LETTER

https://doi.org/10.1038/s41586-019-1559-7

Deep learning for multi-year ENSO forecasts

Yoo-Geun Ham[1]*, Jeong-Hwan Kim[1] & Jing-Jia Luo[2,3]

Variations in the El Niño/Southern Oscillation (ENSO) are associated with a wide array of regional climate extremes and ecosystem impacts[1]. 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[2]. 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[3] 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.

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[4]. Although ENSO forecasts using atmosphere–ocean coupled models[5,6] generally outperform those of current statistical models[1], 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[1].

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[5,7]. Interestingly, equatorial Pacific anomalies during several La Niña events lingered for several years[8]. The high-frequency equatorial winds are less predictable[9], but the slowly varying component of the equatorial winds coupled with underlying sea surface temperatures (SST)[10] 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[11,12]. These studies imply that there is still room for improvement in ENSO prediction, although current methodologies might not be appropriate for this.

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[13]. 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

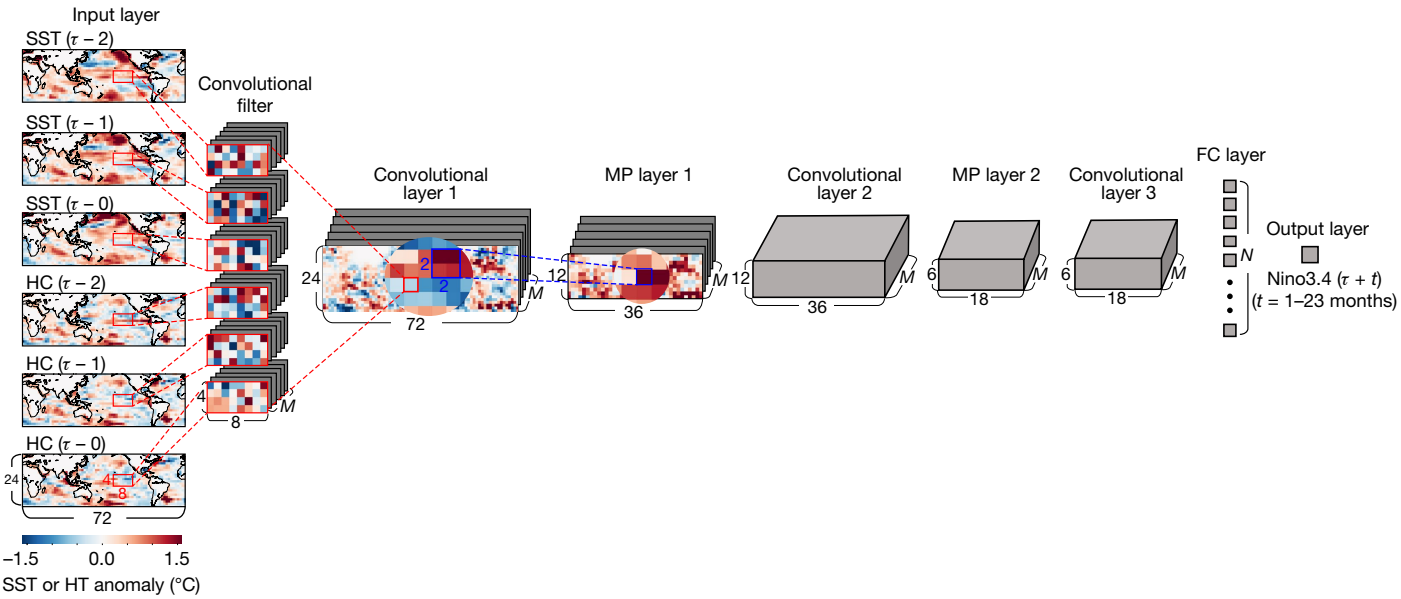


Fig. 1 | Architecture of the CNN model used for the ENSO forecasts. The CNN model consists of one input layer (the predictor), three convolutional layers, two max-pooling (MP) layers, one fully connected (FC) layer, and one output layer (the predictand). The variables of the input layer correspond to the SST (in units of °C) and the oceanic heat content (HC, in units of °C) anomaly maps from time $\tau - 2$ months to time τ (in months), between 0°–360° E and 55° S–60° 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[31]. The *x* and *y* dimensions of the map in each layer are denoted (6, 12, 18, 24, 36, 72) next to the map.

[1]Department of Oceanography, Chonnam National University, Gwangju, South Korea. [2]Institute for Climate and Application Research (ICAR)/CICFEM/KLME/ILCEC, Nanjing University of Information Science and Technology, Nanjing, China. [3]SKLLQG, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, China. *e-mail: ygham@jnu.ac.kr

