

# Deep learning for multi-year ENSO forecasts

밑줄친 항목: 용어사전에 있음

기울인 항목: 추가자료 있음 (트렐로)

## Abstract

Variations in the El Niño/Southern Oscillation (ENSO) are associated with a wide array of regional climate extremes and ecosystem impacts<sup>1</sup>. Robust, long-lead forecasts would therefore be valuable for managing policy responses. But despite decades of effort, forecasting ENSO events at lead times of more than one year remains problematic<sup>2</sup>. Here we show that a statistical forecast model employing a deep-learning approach produces skilful ENSO forecasts for lead times of up to one and a half years. To circumvent the limited amount of observation data, we use transfer learning to train a convolutional neural network (CNN) first on historical simulations<sup>3</sup> and subsequently on reanalysis from 1871 to 1973. During the validation period from 1984 to 2017, the all-season correlation skill of the Nino3.4 index of the CNN model is much higher than those of current state-of-the-art dynamical forecast systems. The CNN model is also better at predicting the detailed zonal distribution of sea surface temperatures, overcoming a weakness of dynamical forecast models. A heat map analysis indicates that the CNN model predicts ENSO events using physically reasonable precursors. The CNN model is thus a powerful tool for both the prediction of ENSO events and for the analysis of their associated complex mechanisms.

## ENSO: 남방진동으로 표기

### [Abstract]

엘니뇨 남방진동의 변화는 다양한 지역적 기후의 극한현상(가뭄, 홍수, 산불, 열파 등) 및 생태계에 미치는 영향과 연관이 있음. 하지만 남방진동을 1년 전에 예측하는 건 여전히 어려운 문제였는데, 우리는 여기서 딥러닝 접근법을 사용하여 1.5년 전에 예측을 해봄. 다만, 관찰 데이터가 적기에 우리는 전이 학습을 사용하여 역사적 요소들에 대해 CNN을 학습시켰고, 1871년부터 1973년까지 데이터에 대해 CNN을 활용하여 다시 분석해보았다.(즉, CNN을 학습시켜놓고 CNN을 사용해서 다시 이게 나온건지 다시 분석한듯함)

자료의 신빙성이 유효한 1984-2017년 자료 사용:

Nino3.4 인덱스를 CNN으로 분석하여 만든 거 상관관계 > 최신 다이내믹 예측 시스템이 찾은 상관관계

= 지금까지 나온 모델들보다 CNN이 분석을 더 잘함

세밀하게 여러 구역으로 나누어진 분포에 대한 바다표면 온도도 CNN이 더 예측 잘함

왜 예측을 더 잘할까?

알기 위해서 heat map 분석을 사용하여 보니, CNN모델이 남방진동을 “물리적으로 논리적인 요소들”을 사용하여 예측하는 것으로 보임.

고로 CNN 모델이 남방진동을 예측하는 데, 그리고 복잡한 메커니즘을 분석하는 데 사용하기 좋은 툴이라 생각됨

## [서론]

The ability to forecast large-scale climate variability, with its effects on global social and environmental systems, is highly dependent on the quality of ENSO predictions<sup>4</sup>. Although ENSO forecasts using atmosphere–ocean coupled models<sup>5,6</sup> generally outperform those of current statistical models<sup>1</sup>, state-of-the-art dynamical forecast systems nevertheless do not provide a skilful prediction of ENSO for lead times longer than one year. The multi-year prediction of ENSO events therefore remains a major challenge<sup>1</sup>.

However, the presence of an oscillating element in ENSO, linked to slowly varying oceanic variations and their coupling to the atmosphere, suggests that multi-year forecasts are possible<sup>5,7</sup>. Interestingly, equatorial Pacific anomalies during several La Niña events lingered for several years<sup>8</sup>. The high-frequency equatorial winds are less predictable<sup>9</sup>, but the slowly varying component of the equatorial winds coupled with underlying sea surface temperatures (SST)<sup>10</sup> is predictable to some extent. SST anomalies outside the equatorial Pacific can lead to an ENSO event with a time-lag longer than a year<sup>11,12</sup>. These studies imply that there is still room for improvement in ENSO prediction, although current methodologies might not be appropriate for this.

지구만큼 커다란 규모의 기후 변동성을 예측하기 위해서, 남방진동 예측에 굉장히 많이 의존하고 있다. 비록 이 남방진동 예측이 대기-해양이 융합된 모델을 사용하면서 기존의 통계학적인 모델을 훨씬 넘어섰지만, 최신의 다이내믹 예측 시스템은 여전히 1년 이전에 남방진동을 예측하는 것이 불가능하다. 그래서 여전히 과학계에서는 몇 년 전에 남방진동을 예측하는 것이 풀리지 않은 큰 숙제로 남아있는 상태이다.

하지만, 남방진동에서 변화하는 요소의 존재가 있는데, 이 존재가 천천히 바뀌고 있는 해양 변동성 혹은 대기변동성과 연관되어있는 것을 보면, 이는 우리가 몇 년 전에 예측하는 것 역시 가능함을 시사한다. 흥미롭게도, 적도 태평양의 변칙(anomalies, 평소와 다른 패턴이 나타남)은 몇번의 라니냐 현상이 나타날동안 몇년간 계속해서 나타나곤 했다.

High-frequency(모르겠음) 적도풍은 예측성이 더 떨어지나, 적도풍의 몇몇 천천히 변화하는 요소들은 바다 표면 온도(SST)와 결부되어 있어 어느정도는 예측이 가능하다.

(바람하나 바꾸는데도 무슨 구름의 양, 습도, 기후, 바람이 생기는 곳 표면의 온도, 기류 등 여러가지 요소가 있는데 그 중에 천천히 바뀌는 요소들)

(즉, 적도에 부는 바람 자체는 예측을 못하나, 그 바람을 일으키는 요소들 중에 바다 표면 온도와 연결된 요소들이 있음. 근데 바다 표면은 언제든지 위성으로 측정이 가능하니까 애네는 대충 이렇게 바뀌겠구나 하고 예측해볼 수 있고, 그러면 적도에 부는 바람이 일어날 확률이 몇퍼센트다~ 이렇게 두루뭉술하게 예측가능함)

적도 태평양 바깥에서의 바다 표면온도 변칙은 추후 남방진동을 일으키는데 1년 정도의 텀을 두고 연관이 있다.

With the advent of the big data era, deep learning has had a dramatic impact on many domains by discovering intricate structures within large datasets<sup>13</sup>. In particular, CNNs have produced outstanding results in the processing of data with multiple-dimensional arrays with spatial structure (for example, used in the identification of objects within colour images)<sup>14,15</sup>. Therefore, CNNs would be appropriate to reveal the links between three-dimensional predictor fields and the predictand index. Here we use a CNN-based statistical model to predict the ENSO index.

CNN은 특히 큰 데이터셋에서 복잡한 구조를 찾아내고, 공간 구조와 다차원의 배열을 처리하는데 뛰어난 결과를 보여준다. 고로 3차원의 예측 분야와 예측될 수 있는 인덱스 사이의 관계를 찾아내는데 CNN은 적합할 것이라 보인다.

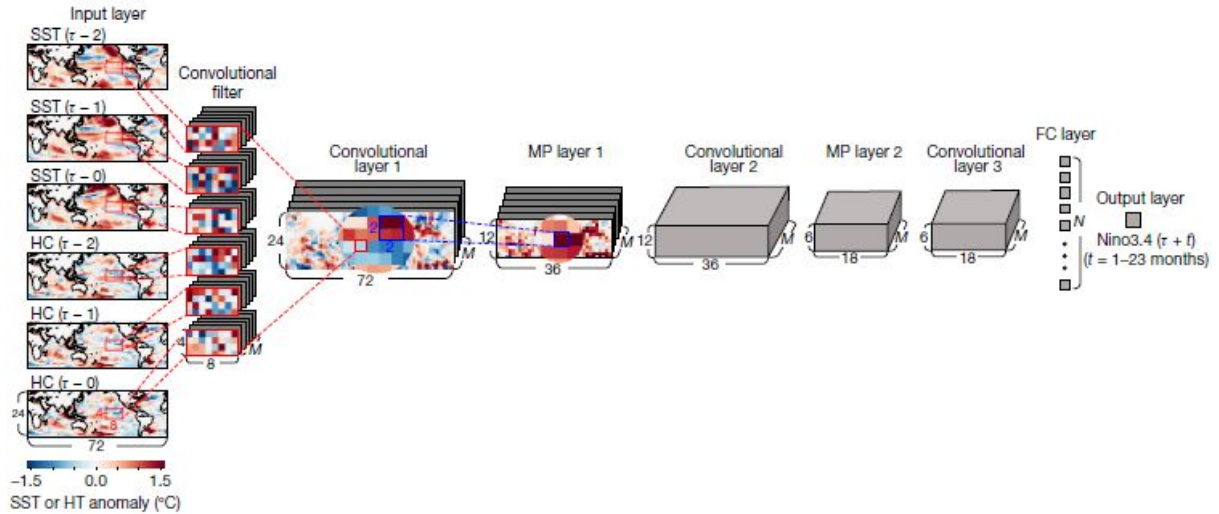
**여기서 우리는 CNN기반으로한 통계 모델을 남방진동 인덱스를 예측하기 위해 사용하였다.**

Our CNN model uses SST and heat content (vertically averaged oceanic temperature in the upper 300 m) anomaly maps over 0°–360° E, 55° S–60° N for three consecutive months as predictors, and the Nino3.4 index (area-averaged SST anomaly over 170°–120° W, 5° S–5° N) as a predictand to be predicted up to two years ahead (Fig. 1, Methods).

