ARTICLE   OPEN

Check for updates

VARENN: graphical representation of periodic data and application to climate studies

Takeshi Ise [1 ✉] and Yurika Oba [1,2]

Analyzing and utilizing spatiotemporal big data are essential for studies concerning climate change. However, such data are not fully integrated into climate models owing to limitations in statistical frameworks. Herein, we employ VARENN (visually augmented representation of environment for neural networks) to efficiently summarize monthly observations of climate data for 1901–2016 into two-dimensional graphical images. Using red, green, and blue channels of color images, three different variables are simultaneously represented in a single image. For global datasets, models were trained via convolutional neural networks. These models successfully classified the rises and falls in temperature and precipitation. Moreover, similarities between the input and target variables were observed to have a significant effect on model accuracy. The input variables had both seasonal and interannual variations, whose importance was quantified for model efficacy. We successfully illustrated the importance of short-term (monthly) fluctuations in the model accuracy, suggesting that our AI-based approach grasped some previously unknown patterns that are indicators of succeeding climate trends. VARENN is thus an effective method to summarize spatiotemporal data objectively and accurately.

INTRODUCTION

In the era of big data, spatiotemporal digital data, such as satellite observations, have been continuously accumulating[1]. Such data should be used to better understand and predict the behavior of the Earth system: from ecosystem to the climate[2]. However, these data are not fully utilized currently, possibly owing to the lack of a suitable framework to extract important information from such data[1,3,4]. Scientists work eagerly to improve simulation models for climatic prediction, but the performance of the models has not be sufficient. Specifically, interannual-to-decadal climatic patterns, such as ENSO (El Nino/southern oscillation), NAO (North Atlantic oscillation), and PDO (Pacific decadal oscillation), are not well reproduced by current climate models (i.e., Earth system models), indicating that the models are not fully optimized with the available data[5,6], especially for decadal teleconnection patterns.

Deep learning has been used in diverse topics recently. For example, convolutional neural networks (CNNs) have been a powerful tool in computer vision for object identification and classification in various fields[7]. Related to Earth system science, computer vision has been applied for satellite-based images[8,9] and field-based pictures[4]. Using visual images, a CNN model successfully detected extreme weather conditions[10]. However, these studies mainly use two-dimensional image data. Because targets for Earth system modeling tend to have a dimension in time, a newer approach that explicitly integrates temporal dynamics is required[1]. For instance, convolutional long short-term memory (LSTM)[11] is one of the methods to explicitly treat spatiotemporal data. Earth system data are characterized by the large size (i.e., big data) with complex spatiotemporal interactions, such as seasonal and interannual climate variability. Ise and Oba[12] approached such data by converting global temperature data into artificial two-dimensional images, and demonstrated that the resultant model successfully summarized temporal characteristics globally, as a classification problem. We believe that this approach

should be augmented and used with various data from atmospheric, oceanic, and terrestrial ecosystems.

In this study, we propose a data-processing method called visually augmented representation of environment for neural networks (VARENN) and apply it to global time series of eight climatic variables from 1901 to 2016 to construct a classification model via supervised learning with a CNN. A few previous studies have reported methods of converting time-series data into images, and performed analyses using CNNs, but they only utilized one variable[12,13]. Here, by using three (red, green, and blue) channels of digital images, we integrate three different time-series variables into a single image. In doing this, we aim to classify combination patterns of decadal climatic variables. We designed a series of experiments to demonstrate that VARENN is an effective method to integrate environmental variables into common CNN frameworks. We also quantified the importance of short-term (monthly) climate fluctuations in the CNN model construction to illustrate that some specific seasonal patterns can be indicators of succeeding climate trends.

RESULTS

VARENN is designed to convert time-series data containing both seasonal and interannual variabilities into two-dimensional images with artificial color assignment. Because a color image is composed of two-dimensional arrays of three color channels, it is possible to assign three different time series to such arrays. This facilitates the analysis of multiple global time-series data integratively via CNN. In this study, using multiple climatic variables for global terrestrial regions, we attempted to construct models to classify future rises and falls in temperature and precipitation, according to information from the multiple time series.

