	74

3. 결과	82
3.1. 자료 확장에 따른 모델 성능의 민감도 분석	82
3.1.1. 강수자료 학습결과	82
3.1.2. Filter 기법에 따른 민감도 분석 결과	82
3.1.3. Wrapper 기법에 따른 민감도 분석 결과	84
3.1.4. Embedded 기법에 따른 민감도 분석 결과	86
3.2. 입출력 시스템 고도화	87
3.2.1. 딥러닝 모델 학습자료 산출 시스템	87
3.2.2. 딥러닝 모델 학습자료 산출 자료 검증	88
3.2. 전처리/초매개변수 기반 다중 인공지능 모델 도출 및 모델간 앙상블	91
3.2.1. 4중 앙상블 기법 구축	92
3.2.2. 4중 앙상블 기법의 학습 성능	94
3.2.3. 앙상블 기법 내 멤버 모델 조합 변화에 따른 학습 성능	99
3.2.4. 격자에 대한 앙상블 기법의 학습 성능	100
3.3. 준지도학습을 통한 MJO 지수 확률예측 인공지능모델 개발	101
3.3.1. 준지도학습 활용 MJO 위상 분류를 위한 인공지능모델 개발	101
3.3.2. 선행시간별 MJO RMM 지수 예측을 위한 인공지능모델 개발	107
3.3.3. 선행시간별 MJO 확률 예측을 위한 인공지능모델 개발	116
3.4. 3-6주 동아시아 기온확률 예측 인공지능 모델 개발	119
3.4.1. 기준 1개월 기온 예측 모델 실험 및 내부 구조 분석	119
3.4.2. 입력 변수 확장을 통한 기온 예측 모델 실험 및 결과	131
3.4.3. 최신 딥러닝 아키텍처를 이용한 기온 예측 모델 실험 및 결과	134
3.4.4. 앙상블 학습 기법을 이용한 기온 예측 모델 실험 및 결과	137
3.4.5. N 스텝 예측 방법을 이용한 기온 예측 모델 실험 및 결과	144
3.4.6. 재학습 방법을 이용한 기온 예측 모델 실험	149
3.4.7. 종합분석	154
3.5. 자료강화기법을 이용한 인공지능모델의 월별 기후예측 개선	157
3.5.1. 자료강화기법 및 훈련자료 활용범위 비교	157
3.5.2. 여름철과 겨울철 평균기온 및 강수(SPI1) 예측	161
3.5.3. 역학모델 결과 활용 효과 검토	164
3.6. 인공지능 기술을 활용한 다중모델 앙상블 강수 역학예측 후처리	165
3.6.1. 설명가능인공지능(XAI) 기술적용을 위한 환경구축 및 시범테스트	165

3.6.2. AI 후처리 접근법 1: 앙상블 선별 (SubSampling, SS) 및 예측 성능 원천 분석	171
3.6.3. AI 후처리 접근법 2: 오차 제거 (DeBiasing, DB) 방식의 아키텍처 민감도 평가	182
3.6.4. AI 후처리 접근법 3: 전이 학습 기반 강수 진단 (Transfer-Learning based Diagnosing, TLD)	185
3.6.5. 기존 통계 편이 보정법: Quantile Mapping 적용성 평가	187
3.6.6. AI-aided 다중모델앙상블 예측 프로토타입 제안	188
4. 요약 및 결론	192
사사	198
References	199
Appendix A. Deep learning models developed for 1-month temperature prediction	206
Appendix B. Examples of the results of top-ranked deep learning-based 3-6 week temperature probability forecast models in 2018-2018 and 2018-2022 evaluation periods.	207

표 차례

Table 2.1. Data description of ERA5.	14
Table 2.2. Overview of the characteristics of ERA5.	14
Table 2.3. Comparison and analysis of the Characteristics of 4 Ensemble Techniques.	29
Table 2.4. Explain the types and main characteristics of the member models that participated in the ensemble technique in this study.	34
Table 2.5. S2S (Sub-seasonal to Seasonal) climate models used in this study, number of ensembles, forecasting time range, length, and period of hindcast of each models.	42
Table 2.6. Information of hyper-parameters and their ranges set in the study.	43
Table 2.7. List of experiments for sensitivity analysis of SSL-based AI model on model learning strategy, loss weights, and data augmentation methods.	51
Table 2.8. Perturbation methods used in this study for MJO RMM probabilistic prediction.	55
Table 2.9. Summary of ALO (Atmosphere-Land-Ocean) data used in this study.	57
Table 2.10. Summary of Dataset.	58
Table 2.11. Dynamical seasonal prediction models used for developing an AI model to enhance EA rainfall forecast accuracy.	73
Table 2.12. Sensitivity test setting for the SS approach.	77
Table 2.13. Sensitivity test setting for the DB approach.	79
Table 3.1. Index and variables of subsets for filter method.	83
Table 3.2. Index and variables of subsets for wrapper method.	85
Table 3.3. Index and variables of subsets for embedded method.	86
Table 3.4. Summary of ECMWF S2S Model and baseline deep-learning-based 1-month temperature prediction models.	120
Table 3.5. Summary of improved deep learning models using feature pair-wise cosine distance analysis.	128
Table 3.6. Summary of deep-learning-based 1-month temperature prediction models with NDVI (land) and SST climatology (ocean) variables.	132
Table 3.7. Summary of modified Stage-N architecture-based models for 1-month temperature prediction.	134
Table 3.8. Summary of deep learning-based ensemble models for 1-month temperature	

prediction.	137
Table 3.9. Summary of deep learning-based n-step prediction models for 1-month temperature prediction.	144
Table 3.10. Summary of 1-month temperature prediction deep learning models.	149
Table A.1. Summary of the ECMWF S2S Model and 1-month temperature prediction deep learning models.	206

그림 차례

Figure 1.1. (a) Supervised learning, (b) semi-supervised learning, (c) unsupervised learning. Colored circles are labeled data and gray ones are unlabeled data.	4
Figure 1.2. Diagram of Pi model (Source: Laine and Aila, 2016).	4
Figure 1.3. Comparison of Pseudo Labels (left) and Meta Pseudo Labels (right) with the feedback from the Student model (Source: Pham et al., 2021).	5
Figure 2.1. NOAA NDVI(Normalized Difference Vegetation Index data form AVHRR satellite. 9	9
Figure 2.2. GPM IMERG product.	10
Figure 2.3. NOAA CMORPH product of daily precipitation.	11
Figure 2.4. The PERSIANN rainfall retrieval algorithm of NOAA CDR program.	11
Figure 2.5. Sea surface temperature data from NOAA.	12
Figure 2.6. ERA5 Maximum temperature at 2 meters.	13
Figure 2.7. Overview of sensitivity analysis and input/output systems.	15
Figure 2.8. Sensitivity analysis of feature selection from in and out data system.	15
Figure 2.9. A block diagram of Attention U-Net architecture(Source: Ozan Oktay et al., 2018). 16	16
Figure 2.10 Schematic of the attention gate(Source: Ozan Oktay et al., 2018)	17
Figure 2.11. A diagram of input data from observation and ECMWF model data for Attention U-Net model.	17
Figure 2.12. Setup of attention U-net model.	18
Figure 2.13. Sensitivity analysis methods for deep learning models.	19
Figure 2.14. Correlation heatmap of observation and ECMWF model variables.	20
Figure 2.15. Scatterplot of correalation of variables.	20
Figure 2.16. A diagram for embedded method for feature selection(Source: Trabelsi et al., 2021).	21
Figure 2.17. The structure of in and out data system.	22
Figure 2.18. GUI of in and out data system.	23
Figure 2.19. Advanced Input/Output System module diagram.	24
Figure 2.20. Refinement data range by calendar python library.	24
Figure 2.21. Libraries of python code for refinement of regions.	25

Figure 2.22. Webbrowser for refinement of regions.	25
Figure 2.23. Map editing Toolbar with save, cancel, and clear all options.	26
Figure 2.24. Refine region by toolbar.	26
Figure 2.25. Export button.	27
Figure 2.26. Data scaling methods for machine learning.	27
Figure 2.27. Data splitting for training datasets.	27
Figure 2.28. Explanation of the voting technique. Data is entered into m4 from various models m1 to m4 and voted on the predicted data (blue circles) to output the final predicted data.	30
Figure 2.29. Explanation of bagging technique. D1 to d4 are random datasets made by bootstrapping segmentation in Data, entered into m1 to m4, voting on predicted data (blue circles) to derive final predicted data.	31
Figure 2.30. Description of boosting technique. Data is input to m1, weights are added to the predicted results (d') and input to the next model m2. Weights (d “) are added again to the predicted results in m2, and the predicted data is derived from the final model.	32
Figure 2.31. Description of stacking technique. Data is input from m1 to m4, predicted results are stacked (Prediction Stack), and the final prediction is output from m5 (meta model).	33
Figure 2.32. Description of the convolutional neural network model (Source: PhilAI.log: “[DL] CNN (Convolution Neural Networks) Concept Explained by a Non-Specialist“ Blog). 35	
Figure 2.33. Diagram of the Long-Short Term Memory convolutional neural network model developed in this study.	36
Figure 2.34. Residual learning block. The left side is the existing method, and the right side is the residual block with skip connection added.	38
Figure 2.35. U-Net model built from previous research. Convolutional layers are wrapped with TimeDistributed layers (source: Chung et al., 2023).	39
Figure 2.36. Attention U-Net developed in a previous study. A model that improves U-Net by inserting an attention block into the expansion process to better extract spatial characteristics. Attention U-Net: (a) is the original Attention U-Net (Source: Oktay et al., 2018), and (b) is the Attention U-Net schematized briefly in the study. ••	40

Figure 2.37. 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.43

Figure 2.38. Steps of data preprocessing for filtered anomaly data.46

Figure 2.39. Composite images of the filtered OLR anomaly (W/m²) for each MJO phase for training (left), validation (center), and test (right) datasets, respectively. The number of days available for each phase are given in the bottom right corner. Negative (Positive) anomalies indicate more (less) rain.47

Figure 2.40. Data split strategy used for semi-supervised learning-based MJO phase classification. (a) Temporal split, (b) Random split method.47

Figure 2.41. Diagram of modified VGG16 model used in this study.48

Figure 2.42. Residual learning structure of ResNets model (Source: He et al., 2016)49

Figure 2.43. Diagram of modified ResNet-18 model and different versions of fully connected layers used in this study.49

Figure 2.44. Diagram of Mean Teacher algorithm (Source: Tarvainen and Valpola, 2018)50

Figure 2.45. Data augmentation methods used in Mean Teacher algorithm.51

Figure 2.46. Diagram of semi-supervised learning-based ResNet-18 model architecture with the version 3 of the fully connected layer used in this study.52

Figure 2.47. An example of perturbed input data using Perlin noise perturbation method.54

Figure 2.48. Schematic diagram of Monte Carlo dropout (Source: Van Katwyk, et al., 2023) 55

Figure 2.49. A schematic of the role of ALO (Atmosphere-Land-Ocean) systems as a source of S2S predictability (Adapted from Mariotti et al. (2018)).56

Figure 2.50. Ratio of training, validation, test data of TDSV32_SHUFFLE dataset.58

Figure 2.51. A Schematic Diagram of the 1-Month Temperature Prediction Model.59

Figure 2.52. The overview of SE block (adapted from Hu et al. (2020)).60

Figure 2.53. The overview of CBAM (adapted from Woo et al. (2018)).60

Figure 2.54. A overlay map of the Korean Peninsula topography and gridded data of ERA-5 reanalysis data: Jeju island (red circle).61

Figure 2.55. Results of NDVI Data Preprocessing: (a) Original NDVI Data, (b) Bilinear Resampling, (c) Average Resampling, (d) Regridding (Including Average Sampling).

Figure 2.56. A Schematic Diagram of the 1-Month Temperature Prediction Model Redesigned

with Two New Input Data Sets.	62
Figure 2.57. A Schematic Diagram of the 1-Month Temperature Prediction Model Redesigned with Two New Input Data Sets.	62
Figure 2.58. A schematic diagram of the 1-month temperature prediction model redesigned with the stage-N block of RDNet.	63
Figure 2.59. A schematic diagram of the 1-month temperature prediction model redesigned with the stage-N mlock of RDNet and two new input data sets.	63
Figure 2.60. Schematic diagrams of ensemble learning models: (a) simple prediction layer (b) CCA prediction layer.	64
Figure 2.61. Example of a schematic diagram of a re-learning model.	65
Figure 2.62. A schematic diagram of re-learning model.	65
Figure 2.63. Visualization of activation map (input images obtained from https://github.com/ismailuddin/gradcam-tensorflow-2).	66
Figure 2.64. An illustrating of 5 points in South Korea.	67
Figure 2.65. Matrix of Forecast and Observed category.	67
Figure 2.66. Structure of the 3D CNN toy model used in this study.	69
Figure 2.67. Structure of the GNN model for Node classification used in this study.	70
Figure 2.68. Structure of the GNN model for Graph classification used in this study.	70
Figure 2.69. AI-based and conventional statistical approaches for enhancing EA seasonal rainfall forecast accuracy. SS: sub-sampling, DB: debiasing, TLD: transfer-learning-based diagnosing, QM: quantile mapping.	75
Figure 2.70. Illustration of the ensemble subsampling (SS) approach: (A) training input, (B) deep-learning architecture, (C) ACC metric value as the label, (D) selection of ensembles with high predicted ACC for the new ensemble mean forecast.	77
Figure 2.71. Illustration of the debiasing (DB) approach: (A) training input, (B) deep-learning architecture, (C) 2-dimensional rainfall field as the label.	78
Figure 2.72. Second DB approach using Stacked Two-models (ST): one is a model for debiasing EA lower-level zonal wind field (left) and the other is a model for debiasing EA rainfall field (right).	80
Figure 2.73. Process of transfer learning for diagnosing EA rainfall.	81
Figure 2.74. Physical variables used as training input for the TLD approach.	81
Figure 3.1. Comparison of the result of Attention U-Net and ECMWF model.	82

Figure 3.2. Trained results obtained from variables based on filter method.	83
Figure 3.3. Datasets of category for wrapper method.	84
Figure 3.4. Comparison of the result of Attention U-Net by Wrapper method.	85
Figure 3.5. Comparison of the result of all variables by Wrapper method.	86
Figure 3.6. Comparison of the result of all variables by Embedded method.	87
Figure 3.7. Advanced Input/Output System.	88
Figure 3.8. The results of advanced input/output system.	89
Figure 3.9 Meta data for verification of output data by advanced input/output system.	89
Figure 3.10. Output data by scaled method.	90
Figure 3.11. Raw data by Advanced input/output system.	90
Figure 3.12. Visualization of output data for verification.	91
Figure 3.13. Steps of outcome by advanced input/output system.	91
Figure 3.14. Voting technique developed in this study.	92
Figure 3.15. Bagging technique developed in this study.	93
Figure 3.16. Boosting technique developed in this study.	93
Figure 3.17. Stacking technique developed in this study.	94
Figure 3.18. Test results of four ensemble techniques (stacking, boosting, voting, bagging) applied to daily maximum temperatures from ECMWF-S2S. The ACC (Anomaly Correlation Coefficient) for the 46-day time scale of the forecast lead-time(left), the PCC (spatial Pattern correlation coefficient) for 1-4 weeks of the forecast lead-time as weekly average(right). ECMWF is ACC and PCC before training. ..	95
Figure 3.19. Test results of four ensemble techniques (stacking, boosting, voting, bagging) applied to daily maximum temperatures from KMA-S2S. The left is the ACC (Anomaly Correlation Coefficient) for the 46-day time scale of the forecast lead-time, the right is the PCC (spatial Pattern correlation coefficient) for 1-4 weeks of the forecast lead-time as weekly average. KMA is ACC and PCC before training.	95
Figure 3.20. For the four ensemble techniques (stacking, boosting, voting, and bagging) applied to the daily maximum temperature of ECMWF-S2S, the spatial distribution of ACC for 1-4 weeks of advance forecasts was calculated, and the difference from the spatial distribution of ACC before training at each time scale was expressed as a heat map.	96

- Figure 3.21. For the five member models (CNN, CNN-LSTM, ResNet, U-Net, Attention U-Net) that participated in the ensemble technique of the daily maximum temperature of ECMWF-S2S, the spatial distribution of ACC was calculated from the results predicted for 1-4 weeks after individual training, and the difference in the spatial distribution of ACC before training was expressed as a heat map.97
- Figure 3.22. For the four ensemble techniques (stacking, boosting, voting, and bagging) applied to the daily maximum temperature of KMA-S2S, the spatial distribution of ACC for 1-4 weeks of advance forecasts was calculated, and the difference from the spatial distribution of ACC before training at each time scale was expressed as a heat map.98
- Figure 3.23. For the four member models (CNN, CNN-LSTM, U-Net, Attention U-Net) that participated in the ensemble technique of the daily maximum temperature of KMA-S2S, the spatial distribution of ACC was calculated from the results predicted for 1-4 weeks after individual training, and the difference in the spatial distribution of ACC before training was expressed as a heat map.99
- Figure 3.24. Comparison of PCC changes according to member model combinations in four ensemble techniques applied for the postprocessing of daily maximum temperatures of ECMWF-S2S (left) and KMA-S2S (right) for 1-4 weeks of the forecast lead-time.100
- Figure 3.25. Comparison of the average ACC at four grid points in South Korea for 1-4 weeks of the forecast lead-time in the ACC spatial distribution of four ensemble techniques applied for the postprocessing of daily maximum temperatures of ECMWF-S2S (left) and KMA-S2S (right).101
- Figure 3.26. Summary of model skill for MJO phase classification for the test data. Colored bars of each panel show the test accuracy (%) of the models of the present study, and the ANN (Martin et al., 2022) and CNN (Delaunay et al., 2022) models are represented by dotted and dashed lines, respectively. Left column panels are the results with RMM input variables (U850, U200, OLR), the center ones with VPM variables (U850, U200, VP), and the right ones with OLR only. The upper row demonstrates the results with temporal split data and the lower one with random split data.102
- Figure 3.27. Ablation experiment results in terms of (a-d) hyperparameters such as batch

size, consistency cost weight, consistency ramp-up, and EMA decay and (e) data augmentation used in the Mean Teacher algorithm in classifying MJO phases in the test data. The ResNet-18 model with the original version of the FC layer was trained on the basis of U850, U200, and OLR with 4500 labeled data.104

Figure 3.28. Performance of semi-supervised learning models for MJO phase classification on test data with different numbers of labeled data with comparison to a supervised Resnet-18 model using all labeled data and previous studies including ANN from Martin et al., 2022 (ANN(M22)) and CNN from Delaunay and Christensen, 2022 (CNN(D22)). The colored lines show the average accuracy of the 10 independent runs with different random seeds with SSL-based Resnet-18 model for three input fields (RMM, VPM, and OLR). The shaded regions denote the standard deviation of the runs.105

Figure 3.29. Monthly and seasonal SSL-based model performance heat maps with the test data for MJO phase classification for (upper row) RMM and (lower row) VPM index. The number of errors in each grid shows the MJO events misclassified by the model.106

Figure 3.30. Monthly and seasonal SSL-based model performances with the test data for MJO phase classification grouped by active (orange) and inactive (blue) MJO cases. Error cases on the y-axis indicate MJO events misclassified by the model.107

Figure 3.31. Comparisons of semi-supervised learning (SSL)-based and supervised learning-based models in predicting MJO with BCOR on test data over lead times with a 5-day interval.108

Figure 3.32. Summary of the model skill (BCOR) on test data in predicting MJO over lead times with a 5-day interval for different predictor sets. In a and b, All (black dashed line) means using all input predictors including both MJO anomaly and Background predictors. a, Raw (red solid line) denotes using only the original raw predictors and Background (yellow solid line) using only the raw ones from which the anomaly was subtracted. b, MJO Anomaly (gray solid line) indicates using only the filtered predictors.109

Figure 3.33. Model performance of bivariate correlation coefficients (BCOR) on test data for winter (November–April) and summer (May–October).111

Figure 3.34. Model performance of bivariate correlation coefficients (BCOR) on test data for

winter (November–April) and summer (May–October) using the ResNet-18 model trained on only MJO anomaly input variables.	111
Figure 3.35. Model performance of bivariate correlation coefficients (BCOR) on test data for moderate (MJO amplitude 1–1.5) and strong MJO events (amplitude > 1.5).	112
Figure 3.36. Errors of MJO amplitude (a) and phase (b) at lead times when applying customized loss function. The error is obtained by subtracting the predicted values from the model and the true values. Original shows the baseline model with no modifications on model architecture and loss function, and Modified the finalized model of the present study.	113
Figure 3.37. Relative contributions of input features for both RMM1 and RMM2 by simple spatial average over lead times with a 5-day interval.	115
Figure 3.38. Attribution maps of input variables (Anomaly and Background variables) for RMM1 at a 15-day lead time.	115
Figure 3.39. Attribution maps of TCWV Anomaly and U850 Background variables for lead times of 5, 10, 15, 20, 25 day.	116
Figure 3.40. Ranked probability score (RPS) for MJO phases for different perturbation methods at lead times of 0 to 30 days.	117
Figure 3.41. Ensemble results with different perturbation methods for MJO RMM probabilistic predictions with 10 ensembles with ERA5 ensemble data. Solid black line indicates the ensemble mean, and the gray line the ERA5 reference data.	117
Figure 3.42. Ensemble results with different perturbation methods for MJO RMM probabilistic predictions with 50 ensembles with Perlin noise perturbation. Solid black line indicates the ensemble mean, and the gray line the ERA5 reference data.	118
Figure 3.43. Ensemble results with different perturbation methods for MJO RMM probabilistic predictions with 50 ensembles with monte carlo dropout perturbation.	118
Figure 3.44. Probabilistic predictions of MJO phases produced by predicted MJO RMMs with 50 ensembles of Perlin noise perturbation.	119
Figure 3.45. Plot of model training results by batch size.	121
Figure 3.46. Plot of the comparison of the ACC by season.	122
Figure 3.47. Plot of the comparison of the ACC by year.	122
Figure 3.48. Summary of the HSS assessment for lead time of 3 to 4 weeks (F1 to F2): (from left) ECMWF S2S model and 6 baseline models (DL-SE-14, DL-SE+BIGDATA-14,	

DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA-14,
DL-CBAM+BIGDATA+FLT2-14).123

Figure 3.49. Summary of the HSS assessment for lead time of 3 week (F1): (from left) ECMWF S2S model and 6 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14).124

Figure 3.50. Summary of the HSS assessment for lead time of 4 week (F2): (from left) ECMWF S2S model and 6 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14).124

Figure 3.51. Line chart of training loss of a deep learning model.125

Figure 3.52. Feature pair-wise cosine distance analysis of a deep learning model.125

Figure 3.53. The results of the feature pair-wise cosine distance analysis of the deep learning models with the SE block.126

Figure 3.54. The results of the feature pair-wise cosine distance analysis of the deep learning models with the CBAM block.126

Figure 3.55. Design of CCA block.127

Figure 3.56. The results of the feature pair-wise cosine distance analysis of the deep learning models with the SE block after changing the bottleneck block.127

Figure 3.57. The results of the feature pair-wise cosine distance analysis of the deep learning models with the CBAM block after changing the bottleneck block.127

Figure 3.58. Summary of the HSS assessment for lead time of 3 to 4 weeks (F1 to F2): (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BLTNCK-14).128

Figure 3.59. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BLTNCK-14).129

Figure 3.60. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during

El Nino events: (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BLTNCK-14).130

Figure 3.61. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during

La Nina events: (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BLTNCK-14).131

Figure 3.62. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from

left) ECMWF S2S model and 10 models with additional input variables (DL-SE-7, DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7, DL-SE+FLT1+BTLNCK-7+SST, DL-SE+FLT2+BTLNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7, DL-CBAM-7+SST+NDVI, DL-CBAM+FLT1+BTLNCK-7).133

Figure 3.63. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during

El Nino events: (from left) ECMWF S2S model and 10 models with additional input variables (DL-SE-7, DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7, DL-SE+FLT1+BTLNCK-7+SST, DL-SE+FLT2+BTLNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7, DL-CBAM-7+SST+NDVI, DL-CBAM+FLT1+BTLNCK-7).133

Figure 3.64. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during

La Nina events: (from left) ECMWF S2S model and 10 models with additional input variables (DL-SE-7, DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7, DL-SE+FLT1+BTLNCK-7+SST, DL-SE+FLT2+BTLNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7, DL-CBAM-7+SST+NDVI, DL-CBAM+FLT1+BTLNCK-7).134

Figure 3.65. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from

left) ECMWF S2S model and 4 new models (NM-SE-7, NM-SE-7+SST+NDVI, NM-SE-7-RELU+SST+NDVI, NM-SE-14).135

Figure 3.66. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during

El Nino events: (from left) ECMWF S2S model and 4 new models (NM-SE-7, NM-SE-7+SST+NDVI, NM-SE-7-RELU+SST+NDVI, NM-SE-14).136

Figure 3.67. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S model and 4 new models (NM-SE-7, NM-SE-7+SST+NDVI, NM-SE-7-RELU+SST+NDVI, NM-SE-14).136

Figure 3.68. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S Model ECMWF S2S model and 5 ensemble models (EM-7, EM-NP-7, EM-NP-FLT2-7, EM-TR-7, EM-NP-14).138

Figure 3.69. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Nino events: (from left) ECMWF S2S Model ECMWF S2S model and 5 ensemble models (EM-7, EM-NP-7, EM-NP-FLT2-7, EM-TR-7, EM-NP-14).138

Figure 3.70. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S Model ECMWF S2S model and 5 ensemble models (EM-7, EM-NP-7, EM-NP-FLT2-7, EM-TR-7, EM-NP-14).139

Figure 3.71. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE+7+SST), and 4 ensemble models (EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7).140

Figure 3.72. Summary of HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Nino events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE+7+SST), and 4 ensemble models (EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7).141

Figure 3.73. Summary of HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE+7+SST), and 4 ensemble models (EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7).141

Figure 3.74. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14), and 1 ensemble model (EM-NP-14).142

Figure 3.75. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Nino events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14), and 1 ensemble model (EM-NP-14).143

Figure 3.76. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14), and 1 ensemble model (EM-NP-14).143

Figure 3.77. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI), and 1 n-step prediction model (, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI).145

Figure 3.78. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Nino events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI), and 1 n-step prediction model (, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI).146

Figure 3.79. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI), and 1 n-step prediction model (, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI).146

Figure 3.80. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7), and 1 n-step prediction model (nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7).148

Figure 3.81. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Nino events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7), and 1 n-step prediction model (nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7).148

Figure 3.82. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7), and 1 n-step prediction model (nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7).149

Figure 3.83. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 1 pretrained model (DL-CBAM-BIGDATA+FLT2+BTLNCK-14), and 1 re-learning model (RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14).150

Figure 3.84. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Nino events: (from left) ECMWF S2S model, 1 pretrained model (DL-CBAM-BIGDATA+FLT2+BTLNCK-14), and 1 re-learning model (RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14).151

Figure 3.85. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S model, 1 pretrained model (DL-CBAM-BIGDATA+FLT2+BTLNCK-14), and 1 re-learning model (RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14).151

Figure 3.86. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 1 pretrained model (DL-SE-14+SST+NDVI), and 1 re-learning model RL-DL-SE-14+SST+NDVI).152

Figure 3.87. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Nino events: (from left) ECMWF S2S model, 1 pretrained model (DL-SE-14+SST+NDVI), and 1 re-learning model (RL-DL-SE-14+SST+NDVI).153

Figure 3.88. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Nina events: (from left) ECMWF S2S model, 1 pretrained model (DL-SE-14+SST+NDVI), and 1 re-learning model (RL-DL-SE-14+SST+NDVI).153

Figure 3.89. Results of deep learning-based models in 2018-2019.155

Figure 3.90. Results of deep learning prediction models in 2018-2022 (refer to figure 3.89 for this caption).155

Figure 3.91. Results of deep learning prediction models during El Nino events in 2018-2022 (refer to figure 3.89 for this caption).156

Figure 3.92. Results of deep learning prediction models during La Nina events in 2018-2022 (refer to figure 3.89 for this caption).156

Figure 3.93. CNN validation accuracy (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.157

Figure 3.94. CNN validation accuracy (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.158

Figure 3.95. GNN Graph classification validation accuracy between the use of months and data augmentation methods.158

Figure 3.96. CNN Heidke Skill Score (HSS) between data pre-processing and augmentation methods.159

Figure 3.97. GNN Node Classification Heidke Skill Score (HSS) between the use of months and data augmentation methods.159

Figure 3.98. GNN Heidke Skill Score (HSS) between the use of months and data augmentation methods.160

Figure 3.99. Class activation map before (left) and after (right) the data augmentation.160

Figure 3.100. Class activation map before (left) and after (right) the data augmentation.161

Figure 3.101. Validation accuracy of TMm (left) and SPI1 (right) during the training phase. 162

Figure 3.102. Entropy of test set prediction results of TMm (left) and SPI1 (right).162

Figure 3.103. HSS of test set prediction results of TMm (left) and SPI1 (right).162

Figure 3.104. PC of test set prediction results of TMm (left) and SPI1 (right).163

Figure 3.105. Confusion matrices of test set prediction results of TMm (upper) and SPI1 (lower).163

Figure 3.106. Confusion matrices of test set prediction results of OBS_short (left) and OBS_long (right).164

Figure 3.107. HSS of test set prediction results of OBS_short (left) and OBS_long (right). ...165

Figure 3.108. 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).167

Figure 3.109. Correlation coefficients between kernel structures.167

Figure 3.110. Visualization of activated values of the first 3D convolution layer.168

Figure 3.111. Visualization of the connection strength of the last hidden layer and the output layer.169

Figure 3.112. HSS of AEM index tercile category forecast from CNN model for (a) December and (b) January target months. Circles and whiskers show the mean and range of 75 samples (k=5 folding ensembles for 15 years) across four combinations of input variables.170

Figure 3.113. (a) Lead-lag regression map of SST and U with respect to DMI adopted from Zhang and Han (2021), (b) GradCAM activation maps of AEM tercile forecasts for December (upper) using preceding three months of ST data and January (lower) using preceding four months of U850 data. Both maps show AN category and are composites of 15 years from 2001 to 2015 with their 5 ensembles (k=5 folding).

171

Figure 3.114. Performance of the first SS approach: comparison of EA rainfall ACCs between

original APCC-MME (black horizontal lines) and five sub-sampling models (circles) averaging ensembles with predicted skill in the highest 10%, 30%, 50%, 70%, 90%. Results are for 12 seasons of the test year 2023.172

Figure 3.115. Evaluation of ensemble subsampling using an ACC estimation model with ResNet18 for MAM 2023: (scatter plot) ACC estimation performance, (bar plot) fraction of bad ensembles and dropout rate, (map plot) EA rainfall anomalies of sub-sampled ensemble means and their ACC skill with respect to ERA5 observation.174

Figure 3.116. Evaluation of ensemble subsampling using an ACC estimation model with ResNet18 for ASO 2023: (scatter plot) ACC estimation performance, (bar plot) fraction of bad ensembles and dropout rate, (map plot) EA rainfall anomalies of sub-sampled ensemble means and their ACC skill with respect to ERA5 observation.175

Figure 3.117. Occlusion sensitivity analysis for three variables: geopotential height at 500hPa, meridional wind at 850hPa, and total precipitation. (Scatter plot, left) Division of ensembles for 12 seasons in 2023 into four groups: Good/Miss, Good/Hit, Bad/Miss, Bad/Hit based on the agreement of original and predicted ACC. (Shading contour plots, right) Occlusion sensitivity composite maps for the four groups, using a 5x5 (25deg by 25deg) occlusion window.177

Figure 3.118. (upper) Occlusion sensitivity map for four ensemble categories for 2023 MJJ. Occlusion applied to all input variables with a 5x5 (25deg by 25deg) window. (lower) Anomaly composite difference between ensemble categories: (left) Good/Miss and Bad/Hit, (right) Bad/Miss and Good/Hit for SLP and U850. 178

Figure 3.119. Monthly timeseries of the WNPSH index: ERA5 observation for the past three months (T-3 to T-1) and dynamical model ensemble mean of four categories (Good/Hit, Bad/Miss, Bad/Hit, Good/Miss) for four months from T0 to T3. Black horizontal solid line shows the observed seasonal mean WNPSH index for MJJ 2023.179

Figure 3.120. Performance of the second SS approach: in addition to the ACC-based approach, results for five subsampling models (reddish circles) averaging ensembles with the smallest U850 anomaly errors over EA. U850 anomalies for the next season estimated using an Attention-UNet architecture, with the error defined as

the departure of EA U850 anomalies from the ensemble mean.	180
Figure 3.121. Comparison of two ensemble subsampling approaches for ASO 2023: (a) based on EA rainfall ACC estimation, (b) based on EA lower-level wind field reproducibility.	181
Figure 3.122. Performance of the DB approach: comparison of EA rainfall ACCs between original APCC-MME (black horizontal lines) and 26 variations of deep-learning models (triangles) with different training inputs, architectures, and versions. Closed yellowish triangles denote three variations of the Stacking Two-models (ST) approach. Results are for 12 seasons of the test year 2023.	183
Figure 3.123. Comparative analysis of EA rainfall anomaly maps among ERA5 observations, APCC-MME forecasts, QM, and outputs from four different DB models for 12 seasons of the test year 2023.	185
Figure 3.124. Comparison of DB and TLD approaches: ACCs of 29 DB models and 3 TLD models relative to the original APCC-MME for EA seasonal rainfall. Results are for 12 seasons of the test year 2023.	187
Figure 3.125. Performance of the QM approach: ACCs of EA rainfall anomalies using QM compared to the original APCC-MME for the test year 2023.	188
Figure 3.126. Comparison of SS, DB, TLD, and QM approaches: ACCs of 10 SS models, 29 DB models, 3 TLD models, and 1 QM model relative to the original APCC-MME for EA seasonal rainfall. Results are for 12 seasons of the test year 2023.	189
Figure 3.127. A prototype of AI-enhanced multi-model ensemble forecasts for EA rainfall.	190
Figure 3.128. Performance of AI-aided MME for EA rainfall: ACCs of individual post-processing model outputs (color circles), their simple composite mean (open black triangle), and the original APCC deterministic MME (black horizontal line).	191
Figure B.1. Example of 3-6 week temperature probability forecast results by the top-ranked model (DL-SE-7+SST+NDVI) during the 2018-2022 evaluation period, specifically from July 30, 2018.	207
Figure B.2. Example of 3-6 week temperature probability forecast results by the top-ranked model (DL-SE-14) during the 2018-2022 evaluation period, specifically from March 12, 2018.	207

1. 서론

최근 NVIDIA, 화웨이, 구글 딥마인드, 유럽중기예보센터(European Centre for Medium-Range Weather Forecasts, 이하 ECMWF) 등 미국, 중국, 유럽 등 세계적 기업과 연구소를 중심으로 15일 이내의 인공지능 기후 예측 모델이 개발되고 있다. NVIDIA는 SFNO (Spherical Fourier Neural Operators)를 적용한 트랜스포머(transformer) 기반 FourCastNet, 화웨이는 Swin-Transformer 기반 Pangu-Weather, ECMWF는 GNN기반의 AIFS (Artificial Intelligence/Integrated Forecasting System)을 개발하였다. 또한, 중국 푸단 대학(Fudan University) 연구진은 U-Transformer 기반의 15일 이내 기후예측모델인 FuXi 모델을 개발하였다(Chen et al., 2023)). 또한, 이들은 계절 내 및 계절예측을 위한 FuXi-S2S 모델을 개발하였다(Chen et al., 2024). 구글 딥마인드는 Graph Neural Network (GNN) 기반 GraphCast를 개발한 데 이어, 기존 모델의 예측성을 증가하는 디퓨전 모델(diffusion model)을 기반으로 15일 이내 머신러닝 기반 날씨 예측 모델(ML weather prediction model)을 개발했다. 이 모델들은 ECMWF 웹사이트를 통해 시범 운영 중에 있다(Bi et al., 2023; Bonev et al., 2023; ECMWF, 2023; Lam et al., 2023; Price et al., 2024). 그러나 대부분의 연구가 예측 기간이 15일 이내로 계절 내 및 계절 시간 규모로의 확장은 아직 미흡하다. 또한, 인공지능 연구에는 대량의 학습 자료가 필요하지만, 현재 세계적 기업, 연구소 등이 사용한 기후자료인 ECMWF ERA-5 (the fifth generation of the ECMWF reanalysis dataset)는 1950년부터 현재까지 약 75년의 자료만 구축되어 있다. 이 자료를 학습자료, 검증자료, 테스트자료로 나누어 사용하면 실제 학습에 사용 가능한 자료는 매우 부족하다.

우리는 이러한 한계를 극복하고 딥러닝 기법을 이용하여 계절 내 및 계절 예측 시간 규모에 적합한 기후 예측 모델을 개발하고자 한다. 이를 위해 기후예측 연구에 필요한 기후자료 구축, 기후자료의 모델 민감도 분석 및 입출력 시스템 개발, 전처리 및 초매개변수 기반 다중 인공지능 모델 도출 및 앙상블 기법을 이용한 예측 모델 구축, 준지도 학습을 통한 MJO 지수 확률예측 인공지능모델 개발, 3-6주 동아시아 기온확률 예측 인공지능 모델 개발, 자료강화기법을 이용한 인공지능모델의 월별 기후예측 개선, 인공지능 기술을 활용한 다중모델 앙상블 강수 역학예측 후처리 기법 연구를 수행하고자 한다.

최근 이상기후와 기후위기의 심화로 인해 단기 예보와 장기 예보 사이를 연결하는 계절 내 예측(3~4주)의 중요성이 점차 부각되고 있다. 이는 특히 아시아 태평양 지역에서 식량 안보, 에너지 관리, 재난 대비 등 다양한 기후 응용 분야에서 실질적인 의사결정을 지원할 수 있는 핵심 도구로 자리 잡고 있다. 이러한 배경에서 전 세계적으로 S2S(Subseasonal to Seasonal) 연구가 활발히 진행되고 있다. 그러나 계절 내 예측은 단기와 장기 기상 변동성을 모두 고려해야 하며, 대규모 대기 현상과 대기-해양 상호작용 등 복잡한 요소들을 모델링에 반영해야 하기 때문에 높은 정확도를 달성하는 데 여전히 많은 어려움이 존재한다.

기존의 기후 모델 기반 접근법은 초기 조건과 경계 조건의 불확실성으로 인해 한계가 있으며, 특히 관측 자료 부족이 주요 문제로 지적된다. 이에 따라 최근에는 딥러닝 기술을 활용하여 계절 내 예측의 정확도를 향상시키려는 시도가 이루어지고 있다. 딥러닝 기술은 비선형적이고 복잡한 기후 데이터를 학습하는 데 강점을 가지며, 순환 신경망 기반 기법은 시계열 데

이더 학습에 유용하다. 또한, 공간적 특성을 반영하고 입력 데이터의 중요한 부분에 가중치를 부여하는 Attention U-Net 모델을 통해 학습 성능을 개선하려는 연구도 활발히 진행되고 있다.

딥러닝 기술은 발전과 함께 다양한 고성능 모델들이 개발되고 있다. 예를 들어, ClimaX는 날씨 및 기후 데이터를 세부적으로 분석하고 실시간에서 계절 예측까지 다양한 분야에 적용 가능한 유연한 딥러닝 모델이다. 또한, NVIDIA의 FourCastNet과 화웨이의 Pangu-Weather는 각각 고해상도 글로벌 날씨 예측과 빠른 계산 속도를 자랑하며 기존 수치예보(NWP) 방법보다 높은 정확도를 제공한다. 그러나 이러한 모델들은 여전히 관측 자료 부족 문제와 리드타임이 1~2주로 제한되는 단점을 가지고 있어 개선이 필요하다.

관측 자료 부족 문제를 해결하기 위해 위성 자료를 활용하여 훈련 데이터를 확장하고 이를 딥러닝 모델에 통합하려는 접근법이 주목받고 있다. 위성 자료는 공간적 및 시간적 데이터의 밀도를 높이는 데 유용하며, 이를 통해 민감도 분석 및 다양한 기후 응용 분야에서 신뢰성 있는 예측 자료를 생산할 수 있다. 본 연구에서는 위성 자료를 활용하여 훈련 데이터를 확장하고, 이를 딥러닝 모델에 효과적으로 민감도 분석에 활용할 수 있는 입출력 시스템을 개발함으로써 계절 내 예측의 정확도를 향상시키고자 한다.

따라서 본 연구는 계절 내 예측의 정확도를 저해하는 주요 요인인 관측 자료 부족 문제를 해결하고, 딥러닝 기반 민감도 방법과 입출력 시스템을 통해 보다 신뢰성 있는 기후 예측 자료를 생산할 수 있는 기반 기술을 개발하고자 한다.

최근 신경망 모델의 발전과 이를 다양한 산업 분야에 적용하려는 시도가 증가하면서, 효과적인 모델 설계와 학습 전략에 대한 연구가 활발히 진행되고 있다. 특히, 초기에는 역전파법(Backpropagation, Rumelhart et al., 1986)을 통해 신경망의 가중치와 바이어스를 조정하는 방법이 제안되었으나, 심층 신경망에서는 학습 과정의 어려움과 최적화 이론의 부재로 인해 한계가 드러났다. 이후, 사전 훈련(pre-training, Hinton et al., 2006) 방법 등이 도입되면서 딥러닝(Deep Learning)으로 발전하였고, 현재는 다양한 분야에서 없어서는 안 될 기술로 자리 잡았다. 주가 예측, 의료, 이미지 및 음성 인식, 자연어 처리, 자율주행 등에서 딥러닝이 활발히 사용되고 있으며, GPU 컴퓨팅과 빅데이터의 발전이 신경망 학습과 계산을 한층 더 가속화하고 있다.

기후 분야에서도 딥러닝은 원격탐사 기반 수질 예측, 미세먼지 예측, 에어로졸 예측 등 다양한 연구에 활용되고 있다(Chae et al., 2021, 2020). 최근에는 기후 변화 예측에서도 딥러닝이 활용되어 조기 경고 신호를 정확히 예측하는 연구가 진행 중이며, 구글 딥마인드와 영국 기상청의 협업으로 초단기 날씨 예측 성능이 크게 향상된 사례도 있다(Reed et al., 2022). 또한, 계절적 예측을 위한 연구로 한국 남부 해역의 해수면 온도 예측 모델 개발이 이루어져 단기 예측 성능을 높이는 데 기여하고 있다(Ju et al., 2022). 반면, S2S(Sub-seasonal to Seasonal) 기후 예측의 경우, 데이터 전처리 기술과 딥러닝 모델 개발이 상대적으로 미흡하며, 예측 자료의 표준화 부족으로 인해 분석 및 활용이 어렵다. 대표적으로 ECMWF와 GloSea5는 서로 다른 기후 변수, 시간 해상도, 데이터 제공 방식 등을 가지고 있어 사용자 입장에서 자료를 처리하는 데 어려움이 있다. 이러한 문제는 예측 성능 개선을 위한 후보정 과정에서도 장애가 되며, 기존 연구에서는 U-Net 기반 모델이 사용되었으나 1-2주 S2S 예측 성능 개선이 뚜렷하지 않은 한계를 보였다(Chung et al., 2023).

이를 해결하기 위해 본 연구는 앙상블 기법(보팅, 배깅, 부스팅, 스택킹)을 활용하여 S2S 예측 모델의 성능을 개선하는 데 초점을 맞췄다. 앙상블 기법은 개별 모델의 예측 결과를 결합해 최종 예측 성능을 향상시키는 데 효과적인 접근법이다. 선정된 각 기법의 특징 및 차이점을 분석하고, 알고리즘 설계와 실험에서 이를 비교 분석하는 데 활용하였다. 구체적으로는 학습 및 예측 방식, 훈련 데이터 샘플링 차이 등을 바탕으로 기법의 특성을 비교하였다. 또한, 앙상블을 위한 멤버 모델로 다양한 특성을 가진 모델들로 구성하고자 하였다. 총 5개의 모델을 선정하였는데, 선례 연구에서 각각 구축된 모델들로 CNN, CNN-LSTM, U-Net, Attention U-Net, Residual U-Net 모델을 수정하여 사용했다. 각각의 앙상블 기법은 최소 3~4개의 멤버 모델로 구성되며 이를 통해 앙상블의 다양성과 성능을 극대화하고자 했다.

보팅(Voting)은 서로 다른 신경망 모델(CNN, CNN-LSTM, U-Net, Attention U-Net, ResNet 등)을 독립적으로 학습시키고, 각각의 예측 결과를 가중합산 방식으로 결합하여 최종 결과를 도출하고자 하였다. 이를 통해 다양한 모델 구조의 강점을 반영하고 예측 정확도를 향상시키고자 하였다. 배깅(Bagging)은 S2S 데이터의 랜덤 샘플링을 통해 보팅에서 활용한 선례 연구의 신경망 모델을 병렬로 학습시킨 후, 모델 결과를 평균화하도록 하였다. 이를 통해 데이터 변동성을 줄이고 과적합(overfitting)을 방지하며, 안정적인 예측 성능을 확보하고자 하였다. 부스팅(Boosting)은 약한 학습기(weak learner)를 반복적으로 학습시키고, 이전 모델의 오차를 보완하는 방식으로 강력한 학습기를 구성하도록 했는데 이를 통해 소규모 S2S 자료에서도 높은 예측 성능을 달성할 수 있을 것으로 기대한다. 스택킹(Stacking)은 다양한 멤버 모델의 예측 결과를 메타 모델(meta-model)에 입력하여 최적화된 결과를 도출하고자 하였다. 따라서 본 연구는 앙상블 기법을 통해 모델의 다변성과 학습 성능을 개선함으로써 S2S 예측 기후 변수의 개선 가능성을 확인하고자 수행하였다. 결과적으로, 기존의 단일 모델 기반 접근법 대비 개선된 예측 성능과 안정성을 제공하였다.

준지도학습(SSL; Semi-Supervised Learning)은 라벨링된 자료와 비라벨링된 자료를 결합하여 학습하는 방법으로 (Figure 1.1), 라벨링 비용과 시간을 절감하면서도 높은 성능을 유지할 수 있다. 지도학습(Supervised Learning)은 라벨링된 자료에서 학습하지만, 일부 문제에서는 모든 자료가 라벨링되지 않은 경우가 많아 자료를 라벨링 하는 데 많은 자원이 소모되는 것이 큰 어려움이다. 준지도학습은 이러한 문제를 해결하기 위해 적은 양의 라벨링된 자료와 다량의 비라벨링된 자료를 활용해 모델 성능을 향상시키는 방법이다. 기후 연구에서는 긴 시간 동안 자료가 축적되었음에도 복잡한 기후 특성으로 인해 충분한 학습 자료를 확보하기 어려운 경우가 많다. 특히, 계절내-계절(Subseasonal to Seasonal; S2S) 예측에서 중요한 인자 중 하나인 Madden-Julian Oscillation(MJO)은 약 30~60일의 주기를 가지며, 발생 빈도가 낮아(1974년 이후 진폭 1 이상 사례 약 62.5%) 딥러닝 모델 학습에 필요한 데이터를 충분히 확보하기 어렵다. 이러한 제한된 자료를 효과적으로 활용하는 한 방법으로 준지도학습이 제안되었다.

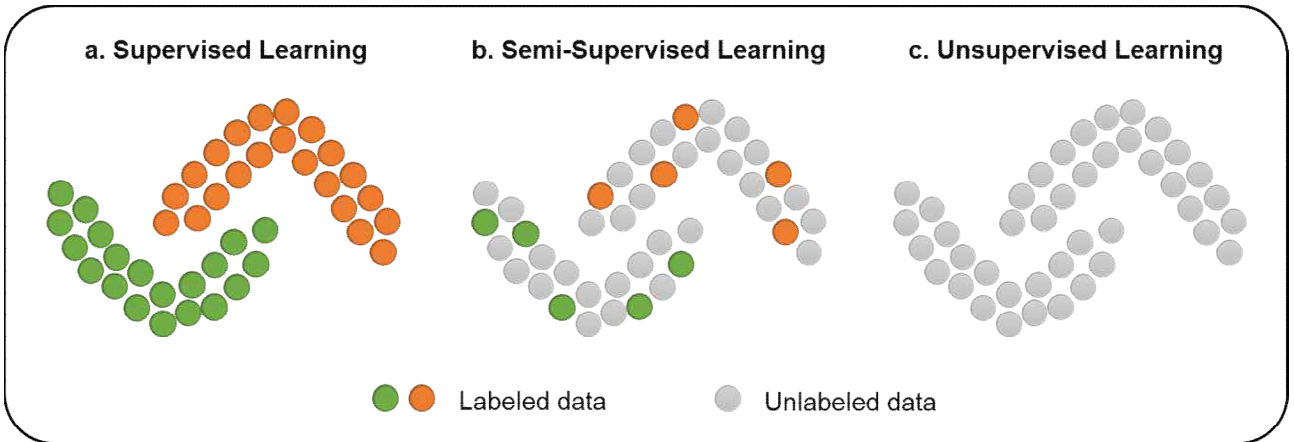


Figure 1.1. (a) Supervised learning, (b) semi-supervised learning, (c) unsupervised learning. Colored circles are labeled data and gray ones are unlabeled data.

준지도학습 기반 알고리즘은 크게 Consistency Regularization, Proxy-label Methods, Generative Models, Graph-based Methods, Entropy Minimization의 다섯 가지 카테고리로 나눌 수 있다. Consistency Regularization은 노이즈나 데이터 증강을 통해 입력 자료를 변형해도 일관된 예측을 유지하도록 한다. Pi-model (Figure 1.2), Mean Teachers, Virtual Adversarial Training 등이 대표적이다(Laine and Aila, 2016). 하지만 기후 데이터와 같이 공간적 특성이 중요한 경우에는 자료증강 기법을 신중히 적용해야 한다. Proxy-label Methods는 초기 라벨링된 자료로 모델을 학습한 후 비라벨링 자료에 pseudo 라벨을 생성하여 신뢰도가 높은 자료를 추가로 학습하는 방식이다. Pseudo-labeling과 Meta Pseudo Labels 방법이 대표적이며 (Figure 1.3), 후자는 확증 편향 문제를 개선해 더 높은 성능을 제공한다 (Pham et al., 2021). 생성형 모델은 자료의 분포를 학습해 새로운 자료를 생성하는 알고리즘이며, 준지도학습 Graph-based Methods는 그래프 구조를 기반으로 노드 간 유사성을 활용해 비라벨링 자료의 라벨을 예측한다. Entropy Minimization은 예측값의 엔트로피를 최소화하여 모델 성능을 높인다.

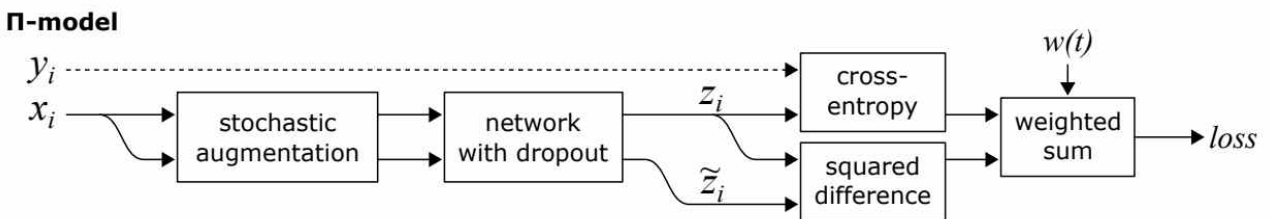


Figure 1.2. Diagram of Pi model (Source: Laine and Aila, 2016).

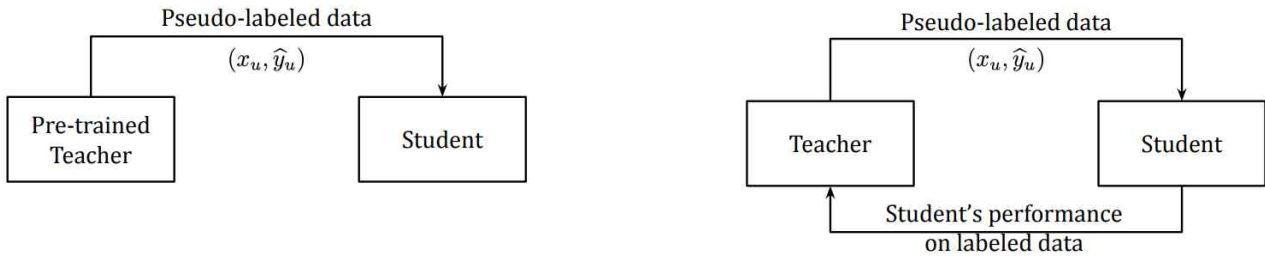


Figure 1.3. Comparison of Pseudo Labels (left) and Meta Pseudo Labels (right) with the feedback from the Student model (Source: Pham et al., 2021).

MJO는 열대지역에서 동쪽으로 이동하는 대규모 대류 시스템으로, 약 30~60일 주기로 발생한다. MJO는 강수, 사이클론, 가뭄 등과 같은 기상 현상뿐만 아니라, 엘니뇨-남방진동(El Niño-Southern Oscillation; ENSO), 준2년 주기 진동(Quasi-Biennial Oscillation; QBO), 북극진동(Arctic Oscillation; AO), 북대서양진동(North Atlantic Oscillation; NAO) 등 대규모 순환 시스템과 상호작용하며 전 세계 기후에 영향을 미친다. MJO의 정확한 예측은 열대 및 아열대 지역의 계절내-계절 예측(S2S) 개선에 중요하다. MJO의 실시간 모니터링에는 상향 장파 복사(OLR; Outgoing Longwave Radiation)와 200hPa 및 850hPa의 수평 바람장에 대해 EOF를 적용하여 첫 번째와 두 번째 주요 주성분(PC1, PC2; Principal Component) 각각을 RMM1과 RMM2로 정의한 RMM(Real-time Multivariate MJO index) 지수가 널리 사용된다. 현재의 기후모델은 예측 기간이 길어질수록 MJO 예측 정확도가 현저히 떨어진다. 딥러닝은 시공간 패턴을 효과적으로 학습할 수 있어 MJO 예측에서 성공적으로 활용되고 있다. 기존의 딥러닝 모델을 활용한 RMM 연구는 RMM 지수를 직접 예측하거나(Shin et al., 2024; Delaunay and Christensen, 2022; Martin et al., 2022; Dasgupta et al., 2020) 기존 기후모델의 예측값을 후처리하는 방식(Kim et al., 2021)으로 수행되었다. 최근 연구에서는 CESM2(Community Earth System Model)와 ERA5 자료(fifth generation ECMWF reanalysis data)와 전이학습(Transfer Learning)을 통한 CNN 기반 모델이 예측 기간을 ~25일(0.5 BCOR)로 연장하며 기존 모델보다 우수한 성능을 보였다(Shin et al., 2024). 최근에는 FuXi-S2S 라고 불리는 트랜스포머(Transformer) 기반 딥러닝 모델이 소개되었다(Chen et al., 2024). 이는 flow-dependent 섭동 방식을 적용하여 좀 더 정확한 자료 분포와 불확실성을 반영함으로써 ECMWF를 훨씬 능가하는 성능을 보여주었다. 하지만 FuXi-S2S와 같은 모델은 해석이 어렵고 계산 비용이 높은 단점이 있다. XAI 기법을 통해 목적변수 예측의 근원을 이해하고자 하는 여러 시도가 있어 왔다. Shin et al. (2024)는 XAI 기법을 개발하여 강수량 및 지표 온도와 같은 장기 MJO 예측의 주요 변수를 확인하였다. 한편 딥러닝 기반 MJO 예측 연구는 대개 필터링을 거친 편차(Anomaly) 정보만을 활용하였다. 편차 정보는 MJO 시그널을 정확하게 탐지하는데 가장 중요한 정보이지만 배경장은 MJO 장기 예측에 중요한 정보로 알려져 있다. 편차 정보에만 집중하고 배경장 상태를 직접적으로 딥러닝 모델에 활용하지 못하여, 공간적인 습도 차 및 저층 동서풍과 같은 MJO 동역학에 중요한 역할을 하는 요인의 배경적인 역할을 고려하지 않는다.

2차년도(2023년)에는 준지도학습 기법을 활용해 MJO 위상(phase)을 분류하는 딥러닝 모델

을 개발했다. 기존 MJO 위상 계산은 기후 자료를 필터링한 편차 자료에 Empirical Orthogonal Function(EOF) 분석을 적용해 1, 2번 주성분(Principal Components)으로 RMM, VPM, OMI 등의 인덱스를 얻는 방식이지만, 인덱스 종류에 따라 MJO 패턴 반영이 다르다는 한계가 있다. 따라서 다양한 인덱스에 대해 모델을 훈련하고, 준지도학습 적용 전후의 성능을 비교했다. 3차년도(2024년) 연구는 2차년도에 개발된 MJO 위상 분류에서 선행시간별 MJO RMM 지수를 예측하는 모델로 확장하였다. 나아가 단정예측 모델을 기반으로 MJO RMM 확률예측 모델을 개발하였다. 다양한 인덱스에 대해 모델을 훈련하고, 준지도학습 적용 전후의 성능을 비교했다. 3차년도(2024년) 연구는 2차년도에 개발된 MJO 위상 분류에서 선행시간별 MJO RMM 지수를 예측하는 모델로 확장하였다. 나아가 단정예측 모델을 기반으로 MJO RMM 확률예측 모델을 개발하였다. 기후 자료를 편차와 배경장을 분리해 CNN 기반 ResNet 모델에 활용함으로써 MJO 예측 성능 향상 실험을 수행하였다. 다양한 모델 아키텍처와 입력 변수를 최적화하기 위한 실험을 수행했으며, 선행시간별로 RMM1과 RMM2에 대한 입력 변수의 기여도를 분석하기 위해 설명가능한 인공지능(eXplainable Artificial Intelligence; XAI)의 Occlusion 민감도 분석을 적용했다. 또한, 선행시간이 길어질수록 MJO 진폭이 감소하는 문제를 해결하기 위해 RMM 값과 진폭을 고려하는 맞춤형 손실함수를 도입해 예측 정확도와 진폭 예측 정확도를 개선했다.

최근 인공지능 분야에서는 2018년부터 시행 중인 유럽연합(EU; European Union)의 개인정보보호(GDPR; General Data Protection Regulation) 규정에 따라 인공지능 알고리즘에 의해 행해진 결정에 대해 질문하고 결정에 관여한 논리에 대해 의미 있는 설명을 요구할 권리를 포함하는 설명을 요구할 권리(Right to explanation)가 주목받고 있다(NIA, 2018). 그리고 2018년 12월에 첫 번째 초안이 공개된 유럽연합의 AI 윤리 가이드라인(Ethics Guidelines for Trustworthy AI)에는 AI 시스템이 신뢰할 수 있는 7가지 주요 요건으로 인간의 권한과 감독(Human agency and oversight), 기술적 견고성 및 안전성(Technical Robustness and safety), 개인정보 보호 및 데이터 거버넌스(Privacy and data governance), 투명성(transparency), 다양성, 비차별, 공정성(Diversity, non-discrimination and fairness), 사회 및 환경적 복지(Social and environmental well-being), 책무(Accountability)가 제시되어 있다. 이 중에서 투명성은 데이터, 시스템, AI 비즈니스 모델의 투명성과 이해관계자에게 적합한 방식으로 결정과정이 설명되어야 한다고 명시되어 있다(EU, 2019). 이러한 이유로 인공지능 분야에서는 블랙박스인 인공지능 모델의 설명가능한 인공지능(XAI, eXplainable Artificial Intelligence) 기법을 적용하고자 하는 노력이 계속되고 있다. 인공지능 모델 중 이미지 분류(image classification), 객체 탐지(object detection) 등에서 활발하게 적용되고 있다.

인공지능 모델은 RNN(Recurrent Neural Network), CNN(Convolutional Neural Network) 등을 토대로 인코더-디코더, 인코더-디코더를 확장한 U-NET(Ronneberger et al., 2015), ResNet(He et al., 2016), VGG(Simonyan and Zisserman, 2014) 등 다양한 아키텍처가 등장하였다. 또한, 학습과정에서 유용한 피쳐에 더 집중할 수 있는 SE 네트워크(squeeze-and-excitation network), CBAM (Convolutional Block Attention Module) 등 어텐션 메커니즘(attention mechanism)도 등장하였다(Hu et al., 2020; Woo et al., 2018).

이러한 딥러닝 아키텍처와 모델을 기반으로 계절 내 및 계절(S2S, Subseasonal to Seasonal) 시간 규모에 적합한 기후 예측 모델을 설계하는 것만큼 모델의 기후 변수 선택에 따라 모델을 구축하는 데 필요한 비용이 다르며 예측성에도 영향을 미친다.

ALO (atmosphere-land surface-ocean) 시스템 관점에서 S2S 시간 규모와 예측성 간 관계를 보면, 짧은 선행시간에서는 대기의 초기 상태를 아는 것이 중요하며, 2주에서 4주의 선행시간에서는 지면의 정보, 30일을 넘어가는 예측에서는 해양에 대한 정보가 중요한 것으로 알려져 있다(Mariotti et al., 2018).

따라서, 딥러닝 아키텍처와 모델뿐만 아니라 ALO 시스템 관점에서 1개월 기온 예측 모델을 구축하고, 동아시아지역 3-6주 기온확률 예측 체계를 구축하고자 한다. 이를 위해 기존 1개월 기온 예측 모델 개발, 입력 기후 변수에 따른 기온 예측 모델 구축, 최신 딥러닝 아키텍처를 이용한 기온 예측 모델 구축, 앙상블 학습 기법을 이용한 기온 예측 모델, N 스텝 예측 방법을 이용한 기온 예측 모델 구축, 재학습 방법을 이용한 기온 예측 모델 구축 등을 수행하고자 한다. 또한, 기존 1개월 기온 예측 모델을 개발과 함께 인공지능 모델의 내부 구조 분석을 통해 모델의 학습 과정에 영향을 미치는 요소들을 찾고 요소들의 조합을 통해 모델의 개선을 진행하고자 한다. 기온 예측 모델을 기반으로 기온확률 예측 체계를 구축 및 실험하고 HSS (Heidke Skill Score)을 이용하여 예측성을 평가하였다.

단순한 인공지능모델에 간결한 자료강화기법을 적용하여도 월별 기후예측을 개선할 수 있을지 가능성을 타진하고자 하였다. 여름철 평균기온 예측을 위한 합성곱신경망모델(CNN)을 구축하여 두 가지 자료강화의 개선 성능을 비교하였으며 모든 월을 사용하는 접근과 목표 월만을 사용하는 접근의 결과 또한 비교하였다. 그리고 모든 월 자료를 훈련에 활용하는 CNN 모델에 대해 여름철과 겨울철, 평균기온과 강수(SPI1) 예측에 대한 자료강화의 개선 성능을 평가하였다.

아태기후센터는 한반도를 포함한 동아시아 지역에 대해 맞춤형 다중모델앙상블 (Multi-Model Ensemble, MME) 예측 서비스를 수행 중에 있다 (www.apcc21.org). 앙상블을 결정론적/확률론적으로 취합하는데 있어서 후보정 유무는 매우 중요하다. 다중모델앙상블 현업 시스템을 처음 구축하고 유지 보수하는 과정에서 후보정 유무의 성능 기여도에 대한 테스트가 충분히 이루어졌고, 전반적이고 유의미한 개선에 대한 힌트를 얻지 못해 APCC 현업에서는 후보정 없이 다중모델앙상블 예측을 생산하고 있다. 하지만 시스템 구축 당시보다 통계에 활용할 수 있는 과거 샘플이 더 늘어나 변화의 가능성이 있고, 특히나 기존 통계가 아닌 딥러닝 기술에 대한 적용성 평가는 이루어지지 않았기에 역학 예측에 대한 후처리 연구 시도는 의미가 있다. 이 연구는 다중모델앙상블 예측 성능을 개선하기 위한 여러 후처리 방식을 시도하고, 이를 위해 기존 통계 기법이 아닌 딥러닝 기법을 적용하는 것으로 계획되었다.

여러 관측 원천간 다소 차이는 있지만, MME 기법을 적용하였음에도 불구하고 동아시아 계절 강수에 대한 역학 모형 예측 성능은 높지 않다. 특히, 집중 호우 등 재난 상황이 빈번한 여름철 성능이 현저히 낮기 때문에 예측 자료 후처리를 통한 강수 예측 개선은 반드시 필요하다. 인공지능 모델은 학습을 위한 다량의 자료를 필요로 하는데, 안타깝게도 계절 시간 규모 예측에 활용될 수 있는 과거 관측 샘플은 극히 제한적이어서 이에 대한 대안이 필요하다. 센터

가 MME 예측 서비스를 본격적으로 시작한 10여년 전부터 축적된 실시간 예측 앙상블 수가 3만개 이상이며 이를 인공지능 모델 훈련을 위한 자료로 활용할 수 있다면 인공지능 기반 후처리 기술 개발 및 강수 예측 성능 개선을 기대할 수 있겠다. 앙상블 자료를 활용한 계절 강수 후처리의 방안으로 양질의 앙상블 선별 방식을 최우선적으로 시도해 볼 수 있을 것이다. 과거 앙상블 자료를 이용해 사후 분석한 결과, 다수의 앙상블 중 ACC 성능이 낮은 일부를 탈락시키고 ACC 성능이 높은 앙상블들을 취할 때 전체적인 평균 예측 성능을 높아지는 것을 알 수 있다. 다만 미리 특정 앙상블의 성능을 추정하는 것은 쉽지 않기 때문에, 이를 구현하는 방법이 딱히 개발되지 못했다. 하지만 역학 모델 예측 성능이 ENSO 시그널이 있을 때 높다는 것은 널리 알려져 있고, 유사한 형태로 동아시아 3-4주 기온 예측 성능이 기후 모드 강도/위상 조합이나 앙상블의 시간 안정성 등에 의존적일 수 있다는 연구도 제시된 바 있다(Kim et al., 2022; 아태기후센터 연구보고서). 만약 인공지능 기술이 선행 기후 모드와 역학 모델 오차간의 관련성 등 성능 추정을 위한 복잡한 정보들을 자료로부터 뽑아낼 수 있다면 우수한 앙상블 선별하는 방식으로 성능 개선이 가능할 것으로 판단하였고 이것이 인공지능 기반 동아시아 계절 강수 후처리 기술 개발의 배경이자 시작점이 되겠다.

2. 방법 및 자료

2.1. 자료 확장에 따른 모델 성능의 민감도 분석 및 입출력 시스템 설계

2.1.1. 자료

2.1.1.1. 관측자료

관측자료는 위성자료를 기반으로 식생지수, 해수면온도, 강수량 자료를 수집하고 이용하였다. 식생지수(NDVI; Normalized Difference Vegetation Index)는 미국 NOAA(National Oceanic and Atmospheric Administration)에서 자료를 제공하고 있으며, NIR(근적외 반사값)과 Red 채널의 반사값을 노말라이즈해서 -1부터 1사이의 값을 가지며 0보다 크면 식생의 상태가 좋은 것을 나타내고 0보다 작으면 물이나 눈과 같은 식물이 아닌 표면을 나타낸다(Figure 2.1). 자료는 1981년부터 서비스 하고 있으며 NOAA 극궤도 위성 AVHRR(Advanced very High Resolution Radiometer), VIIRS(the Visible Infrared Imaging Radiometer Suite)센서 기반으로 만들어진 자료이며 NOAA-7,-9,-11,-14,-16,-17,-18,-19)에서 얻은 데이터를 사용하여 공간해상도는 0.05°의 해상도로 일자료를 제공하고 있기 때문에 얻어진 자료를 1.5° 일자료로 변환하였다. 데이터셋은 GSFC(NASA Goddard Space Flight Center)와 UMD(University of Maryland)의 지표면 기후데이터셋의 하나로 지표면 반사값, 관측시간, 위,경도 자료등을 netCDF-4 포맷자료로 제공하고 있기 때문에, 이러한 내용을 바탕으로 전처리하였다.

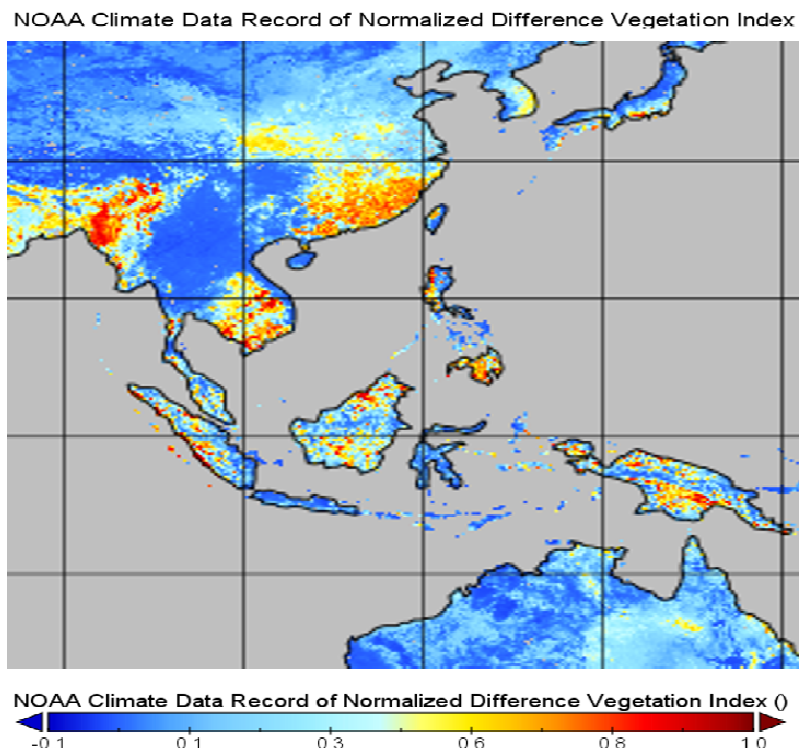


Figure 2.1. NOAA NDVI(Normalized Difference Vegetation Index) data form AVHRR satellite.

GPM IMERG(Daily Global Precipitation Measurement Integrated Multi-satellitE Retrievals for GPM)자료는 일별로 측정된 전 세계 강수량 데이터를 제공하는 product로 데이터는 GPM 코어 위성과 협력 위성에서 수집한 자료를 결합하여 고해상도 강수량 분포와 지구의 강수 패턴 및 변동성 정보를 제공한다. GPM(Global Precipitation Measurement) 위성은 강수의 양과 강수 패턴을 측정하기 위해 Ku, Ka 레이더센서와 마이크로파 이미저 센서를 사용하여 다양한 데이터를 생산한다. 이 연구에서 사용된 IMERG Final 데이터는 2000년 6월부터 자료를 수집하여 ECMWF와 같은 공간해상도인 1.5° 로 전처리하여 입력자료로 이용하였다(Figure 2.2).

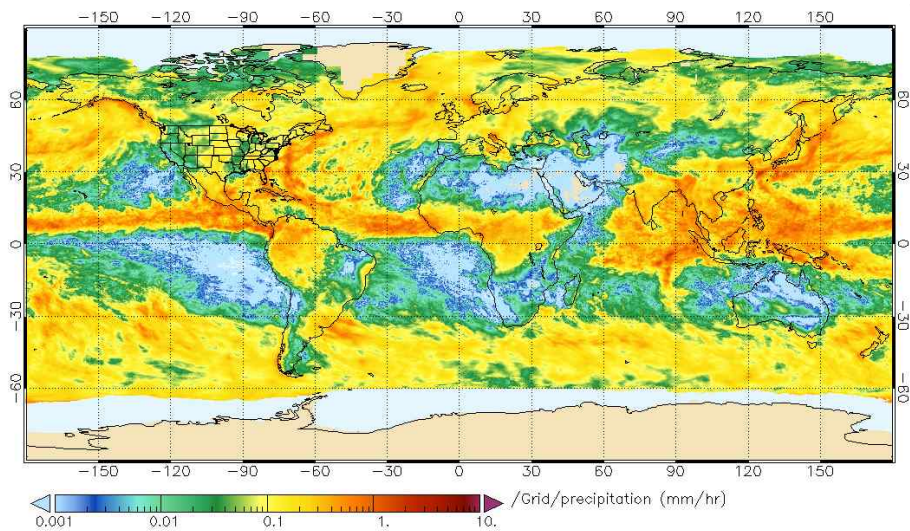


Figure 2.2. GPM IMERG product.

GPM IMERG자료이외에 관측 강수량자료 NOAA CDR(Climatic Data Records)프로그램으로 제공하고 있는 CMORPH 자료를 수집하였다. NOAA의 CMORPH의 자료는 Monthly자료는 제공하고 있지 않기 때문에 daily 자료를 전처리하고 Monthly자료를 변환하여 최종적으로 daily, monthly 으로 데이터베이스 하였다(Figure 3). CMORPH 자료는 TMI, AMSR, SSMIS, MHS, ASU, MWRI 위성 센서자료와 정지궤도 적외채널 자료를 결합하고, GPCP자료로 보정하여 산출한 자료로서 1998년 6월부터 2023년 4월까지 제공하고 있으며, 25km의 공간해상도로 NetCDF자료로 제공한다. 다른자료와 달리 위도는 -60N ~ 60N까지 자료를 제공하는데 ERA5자료에 맞게 위, 경도를 변경하여 전처리하고 입출력 시스템 고도화를 위한 연계하였다. GPM IMERG, CMORPH 데이터 이외에 PERSIANN데이터를 수집하고 전처리하였으며, PERSIANN-CDR 데이터는 0.25도의 고해상도 공간해상도를 가진 자료로서 1983년부터 현재까지 데이터 제공중에 있다.

NOAA Climate Data Record of PERSIANN-CDR daily precipitation

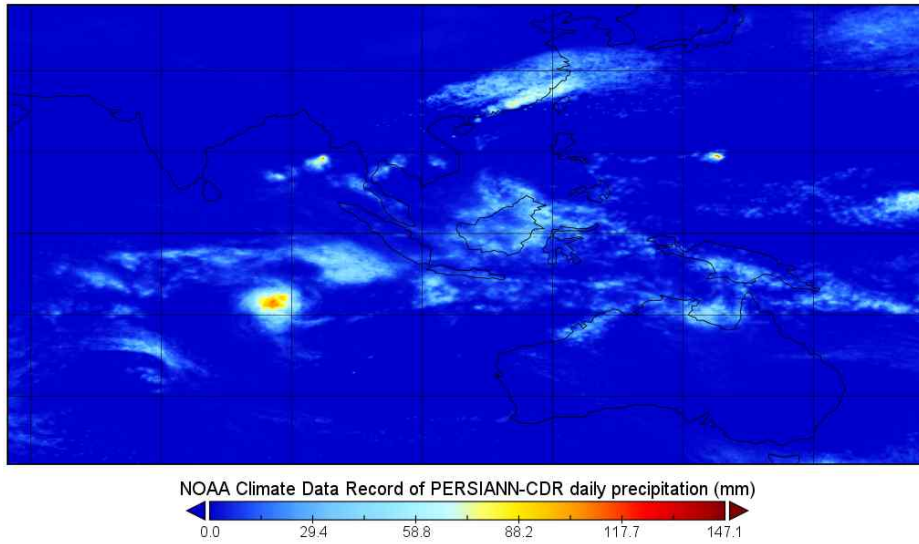


Figure 2.3. NOAA CMORPH product of daily precipitation.

PERSIANN-CDR 데이터는 daily자료이기 때문에 Monthly자료로 활용하기 위해서 upscaling 데이터 전처리하여 자료를 수집하였다. 자료 변환 알고리즘은 Gridsat-B1(3-hourly IR)를 기반으로 GPCP(monly 강수)자료를 보정하여 자료를 제공하는 자료를 수집하였다(Figure 2.4).

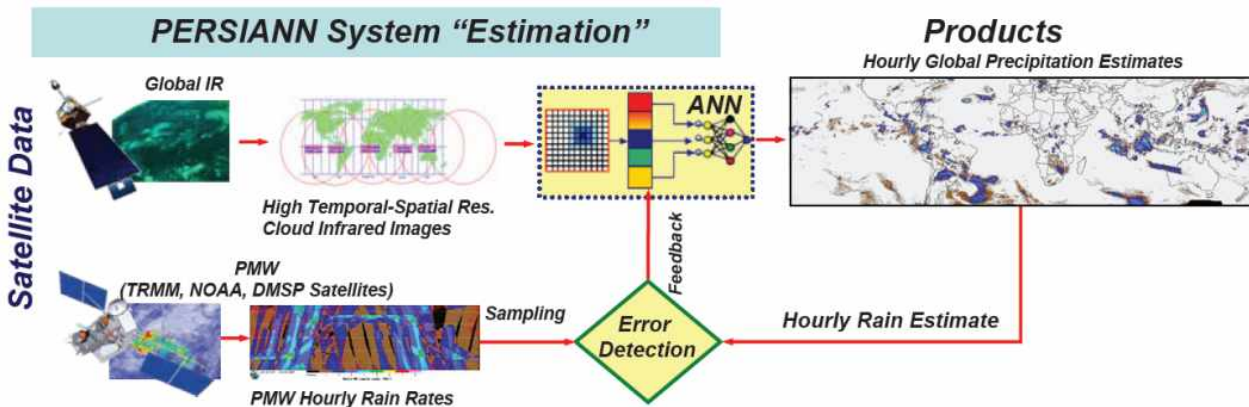


Figure 2.4. The PERSIANN rainfall retrieval algorithm of NOAA CDR program.

자료는 위도 -60N ~ 60N, 경도 -180 ~ 180의 범위로 자료로 NetCDF포맷으로 자료를 수집하여 전처리하였다.

해수면 온도는 NOAA OISST(Optimum Interpolation Sea Surface Temperature)자료를 수집하여 전처리하였다. 해수면온도 자료는 위성에서 산출한 해면 온도와 해양부이 및 선박관측자료를 최적보간 기법으로 사용하여 전지구 일일 해면 온도를 산출하고 있다. 미국 국립기후데이

터센터(NCDC)에서 1981부터 일자료를 산출하고 있으며 최근 Version 2의 데이터를 수집하였다. OISST자료는 “AVHRR-only”와 “AVHRR+AMSR”의 두 가지 버전이 있으며, 공간해상도는 $0.25^{\circ} \times 0.25^{\circ}$ 의 해상도로 자료포맷은 NetCDF4 자료로 제공하고 있으며, AVHRR-only자료만 30년간의 장기간 기후자료를 제공하고 있다. (Reynolds et al;2002). 사용할 수 있는 변수는 Sea Ice Concentration, Sea Surface Temperature, Sea Surface Temperature Anomalies들이고, 여기서 Sea Surface Temperature 자료를 1998년부터 2017년자료를 시공간 보간하여 사용하였다(Figure 2.5).

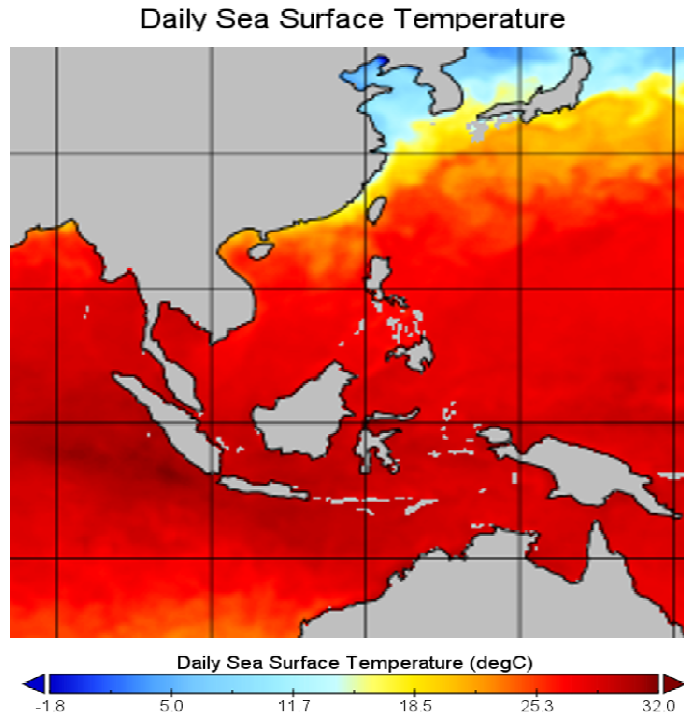


Figure 2.5. Sea surface temperature data from NOAA.

2.1.1.2 모델자료

계절내 기후자료는 ECMWF 자료는 계절내에서 계절까지의 시간대에 예보 정확도와 이해를 높이기 위하여 자료를 생산하고 있으며, WMO(World Meteorological Organization)의 계절내 예측 프로젝트 기반하여 다양한 기후모델 자료(NCEP, BoM, JMA, ECMWF)를 제공하고 있다. 이 연구에서는 정확도 높은 ECMWF 자료는 0-46일의 자료 중에서 리드타임 3주 자료를 1998년부터 2017년까지 모든 변수 자료를 수집하고 전처리하였다. 변수는 MSLP, OLR, PREC, Q700, Q850, SST, T2M, TMAX, TMIN, U50, U850, V200, W850, Z200, Z500의 자료를 민감도 분석 학습자료로 맞춰 전처리하였다.

2.1.1.3 재분석자료

ERA5는 ECMWF(European Center for Medium-Range Weather Forecasts)에서 제공하는 대기 재분석 데이터셋의 5번 세대의 자료로 과거 관측 데이터를 일관되게 전처리하여 기후자료로 종합적이고 일관된 자료를 제공하고 있다(Figure 6). ERA-5는 이전 버전인 ERA-Interim보다 고해상도 및 시간 해상도를 높여 제공하고 있으며, 물리 모델과 데이터 동화기술을 개선하여 제공하고 있다. ERA5 재분석자료는 1979년부터 현재까지 전지구적으로 자료를 제공하고 있으며, 연직프로파일은 137개로 지표면에서 80km까지의 포함하고 있으며, 온도, 습도, 풍속, 최고/최저 온도 등 다양한 변수를 제공한다. 1998년부터 2017년까지 시간자료를 일자료로 전처리하고 공간해상도도 1.5° 로 변환하여 사용하였다(Table 2.1).

Maximum temperature at 2 metres

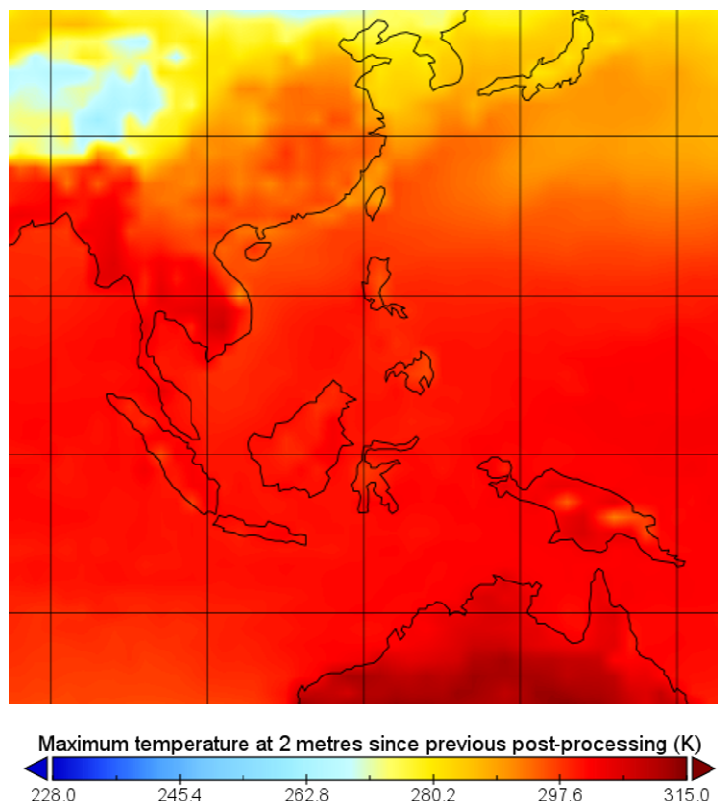


Figure 2.6. ERA5 Maximum temperature at 2 meters.

ERA5는 관측자료인 NDVI, SST, ECMWF 모델 자료와 같은 기간 시/공간 해상도를 맞추어 타겟기간인 3주차에 맞춰 입력자료를 구성하였다. 타겟 변수는 강수량을 먼저 입력자료로 구성하고, 최대 온도값을 민감도 분석 입력자료로 전처리하였다(Table 2.2).

Table 2.1. Data description of ERA5.

Data Description	
Data type	Gridded
Projection	Regular latitude–longitude grid
Horizontal Coverage	Global
Horizontal resolution	Reanalysis: 0.25° x 0.25° (atmosphere), 0.5° x 0.5° (ocean waves) Mean, spread and members: 0.5° x 0.5° (atmosphere), 1° x 1° (ocean waves)
Temporal coverage	1959 to present
Temporal resolution	Hourly
File format	GRIB
Update frequency	Daily

Table 2.2. Overview of the characteristics of ERA5.

ERA5	
Period covered	1950–present
Production Period	Jan 2016–end 2017, then continued in near real-time
Model version	IFS Cycle 41r2
Assimilation system	IFS Cycle 41r2 4D-Var
Spatial resolution	31km globally, 62km for the Ensemble of Data Assimilation(EDA), 137 levels to 0.01hpa
Output frequency (temporal resolution)	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)
Uncertainty estimates	From a 10-member Ensemble of Data Assimilations (EDA) at 63 km resolution
Model input	Appropriate for climate (e.g. CMIP5 greenhouse gases, volcanic eruptions, SST and sea-ice cover)
Input observations	In addition, various newly reprocessed datasets and recent instruments that could not be ingested in ERA-Interim
Variational bias scheme	Also ozone, aircraft and surface pressure data
Satellite data	RTTOV-11, all-sky for various components
New parameters	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.
Parameters removed	A few parameters present in ERA-Interim are not available in ERA5. For availability of specific parameters please see the technical documentation linked below.
Handling of accumulated parameters	Accumulated from previous post-processing

2.1.2. 방법

2.1.2.1 민감도 분석 및 입출력 시스템 개요

계절 내 예측자료의 정확도 개선을 위해서는 모델 자료 이외에 관측자료의 확장하고 확장된 자료의 변수를 다양하게 선택하여 학습하는 민감도 분석을 통하여 확장자료의 인공지능 모델 입력자료로서의 유용성과 정확도 개선을 제고하였다(Figure 2.7). 많은 학습자료와 변수들 중에서 계절 내 예측성에 필요한 자료를 알아내기 위해서는 많은 학습 케이스가 필요하다. 따라서 민감도 분석을 위한 입출력 시스템을 개발함으로써 다양한 학습을 통한 딥러닝 모델을 산출하고자 한다(Figure 2.8).

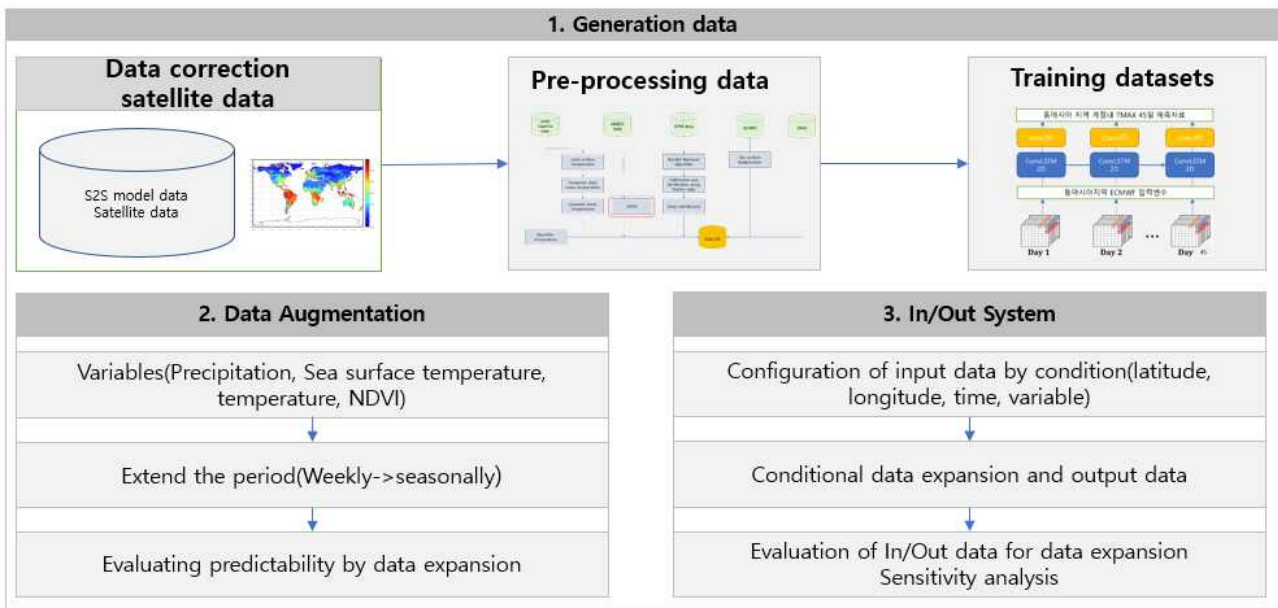


Figure 2.7. Overview of sensitivity analysis and input/output systems.

2.1.2.2 딥러닝 모델

자료 입력에 따른 민감도 분석을 하기 위하여 딥러닝 모델은 U-Net의 개선 버전인 Attention U-Net 모델(Ozan Oktay et al, 2018)을 이용하였다. U-Net 모델은 이미지 분석 및 예측에 있어서 신뢰할수 있는 모델이며, 합성곱 신경망(CNN)의 대표적인 모델이다. U-Net 모델의 아키텍처는 인코더와 디코더 구조를 가지며, 인코더는 계층적 특징을 추출하고 디코더는 이를 사용하여 원래 이미지의 고해상도 세그멘테이션을 추출하기 위해서 사용된다. U-Net 모델의 특징 중의 하나는 인코더에서 추출한 피쳐의 특성을 디코더의 동일한 해상도의 층과 결합하여 해상도 높은 이미지에서 산출가능한 정보와 저해상도에서 산출가능한 정보를 모두 사용하여 보다 정확한 세그멘테이션 맵을 산출가능하게 해준다. 계절내 예측자료의 시계열 입력자료의 정확도 향상에 효과적으로 이용하기 위하여 U-net 모델이 이용되어왔다(Figure 2.9). Attention

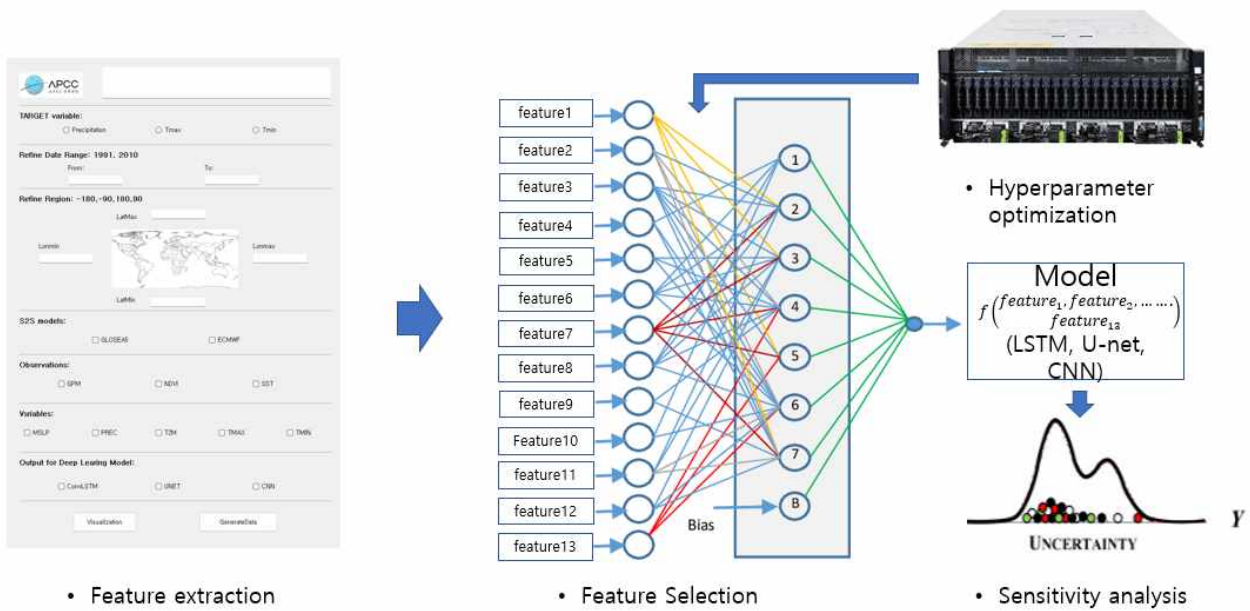


Figure 2.8. Sensitivity analysis of feature selection from in and out data system.

U-Net모델은 U-Net모델에서 어텐션 매커니즘을 추가하여 보다 정확도를 향상시킨 모델이다. 어텐션 매커니즘은 입력자료에서 중요한 부분에 가중치를 더 부여하여 모델이 좀더 중요한 부분에서 깊게 학습하도록 해주기 때문에 단독 U-Net 모델을 이용한 것보다 보다 정확도를 향상시킬 수 있기 때문에 본 연구에서 민감도 분석을 위하여 Attention U-Net 모델을 사용하였다 (Figure 2.10).

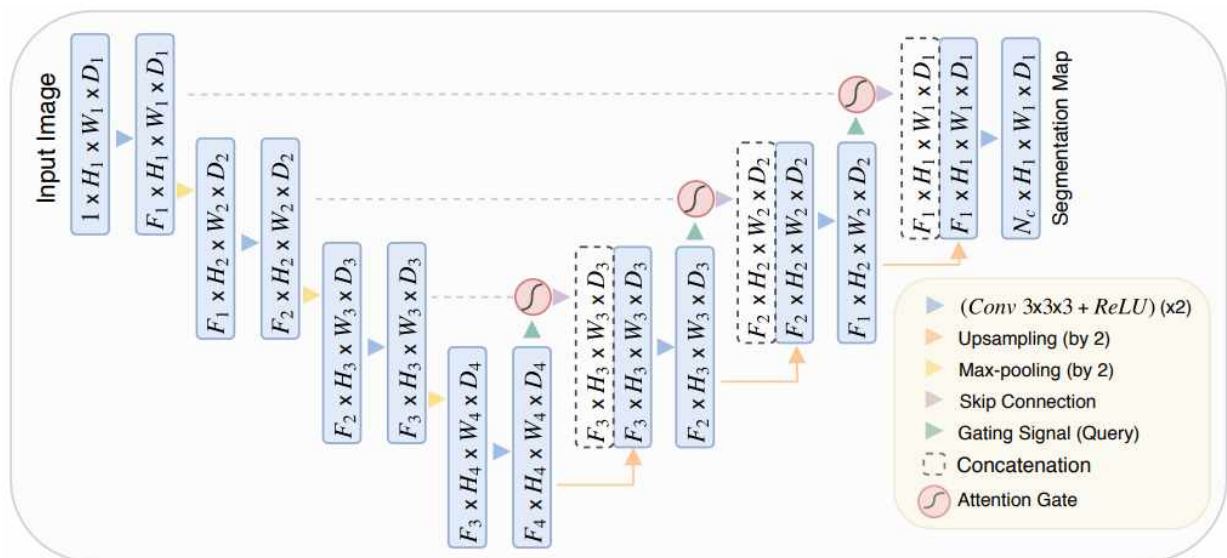


Figure 2.9. A block diagram of Attention U-Net architecture(Source: Ozan Oktay et al., 2018).

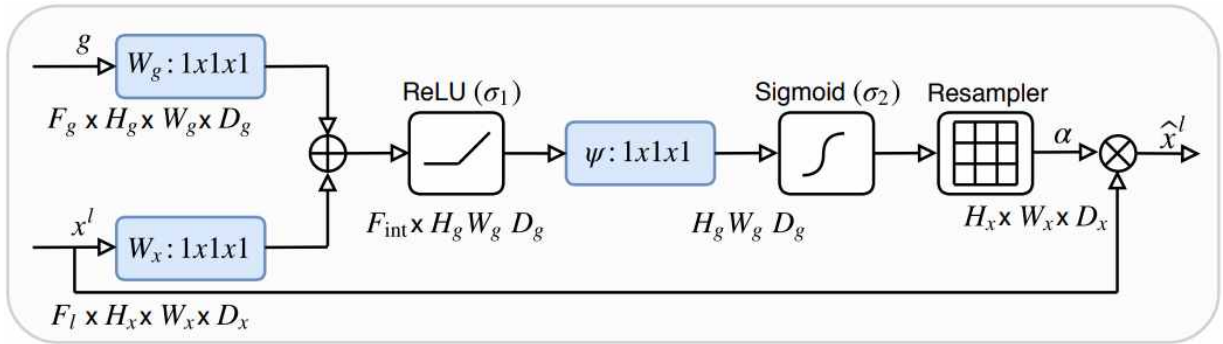


Figure 2.10 Schematic of the attention gate(Source: Ozan Oktay et al., 2018)

인코딩 과정에서는 Attention U-Net 민감도 입력자료인 모델자료와 관측자료에서 행렬의 주변 픽셀을 고려하여 범위를 넓혀가며 feature의 특징을 추출하고 3×3 합성곱을 수행할 때마다 채널정보를 증가시키면서 레이어를 추가시켰다. 디코딩 과정에서는 인코딩에서 산출된 feature의 특징중에서 중요한 자료에 가중치를 더하고 upsampling을 하면서 최종적으로 온도 3주치의 7일자료를 산출하였다. 하이퍼파라미터로 optimizer는 adam 함수를 Loss는 MSE 함수를 이용하였고, Epoch는 300, 배치 사이즈는 8로 고정하여 입력자료만 변화를 주고 학습하였다. 입력자료의 구성은 ECMWF 모델자료와 관측자료를 Attention U-Net의 학습자료로 이용하기 위하여 Figure 2.11과 같이 입력자료를 구성하였다. 모델자료 예측일인 2월 1일 기점으로 관측 자료는 2월 1일보다 일주일 전 자료를 이용하였고, 예측자료인 ECMWF 모델자료는 리드타임 3주치의 7일 자료를 이용하여 학습자료를 구성하였고, 타겟변수인 ERA5의 자료는 3주차 최고 온도 7일 자료를 타겟자료로 활용하였다.

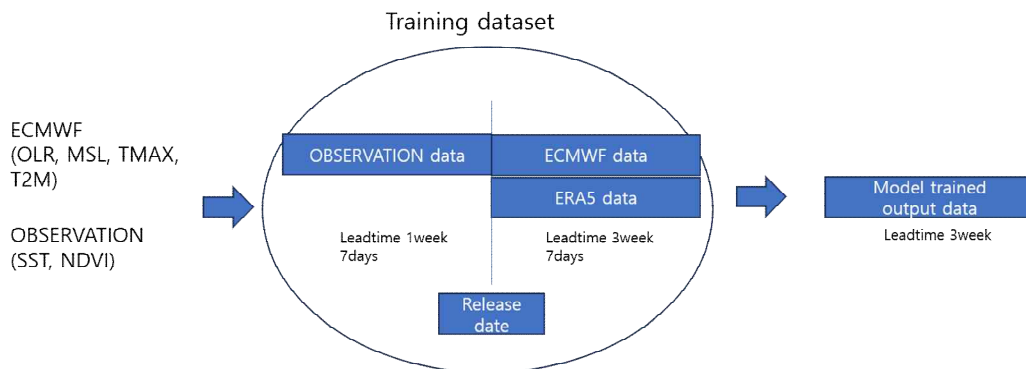


Figure 2.11. A diagram of input data from observation and ECMWF model data for Attention U-Net model.

민감도 분석을 위한 설계는 Attention U-Net 모델을 최적화된 파라미터를 기반으로 고정하고 학습자료는 예측자료인 ECMWF와 관측자료인 NDVI, SST 자료를 기준일 기준으로 관측자료는 과거 일주일 자료로 예측자료인 ECMWF 자료는 리드타임 3주차 일주일 자료로 타겟변수인

강수량, 최고온도에 대해서 ERA5자료는 리드타임 3주차 일주일 자료로 구성하였다. 민감도 분석을 위하여 입력자료는 Filter, Wrapper, Embedded 기법에 맞춰 입력자료에 따른 정확도 개선을 보고자 하였다(Figure 2.12).

