ARTICLE   OPEN

Check for updates

Deep learning reveals moisture as the primary predictability source of MJO

Na-Yeon Shin[1], Daehyun Kim [ID][2,3 ✉], Daehyun Kang [ID][4], Hyemi Kim[5,6] and Jong-Seong Kug[1,7 ✉]

The Madden-Julian Oscillation (MJO) is the dominant mode of tropical intraseasonal variability that interacts with many other Earth system phenomena. The prediction skill of the MJO in many operational models is lower than its potential predictability, partly due to our limited understanding of its predictability source. Here, we investigate the source of MJO predictability by combining machine learning (ML) with a 1200-year-long Community Earth System Model version 2 (CESM2) simulation. A Convolutional Neural Network (CNN) for MJO prediction is first trained using the CESM2 simulation and then fine-tuned using observations via transfer learning. The source of MJO predictability in the CNN is examined via eXplainable Artificial Intelligence (XAI) methods that quantify the relative importance of the input variables. Our CNN exhibits an enhanced prediction skill over previous ML models, achieving a skill level of about 25 days. This level of performance is slightly superior or comparable to most operational models participating in the S2S project, although a few dynamical models surpass it. The XAI methods highlight precipitable water anomalies over the Indo-Pacific warm pool as the primary precursors of the subsequent MJO development for 1–3 weeks forecast lead times. Our results suggest that realistic representation of moisture dynamics is crucial for accurate MJO prediction.

INTRODUCTION

The Madden-Julian Oscillation (MJO), the dominant mode of tropical intraseasonal variability, is a planetary-scale, eastward-propagating disturbance in the tropics with a period of 30–90 days[1,2]. MJO is known to influence, among others, the frequency and intensity of high-impact weather events in the midlatitudes such as flooding and tornadoes[3–5]. Therefore, an accurate prediction of the MJO is a prerequisite for a reliable outlook of high-impact weather events with more than 2 weeks of forecast lead time.

Despite the steady improvement in the MJO prediction skill of the dynamical models over the past few decades[6], many operational models still struggle with a rather poor prediction skill[7,8]. Moreover, the MJO prediction skill of the operational models fall short of the estimated potential predictability of the MJO[9–11], suggesting that model errors are a barrier to further enhancing MJO prediction skill. The persistent errors in the model representation of the MJO[12–15] partly come from our limited understanding of the phenomenon[16,17].

One of the least understood aspects of the MJO is its predictability source. Even though there are models that exhibit a superior skill in forecasting the MJO than other models (e.g., ECMWF integrated forecasting system CY43R3[7]), understanding why the particular model performs better than others are a non-trivial task. In particular, it is not well understood where the predictability comes from. A systematic assessment of the source of MJO predictability would be helpful to better understand the nature of the MJO and eventually help improve its prediction skill.

Recent applications of Artificial Intelligence (AI) and Machine Learning (ML) techniques to weather and climate prediction have demonstrated their high potential[18–20]. In addition, attempts have been made to understand ML results by utilizing 'eXplainable AI' (XAI) tools[18,21,22], which aid human interpretation of the ML predictions[23].

Although ML methods and XAI tools have been also applied to MJO prediction in recent studies[24–27], their prediction skill (~20 days) has remained lower than that of most operational forecasts. Thus, the sources of MJO predictability using XAI for relatively long forecast lead times (>15 days) have yet to be thoroughly investigated and quantified. In addition, while the XAI methods they used have provided information about the specific regions within the input fields that influenced the ML model's prediction, they did not quantify the impact of changes to the input field on the predicted output value[24–27]. A shortcoming of the previously developed ML models for MJO prediction is that they used only observations, which limits the number of MJO events used in the training to capture the diverse nature of the MJO[28].