우리의 CNN 모델은 바다 표면 온도와 열용량(바다 표면에서 300m 들어간 온도의 평균)의 anomaly map (용어정리 사진 4 참조)을 0°–360° E, 55° S–60° N 여기 위치에 연속하는 3개월동안 예측변수(독립변수)로 표기해 보았다. 그리고 Nino3.4 인덱스 (지역-평균 바다 표면 온도 anomaly 를 2년 후를 예측할 수 있는 피예측량으로 삼았다. (anomaly 역주: 지금까지 평균적인 바다 표면 온도 및 열용량을 기준으로 지도를 만들고, 거기서 얼마나 뜨거워졌고 차가워졌는지를 이미지로 표기한 맵)

predictand: 피예측량, 예측을 해야하는 함수

이 말을 더 잘 이해하기 위해서는 Fig 1. Architecture of the CNN model used for ENSO forecasts. 참조



**Fig. 1 | Architecture of the CNN model used for the ENSO forecasts.** The CNN model consists of one input layer (the predictor), three convolutional layers, two max-pooling (MP) layers, one fully connected (FC) layer, and one output layer (the predictand). The variables of the input layer correspond to the SST (in units of  $^{\circ}\text{C}$ ) and the oceanic heat content (HC, in units of  $^{\circ}\text{C}$ ) anomaly maps from time  $\tau - 2$  months to time  $\tau$  (in months), between  $0^{\circ}$ – $360^{\circ}$  E and  $55^{\circ}$  S– $60^{\circ}$  N. The three-month-averaged Nino3.4 index from time  $\tau + 1$  month to  $\tau + 23$  months is used as a variable for the

output layer. The red boxes and lines highlight the convolutional filter and convolutional process, respectively; and the blue box and lines indicate the max-pooling window and the max-pooling process, respectively.  $M$  denotes the number of feature maps, while  $N$  denotes the number of neurons in the FC layer, which are set to be either 30 or 50 in this study. The global map is generated in Matplotlib<sup>31</sup>. The x and y dimensions of the map in each layer are denoted (6, 12, 18, 24, 36, 72) next to the map.

CNN 모델은 하나의 인풋 레이어 (예측 변수(독립 변수)), 3층의 convolutional 층, 2개의 맥스-풀링 층, 1개의 fully connected 층, 그리고 하나의 아웃풋 레이어(피예측량)으로 구성. 인풋 페이지의 변수들은 바다 표면 온도, 바다 열용량을 두 달 anomaly map로 만든 것과 동일하다. 연속된 3달 평균의 Nino3.4 인덱스(Nino 3.4 지역의 해수면 온도)에 있어, t+1부터 t+23으로 아웃풋 레이어를 뽑아내었다.

붉은 박스 = 컨볼루션 필터, 붉은 라인 = 컨볼루션 프로세스를 각각 강조한다. 그리고 파란박스 = 맥스-풀링 윈도우, 파란 라인 = 맥스-풀링 프로세스를 강조한다. M은 피쳐맵의 수를 나타냈고, N은 FC 층의 뉴런의 수(보통 30 혹은 50개)를 나타낸다. 글로벌 맵은 Matplotlib를 통해 그렸다. 합성곱 레이어 1층부터 MP 레이어 포함 3층 합성곱 레이어까지 들어오는 이미지들의 가로(x), 세로(y) 수치가 6,12, 18, 24, 36,72 이다.

One of the biggest limitations in applying deep learning to climate forecasts is that the observation period is too short to achieve proper training. Observations of global oceanic temperature distributions are available from 187116. This means that, for each calendar month, the number of samples is less than 150. To greatly increase the number of training data, we utilize the output of the climate models that participated in the Coupled Model Intercomparison Project phase5 (CMIP5), in which the ENSO is realistically simulated to some extent17 (Extended Data Table 1).

현재 문제는 데이터가 너무 적음. 1871년 부터 지구적으로 바다 표면 온도 분배에 대한 관측치가 있는데, 이 말은 매달 데이터가 150개보다 적은 수의 샘플만 있다는 거. 트레이닝 데이터의 수를 늘리기 위해서, 우리는 지구 변화 모델 - CMIP5라 불리는 모델을 사용할건데, 이 모델은 시뮬레이션을 돌려봤을 때 남방진동을 어느정도로 사실적으로 검증해냈기에 이 모델의 아웃풋을 사용함. (extended data table 1) - 아래 첨부. 이 표에 대한 이해도가 낮으니, 혹시 추가의견 있는 분이 있다면 추가 바랍니다

Extended Data Table 1 | Details of the CMIP5 models

CMIP ID	Modeling Group	Integration Period	Number of ensemble members
BCC-CSM1.1-m	Beijing Climate Center, China Meteorological Administration	JAN1850 - DEC2012	1
CanESM2	Canadian Centre for Climate Modelling and Analysis	JAN1850 - DEC2005	5
CCSM4	National Center for Atmospheric Research	JAN1850 - DEC2005	1
CESM1-CAM5	Community Earth System Model Contributors	JAN1850 - DEC2005	1
CMCC-CM	Centro Euro-Mediterraneo per i Cambiamenti Climatici	JAN1850 - DEC2005	1
CMCC-CMS			1
CNRM-CM5	Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	JAN1850 - DEC2005	5
CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	JAN1850 - DEC2005	5
FIO-ESM	The First Institute of Oceanography, SOA, China	JAN1850 - DEC2005	1
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory	JAN1861 - DEC2005	1
GISS-E2-H	NASA Goddard Institute for Space Studies	JAN1850 - DEC2005	5
HadGEM2-AO	National Institute of Meteorological Research/Korea Meteorological Administration	JAN1860 - DEC2005	1
HadCM3		DEC1859 - DEC2005	1
HadGEM2-CC	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)	DEC1859 - NOV2005	1
HadGEM2-ES		DEC1859 - NOV2005	4
IPSL-CM5A-MR	Institut Pierre-Simon Laplace	JAN1850 - DEC2005	1
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	JAN1850 - DEC2012	1
MPI-ESM-LR			
MRI-CGCM3	Max-Planck-Institut für Meteorologie (Max Planck Institute for Meteorology)	JAN1850 - DEC2005	3
MRI-CGCM3	Meteorological Research Institute	JAN1850 - DEC2005	1
NorESM1-M	Norwegian Climate Centre	JAN1850 - DEC2005	1
NorESM1-ME			1

The list of CMIP5 models used to train the CNN model.

Reanalysis data from 1871 to 1973 is also used to train the CNN model<sup>16</sup>. The period for validating the forecast skill is from 1984 to 2017 (Extended Data Table 2). We leave a ten-year gap between the latest year in the training period and the earliest year in the validation period to remove the possible influence of oceanic memory in the training period on the ENSO in the validation period.

1871년부터 1973년까지의 데이터를 다시분석한 것은 또한 CNN model을 훈련시키기 위해 사용됨. 예측 결과가 맞는지 퍼포먼스를 확인하기 위한 기간은 1984년부터 2017년까지로 잡았음.(내 생각에, 백테스팅을 한듯) \*3 우리는 트레이닝 기간 중 제일 늦은 연도와 확인용 기간 중 제일 빠른 연도 사이에 10년의 갭이 있는데, 이는 확인용 기간에서 남방진동에 대해 트레이닝 기간에서의 해양 기억?에 대해 가능한 영향을 지우기 위해서임.