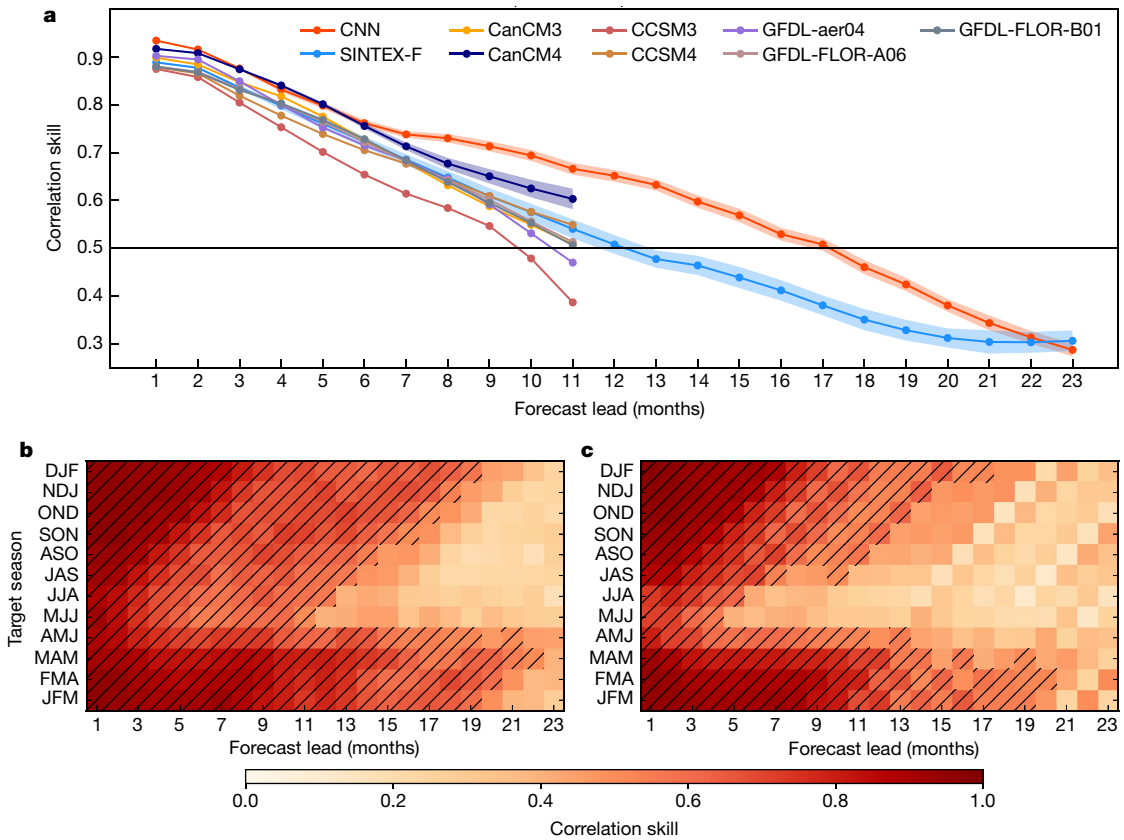


Fig. 2 | ENSO correlation skill in the CNN model. a, The all-season correlation skill of the three-month-moving-averaged Nino3.4 index as a function of the forecast lead month in the CNN model (red), SINTEX-F dynamical forecast system (blue), and dynamical forecast systems included in the North American 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.

colour images)[14,15]. 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.

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).

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 1871[16]. 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 extent[17] (Extended Data Table 1). Reanalysis data from 1871 to 1973 is also used to train the CNN model[16]. 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.

We apply the transfer learning technique[18] 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.

Figure 2a shows the all-season correlation skill of the three-month-moving-averaged Nino3.4 index from 1984 to 2017. The forecast skill of the Nino3.4 index in the CNN model is systematically superior to all state-of-the-art dynamical prediction systems at lead times longer than six months. 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-F[5], 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. 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.

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).

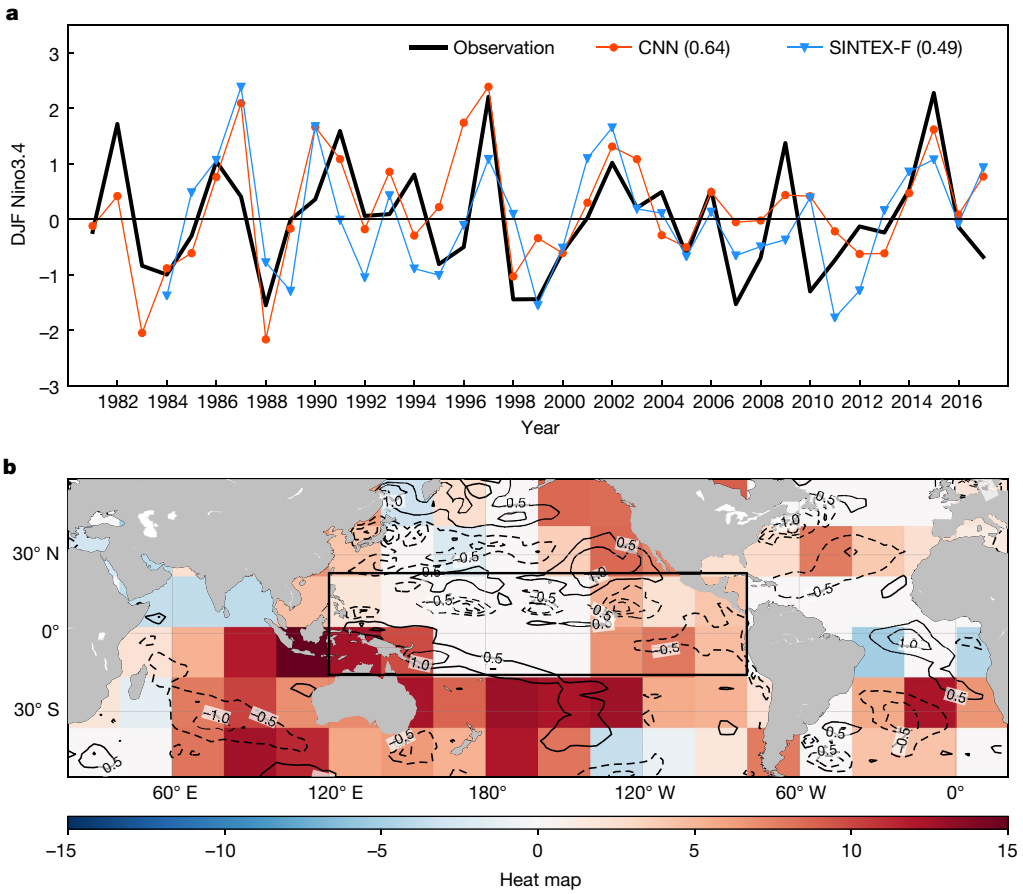


Fig. 3 | Physical interpretation of the CNN model ENSO forecasts. **a**, Time series of DJF season Nino3.4 indexes for an 18-month-lead forecast using the CNN model (red) and the SINTEX-F model (blue). The observed Nino3.4 index is also shown (black). **b**, The heat map (shading), and SST or heat content anomalies (contours; dashed contours denote the negative values of SST or HC anomalies and solid contours denote the positive values of SST or HC anomalies) for the MJJ season in 1996 used for the prediction of the DJF season during the 1997/98 El Niño event. Heat content anomalies (in units of °C) are shown over the tropical Pacific (within the black box), while SST anomalies (in units of °C) are denoted outside the tropical Pacific. Only the values with over 95% confidence level based on Student's *t*-test using the standard deviation of the heat map during 1984–2017 are shaded. The global map is generated in Matplotlib[31].

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[19].

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). To understand how the CNN model can successfully predict the Nino3.4 index for such long lead times, we produce a heat map[20] for the 18-month-lead forecast of the 1997/98 El Niño event (Fig. 3b). 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). 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[21], Indian Ocean[11] and subtropical Atlantic[12] (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[21]. A negative SST anomaly over the southwestern Indian Ocean contributes to a negative Indian Ocean Dipole (IOD) during the 1996 boreal autumn[11] (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[22]. 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[12] (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[23]. 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. In this model, the predictand corresponds to the percentage occurrence of three El Niño categories[24]: 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. 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[23]. 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.