In this study, we use a long-term (1200-year) climate simulation made with Community Earth System Model version 2 (CESM2)[29], which is known to simulate a realistic MJO[30,31], to build a robust regression CNN model for MJO prediction. We then fine-tune the parameters of the CESM2-data-trained CNN with observations via the so-called 'transfer learning' technique to address the systematic bias of CESM2[18,32]. Although transfer learning has been successfully utilized in several climate studies[18,21], it has not been applied to MJO research. It will be shown that our model is skillful at forecasting the observed MJO and that atmospheric water vapor anomalies are identified as the primary source of MJO predictability for forecast lead times within 3 weeks.

[1]Division of Environmental Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang, South Korea. [2]School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea. [3]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA. [4]Center for Sustainable Environment Research, Korea Institute of Science and Technology, Seoul, South Korea. [5]Department of Science Education, Ewha Womans University, Seoul, South Korea. [6]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA. [7]Institute for Convergence Research and Education in Advanced Technology, Yonsei University, Seoul, South Korea. ✉email: daehyun@snu.ac.kr; jskug@postech.ac.kr

RESULTS

Performance of the CNN-based MJO prediction model

Figure 1 shows the bivariate correlation coefficients (BCOR)[6,33,34] skill scores of our CNN-based MJO prediction model as a function of the forecast lead time. The model is trained to predict the values of RMM1 and RMM2 indices with the daily anomaly maps over 30°S-30°N and 0-360°E of the following five input variables: outgoing longwave radiation (OLR), zonal winds at 200hPa (U200) and 850hPa (U850), precipitable water (PW), and surface temperature (TS). We use data from boreal winter (December-January-February-March) during which MJO activity peaks[35]. See Methods for the details. Hereafter, the CNN model before and after the transfer learning will be referred to as CNN_{CESM2} and CNN_{OBS}, respectively.

The prediction skill of the CNN_{OBS} is above 0.5 up to ~25 days of forecast lead time, which is superior to those of the previous ML models developed by Martin et al.[26] (~19 days) and Delaunay & Christensen[24] (~21 days). When measured in a similar way, the ECMWF model shows the highest prediction skill (~36 days based on NDJFM season and initial MJO amplitude>1), which is followed by several models, including CESM2 as a dynamical model (red line in Supplementary Fig. 1), that exhibit the prediction skill of about 24–27 days[7,8,16] as shown in Supplementary Fig. 1. Some dynamical models show skills lower than ~21 days (Supplementary Fig. 1). Therefore, the CNN_{OBS} is capable of tapping the predictability of the MJO embedded in the input data to achieve forecast skill that is slightly better than or comparable to that of most dynamical models in the S2S archive, except for the ECMWF model. However, there are other dynamical models that still outperform CNN_{OBS}[36,37].

Refining the CNN_{CESM2} with observations is found to have statistically significant impacts on the prediction skill of the model up to 3 weeks forecast lead time ($p < 0.05$) (Supplementary Fig. 1). The increase in the BCOR skill score after the transfer learning is most pronounced for weeks 2 and 3, with the increases being larger than 0.1. It is worthwhile to note that the period of the CESM2-simulated MJO is on the shorter side of the observed range (Supplementary Figs 2a, b), which may limit the CNN_{CESM2} in detecting the source of MJO predictability in observation within 2 weeks. The notable increase in the skill score after the transfer learning, therefore, suggests that the CNN_{OBS} is appropriately calibrated to lessen the effect of the systematic bias of CESM2 (i.e., the shorter-than-observed periodicity) for the forecast lead time within 3 weeks through the transfer learning.

Investigation of the MJO's predictability source

The high skill of the CNN_{OBS} demonstrated in Fig. 1 warrants investigating its origin, namely, the source of MJO predictability. As a rough, but simple way to evaluate the relative importance of the input variables, we repeated the prediction with the CNN_{OBS} by zeroing a single variable at a time. Through this process, the importance of a variable is assumed to be proportional to the degree of skill degradation when the effect of a single variable is eliminated.