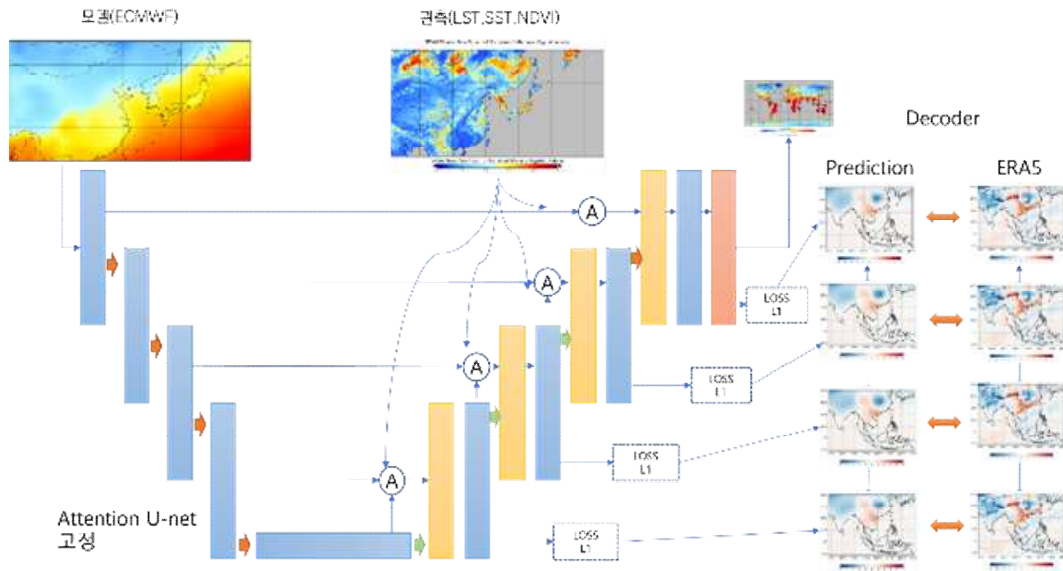


Figure 2.12. Setup of attention U-net model.

2.1.2.3 민감도 분석 방법

민감도 분석 방법은 Filter, wrapper, embedded 방법이 있으며, 입출력 시스템의 Feature Selection(특성 선택)을 수행하는 방법이며, 그중에서 filter 방법은 데이터의 특성 간의 통계적 속성을 기반으로 중요도를 평가하여 불필요한 특성을 제거하거나 중요한 특성을 선택하는 방식이다. 이 방법은 특정 딥러닝 모델을 사용하지 않고 데이터 자체의 통계적 속성을 기반으로 특성을 선택하기 때문에 자원활용에 좋으며, 빅데이터에서도 효율적으로 사용할 수 있다. 반면에 단순한 통계적 기준을 사용하기 때문에 다른 방법(Wraper, Embedded)에 비해 예측 성능이 낮을 수 있다. filter Method의 주요 기법은 각 특성의 분산을 계산하여, 분산이 낮은 특성을 제거하는 분산 기법과 특성 간 상관 계수를 계산하여 높은 상관 관계를 가진 특성 중 하나를 제거하는 상관관계 기법등의 방법이 있다. wrapper 방법은 예측 정확도를 극대화할 수 있는 최적의 특징 조합을 찾아내는 방식으로 모델 학습 과정에서 다양한 특징 조합을 평가하여 가장 성능이 좋은 특징 집합을 선택하는 방법이다. wrapper 방법의 특징은 이 연구에서 사용된 U-net 모델을 활용하여 특징 조합으로 계절 내 예측 성능을 반복적으로 평가한다. 모델의 예측 정확도를 기준으로 변수의 중요도를 판단하게 된다. 이 방법은 모든 가능한 특징 조합을 테스트하거나 반복적으로 학습해야 하므로 자원 활용 비용과 시간이 많이 소요되는 단점이 있다. 그럼에도 불구하고, 최종적으로 모델 성능이 가장 뛰어난 특징 집합을 도출할 가능성이 높기 때문에 본 연구에서 wrapper 방법에 많은 학습 시간을 할당하였다. 구체적인 방법은 전진 선택법으로 초기 중요도가 낮은 변수부터 시작하여 반복적으로 가장 중요한 특징을 추가하는 방법이 있으며, 모든 특징을 포함한 상태에서 시작하여 중요도가 낮은 변수부터 하나씩 제거하여 민감도 분석을 실시하는 후진 소거법이 있고 전진 선택법과 후진 소거법을 결합하여 추가와

제거를 번갈아가며 최적의 조합을 찾는 단계적 선택법이 있다. 마지막으로 민감도 분석에서 embedded 방법은 입력 변수의 중요도를 평가하기 위해 학습 모델과 변수 선택을 결합한 방법으로 변수의 중요도를 통해 변수 중요도를 평가하는 방법으로 결정 트리나 랜덤 포레스트의 특징을 가진다. 딥러닝 모델의 성능을 저하시키지 않으면서 중요한 변수를 효과적으로 선택할 수 있으며 계산 효율성이 높은 장점을 가진다. 반면에 특정 머신러닝 모델에 의존적이므로 다른 유형의 모델로 일반화하기 어려울 수 있고 있는 어려움이 있다(Figure 2.13).

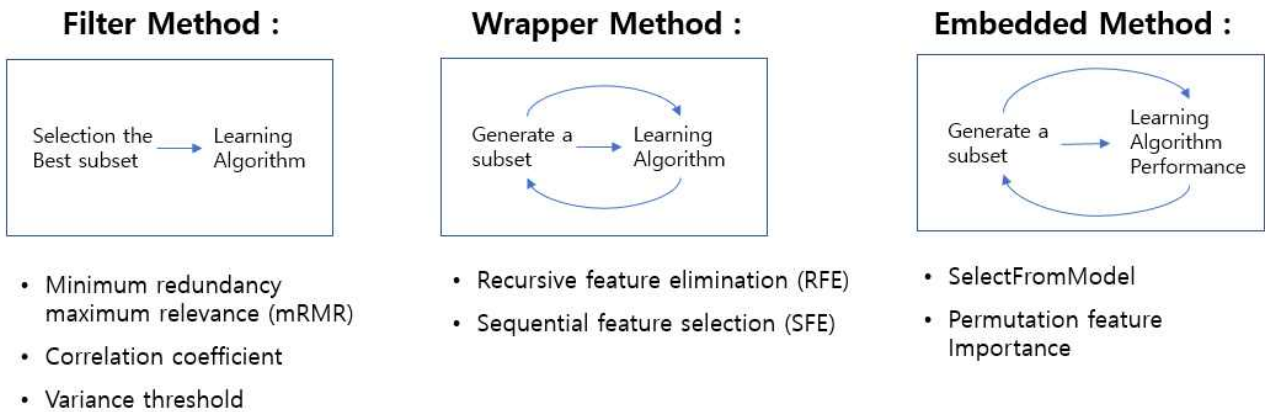


Figure 2.13. Sensitivity analysis methods for deep learning models.

2.1.2.4 Filter 기법

민감도 분석 기법중의 하나인 Filter 기법은 주로 신호 처리, 영상 처리, 그리고 기계 학습 분야에서 사용되는 기법중의 하나로 이미지 프로세싱에서의 필터인 이동 평균 필터는 시계열 데이터나 영상에서 잡음을 제거하거나 부드럽게 만들기 위해 사용되고 각 픽셀에 대해 주변 픽셀의 평균 값을 사용하여 중앙 픽셀 값을 대체한다. 가우시안 필터는 이미지에서 높은 주파수 성분을 감소시키기 위해 사용되는 필터로, 이미지를 부드럽게 만들거나 잡음을 감소시키는 역할을 하고, 신호 처리에서의 저패스 필터는 신호에서 낮은 주파수 성분만을 통과시키고 높은 주파수를 차단하는 필터방법이다. 기계 학습에서의 필터 민감도 분석 방법은 입력 데이터의 특징을 강조하거나 변형하여 정확도를 향상시키는데 사용된다. Attention U-Net의 바탕이 되는 합성곱 신경망(Convolutional Neural Network)에서 사용되는 컨볼루션 필터방법이 입력자료에서 특징을 추출하고 특성 중요도를 측정하여 중요한 특성만을 선택하여 정확도를 평가하고자 하였다. 따라서 ECMWF 모델 변수와 관측자료 변수에서 중요한 특정 패턴을 추출하여 정확도를 평가하고자 하였다. ECMWF의 MSLP, OLR, PREC, Q700, Q850, SST, T2M, TMAX, TMIN, U50, U850, V200, W850, Z200, Z500변수와 관측자료의 NDVI,SST을 ERA5의 최고온도와의 상관성을 변수별로 구해서 상관성이 높은 변수만을 선택하여 학습자료로 구성하고 Attention U-Net모델로 정확도 성능을 비교 분석하기 위하여 Filter 기법을 사용하였다.

모델/관측자료의 모든 변수와 재분석자료인 ERA5의 최고온도와의 상관성을 구한 결과 TMAX, TMIN, T2M, Z200, Z500, Q850, Q700, MSL, PRE, NDVI, SST의 순서대로 상관성이 높게 나와서 각각의 특성에 맞게 순서대로 입력자료를 만들어 Attention U-Net모델의 입력자료로 활용하였다(Figure 2.14: Figure 2.15).

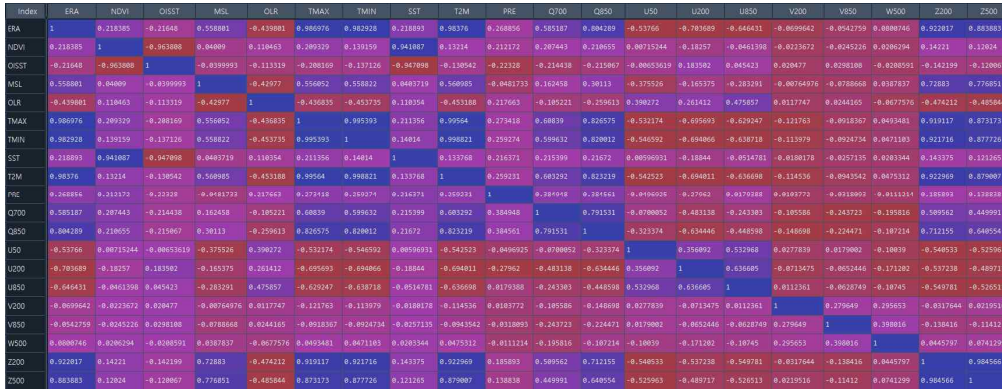


Figure 2.14. Correlation heatmap of observation and ECMWF model variables.

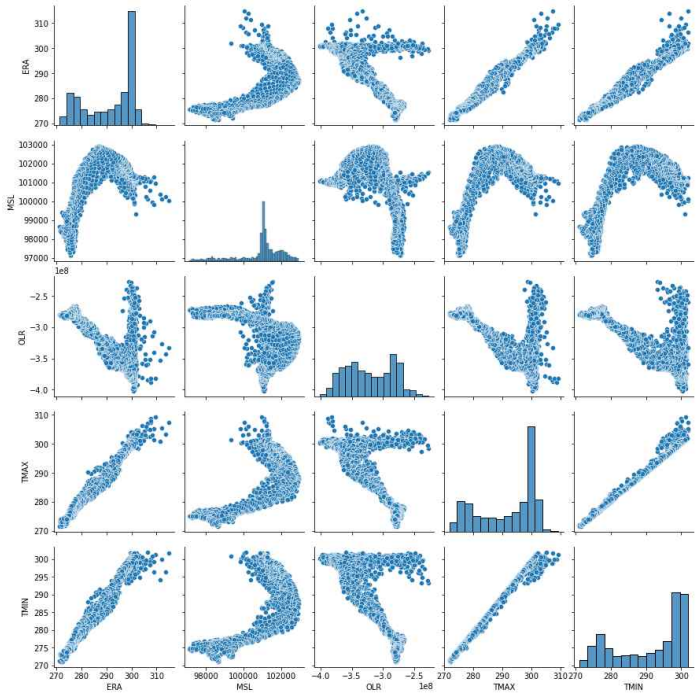


Figure 2.15. Scatterplot of correlation of variables.

2.1.2.5 Wrapper 기법

기계 학습에서 민감도 분석방법 중의 하나인 Wrapper기법은 단어에 주어진 의미와 같이 입력자료를 감싸거나 래핑하여 새로운 기능을 추가하거나 정확도 향상 민감도 분석을 나타내는 기법이다. 본 연구에서는 Attention U-Net모델에서 특성있는 변수의 특징을 알고자 이용하였다. 많이 이용되는 Wrapper 기법중에서 Feature Selection는 변수 특성 선택 알고리즘을 사용하여 가장 중요한 변수를 선택하고, 선택된 변수들로 Attention U-Net모델을 훈련하였다. Wrapper방법중에는 앙상블 기법을 이용하여 다양한 모델을 결합하여 정확도를 평가하는 방법

이나 여러 모델을 사용하여 각각의 예측을 다른 모델에 입력으로 사용하여 최종 예측을 생성하는 방식인데 본 연구에서는 딥러닝 모델의 Attention U-Net 단독모델을 이용하였기 때문에 모델별 특성을 이용한 민감도 분석은 사용하지 않았다.

2.1.2.6 Embedded 기법

Embedded 기법은 Filter와 Wrapper의 장점들을 결합한 방법으로 각각의 변수를 직접 학습하며, Attention U-Net 모델의 정확도에 기여하는 변수를 선택하여 학습자료의 입력변수로 선택하여 분석하였다. 본 연구에서는 회귀식을 통한 회귀계수의 변수가 높은 순서대로 각각의 학습결과를 비교분석하였다(Figure 2.16).

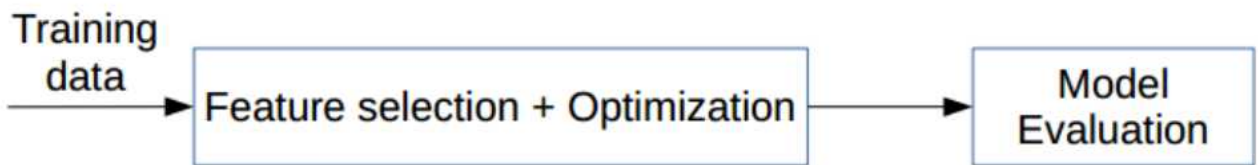


Figure 2.16. A diagram for embedded method for feature selection(Source: Trabelsi et al., 2021).

2.1.2.7 분석 및 검증

민감도 분석방법은 다양한 입력자료에 따른 변수 선택에 따른 정확도를 분석하기 위하여 Attention U-Net 모델을 고정하고 Filter, Wrapper, Embedded 방법에 따른 변수선택을 하여 학습자료를 구성하고 정확도를 비교 검증하였다. 다양한 입력자료들의 학습시간을 단축하기 위하여 파라미터를 최적화할 때 배치파일을 최대한 크게 잡아 학습하였다.

예측의 정확도를 평가하기 위하여 입력자료는 과거 관측자료와 리드타임 3주차 모델자료를 구성하여 ERA5 최고온도에 자료를 매치시키고 정확도를 평가하였다. 정확도를 평가하기 위하여 ACC를 통계방법을 이용하였다(Eq. [2.1]).

$$ACC(t) = \frac{\frac{1}{n} \sum_{y=1}^n (S'_{yt} - \bar{S}_t)(O_{yt} - \bar{O}_t)}{\sqrt{\frac{1}{n} \sum_{y=1}^n (S'_{yt} - \bar{S}_t)^2} \sqrt{\frac{1}{n} \sum_{y=1}^n (O_{yt} - \bar{O}_t)^2}} \quad [2.1]$$

여기서 S 는 Attention U-Net 모델에서 산출된 Filter, Wrapper, Embedded 기법에 활용된 변수들의 의해 학습된 최고온도자료이며, O 는 ERA5의 최고온도 자료이다.

2.1.2.8 입출력 시스템 설계

입출력 시스템은 계절 내 예측자료의 정확도 향상을 위한 자료확장과 자료의 변수의 민감도 분석을 쉽게 직관적으로 할 수 있도록 모델 학습에 필요한 데이터를 생성, 처리, 관리하는 시스템을 개발하고자 하였다. 이 시스템은 실제 데이터의 부족, 데이터 수집의 비용과 시간 문제를 해결하며, 딥러닝 모델의 성능 향상과 앞으로 다양한 응용 분야에서 활용될 수 있다. 데이터는 자료 확장으로 기존 모델데이터 이외에 새로운 관측자료를 추가하였으며, 데이터 전처리에서 입력 데이터의 정규화, 결측값 처리 등을 통해 학습에 적합한 형태로 변환하게끔 시스템을 설계하였다. 수집된 데이터를 기반으로 딥러닝 모델의 입력자료를 산출하도록 구성하였으며, 이를 바탕으로 생성된 자료가 모델의 예측성에 미치는 영향을 민감도 분석이 도움이 될 수 있도록 시스템을 구축하였다. 이 시스템은 대량의 데이터 확보가 어려운 상황에서 효율적으로 학습자료를 산출할 수 있으며, 다양한 시나리오와 조건에 대한 민감도 분석이 가능한 장점이 있다.

초기의 입출력 시스템은 feature selection하는 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 2.17).

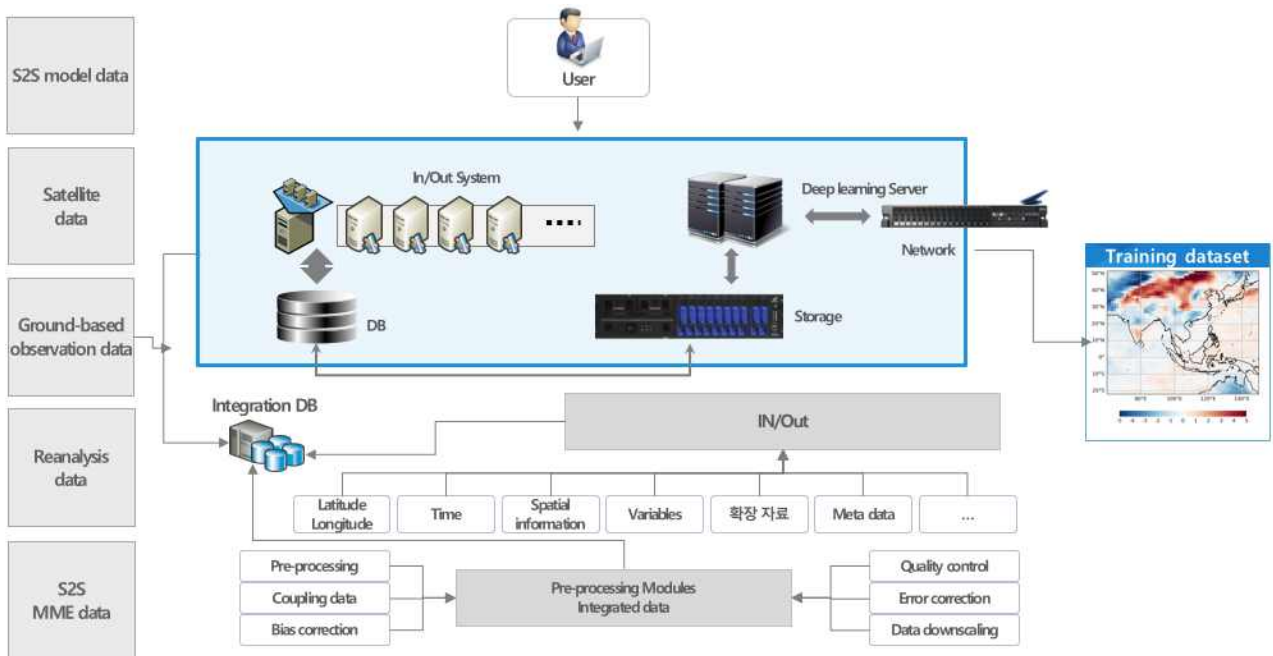


Figure 2.17. The structure of in and out data system.

입출력 시스템 모듈은 Clipping, 기간선정, 역학 모델 선택, 관측자료 선택, 최종자료 산출 가시화와 같은 모듈을 GUI환경과 연계되어 객체지향적으로 수행되는 시스템이다. GUI 환경은 상태바에서 “Processing” 으로 시작해서 모든 모듈이 돌아가서 자료를 산출하면 “End” 상태를 표시된다. 타겟변수는 강수, 최고/최저 온도를 선택가능하며, 기간은 1991년부터 2010년까지 20년간 자료를 선정할 수 있도록 표시하였고, 더불어 위, 경도를 입력하면 입력된 자료에 맞게 입력자료가 산출되도록 GUI를 구성하였다. 역학모델은 Glosea6, ECMWF을 선택가능하면 관측자료는 GPM, NDVI, SST등의 위성자료 선택 가능하다. Feature변수를 선택하고 학습할 딥러닝 모델을 선택하여 GenerateData 버튼을 클릭하면 최종 학습할 자료가 지정된 디렉토리에 생성된다. Visualization을 클릭하면 최종 학습자료가 잘 산출되었는지 자료 가시화를 통해서 확인 가능하다(Figure 2.18).



Figure 2.18. GUI of in and out data system.

2.1.2.9 입출력 시스템 고도화

초기의 입출력시스템에서 보다 직관적이고 사용자 중심의 입출력 시스템을 개발하기 위하여 다양한 모듈을 추가하여 연계하였다(Figure 2.20). 먼저 Daily 자료에서 Monthly 자료를 추가하였고, 학습을 위한 기간설정은 기 시스템의 사용자가 직접 입력하는 시스템에서 달력을 추가하여 직관적으로 기간이 선택할 수 있도록 모듈을 추가하였다. 학습지역 선택 역시 사용자가 지역을 마우스로 선택하여 위,경도 값을 받아 영역 선택이 가능하게 하였고, 관측자료는 GPM IMERG 이외에 NOAA CMORPH, PERSIANN 값을 추가하였다. 데이터 스케일링 기능을 추가하여 데이터를 정규화하였고, 입력자료를 학습자료와 테스트 자료로 나누어 주는 기능을 추가하여 고도화하였다.(Figure 2.19).

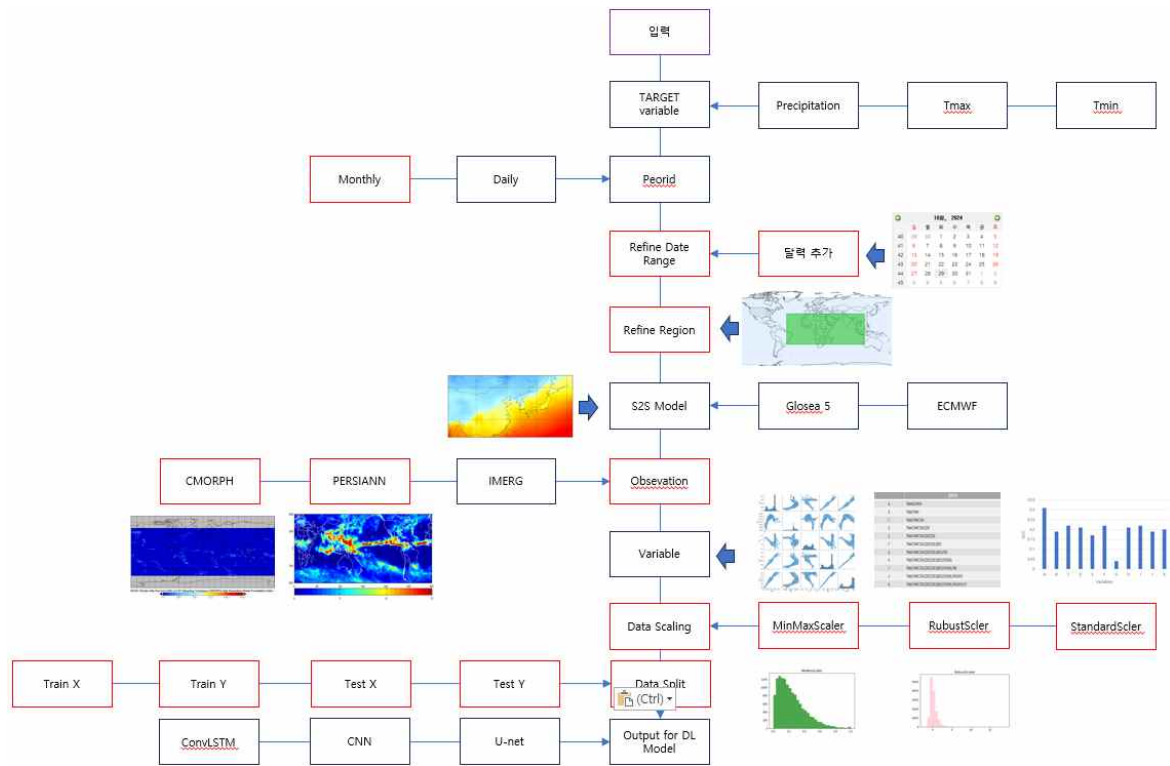


Figure 2.19. Advanced Input/Output System module diagram.

데이터 기간을 산정할 때 현재는 데이터를 직접적으로 입력을 해줘야 하는데 달력을 이용하여 데이터 입력을 받을 수 있도록 변경하였다. 파이썬의 calendar 라이브러리를 이용하여 기간을 지정할 수 있도록 모듈을 연계하여 직관적으로 기간이 선택되도록 수정하였다(Figure 2. 20).



Figure 2.20. Refinement data range by calendar python library.

기존의 텍스트 기반으로 위, 경도를 숫자를 직접 입력하는 방법에서 Map에서 커서를 이용하여 이용하고자 하는 위치를 사각형으로 드래그해서 만든 후 Clipping하여 영역선택이 가능하게 하였다. 주요 라이브러리는 tkinter과 관련 모듈을 설치하였는데 파이썬의 GUI 라이브러리로 활용하였다. 지도 시각화를 위한 라이브러리로 folium을 설치하여 대화형 지도를 생성하고 영역선택이 가능하게 하였다. from folium.plugins import Draw은 Folium의 플러그인 중 하나인 Draw를 임포트하게 코딩하였다. 이 플러그인은 지도 위에 도형(폴리곤, 선 등)을 그릴 수 있는 기능을 제공하기 때문에 이 플러그인을 이용하여 마우스를 이용하여 영역 폴리곤을 선택할 수 있도록 연계하였다. import webbrowser로 기본 웹 브라우저를 열어서 전세계 지도가 가시화 되게 코딩하였다. 선택된 위, 경도 정보들이 저장될 수 있도록 JSON과의 형식으로 위치 정보가 저장되도록 하였다(Figure 2.21).

```
import tkinter as tk
from tkinter import filedialog
import folium
import webbrowser
import json
import os
from folium.plugins import Draw
```

Figure 2.21. Libraries of python code for refinement of regions.

모듈로 제작된 영역선택 모듈을 실행하면 webbrowser 모듈에 의해서 기본 웹 브라우저가 실행되어 전세계 맵이 열리게 된다(Figure 2.22).

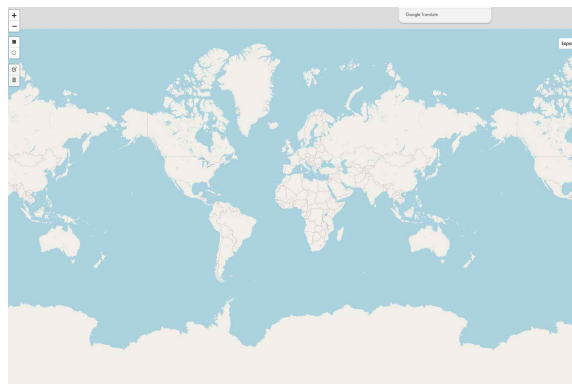


Figure 2.22. Webbrowser for refinement of regions.

실행버튼은 왼쪽바에 위치하며, 사각형, 원형으로 위,경도를 Clipping할수 있으며, 선택된 영역을 취소할 수 있도록 모듈이 코딩되었다(Figure 2.23).



Figure 2.23. Map editing Toolbar with save, cancel, and clear all options.

checkbox를 커서로 만들 수 있으며, 취소하고 재설정이 가능하도록 코딩되었으며, 체크박스를 조금 더 자세하게 미세조정하여 위,경도를 선정할 수 있는 메뉴가 있어 보다 정밀하게 위도, 경도를 설정가능하게 하였다(Figure 2.24).

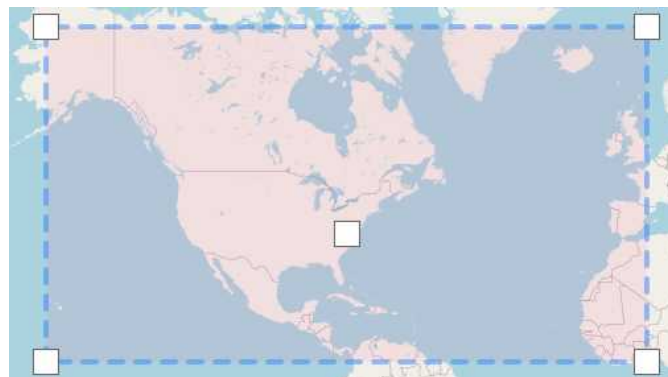


Figure 2.24. Refine region by toolbar.

최종적으로 체크박스과 원형으로 원하는 자료를 잘라낼수 있도록 코딩을 완료하고 원하는 지역을 커서로 선택하고 오른쪽 상단의 export 항목을 선택하면 위, 경도의 정보가 geojson 파일로 저장되어 출력되게 된다(Figure 2.25).



Figure 2.25.
Export
button.

최종적으로 저장된 geojson 파일을 찾아서 열게 되면 사용자가 지정한 위,경도 정보가 가시화되고, 정보를 x1, x2, y1, y2 변수로 자료를 받아서 Clipping 학습자료 산출에 이용하게 된다. 또한 데이터 전처리에서 많이 이용하게 되는 스케일링 방법을 모듈화하고 적용하였다. MinMaxScaler는 데이터 데이터 전처리 과정에서 사용되는 스케일링 기법으로, 주어진 데이터를 특정 범위(주로 0과 1 사이)로 변환한다. 이는 데이터의 범위를 조정하여 계절 내 모델 학습 시 각각의 변수들이 동일한 중요도를 갖도록 하고, 학습 속도를 높이며, 수렴을 안정적으로 만드는 데이터 전처리 기법으로 이렇게 만든 데이터 정규화는 딥러닝 모델이 특정 변수에 의존하지 않도록 하며, 스케일링을 통해 경사하강법의 수렴 속도를 개선해 전체적으로 자원 효율이 높아진다. 또한 많이 사용하는 standardscaler는 데이터의 변수를 표준화하여 평균을 0, 표준편차를 1로 맞추어 사용하게 되는데, 딥러닝 모델이 서로 다른 변수들로 인해 영향을 받지 않도록 한다. 이러한 스케일링 기법을 모듈화하고 입출력시스템에 연계하였다(Figure 2.26).

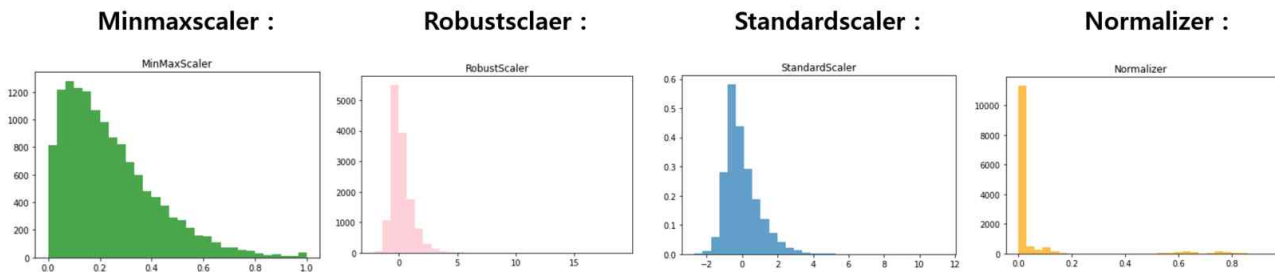


Figure 2.26. Data scaling methods for machine learning.

최종적으로 다양한 변수선택과 기간선정, 지역선택, 스케일링, 산출될 딥러닝 모델을 선택하고 산출된 자료가 학습자료와 테스트 자료로 이용될 수 있도록 사용자가 몇퍼센트로 나눌지 선택할 수 있도록 모듈화하고 연계하여 데이터가 학습자료와 테스트자료로 산출 될 수 있도록 개발하였다(Figure 2.27).



Figure 2.27. Data splitting for training datasets.

2.2. 전처리/초매개변수 기반 다중 인공지능 모델 도출 및 모델간 앙상블

신경망 모델에 훈련자료를 입력하기 전 훈련자료에 대한 가공이 필요한데 이 중 전처리(preprocessing)과정은 매우 중요하다. 흔히 전처리는 자료 전처리(data preprocessing)를 의미한다. 자료를 모델이 이해할 수 있는 형태로 변환하거나 자료의 품질을 올리는 일련의 과정으로, feature tuning이라고도 한다. 전처리는 여러 단계로 이루어지는데, 단계별로 처리된 자료는 3가지 형태로 명명할 수 있다. 첫 번째가 원시자료(raw data)로 초기에 수집된 원본 데이터로 분석이나, 머신러닝 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. 따라서 본 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 Engineering이라고 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 과정이다.

초매개변수 (hyperparameter, 초매개변수)는 모델을 생성할 때 사용자가 직접 설정하는 변수이고, 파라미터(parameter, 매개변수)는 모델의 학습 과정에서 생성되는 변수이다. 신경망 모델에서 에포크 (epoch, 학습 회수), 배치 사이즈 (batch size, 한 번에 학습하는 양), 학습률 (learning rate), 활성화 함수 (activation) 등이 초매개변수에 속한다 (Table 2.6). 그리드 탐색 (Grid Search)은 초매개변수를 일정한 간격으로 변경하며 최적의 파라미터를 찾아가는 기법이 그리드를 생성한다고 하여 붙여진 것이다. 그리드 탐색은 특정 범위 내에서 초매개변수를 일정 값만큼 일일이 변경, 즉 순차적으로 모두 입력한 뒤 가장 높은 성능을 보이는 출력값을 비교하는 방식으로 동작한다. 즉, 가장 높은 성능을 가진 값을 찾기 위해 가능한 모든 초매개변수 조합을 검색하는 기법이기에 때문에 계산 비용이 많이 들기도 하지만 모델에 가장 적합한 초매개변수를 찾는 데에는 가장 체계적이고 신뢰할 만한 접근 방식으로 알려져 있다. 그리드 탐색은 2가지 한계점이 있는데, 첫째는 최적의 해를 찾을 수 있다는 보장이 없다는 것이다. 그리드 탐색은 최적의 해를 찾지 못하는 경우가 생길 수 있다. 둘째는 동작 방식이 비효율적이다. 지정한 모든 구간을 탐색해야 하기 때문에 굳이 탐색하지 않아도 되는 초매개변수 값에 대해서도 탐색하는 경우가 생긴다. 예를 들어, 최댓값을 구하는 문제에서는 탐색할 필요가 없는 구간임에도 불구하고 그리드 탐색에서는 이러한 구간까지 탐색하기 때문에 탐색 과정 (시간 등)이 비효율적이라고 할 수 있기 때문이다.

선례 연구에서 도출된 전처리 방법과 초매개변수를 기반으로 당해연도(24년)는 앙상블 기법 적용을 통해 선행예측기간 1-2주의 미흡한 예측 성능을 개선하고 3-4주까지의 예측 성능을 유지할 수 있는 모델을 개발하기 위해 앙상블 기법을 도입하였다. 앙상블 학습을 통한 분류는

여러 개의 분류기(Classifier)을 생성하고 그 예측을 결합함으로써 보다 더 정확한 최종 예측을 도출하는 기법을 말한다. 특히, XGboost나 LightGBM 등과 같은 앙상블 모델을 정형 데이터의 분류나 회귀 분야에서 예측 성능이 매우 뛰어난 학습 모델을 만드는데 용이하다.

앙상블의 특징은 단일 모델의 약점을 다수의 모델들을 결합하여 보완하는 것이다. 서로 다른 유형이나 다른 계열의 모델들을 섞으면 오히려 더 성능이 나아지기도 한다. 예를 들어, 랜덤 포레스트나 뛰어난 부스팅 알고리즘들은 모두 결정 트리 기반 알고리즘으로 적용되는데, 결정 트리는 weak learner로써 과적합이라는 단점이 있지만, 이를 앙상블로 보완할 수 있기 때문이다. 앙상블의 유형에는 보팅, 배깅(예, 랜덤 포레스트), 부스팅(예, AdaBoost, 그라디언트 부스트, XGBoost 등), 스택킹 등이 있다.

Table 2.3. Comparison and analysis of the Characteristics of 4 Ensemble Techniques.

	Voting	Bagging	Boosting	Stacking
Learning/ Prediction Scheme	<ul style="list-style-type: none"> The final prediction is made by aggregating the prediction results of individual member models (e.g., majority vote) 	<ul style="list-style-type: none"> The final prediction is made by averaging the prediction results of individual member models 	<ul style="list-style-type: none"> Sequentially train individual member models while increasing the weights for samples that the previous model predicted incorrectly (Note: Weighting) 	<ul style="list-style-type: none"> The prediction results of the member model are trained in the meta model as new features to make the final prediction
Learning Materials	<ul style="list-style-type: none"> Use the same training data for each member model 	<ul style="list-style-type: none"> Provide different data to each member model through bootstrap sampling 	<ul style="list-style-type: none"> Input the same training data to each member model 	<ul style="list-style-type: none"> Input the same training data to each member model
Member Model Interaction	<ul style="list-style-type: none"> Interaction between member models 	<ul style="list-style-type: none"> Lower interaction between member models than other ensemble techniques 	<ul style="list-style-type: none"> Interaction between member models 	<ul style="list-style-type: none"> Interaction between member models
Weighting	<ul style="list-style-type: none"> Treated equally 	<ul style="list-style-type: none"> Treated equally 	<ul style="list-style-type: none"> Giving high weighting to data from models that predict incorrectly 	<ul style="list-style-type: none"> High weighting of high performing member models

본 연구에서 선정한 기법은 보팅, 배깅, 스택킹 및 부스팅 기법 4종을 선택하였으며, 각 기법의 멤버 모델로는 5단계에서 구축된 CNN 및 CNN-LSTM(2019-2021년), 6단계의 22년과 23

년에 구축된 U-Net 및 Attention U-Net을 수정하여 활용하였다. 앙상블 기법에서는 다양한 모델 구성이 중요하다. 그래서 선례 연구에서 구축되었던 Residual U-Net을 수정하여 ResNet 모델도 멤버 모델로 추가 구성하여 총 5개의 멤버 모델을 준비하였다.

2.2.1. 앙상블 기법 선정 및 특징 분석

앙상블 기법 특성 및 차이점을 분석하였다. 기법 간 차이, 즉 장단점 확인하여 기법 간 학습 및 예측 차이 및 훈련자료 샘플링 차이 등의 기법 특성 비교는 알고리즘 구축 및 예측 차이 등의 비교 분석에 활용하였다(Table 2.3).

(1) Voting

일반적 보팅(Voting)은 동일한 자료의 100%를 활용해서 모두 다른 모델로 학습하고 각 모델의 결과를 투표(Voting)를 통해서 결정하는 기법이다. 다시 말해, 보팅은 여러 개의 모델이 투표를 통해 최종 예측 결과를 결정하는 방식으로 서로 다른 알고리즘을 가진 모델(예, 분류기)가 같은 자료를 기반으로 학습되고 결합하는 것이다(Figure 2.28). 본 연구에서 선정한 기법 중 가장 간단한 앙상블 기법이다.

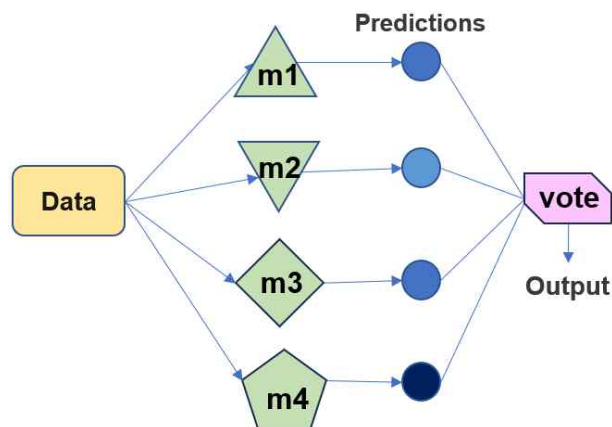


Figure 2.28. Explanation of the voting technique. Data is entered into m4 from various models m1 to m4 and voted on the predicted data (blue circles) to output the final predicted data.

(2) Bagging

배깅(Bagging)은 훈련자료에서 샘플링하는 방식, bootstrap aggregation의 줄임말로, bootstrap 통해 랜덤 샘플링한 자료를 같은 모델로 학습하고 각 모델의 결과를 투표(Voting)를

통해서 결정하는 기법이다. 그런데, 본 연구에서는 bootstrap 통해 학습 모델 구성은 일반적 배경과 유사하지만, 같은 멤버 모델을 사용하지 않고 다른 멤버 모델로 구성하였다(Figure 2.29).

배깅의 핵심은 자료의 샘플링이다. 전체 학습 자료에서 개별 하위 자료들로(sub-dataset) 샘플링하고, 대부분 결정 트리 알고리즘을 사용한다. 즉, 배깅은 같은 알고리즘 유형의 모델들이 이지만 자료 샘플링을 다르게 하여 학습 자료가 각각 다르다. 단, 교차 검증과 다르게 자료의 서브 세트간 중첩을 허용한다. 즉, 배깅은 훈련자료에서 중복을 허용하여 샘플링하는 방식, 다시 말해서, bootstrap aggregating의 줄임말이다. 샘플링된 개별 자료 서브 세트를 가지고 같은 결정 트리의 분류기(예, classifier)가 학습하고, 각각의 class에 대한 확률을 평균 내서 최종 예측값을 도출한다. 배깅의 대표적인 결정 트리 알고리즘은 랜덤 포레스트(Random Forest)이다. 랜덤 포레스트 알고리즘은 앙상블 알고리즘들 중에서 비교적 빠른 수행 속도를 보이며, 다양한 영역에서 높은 예측 성능을 보인다. 랜덤 포레스트의 자료 샘플링 방식은 전체 자료에서 일부가 중첩되게 샘플링된 자료, 이것을 부트스트래핑(Bootstrapping) 분할 방식이다. 현재 있는 표본에서 추가적으로 표본을 복원 및 추출하고 각 표본에 대한 통계량과 모델을 다시 계산하는 것이며, 데이터나 표본 통계량이 정규분포를 따라야 한다는 가정은 꼭 필요하지 않다. 즉, 원래 표본을 수천, 수백만 번 복제하는 것이라고 할 수 있다. 이를 통해 원래 표본으로부터 얻어지는 모든 정보를 포함하는 가상 모집단을 얻을 수 있다. 부트스트랩은 표본 통계량의 변동성을 평가하는 강력한 도구이다. 부트스트랩 재표본 추출 알고리즘은 첫째 샘플 값을 하나 뽑아서 기록하고 제자리에 놓고, n번 반복한다. 그 후 재표본 추출된 값의 평균을 기록한다. 이 단계를 R번 반복한다. R개의 결과를 사용하여 표준편차, 히스토그램, 신뢰구간 등을 찾는다.

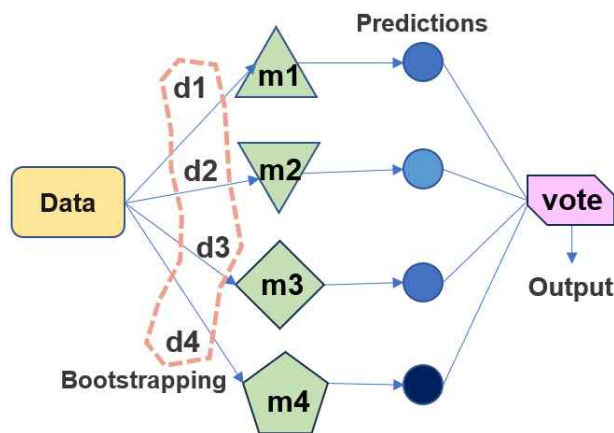


Figure 2.29. Explanation of bagging technique. D1 to d4 are random datasets made by bootstrapping segmentation in Data, entered into m1 to m4, voting on predicted data (blue circles) to derive final predicted data.

그래서 부트스트래핑(중복을 허용한 리샘플링)은 각 예측기(예, 모델)가 학습하는 자료에

다양성을 증가시키므로 배깅이 페이스팅(중복을 허용하지 않는 리샘플링 혹은 교차 검증)보다 편향이 조금 더 높다. 다양성을 추가한다는 것은 예측기들의 상관관계를 줄이므로 앙상블의 분산을 감소시킨다. 전반적으로 배깅이 더 나은 모델을 만들기 때문에 일반적으로 더 선호한다.

(3) Boosting

부스팅(Boosting)은 배깅의 변형으로 이전 모델이 예측하지 못한 자료에 대해 가중치를 두어서 다음 모델이 더 잘 학습할 수 있도록 하는 방법을 통해 bias를 감소시키는 기법이다. 약한 학습기를 여러 개 연결하여 강한 학습기를 만드는 앙상블 방법, 앞의 모델을 보완해 나가면서 일련의 예측기를 학습시킨다. 이전 모델이 과소적합했던(overfitting or underfitting) 훈련자료의 가중치를 더 높여서 이전 예측기를 보완하는 새로운 예측기를 만든다. 이렇게 하면 예측기는 학습하기 어려운 샘플에 점점 맞춰지게 된다.

Figure 2.30을 보면 알고리즘이 기반이 되는 첫 번째 분류기(예로 결정 트리)를 훈련자료에서 훈련을 시키고 예측을 만든다. 그 다음에 알고리즘이 잘못 분류된 훈련자료의 가중치를 상대적으로 높인다. 두 번째 분류기는 업데이트된 가중치를 사용해 훈련자료에서 훈련하고 다시 예측을 만든다. 그 다음에 다시 가중치를 업데이트한다. 즉, 각각 개별 약한 학습기(예, 멤버 모델)는 가중치를 부여해 결합한다. 그라디언트 부스트는 가중치보다는 이전 멤버 모델의 오차를 보정하도록 예측기를 순차적으로 추가한다. 에이다부스트처럼 반복마다 샘플의 가중치를 수정하는 대신 이전 예측기가 만든 잔여 오차에 새로운 예측기를 학습시킨다. 과대적합시에는 학습률을 감소시키거나 예측기 수(예, 멤버 모델의 수)가 너무 많으면 알맞은 개수를 찾기 위해 조기 종료 기법을 사용할 수 있다.

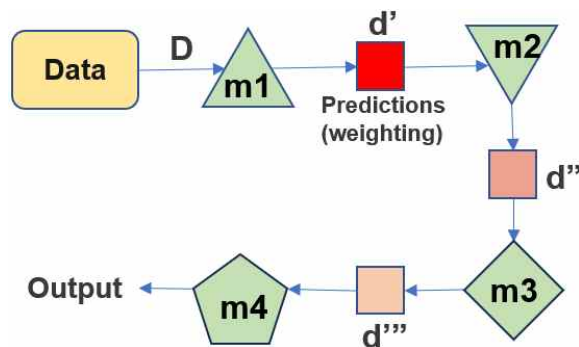


Figure 2.30. Description of boosting technique. Data is input to m1, weights are added to the predicted results (d') and input to the next model m2. Weights (d'') are added again to the predicted results in m2, and the predicted data is derived from the final model.

(4) Stacking

스태킹(Stacking)은 모든 멤버 모델의 예측 결과를 취합하는 간단한 함수를 사용하는 대신 취합하는 모델을 훈련시킬 수 없을까? 라는 기본 아이디어로 출발했다. 스택킹은 여러 멤버 모델들의 학습 결과가 메타 학습을 거치기 때문에 학습 데이터에 한하여 우수한 성능을 보일 수 밖에 없다. 그래서 주로 캐글 경진 대회에서 마지막 성능을 올리기의 편법(치트키)로 쓰이는 기법들 중 하나이다. 그러나 학습 자료와 학습 도메인(domain; 예, 학습 지역)이 다르거나 분산이 다른 자료가 들어오게 된다면 모델의 예측 및 분류 성능은 현저히 떨어지게 되는 것이 이 기법의 단점이다. 왜냐하면, 스택킹은 두 단계, 즉 첫 번째는 n개의 멤버 모델로부터 학습 결과를 생성하고, 두 번째 n개의 모델 결과로 메타 학습을 거치게 되므로 학습 데이터의 과적합(overestimating) 될 수 있기 때문이다. 그래서 스택킹의 장점은 단일 모델로 했을 때보다 성능은 향상되지만, 단점은 과적합이 될 수 있다.

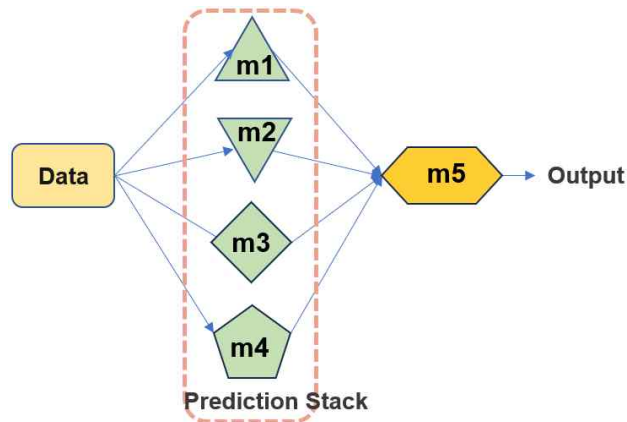


Figure 2.31. Description of stacking technique. Data is input from m1 to m4, predicted results are stacked (Prediction Stack), and the final prediction is output from m5 (meta model).

스태킹은 개별적인 여러 알고리즘을 서로 결합해 예측 결과를 도출한다는 점에서 배깅, 부스팅과 공통점을 가지고 있다. 하지만 가장 큰 차이점은 개별 알고리즘으로 예측한 자료를 기반으로 다시 예측을 수행한다는 것이다. 즉, 개별 알고리즘의 예측 결과 자료를 최종적인 메타 자료로 만들어 별도의 머신러닝 알고리즘으로 최종 학습을 수행하고 테스트 자료를 기반으로 다시 최종 예측을 수행하는 방식이다(Figure 2.31). 스택킹 모델은 두 종류의 모델이 필요하다. 개별적인 기반 모델, 즉 훈련을 수행할 개별 멤버 모델과 이 개별 기반 모델의 예측 자료를 학습 자료로 만들어서 다시 학습하는 최종 메타 모델이 필요하다. 다시 말해서, 스택킹 모델의 핵심은 여러 개별 모델의 예측 데이터를 각각 스택킹 형태로 결합해 최종 메타 모델의 학습용 피쳐 자료와 테스트용 피쳐 자료를 만드는 것이다.

2.2.2. 앙상블 기법 위한 멤버 모델 구성

앙상블 적용을 위한 멤버 모델의 수는 적어도 세 개에서 네 개의 멤버 모델 필요하다. 본 연구에서 ECMWF의 경우, 5개 멤버 모델 모두 참여시킨 경우 1개와 최소 4개 멤버 모델을 유지한 조합 실험 5개, 총 6개의 조합 실험 수행하였다. KMA(GloSea5)의 경우 4개 멤버 모델을 모두 참여시킨 경우 1개와 3개 멤버 모델로 구성된 조합 4개, 총 5개의 조합 실험을 수행하였다(Table 2.4).

Table 2.4. Explain the types and main characteristics of the member models that participated in the ensemble technique in this study.

	Key Features	Individual climate models participating as member models
CNN (Convolution Network)	A neural network developed to process 2D data such as images , and is effective in extracting and classifying features using convolutional layers and pooling layers	ECMWF, KMA
CNN-LSTM	It is a useful model for analyzing time series data or images by extracting spatial features of images or images with CNN and learning temporal order or continuous patterns with LSTM (Long Short Term-Memory)	ECMWF, KMA
ResNet* (Residual Network)	In order to solve the problem of gradient loss occurring in very deep neural networks, it is a neural network model that efficiently learns using a 'skip connection' structure that connects input directly to output	ECMWF
U-Net	A convolutional neural network mainly used for medical image segmentation, a model that preserves location information through symmetric skip connections with encoder-decoder structures while enabling fine-grained predicti	ECMWF, KMA
Attention U-Net	A model that enables more sophisticated and accurate medical image segmentation by adding an attention mechanism to the U-Net to focus on the features of critical locations	ECMWF, KMA

* ResNet does not participant as a member model of KMA

(1) 합성곱 신경망 (Convolution Neural Networks, CNN)

합성곱 신경망(Convolution Neural Networks, CNN)은 Convolution Neural Networks의 약자로 컴퓨터 비전 분야에서 이미지나 영상 데이터를 처리할 때 많이 이용되는 모델이며, 5단계(2019-2020년) 연구에서 구축 및 활용한 모델이다(APEC기후센터, 2020). CNN은 다층 퍼셉트론(Multilayer Perceptron, MLP)이 가진 한계점을 해결하기 위해 개발된 모델이다. MLP는 이미지와 같은 2D 구조의 자료를 평평하게 펼쳐서 처리할 수는 있지만, 이것 때문에 이미지의 공간 구조와 지역적 정보가 손실될 수 있다. 즉, 이미지 내의 패턴, 텍스처, 공간적인 관계 등의 중요한 정보가 무시되며, 이것으로 인해 MLP를 사용한 학습은 이미지 처리 작업에 비효율적일 수 있다. CNN은 이러한 문제를 극복하기 위해 등장한 모델이다. 일반적인 인공신경망(Artificial neural network, ANN)은 fully-connected 계층을 여러 층 쌓은 구조, 즉 기존의 ANN 구조 앞에 Convolutional Layer와 Pooling Layer를 추가한 구조라고 생각하면 된다. Fully connected layer만으로 구성된 ANN의 입력 자료는 1차원(배열) 형태로 한정된다. 그런데, 한 장의 컬러 사진은 3차원 자료이다. 배치 모드에 사용되는 여러 장의 사진은 4차원 자료이다. 따라서 사진 자료로 전 연결(FC, fully connected) 신경망을 학습시켜야 할 경우, 이 3차원 사진 자료를 1차원으로 평면화시켜야 한다. 사진 자료를 평면화시키는 과정에서 공간정보가 손실될 수밖에 없다. 결과적으로 이미지 공간정보 유실로 인한 정보 부족으로 ANN이 특징을 추출 및 학습이 비효율적이고 정확도를 높이는 것에 한계가 있다. 이미지의 공간정보를 유지한 상태로 학습한 모델이 바로 CNN이다.

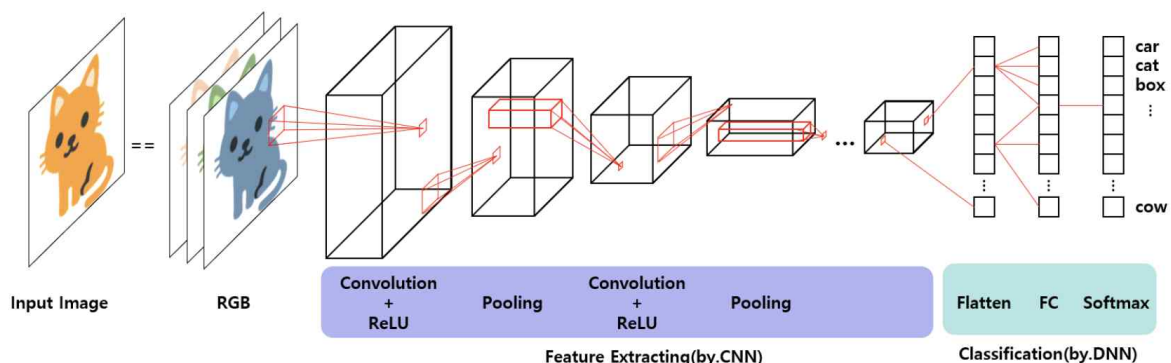


Figure 2.32. Description of the convolutional neural network model (Source: PhilAI.log: “[DL] CNN (Convolution Neural Networks) Concept Explained by a Non-Specialist” Blog).

CNN은 이미지를 입력 그대로 받아서 공간적/지역적 정보를 보존하면서 특성의 계층을 구축하는 것에 중점을 둔다. 이는 주로 두 가지 개념인 “합성곱 연산”과 “풀링”을 활용하여 이루어진다. CNN의 핵심 아이디어는 이미지를 작은 부분들로 나누어 보는 것이며, 이렇게 작은 부분을 통해 이미지의 지역적인 패턴과 특징을 감지하고 학습한다. 이를 통해 이미지의 한 픽셀과 그 주변 픽셀들간의 연관성을 살리고 이미지 내의 구조를 이해하는 데에 도움을 준다. CNN은 Figure 2.32와 같이 이미지의 특징을 추출하는 부분과 클래스를 분류하는 부분으로 나눌 수

있다. 특징 추출(Feature extraction/learning) 영역은 Convolution Layer와 Pooling Layer를 여러 겹 쌓는 형태로 구성된다. Convolution Layer는 입력 자료에 필터를 적용 후 활성화 함수를 반영하는 필수 요소이다. Convolution Layer 다음에 위치한 Pooling Layer는 선택적인 Layer이다. CNN 마지막 부분에는 이미지 분류(Classification)를 위한 Fully Connected Layer가 추가된다. 이미지의 특징을 추출하는 부분과 이미지를 분류하는 부분 사이에 이미지 형태의 자료를 배열 형태로 만드는 Flatten Layer가 위치한다.

(2) 장단기 기억 합성곱 신경망 (CNN-LSTM)

장단기 기억 합성곱 신경망(Convolutional LSTM, CNN-LSTM)도 5단계(2020-2021년) 연구에서 구축 및 활용한 모델로, 2020년에 구축한 CNN에 모델에 2019년에 구축한 장단기 기억(Long Short Term Memory, LSTM) 기술을 수정하여 결합한 모델이다(APEC기후센터, 2020). LSTM은 시계열 자료(Sequential data)와 관련한 머신러닝 문제를 해결하는 데 효과적이다. 시계열 자료의 예시는 텍스트인데, 예를 들어, 문장에서 각 단어는 앞의 단어에 종속된다. LSTM 모델은 이러한 순차적 종속성을 구현하기 위한 모델이다. 앞서 설명한 CNN은 이미지 분류, 객체 탐지, 분할 등 이미지나 동영상과 관련된 머신러닝 문제를 해결하는 일종의 딥러닝 모델로, convolution layer라는 특수한 유형의 layer를 사용하기 때문에 테두리나 윤곽선 등 이미지에서 학습한 패턴은 이미지의 픽셀 위치에 독립적인 것으로 가정하므로 가중치나 매개변수를 공유할 수 있다.

따라서 CNN-LSTM은 CNN과 LSTM을 연결해서 이미지나 동영상을 가져와 텍스트를 출력하는 하이브리드 모델을 구성할 수 있다(Figure 2.33). CNN-LSTM은 입력 자료(예, 이미지)에서 특징을 추출하는 합성곱 계층에 뒤이어 순차적 예측을 수행하는 LSTM 계층이 나오는 형태로 구성된다. 이 모델에서 CNN은 주로 입력 이미지를 가져와 고차원 특징이나 임베딩을 출력하는 인코더로 사용되고, LSTM은 텍스트를 생성하는 디코더로 사용된다고 할 수 있다. 그래서 CNN-LSTM 기술은 학습 자료를 CNN에 입력 값으로 넣어 특징을 추출하고, 추출된 특징값을 LSTM의 입력 값으로 넣는 방식으로 동작한다. 이 기술은 전력 예측, 코로나19 확진자 수 예측 등 주로 시계열 예측에 많이 사용된다.

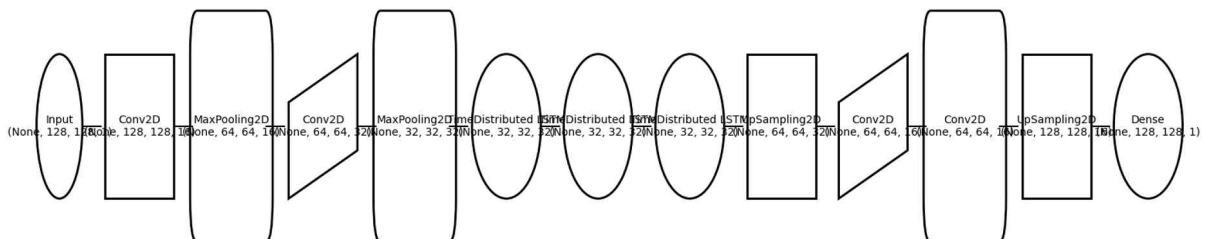


Figure 2.33. Diagram of the Long-Short Term Memory convolutional neural network model developed in this study.

(3) 잔차 신경망 (Residual Network, ResNet)

6단계(2022년) 연구에서 U-Net 모델을 확장하기 위해 개발한 Residual U-Net에서 U-Net과 결합하기 전 개발해 둔 잔차 신경망(Residual Network, ResNet)을 수정하여 활용했다(APEC기후센터, 2022). ResNet은 컴퓨터 비전과 이미지 처리에 있어서 CNN은 필수적인 모델이다. 이미지 분류를 위해 아주 깊은 모델이 만들어졌고 깊은 모델일수록 그 성능은 얇은 모델보다 뛰어났다. 어찌 생각하면 깊은 모델일수록 feature representation을 잘할 것이니 당연한 결과이기도 했다. 그러나 모델이 너무 깊어져서 10개 이상의 layer, 즉 수십 개의 layer를 가진 모델의 성능이 오히려 낮아지는 현상이 발생했다. 이 현상이 바로 degradation 현상이다. 과적합(overfitting)과 Degradation의 차이점은, overfitting은 딥러닝 모델이 훈련자료에 너무 적합하게 훈련되어 있어 막상 테스트 자료에서는 훈련자료에서 보여주었던 성능을 보여주지 못하는 경우를 말한다. 반면에 degradation은 훈련자료와 테스트 자료 모두에서 적은 layer로 구성되어 있는 네트워크보다 높은 오차를 보여주는 현상을 말한다. 그렇다면, layer가 너무 깊어져서 훈련자료에 overfitting이 되어 성능이 안 좋아진 것일까? 모델이 깊어져서 overfitting이 일어났다면, 깊은 모델의 훈련 손실이 얇은 모델에서 더 낮아야 하는데, 테스트 손실은 물론이고 훈련 손실에서까지 얇은 모델보다 깊은 모델이 성능이 안 좋았다. 그래서 내린 결론은 모델이 깊어질수록 성능이 안 좋아지는 이유는 overfitting 때문이 아니라 “모델 최적화(optimization)가 어렵기 때문이다.” 라는 결론을 내렸다. 모델이 깊어질수록 즉, gradient vanishing 혹은 gradient exploding 현상이 발생하기 때문이다. Optimization이 문제라면 새로운 Optimizer를 만들거나 깊어지더라도 쉽게 Optimization을 할 수 있는 새로운 아키텍처를 만들어 문제를 해결할 수 있다. 그러나 Optimizer를 새로 만드는 것은 매우 어려운 문제이기 때문에 새로운 Network를 만드는 데 집중했다. 그것이 Residual block의 skip connection이다.

예를 들어, Figure 2.34(아래)에서처럼 일반적인 CNN에서 나타나는 “main path”와 “Skip connection”에 의해 연결되는 “short cut”을 가지게 된다. 쉽게 Optimal depth를 알 수가 없다. 20층의 layer가 optimal인지, 30층이 optimal인지, 100층이 optimal인지 아무도 모른다. 하지만, degradation problem은 바로 이렇게 알 수 없는 optimal depth를 넘어가면 바로 일어난다. ResNet은 엄청나게 깊은 네트워크를 만들어주고, “Optimal depth”에서의 값을 바로 Output으로 보내버릴 수 있다. 이게 어떻게 가능할까? 바로 “Skip connection” 때문이다. ResNet은 Skip connection이 존재하기 때문에 Main path에서 Optimal depth 이후의 Weight와 Bias가 전부 0에 수렴하도록 학습된다면 Optimal depth에서의 Output이 바로 Classification으로 넘어갈 수 있다. 다시 말해서, Optimal depth 이후의 block은 모두 빈칸통이다.

요약하면, degradation을 해결하려는 방법으로 제시한 것이 잔차 학습(residual learning)으로 skip connection을 가진 Residual block을 만든 것이다. 그래서 Residual block은 기존 신경망은 $H(x)$ 가 정답값 y 에 정확히 매핑되는 함수를 찾는 것을 목적으로 했다. 즉, 신경망은 학습을 하면서 $H(x) - y$ 의 값을 최소화시키면서 결국 $H(x)=y$ 가 되는 함수를 찾았다. Figure 2.34에서 $H(x)$ 는 Identity를 매핑해 주는 함수이기 때문에 $H(x)-x$ 를 최소화하면서 $H(x)=x$ 가 되는 것을 목표로 한다(Figure 2.34). 기존 신경망이 $H(x)-x=0$ 을 만들려 했다면 ResNet은 $H(x)-x=F(x)$ 로 두어 $F(x)$ 를 최소화시키려고 한다. 즉, $F(x)=0$ 이라는 목표를 두고 학습을 진행한다. 이렇게 학습을 진행하면 $F(x)=0$ 이라는 목표 값이 주어지기 때문에 학습이 더 쉬워진다. 결국 $H(x)=F(x)+x$ 가 되는

데 이때 입력값인 x 를 사용하기 위해 쓰는 것이 Skip Connection이다. 즉, Skip Connection은 입력 값이 일정 층들을 건너뛰어 출력에 더할 수 있게 하는 역할을 한다.

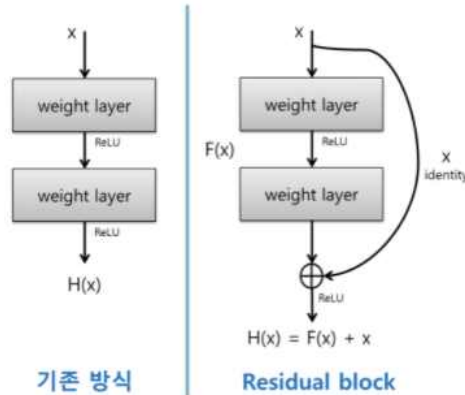


Figure 2.34. Residual learning block. The left side is the existing method, and the right side is the residual block with skip connection added.

(4) 유넷 (U-Net)

기본 U-Net은 2015년 Ronneberger et al. (2015)에 의해 제안되었으며 6단계의 1차원(22년) 연구의 기본 모델이다(APEC기후센터, 2022, 2023). U-Net은 이미지의 전반적인 특징 정보를 얻기 위한 수축 형태 (Contracting Path)의 네트워크와 정확한 지역화를 위한 팽창 형태 (Expanding Path)의 네트워크를 U자형의 대칭 형태로 갖는데, 기후 분야에서 활용도가 높은 모델이다 (Weyn et al., 2021). 이미지 세그멘테이션(image segmentation)이란 이미지의 모든 픽셀이 어떤 카테고리(예를 들면 자동차, 사람, 도로 등)에 속하는지 분류하는 것을 말한다. 이미지 전체에 대해 단일 카테고리를 예측하는 이미지 분류(image classification)와는 달리, 이미지 세그멘테이션은 픽셀 단위의 분류를 수행하므로 일반적으로 더 어려운 문제로 인식되고 있다.

U-Net은 이미지의 전반적인 특징 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), 선례 연구에서는 TimeDistributed 레이어를 U-Net의 Convolution 레이어에 추가하여 시계열 효과를 주고자 했다. TimeDistributed는 차원 변경 없이 간단하게 2D convolution을 사용할 수 있으므로 5차원 배열의 훈련자료가 바로 입력된다.

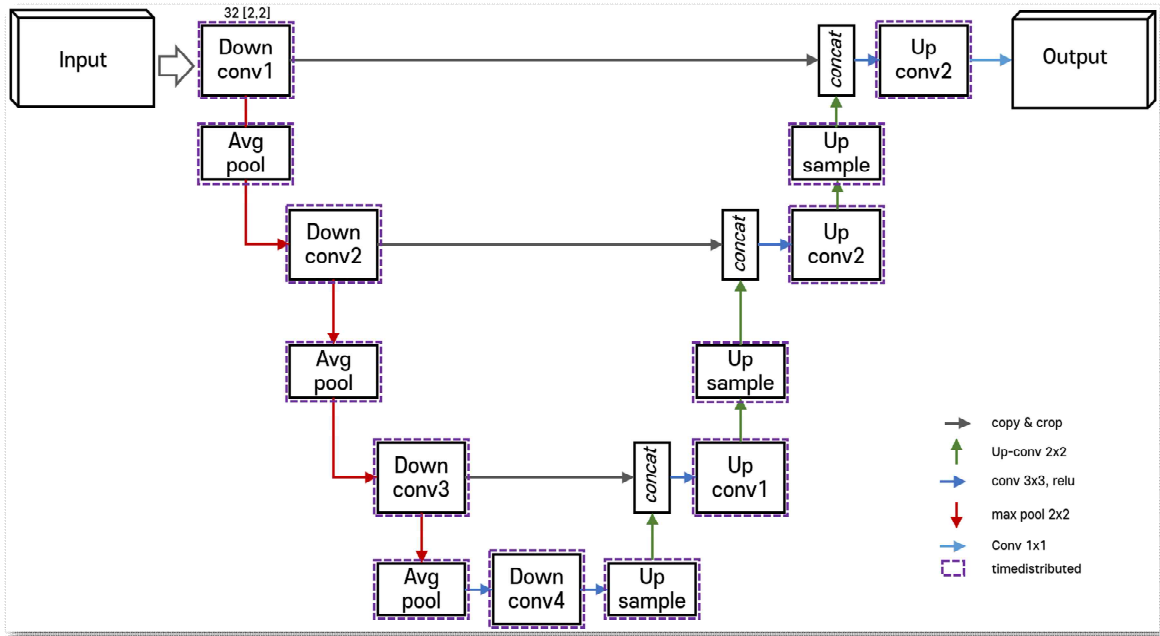


Figure 2.35. U-Net model built from previous research. Convolutional layers are wrapped with TimeDistributed layers (source: Chung et al., 2023).

Figure 2.35에서 semantic segmentation은 이미지 내에 있는 객체들을 의미 있는 단위로 분할해 내는 것이고, instance segmentation은 같은 카테고리에 속하는 서로 다른 객체까지 더 분할하여 semantic segmentation 범위를 확장한 것이다. 이미지 세그멘테이션은 의료 이미지 분석(종양 경계 추출 등), 자율주행 차량(도로면, 보행자 감지 등) 및 증강현실과 같은 광범위한 분야에서 사용되고 있다. U-Net은 'U-Net: Convolutional Networks for Biomedical Image Segmentation'이라는 논문에서 제안한 구조로서 매우 적은 수의 학습 데이터로도 정확한 이미지 세그멘테이션 성능을 보여주었으며 ISBI 세포 추적 챌린지 2015에서 큰 점수 차이로 우승했다고 한다. U-Net은 오토인코더(auto-encoder)와 같은 인코더-디코더(encoder-decoder) 기반 모델에 속한다. 보통 인코딩 단계에서는 입력 이미지의 특징을 포착할 수 있도록 채널의 수를 늘리면서 차원을 축소해 나가며, 디코딩 단계에서는 저차원으로 인코딩된 정보만 이용하여 채널의 수를 줄이고 차원을 늘려서 고차원의 이미지를 복원한다. 하지만 인코딩 단계에서 차원 축소를 거치면서 이미지 객체에 대한 자세한 위치 정보를 잃게 되고, 디코딩 단계에서도 저차원의 정보만을 이용하기 때문에 위치 정보 손실을 회복하지 못하게 된다.

(5) 집중 유넷 (Attention U-Net)

앞서 언급한 바와 같이 U-Net 모델은 image segmentation을 목적으로 제안된 fully-convolutional 기반 모델 형태가 U자형으로 context 포착과 세밀한 localization을 위해서 대칭으로 구성되었다. 이에 반해 CNN은 인간의 시신경을 통해서 모습이 보이는 것에서 착안하여 만든 딥러닝 모델이다 (Oktay et al., 2018). 우리의 눈은 전체를 보지만 일부를 집중해서 보는 것과 같이 겹치는 부분이 있는데, 이처럼 CNN은 이미지 처리에 특히 특화되어 있으며

convolutional layer, pooling layer가 있다. U-Net, U-Net++, VB-Net 등이 다양하게 사용되다가, 기존 U-Net의 학습 정확도를 개선하기 위해 Attention Gate(이하, AG) 장치가 장착된 Attention U-Net을 제안하였다 (Figure 2.36). 그래서 6단계 2차원에서 AG가 장착된 U-Net을 구축하였다 (APEC기후센터, 2023).

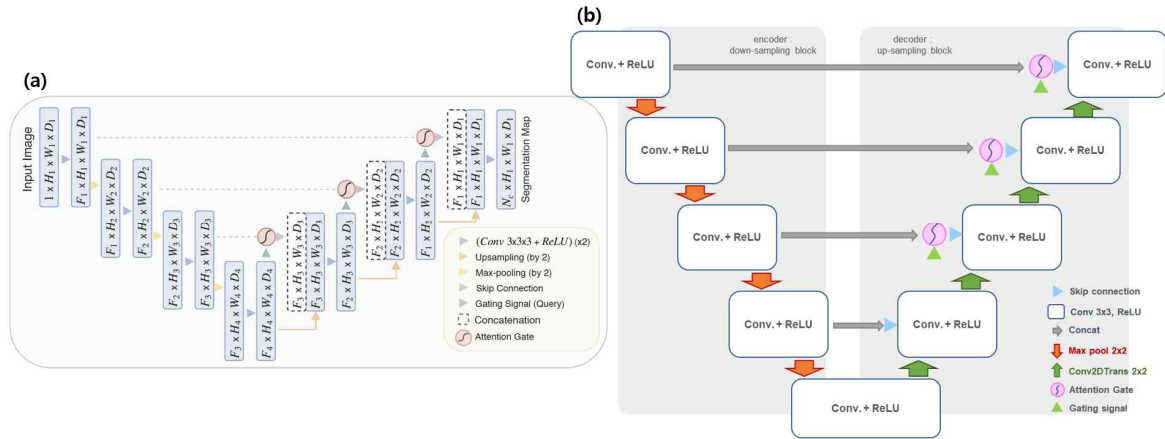


Figure 2.36. Attention U-Net developed in a previous study. A model that improves U-Net by inserting an attention block into the expansion process to better extract spatial characteristics. Attention U-Net: (a) is the original Attention U-Net (Source: Oktay et al., 2018), and (b) is the Attention U-Net schematized briefly in the study.

Attention이란 이미지 세그멘테이션에서 Attention은 학습할 때 관심 있는 영역 혹은 관련된 영역을 강조하는 방법이다. 예측이나 분류할 정답이 관련 없는 영역들을 무시하거나 관심을 덜 두기 때문에 계산 비용이 줄어든다는 특징이 있는데, 이미지의 특정 영역에만 집중 (Attention)한다는 것이다. 여기에 U-Net의 주요 특징은 Skip Connection이 있다는 것이다. 업샘플링이 진행되는 expanding path에서 소실되었던 공간 정보(다운 샘플링=contracting path에서 공간정보가 소실)를 다시 만드는데, 문제는 이 정보가 부정확하다는 것이다. 그래서 U-Net에서는 Skip Connection을 이용해서 다운 샘플링하는 contracting path의 공간정보를 업샘플링 path에 더해 이 문제를 어느 정도 해결했다. 하지만 Skip Connection을 사용함으로써 추출된 특징들이 중복된다는 부가적인 문제가 발생했다. 신경망의 초기 layer들의 특징 표현이 조잡(poor)하기 때문이다. Attention 메커니즘은 2016년에 computer vision에서 사용되었다(Bahdanau et al., 2016). Neural Network에서 prediction을 만드는 것이 무엇인지 이해하기 위해서 사용되었다. Attention 메커니즘은 입력에 따른 다음 결과를 예측하는 작업에서 특정 정보에 더 집중해야 하기에 encoder에서의 어떤 값에 집중해야 하는지를 찾는 방법이다.

AG(Attention Gate)를 다시 설명하면 CNN 아키텍처와 쉽게 통합할 수 있고, 최소한의 계산량 증가와 함께 모델의 민감도와 예측 정확도를 향상될 수 있다. 이것은 U-Net도 Convolutional layer에 ReLU 함수가 결합한 layer가 한 쌍으로 CNN에서 파생되었으므로 CNN과 잘 맞을 것이며, dimension이 줄어들기 때문에 비용(예, 계산 속도 등)이 감소한다는 것은 즉, 계산량이 작은 상태에서도 정확도를 높일 수 있으니 효율적이다. 이것은 layer가 깊어지면 모델 매개변수(parameter)가 과도하게 중복되고 유사한 저수준의 피쳐(feature)가 반복적으로 추출되므로 계산 리소스도 많이 들기 때문이다. 이것을 개선하려고 시도한 것이 AG이다. Hard

attention과 Soft attention이 있는데, Hard attention은 외재적으로 하이라이트할 영역을 지정해서 강조하며 미분은 불가능하고 강화학습을 필요로 하지만, Soft attention은 학습 도중에 내재적으로 task에 유리한 정보를 강조하고 불필요한 정보를 덜 강조하는 효과를 주는 것이 특징이다. Oktay et al. (2018)에서 제안된 AG들은 표준 U-Net에 통합되어 두드러진 점을 강조하고, AG는 정방향 패스와 역방향 패스에서 모든 뉴런의 활성화를 필터링한다. 배경 영역에서 발생하는 gradient는 역방향 패스 동안 가중치가 낮다. 이것은 주어진 작업이 관련된 공간 영역을 기반으로 얇은 레이어의 모델 매개변수를 업데이트할 수 있도록 해주는 것이다. 또한, 저수준의 feature map, 즉 첫 번째 skip-connection은 AG 기능에 사용되지 않는다. 즉, 고차원 공간에서 입력 데이터를 나타내지 않는다. deep supervision을 사용하여 중간 feature map이 각 이미지 스케일에서 의미론적으로 구별되도록 한다. skip connection의 작은 부분집합에서 조밀한 예측이 재구성되는 것을 방지하는 역할을 하게 된다. 1차원으로만 적용 가능한데, time-series 별로 혹은 feature 별로만 적용할 수 있었다. 이 때문에, Dual-Stage Attention 아이디어가 나오기도 했다(Liu et al., 2019; Qin et al., 2017). 또한, Attention은 길이 다시 말하면 time-sequence가 길어질수록, 내적 연산에 따른 속도 저하가 매우 크다고 알려져 있다. Beam-Search 알고리즘 등을 사용하면 속도 저하 문제를 조금 타파할 수는 있겠지만, 분류 문제에서 선택 가능하고, 회귀 문제에서는 선택 불가능하다. 예를 들면, Encoder와 Decoder의 time-series가 무척 긴 경우(예를 들면, 약 200개 이상씩 될 때), 매 과정에서 찾아 나가서 결정하다 보니 느려지는 경향이 있다고 한다. Attention은 정해진 Time-Sequence에서만 집중할 수 있다. 그래서 Attention 시계열 데이터 예측에 대한 연구가 비교적 최근부터 시작되었다. 하지만 현실의 문제는 다르다, 실제 데이터가 100일 범위로만 결정될 수 있을 것으로 생각했지만, 특정 어떤 데이터는 200일 전 데이터가 Attention이 더 잘 될 수도 있기 때문이라고 Qin et al. (2017)은 언급하였다. 이런 데이터 구성의 경우 Attention이 잘 동작하지 않는다. 또한, 대부분의 문제 해결 방식이 CNN을 이용하고 Attention을 진행하기 때문에, CNN을 적용하는 부분이 데이터셋과 적합한 모델 아키텍처 구조인지도 판단해 봐야 한다고도 언급하였다(Shih et al., 2019). 그 데이터의 카테고리가 매우 제한적이며 실제 사례에 적용된 레퍼런스를 찾아볼 수 없다. Shih et al. (2019)에 따르면 Attention을 이용하여 Time-Series Forecasting을 하는 것은 시기상조라고한다. Attention은 어떤 한 시점을 기준으로 유한한 단어를 학습시키기 위해 최적화 되어있고, 현실 세계에서 주가와 같은 무한한 실수 및 정수 예측 문제에서는 활용할 수 없다는 것이다. 대부분 %와 같은 상대적인 유한한 값으로 치환하여 사용한다든지 하는 형태로 활용만 하고 안정성이 떨어진다고 한다. 또한, Attention 기반의 Time-Series Forecasting은 현재로서는 불안정하고 아직 예측성이 부족하다고 언급하였다. 그래서 seq2seq나 Encoder-Decoder 기반의 BI-LSTM 형태로 구성하는 것이 현재로서는 가장 합리적인 Time-Series Forecasting 방법이라고 제안하였다.

2.2.3. 계절내 기후 예측 훈련자료

본 연구에서는 기존 연구에서 사용한 훈련자료(버전 2018)를 활용하지 않고 새로운 2022년 버전으로 교체하였다(Table 2.5). 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)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)로 기존 연구에서 활용한 6개 개별 기후모델 중 2개의 개별 기후모델만 이용하였다.

당해연도 연구에서 개별 기후모델의 기후 변수는 22년 연구 결과에서 계절내 일 최고기온의 딥러닝 훈련 및 예측에서 영향을 미치는 기후 변수 중심으로 선정하였다. 그래서, 올해 훈련자료에 입력한 개별 기후모델의 기후 변수는 TMAX 및 TMIN(Maximum and minimum air temperature at 2m), T2M(2 meter air temperature), SST(Sea surface temperature), OLR(Top net thermal radiation), Z200 및 Z500(Geopotential height at 200 and 500hPa) 등 약 7개이다 (Table 2.5). 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(Figure 2.37).

또한, 앙상블 수, 과거 재현 예측(이하, Hindcast) 자료의 생산 방법, 선행예측시간(lead-time), Hindcast의 기간, 예측 자료의 생산 주기 등이 모두 다른데, S2S 예측 자료의 앙상블은 기후모델의 기후 변수가 증가하면 (그) 앙상블도 증가하여 기존 훈련자료에서는 기후모델의 기후 변수의 앙상블들을 모두 평균하여 입력했지만, 2021버전의 훈련자료에서는 Hindcast의 앙상블을 그대로 유지하였다. 선행예측시간도 각각 46일과 60일로 구축하였으며, S2S 예측 자료들은 학습에 활용하기 쉽게 다차원 배열 (예, 시간, 선행예측시간, 위도, 경도, 변수로 구성된 5차원 배열) 데이터 구조로 구성하였다. 또한, 훈련자료의 전처리는 Chung et al.(2023)에서 일 최고기온 훈련에 영향을 미치는 전처리 방법, 즉 표준화 기법으로 처리하였다.

Table 2.5. S2S (Sub-seasonal to Seasonal) climate models used in this study, number of ensembles, forecasting time range, length, and period of hindcast of each models.

Climate model	Producing center	Hindcast ensemble size	Hindcast length	Forecasting frequency	Forecasting time range
KMA (GloSea5)	Seoul	3	1993-2016	4/month (1,9,17,25)	60 days
ECMWF	ECMWF	11	2001-2020	2/week (Mon, The)	46 days

2.2.4. 구동 환경, 연구 지역 및 평가 방법

전체 훈련자료에 대한 훈련 기간과 테스트 기간은 개별 기후모델의 Hindcast 기간에 따라 각각 달라지는데 Hindcast 기간 전체에 대해 80%는 훈련 기간, 20%는 테스트 기간으로 계산된다. 예를 들면, GloSea5의 Hindcast 기간이 1993년부터 2016년이므로 1995년부터 2010년까지 16년 (80%)은 훈련 기간으로, 2011년부터 2014년까지 4년(20%)은 테스트 기간이 되며 교차 검증을 수행하였다.

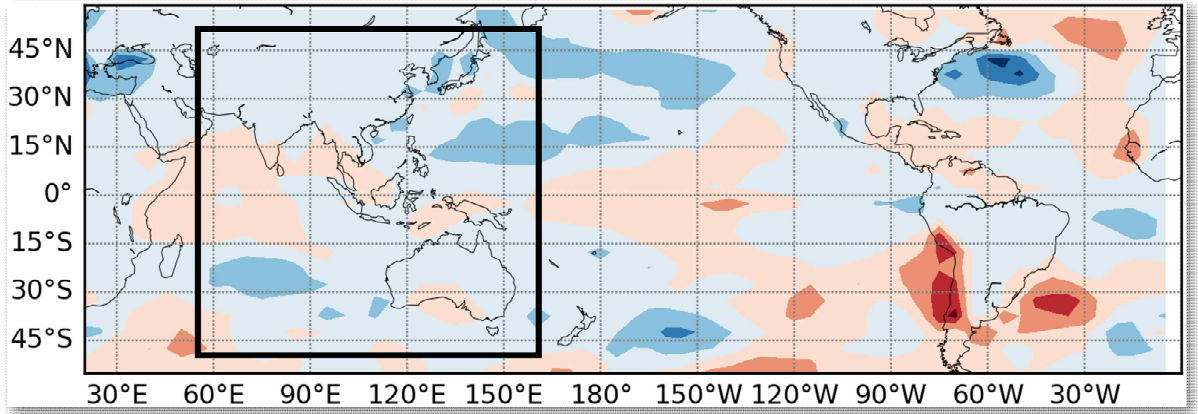


Figure 2.37. The domain of the constructed training data of Sub-seasonal to seasonal. The black box indicates the input and output domains of the deep learning training model in this study.

연구 지역은 합성곱 신경망 규모 등을 고려할 때 교차 검증(k-fold cross validation)은 시간이 많이 걸리지만, 올해 연구에서도 기존 연구와 같이 이런 점을 고려하여 4-fold (4겹) 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)로 72의 격자를 가진 동아시아 지역으로 설정하였다(Figure 2.37).

Table 2.6. Information of hyper-parameters and their ranges set in the study.

Hyper-parameter	Setting of the study
epoch	100
batch size	64
learning rate	0.0001(1e-04)
Optimizer	Adam
Activation function	ReLu
Filter size (Kernel size)	16, 32 (CNN, CNN-LSTM, U-Net, Attention U-Net) 32, 64, 128, 256, 512 (ResNet)

양상블 내 멤버 모델은 케라스 (Keras) 및 텐서플로우 (TensorFlow) 프레임워크를 사용하였으며, 각각의 멤버 모델의 합성곱 신경망 모델들의 최적화를 위한 손실함수 (loss function)는 평균절대오차 (Mean Square Error, MSE)로 계산되었으며, 측정함수 (metrics)로는 일 최고기온에 대한 오차와 정확도 평가를 위해 평균평방근오차 (Root Mean Square Error, RMSE)를 계산하여 학습하였다. 예를 들면, 예측 일 최고기온이 생성되면 예측 시점의 관측 일 최고기온(예,

ERA5 재분석 자료)의 편차를 나타내는 손실함수가 계산되는 것이다. 본 연구에서 사용한 초매개변수는 Table 2.6이다. 합성곱 레이어(Convolutional layer)에 적용할 필터 크기(Filter size, 커널 크기(Kernel size))의 경우 ResNet 모델의 경우는 일반적으로 가장 많이 사용되는 필터 크기, 32, 64, 128, 256, 512로 구축하였고, CNN, CNN-LSTM, U-Net, Attention U-Net 모델의 필터는 선례 연구에서 도출된 16, 32 사용하였다.

앙상블 기법의 학습 및 예측 결과의 공간적 예측 정확성 및 시간적 변동성은 편차상관계수 (Temporal anomaly Correlation Coefficient, TCC) 및 공간편차상관계수 (spatial Pattern anomaly Correlation Coefficient, PCC)를 계산하여 평가하였다 (Eq. 2.2 및 Eq. 2.4).

$$TCC_i = \frac{\sum_{j=1}^N (x_{i,j} - \bar{x}_i) \times (y_j - \bar{y})}{\sqrt{\sum_{j=1}^N (x_{i,j} - \bar{x}_i)^2} \times \sqrt{\sum_{j=1}^N (y_j - \bar{y})^2}} \quad [2.2]$$

여기서 x_i 는 학습 전 혹은 학습 후 각각의 개별 기후모델(KMA 및 ECMWF) 기반의 S2S 예측 기후자료의 예측 날짜 i 일째의 값이고, \bar{x}_i 는 훈련 기간 동안 개별 기후모델 기반 계절내 예측 기후자료의 평균, 즉 모델의 기후 평균값을 나타낸다. y_i 는 i 일째에서의 관측 자료, \bar{y}_i 는 훈련 기간 동안 관측 자료의 평균, 즉 기후값을 나타낸다. j 는 선행예측시간, N 은 예측 개수를 각각 나타낸다.

또한, 공간적으로 계산된 TCC의 차이를 평가하였다. 각각의 앙상블 기법의 TCC에서 관측 자료의 ACC와의 차이를 계산하였다 (Eq. 2.3).

$$Difference = \frac{ACC_{pred} - ACC_{raw}}{ACC_{raw}} \quad [2.3]$$

여기서, ACC_{pred} 는 훈련 모델에 의해 후보정된 목표변수의 예측 성능을 평가한 지수 (훈련 후 평가된 ACC)이고, ACC_{raw} 는 훈련 모델에 입력하기 전, 즉 학습되지 않은 목표변수의 원시자료 예측 성능을 평가한 지수 (훈련 전 평가된 ACC)이다.

$$PCC_j = \frac{\sum_{i=1}^N \Delta x_{i,j} \Delta y_{i,j}}{\sqrt{\sum_{i=1}^N \Delta x_{i,j}^2} \sqrt{\sum_{i=1}^N \Delta y_{i,j}^2}} \quad [2.4]$$

여기서, j 는 예측 선행기간이고 i 는 격자이다. $\Delta x_{i,j}$ 와 $\Delta y_{i,j}$ 은 격자 i 에서 선행 예측기간 j 의 평균 아노말리(anomaly)이다. N 은 연구지역 전체 격자 개수이다.

2.3. 준지도학습을 통한 MJO 지수 확률예측 인공지능모델 개발

2.3.1. 사용된 자료 및 전처리 방법

준지도학습을 활용한 MJO 위상 분류 연구에서는 딥러닝 모델에 사용할 입력 변수로 널리 활용되고 있는 MJO 인덱스인 RMM 계산에 사용되는 Outgoing Longwave Radiation (OLR), 850hPa 및 200hPa의 상하층 동서 방향 바람장(U-component winds at 850 and 200 hPa; U850, U200)과 VPM 인덱스에 사용되는 Velocity Potential (VP)을 사용하였다. U850 및 U200 자료는 센터에 구축된 ECMWF ERA5 재분석자료(Hersbach et al., 2020)로부터 계산된 1.5° 공간해상도의 일일 평균 자료를 사용하였다. OLR은 센터에 구축된 NOAA Interpolated OLR 자료를 활용하였다 (Liebmann and Smith, 1996). NOAA OLR 자료가 2.5° 공간해상도로 제공되고 있으므로 사용한 나머지 변수들을 선형 보간 방법으로 평균하여 같은 해상도로 맞추어 활용하였다. VP는 200hPa 동서 및 남북 바람장을 사용하여 계산하였다 (Dawson, 2016). MJO 위상 분류 모델링은 목적변수로 총 9개의 클래스(class), 8개의 MJO 위상과 Inactive MJO (Amplitude 1 이하)를 사용한다. 자료 전처리는 기본적으로 Wheeler and Hendon, 2004에서 제시된 방법을 따른다. MJO 이외의 파동 영향을 제거하기 위해 수행된 세 가지의 장기 계절 및 경년 시그널 제거 전처리 절차들은 Figure 2.38과 같다. Climatology mean, first three harmonics가 제거되며, 이전 120일 평균 제거 절차들이 적용된다. 전처리 된 변수들을 기반으로 EOF 분석이 수행되고, PC1 및 PC2가 각각 RMM1 및 RMM2로 사용된다. RMM1 및 RMM2의 강도가 1 이상인 경우 active MJO라고 정의하였다. 자료 누수(Data leakage) 문제를 예방하기 위해 모든 전처리 및 정규화 계산은 훈련 기간에 대해서 수행하였으며 훈련 기간에서 산출된 중간 통계치를 검증 및 테스트 자료에 적용하여 구축하였다. Figure 2.39은 최종적으로 구축된 각 MJO 위상의 OLR에 대한 아노말리 합성장(Composites)을 보여준다. 각 위상의 이미지 오른쪽 하단은 위상별로 가용한 이미지 자료의 개수를 보여준다.

딥러닝 모델을 학습하기 위해서 훈련(Training) 자료와 훈련 중에 모델의 과적합(over-fitting)을 방지하기 위해 검증(Validation) 자료를 사용하였고, 학습된 최종 모델을 평가하기 위해서 테스트(Test) 자료가 사용된다. 테스트 자료는 모델 훈련에 사용되지 않은 완전히 독립적으로 구분된 자료셋이다. 준지도학습 기반 MJO 위상 분류 연구에서는 자료 분리 방법에 따른 모델 성능을 테스트하기 위해 시간적 분리(Temporal Split) 및 무작위 분리(Random Split) 방법을 사용하였다 (Figure 2.40). 시간적 분리는 전체 시계열 자료를 훈련, 검증, 테스트 자료를 각각 1975-2005 (31년), 2006-2013 (8년), 2014-2022 (9년)으로 나누었고, 무작위 분리 방법은 훈련과 검증 기간에 대해서 랜덤하게 각각 67% 및 33%로 나누어 활용하였다.

선행시간별 MJO RMM 지수 예측 연구에서는 주요 입력 변수로 850 및 200 hPa의 U-성분

바람(U850, U200), OLR, 200 hPa VP이다. 이 변수들은 MJO의 대규모 대류 순환을 반영한다 (Wheeler & Hendon, 2004). 추가로 표면 온도(TS), TCWV(Total Column Water Vapor), dqdt(습도 경향성) 및 horizontal advection(hadvect; 습도 수평 이류)이 고려되었다. hadvect는 dqdx와 dqdy의 합산 값으로 계산되었다. 입력 변수의 편차 자료는 MJO 시그널을 강조하는 데 사용되었으며, 배경장 상태를 반영하기 위해 배경장 입력 변수로 사용하였다. 배경장 상태는 원자료에서 편차를 빼는 방식으로 구축하였다. 자료 수집과 전처리는 앞서 설명한 것과 동일하게 수행되었다. EOF 분석은 1979-2001년의 수평 바람장과 OLR 필드에 적용되어 RMM1과 RMM2를 구하였다. 1979-2021년의 전체 자료셋은 훈련(1979-2012), 검증(2013-2016), 테스트(2017-2021) 데이터로 나누었다. 데이터 정규화는 훈련 데이터의 파라미터를 기반으로 수행되어 데이터 누수 문제를 방지했다.

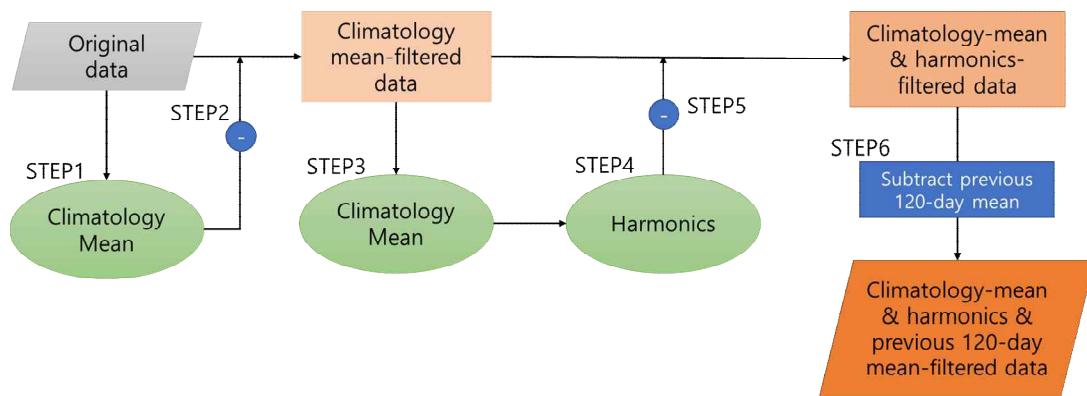


Figure 2.38. Steps of data preprocessing for filtered anomaly data.

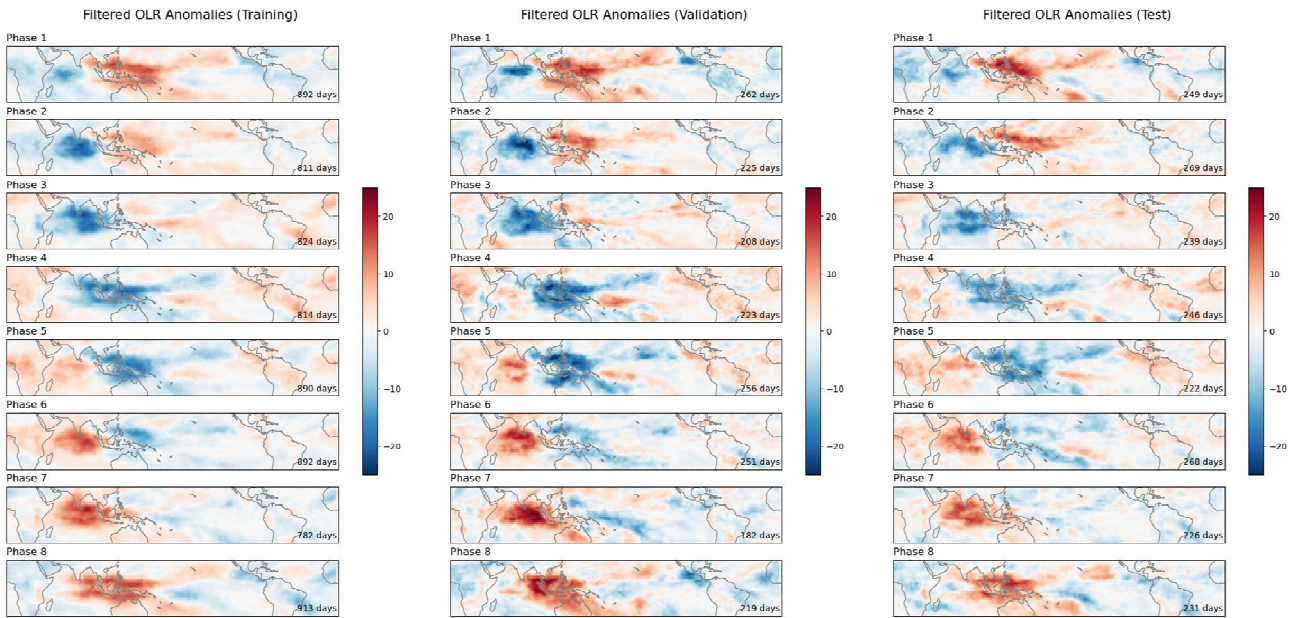


Figure 2.39. Composite images of the filtered OLR anomaly (W/m²) for each MJO phase for training (left), validation (center), and test (right) datasets, respectively. The number of days available for each phase are given in the bottom right corner. Negative (Positive) anomalies indicate more (less) rain.

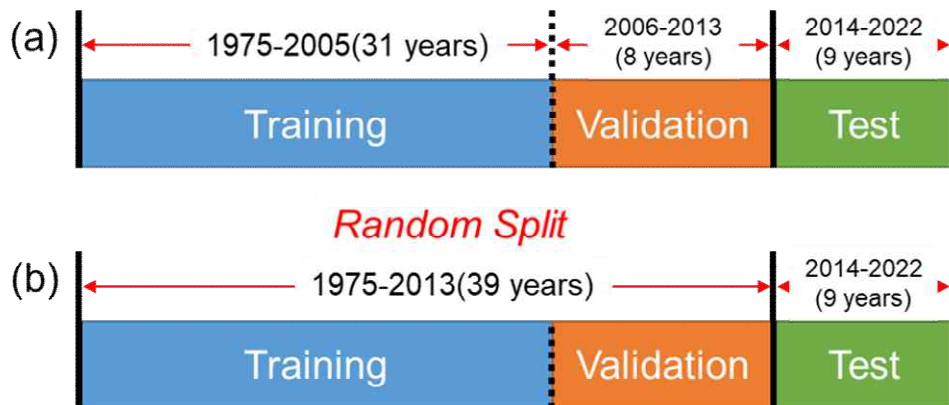


Figure 2.40. Data split strategy used for semi-supervised learning-based MJO phase classification. (a) Temporal split, (b) Random split method.

2.3.2. 준지도학습을 적용한 MJO 위상 분류 모델 개발 방법

2.3.2.1. 준지도학습 기준 모델 구축 방법

준지도학습 적용 MJO 위상 분류 모델에서는 준지도학습을 위한 기준 모델을 선정하기 위해서 다양한 이미지 기반 모델을 평가하였다. Table 2.1은 다양한 변수 조합, 자료 분리 방법, 모델 종류 및 모델의 Fully Connected Layer(FC 층)에 따른 모델의 민감도 실험을 위한 다양한 조건들을 보여준다. 기타 설정으로는 Adam 최적화 함수를 사용하였고(Kingma and Ba, 2014), 초기 학습률(Learning rate)은 $5e-3$ 으로 설정하였으며, 배치 크기(Batch size)는 32, 에포크

(Epoch)는 50으로 설정하였고, 과적합 방지를 위한 조기 종료(Early Stopping)는 10 epoch로 설정하였고, 사용된 모델들은 파이토치(Pytorch)의 토치비전(Torchvision)에서 제공되는 사전 학습(Pretrained)된 모델을 사용하였다. 사전 학습된 모델이란 ImageNet 자료셋에 대해 사전 학습된 모델을 말한다. 사전 학습된 모델을 활용하는 것이 더 높은 성능을 보여주었다.