[1]Field Science Education and Research Center (FSERC), Kyoto University, Kyoto, Japan. [2]Center for the Promotion of Interdisciplinary Education and Research, Kyoto, Japan. ✉email: ise@kais.kyoto-u.ac.jp

Table 1. Eight climatic variables used in this study, from CRU TS 4.01. All data are prepared as monthly datasets for 1901–2016, and the spatial resolution is 0.5°.

Label	Variable	Units
cld	Cloud cover	%
dtr	Diurnal temperature range	°C
frs	Frost-day frequency	days
pet	Potential evapotranspiration	$\mathrm{mm\,d^{-1}}$
pre	Precipitation	$\mathrm{mm\,mo^{-1}}$
tmp	Daily mean temperature	°C
vap	Vapor pressure	hPa
wet	Wet-day frequency	days

In this study, eight spatiotemporal climatic variables were obtained from the Climatic Research Unit Time-Series v. 4.01[14,15] (CRU TS4.01, Table 1). CRU TS4.01 is a high-resolution gridded dataset for terrestrial areas divided into 0.5° grids. It contains monthly climatic data for the period of 1901–2016. From the eight climatic variables, we systematically selected up to three variables and assigned them to color channels (red, green, and blue (RGB)) to construct artificial color images that conveyed varied climatic information.

To visualize seasonal and interannual variations and trends, a 30-year window (training period: duration of time where we obtained data for VARENN images) was systematically selected from the 116 years of the dataset. We created graphical images of 60×60 pixels from the 30-year window data (Fig. 1). The climatic variables were scaled to 0–1 to fit the color spaces. Monthly data for a given year were vertically aligned from the top, and the yearly data were horizontally aligned from the left.

To systematically analyze the model performance and find the optimal combination of climatic variables, 92 combinatorial experiments were executed (Supplementary Table 1): eight models used one climatic variable (1-VAR experiments), 28 models used two climatic variables (2-VAR experiments), and 56 models used three climatic variables (3-VAR experiments). This combinatorial analysis was conducted independently for temperature change experiments (TMP-EX) and precipitation change experiments (PRE-EX). In contrast to the previous study with one variable[12], here we superimpose up to three different climatic variables into color (RGB) channels. For 1-VAR models, the selected climatic variable was allocated to the R channel. For 2-VAR models, the R and G channels were used. The vacant channel(s) was/were filled with zeroes. For 3-VAR models, the R, G, and B channels were all used.

The images are annotated with five categories based on decadal climatic trends according to the following relationships:

$$F_T = \begin{cases} T_1, & 5 \le \mu_{10} - \mu_{30} \\ T_2, & 2.5 \le \mu_{10} - \mu_{30} < 5 \\ T_3, & 0 \le \mu_{10} - \mu_{30} < 2.5 \\ T_4, & -2.5 \le \mu_{10} - \mu_{30} < 0 \\ T_5, & \mu_{10} - \mu_{30} < -2.5 \end{cases} \quad (1)$$

$$F_P = \begin{cases} P_1, & 30 \le \mu_{10} - \mu_{30} \\ P_2, & 10 \le \mu_{10} - \mu_{30} < 30 \\ P_3, & -10 \le \mu_{10} - \mu_{30} < 10 \\ P_4, & -30 \le \mu_{10} - \mu_{30} < -10 \\ P_5, & \mu_{10} - \mu_{30} < -30 \end{cases} \quad (2)$$

where F_T and F_P are the classification categories for TMP-EX and PRE-EX, respectively, μ_{30} is the mean of the target variable (tmp or

pre) for the 30-year training period, and μ_{10} is the mean of the target variable for the labeling period (a 10-year window following the training period). For TMP-EX, the unit is °C. For PRE-EX, the unit is mm. F_T or F_P categories were assigned as annotation for supervised learning.

We conducted 92 combinatorial experiments for both TMP-EX and PRE-EX (Fig. 2, Supplementary Table 1) with randomly chosen images ~100,000. The test accuracy tended to be high when the target variable (tmp or pre) was in the input variables. The test accuracy considerably varied among experiments with different input variables. For instance, in experiments for 3-VAR TMP-EX, the highest test accuracy (0.829) was achieved with pet, tmp, and vap, whereas the lowest accuracy (0.739) was achieved with frs, pre, and wet.