When compared to the skill of the 'all-variable' forecasts, it is found that the greatest decrease in the skill score occurs when PW is removed, particularly within 20 days of forecast lead times, suggesting that PW anomalies are the dominant predictability source for up to 3 weeks (Fig. 1). At the 10-day forecast lead time, for example, the BCOR skill score drops by about 0.15 when the effect of PW is removed from the input. Small, but noticeable reductions in BCOR are also observed with the removal of U200 and TS. Interestingly, the effects of U200 on the skill score are largest at shorter forecast lead times (5–10 days), while the reduction in skills after removing TS is greatest at forecast lead times longer than 25 days. Overall, U850 or OLR has minimal effects on the prediction skill.

While Fig. 1 shows the relative importance of each input variable on the skill of the CNN_{OBS}, it lacks information about the specific features in the input data that the ML model 'sees' when making predictions. To identify the features in the input data that have a relatively large impact on the predicted RMM indices, we use an XAI method called the Signed-Contribution (SC) map (see Methods).

Figure 2 shows the SC maps of each input variable for the models that forecast RMM1 (Fig. 2a-e) and RMM2 (Fig. 2f-j) values 15 days after the initial date. As consistent with Fig. 1, the magnitude of the signals presented in the SC maps is the largest in PW, suggesting that PW is the most important input field for the MJO prediction. The corresponding SC maps of TS and upper and lower-level zonal winds show signals that are comparable in magnitude, but much weaker than the PW's, and a negligible signal is found in OLR. Note that the results shown in Fig. 2 are consistent with those from the saliency map[38] (Supplementary Fig. 3), a well-known XAI method that indicates the sensitivity of the output to each input variable.