Extended Data Table 2 | The dataset for training and validating the CNN model

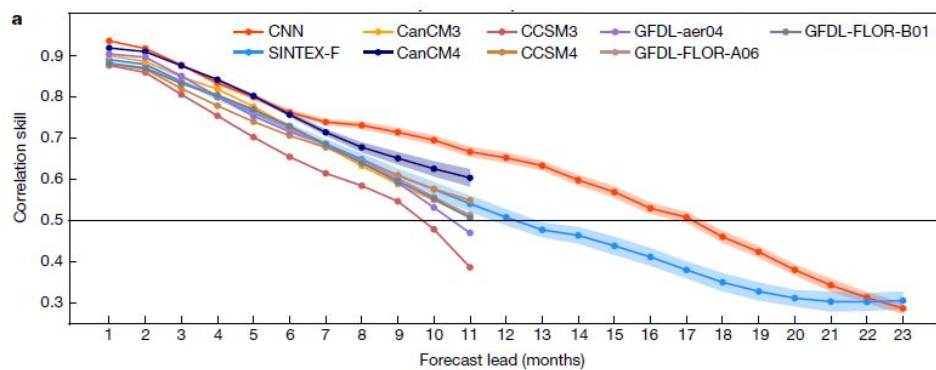
	Data	Period
Training dataset	CMIP5 historical run	1861-2004
	Reanalysis (SODA)	1871-1973
Validation dataset	Reanalysis (GODAS)	1984-2017

The dataset and period used for the training and validation of the CNN model. We note that the year in the CMIP5 models is solely dependent on the prescribed greenhouse gas forcing, so no observational information was added to the CMIP5 historical simulations.

We apply the transfer learning technique<sup>18</sup> to optimally train the CNN using both the CMIP5 output and the reanalysis data for the training period. This technique uses the knowledge acquired from a similar task with a larger number of samples for performing the target task. In this study, the CNN model is first trained using the CMIP5 output, and then the trained weights are used as initial weights to formulate the final CNN model with the reanalysis. The systematic errors in the CNN, reflecting those of the CMIP5 samples, are corrected after the second training period using reanalysis.

CMIP5 아웃풋과 학습 기간을 위한 데이터 재분석한 데이터를 사용하여 잘 학습시키기위해, 우리는 전이학습을 사용하였다. (역자 번역) 이는 어떤 학습을 하려고, 위와 같은 사례처럼 (직접적이지 않은) 수많은 샘플로 구성하여 수행했던 비슷한 태스크에서 얻은 방법을 사용한 것이다. 이 연구에서 CNN 모델은 CMIP5 아웃풋을 사용하여 첫번째로 학습되었고, 그 뒤에 학습된 가중치는 재분석과 함께들어가는 마지막 CNN모델을 수식화하기위한 초기 가중치값으로 사용하였다. 이러한 CMIP5 샘플을 반영한 CNN의 시스템적인 에러는 재분석을 사용한 두번째 학습 기간이 지나면 고쳐질 것이다.

Fig 2의 a는 1984년에서 2017년까지의 Nino3.4가 3개월마다 움직인 평균과 상관관계 퍼포먼스를 보여준다. CNN 모델에 있는 Nino3.4 인덱스의 예측 퍼포먼스는 특히 6개월 이상 전에 예측할 수 있고, 이는 시스템적으로 최신의 다이내믹 예측 시스템에 비해 훨씬 낫다는 것을 보여준다.





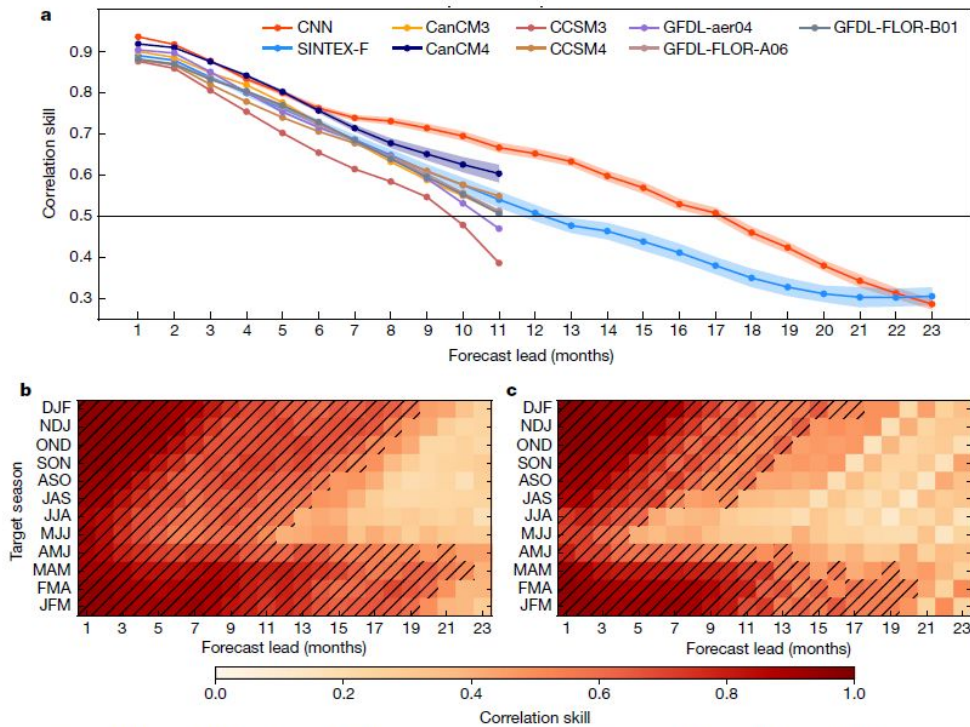


Fig. 2 | ENSO correlation skill in the CNN model. **a**, The all-season correlation skill of the three-month-moving-averaged Nino3.4 index as a function of the forecast lead month in the CNN model (red), SINTEX-F dynamical forecast system (blue), and dynamical forecast systems included in the North American Multi-Model Ensemble (NMME) project (the other colours). The validation period is between 1984 and 2017. The shading around the lines for CNN, CanCM4 and SINTEX-F denotes the 95% confidence interval, based on the bootstrap method. **b**, **c**, The correlation skill of the Nino3.4 index targeted to each calendar month in the CNN model (**b**) and the SINTEX-F dynamical forecast system (**c**). Hatching highlights the forecasts with correlation skill exceeding 0.5.

The CNN model is one of two models with the best forecasting skills for the first six forecast lead months. The all-season correlation skill of the Nino3.4 index in the CNN model is above 0.5 for a lead of up to 17 months, whereas it is 0.37 at a lead of 17 months in the SINTEX-F5, a leading dynamical forecasting system. We conclude that the CNN model provides a skilful forecast of ENSO events up to 1.5 years in advance: a result that is not possible using any of the state-of-the-art forecast systems.

CNN모델은 6달 전에 예측하는 모델중에서도 가장 퍼포먼스가 잘나는 모델 2개 중 하나이다. CNN모델에 있는 Nino3.4 인덱스의 사계절 상관관계는 17개월 전까지 0.5가 넘는 값이 나왔고, 반면에 다이내믹 예측 시스템 중 잘되는 모델인 SINTEX-F5은 17개월을 앞서자 0.37이 나왔다. 우리는 다른 최신의 예측 시스템이 할 수 없었던 남방진동을 1.5년 전에 예측하는 것을 CNN모델을 통해 잘 수행할 수 있다는 결론을 내렸다.

In addition to the superiority of the deep-learning algorithm compared to the previous statistical method (Extended Data Fig. 1 and Extended Data Table 3), the skill improvement in the CNN model is due to a large number of CMIP5 samples and the successful application of the transfer learning technique to some extent (Extended Data Fig. 2). The skill of the CNN model is not much affected by changes in the training dataset (Supplementary Fig. 1). The CNN model is even

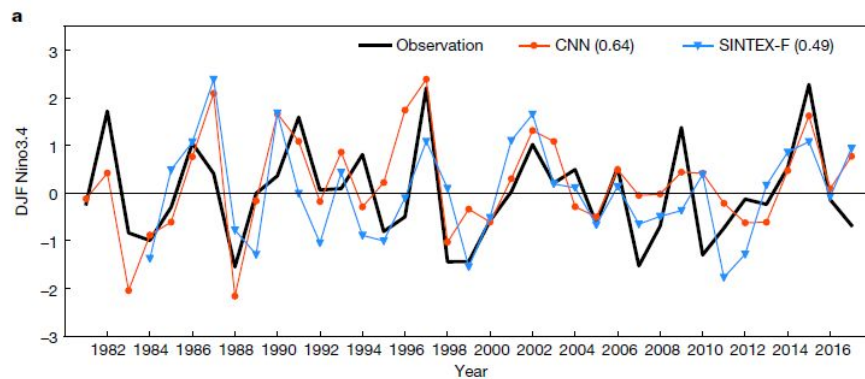
successful (with a correlation skill exceeding 0.5 for lead times up to 1.5 years) in predicting the modelled ENSO index in some of the CMIP5 models that capture realistic ENSO dynamics (Supplementary Figs. 2–4). **The skill uncertainty produced by changing the training and validation dataset is small, indicating that the CNN can provide skilful real-time forecasts.**

이전의 통계학적 방법에 비교하여 딥러닝 알고리즘이 자체가 더 우월한 것 뿐만 아니라, CNN모델이 증진된 이유는 많은 수의 CMIP5 샘플을 사용한 것과 어느정도로 전이학습이 성공적으로 활용된 것 덕분이다.