Figure 2.41은 본 과제에서 수정 활용한 VGG16 모델의 구조를 보여준다. VGG16은 Simonyan and Zisserman (2015)에 의해 개발된 모델로 단순히 모델 구조의 깊이(depth)를 증가시켰을 때 성능이 어떻게 증가하는지 실험하기 위해 제안된 방법이다. 이러한 접근법은 여러 이미지 분류 문제의 정확도를 높일 수 있었다. 하지만 층이 깊어질수록 gradient vanishing, exploding gradient, 과적합 등의 문제가 발생하게 되어 깊이가 너무 깊어지면 오히려 성능이 저하되는 것을 보였다. 또한 VGG 모델은 학습해야 할 파라미터의 수가 많아 학습에 많은 시간이 소요되는 문제가 있다. VGG 모델의 한계점을 개선하기 위해서 He et al. (2016)는 ResNets (Residual Nets) 모델을 제안하였다. Figure 2.42는 ResNet 모델의 Residual 학습 구조를 보여준다. 모델은 skip connection이라 부르는 층의 입력(x)을 출력에 연결하여 잔차(Residual)를 학습하는 개념을 도입하였다. 이러한 접근법은 정확도를 획기적으로 올릴 수 있었을 뿐만 아니라 학습 시간도 현저하게 줄었다. Figure 2.43는 본 과제에서 수정 활용한 ResNet 모델 구조를 보여준다. 또한 FC 층에 대해서 층 내의 노드의 개수와 층의 수를 달리하여 네 가지의 다양한 변형 버전의 FC를 테스트하였다.

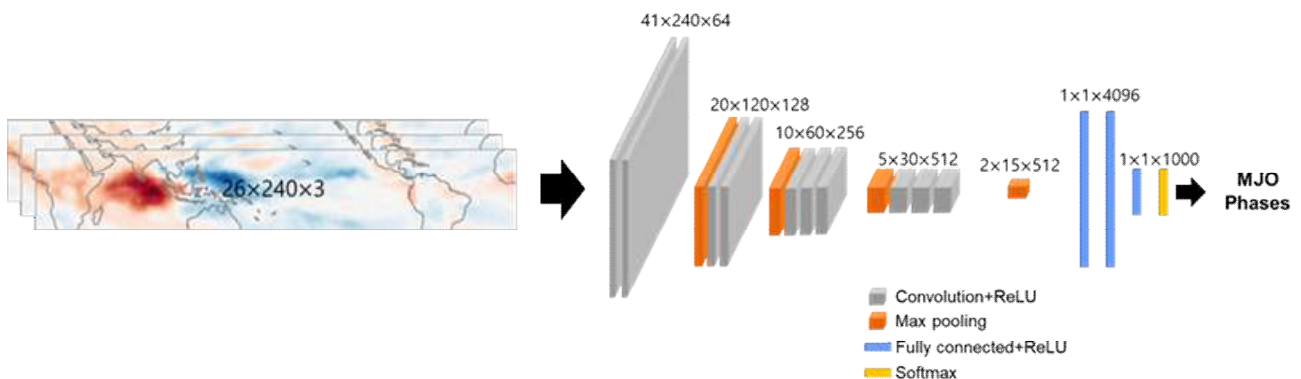


Figure 2.41. Diagram of modified VGG16 model used in this study.

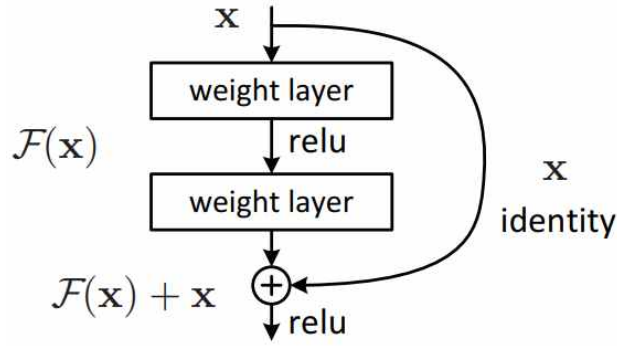


Figure 2.42. Residual learning structure of ResNets model (Source: He et al., 2016)

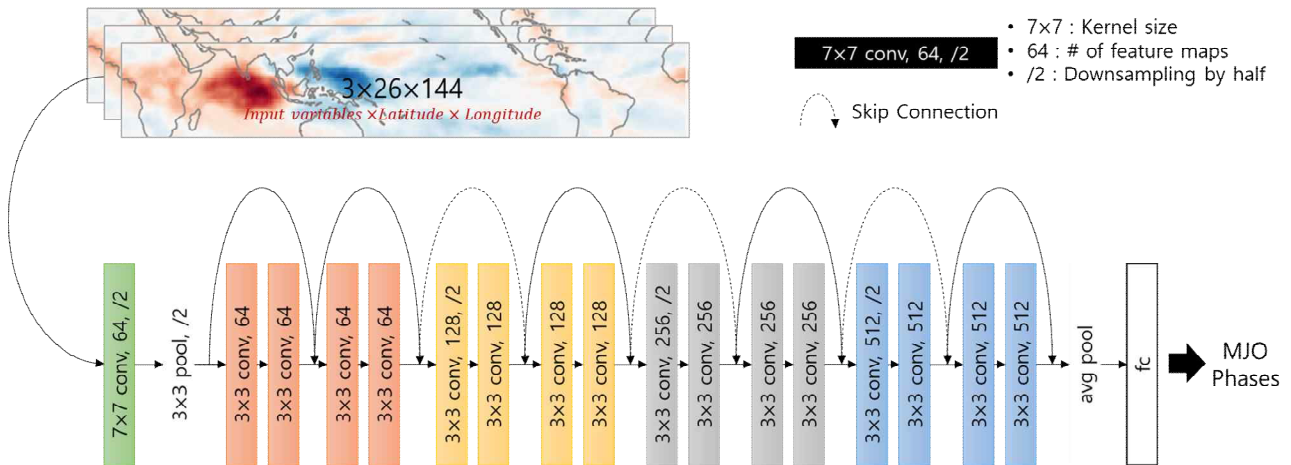


Figure 2.43. Diagram of modified ResNet-18 model and different versions of fully connected layers used in this study.

2.3.2.2. 준지도학습 기반 MJO 위상 분류 모델 구축 및 민감도 테스트 방법

Figure 2.44은 Mean Teacher 준지도학습 방법의 모식도를 보여준다 (Tarvainen and Valpola, 2018). Student 모델은 지도학습, Teacher 모델은 비지도학습 방식으로 학습된다. Student 모델의 출력값은 기존의 라벨 자료와 비교하여 Classification cost (Supervised loss; 지도학습 손실)를 계산한다. Teacher 모델은 지수 이동 평균(EMA; Exponential Moving Average) (Eq. [2.5])을 통해 Student 모델 파라미터의 일정 부분이 넘어가는 방식으로 모델의 파라미터가 업데이트된다(즉, Backpropagation이 수행되지 않음). Teacher 모델은 비라벨링 자료에 대해서 추론(inference)하고 이를 Student 모델과 비교하여 Consistency loss를 계산한다. 전체적인 모델은 Classification cost와 Consistency cost를 함께 고려하여 학습된다.

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t \quad [2.5]$$

Figure 2.45은 Mean Teacher 알고리즘 내에 기존의 자료 증강 방법을 보여준다. 본 과제에서는 MJO의 수평 이동하는 특성에 따라 수평 반전이나 값 자체를 흐리게(blur)하거나 무작위 섭동을 주는 자료 증강 방법들을 테스트하였다. 본 연구에서는 준지도학습 알고리즘의 여러 파라미터에 따른 실험을 수행하였다. 배치크기(Batch size), Labeled batch size, Consistency cost weight, Consistency rampup, EMA decay, 자료증강(Data Augmentation) 등의 파라미터에 대한 Ablation 실험을 수행하였다. Batch size 및 Labeled batch size는 각각 전체 자료에 대한 배치 크기와 라벨링된 일부 자료에 대한 배치 크기를 조정한다. Consistency loss weight는 전체 cost 함수의 consistency cost 항에 주는 가중치이다. Consistency ramp-up은 준지도학습 모델을 처음 시작할 때 초기 몇 에폭은 라벨링된 자료로 학습한다. 이는 초기부터 비라벨링 자료의 정보가 사용되면 오히려 학습을 크게 저하시킬 수 있기 때문이다. 본 과제에서는 Mean Teacher 준지도학습 알고리즘을 적용한 결과에 집중한다. Pseudo-labeling 알고리즘을 테스트했으나 크게 성능 향상을 보이지 않았다. Figure 2.46은 최종적인 준지도학습 기반 MJO 위상 분류 모식도를 보여준다.

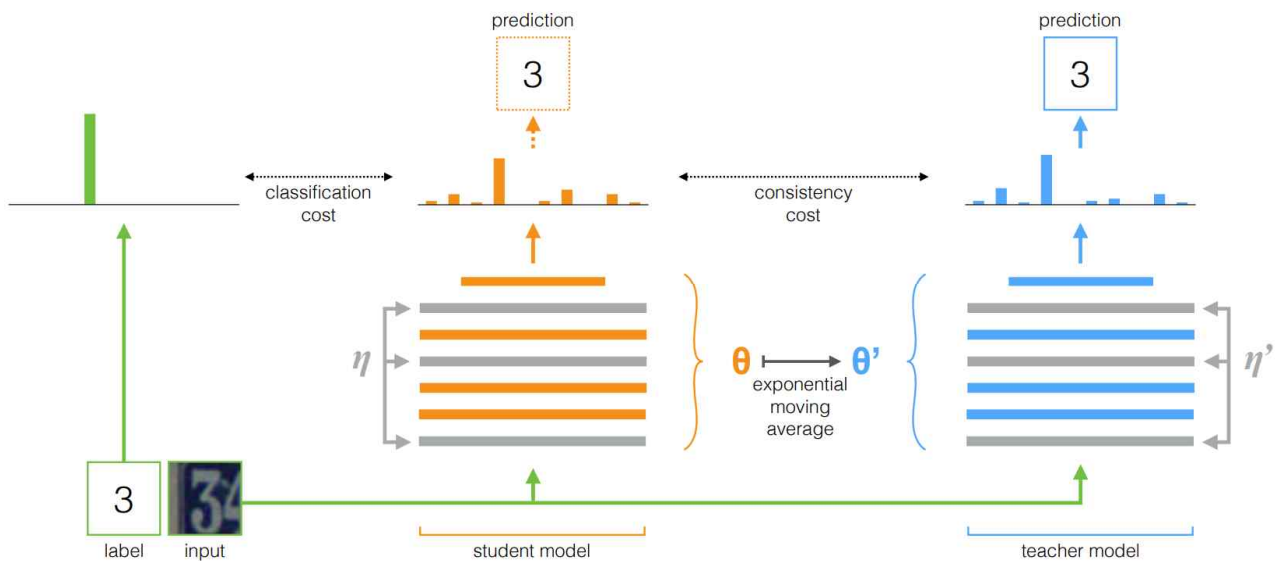


Figure 2.44. Diagram of Mean Teacher algorithm (Source: Tarvainen and Valpola, 2018)

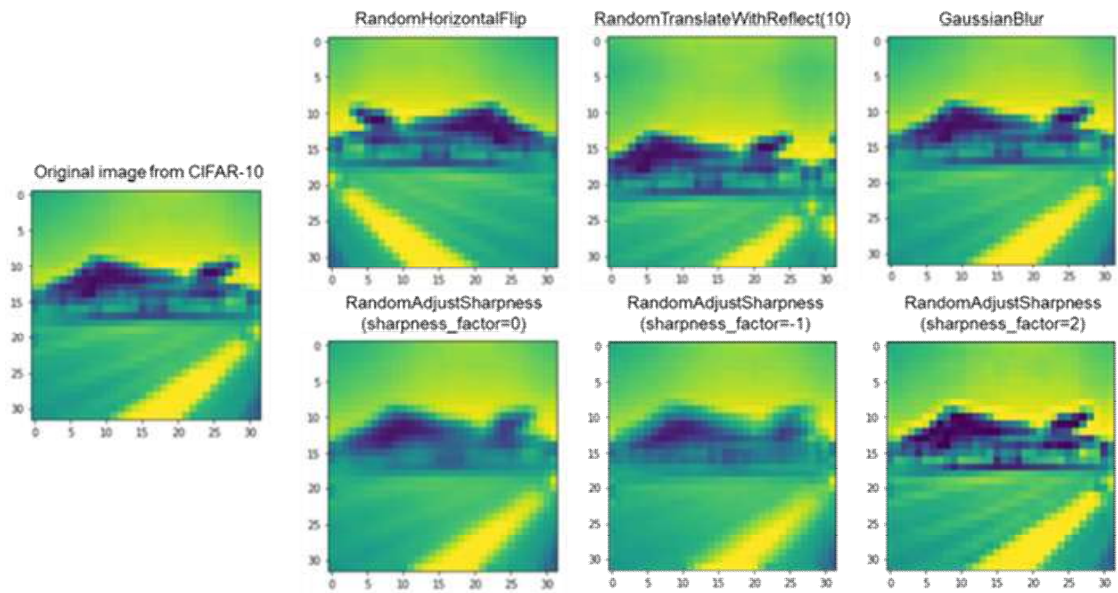


Figure 2.45. Data augmentation methods used in Mean Teacher algorithm.

Table 2.7. List of experiments for sensitivity analysis of SSL-based AI model on model learning strategy, loss weights, and data augmentation methods.

Test	1	2	3	4
Batch Size	32	16	16	32
Labeled Batch Size	16	8	8	16
Consistency loss weight	100	3	3	100
Consistency ramp-up	5	5	5	10
DA method	Random vertical flip	Random vertical flip	No DA	Random vertical flip

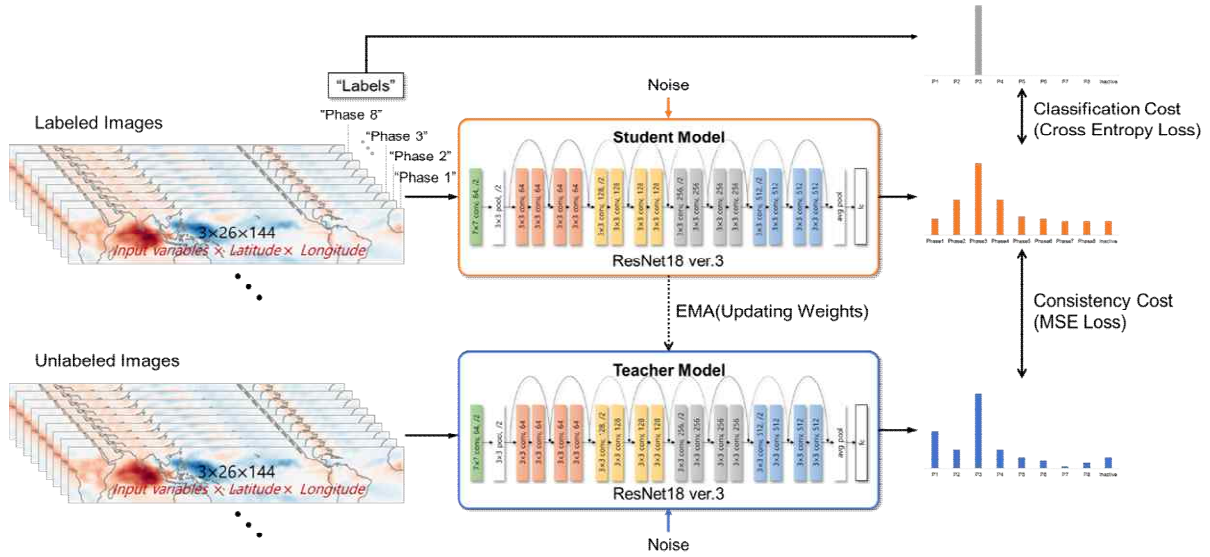


Figure 2.46. Diagram of semi-supervised learning-based ResNet-18 model architecture with the version 3 of the fully connected layer used in this study.

2.3.3. 선행시간별 MJO RMM 지수 확률예측 인공지능 모델 개발 방법

2.3.3.1. 선행시간별 MJO RMM 지수 예측 인공지능 모델 구축 방법

준지도학습 기반 MJO 위상 분류에서 선행시간별 MJO RMM 지수 예측 모델로의 확장을 위해서 0~30일 선행시간별 RMM1과 RMM2 예측 CNN 기반의 ResNet-18 모델이 개발되었다. 기존 ResNet-18 모델 구조에서 과적합을 방지하기 위해 드롭아웃(Dropout) 층(0.2 및 0.5 확률 테스트)을 추가하고 일부 층을 삭제하여 모델 구조를 수정하였다. 손실함수로는 평균 제곱 오차(MSE)를 사용했다. 추가로 선행시간이 지남에 따라 MJO 진폭 에러가 증가하는 기존의 문제점을 개선하는 한 방안으로 MJO 진폭에 대한 항을 손실함수에 포함하여 다음과 같은 맞춤형 (Custom) 손실함수를 설계하였다.

$$L = L_{RMM1} + L_{RMM2} + L_{Amp} \quad [2.6]$$

여기서 L_{RMM1} 은 RMM1에 대한 손실함수, L_{RMM2} 는 RMM2에 대한 손실함수, L_{Amp} 는 MJO Amplitude에 대한 손실함수를 의미하며, 최종 손실함수는 모든 손실함수의 합으로 계산된다.

훈련은 Adam 옵티마이저(Optimizer) 최적화 함수를 사용하여 학습률 $1e-3$, 배치 크기 32, 10번의 검증 정확도 개선이 없을 시 학습을 조기 종료하였으며 최대 에포크 수 500으로 설정하였다. 모델 성능은 아래와 같이 이변량 상관 계수(BCOR)와 이변량 평균 제곱 오차(BMSE)를 사용하여 평가했다 (Kim et al., 2019; Vitart et al., 2017).

$$BCOR(\tau) = \frac{\sum_{t=1}^{t=N} [O_1(t)P_1(t,\tau) + O_2(t)P_2(t,\tau)]}{\sqrt{\sum_{t=1}^{t=N} [O_1^2(t) + O_2^2(t)]} \sqrt{\sum_{t=1}^{t=N} [P_1^2(t) + P_2^2(t)]}} \quad [2.7]$$

$$BMSE(\tau) = \sqrt{\frac{1}{N} \sum_{t=1}^{t=N} [|O_1(t) - P_1(t,\tau)|^2 + |O_2(t) - P_2(t,\tau)|^2]} \quad [2.8]$$

여기서 $O_1(t)$ 와 $O_2(t)$ 는 RMM1과 RMM2의 시간 t 에서의 관측값이다. $P_1(t, \tau)$ 와 $P_2(t, \tau)$ 는 시간 t 와 선행시간 τ 에서 예측된 RMM1과 RMM2이다. N 은 MJO 날수이다. MJO의 진폭(Amplitude)와 위상(Phase)는 아래와 같이 계산된다 (Lim et al., 2018).

$$A_o(t) = \sqrt{O_1^2(t) + O_2^2(t)} \quad [2.9]$$

$$A_p(t,\tau) = \sqrt{P_1^2(t) + P_2^2(t)} \quad [2.10]$$

$$\Phi_o(t) = \tan^{-1} \left[\frac{O_2(t)}{O_1(t)} \right] \quad [2.11]$$

$$\Phi_p(t,\tau) = \tan^{-1} \left[\frac{P_2(t,\tau)}{P_1(t,\tau)} \right] \quad [2.12]$$

여기서 A_o , Φ_o 는 관측 RMM의 진폭과 위상을, A_p , Φ_p 는 예측된 RMM의 진폭과 위상이다. MJO 진폭과 위상의 오차는 다음과 같이 계산되었다 (Lim et al., 2018).

$$\overline{AE} = \frac{1}{N} \sum_{t=1}^{t=N} [A_p(t,\tau) - A_o(t)] \quad [2.13]$$

$$\overline{PE} = \frac{1}{N} \sum_{t=1}^{t=N} [\Phi_p(t,\tau) - \Phi_o(t)] \quad [2.14]$$

2.3.3.2. 선행시간별 MJO RMM 지수 예측 인공지능 모델 해석 방법

입력 변수의 기여도를 분석하기 위해, 본 과제에서는 Occlusion 민감도 맵을 사용했다 (Zeiler and Fergus, 2013). 3×3 크기의 커널을 한 칸씩 이동시키며 입력 이미지를 마스킹하는데 마스킹된 영역은 각 변수의 기준값으로 교체했다. 여기서 기준값은 원 값에서 표준편차를 빼 값으로 정의하였다. 기준값으로 대체되었을 때 원 RMM 값에서 예측된 RMM을 빼는 방식으로 각 영역의 기여도를 계산하였다. 양의 기여도(빨간색 영역)는 해당 영역이 마스크 처리될 때 예측된 RMM 값이 감소하는 것을 의미하며, 음의 기여도(검은색 영역)는 해당 영역이 마스크 처리될 때 예측된 RMM 값이 증가하는 것을 의미한다. 테스트 기간 동안 각 그리드 지점에서

기여도의 합이 최종 출력값으로 제공된다.

2.3.4. MJO 확률예측을 위한 인공지능 모델 구축 방법

선행시간별 MJO 확률 예측을 위한 인공지능모델 개발을 위해서 기 개발된 RMM 지수 단정예측 모델을 기반으로, 확률 추정을 위한 앙상블 기법들을 적용하였다. 이는 ECMWF의 ensemble run과 비슷한 방식으로 초기장(initial conditions)에 섭동을 주거나 모델 역학에 섭동을 주는 방식을 차용하여 (Chen et al., 2024), 본 연구에서도 1) 초기장 섭동 방법과 2) 모델 파라미터 섭동 방법을 테스트하였다. 초기장 섭동 방법으로 먼저 ECMWF에서 제공하는 10개의 앙상블을 가진 ERA5 재분석자료 자료동화 자료를 활용하여 테스트하였다. 초기장 섭동의 다른 방법으로 Perlin noise라고 하는 랜덤 노이즈를 입력 자료에 추가하는 방식으로 50개의 앙상블 자료를 만들어 테스트하였다. Perlin noise는 perlin-numpy Python 패키지를 활용하였다. Figure 2.47는 Perlin noise로 입력 변수들의 초기장에 섭동을 주는 방식으로 생산한 예시를 보여준다. 모델 파라미터 섭동 방법으로는 MC dropout을 적용하였다. 이는 테스트 자료에 대한 추론(inference 혹은 예측) 과정에서도 dropout을 활성화하여 예측값의 분포를 산출하는 방법이다 (Figure 2.48). 초기장 섭동 방법과 마찬가지로 50번 반복 수행하여 50개의 앙상블 결과를 산출하였다.

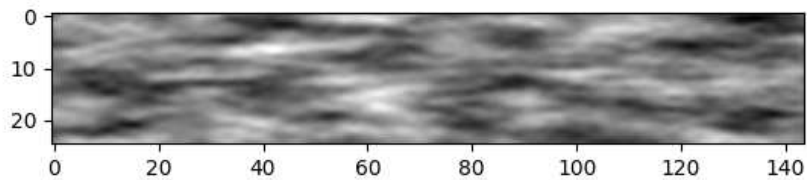


Figure 2.47. An example of perturbed input data using Perlin noise perturbation method.

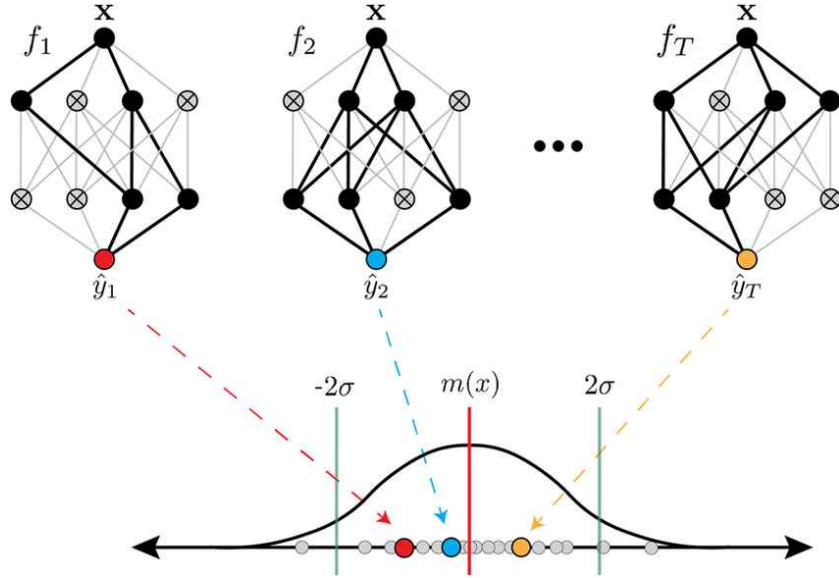


Figure 2.48. Schematic diagram of Monte Carlo dropout (Source: Van Katwyk, et al., 2023)

여러 가지 섭동 방법을 적용하여 예측된 RMM 값들을 기반으로, MJO RMM 앙상블 예측 결과를 MJO 위상으로 변환하여 결과를 분석하였다. 먼저 MJO 위상 확률 예측값에 대한 정량적인 성능 지표 결과를 확인하였다. 이를 위해 RPS(Ranked Probability Score)라고 하는 다중분류 확률예측 평가지표를 활용하였다 (Marshall et al., 2016; Martin et al., 2022). RPS는 아래와 같이 계산된다.

$$RPS = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{M-1} \sum_{m=1}^M \left[\left(\sum_{c=1}^m P_c \right) - \left(\sum_{c=1}^m O_c \right) \right]^2 \right\} \quad [2.15]$$

여기서 N은 예측 일수, M은 MJO 위상 개수, P_c 는 한 MJO 위상에서의 예측 확률, O_c 는 관측 확률, m은 MJO 위상 카테고리(0: inactive MJO, 위상1~8)를 나타낸다. RPS가 0이면 예측이 완전히 정확하고, RPS가 1이면 완전히 틀렸음을 의미한다. 따라서 RPS는 값이 낮을수록 정확한 것으로 해석한다. Table 2.8는 본 연구에서 실험한 3가지의 섭동 실험에 대한 실험 방법과 해당 실험의 섭동(ensemble) 수를 보여주고 있다.

Table 2.8. Perturbation methods used in this study for MJO RMM probabilistic prediction.

Experiment	Perturbation method	Number of perturbations
Noise to initial conditions (초기장 섭동)	ERA5 ensemble	10
	Perlin random noise	50
Perturbing the model parameters (모델 섭동)	Monte Carlo dropout	50

2.4. 3-6주 동아시아 기온확률 예측 인공지능 모델 개발

2.4.1. 연구자료

2.4.1.1. ALO (Atmosphere-land surface-ocean) 자료

본 연구에서는 딥러닝 기법을 이용하여 T2M (2 meter temperature) 기온자료를 예측하는 모델을 개발하고 개발한 모델을 토대로 3~6주 기온확률 예측 인공지능 모델을 개발하고자 한다.

ALO (atmosphere-land surface-ocean) 시스템에서 S2S 시간규모와 예측성 간 관계를 보면 대기는 짧은 선행 시간에서는 대기의 초기 상태를 아는 것이 중요하며, 2주에서 4주의 선행 시간에서는 지면의 정보, 30일을 넘어가는 예측에서는 해양에 대한 정보가 중요하다(Mariotti et al. (2018), Figure 2.49).

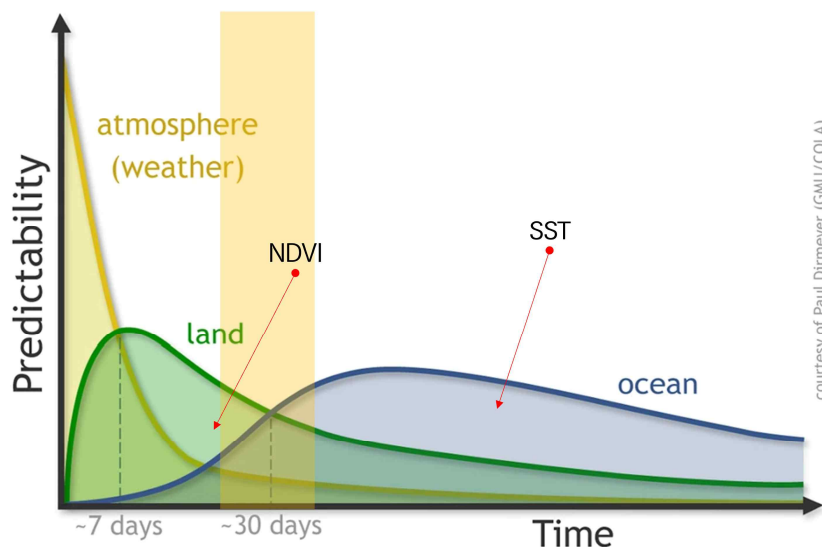


Figure 2.49. A schematic of the role of ALO (Atmosphere-Land-Ocean) systems as a source of S2S predictability (Adapted from Mariotti et al. (2018)).

본 연구에서는 T2M 변수를 포함하여 지면 정보로 정규 식생 지수(Normalized difference vegetation index, NDVI), 해양 정보는 SST 변수를 이용하였다.

T2M와 SST 변수는 유럽중기예보센터(이하 ECMWF, European Centre for Medium-Range Weather Forecasts)에서 제공하는 ERA-5 재분석자료를 이용하였다. ERA-5 재분석자료는 $0.25^{\circ} \times 0.25^{\circ}$ 해상도의 격자 자료를 제공하며, 기존에 제공하던 ERA-Interim reanalysis 자료를 개선한 자료이다. 6시간 단위의 T2M (2 meter temperature), SST (Sea Surface

Temperature) 변수자료를 수집하고 일 단위자료로 변환하여 사용하였다.

NDVI 자료는 Parente와 Hengl (2020)이 MODIS와 AVHRR 자료를 이용하여 공간해상도 5km로 구축한 월 단위 전지구 NDVI를 이용하였다(Table 2.1).

Table 2.9. Summary of ALO (Atmosphere–Land–Ocean) data used in this study.

Variable	Long name	Unit	Period	Frequency
T2M	2 meter temperature	K	1965 to 2022	Daily
SST	Sea surface temperature	K	1965 to 2022	Daily
NDVI	Normalized Difference Vegetation Index		1982 to 2019	Monthly

2.4.1.2. 딥러닝 학습자료

본 연구에서는 ALO 자료를 토대로 딥러닝 학습자료를 구축하였다(Table 2.2). 학습자료는 X 입력자료 구성은 T2M, T2m climatology, SST climatology, NDVI monthly 자료, Y 입력자료 구성은 T2M 자료를 이용하였다. X 입력자료는 학습자료 기간의 자료를 대상으로 순서를 섞은 다음에 학습자료(train_ratio)와 검증자료 비율(validation ratio)로 구축하였다. 학습자료는 TDSV31-1_SHUFFLE, TDSV32_SHUFFLE, TDSV61_SHUFFLE, TDSV61-2_SHUFFLE, TDSV61-3_SHUFFLE로 총 5개를 구축하였다.

TDSV31-1_SHUFFLE 학습자료는 X 입력자료로 T2M, T2M Climatology 15년, Y 입력자료는 입력기간 14일, 예측기간 1일로 하는 1986년에서 2015년 기간의 자료를 7 대 3 비율로 학습자료와 검증자료를 구성하고 2016년에서 2022년 기간의 자료로 테스트자료를 구축하였다.

TDSV32_SHUFFLE 학습자료는 X 입력자료로 T2M, T2M Climatology 15년, Y 입력자료는 입력기간 14일, 예측기간 1일로 하는 1965년에서 2015년 기간의 자료를 7 대 3 비율로 학습자료와 검증자료를 구성하고 2016년에서 2022년 기간의 자료로 테스트자료를 구축하였다.

TDSV61_SHUFFLE 학습자료는 X 입력자료로 T2M, T2M Climatology 15년, Y 입력자료는 입력기간 14일, 예측기간 1일로 하는 1982년에서 2015년 기간의 자료를 6 대 4 비율로 학습자료와 검증자료를 구성하고 2016년에서 2019년 기간의 자료로 테스트자료를 구축하였다.

TDSV61-2_SHUFFLE 학습자료는 X 입력자료로 T2M, T2M Climatology 15년, SST Climatology 15년, 월단위 NDVI, Y 입력자료는 입력기간 7일, 예측기간 1일로 하는 1982년에서 2015년 기간의 자료를 6 대 4 비율로 학습자료와 검증자료를 구성하고 2016년에서 2019년 기간의 자료로 테스트자료를 구축하였다.

TDSV61-3_SHUFFLE 학습자료는 X 입력자료로 T2M, T2M Climatology 15년, SST Climatology 15년, 월단위 NDVI, Y 입력자료는 입력기간 14일, 예측기간 1일로 하는 1982년에

서 2015년 기간의 자료를 6 대 4 비율로 학습자료와 검증자료를 구성하고 2016년에서 2019년 기간의 자료로 테스트자료를 구축하였다. 데이터셋 TDSV32_SHUFFLE의 학습자료의 비율을 예시로 들면 Figure 2.50와 같다.

Table 2.10. Summary of Dataset.

Dataset name	input_x1_var train_ratio test_period_start	input_x2_var validation_ratio test_period_end	input_x3_var test_ratio input_time_step	input_x4_var train_period_start output_time_step	input_y_var train_period_end
TDSV31-1_SHUFFLE	T2M 0.7 2016	T2M Climatology 15y 0.3 2022	1.0 14	1986 1	T2M 2015
TDSV32_SHUFFLE	T2M 0.7 2016	T2M Climatology 15y 0.3 2022	1.0 14	1965 1	T2M 2015
TDSV61_SHUFFLE	T2M 0.6 2016	T2M Climatology 15y 0.4 2019	1.0 7	1982 1	T2M 2015
TDSV61-2_SHUFFLE	T2M 0.6 2016	T2M Climatology 15y 0.4 2019	SST Climatology 15y 1.0 7	NDVI Monthly 1982 1	T2M 2015
TDSV61-3_SHUFFLE	T2M 0.6 2016	T2M Climatology 15y 0.4 2019	SST Climatology 15y 1.0 14	NDVI Monthly 1982 1	T2M 2015

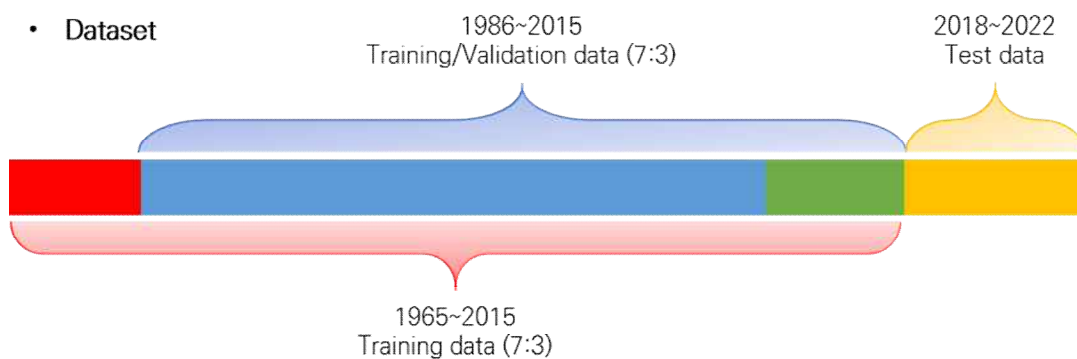


Figure 2.50. Ratio of training, validation, test data of TDSV32_SHUFFLE dataset.

2.4.2. 3-6주 동아시아 기온확률 예측 인공지능 모델 개발

2.4.2.1. 기준 1개월 기온 예측 모델 구축

1개월 기온 예측 모델은 Azad et al.(2019)가 제안한 U-NET 기반의 모델을 이용하여 개발하였다. U-NET은 축소 경로(contracting path), 병목 경로(bottleneck), 확장 경로(expansive path)로 구성된다. 축소 경로는 입력 자료를 다운 샘플링하면서 컨텍스트(context)를 포착하며, 확장 경로는 채널수를 늘리면서 원래 크기로 업샘플링(upsampling)한다. 스킵 연결(skip connection)은 축소 경로의 특징(feature)를 업샘플링 과정에 포함하여 축소 경로 과정에서 소실된 정보를 보완하는 역할을 한다. U-NET은 세그멘테이션을 위해 고안된 모델로 자료의 특징 중 약한 시그널이 사라지는 현상이 있다. 이를 보완하기 위해 prediction layer에 자료를 추가(input 2)하여 이를 보완하였다. 추가 입력자료는 prediction layer 전에 연결된다. 축소 경로와 확장 경로의 기본 합성곱 레이어 블록은 conv2d + conv2d + attention block으로 구성되어 있다. 일반적으로 사용하는 conv2d + conv2d + BN (batch normalization) + activation 구조에서 BN와 activation을 빼고 attention block으로 수정한 구조이다. 어텐션 블록은 Hu et al.(2018)이 제안한 squeeze-and-excitation networks (이하 SEblock)와 Woo et al. (2018)이 제안한 CBAM (Convolutional Block Attention Module)을 이용하였다. SEblock은 채널 어텐션 메커니즘(channel attention mechanism)을 바탕으로 설계되었으며, 채널 단위로 가중치를 부여하고 의미 있는 채널의 정보를 다음 네트워크로 넘길 수 있는 장점이 있다(Figure 2.52). CBAM은 채널 어텐션 메커니즘과 공간 어텐션 메커니즘을 결합한 구조로 되어 있다(Figure 2.53). 1개월 기온 예측 모델은 T2M 변수와 T2M 변수로 생산한 기후평년값(climatology)을 입력 자료로 이용한다.

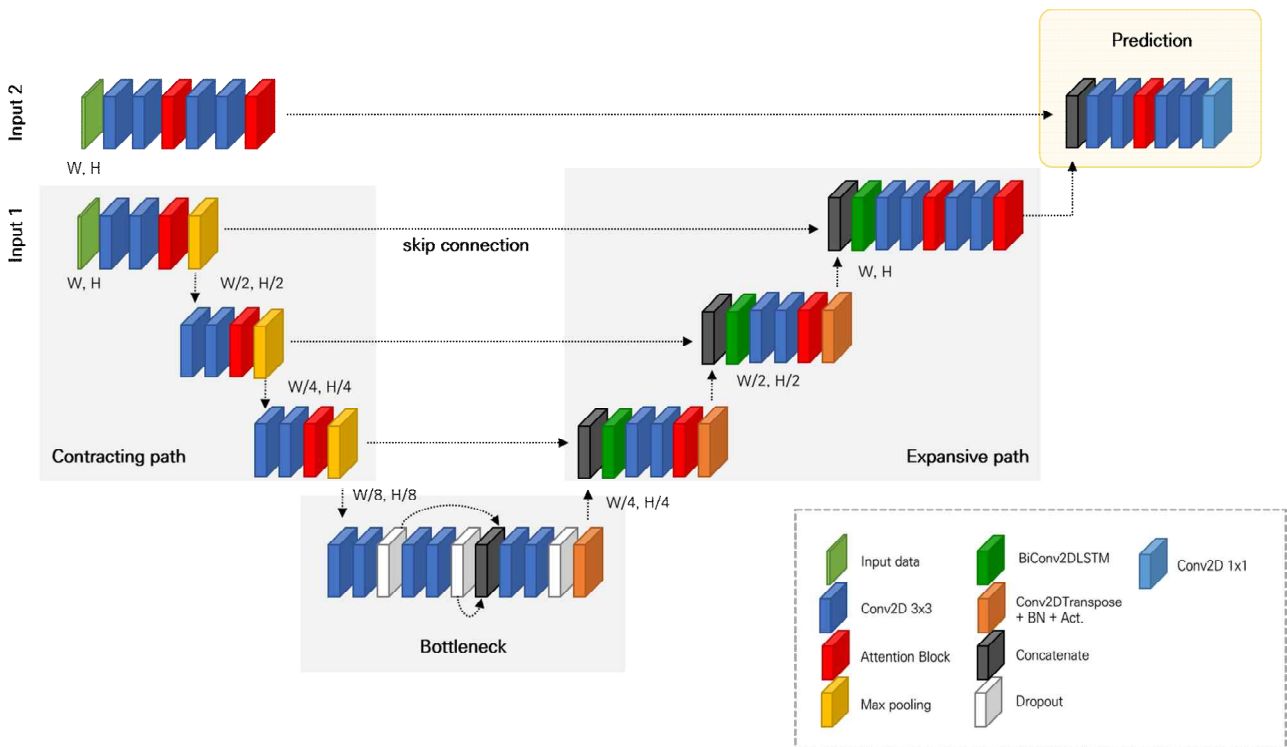


Figure 2.51. A Schematic Diagram of the 1-Month Temperature Prediction Model.

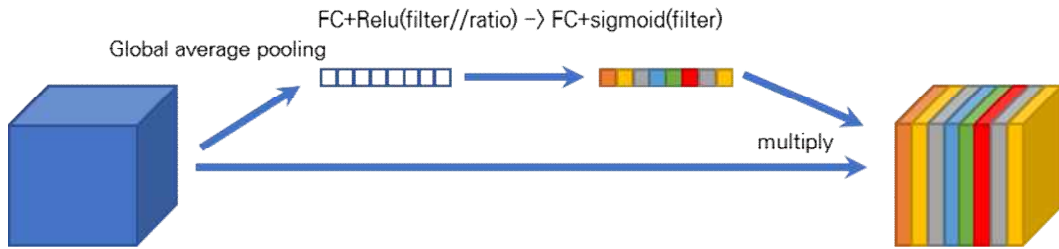


Figure 2.52. The overview of SE block (adapted from Hu et al. (2020)).

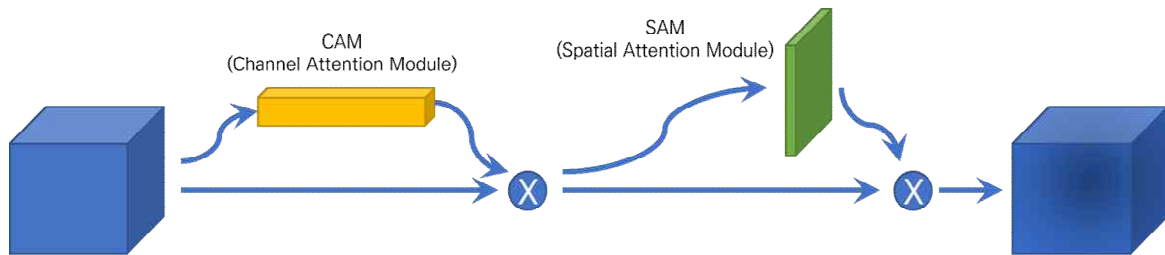


Figure 2.53. The overview of CBAM (adapted from Woo et al. (2018)).

2.4.2.2. 입력 변수 확장을 통한 기온 예측 모델 구축

T2M 변수만을 이용하는 1개월 기온확률 예측 모델(Figure 2.51)을 바탕으로 ALO (Atmosphere-Land-Ocean) 시스템을 고려하여 T2M 변수 외에 지면자료로 NDVI, 해양정보로 SST 변수를 추가하여 딥러닝 모델을 구축하였다.

SST 변수는 T2M 변수와 같이 ERA-5 재분석자료를 이용하였다. T2M 변수는 T2M 자료와 T2M 자료를 이용하여 구축한 15년 단위의 기후평년자료를 이용하였다. SST 자료는 SST 변수 자료가 아닌 T2M 기후평년자료와 같이 15년 단위의 기후평년자료를 구축하여 이용하였다. NDVI 자료(Parente와 Hengl, 2020)는 ERA-5 재분석자료와 공간해상도와 격자의 중심좌표가 다르기 때문에 ERA-5 재분석자료(Figure 2.54)의 격자정보를 중심으로 공간해상도와 중심좌표를 보정하는 전처리를 수행하였다.

NDVI 자료는 공간해상도 5km로 구성되어 있다(Figure 2.55(a)). ERA-5 재분석자료와 같은 공간해상도를 맞추기 위해 리샘플링(resampling) 방법을 이용하였다. NDVI 자료가 ERA-5 재분석자료보다 고해상도(high resolution)이고 해양자료(ocean)는 없어서 리샘플링 방법에 따라서는 섬지역과 해양과 맞닿은 지역은 제대로 처리되지 않는다(Figure 2.55(b)). 본 연구에서는 bilinear, average 등 리샘플링 방법의 실험을 통해 average 방법을 리샘플링 방법으로 선택하였다. 선정된 리샘플링 방법으로 자료 전처리를 수행하고 격자의 중심좌표를 ERA-5 재분석 자료와 좌표를 맞추는 리그리딩(regridding)을 수행하였다(Figure 2.55(d)).

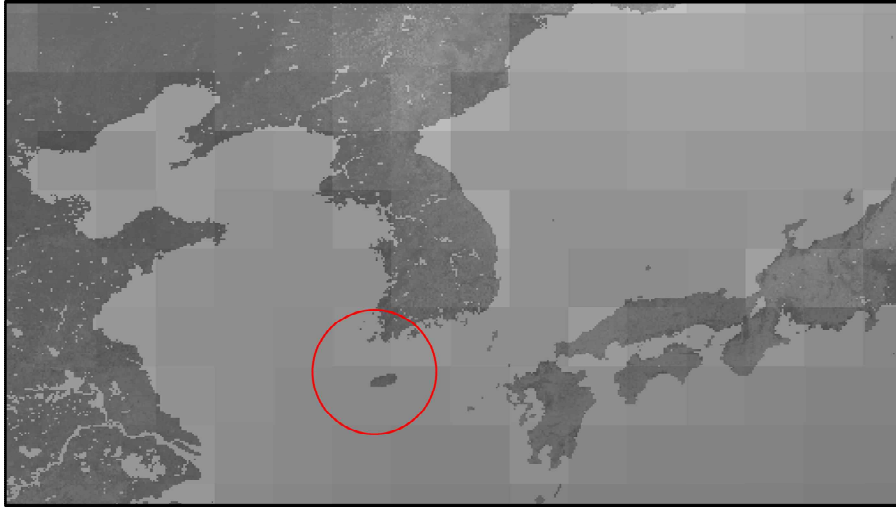


Figure 2.54. A overlay map of the Korean Peninsula topography and gridded data of ERA-5 reanalysis data: Jeju island (red circle).

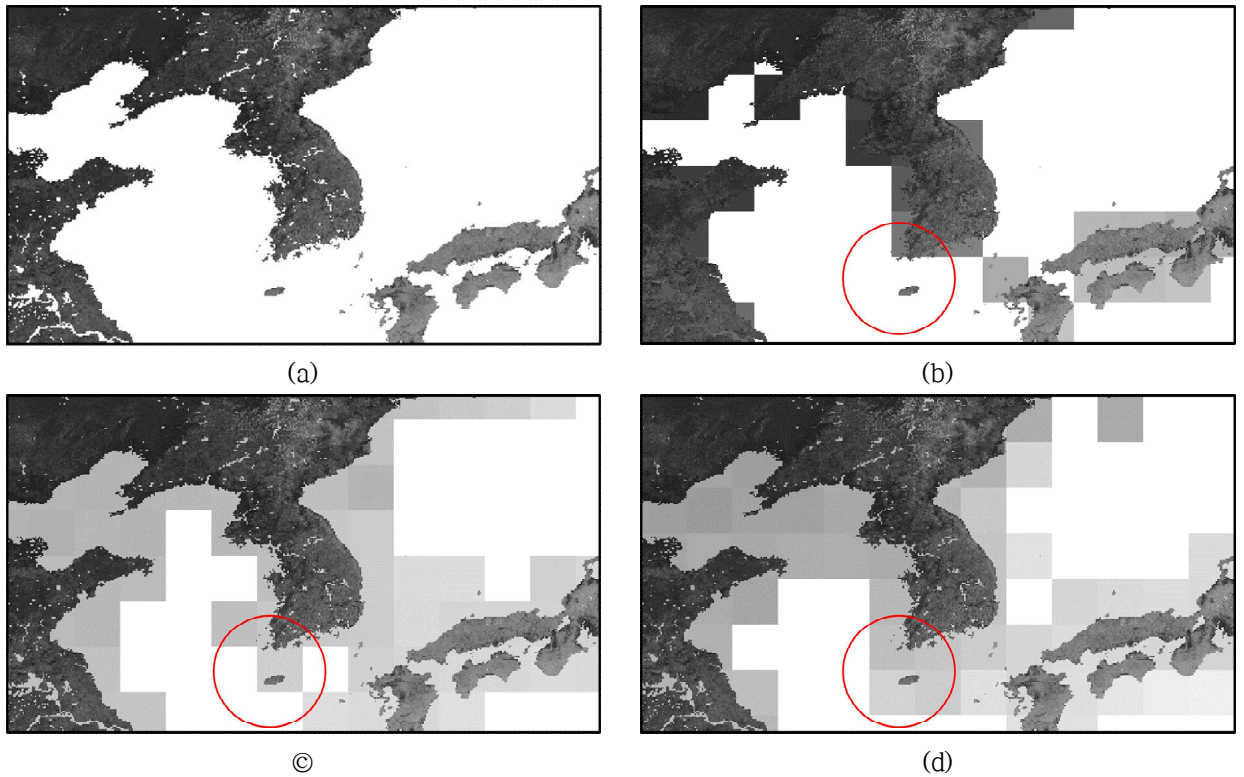


Figure 2.55. Results of NDVI Data Preprocessing: (a) Original NDVI Data, (b) Bilinear Resampling, (c) Average Resampling, (d) Regridding (Including Average Sampling).

SST 기후평년값, NDVI 자료 추가로 1개월 기온 예측 모델(Figure 2.51)의 구조를 개선하였다. 모델의 Input 2의 구간을 Figure 2.56과 2.57과 같이 2가지 경우로 구축하였다. 첫 번째 방법은 기존 Input 2와 같이 개별 CCA (Conv2D, Conv2D, Attention Block) 블록을 수행 후 Input

2, 3, 4와 연결 레이어(concatenate layer)를 이용하여 병합 후 prediction 파트의 입력자료로 이용한다. 두 번째 방법은 Input 2, 3, 4를 먼저 연결 레이어를 이용하여 병합 후 CCA 블록을 수행하고 prediction 파트의 입력 자료로 이용한다.

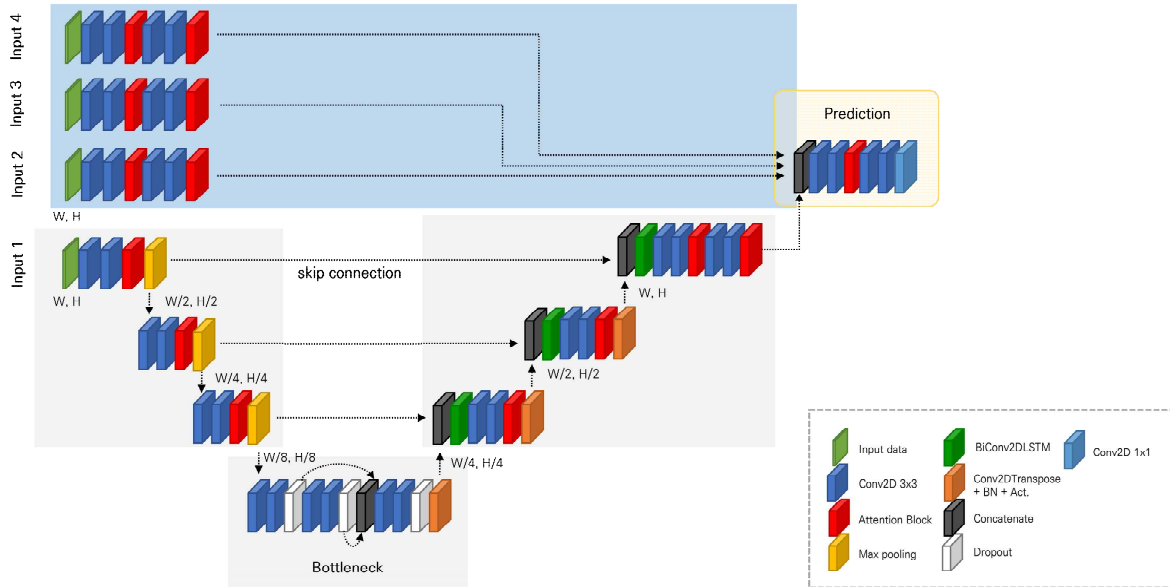


Figure 2.56. A Schematic Diagram of the 1-Month Temperature Prediction Model Redesigned with Two New Input Data Sets.

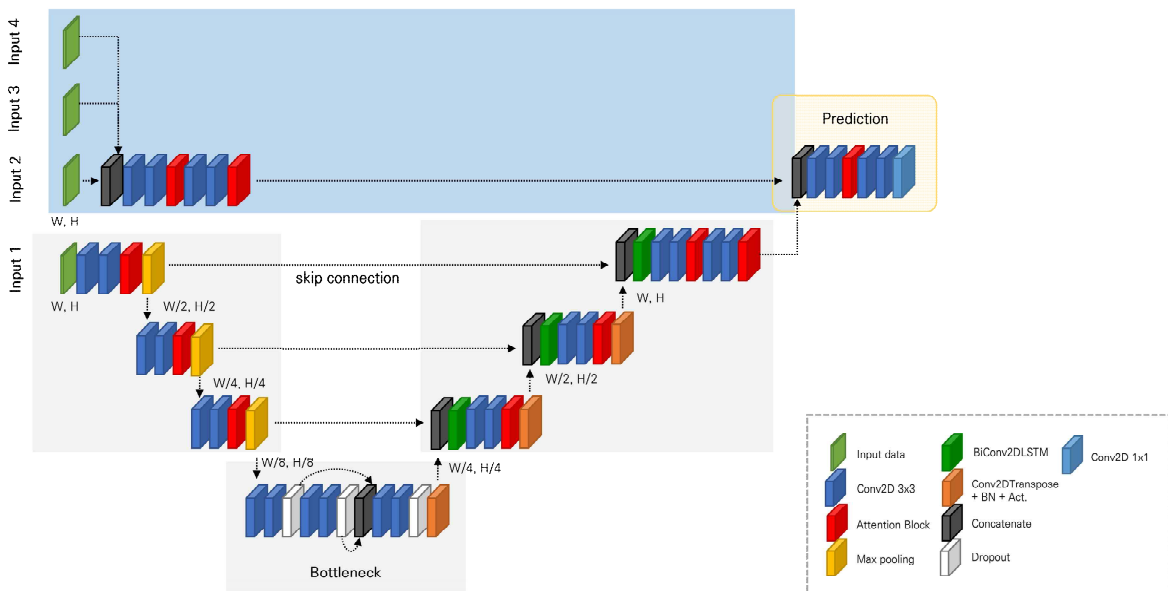


Figure 2.57. A Schematic Diagram of the 1-Month Temperature Prediction Model Redesigned with Two New Input Data Sets.

2.4.2.3. 최신 딥러닝 아키텍처를 이용한 기온 예측 모델 구축

기본 1개월 기온 예측 모델의 아키텍처 중 Kim 등(2024)이 제안한 RDNet(DenseNets Reloaded)의 Stage-N block을 이용하여 딥러닝 모델을 구축하였다(Figure 2.11). 기존 CCA (conv2d + conv2d + attention block) 블록을 RDNet의 Stage-N block을 응용하여 conv2d, LayerNormalization(LN), conv2d, GELU, conv2d, attention block으로 CLCGCA 블록구조로 변경하였다. CLCGCA 블록을 이용하여 변수 확장을 고려한 모델을 토대로 Figure 2.58과 2.59와 같이 수정하였다.

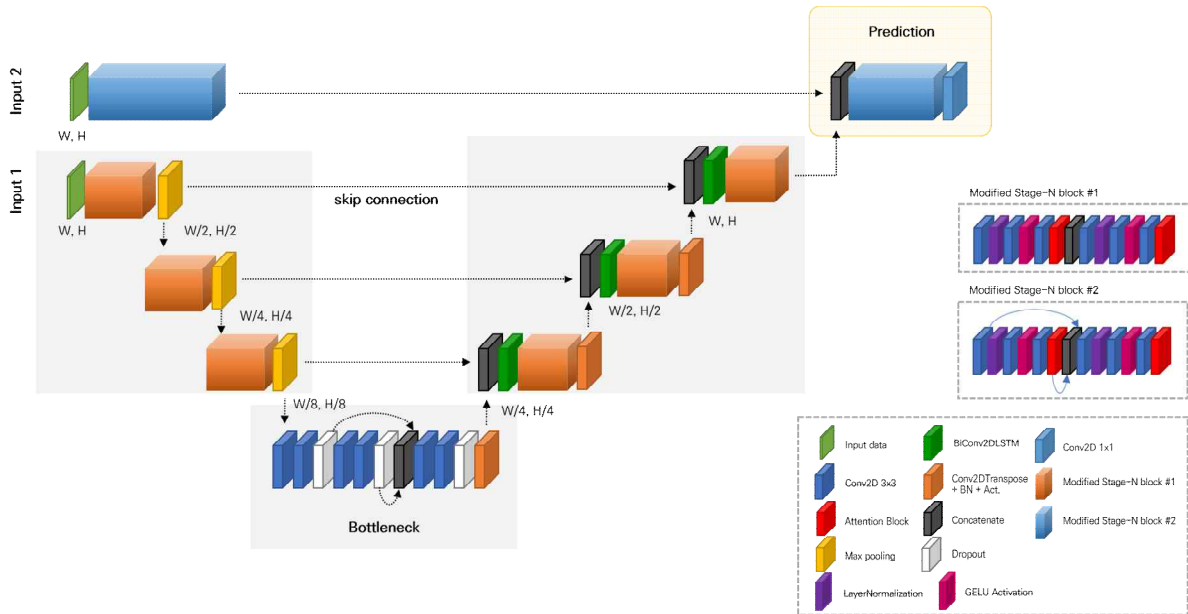


Figure 2.58. A schematic diagram of the 1-month temperature prediction model redesigned with the stage-N block of RDNet.

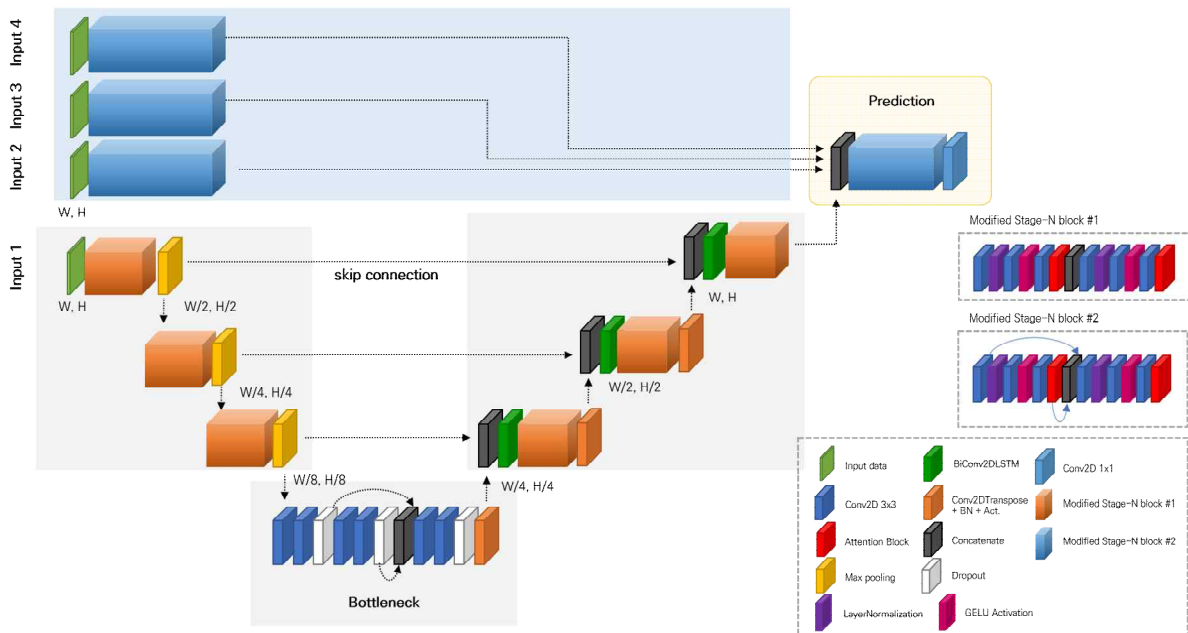


Figure 2.59. A schematic diagram of the 1-month temperature prediction model redesigned with the stage-N block of RDNet and two new input data sets.

2.4.2.4. 앙상블 학습 기법을 이용한 기온 예측 모델 구축

스태킹 앙상블 기법을 이용하여 1개월 기온 예측 모델로 기온 예측 모델을 구축하는 앙상블 학습 모델을 구축하였다(Figure 2.12). 스태킹 앙상블 기법은 개별 학습된 모델의 결과를 연결(concatenate)하여 prediction layer를 통해 최종 예측하는 방법이다. 앙상블 학습 모델은 prediction layer를 2가지 종류로 구축하였다. 첫 번째는 prediction layer를 단순히 연결하고 커널 크기가 1×1인 합성곱 레이어를 이용하여 최종 예측한다. 두 번째는 prediction layer를 기존 1개월 기온 예측 모델에 사용된 합성곱 레이어 2개와 어텐션 레이어로 구성하고 최종 예측한다. 또한, 앙상블 학습 모델은 모델 학습 과정에서 개별 모델의 bias와 weight를 고정하거나 업데이트하는 방법과 prediction layer는 bias와 weight가 업데이트되는 방법으로 구성되었다.

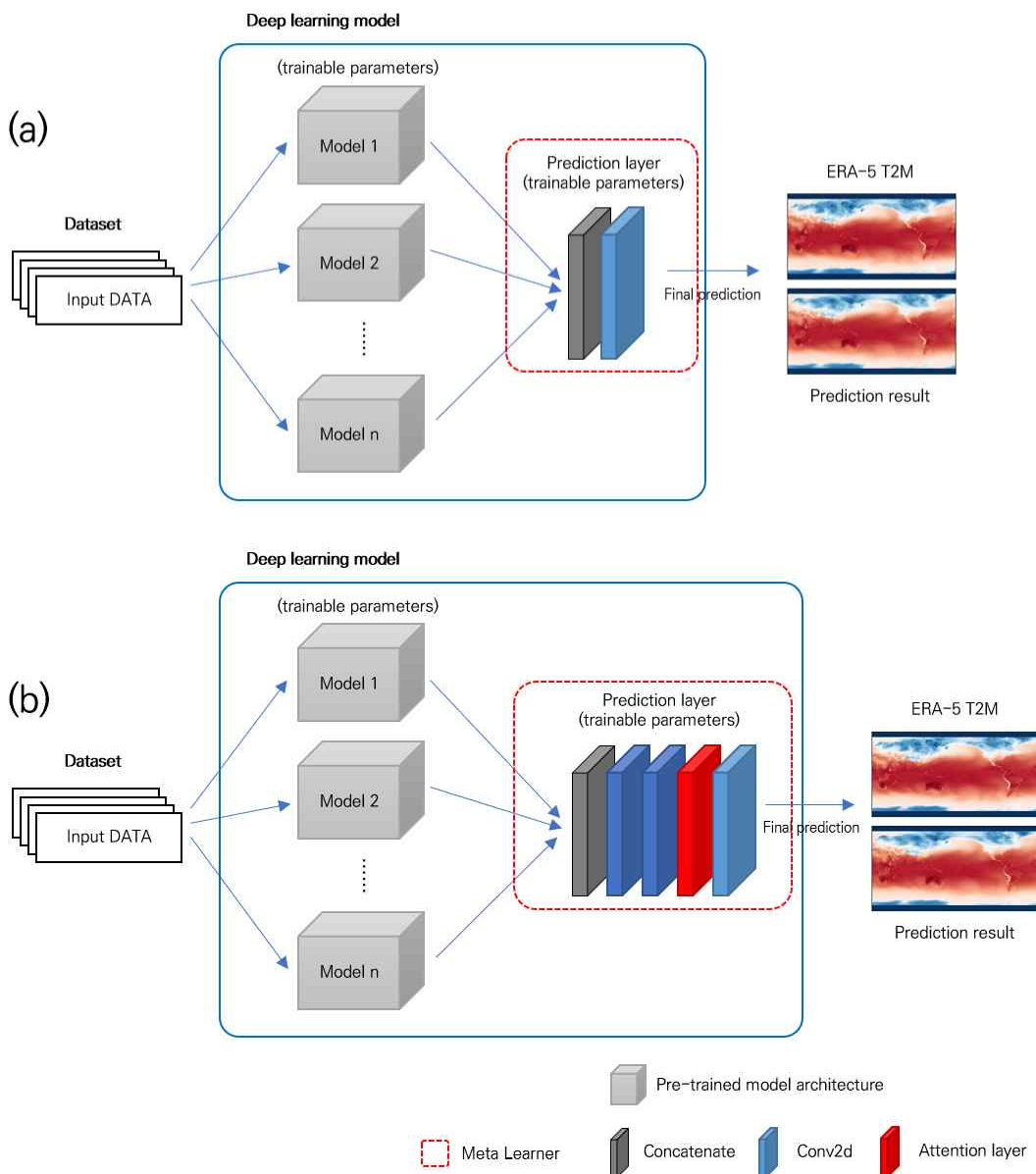


Figure 2.60. Schematic diagrams of ensemble learning models: (a) simple prediction layer (b) CCA prediction layer.

2.4.2.5. N 스텝 예측 방법을 이용한 기온 예측 모델 구축

N 스텝 예측 방법을 이용한 기온 예측 모델은 3-6주 예측 결과에서 예측 주마다 예측성이 우수한 모델을 N개의 딥러닝 모델을 순차적으로 연결하여 기온 예측을 하는 모델이다. Figure 2.61은 3-5주 예측성이 높은 A 모델과 6주 예측성이 높은 B 모델을 이용하여 3-6주 예측하는 과정 모식도이다.

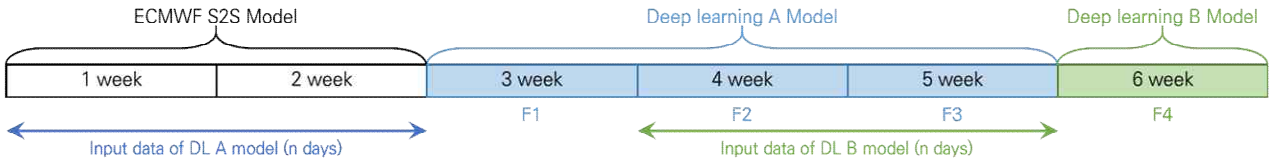


Figure 2.61. Example of a schematic diagram of a re-learning model.

2.4.2.6. 재학습 방법을 이용한 기온 예측 모델 구축

재학습 방법은 딥러닝 모델 학습 과정에서 일정 수준 이상으로 손실값이 수렴하면 학습을 중단하는 조기종료(earlystopping) 방법에 따라 학습이 중단된 모델을 학습률을 전이 학습(transfer learning)과 같이 이전 학습률보다 낮게 설정하고 모델 학습을 이어가는 것이다 (Figure 2.62). 딥러닝 모델은 일정 이상 손실값이 개선되지 않으면 학습률을 낮추는 학습률 스케줄러(learning rate scheduler)를 이용한다. 학습률 스케줄러가 딥러닝 모델 학습에 과정에서 학습률을 적절하게 조정하는 중요한 역할을 수행하지만, 학습률의 변화가 데이터셋 등에 따라 적절하게 조정되지 않고 조기종료 방법에 의해 학습이 종료될 수도 있다. 그러므로 재학습 방법을 통해 손실값의 최소화를 통해 기온 예측 모델의 예측성을 높일 수 있는지를 보고자 한다.

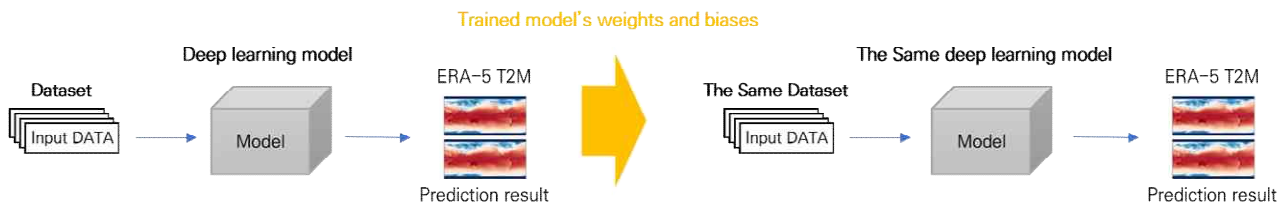


Figure 2.62. A schematic diagram of re-learning model.

2.4.3. 피쳐붕괴 분석을 통한 딥러닝 모델 분석

합성곱 신경망(Convolutional neural network, CNN) 구조의 딥러닝 모델은 필터(filter)를 통과하는 과정에서 뉴런(neuron)이 죽거나(dead) 포화(saturation)하는 현상에 의해 피쳐붕괴(feature collapse) 현상이 발생한다(Figure 2.63). 모델의 예측성을 높이기 위해서는 입력 자료가

convolutional layer 등을 통과 후 죽거나 포화되는 뉴런이 적을수록 모델의 예측성능을 높일 수 있다. 즉, 피쳐붕괴 현상이 발생하는 layer가 많을수록 딥러닝 모델이 학습과정에서 유용한 정보를 생성하는데 방해가 되며 결과적으로는 예측성능이 떨어지는 요인이 된다.

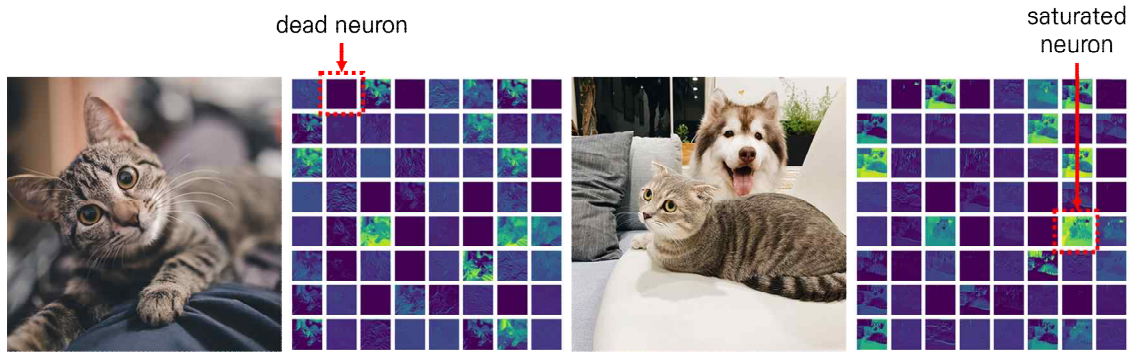


Figure 2.63. Visualization of activation map (input images obtained from <https://github.com/ismailuddin/gradcam-tensorflow-2>).

피쳐붕괴 현상을 Activation layer의 시각화를 통해 확인하는 것은 비효율적인 방법으로, 본 연구에서는 Woo et al.(2023)이 CNN 레이어의 성능 분석에 사용한 feature pair-wise cosine distance (Eq. [2.16])를 이용하여 딥러닝 모델의 주요 구간의 피쳐붕괴 현상을 분석하였다. cosine distance는 자연어처리 분야에서 문서 분류에도 사용된다.

$$\frac{1}{C^2} \sum_i^C \sum_j^C \frac{1 - \cos(X_{F_i}, Y_{F_j})}{2} \quad [2.16]$$

여기서, C 는 채널(channel) 수, F_i 와 F_j 는 convolutional layer, activation layer 등을 통과한 i 와 j 번째 피쳐(feature), i 와 j 는 전체 채널의 순서를 의미한다. 코사인 거리가 길수록 붕괴현상이 적고 짧을수록 붕괴현상이 심한 것을 알 수 있다.

2.4.4. 기온확률 예측 평가 방법

기온확률 예측은 남한지역 5개 지점의 평균값을 이용하여 3분위 예측을 수행하고 HSS (Heidke Skill Score) 방법을 이용하여 평가하였다(Figure 2.64).

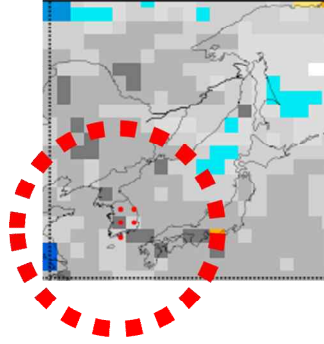


Figure 2.64. An illustrating of 5 points in South Korea.

확률예측은 딥러닝 모델을 rolling prediction 방법으로 ECMWF S2S 모델의 1-2주 자료를 입력자료로 이용하여 3-6주의 일단위 기온을 예측한다. ECMWF S2S 모델의 forecast와 hindcast 자료를 대상으로 수행하여 딥러닝 모델 기반의 forecast와 hindcast 자료가 생산된다. hindcast 자료를 이용하여 ECMWF와 딥러닝 예측 자료별로 3분위 예측의 AN (above normal), BN (below normal)의 임계값을 계산한다. 임계값은 평균과 분산을 이용하여 CDF (cumulative normal distribution) 기법으로 계산한다.

계산된 hindcast AN, NN, BN 값 범위는 forecast 자료를 이용하여 예측(forecast)와 관측(observation) 매트릭스(Figure 2.65)로 분류하고 Eq. [2.17]을 이용하여 HSS를 계산한다. HSS 값은 $-\infty < HSS \leq 1$ 범위이며, 0이면 랜덤모형과 같이 스킬이 없음을 의미한다

		Observed category			
		$O_1(AN)$	$O_2(NN)$	$O_3(BN)$	
Forecast category	$F_1(AN)$	$n(F_1, O_1)$	$n(F_1, O_2)$	$n(F_1, O_3)$	$N(F_1)$
	$F_2(NN)$	$n(F_2, O_1)$	$n(F_2, O_2)$	$n(F_2, O_3)$	$N(F_2)$
	$F_3(BN)$	$n(F_3, O_1)$	$n(F_3, O_2)$	$n(F_3, O_3)$	$N(F_3)$
		$N(O_1)$	$N(O_2)$	$N(O_3)$	N

Figure 2.65. Matrix of Forecast and Observed category.

$$HSS = \frac{\left(\frac{\sum_{i=1}^3 n(F_i, O_i)}{N} - \frac{\sum_{i=1}^3 N(F_i)N(O_i)}{N^2} \right)}{\left(1 - \frac{\sum_{i=1}^3 N(F_i)N(O_i)}{N^2} \right)} \quad [2.17]$$

모델의 평가 기간은 2018년 1월부터 2019년 12월, 2018년 1월부터 2022년 12월로 기간을 나누어 선정하였다. 지면 자료인 NDVI자료가 1982년부터 2019년까지만 구축되었기 때문에 지면 자료가 3~6주 기온 확률 예측에 미치는 영향을 분석하기 평가 기간을 나누었다. 2022년 12월까지 기간을 확장한 평가는 NDVI 자료를 2020년부터의 NDVI 자료는 2019년 자료에서 해당 월의 자료를 이용하였다.

엘니뇨(El Niño)와 라니냐(La Niña) 발생 시기의 기온 확률 예측성을 분석하였다. 엘니뇨는 2018년 9월~2019년 6월과 2019년 11월~2020년 3월, 라니냐는 2020년 8월~2021년 5월, 2021년 8월~2023년 1월에 발생한 기간의 자료만을 이용하여 시기별로 분석하였다. 2018~2019년 기간의 분석은 자료 부족으로 엘니뇨와 라니냐 분석을 제외하였다.

2.5. 자료강화기법을 이용한 인공지능모델의 월별 기후예측 개선

2.5.1. 자료 수집 및 전처리

모델의 입력 변수로 강수, 해수면온도, 하층기온, 지위고도, 해면기압을 이용하였다. 관측 및 재분석 자료 중 일별 강수는 CPC Global 육상 강수, 월별 강수는 CMAP 자료를 이용하였으며(ftp.cdc.noaa.gov/Datasets/cpc_global_precip 및 <ftp.cpc.ncep.noaa.gov/precip/cmap/>), 해수면온도는 NOAA OISST 고해상도 자료(<ftp.cdc.noaa.gov/Datasets/noaa.oisst.v2.highres>), 하층기온(850hPa), 지위고도(500hPa), 해면기압은 NCEP/DOE Reanalysis II (NCEP2) 자료를 활용하였다(<ftp.cdc.noaa.gov/Datasets/ncep.reanalysis2.derived/> 및 <ncep.reanalysis2.dailyavgs/>). 관측 자료가 부재한 기간에 대해서는 역학모델의 예측자료를 활용하였는데, 동일한 다섯 변수에 대해 APCC MME 자료를 수집하여 이용하였다.

합성곱신경망모델(CNN)의 입력으로 사용하기 위해서 모든 자료를 2.5° x2.5° 해상도로 리샘플하고 공통 영역인 72x144 격자로 잘라내었다. 그래프신경망모델(GNN)을 위해서는 15° x15° 해상도로 리샘플하고 공통 영역인 12x24 격자로 잘라내었다. 시간적으로는 모든 자료를 각 격자에 대하여 기준기간(1991-2020)에 대해 표준화 시계열 전처리하였다.

2.5.2. 자료강화기법 적용

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. Semi-monthly 방법에서는 일별 자료를 전월, 상반기(1-15일), 하반기(16-말일)에 대해 합산(강수) 또는 평균하여 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. Cutmix 방법(Yun et al. 2019)은 모델 입력을 위한 4D 자료에 적용하되, 변수에 대해서는 고정하고 (위도, 경도, 기간)의 3D 부피에 대해 치환하여 부피비로 각 카테고리의 가중치를 곱하여 레이블을 산정하였다. 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. 입력자료의 수가 적으므로, 훈련연도의 모든 조합에 대해 Cutmix를 적용하였다.

2.5.3. 인공지능모델 및 실험 구성

모델의 아키텍처 개선에 대한 실험은 아니므로 최대한 간단한 모델을 구축하였다. CNN 모델은 두 개의 합성곱신경망 블록으로 이루어진 특징 추출 네트워크와 완전연결층 및 Softmax 함수의 분류 네트워크로 구성하였다. 입력자료 차원은 (위도, 경도, 기간, 변수)의 (72, 144, 7, 5), 목표는 우리나라 평균기온 및 강수(SPI1) 삼분위 예측 카테고리 3개로 정하였으며 모델의 최적화 함수로 Adam, 손실함수는 categorical cross entropy를 사용하였다. 불확실성 분석을 수행하는 경우에 대하여 두 합성곱신경망 블록과 완전연결층에 대해 0.2의 dropout을 적용하였다(Figure 2.66).

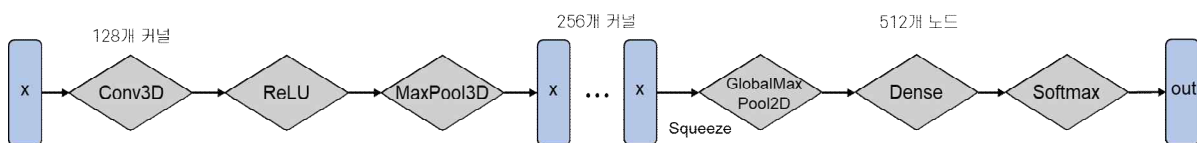


Figure 2.66. Structure of the 3D CNN toy model used in this study.

GNN 모델은 노드 분류를 위한 모델(Figure 2.67)과 그래프 분류를 위한 모델(Figure 2.68)을 각각 구성하였다. 노드 분류 모델은 세 개의 그래프합성곱신경망 블록으로 이루어진 특징 추출 네트워크와 완전연결층 및 Softmax 함수의 분류 네트워크로 이루어져 있으며 각 월이 하나의 노드를 구성한다(ex. 2017년 7월). 노드 특징(Node features)은 (위도, 경도, 기간, 변수)를 풀어 1D 벡터로 구성된 입력자료 값이며 노드 라벨(Node label) 우리나라 평균기온 삼분위 예측 카테고리 3개이다. 엣지(Edge)는 노드 사이의 연결 및 연결 강도를 나타내는데 엣지 연결 및 가중치 초기값으로 각 노드쌍에 대해 노드 특징 벡터 사이의 코사인 유사도를 구하고 그 값이 0.1 초과인 경우(대략 상위 25% 정도에 해당됨) 엣지를 연결하여 그 코사인 유사도를 가

중치 초기값으로 설정한다. �지의 가중치를 학습가능한 매개변수로 지정하여, 모델의 학습과정에서 함께 학습될 수 있도록 하였다. 모델의 최적화 함수와 손실함수는 CNN 모델과 같다. 노드 임베딩의 인코더로 GCN (Graph Convolutional Network) 모델을 사용하였다.

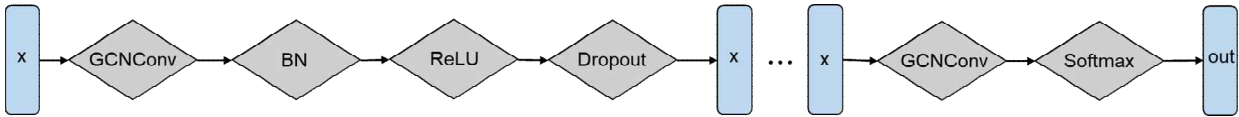


Figure 2.67. Structure of the GNN model for Node classification used in this study.

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 카테고리가 된다. 그래프 안의 각 노드는 공간 격자(위도, 경도)로 이루어지며 노드 특징은 (기간, 변수)를 푼 1D 벡터로 구성된 입력자료 값이다. �지(edge)는 각 노드쌍에 대해 노드 특징들 사이의 코사인 유사도를 구하고 그 값이 0.1 초과인 경우 또는 인접한 그리드인 경우에 대해 연결하며 가중치는 사용하지 않았다. 모델의 최적화 함수와 손실함수는 CNN 모델과 같다.

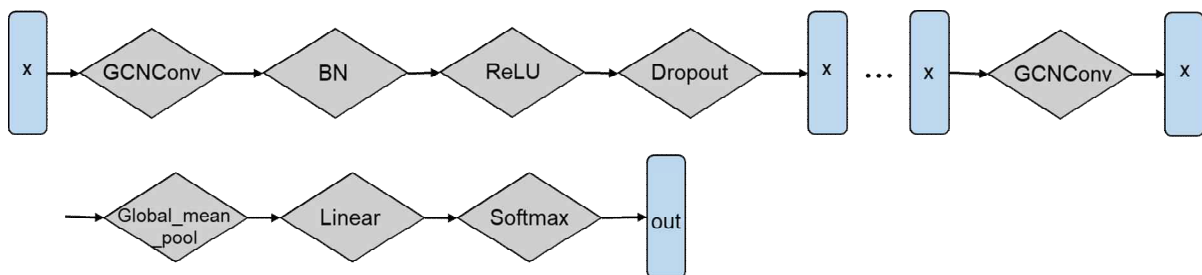


Figure 2.68. Structure of the GNN model for Graph classification used in this study.

여름철(6월 LT1, 7월 LT2, 8월 LT3)과 겨울철(12월 LT1, 1월 LT2, 2월 LT3) 평균기온 및 강수(SPII) 예측 성능 개선이 목표인데 월별, 선행시간별로 모델을 따로 구성하는 경우 가용 자료의 수가 매우 적다. 따라서 여름철 평균온도 예측에 대해 모델 훈련 시 모든 월의 자료를 다 사용하는 경우(Month-Agnostic)와 목표 월만 사용하는 경우(Target Month Only)를 비교하였다. 전체 기간에서 모델의 성능을 최종 평가할 테스트셋을 제외하고 나머지 기간을 3개의 폴드로 나누어 교차검증을 수행하였다.

2.6. 인공지능 기술을 활용한 다중모델 앙상블 강수 역학예측 후처리

2.6.1. 훈련 자료 및 전처리 방법

현재 11개국 15개 모델이 APCC-MME 전지구 예측에 참여 중이며 (Table 2.11), 이러한 국가/기관의 다양성은 개별 모델 독립성과 충분히 넓은 앙상블 분포를 요하는 MME방식에 부합한다고 볼 수 있다. APCC에서 계절 예측 서비스가 본격적으로 시작된 2013년 1월부터 2023년 12월까지 11년 동안 수집된 모형 중에서 QM 후보정이 가능할 만큼 충분한 자료량을 가진 모형은 37개로 목록화 할 수 있으며, 기간 내 모든 모형 Real-Time Forecast(RTF) 전체 앙상블 숫자는 34602이다. MME 참여 모형 개수는 2013년 1월 6개에서 2023년 12월 13개까지 점차 증가했고, 개별 모델링 센터에서도 최신 버전/기간으로 갈수록 앙상블 개수를 늘려왔다. 따라서, MME에 활용되는 총 앙상블 개수도 최신으로 갈수록 전반적으로 늘어나는 경향을 보인다. 2023년 12월 기준으로 가장 많은 앙상블을 가진 모형은 KMA GloSea6-GC3.2(84개)이고 그 다음 그룹은 METFR SYS8(51개), CMCC SPS3(50개), JMA MRI-CPS3(50개)이다. 반대로 가장 적은 앙상블을 가진 모형은 NASA GOES-S2S-2.1, APCC SCoPS로 10개의 앙상블을 가지고 있다. 개별 예측 시스템 및 모형 개요, 모델 변경 이력 등의 정보는 아태 기후센터 홈페이지를 통해 확인 가능하다 (<https://apcc21.org/prediction/global/model?lang=ko>).

이 연구에서는 2013년부터 2023년까지 11년 동안의 모든 역학 모델 RTF 앙상블 예측장을 인공지능 후처리를 위한 학습자료로 활용하였다. 이 중 2013년부터 2022까지 10년에 대한 앙상블을 train/validation을 위해 4/1의 비율로 랜덤하게 나눠 훈련을 진행하고, 마지막 2023년 12개월에 대해 성능을 테스트 하였다. 훈련용 자료는 월자료로, 목적 변수는 계절 시간 규모 (3개월 평균)로 frequency를 분리한 것이 이 연구의 특징으로 볼 수 있겠다. 역학 모델 후처리의 특성상 훈련에 사용되는 모델 앙상블 샘플은 많지만 이에 상응하는 관측 realization은 매우 한정적이다. 어쩔 수 없이 모든 계절을 동시에 고려하는 방식으로 (season-agnostic) 모델을 개발하게 된다. 이 과정에서 계절 사이클의 의존성을 회피하기 위해서 모든 모델 훈련 직전에 학습 샘플을 셔플해서 계절성을 제거하는 방식으로 코드를 구성하였다.

APCC MME에 활용되는 모든 앙상블에 대해 8개의 대기 변수장(강수(PRECIP), 하층 기온(T2M), 상하층 동서/남북 바람장(U850, V850, U200, V200), 중층 지위고도장(Z500), 해면기압(SLP))과 해수면 온도 (SST)를 추출해서 학습에 활용하였다. 이에 상응하는 ERA5 (Hersbach et al., 2020) 재분석 자료 변수 역시 추출해서 활용하였다. 인공지능 후처리 모델의 목적 변수인 동아시아 계절 강수에 대해 다양한 관측 원천 옵션이 있을 수 있으나 이 연구에서는 ERA5 강수장을 선택하였다. 입력 자료로 들어가는 다른 변수들과의 일관성과 후처리 모델 개발 후 현업적인 활용을 고려했을 때, ERA5 강수장으로 세팅하는 것이 가장 합당하다고 판단하였기 때문이다. 동아시아 지역은 75~150° E, 15~60° N 영역으로 설정하였고, 2.5도 공간 해상도 자료를 활용하였다.

평년값 기준 기간은 개별 모델링 센터들의 정책에 따라 달라지며 MME 예측에서는 그 공통기간을 평년값 기준으로 잡게 된다. MME를 추출하기 위한 평년 기간은 2021년 1월을 기준으로 1983-2005에서 1991-2010으로 변경되었다. 안타깝게도 WMO가 추천하는 30년 평년값 기

준에 맞출 수 없었는데 이는 개별 센터들 각자의 과거 재현 자료(Hindcast) 생산 기간이 다르
고, 매해/매월 on-the-fly 형태로 최신 시기로 업데이트되면서 모델간 공통 기간이 계속적으로
짧아지기 때문이다. 본 연구에서는 MME 기준에 부합하는 평년 기간과 일관되게 아노말리장을
산출하고 이 자료를 인공지능 모델 학습에 활용함과 동시에 실제 성능 평가에도 적용하였다.

현업 계절 예측 모델 자료 후처리 보안을 위해 전이 학습을 적용할 수 있고, 이를 위해
외부 모델 자료 CESM2-LENS (Rodgers et al., 2021)를 수집하여 학습에 활용하였다. 1850년에
서 2014년까지 165년에 대해 100개 대량 앙상블 historical simulation 실험 자료로서 앙상블 실
험 설계에 대한 세부 설명은 홈페이지에서 확인할 수 있다
(<https://www.cesm.ucar.edu/community-projects/lens2>). 앞서 제시한 9개 대기/해양 변수 이외에
하층 specific humidity (Q850), 하층 연직 속도 (OMEGA850) 변수를 추가하여 다양한 입력 변수
조합에 대한 모델 실험을 수행하였다.

Table 2.11. Dynamical seasonal prediction models used for developing an AI model to enhance EA rainfall forecast accuracy.

Abbr.	Organization/Nation	Model system name	Main reference/URL
APCC	APEC Climate Center/Korea	SCoPS (Seamless Coupled Prediction System)	Ham et al. (2019)
BCC	Beijing Climate Center/China	BCC_CSM1.1m	Wu et al. (2010)
BOM	Bureau of Meteorology/Australia	ACCESS-S2 (Seamless seasonal prediction System)	To be provided.
CMCC	Centro Euro-Mediterraneo sui Cambiamenti Climatici/Italy	CMCC-SPS3.5 (CMCC operational Seasonal Prediction System version 3.5)	https://www.cmcc.it/it/publications/tn0288-the-new-cmcc-operational-seasonal-prediction-system-sps3-5
CWA	Central Weather Administration/Chinese Taipei	TCWA1Tv1.1	Paek et al. (2015)
ECCC	Environment Climate Change Canada/Canada	CanSIPSv2.1 (The Canadian Seasonal to Interannual Prediction System ver. 2.1)	https://collaboration.cmc.ec.gc.ca/cmcc/cmci/product_guide/docs/tech_notes/technote_cansips-300_e.pdf
HMC	Hydrometeorological Centre of Russia/Russia	SL-AV-072L96	http://neacc.meteoinfo.ru/forecast/documentation
JMA	Japan Meteorological Agency/Japan	MRI-CPS3	https://ds.data.jma.go.jp/tcc/tcc/products/model/outline/cps3_description.html
KMA	Korea Meteorological Administration/Korea	GloSea6GC3.2 (Global Seasonal Forecasting System ver 5)	Kim et al. (2021)
METFR	Météo-France/France	MF Sys 8	http://www.umr-cnrm.fr/IMG/pdf/system8-technical.pdf
MGO	Voeikov Main Geophysical Observatory/Russia	MGOAM-2	Модель общей циркуляции атмосферы ГГО (версия MGO-03 T63L25) // Труды ГГО. Вып. 571. С. 5–87.
NASA	National Aeronautics and Space Administration/USA	GEOS-S2S-2.1(Observing System Atmosphere-Ocean General Circulation Model and Data Assimilation System Version S2S-2_1)	https://gmao.gsfc.nasa.gov/cgi-bin/products/climateforecasts/geos5/S2S_2/index.cgi
NCEP	National Centers for Environmental Prediction/USA	CFSv2 (Climate Forecast System version 2)	https://cfs.ncep.noaa.gov/
PNU-RDA	Pusan National University-Rural Development Administration/Korea	PNU-RDA CGCM v2.0 (Pusan National University-Rural Development Administration (National Institute of Agricultural Sciences) Coupled General Circulation Model version 2.0)	Ahn et al. (2018)
UKMO	United Kingdom Met Office/United Kingdom	GloSea6 (Global Seasonal Forecasting System version 6)	https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/gpc-outlooks/user-guide/global-seasonal-forecasting-system-glosea6

2.6.2. 후처리 대상 지역 및 성능 평가 방법

본 연구에서는 동아시아 강수 예측장과 관측장 사이의 공간분포가 얼마나 유사한가를 평가하기 위해 ACC(Anomaly Correlation Coefficient, <https://confluence.ecmwf.int/display/FUG/Section+6.2.2+Anomaly+Correlation+Coefficient?form=MG0AV3>) 지수를 적용하였다. 예측장과 관측장에서 각각의 기후장을 제거해 편차장을 얻고, 위도 별로 줄어드는 면적에 대한 가중치를 고려해 두 공간분포에서 동일한 위치에 있는 값들 간에 상관계수를 계산하게 되며, 수식은 아래와 같다.

$$ACC = \frac{\sum_{i=1}^n w_i (f_i - \bar{f}_i)(o_i - \bar{o}_i)}{\sqrt{\sum_{i=1}^n w_i (f_i - \bar{f}_i)^2 \sum_{i=1}^n w_i (o_i - \bar{o}_i)^2}}, \quad w_i = \cos \varphi_i \quad [2.18]$$

여기서 n은 격자수를 φ_i 는 위도를 의미하고, $f_i, \bar{f}_i, o_i, \bar{o}_i$ 는 격자별 예측값, 예측자료의 기후값, 관측값, 관측자료의 기후값을 나타낸다. 양의 부호로 값이 클수록, 예측 편차장과 관측 편차장의 패턴이 유사하다고 판단할 수 있다.

2.6.3. 역학 모형 후처리 접근 방식 및 덤러닝 모형

최종적으로 선정된 세 가지 후처리 방식은 아래 Figure 2.69과 같다. 첫 번째는 실시간 예측 앙상블 중에서 양질의 앙상블을 선별(SubSampling, SS)하는 방식이다. 매달 300개가 넘는 앙상블들이 존재하고 그 중에서는 실제와 매우 다른 반대의 패턴을 보이는 경우도 다수이다. 이러한 불량 앙상블들을 미리 선별해서 탈락시키고 가능한 양질의 앙상블들만을 취하게 된다면 다소 앙상블의 개수가 줄어들긴 하지만 특정 모형을 선택해서 탈락시키는 것이 아니기에 여전히 독립성은 유지하게 되고 따라서 기존 다중모형앙상블 기법의 장점을 유지할 수 있을 것이다. ACC 값에 따라 앙상블을 나열하고 상위 10, 30, 50, 70, 90%에 해당하는 앙상블만을 취하는 분석에서 ACC값은 전체 앙상블 평균 대비 높아지는 것으로 나타난다. 기준 %tile이 높을수록 최종 성능 역시 비례적으로 올라가고, 계절별, 연도별 분석에서도 일관되게 나타나는데, 이러한 결과는 앙상블 선별의 잠재력을 제시한다고 볼 수 있다. 기존 연구(Kim et al., 2022; 아태 기후센터 연구보고서) “동아시아 겨울 기온 선행3주 예측 성능 추정을 위한 가이드라인”에 따르면, 기후 변동 모드의 위상, 앙상블 예측의 시간 변동폭, 그리고 이전 달 이슈된 예측의 평가 성능 등의 정보를 취합해 이번 달에 이슈된 예측 앙상블의 성능을 선추정하는 것이 일정 정도 가능하다는 것을 보인 바 있다. 또한, 계절 예측에서 대기 변수 성능이 해양 변동모드 ENSO의 위상에 크게 의존적인 것은 매우 잘 알려져 있다. 이러한 앙상블 성능 의존 조건을 덤

러닝 모형이 학습할 수 있게 함으로써 300개가 넘는 앙상블 중에서 양질의 앙상블을 선별할 수 있게 되고, 결과적으로 좀 더 나은 다중모델앙상블 결과를 도출할 수 있을 것으로 판단하였다.

두 번째 후처리 방식은 예측장과 관측값을 one-to-one (혹은, multi-to-one) 학습하도록 함으로써 동아시아 강수 예측의 오차를 제거 (DeBiasing, DB)시키게 되며, 기존의 연구들에서 주로 활용하는 방식이다. 마지막 후처리 방식은 전이 학습 기반의 진단 방식(Transfer-Learning based Diagnosing, TLD)이 되겠다. 현업 앙상블들을 학습하는 경우 샘플을 많이 확보할 수 있지만 상응하는 과거 관측 사례는 11년, 12 계절에 불과해 이제까지 발생하지 않은 새로운 케이스에 대한 예측에 대응하기 어렵다. 따라서 과거의 사례를 늘리는 것이 가능한 외부 모델을 차용해 학습을 진행하는 것에 대한 방안을 모색하였고, 165년의 장기, 100개의 대량 앙상블 조건을 갖춘 CESM2-LENS 모델 자료를 활용해 전이 학습을 진행하였다. 다만 여기서 목적 변수인 강수를 입력장에서 제외시켜 기존 DB방식과 차별화를 두었다, 이는 NVIDIA FourCastNet (Pathak et al., 2022)에서 강수에 대해 prognostic 방식이 아닌 다른 변수로부터의 강수를 diagnose하는 방식에서 힌트를 얻은 것이다. 세 가지 딥러닝 기반 후처리 방식 시도가 의미를 갖기 위해서는 기존의 통계 방식 후보정 대비 더 나은 성능을 담보할 수 있어야 한다. 이를 위해 보편적으로 가장 많이 쓰이는 통계 오차 보정 기법인 Quantile Mapping(QM)을 적용한 동아시아 계절 강수 결과를 기초선으로 평가를 진행하였다.

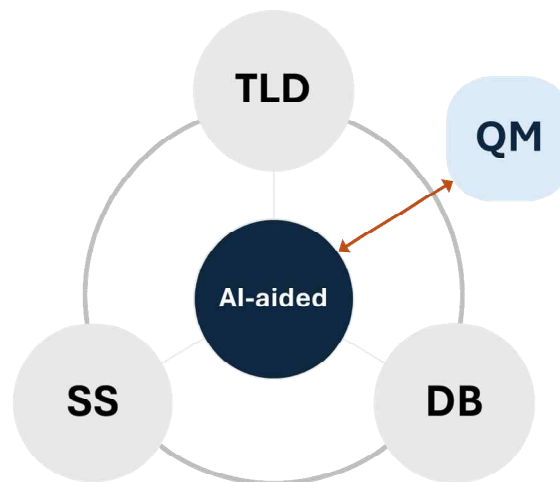


Figure 2.69. AI-based and conventional statistical approaches for enhancing EA seasonal rainfall forecast accuracy. SS: sub-sampling, DB: debiasing, TLD: transfer-learning-based diagnosing, QM: quantile mapping.

2.6.3.1. 앙상블 선별 모델 구축 및 민감도 실험 방법

앙상블 선별 모델의 구조는 아래와 Figure 2.70와 같다. 예측을 수행하는 월을 기준으로 T=0로 두고, 과거 3개월의 관측(재분석)장과 미래 3개월의 앙상블 예측장을 이어 붙이고 최대

8개까지의 동적 물리 변수와 land-sea mask(LSM) 등 정적 변수를 쌓아 다수의 입력 채널을 구성하게 된다. 예측 수행 월(T=0)에 대한 앙상블 값을 이전 달의 lead=2에 해당하는 앙상블 예측을 이어 붙여 시간적인 연속성을 고려하는 민감도 실험도 수행하였다. 강수라는 물리 변수가 아닌 개별 앙상블 ACC 값을 목적 변수로 둔 것이 기존 연구와 차별화되는 점이라 볼 수 있다. 이렇게 최신 대기상태에 대한 정보를 학습해서 개별 앙상블의 ACC를 추정하고, 추정된 스کیل 값 기준으로 상위 앙상블만을 선별해 동아시아 계절 강수값을 얻는 구조이다. 이 과정에서 ResNet 아키텍처(He et al., 2016)를 적용했다. 이러한 앙상블 선별 모델 설계는 과거 관측에서부터 앙상블 예측으로 이어지는 시·공간 진화가 앙상블 예측 성능 (ACC)와 일정 정도 관련성이 있을 수 있다는 가정에 기반한다. 대기장이 제시하는 기후 변동성의 위상이나 강도 조건뿐만 아니라, 관측에서 예측으로 이어지는 대기장 진화/방향성이 성능에 영향을 줄 수 있기 때문이다. 입력장의 선행 시점을 기준으로 봤을 때, 과거 3개월의 재분석장 정보는 딥러닝 모델 학습 과정에서 예측(forecasting), 미래 3개월의 앙상블 예측장 정보는 오차 제거(debiasing)라는 방식으로 ACC 추정 모델에 기여한다고 이해할 수 있을 것이다.