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

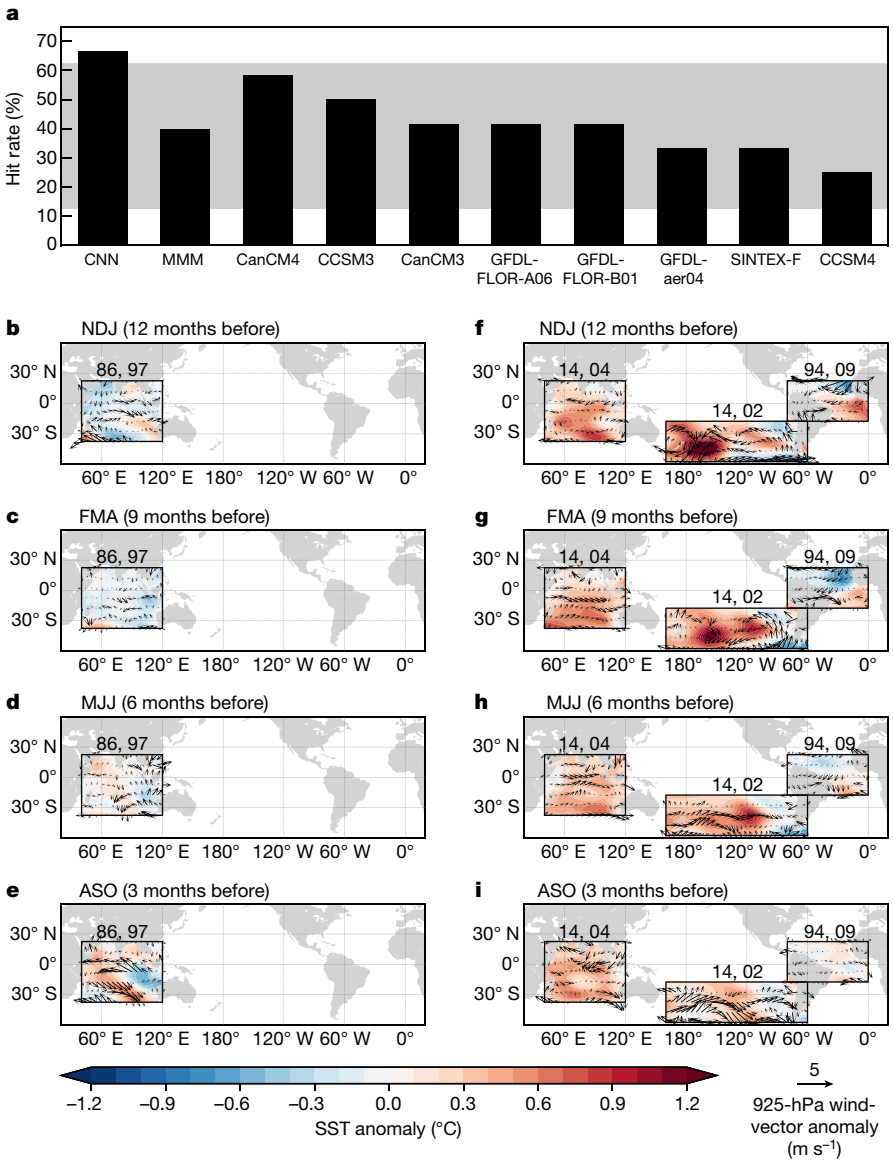


Fig. 4 | CNN forecast system for the prediction of El Niño type. a, The hit rate of the 12-month-lead prediction of El Niño types (EP-type, CP-type or mixed-type) during the DJF season using the CNN model and the SINTEX-F model, and models in the NMME project, for the validation period 1984–2017. The multi-model mean (MMM) of the hit rate is also shown. The grey shading denotes the 95% confidence interval of the random forecasts. **b**–**e**, The SST and 925-hPa wind-vector anomalies for the NDJ (**b**), FMA (**c**), MJJ (**d**) and ASO (**e**) seasons during the El Niño events when the area-averaged heat map value is greatest for the EP-type El Niño. **f**–**i**, As for panels **b**–**e**, but for the CP-type El Niño. The numbers above each ocean basin box denote the El Niño years selected for drawing the anomalies in each basin, and for each type of El Niño event. The anomalies are visible only when the area-averaged heat map value in any ocean basin exceeds one standard deviation for El Niño events for all ocean basins (see Extended Data Fig. 5 for details). The global map is generated in Matplotlib[31].

model is 66.7% during the validation period (1984–2017) (Fig. 4a and Extended Data Table 4). 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[25]. This would indicate that the deep-learning-based model can predict the spatial complexity of the El Niño events with great precision[16,26].

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). 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.

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. 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[27] (Fig. 4e). For the CP-type El Niño, SST cooling over the north tropical Atlantic is shown (Fig. 4f, g). Our results are consistent with those of previous studies[11,28]. The CP-type El 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 El Niño event (Extended Data Fig. 6). This indicates that the CNN can be a powerful tool to reveal complex ENSO mechanisms. 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[29]. 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[30]. 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1559-7.

Received: 20 November 2018; Accepted: 10 July 2019;
Published online 18 September 2019.

1. McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. ENSO as an integrating concept in Earth science. *Science* **314**, 1740–1745 (2006).
2. Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S. & DeWitt, D. G. Skill of real-time seasonal ENSO model predictions during 2002–11: is our capability increasing? *Bull. Am. Meteorol. Soc.* **93**, 631–651 (2012).
3. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
4. Cane, M. A., Zebiak, S. E. & Dolan, S. C. Experimental forecasts of El Niño. *Nature* **321**, 827–832 (1986).
5. Luo, J.-J., Masson, S., Behera, S. K. & Yamagata, T. Extended ENSO predictions using a fully coupled ocean–atmosphere model. *J. Clim.* **21**, 84–93 (2008).
6. Tang, Y. et al. Progress in ENSO prediction and predictability study. *Natl Sci. Rev.* **5**, 826–839 (2018).
7. Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E. & Huang, D. Predictability of El Niño over the past 148 years. *Nature* **428**, 733–736 (2004).
8. Gao, C. & Zhang, R. H. The roles of atmospheric wind and entrained water temperature (Te) in the second-year cooling of the 2010–12 La Niña event. *Clim. Dyn.* **48**, 597–617 (2017).
9. Hu, S. & Fedorov, A. V. Exceptionally strong easterly wind burst stalling El Niño of 2014. *Proc. Natl Acad. Sci. USA* **113**, 2005–2010 (2016).
10. Gebbie, G. & Tziperman, E. Predictability of SST-modulated westerly wind bursts. *J. Clim.* **22**, 3894–3909 (2009).
11. Izumo, T. et al. T. Influence of the state of the Indian Ocean Dipole on the following year's El Niño. *Nat. Geosci.* **3**, 168–172 (2010).
12. Park, J. H., Kug, J. S., Li, T. & Behera, S. K. Predicting El Niño beyond 1-year lead: effect of the Western Hemisphere warm pool. *Sci. Rep.* **8**, 14957 (2018).
13. LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
14. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012).
15. Oquab, M., Bottou, L., Laptev, I. & Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. In *Proc. IEEE Conf. on Computer Vision And Pattern Recognition* 1717–1724 (IEEE, 2014).
16. Giese, B. S. & Ray, S. El Niño variability in simple ocean data assimilation (SODA), 1871–2008. *J. Geophys. Res. Oceans* **116**, https://doi.org/10.1029/2010JC006695 (2011).
17. Bellenger, H., Guilyardi, É., Leloup, J., Lengaigne, M. & Vialard, J. ENSO representation in climate models: from CMIP3 to CMIP5. *Clim. Dyn.* **42**, 1999–2018 (2014).
18. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **27**, 3320–3328 (2014).
19. Webster, P. J. & Yang, S. Monsoon and ENSO: selectively interactive systems. *Q. J. R. Meteorol. Soc.* **118**, 877–926 (1992).
20. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2921–2929 (IEEE, 2016).
21. Anderson, B. T. On the joint role of subtropical atmospheric variability and equatorial subsurface heat content anomalies in initiating the onset of ENSO events. *J. Clim.* **20**, 1593–1599 (2007).
22. Kug, J. S. & Kang, I. S. Interactive feedback between ENSO and the Indian Ocean. *J. Clim.* **19**, 1784–1801 (2006).
23. Yeh, S. W. et al. El Niño in a changing climate. *Nature* **461**, 511 (2009).
24. Zhang, Z., Ren, B. & Zheng, J. A unified complex index to characterize two types of ENSO simultaneously. *Sci. Rep.* **9**, 8373 (2019).
25. Pillai, P. A. et al. How distinct are the two flavors of El Niño in retrospective forecasts of Climate Forecast System version 2 (CFSv2)? *Clim. Dyn.* **48**, 3829–3854 (2017).
26. Johnson, N. C. How many ENSO flavors can we distinguish? *J. Clim.* **26**, 4816–4827 (2013).
27. Webster, P. J., Moore, A. M., Loschnigg, J. P. & Leben, R. R. Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. *Nature* **401**, 356–360 (1999).
28. Ham, Y. G., Kug, J. S. & Park, J. Y. Two distinct roles of Atlantic SSTs in ENSO variability: north tropical Atlantic SST and Atlantic Niño. *Geophys. Res. Lett.* **40**, 4012–4017 (2013).
29. Zeiler, M. D. & Fergus, R. Visualizing and understanding convolutional networks. In *Eur. Conf. On Computer Vision* 818–833 (Springer, 2014)
30. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
31. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