CNN 모델의 퍼포먼스는 트레이닝 데이터셋을 변화한 거에 그렇게 많은 영향을 받지 않았다. CNN 모델은 심지어 현실에서 나타난 남방진동 다이내믹스를 잡아낸 CMIP5 모델 중 몇몇 모델화된 남방진동 인덱스를 예측하는 데 역시 1.5년 전에 예측한 상관관계가 0.5이상 나올정도로 성공적이었다. **\*5 이렇게 학습 데이터셋을 바꾸는 것과 밸리데이션 데이터셋이 작아서 생기는 퍼포먼스의 불확실성은 CNN이 실제로 예측을 함에 있어 상당히 좋은 퍼포먼스를 낼 수 있다고 나타낼 수 있다.**

The CNN model also shows a higher correlation skill of the Nino3.4 index for almost all targeted seasons, compared to SINTEX-F (Fig. 2b, c). The correlation skill improvement is especially robust for predictions targeting the seasons between the late boreal spring and autumn. For example, the forecasts targeting the May–June–July (MJJ) season have correlation skill exceeding 0.5 only up to a lead of four months in the SINTEX-F, compared to a lead of up to 11 months in the CNN. **This reduces the gap in prediction skills of the CNN between target seasons, and we conclude that the CNN model is less affected by spring predictability barriers<sup>19</sup>.**

CNN 모델은 또한 SINTEX-F와 비교하여 거의 사계절을 타겟으로 한 Nino3.4 인덱스와 높은 상관관계를 나타내는 것을 보여주었다. 상관관계 퍼포먼스의 증진은 특히 늦봄(4~5월 봄중에서도 특히 늦봄)과 가을사이를 타겟팅한 계절을 예측하는데 강했다. 예를 들어, 5월-6월-7월 계절을 타겟한 예측은 SINTEX-F에서는 오직 4개월 전에만 상관관계가 0.5를 넘었는데, CNN에서는 11개월 전에 예측한 것도 0.5를 넘었다. **이는 CNN 모델이 봄 예측의 장벽에 덜 영향을 받는다는 것으로 결론지었다.**



DJF = 12월-1월-2월

The Nino3.4 index for the December–January–February (DJF) season for the 18-month-lead forecast demonstrates that the CNN model correctly predicts the ENSO amplitude (Fig. 3a).

18개월 앞선 예측을 하기 위한 DJF(12월-1월-2월)에 대한 Nino3.4 인덱스는 CNN모델이 정확하게 ENSO의 진폭(amplitude)을 예측하는 것을 나타냈다.

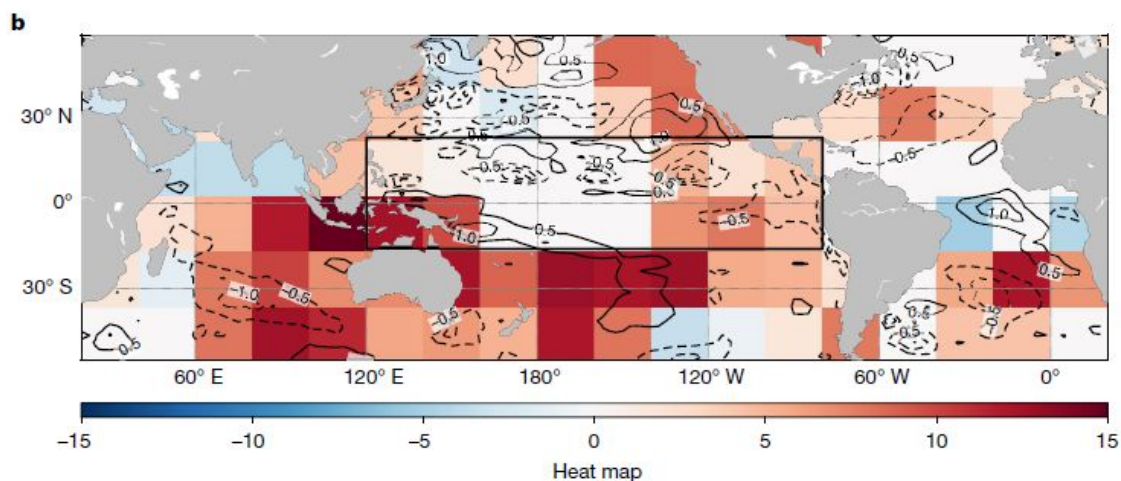
(이 부분에 대해 아는 분이 있다면 언제든지 추가해주세요!

저는 진폭 해석(Amplitude Analysis)에서 온 것으로 이해했습니다. 진폭 해석: 신호의 진폭을 특정 레벨과 비교해서 대소의 크기에 따라 신호를 분별하는 것 또는 잡음 등의 작은 신호를 제거하는 것)

To understand

how the CNN model can successfully predict the Nino3.4 index for such long lead times, we produce a heat map for the 18-month-lead forecast of the 1997/98 El Niño event (Fig. 3b).

CNN모델이 어떻게 성공적으로 Nino3.4를 이렇게 긴 기간 전에 예측했는지 이해하기 위해서, 우리는 1997/98 엘니뇨 사건의 18개월 전 예측 모델의 히트맵을 만들어보았다.



The heat map quantifies the contributions of the predictors at each grid point to the predictand; positive (negative) values in the heat maps denote predictors over certain regions contributing to the prediction of a positive (negative) Nino3.4 (Methods).

이 히트맵은 피예측량에 대한 각각 한 네모 내의 예측값 분배를 정량화 했다: 특정 지역의 예측량 분배량을 기준으로 잡고 히트맵에 있는 + 혹은 - 값으로 예측값을 나타내었다. (해석 어려움)

The heat map of the predictors (an anomaly during the MJJ season in 1996) for the DJF season in 1997/98 indicates that the anomalies over the tropical western Pacific<sup>21</sup>, Indian Ocean<sup>11</sup> and subtropical Atlantic<sup>12</sup> (red shadings in Fig. 3b) are the main contributors to the successful prediction of the 1997/98 El Niño.

The oceanic signals highlighted by the heat map can induce the 1997/98 El Niño on the basis of the physical linkages. A positive heat content anomaly over the southern tropical western Pacific denotes the recharged heat contents for successive El Niño developments<sup>21</sup>.

1997/1998의 12월-1월-2월을 위한 1996년의 5월-6월-7월에 동안의 특변화점을 예측한 값을 그린 히트맵을 보면 서태평양 적도지역, 인도양, 북태평양 아열대 지역의 변화가 1997/98 엘니뇨를 성공적으로 예측한 데 큰 역할을 한 요소라고 볼 수 있다.

히트맵에 의해 잘보이는 바다의 신호들 (위에서 서술했듯 서태평양 적도지역, 인도양, 북태평양 아열대지역의 변화)이 물리적 연결고리에 기반하여 97/98년 엘니뇨를 만들어낼 수 있다. 서남부 열대 태평양에 대해서 열용량이 더 많아진 것은 (더 더워진 것은) 계속된 엘니뇨의 발생을 위해 열용량이 재충전되고 있다는 것을 보여준다.

A negative SST anomaly over the southwestern Indian Ocean contributes to a negative Indian Ocean Dipole (IOD) during the 1996 boreal autumn<sup>11</sup> (Extended Data Figs. 3a and 4b).

The negative IOD causes negative SST anomalies to develop over the entire Indian Ocean during the subsequent season (Extended Data Figs. 3b and 4d), which triggers the westerly over the equatorial western Pacific, thus inducing an El Niño event after one year<sup>22</sup>.