The SC map of PW for RMM1 indicates that positive PW anomalies in the western Indian Ocean (WIO) are related to an increase in the RMM1 forecast with a 15-day forecast lead time (Fig. 2a). Likewise, positive PW anomalies in the western Pacific cause a decrease in the value of RMM1 forecast (Fig. 2a), while increasing the value of RMM2 forecast (Fig. 2b). The corresponding lag-correlation map (Supplementary Fig. 4) indicates that the strong PW signals in the SC map of the 15-day forecast overlap the anomalies that are most pronounced 15 days before RMM1 or RMM2 peaks. This suggests that the CNN_{OBS} makes skillful predictions by carrying the predictability embedded in the input moisture anomaly field through 15 days. Similarly, the SC map and
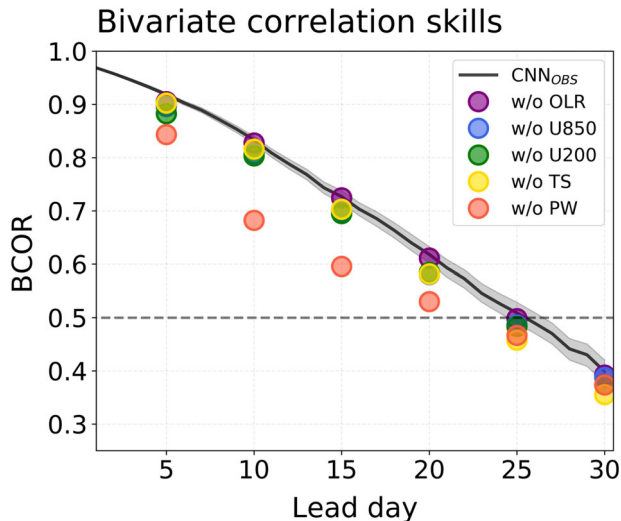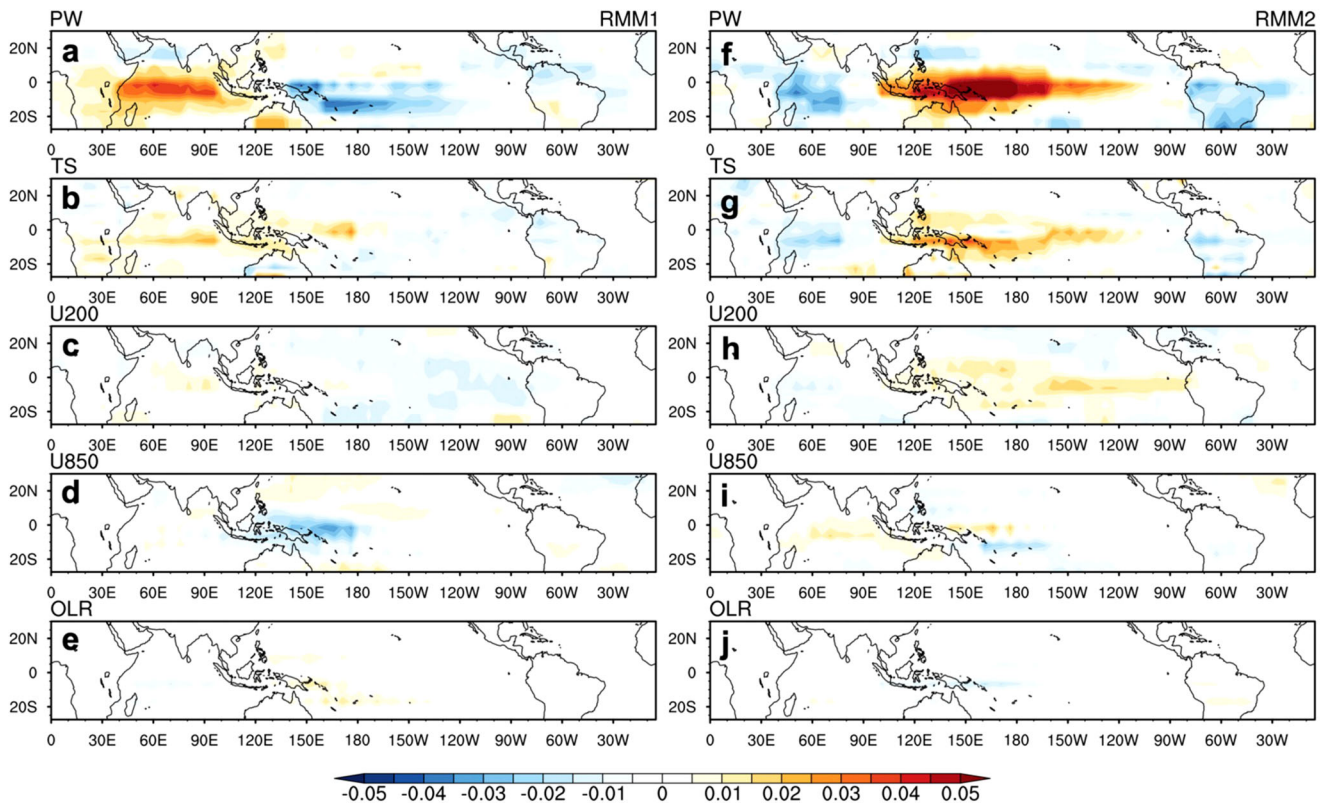

Fig. 1 The skills of the CNN_{OBS} and relative importance among input variables based on skill. The bivariate correlation (BCOR) skills between the target and the CNN_{OBS} output variables (i.e., RMM indices) for observation data. The BCOR is calculated by the mean for 20 networks. The black solid line is calculated for all test datasets of observation when the initial MJO amplitude $\left(\sqrt{(\mathbf{RMM1})^2 + (\mathbf{RMM2})^2}\right)$ is strong (>1). The shaded areas indicate the 95% confidence level using the bootstrap method. The gray dashed line indicates a BCOR skill at 0.5. The dots represent the BCOR skills between the target and the prediction after removing the effect of a single variable using the CNN_{OBS} for the 5, 10, 15, 20, 25, and 30-day forecast lead time. The red, yellow, green, blue, and purple dots indicate the prediction skills with PW, TS, U200, U850, and OLR removed, respectively.