민감도 실험은 기본적으로 모델 깊이, 입력변수, 초매개변수 조정에 대해 이루어졌다 (Table 2.12). ResNet 레이어 깊이를 18/34/50/101로 늘려가며 성능 변화와 안정도를 확인하였고 가장 단순한 ResNet18과 가장 복잡한 ResNet101이 가장 좋은 결과를 보여, 이 두 아키텍처 중심으로 추가 실험이 이루어졌다. 입력장의 공간 해상도를 2.5도에서 5도로 낮췄을 때의 성능 차이가 없음을 확인했고, 전체적인 입력 변수 해상도를 낮추는 방식으로 전환해 메모리 이슈를 해결할 수 있었다. 입력변수에 SST 추가 여부, LSM, 위도, 지형 변수 추가 여부, 이전 월에 이슈된 예측 수행 시점(T=0) 앙상블 예측 추가 여부, 1차원 기후 지수들을 multi-modal input 형태로 입력에 추가할지 여부 등에 따른 민감도를 다양한 조합으로 실험하게 된다. 육지를 zero padding 처리한 SST 자료를 입력에 추가했을 때 loss를 줄이지 못해 학습이 멈추는 경우가 나타나 정제된 실험에서는 SST를 제외하게 된다. static 변수로는 지형 변수를 제외한 LSM과 위도 조합이 성능 개선에 유의하게 기여하는 것으로 판단해 최종 입력변수로 확정했다. T=0 예측장 추가에 대해서는 일관된 결과를 확인할 수 없었고, 기후 지수들을 추가하는 실험 역시도 다른 실험 조건과 조합했을 때 들쭉날쭉한 결과를 보여서 최종적으로는 제외하게 된다. 초매개변수에서는 learning rate, batch size, epoch tolerance 등 조정하면서 최적 세팅을 찾게 되는데, 최종 모델 확정시에는 learning rate는 5×10^{-3} , 1×10^{-3} , batch size=8, epoch tolerance=20,30으로 세팅을 진행하였다. 특히, batch size 줄이는 것은 훈련 속도를 높일 뿐만 아니라 테스트 성능까지 높일 수 있는 방안이라는 팀 동료의 제안을 참고했다. batch size를 줄이는 대신 가용 gpu memory에 맞춰 입력 자료 종류를 늘리는 방식으로 실험을 확대할 수 있었다. 비용 함수는 기존 연구들처럼 주로 MSE를 활용했다. custom 방식으로 위도에 가중치를 주는 MSE 역시 적용해 보았으나, 성능이 오히려 감소하는 경향을 보여 최종적으로는 모델 세팅에서 제외하였다. 성능 추정에 실제 기여하는 지역의 범위가 중위도 좁은 영역으로 한정되다 보니, 저위도에 높은 가중치를 주는 방식이 잘못된 방향으로 학습을 유도하게 되는 것이 아닌가 가늠할 수 있었다. 이러한 민감도 실험에 대한 세부적인 결과는 과제 보고서에 포함하지 않았고, 최종적으로 선정된 모델 결과 위주로 기술하였다.

로 설계하지는 않았다. 앞선 앙상블 선별 모델 설계에서와 마찬가지로 예측 시점 T=0 기준으로 과거 3개월 관측장과 미래 3개월 앙상블 예측장을 연결해서 붙여서 입력장을 구성하게 되고, 목적 변수는 우리가 원하는 3개월 평균 강수 아노말리장으로 세팅하게 된다(Figure 2.71). 강수 이외의 여러 대기 물리 변수, 정적 변수들을 고도에 따라 적층시키고 각 선행 시간 자료를 연속으로 쌓는 방식으로 수 십개의 채널을 구성하게 된다. 입력장은 여러 물리장을 포함하게 되는데 이러한 강수장을 포함한 최신 가용 대기장을 학습해서 다음 동아시아 계절 강수장을 추정하는 방식으로 오차 제거 모델을 설계하였으며, 개별 앙상블에 대한 결과에 대한 합성장으로 최종 결과를 도출하게 된다. 딥러닝 아키텍처로 기본 UNET (Ronneberger et al., 2015) 이외에 attention module을 포함하는 ATTUNET (Oktay et al., 2018), Residual net 기반 RESUNET (Diakogiannis et al., 2020), 마지막으로 합성곱 신경망을 swin-transformer로 대체한 SWINUNET (Zhou et al., 2022)으로 변형 실험을 수행하였다. 이러한 오차 제거 후처리 모델 설계는 과거 관측에서부터 앙상블 예측으로 이어지는 시·공간 진화가 앙상블 예측 오차(bias)와 일정 정도 관련성이 있을 수 있다는 가정에 기반한다.

Spatiotemporal Evolution (from OBS to ENS) ~ DSP BIAS

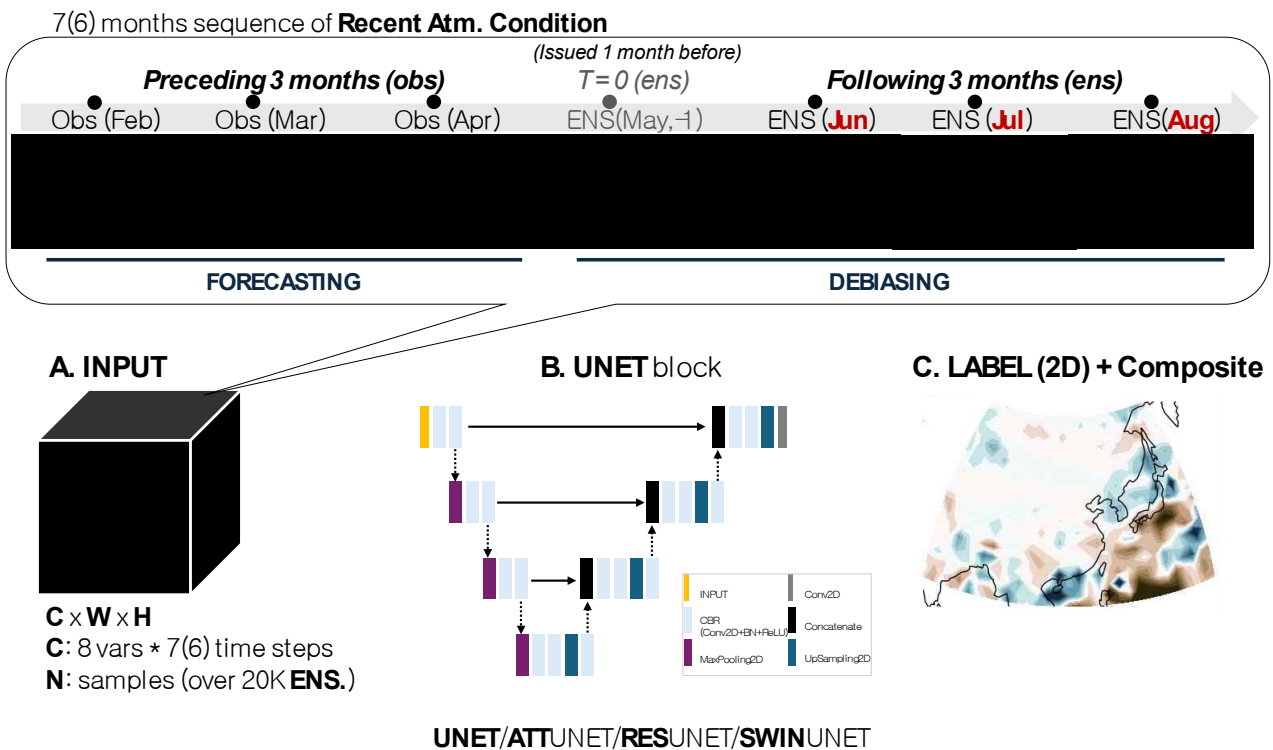


Figure 2.71. Illustration of the debiasing (DB) approach: (A) training input, (B) deep-learning architecture, (C) 2-dimensional rainfall field as the label.

UNET 아키텍처를 기본으로 민감도를 테스트하였다(Table 2.13). 기본 UNET과 그 변형 아키텍처 ATTUNET, RESUNET, SWINUNET에 대해 합성곱 신경망 layer depth와 down/up sampling 층 수를 조절하는 실험을 수행하였고, 안정적이고 유의한 성능을 보이는 깊이를 최종

모델에 적용하였다. 입력장 관련, 물리 변수와 선행 시간에 대한 여러 조합에 대해, 정적 변수의 추가 여부에 대해, 입력장 해상도에 대한 실험하였다. 최종적으로는 'PRECIP', 'T2M', 'U850', 'V850', 'Z500', 'U200', 'V200', 'SLP' 8개 물리 변수를 T-3에서 T+3까지 6 (7)개 선행시간을 입력장에 적용하였다. 선행시간에 대해 고정인 정적변수는 LSM, 위도 2개를 추가하는 것이 가장 안정적이어서 최종 모델에 적용하였다. 5deg 저해상도 자료로 훈련하였을 때도 성능 차이가 크지 않아 저해상도 세팅으로 정하였다. 초매개변수는 민감도 테스트 결과 앙상블 선별에서와 유사하게 batch size=8, epoch tolerance=20,30으로 세팅하였다. 학습율의 경우는 아키텍처별로 다르게 세팅했으며, 특히 SWINUNET의 경우는 그 값을 1×10^{-4} 혹은 그 이하로 세팅하였다.

Table 2.13. Sensitivity test setting for the DB approach.

ARCHITECTURE	INPUT	HYPERPARAMETER
- basic UNET and their variants (ATTUNET, RESUNET, SWINUNET)	- spatial resolution (2.5 or 5deg)	- Learning rate (5x10 ⁻³ , 1x10 ⁻³ , 5x10 ⁻⁴ , 1x10 ⁻⁴)
- layer depth	- different variables and lead-time combinations	- batch size (8,16,32)
	- adding time-invariant variables or not (LSM, LAT, Orography)	- Epoch tolerance (15,20,25,30)

앞서 제시한 오차 제거 방식 UNET 모델 2개를 적층하는 변형 방식의 오차 제거 실험도 추가하였다. Yang et. al. (2023)은 북서태평양아열대고기압(WNPSH) 예측 모델 결과를 강수 예측 모델 입력장으로 추가하는 형태로 2개를 적층하는 방식(Two-step U-Net model)을 통해 양쯔강 중하류의 여름철 계절 강수 예측이 개선됨을 제시하였다. 이 연구를 벤치마킹해 동아시아 하층 바람장 DB 모델과 동아시아 강수 DB 모델 2개를 적층하는 방식으로 실험을 수행하였다 (Figure 2.72). 이러한 시도의 근거는 딥러닝 모델 출력으로 얻어진 개선된 하층 바람장을 입력으로 받아 강수 DB 모델을 수행하게 되면 최종 강수장 개선을 기대할 수 있다고 보기 때문이다. 기본 실험 설계는 미리 훈련된 하층 바람장 DB 모델 결과를 강수 DB 모델의 2번째 layer 입력에 추가해 강수 모델을 훈련하는 것이다. 또 다른 시도는 기존 모델의 바람장/강수장 결과를 새로운 강수 모델 초기 입력으로 넣어 학습함으로써 수정된 강수장을 얻는 방식이다. 두 모델 구조에 대하여, 입력 변수 다변화, 초매개 변수 수정 등의 민감도 실험을 수행하였으며, 최종적으로 유의한 성능을 보인 모델을 선정하고 그 결과를 다음 장에서 기술하겠다.

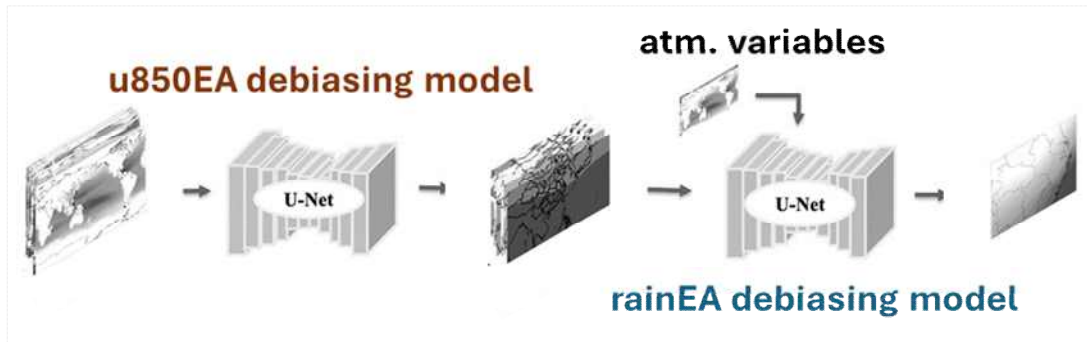


Figure 2.72. Second DB approach using Stacked Two-models (ST): one is a model for debiasing EA lower-level zonal wind field (left) and the other is a model for debiasing EA rainfall field (right).

2.6.3.3. 전이 학습 모델 구축 및 민감도 실험 방법

현업 계절 예측 자료를 활용한 훈련의 가장 큰 한계는 최근 10여년이라는 짧은 자료 기간 때문에 학습할 수 있는 과거 realization의 수가 현저히 부족하다는 점이다. 이 때문에, 과거에 존재하지 않은 패턴/현상에 대해서는 예측 정확도를 기대하기 어렵다. ERA5 자료가 1940년 과거까지 확장되어 관측 대조군으로는 80년 넘는 기간이 있지만, 역학 예측 후처리라는 연구 목적에 부합하는 현업 계절 예측 기간을 역지로 연장하는 것은 불가능하다. 과거 재현 자료(hindcast)를 활용하는 방법도 시도할 수 있지만, hindcast는 기후장르 생산하기 위한 모델 세팅으로, 정제된 재분석장을 초기값으로 취하기에 real-time forecast에서와 같은 형태의 오차 추이를 가질 수 없다는 점, 각 월에 대해 다른 버전의 hindcast를 선별하고 종합하는 과정이 복잡하다는 점 때문에 논의에서 제외시켰다. 대신, 장기(165년)의 대량 앙상블(100개)을 가진 외부 모델 CESM2-LENS 자료를 이용해서 pre-train하고 이를 재분석장으로 fine-tuning하는 전이 학습 방식의 모델을 설계하게 된다(Figure 2.74). CESM2에서 훈련된 최신 대기장과 3개월 강수장 간의 관련성이 재분석장을 거치면서 관측 기반으로 조정되고 이렇게 얻어진 모델에 최종적으로 MME 자료에 대해 테스트하는 실험을 수행하였다. 개별 계절에 대해 16,500개의 pre-training 훈련 샘플을 확보할 수 있기에, 모든 계절을 하나의 모델로 훈련하는 앞선 두 후처리 방식과 다르게 계절별로 독립된 모델을 훈련시켰다. 입력 변수 조합에 대한 민감도 실험을 수행하였고(Figure 2.75), 과거 3개월 관측 재분석장에 대해서는 모델 변수 이외에 850hPa specific humidity (Q850), vertical velocity (OMEGA850)을 추가하였다. 미래 3개월 앙상블 예측장에 대해서는 강수 변수를 제외함으로써 변수에 대한 one-to-one 직접 보정이 아닌 다른 대기 변수로부터 진단(diagnose) 방식으로 후보정 전략을 선택하였다. 이는 NVIDIA FourCastNet(Pathak et al., 2022)의 강수 모듈 전략을 일부 차용한 것이다. ATTUNET, SWINUNET을 기본 아키텍처로 사용하였고, 학습율을 고정시킨 이전 실험들과 달리 step-wise reduction, loss 감소에 따른 동적 조정 방식을 pre-train과 fine-tuning 각각에 적용하였다.

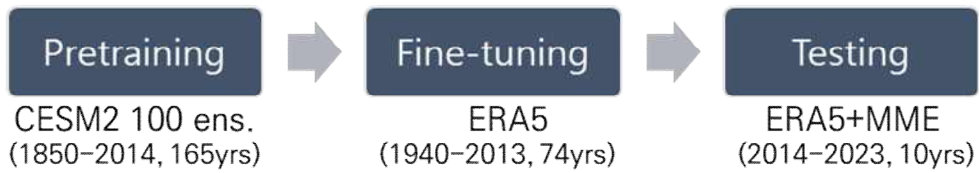


Figure 2.73. Process of transfer learning for diagnosing EA rainfall.

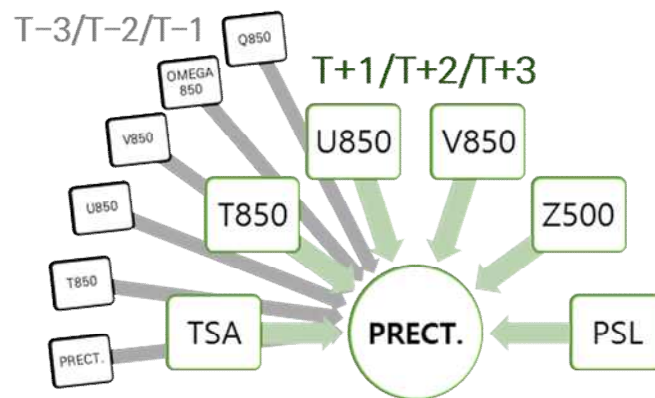


Figure 2.74. Physical variables used as training input for the TLD approach.

3. 결과

3.1. 자료 확장에 따른 모델 성능의 민감도 분석

3.1.1. 강수량 자료 학습결과

자료확장에 따른 모델 성능의 정확도를 빠르게 평가하기 위하여 정확도 개선이 어려운 강수량에 대해서 Attention U-Net모델의 입력자료로 ECMWF의 강수량자료와 관측자료인 GPM 강수량자료를 3D입력자료(20년*7일, 위도, 경도, 변수)로 구성하고, 이 중에서 15년치는 학습자료로 나머지 5년치는 테스트 자료로 훈련셋을 구성하였다. Attention U-Net모델에서 convLSTM2D함수를 이용하여 합성곱으로 레이어를 구성하고 activation은 relu 함수를 사용하여 모델을 구성하였다. 손실함수는 MSE(Mean Square Error)함수로, optimizer 함수는 Adam함수로 이용하고 batch_normalization을 적용하지 않고 최종적으로 리드타임 3주차 7일 강수량자료가 나오도록 ERA5자료를 두고 학습을 진행하였다. Figure 3.1에서 보듯이 두 개의 변수만으로 학습한 결과는 극치가 평균화되어 잘 표현되지 못하는 것을 볼 수 있다. 강수량의 경우에는 Attention U-Net모델에서 강수 특성이 리드타임 3주까지 전달되지 않는 것을 볼 수 있다.

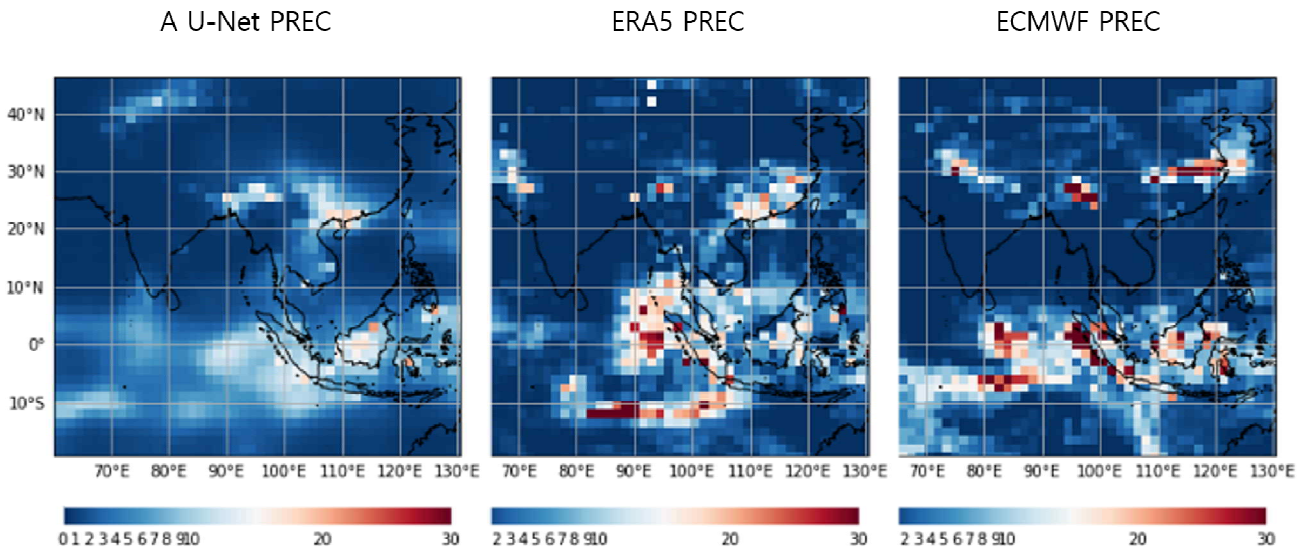


Figure 3.1. Comparison of the result of Attention U-Net and ECMWF model.

3.1.2. Filter 기법에 따른 민감도 분석 결과

관측자료인 NDVI, SST와 ECMWF의 모든 변수를 ERA5와의 상관성을 조사하였다. 각각의 변수별과 ERA5의 최고온도에 상관성 중에서 가장 상관성이 높은 변수를 중심으로 입력자료를 구성하였다(Table 3.1). 각각의 변수는 TMAX, TMIN, T2M, Z200, Z500, Q850, Q700, MSL, PRE, NDVI(Obs), SST(Obs)의 순서대로 상관성이 높았다. 상관성이 높은 변수를 선택하고 각각의 변

수중에서 상관성이 높은 순서대로 Attention U-Net모델에 입력자료로 구성하여 학습을 진행하였다.

Table 3.1. Index and variables of subsets for filter method.

Description of subsets for Filter method	
A	TMAX(ECMWF)
B	TMAX,TMIN
C	TMAX,TMIN,T2M
D	TMAX,TMIN,T2M,Z200
E	TMAX,TMIN,T2M,Z200,Z500
F	TMAX,TMIN,T2M,Z200,Z500,Q850
G	TMAX,TMIN,T2M,Z200,Z500,Q850,Q700
H	TMAX,TMIN,T2M,Z200,Z500,Q850,Q700,MSL
I	TMAX,TMIN,T2M,Z200,Z500,Q850,Q700,MSLPRE
J	TMAX,TMIN,T2M,Z200,Z500,Q850,Q700,MSLPRE,NDVI(OBS)
K	TMAX,TMIN,T2M,Z200,Z500,Q850,Q700,MSLPRE,NDVI(OBS),SST(OBS)

Filter 기법 분석결과 Table 2.1에서 보듯이 A의 카테고리인 ECMWF 계절내 모델의 TMAX를 단독으로 사용하여 ERA5와의 리드타임 3주차 ACC값이 다른 모든 상관성이 높은 변수별로 학습한 최고온도와 ERA5의 최고온도가 ACC보다 높았다. ECMWF의 TMX값의 ACC값이 0.31로 ACC 0.2대의 상관성 높은 변수들의 학습결과 보다 정확도가 높았는데 ECMWF 모델 자체의 예측성능이 좋은 것으로 판단된다(Figure 3.2).

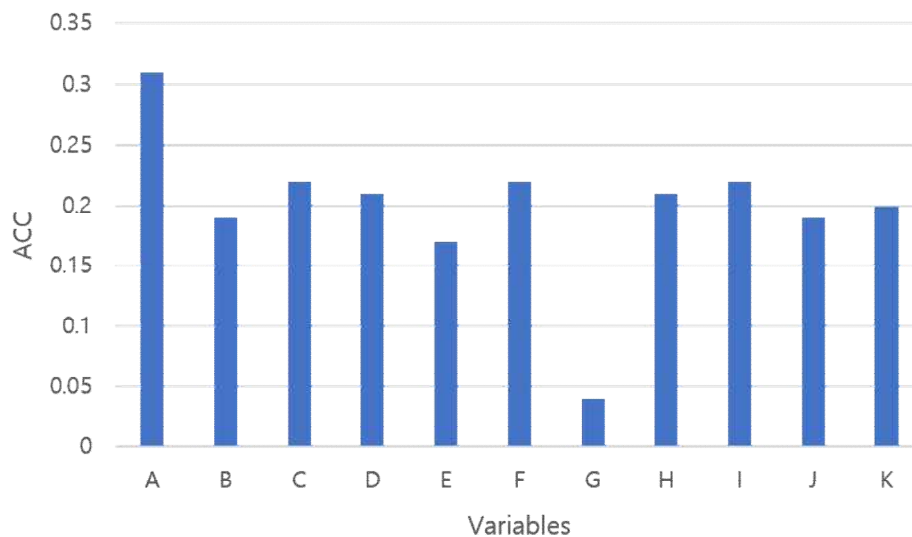


Figure 3.2. Trained results obtained from variables based on filter method.

3.1.3. Wrapper 기법에 따른 민감도 분석 결과

Wrapper 기법의 결과를 분석하기 위하여 관측자료 NDVI, SST을 기반으로 ECMWF 모델 변수들로 이용하여 Attention U-Net 모델을 산출한 최고온도자료와 ERA5 최고온도 및 ECMWF 단독 변수 최고온도 자료를 가시화하여 산출하였다.

위성자료인 NDVI(OBS), SST(OBS)와 ECMWF 모델의 OLR, MSL, TMAX변수의 조합으로 입력자료를 구성하고(Figure 3.3) Attention U-Net모델의 학습한 최고온도를 ERA5의 최고온도와 비교한 결과가 ECMWF 모델의 최고온도 단독으로 ERA5와 비교한 결과보다 ACC가 높았다는 것을 볼 수 있다(Figure 3.4). 이 외에 관측자료 NDVI(OBS), SST(OBS)와 ECMWF 모델의 OLR, MSL, TMAX, T2M, SST와의 모든 조합의 민감도 분석에서 단독의 ECMWF 모델의 최고온도보다 ACC가 높은 구성을 보여주지 못한다. 여전히 3주차에 대한 딥러닝 모델의 학습의 어려움을 보여준다.

NDVI(OBS), SST(OBS), OLR, MSL, TMAX, T2M, SST 변수의 조합으로 학습자료를 구성하여 학습한 결과 카테고리 D인 NDVI(OBS), SST(OBS), OLR, MSL, TMAX의 입력변수일 때 가장 ACC가 높다는 것을 확인하였다. 더불어 Q700, Q850, U50, U200, U850, V200, W500, Z200, Z500의 변수를 덧붙여 학습결과를 비교하고자 하였다. 따라서 최종입력자료를 NDVI(OBS), SST(OBS), OLR, MSL, TMAX, T2M, SST, Q700, Q850, U50, U200, U850, V200, W500, Z200, Z500 입력자료를 구성하고 모든 변수에 대한 학습을 진행하였다(Table 3.2).

	입력자료
A	TMAX(ECMWF)
B	NDVI(obs),OLR,MSL
C	NDVI(obs),OLR,MSL,TMAX
D	NDVI(obs),SST(obs),OLR,MSL,TMAX
E	NDVI(obs),SST(obs),OLR,MSL,TMAX,T2M
F	NDVI(obs),SST(obs),OLR,MSL,TMAX,T2M,SST
G(Model)	OLR,MSL,TMAX,T2M,SST
H(Obsevation)	NDVI,SST

Figure 3.3. Datasets of category for wrapper method.

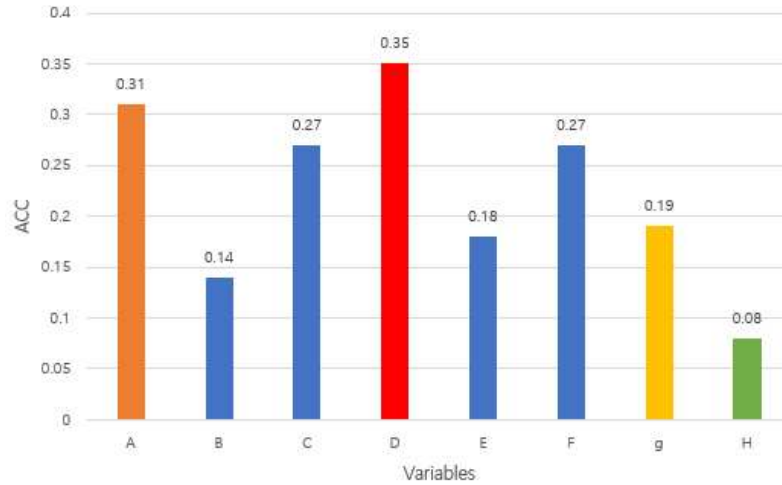


Figure 3.4. Comparison of the result of Attention U-Net by Wrapper method.

Table 3.2. Index and variables of subsets for wrapper method.

Description of subsets for wrapper method	
A	TMAX(ECMWF)
B	NDVI(OBS),OLR,MSL
C	NDVI(OBS),OLR,MSL,TMAX
D	NDVI(OBS),SST(OBS),OLR,MSL,TMAX
E	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M
F	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,TMAX,T2M,SST
G	OLR,MSL,TMAX,SST
H	NDVI(OBS),SST(OBS)
I	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700
J	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850
K	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850,U50
M	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850,U50,U200
N	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850,U50,U200 U850
O	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850,U50,U200 U850,V200
P	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850,U50,U200 U850,V200,W500
Q	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850,U50,U200 U850,V200,W500,Z200
R	NDVI(OBS),SST(OBS),OLR,MSL,TMAX,T2M,SST,Q700,Q850,U50,U200 U850,V200,W500,Z200,Z500

모든 변수의 각각의 A에서 R까지 카테고리에서 ECMWF 단독 최고온도보다 ACC가 높은 경우는 카테고리 D인 NDVI(OBS), SST(OBS), OLR, MSL, TMAX 변수의 조합으로만 학습결과가 높았다(Figure 3.5).

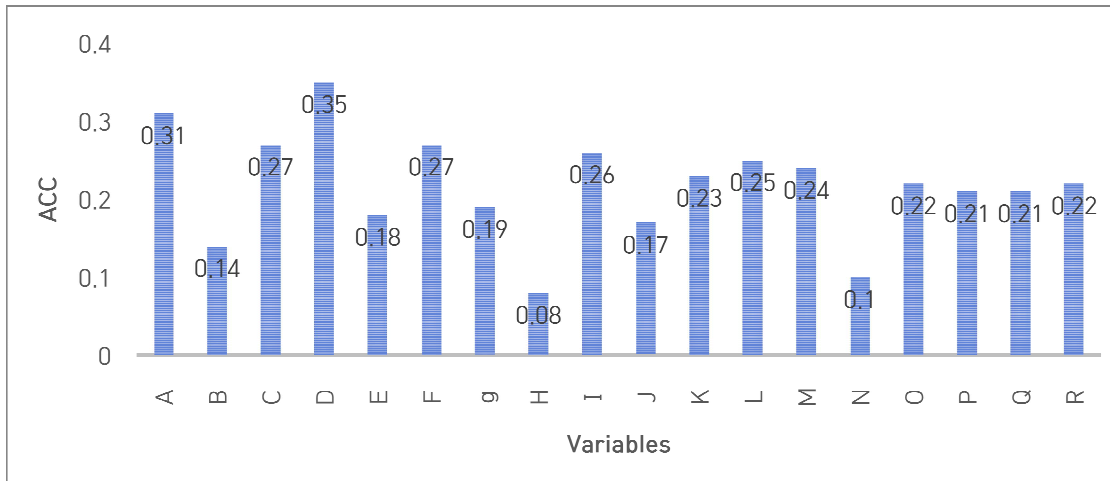


Figure 3.5. Comparison of the result of all variables by Wrapper method.

3.1.4. Embedded 기법에 따른 민감도 분석 결과

Embedded 기법은 ECMWF의 모든 변수를 이용하여 리드타임 3주차 7일에 대한 ERA5의 최고온도에 대해서 회귀식을 이용하여 회귀계수를 산출하였다. 회귀계수중에서 가장 계수가 높은 순서대로 변수를 추출하고 선택하였다. 관측자료인 NDVI와 SST는 관측자료가 없는 지역의 NaN값으로 본 연구에서는 변수에서 제외하였다.

회귀계수 중요도에서 보면 변수중에서 U850, U200, V200, Z500, Z200, TMAX, TMIN, T2M, V850의 변수들의 중요도가 높아서 선택을 하였고 중요 변수들의 모든 조합에 따른 학습 결과를 산출하여 비교하였다(Table 3.3).

Table 3.3. Index and variables of subsets for embedded method.

Description of subsets for embedded method	
A	TMAX(ECMWF)
B	U850,U200
C	U850,U200,V200
D	U850,U200,V200,Z500
E	U850,U200,V200,Z500,Z200
F	U850,U200,V200,Z500,Z200,TMAX
G	U850,U200,V200,Z500,Z200,TMAX,TMIN
H	U850,U200,V200,Z500,Z200,TMAX,TMIN,T2M
I	U850,U200,V200,Z500,Z200,TMAX,TMIN,T2M,V850

비교분석 결과 Filter기법과 같이 ECMWF의 모델 단독 최고온도 ACC값이 가장 좋음을 알 수 있다. 이의 결과는 변수선택과 더불어 딥러닝 모델의 기후분야의 지속적인 발전이 필요한 것으로 판단된다(Figure 3.6).

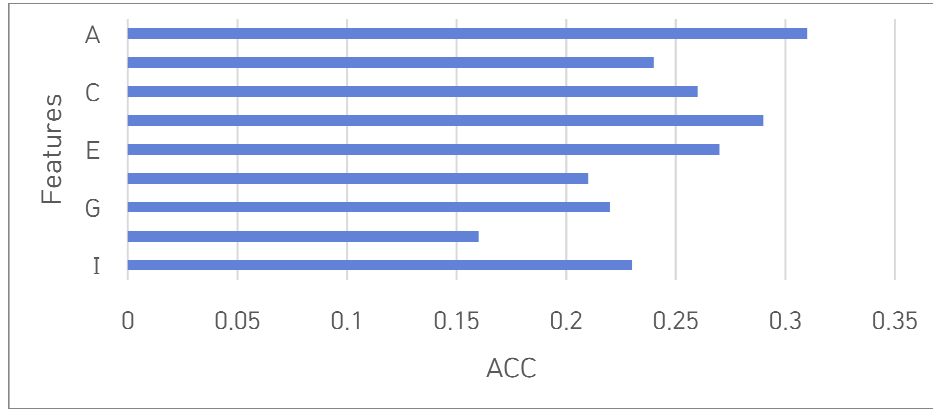


Figure 3.6. Comparison of the result of all variables by Embedded method.

3.2. 입출력 시스템 고도화

3.2.1. 딥러닝 모델 학습자료 산출 시스템

입출력 시스템의 고도화는 학습자료 산출 시스템의 사용자 인터페이스가 고도화 되었음을 보여주며, 학습 데이터 선택, 처리, 및 출력 과정을 직관적으로 사용자가 이용할 수 있도록 최종적으로 시스템을 개발하였다. 주요 기능은 다음과 같다.

- 지역 선택

사용자는 지도 또는 좌표를 통해 특정 지역을 선택할 수 있다

위도와 경도를 입력하거나, 지도를 클릭하여 관심 지역을 설정할 수 있다

- 맵 기반 지역 선택

지도 인터페이스를 활용해 직관적으로 지역을 선택할 수 있다

시각적 가시화를 통해 사용자가 선택한 영역을 확인할 수 있다

- 기간 선택

사용자는 분석할 데이터의 시간 범위를 설정한다

달력 형식으로 제공되어 시작 시점과 종료 날짜를 쉽게 지정할 수 있다

- 관측 자료 선택

변수선택을 통하여 다양한 데이터를 지정할 수 있다

기후모델에서 예측 데이터를 가져올 수 있도록 조건 선택이 가능하다

- 자료 스케일링

데이터를 정규화하거나 표준화하는 등의 스케일링 옵션이 제공된다

데이터 전처리를 통해 학습 모델의 성능을 향상시킬 수 있도록 지원한다

- 학습자료 Split

학습 데이터와 테스트 데이터를 분리하는 비율을 설정할 수 있습니다(예: 80:20).

이는 머신러닝 모델의 학습과 검증에 사용된다

- 가시화

데이터를 시각화하여 선택한 변수와 지역에 대한 정보를 그래프로 확인 가능하다

데이터의 패턴과 이상치를 확인하는 데 유용하다

- 최종 학습 자료 산출

“Generate“ 버튼을 통해 최종적으로 처리된 학습 데이터를 산출한다.

이 데이터는 머신러닝 모델 학습에 바로 사용할 수 있는 형태로 출력된다.

이 시스템은 사용자 중심적인 인터페이스와 다양한 데이터 처리 옵션을 제공하여 학습 자료 산출에 보다 효과적이며, 기후예측 자료를 분석하고 학습 자료로 변환하는 데 최적화되도록 개발되었다(Figure 3.7).

3.2.2. 딥러닝 모델 학습자료 산출 자료 검증

최종 자료는 학습자료 X와 테스트 자료 Y로 산출되며, 데이터 스케일 하기전의 학습자료와 스케일된 학습자료가 산출되도록 개발하였다. 산출자료는 가장 많이 이용되고 있는 사용자 포맷인 NetCDF형식으로 출력이 되도록 개발되었다(Figure 3.8).

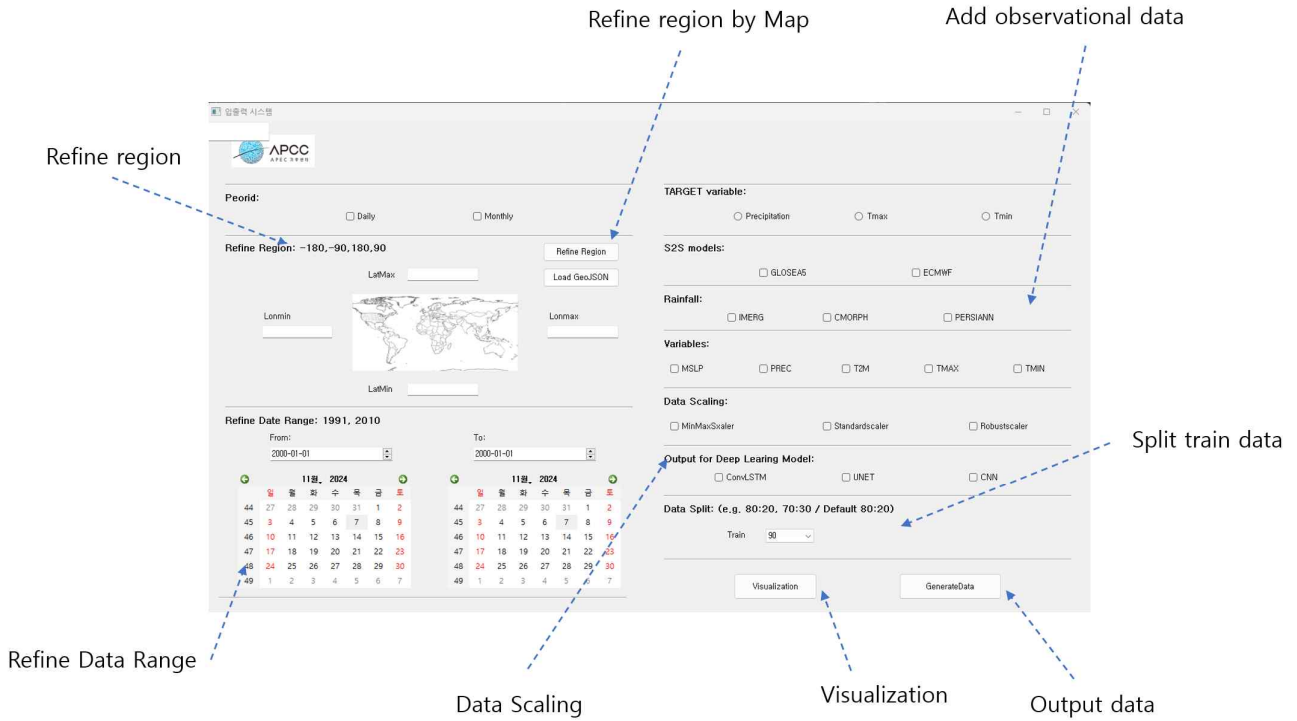


Figure 3.7. Advanced Input/Output System.

이름	수정한 날짜	유형	크기
Train_x_2001_2002.nc	2024-11-01 오후 11:22	NC 파일	390KB
Train_x_scaled_2001_2002.nc	2024-11-01 오후 11:22	NC 파일	769KB
Train_y_2001_2002.nc	2024-11-01 오후 11:22	NC 파일	140KB
fig	2024-11-01 오후 11:11	파일 폴더	

Figure 3.8. The results of advanced input/output system.

산출된 자료가 제공이 되었는지 확인하기 위하여 산출된 자료 NetCDF자료의 메타자료를 이용하여 검증한 결과 초기 선택된 변수와 기간, 지역들이 맞게 산출된 것을 확인할 수 있었다 (Figure 3.9).

```

Variable "msl"
In file "Split_Train_x_2001_2002.nc"
double msl(variable=3, sample=2, time=19, lat=17, lon=34);
: _FillValue = NaN; // double
  
```

Figure 3.9 Meta data for verification of output data by advanced input/output system.

또한 산출된 자료는 NetCDF 뷰어를 통해서 영역선택에 맞게 자료가 산출되었는지(Figure

3.10) 자료가 선택한 스케일링 전처리에 맞추어 생산되었는지 확인한 결과 스케일링 된 자료는 0에서 1로 잘 정규화된 것을 확인할 수 있었다(Figure 3.11).

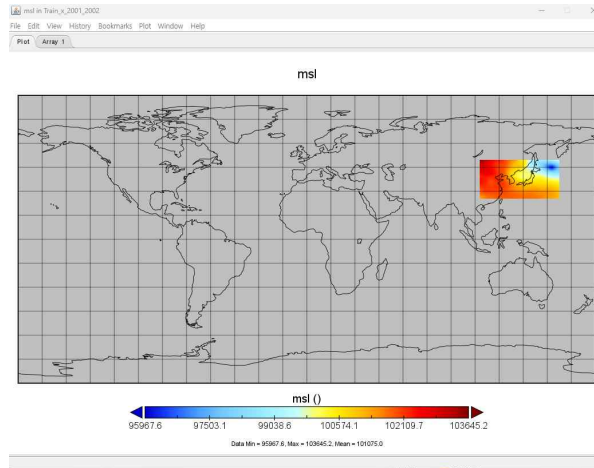


Figure 3.10. Output data by scaled method.

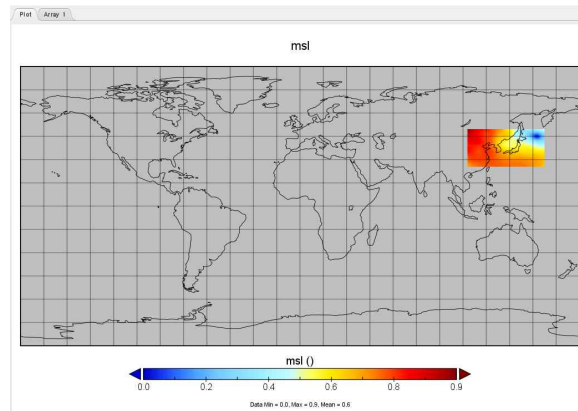


Figure 3.11. Raw data by Advanced input/output system.

최종적으로 데이터 시각화를 위해서 matplotlib, cartopy 라이브러리를 설치하고 이를 이용한 지도 기반 데이터 시각화를 이용하고 사용자 기반의 정의 색상과 레이블을 선택하여 지원 가능하게 가시화 되도록 개발이 완료되었다(Figure 3.12).

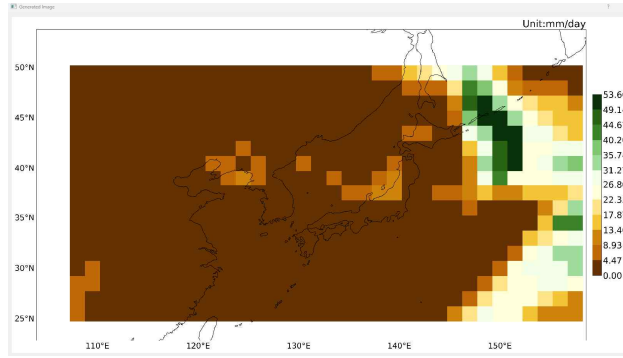


Figure 3.12. Visualization of output data for verification.

입출력 시스템에서 단계적으로 변수를 지정하고 데이터를 선택하면 최종적으로 딥러닝 기반의 입력자료가 산출되는 시스템을 모식화 하였다(Figure 3.13).

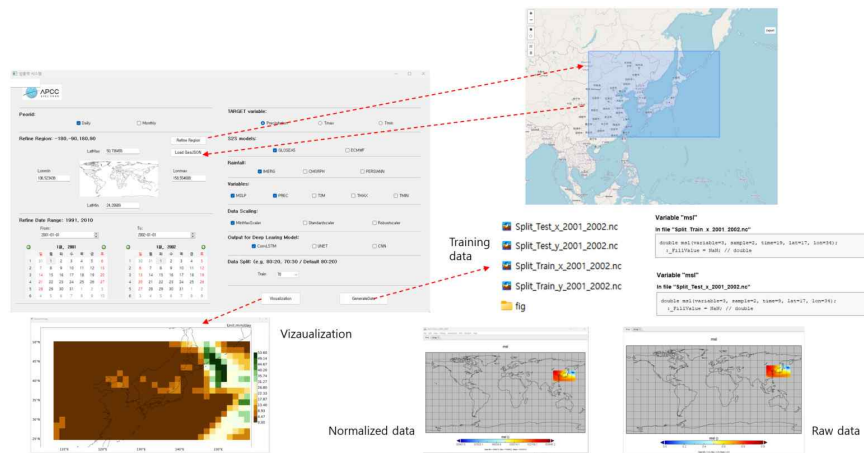


Figure 3.13. Steps of outcome by advanced input/output system.

3.2. 전처리/초매개변수 기반 다중 인공지능 모델 도출 및 모델간 앙상블

본 연구에서 전처리(22년, 6단계 1차년) 및 초매개변수(23년, 6단계 2차년)는 선례 연구 결과에 따라 처리하였다. 예를 들면, Chung et al. (2023)은 S2S 예측 자료의 딥러닝 적용 예측 성능 평가에서 다양한 전처리 기법에 따른 예측 성능 변화를 평가하였는데, S2S 일 최고기온은 표준화 방법에 의한 전처리가 선행예측 3-4주 시간 규모의 개선에 뚜렷하게 기여하는 것을 확인하였다. 또한, 1차년에 개발한 U-Net에 잔차 학습 (Residual block) 혹은 Attention block을 U-Net에 추가한 다양한 확장 U-Net을 구축하고 최적화된 초매개변수로 셋팅된 확장 U-Net의 예측 성능을 평가하였다(APEC기후센터, 2023). Attention block이 추가된 U-Net에서 목표 선행예측 시간 규모의 예측 성능이 개선되었다. 그러나 목표 선행예측 시간 규모 3-4주는 개선됨을 확인하였으나 1-2주에서는 오히려 개선되지 않았다. 그러나 당해연도(24년, 3차년) 연구에서는 선례 연구에서 도출된 S2S 일 최고기온의 예측 성능 개선에 적합한 전처리 기법과 각각의 신

경망 모델의 초매개변수를 기반으로 다양한 앙상블 기법 구축 및 적용을 통해 S2S 일 최고기온에 대한 목표 선행예측 시간 규모(3-4주)의 개선은 물론 미흡했던 1-2주의 예측 성능을 개선하였다.

3.2.1. 4종 앙상블 기법 구축

본 연구에서 구축한 앙상블 기법은 4종이며(Table 2.1), 앙상블 기법에 포함할 멤버 모델, 즉 신경망 모델은 5종인데(Table 2.2) ECMWF에는 5종 모두가 멤버 모델로 참여하였지만 KMA(GloSea5)에는 ResNet을 제외한 4종의 멤버 모델이 참여하였다.

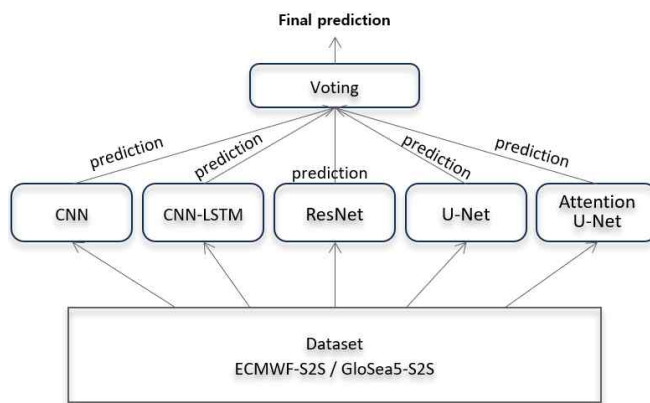


Figure 3.14. Voting technique developed in this study.

먼저, 보팅(Voting)은 서로 다른 신경망 모델(CNN, CNN-LSTM, U-Net, Attention U-Net, ResNet)을 독립적으로 학습시키고, 각각의 예측 결과를 가중합산 방식으로 결합하여 최종 결과를 도출하였다(Figure 3.14).

배깅(Bagging)은 훈련 세트에서 샘플링하는 방식, bootstrap aggregating의 줄임말로, bootstrap 통해 랜덤 샘플링한 훈련자료를 같은 모델(같은 알고리즘)로 학습하고 각 모델의 결과를 투표(voting)를 통해서 결정한다. 그런데, 본 연구에서는 bootstrap 통해 학습 모델 구성은 일반적 배깅과 유사하게 구축하였지만 같은 멤버 모델을 사용하지 않고 다른 멤버 모델, 즉 앞서 보팅에서 활용한 선례 연구의 신경망 모델들을 병렬로 학습시킨 후, 모델 결과를 평균화하였다(Figure 3.15).

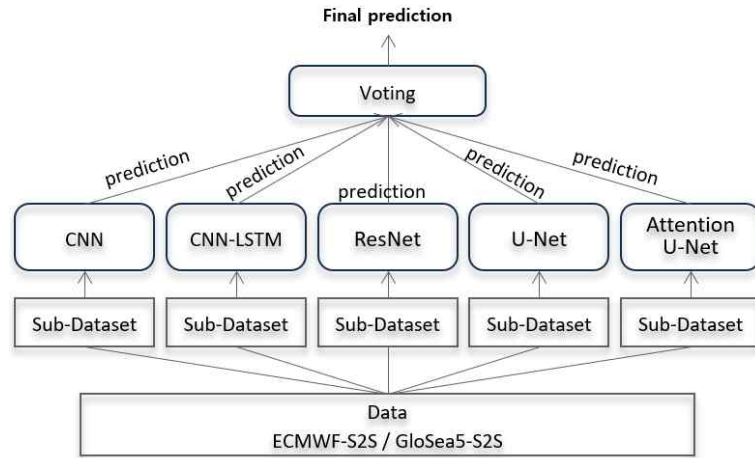


Figure 3.15. Bagging technique developed in this study.

부스팅(Boosting)은 배깅의 변형으로 이전 모델이 예측하지 못한 훈련자료에 대해 가중치를 두어서 다음 모델이 더 잘 학습 할 수 있도록 하는 방법을 통해 bias를 감소시키는 기법이다. 약한 학습기(weak learner), 본 연구에서는 CNN을 가장 약한 신경망 모델이라고 가정하여 CNN-LSTM, ResNet, U-Net, Attention U-Net 순으로 Attention U-Net을 강한 학습기로 가정하여 구성하였다(Figure 3.16).

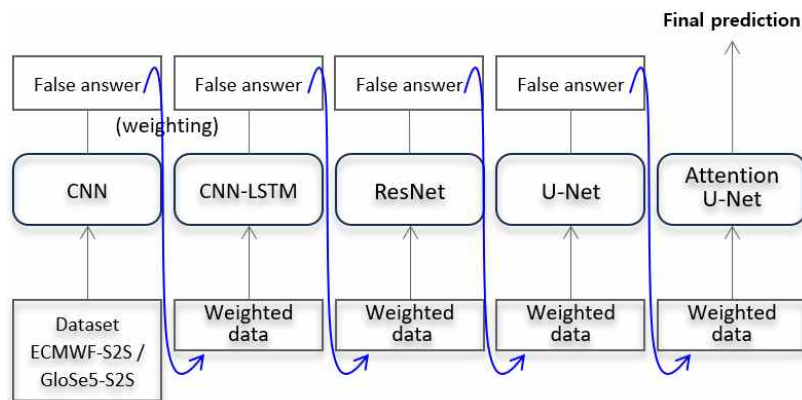


Figure 3.16. Boosting technique developed in this study.

스태킹(Stacking)은 여러 모델들의 학습 결과가 메타 학습을 거치기 때문에 훈련(학습)자료에 한에서 우수한 성능을 보일 수밖에 없다. 그래서 주로 캐글(kaggle) 경진 대회에서 마지막 성능 올리기 위한 편법(치트키)로 쓰이는 기법들 중 하나이다. 그러나 훈련자료와 도메인이 다르거나 분산이 다른 자료가 들어오게 된다면 모델의 예측 혹은 분류 성능이 현저히 떨어지게 된다. 왜냐하면 스택킹은 두 단계, 즉 첫 번째는 n개의 모델을 생성하고, 두 번째 n개의 모델 결과로 메타 학습을 거치게 되므로 훈련 자료에 과적합(overfitting) 될 수 있기 때문이다. 그래서 스택킹의 장점은 단일 모델로 했을 때 보다 성능은 향상되지만, 단점은 과적합이 될 수 있

다. 본 연구의 스택킹 기법에서는 앞서 언급한 멤버 모델의 예측 결과를 메타 모델 (meta-model), 본 연구에서는 Linear regression을 메타 모델로 선정하고 Linear regression에 다른 신경망 모델의 예측(훈련) 결과를 입력하여 최종 예측 결과를 도출했다(Figure 3.17).

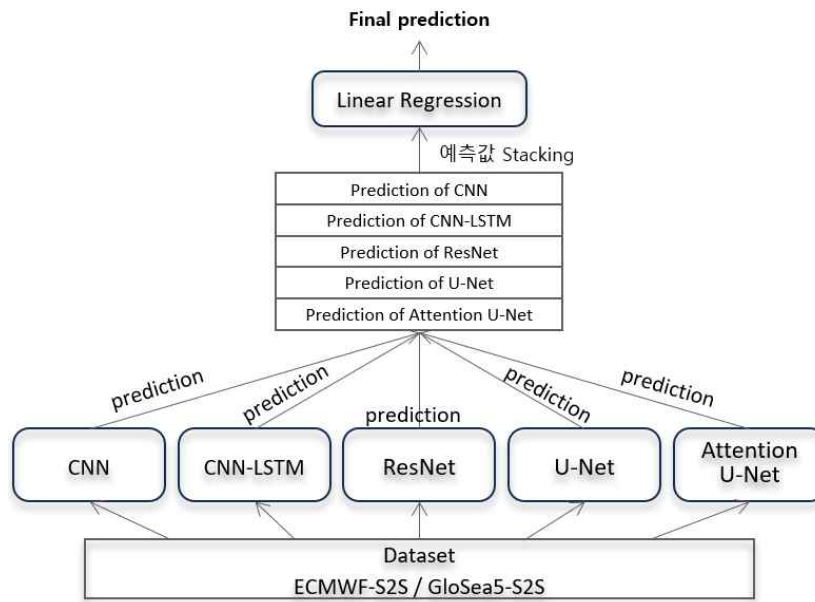


Figure 3.17. Stacking technique developed in this study.

3.2.2. 4종 앙상블 기법의 학습 성능

Figure 3.18와 Figure 3.19은 ECMWF-S2S 및 KMA-S2S 일 최고기온에 대해 앙상블 기법 4종을 적용한 후 전체 예측 기간(예, 46일 혹은 60일) 동안의 편차상관계수(Anomaly Coefficient Correlation, ACC)와 선행예측 시간 규모(1주-4주)의 공간편차상관계수(Pattern Coefficient Correlation)를 평가한 결과이다. ECMWF-S2S의 일 최고기온에서는 앙상블 기법 4종(컬러) 모두 훈련 전 ECMWF-S2S의 ACC(회색)보다 향상되었다(left of Figure 3.18). 스택킹 기법의(파란색) 특징 중 우수한 성능의 데이터를 스캐킹 기법으로 학습하면 우수한 결과가 나오는 특징에 맞게, ECMWF-S2S의 일 최고기온은 현재 S2S 예측 자료 중에서 성능이 우수한 자료이기 때문에 우수한 훈련자료에 앙상블 기법을 적용했기 때문으로 추정된다.

미세한 차이지만 배경이(핑크색) 다른 기법들 보다 더 향상된 것은 첫 번째 일반적 배경에서는 같은 멤버 모델을 구성한다고 하는데, 본 연구에서는 다른 멤버 모델들로 구성된 것과 우수한 ECMWF-S2S의 일 최고기온 예측 자료를 bootstrap으로 랜덤 샘플링한 일부 훈련자료가 학습에 잘 맞아 효과으로 학습된 것으로 판단된다.

마찬가지로 PCC 결과도 ACC와 유사하게 나타났는데 ECMWF-S2S의 일 최고기온에서 앙상블 4종의 PCC는 선례 연구에서 미흡했던 1-2주 예측 성능도 개선되었을 뿐만 아니라 목표

선행예측 3-4주 시간 규모의 예측 성능 개선도 유지되었다(right of Figure 3.18).

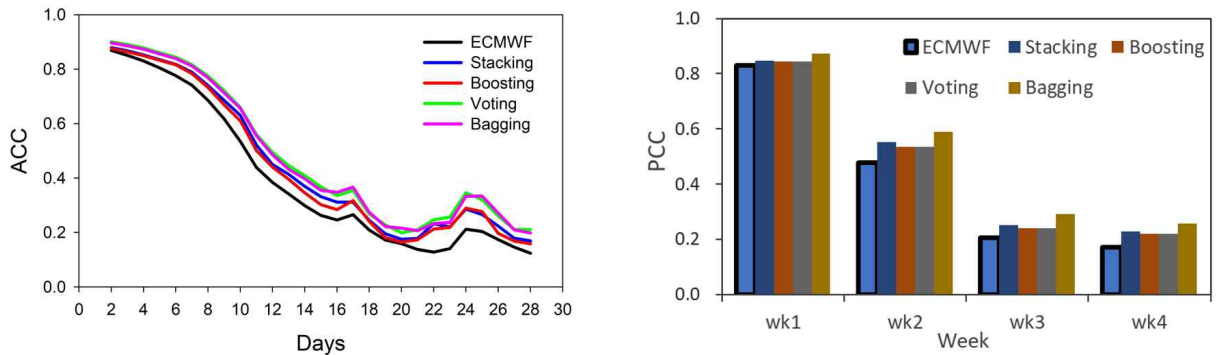


Figure 3.18. Test results of four ensemble techniques (stacking, boosting, voting, bagging) applied to daily maximum temperatures from ECMWF-S2S. The ACC (Anomaly Correlation Coefficient) for the 46-day time scale of the forecast lead-time(left), the PCC (spatial Pattern correlation coefficient) for 1-4 weeks of the forecast lead-time as weekly average(right). ECMWF is ACC and PCC before training.

반면, KMA(GloSea5)-S2S 일 최고기온에서는 스택킹(파랑색)과 부스팅(빨강색) 앙상블 기법은 훈련 전 S2S의 ACC(회색)보다 향상되었지만, 보팅(초록색)과 배깅(핑크색)은 4주(20일 부터)에서만 향상되었다(left of Figure 3.19). 보팅과 배깅을 적용했을 때의 ACC 개선이 미흡한 이유를 추정해보면 멤버 모델들의 보팅 점수가 낮았거나 KMA-S2S 일 최고기온의 랜덤 훈련자료가 학습에 적합할 정도로 충분하게 커버하지 못한, 즉 KMA-S2S 일 최고기온 예측 자료를 일부 훈련자료로 학습할 경우는 학습이 잘 안될 수 있으므로 추정된다.

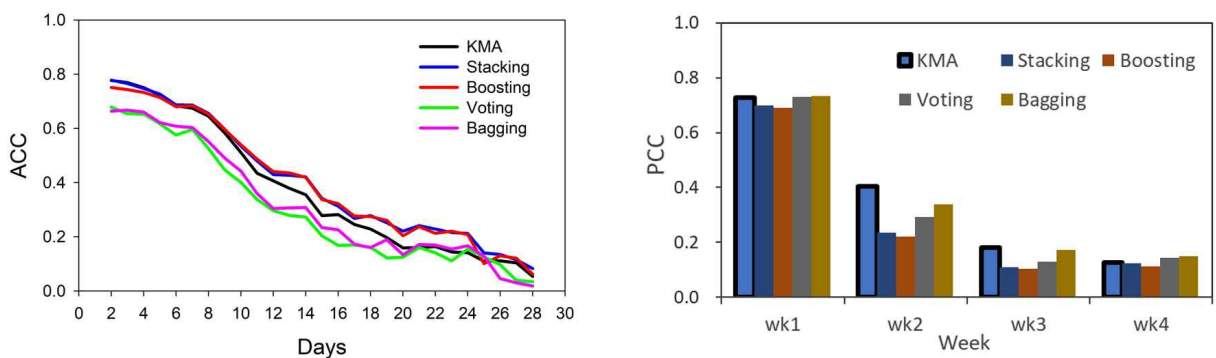


Figure 3.19. Test results of four ensemble techniques (stacking, boosting, voting, bagging) applied to daily maximum temperatures from KMA-S2S. The left is the ACC (Anomaly Correlation Coefficient) for the 46-day time scale of the forecast lead-time, the right is the PCC (spatial Pattern correlation coefficient) for 1-4 weeks of the forecast lead-time as weekly average. KMA is ACC and PCC before training.

Figure 3.19(right)는 KMA-S2S 일 최고기온의 PCC 결과인데, ACC 결과(left of Figure 3.19)와 다르게 나타났으며, ECMWF 결과와도 다르게 나타났다. ACC에서는 보팅과 배킹이 낮게 평가되었는데, PCC에서 보팅과 배킹은 1주에서 개선되었다. 그러나 선행예측 2-3주 시간 규모에서 4종 앙상블 기법은 훈련 전 S2S(하늘색 막대)보다 개선되지 않았으며 4주에서는 보팅과 배킹(스태킹과 부스팅 제외)이 훈련 전 S2S 보다 유사하거나 다시 개선되는 것으로 나타났다. ACC와 PCC 평가 결과의 차이는 시간과 공간의 차이일 것으로 추정되는데 KMA 기후 모델이 공간적으로 영향을 받으며 배킹 기법이 ACC에서는 효과가 크지 않았지만 PCC에서는 개선되는 것으로 나타나서 랜덤 훈련자료가 공간적 영향을 받은 것으로 판단된다.

Figure 3.20과 Figure 3.21은 ECMWF-S2S 일 최고기온에 적용한 4종 앙상블 기법에 대해 ACC를 선행예측 시간 규모 1주에서 4주에 대해 공간적으로 평균하고 훈련 전 ACC 공간 분포와의 차이를 비교한 것이다. 다시 말해서, ECMWF-S2S의 일 최고기온의 훈련 후 선행예측 1주에서 4주에 대해 계산된 ACC 공간 분포에서 각각의 선행예측에 대한 훈련 전 S2S의 ACC 공간 분포를 뺀 차이를 Heat map으로 표현하였다.

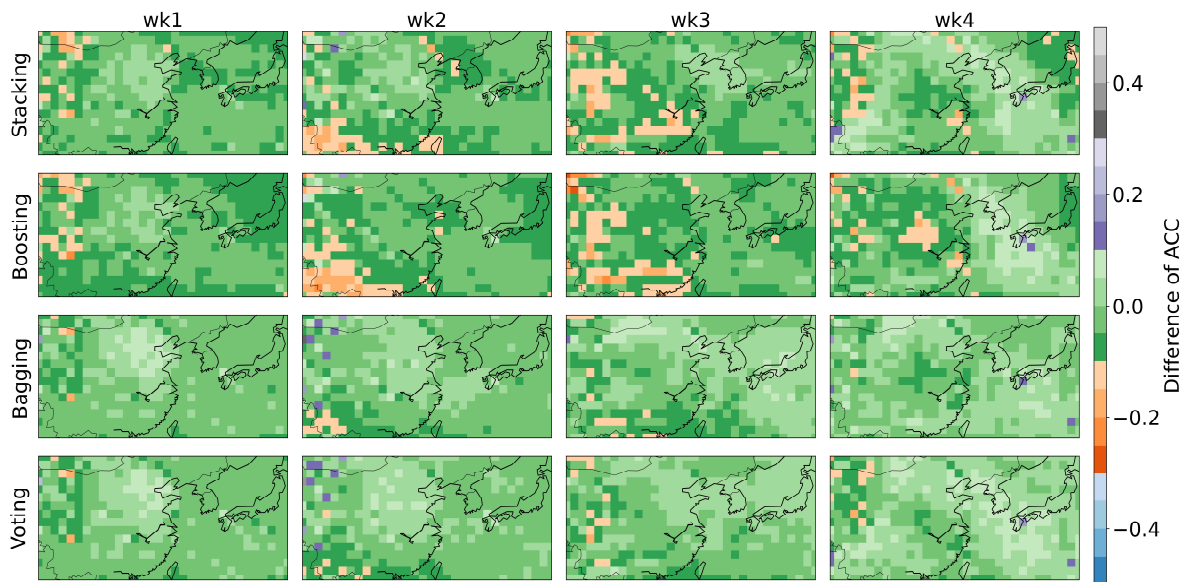


Figure 3.20. For the four ensemble techniques (stacking, boosting, voting, and bagging) applied to the daily maximum temperature of ECMWF-S2S, the spatial distribution of ACC for 1-4 weeks of advance forecasts was calculated, and the difference from the spatial distribution of ACC before training at each time scale was expressed as a heat map.

Figure 3.20은 ECMWF-S2S의 일 최고기온에 적용한 4종의 앙상블 기법에 대한 ACC 공간 분포인데 앞서 본 46일 기간의 ACC 결과와 유사하게 모든 앙상블 기법의 1주에서 4주까지 훈련 전 ACC와 훈련 후 ACC의 차이가 0에 가깝게 나타났으며 보팅과 배킹에서 공간 분포가 훈련 전 ACC와 거의 차이가 없는 것으로 나타났다.

Figure 3.21은 개별 멤버 모델들을 독립적으로 훈련 후 계산한 ACC 공간 분포에서 훈련 전 ACC 공간 분포 차이를 비교한 것이다. 여기에서 선례 연구에서 1-2주 예측이 미흡했던 것이 앙상블 기법에서는 보완된 이유를 어느 정도 설명된다. 다시 말해서, 단일 신경망 모델로는 1-2주 개선이 미흡했지만, U-Net과 CNN-LSTM 및 Attention U-Net에서 개선을 보였기 때문에 상호보완(trade-off)이 발생한 것으로 추정된다. 또한, ECMWF에서 앙상블 기법 4종 모두 개선된 것은 멤버 모델 5개 중에서 3개가 전반적으로 우수한 훈련 및 예측 결과를 보였기 때문으로 사료된다. 따라서, 본 연구에서 ECMWF-S2S의 일 최고기온의 예측 성능 개선을 위해 참여시킨 멤버 모델의 구성이 다양하여 서로 상호보완이 효과적으로 작용했기 때문으로 판단된다.

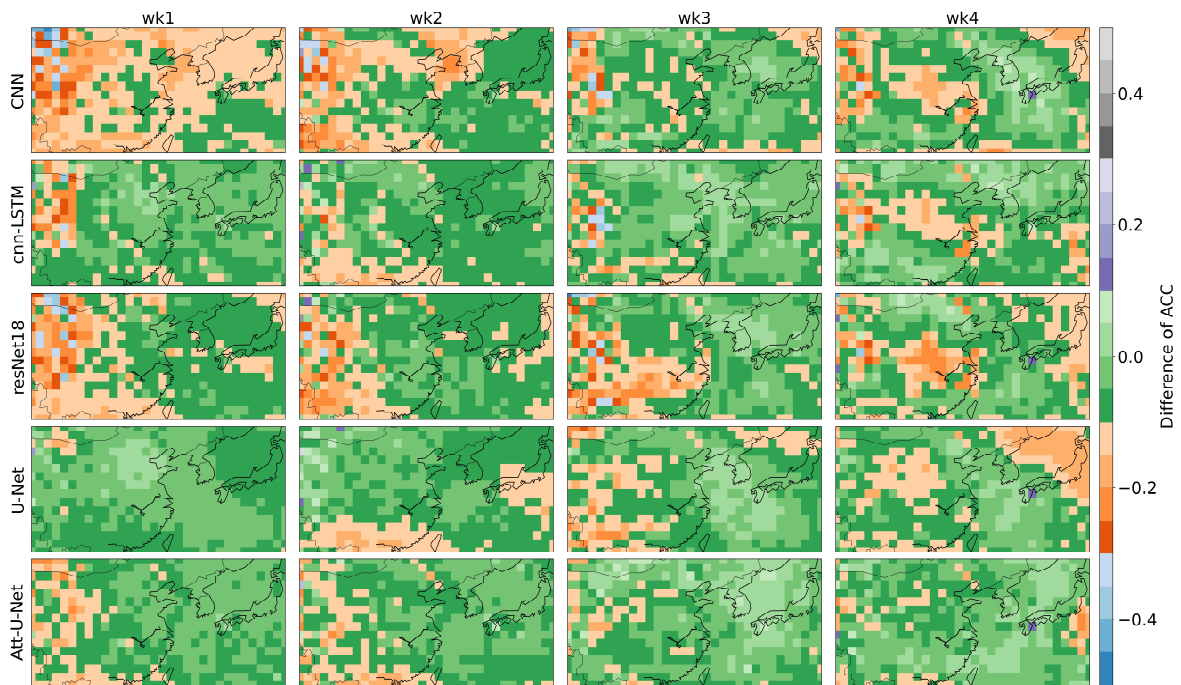


Figure 3.21. For the five member models (CNN, CNN-LSTM, ResNet, U-Net, Attention U-Net) that participated in the ensemble technique of the daily maximum temperature of ECMWF-S2S, the spatial distribution of ACC was calculated from the results predicted for 1-4 weeks after individual training, and the difference in the spatial distribution of ACC before training was expressed as a heat map.

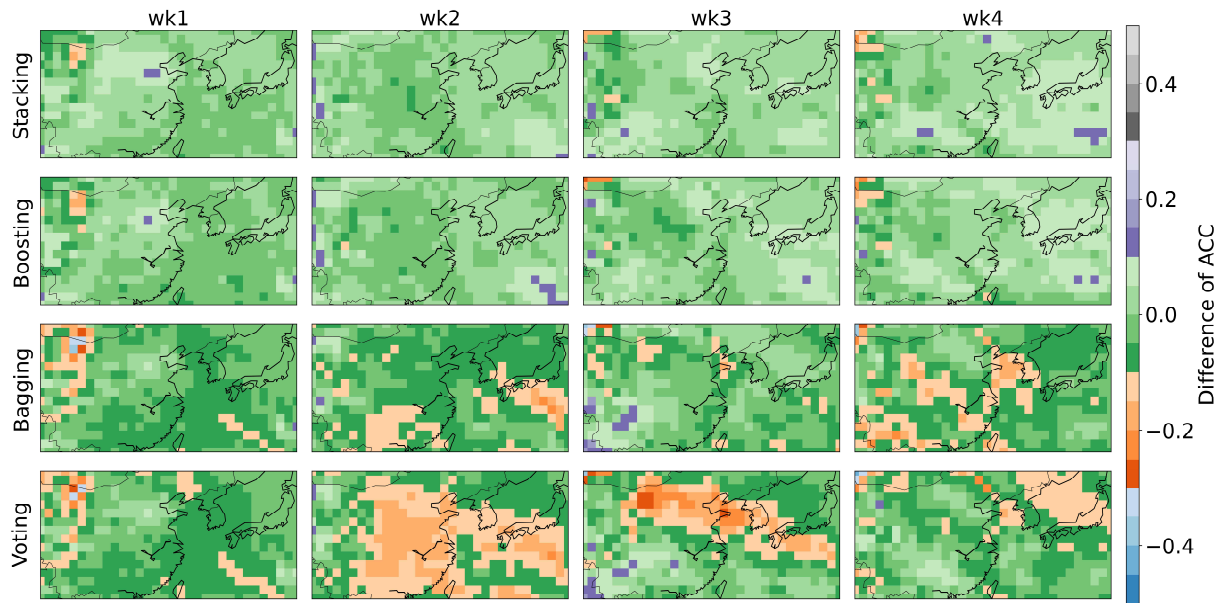


Figure 3.22. For the four ensemble techniques (stacking, boosting, voting, and bagging) applied to the daily maximum temperature of KMA-S2S, the spatial distribution of ACC for 1-4 weeks of advance forecasts was calculated, and the difference from the spatial distribution of ACC before training at each time scale was expressed as a heat map.

Figure 3.22와 Figure 3.23은 KMA-S2S 일 최고기온에 대해 앙상블 기법과 각각의 개별 멤버 모델에 대한 ACC 공간 분포 차이를 비교한 것이다. Figure 3.22는 KMA-S2S의 일 최고기온에 적용한 4종의 앙상블 기법에 대한 ACC 공간 분포인데 앞서 본 60일 기간의 ACC 결과와 유사하게 스택킹과 부스팅에서 1주에서 4주까지 훈련 전 ACC와 훈련 후 ACC의 차이가 0에 가깝게 나타났다. 그리고 보팅과 배깅에서 공간 분포가 훈련 전 ACC는 음의 차이가 관찰되었다. 다만, 앞서 본 PCC는 공간적 편차상관으로 ACC를 공간 분포로 계산한 결과와 유사할 것으로 가정하였지만 ACC 공간 분포도 PCC와 달랐다. 이것은 ACC와 PCC 평가 지수를 활용할 때 보다 유의할 사항으로 판단된다.

Figure 3.23은 KMA-S2S의 일 최고기온을 통해 개별 멤버 모델들을 독립적으로 훈련 후 계산한 ACC 공간 분포에서 훈련 전 ACC 공간 분포 차이를 비교한 것이다. KMA의 경우 ResNet 모델을 제외한 4종의 신경망 모델을 활용하였는데, 1주에서 개선이 관찰되었던 이유는 ECMWF의 이유와 유사한 것으로 판단된다. 그런데 KMA의 경우 2주에서 멤버 모델 중 Attention U-Net을 제외하고 모델의 훈련 후 ACC 공간 분포가 훈련 전 ACC 공간 분포와 많은 차이를 보인 것이 앞서 본 PCC의 2주와 3주에서 급격히 나빠진 이유로 유추할 수 있을 것으로 판단된다. 반면, 4주에서 다시 PCC가 좋아진 것은 그림 20에서 4주의 ACC 공간 분포를 보면 유추할 수 있다. 즉, 각각의 멤버 모델의 4주의 ACC 공간 분포에서 0에 가까운 격자들이 많아진 것으로 확인할 수 있다.

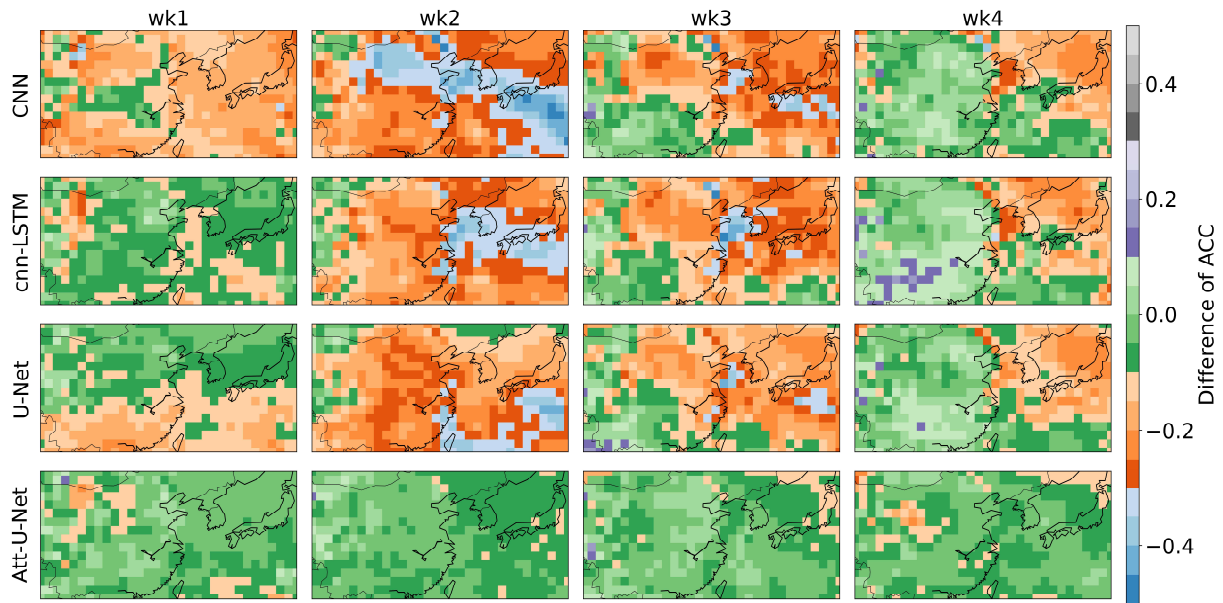


Figure 3.23. For the four member models (CNN, CNN-LSTM, U-Net, Attention U-Net) that participated in the ensemble technique of the daily maximum temperature of KMA-S2S, the spatial distribution of ACC was calculated from the results predicted for 1-4 weeks after individual training, and the difference in the spatial distribution of ACC before training was expressed as a heat map.

3.2.3. 앙상블 기법 내 멤버 모델 조합 변화에 따른 학습 성능

Figure 3.24은 4종의 앙상블 기법에서 멤버 모델의 조합에 따라 PCC의 변화를 비교한 것이다. ECMWF의 경우 참여한 멤버 모델이 5개인데, 5개 멤버 모델이 모두 참여한 경우와 4개 멤버 모델만 참여하되 매회 마다 1개 멤버 모델을 뺀, 예를 들면, CNN, CNN-LSTM, ResNet, U-Net 4개 멤버 모델만 참여시켜서 훈련한 경우와 다음 훈련에서는 CNN을 빼고 이전에서 제외한 Attention U-Net을 참여시켜 훈련하였다. 마찬가지로 KMA도 같은 방법으로 실험하는데, KMA의 경우는 ResNet이 제외되었으므로 4개 멤버 모델이 모두 참여한 경우와 3개 멤버 모델이 참여한 경우로 총 5개의 결과가 도출되었는데, ECMWF의 총 6개의 실험 결과와 KMA의 총 5개의 결과를 각각 합쳐서 비교하였다.

박스 그래프에서 박스의 크기는 PCC 값의 변화를 나타내는데, ECMWF-S2S의 일 최고기온에서(left of Figure 3.24), 스택킹과 부스팅 기법의 박스는 거의 변화가 없고, 보팅과 배깅은 2주에서 살짝 커졌다가 4주에서 다시 커지는 것이 관찰되었다. 전체적으로 ECMWF는 앙상블 기법에서 멤버 조합 변화가 성능, 즉 PCC 변화에 영향을 미치지 않는다는 것을 시사한다. 그러나 기법 간 차이를 비교하면, 스택킹과 부스팅은 멤버 조합의 영향이 크지 않지만 보팅과 배깅은 멤버 조합이 영향을 미칠 수 있음을 알 수 있다.

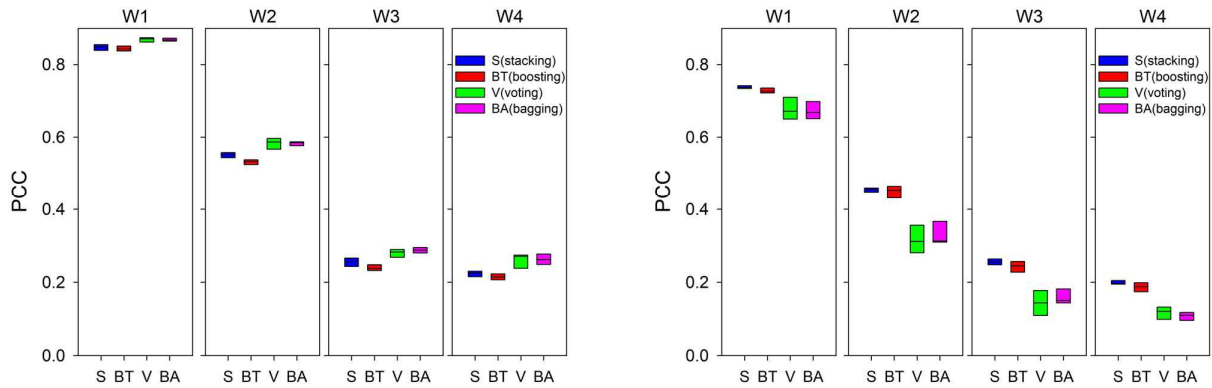


Figure 3.24. Comparison of PCC changes according to member model combinations in four ensemble techniques applied for the postprocessing of daily maximum temperatures of ECMWF-S2S (left) and KMA-S2S (right) for 1-4 weeks of the forecast lead-time.

반면, KMA-S2S의 일 최고기온(right of Figure 3.24)에서는 스택킹은 1주에서 4주까지 멤버 모델 변화에 따른 PCC 변화가 크지 않지만, 부스팅, 보팅, 배깅은 커졌다(박스의 크기가 커짐). 특히, 보팅과 배깅은 1주부터 멤버 모델 변화에 PCC 변화가 큰데, 이것은 앞서 60일 기간 동안 ACC 결과에서 보팅과 배깅이 스택킹 및 부스팅 보다 개선이 낮게 나타난 이유와 관련이 있으며 3.2.2. 섹션의 ACC 공간 분포 결과와도 유사하였다. 그러나 3.2.2 섹션에서 Figure 3.19의 좌의 PCC 결과와는 상반된 결과를 보였다. 이것은 Figure 3.19의 결과는 멤버 모델 전체 즉, KMA에서 4개 멤버 모델이 참여한 경우에 대한 PCC 결과로, 멤버 모델 4개가 모두 기여할 경우 보팅과 배깅이 KMA-S2S의 일 최고기온을 더 개선하였다고 볼 수 있다. 그러나 ECMWF의 경우 멤버 모델 조합 즉 5개 혹은 4개에 대한 영향이 작아 KMA와 다른 결과를 보인 것으로 판단된다.

3.2.4. 격자에 대한 앙상블 기법의 학습 성능

앙상블 기법이 적용된 ECMWF-S2S와 KMA-S2S의 일 최고기온에 대한 ACC 공간 분포에서 남한에 해당된 제주도를 제외한 격자 4지점의 ACC를 추출하여 그 평균 ACC를 비교하였다(Figure 3.25). ECMWF-S2S의 일 최고기온의 경우 앞서 본 결과와 다르게 관찰되었다(left of Figure 3.25). 남한 격자 4지점의 평균을 대상으로 했을 때, 선행예측 1-2주에서 훈련 전(검정색 막대)보다 4종 앙상블 기법의 ACC는 개선되지 않았다. 그러나 3주에서는 부스팅 기법만(빨강색 막대) 제외하고 스택킹, 보팅, 배깅에서는 개선되었고, 4주에서는 모든 기법에서 훈련 전보다 개선되었다. KMA-S2S의 일 최고기온의 경우는 ECMWF와 다른 결과를 보였다(right of Figure 3.25). 보팅과 배깅을 제외하고 남한 격자 4지점의 평균 ACC 평균은 전체 선행예측 1-4주 시간 규모에서 개선되었는데, 개선이 나타난 앙상블 기법은 스택킹과 부스팅으로 앞서 본 결과와 동일한 결과를 보였다.

이것이 의미(시사)하는 점은 전체 및 관심 격자를 함께 고려하기는 어렵다는 것이다. 이미 알려진 발견(finding)으로, 전체적으로 개선되었지만 관심 격자(혹은 지역)로 확대하면, 효과를

보였던 그 기법(들)이 임의 지점에서 효과를 동일하게 보이는 것이 아니라는 것이다. 반대의 경우도 마찬가지이다. KMA와 같이 앙상블 기법을 적용했음에도 ECMWF와 다르게 전체적인 평가에서 선례 연구에서처럼 시간 규모 1-2주에서는 낮은 개선을 보였지만, 남한 격자만을 보면 양호한 개선 결과를 보인 것과 같이 전체적으로는 낮은 개선을 보일 수 있지만 임의 관심 격자에서는 좋은 개선을 보일 수 있다. 반대로 ECMWF는 전체적으로는 개선된 것으로 나타났지만 임의 격자에서는 낮은 개선이 관찰될 수 있다.

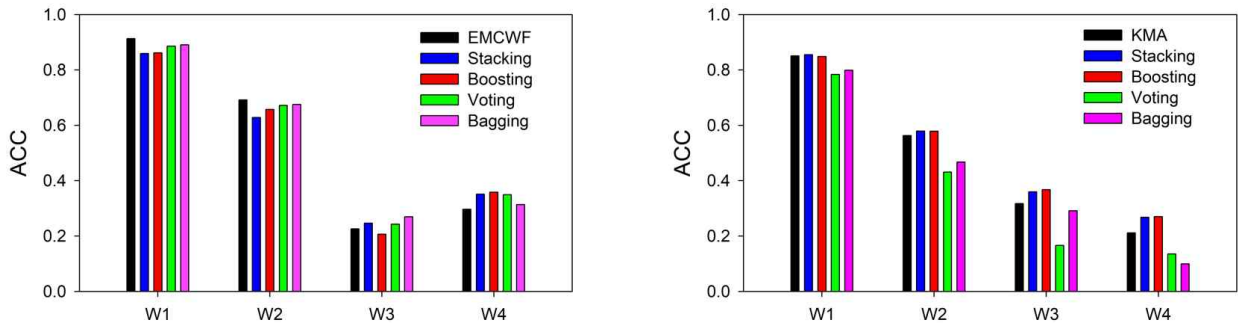


Figure 3.25. Comparison of the average ACC at four grid points in South Korea for 1-4 weeks of the forecast lead-time in the ACC spatial distribution of four ensemble techniques applied for the postprocessing of daily maximum temperatures of ECMWF-S2S (left) and KMA-S2S (right).

3.3. 준지도학습을 통한 MJO 지수 확률예측 인공지능모델 개발

3.3.1. 준지도학습 활용 MJO 위상 분류를 위한 인공지능모델 개발

3.3.1.1. 다양한 변수 조합, 자료 분리 방법, 모델 구조에 따른 모델 민감도 테스트 결과

Figure 3.26은 다양한 변수 조합, 데이터 분할 방법, 모델 구조에 따른 테스트 데이터에서의 지도학습 모델 정확도를 보여준다. 가장 왼쪽 컬럼은 RMM 인덱스, 중간은 VPM 인덱스, 가장 오른쪽 컬럼은 OLR 인덱스 결과이며, 위 행은 시간적 분리 결과, 아래 행은 랜덤 분리 결과를 보여준다. 결과 그래프 내 컬러바는 테스트한 각 모델들(VGG-16, ResNet-18, ResNet-50, WideResNet-50, WideResNet-101)의 MJO 위상 분류 테스트 정확도를 나타내며, 가로 점선(dotted line)은 Martin et al. (2022)에서 제안한 ANN(Artificial Neural Networks) 모델을 구현하여 산출한 정확도 값이며, 가로 파선(dashed line)은 Delaunay and Christensen (2022)에서 제안한 CNN 모델을 구현하여 산출한 정확도 값을 보여준다. 지도학습에서는 ResNet-18 모델이 일관되게 높은 정확도를 기록했지만, VGG-16과 다른 ResNet 모델들은 낮고 불안정한 정확도를 보였다. 앞서 설명한 것처럼 VGG는 ResNet 모델에 비해 성능이 떨어질 것으로 예상됐으며, VGG의 학습 시간은 ResNet보다 훨씬 길었다. 또한 ResNet-18 모델은 더 깊은 ResNet-50이나 더 넓은 WideResNet 모델보다 더 좋은 성능을 보였다. 이는 제한된 양의 학습 데이터에 대해 모델 복잡도를 증가시키기에 따른 과적합 문제에 의한 성능 저하로 보인다. VPM 지수 관련 변수 조합(U850, U200, VP)이 RMM 변수 조합(U850, U200, OLR)보다 전반적으로 더 나은 결과를 보

였다. 이는 VP가 상층 발산을 더 잘 표현하며, MJO와 직접적으로 연관된 대류적 특성을 잘 반영했기 때문이라는 것을 시사한다 (Ventrice et al., 2013). 랜덤 분할이 시간적 분할보다 약간 더 나은 성능을 보였으며, 이는 랜덤 분할된 데이터로 학습할 때 모델 일반화가 향상되어 정확도가 개선된 것으로 판단된다. 독립적으로 구성된 테스트 데이터에서 일관된 성능은 모델이 보지 못한 데이터 분포에서도 합리적인 결과를 낼 수 있음을 보여준다.

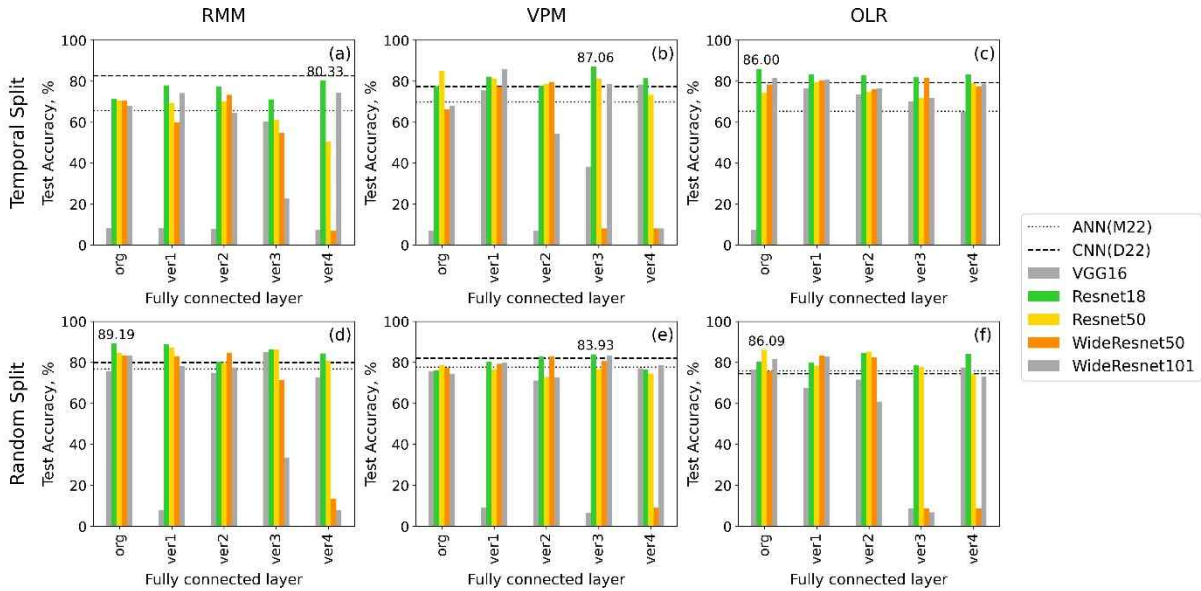


Figure 3.26. Summary of model skill for MJO phase classification for the test data. Colored bars of each panel show the test accuracy (%) of the models of the present study, and the ANN (Martin et al., 2022) and CNN (Delaunay et al., 2022) models are represented by dotted and dashed lines, respectively. Left column panels are the results with RMM input variables (U850, U200, OLR), the center ones with VPM variables (U850, U200, VP), and the right ones with OLR only. The upper row demonstrates the results with temporal split data and the lower one with random split data.

3.3.1.2. 준지도학습 기반 MJO 위상 분류 모델 민감도 테스트 결과

Figure 3.27은 준지도학습 알고리즘의 여러 파라미터에 따른 준지도학습 기반 MJO 위상 분류에 대한 모델의 성능 변화를 보여준다. 실험에 사용된 모델은 ResNet-18 모델이며 U850, U200, OLR 변수의 4500개의 라벨링된 자료 (대략 전체의 57%의 자료가 라벨링됨. 나머지는 비라벨링)에 대해서 훈련되었다. Figure 3.27a는 배치 크기와 labeled 배치 크기 변화에 따른 모델 성능 변화를 보여준다. 본 연구에서는 labeled 배치 크기가 작은 경우에는 전체 배치 크기가 클수록 약간의 성능 개선이 있었다. Labeled 배치 크기가 커질수록 전체 배치 크기에 따라서 성능 변동이 크다. 배치 크기 조합은 Labeled 배치 크기가 16일 때 전체 배치 크기가 64인 경우 가장 높은 모델 성능을 보였다. Figure 3.27b는 Consistency cost weight 파라미터 변화에 따른 모델 성능 변화를 보여준다. 대체로 값이 커질수록 정확도가 올라가는 경향을 보이지만, consistency cost weight가 60일 때 가장 높은 정확도를 보인다. Figure 3.27c는 Consistency

rampup 파라미터 변화에 따른 성능 변화를 보여준다. 전체적으로 큰 성능 차이를 보이지 않지만, 학습 초기의 rampup은 짧을수록 성능이 높아지며 본 연구에서는 3 에포크의 rampup이 가장 높은 성능을 보였다. Figure 3.27d는 EMA decay 파라미터 변화에 따른 성능 변화를 보여준다. 대개 값이 커질수록 정확도가 높아지는 경향이며 본 연구에서는 0.99의 EMA decay에 대한 설정값이 가장 높은 성능을 보였다. Figure 3.27e는 자료 증강 방법에 따른 성능 변화를 보여준다. 자료 증강 방법의 변화는 실험 조건 중에서 성능 변화가 가장 크게 나타났다. 본 연구에서는 대개 Vertical Flip을 적용했을 때 높은 정확도를 보였고, Vertical Flip만 적용한 모델과 Vertical Flip과 Gaussian Blur를 이용한 자료 증강을 함께 적용한 모델이 비슷한 성능을 보였다. 이는 가로로 전파되는 MJO 특성 때문에 가로 방향 데이터 변환은 정확도를 감소시키는 반면, 수직 반전은 성능 향상에 기여한 것으로 판단된다. 이 실험은 목표 변수의 특성에 맞는 데이터 증강 방법을 적용하는 것이 중요하다는 점을 보여준다.

Figure 3.28는 비율에 따라 전체 자료 중 일부를 라벨링한 자료로 활용했을 때의 준지도 학습 기반 MJO 위상 분류 모델의 성능 변화를 보여준다. 가로축은 전체 샘플 중에서 라벨링된 자료의 비율을 보여주며, 세로축은 테스트 자료에 대한 MJO 위상 분류 정확도를 보여준다. 초록색은 RMM 지수, 주황색은 VPM 지수, 회색은 OLR 지수 결과를 보여주며 음영 처리된 부분은 같은 모델 모형을 10개의 다른 초기 웨이트 값으로 산출한 결과의 1*STD 값을 보여주며 음영 지역 중간의 굵은 실선은 10개 모델의 평균값을 보여준다. 파선(dashed line)은 지도학습 기반 ResNet-18 모델의 정확도이며, 1점 쇄선(dash-single dotted line)은 Delaunay and Christensen (2022)의 CNN 모델 정확도, 점선(dotted line)은 Martin et al. (2022)의 ANN 모델 정확도를 보여준다. 대체로 적은 라벨링된 자료를 사용할 경우 정확도가 낮아지는 것을 보였다. 라벨링된 자료의 비율이 높아질수록 모델 성능이 좋아지며 한편 3600개의 라벨링된 자료 결과에서는 OLR과 RMM 지수의 결과가 서로 비슷한 것을 보여준다. 이는 제한된 데이터로도 준지도학습 환경에서 지도학습과 유사한 성능을 낼 수 있음을 의미한다. 향후 가용한 자료가 제한적인 경우 일부 라벨링 데이터만으로도 비슷한 성능을 낼 수 있다는 점에서 중요한 시사점을 제공한다.

Figure 3.29은 월별 및 계절별 MJO 위상 분류에 대한 준지도학습 기반 모델의 테스트 자료에 대한 성능을 보여준다. 왼쪽 컬럼은 각 연도별 각 월의 오분류된 MJO 날수를 보여주며 오른쪽 컬럼은 각 연도별 각 계절의 오분류된 MJO 날수를 보여준다. 위 행은 RMM 지수, 아래 행은 VPM 지수의 결과를 보여준다. 빨간색으로 갈수록 더 많은 오분류를 의미한다. 두 지수 모두 전반적으로 여름철에 오분류된 경우가 많았다. 이는 겨울철 MJO 발생 빈도가 높기 때문에 여름철 정확도가 낮을 것이라는 예상과 일치한다. 계절별 결과에서 연도별로 성능이 다르지만, VPM의 경우 2017년 여름철에 성능 저하가 두드러지며 RMM 지수는 2018-2020년에 더 많은 오분류가 관찰된다. Figure 3.30은 월별 및 계절별 MJO 위상 분류에 대한 inactive와 active의 MJO 사례를 나눈 결과를 보여준다. 위 행은 RMM 인덱스, 아래 행은 VPM 인덱스 결과를 보여준다. 왼쪽 컬럼은 월별 모델에 의해 잘못 분류된 MJO 사례를 inactive와 active MJO 사례로 나누어 나타내었다. 오른쪽 컬럼은 계절별 모델에 의해 잘못 분류된 MJO 사례를 inactive와 active MJO 사례로 나누어 나타내었다. 전체 결과는 8100개의 라벨링된 자료를 사용한 모델의 결과를 나타낸다. 이 결과에서도 마찬가지로 여름철에 성능 저하가 두드러진다. 또한, inactive MJO의 경우 active MJO보다 오류가 높게 나타났다. RMM의 겨울철 성능 저하는 VPM에 비해

상대적으로 낮은 모델 성능 때문일 가능성이 높다.

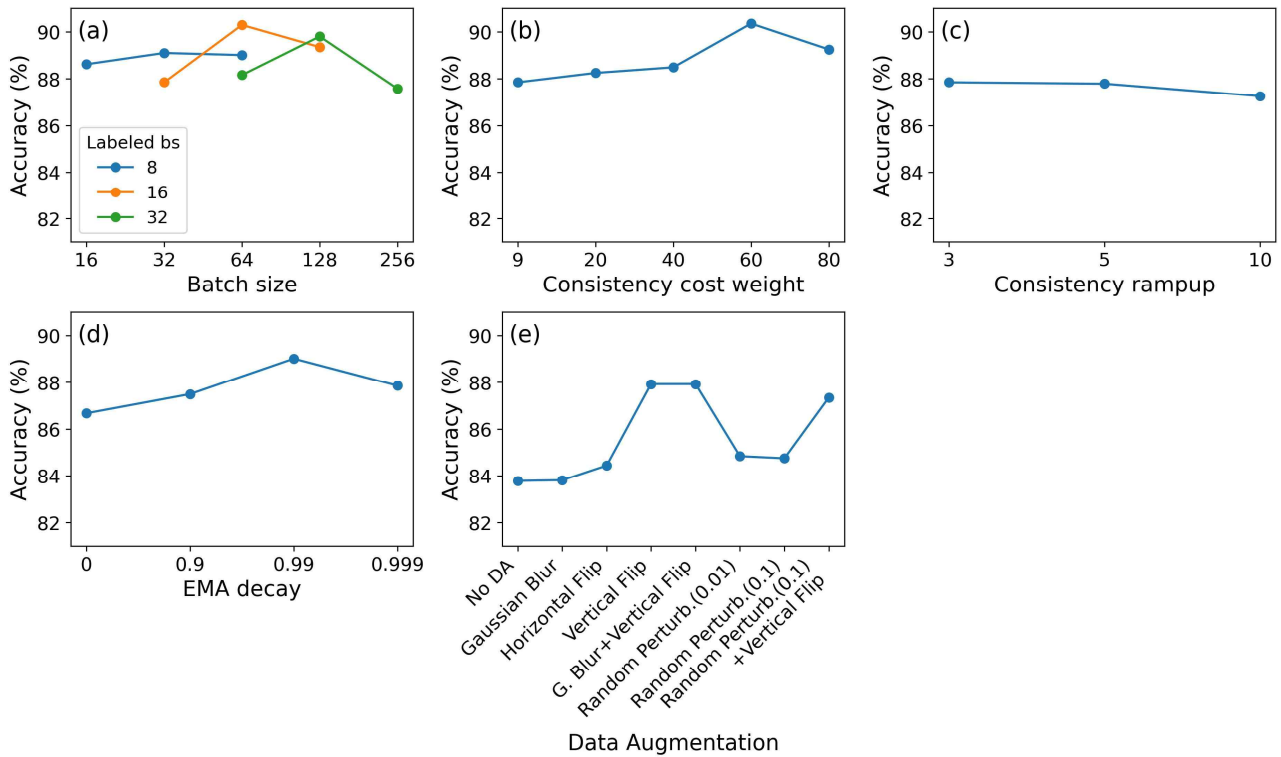


Figure 3.27. Ablation experiment results in terms of (a-d) hyperparameters such as batch size, consistency cost weight, consistency ramp-up, and EMA decay and (e) data augmentation used in the Mean Teacher algorithm in classifying MJO phases in the test data. The ResNet-18 model with the original version of the FC layer was trained on the basis of U850, U200, and OLR with 4500 labeled data.