Using the nonparametric Kruskal–Wallis test, we observe that there are significant differences among the accuracies of the 1-VAR, 2-VAR, and 3-VAR experiments ($p < 0.05$) for both TMP-EX and PRE-EX. The result shows that the number of climatic variables had an obvious effect on the accuracy of the classification model. Then, with the Mann–Whitney U test, we check the differences in test accuracies for each pair of the three experiments. In TMP-EX, the accuracy differences of 1-VAR and 3-VAR were significant ($p < 0.05$). In PRE-EX, the accuracy differences of 1-VAR, 2-VAR, and 3-VAR were significant ($p < 0.05$).

Because the eight climatic variables used in this study have varying levels of similarity (Supplementary Fig. 1), we examine the effects of the relationships between the target variables (i.e., tmp or pre) and input variables (i.e., cld, dtr, frs, pet, pre, tmp, vap, and wet). To quantify the similarity among variables, we calculated the Euclidean distances from each target variable to the input variable. Then, we analyzed the relationship between the similarity of variables and test accuracy by using linear regression. As a result, we found that the accuracy of models with variables that are closely related to the target tends to be higher. The regression lines in Fig. 2 are all statistically significant ($p < 0.05$), except 1-VAR PRE-EX. The eight climatic variables used in this study have varied levels of similarity (Supplementary Fig. 1), and we examine the effects of the relationships among the variables that were assigned to R, G, and B to the model performance.

The best model in 3-VAR was the combination of pet, tmp, and vap for TMP-EX (#86 in Supplementary Table 1), and the combination of frs, pre, and wet for PRE-EX (#48). To maximize the performance, we increased the number of training images to ~1.5 million. For these high-performance models of TMP-EX #86 and PRE-EX #48, the resultant test accuracies were 0.937 and 0.828, and the weighted Kappa statistics were 0.959 and 0.867, respectively. These results suggest that the VARENN framework is scalable; the greater the amount of training data, the higher the accuracy. Using these high-performance models, we visualized the model classifications and errors (Fig. 3). Basically, the models were able to assign correct categories globally. However, classification errors were concentrated on the border of F_T or F_P categories. This is reasonable because VARENN images of border grids tend to contain characteristics of both categories across the border. It should be noted that the partitioning of categories is arbitrary; if classification schemes other than F_T or F_P were used, the erroneous borders would change positions. Comparing TMP-EX and PRE-EX, it should be noted that the distribution patterns of five categories are more complex for the latter. This would increase the length of borders and result in lower accuracies found in PRE-EX than TMP-EX.

To check if seasonal and/or interannual variations are critical for constructing models with high accuracies, we designed "knock-out" experiments, in which we intentionally "inactivate" seasonal or interannual variations (Table 2). When we remove interannual variations (i.e., monthly averaged climatic variables for 30 years), the VARENN images resemble squares with horizontal stripes. On the other hand, when we remove seasonal variations (i.e., annual
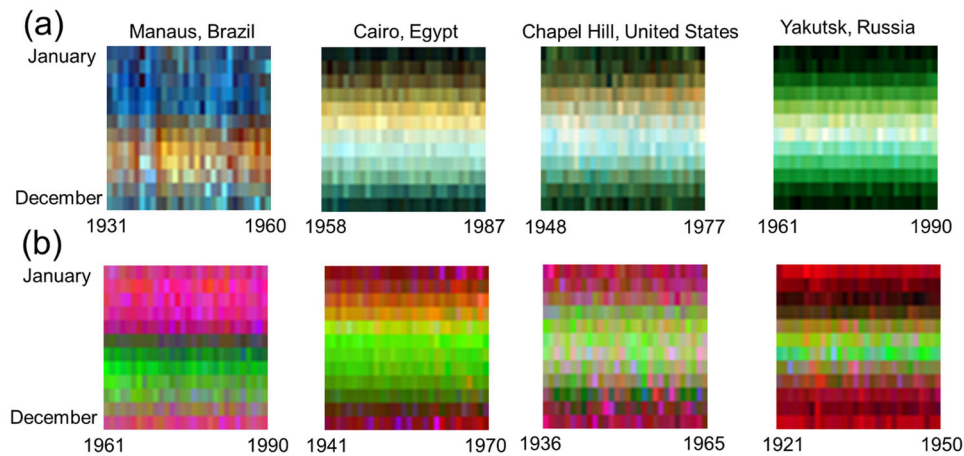
Fig. 1 Examples of VARENN images. a TMP-EX with pet, tmp, and vap assigned to R, G, and B channels, respectively, and **b** for PRE-EX with cld, pet, and preassigned to R, G, and B channels, respectively.
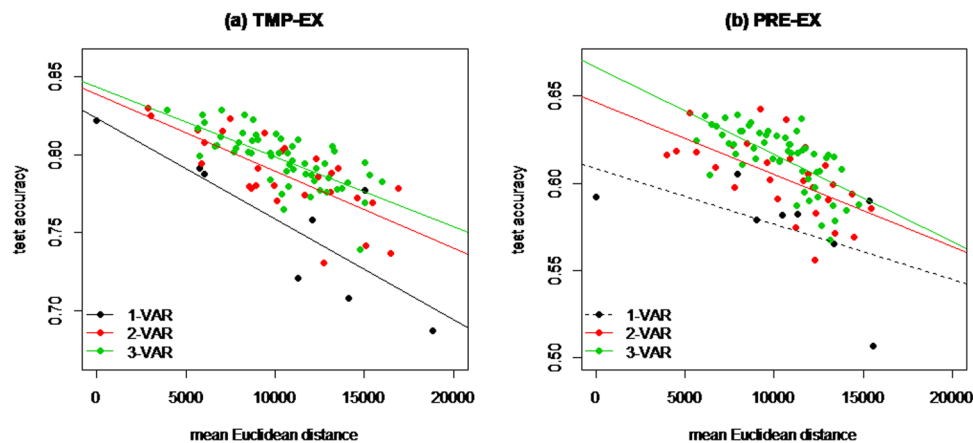