Fig. 2 The signed-contribution (SC) map for a 15-day forecast lead time. The SC maps of observation for (**a-e**) RMM1 and (**f-j**) RMM2 at a 15-day forecast lead time. (**a,f**) PW, (**b,g**) TS, (**c,h**) U200, (**d,i**) U850, and (**e,j**) OLR. The map indicates an averaged SC map of the individually obtained SC map for each network, initialized with different weights.

the lag-correlation map for RMM2 show a reasonable match, albeit being somewhat weaker than for RMM1 (Supplementary Fig. 4).

Interestingly, Figs. 1 and 2 show that OLR anomalies contain almost no memory of the future MJO evolution, although PW and OLR are both strongly correlated with RMM indices (Supplementary Fig. 4). The lack of memory in the OLR field suggests that the moisture anomaly largely determines the distribution of convective activity[39,40]. While the OLR SC maps show much stronger signals when the CNN_{OBS} is trained without PW, the resulting model exhibits lower prediction skill (not shown). This again suggests that the memory of OLR variability in intraseasonal timescale largely comes from that in the moisture field.

The patterns of the PW signals in the SC maps vary notably with the forecast lead time. Figure 3 shows the SC and lag-correlation maps of PW for RMM1 at the 5, 10, 15, 20, and 25-day forecast lead times. Overall, strong signals are mostly concentrated in the Indo-Pacific warm pool region. At a longer lead time (e.g., 25-day), positive signals are found mainly in the far western Indian Ocean (Fig. 3a). As the forecast lead time decreases, the dipole of positive and negative signals gradually moves to the east, and the strong signals appear in the Maritime Continent, northern Australia, and the west Pacific at 5-day forecast lead time (Fig. 3e). These regions are close to the areas of negative OLR anomalies associated with RMM1 (Supplementary Fig. 2c).

For all forecast lead times considered, the PW SC maps (Fig. 3a-e) favorably match the corresponding lag-correlation map (Fig. 3f-j). This again suggests that, for a given moisture anomaly field, part of the subsequent evolution of the moisture field is predictable, and the CNN_{OBS} learns the pattern of that evolution. In other words, moisture anomalies contain memories that last longer than the weather time scale, and the CNN_{OBS} is able to use the memory for the prediction of RMM indices. The smooth transition of the signals in the zonal direction as a function of the forecast lead time also

strongly suggests that, although the CNN_{OBS} is trained independently for each lead time, the CNN_{OBS} captures a common feature in the dynamics of the real MJO.

The signals in the TS SC maps look similar to that of the PW SC maps, while the TS signals are located slightly to the east of the PW signals (Supplementary Fig. 5). Note that the signals in the TS SC maps show a zonally-elongated pattern along the equator across the western Pacific, which is absent in the corresponding lag-correlation map. This phase lag between PW and TS is presumably due to that SST anomalies tend to be highest to the east of enhanced MJO convection after the ocean received enhanced shortwave under the suppressed phase and before enhanced surface flux and turbulent mixing cool the ocean surface during the active phase. Our results are consistent with the studies that highlighted the pre-conditioning role of SST anomalies in the propagation of the MJO[41].

Relatively large signals in the western hemisphere are found in the SC and linear correlation maps of U200 (Supplementary Fig. 6), unlike in the corresponding PW or TS maps, where the signals are mostly concentrated in the Indo-Pacific warm pool. The U200 anomalies in the western hemisphere are associated with the 'dry' component of the MJO[42,43], which has been suggested to sometimes help initiate the MJO in the Indian Ocean[44]. Our model seems to capture the dry component of the MJO as a predictability source of the MJO.