인도양 다이폴은 용어해석에 자세히 설명해둬!

남서 인도양의 바다 표면 온도가 유독 낮아진 현상은 1996년도 늦가을동안 음성 인도양 다이폴 현상이 일어나는 데 기여를 했다.

음성 인도양 다이폴은 인도양 전체를 걸쳐 바다 표면온도가 차가워지는 현상을 발달시키게 되고, 이로인해 서태평양 적도지역의 서풍을 야기시켜서, 1년 뒤에 엘니뇨 현상이 일어나게끔 한다.

Moreover, a negative SST anomaly over the northern subtropical Atlantic during the MJJ 1996 season contributes

to the 1997/98 El Niño event by exciting mid-latitude Pacific variability<sup>12</sup> (Extended Data Figs. 3c, d and 4a).

In addition to the ENSO amplitude, the global impacts of El Niño events vary greatly according to the detailed zonal distributions of El Niño SST anomalies: central-Pacific-type (CP-type) and eastern-Pacific-type (EP-type) El Niños<sup>23</sup>.

Therefore, successful prediction of the types of El Niño based on the zonal locations of the SST anomaly is essential to improve the quality of global climate forecasts. For this purpose, we set up an additional CNN model to predict the type of El Niño.

더욱이, 1996년의 5월-6월-7월동안 북대서양 아열대에서 일어난 바다 표면 온도가 낮아지는 이상현상은 중위도의 태평양 변동성을 자극하며 1997/1998년 엘니뇨 현상에 기여하였다. 남방진동 진폭에 대해 추가하자면, 엘니뇨 현상의 전지구적 영향은 엘니뇨 바다표면온도 이상현상이 각 지역에 따라 어떻게 분배되어있느냐에 따라 크게 변동한다. 중앙태평양 타입과 동태평양타입의 엘니뇨를 볼 것  
그러므로, 해양표면온도 이상현상의 지역별 위치에 기반하여 엘니뇨의 타입을 성공적으로 예측하는 것은 전지구적인 기후 예측 퀄리티를 증가시키는 데 필수적이다. 이러한 목적으로 우리는 엘니뇨의 타입을 예측하는 추가 CNN 모델을 구축하였다.

In this model, the predictand corresponds to the percentage occurrence of three El Niño categories<sup>24</sup>: CP-type, EP-type and a mixture of the two. The category with the highest probability of the three types is considered as the final forecast.

이 모델에서, 피예측량은 중앙태평양타입(CP-type), 동태평양타입(EP-type), 이 두개가 섞인 타입 이렇게 3개의 엘니뇨 카테고리의 발생 확률과 같다. 이 세개의 타입중에 가장 높은 확률의 카테고리가 최종 예측값으로 나오게 된다.

We note that the reanalysis

data are not used to train the CNN model for predicting the El Niño types, because the El Niño events during the reanalysis training period are known to be of a single type<sup>23</sup>. Hence, we train the CNN model to predict the type of El Niño using only the CMIP5 model outputs and we do not apply the transfer learning technique.

우리는 엘니뇨 타입을 예측하기 위한 CNN모델을 사용하는데 재분석한 데이터가 사용되지 않았음을 명시하고자 한다. 왜냐하면 재분석 트레이닝 기간에 있는 엘니뇨 현상은 하나의 타입으로 알려져 있었기 때문이다. 고로, 우리는 오직 CMIP5 모델의 아웃풋만 사용하여 CNN모델이 엘니뇨의 타입을 예측하게끔 했고, 전이학습 기술은 활용하지 않았다.



A series of hindcast experiments are conducted to predict the types of El Niño events 12 months in advance, and the hit rate of the CNN model is 66.7% during the validation period (1984–2017) (Fig. 4a and Extended Data Table 4).

엘니뇨의 타입을 12개월 전에 예측하는 일련의 백테스팅(=hindcast) 실험이 시행되었고, CNN모델은 1984년부터 2017년까지 66.7퍼의 적중률을 달성했다.

The hit rate in the random forecast with 95% confidence interval is between 12.5% and 62.5%, so the CNN hit rate of 66.7% is significantly superior with a P value of 0.016. By contrast, none of the dynamical forecast models exhibit statistically significantly better forecast skills than do random forecasts, implying that the CNN model overcomes a long-standing weakness of the state-of-the-art forecast models<sup>25</sup>. This would indicate that the deep-learning-based model can predict the spatial complexity of the El Niño events with great precision<sup>16,26</sup>.

랜덤 포레스트모델에서 적중률은 95%의 콘피던스 인터벌로 12.5%에서 62.5% 사이였다, 그래서 CNN 모델의 66.7% 적중률은 0.016의 P value와 함께 특히 뛰어남을 나타낸다.

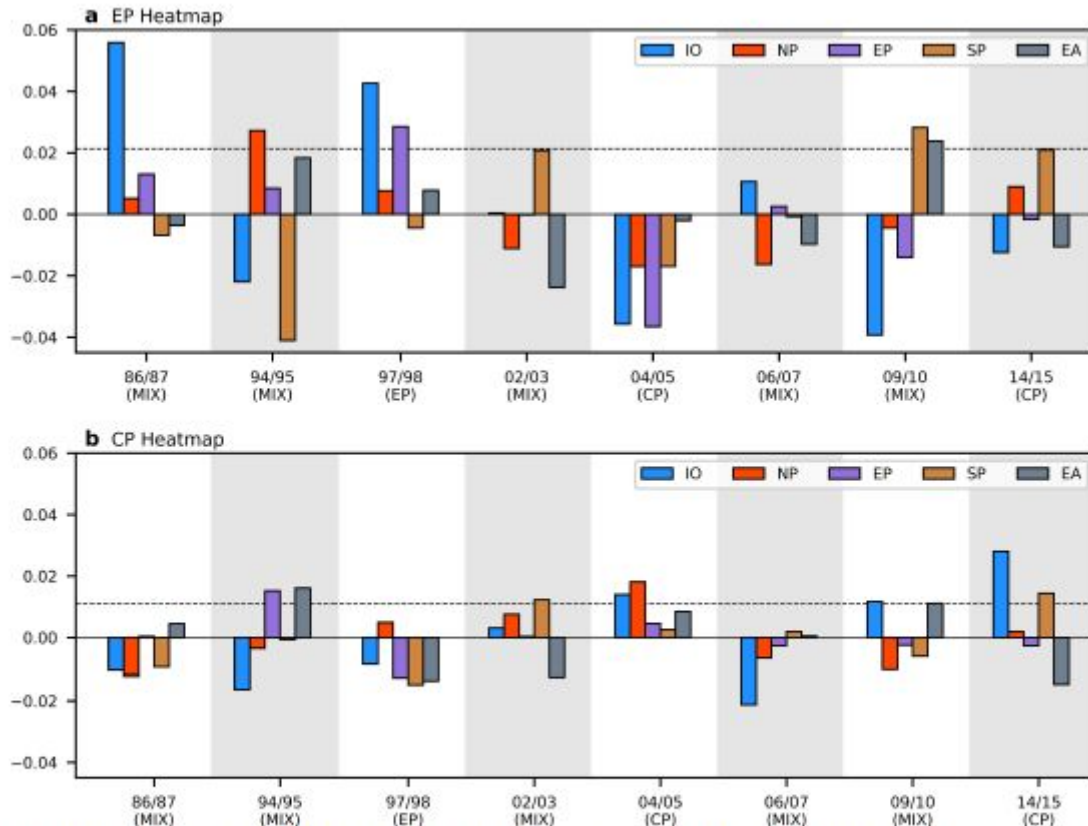
반면에 다이내믹 예측 모델 중 그 어떤것도 랜덤 포레스트보다 더 나은 퍼포먼스를 보이지 않았다는 것은 CNN모델이 오랜시간 지속되었던 최신 모델의 약점을 극복했음을 의미한다. 이 사실은 딥러닝 기반의 모델이 엘니뇨 현상의 공간 복잡성을 높은 정확도로 잘 예측한다는 것을 보여준다.

Apart from forecasting the ENSO strength and type, the CNN model also allows us to identify which SST signals lead to EP-type or CP-type El Niño events. For this, we compute an area-averaged heat map value over five ocean basins (Extended Data Fig. 5).

남방진동의 강도와 타입을 예측하는 것과는 별도로, CNN 모델은 또한 어떤 바다표면온도 변화 시그널이 동태평양타입(EP-type)과 중앙태평양타입(CP type)의 엘니뇨 현상을 예측할 수 있는 지 알아낼 수 있다는 것을 알려준다. 이것을 위해, 우리는 지역-평균 히트맵 값을 다섯개의 해양 분지 지역에 대해 계산하였다.