Fig. 2 Plots to compare performances of 1-VAR, 2-VAR, and 3-VAR experiments. a Summary for TMP-EX. **b** Summary for PRE-EX.

average climatic variables), the images have vertical stripes. When we construct CNN models with these images, the reduction in information due to averaging lowered test accuracies. This experiment strongly suggests that both seasonal and interannual variabilities are critical for constructing models that appropriately classify climatic patterns. It is particularly interesting that the data in seasonal patterns added more accuracy to the CNN models. This implies that particular short-term (monthly) climatic fluctuations could be important predictors of climate change.

In addition, to quantify the effect of the length of the training period, we carried out experiments with training period of 10 years and compared test accuracies with the 30-year experiments (Table 2). The reduction in the duration of the training period lowered model accuracy considerably, indicating that the multi-decadal climate patterns in the 30-year training period convey information to classify climatic patterns more accurately.

DISCUSSION

In the series of experiments, we successfully illustrated that VARENN can effectively represent global time-series data of climate, with seasonal and interannual variations. Applying this method to gridded data, we were able to classify climatic patterns in the time series. We examined 92 combinations of variables for TMP-EX and PRE-EX. As a result, the number of climatic variables had a clear effect on the accuracy of the classification model; the greater the number of variables embedded in training images, the higher the test accuracy obtained. We also found that the model

performance is related to the similarity between the target and input variables.

Our approach, VARENN, has some similarities with convolutional LSTM[11]. Both approaches are designed to explicitly treat spatiotemporal data. However, the former has several unique characteristics: (1) two time-series trends (i.e., seasonal and interannual) are graphically represented, (2) arbitrary combinations of multiple variables are accepted, and (3) in the future, an application for guided Grad-CAM[16] would help researchers find particular signatures concerning climatic trends. Thus, VARENN can be an effective analysis tool for environmental and earth sciences. On the other hand, convolutional LSTM explicitly captures spatial information, which is not fully considered in VARENN. A comparison study of these methods is needed.

The conversion of times-series signals into artificial, two-dimensional images to detect anomalies using CNN has been suggested previously[14]. However, the signals used in that study did not have well-defined periodicity, such as seasonality, and thus the dimensions of the image were set arbitrarily. In this study, in contrast, we have set the dimensions according to the explicit seasonality, and thus VARENN images exhibit clear patterns. Superimposed multiple signals in RGB color channels are another unique characteristic of VARENN. Thus, we believe that our approach is effective, especially for large-scale spatiotemporal environmental data.