METHODS

Architecture of the CNN model applied to ENSO forecasts. The CNN-based[32] 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 × 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[33]. 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.

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[34].

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

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

where P_i and Q_i denote the zonal and meridional dimensions of the convolutional filter for the ith convolutional layer, respectively. A hyperbolic tangent function (tanh) is used as the activation function. The dimensions of the convolutional filter were set at 8 × 4 during the first convolutional process (that is, $P_1 = 8$; $Q_1 = 4$), and at 4 × 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 mth feature map in the $(i-1)$th layer to the jth feature map in the ith convolutional layer. Additionally, $v_{(i-1),m}^{(x+p-P_i/2,\, y+q-Q_i/2)}$ denotes the value of the mth 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 jth feature map in the ith convolutional layer. To ensure that the horizontal dimension of the ith layer corresponds to the horizontal dimension of the $(i-1)$th layer, the empty spaces were filled with zeroes using a padding technique.

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 mth feature map in the last convolutional layer L to the nth neuron in the fully connected layer F), $v_{L,m}^{x,y}$ denotes the value of the mth feature map of the last convolutional layer L at grid point (x, y), $b_{F,n}$ denotes the bias of the nth neuron in the fully connected layer F; $W_{O,n}$ denotes the weight (used to link the nth 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:

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

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[1] of the forecast lead months l:

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

Here, Y and P denote the observed and the predicted values, respectively. $\overline{Y}_m$ and $\overline{P}_{m,l}$ 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.

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[24]. UCEI denotes a unified complex ENSO index.

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

where

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

and

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

N_3 and N_4 denote the Nino3 and Nino4 indices, respectively. The EP-type El Niño is defined for $\theta > 15°$ and $\theta < 90°$, while the CP-type El Niño is defined for $\theta > -90°$ and $\theta < -15°$. Finally, the mixed-type El Niño is defined for $\theta > -15°$ and $\theta < 15°$. 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.

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[35]. 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 2.2.4[16] 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)[36], while the horizontal wind vector at 925 hPa and the precipitation data were obtained from the ERA-Interim archives (1984–2017)[37]. Eight models of the North American Multimodel Ensemble phase 1 (1984–2017)[38] and SINTEX-F (1984–2017)[5,39] 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° × 2.5° for the analysis, and at a resolution of 5° × 5° for the training of the CNN model, in order to reduce the number of weighting coefficients.

Data availability

Data related to this paper can be downloaded from: SODA version 2.2.4, https://climatedataguide.ucar.edu/climate-data/soda-simple-ocean-data-assimilation; GODAS, https://www.esrl.noaa.gov/psd/data/gridded/data.godas.html; ERA-Interim, https://apps.ecmwf.int/datasets/data/interim-full-daily; NMME phase 1, https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/; and The CMIP5 database, https://esgf-node.llnl.gov/projects/cmip5/.

Code availability

TensorFlow (https://www.tensorflow.org) libraries were implemented to formulate the statistical forecast model using the CNN. The code for the CNN model can be downloaded at https://doi.org/10.5281/zenodo.3244463.

32. Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep Learning* (MIT Press, 2016).
33. Yoo, J. H. & Kang, I. S. Theoretical examination of a multi-model composite for seasonal prediction. *Geophys. Res. Lett.* **32**, L18707 (2005).
34. Kalchbrenner, N., Grefenstette, E. & Blunsom, P. A convolutional neural network for modelling sentences. In *Proc. 52nd Ann. Meet. Association for Computational Linguistics* 655–665 (Association for Computational Linguistics, 2014).
35. Wu, A., Hsieh, W. W. & Tang, B. Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Netw.* **19**, 145–154 (2006).
36. Behringer, D. W. & Xue, Y. Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. In *Proc. Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface* (AMS 84th Annual Meeting) (AMS, 2004).
37. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
38. Kirtman, B. P. et al. The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Am. Meteorol. Soc.* **95**, 585–601 (2014).
39. Luo, J.-J., Liu, G., Hendon, H., Alves, O. & Yamagata, T. Inter-basin sources for two-year predictability of the multi-year La Niña event in 2010–2012. *Sci. Rep.* **7**, 2276 (2017).