(해양 분지)



**Extended Data Fig. 5 | The area-averaged heat map values for El Niño events. a, b.** The area-averaged heat map for EP-type El Niño events (a), and CP-type El Niño events (b) among all El Niño events over five ocean domains (that is, south Pacific, equatorial Pacific, north Pacific, Indian Ocean and equatorial Atlantic). These areas are defined as: south Pacific, [160° E–60° W, 57.5°–17.5° S]; equatorial Pacific, [120° E–80° W, 17.5° S–22.5° N]; north Pacific [120° E–100° W, 22.5–62.5° N]; Indian Ocean, [40°–120° E, 37.5° S–22.5° N]; and equatorial Atlantic [60°–0° W, 17.5° S–22.5° N]. The horizontal dashed line denotes one standard deviation of the heat map value for the displayed El Niño events for five ocean basins. We note that only the heat maps of the El Niño events for which the type is correctly predicted in the CNN are analysed.

Then, we select two

cases having the biggest heat map value for each ocean basin, which can be considered as the most favourable patterns for the development of EP-type or CP-type El Niño events.

그래서, 우리는 각각의 해양 분지에 대해서 가장 큰 히트맵 값을 가진 (EP-type 혹은 CP-type의 엘니뇨 현상의 발전을 위해 가장 선호되는 패턴인) 두 케이스를 골라보았다.

Even though some of the patterns were selected from the years when the mixed-type El Niño occurred, the existing literature and additional analysis show that the selected patterns can lead to EP-type or CP-type El Niños.

비록 엘니뇨 믹스 패턴이 발생한 연도에 정해진 몇몇 패턴이 선택되었을지라도, 현존하는 논문과 추가 조사가 이렇게 정해진 패턴 역시 EP-type 혹은 CP타입 엘니뇨로 이어질 수 있다고 말한다.

For the EP-type El Niño, the selected heat content anomalies at one year before the target season can induce a positive IOD that is one season before the El Niño mature phase<sup>27</sup> (Fig. 4e).

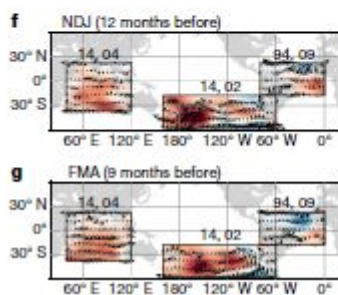
- 1) 동태평양 타입의 엘니뇨에 있어서, 타겟 계절 바로 전 1년 중 나타난 열용량 이상현상이 엘니뇨 성숙단계 이전의 시즌인 양성 인도양 다이폴을 일으킬 수 있다.
- 2) EP(동태평양) 타입의 엘니뇨에 있어서, 타겟 계절 전년도에 일어나는 특정하게 선택된 열용량의 이상변화가 엘니뇨 성숙 단계 이전의 계절인 양성 인도양 다이폴을 일으킬 수 있다.

For the CP-type EI

Niño, SST cooling over the north tropical Atlantic is shown (Fig. 4f, g).

Our results are consistent with those of previous studies<sup>11,28</sup>. The CP-type EI Niño precursors in the South Pacific and the Indian Ocean have not been reported before, and the additional analysis shows that the identified precursors can lead the CP-type EI Niño event (Extended Data Fig. 6). This indicates that the CNN can be a powerful tool to reveal complex ENSO mechanisms.

중양태평양 타입의 엘니뇨의 경우, 북대서양 열대 지역에서 바다표면온도가 낮아지는 현상이 보였다. 남태평양과 인도양에 있는 중양태평양 타입의 엘니뇨 이전에 미리 알려주는 전조는 그 전에 보이지 않았다. 그리고 추가적으로 시행된 분석은 이제 보여지는 이 전조들이 CP-type 엘니뇨 현상으로 이어질 수 있다고 보여주었다. 이는 CNN이 복잡한 남방진동 메커니즘을 파악하는데 좋은 툴이 될 수 있을 것이라는 것을 알려주었다.



However, future studies are warranted to explore the physical mechanisms of the statistical relationship revealed by the CNN model with the limited sample size.

The superiority of CNN over previous models derives from the successful extraction of features in input variables by using a convolutional process. The CNN identifies the basic shape that can be used to encode a variety of different shapes, thus exhibiting a partial invariance to translation and deformation<sup>29</sup>.

하지만, 미래의 연구는 한정된 샘플 사이즈를 가진 CNN모델이 찾아낸 통계적 관계의 물리 메커니즘을 알아내야함이 확실하다.

그 전 모델들과 비교하여 CNN의 우월성은 인풋 변수에서 컨볼루션 절차를 사용하여 성공적으로 추출해낸 특징들에서 기인한다. CNN은 다양한 형태의 변화를 해석해내는 데 사용되는 기초 형태를 잘 찾아내었고, 고로 해석과 변형에 대한 부분적인 불변성이 보인다.



**불변성:** 통계적 추론의 구성요소인 모수치, 측정치, 가정된 분포 등의 구조가 어떤 변환에 의하여 변하지 않는 성질이다.

Hence, within the CNN, the precursor signals can properly affect the predictand even if the detailed spatial distribution has been shifted or deformed from typical precursors. In addition, the convolutional process allows the CNN model to be properly trained using a relatively small number of climate samples.

Deep learning has led to progress in forecasting many Earth system components, but its applications in climate prediction so far have been rare<sup>30</sup>. The successful application of deep learning by using transfer learning and heat map analysis for predicting and understanding a climate phenomenon as reported here could boost interdisciplinary research between engineering and the Earth sciences.

그러므로, CNN 내에서, 비록 전형적인 전조에서 세부적인 공간 분배가 바뀌거나 변형되었을지라도, 전조 신호는 적절히 피예측량에 영향을 준다. 추가적으로, 컨벌루션 프로세스는 CNN 모델이 상대적으로 작은 기후 샘플을 가지고도 적절하게 학습되게 해준다. 딥러닝은 지구 시스템 많은 요소들은 예측하는 데 발전할 수 있도록 이끌었으나, 기후 예측에 대한 응용은 지금까지 거의 없었다. 여기서 서술되었듯 기후 현상을 이해하고 예측하기 위해 히트맵 분석과 전이학습을 사용하여 성공적인 딥러닝의 활용을 이끌어낸 것은 엔지니어링과 지구과학간의 학제간 연구를 더욱 발전시켜나갈 수 있을 것이다.

[Methods]

DAY1

### Architecture of the CNN model applied to ENSO forecasts.

The CNN-based<sup>32</sup>

statistical model used for the ENSO forecasts has three convolutional layers and two max-pooling layers between the layers. The max-pooling process extracts the largest value from each  $2 \times 2$  grid. A third convolutional layer is linked to the neurons in the fully connected layer, which is linked to the final output. The dimension of the output is one, and the CNN model is formulated separately for each forecast lead month and target season. The total number of convolutional filters and neurons in the fully connected layer is either 30 or 50. Therefore, there are four combinations of CNN models (C30H30, C30H50, C50H30 and C50H50, where the numbers after C and H denote the number of convolutional filters and neurons in the fully connected layer, respectively). The total number of parameters for C30H30 is 117,511, for C30H50 is 182,351, for C50H30 is 211,811 and for C50H50 is 319,851, respectively. The predicted Nino3.4 indices from four CNN models with a different number of convolutional filters and neurons are averaged to obtain the final forecast results. This averaging leads to a slight systematic increase in the forecast skills by cancelling out the forecast errors in individual CNN models<sup>33</sup>. The size of the mini batch for each epoch is set to 400, and the number of an epoch is 700 for the first training using CMIP5 output. The different numbers of the epoch from 600 to 1,000 do not affect the ENSO forecast skill at all. For the second training using the reanalysis, the number of an epoch is set to 20. The learning rate is fixed to 0.005, and it is not changed throughout the iterations: a learning rate schedule is not used.