VARENN-based models successfully highlighted the importance of short-term (monthly) fluctuations in classification of decadal climate. This implies that the CNN models constructed by VARENN
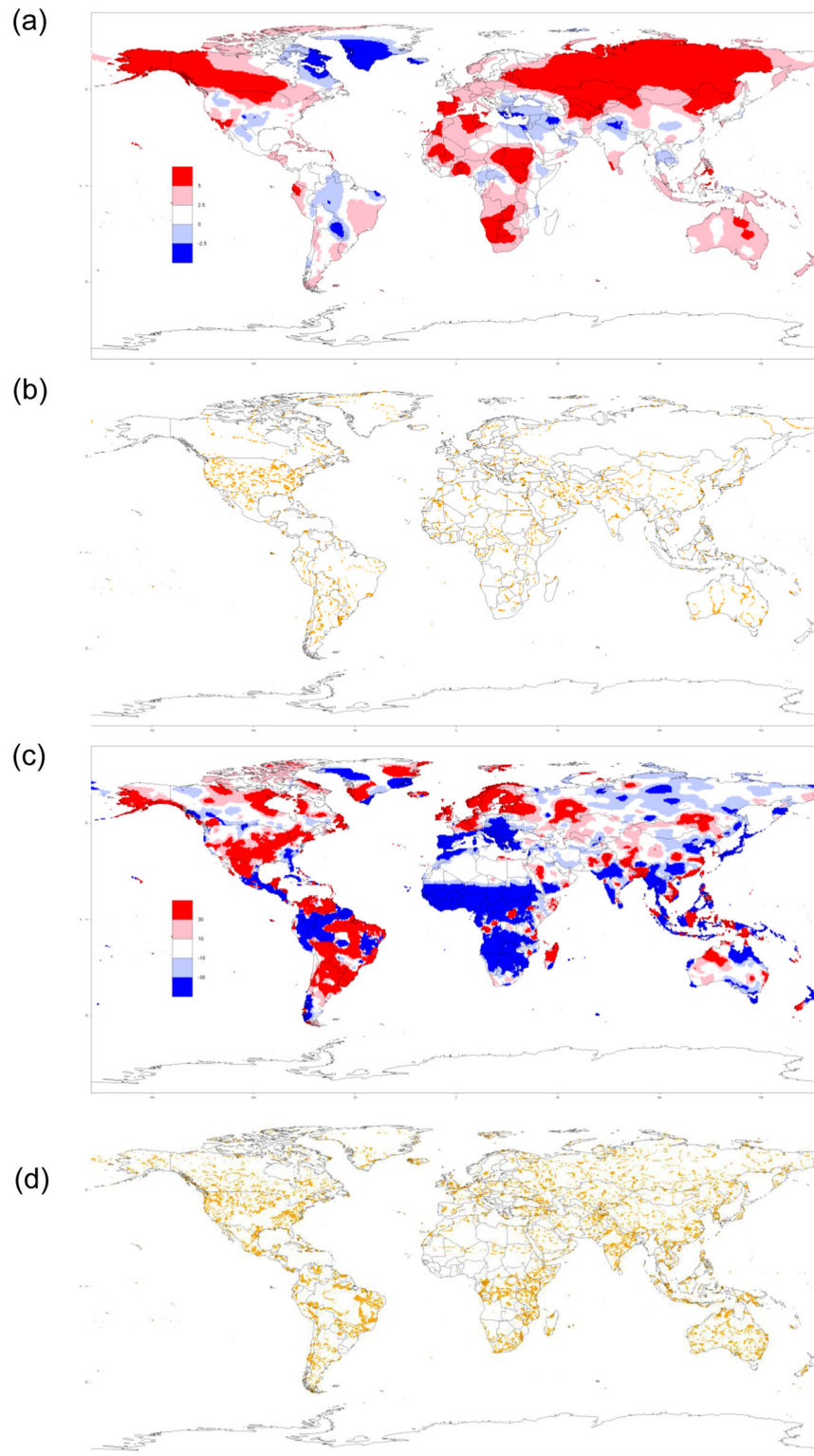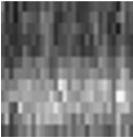
Fig. 3 Visualization of classification by high-performance models. a TMP-EX #86. **b** PRE-EX #48. **c** Classification errors (indicated in orange) for TMP-EX #86. **d** Classification errors for PRE-EX #48.

have grasped some temporal signatures in short-term climate that are an indicator of the future climate change. Since the AI-based approach has the capacity to discern patterns that are previously unknown to researchers, these findings of this study have important implications for other research areas, such as physics-based climate modeling.