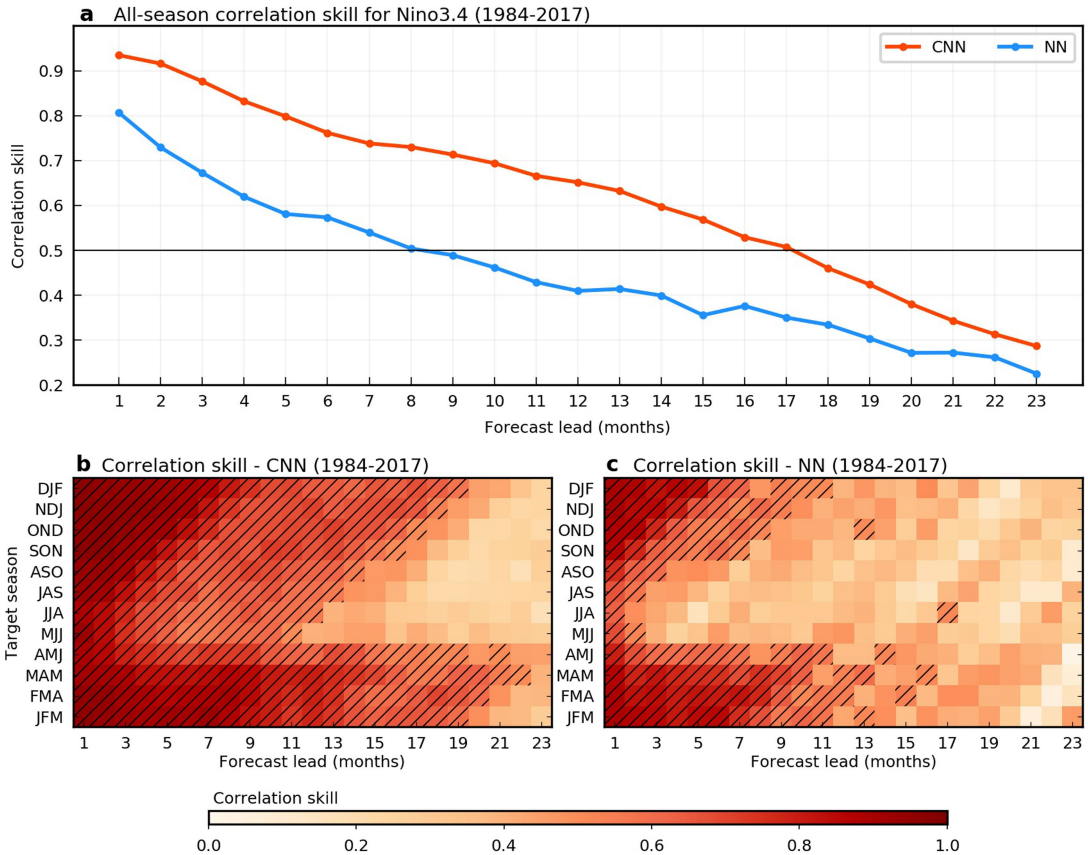
Acknowledgements This study is funded by the Korea Meteorological Administration Research and Development Program under grant KMI2018-03214. Y.-G.H. was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1A6A1A03012647). J.-J.L. is supported by 'The Startup Foundation for Introducing Talent' of NUIST. We are grateful to W. Merryfield for providing comments and to T. Doi for providing part of the SINTEX-F hindcast data used in the validation.

Author contributions Y.-G.H. and J.-H.K. designed the experiments and conducted the analyses. Y.-G.H. wrote most of the manuscript. J.-H.K. and Y.-G.H. performed the CNN hindcast experiments. J.-J.L. conducted the SINTEX-F hindcast experiments and reported the results. All authors discussed the study results and reviewed the manuscript.

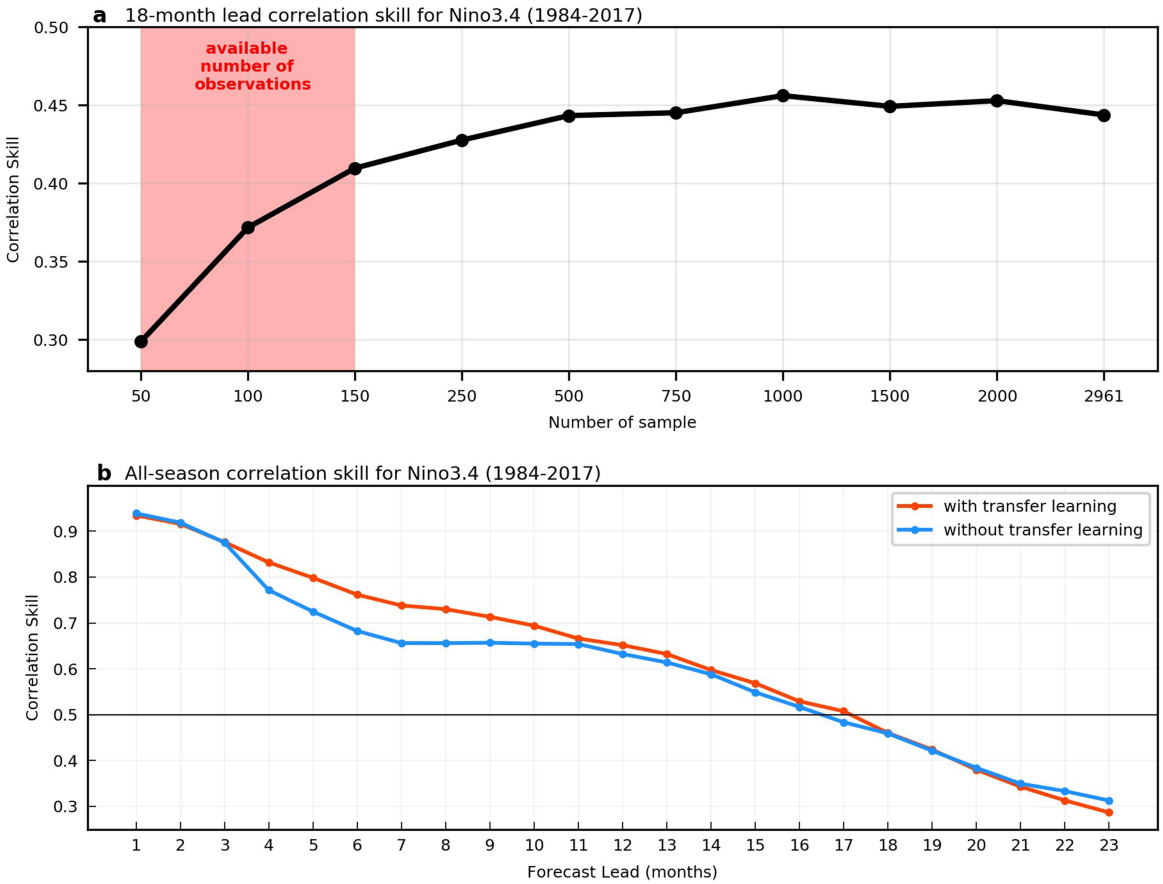
Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1559-7.