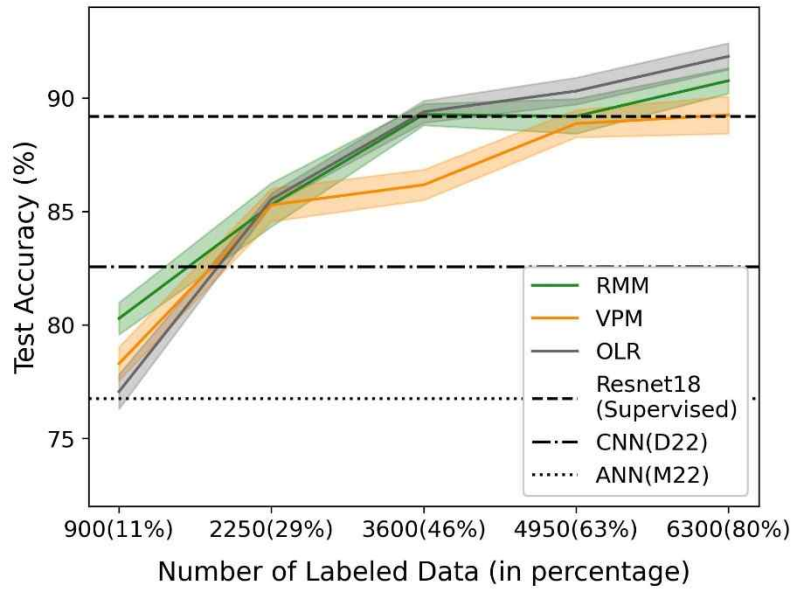


Figure 3.28. Performance of semi-supervised learning models for MJO phase classification on test data with different numbers of labeled data with comparison to a supervised Resnet-18 model using all labeled data and previous studies including ANN from Martin et al., 2022 (ANN(M22)) and CNN from Delaunay and Christensen, 2022 (CNN(D22)). The colored lines show the average accuracy of the 10 independent runs with different random seeds with SSL-based Resnet-18 model for three input fields (RMM, VPM, and OLR). The shaded regions denote the standard deviation of the runs.

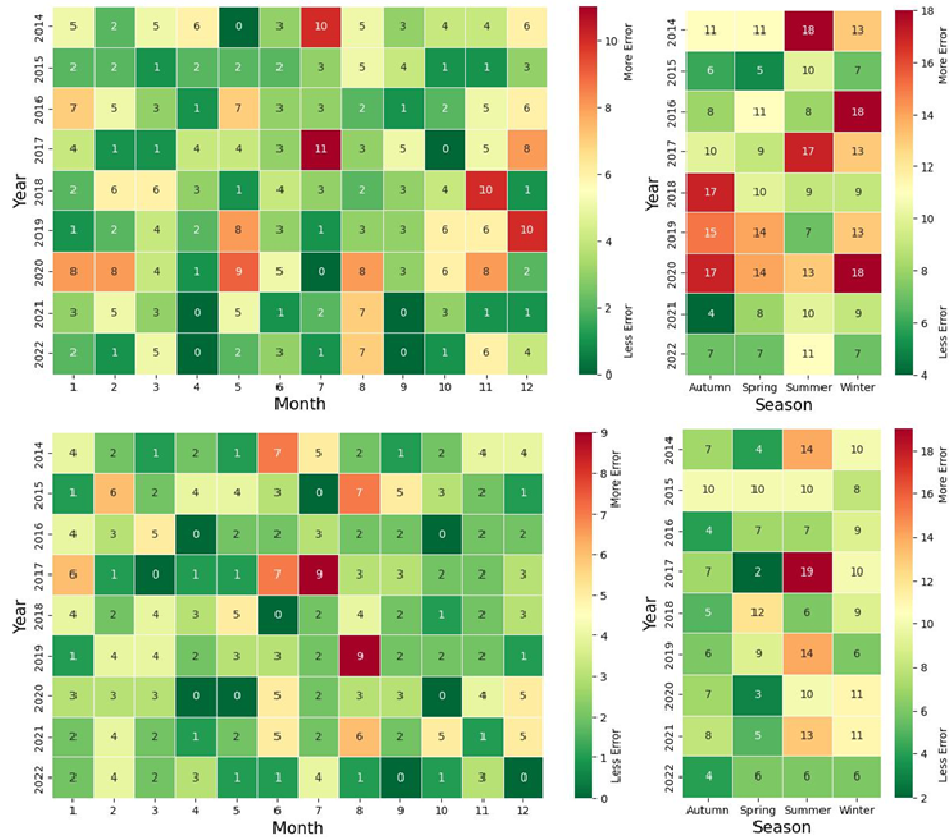


Figure 3.29. Monthly and seasonal SSL-based model performance heat maps with the test data for MJO phase classification for (upper row) RMM and (lower row) VPM index. The number of errors in each grid shows the MJO events misclassified by the model.

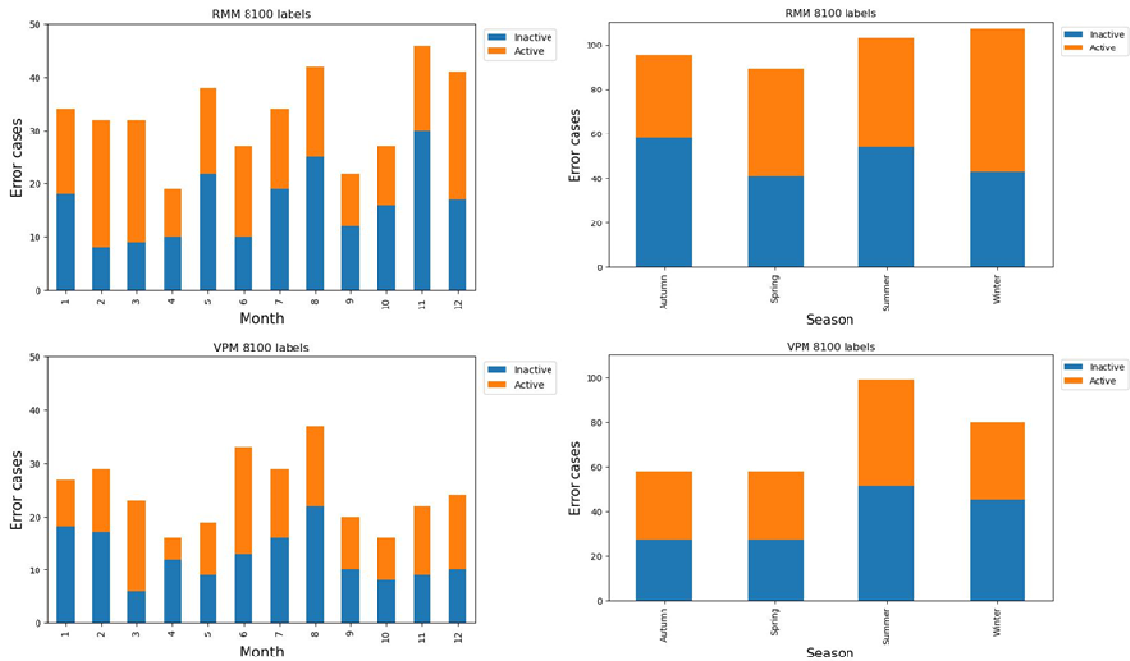


Figure 3.30. Monthly and seasonal SSL-based model performances with the test data for MJO phase classification grouped by active (orange) and inactive (blue) MJO cases. Error cases on the y-axis indicate MJO events misclassified by the model.

3.3.2. 선행시간별 MJO RMM 지수 예측을 위한 인공지능모델 개발

3.3.2.1. MJO RMM 지수 예측 인공지능모델 성능 평가

준지도학습 기반 MJO 위상 분류 연구에서 선행시간별 MJO RMM 지수 예측으로 인공지능 모델을 확장하고 개발하였다. Figure 3.31는 준지도학습 기반 모델(SSL)과 지도학습 모델들(Supervised)을 비교하였다. 가로축은 0일부터 5일 간격의 선행시간이며 세로축은 테스트 자료에 대한 모델의 BCOR 정확도를 나타낸다. 전체 모델은 선행시간이 길어질수록 정확도가 낮아지는 것을 보인다. 준지도학습 기반 모델 간에는 라벨링된 자료의 비율에 따라 성능 차이가 크지 않는 것을 보인다. 하지만 예상대로 전체 자료를 라벨링한 자료로 활용한 지도학습 모델이 준지도학습 기반 모델에 비해 더 높은 정확도를 보였다. 따라서 이후로는 지도학습을 기반으로 한 선행시간별 MJO RMM 지수 예측 인공지능 모델을 구축하고 나아가 RMM 확률 예측 결과에 집중하였다.

선행시간별 MJO RMM 지수 예측 인공지능 모델의 최종 모델은 여러 기후 자료(상하층 동서 바람장, OLR, VP, TS, TCWV, hadvect 포함)의 편차(Anomaly) 및 배경장(Background) 정보를 포함한 14개의 입력 변수를 사용하여 RMM 지수를 예측하도록 설계하여 학습되었다. 네 가지 실험(All, Raw, Background, MJO Anomaly)은 원시(Raw), 배경(Background) 및 편차(Anomaly) 데이터의 조합을 달리하여 수행되었다. Figure 3.32은 각 실험에 대한 선행시간별 모델 성능 결과를 보여준다. 가로축은 0일부터 30일까지 5일 간격의 선행시간이며 세로축은 테스트 자료에 대한 BCOR 정확도를 보여준다. All 실험은 사용된 입력 변수의 편차와 배경장 변수를 함께

쓴 모델 결과, Raw 실험은 입력 변수의 Raw 자료만 사용한 모델, Background 실험은 입력 변수의 배경장 자료만 사용한 모델, MJO Anomaly 실험은 입력 변수의 편차 자료만 사용한 모델의 결과이다. All 실험은 Raw, Background, MJO Anomaly 실험에 비해 예측 성능을 크게 향상시켰다. All 실험 모델은 다른 실험들과 비교해 22일 이후까지 예측 성능을 유지했다. 배경장 자료만 사용한 Background 실험 모델은 MJO 시그널을 탐지할 만한 충분한 정보가 없을 것이며 결과적으로 다른 모델에 비해 크게 예측력 떨어진다. 원시 데이터만 사용하는 Raw 실험 결과는 All 실험에 비해 예측력이 떨어지는데 이는 편차와 배경장 상태를 분리하여 함께 사용하는 것이 MJO 예측에 중요한 정보가 더 효과적으로 학습이 되는 것을 의미한다.

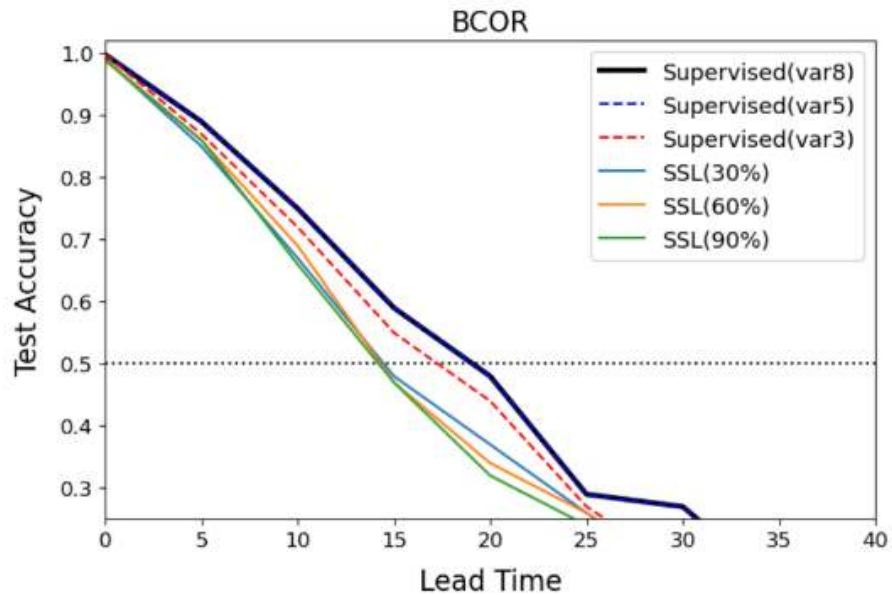


Figure 3.31. Comparisons of semi-supervised learning (SSL)-based and supervised learning-based models in predicting MJO with BCOR on test data over lead times with a 5-day interval.

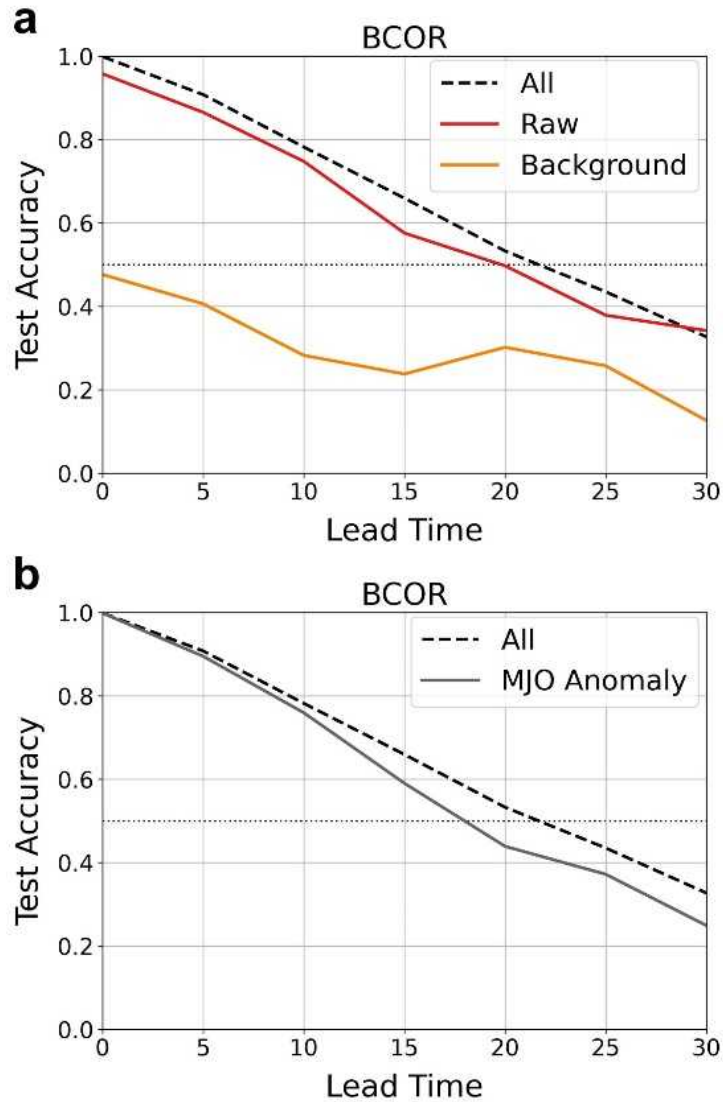


Figure 3.32. Summary of the model skill (BCOR) on test data in predicting MJO over lead times with a 5-day interval for different predictor sets. In a and b, All (black dashed line) means using all input predictors including both MJO anomaly and Background predictors. a, Raw (red solid line) denotes using only the original raw predictors and Background (yellow solid line) using only the raw ones from which the anomaly was subtracted. b, MJO Anomaly (gray solid line) indicates using only the filtered predictors.

Figure 3.33-34는 겨울(11월-4월)과 여름(5월-10월)으로 나눈 계절에 따른 모델 성능을 비교한 결과를 보여준다. Figure 3.33는 편차 및 배경장 변수를 같이 사용한 최종 All 실험 모델의 결과이며, Figure 3.34는 편차 변수만 사용한 MJO Anomaly 실험 결과이다. All 실험에서 예측 성능은 두 시즌 모두 안정적이고 계절 간의 성능 차이가 크지 않은 것으로 확인된다. 이는

편차 및 배경 상태를 모두 반영한 모델이 계절적 변화와 무관하게 학습할 수 있음을 시사한다. 계절적 역학은 MJO의 전파와 빈도에 영향을 미치지만, 편차 및 배경 상태를 모두 포함한 통합 모델 접근법이 두 시즌 모두 효과적으로 작용하는 것으로 나타났다. 이는 계절적인 특성에 따라 계절별 모델을 개발하는 기존의 연구들과 달리 전 계절에 통합적으로 사용할 수 있는 모델로 작동할 수 있음을 의미할 수 있다. 반면 Figure 3.34에서 보이는 것처럼 배경장 상태를 제외한 경우 계절 간 성능 차이가 더 뚜렷하게 나타났고, 이는 배경 상태를 포함하는 것이 정확한 예측에 중요한 역할을 한다는 것을 보여준다.

Figure 3.35은 보통의 MJO와 강한 MJO 사례의 모델 성능 차이를 보여준다. 예상대로 모델은 강한 MJO 사례를 예측할 때 더 높은 정확도를 보였으며, 이는 강한 MJO가 더 구조와 패턴이 뚜렷하여 강한 시그널을 나타내기 때문에 모델이 이를 더 잘 포착할 수 있기 때문인 것으로 판단된다. 강한 MJO는 또한 동쪽으로 더 잘 전파됨으로써 이러한 특성은 모델이 MJO 예측을 장기적으로 더욱 정확히 할 수 있도록 도와준다. 반면, 중간 정도의 약한 MJO는 신호 대 잡음 비율이 낮아 예측이 어려운 경우가 많은 것으로 판단된다. 이는 강한 MJO에 비해서 약한 MJO는 그 구조가 뚜렷하지 않고 시그널이 약하기 때문에 예측이 더 어렵다.

Figure 3.36은 선행시간에 따른 MJO 진폭 감소 문제를 해결하기 위해 개발한 맞춤형 손실 함수(CLF)의 적용 전후 모델 성능 결과 비교를 보여준다. 이 함수는 MJO 진폭과 RMM 값을 동시에 최적화하는 방식으로, 손실함수에 진폭 오류 항을 추가하는 방식으로 구성되었다. 그림에서 가로축은 0일에서 30일까지의 선행시간을 보여주며 결과는 1일 간격으로 생산되었다. 세로축은 예측된 RMM 값으로 산출된 MJO 진폭을 예측된 값에서 실제 값을 뺀 에러를 나타낸다. Original은 CLF를 적용하지 않은 RMM 값에 대해서만 손실이 계산된 모델이며, Modified는 CLF를 적용한 모델의 결과이다. Figure 3.36a에서 CLF를 적용하는 것이 전반적으로 MJO 진폭 오류를 줄이는 데 효과적인 것을 알 수 있다. 기존 연구와 달리 MJO 예측에서 진폭을 최적화하는 것이 모델 성능을 향상시키는 것을 보여준다. Figure 3.36b는 CLF 전후 MJO 위상에 대한 에러 결과를 보여준다. CLF의 적용은 MJO 위상 오류를 약간 개선했지만, 주요한 효과는 진폭 오류 감소에서 더 뚜렷하게 나타난다.

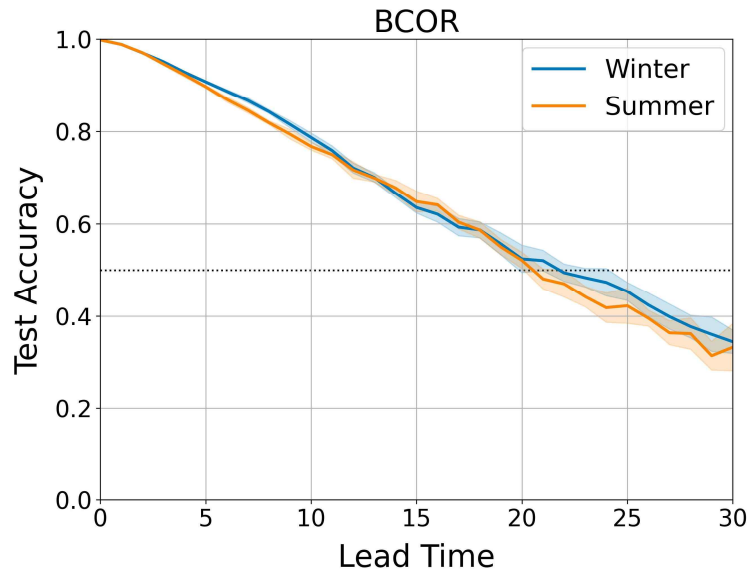


Figure 3.33. Model performance of bivariate correlation coefficients (BCOR) on test data for winter (November–April) and summer (May–October).

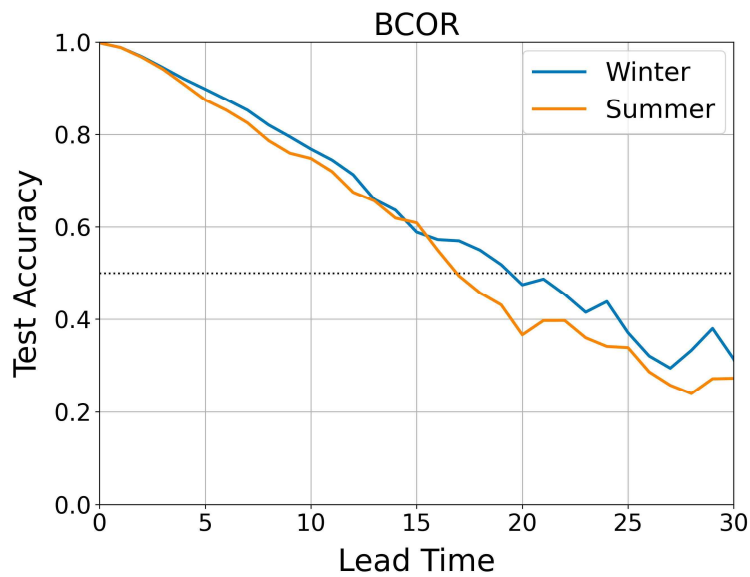


Figure 3.34. Model performance of bivariate correlation coefficients (BCOR) on test data for winter (November–April) and summer (May–October) using the ResNet-18 model trained on only MJO anomaly input variables.

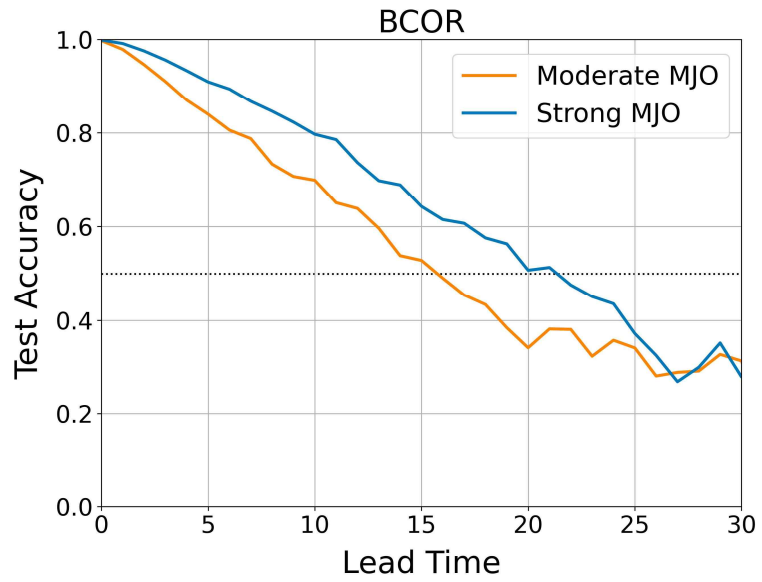


Figure 3.35. Model performance of bivariate correlation coefficients (BCOR) on test data for moderate (MJO amplitude 1-1.5) and strong MJO events (amplitude > 1.5).

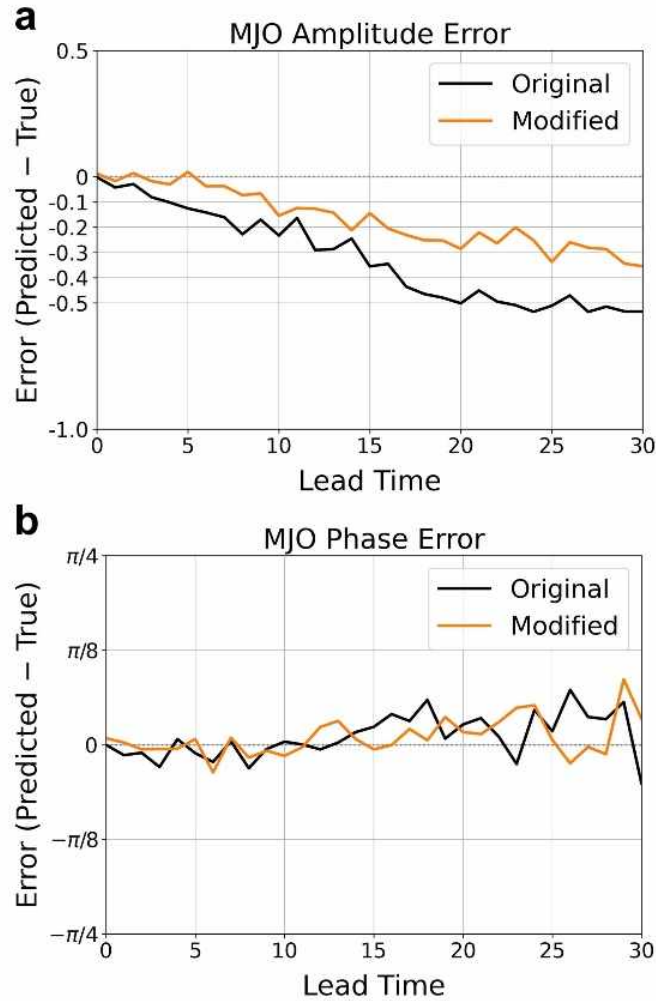


Figure 3.36. Errors of MJO amplitude (a) and phase (b) at lead times when applying customized loss function. The error is obtained by subtracting the predicted values from the model and the true values. Original shows the baseline model with no modifications on model architecture and loss function, and Modified the finalized model of the present study.

3.3.2.2. MJO RMM 지수 예측 인공지능 모델의 해석

선행시간별 MJO RMM 지수 예측에 대한 각 변수의 기여도를 확인하기 위해 Occlusion 민감도 맵을 사용했다. Occlusion 민감도 맵은 3×3 영역(window 또는 kernel)으로 입력 이미지를 훑어가면서 기준값으로 대체해 모델 출력 변화량을 보여준다. 본연구에서는 대체할 기준값을 각 변수의 원 값에서 표준편차를 뺀 값으로 설정했다. 즉 항상 더 낮은 값으로 대체되는 것을 의미한다. 양의 기여도(빨간 지역)는 해당 영역이 대체되면 RMM 값이 감소함을, 음의 기여도(검은색 지역)는 RMM 값이 증가함을 나타낸다. 예를 들어, 낮은 OLR 값(강한 대류와 연관)이

대체되면 음의 기여도가 나타나 더 강한 MJO 활동을 의미한다. 마찬가지로, 활발한 대류 지역에서 TCWV가 더 낮은 값으로 대체되면 양의 기여도가 나타나는 것으로 해석한다.

Figure 3.37은 선행시간에 따른 각 변수의 상대적 기여도를 보여준다. 위쪽 행은 사용된 입력 변수의 편차(prime)를 나타내며 아래는 배경장(overbar)을 나타낸다. 각 세로 바는 각 선행시간에 따라 다르며, 세로축의 값은 RMM 예측에 대한 상대적인 기여도를 나타낸다. 그림의 결과는 RMM1과 RMM2의 평균 결과이다. 0일(선행 없음; 남색 바)일 때는 예상대로 U200', U850', OLR'의 기여도가 가장 높았다. 이는 실시간의 대기 상태는 이들 변수에서 가장 잘 보이기 때문이다. TCWV'는 초기에는 낮은 기여도를 보이지만, 이후 선행시간이 길어질수록 가장 중요한 변수로 확인된다. TS'는 리드 타임이 길어질수록 기여도가 커지며, 변화가 적은 SST 같은 변수가 장기 예측 성능을 높인다는 기존 연구를 뒷받침한다. 전체적으로 배경장 변수는 편차 변수에 비해 기여도가 적지만 (y축의 범위가 작음), U850, OLR, TCWV, hadvect의 배경장의 기여도가 일부 MJO 예측에 기여하는 것으로 보인다. 이는 선행시간에 따라 그 기여도 차이를 보이는데 추가적인 MJO 예측성 원천은 입력 변수의 배경장으로부터 기인할 수 있다. OLR, U850의 배경장은 저층 수분 이류와 관련되어 MJO 예측에 기여할 수 있으며 이러한 특성은 hadvect 배경장 변수에 반영되어 선행시간이 늘어남에 따라 hadvect 배경장 변수의 중요도가 증가하는 것을 확인할 수 있다.

Figures 3.38는 RMM1 예측에 대한 15일 선행시간에서의 각 입력 변수의 기여도 맵을 보여준다. RMM1은 MC(maritime continents) 지역에서 convection peak가 보이며, RMM2는 WP(western pacific) 지역에서 convection peak가 보이는 것으로 알려져 있다. 이러한 사실을 기반으로 15일 선행시간을 고려하면, OLR과 TCWV anomaly 변수가 RMM1에 대해서는 Indian Ocean에서 active convection을 보이고 (즉, 15일 전 Anomaly의 convection 중심) WP에서 suppression이 있음을 알 수 있다. 이는 잘 알려진 MC 지역 convection의 선행조건과 MJO의 전과 역학과 관련된다. OLR Anomaly의 인도양에서의 강한 대류 지역(파란색 동그라미; 검은 지역; 음의 기여도)은 U200 Anomaly 변수에서의 발산과 U850 Anomaly 변수에서의 수렴과 일치한다. TS Anomaly 변수에서는 호주 북부의 티모르해(Timor Sea)에서의 SST 증가는 표면 잠열 유량(latent heat flux) 증가와 관련되어 이는 수증기 공급 증가로 연결될 수 있다 (빨간색 동그라미). 이는 호주 방향으로의 남동향의 하층 수증기 수송과 관련되어 TCWV Anomaly 변수에서 강한 대류 지역(빨간색 화살표)과 일치한다. U850 배경장 변수에서 호주 북부지역에서 음의 기여도(동풍)를 보이며 이는 OLR 배경장 변수에서 인도양에서의 음의 기여도(Wet)와 MC 지역에서의 양의 기여도(Dry)와 관련된다. 이는 TCWV 배경장에서의 인도양의 양의 기여도(Wet)과 MC 지역의 음의 기여도(Dry) 패턴과 일치한다. U850과 OLR 배경장 변수는 TS 배경장에서의 호주 내륙 가열과 관련될 수 있다. 호주 내륙이 가열됨에 따라 하층대기 순환이 강화될 수 있고 남북방향으로의 수증기 수송이 강화될 수 있다. 이러한 특성은 V 성분의 바람장에서 더 잘 보일 수 있을 수 있으며 향후 연구에서는 V 성분 남북 바람장을 활용하는 것이 도움이 될 수 있을 것으로 보인다. Figure 3.39은 MJO RMM 예측에 주요 변수인 TCWV anomaly와 U850 배경장을 함께 보여주고 있다. 선행시간이 지남에 따라서 TCWV에서 active convection 지역이(빨간색 지역) 점차 왼쪽으로 이동하는 것을 잘 보여주고 있다. U850 배경장은 RMM1에 대해서는 MC 지역에서 하층 바람장의 역할을 잘 보여주고 있다.

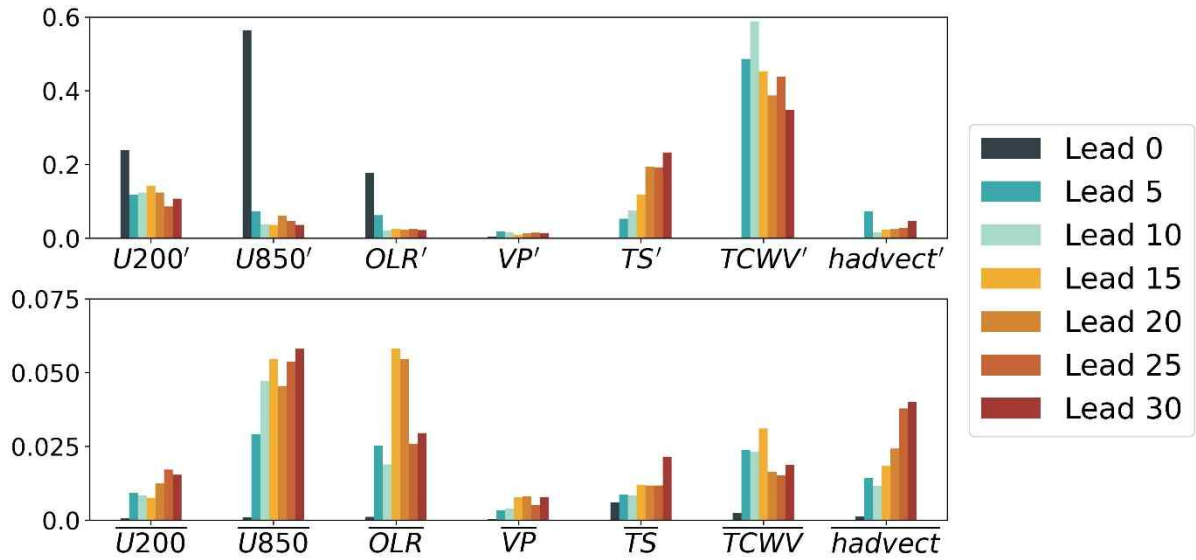


Figure 3.37. Relative contributions of input features for both RMM1 and RMM2 by simple spatial average over lead times with a 5-day interval.

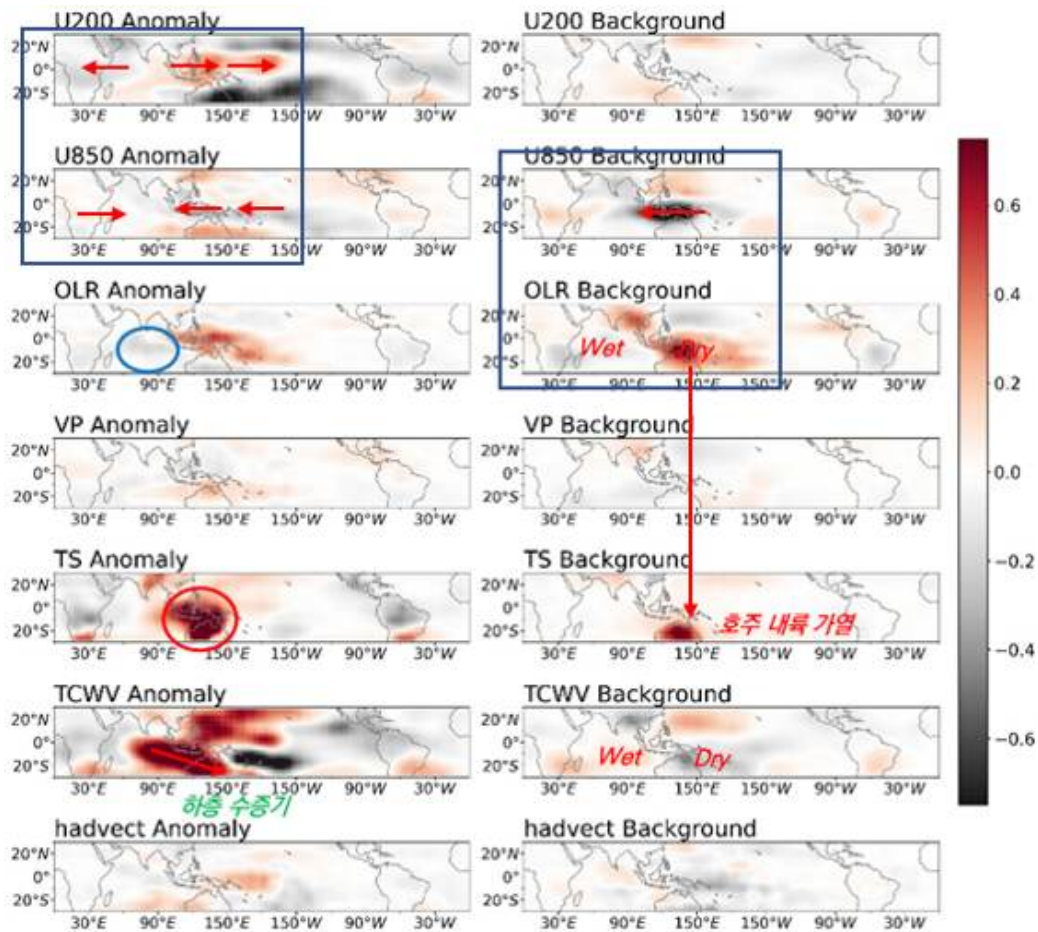


Figure 3.38. Attribution maps of input variables (Anomaly and Background variables) for RMM1 at a 15-day lead time.

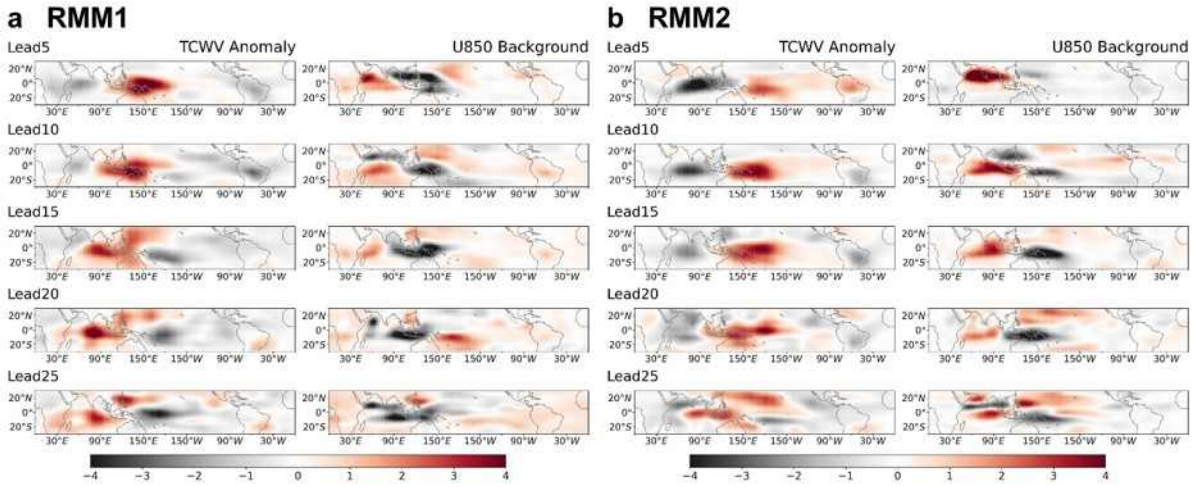


Figure 3.39. Attribution maps of TCWV Anomaly and U850 Background variables for lead times of 5, 10, 15, 20, 25 day.

3.3.3. 선행시간별 MJO 확률 예측을 위한 인공지능모델 개발

개발된 선행시간별 MJO RMM 지수 단정예측 인공지능모델을 기반으로 확률 예측을 위한 인공지능 모델을 개발하여 몇 가지 MJO 사례에 대한 분석을 수행하였다. Figure 3.40은 예측된 MJO RMM을 위상으로 변환한 후 3가지 섭동 실험에 대한 RPS 스코어 결과 보여주고 있다. 선행시간이 길어질수록 RPS 값이 높아지는 것을 볼 수 있다. 이는 Marshall et al. (2016)에서 보고된 MJO 위상에 대한 RPS 결과와 비슷한 결과를 보인다. ERA5 앙상블 다른 섭동 방법에 비해서 더 큰 변동성을 보이는데 이는 다른 섭동 방법에 비해서 앙상블 개수가 적은 데에서 기인하는 것으로 보인다.

Figures 3.41-43은 3가지 특정 MJO 사례에 대한 각 섭동 방법에 따른 앙상블 적용 결과를 보여주고 있다. Figure 3.41는 10개의 ERA5 앙상블 자료를 사용하여 생산된 MJO RMM 앙상블 예측 결과를 보여준다. ERA5 앙상블 자료로 생산한 10개의 예측 결과는 선행시간에 따른 앙상블 분산이 크지 않다. 이는 초기 조건에서의 자료 섭동만으로는 큰 앙상블 확산을 생산하기 어렵다는 것을 의미한다. Figure 3.42은 Perlin noise 섭동 방법을 통해 생산된 50개의 MJO RMM 앙상블 예측 결과를 보여준다. 세 가지의 섭동 방법 중 Perlin noise 방식이 선행시간이 지남에 따라 앙상블 분산이 가장 커지는 것을 보여준다. Perlin noise 방식은 기존 인공지능 기반 연구에서 앙상블 결과를 생산할 때 여전히 많이 활용되고 있지만 최근 연구에서는 예측 향상에 크게 기여하지 못한다는 결과가 보고된 바가 있다 (Chen et al., 2024). Figure 3.43는 MC dropout을 적용하여 모델 파라미터에 대한 섭동을 적용한 50개의 MJO RMM 앙상블 예측 결과를 보여준다. ERA5 앙상블 자료 적용에 비해서 더 큰 앙상블 분산을 만들어 내는 것을 볼 수 있다. 초기장과 모델 섭동 확률 예측 결과를 통해 평균에 수렴하는 인공지능 모델의 특성상 초기 자료 간의 분포의 차이만으로 큰 앙상블 스프레드를 도출하기 어려워 보이며, 모델 파라미터 섭동 방식도 MJO 사례별로 항상 큰 앙상블 스프레드를 산출하지 못하는 것으로 판단된다. Figure 3.44는 앞선 3가지 MJO 사례에 대한 선행시간에 따른 MJO 위상 확률 예측 결과를 보여준다. 이는 가장 큰 앙상블 분산을 보여준 Perlin noise 섭동 방식을 적용한 모델에 대한 결

과이다. 2021년 3월 17일(맨 위) 및 5월 1일(맨 아래)의 MJO 사례의 경우에 선행시간이 길어질 수록 앙상블 분산이 커지면서 각 MJO 위상의 확률이 분산되는 것을 보인다. 하지만 중간의 MJO 사례에서는 모든 앙상블 결과가 한가지 MJO 위상에 집중되고 있는 것을 볼 수 있다.

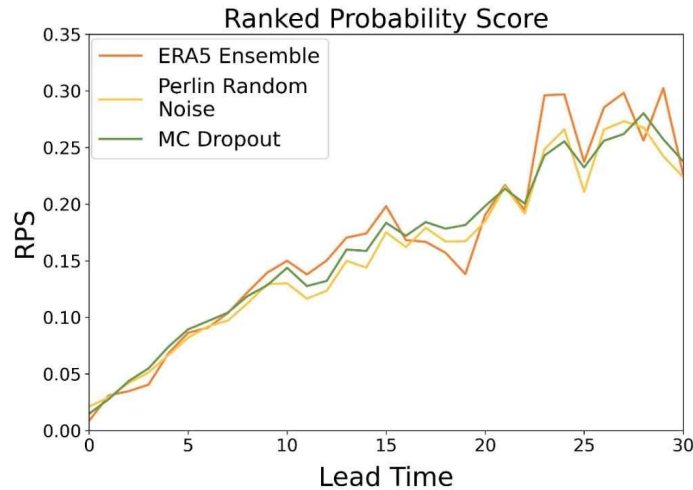


Figure 3.40. Ranked probability score (RPS) for MJO phases for different perturbation methods at lead times of 0 to 30 days.

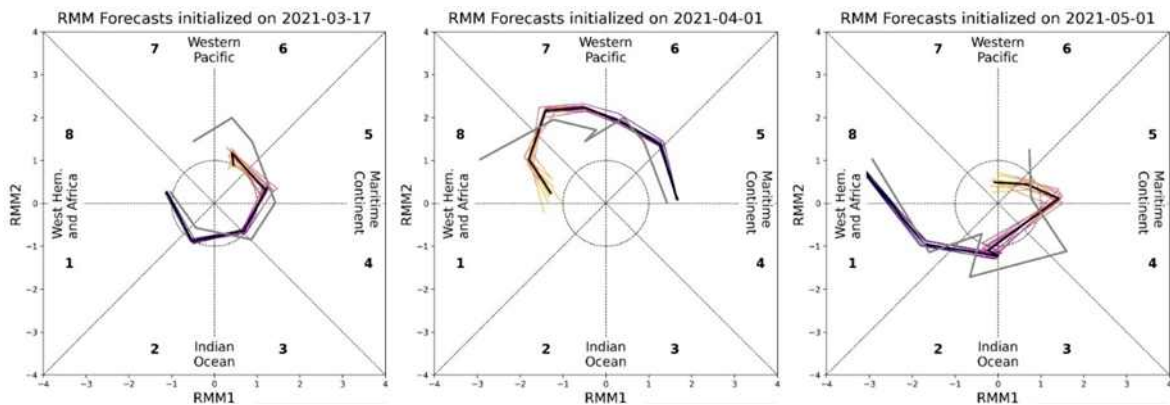


Figure 3.41. Ensemble results with different perturbation methods for MJO RMM probabilistic predictions with 10 ensembles with ERA5 ensemble data. Solid black line indicates the ensemble mean, and the grey line the ERA5 reference data.

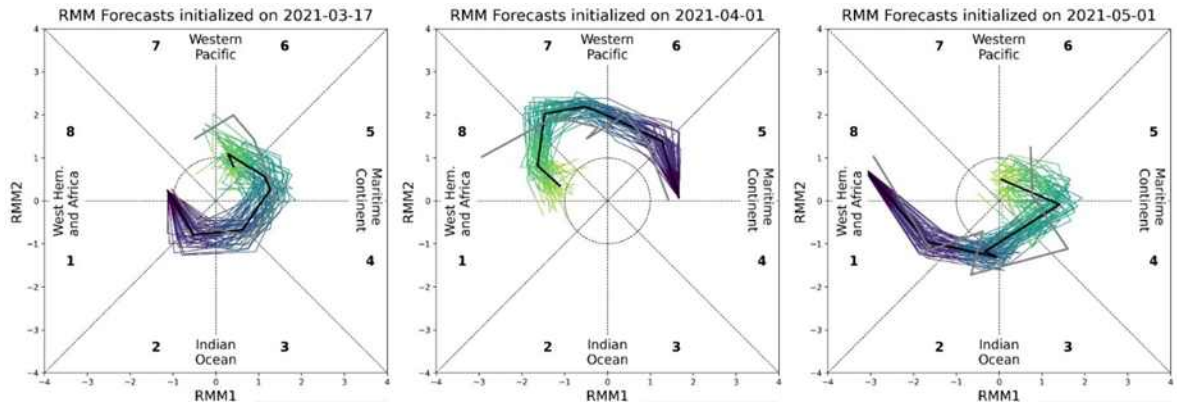


Figure 3.42. Ensemble results with different perturbation methods for MJO RMM probabilistic predictions with 50 ensembles with Perlin noise perturbation. Solid black line indicates the ensemble mean, and the gray line the ERA5 reference data.

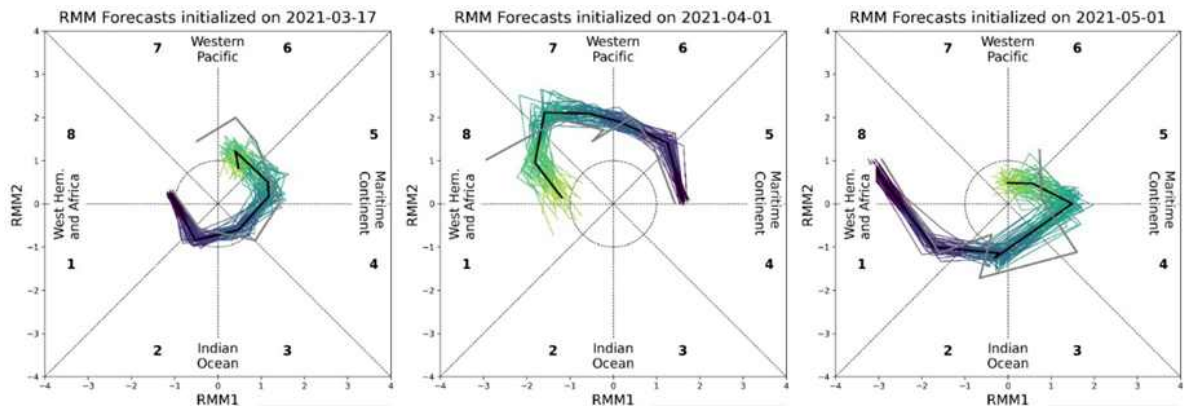


Figure 3.43. Ensemble results with different perturbation methods for MJO RMM probabilistic predictions with 50 ensembles with monte carlo dropout perturbation.

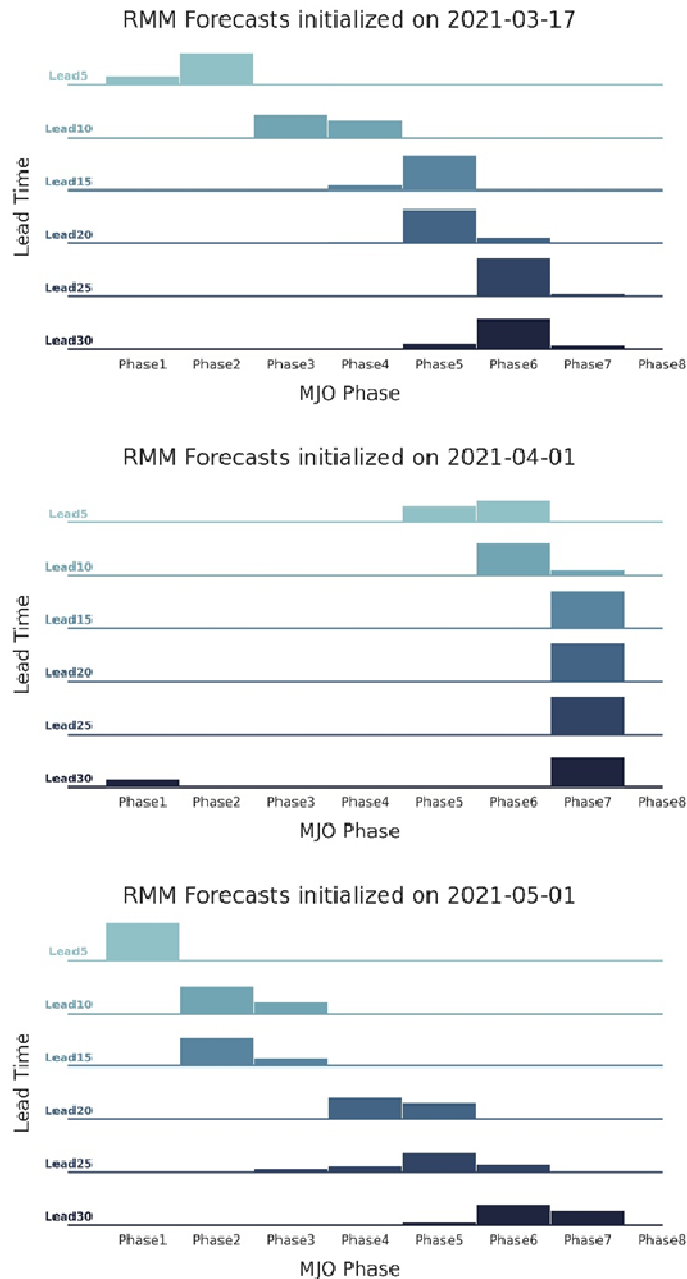


Figure 3.44. Probabilistic predictions of MJO phases produced by predicted MJO RMMs with 50 ensembles of Perlin noise perturbation.

3.4. 3-6주 동아시아 기온확률 예측 인공지능 모델 개발

3.4.1. 기준 1개월 기온 예측 모델 실험 및 내부 구조 분석

실험에 사용된 1개월 기온확률 예측 모델은 ECMWF S2S Model, DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14,

DL-CBAM+BIGDATA+FLT2+BTLNCK-14 이며, 각 모델의 Dataset, Epoch, Loss, Validation Loss, Input length는 Table 3.4와 같다. 여기서, 사례명(case name)의 정의는 DL이 딥러닝 모델, SE와 CBAM이 어텐션 블록 메커니즘 방법, BIGDATA가 학습데이터 기간 확장, FLT2이 합성곱 레이어의 필터 수 2배, 14가 입력자료의 길이를 의미한다. 데이터셋은 Table 2.10를 참조하면 된다. DL-SE+BIGDATA+FLT2+BTLNCK-14와 DL-CBAM+BIGDATA+FLT2+BTLNCK-14 모델은 피쳐 붕괴 분석 방법을 이용하여 BIGDATA와 FLT2가 적용된 모델을 분석하고 예측성에 영향을 주는 구간을 개선한 모델이다.

Table 3.4. Summary of ECMWF S2S Model and baseline deep-learning-based 1-month temperature prediction models.

Case name	Dataset	Epoch	Loss	Val loss	Input length
ECMWF	ECMWF S2S Model				
DL-SE-14	TDSV31-1_SHUFFLE	18	0.00189	0.00326	14
DL-SE+BIGDATA-14	TDSV32_SHUFFLE	21	0.00155	0.00306	14
DL-SE+BIGDATA+FLT2-14	TDSV32_SHUFFLE	17	0.00139	0.00303	14
DL-SE+BIGDATA+FLT2+BTLNCK-14	TDSV32_SHUFFLE	19	0.00101	0.00292	14
DL-CBAM-14	TDSV31-1_SHUFFLE	16	0.00167	0.00328	14
DL-CBAM+BIGDATA-14	TDSV32_SHUFFLE	20	0.00156	0.00298	14
DL-CBAM+BIGDATA+FLT2-14	TDSV32_SHUFFLE	19	0.00247	0.00322	14
DL-CBAM+BIGDATA+FLT2+BTLNCK-14	TDSV32_SHUFFLE	19	0.00093	0.00300	14

실험을 위해 딥러닝 기반 1개월 기온 예측 모델의 구조를 어텐션 메커니즘, 학습 데이터셋, 레이어 필터 수 확장 등을 고려하여 설계하였다. 어텐션 메커니즘을 중심으로 SE, CBAM, 그리고 학습데이터 확장, 레이어 필터 수 확장을 반영하였다. 모델은 DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14이다.

모델은 과거 14일 자료를 이용하여 1일 예측하는 모델로 구성하였다. 활성화 함수(activation function)는 Relu, 옵티마이저(optimizer)는 ADAM, 학습률(learning rate)은 1e-4, 손실함수(loss function)는 MSE(mean square error), 커널 사이즈(kernel size)는 3×3, 레이어 풀링(layer pooling)은 MaxPooling, 드롭아웃률(dropout rate)은 0.5를 이용하였다.

모델의 평가는 3주와 4주의 예측자료에서 남한지역의 5개 격자를 평균하여 계산한 확률예측값과 ASOS(Automatic Synoptic Observing System) 관측자료로 계산한 3분위값(above normal, near normal, below normal)을 HSS(Heidke Skill Score)를 이용하여 평가하였다. 딥러닝 기반 1개월 기온 예측 모델을 이용하여 rolling prediction method로 ECMWF S2S 예보자료의 1-2주 자료를 입력자료로 이용하여 3주와 4주를 재예측(reforecast)하였다. 재예측은 forecast, hindcast 자료 모두에 적용하였다. 학습자료에서 2016년과 2017년은 ASOS 관측자료로 계산한 3분위 자료가 없어서 제외하였다.

1개월 기온 예측 모델을 기초로 어텐션 메커니즘, 데이터확장, 레이어 필터 수 증가 등을 고려하여 ECMWF S2S 예측자료와 HSS 기법을 통해 예측성을 분석하였다. 실험에 사용한 딥러닝 모델은 어텐션 메커니즘을 중심으로 구분이 되며, DL-SE, DL-SE+BIGDATA, DL-SE+BIGDATA+FLT2, DL-CBAM, D-CBAM+BIGDATA, DL-CBAM+BIGDATA+FLT2로 구성하였다. SE와 CBAM은 어텐션 메커니즘, BIGDATA는 데이터 학습자료 기간 확장, FLT2는 레이어 필터 수 2배를 의미한다. 실험은 2018년~2022년 기간의 ECMWF S2S 예측자료의 1-2주 자료를 입력자료로 이용하여 예측 1주(F1, W+3), 예측 2주(F2, W+4)를 예측하였다. ECMWF S2S 예측자료와 딥러닝 예측자료를 KMA ASOS 자료(관측값)와 3분위(above normal, near normal, below normal) 확률예측을 수행하고 비교하였다.

최근 연구에서는 모델의 파라미터수가 기하급수적으로 늘어나면서 그리드 탐색(grid search) 알고리즘 등과 같은 기법을 이용한 최적 하이퍼파라미터를 찾는 연구는 줄어드는 추세이다. 그러나 그리드 탐색 알고리즘 등은 여전히 최적 하이퍼파라미터를 찾는 중요한 기법이다. 본 연구에서는 앞선 연구에서 수행한 그리드 탐색 알고리즘 대신 딥러닝 모델의 하이퍼파라미터 중 배치크기만을 고려하여 앞서 설계된 모델 중 CBAM 어텐션 메커니즘이 적용된 모델을 이용하여 학습하고 예측성을 분석하였다. 실험에 사용한 모델은 딥러닝 기반 1개월 기온 예측 모델이다. 실험 결과, 배치 크기가 증가할수록 validation loss가 증가하는 것을 보였다 (Figure 3.45). 배치크기가 증가함에 따라 학습에 소요되는 시간은 줄어들지만 예측성은 낮아지는 것을 보였다. 이 실험의 결과를 토대로 앞서 설계된 6개월 모델의 학습에는 배치크기 2를 적용하였다.



Figure 3.45. Plot of model training results by batch size.

학습데이터셋의 테스트기간 중 2018-2020년 총 3년에 대해 ACC 평가를 수행하였다. 계절

별로는 JJA가 다른 계절에 비해 낮은 ACC를 보였다(Figure 3.46). 연도별 ACC는 큰 차이를 보이지 않았다(Figure 3.47). 그러나 배치 크기가 커질수록 예측성은 낮아지는 것으로 나타났다.

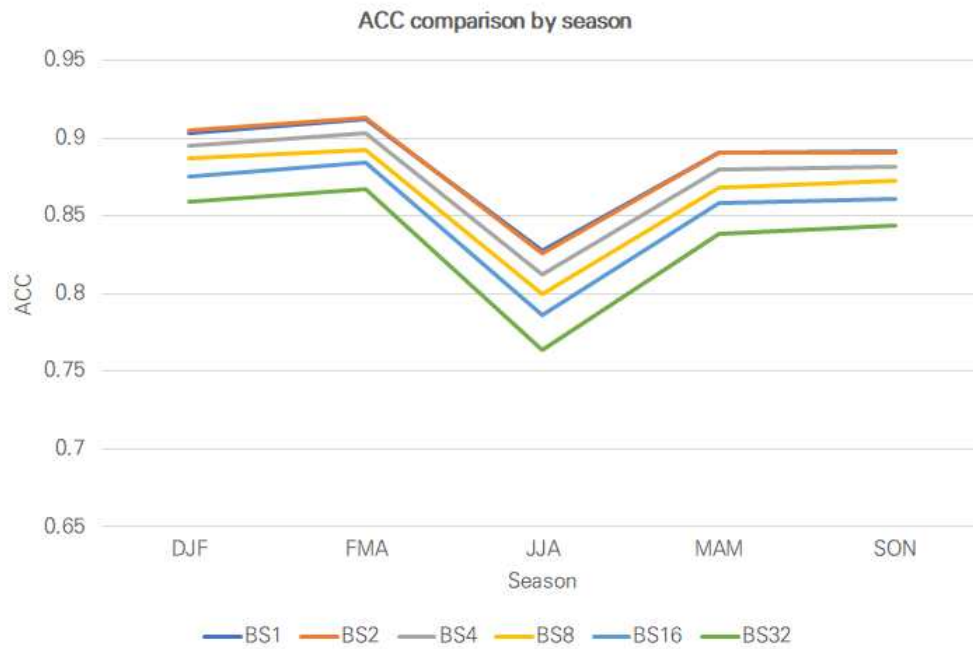


Figure 3.46. Plot of the comparison of the ACC by season.

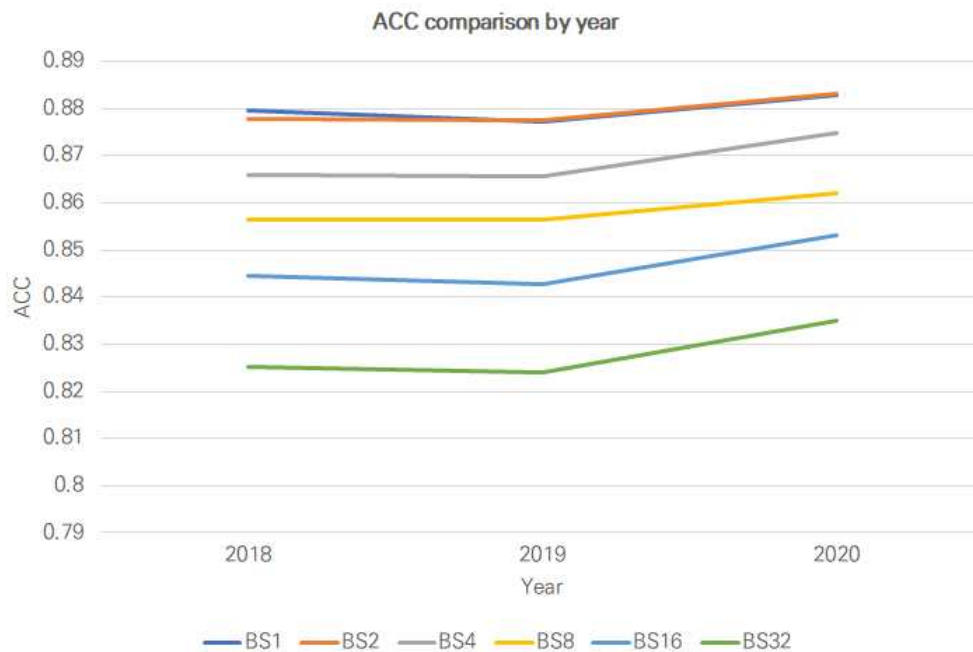


Figure 3.47. Plot of the comparison of the ACC by year.

전체 평가기간의 HSS 분석 결과, ECMWF는 F1과 F2에서 각 0.18, 0.08을 보였다. 딥러닝

모델은 F1에서 DL-SE 모델을 제외한 나머지 모델들이 ECMWF보다 좋은 성능을 보였다. F2에서는 DL-SE+BIGDATA 모델을 제외한 나머지 모델들이 ECMWF보다 낮은 성능을 보였다(Figure 3.4). 연도별 HSS 분석결과, F1에서 2018년, 2019년, 2021년은 DL-SE 모델을 제외한 모든 모델들이 높은 예측성을 보였으며, 2020년은 모든 딥러닝 모델이 좋은 성능을 보였다. 2022년은 모든 딥러닝 모델이 ECMWF보다 낮은 성능을 보였다(Figure 3.5). F2에서 2018년은 모든 딥러닝 모델이 높은 예측성을 보였으며, 2019년은 DL-SE, 2020년은 모든 딥러닝 모델이 낮은 성능을 보였다. 2021년은 DL-SE+BIGDATA-14와 DL-CBAM-14 모델이 ECMWF보다 높은 예측성을 보였으며, 2022년은 모든 딥러닝 모델이 낮은 예측성을 보였다(Figure 3.6).

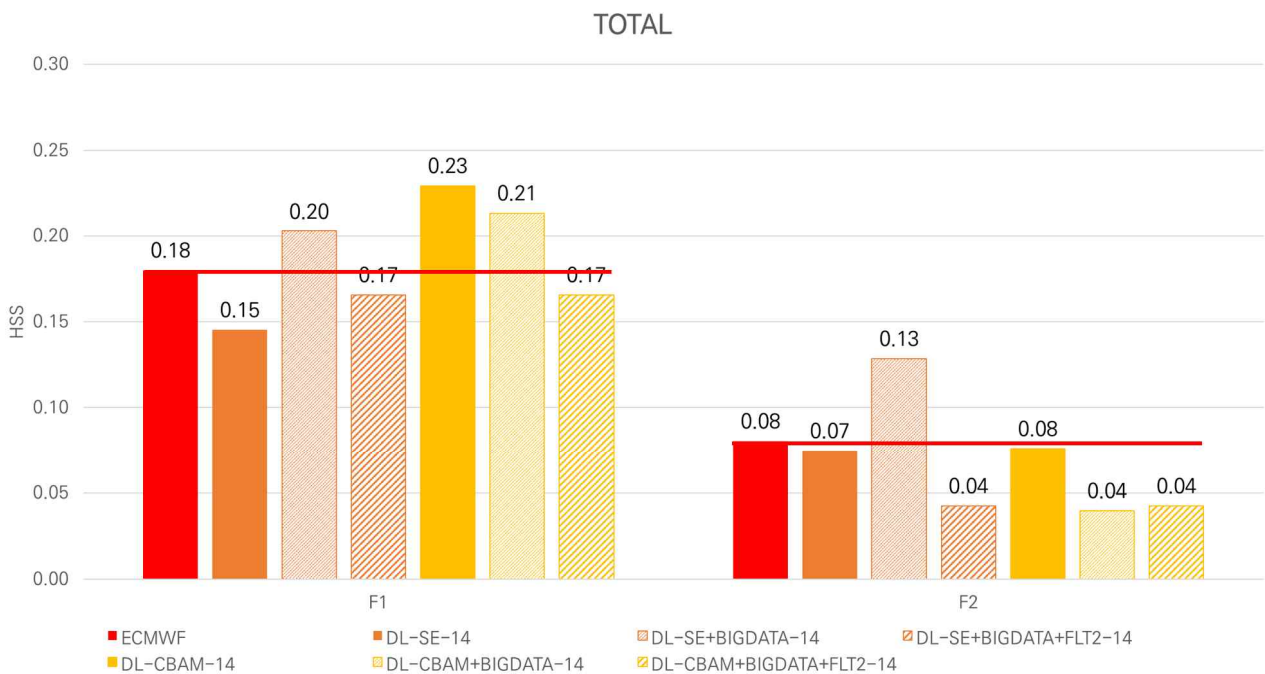


Figure 3.48. Summary of the HSS assessment for lead time of 3 to 4 weeks (F1 to F2): (from left) ECMWF S2S model and 6 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14).

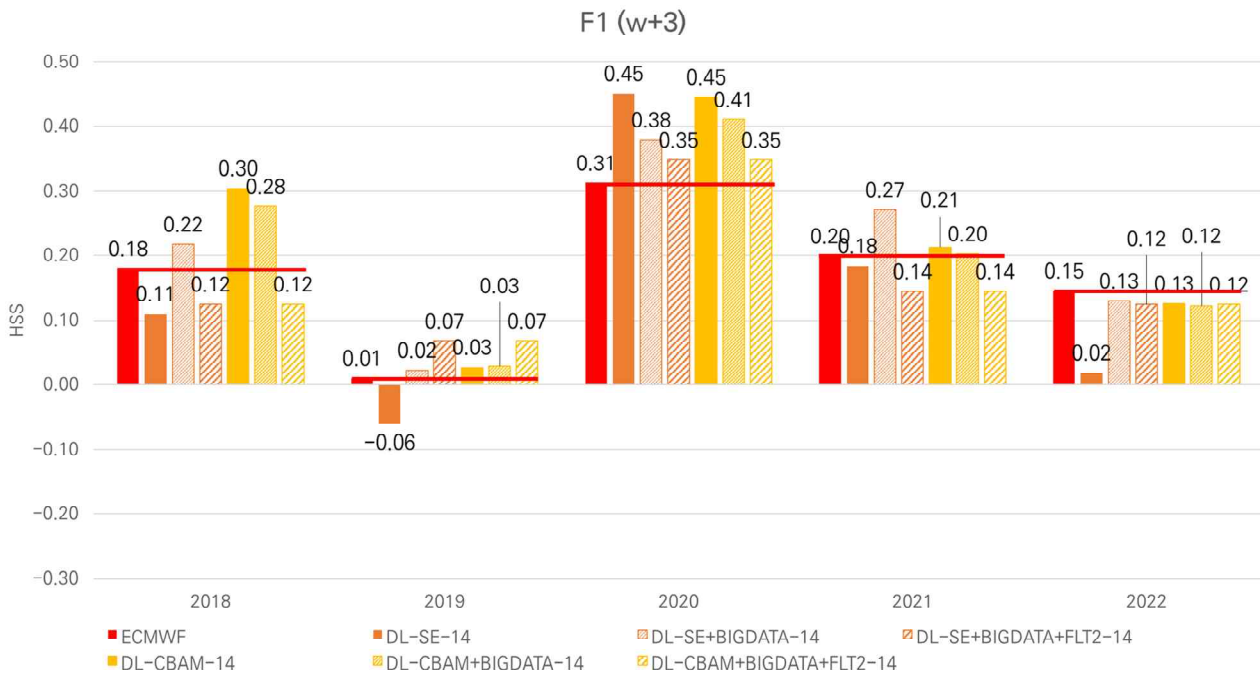


Figure 3.49. Summary of the HSS assessment for lead time of 3 week (F1): (from left) ECMWF S2S model and 6 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14).

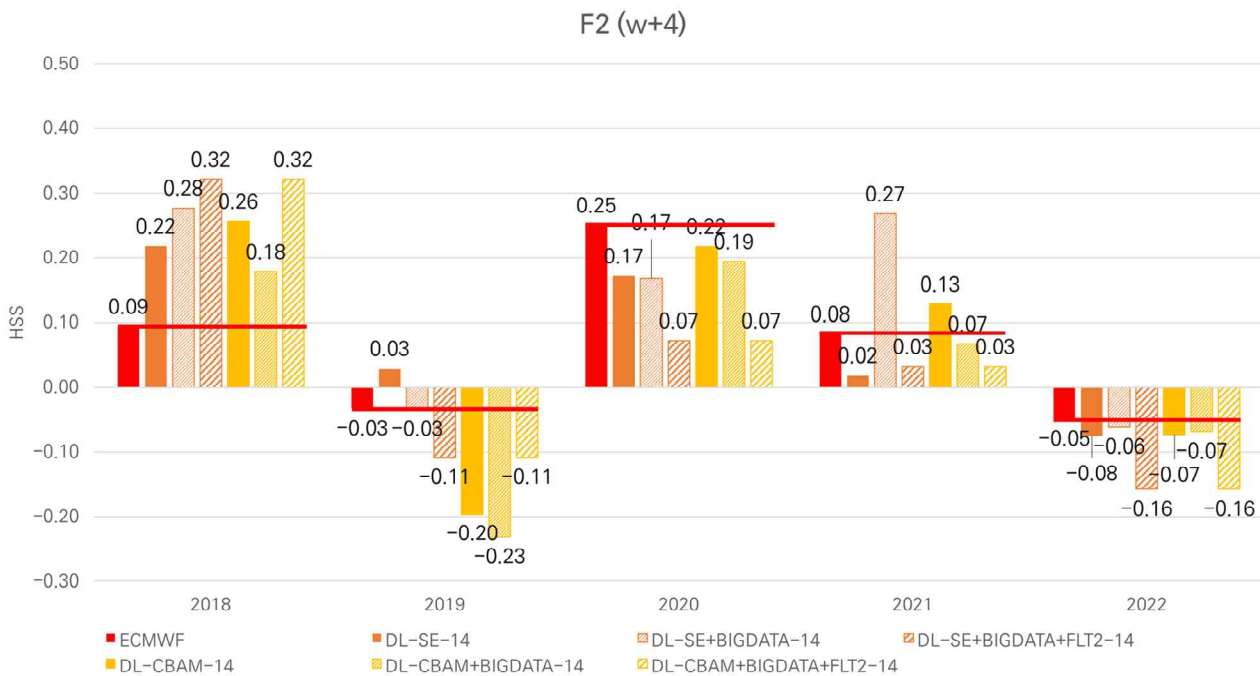


Figure 3.50. Summary of the HSS assessment for lead time of 4 week (F2): (from left) ECMWF S2S model and 6 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14).

딥러닝 모델의 학습데이터셋, 어텐션 메커니즘, 레이어 필터 수 증가에 따른 모델별 피쳐 붕괴(feature collapse) 분석을 통해 모델의 정확도에 영향을 미치는 구간을 분석하였다. 모델의 분석에 앞서, 모델의 에포크 증가에 따른 모델의 피쳐붕괴 분석을 통해 학습 손실률의 감소와 피쳐 붕괴 간 상관관계에 대해 분석하였다. 분석 결과, 딥러닝 모델의 손실률 감소에 따라 모델의 피쳐 붕괴는 낮아지는 것으로 나타났다(Figure 3.51 및 3.52). 이는 피쳐 붕괴가 모델 예측성에 영향을 미치는 것을 의미한다.

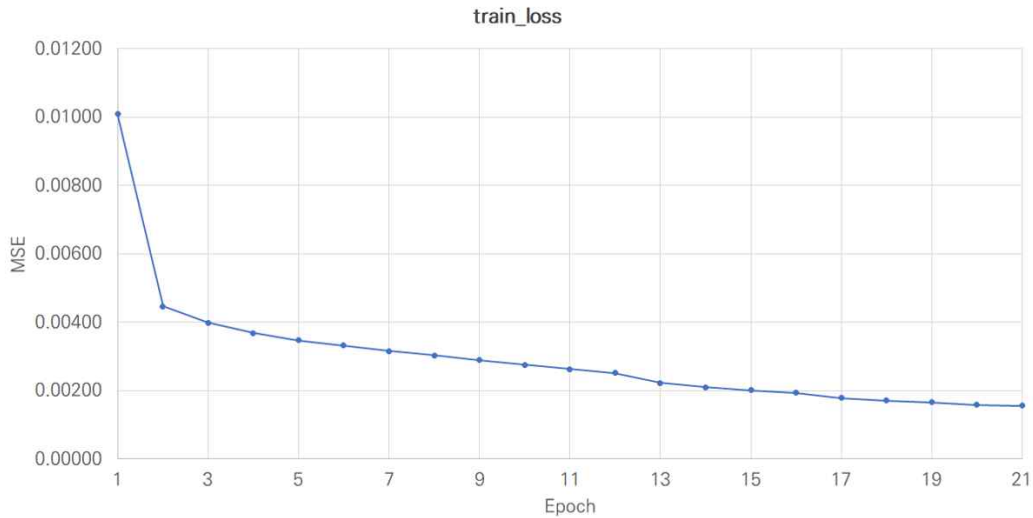


Figure 3.51. Line chart of training loss of a deep learning model.

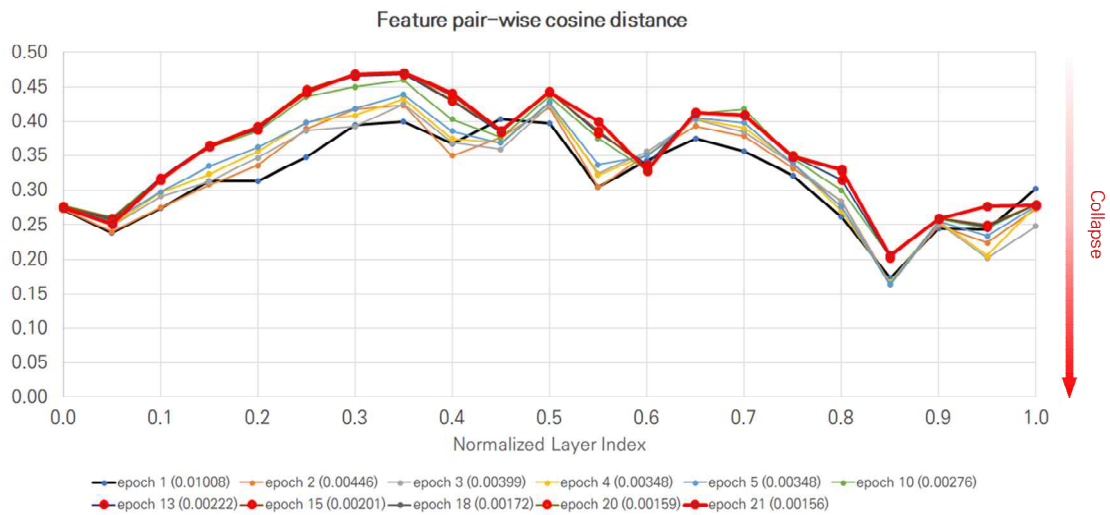


Figure 3.52. Feature pair-wise cosine distance analysis of a deep learning model.

앞서 실험에 사용한 6개 모델의 주요 구간(2차원 합성곱 레이어, 활성화 레이어 등)에 대해 피쳐 붕괴 분석을 수행하였다(Figure 3.53 및 3.54). 피쳐 붕괴 분석에는 피쳐 코사인 거리

(feature cosine distance)가 사용되며, 거리가 가까울수록 피쳐 붕괴가 심화되는 것을 의미한다. 분석 결과, 어텐션 메커니즘 SE와 CBAM 모두 U-NET 아키텍처의 bottleneck 구간에서 일어나는 것으로 나타났다. 그래서 bottleneck 구간을 contrasting path와 expansive path에서 사용한 CCA(conv2d+conv2d+attention block) 블록을 이용하였다(Figure 3.55). CCA 블록은 FLT2가 적용된 두 모델(DL-SE+BIGDATA+FLT2-14, DL-CBAM-BIGDATA+FLT2-14)에 적용하고 학습 후 피쳐 붕괴 분석을 하였다. Bottleneck 개선 전과 후의 피쳐 코사인 거리가 크게 개선되는 것을 보였다(Figure 3.56 및 3.57). Table 3.5은 bottleneck 구간이 개선된 모델이다.

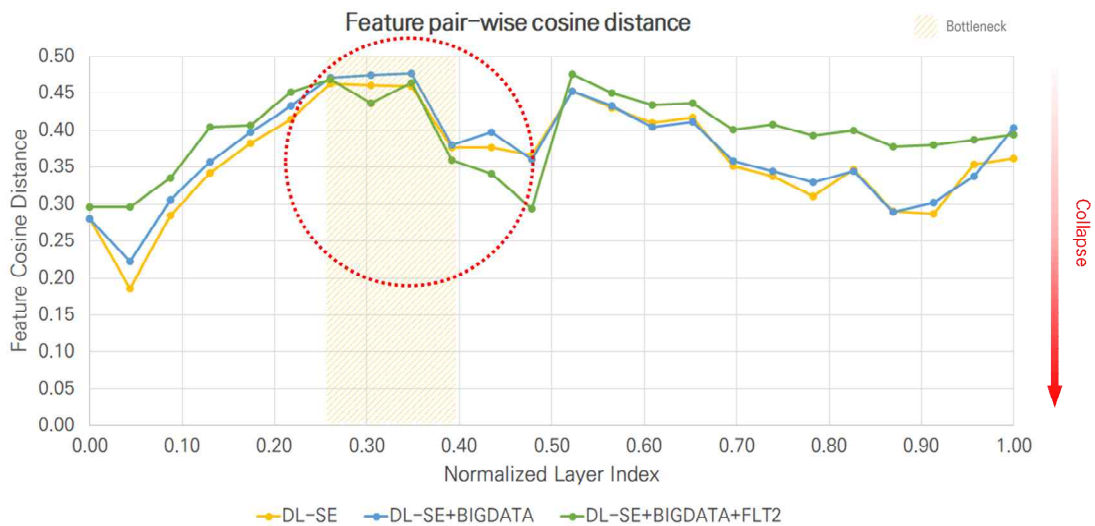


Figure 3.53. The results of the feature pair-wise cosine distance analysis of the deep learning models with the SE block.

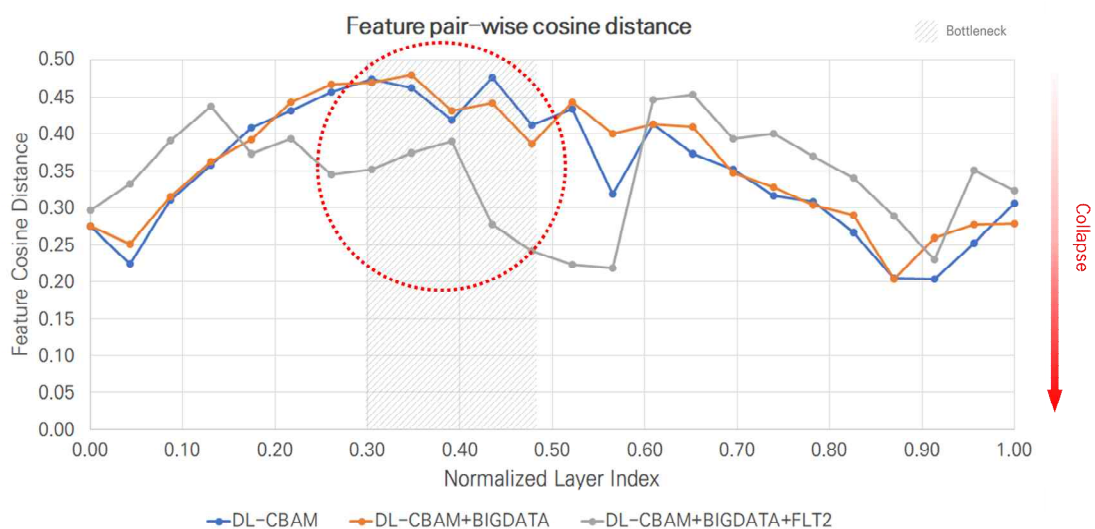


Figure 3.54. The results of the feature pair-wise cosine distance analysis of the deep learning models with the CBAM block.

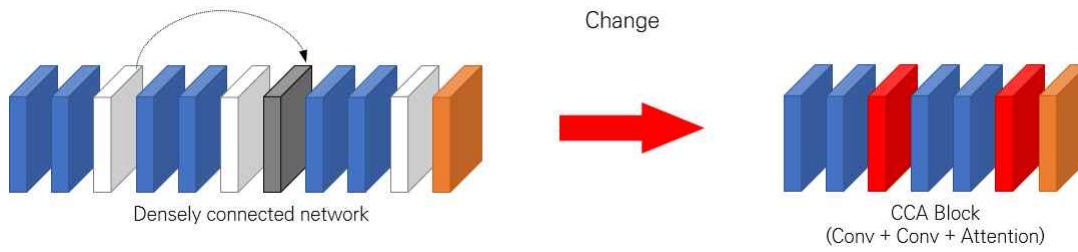


Figure 3.55. Design of CCA block.

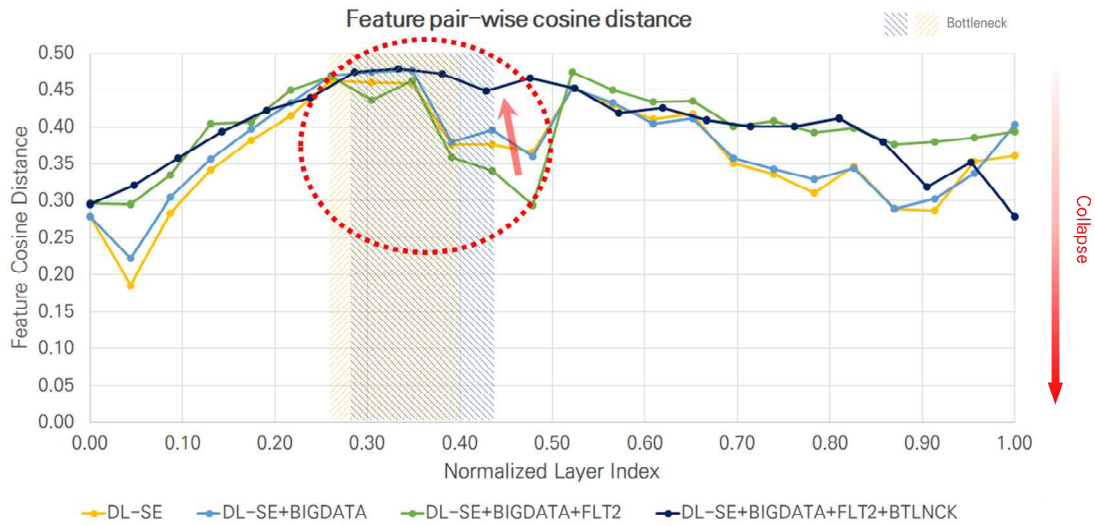


Figure 3.56. The results of the feature pair-wise cosine distance analysis of the deep learning models with the SE block after changing the bottleneck block.

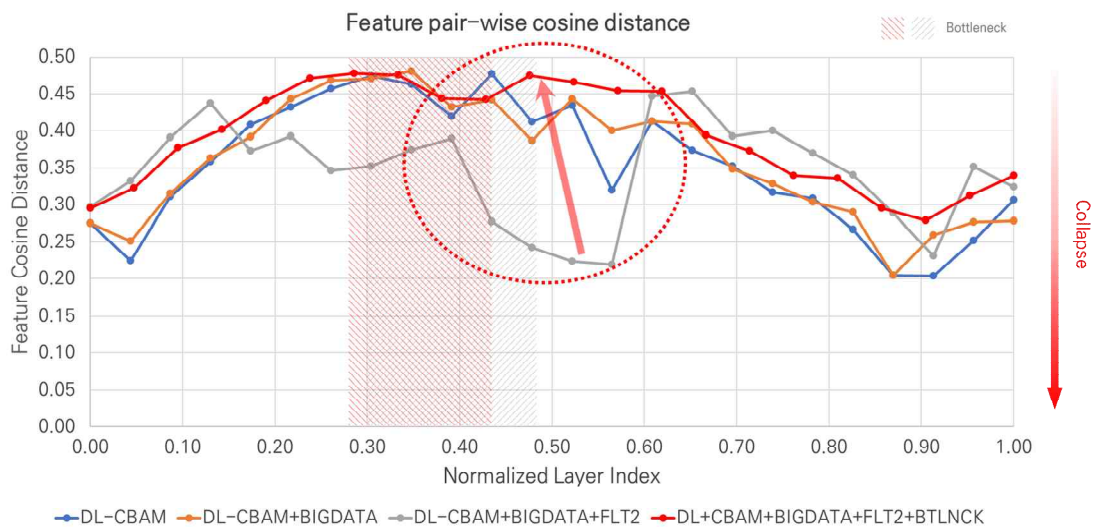


Figure 3.57. The results of the feature pair-wise cosine distance analysis of the deep learning models with the CBAM block after changing the bottleneck block.

Table 3.5. Summary of improved deep learning models using feature pair-wise cosine distance analysis.

Case name	Dataset	Epoch	Loss	Val loss	Input length
DL-SE+BIGDATA+FLT2+BTLNCK-14	TDSV32_SHUFFLE	19	0.00101	0.00292	14
DL-CBAM+BIGDATA+FLT2+BTLNCK-14	TDSV32_SHUFFLE	19	0.00093	0.00300	14

Bottleneck이 개선된 모델을 이용하여 HSS 평가를 수행하였다. 전체 기간에 대해서는 F1에서 DL-SE+BIGDATA+FLT2-14는 0.17에서 0.25, DL-CBAM+BIGDATA+FLT2-14는 0.17에서 0.20, F2에서 DL-SE+BIGDATA-FLT2-14는 0.04에서 0.08, DL-CBAM+BIGDATA+FLT2-14는 0.04에서 0.05로 개선되었다(Figure 3.15). 예측 1주(F1, W+3)에서 2021년도 DL-CBAM+BIGDATA+FLT2-14 모델을 제외하고는 모두 개선되었으며, 2022년 DL-SE+BIGDATA+FLT2-14모델은 0.12에서 0.33으로 2배 이상 개선되었다(Figure 3.45). 예측 2주(F2, W+4)에서는 2020년 DL-SE+BIGDATA+FLT2-14, 2021년 DL-SE+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2-14, 2022년 DL-SE+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2-14 모델에서는 개선되었지만 그 외에서는 기존 모델보다 예측성이 떨어지는 경향을 보였다(Figure 3.58). 예측 2주는 ECMWF S2S 모델의 1-2주를 입력자료로 이용하여 14일을 예측한 결과 중 8~14일의 기간을 사용한다. 피쳐 붕괴 현상을 줄임으로써 예측 1주의 예측성을 높일 수 있지만 예측 기간이 길어질수록 모델의 예측성이 떨어지는 현상이 나타났다. 장기간의 예측성을 높일 수 있는 모델의 설계가 필요할 것으로 판단된다.

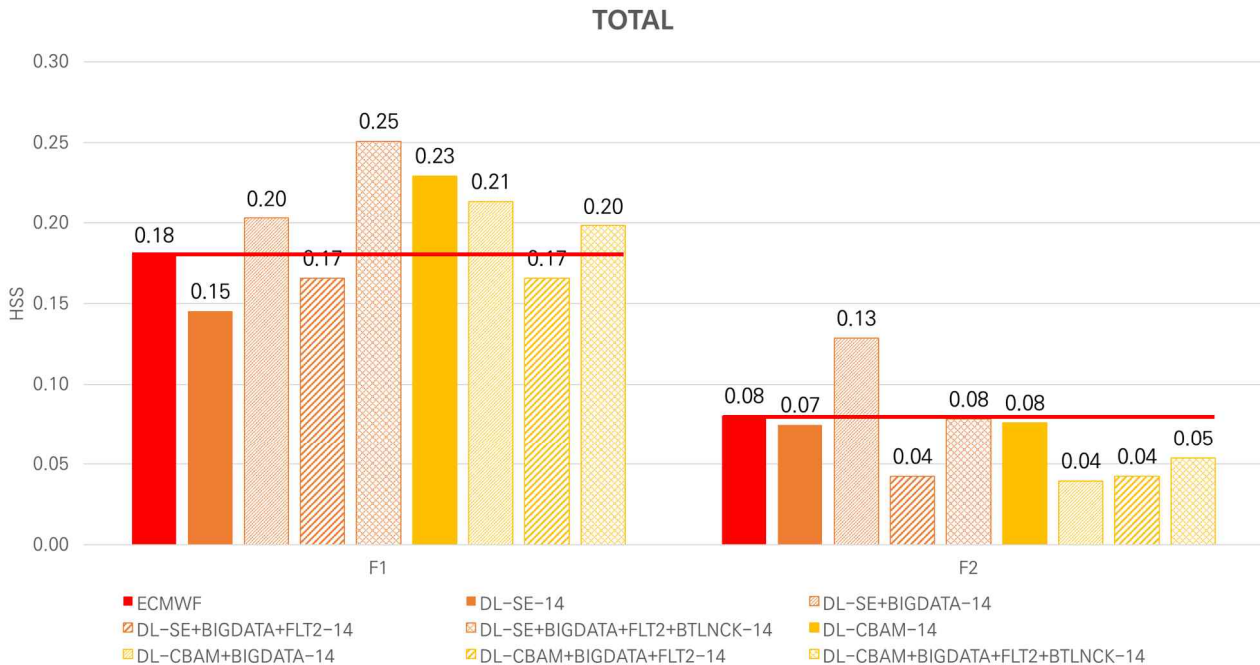


Figure 3.58. Summary of the HSS assessment for lead time of 3 to 4 weeks (F1 to F2): (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14).

동아시아지역 3-4주 기온확률 예측 결과를 이용하여 모델의 예측 가능성을 평가하고 피쳐 붕괴 분석을 통해 모델을 개선하였다. 기준 1개월 기온 예측 모델을 바탕으로 5-6주까지 기온 확률 예측을 확장하고 예측성을 분석하였다. F1은 예측 1주, 즉 3주를 의미한다. 또한 엘니뇨 (El Niño)와 라니냐(La Niña) 발생 시기의 기온확률 예측성을 분석하였다.

기준 1개월 기온 예측 모델은 ECMWF, DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BLTNCK-14 등 총 9개이다. 5주(F3) 예측에서는 8개의 딥러닝 모델이 ECMWF보다 좋은 예측성을 보였다. 6주(F4) 예측에서는 DL-SE-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM+BIGDATA+FLT2-14 등 4개의 딥러닝 모델이 ECMWF보다 좋은 예측성을 보였다. 3-6주 예측 결과를 종합하면 1위 DL-SE+BIGDATA+FLT2+BLTNCK-14, 2위 DL-CBAM-14, 2위 DL-SE+BIGDATA-14 순으로 예측성이 좋았으며 ECMWF는 9개 모델을 6위의 예측성을 보였다.

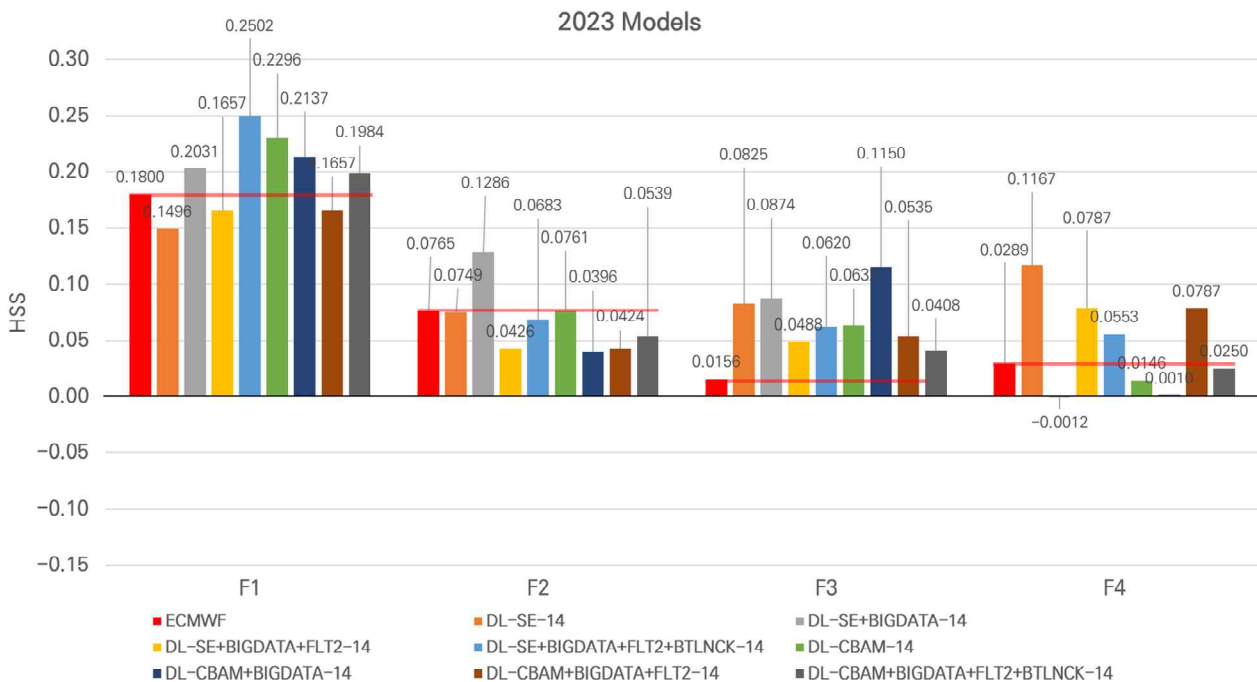


Figure 3.59. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BLTNCK-14).

엘니뇨 발생 시기의 3주(F1) 예측은 DL-SE-14, DL-SE+BIGDATA-14 모델을 제외한 나머지 6개 모델이 ECMWF보다 좋은 예측성을 보였다(Figure 3.60). 4주(F2) 예측은 DL-SE-14,

DL-SE+BIGDATA+FLT2-14, DL-CBAM-14, DL-CBAM+BIGDATA+FLT2-14 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 ECMWF보다 좋은 예측성을 보인 모델이 없었으며 6주(F4) 예측에서는 DL-CBAM+BIGDATA+FLT2+BLTNCK-14 모델만 ECMWF보다 예측성이 높았다. 엘니뇨 발생 시기의 예측 순위는 1위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 2위 ECMWF, 3위 DL-CBAM-14 순으로 나타났다. 라니냐 발생 시기의 3주(F1) 예측은 DL-SE-14, DL-SE+BIGDATA+FLT2-14 모델을 제외한 6개 모델이 ECMWF보다 좋은 예측성을 보였다 (Figure 3.61). 4주(F2) 예측은 DL-SE+BIGDATA-14, 5주(F3) 예측은 DL-SE+BIGDATA-14, DL-CBAM+BIGDATA-14 모델이 ECMWF보다 예측성이 높았다. 그리고 6주(F4) 예측은 DL-SE-14, DL-SE+BIGDATA+FLT2-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14 모델이 ECMWF보다 예측성이 높았다. 라니냐 발생 시기의 예측 순위는 1위 DL-CBAM+BIGDATA-14, 2위 DL-SE+BIGDATA-14, 3위는 DL-SE-14 순이었다.

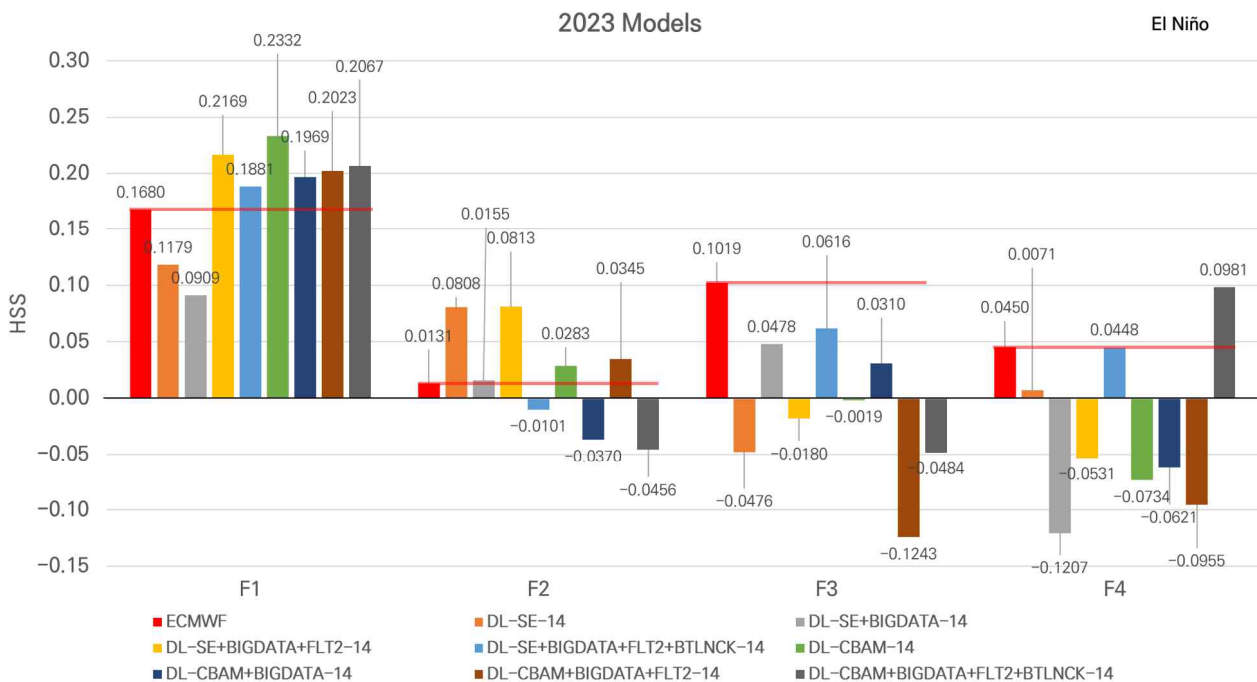


Figure 3.60. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14).