Figure 4 summarizes our investigation of MJO predictability source in a quantitative manner. In Fig. 4, the relative role of each variable is quantified as the bivariate mean-squared difference[45] (BMSD) between the prediction results with and without a single variable (see Methods). Figure 4 quantifies the relative influence of each variable on the changes in RMM indices themselves, whereas Fig. 1 shows the effect of each input variable on the skill score. Therefore, high (low) BMSD values indicate that the input variables
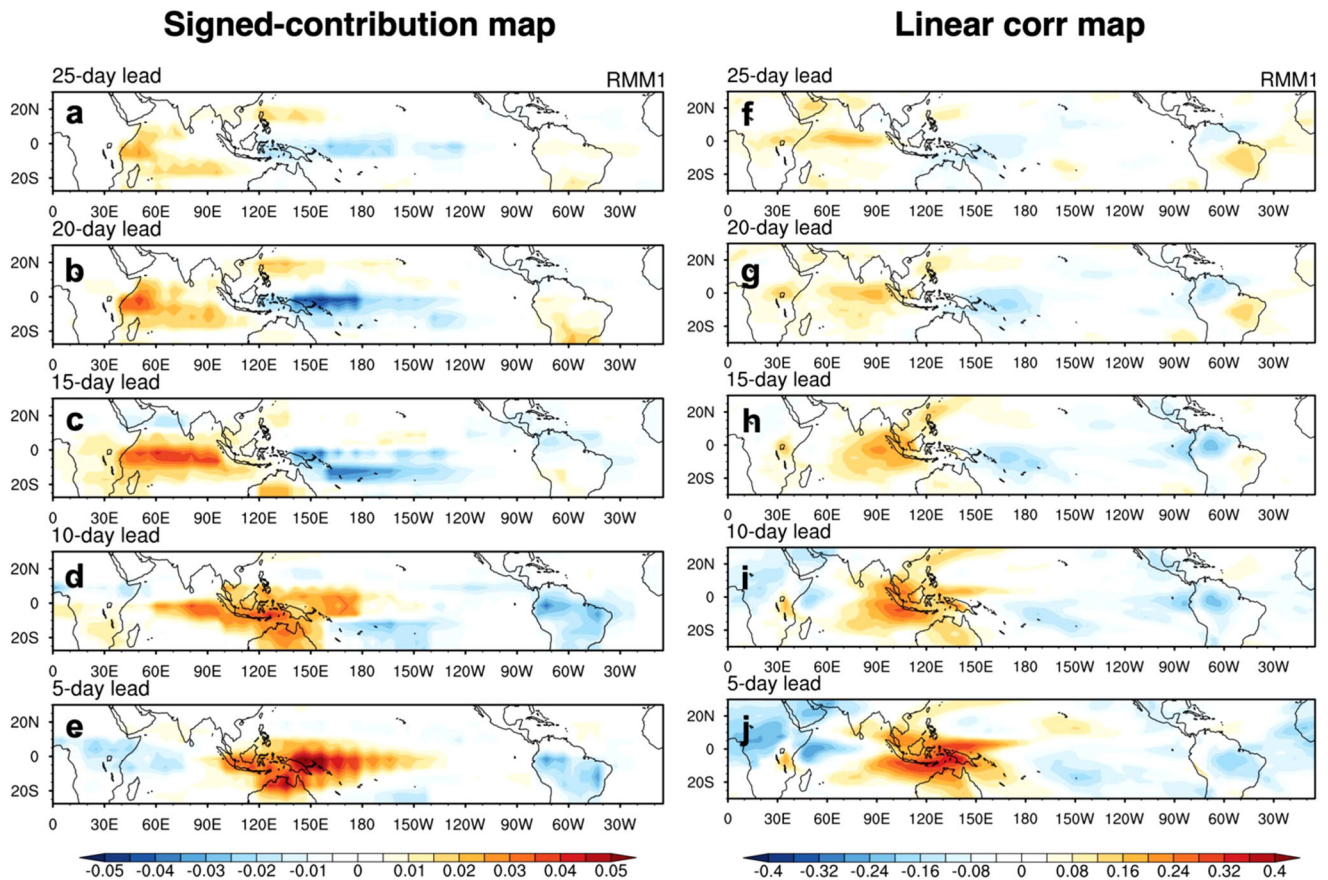
Signed-contribution map