남방진동 예측에 사용된 CNN 모델의 구조

남방진동 예측에 사용된 통계 모델에 기반한 CNN 모델은 3개의 컨볼루션 층과 그 사이를 이루는 두 개의 맥스풀링층으로 이루어져있다. 맥스풀링 프로세스는 각각  $2 \times 2$  그리드에서 가장 큰 값을 추출해낸다. 세 번째 컨볼루션 층은 최종 아웃풋값에 연결된 풀리 커넥티드 층에 있는 뉴런과 연결되어 있다. 그 아웃풋의 차원은 하나이고, CNN 모델은 예측하기 위해 선정된 앞선 달과 타겟 시즌에 대해 각각 나누어져 수식화되어있다. 이 컨볼루션 필터와 풀리 커넥티드 층을 이루는 뉴런의 개수는 30개나 50개이다. 그러므로, 네가지의 융합된 CNN모델들이 있다. C30H30, C30H50, C50H30, C50H50 - C 뒤에 있는 숫자는 컨볼루션 필터의 숫자이고 H 뒤에 있는 숫자는 풀리커넥티드층의 뉴런 개수이다. 모수의 전체 수는 각각 C30H30의 경우 117,511, C30H50의 경우 182,351, C50H30의 경우 211,811, C50H50의 경우 319,851이다. 이 네가지 CNN모델에서 나온 예측된 니노3.4 인덱스는 최종 예측값을 얻기위해 평균화되었다. 이 평균화한 값은 예측 퍼포먼스에서 각각의 CNN모델이 가지는 예측 에러를 줄임으로써 약간의 시스템적인 증진을 일으킨다. 각각 에폭의 미니배치 사이즈는 400으로 설정되어있고, CMIP5 아웃풋을 사용한 첫번째 트레이닝을 위한 에폭의 수는 700이다. 에폭의 수는 600에서 1000으로 달랐으나 이는 남방진동 예측 퍼포먼스에 영향을 전혀 미치지 않았다. 재분석을 사용한 두번째 트레이닝에서는 에폭의 수가 20으로 설정되었고, 러닝레이트는 0.005에 고정하였다. 러닝레이트스케줄은 사용되지 않았으며, 초기값에서 전혀 바뀌지 않았다.

### Convolutional process.

The convolutional process of the CNN involves the extraction of local characteristics from the global maps, and the calculation of dot products between values in the convolutional filter and those in the input layer. The output of the convolutional process is then translated into a feature map. The values of the convolutional filter are determined automatically by iteration, in order to minimize the cost function, defined as the mean squared difference between the predicted and true distributions.

### 콘볼루션 프로세스

CNN의 콘볼루션 프로세스는 글로벌 맵에서 지역적인 특성을 추출해내는 것과 연관이 있는데, 콘볼루션 필터의 값과 인풋 레이어의 값을 내적하여 계산한 값이다. 그러면 콘볼루션 프로세스의 아웃풋은 피쳐맵으로 해석된다. 콘볼루션 필터의 값은 예측값과 실제값의 평균 제곱의 차로 정의되는 코스트 함수를 최소화 시키기 위해 자동으로 초기화된다.

MSE: <https://blog.naver.com/tommybee/221750403141>

The value of the  $j$ th feature map in the  $i$ th convolutional layer at grid point  $(x, y)$  (denoted  $\mathbf{v}_{i,j}^{x,y}$ ) is calculated using the following equation:

$x, y$  포인트에 있는  $i$ 번째 콘볼루션 층에 있는  $j$ 번째 피쳐의 값은 ( $v$  어찌고로 표기됨) 아래 방정식을 사용하여 계산할 수 있다.

$$\mathbf{v}_{i,j}^{x,y} = \tanh \left( \sum_{m=1}^{M_{i-1}} \sum_{p=1}^{P_i} \sum_{q=1}^{Q_i} w_{i,j,m}^{p,q} v_{(i-1),m}^{(x+p-P_i/2, y+q-Q_i/2)} + b_{i,j} \right)$$

where  $P_i$  and  $Q_i$  denote the zonal and meridional dimensions of the convolutional filter for the  $i$ th convolutional layer, respectively. A hyperbolic tangent function ( $\tanh$ ) is used as the activation function. The dimensions of the convolutional filter were set at  $8 \times 4$  during the first convolutional process (that is,  $P_1 = 8$ ;  $Q_1 = 4$ ), and at  $4 \times 2$  during the second and third convolutional processes.  $M_{i-1}$  denotes the number of feature maps in the  $(i-1)$ th layer. On the other hand,  $w_{i,j,m}^{p,q}$  denotes the weight at grid point  $(p, q)$  in the convolutional filter; this is used to link the  $m$ th feature map in the  $(i-1)$ th layer to the  $j$ th feature map in the  $i$ th convolutional layer. Additionally,  $v_{(i-1),m}^{(x+p-P_i/2, y+q-Q_i/2)}$  denotes the value of the  $m$ th feature map for the  $(i-1)$ th convolutional layer at grid point  $(x+p-P_i/2, y+q-Q_i/2)$ , while  $b_{i,j}$  denotes the bias of the  $j$ th feature map in the  $i$ th convolutional layer. To ensure that the horizontal dimension of the  $i$ th layer corresponds to the horizontal dimension of the  $(i-1)$ th layer, the empty spaces were filled with zeroes using a padding technique.

$i$ 는 컨볼루션층을 이야기하는데,  $P_i$ 는  $i$ 번째 컨볼루션 층에 속한 지역적 컨볼루션 필터의 차원들 그리고  $Q_i$ 는  $i$ 번째 컨볼루션 층에 속한 1)meridional (남쪽의?) 컨볼루션 필터의 차원들을 뜻한다. 하이퍼볼릭 탄젠트 함수는 활성화 함수로 사용된다. 컨볼루션 필터의 차원은 첫번째 컨볼루션 프로세스에서는  $8 \times 4$ (즉,  $P_1 = 8$ ,  $Q_1 = 4$ )로 설정되었고, 두번째와 세번째 컨볼루션 프로세스에서는  $4 \times 2$ ( $P_2, P_3 = 4$ ,  $Q_2, Q_3 = 2$ )로 설정되었다.  $M_{i-1}$ 은  $i-1$ 번째 레이어의 피쳐맵의 숫자를 뜻한다. 반면에,  $w_{i,j,m}^{p,q}$ 은 컨볼루션 필터 속  $p, q$  포인트의 가중치를 의미한다; 이는  $i$ 번째 컨볼루션 층에 있는  $j$ 번째 피쳐맵에 대한  $i-1$ 번째 레이어에

있는  $m$ 번째 피쳐맵을 잇는데 사용된다. 추가적으로,  $v_{(i-1),m}^{(x+p-P_i/2, y+q-Q_i/2)}$ 는  $(x+p-P_i/2, y+q-Q_i/2)$  좌표에 있는  $i-1$ 번째 컨볼루션 층을 위한  $m$ 번째 피쳐맵의 값을 의미하고,  $b_{i,j}$ 는  $i$ 번째 컨볼루션층에 있는  $j$ 번째 피쳐맵의 바이어스를 뜻한다.  $i-1$ 번째 층의 수평 차원에 대응하여  $i$ 번째 층의 수평( $x$ 축) 차원을 확실히 하기 위해서, 빈 공간은 패딩 기술을 이용한 0으로 채워진다.

## DAY 2

**Heat map analysis.** To decipher the heat map analysis, the output variable is first mentioned as follows:

$$V = \sum_{n=1}^N \left\{ \tanh \left[ \sum_{m=1}^{M_L} \sum_{y=1}^{Y_L} \sum_{x=1}^{X_L} ((W_{F,m,n}^{x,y} v_{L,m}^{x,y})) + b_{F,n} \right] W_{O,n} \right\} + b_O$$

where  $V$  denotes the output neuron (that is, the predictand), while  $X_L$  and  $Y_L$  denote the dimensions of the feature map in the third convolutional layer (that is,  $X_L = 18$ ;  $Y_L = 6$ ).  $N$  denotes the number of neurons in the fully connected layer,  $W_{F,m,n}^{x,y}$  denotes the weight at grid point  $(x, y)$  (used to link the  $m$ th feature map in the last convolutional layer  $L$  to the  $n$ th neuron in the fully connected layer  $F$ ),  $v_{L,m}^{x,y}$  denotes the value of the  $m$ th feature map of the last convolutional layer  $L$  at grid point  $(x, y)$ ,  $b_{F,n}$  denotes the bias of the  $n$ th neuron in the fully connected layer  $F$ ;  $W_{O,n}$  denotes the weight (used to link the  $n$ th neuron in the fully connected layer to the output layer  $O$ ), and  $b_O$  denotes the bias of the output layer  $O$ . Through this process, we calculate the sum of the products between the last convolutional layer and the weights; hence, all spatial information is lost.