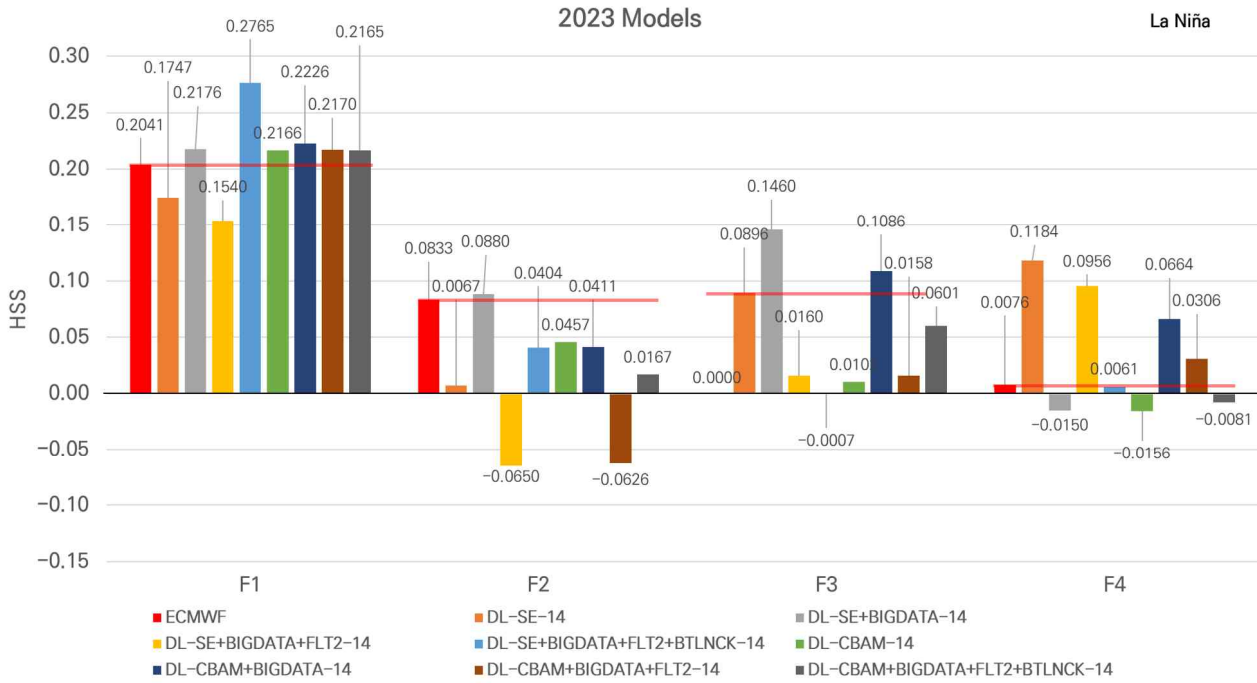


Figure 3.61. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model and 8 baseline models (DL-SE-14, DL-SE+BIGDATA-14, DL-SE+BIGDATA+FLT2-14, DL-SE+BIGDATA+FLT2+BLTNCK-14, DL-CBAM-14, DL-CBAM+BIGDATA-14, DL-CBAM+BIGDATA+FLT2-14, DL-CBAM+BIGDATA+FLT2+BLTNCK-14).

3.4.2. 입력 변수 확장을 통한 기온 예측 모델 실험 및 결과

기존 1개월 기온 예측 모델을 기반으로 한 동아시아지역 1개월 기온 확률 예측모델은 T2M 변수만을 이용하여 구축되었다. 5-6주 예측에서는 대기-지면-해양(ALO) 시스템에 따라 초기장으로 지면, 해양이 중요한 역할을 한다(Mariotti et al., 2018). 기존 1개월 기온 예측 모델에 지면 자료로 NDVI, 해양자료로 SST 기후평년자료를 추가하고 DL-SE-7, DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BLTNCK-7, DL-SE+FLT1+BLTNCK-7+SST, DL-SE+FLT2+BLTNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7, DL-CBAM-7+SST+NDVI, DL-CBAM+FLT1+BLTNCK-7 등 10개 모델을 구축하였다. 각 사례에서 사용한 데이터셋, 학습 종료 에포크, 손실값, 검증 손실값, 입력자료 길이는 Table 3.6와 같다. 사례명(case name)은 앞서 설명한 것과 같으며 SST, NDVI는 기존 1개월 기온 예측 모델에서 추가한 변수를 의미한다. 모델의 입력자료 길이는 기존 모델과 다르게 7일을 이용하였다. 실험 평가는 TDSV32_SHUFFLE 데이터셋의 테스트 자료를 이용하여 2018~2022년을 분석하였다. NDVI 자료가 1982~2019년까지만 구축되어 있기 때문에 2020~2022년은 2019년도 자료를 이용하였다. NDVI 자료가 구축된 2018-2019년 분석 및 구축된 모델 전체의 동아시아지역 5-6주 기온확률예측 종합분석은 3.4.7. 종합분석에서 다뤘다.

Table 3.6. Summary of deep-learning-based 1-month temperature prediction models with NDVI (land) and SST climatology (ocean) variables.

Case name	Dataset	Epoch	Loss	Val loss	Input length
DL-SE-7	TDSV61_SHUFFLE	29	0.00268	0.00371	7
DL-SE-7+SST	TDSV61-2_SHUFFLE	20	0.00170	0.00319	7
DL-SE-7+SST+NDVI	TDSV61-2_SHUFFLE	26	0.00185	0.00343	7
DL-SE+FLT1+BTLNCK-7	TDSV61_SHUFFLE	25	0.00279	0.00384	7
DL-SE+FLT1+BTLNCK-7+SST	TDSV61-2_SHUFFLE	20	0.00115	0.00324	7
DL-SE+FLT2+BTLNCK-7	TDSV61_SHUFFLE	19	0.00178	0.00356	7
DL-SE-14+SST+NDVI	TDSV61-3_SHUFFLE	23	0.00140	0.00314	14
DL-CBAM-7	TDSV61_SHUFFLE	30	0.00274	0.00381	7
DL-CBAM-7+SST+NDVI	TDSV61-2_SHUFFLE	17	0.00215	0.00354	7
DL-CBAM+FLT1+BTLNCK-7	TDSV61_SHUFFLE	25	0.00279	0.00384	7

평가 결과, 3주(F1) 예측은 ECMWF보다 좋은 예측성을 보인 모델이 없었으며, 4주(F2) 예측은 DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7+SST, DL-SE+FLT2+BTLNCK-7 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 9개 모델 모두 ECMWF보다 좋은 예측성을 보였으며, 6주(F4) 예측은 DL-SE-7, DL-SE-7+SST+NDVI, DL-SE-14+SST+NDVI, DL-CBAM-7 모델이 ECMWF보다 좋은 예측성을 보였다(Figure 3.62). 전체 기간의 예측 순위는 1위 DL-SE-14+SST+NDVI, 2위 DL-SE+FLT1+BTLNCK-7+SST, 3위 DL-SE-7+SST+NDVI 순이었다.

엘니뇨 발생 시기의 3주(F1) 예측은 DL-SE-14+SST+NDVI 모델이 ECMWF보다 좋은 예측성을 보였다(Figure 3.63). 4주(F2) 예측은 DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7+SST 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 DL-SE+FLT1+BTLNCK-7+SST 모델이 ECMWF보다 좋은 예측성을 보였으며, 6주(F4) 예측에서는 DL-SE-7, DL-SE-7+SST+NDVI, DL-CBAM-7 모델이 ECMWF보다 예측성이 높았다. 엘니뇨 발생 시기의 예측 순위는 1위 ECMWF, 2위 DL-SE-7+SST+NDVI, 3위 DL-SE+FLT1+BTLNCK-7+SST 순으로 나타났다. 라니냐 발생 시기의 3주(F1) 예측은 10개 모델 모두 ECMWF보다 낮은 예측성을 보였다(Figure 3.64). 4주(F2) 예측은 DL-SE-7+SST, 5주(F3) 예측은 DL-SE+FLT2+BTLNCK-7 모델이 ECMWF보다 예측성이 높았다. 그리고 6주(F4) 예측은 DL-SE+FLT2+BTLNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7 모델이 ECMWF보다 예측성이 높았다. 라니냐 발생 시기의 예측 순위는 공동 1위 DL-SE-14+SST+NDVI, 공동 1위 DL-SE+FLT2+BTLNCK-7, 3위는 ECMWF 순이었다.

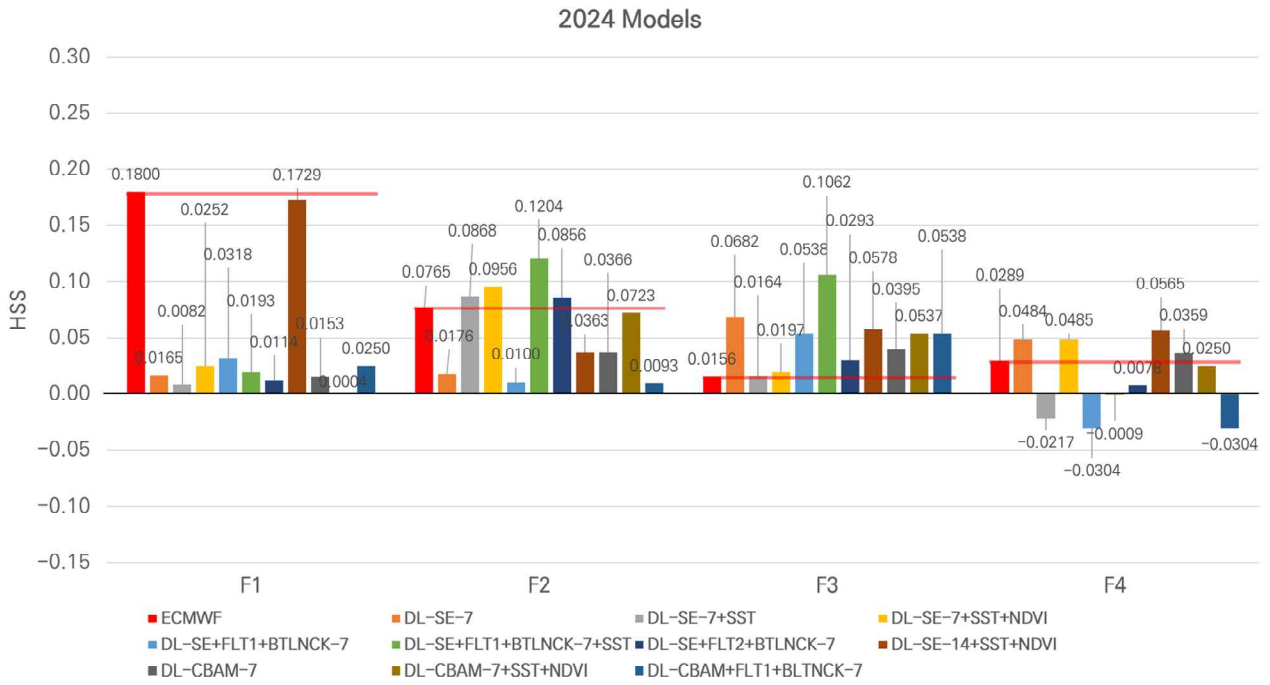


Figure 3.62. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model and 10 models with additional input variables (DL-SE-7, DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7, DL-SE+FLT1+BTLNCK-7+SST, DL-SE+FLT2+BTLNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7, DL-CBAM-7+SST+NDVI, DL-CBAM+FLT1+BLTNCK-7).

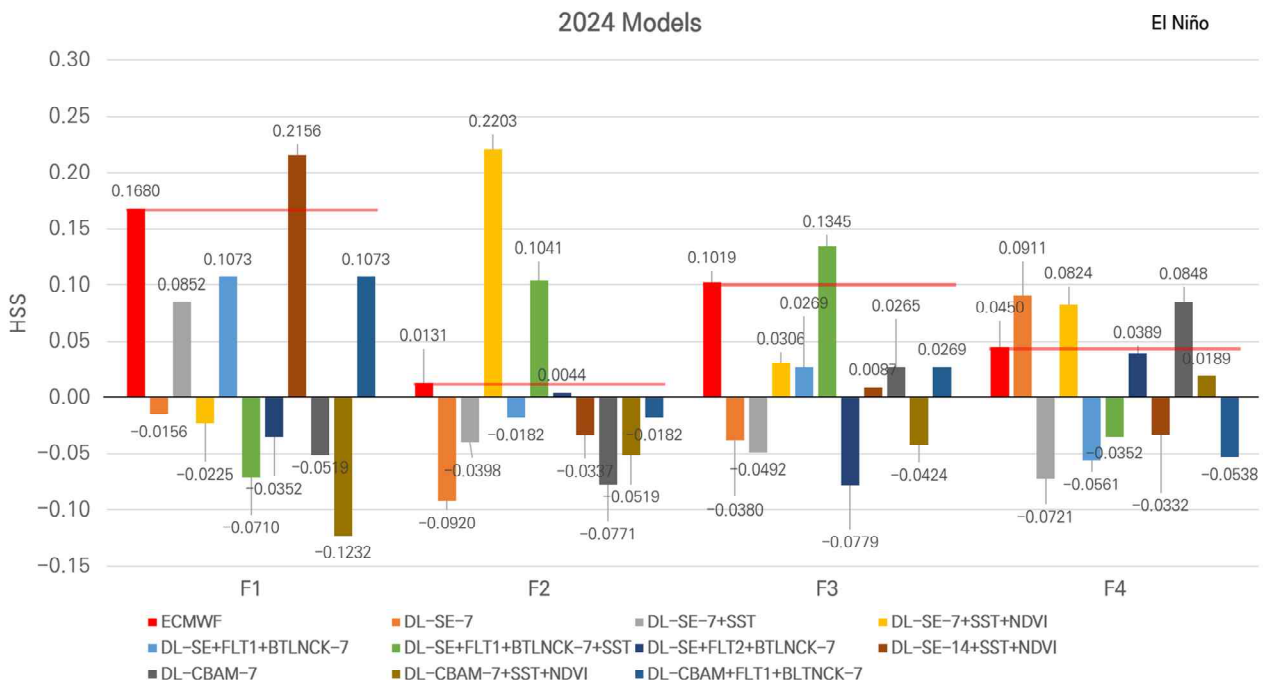


Figure 3.63. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model and 10 models with additional input variables (DL-SE-7, DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7, DL-SE+FLT1+BTLNCK-7+SST, DL-SE+FLT2+BTLNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7, DL-CBAM-7+SST+NDVI, DL-CBAM+FLT1+BLTNCK-7).

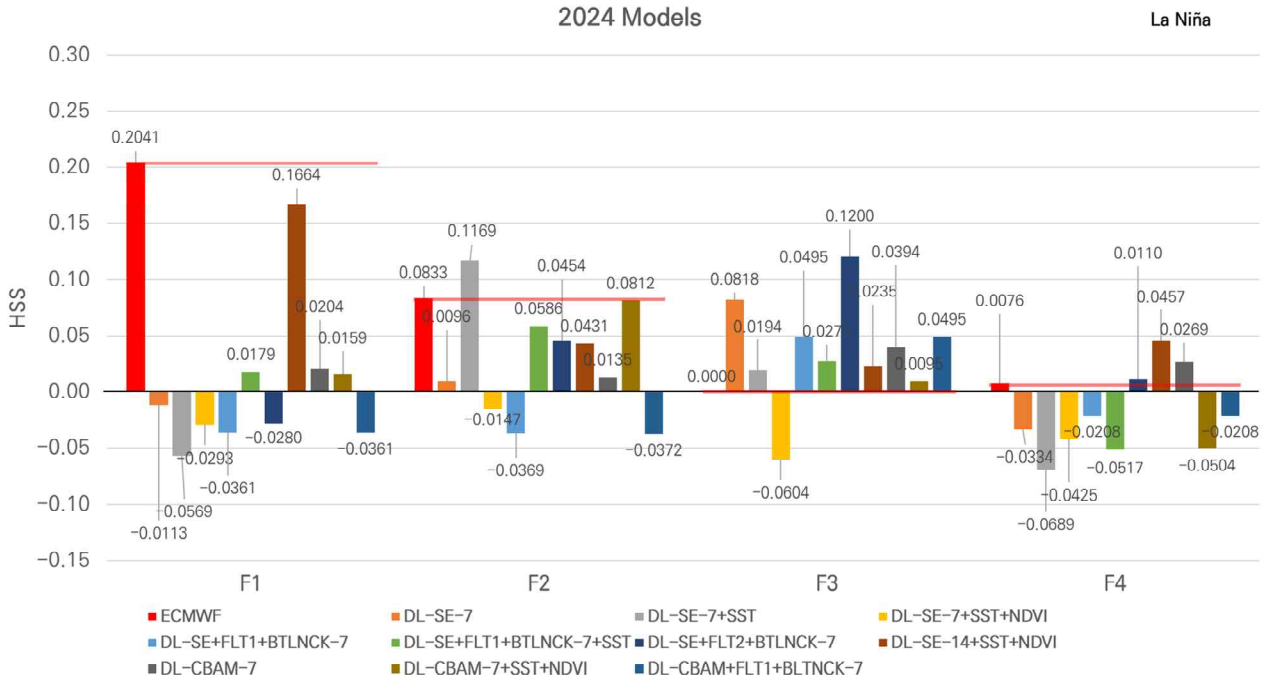


Figure 3.64. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model and 10 models with additional input variables (DL-SE-7, DL-SE-7+SST, DL-SE-7+SST+NDVI, DL-SE+FLT1+BTLNCK-7, DL-SE+FLT1+BTLNCK-7+SST, DL-SE+FLT2+BTLNCK-7, DL-SE-14+SST+NDVI, DL-CBAM-7, DL-CBAM-7+SST+NDVI, DL-CBAM+FLT1+BTLNCK-7).

3.4.3. 최신 딥러닝 아키텍처를 이용한 기온 예측 모델 실험 및 결과

Kim 등(2024)이 제안한 RDNet(DenseNets Reloaded)의 Stage-N block을 이용하여 NM-SE-7, NM-SE-7+SST+NDVI, NM-SE-7-RELU+SST+NDVI, NM-SE-14 등 입력 변수 확장을 통한 기온 예측 모델에서 높은 예측성을 보인 모델을 중심으로 4개 모델을 구축하였다(Table 3.7). NM-SE-7-RELU+SST+NDVI 모델을 제외한 3개 모델의 기본 활성화 함수로 RDNet의 기본 활성화 함수인 GELU를 이용하였다.

Table 3.7. Summary of modified Stage-N architecture-based models for 1-month temperature prediction.

CASE NAME	Dataset	Epoch	Loss	Val loss	Input length
NM-SE-7	TDSV61-2_SHUFFLE	27	0.00140	0.00331	7
NM-SE-7+SST+NDVI	TDSV61-2_SHUFFLE	23	0.00165	0.00336	7
NM-SE-7-RELU+SST+NDVI	TDSV61-2_SHUFFLE	24	0.00159	0.00342	7
NM-SE-14	TDSV32_SHUFFLE	21	0.00164	0.00312	14

평가 결과, 3주(F1) 예측은 NM-SE-14 모델이 ECMWF보다 좋은 예측성을 보였으며, 4주(F2) 예측은 4개 모델 모두 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 NM-SE-14 모델이 ECMWF보다 좋은 예측성을 보였으며, 6주(F4) 예측은 NM-SE-7+SST+NDVI 모델이 ECMWF보다 좋은 예측성을 보였다(Figure 3.65). 전체 기간의 예측 순위는 1위 ECMWF, 2위 NM-SE-14, 3위 NM-SE-7+SST+NDVI 순이었다.

엘니뇨 발생 시기의 3주(F1) 예측은 모델 모두 ECMWF보다 낮은 예측성을 보였다(Figure 3.66). 4주(F2) 예측은 NM-SE-7-RELU+SST+NDVI 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3)와 6주(F4) 예측에서는 모든 모델이 ECMWF보다 낮은 예측성을 보였다. 엘니뇨 발생 시기의 예측 순위는 1위 ECMWF, 공동 2위 NM-SE-7+SST+NDVI, 공동 2위 NM-SE-7-RELU+SST+NDVI 순으로 나타났다. 라니냐 발생 시기의 3주(F1) 예측은 NM-SE-14 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.67). 4주(F2) 예측은 모든 모델이 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 NM-SE-7, NM-SE-7-RELU+SST+NDVI 모델, 6주 예측은 NM-SE-7+SST+NDVI 모델이 ECMWF보다 예측성이 높았다. 라니냐 발생 시기의 예측 순위는 1위 ECMWF, 2위 NM-SE-7+SST+NDVI, 3위 NM-SE-7 순이었다.

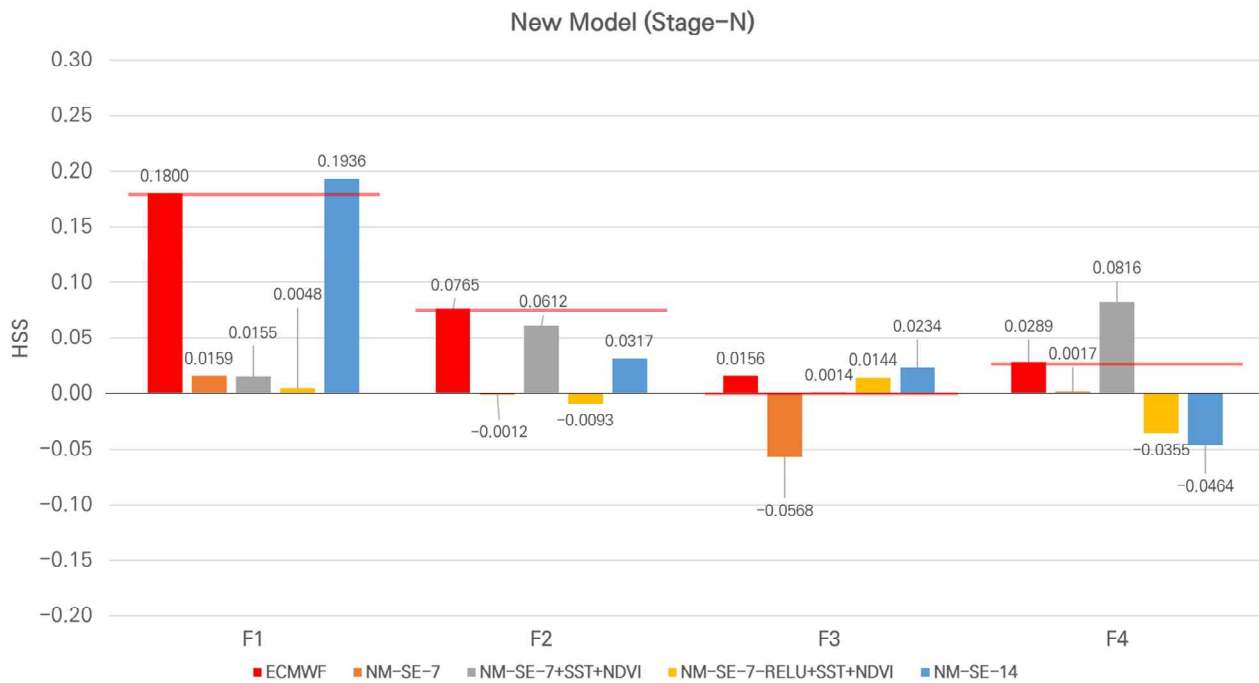


Figure 3.65. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model and 4 new models (NM-SE-7, NM-SE-7+SST+NDVI, NM-SE-7-RELU+SST+NDVI, NM-SE-14).

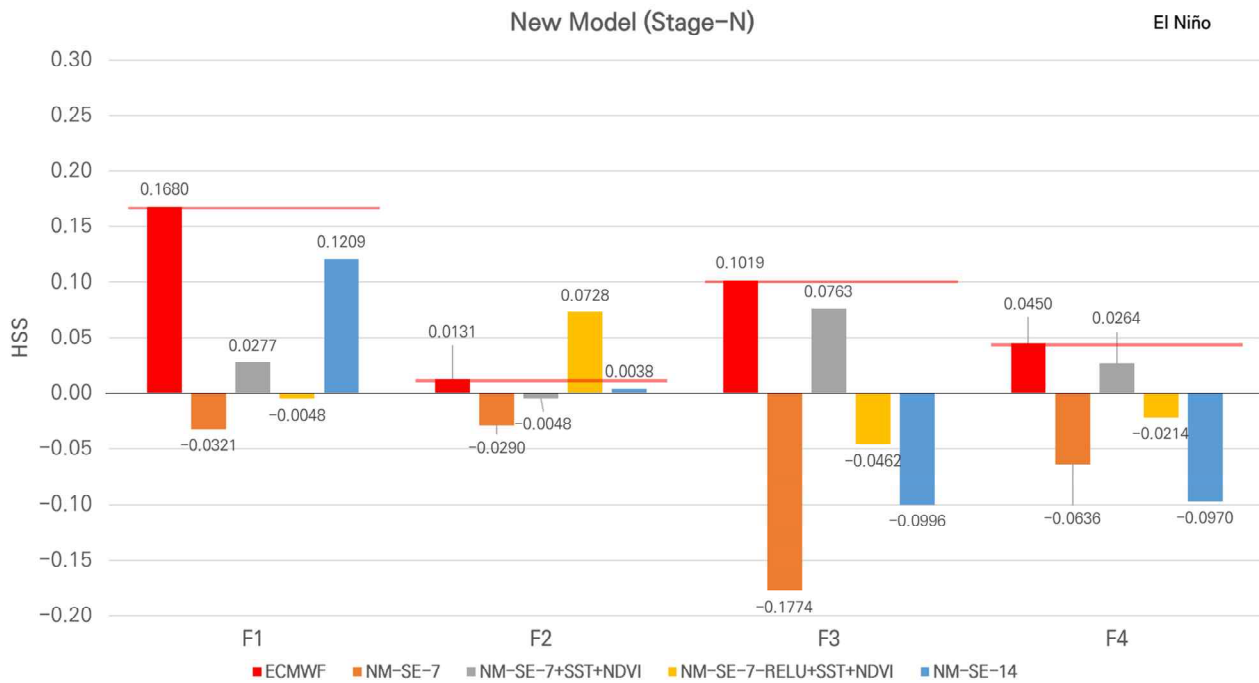


Figure 3.66. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model and 4 new models (NM-SE-7, NM-SE-7+SST+NDVI, NM-SE-7-RELU+SST+NDVI, NM-SE-14).

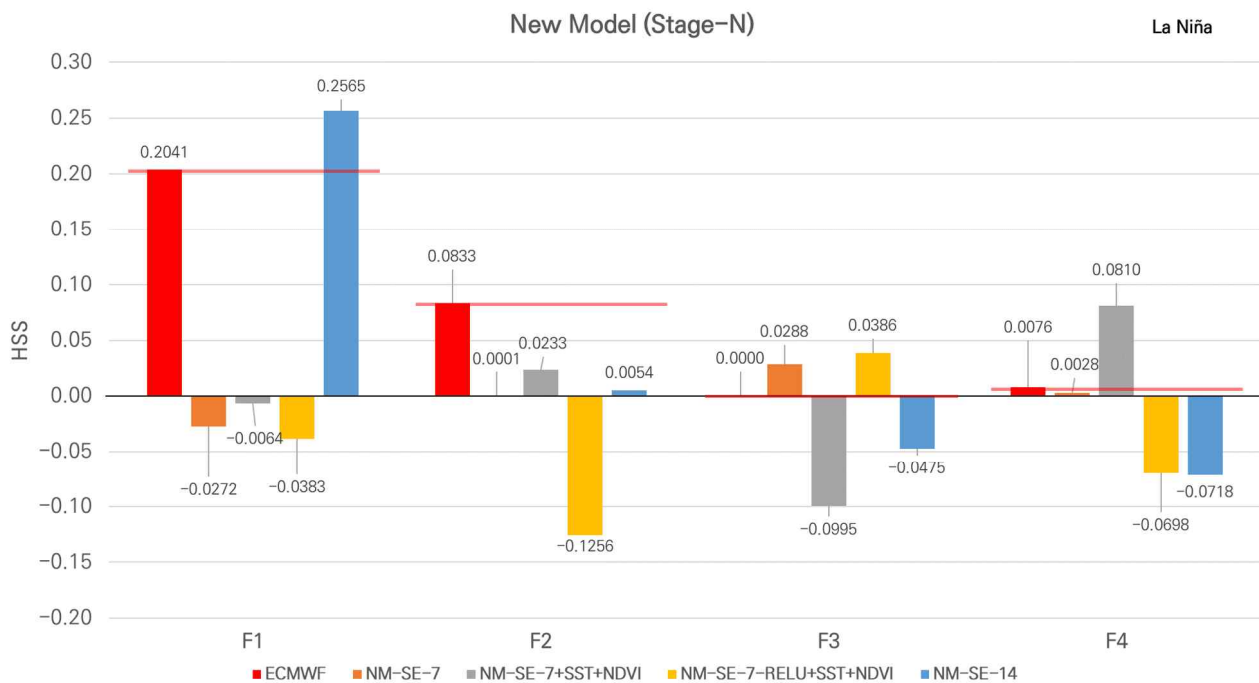


Figure 3.67. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model and 4 new models (NM-SE-7, NM-SE-7+SST+NDVI, NM-SE-7-RELU+SST+NDVI, NM-SE-14).

3.4.4. 앙상블 학습 기법을 이용한 기온 예측 모델 실험 및 결과

기온 1개월 기온 예측 모델과 입력변수 확장을 통한 기온 예측 모델 중 예측성이 낮은 모델을 이용하여 앙상블 학습 기법을 통해 기온 예측 모델의 예측성을 높이는 방법을 모색하기 위해 앙상블 모델을 구축하고 실험하였다(Table 3.8). 앙상블 기법은 스택킹 방식을 이용하였으며 EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7, EM-NP-14 등 총 5개의 모델을 구축하였다. EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7 등 4개 모델은 DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST 모델을 이용하였으며 EM-NP-14 모델은 DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK, DL-CBAM-14 모델을 이용하였다. 모델의 이름에서 EM은 앙상블 모델(ensemble model), NP는 new prediction layer, TR은 trainable parameter를 의미한다.

ECMWF와 앙상블 모델들의 HSS 평가 결과, 3주(F1) 예측은 EM-NP-14 모델, 4주(F2) 예측은 EM-NP-7 모델이 이 ECMWF보다 높은 예측성을 보였다. 5주(F3) 예측은 EM-TR-7 모델, 6주(F4) 예측은 EM-NP-14 모델을 제외한 4개 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.68). 전체 기간의 예측 순위는 1위 EM-NP-7, 2위 EM-NP-14, 공동 3위 ECMWF와 EM-7 순이었다. 엘니뇨 발생 시기의 3주(F1) 예측은 EM-NP-14 모델이 ECMWF보다 좋은 예측성을 보였다(Figure 3.69). 4주(F2) 예측은 EM-NP-7 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 EM-NP-14 모델이 ECMWF보다 좋은 예측성을 보였으며, 6주(F4) 예측에서는 EM-7 모델이 ECMWF보다 예측성이 높았다. 엘니뇨 발생 시기의 예측 순위는 1위 ECMWF, 2위 EM-NP-14, 3위 EM-NP-7 순으로 나타났다. 라니냐 발생 시기의 3주(F1) 예측은 EM-NP+FLT2-7 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.70). 4주(F2) 예측은 5개 모델 모두 ECMWF보다 낮은 예측성을 보였으며, 5주(F3) 예측은 EM-7, EM-NP-14 모델이 ECMWF보다 예측성이 높았다. 그리고 6주(F4) 예측은 EM-7 모델이 ECMWF보다 예측성이 높았다. 라니냐 발생 시기의 예측 순위는 1위 ECMWF, 2위 EM-7, 3위 EM-NP-7 순이었다.

Table 3.8. Summary of deep learning-based ensemble models for 1-month temperature prediction.

CASE NAME	Dataset	Epoch	Loss	Val loss	Input length
Pre-trained model					
EM-7	TDSV61-2_SHUFFLE	87	0.00182	0.00329	7
DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST					
EM-NP-7	TDSV61-2_SHUFFLE	63	0.00114	0.00323	7
DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST					
EM-NP+FLT2-7	TDSV61-2_SHUFFLE	43	0.00110	0.00321	7
DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST					
EM-TR-7	TDSV61-2_SHUFFLE	34	0.00074	0.00330	7
DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST					
EM-NP-14	TDSV32_SHUFFLE	43	0.00090	0.00299	14
DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK, DL-CBAM-14					

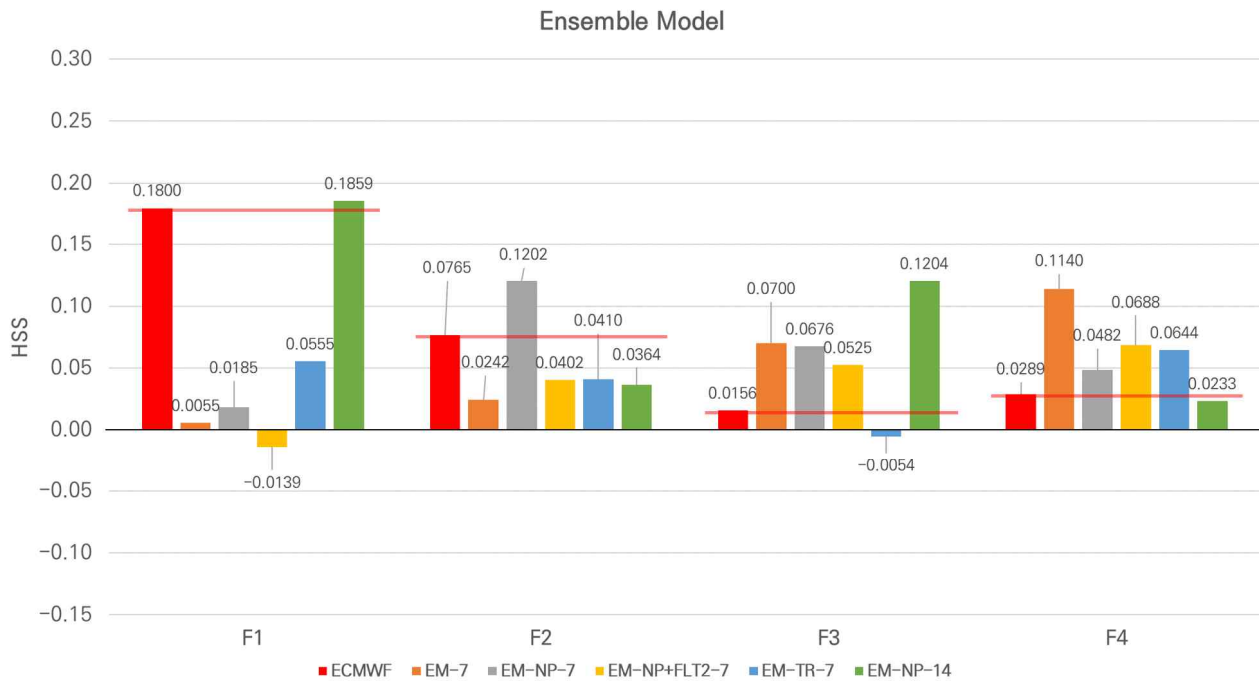


Figure 3.68. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S Model ECMWF S2S model and 5 ensemble models (EM-7, EM-NP-7, EM-NP-FLT2-7, EM-TR-7, EM-NP-14).

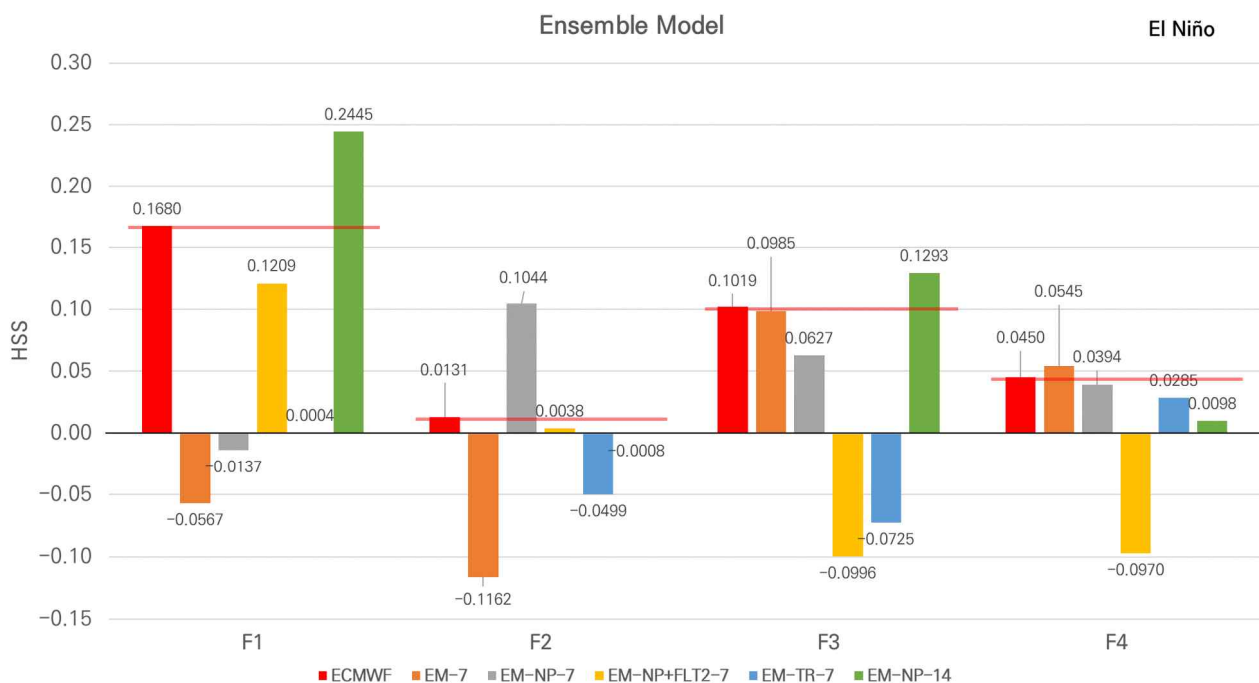


Figure 3.69. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S Model ECMWF S2S model and 5 ensemble models (EM-7, EM-NP-7, EM-NP-FLT2-7, EM-TR-7, EM-NP-14).

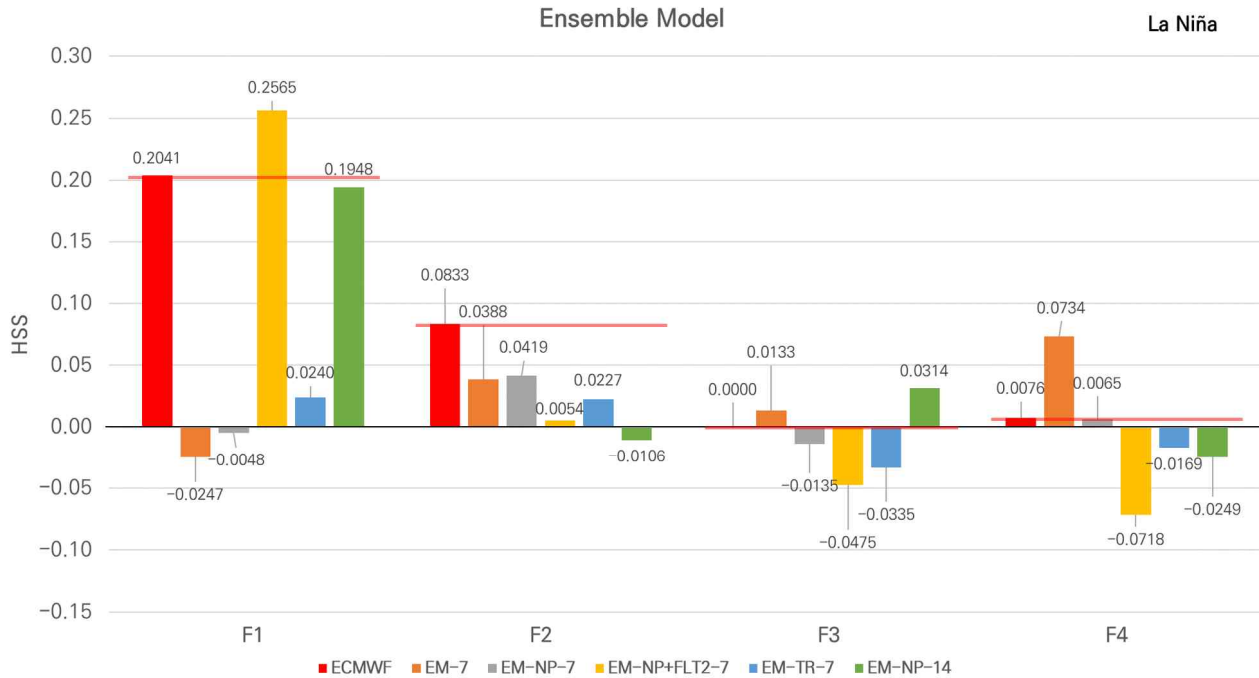


Figure 3.70. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S Model ECMWF S2S model and 5 ensemble models (EM-7, EM-NP-7, EM-NP-FLT2-7, EM-TR-7, EM-NP-14).

같은 학습 모델을 이용하는 EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7 모델과 EM-NP-14 모델을 앙상블 모델에 사용한 사전학습 모델과 각각 비교 분석하였다.

EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7 모델과 DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE-7+SST 사전학습 모델의 전체 기간 분석 결과, 3주(F1) 예측은 7개 모델 모두 ECMWF보다 예측성이 낮았으며, 4주(F2) 예측은 DL-SE+FLT1+BTLNCK-7+SST, DL-SE-7+SST, EM-NP-7 모델이 ECMWF보다 높은 예측성을 보였다. 5주(F3) 예측은 EM-TR-7 모델을 제외한 6개 모델이 ECMWF보다 예측성이 높았으며, 6주(F4) 예측은 EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.71). 예측성 순위는 1위 DL-SE+FLT1+BTLNCK-7+SST, 2위 EM-NP-7, 공동 3위는 EM-7와 ECMWF 순이었다.

엘니뇨 발생 시기의 3주(F1) 예측은 7개 모델 모두 ECMWF보다 낮은 예측성을 보였다(Figure 3.72). 4주(F2) 예측은 DL-SE-FLT1-BTLNCK-7+SST, EM-NP-7 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 DL-SE-FLT1-BTLNCK-7+SST 모델이 ECMWF보다 높은 예측성을 보였으며, 6주(F4) 예측에서는 EM-7 모델이 ECMWF보다 높은 예측성을 보였다. 엘니뇨 발생 시기의 예측 순위는 1위 ECMWF, 2위 DL-SE+FLT1+BTLNCK-7+SST, 공동 3위 DL-CBAM-7+SST+NDVI, EM-7 순으로 나타났다.

라니냐 발생 시기의 3주(F1) 예측은 EM-NP+FLT2-7 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.73). 4주(F2) 예측은 DL-SE-7+SST 모델이 ECMWF보다 높은 예측성을 보였으며, 5주(F3) 예측은 DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE-7+SST, EM-7 모델이 ECMWF보다 예측성이 높았다. 그리고 6주(F4) 예측은 EM-7 모델이 ECMWF보다 높은

예측성을 보였다. 라니냐 발생 시기의 예측 순위는 1위 ECMWF, 2위 EM-7, 3위 EM-NP-7 순이었다.

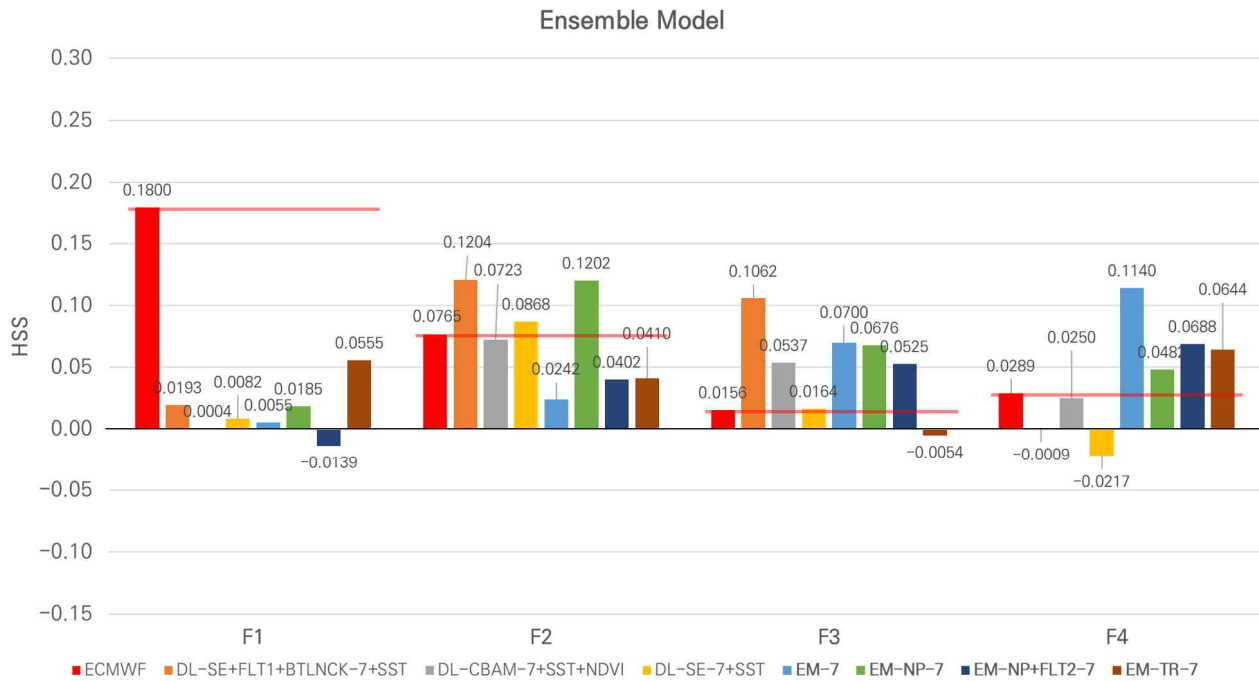


Figure 3.71. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE+7+SST), and 4 ensemble models (EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7).

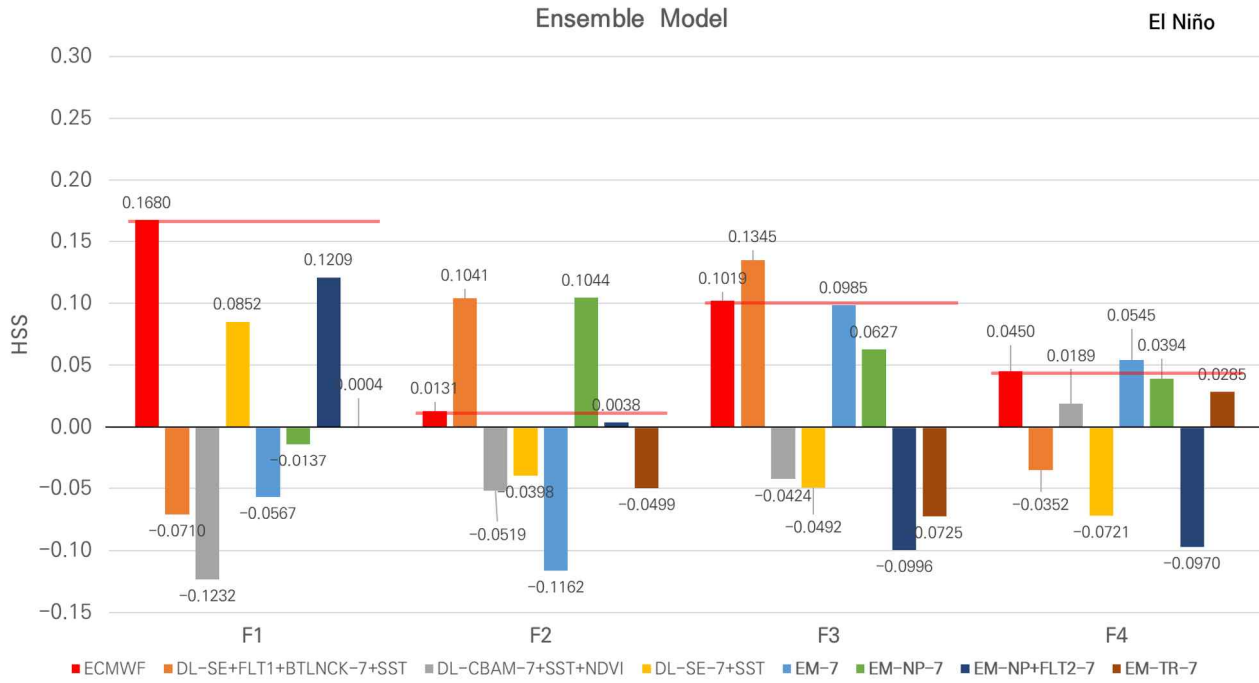


Figure 3.72. Summary of HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE+7+SST), and 4 ensemble models (EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7).

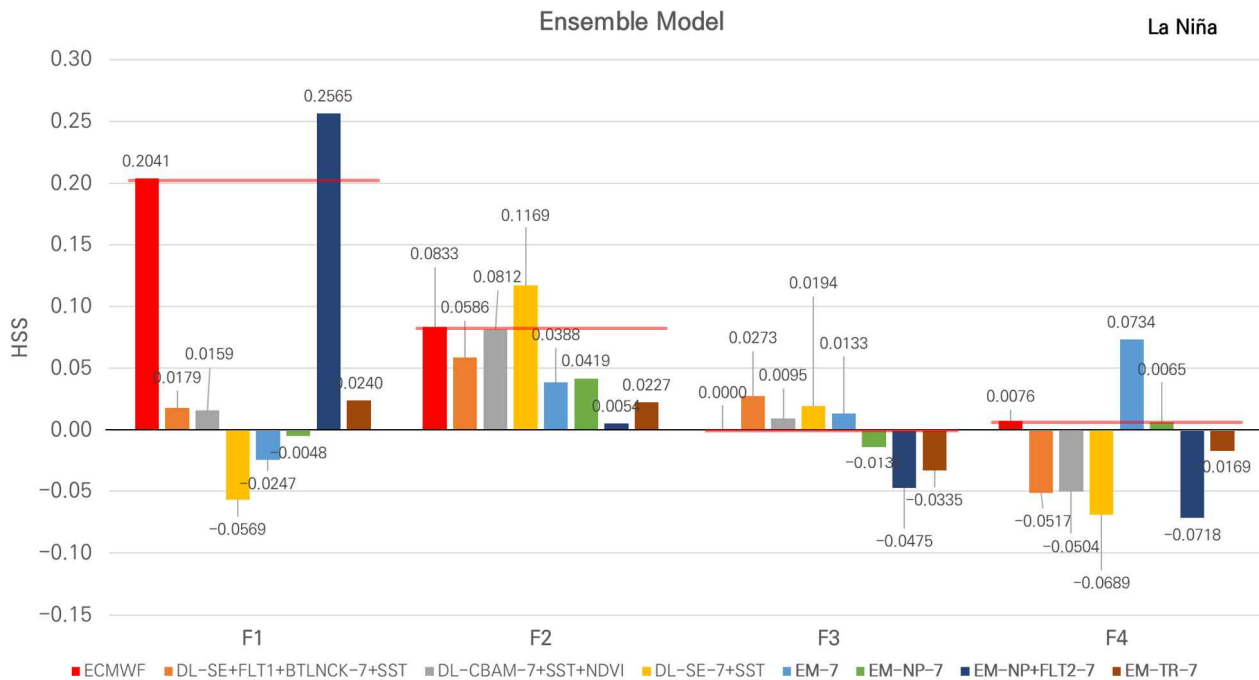


Figure 3.73. Summary of HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+FLT1+BTLNCK-7+SST, DL-CBAM-7+SST+NDVI, DL-SE+7+SST), and 4 ensemble models (EM-7, EM-NP-7, EM-NP+FLT2-7, EM-TR-7).

EM-NP-14 모델과 DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14 등 사용된 사전 학습모델의 전체 기간 분석 결과, 3주(F1) 예측은 DL-CBAM-14 모델을 제외한 3개 모델이 ECMWF보다 예측성이 높았으며, 4주(F2) 예측은 DL-SE+BIGDATA-14 모델이 ECMWF보다 높은 예측성을 보였다. 5주(F3) 예측은 4개 모델 모두 ECMWF보다 예측성이 높았으며, 6주(F4) 예측은 DL-CBAM-14 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.74). 예측성 순위는 1위 DL-SE+BIGDATA-14, 2위 DL-CBAM+BIGDATA+FLT2+BTLNCK-14, 공동 3위는 ECMWF와 EM-NP-14 순이었다.

엘니뇨 발생 시기의 3주(F1) 예측은 DL-SE+BIGDATA-14 모델을 제외한 3개 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.75). 4주(F2) 예측은 DL-SE+BIGDATA-14, DL-CBAM-14 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 EM-NP-14 모델이 ECMWF보다 높은 예측성을 보였으며, 6주(F4) 예측에서는 DL-CBAM+BIGDATA+FLT2+BTLNCK-14 모델이 ECMWF보다 높은 예측성을 보였다. 엘니뇨 발생 시기의 예측 순위는 1위 EM-NP-14, 공동 2위 ECMWF와 DL-CBAM-14 순으로 나타났다.

라니냐 발생 시기의 3주(F1) 예측은 EM-NP-14 모델을 제외한 3개 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.76). 4주(F2) 예측은 DL-SE+BIGDATA-14 모델이 ECMWF보다 높은 예측성을 보였으며, 5주(F3) 예측은 4개 모델 모두 ECMWF보다 예측성이 높았다. 그리고 6주(F4) 예측은 4개 모델 모두 ECMWF보다 낮은 예측성을 보였다. 라니냐 발생 시기의 예측 순위는 1위 DL-SE+BIGDATA-14, 2위 DL-CBAM+BIGDATA+FLT2+BTLNCK-14, 3위 ECMWF 순이었다.

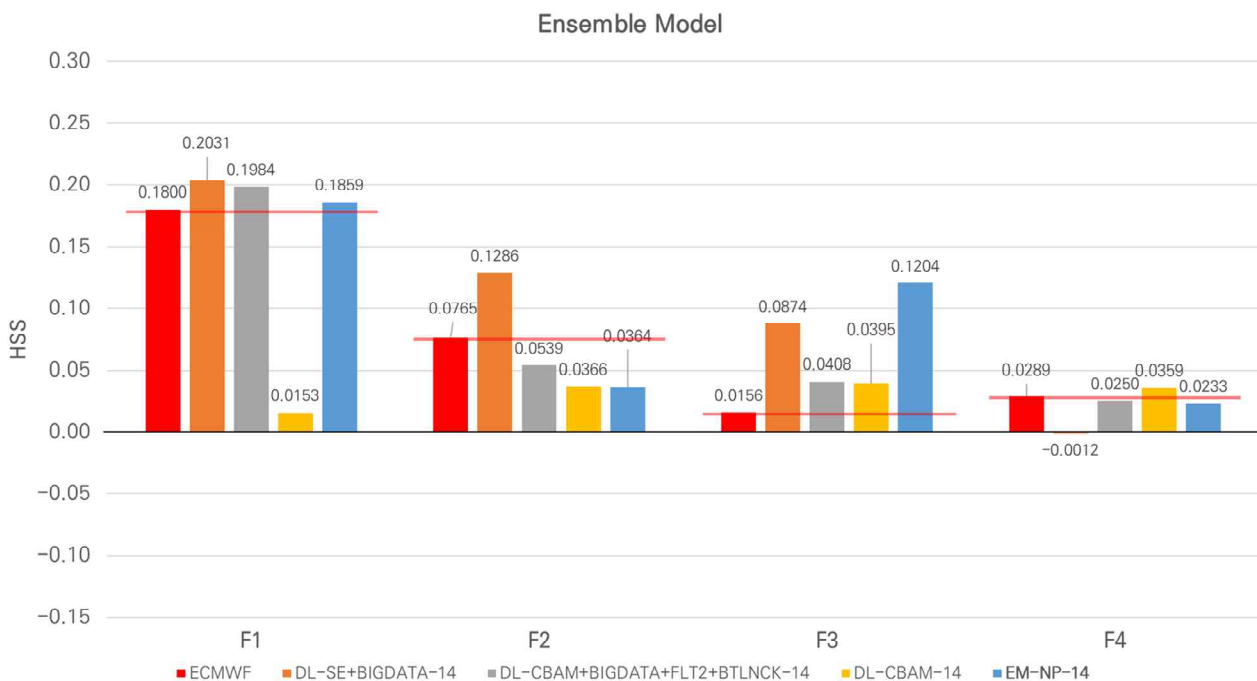


Figure 3.74. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14), and 1 ensemble model (EM-NP-14).

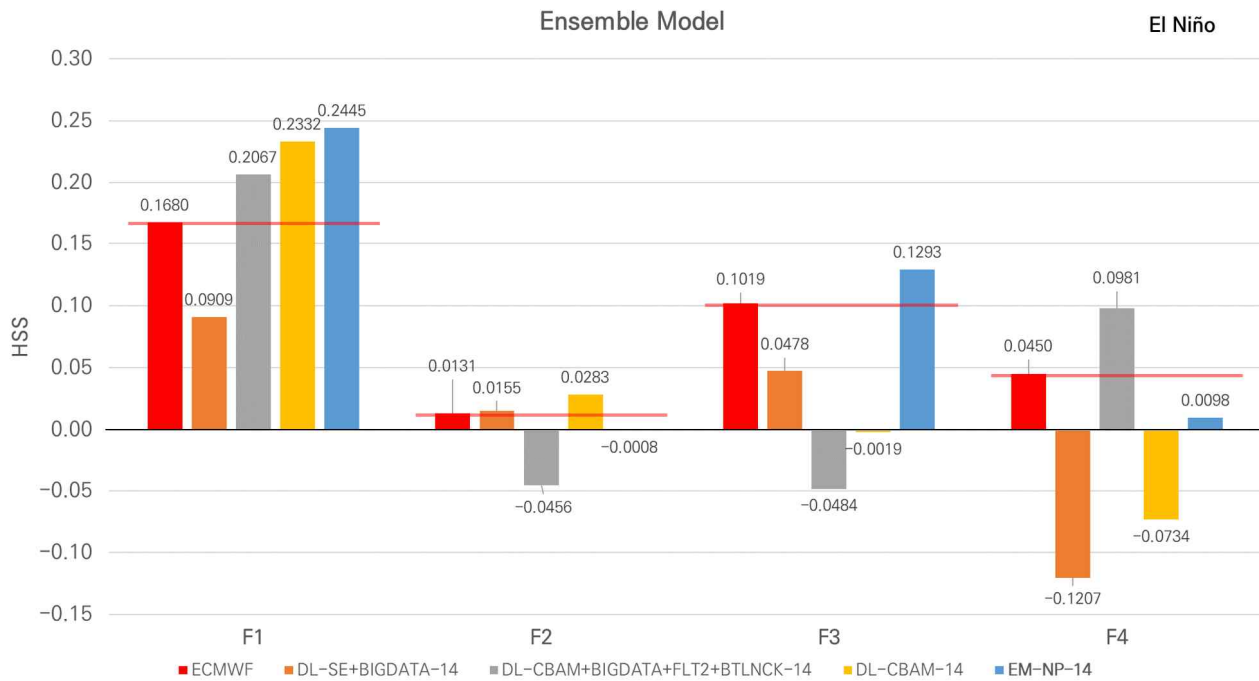


Figure 3.75. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14), and 1 ensemble model (EM-NP-14).

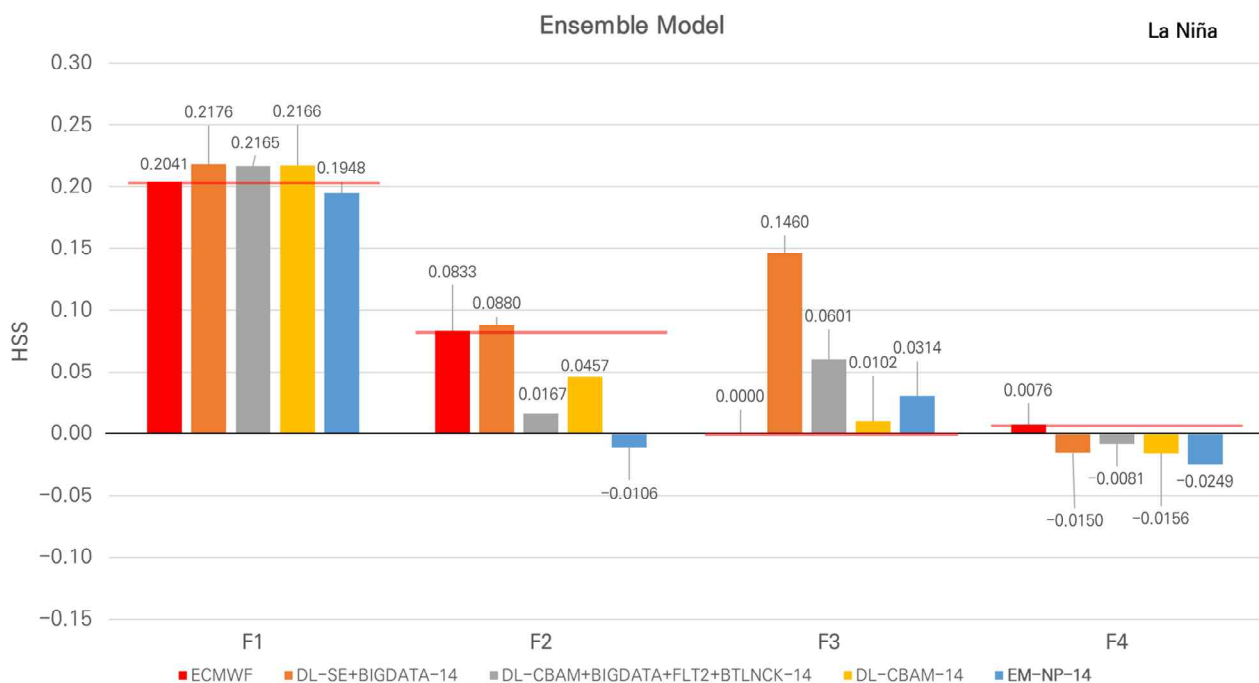


Figure 3.76. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model, 3 ensemble member models (DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK-14, DL-CBAM-14), and 1 ensemble model (EM-NP-14).

3.4.5. N 스텝 예측 방법을 이용한 기온 예측 모델 실험 및 결과

N 스텝 예측 모델은 기준 1개월 기온 예측 모델, 입력 자료 확장을 고려한 1개월 기온 예측 모델 중에서 3주 예측성이 높은 모델과 4-6주 높은 예측 모델을 순차적으로 연결한 모델이다. 즉, A 예측모델로 3주를 예측하고, 그 결과를 B 예측모델의 입력 자료로 활용하여 4-6주를 예측한다.

N 스텝 예측 모델은 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 등 2개 모델을 구축하였다. nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델은 DL-SE+BIGDATA+FLT2+BTLNCK-14와 NM-SE-7+SST+NDVI 모델, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델은 DL-SE+BIGDATA+FLT2+BTLNCK-14와 NM-SE-7 모델을 이용한다.

Table 3.9. Summary of deep learning-based n-step prediction models for 1-month temperature prediction.

CASE NAME	Dataset	Epoch	Loss	Val loss	Input length
Pre-trained model					
nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI	TDSV32_SHUFFLE	30	0.00989	0.09943	14
DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI					
nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7	TDSV32_SHUFFLE	30	0.00125	0.00321	14
DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7					

먼저, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델과 N 스텝 모델에 사용한 2개 모델의 비교 결과, 3주(F1) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델이 ECMWF보다 좋은 예측성을 보였으며, 4주(F2) 예측은 3개 모델 모두 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델이 ECMWF보다 좋은 예측성을 보였으며, 6주(F4) 예측은 3개 모델 모두 ECMWF보다 좋은 예측성을 보였다(Figure 3.77). 전체 기간의 예측 순위는 1위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 2위 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI, 3위 ECMWF 순이었다.

엘니뇨 발생 시기의 3주(F1) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14와 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.78). 4주(F2) 예측은

nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 모든 모델이 ECMWF보다 낮은 예측성을 보였으며, 6주(F4) 예측에서는 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델이 ECMWF보다 높은 예측성을 보였다. 엘니뇨 발생 시기의 예측 순위는 1위 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI, 2위 ECMWF, 3위 DL-SE+BIGDATA+FLT2+BTLNCK-14 순으로 나타났다.

라니냐 발생 시기의 3주(F1) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14와 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델이 ECMWF보다 높은 예측성을 보였다(Figure 3.79). 4주(F2)와 5주(F3) 예측은 모든 모델이 ECMWF보다 낮은 예측성을 보였다. 6주 예측은 NM-SE-7+SST+NDVI 모델이 ECMWF보다 예측성이 높았다. 라니냐 발생 시기의 예측 순위는 1위 ECMWF, 2위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 공동 3위 NM-SE-7+SST+NDVI, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 순이었다.

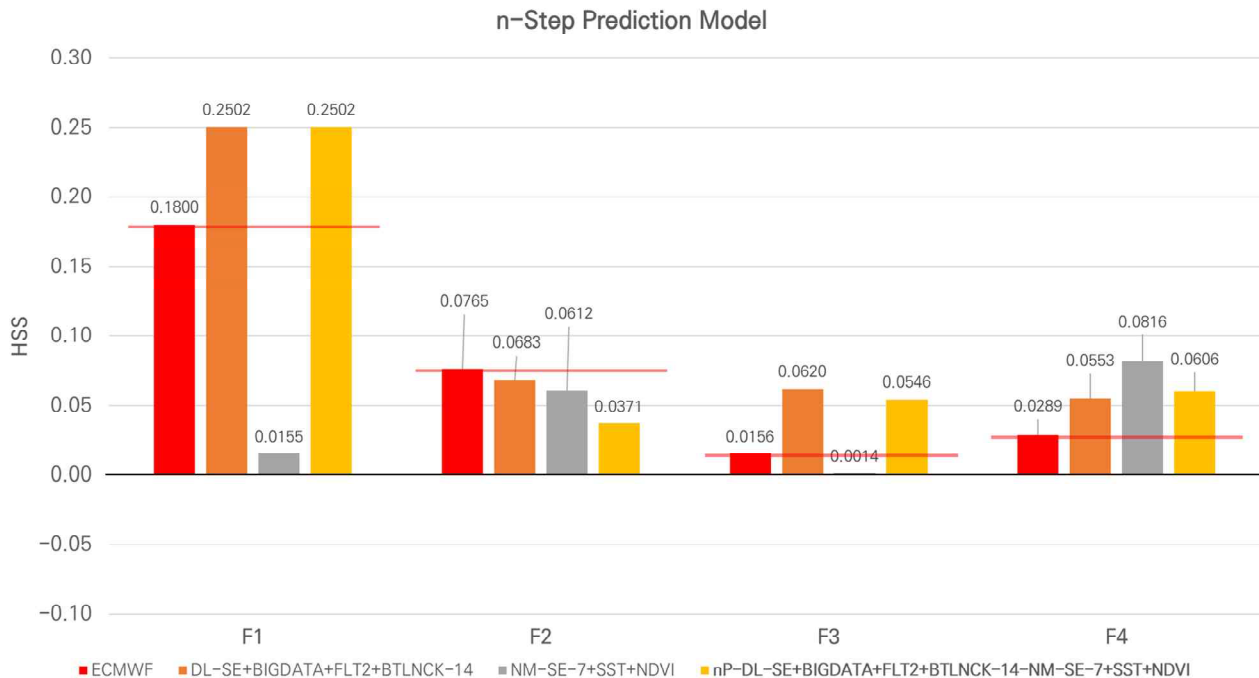


Figure 3.77. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI), and 1 n-step prediction model (, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI).

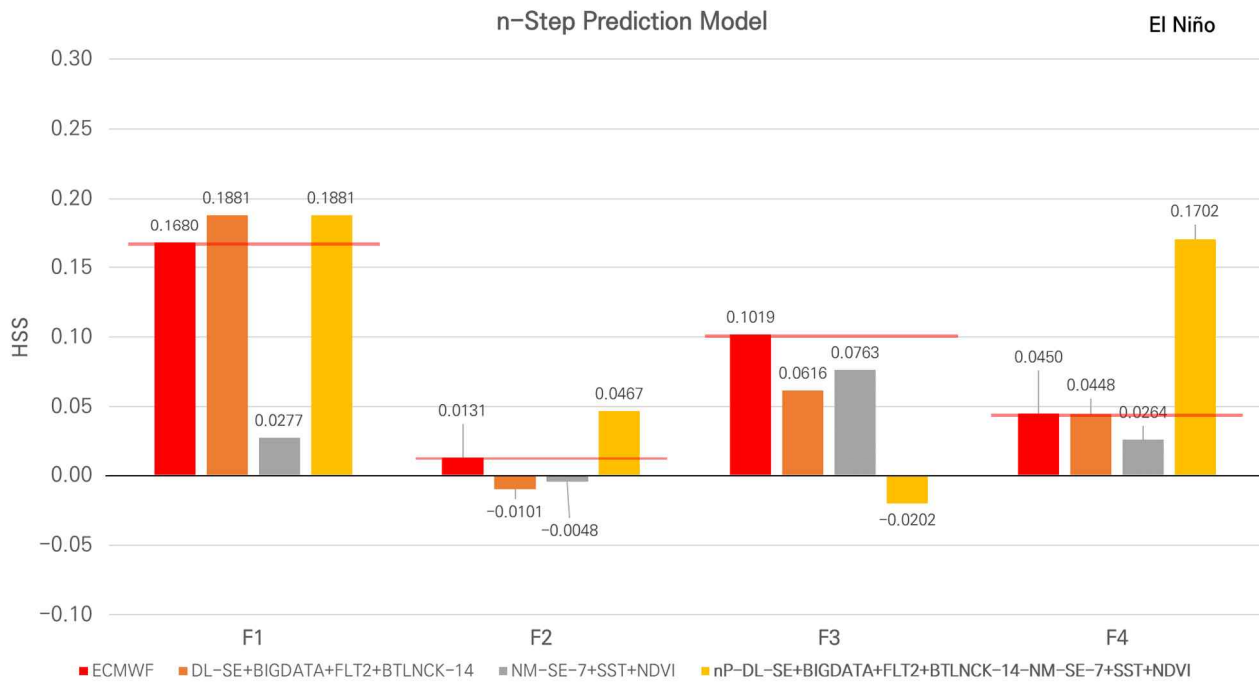


Figure 3.78. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI), and 1 n-step prediction model (, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI).

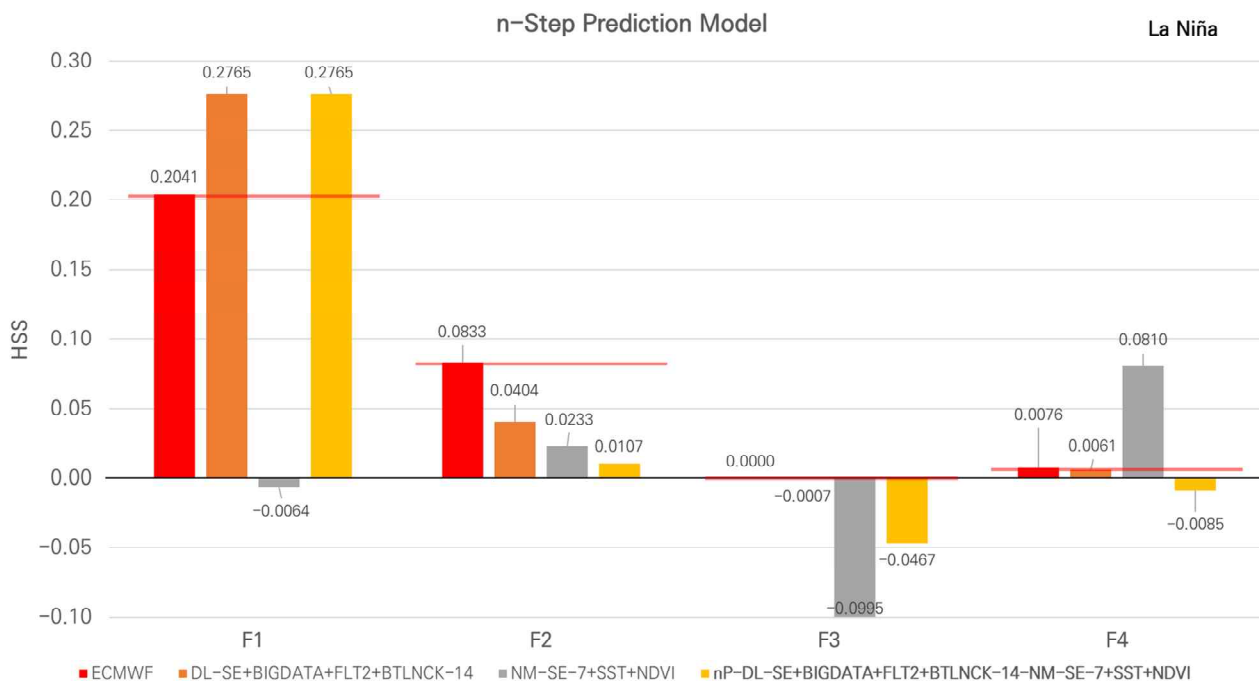


Figure 3.79. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI), and 1 n-step prediction model (, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI).

nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델과 N 스텝 모델에 사용한 2개 모델의 비교 결과, 3주(F1) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델이 ECMWF보다 좋은 예측성을 보였으며, 4주(F2) 예측은 3개 모델 모두 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14, nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델이 ECMWF보다 좋은 예측성을 보였으며, 6주(F4) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14와 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델이 ECMWF보다 좋은 예측성을 보였다 (Figure 3.80). 전체 기간의 예측 순위는 1위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 2위 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7, 3위 ECMWF 순이었다.

엘니뇨 발생 시기의 3주(F1) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14와 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델이 ECMWF보다 높은 예측성을 보였다 (Figure 3.81). 4주(F2) 예측은 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3)와 6주(F4) 예측은 모든 모델이 ECMWF보다 낮은 예측성을 보였다. 엘니뇨 발생 시기의 예측 순위는 1위 ECMWF, 2위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 3위 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 순으로 나타났다.

라니냐 발생 시기의 3주(F1) 예측은 DL-SE+BIGDATA+FLT2+BTLNCK-14와 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7 모델이 ECMWF보다 높은 예측성을 보였다 (Figure 3.82). 4주(F2) 예측은 모든 모델이 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 NM-SE-7 모델이 ECMWF보다 높은 예측성을 보였으며, 6주(F4) 예측은 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 모델이 ECMWF보다 예측성이 높았다. 라니냐 발생 시기의 예측 순위는 1위 ECMWF, 2위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 3위 nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI 순이었다.

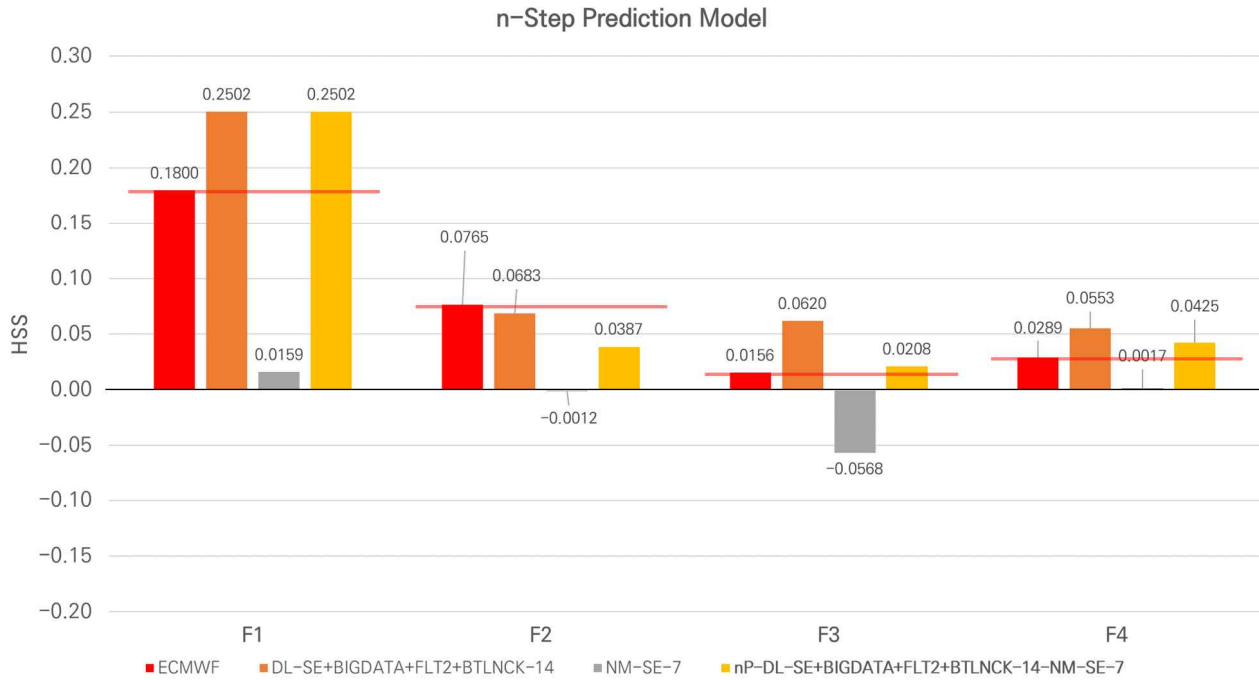


Figure 3.80. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7), and 1 n-step prediction model (nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7).

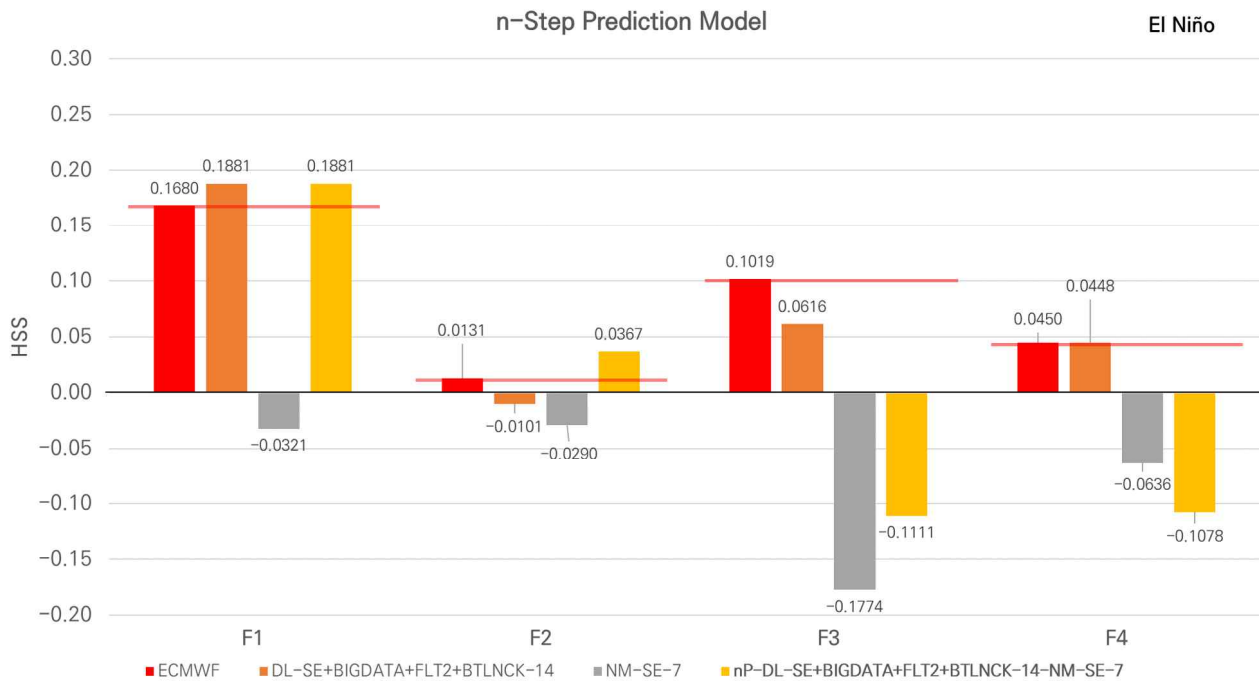


Figure 3.81. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7), and 1 n-step prediction model (nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7).

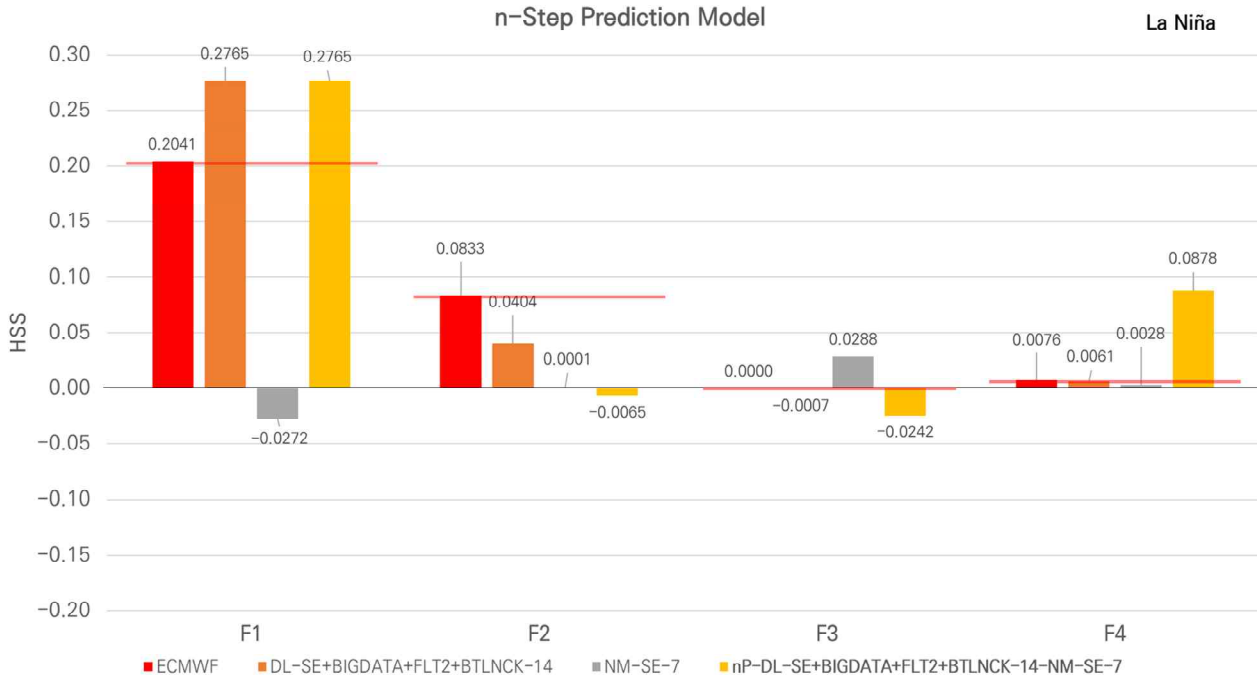


Figure 3.82. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model, 2 n-step prediction member models (DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7), and 1 n-step prediction model (nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7).

3.4.6. 재학습 방법을 이용한 기온 예측 모델 실험

재학습 방법을 이용한 기온 예측 모델은 기준 1개월 기온 예측 모델과 입력자료 확장을 고려한 기온 예측 모델 중에서 모델을 선정하고 학습률(learning rate)를 조정하여 재학습을 통해 구축한 모델이다. RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14, RL-DL-SE-14+SST+NDVI 등을 2개 모델을 구축하고 재학습에 사용된 모델과 비교 평가하였다.

Table 3.10. Summary of 1-month temperature prediction deep learning models.

CASE NAME	Dataset	Epoch	Loss	Val loss	Input length
Pre-trained model					
RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14	TDSV32_SHUFFLE	30	0.00989	0.09943	14
DL-CBAM+BIGDATA+FLT2+BTLNCK-14					
RL-DL-SE-14+SST+NDVI	TDSV61-3_SHUFFLE	30	0.00125	0.00321	14
DL-SE-14+SST+NDVI					

RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14 모델의 전체 기간 평가 결과, 3주(F1) 예측에서는 2개 모델 모두 ECMWF보다 높은 예측성을 보였으며, 4주(F2) 예측은 2개 모델 모두

ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 2개 모델 모두 ECMWF보다 높은 예측성을 보였으며, 6주(F4) 예측은 RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14 모델이 ECMWF보다 높은 예측성을 보였다. 예측 순위는 1위 RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14, 공동 2위 ECMWF, DL-CBAM+BIGDATA+FLT2+BTLNCK-14 순이었다(Figure 3.39).

엘니뇨 발생 시기의 3주(F1) 예측은 2개 모델 모두 ECMWF보다 높은 예측성을 보였다(Figure 3.40). 4주(F2) 예측은 RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14 모델이 ECMWF보다 좋은 예측성을 보였다. 5주(F3) 예측은 2개 모델 모두 ECMWF보다 낮은 예측성을 보였으며, 6주(F4) 예측은 DL-CBAM+BIGDATA+FLT2+BTLNCK-14 모델이 ECMWF보다 예측성이 높았다. 엘니뇨 발생 시기의 예측 순위는 ECMWF S2S 모델을 포함해 모두 1위로 나타났다.

라니냐 발생 시기의 3주(F1) 예측은 2개 모델 모두 ECMWF보다 낮은 예측성을 보였다(Figure 3.41). 4주(F2) 예측은 모든 모델이 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 DL-CBAM+BIGDATA+FLT2+BTLNCK-14 모델이 ECMWF보다 높은 예측성을 보였으며, 6주(F4) 예측은 2개 모델 모두 ECMWF보다 예측성이 낮았다. 라니냐 발생 시기의 예측 순위는 공동 1위 ECMWF와 DL-CBAM+BIGDATA+FLT2+BTLNCK-14, 2위 RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14 순이었다.

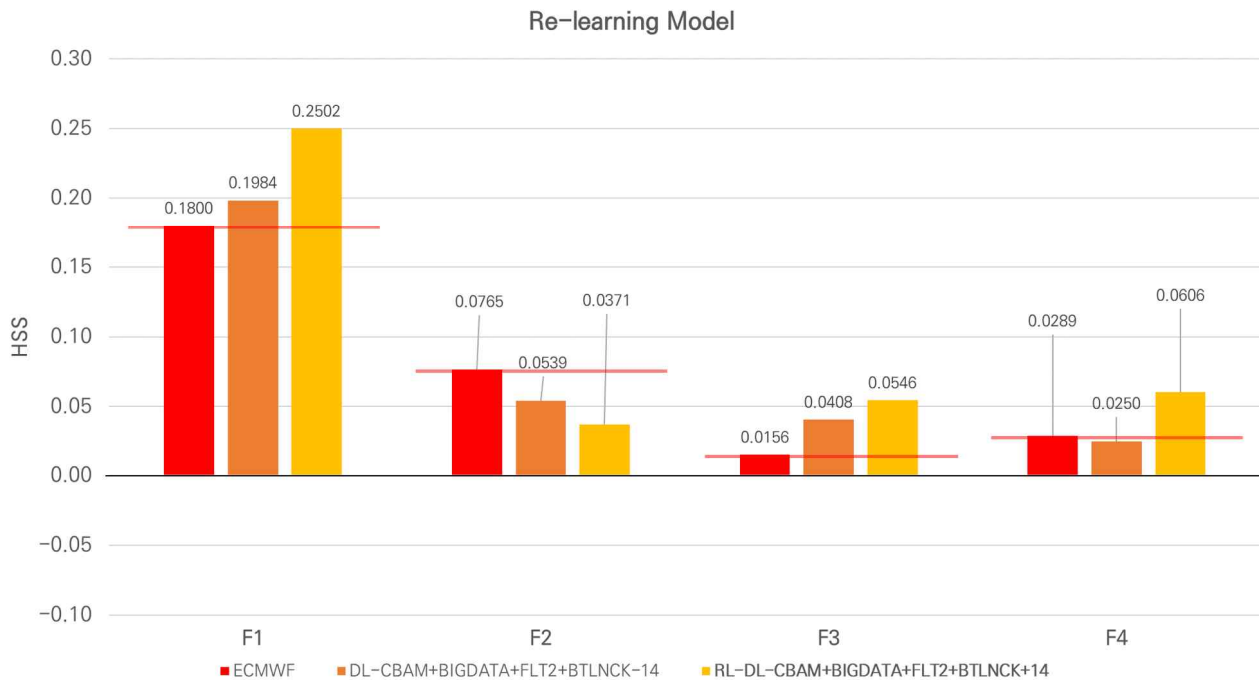


Figure 3.83. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 1 pretrained model (DL-CBAM-BIGDATA+FLT2+BTLNCK-14), and 1 re-learning model (RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14).

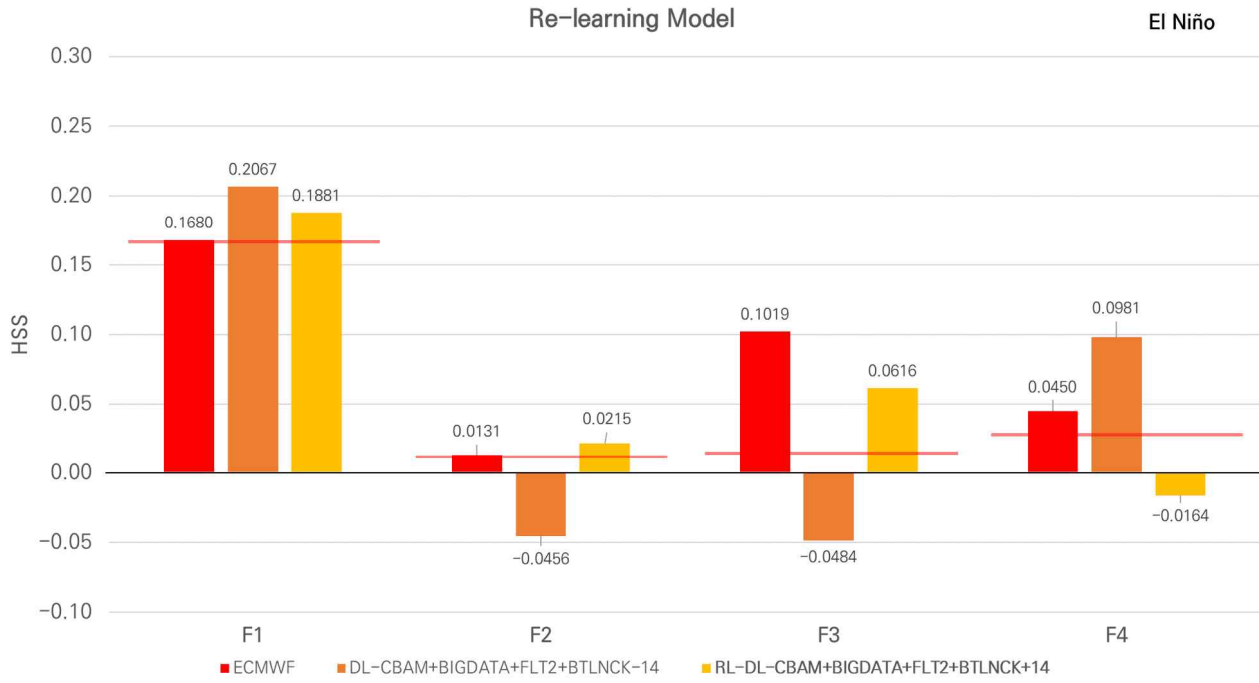


Figure 3.84. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model, 1 pretrained model (DL-CBAM-BIGDATA+FLT2+BTLNCK-14), and 1 re-learning model (RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14).

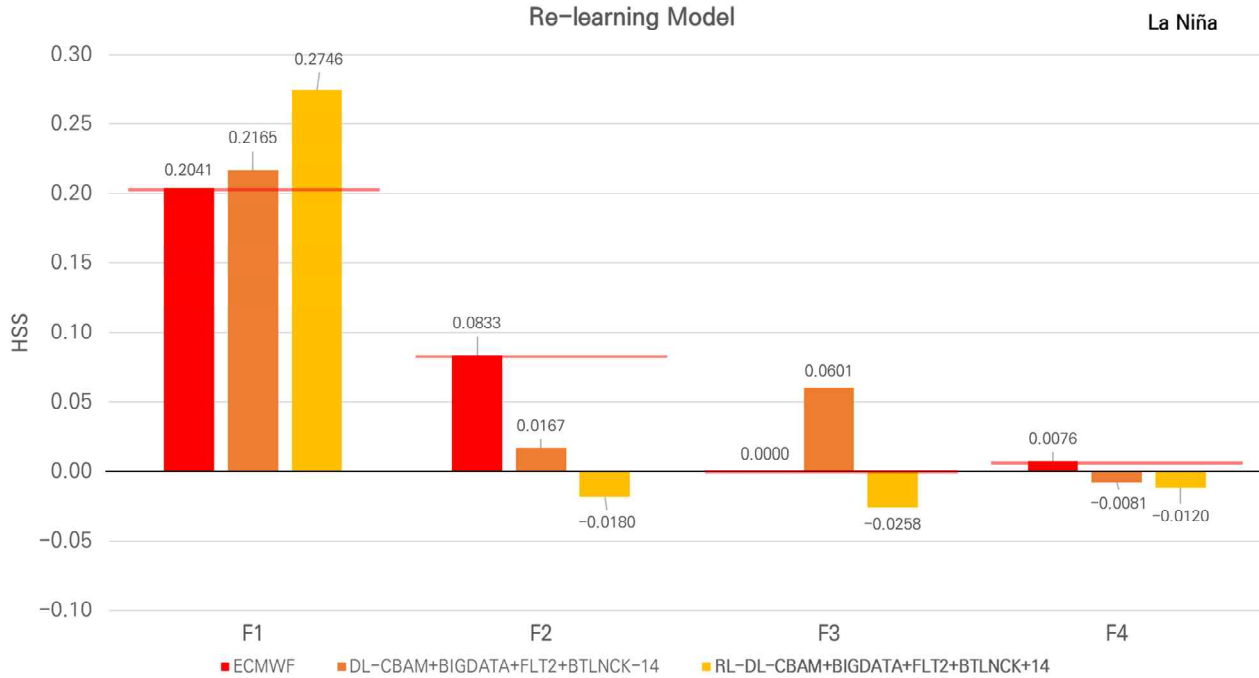


Figure 3.85. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model, 1 pretrained model (DL-CBAM-BIGDATA+FLT2+BTLNCK-14), and 1 re-learning model (RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14).

RL-DL-SE-14+SST+NDVI 모델의 전체 기간 평가 결과, 3주(F1) 예측에서는 2개 모델 모두 ECMWF보다 높은 예측성을 보였으며, 4주(F2) 예측은 2개 모델 모두 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 2개 모델 모두 ECMWF보다 높은 예측성을 보였으며, 6주(F4) 예측은 2개 모델 모두 ECMWF보다 낮은 예측성을 보였다. 예측 순위는 1위 DL-SE-14+SST+NDVI, 2위 ECMWF, 3위 RL-DL-SE-14+SST+NDVI 순이었다(Figure 3.86).

엘니뇨 발생 시기의 3주(F1) 예측은 2개 모델 모두 ECMWF보다 높은 예측성을 보였다(Figure 3.87). 4주(F2) 예측은 2개 모델 모두 ECMWF보다 낮은 예측성을 보였다. 5주(F3) 예측은 RL-DL-SE-14+SST+NDVI 모델이 ECMWF보다 낮은 예측성을 보였으며, 6주(F4) 예측은 2개 모델 모두 ECMWF보다 예측성이 높았다. 엘니뇨 발생 시기의 예측 순위는 1위 RL-DL-SE-14+SST+NDVI, 2위 ECMWF, 3위 DL-SE-14+SST+NDVI 모델 순이었다.

라니냐 발생 시기의 3주(F1)와 4주(F2) 예측은 2개 모델 모두 ECMWF보다 낮은 예측성을 보였다(Figure 3.88). 5주(F3) 예측은 2개 모델 모두 ECMWF보다 높은 예측성을 보였으며, 6주(F4) 예측은 DL-SE-14+SST+NDVI 모델이 ECMWF보다 예측성이 낮았다. 라니냐 발생 시기의 예측 순위는 1위 ECMWF, 2위 DL-SE-14+SST+NDVI, 3위 RL-DL-SE-14+SST+NDVI 순이었다.

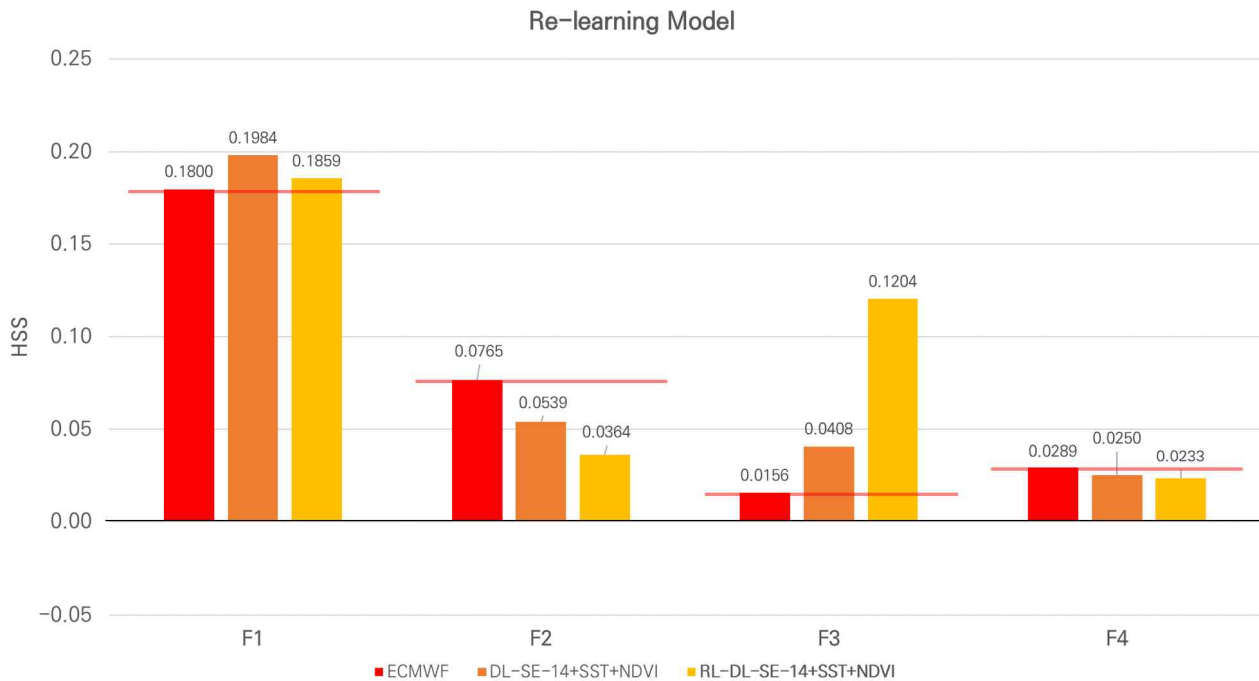


Figure 3.86. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4): (from left) ECMWF S2S model, 1 pretrained model (DL-SE-14+SST+NDVI), and 1 re-learning model RL-DL-SE-14+SST+NDVI).



Figure 3.87. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during El Niño events: (from left) ECMWF S2S model, 1 pretrained model (DL-SE-14+SST+NDVI), and 1 re-learning model (RL-DL-SE-14+SST+NDVI).

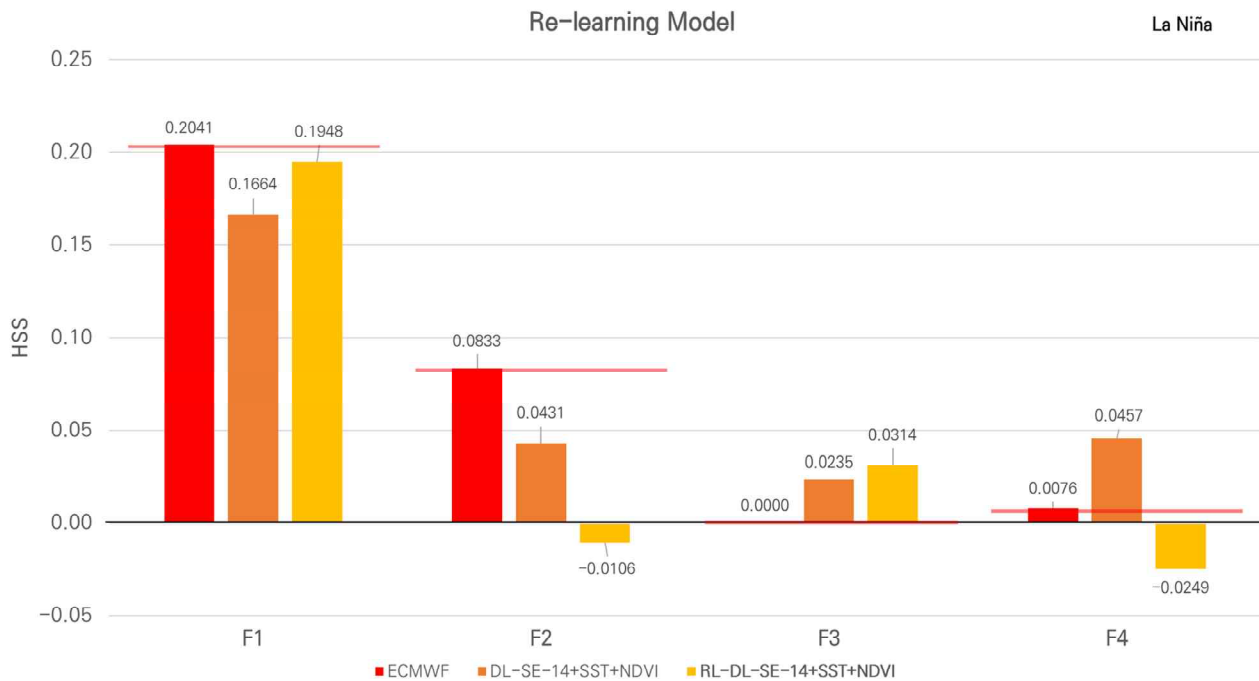


Figure 3.88. Summary of the HSS assessment for lead time of 3 to 6 weeks (F1 to F4) during La Niña events: (from left) ECMWF S2S model, 1 pretrained model (DL-SE-14+SST+NDVI), and 1 re-learning model (RL-DL-SE-14+SST+NDVI).