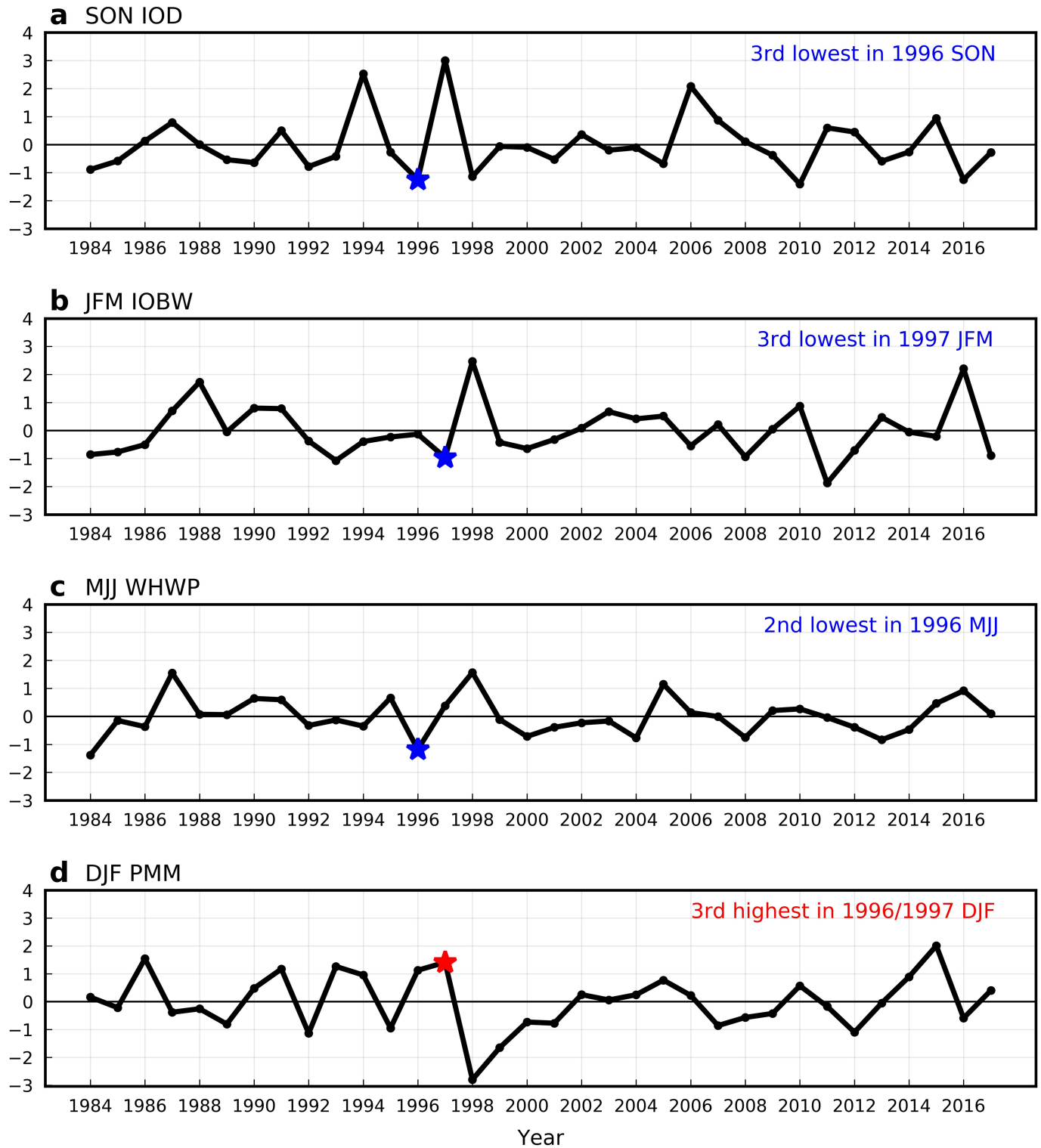
Correspondence and requests for materials should be addressed to Y.-G.H.



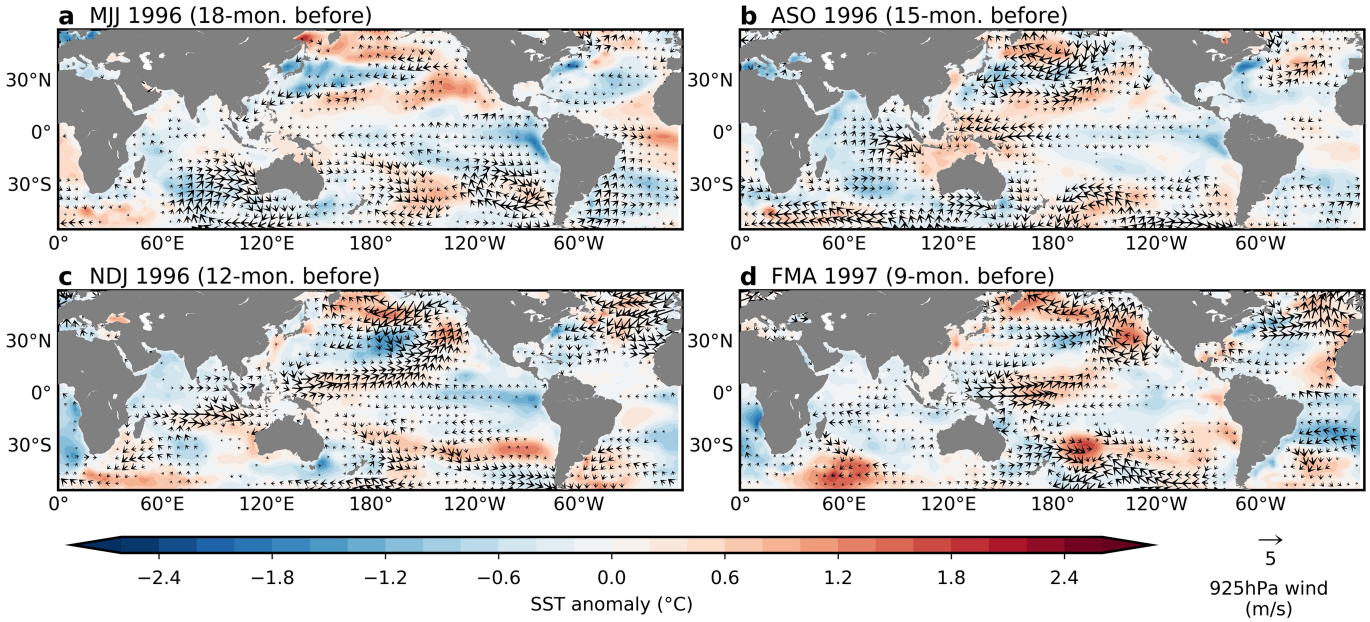
Extended Data Fig. 1 | Comparison of the ENSO correlation skill between the CNN model and the feed-forward neural network model. **a**, The all-season correlation skill of the three-month-moving-averaged Nino3.4 index as a function of the lead months of the forecast in the CNN model (red) and the feed-forward neural network model (blue). The validation period is between 1984 and 2017. **b**, **c**, The correlation skill of the Nino3.4 index targeted to each calendar month in the CNN model (**b**) and in the feed-forward neural network model (**c**). Hatching highlights the forecasts with correlation skill exceeding 0.5.