However, the heat map is calculated considering the contribution of each grid point to the output neuron: this is expressed by the weights used to link the last convolutional layer to the fully connected layer (that is,  $W_{F,m,n}^{x,y}$ ). In other words, the heat map is calculated by avoiding the summation of the horizontal dimensions. The heat map value for the neuron of the output layer at grid point  $(x, y)$  (indicated as  $h^{x,y}$ ) was finally calculated using the following equation:

$$h^{x,y} = \sum_{n=1}^N \left\{ \tanh \left[ \sum_{m=1}^{M_L} (W_{F,m,n}^{x,y} v_{L,m}^{x,y}) + \frac{b_{F,n}}{X_L Y_L} \right] W_{O,n} \right\} + \frac{b_O}{X_L Y_L}$$



DAY 3

**ENSO-index forecasts.** The Nino3.4 index (the averaged SST anomaly over the area 170°–120°W, 5° S–5° N) was used as the predictand in this study. The start of the predicted period corresponds to the time of the latest available observed data. For example, the output of the dynamical forecasts starting from 1 January was compared with the CNN model forecasts using the predictors of the previous year's OND period. The lead time is defined as the number of months between the latest available observed data and the middle of the three-month forecast target period. The target periods for all forecast lead months are included between January 1984 and December 2017.

The forecast skill of the Nino3.4 index was evaluated using the temporal anomaly correlation coefficient  $C$  as a function of the forecast lead months  $l$ :

$$C_l = \sum_{m=1}^{12} \frac{\sum_{y=s}^e (Y_{y,m} - \bar{Y}_m) ((P_{y,m,l} - \bar{P}_{m,l}))}{\sqrt{\sum_{y=s}^e ((Y_{y,m} - \bar{Y}_m))^2 \sum_{y=s}^e ((P_{y,m,l} - \bar{P}_{m,l}))^2}}$$

Here,  $Y$  and  $P$  denote the observed and the predicted values, respectively.  $\bar{Y}_m$  and  $\bar{P}_{m,l}$  denote the temporal climatologies with respect to the calendar month  $m$  (from 1 to 12) and the forecast lead months  $l$ . The label  $y$  denotes the forecast target year. Finally,  $s$  and  $e$  denote the earliest (that is, 1984) and the latest year (that is, 2017) of the validation, respectively.

The confidence interval of the forecast skills for the CNN and the dynamical forecast system is calculated using the bootstrap method. First, we randomly select the  $N$  ensemble members.  $N$  is a number of ensemble members for each forecast systems (for example, for the CNN model,  $N$  is 40). During the random selection process, overlapping is allowed; the selected ensemble member can be selected again. The forecast skill of the ensemble-averaged value is then calculated. This procedure was repeated 10,000 times: the 250th highest and lowest values of the forecast skill are used to define the 95% confidence interval.

DAY 4

**El Niño type forecasts.** The EP-type, CP-type and mixed-type El Niño events were defined using the Nino3 index (SST anomaly averaged over 150°–90° W, 5° S–5° N) and the Nino4 index (SST anomaly averaged over 160° E–150° W, 5° S–5° N) as follows<sup>24</sup>. UCEI denotes a unified complex ENSO index.

$$\text{UCEI} = (N_3 + N_4) + (N_3 - N_4)i = re^{\theta i}$$

where

$$r = \sqrt{(N_3 + N_4)^2 + (N_3 - N_4)^2}$$

and

$$\theta = \begin{cases} \arctan \frac{(N_3 - N_4)}{(N_3 + N_4)}, & \text{when } N_3 + N_4 > 0 \\ \arctan \frac{(N_3 - N_4)}{(N_3 + N_4)} + \pi, & \text{when } N_3 + N_4 < 0; N_3 - N_4 > 0 \\ \arctan \frac{(N_3 - N_4)}{(N_3 + N_4)} - \pi, & \text{when } N_3 + N_4 < 0; N_3 - N_4 < 0 \end{cases}$$

$N_3$  and  $N_4$  denote the Nino3 and Nino4 indices, respectively. The EP-type El Niño is defined for  $\theta > 15^\circ$  and  $\theta < 90^\circ$ , while the CP-type El Niño is defined for  $\theta > -90^\circ$  and  $\theta < -15^\circ$ . Finally, the mixed-type El Niño is defined for  $\theta > -15^\circ$  and  $\theta < 15^\circ$ . Overall, the El Niño events are defined when the  $r$  value during the DJF season is greater than their standard deviation. The El Niño events in the CMIP5 output are classified using the same methodology. The number of samples used to train the CNN model is 872.

The significance test for the El Niño-type prediction was performed on the basis of random forecasts. After the El Niño events are classified randomly, the hit rate of the random forecasts is calculated. The occurrence ratio of each type of El Niño is obtained from the CMIP5 historical simulation (that is, 30%, 26% and 44% for EP-type, CP-type and mixed-type El Niños, respectively). We note that the random forecasts with climatological occurrence ratio from CMIP5 archives give a higher hit rate than that of the random forecast with an equal occurrence ratio. This procedure is repeated 10,000 times. Afterward, the 250th highest and lowest values of the hit rate are used to define the 95% confidence interval.

## DAY 5

**The feed-forward neural network model for ENSO forecasts.** To compare the ENSO forecast skills in the CNN model, we formulated a nonlinear statistical model based on the feed-forward neural network method<sup>35</sup>. The empirical orthogonal function (EOF) principal components of the SST and the heat contents were used as predictors of the neural network model. In addition to the EOF principal components over Indo-Pacific regions (40° E–100° W, 20° S–20° N), the EOF principal components over the Atlantic (100–0° W, 30° S–30° N), and North Pacific (120° E–100° W, 20–60° N) were obtained separately, and those were used as predictors to consider the remote impact of Atlantic and North Pacific climate variability on the ENSO. The predictand is the Nino3.4 index. In addition to the reanalysis data from 1871 to 1973, we used the CMIP5 model output to train the neural network model, just as for the CNN in this study. The principal component time series for the CMIP5 model output were obtained by calculating the pattern projection coefficients onto the observed EOF eigenvectors. The number of hidden layers is two, and a number of hidden neurons at each layer is set to 20. The activation function is a hyperbolic tangent function.

A series of sensitivity hindcast experiments were performed with a different number of predictors to predict the DJF Nino3.4 index at a lead of 18 months (Extended Data Table 3). The forecast skill of the best neural network model is 0.52 with 9 EOF principal components for Indo-Pacific and 7 EOF principal components for the Atlantic and North Pacific as predictors, compared to 0.64 for the CNN. The correlation skills of the neural network models strongly vary with a slightly different number of predictors. Therefore, a correlation skill similar to that of the best neural network model is difficult to achieve in any other neural network model with a different setting. With the best setup of the neural network model, we performed hindcast experiments to calculate the all-season correlation of Nino3.4 for leads of up to 24 months (Extended Data Fig. 1). Comparison with the all-season correlation skill of Nino3.4 demonstrates that the ability to predict an ENSO in the CNN is systematically superior to that in the neural network model.

**Observed dataset and CMIP5 outputs.** The historical simulations produced by the 21 CMIP5 models were used to train the CNN model. Details of the establishments where the models were run and about the integration periods are given in Extended Data Table 1. A single ensemble member for all CMIP5 models is used to train the CNN model. As a result, for each target month, the total number of CMIP5 samples to train for the CNN model is 2,961. Also, 103 years of the reanalysis data (between years 1871 and 1973) from the Simple Ocean Data Assimilation version 2.2.416 were used to train the CNN model. The reanalysis product after the year 1974 was not used in any training process, to ensure that the training and validation periods are independent of each other. To validate the performance of the model by comparison with the observed values, monthly mean SST and heat content data were collected from the Global Ocean Data Assimilation System (GODAS) reanalysis (1984–2017)<sup>36</sup>, while the horizontal wind vector at 925 hPa and the precipitation



data were obtained from the ERA-Interim archives (1984–2017)<sup>37</sup>. Eight models of the North American Multimodel Ensemble phase 1 (1984–2017)<sup>38</sup> and SINTEX-F (1984–2017)<sup>5,39</sup> were employed to compare the prediction performance of the Nino3.4 index with the CNN model. The forecasts were performed for lead times of up to 12 months using NMME, and for lead times of up to 24 months using SINTEX-F. The anomalies of the dynamical forecasts are calculated by subtracting the climatology of each system with respect to the forecast lead months. The spatial resolution was interpolated at a resolution of  $2.5^{\circ} \times 2.5^{\circ}$  for the analysis, and at a resolution of  $5^{\circ} \times 5^{\circ}$  for the training of the CNN model, in order to reduce the number of weighting coefficients.