3.4.7. 종합분석

동아시아지역 3-6주 기온 예측 모델 개발을 위해 기준 1개월 기온 예측 모델(Table 3.4), 입력 변수 확장을 통한 기온 예측 모델(Table 3.6), 최신 딥러닝 기법을 이용한 기온 예측 모델(Table 3.7), 앙상블 학습 기법을 이용한 기온 예측 모델(Table 3.8), N 스텝 예측 방법을 이용한 기온 예측 모델(Table 3.9), 재학습 방법을 이용한 기온 예측 모델(Table 3.10)을 구축하였다. NDVI 자료 기간을 고려하여 2018~2019년과 2018~2022년으로 평가 기간을 구분하고 전체 모델을 ECMWF와 비교하였다(Table A.1, Figure 89, 90, 91, 92). 각 평가 기간의 예측성 순위로 2018~2019년은 1위 DL-SE-7+SST+NDVI (Appendix B.1), 공동 2위 DL-SE+FLT2+BTLNCK-7, NM-SE-7, EM-NP-14, nP-DL-SE+BIGDAA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI, RL-DL-SE-14+SST+NDVI였으며, ECMWF는 25위였다. 2018~2019년은 엘니뇨와 라니냐 기간은 평가 기간 내 엘니뇨와 라니냐 발생 자료가 부족하여 제외하였다. 2018~2022년은 전체 평가 기간에서 1위 DL-SE-14 (Appendix B.2), 2위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 3위 DL-SE+BIGDATA-14였으며, ECMWF는 18위였다. 엘니뇨 발생 시기는 공동 1위 DL-SE-7+SST+NDVI, DL-SE+FLT2+BTLNCK-7. 3위 DL-SE+BIGDATA-14이며 ECMWF는 공동 7위였다.

2018~2019년은 ALO 시스템을 고려한 입력자료 확장을 고려한 모델들이 주로 좋은 예측성을 보였다. 이는 지면자료(NDVI)와 해양자료(SST)가 4~6주 예측에 긍정적인 영향을 준 것으로 판단된다. 반면 2018~2022년은 NDVI 자료 구축 기간이 2018~2019년으로 2020년부터는 2019년 자료를 이용하였기 때문에 T2M 자료만 이용한 모델에 비해 상대적으로 예측성이 떨어지는 경향을 보였다.

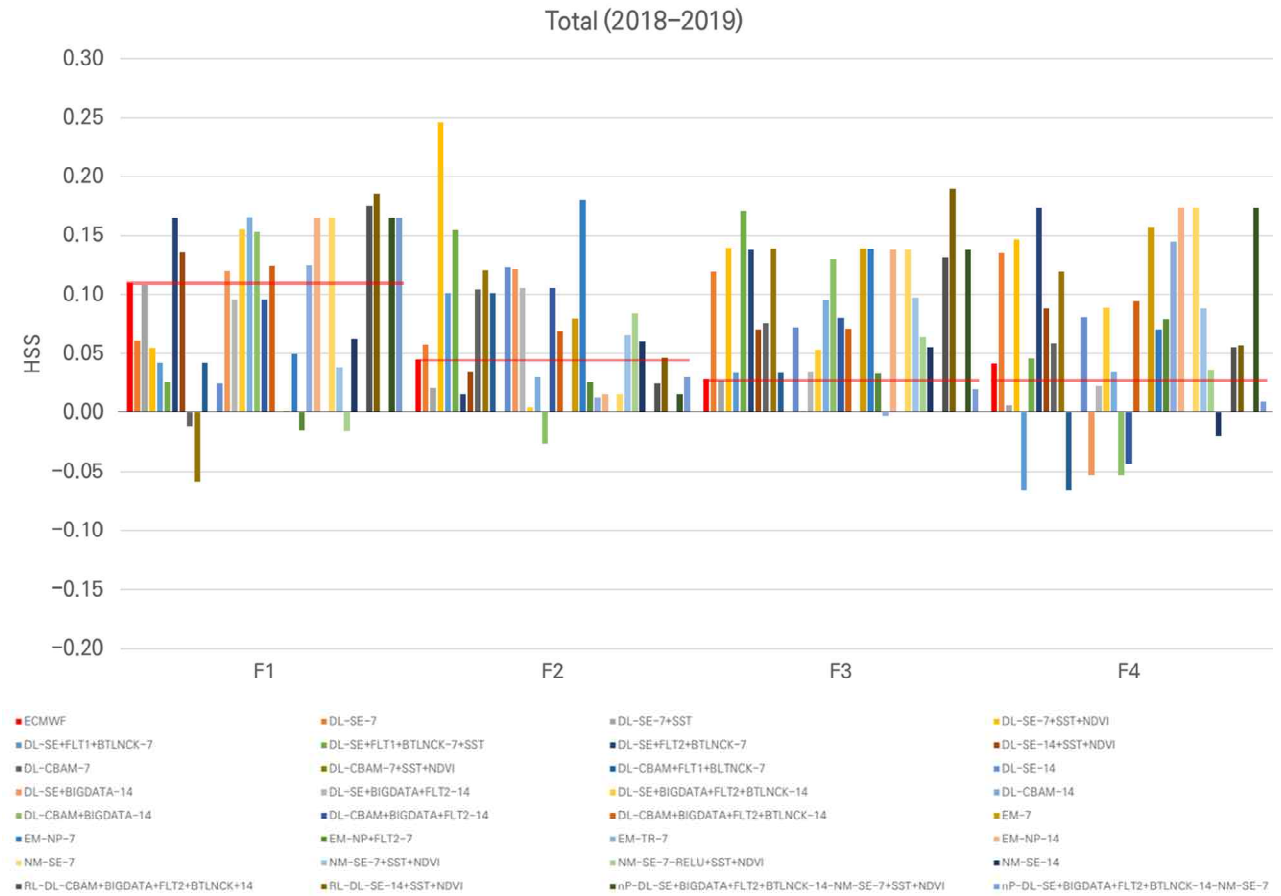


Figure 3.89. Results of deep learning-based models in 2018-2019.

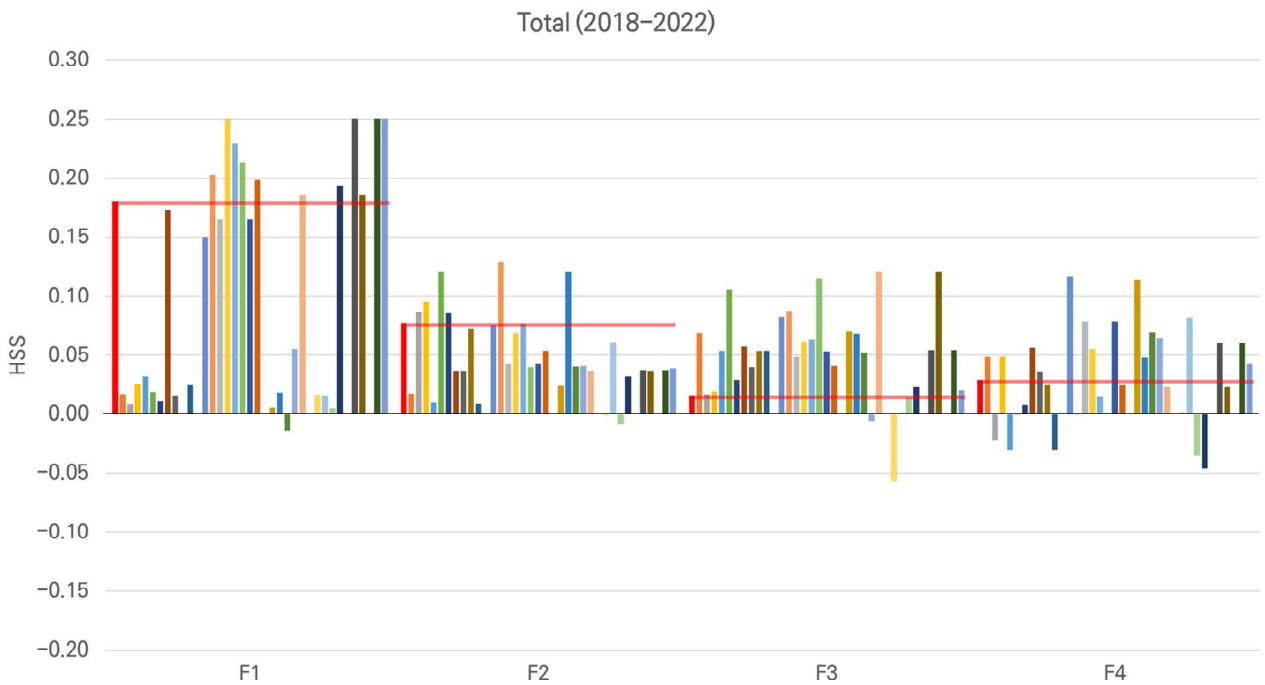


Figure 3.90. Results of deep learning prediction models in 2018–2022 (refer to figure 3.89 for this caption).

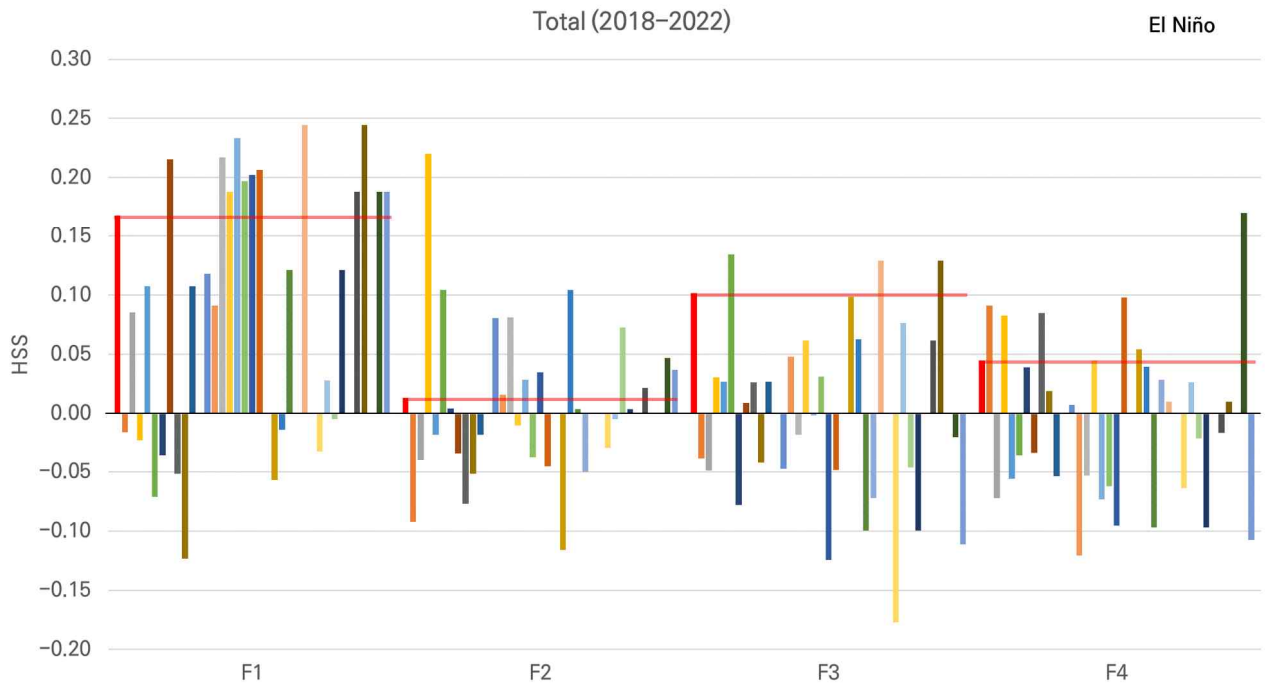


Figure 3.91. Results of deep learning prediction models during El Niño events in 2018-2022 (refer to figure 3.89 for this caption).

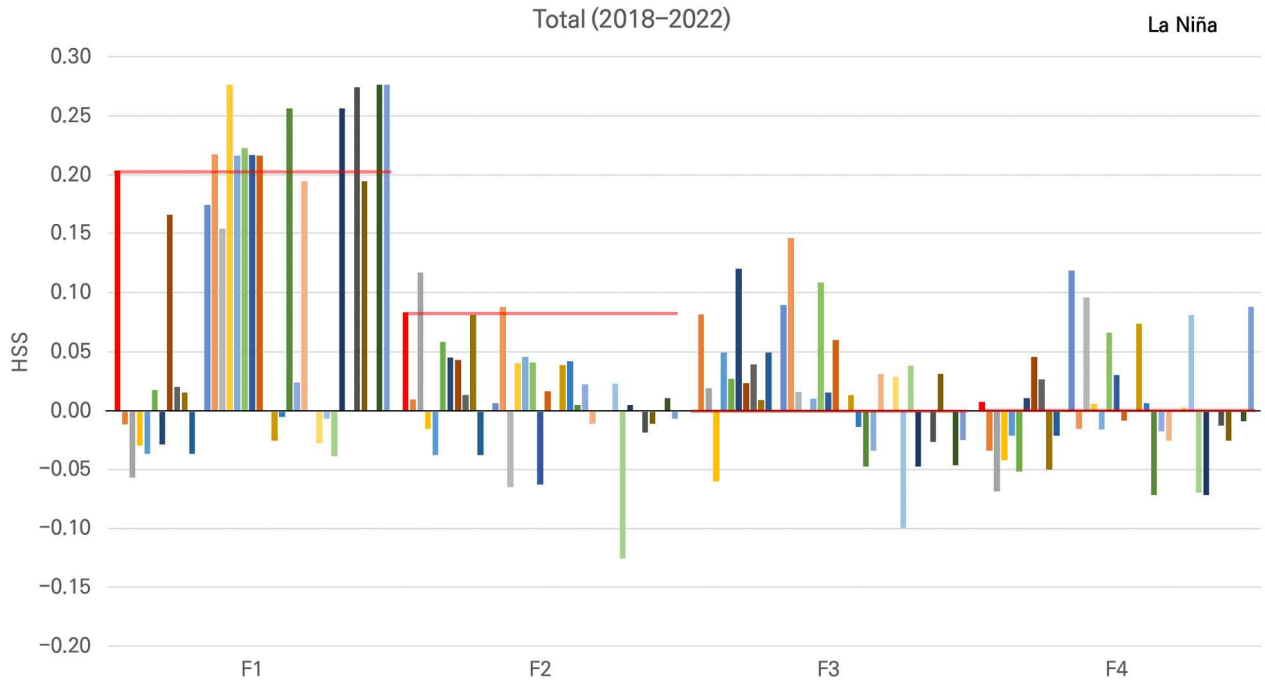


Figure 3.92. Results of deep learning prediction models during La Niña events in 2018-2022 (refer to figure 3.89 for this caption).

3.5. 자료강화기법을 이용한 인공지능모델의 월별 기후예측 개선

3.5.1. 자료강화기법 및 훈련자료 활용범위 비교

여름철 평균기온 예측에 대하여 CNN 모델의 Semi-monthly와 Cutmix 자료강화의 개선 성능을 비교하였으며 모든 월을 사용하는 Month-Agnostic 방안과 목표 월만을 사용하는 Target Month Only 방안의 결과 또한 비교하였다. Semi-monthly 방법을 적용한 경우 자료강화하지 않은 경우와 같이 epoch에 따라 모델의 검증 손실이 지속적으로 증가하여 자료강화 방법으로서 적합하지 않음을 시사하였으며, 훈련 손실을 살펴보면 충분히 학습되지 않은 상태에서 검증 손실 정체/증가로 학습이 중단되어 모델이 과소적합되었음을 확인할 수 있었다. Cutmix 방법을 적용한 경우 Month-Agnostic으로 모델 훈련을 하였을 때, 검증 손실이 epoch에 따라 그다지 증가하지 않음을 볼 수 있었으며 훈련이 충분히 이루어지기 전에 학습이 중단되는 상황도 관찰되지 않고 모델의 검증 정확도가 개선되었다(Figure 3.93). 그러나 Target month만을 사용한 경우는 과소적합이 발생하였으며 모델의 검증 정확도 또한 개선되지 않았다(Figure 3.95).

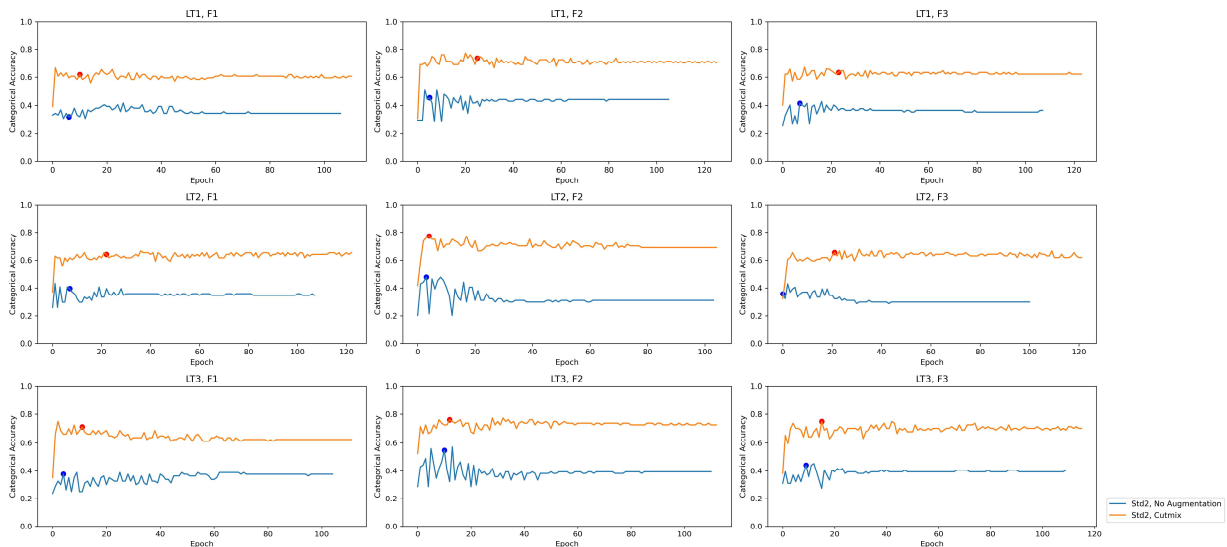


Figure 3.93. CNN validation accuracy (Month-Agnostic): Augmentation (Cutmix) vs. No Augmentation.

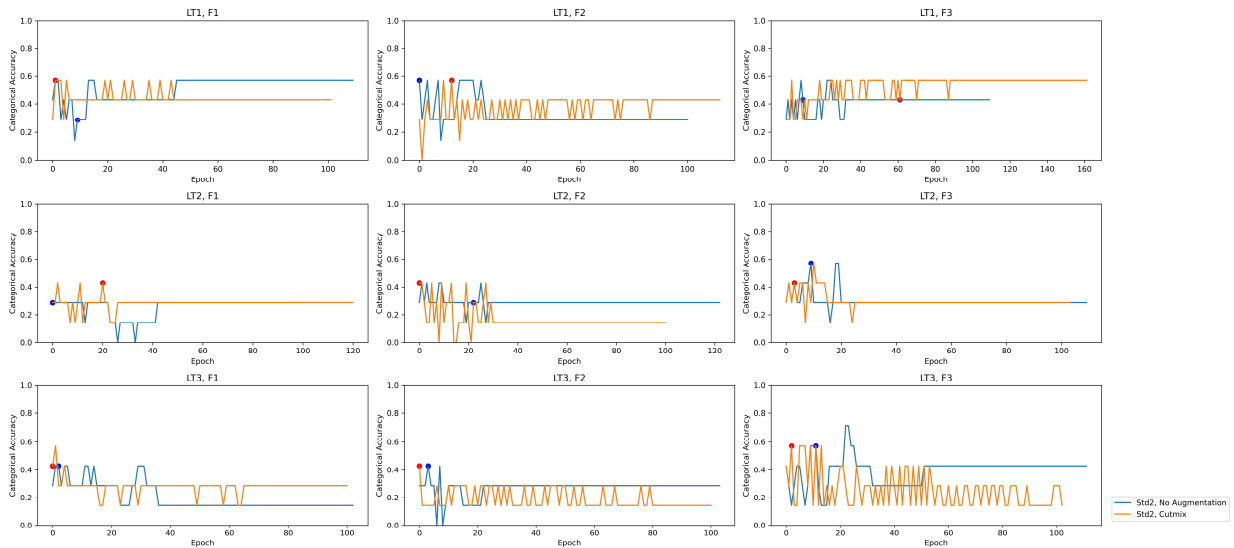


Figure 3.94. CNN validation accuracy (Target Month Only): Augmentation (Cutmix) vs. No Augmentation.

GNN의 경우에는 노드 분류 모델 및 그래프 분류 모델 모두 목표 월만 사용하고 자료강화하지 않은 경우의 훈련 손실이 가장 작았는데, 검증 손실을 보면 개선 여부가 일관적으로 나타나지 않았다. 그러나 그래프 분류 모델의 검증 정확도를 보면 6월(LT1)의 경우 목표 월만 사용하고 자료강화 하지 않은 경우가 정확도 가장 높으나 7월(LT2)의 경우에는 매우 낮고 오히려 모든 월을 사용이나 자료강화한 경우 다른 경우와 비교해 정확도가 높은 것을 알 수 있었다(Figure 3.95).

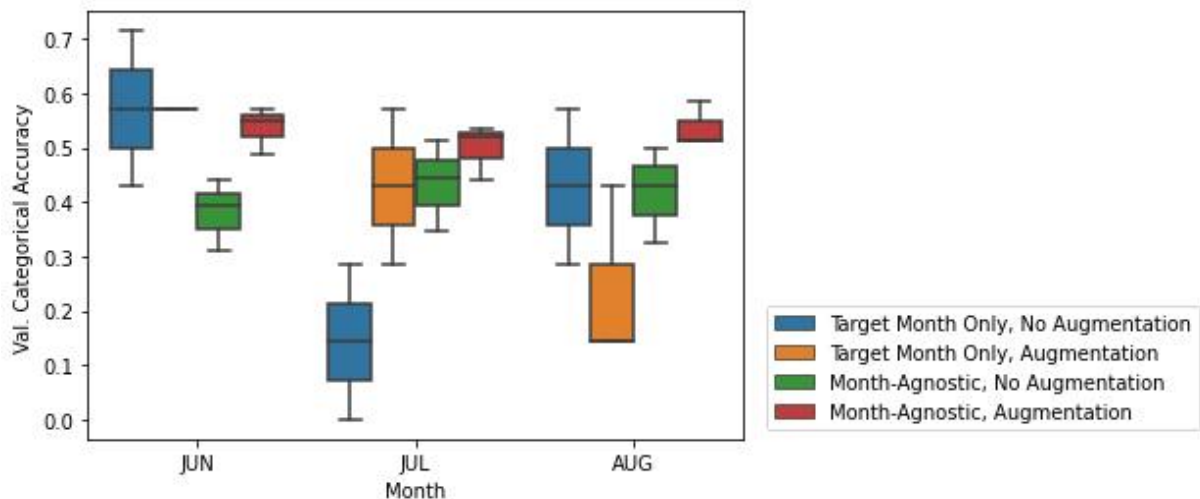


Figure 3.95. GNN Graph classification validation accuracy between the use of months and data augmentation methods.

테스트셋에 대한 지도학습 성능평가 결과를 살펴보면, CNN 모델의 경우 6월(LT1)과 7월(LT2)의 경우 모두 자료 강화의 개선 효과를 확인할 수 있었으나 8월(LT3)의 경우는 개선되지 않았다(Figure 3.96). GNN 노드 분류 모델의 경우, 7월(LT2)의 경우 목표 월만 사용하고 자료강화하지 않은 경우의 성능이 워낙 낮아 모든 월을 사용한 경우와 자료강화한 경우의 개선 효과가 뚜렷이 나타났으나, 그 밖의 경우에는 개선의 효과가 나타나지 않았다(Figure 3.97). GNN 그래프 분류 모델의 경우 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(Figure 3.98).

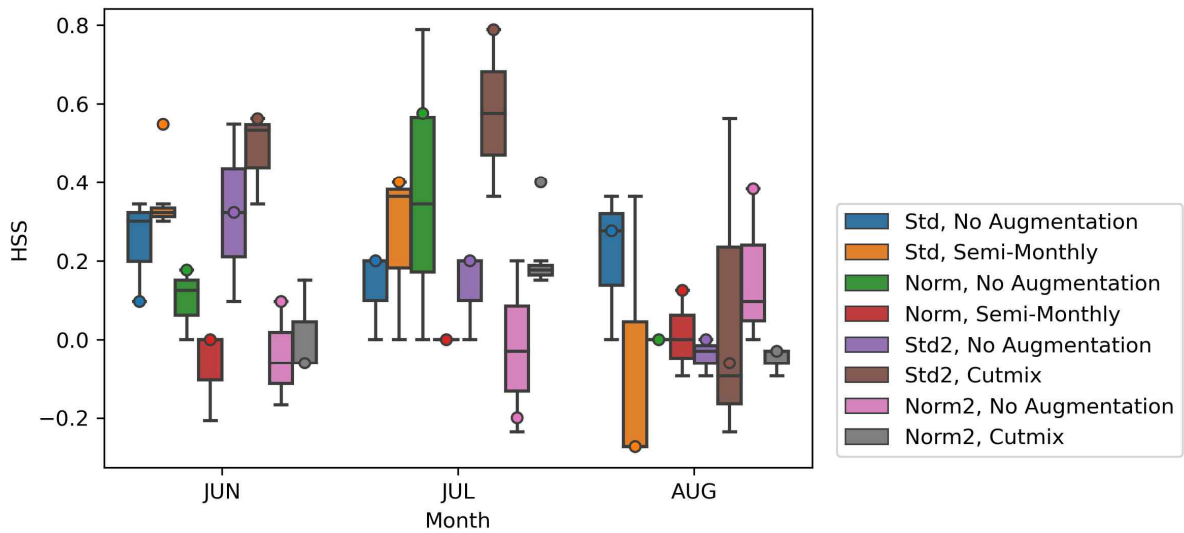


Figure 3.96. CNN Heidke Skill Score (HSS) between data pre-processing and augmentation methods.

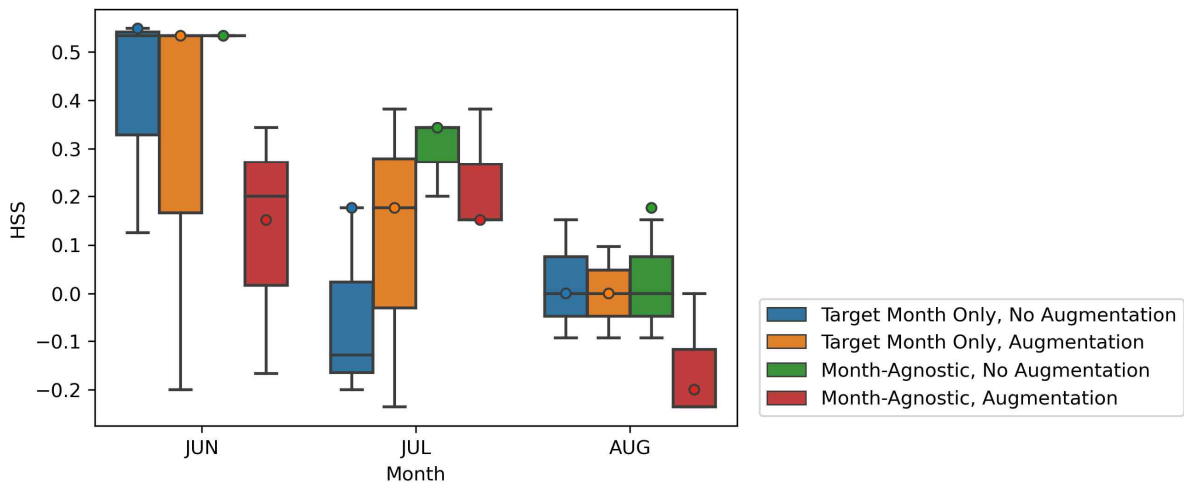


Figure 3.97. GNN Node Classification Heidke Skill Score (HSS) between the use of months and data augmentation methods.

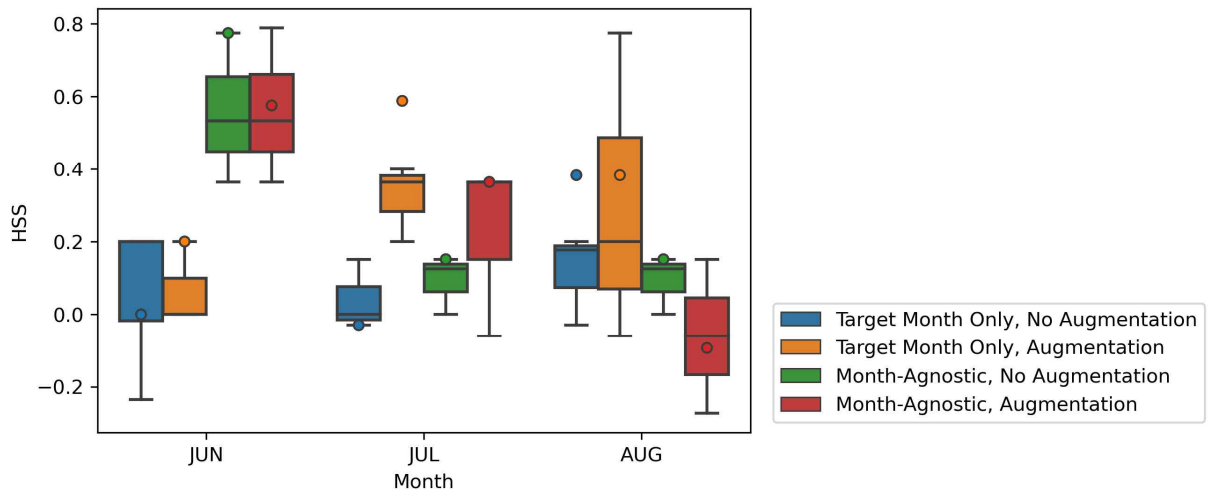


Figure 3.98. GNN Heidke Skill Score (HSS) between the use of months and data augmentation methods.

Cutmix 자료강화를 적용한 CNN 모델 결과에 대해 월별, 폴드별로 활성화맵을 그려 입력자료에서 중요하게 기여한 지리적 위치를 도출하여 자료강화 전후를 비교하였다. 이때 활성화맵은 완전연결층 직전의 공간정보를 이용하여 도출하였으며 Cutmix 자료강화 통해 개선된 사례 중 삼분위 카테고리 참값이 NN이 아니고 2018년 이후인 경우의 활성화맵을 살펴보았다.

2018년 7월(F3모델): BN로 낮게 예측 → AN로 개선

2018년 7월 우리나라 대기 상하층에 정체된 고기압으로 인해 하강기류에 의한 기온 상승과 일사량 증가로 고온이 지속되면서 폭염이 발생: 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 3.99).

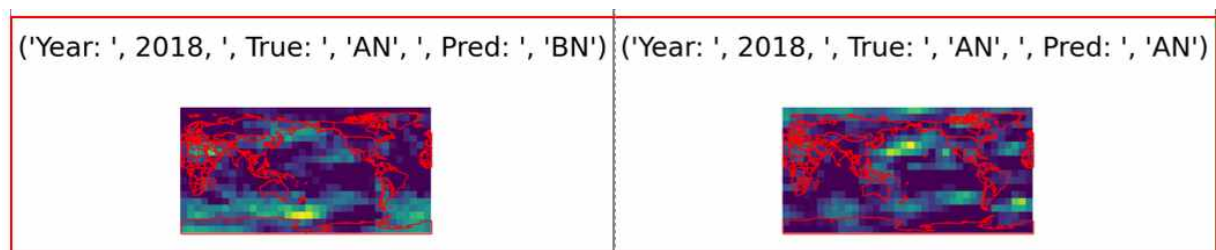


Figure 3.99. Class activation map before (left) and after (right) the data augmentation.

2021년 7월(F3모델): BN로 낮게 예측 → AN로 개선

2021년 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 3.100).

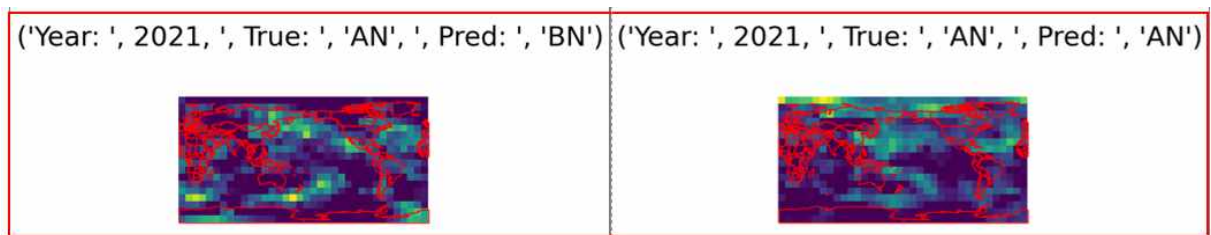


Figure 3.100. Class activation map before (left) and after (right) the data augmentation.

3.5.2. 여름철과 겨울철 평균기온 및 강수(SPII) 예측

앞서 여름철 평균기온 예측에 대하여 CNN 모델과 GNN 모델(노드 분류 및 그래프 분류)을 구축하고, CNN 모델에 대해 Semi-monthly와 Cutmix 자료강화의 개선 성능을 비교하고 모든 월을 사용하는 Month-Agnostic 방안과 목표 월만을 사용하는 Target Month Only 방안의 결과 또한 비교한 결과를 반영하여, Month-Agnostic 접근을 활용하는 CNN 모델에 대해 여름철과 겨울철, 평균기온과 강수(SPII) 예측에 대한 Cutmix 자료강화 기법의 개선 성능을 평가하였다.

자료강화한 경우 훈련 과정에서의 검증 정확도는 여름철, 겨울철 그리고 평균기온 및 강수(SPII)에 대해 모두 증가하였으며(Figure 3.101), Monte Carlo Dropout 적용하여 살펴본 결과 모델 결과의 epistemic 불확실성은 감소하였다(Figure 3.102). 그러나 테스트셋의 개별적인 예측 결과는 월별로 보았을 때 평균기온은 1월(LT2), 2월(LT3), 8월(LT3), 그리고 강수(SPII)는 7월(LT2), 8월(LT3), 12월(LT1)의 경우에만 개선이 관찰되었다(Figures 3.103-3.104).

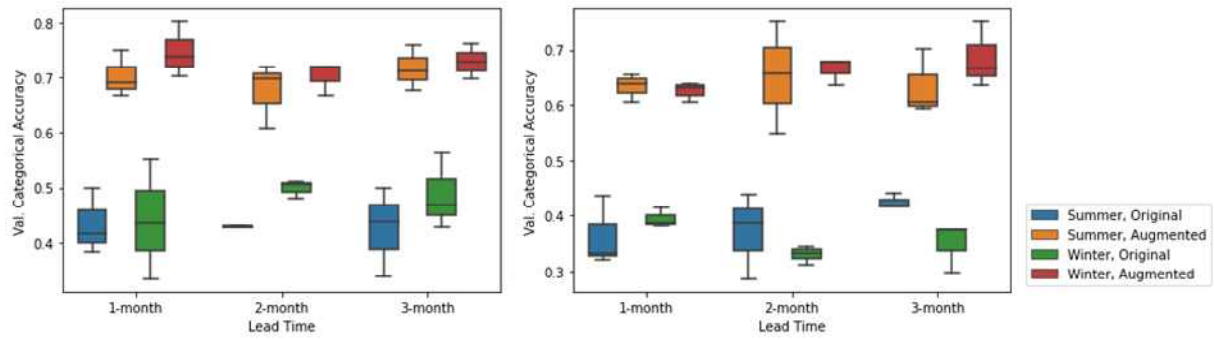


Figure 3.101. Validation accuracy of TmM (left) and SPI1 (right) during the training phase.

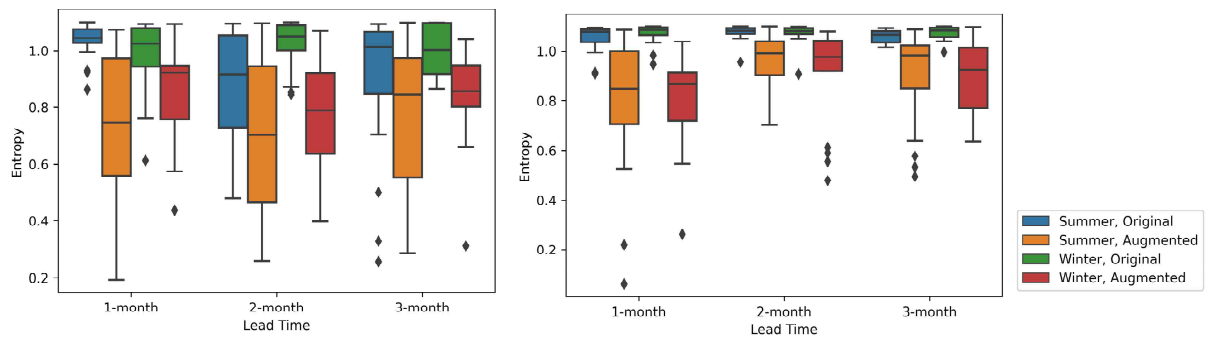


Figure 3.102. Entropy of test set prediction results of TmM (left) and SPI1 (right).

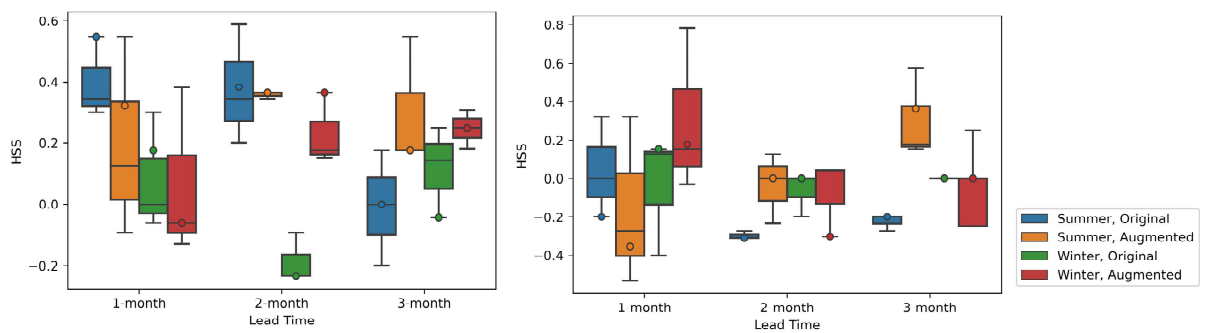


Figure 3.103. HSS of test set prediction results of TmM (left) and SPI1 (right).

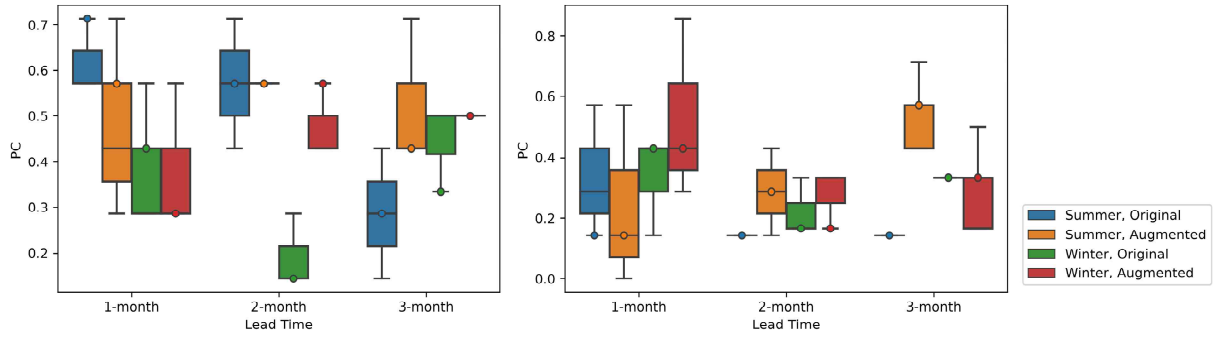


Figure 3.104. PC of test set prediction results of TmM (left) and SPI1 (right).

Confusion matrix를 통해 개별 예측 결과를 검토하였다(Figure 3.105). 좌상단과 우하단을 잇는 대각선을 따라 이루어진 예측이 명중한 것으로, 자료강화하지 않은 경우(왼쪽)와 자료강화한 경우(오른쪽)를 비교하면 대부분의 경우 적중률이 증가하였음을 볼 수 있다.

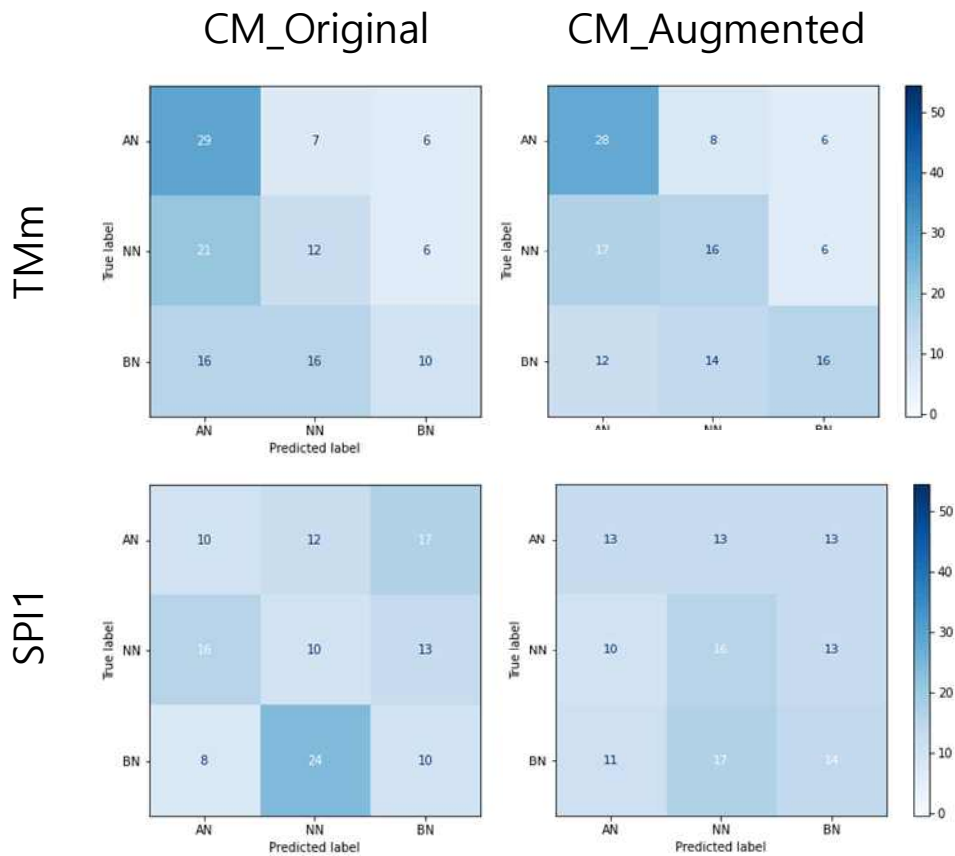


Figure 3.105. Confusion matrices of test set prediction results of TmM (upper) and SPI1 (lower).

3.5.3. 역학모델 결과 활용 효과 검토

모델 훈련에 있어 Community Earth System Model (CESM) 자료를 포함하여 전이학습을 하든지 ERA5나 CSF-20C 등의 긴 재분석 자료를 활용한다면 좀 더 많은 훈련자료를 확보할 수 있다. 그 경우 우리나라 평균기온이나 강수(SPI1) 값은 해당 위치의 격자 값에서 도출해야 한다. 이 연구는 단순한 인공지능모델에서 간결한 자료강화기법이 가지는 효과를 분석하는 것을 목적으로 하므로 그러한 방안은 고려하지 않았으며 평균기온 및 강수(SPI1) 예측의 reference로 62개 ASOS에서 도출한 값을 사용하고 있다. 사용한 ASOS의 개수를 62개 정도로 확보하려면 1973년 이후의 자료를 활용해야 하는데, 이보다 더한 제약은 APCC MME의 hindcast 기간이 1991년 이후라는 점이 된다. 실제 예측 적용 시 관측 없는 시기에 대해 역학모델의 결과를 활용하기 위함인데, 그로 인해 가용한 훈련자료의 기간이 짧아지는 trade-off가 생기는 셈이다.

역학모델 결과의 활용 효과를 검토하기 위해 같은 기간에 대해 MME를 제외한 관측 및 재분석 자료만을 이용하는 경우(OBS_short)와, MME를 제외함으로 인해 확보된 자료를 좀 더 포함하여 1981년 이후의 자료를 활용한 경우(OBS_long)의 예측 결과를 동일한 테스트셋에 대하여 비교하였다(Figures 3.106, 3.107). 전체적으로 Confusion matrix를 살펴보았을 때 평균기온의 경우 MME 및 OBS_long 우수, SPI1은 OBS_short가 우수한 편으로 해당 실험 세팅에서 역학모델 활용의 큰 이점을 찾아보기는 힘들었다.

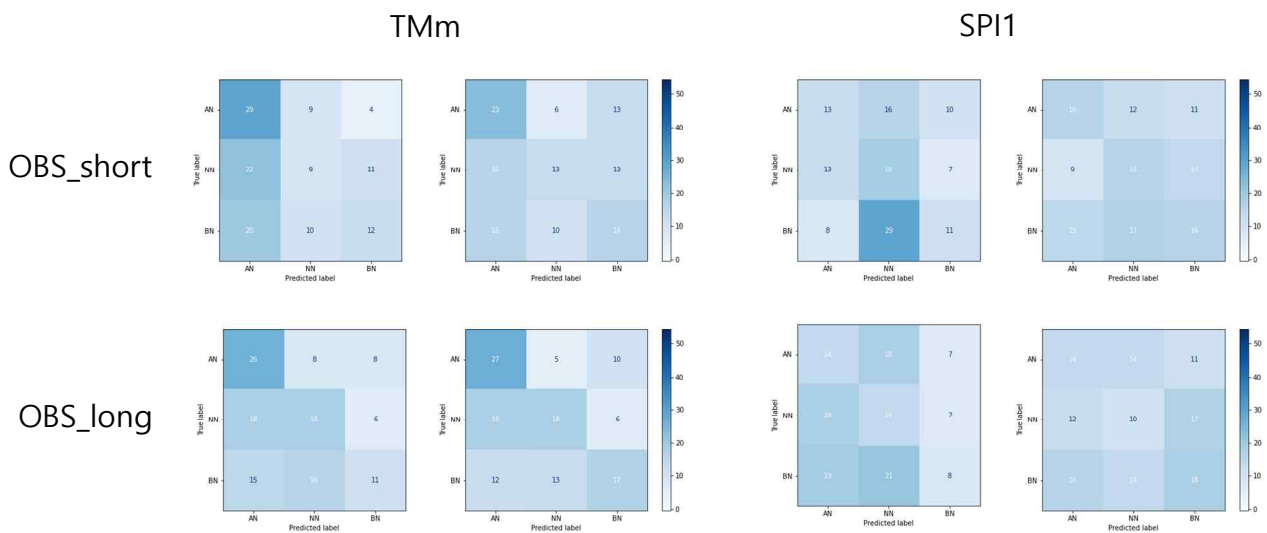


Figure 3.106. Confusion matrices of test set prediction results of OBS_short (left) and OBS_long (right).

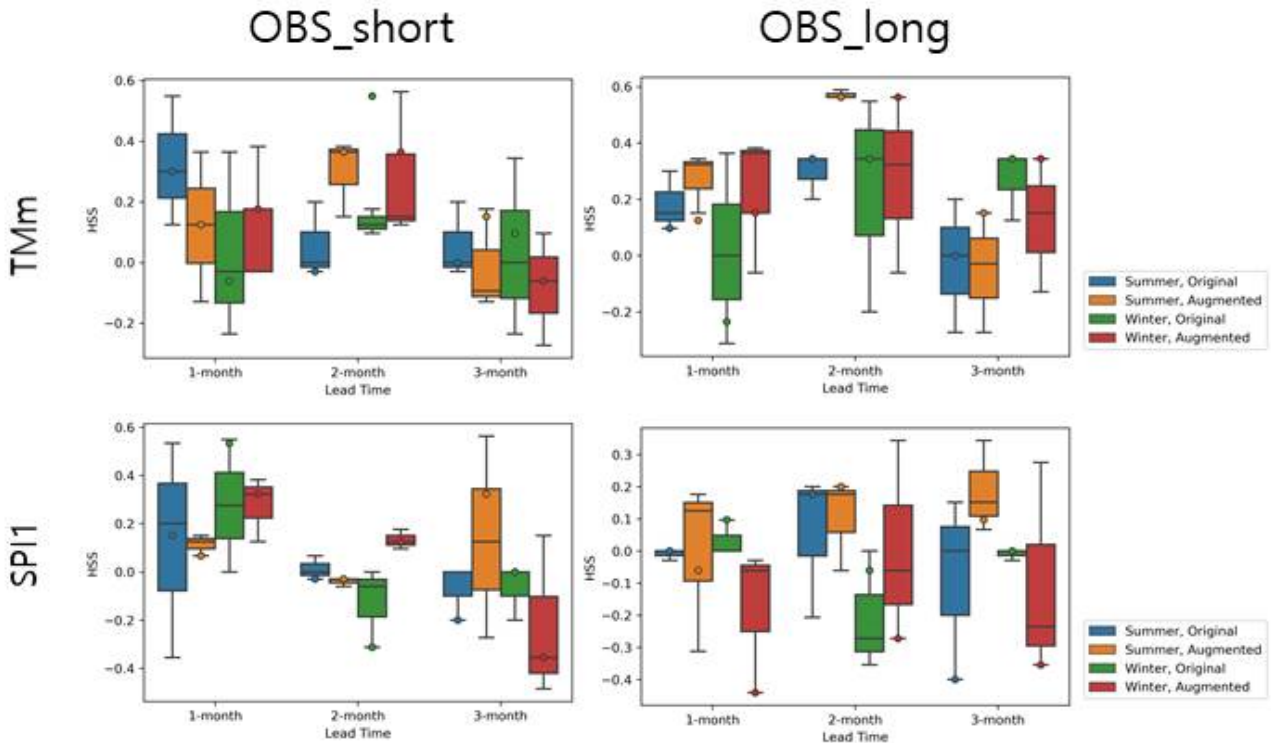


Figure 3.107. HSS of test set prediction results of OBS_short (left) and OBS_long (right).

3.6. 인공지능 기술을 활용한 다중모델 앙상블 강수 역학예측 후처리

3.6.1. 설명가능인공지능(XAI) 기술적용을 위한 환경구축 및 시범테스트

이 연구의 목표는 기후 자료 처리 및 기후 예측에 적합하고 예측성이 향상된 인공지능 기반 인공지능 기후 예측 후처리 시스템을 개발하는 것이다. 이를 위해 제 1 차년도 과제 수행에서는 설명가능한 인공지능 기술(XAI; eXplainable Artificial Intelligence)을 도입하여, 인공지능 시스템의 퍼포먼스 뿐만 아니라, 인공지능 내부 파라미터의 분배 및 자료의 처리 과정을 살펴보고자 하였다. 이를 통하여 인공지능 모형의 구조적 적합성 여부와 인공지능의 결과 도출 과정에서 기후 자료가 물리/역학적으로 의미있는 방향으로 처리되고 있는지 확인하였다. 인공지능 내부 및 자료 처리 과정 진단을 통해 인공지능 모델이 실제로 어떻게 학습을 수행하는지, 실제 학습 과정에서 인공지능 모델이 어떠한 정보를 활용하는지에 대해서 면밀하게 확인하였다. 나아가 기존의 물리/역학적인 매커니즘에 기반하는 학습을 수행하지에 대한 예시를 제시함으로써 “How AI learn” 과 “What AI see” 에 대한 대답 및 가능성을 확인하고자 하였다.

3.6.1.1. 인공지능 모형 훈련과정 평가: 커널 시각화 및 피쳐맵 분석

설명가능한 인공지능(XAI)를 통한 계절 예측에 적합한 후처리 시스템 개발을 위한 첫 단계로서 기존 인공지능 모델을 구현하고 설명 가능한 평가 환경을 구축하고자 하였다. 인공지능 기반 계절 예측 후처리 모형의 성능을 평가하기 위하여 사전 학습과 미세 조정 학습의 두 가지 단계를 통해 모형을 훈련하고 학습 과정을 평가하였다. 구체적으로, 사전 학습을 위해서는 비교적 장기 기후 정보가 제공되는 NCEP1 재분석 자료를 이용하였고, 미세 조정 학습을 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 모형으로 미세 조정 학습을 수행하였다.

합성곱 커널의 가시화를 통해서 인공지능망 자체가 어떻게 구성되어있는지 확인하고 이를 기반으로 모델을 개선하고자, 3-차원 합성곱을 통해 다변수 입력을 처리하는 층의 커널 모습 (Figure 3.109)를 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 상관도(Figure 3.110)을 살펴보았을 때, 변수간 분포가 비슷한 구조를 띄는 경우(예를 들어, 2-5 혹은 6-8 커널)가 종종 관찰된다. 커널 크기 조정을 통해 기후 변수간의 물리/역학적 관계를 검토하거나, 커널간 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)을 사용하였으나, 이는 중위도에서 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.

또한, 각 층의 활성화 정도와 각 층을 통과하는 정보량을 이용하여 인공지능망의 복잡도를 일부 평가해볼 수 있겠다. 앞서 살펴본 활성화맵과 비슷한 모습을 보이고는 있는 위 그림 (Figure 3.111)은, 실제 인공지능망 안에서 첫번째 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.

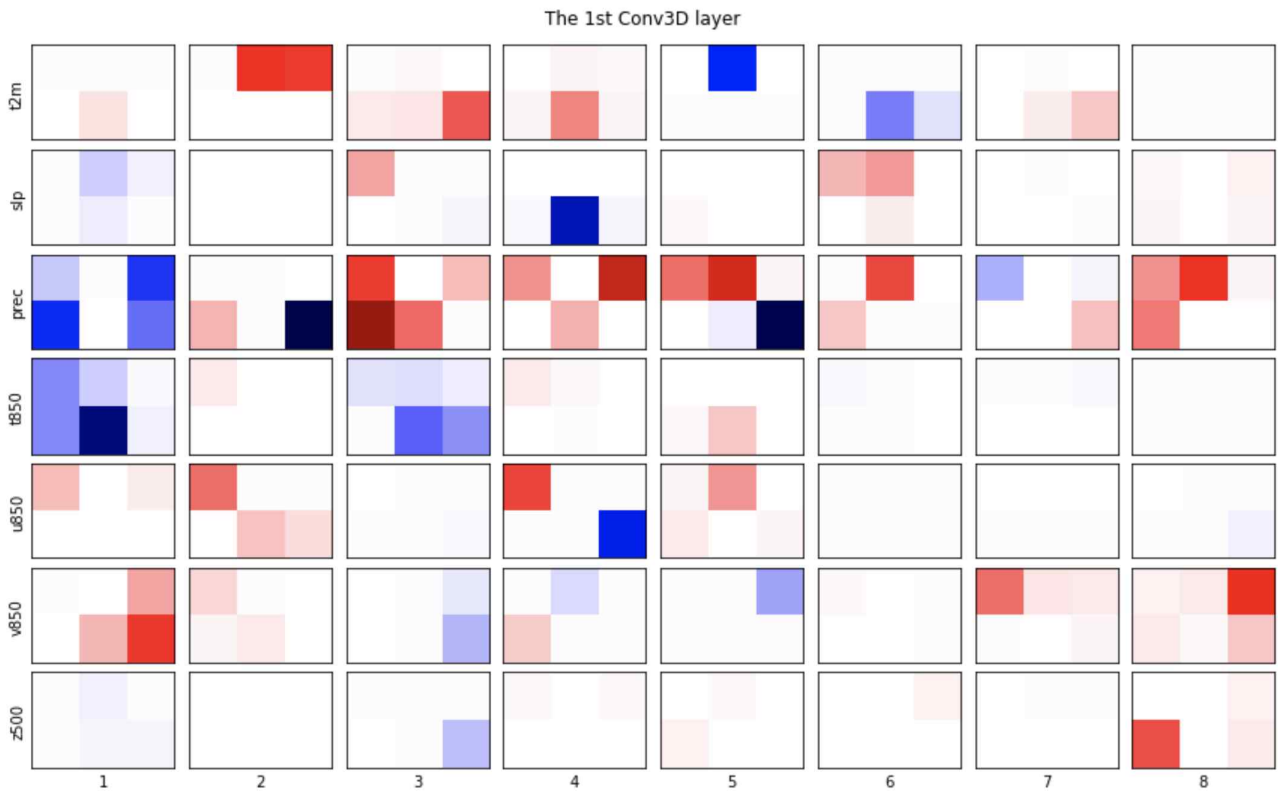


Figure 3.108. Weights of the 3D convolution kernels of the input layer with the dimension of 2(latitude) x 3(longitude) x 7(variable) x 8(number of kernels).

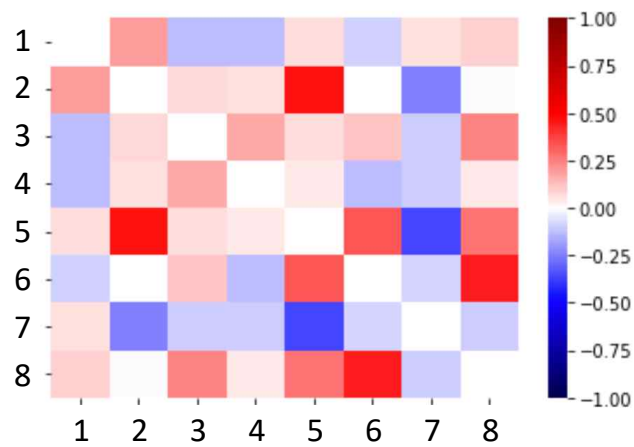


Figure 3.109. Correlation coefficients between kernel structures.

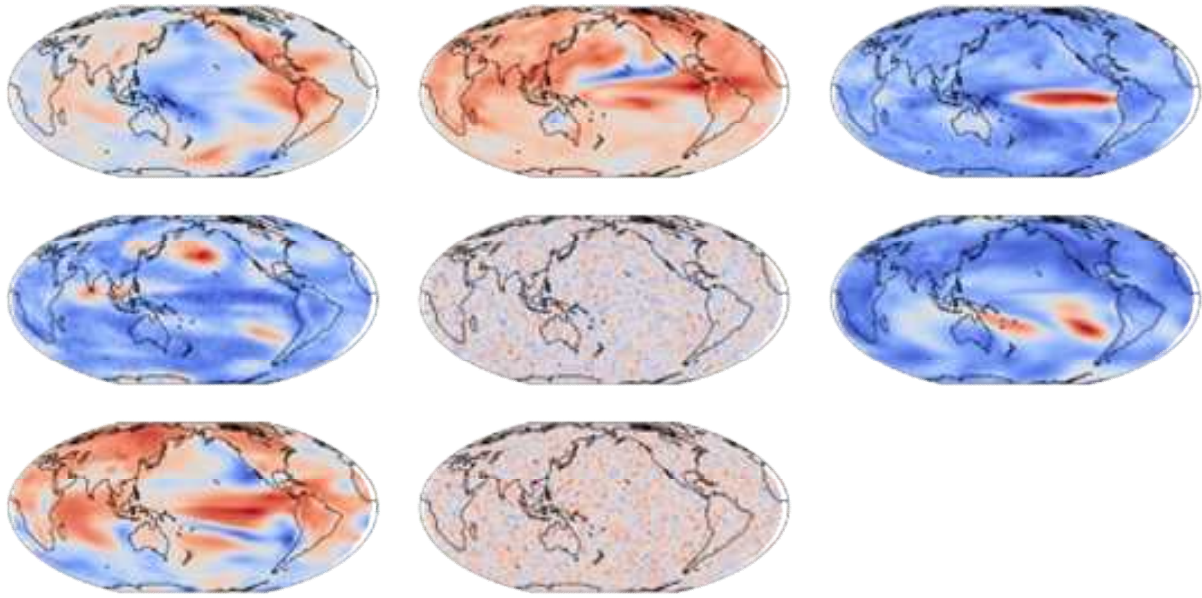


Figure 3.110. Visualization of activated values of the first 3D convolution layer.

정보를 입력받는 초기층 뿐만 아니라, 마지막 층에 대한 분석도 인공지능망 모델 개선에 흥미로운 정보를 줄 수 있다. 이 그림(Figure 3.112)은 관측재분석 자료를 이용해 사전 전이 학습을 수행한 인공지능망 모델의 맨 마지막 완전 연결층의 연결 강도와 맨 마지막 완전 연결층을 제외한 인공지능망 모델은 고정하고 실제 자료(예를 들어 MME-관측소 자료의 조합)를 이용하여 추가 미세 조정 학습을 수행한 후 변화한 완전 연결층의 연결 강도 변화를 보여주고 있다. 이 과제에서는 적은 수의 자료를 이용하여 복잡한 기후 현상을 예측하기 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 맞는 현상 사이에 큰 차이가 존재하며, 최종 미세 학습 과정에서 인공지능망 모형이 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 인공지능망 모델을 구성하는 데 활용할 수 있을 것으로 보인다.

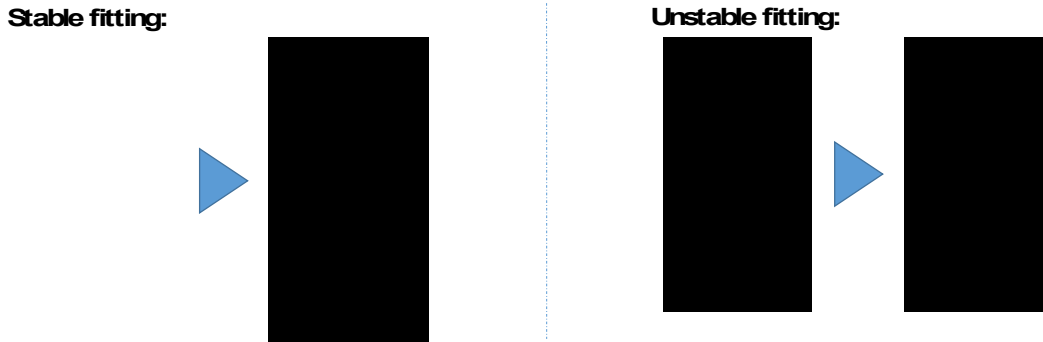


Figure 3.111. Visualization of the connection strength of the last hidden layer and the output layer.

3.6.1.2. 인공지능망의 역학/물리 과정 재현 예시: 대서양 온난 현상 예측 원천 평가

설명가능 인공지능 기술 습득 및 환경 구축 과정에서 단순 2차원 합성곱신경 모델을 활용해 대서양 적도 모드(Atlantic Equatorial Mode, AEM) 지수에 대한 삼분위 카테고리 확률 예측을 시범적으로 수행해 보았다. 대서양이 가열되고 차가워지는 장기 변동을 인공지능 모델로 구현했을 때, 어떠한 정보를 활용해서 합성곱신경망이 예측을 수행하는지, 알려진 역학/물리 과정을 재현하고 있는지에 대해서 GradCAM (Selvaraju et al., 2019) 설명가능 인공지능 기술을 적용해 분석하였다. 이 실험의 목표는 “what AI see” 정보를 도출하고 기후 분야에 있어 활성맵의 효율적 가시화 방안을 찾는 것이다. 또한 역학 기반 예측 원천 학습 여부를 확인함으로써 인공지능 모델의 가능성을 탐색해 보고자 함이다.

대표 재분석장인 ERA5의 자료기간이 짧기 때문에, 1870년부터 2015년까지의 20CR & Hadley 재분석 자료를 활용하였다. 11개의 입력 변수(T2M, SLP, PRECIP, T850, U850, V850, Z500, SICE, VP200, VP850, ST) 조합에 대해서 민감도 실험을 수행하였다. 여기서 ST는 해양에서의 해수면온도 (SST)와 육지에서의 토양온도(soil temperature) 자료를 각각의 missing 위치에 채워 넣어 missing 없이 전구 surface temperature 형태로 재구성한 자료이다. 아래 Figure 3.113은 몇 가지 입력 변수 조합에 따른 HSS 카테고리 예측 성능을 제시한다. 선행 3개월 입력 자료를 활용한 12월 AEM 지수 예측과 (Figure 3.113a), 선행 4개월 입력 자료를 활용한 1월 AEM 지수 예측 (Figure 3.113b) 모두에서 ST 변수, 즉 해양에서의 SST와 지면에서의 토양온도에 대한 훈련만으로도 가장 높은 성능을 보여준다. 즉, 대서양 지수 예측에 있어서 대기 변수를 추가하는 것이 효율성이 높지 않고 오히려 성능을 떨어뜨림을 알 수 있다. 입력 변수에 따른 성능 변화 추이는 두 실험 모두에서 비슷한데, 하층 동서 바람장(U850) only이나 11개 모든 변수 조합에서 성능이 저조하다. 하지만, ST+U850+VP200 즉 해수면온도와 토양온도, 하층 동서 바람장, 상층 속도포텐셜 정보 조합에 대한 성능이 매우 높아지는 특성을 확인할 수 있다. 즉, 해양/육지/대기순환 정보를 적절히 조합했을 때 유의미한 정보기반 효율적인 훈련이 이루어졌음을 의미한다. 현재는 ST 변수만으로 최고 성능을 보이지만 추가적인 입력 변수 조정과 아키텍처 조정 실험을 통해 충분한 성능 개선이 가능할 것으로 예상할 수 있다.

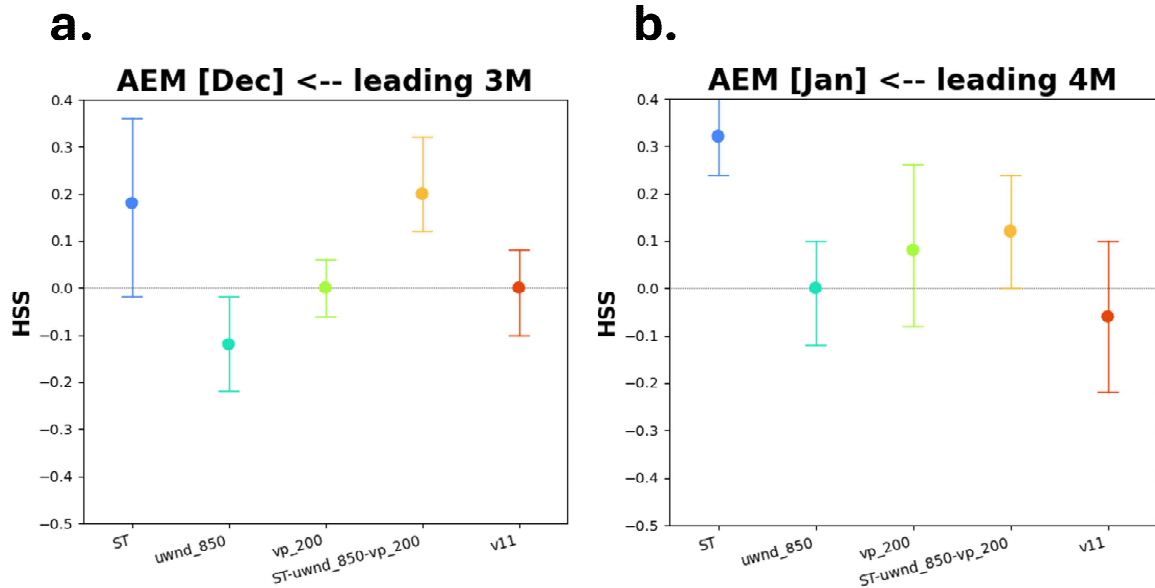


Figure 3.112. HSS of AEM index tercile category forecast from CNN model for (a) December and (b) January target months. Circles and whiskers show the mean and range of 75 samples ($k=5$ folding ensembles for 15 years) across four combinations of input variables.

Zhang and Han (2021)은 인도양 쌍극자 패턴 (Indian Ocean Dipole, IOD)이 대서양 온난화(Atlantic Nino)를 유도한다는 대양간 대기-해양 상호작용에 대한 연구 결과를 제시한 바 있다. 구체적으로 양의 IOD 시기에 강화된 적도 서인도양 강수 활동이 적도 대서양의 무역풍을 약화시키고, 그에 따라 대서양 중앙/동쪽의 온난화를 유도하는 일련의 과정을 통해 대략 3-5개월에 걸쳐 Atlantic Nino가 발생하게 됨을 제시하였다 (Figure 3.114a, Zhang and Han (2021)에서 발췌). 2001-2015년 15년 테스트 기간 동안 모든 $k=5$ folding 앙상블에 대해 AN (Above Normal) category에 대한 GradCAM 활성화맵을 구하고, 전체 합성장을 표출해 보았다. 12월 AEM 지수 예측의 경우 인도양 지역에서 강한 활성화 시그널을 보이는데(Figure 3.114b), 입력 변수가 ST인 것을 고려하면 선행하는 인도양 쌍극자 모드의 영향으로 해석할 수 있다. 온난 1월 AEM 지수 예측의 경우 동대서양~서아프리카 연안에서 강한 활성화 시그널을 보이는데(Figure 3.114c), 이 때 입력 변수가 U850이므로 이 역시 대서양 온난화를 유도하는 적도 대서양 무역풍 약화라는 역학적인 매커니즘으로 해석할 수 있다. 12월/1월 각 합성곱 모델의 입력장에서 어느 선행 시점에서 가장 활성화 시그널을 보이는데 구체적으로 알지 못해 Zhang and Han (2021)이 제시한 선행 시기와 완벽하게 일치하는 것으로 볼 수는 없다. 그럼에도 불구하고, 선행 3개월/4개월 내에서 AI가 가장 활발하게 예측에 참고하는 지역이 어디인지, 그 변수가 어디인지에 대한 적절한 이해에 도달할 수 있었다. 또한 그 결과를 통해 합성곱 신경망의 학습 과정이 대서양 온난화 과정에 대한 기존 역학/물리 과정을 유의미하게 재현하는 것으로 판단할 수 있다. 이 절은 간단한 설명가능 인공지능 기술 (GradCAM 등)을 활용해 “What AI see”에 대한 질문에 대한 대답을 얻는 과정에 대한 예시를 제시하고 있다. 최종 계절예측 후처리 모델 개발 과정에서 이러한 XAI 기술을 적극적으로 활용해 AI가 재현하는 역학/물리 과정에 대해 좀 더 깊이 있는 분석을 시도하였다.

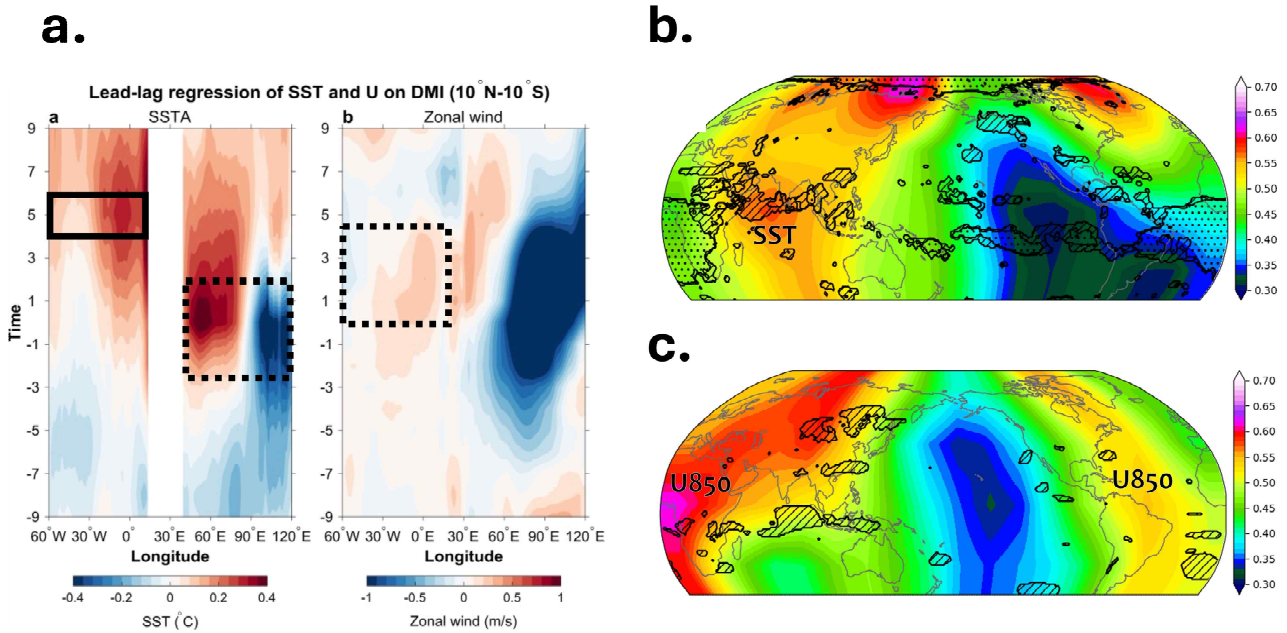


Figure 3.113. (a) Lead-lag regression map of SST and U with respect to DMI adopted from Zhang and Han (2021), (b) GradCAM activation maps of AEM tercile forecasts for December (upper) using preceding three months of ST data and January (lower) using preceding four months of U850 data. Both maps show AN category and are composites of 15 years from 2001 to 2015 with their 5 ensembles ($k=5$ folding).

3.6.2. AI 후처리 접근법 1: 앙상블 선별 (SubSampling, SS) 및 예측 성능 원천 분석

이 절부터는 실제 동아시아 계절 강수 역학 예측 후처리를 위한 인공지능 모델 개발 과정에 대해 기술하겠다. 첫 번째로 다중모델앙상블에 사용되는 다량의 앙상블들 중에 양질의 것만을 선별하는 접근법에 대한 후처리 결과이다.

3.6.2.1. ACC 스킬 추정 기반 앙상블 선별 모델 평가

앙상블 선별을 위한 성능 추정용 기준 ResNet 아키텍처를 설계하고 Figure 3.115에서 제시한 다양한 민감도 실험을 통해 가장 우수한 성능을 보인 ResNet18모델에 대한 결과를 주로 제시한다. 이 모델에서는 8개 대기 변수 이외에 land-sea mask, latitude 정적 변수를 입력으로 넣고, 이에 대한 T-3, T-2, T-1, T1, T2, T3 6개 선행월 자료를 입력으로 적층하였다. mse 비용함수, early stopping-용 epoch tolerance = 20, batch size=16, learning rate=0.001를 세팅하였다. 현업 계절 예측 모델로부터 가용한 모든 앙상블 중에서 ACC 추정모델의 출력값을 기준으로 상위 10, 30, 50, 70, 90%에 해당하는 것들만을 모아서 동아시아 계절 강수 단순 합성맵을 얻고 ACC 평가 메트릭을 적용해 관측 기준(ERA5)의 패턴과 비교했을 때 결과는 아래 Figure 3.115와 같다. 2023년 12 계절 중 늦겨울부터 초여름까지의 계절에서 5개의 앙상블 선별 후처리 결과의 대부분이 기존 다중모델앙상블 대비해서 더 나은 성능을 보였다. 다수의 불량 앙상블들이 제외되고 양질의 앙상블들이 선별되어 얻어진 결과로 판단된다. 하지만 예상과 다르게 몇몇 계절

(MAM, MJJ, JJA, DJF)을 제외하고 선별 기준 %tile에 대해서 비례적으로 결과가 좋아지는 것은 아니어서, ACC 추정 모델이 이상적으로 앙상블들을 나레비 세우지는 못하는 것으로 이해할 수 있다. 늦여름부터 가을까지 ASO, SON, OND 계절의 경우는 5개 선별 결과 모두 기존 다중모델 앙상블보다 성능이 떨어지고, 오히려 선별 기준 %tile에 대해 반비례하게 성능이 감소하는 것으로 나타났다.

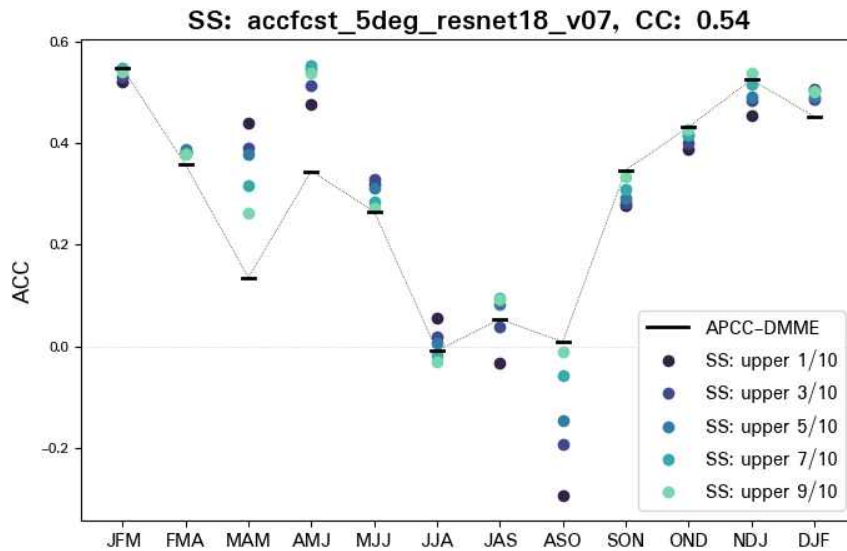


Figure 3.114. Performance of the first SS approach: comparison of EA rainfall ACCs between original APCC-MME (black horizontal lines) and five sub-sampling models (circles) averaging ensembles with predicted skill in the highest 10%, 30%, 50%, 70%, 90%. Results are for 12 seasons of the test year 2023.

앙상블 선별에 대한 성능 개선이 뚜렷한 2023년 MAM에 대한 후처리 모델 평가 결과는 Figure 3.116과 같다. 산포도에서 앙상블 분포가 $y=1$ 대각선에 근접하게 나타나, ResNet 모델이 실제 ACC를 거의 이상적으로 추정하는 것을 알 수 있다. 모델 출력값을 기준으로 상위 90, 70, 50, 30, 10% 앙상블들을 밝은 푸른색에서 짙은 푸른색으로 표시했을 때, 이상적으로 불량 앙상블들이 단계적으로 탈락하는 것을 알 수 있다. 가운데 타일 그래프는 각 기준에 대해 선별된 앙상블들을 표시한 것으로 y축은 실제 ACC 값에 대한 내림차순 앙상블 정렬을 나타내고, 각 앙상블에 표시된 컬러는 선별된 앙상블의 추정 ACC 값이다. 흰색으로 표시된 탈락 앙상블들이 아래쪽에 집중적으로 분포하고, 컬러 표시가 전반적으로 위에서 아래로 갈수록 짙어지는 양상을 보여 상대적 ACC 추정이 적절히 이루어졌음을 알 수 있다. 아래의 막대그래프와 표는 $ACC=0$, 하위 25% ACC값 이 두 가지 기준에 대한 전체 샘플에서 불량 앙상블의 비율, 탈락 샘플 중 불량 앙상블 비율을 각각 의미한다. 선별 %tile이 높아질수록 불량 앙상블 비율이 늘어나고 탈락율도 높아지는 것은 앙상블 선별이 이상적으로 이루어지는 것을 의미한다. 선별된 앙상블들만을 모아 얻어진 최종 단순 합성장은 우측 패널에서 확인할 수 있다. 중앙 상단의 기존 다중모델앙상블의 $ACC=0.14$ 이 비해 5개 선별 결과 모두 개선된 성능을 보였고, 특히 상위 10%

선별 결과는 ACC=0.44로 압도적으로 좋은 성능을 나타냈다. 이는 관측 재분석장(좌상단)에 두드러진 남중국, 방글라데시의 건조 패턴 강도를 상위 10% 선별에서 유의미하게 예측하는 점에 기인한다고 볼 수 있겠다. 또한, 동남아시아 영역에서 습윤 패턴이 두드러진 역학 다중모델 앙상블의 결과를, 후처리를 통해 건조 패턴으로 수정한 점 역시 두드러진 성능 개선 포인트이다. 2023년 MAM 계절에서 앙상블 선별 후처리가 효과적일 수 있는 기저 원인은 앙상블 분산값 분석을 통해 확인해 볼 수 있다 (좌하단). 관측 재분석장에서 건조 시그널이 두드러진 지역들(방글라데시, 남중국)에서 앙상블들간의 변동성이 크고, 따라서 앙상블들을 적절히 선별했을 때 이 지역들에 대한 개선 가능성이 있기 때문이다.

반대로 앙상블 선별에 실패한 2023년 ASO 계절에 대한 후처리 모델 평가 결과는 Figure 3.117과 같다. 산포도가 제시하듯, ACC 추정 모델이 실제 ACC에 대비해 앙상블을 반대로 나열함으로써 최종 후처리 과정에서 불량 앙상블 대신 우량 앙상블을 탈락시키는 결과를 얻게 되는 것이다. 가운데 타일 그래프에서 보듯이 실제 ACC 높은 위쪽에서 앙상블 탈락이 이루어지고, 막대 그래프와 표가 제시하듯 기준 %tile이 상위 10%인 경우가 상위 90%보다 불량 앙상블 비율이 더 높다. 그리고 전체적으로 불량 앙상블 탈락율 수치가 2023 MAM에 비해 현저히 낮다. 따라서, 선별된 앙상블 단순 합성장(우측 패널)이 기존 다중모델 앙상블의 성능을 넘어서지 못하게 된다. (좌하단) 앙상블 간의 분산 패턴에서 아열대 태평양 해양에서의 값이 두드러지는 등 실제 강수 관측 패턴과 괴리가 있는데, 따라서 앙상블 선별로 얻은 합성장의 차이가 관심 강수 지역이 아닌 해양에서 나타날 가능성이 높고 이는 앙상블 선별의 효과를 기대하기 어렵게 한다. (좌상단) 관측 재분석장에서 보듯 동아시아 여름 강수는 복잡한 공간 분포를 보이고 국지적 물리 매커니즘에 크게 영향을 받는다. 3개월 평균, 2.5도의 시·공간 규모에도 불구하고 이러한 공간 복잡성을 갖는데, 역학 모델은 이러한 국지적 강수 분포를 적절하게 재현하지 못한다. 딥러닝 모델을 적용하더라도 복잡한 패턴 즉 복잡한 매커니즘을 갖는 여름 대기 상태로부터 숨겨진 관련성을 찾아내는 것이 쉽지 않으며, 따라서 앙상블 선별 후처리를 하더라도 다음 계절의 강수 성능을 추정하는 데에는 한계가 있다고 판단할 수 있겠다. ACC 추정 모델에서 과거 3개월 앙상블 예측이 입력장으로 들어가게 되는데, 여름철 성능 자체가 극도로 낮다는 점도 늦여름부터 초가을의 ACC 추정 모델 실패, 나아가 앙상블 선별 후처리 성능 하락의 원인이 될 수 있을 것이다.

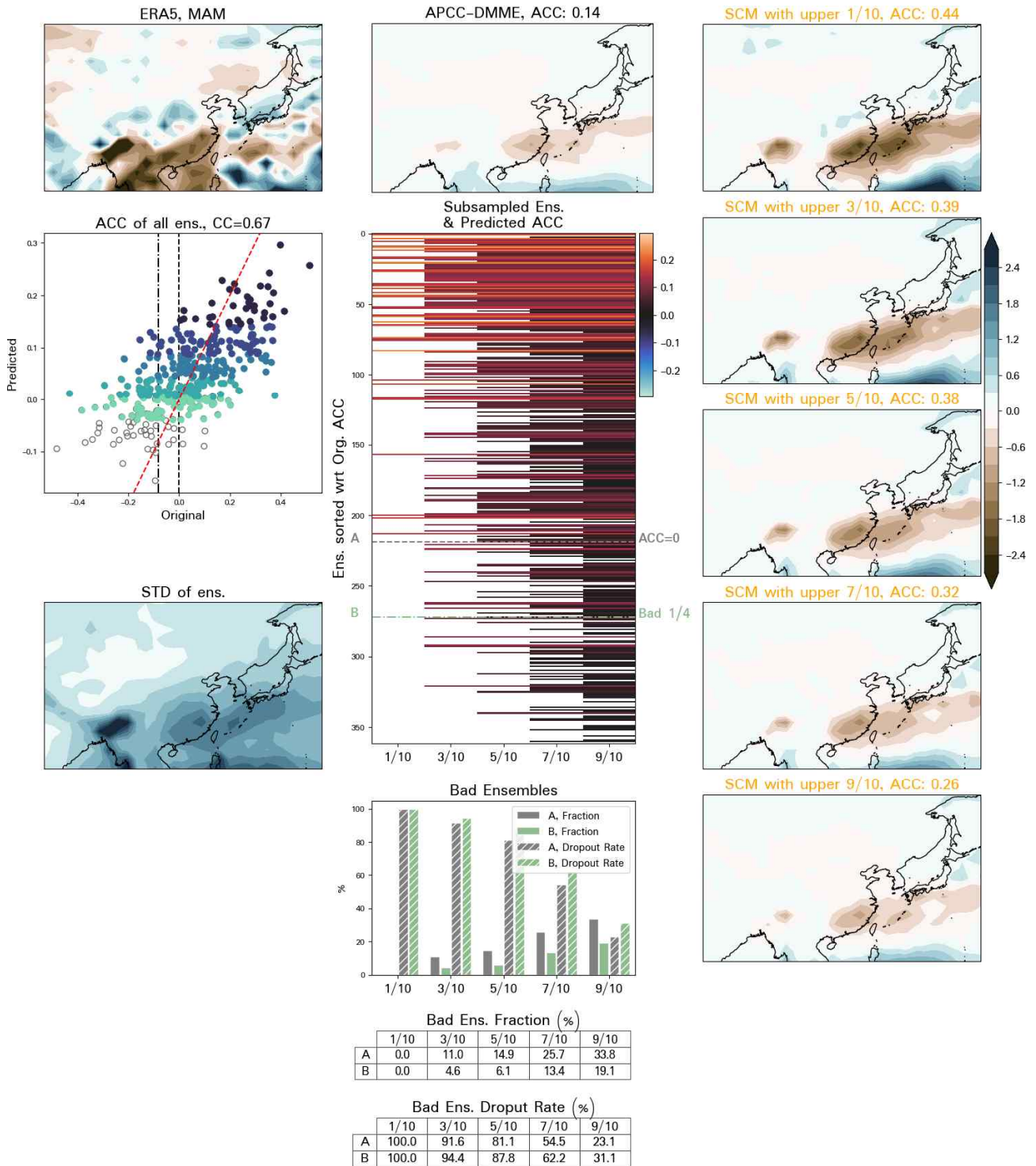


Figure 3.115. Evaluation of ensemble subsampling using an ACC estimation model with ResNet18 for MAM 2023: (scatter plot) ACC estimation performance, (bar plot) fraction of bad ensembles and dropout rate, (map plot) EA rainfall anomalies of sub-sampled ensemble means and their ACC skill with respect to ERA5 observation.

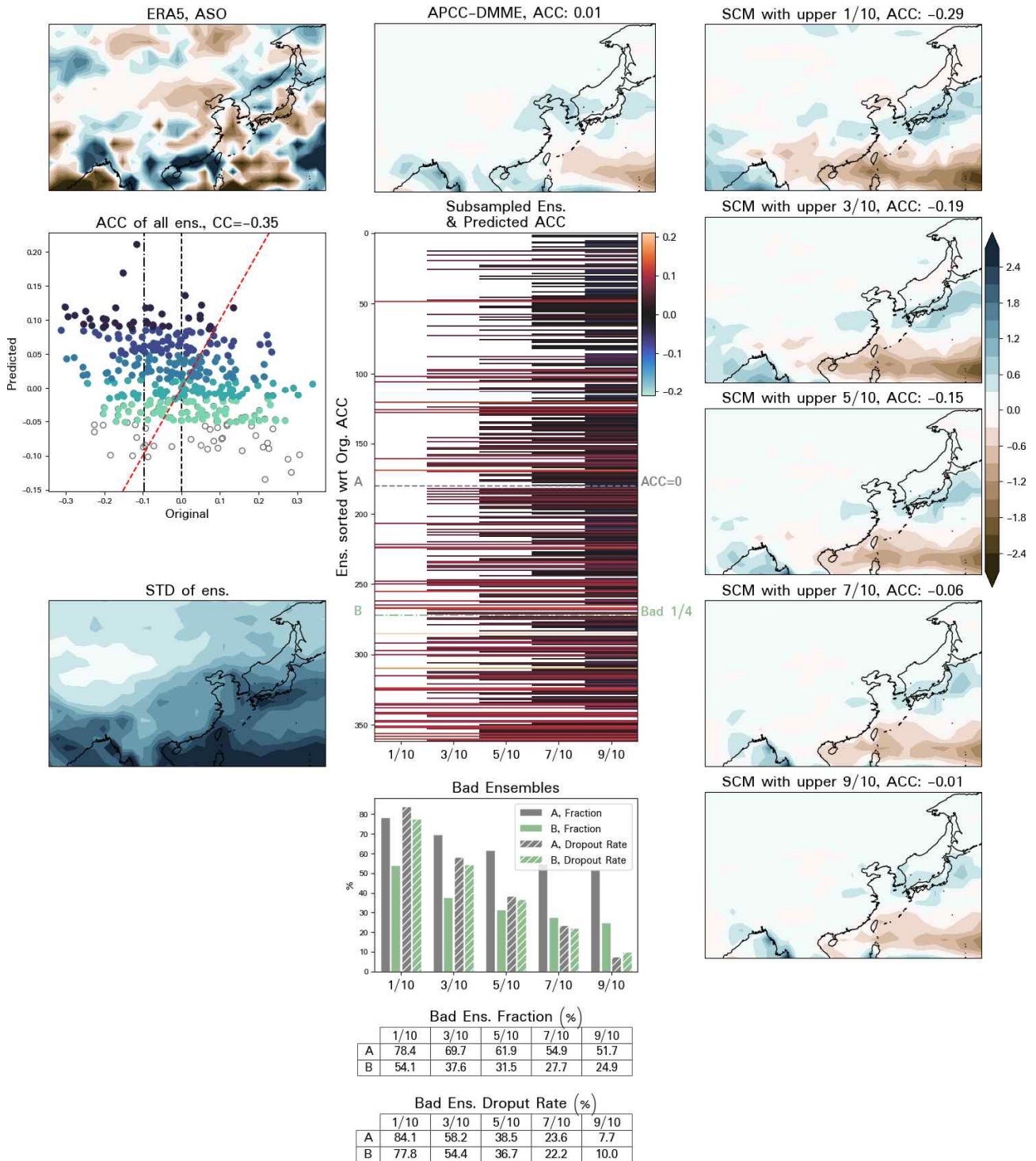


Figure 3.116. Evaluation of ensemble subsampling using an ACC estimation model with ResNet18 for ASO 2023: (scatter plot) ACC estimation performance, (bar plot) fraction of bad ensembles and dropout rate, (map plot) EA rainfall anomalies of sub-sampled ensemble means and their ACC skill with respect to ERA5 observation.

3.6.2.2. 예측 성능 원천 (변수/지역) 진단

2023년 전체 계절, 모든 앙상블에 대한 ACC 추정용 RESNET 모델의 결과는 Figure 3.118

좌측 산포도와 같다. 추정된 ACC와 실제 ACC 사이의 Pearson Correlation Coefficient 값이 0.54로 높으나 성능의 계절 cycle이 이 안에 녹아 있어 앞선 분석들처럼 계절별 특성을 면밀히 살펴볼 필요가 있다. 이 산포도를 바탕으로 실제 ACC와 추정 ACC에 대한 contingency table을 만들고, Good/Miss(light brown), Good/Hit(dark blue), Bad/Miss(light blue), Bad/Hit(dark brown)의 4개의 앙상블 그룹으로 분류해 보았다. x,y 축에서 1/4, 3/4 quartile에 해당하는 값을 기준으로 하였고 경계에 있는 marginal 앙상블들은 분석에서 제외하였다. 각 그룹은 실제 성능이 좋은 앙상블 중에서 ACC 추정 모델이 실패한 경우와 성공한 경우, 실제 성능이 나쁜 앙상블 중에서 ACC 추정이 실패한 경우와 성공한 경우, 이렇게 나눠서 해석할 수 있다. ACC 추정 모델에서 어떠한 입력 변수와 지역이 모델 결과값에 어떠한 방향으로 기여하게 되는지를 Occlusion Sensitivity(OS) 분석을 통해 확인해 보았다. Occlusion Window 사이즈는 25km로 세팅하였으며, 해당 격자의 별수별 입력값을 0으로 처리하는 경우와의 차이값을 이용하는 기존의 방식을 따랐다. 성능 추정 기여도가 크고 유의미한 해석이 가능한 Z500, V850, PREC 세 변수에 대한 결과는 Figure 3.118 우측 2차원 공간맵과 같다. 산포도 오른쪽 앙상블의 경우(bluish box), 뱅골만 지역, 500hPa 지위고도장에서 강한 민감도를 보여 이 지역 입력 지위고도장에 시그널이 존재하는 경우 RESNET 모델이 성능을 높게 추정할 가능성을 점쳐볼 수 있다. HIT case인 1/3 사분면에 해당하는 앙상블들에 대한 결과를 비교해 보면, 붉은 박스로 표시된 동아시아 영역 안, 중국/동남아시아 ‘내륙’에서의 입력 강수 시그널에 직접적인 영향을 받는 것을 알 수 있다. 반대로 MIS case인 2/4 사분면에 해당하는 앙상블들에 대한 결과를 비교해 보면, 붉은 박스 바깥의 ‘해양’에서의 지위고도 입력장에 대한 민감도가 상당히 큰 것으로 보인다. 풍상측 인도양에서 시그널이 있을 때는 실제 낮은 ACC보다 높게 추정하고, 풍하측 서태평양에서 시그널이 있을 때는 실제 높은 ACC를 낮게 추정하는 경향을 확인할 수 있다. 이를 통해 동아시아 영역 바깥 해양에서의 지위고도 시그널을 ACC 추정 실패의 원인으로 작용할 수 있다는 해석이 가능하다. 하지만 이 연구에서 사용한 OS방식은 해당 격자 해당 입력 변수의 부호에 대한 방향성, 강도에 대한 민감도를 제시하지는 못하기 때문에, 추후 이를 보완하는 방식으로의 분석 방식 변경이 필요하다고 생각한다.

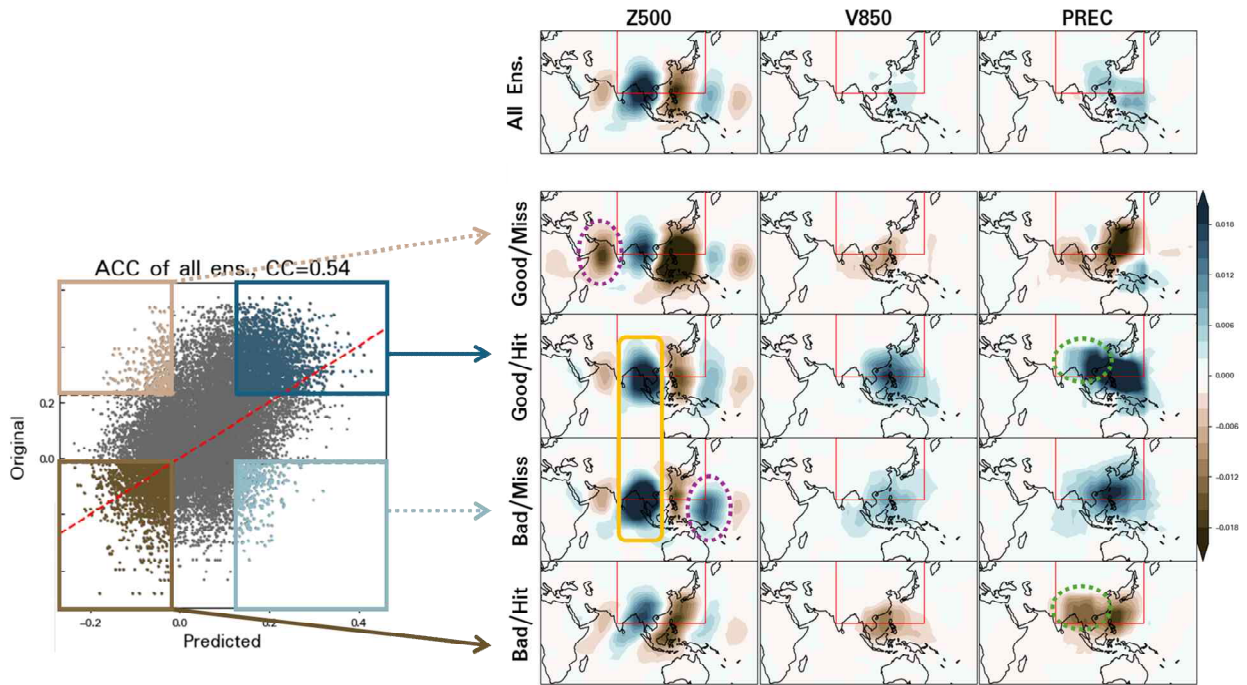


Figure 3.117. Occlusion sensitivity analysis for three variables: geopotential height at 500hPa, meridional wind at 850hPa, and total precipitation. (Scatter plot, left) Division of ensembles for 12 seasons in 2023 into four groups: Good/Miss, Good/Hit, Bad/Miss, Bad/Hit based on the agreement of original and predicted ACC. (Shading contour plots, right) Occlusion sensitivity composite maps for the four groups, using a 5x5 (25deg by 25deg) occlusion window.

2023년 MJJ에 대한 OS 분석 결과는 Figure 3.119의 상단과 같다. 전체적인 ACC 추정 성능은 Pearson Correlation Coefficient = 0.34 정도이고, 앞서와 같은 contingency table 형태로 앙상블 그룹을 나누어 분석해 보았다. 4개 그룹 대부분 붉은 박스, 즉 동아시아 내부 영역의 민감도가 큰 것으로 나타났다. 북서태평양 해양에서 ACC 추정 기여도가 뚜렷하고, 추가적으로 동인도양 시그널이 음의 ACC값 추정(2/3사분면)에 기여하는 것으로 보인다. 모형이 ACC를 높게 추정하는 앙상블 그룹(1/4사분면)에서 실제 ACC와 합치 여부는 방글라데시 지역 육지 시그널이 결정적인 것으로 보인다. 마찬가지로 모형이 ACC를 낮게 추정하는 앙상블 그룹(2/3사분면)에서도 중국 남부 내륙의 기여도가 반대로 나타나 육지 시그널이 중요한 것으로 판단된다.

4개 앙상블 그룹별로 SLP, U850 입력 변수 합성장에 대해 Miss case 와 Hit case 간의 차이는 Figure 3.119 하단과 같다. RESNET 모형이 ACC를 낮게 추정한 두 앙상블 그룹(2/3사분면) 사이의 차이를 살펴보면(Figure 3.119 하단 좌측), 선행시간 T=1에서 북인도양에서 고기압성 패턴이 뚜렷하고, 선행시간 T=2에서는 한반도 이남 해양 지역에서 Pacific-Japan(PJ) 혹은 북서태평양아열대고기압(Western North Pacific Subtropical High, WNPSH) 패턴처럼 남북방향 쌍극자 형태가 나타난다. RESNET 모형이 낮은 ACC 성능을 높게 추정하는 경우에, T=2 입력장에서 상대적인 양의 위상 WNPSH 패턴이 확인되며, 따라서 동아시아 강수와 관련성이 크다고 알려진 WNPSH 모드가 ACC 추정에 기여한다고 판단할 수 있다. 반대로 ACC를 높게 추정한 두 앙상블 그룹(1/4사분면) 사이의 차이를 살펴보면(Figure 3.119 하단 우측), 모형이 높은 ACC 성능을 낮게 추정한 경우에는 선행시간 T=3에서 상대적으로 음의 위상의 WNPSH와 유사한 패

턴이 나타난다.