Process-based simulation models will remain as the core for projections of the climate because statistical methods, including AI, have limited ability to predict the future in novel conditions where it is difficult to obtain training data. We suggest using AI-based methods such as VARENN to improve process-based models. For example, in this study, the model successfully

Table 2. Test accuracy of models of "knockout" experiments of temporal variations and "shortened" experiments (sample images here have been converted to greyscale).

Experiments	Sample images	Test accuracies	
		TMP-EX#86	PRE-EX#48
Default (with both seasonal and interannual variations, 30-year training)		0.829	0.638
Seasonal variations only		0.569	0.489
Interannual variations only		0.766	0.517
10-year training		0.742	0.562

classified interannual-to-decadal climatic patterns. We selected the window size (30 years for training and 10 years for labeling) to capture the temperature and precipitation patterns occurring in the timescale of teleconnection patterns such as ENSO (Supplementary Fig. 2). By comparing the results of AI-based (statistical) studies and process-based (physical) simulations, the accuracy and precision of the simulations can be evaluated appropriately. Objective parameterization, such as data assimilation of process-based simulation models, is suggested[3,17], but feature extraction with deep learning is unique because it proposes the model itself, not just parameters, objectively.

The VARENN approach aims to find climatic patterns as a classification problem. We believe that our approach has successfully captured multidecadal climatic signatures because the experiments with the 30-year training period had better performance than 10-year experiments. Our study should be augmented in the near future. For example, by using a visualization technique, it is possible to obtain visual explanations for the classification done by the model. By doing this, we may determine which part of the image is important for the AI decision, whether such decisions differ among different climatic regions, how much time lag is considered, and if a pattern can be identified for extreme weather events. In summary, the method to analyze spatiotemporal big data will be a promising tool for improving studies and projections concerning climate change.

METHODS

In this study, a supervised learning system with CNN was constructed with an emphasis on cost performance and easy operation. The hardware used was a consumer-level PC with Intel Core i7-9700K CPU, 32-GB RAM, and NVIDIA GeForce RTX 2080 Ti GPU, and the operating system was Ubuntu 16.04 LTS. The backend used for the CNN was TensorFlow 1.12.0, implemented on NVIDIA DIGITS 6.1[18] (provided as NVIDIA GPU Cloud image DIGITS Release 18.12). We chose DIGITS 6.1 for its easy operation, wherein it allowed adjusting most parameter settings with a graphical user

interface. We employed a CNN with LeNet[19] (Supplementary Fig. 3). The number of training epochs was 30, and the solver type was stochastic gradient descent. In preliminary experiments with our supervised training framework with DIGITS 6.1, we found that training images in a square shape significantly reduce computational time. Thus, we manually constructed all images in 60×60 squares to maximize efficiency. One of the advantages of NVIDIA DIGITS 6.1 was the easy implementation of gradually changing learning rates, which was utilized when we constructed the high-performance models with ~1.5 million images. In our environment, the time consumed for learning was ~5 min for ~100,000 images and ~1 h for ~1.5 million images. With the test datasets, we constructed confusion matrices and calculated the test accuracy. We also calculated the weighted Kappa statistic for these high-performance models. The relationships among 1-VAR, 2-VAR, and 3-VAR experiments were analyzed by nonparametric significance test of Kruskal–Wallis test and Mann–Whitney U test with Bonferroni adjustment. VARENN image generation and statistical analyses were performed using R 3.5.1[20].

Selecting images and supervised learning by CNN

Globally, there are ~67,000 terrestrial grid cells with a resolution of 0.5°. The length of the time series is 116 years. When we systematically shift the window of 40 years (30 years for training periods and 10 years for labeling periods), there can be $116 - 40 + 1 = 77$ images in one grid. Thus, the maximum number of images is $67,000 \times 77 \approx 5,200,000$. However, owing to limitations in computational resources, using the whole images would significantly deter the study. Thus, we decided to reduce the number of images to increase the efficiency of the study design. We assigned a random number c from 0 to 1 for each grid. When $c > c_t$, where c_t is the threshold, the grid is selected for Image Set, which is the dataset used in the experiments. By changing c_t, we could balance the efficiency and model performance. For example, to perform a 2×92 combinatorial analysis, we set $c_t = 0.98$. This limited the number images in Image Set to ~100,000. When we constructed high-performance models, we set $c_t = 0.7$ to use ~1.5 million images. By using the built-in exponential decay function in DIGITS 6.1 to prevent excessive overfitting, we gradually reduced the learning rates as the epoch progressed. Image Set was partitioned into training, validation, and test subsets in portions of 75%, 20%, and 5%, respectively. Because our study concerns time series, we assign the label training, validation, or test to each grid, rather than each image, to prevent partial duplications of images, which may induce overestimation of