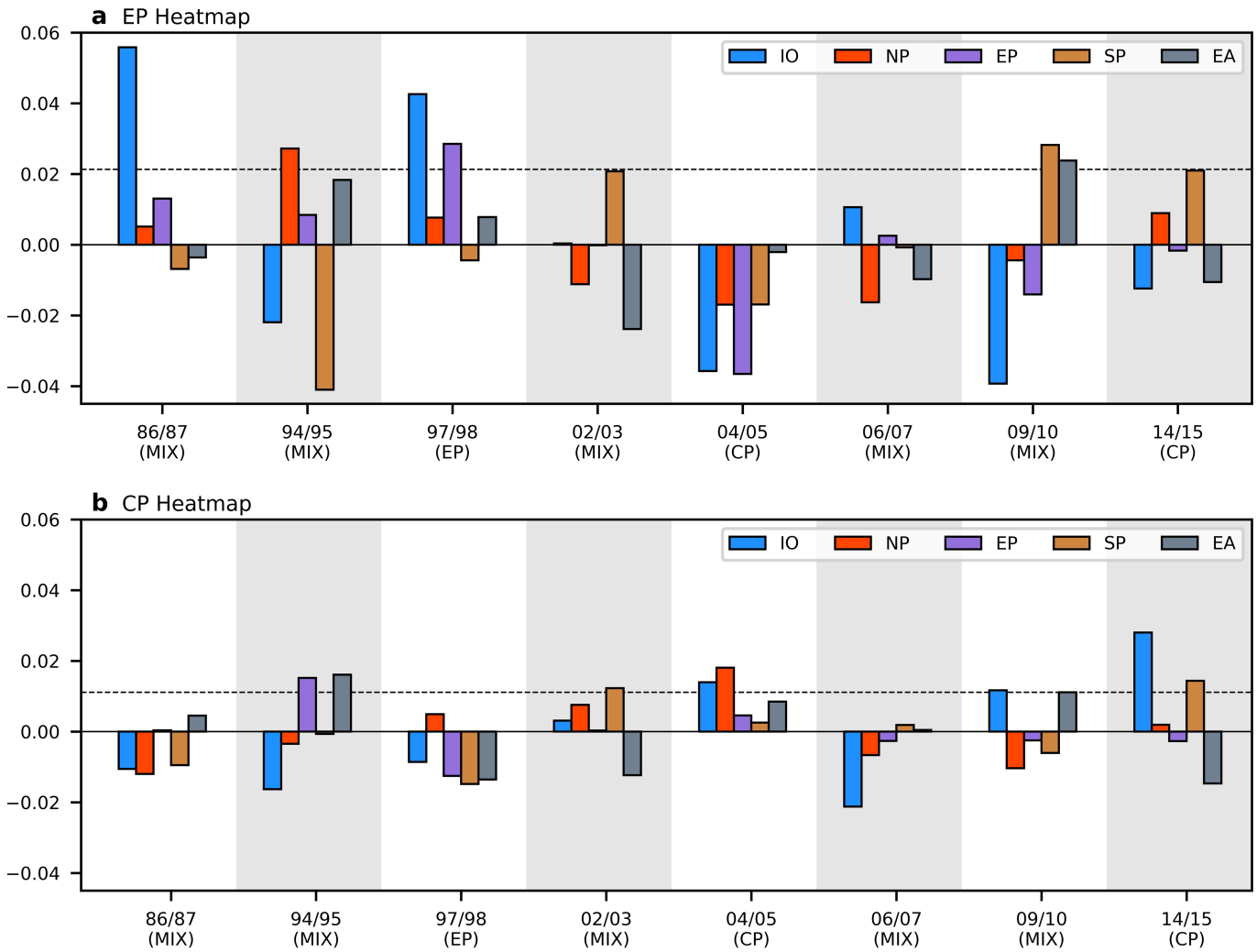
Extended Data Fig. 2 | The improvement in the skill of the CNN model due to the CMIP5 dataset and the transfer learning. a, The all-season correlation skill of Nino3.4 index at a lead of 18 months with a different number of CMIP5 samples. The red area denotes the number of available observed samples. We note that transfer learning is not applied to the series of sensitivity tests (that is, observations during the training period are not used to set up the CNN model). **b**, The all-season Nino3.4 correlation skill as a function of the lead months of the forecast with and without the transfer learning. The CNN model without the transfer learning is formulated by using all the CMIP5 and the observed samples during the training period (that is, 1871 to 1973) in a single training period. Therefore, the number of samples for the CNN model without the transfer learning is exactly the same with those with the transfer learning.