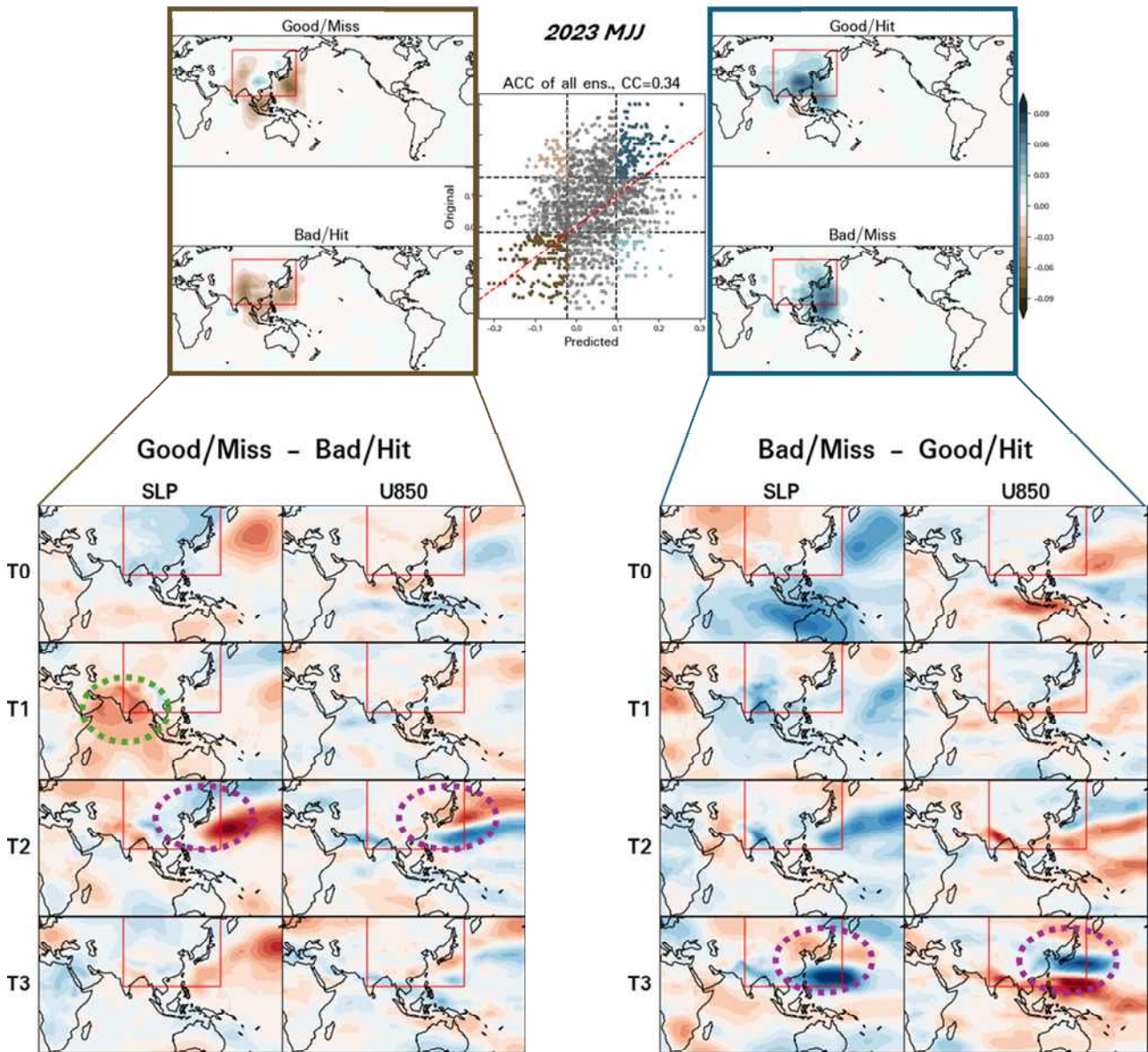


Figure 3.118. (upper) Occlusion sensitivity map for four ensemble categories for 2023 MJJ. Occlusion applied to all input variables with a 5x5 (25deg by 25deg) window. (lower) Anomaly composite difference between ensemble categories: (left) Good/Miss and Bad/Hit, (right) Bad/Miss and Good/Hit for SLP and U850.

이러한 개연성은 WNPSH 지수값에 대한 비교를 통해 좀 더 명확하게 이해할 수 있는데, He and Zhou (2015)의 정의(Eq. [3.1])를 이용해 입력장 하층바람장(U850)으로부터 생산된 WNPSH 지수를 각 앙상블 그룹에 대해 평균한 결과는 아래 Figure 3.120과 같다.

$$\int_{120^{\circ} E}^{150^{\circ} E} \int_{25^{\circ} N}^{35^{\circ} N} u850 d\theta d\phi - \int_{130^{\circ} E}^{150^{\circ} E} \int_{10^{\circ} N}^{20^{\circ} N} u850 d\theta d\phi \quad [3.1]$$

T=-3부터 T=-1까지의 관측 월별 WNPSH 지수 시계열은 앙상블 그룹에 관계없이 모두 같으나 앙상블 예측장을 입력으로 두는 T=0부터의 지수는 앙상블 그룹간에 큰 차이를 보인다. 예측 목표인 2023년 MJJ 평균 WNPSH 관측 지수값(검정 실선)은 0에 가까워 북서태평양에서의 순환장에서 뚜렷한 특이점은 없다고 볼 수 있다. HIT 앙상블 그룹(짙은 실선)은 강수 예측 목표 관측 지수값과 매우 유사하고 시간적으로 안정된 양상을 보인다. 이는 HIT 앙상블 그룹 입력장이 적절한 북서태평양 대규모 순환장 정보를 포함하고 있다고 이해할 수 있을 것이다. 하지만 점선으로 표현된 MISS 앙상블 그룹은 관측 지수값으로부터 크게 벌어지고 시간적으로 불안정한 양상을 보인다. 구체적으로 서술하면, 입력장 선행시간이 길어짐에 따라 Good/Miss 앙상블 그룹은 양의 위상으로 벌어지고, Bad/Miss 앙상블 그룹은 음의 위상으로 크게 벌어지게 된다. 이러한 Miss 그룹 앙상블들은 부적절한 북서태평양에서의 대규모 순환장, 즉 WNPSH 정보가 입력장에 포함되어 있음을 알 수 있다. 북서태평양 지역의 대규모 순환장의 동아시아 강수에의 기여도가 큰 것은 기후학적으로 잘 알려져 있고, ACC 추정용 RESNET 모델이 이러한 역학 과정을 적절히 이용해 훈련을 진행한 것으로 보인다. 북서태평양 대규모 순환장 자체를 후처리해서 그것으로부터 동아시아 강수를 추정하는 방식으로라도 예측성을 개선할 수 있지 않을까에 대한 합리적 의심이 가능하며, 이를 바탕으로 추가적인 실험을 진행하게 된다.

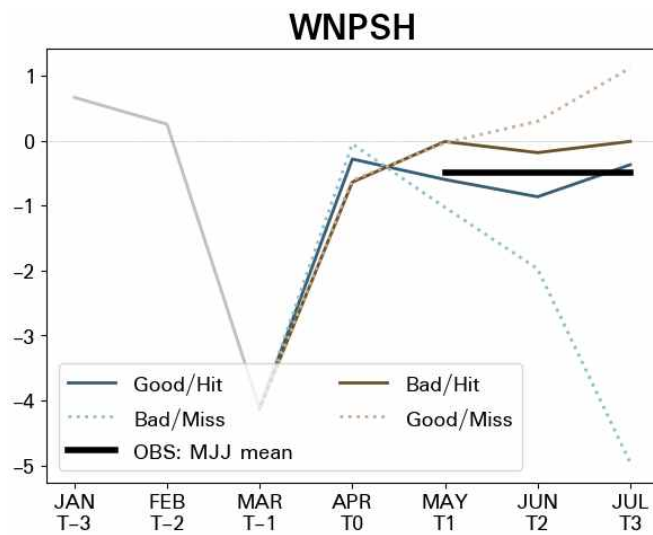


Figure 3.119. Monthly timeseries of the WNPSH index: ERA5 observation for the past three months (T-3 to T-1) and dynamical model ensemble mean of four categories (Good/Hit, Bad/Hit, Bad/Miss, Good/Miss) for four months from T0 to T3. Black horizontal solid line shows the observed seasonal mean WNPSH index for MJJ 2023.

3.6.2.3. 동아시아 하층 바람장 재현성 기반 앙상블 선별 모델 평가

WNPSH를 제대로 재현하는 앙상블이 높은 예측성을 가질 수 있다는 앞선 분석을 바탕으로 WNPSH 지수 재현성에 기반해서 또 다른 앙상블 선별 모델을 설계하게 된다. 초기에는 단순하게 1차원 목적 변수를 ACC 대신 WNPSH 지수 시계열로 대체해 RESNET 모델을 훈련하고, 전체 앙상블 평균에 가까운 앙상블을 재현성이 높다고 가정해 Root Mean Squared Error(RMSE) 값이 낮은 순서로 10, 30, 50, 70, 90%의 앙상블들을 뽑아내고 그 때의 평균 강수장의 예측성을 ACC로 평가했다. 결과적으로는 기존 ACC 추정 기반 앙상블 선별에 비해 유의한 성능개선은 없었다. 단순 지수값만으로 강수에 관련되는 북서태평양 순환장 정보를 추출하기 어렵다는 판단하에 추가 실험을 수행했다. 북서태평양 대규모 순환장 정보를 1차원 기후 지수가 아닌 하층 바람장(U850) 2차원 공간장으로 대체해서 목적 변수로 두고, RESNET이 아닌 UNET 계열 아키텍처로 수정해 모델을 훈련해 보았다.

출력된 전체 앙상블 평균을 참값으로 가정하고, 개별 앙상블의 U850 공간장 RMSE 값이 작은 순서로 10, 30, 50, 70, 90%의 앙상블들을 뽑아내 그 때의 평균 강수장의 예측성을 ACC로 평가해 보았다 (Figure 3.121). MAM과 AMJ를 제외하고 늦여름부터 겨울까지 (JAS to JFM) 성능이 성능 추정 기반 접근법 대비 개선된 것을 알 수 있다. 특히 늦여름철의 성능 개선이 두드러지는데 JAS, ASO의 경우 ACC 성능이 양의 값으로 올라올 뿐만 아니라 기준 %tile이 커짐에 따라 성능이 개선되는 이상적인 모습을 보였다.

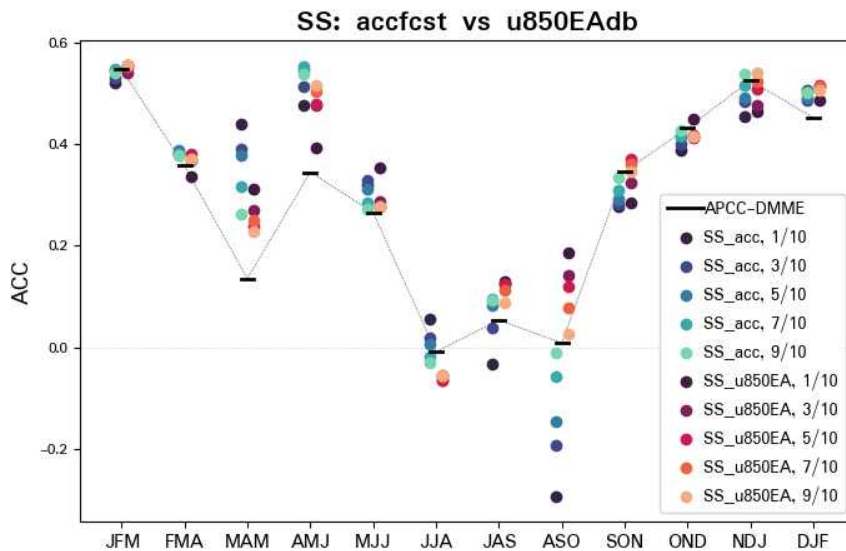


Figure 3.120. Performance of the second SS approach: in addition to the ACC-based approach, results for five subsampling models (reddish circles) averaging ensembles with the smallest U850 anomaly errors over EA. U850 anomalies for the next season estimated using an Attention-UNet architecture, with the error defined as the departure of EA U850 anomalies from the ensemble mean.

ASO 계절에 대한 성능 추정 결과를 자세히 살펴보면 Figure 3.122와 같다. 기존 성능 추정 기반 모델에서는 산포도에서 추정 ACC와 실제 ACC가 음의 상관관을 보였으나 (Figure 3.122 a) 하층바람장 재현성 기반 모델에서는 다소 비례하는 형태로 전환된 것을 알 수 있다 (Figure 3.122b). 5개 개별 기준에 대해서 추정된 2차원 강수장의 공간 분포를 보면, 성능 추정 방식에서는 남북방향 쌍극자 형태로 단순한데 비해 하층 바람장 추정 기반 방식에서는 복잡한 여름철 공간 특성이 일정 수준 재현되는 것으로 보인다. 특히 상위 1/10 (90%tile) 기준 결과는 인도 서쪽, 중국 내륙 북쪽과 필리핀 동쪽 해양의 건조 시그널과 한반도부터 러시아까지의 대륙 동안과 미얀마 지역의 습윤 시그널을 보여, APCC-MME나 성능 추정 기반 인공지능 모델 결과에 비해 관측에 부합하는 특성을 확인할 수 있다. 다소 낮은 10%tile 기준에서도 APCC-MME보다 높은 ACC를 보인다는 점에서, 불량한 앙상블들이 선별적으로 탈락되었다는 것을 알 수 있고, 하층 바람장에 대한 강수 의존성을 고려해 모델을 설계한 전략이 매우 유용했다는 것을 나타낸다.

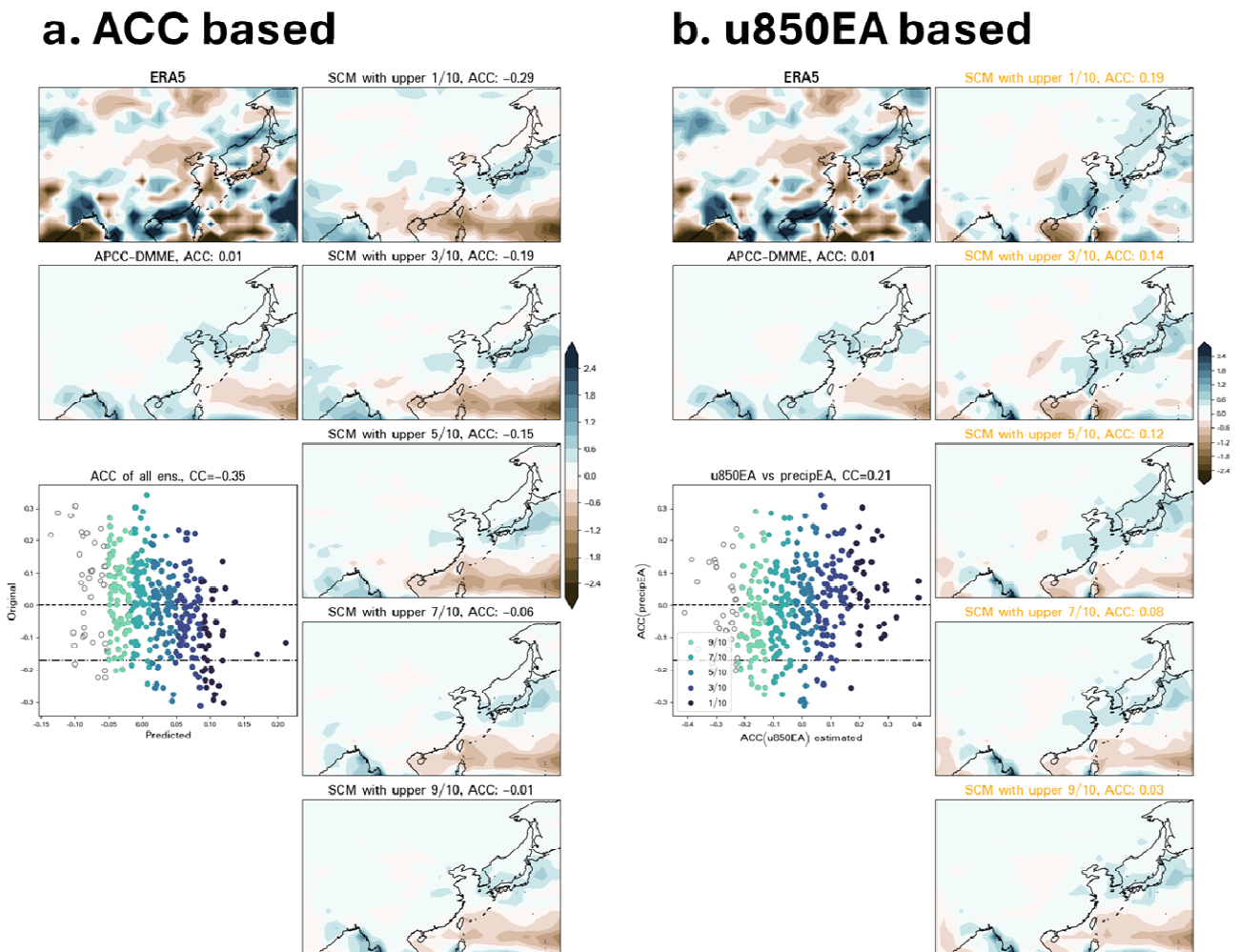


Figure 3.121. Comparison of two ensemble subsampling approaches for ASO 2023: (a) based on EA rainfall ACC estimation, (b) based on EA lower-level wind field reproducibility.

3.6.3. AI 후처리 접근법 2: 오차 제거 (DeBiasing, DB) 방식의 아키텍처 민감도 평가

역학 모델 자료를 입력으로 넣고 실제 관측값으로 맞추어 나가는 ‘오차 제거’ 방식은 ‘오차 보정’ 이라고도 불리며 인공지능을 활용한 모델 자료 후처리에서 주로 쓰이고 있다. 입력장에 다양한 변수와 다양한 선행시간 자료를 추가하는 형태로 변형을 주어 오차 제거 방식을 수행해 보았다. 또한 기본 UNET에 더해 ATTUNET, RESUNET, SWINUNET 등 변형된 아키텍처에 대한 민감도 실험을 주로 수행하였다. 이에 더해서 Table 2.13에서 제시한 다양한 입력 변수, 초매개변수에 대한 세부 조정 실험 역시 수행하였는데 개별 아키텍처별로 learning rate에 대한 민감도가 다소 높게 나오는 것을 알 수 있었다. 개별 실험마다 공통적으로 입력변수는 8개 대기 변수에 LSM, 위도 정적 변수가 추가된 형태이고, mse 비용함수를 사용하였다. 선행 시간 T0를 추가한 경우와 추가하지 않은 경우로 나뉘고, early stopping용 epoch tolerance, batch size, learning rate도 실험마다 다르게 세팅되었다.

총 29개 변형 실험에 대한 오차 제거 성능은 Figure 3.123과 같다. 전체적으로 JFM부터 ASO까지의 계절에서는 아주 개선된 결과를 보이는데 반해, SON부터 DJF까지 4개 계절에 대해서는 오히려 성능이 크게 하락하는 결과를 보인다. 각 계절별로 최적의 성능을 보이는 아키텍처가 달라지는 특성이 보이는데, 가장 많은 계절에서 SWINUNET이 최고 성능 모델로 뽑히고, 다음으로 기본 UNET이 뽑혔다. RESUNET은 전반기 계절에서는 성능이 다른 아키텍처에 비해 살짝 떨어지는 경향이 있지만, 반면에 하반기에서 성능 하락이 다소 적어 전체 계절 평균 측면에서는 상당히 우수하다고 판단할 수 있다. 하층 바람장 DB 모델과 강수 DB 모델을 적층하는 ST 모델들(yellowish)의 경우 상반기 계절에서 성능은 다소 떨어지지만 하반기에서는 다른 오차 제거 모델들에 비해서 성능이 높고 DJF 계절에서는 가장 우수한 모델로 뽑혔다. 그럼에도 불구하고 가을철 전반에서는 APCC-MME 성능을 극복하지 못한다.

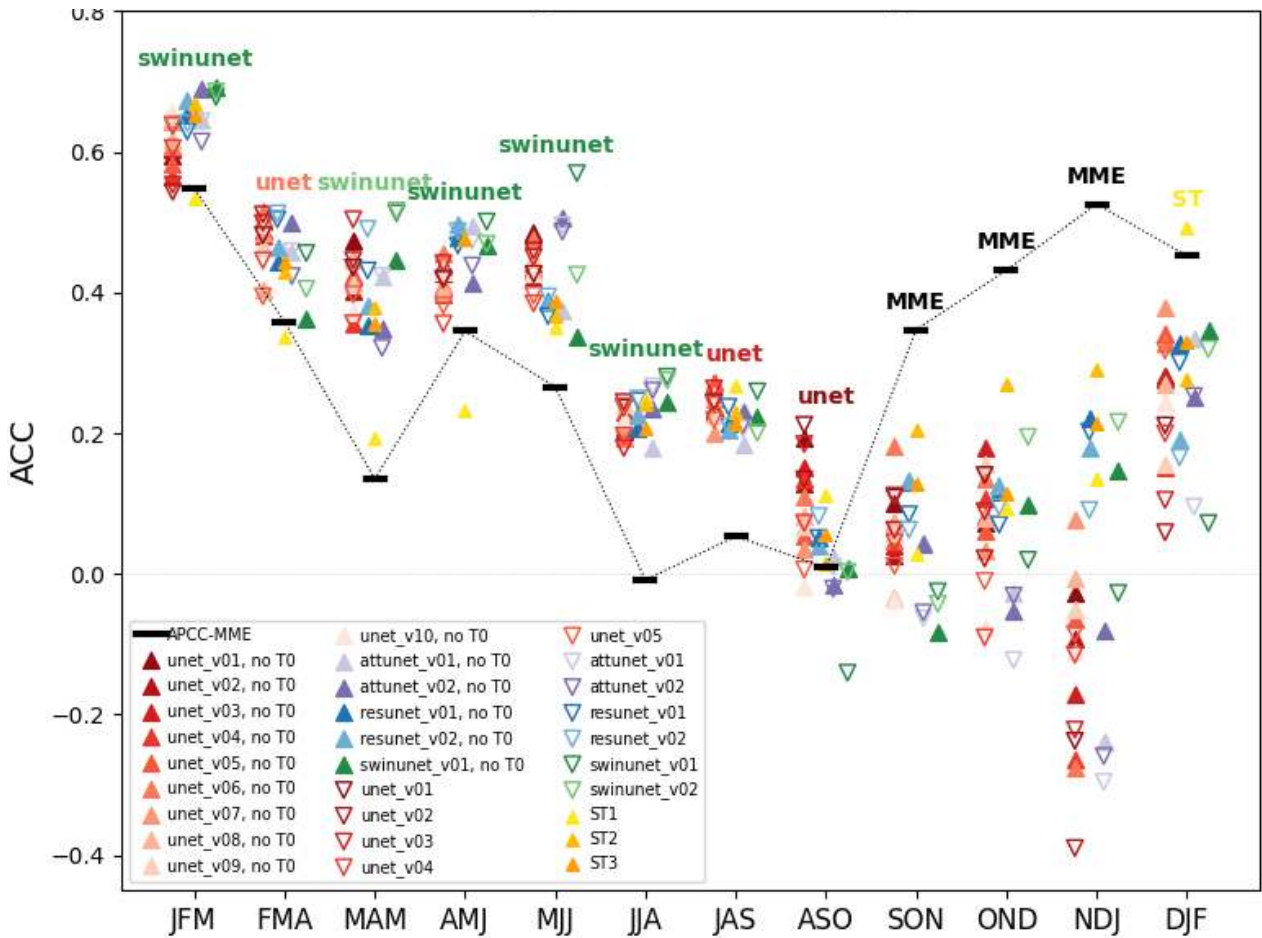


Figure 3.122. Performance of the DB approach: comparison of EA rainfall ACCs between original APCC-MME (black horizontal lines) and 26 variations of deep-learning models (triangles) with different training inputs, architectures, and versions. Closed yellowish triangles denote three variations of the Stacking Two-models (ST) approach. Results are for 12 seasons of the test year 2023.

2023년 12 계절 동아시아 강수 분포에 대한 관측장과 인공지능 모델들의 재현장을 비교해 보았다 (Figure 3.124). 강수의 경우 계절 평균에서도 매우 복잡한 지역 분포를 보여 이에 대한 역학 모델의 공간 재현성은 매우 낮을 수 밖에 없다. 한반도를 포함한 극동 아시아 지역에서는 2023 JFM/FMA 시기에 북쪽이 건조하고 남쪽이 습윤한 쌍극자 형태의 시그널을 보이다가 그 이후 한반도 위도대로 습윤한 강수대가 올라오는 패턴으로 전환되고, OND/NDJ/DJF 하반기에는 북쪽이 습윤하고 남쪽이 건조한 쌍극자 형태로 뒤집어지는 것을 알 수 있다. 이렇게 JFM과 DJF 패턴이 뒤집히는 특성은 23/24 엘니뇨 발달에 따른 지역 반응이라고 판단할 수 있다. 역학 모델들은 계절에 따라 패턴이 뒤집히는 것을 적절히 재현하고 있는데 이는 ENSO 지수 및 지역 반응에 대한 예측성이 높기 때문으로 판단된다. 인공지능 모델에서는 기본적으로 입력 샘플에 기반해 학습이 이루어진다. 본 연구에서는 APCC-MME real-time forecast 자료가 존재하는 2013-2022년 10년에 대한 샘플들만을 학습시키기 때문에 딱 10년에 대한 계절 사이클 realization이 존재한다고 볼 수 있다. 이 시간 내에 15/16 엘니뇨 한 번만 발생했고 이 엘니뇨 시기에 다른 주요 지수 상태는 23/24년과 매우 다르다. 성층권 상태를 나타내는 QBO, 태평양

전제 장기 변동성을 나타내는 PDO, 그리고 엘니뇨의 슬레이브로 알려진 인도양의 DMI 지수가 두 케이스 초반 즉, 엘니뇨 발달시기에 매우 다른 양상을 보이기 때문이다. 특히나 과거 엘니뇨 발달시기 여름/가을 합성장에서 동아시아 특히 한반도 지역 아래의 건조 시그널이 뚜렷한데 반해, 2023년의 경우 그 반대의 특성을 보인다. 즉, 23/24 엘니뇨에서 동아시아 지역 강수 영향은 15/16 엘니뇨 시기와 상이할 수 밖에 없고, 인공지능은 과거에 없던 사례에 대한 창조적인 예측이 불가하기 때문에 강수 패턴의 시공간 변동성이 제대로 예측되지 않은 것으로 해석할 수 있겠다.

실제 예측된 4개의 오차 제거 후처리 모델들이 제시하는 강수 계절 진화 형태는 조금씩 다르지만, 하나같이 23년 하반기 계절에서 한반도로부터 연장되는 중위도 건조 시그널과 저위도 습윤 시그널의 쌍극자 형태로 관측과 반대되는 강수 패턴을 보인다. 상반기에서는 중국 남부와, 방글라데시 벵갈만 주변의 건조 시그널의 잘 캐치해 성능 개선이 뚜렷하다. 반대로 하반기에는 한반도와 남중국해 주변의 패턴을 뒤집어서 추정하기에 성능 하락이 뚜렷하다. 다만, UNET에 비해 RESNET과 SWINUNET의 패턴이 다소 희석되어 나타나는 특징을 보인다. 이러한 결과는 역할 모델 결과 보정에 있어서 샘플의 중요성을 재확인시켜준다. RTF 샘플 기간을 확장하는 것이 현실적으로 불가능하다는 점이 오차 제거 방식 후처리 모델의 매우 심각한 한계로 지적될 수 있겠다. HINDCAST 자료를 추가하여 오차 제거 모델의 샘플을 늘릴 수도 있으나, 재분석 자료를 초기장으로 세팅하기 때문에 모델 예측장을 초기장으로 세팅하는 RTF와 오차의 양상이 다를 것으로 보이고, 따라서 같은 샘플 그룹에 추가하는 것이 적절하지 않다고 판단하였다. 또한 대략 30년 정도인 HINDCAST realization 역시 인공지능 모델이 과거를 재현함에 넉넉한 샘플을 제공하는 것은 아니기 때문에 그 실효성에 있어 의문이 있을 수 있다.



Figure 3.123. Comparative analysis of EA rainfall anomaly maps among ERA5 observations, APCC-MME forecasts, QM, and outputs from four different DB models for 12 seasons of the test year 2023.

3.6.4. AI 후처리 접근법 3: 전이 학습 기반 강수 진단 (Transfer-Learning based Diagnosing, TLD)

앞선 오차 제거 방식의 후처리에서 드러난 한계는 학습에 활용되는 과거 샘플이 부족하다는 점이다. 이러한 한계를 극복하기 위한 방안으로 전이 학습이 널리 활용된다. 실제 관측 자

료를 확장하는 것이 물리적으로 불가하니 개별 모델의 free run 장기 자료를 통해서 변수간의 역학적인 관련성을 선학습하고 (pre-training) 이를 관측 샘플을 통해 미세 조정 (fine-tuning)하는 방식이다. 본 연구에서는 CESM2-LENS 장기/대량 앙상블 자료를 사용하여 전이 학습을 수행하였다. historical 실험으로 165년 과거 샘플에 대해서 100개의 앙상블을 모두 이용하였기 때문에 확장된 과거 realization에 대해 변형된 자료까지 많은 자료에 대한 학습이 가능하다고 판단했다. 개별 계절에 대해 16,500개의 샘플에 대한 학습이 가능하기 때문에 전이 학습 방식에서는 season-agnostic이 아닌 season-specific하게 모델을 훈련시켰다. 앞 절에서 언급했던 오차 제거 방식의 가을철 성능 하락은 인공지능 모델이 계절 특성을 온전히 학습하지 못해서이고, 계절별로 독립적으로 모델을 훈련시키면 이러한 성능 하락을 다소 완화시킬 수 있을 것이라는 동료의 제안에 대한 대응이기도 하다. ATTUNET과 SWINUNET 아키텍처를 주로 사용하고 변수 조합과 초매개변수 변형을 통해 몇 가지 민감도 실험을 수행하였다. 특히, T-3/T-2/T-1 선행에 대해서는 변수를 다양화하고, 대신 T1/T2/T3에 대해서는 강수 변수를 제외함으로써 강수에 대한 diagnostic한 접근 방식으로 입력 변수를 세팅해 민감도를 실험했다. 12 계절에 대해 유의한 개선을 보인 3가지 모델 실험들(기존 변수에 대한 SWINUNET 모델, 새로운 변수 조합에 대한 ATTUNET, SWINUNET 모델)에 대한 ACC 값을 오차 제거 방식에 더해서 도식화 하였다 (Figure 3.125). 안타깝게도 2023년 12 계절에 대해서는 전이 학습 방식이 오차 제거 방식을 온전히 극복하는 결과를 보이지는 않았다. 특히, 상반기에 성능은 겨우 APCC-MME를 넘어서는 정도로 개선이 크지 않았다. 하지만 가을철을 포함한 하반기에서의 성능은 오차 제거 방식의 성능 하락을 일정 정도 완화시키고, 특히 DJF 계절에 대해서는 최고 성능을 보여 전이 학습 방식의 가능성을 보여주기도 했다. CESM2-LENS 자료의 경우 historical 실험을 가져왔기 때문에 학습할 수 있는 realization이 165년에 그친다는 한계가 있을 수 있다고 생각한다. 향후 장기 free run 실험 자료를 가져와서 좀 더 다양한 realization에 대한 역학적 관련성을 학습할 수 있다면 추가적인 성능 개선이 가능할 것이라 판단한다.

DB vs TLD

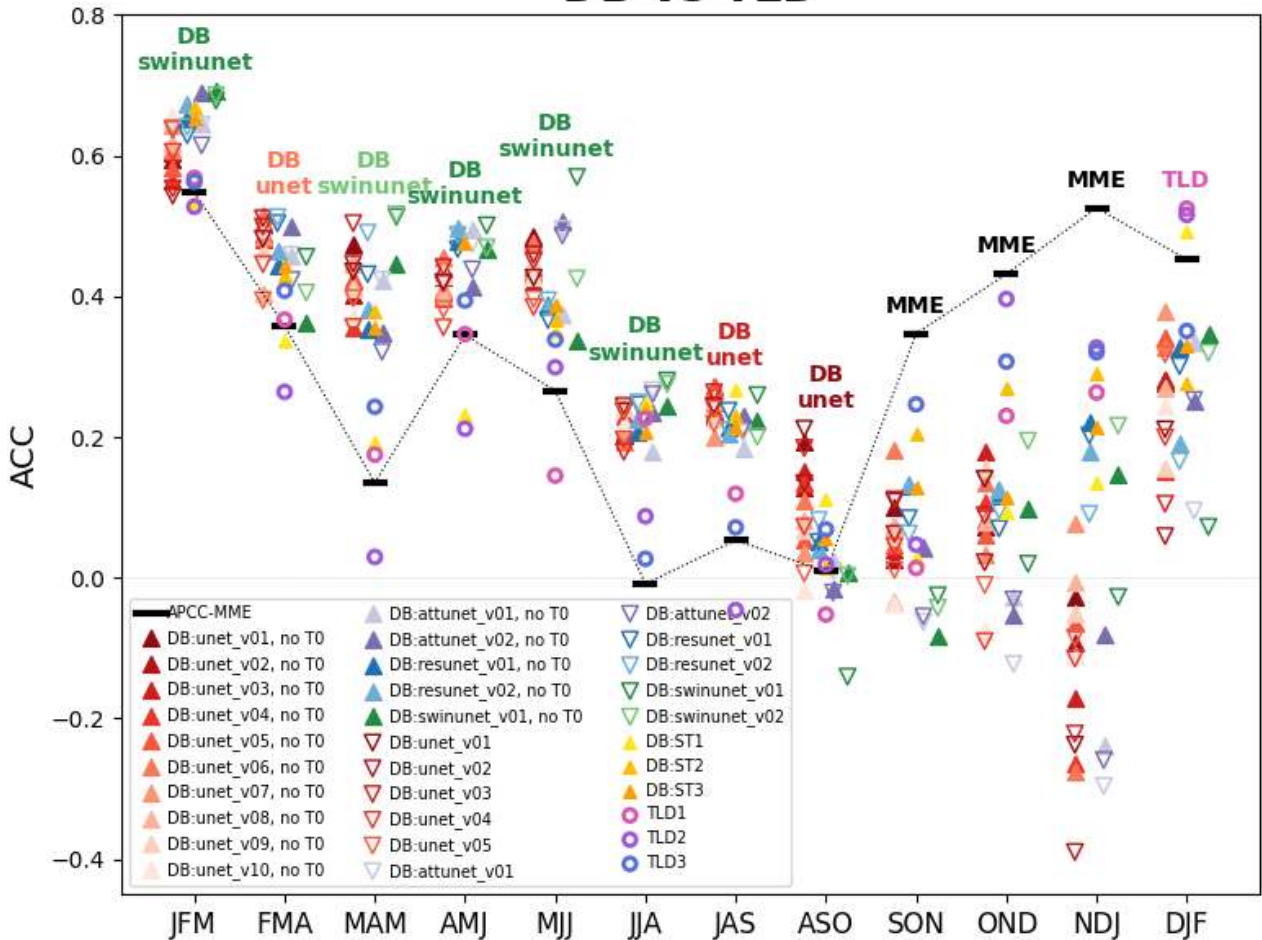


Figure 3.124. Comparison of DB and TLD approaches: ACCs of 29 DB models and 3 TLD models relative to the original APCC-MME for EA seasonal rainfall. Results are for 12 seasons of the test year 2023.

3.6.5. 기존 통계 편이 보정법: Quantile Mapping 적용성 평가

인공지능을 활용한 후처리 모델 개발이 의미를 갖기 위해서는 기존 통계 방식의 후처리 결과 대비 개선이 담보되어야 한다. 인공지능 후처리의 성능 기초선으로 APCC-MME 모델에 더해 편이 보정 (Quantile Mapping, QM)을 적용한 결과를 적용하였다. QM은 강수 예측 자료의 보정에 널리 사용되고 있으며, 본 연구에서는 Cannon et al (2015)와 Tong et al (2021)의 Quantile Mapping, Detrended Quantile Mapping (DQM), Quantile Delta Mapping (QDM) 방식을 추가적으로 적용하였다. DQM과 QDM은 과거 분포와 예측 시점의 분포가 다르다고 가정하는 방식으로 기후 변화에 따른 분포 특성 변화까지 고려할 수 있는 보정법이다. 미래의 분포의 변화를 고려하는 방식으로 평균 차이를 더해주느냐 (+, additive), 비율을 곱해주느냐로 (*, multiplicative) 분리해서 방법론을 다양화 하였다. 2023년 12 계절에 대한 5개의 QM 방식을 적용한 성능 평가 결과는 Figure 3.126과 같다. QM을 적용했을 때, 겨울, 봄, 초여름까지 총 7개 계절에서 APCC-MME 대비 성능 개선을 보였고, 전체적으로 ACC 9% (0.288 --> 0.314) 상승 효과를 얻었다. 이러한 계절별 성능 개선은 앞서 언급한 인공지능 기반 후처리 모델들이 보여

주는 계절별 성능 개선 추이와 거의 유사하다. 즉 늦여름~가을철에 대해서는 방법론에 무관하게 성능 개선이 쉽지 않은데, 2023년의 계절 추이가 이전에 없던 새로운 사례이기 때문으로 보여진다. 기본 QM 이외에 DQM multiplicative 방식과 QDM additive 방식 역시 ACC 수치상으로는 개선된 성능을 보이지만 실제 보정된 공간장을 확인했을 때, 동아시아 모든 영역이 건조 시그널로 가득차 원하는 형태의 보정이 이루어지지 않았음을 알 수 있었다 (not shown). 이는 ACC 지표의 한계를 보여주는 것으로, 평가는 다양한 측면에 대해 다각적인 방식으로 이루어져야 할 필요성을 강조한다.

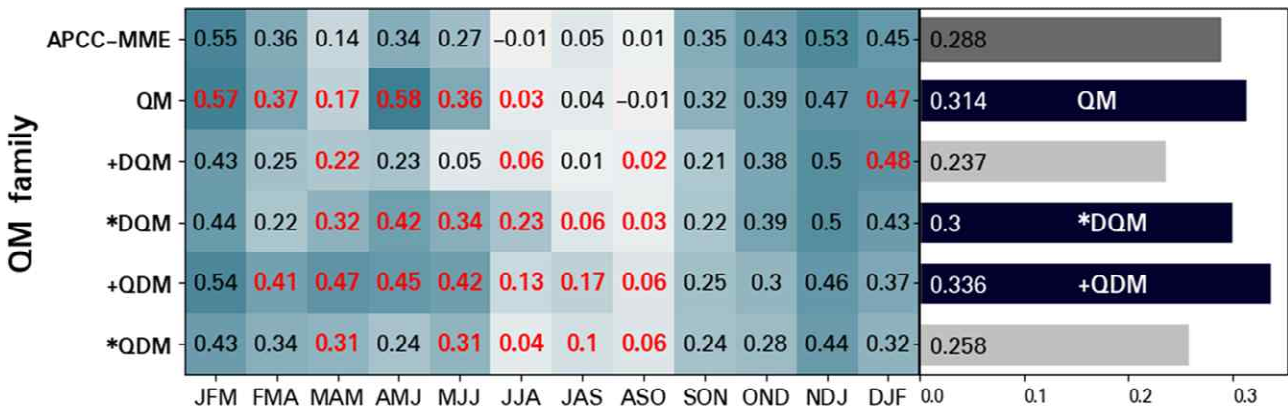


Figure 3.125. Performance of the QM approach: ACCs of EA rainfall anomalies using QM compared to the original APCC-MME for the test year 2023.

3.6.6. AI-aided 다중모델앙상블 예측 프로토타입 제안

동아시아 계절 강수 예측 개선을 위해 본 연구에서 개발한 인공지능 기반 후처리 모델들과 기존 통계 편이 보정법의 성능을 종합적으로 평가해 보았다 (Figure 3.127). 2023년 12 계절에 대해 10개의 SS 모델들, 29개의 DB 모델들, 3개의 TLD 모델들, 1개의 QM 적용한 결과의 ACC 값을 중첩해 도식화했고, 각 계절에서 가장 높은 ACC 성능을 보이는 후처리 방식/아키텍처를 상단에 표출하였다. 12 계절 중에서 7개 계절에서 DB 방식이, 3개 계절에서 SS 방식이, 나머지 2개 계절에서 TLD, QM 방식이 최고 성능을 보인다. 이는 역학 예측 후처리에 있어 인공지능 기술의 잠재력과 이러한 접근 방식의 타당성을 제시한다고 볼 수 있겠다. 그럼에도 불구하고 인공지능 기반 후처리에 있어 몇 가지 한계를 확인할 수 있었다. 안타깝게도 하반기 가을철 시기에 많은 인공지능 모델들이 성능 하락을 보이는데 이는 인공지능 모델이 학습에 활용한 자료가 부족하기 때문으로 현실적으로 이에 대한 극복이 쉽지 않다. 여름 계절의 경우에는 ACC 지표에서는 APCC-MME 대비 성능이 개선되었다고 볼 수 있지만, 관측에서 나타나는 복잡한 공간 해상도를 인공지능 모델도 거의 재현하지 못한다. 마지막으로 후처리 접근법별, 모델 버전별, 계절별로 결과가 상이해 성능 spread가 매우 크다는 것은 그만큼 민감도가 크다는 것을 의미하고, 따라서 실제 현업에 적용한다고 가정할 때 하나의 모델을 선택함으로써 감내해야 할 불확실성이 클 수 밖에 없다.

SS vs DB vs TLD vs QM

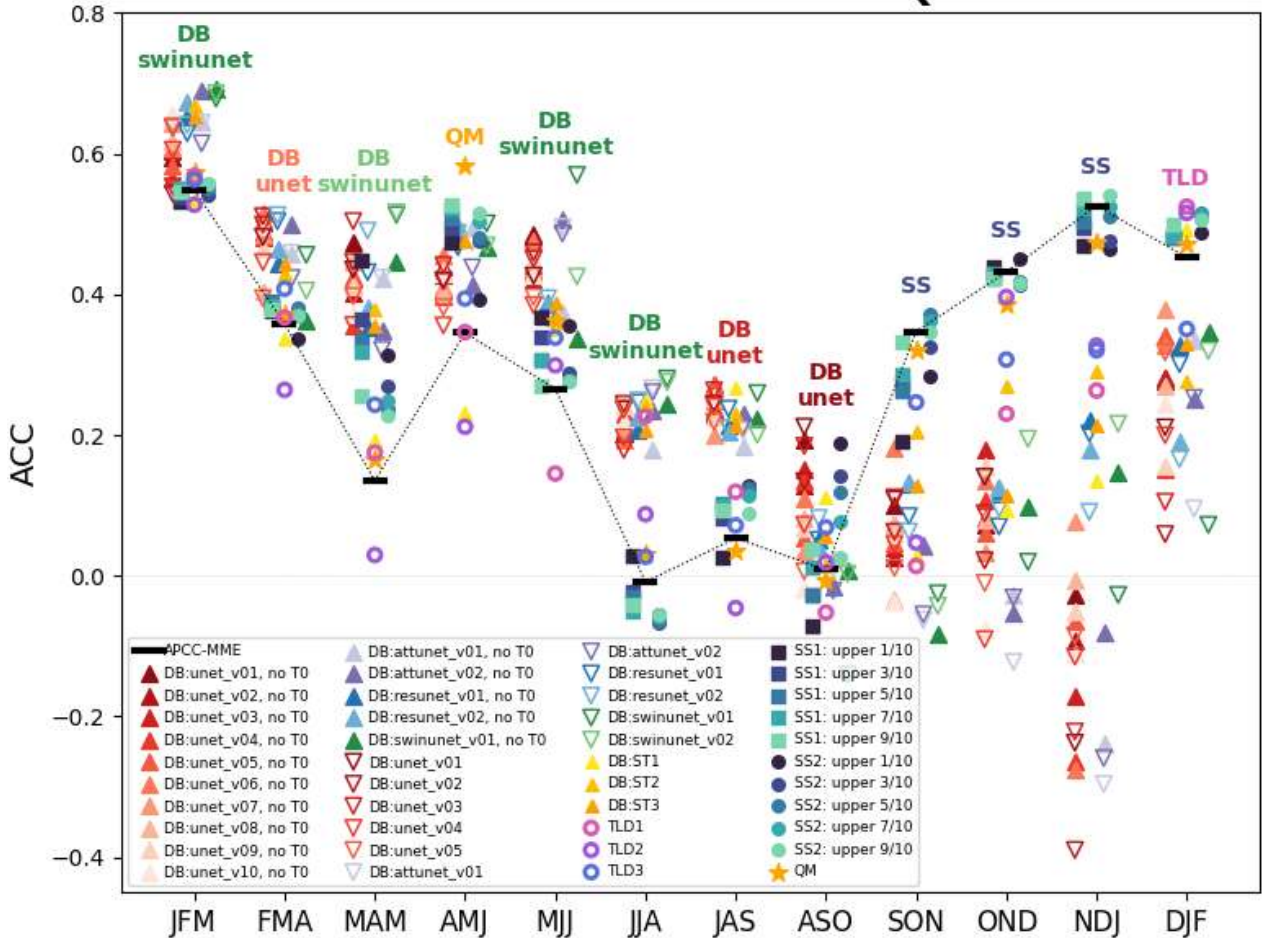


Figure 3.126. Comparison of SS, DB, TLD, and QM approaches: ACCs of 10 SS models, 29 DB models, 3 TLD models, and 1 QM model relative to the original APCC-MME for EA seasonal rainfall. Results are for 12 seasons of the test year 2023.

마지막에 제시한 인공지능 모델간 큰 변동성을 극복하기 위해 MME 기법을 후처리 결과들에 대해서도 적용해 보았다. Figure 3.127에서 표출한 모델들 중 계절별로 안정적이고 우수한 성능을 보이는 3개의 DB 모델, 1개의 TLD 모델, 4개의 SS 모델을 선별하고, 전체 역학 모델 앙상블을 단순 합산하는 Grand Ensemble Average와 Quantile Mapping 결과까지 총 10개 모델 결과를 합산하는 방식으로 AI가 도와주는 다중 후처리 모델 앙상블에 대한 prototype을 제시하였다 (Figure 3.128). 10개 모형에 대한 단순 평균을 AI-aided DMME으로, k=5 앙상블까지 모두 포함하는 샘플에 대해 0을 기준으로 양과 음의 개수를 세는 방식의 median probability forecast를 AI-aided PMME로 정의하였다. 10개 후처리 모델 공간장에 더해 AI-aided MME의 결과를 기존 APCC의 DMME/PMME 결과와 비교할 수 있도록 표출 모듈 개발하였다. 개별 후처리 모델과 AI-aided/APCC DMME의 과거 누적 성능표를 제시함으로써 예보 당사자가 현재 시점 예측을 정성적으로 판단하는 manual forecast 가능성을 염두해 두고 표출 창의 레이어를 구성하였다. Figure 3.128은 2023 AMJ 계절에 대한 결과로서, APCC-MME가 놓치고 있는 아시아 남부지역의 ERA5의 건조 시그널을 AI-aided DMME/PMME가 유의미하게 재현하고 있음을 알 수 있다. 좌측 표에서 ACC 값과 최종 순위를 보면 'A' 라고 표시된 AI-aided DMME 가 1위를

차지해 최고 성능을 보여준다는 점에서 다중 후처리 모델 앙상블 기법의 가능성을 확인할 수 있다.

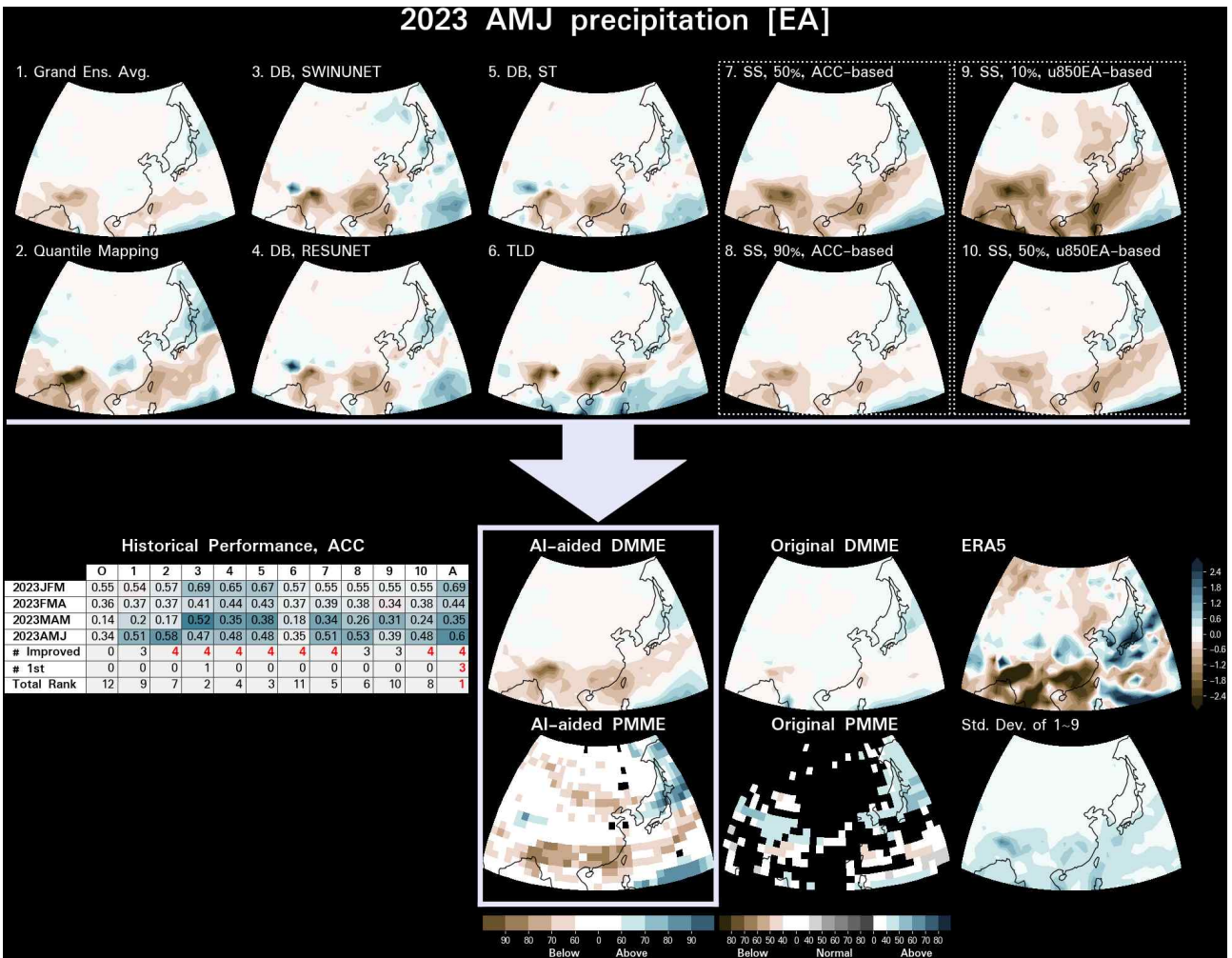


Figure 3.127. A prototype of AI-enhanced multi-model ensemble forecasts for EA rainfall.

다중모델앙상블 기법은 개별 역학 모델들이 가지고 있는 불확실성을 감소시킬 뿐만 아니라 전반적인 성능을 개선하는 것으로 알려져 있다. 인공지능 기반 후처리 모델들을 같은 방식으로 합산함으로써 이러한 이점을 가질 뿐만 아니라, Figure 3.127에서 보이는 모델간 큰 변동성을 극복할 수 있을 것이다. 2023년 12 계절에 대한 AI-aided DMME의 성능은 Figure 3.129에서 확인할 수 있다. 세모로 표시된 AI-aided DMME가 많은 계절에서 개별 후처리 모델 대비 가장 높거나 상위권 순위를 차지한다. 또한 NDJ를 제외한 모든 계절에서 APCC-MME 성능을 넘어선다. 인공지능 후처리 모델 결과를 합산함으로써 새로운 예측에서 극단적으로 잘못된 예측을 할 가능성을 줄이고 양호한 성능 개선을 보일 것이라 기대할 수 있으며, 이는 이 연구에서 제시하는 AI-aided MME 즉, AI가 도와주는 다중 후처리 모델 앙상블 prototype의 잠재력 및 향후 활용 가능성을 제시한다.

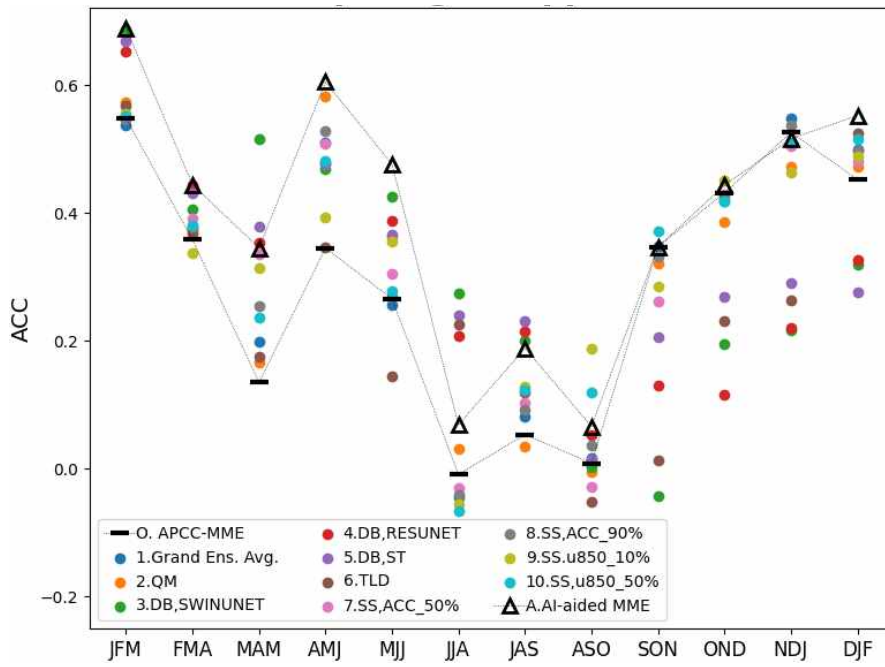


Figure 3.128. Performance of AI-aided MME for EA rainfall: ACCs of individual post-processing model outputs (color circles), their simple composite mean (open black triangle), and the original APCC deterministic MME (black horizontal line).

4. 요약 및 결론

계절 내 및 계절 기후 예측은 농업, 에너지, 재난 관리와 같은 다양한 분야에 중요한 영향을 미치지만, 예측 정확도는 여전히 제한적이다. 본 연구에서는 이러한 예측 정확도의 한계를 극복하기 위해 딥러닝 기반의 모델링 접근법, 데이터 전처리 기술, 준지도 학습 기법, 데이터 증강 기법, 설명 가능한 인공지능(XAI, eXplainable Artificial Intelligence) 기술, 앙상블 기법 등을 활용하여 기후 분야의 주요 변수(기온, 강수)뿐만 아니라 MJO 지수 등의 예측성을 향상시키기 위해 연구를 수행하였다.

본 연구는 동아시아 태평양 지역의 3~4주 리드타임 계절 내 예측 자료의 정확도를 개선하기 위해 딥러닝 모델을 활용하고, 자료 확장 및 민감도 분석을 수행하였다. 관측자료와 기간 확장을 통해 부족한 데이터를 보완하고, ECMWF, GloSea5 예측모델과 GPM 위성자료, ERA5 재분석 자료를 활용하여 훈련셋을 구축하였다. 딥러닝 모델로는 ConvLSTM과 U-Net 아키텍처를 기반으로 학습을 진행하였으며, 그 결과 ConvLSTM은 데이터 부족 문제로 성능 개선에 한계가 있었으나 U-Net 모델에서는 성능이 개선되는 것을 확인하였다. 이를 바탕으로 GUI 기반의 입출력 시스템을 개발하여 사용자가 입력자료를 쉽게 생성하고 민감도 분석을 수행할 수 있도록 하였다.

민감도 분석은 Filter, Wrapper, Embedded 기법을 활용하여 변수 선택의 효과를 평가하였다. 분석 결과, ECMWF 단독 최고온도 변수가 가장 높은 정확도를 보였으나 변수 선택만으로는 예측 성능을 크게 개선하기 어려운 한계가 있었다. 또한 U-Net 모델에 어텐션 메커니즘을 추가한 Attention U-Net 모델을 구성하여 입력자료의 중요한 부분에 가중치를 부여함으로써 모델의 정확도를 더욱 향상시킬 수 있었다.

본 연구는 동아시아 태평양 지역의 계절 내 예측 정확도를 높이기 위한 사례로서, 관측자료 확장과 전처리를 통해 훈련셋의 품질을 높이고 딥러닝 모델의 성능을 비교하였다. 특히 Attention U-Net 모델은 어텐션 메커니즘을 통해 중요한 정보에 집중하여 학습 성능을 향상시켰다. 또한 GUI 기반 입출력 시스템은 사용자가 쉽게 딥러닝 모델 입력자료를 생성하고 민감도 분석을 수행할 수 있도록 지원하며, 최적화된 인공지능 기법 개발에 기여할 수 있는 가능성을 보여주었다.

결론적으로, 본 연구는 동아시아 태평양 지역의 계절 내 예측 정확도를 향상시키기 위한 기초 자료로 활용될 수 있으며, 향후 리드타임 3~4주 기간에 적합한 딥러닝 모델의 추가적인 개선과 데이터 부족 문제 해결 방안을 모색해야 할 필요가 있다.

S2S 훈련자료에 적절한 전처리 기법을 적용하면 예측 성능이 개선됨을 Chung et al.(2023)에서 확인하였다. 특히 일 강수의 경우 Standard와 Robust 기법을 통해 모든 선행예측 시간 규모(1주-4주)에서 예측 성능이 향상되었다. 일 최고 및 최저기온은 3-4주 예측 성능이 개선되었지만, 1주 예측 개선은 여전히 미흡하였다. 차원 감소를 위한 피쳐 선택 기법은 일 강수 예측 성능에 큰 영향을 미치지 않았다. 그러나 일 최고 및 최저기온의 경우, 입력 변수가 기준 이하(예: 7개 이하)로 줄어들면 예측 성능이 저하될 가능성이 있었다. 특히 기온의 1주 예측 개선에

는 피쳐 선택 기법의 효과가 제한적이었다. MME 기반 S2S 훈련자료를 구축하고 이를 딥러닝으로 보정하여 선행예측시간 2-4주의 예측 성능을 향상시켰다. 이는 MME 기법이 S2S 예측 성능 개선에 효과적임을 입증한다. 기온의 1주 예측 성능 개선에는 전처리 기법만으로는 한계가 존재했으며, 이를 극복하기 위해 새로운 접근 방식이 필요하다(APEC기후센터, 2022). 예를 들어, 딥러닝 모델 구조 최적화 및 파라미터 조정 등이 필요했다. 따라서 기존 U-Net에 Attention 및 Residual 장치를 추가하여 다양한 확장 U-Net 아키텍처를 구축하고, 이를 S2S 기후변수(최고기온 및 강수) 예측 개선에 적용하였다. Attention U-Net, Residual U-Net, 그리고 두 장치를 결합한 Attention based-on Residual U-Net을 포함한 네 가지 아키텍처의 초매개변수를 그리드 탐색(Grid Search) 알고리즘으로 최적화한 결과, 각 변수의 특성에 따라 예측 성능이 향상됨을 확인하였다. 최고기온 예측에서는 Residual 장치가 포함된 아키텍처가 연속적인 기온 특성에 적합하여 성능이 개선되었고, 특히 Attention과 Residual 장치를 결합한 모델에서 ACC 향상이 높았다(APEC기후센터, 2023). 반면, 강수 예측에서는 지역적이고 불연속적인 특성을 반영한 Attention U-Net에서 높은 ACC 개선을 보여, 변수 특성에 따라 적합한 아키텍처 설계의 중요성을 강조하였다. 그러나 최고기온의 단기 예측(1주)과 강수의 중기 예측(2-3주) 성능 개선은 한계가 여전히 남아 있었다. 그래서 24년 연구는 앙상블 기법(보팅, 배킹, 부스팅, 스택킹)을 활용하여 기존 구축된 S2S 예측 모델의 성능을 개선하는 데 초점을 맞췄다. 다양한 신경망 모델(CNN, CNN-LSTM, U-Net, Attention U-Net, Residual U-Net)을 멤버 모델로 구성하고, 각 앙상블 기법의 특성과 차이를 비교 분석하여 모델의 다변성과 개선이 미흡한 선행예측 시간 규모의 예측 성능을 개선하고자 했다. 이를 통해 기존 단일 모델 접근법 대비 향상된 예측 정확도와 안정성을 달성했다. 특히, 배킹은 과적합을 방지하고 안정성을 높이는 데, 부스팅은 소규모 데이터에서도 강력한 성능을 발휘하는데, 스택킹은 메타 모델 기반 최적화를 통해 예측 성능을 개선하는데 기여했다. 결과적으로, 기존의 단일 모델 기반 접근법 대비 개선된 예측 성능과 안정성을 제공하였다. 그러나 딥러닝 특성상 평균값에 가까워지도록 훈련 및 예측하기 때문에 이상치 즉 이상고온에 대한 훈련 및 예측이 어렵다. 이에 7단계 후속 연구에서는 S2S 예측 기후자료에서 이상값 훈련을 위한 딥러닝 아키텍처를 개발 혹은 기존 연구에서 개발한 기준 모델을 수정하여 이상고온 예측 가능성을 탐색함으로써 S2S 예측의 실질적 활용도를(예, 선행예측 시간 규모 3-4주에서 예측성이 개선된 이상고온 자료를 활용한 작황 예측 등) 높이는 데 기여할 수 있을 것으로 기대된다.

본 연구 과제에서는 먼저 준지도학습 기법을 적용하여 MJO 위상을 분류하는 이미지 기반 인공지능 모델을 개발하였다. 준지도학습 기법을 적용하기 전 지도학습 기반 MJO 위상을 분류하는 이미지 기반 인공지능 모델 개발을 위해 MJO 인덱스 종류에 따른 입력 변수 조합, 인공지능 모델 구조, 훈련 자료 분리 방법 등을 테스트하여 MJO 위상을 분류하는 최적의 인공지능 모델을 개발하고 준지도학습 환경에서 MJO 위상 분류 인공지능 모델을 개발하였다. MJO 인덱스 종류에 따른 입력 변수 조합, 라벨링된 자료의 개수 차이, 모델 학습 방법 등에 따라 준지도학습 기반 모델의 민감도를 분석하였고 지도학습 기반의 모델과 비교 평가하였다. 준지도학습의 적용은 MJO 위상 분류 문제에 유용한 것으로 확인되었다. 라벨링된 자료의 양을 줄인 후에도 MJO 위상 분류에 중요한 기후 패턴이나 특징들을 효과적으로 학습하고 더 많은 라벨링을 사용하거나 모든 라벨링된 자료를 사용하는 지도학습의 결과에 비해서 비슷하거나 더 높은 정확도를 보이는 사례도 확인되었다. VPM 인덱스 기반 결과에서 여름철 기간에 에러가 증가하

는 경향을 보였는데 이는 MJO 인덱스의 종류에 따라서 MJO 위상 분류에 여러 기후 변동성을 학습하는데 차이가 있는 것을 확인하였고 이는 딥러닝 기반 MJO 관련 연구를 수행할 때 이러한 요소들을 고려하는 것이 필요함을 시사한다. 준지도학습 알고리즘 내 여러 자료 증강 방법에 따른 민감도 실험에서 수평 이동하는 MJO의 특성에 따라 수평적으로 이미지를 반전하는 자료 증강 방법은 오히려 정확도를 낮추는 결과를 보였고 한편 수직 반전 방법은 성능 향상에 도움이 됐다. 입력 자료의 이미지를 변형하는 다양한 증강 방법들은 공간적인 특징이 중요한 기후 자료에 적용할 때 목적변수의 특성에 맞게 적용될 필요가 있으며 가용한 자료가 제한적일 때 유용하게 활용될 수 있을 것이다. 연구 결과는 MJO 위상을 분류하는 데 있어 더 적은 이미지 자료만을 활용하는 준지도학습 기술의 적용 가능성을 보여주었다. 이는 자료의 양이 제한적인 기후 분야의 다양한 연구에 적합한 기술이 될 수 있을 것이다.

준지도학습 기반 MJO 위상 분류 연구는 이후 선행시간별 MJO RMM 지수를 예측하는 인공지능 모델 개발로 확장하였다. ResNet-18 모형을 기반으로 다양한 아키텍처와 변수 조합들을 테스트하여 수정된 모형을 구축하였다. 최종적으로 U200, U850, OLR, VP, TS, TCWV, hadvect의 기후 자료를 기반으로 기후 변수들의 편차와 배경장을 함께 입력 변수로 활용하여 선행시간별 RMM1과 RMM2 지수를 예측하는 인공지능 모델을 개발하였다. 결과적으로, 이 모델은 BCOR 0.5 기준 최대 대략 22일까지의 RMM 예측 성능을 보여주었다. 본 연구의 주요 기여는 두 가지이다. (1) 편차 정보뿐만 아니라 배경장 변수를 통합하여 모델 성능을 더욱 개선했으며, 이를 통해 MJO의 특성을 더 잘 포착할 수 있게 되어 예측성 향상에 기여했다. (2) MJO의 물리적 특성, 특히 MJO 진폭을 맞춤형 손실함수에 추가하여 MJO 변동성을 더 잘 표현하고 진폭 관련 오류를 상당히 줄였다. MJO 진폭을 손실함수에 포함시키는 방법은 인공지능 기반 MJO 연구에서 도메인에 특화된 모델의 조정이 중요함을 보여주며, 이는 MJO 예측에 있어 주요 진폭 감소 문제를 해결하는 새로운 접근 방식을 제시하였다. 또한, 연구에서는 RMM 예측에 대한 모델의 의사결정을 해석하기 위해 해석가능한 인공지능 기법을 적용하였다. 이전 연구들이 수증기의 중요성에 초점을 맞춘 것과 일치하는 결과를 보였으며, 이 연구는 수증기 수송과 함께 바람장의 배경 상태가 MJO 예측에 미치는 영향을 추가적으로 분석하였다. 기상 역학에 따른 배경 상태 정보와 수분 편차를 결합하는 것이 MJO 예측에 중요한 역할을 한다는 점은 향후 MJO 연구의 더욱 깊이 있는 이해와 개선에 중요한 기여를 할 수 있다.

개발된 선행시간별 MJO RMM 지수의 단정예측 인공지능 모델을 기반으로 다양한 섭동 적용 방법을 통해서 확률예측 인공지능 모델을 개발하였다. 초기장 섭동 방법으로 우선 ERA5에서 제공하는 10개의 앙상블 자료를 활용하였고, 기존의 인공지능 기반 기후 연구에서 많이 활용된 Perlin noise 방법을 적용하여 50개의 앙상블 확률예측 결과를 테스트하였다. 모델 파라미터 섭동은 MC dropout 방법을 적용하여 50개의 앙상블 결과를 산출하였다. 전체적으로 Perlin noise 방법이 선행시간이 길어질수록 앙상블 분산이 커지는 것을 확인할 수 있었다. 하지만 몇 가지 MJO 사례 분석을 통해서 모든 경우에 앙상블 분산이 커지지 않는 것을 확인하였다. 이는 인공지능 모델은 오차를 최소화하는 방향으로 학습되기 때문에 대개 평균에 수렴하는 특성을 가짐으로 초기장의 약간의 분포 차이나 모델 파라미터 섭동 방식으로 확실하게 앙상블 분산을 생성하기 어려울 수 있다.

향후 연구에서는 복잡하지만 성공적인 모델 결과를 산출하는 것으로 보고되는 생성형 인

공지능 모델 등의 모델로 확장할 수 있다. 생성형 모델은 이미 여러 기후 연구에서 성공적으로 적용된 바가 있고, 이번 과제에서 발굴된 배경장 정보 활용의 중요성을 고려하여 편차뿐만 아니라 배경장 자료를 활용하여 다양한 최신 생성형 모델에 통합 활용하는 방안을 더 탐구할 수 있을 것이다. 또한, MJO의 공간 패턴을 더욱 정밀하게 탐지할 수 있는 고해상도 자료를 활용하여 MJO의 위상과 진폭을 더 정확하게 예측할 수 있다. 현재의 인공지능 기반 MJO 예측 모델은 Hindcast 기간의 재분석 데이터를 기반으로 하고 있으므로, 향후 실시간으로 활용될 수 있게 하기 위해서는 예측 시 자료의 가용성을 고려한 모델의 수정과 개선이 필요할 것이다.

본 연구에서는 동아시아지역 3-6주 기온확률예측을 위해 기준 1개월 기온 예측 모델, 입력 기후 변수에 따른 기온 예측 모델, 최신 딥러닝 아키텍처를 적용한 기온 예측 모델, 앙상블 학습 기법을 이용한 기온 예측 모델, N 스텝 예측 방법을 이용한 기온 예측 모델, 재학습 방법을 이용한 기온 예측 모델 등을 구축하였다. 1개월 기온 예측 모델은 U-NET 아키텍처를 기반으로 어텐션 메커니즘 등 딥러닝 기법을 이용하여 기준 모델을 구축하였다. 또한, 합성곱 레이어의 필터 수 조정, 빅데이터 관점에서 학습자료양에 따른 예측 모델의 정확도, 모델 내부 구조 분석을 통해 모델의 정확도 개선 등의 실험을 통해 모델의 정확도 개선을 도모하였다. 학습자료는 ERA5 재분석자료를 이용하여 구축하였으며, T2M 변수 하나만을 이용하여 14일 입력 자료로 미래 1일을 예측하도록 설계하였다. 이 모델을 바탕으로 모델의 예측자료를 입력자료로 활용하는 순환 예측(rolling prediction) 기법을 이용하여 ECMWF S2S 모델의 1-2주 forecast와 hindcast 예측자료를 입력자료로 하는 3-6주 기온확률 예측 체계를 구축하였다. 기온확률예측 결과는 KMA ASOS 관측자료로 계산한 남한지역의 3분위 확률값을 ECMWF S2S 예측자료, 딥러닝 모델의 예측자료와 각각 HSS (Heidke Skill Score) 기법을 이용하여 분석하고 비교하였다.

그리고 ALO (atmosphere-land surface-ocean) 시스템 관점에서의 계절 내에서 계절로 넘어가는 기간의 예측성을 높이기 위해 대기자료(T2M) 외에 지면자료(NDVI), 해양자료(SST) 등을 입력 변수로 활용하는 기온 예측 모델을 구축하였다. 또한, 최신 딥러닝 아키텍처를 이용한 기온 예측 모델, 앙상블 학습 기법을 이용한 기온 예측 모델, N 스텝 예측 방법을 이용한 기온 예측 모델, 재학습 방법을 이용한 기온 예측 모델 등을 구축하고 예측성 평가를 통해 예측성이 낮은 모델의 개선 가능성을 평가하였다.

모델의 입력변수 중 NDVI 자료는 T2M과 SST 변수는 달리 1982~2019년 자료만 구축된 데이터셋을 이용하였기 때문에 모델의 평가기간을 2018~2019년과 2018~2022년으로 나누어 수행하였다. 평가기간 2018~2022년의 경우, NDVI 자료는 2020년 이후가 없기 때문에 2019년 자료를 이용하였다. 평가 결과, 2018~2019년은 1위 DL-SE-7+SST+NDVI, 공동 2위 DL-SE+FLT2+BTLNCK-7, NM-SE-7, EM-NP-14, nP-DL-SE+BIGDAA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI, RL-DL-SE-14+SST+NDVI였으며, ECMWF는 25위였다. 2018~2022년은 전체 평가 기간에서 총 32개 모델(ECMWF 포함) 중 1위 DL-SE-14, 2위 DL-SE+BIGDATA+FLT2+BTLNCK-14, 3위 DL-SE+BIGDATA-14였으며, ECMWF는 18위였다. 엘니뇨 발생 시기는 공동 1위 DL-SE-7+SST+NDVI, DL-SE+FLT2+BTLNCK-7. 3위 DL-SE+BIGDATA-14이며 ECMWF 는 공동 7위였다. 2018~2019년은 ALO 시스템을 기초로 입력 자료 확장을 고려한 모델들이 주로 좋은 예측성을 보였다. 이는 지면자료(NDVI)와 해양자료(SST) 자료가 4~6주 예측에 긍정적인 영향을 준 것으로 판단된다. 반면 2018~2022년은 NDVI 자

료 구축 기간이 2018~2019년으로 2020년부터는 2019년 자료를 이용하였기 때문에 T2M 자료만 이용한 모델에 비해 상대적으로 예측성이 떨어지는 경향을 보였다. ALO 시스템 관점에서 계절 내 및 계절 기온 예측에서 SST와 NDVI 자료가 모델의 예측성에 긍정적인 영향을 미치는 것으로 나타났지만, 현업에 모델을 활용하기 위해서는 추가적인 자료 구축과 함께 예측성을 평가하는 후속 연구가 필요할 것으로 판단된다.

단순한 인공지능모델에 간결한 자료강화기법을 적용하여도 월별 기후예측을 개선할 수 있을지 가능성을 타진하고자 하였다. 여름철 평균기온 예측을 위한 합성곱신경망모델(CNN)을 구축하여 Semi-monthly와 Cutmix 자료강화의 개선 성능을 비교하였으며 모든 월을 사용하는 Month-Agnostic 접근과 목표 월만을 사용하는 Target Month Only 접근의 결과 또한 비교하였다. Semi-monthly 방법을 적용한 경우 자료강화하지 않은 경우와 같이 epoch에 따라 모델의 검증 손실이 지속적으로 증가하여 자료강화 방법으로 적합하지 않음을 확인하였으며, 훈련 손실을 살펴보면 충분히 학습되지 않은 상태에서 검증 손실 정체/증가로 학습이 중단되어 모델이 과소적합되었음을 확인할 수 있었다. Cutmix 방법을 적용한 경우 Month-Agnostic으로 모델 훈련을 하였을 때, 검증 손실이 epoch에 따라 그다지 증가하지 않음을 볼 수 있었으며 훈련이 충분히 이루어지기 전에 학습이 중단되는 상황도 관찰되지 않고 모델의 검증 정확도가 개선되었다. 그러나 Target month만을 사용한 경우는 과소적합이 발생하였으며 모델의 검증 정확도 또한 개선되지 않았다.

Month-Agnostic 접근을 활용하는 CNN 모델에 대해 여름철과 겨울철, 평균기온과 강수(SPI1) 예측에 대한 Cutmix 자료강화의 개선 성능을 다시금 평가하였다. 자료강화한 경우 훈련 과정에서의 검증 정확도는 여름철, 겨울철 그리고 평균기온 및 강수(SPI1)에 대해 모두 증가하였으며, Monte Carlo Dropout 적용하여 살펴본 결과 최종 예측 결과의 epistemic 불확실성은 감소하였다. 그러나 월별로 보았을 때 평균기온은 1월(LT2), 2월(LT3), 8월(LT3), 그리고 강수(SPI1)는 7월(LT2), 8월(LT3), 12월(LT1)의 경우에만 개선이 관찰되었다. Confusion matrix를 통해 개별 예측 결과를 전체적으로 검토하였는데, 자료강화한 경우 대부분의 경우 적중률이 증가하였음을 확인하였다. 결론적으로 평균기온이나 강수(SPI1) 예측 부진의 주요 원인이 자료 부족이더라도 지나치게 단순한 모델로는 자료강화의 효과가 제한적일 수 있으나 전반적으로 개선의 효과를 확인할 수 있었다. 활용할 수 있는 최대한의 모델 아키텍처를 고려한 SOTA 모델 활용과 그에 적합한 자료강화기법의 적용을 수반한 후속 연구가 필요하다.

센터는 동아시아 지역에 대해 맞춤형 다중모델앙상블(Multi-Model Ensemble, MME) 예측 서비스를 수행하고 있으나, 동아시아 계절 강수에 대한 역학 모형 예측 성능은 기대에 미치지 못한다. 특히, 집중 호우 등 재난 상황이 빈번한 여름철 성능이 현저히 낮는데 이는 MME 과정에서 국지성을 띄는 강수 현상을 재현할 만한 공간 해상도를 잃어버린다는 점이 문제로 지적될 수 있고, 따라서 역학 모형 자료 후처리를 통한 강수 예측 개선은 반드시 필요하다. 이를 위해 본 연구는 단기 날씨 예측뿐만 아니라 전지구 순환 모델에서 우수성을 증명하고 있는 인공지능 기술의 역학 예측 후처리에 대한 적용성을 탐색해 보고자 하였다.

센터가 본격적으로 MME 서비스를 시작한 2013년부터 축적된 3만 개가 넘는 real-time forecast 앙상블 자료들의 활용성을 높이고 절대적으로 부족한 계절 시간 규모 학습 샘플을 극

복하기 위해서, 양질의 앙상블만을 선별하는 방식의 인공지능 기술 적용 가능성을 탐색하게 되었다. 앙상블의 성능을 선제적으로 추정하는 것은 이제껏 시도되지 않은 매우 도전적인 문제로, 이는 과거 관측에서 앙상블 예측으로 이어지는 입력장의 시공간 진화가 강수 역학 예측의 오차/스킬과 유의한 관련성을 가진다는 가정과, 인공지능이 수많은 자료 안에 숨어있는 복잡한 관련성 정보들을 뽑아낼 수 있다는 기대를 기반으로 추진되었다. 특히, 물리 변수가 아닌 예측 스킬(ACC)를 목적 변수로 두고 이를 추정하는 딥러닝 모델을 설계했다는 점이 이 연구의 독창성/차별점이라 할 수 있다. 또한 XAI 기법 적용을 통해 활성 변수 및 지역을 확인해 유의미한 예측 성능 원천을 파악할 수 있었고, 이를 기반으로 동아시아 하층 바람 재현성 기반 앙상블 선별 모델을 추가적으로 설계할 수 있었다. 앙상블 선별에 더해, 현업 모델 앙상블 후처리 방식을 오차 제거와 전이 학습 기반 진단 등으로 다양화하고, 입력장 및 초매개변수 변형 실험을 통해 여러 버전의 인공지능 모델을 개발하였다. 또한, 유의한 성능을 나타내는 다수의 인공지능 후처리 모델을 종합하는 방식으로 새로운 차원의 동아시아 다중모델앙상블 강수 예측 시스템의 원형을 제안하였다. 이를 위해 총 10개 후처리 모델에 대한 단정론/확률론적 합성 코드를 구현하고, 공간 예측장 뿐만 아니라 과거 성능 평가 정보에 대한 통합 표출 모듈을 개발하였다. 2023년 12 계절 테스트 결과, 현업 APCC-DMME(ACC=0.288) 대비 후처리 모델 SCM(ACC=0.403)이 평균 40% 정도 개선된 성능을 보였다. 기존 MME를 보완하는 형태의 역학-AI 하이브리드 방식 다중모델앙상블을 도입함으로써 예측 원천의 다양성을 확보함과 동시에 신뢰도 높은 동아시아 계절 강수 전망 제공을 가능하게 하고, 홈페이지를 통한 기후정보 서비스 콘텐츠를 확장할 수 있을 뿐만 아니라, 향후 대상 지역/목적 변수 변경 등을 통해 역학 예측 후처리를 인공지능 기술 저변을 확대해 APCC 계절 예측 역량을 강화에 기여할 것으로 기대 된다.

사사

본 연구는 APCC의 지원을 통해 수행되었습니다.

References

- 김가영, 김해정, 오지현, 유태우, 이운영, 정여민, 정유란, 정유림, 함수련 기후분석과 운순조, 2019: 다중모델 기반 계절내 예측 활용 가이드라인 개발, 2019 아태기후센터 연구보고서.
- APEC기후센터, 2020: 2020년도 연차보고서: 예측성 향상을 위한 다중모델기반 기후예측기술 개발. p80-107.
- APEC기후센터, 2022: 2022년도 연차보고서: 객관적 기후예측에 적합한 인공지능활용 기반기술 개발(I). p23-33.
- APEC기후센터, 2023: 2023년도 연차보고서: 객관적 기후예측에 적합한 인공지능활용 기반기술 개발(II). p19-32.
- NIA, 2018: EU의 인공지능 新 규제메카니즘: ‘설명가능 인공지능(XAD)’ , NIA Special Report 2018-3, 2018.3.15.
- Ahn, J. B., J. L. Lee, and S. R. Jo, 2018: Evaluation of PNU CGCM Ensemble Forecast System for boreal winter temperature over South Korea. *Atmosphere*, **28**, (4), 509-520.
- Bahdanau, D., K. Cho, and Y. Bengio, 2016: Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473, <https://doi.org/10.48550/arXiv.1409.0473>.
- Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, **619**, 533-538 (2023). <https://doi.org/10.1038/s41586-023-06185-3>.
- Bonev, B., T. Kurth, C. Hundt, J. Pathak, M. Baust, K. Kashinath, A. Anandkumar, 2023: Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere. <https://doi.org/10.48550/arXiv.2306.03838>.
- Chae, S., S. Jang, S. Lee, and D. Lee, 2020: Complex System Analysis of Korean Peninsula Earthquake Data. *Scientific Reports*, **10**, 2686, <https://doi.org/10.1038/s41598-020-59619-7>.
- Chae, S., J. Shin, S. Kwon, S. Lee, S. Kang, and D. Lee, 2021: PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. *Scientific Reports*, **11**, 11952, <https://doi.org/10.1038/s41598-021-91253-9>.
- Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., 2023: Fuxi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, **6**, 190. <https://doi.org/10.1038/s41612-023-00512-1>.
- Chen, L., Zhong, X., Li, H., Wu, J., Lu, B., Chen, D., Xie, S.P., Wu, L., Chao, Q., Lin, C., Hu, Z., and Qi, Y., 2024: A machine learning model that outperforms conventional global subseasonal forecast models. *Nature Communications*, **15(1)**, 6425.

- Chung, U., J. Rhee, M. Kim, S.-J. Sohn, 2023: Investigating data preprocessing algorithms of a deep learning postprocessing model for improvement of sub-seasonal to seasonal climate predictions. *Korean Journal of Agricultural and Forest Meteorology*, **25**, 80–98, <https://doi.org/10.5532/KJAFM.2023.25.2.80>.
- Dasgupta, P., Metya, A., Naidu, C. V., Singh, M., and Roxy, M. K. 2020: Exploring the long-term changes in the Madden Julian Oscillation using machine learning. *Scientific reports*, **10(1)**, 18567.
- Dawson, A., 2016: Windspharm: A high-level library for global wind field computations using spherical harmonics. *Journal of Open Research Software*, **4(1)**, <https://doi.org/10.5334/jors.129>.
- de Andrade, F. M., C. A. S. Coelho, and I. F. A. Cavalcanti, 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>.
- Delaunay, A., and Christensen, H. M. 2022: Interpretable deep learning for probabilistic MJO prediction. *Geophysical Research Letters*, **49(16)**, e2022GL098566.
- Diakogiannis, F. I., F. Waldner, P. Caccetta, and C. Wu, 2020: ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **162**, 94–114. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>.
- Hall, M. 1999: Correlation-based Feature Selection for Machine Learning (PhD thesis). University of Waikato
- Ham, Y.-G., J.-S. Kug, G.-H. Lim, J. Park, and D. Kim, 2019: Correction to: A newly developed APCC SCoPS and its prediction of East Asia seasonal climate variability. *Climate Dynamics*, **53**, 3703–3704. <https://doi.org/10.1007/s00382-019-04894-y>.
- He, C., and T. Zhou, 2015: Responses of the Western North Pacific Subtropical High to Global Warming under RCP4.5 and RCP8.5 Scenarios Projected by 33 CMIP5 Models: The Dominance of Tropical Indian Ocean?Tropical Western Pacific SST Gradient. *Journal of Climate*, **28(1)**, 365–380. <https://doi.org/10.1175/JCLI-D-13-00494.1>.
- He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). *IEEE*. <https://doi.org/10.1109/CVPR.2016.90>.
- He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, 770–778, <https://doi.org/10.1109/CVPR.2016.90>.

- Hersbach, H., B. Bell, P. Berrisford, P. Dahlgren, A. Horanyi, J. Muñoz-Sabater, J. Nicolas, R. Radu, D. Schepers, A. Simmons, and C. Soci, 2020: The ERA5 global reanalysis: Achieving a detailed record of the climate and weather for the past 70 years. *Quarterly Journal of the Royal Meteorological Society*, **146(727)**, 1999–2049. <https://doi.org/10.1002/qj.3803>.
- Hinton, G. E., S. Osindero, and Y.-W. Teh, 2006: A fast learning algorithm for deep belief nets. *Neural Computation*, **18**, 1527–1554, <https://doi.org/10.1162/neco.2006.18.7.1527>.
- Hu, J., L. Shen, S. Albanie, G. Sun, E.H., Wu, 2020: Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42(8)**, 2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
- Ju, H.-J., J.-Y. CHAE, E.-J. LEE, Y.-T. KIM, and J.-H. Park, 2022: Impacts of Seasonal and Interannual Variabilities of Sea Surface Temperature on its Short-term Deep-learning Prediction Model Around the Southern Coast of Korea. *The Sea. Journal of the Korean Society of oceanography*, **27**, 49–70. <https://doi.org/10.7850/JKSO.2022.27.2.049>
- Kim, D., Heo, B., and Han, D., 2024: DenseNets Reloaded: Paradigm Shift Beyond ResNets and ViTs. arXiv:2403.19588v2. <https://doi.org/10.48550/arXiv.2403.19588>
- Kim, D.-H., G.-H. Han, E.-C. Park, J.-H. Jeong, and K.-E. Kim, 2021: The KMA Global Seasonal Forecasting System (GloSea6) ? Part 1: Operational system and improvements. (In Korean).
- Kim, H., Ham, Y. G., Joo, Y. S., and Son, S. W. 2021: Deep learning for bias correction of MJO prediction. *Nature Communications*, **12(1)**, 3087.
- Kim, H., Richter, J. H., and Martin, Z. 2019: Insignificant QBO-MJO prediction skill relationship in the SubX and S2S subseasonal reforecasts. *Journal of Geophysical Research: Atmospheres*, **124(23)**, 12655–12666.
- Kingma, D. P., and Ba, J., 2014: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, <https://doi.org/10.48550/arXiv.1412.6980>.
- Laine, S., and T. Aila, 2016: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242, <https://doi.org/10.48550/arXiv.1610.02242>.
- Lam, R., A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, and P. Battaglia, 2023: Learning skillful medium-range global weather forecasting. *Science*. <https://doi.org/10.1126/science.adi2336>.
- Liebmann, B., and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society*, **77(6)**,

1275–1277.

- Lim, Y., Son, S. W., and Kim, D. 2018: MJO prediction skill of the subseasonal-to-seasonal prediction models. *Journal of Climate*, **31(10)**, 4075–4094.
- Liu, Y., C. Gong, L. Yang, and Y. Chen, 2019: DSTP-RNN: a dual-stage two-phase attention-based recurrent neural networks for long-term and multivariate time series prediction. arXiv:1904.07464. <https://doi.org/10.48550/arXiv.1904.07464>.
- Maclachlan, C., and Coauthors, 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>.
- Mariotti, A., Ruti, P. M., and Rixen, M., 2018: Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate and Atmospheric Science*, **1(4)**, 1–4. <https://doi.org/10.1038/s41612-018-0014-z>
- Marshall, Andrew G., Harry H. Hendon, and Debra Hudson. 2016: Visualizing and verifying probabilistic forecasts of the Madden-Julian Oscillation. *Geophysical Research Letters*, **43.23**, 12–278.
- Martin, Z. K., E. A. Barnes, and E. Maloney, 2022: Using Simple, Explainable Neural Networks to Predict the Madden-Julian Oscillation. *Journal of Advances in Modeling Earth Systems*, **14(5)**, <https://doi.org/10.1029/2021MS002774>.
- Oktay, O., and Coauthors, 2018: Attention U-Net: Learning Where to Look for the Pancreas. arXiv:1804.03999, <https://doi.org/10.48550/arXiv.1804.03999>.
- Oktay, O., J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, 2018: Attention U-Net: Learning where to look for the pancreas. In *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2018* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-76578-7_28.
- Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert. 2023: Attention U-Net: Learning Where to Look for the Pancreas. <https://doi.org/10.48550/arXiv.1804.03999>
- Paek, H., J.-Y. Yu, and Y.-H. Ryu, 2015: A source of AGCM bias in simulating the Western Pacific Subtropical High: Different sensitivities to the two types of ENSO. *Monthly Weather Review*, **143**, 2348–2362.
- Parente, L., and Hengl, T., 2020: Monthly Global NDVI at 5 km based on MODIS and AVHRR

products - 1982 to 2019 (v1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4305975>

- Pathak, J., S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, P. Hassanzadeh, K. Kashinath, and A. Anandkumar, 2022: FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214. <https://doi.org/10.48550/arXiv.2202.11214>.
- Pham, H., Z. Dai, Q. Xie, and Q. V. Le, 2021: Meta pseudo labels. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 11557–11568.
- Price I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P. Lam, R., Willson, M., 2024: Probabilistic weather forecasting with machine learning, *Nature*, <https://doi.org/10.1038/s41586-024-08252-9>
- Qin, Y., D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, 2017: A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. arXiv:1704.02971, <https://doi.org/10.48550/arXiv.1704.02971>.
- Reed, S., and Coauthors, 2022: A Generalist Agent. arXiv:2205.06175, <https://doi.org/10.48550/arXiv.2205.06175>.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, doi:10.1175/1520-0442(2002)015,1609:AIISAS.2.0.CO;2.
- Rodgers, K. B., S.-S. Lee, N. Rosenbloom, A. Timmermann, G. Danabasoglu, C. Deser, J. Edwards, J.-E. Kim, I. R. Simpson, K. Stein, M. F. Stuecker, R. Yamaguchi, T. B?dai, E.-S. Chung, L. Huang, W. M. Kim, J.-F. Lamarque, D. L. Lombardozzi, W. R. Wieder, and S. G. Yeager, 2021: Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, **12(4)**, 1393–1411. <https://doi.org/10.5194/esd-12-1393-2021>.
- Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015 (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning Internal Representations by Error Propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, MIT Press, 318–362.
- Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, **41**, 2514–2519, <https://doi.org/10.1002/2014GL059637>.

- Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, 2019: Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, **128(3)**, 336–359. <https://doi.org/10.1007/s11263-019-01228-7>.
- Shih, S. Y., F. K. Sun, and H. yi Lee, 2019: Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, **108**, 1421–1441, <https://doi.org/10.1007/s10994-019-05815-0>.
- Shin, N. Y., Kim, D., Kang, D., Kim, H., and Kug, J. S. 2024: Deep learning reveals moisture as the primary predictability source of MJO. *npj Climate and Atmospheric Science*, **7(1)**, 11.
- Simonyan, K., and Zisserman, A., 2014: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, <https://doi.org/10.48550/arXiv.1409.1556>.
- Tarvainen, A., and H. Valpola, 2017: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, **30**.
- Trabelsi, I., Zeddini, B., Zolghadri, M., Barkallah, M. and Haddar, M. 2021: Obsolescence Prediction based on Joint Feature Selection and Machine Learning Techniques. DOI: 10.5220/0010241407870794
- Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K. Gupta, Aditya Grover, 2023: ClimaX: A foundation model for weather and climate. <https://doi.org/10.48550/arXiv.2301.10343>
- Van Katwyk, P., Fox-Kemper, B., Seroussi, H., Nowicki, S., and Bergen, K. J. 2023: A variational LSTM emulator of sea level contribution from the Antarctic ice sheet. *Journal of Advances in Modeling Earth Systems*, 15.12, e2023MS003899.
- Ventrice, M. J., M. C. Wheeler, H. H. Hendon, C. J. Schreck, C. D. Thorncroft, and G. N. Kiladis, 2013: A modified multivariate Madden–Julian oscillation index using velocity potential. *Monthly Weather Review*, **141(12)**, 4197–4210, <https://doi.org/10.1175/MWR-D-12-00327.1>.
- Vitart, F., and A. W. Robertson, 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>.
- Weyn, J. A., D. R. Durran, R. Caruana, and N. Cresswell-Clay, 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>.

- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly weather review*, **132(8)**, 1917–1932, [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- Woo, S., J. Park, J. Lee, I.S. Kweon, 2018: CBAM: Convolutional block attention module. In: Computer Vision – ECCV 2018. *Lecture Notes in Computer Science*, **11211**. 32–19. https://doi.org/10.1007/978-3-030-01234-2_1
- Woo, S., S. Debnath, R. Hu, X. Chen, Z. Liu, I.S. Kweon, and S. Xie, 2023: ConvNeXt V2: Co-designing and Scalling ConvNets with Masked Autoencoders, arxiv:2301.00808.
- Wu, T., et al., 2014: An overview of BCC climate system model development and application for climate change studies. *Journal of Meteorological Research*, **28(1)**, 34–56.
- Yang, S., F. Ling, Y. Li, and J.-J. Luo, 2023: Improving seasonal prediction of summer precipitation in the middle-lower reaches of the Yangtze River using a TU-Net deep learning approach. *Artificial Intelligence for the Earth Systems*, **2(2)**, 123–135. <https://doi.org/10.1175/AIES-D-22-0078.1>.
- Zeiler, M. D. 2014: Visualizing and Understanding Convolutional Networks. In European conference on computer vision/arXiv (Vol. 1311).
- Zhang, L., and W. Han, 2021: Indian Ocean Dipole leads to Atlantic Niño. *Nature Communications*, **12(1)**, 2623. <https://doi.org/10.1038/s41467-021-26223-w>.
- Zhou, D., J. Wang, J. Shen, and M. Sun, 2022: Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. arXiv preprint arXiv:2201.04676.

Appendix A. Deep learning models developed for 1-month temperature prediction

Table A.1. Summary of the ECMWF S2S Model and 1-month temperature prediction deep learning models.

CASE NAME	Pretrained models	Predictability Rank	Predictability Rank 2018-2022		
		2018-2019	Total	El Niña	La Niña
ECMWF		25	18	3	7
DL-SE-14		14	1	13	4
DL-SE+BIGDATA-14		22	3	18	3
DL-SE+BIGDATA+FLT2-14		19	9	11	18
DL-SE+BIGDATA+FLT2+BTLNCK-14		21	2	8	13
DL-CBAM-14		14	6	18	10
DL-CBAM+BIGDATA-14		25	9	24	1
DL-CBAM+BIGDATA+FLT2-14		18	8	27	21
DL-CBAM+BIGDATA+FLT2+BTLNCK-14		11	16	17	7
DL-SE-7		13	22	18	15
DL-SE-7+SST		32	27	31	12
DL-SE-7+SST+NDVI		1	16	1	32
DL-SE+FLT1+BTLNCK-7		29	28	22	21
DL-SE+FLT1+BTLNCK-7+SST		9	11	3	13
DL-SE+FLT2+BTLNCK-7		2	25	16	1
DL-SE-14+SST+NDVI		16	14	23	5
DL-CBAM-7		19	26	15	6
DL-CBAM-7+SST+NDVI		9	24	25	20
DL-CBAM+FLT1+BTLNCK-7		29	30	21	23
NM-SE-7		2	31	32	16
NM-SE-7+SST+NDVI		17	20	10	16
NM-SE-7-RELU+SST+NDVI		28	32	14	29
NM-SE-14		27	28	29	30
EM-7	DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST	8	19	11	7
EM-NP-7	DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST	7	7	1	10
EM-NP+FLT2-7	DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST	31	23	29	30
EM-TR-7	DL-SE-FLT1-BTLNCK-7+SST, DL-CBAM-7+SST, DL-SE-7+SST	24	21	26	26
EM-NP-14	DL-SE+BIGDATA-14, DL-CBAM+BIGDATA+FLT2+BTLNCK, DL-CBAM-14	2	12	6	24
nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7+SST+NDVI	DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7+SST+NDVI	2	4	5	26
nP-DL-SE+BIGDATA+FLT2+BTLNCK-14-NM-SE-7	DL-SE+BIGDATA+FLT2+BTLNCK-14, NM-SE-7	22	14	28	19
RL-DL-CBAM+BIGDATA+FLT2+BTLNCK+14	DL-CBAM+BIGDATA+FLT2+BTLNCK-14	11	4	8	28
RL-DL-SE-14+SST+NDVI	DL-SE-14+SST+NDVI	2	12	6	24

Appendix B. Examples of the results of top-ranked deep learning-based 3 - 6 week temperature probability forecast models in 2018-2018 and 2018-2022 evaluation periods.

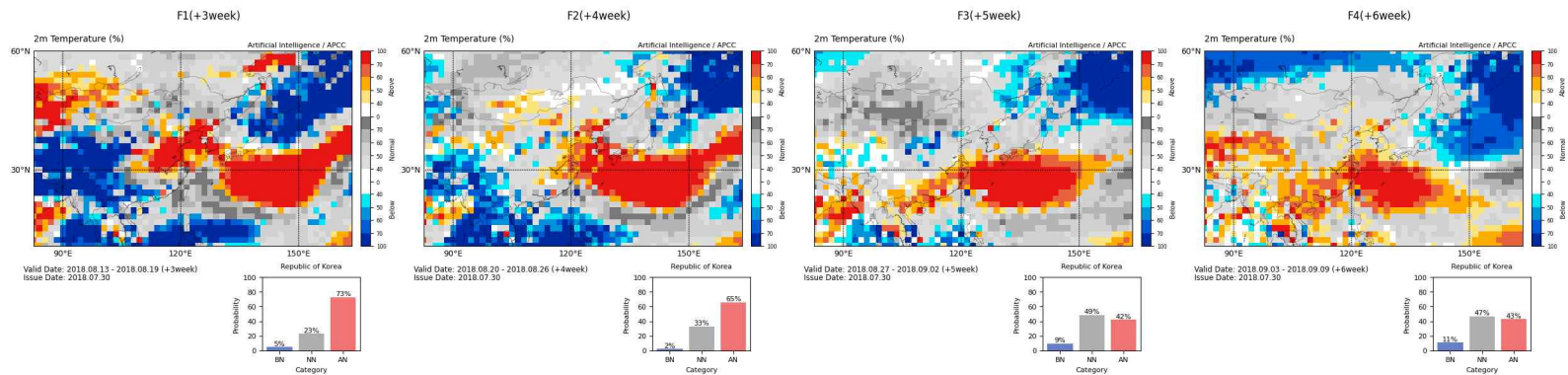


Figure B.1. Example of 3-6 week temperature probability forecast results by the top-ranked model (DL-SE-7+SST+NDVI) during the 2018-2022 evaluation period, specifically from July 30, 2018.

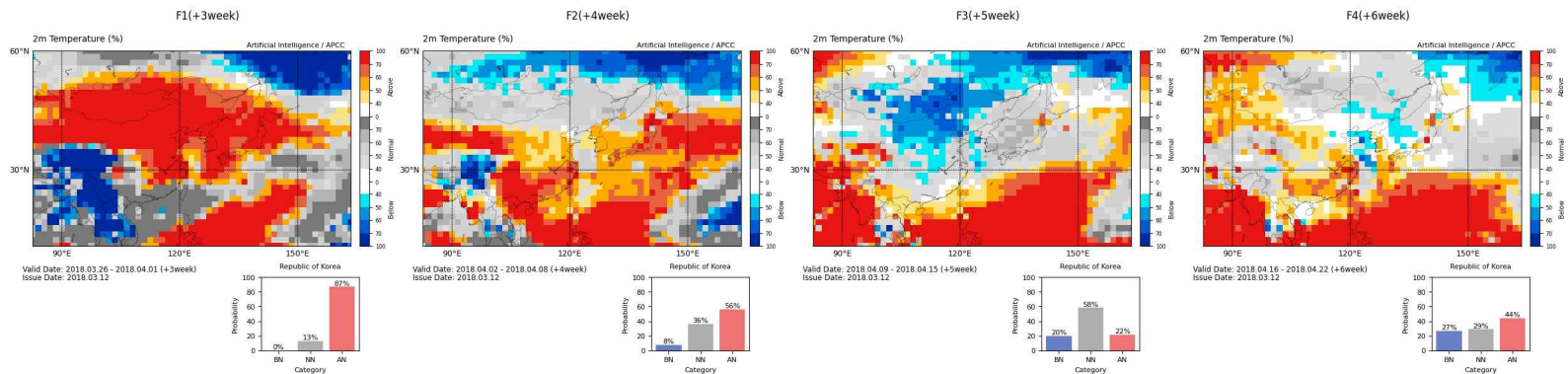


Figure B.2. Example of 3-6 week temperature probability forecast results by the top-ranked model (DL-SE-14) during the 2018-2022 evaluation period, specifically from March 12, 2018.