Linear corr map

Fig. 3 The evolution of SC and lag-correlation maps with decreasing forecast lead times. The evolution of the (**a–e**) SC maps of CNN$_{OBS}$ and linear (f-j) lag-correlation maps of observation for (**a,f**) 25, (**b,g**) 20, (**c,h**) 15, (**d,i**) 10, and (**e,j**) 5-day forecast lead time. The maps are calculated for PW and RMM1 index. The SC map in each panel indicates an averaged SC map of the individually obtained SC map for each network, initialized with different weights. The shadings of lag-correlation maps indicate a 95% confidence level.

exert a large (small) influence on the variation of the RMM indices, suggesting a wealth (scarcity) of information that can be utilized as sources of MJO predictability. In general, forecasts at shorter lead times tend to be affected more strongly by input values (initial conditions), whereas the effect of input values diminishes as the forecast lead time increases. Interestingly, however, the total BMSD (gray bars in Fig. 4) increases until the 10-day forecast lead time before decreasing at the longer lead times. This may imply that the factors governing the predictability in the lead times shorter than 10 days are not sufficiently considered in the CNN$_{OBS}$. Thus, the potential to improve the skill score of the early forecast lead times (< 10 days) remains. Note that the corresponding results obtained by comparing other all possible combinations of the input variables show consistent results (Supplementary Methods and Supplementary Fig. 7).

The lines in Fig. 4 show that the relative influence of PW is the largest for the forecast lead times shorter than 3 weeks, again strongly suggesting that PW is the dominant source of MJO predictability at those forecast lead times. The relative contribution of PW reaches up to about 50% at the 10-day lead and reduces as the forecast lead times increase. Therefore, up to the forecast lead time of about 3 weeks, PW is not only most conducive to skill improvement (Fig. 1) but also the variable that has the most control over the predicted RMM indices (Fig. 4).

The next largest contributions to the total BMSD come from U200 and TS, followed by that of U850 and then OLR. The U200's relative impact, which has the second largest proportion up to about 2 weeks, reduces from about 23% (at 5-day lead) to 17% (at 15-day lead) and then increases to about 32% (at 30-day lead).

This result agrees with the U200 showing the second largest impact on a skill within about 2 weeks (Fig. 1). The relative effects of TS show an increasing trend as the forecast lead time increases. At the 25-day forecast lead time, the proportion of TS to the total BMSD is more than two times that at the 5-day forecast lead time. This suggests that TS has a relatively large memory for MJO predictability on longer timescales. The effects of U850 and OLR, which occupy a small proportion with relatively weak influence, decrease monotonically as the forecast lead time increases.

DISCUSSION

Once successfully trained, making predictions with ML-based weather and climate prediction models takes only a tiny fraction of the time and computing power that are required to numerically solve the governing equations of the Earth system components for the equivalent prediction exercise. If the prediction skill of the ML models is comparable to that of the dynamical models, the low-cost aspect makes them an attractive tool for weather and climate prediction that can be used together with the dynamical models. By overcoming the issue of limited data availability with the use of a multi-century CESM2 simulation, the CNN$_{OBS}$ showed a remarkable MJO prediction skill of about ~25 days, which was superior to the skill of previous ML models as well as many dynamical models.