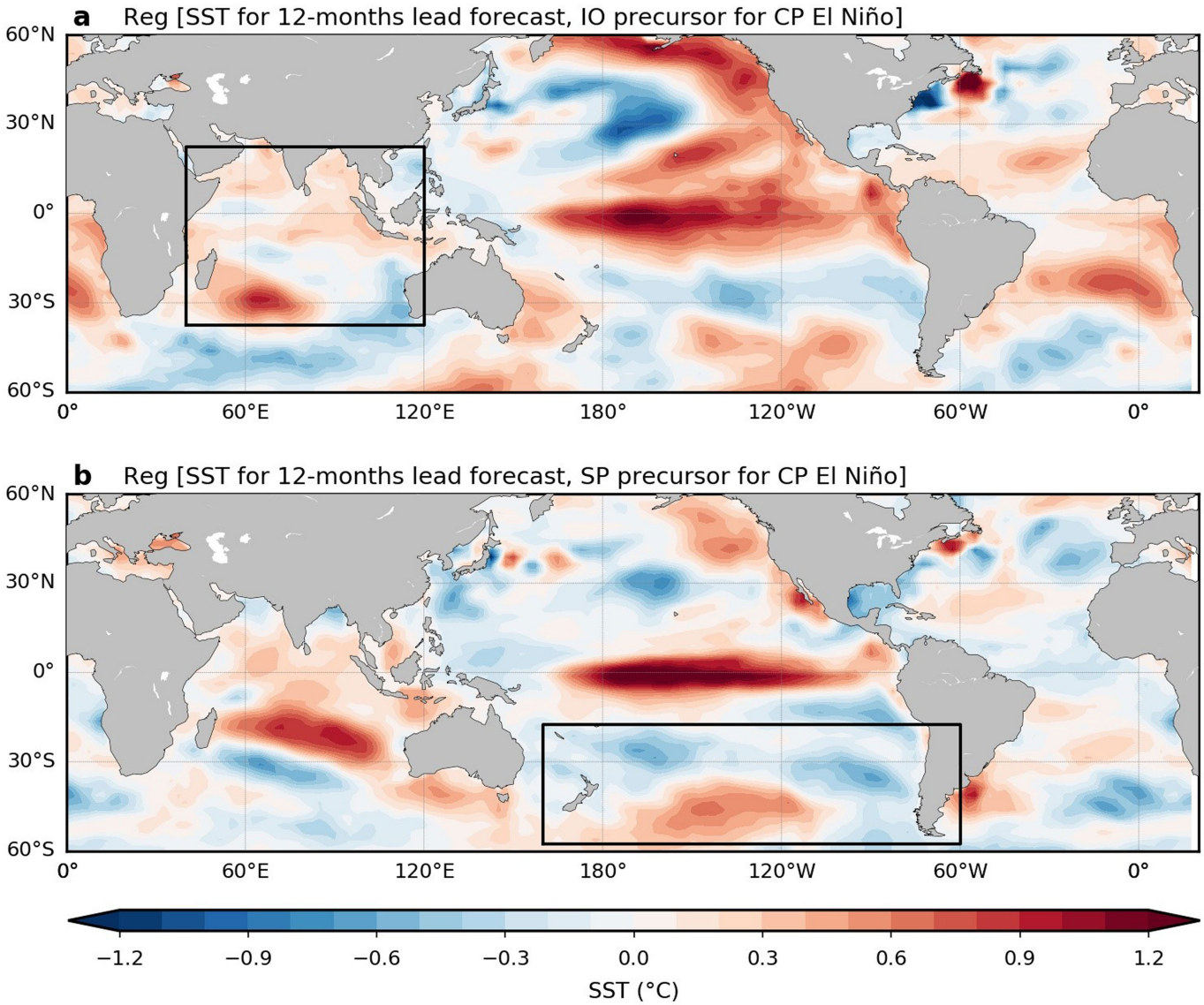
Extended Data Fig. 3 | The time series of climate indices. a–d, The time series of the IOD index (difference of the area-averaged SST over 50–70° E, 10° S–10° N from that over 90–110° E, 15°–0° S) during the SON season (**a**), the Indian Ocean Basin-wide warming (IOBW) index (area-averaged SST over 40–110° E, 15° S–10° N) during the JFM season (**b**), the Western Hemispheric Warm Pool (WHWP) index (area-averaged SST over 60–105° E, 10–35° N) during the MJJ season (**c**), and the Pacific Meridional Mode (PMM) index (first Maximum Covariance Analysis (MCA) principal components over the 175° E–95° W, 21° S–32° N using SST and 10-m winds) during the DJF season (**d**). The values preceding the 1997/98 El Niño event are denoted by the red star when the value is positive and the blue star when the value is negative.



Extended Data Fig. 4 | The time evolution of the 1997/98 El Niño event. The SST (shading) and 850-hPa wind vector (vectors) at: **a**, MJJ 1996; **b**, ASO 1996; **c**, NDJ 1996; and **d**, FMA 1997. The global map is generated in Matplotlib[31].



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.



Extended Data Fig. 6 | The SST pattern developed by precursors for CP El Niño. a, **b**, The SST anomalies for forecasts with leads of 12 months regressed onto the pattern regression index for CP-type El Niño precursors over the Indian Ocean (**a**) and the south Pacific (**b**) at the NDJ season. The pattern regression index for CP-type El Niño precursors is obtained by calculating the pattern regression of the NDJ SST and heat content anomalies onto the selected anomaly for CP El Niño event in Fig. 4f. The black box denotes the region in which the pattern regression index was calculated. We note that both regressed SST patterns are classified as the CP-type El Niño[24]. The global map is generated in Matplotlib[31].

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	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)	DEC1859 - DEC2005	1
HadGEM2-CC		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	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.

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.

Extended Data Table 3 | The skill of the feed-forward neural network model with a different number of predictors

	Atlantic, North Pacific					
Indian-Pacific	Number of EOF PCs	5	6	7	8	9
	5	0.43	0.44	0.50	0.41	0.41
	6	0.42	0.43	0.50	0.41	0.40
	7	0.43	0.41	0.49	0.40	0.38
	8	0.42	0.41	0.48	0.39	0.38
	9	0.46	0.47	0.52	0.46	0.45
	10	0.21	0.30	0.38	0.23	0.39

The correlation skill of the Nino3.4 index for the DJF season with an 18-month lead in the feed-forward neural network model with a different number of EOF principal components (PCs) as predictors. The validation period is from 1984 to 2017. The correlation skill of the CNN targeting for the DJF Nino3.4 index for an 18-month lead is 0.64.

Extended Data Table 4 | 12-month lead forecast results of predicting the type of El Niño

Year	OBS	CNN	SINTEX-F	CanCM3	CanCM4	CCSM3	CCSM4	GFDL-aer04	GFDL-FLOR-A06	GFDL-FLOR-B01
1976	EP	EP	-	-	-	-	-	-	-	-
1977	CP	CP	-	-	-	-	-	-	-	-
1979	MIX	MIX	-	-	-	-	-	-	-	-
1982	EP	EP	-	-	-	-	-	-	-	-
1986	MIX	MIX	MIX	MIX	MIX	MIX	MIX	EP	CP	CP
1987	MIX	CP	MIX	EP	EP	EP	CP	EP	EP	EP
1990	CP	MIX	MIX	MIX	MIX	CP	MIX	MIX	MIX	MIX
1991	MIX	CP	CP	MIX	MIX	MIX	EP	EP	EP	EP
1994	MIX	MIX	EP	EP	EP	EP	EP	EP	EP	EP
1997	EP	EP	MIX	MIX	MIX	MIX	MIX	MIX	MIX	MIX
2002	MIX	MIX	MIX	EP	MIX	MIX	EP	MIX	MIX	MIX
2004	CP	CP	MIX	EP	CP	EP	EP	EP	CP	CP
2006	MIX	MIX	MIX	MIX	MIX	MIX	MIX	MIX	MIX	MIX
2009	MIX	MIX	EP	MIX	MIX	MIX	MIX	MIX	MIX	MIX
2014	CP	CP	MIX	MIX	MIX	MIX	MIX	CP	MIX	MIX
2015	MIX	CP	CP	MIX	MIX	EP	EP	CP	MIX	MIX
Hit rate (%)	-	66.67 (75.00)	33.33	41.67	58.33	50	25	33.33	41.67	41.67

The prediction result of a 12-month-lead forecast of El Niño type during the DJF season for EP-type, CP-type and mixed-type El Niño events in the CNN and dynamical forecast systems. The hit rate during the validation period (from 1984 to 2017) is denoted in the bottom line. Values in parentheses denote the hit rate of the CNN from 1976 to 2017. Green shading shows when the forecast is correct.