accuracy, in the three subsets. Selected DIGITS 6.1 outputs for the supervised learning are shown in Supplementary Fig. 4.

DATA AVAILABILITY

The datasets analyzed during the current study are available in the Climatic Research Unit Time-Series v. 4.01[14,15] (CRU TS4.01, Table 1), https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.01/.

REFERENCES

1. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
2. Hampton, S. E. et al. Big data and the future of ecology. *Front. Ecol. Environ.* **11**, 156–162 (2013).
3. Rodell, M. et al. The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
4. Ise, T., Minagawa, M. & Onishi, M. Classifying 3 moss species by deep learning, using the "Chopped Picture" method. *Open J. Ecol.* **8**, 166–173 (2018).
5. Lu, Z., Fu, Z., Hua, L., Yuan, N. & Chen, L. Evaluation of ENSO simulations in CMIP5 models: a new perspective based on percolation phase transition in complex networks. *Sci. Rep.* **8**, 14912 (2018).
6. Roy, I., Gagnon, A. S. & Singh, D. Evaluating ENSO teleconnections using observations and CMIP5 models. *Theor. Appl. Climatol.* **136**, 1085–1098 (2019).
7. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
8. Pan, B., Shi, Z. & Xu, X. MugNet: deep learning for hyperspectral image classification using limited samples. *ISPRS J. Photogramm. Remote Sens.* **145**, 108–119 (2018).
9. Mou, L., Bruzzone, L. & Zhu, X. X. Learning spectral-spatial temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **57**, 924–935 (2019).
10. Racah, E., Beckham, C., Maharaj, T. & Prabhat, C. P. Semi-supervised detection of extreme weather events in large climate datasets. In *Proc 31st Conference on Neural Information Process. Syst. (NIPS 2017), Long Beach, CA, USA.* Vol 3, 1–13 (2017)
11. Shi, X. et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In *Proc Advances in Neural Information Processing System* Vol 28 (eds Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 802–810 (Curran Associates, Inc., 2015).
12. Ise, T. & Oba, Y. Forecasting climatic trends using neural networks: an experimental study using global historical data. *Front. Robot.* https://doi.org/10.3389/frobt.2019.00032 (2019)
13. Wen, L., Li, X., Gao, L. & Zhang, Y. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* **65**, 5990–5998 (2018).
14. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
15. Harris, I. & Jones, P. D. CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2016). https://doi.org/10.5285/58a8802721c94c66ae45c3baa4d814d0 (Centre for Environmental Data Analysis, 2017).
16. Selvaraju, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proc IEEE International Conference on Computer Vision (ICCV), Venice, Italy*. 618–626. https://doi.org/10.1109/ICCV.2017.74 (2017).
17. Ise, T., Ikeda, S., Watanabe, S. & Ichii, K. Regional-scale data assimilation of a terrestrial ecosystem model: leaf phenology parameters are dependent on local climatic conditions. *Front. Environ. Sci.* https://doi.org/10.3389/fenvs.2018.00095 (2018).
18. Yeager, L., Julie, B., Allison, G. & Houston, M. DIGITS: the deep learning GPU training system. *ICML AutoML Work.* https://doi.org/10.1016/j.pce.2003.11.004 (2015).
19. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2323 (1998).
20. R Core Team. *R: A Language and Environment for Statistical Computing.* https://doi.org/10.1108/eb003648 (R Foundation for Statistical Computing, Vienna, Austria, 2018).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

T.I. prepared environments for the study. Y.O. performed experiments and summarized the results. T.I. and Y.O. contributed equally and are considered as co-first authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41612-020-0129-x.

Correspondence and requests for materials should be addressed to T.I.

Reprints and permission information is available at http://www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.