Although the skill of the CNN$_{OBS}$ is still behind that of a few dynamical models, our results demonstrated the potential that the ML models can be further improved by fully utilizing high-quality, high-volume data. Interestingly, the skill of the CNN$_{OBS}$ was
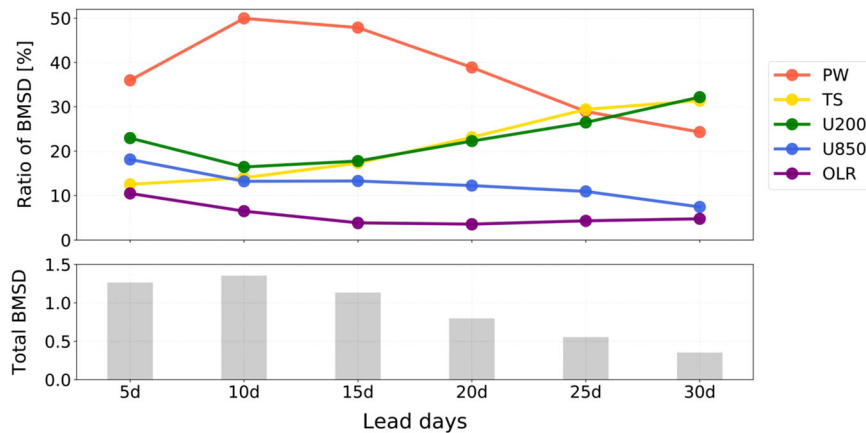
Fig. 4 The quantified contribution (bivariate-mean-square difference; BMSD) of each input variable for 5, 10, 15, 20, 25, and 30-day forecast lead time. The BMSDs are calculated between the original prediction and the prediction after removing the effect of a single variable using the CNN$_{OBS}$ for each forecast lead time. The gray bars indicate the sum of the BMSDs for all variables. The red, yellow, green, blue, and purple bars indicate the proportion of BMSD for PW, TS, U200, U850, and OLR, respectively, to the total sum. The x-axis is the lead days.

comparable to that of the dynamical forecasts made with CESM2 (~24 days; red line in Supplementary Fig. 1)[46]. Considering that the CNN$_{OBS}$ is trained with the CESM2 long-term simulation, our results suggest that the skill of ML models may further improve if a GCM that better simulates the MJO is used to create the long-term training simulation data for the training of an ML model. In addition, our methods can be applied to multi-model simulation results by training deep learning models separately for each simulation, which would facilitate comparisons of MJO predictability and its sources among different climate models.

The current study also demonstrated that skillful ML models could be used to reveal the nature of the target phenomenon. Our results highlighted the large-scale moisture anomalies as the dominant source of MJO predictability for up to 3 weeks of forecast lead time. This strongly supports the 'moisture mode' theoretical framework for the MJO[47–50], which explains the maintenance and propagation of MJO convection by those of moisture anomalies. Our results also suggest that moisture dynamics is a crucial process for successful MJO prediction with the dynamical models. While we used the column-integrated water vapor as input, it will be of interest to further examine whether water vapor anomalies in different layers (boundary layer vs. lower free-troposphere) play different roles as the MJO's predictability source.

All current ML models for MJO prediction, including our CNN model, take the 'time-skipping' approach, in which the models take input variables of the initial time and directly predict output variables at a target future time. Therefore, the time-skipping ML models are constructed independently for each forecast lead time. While our results indicated that the ML models constructed for different target forecast lead times exhibited consistency (Fig. 3), the use of time-skipping models may not be suitable to represent the non-linear evolution of errors in the initial conditions, limiting the effectiveness of the ensemble prediction. Meanwhile, recursive ML models have been successfully constructed and used for weather and subseasonal prediction (FourCastNet[51], DLWP[52]). It will be worthwhile to design the recursive ML models (e.g., recurrent neural network, long-short term memory, transformer, etc.) specifically for MJO prediction and assess their MJO prediction skills. Furthermore, the application of various DL techniques, including graphical neural network and generative adversarial network, holds promise for further enhancing MJO prediction skill and understanding the uncertainty associated with it, taking advantage of the inherent capabilities offered by each of these techniques.

METHODS

Data

We used a 1200-year-long pre-industrial simulation made with CESM2[29]. CESM2 shows a realistic representation of the MJO[30], which qualifies the model as a reasonable tool to understand the predictability of the MJO. The daily mean of outgoing longwave radiation (OLR), zonal winds at 250hPa (U250) and 850hPa (U850), precipitable water (PW), and surface temperature (TS) over 30°S-30°N and 0-360°E were used. All variables were interpolated onto a 2.5° × 2.5° grid. The daily mean fields were pre-processed as anomalies by removing the mean and the first three harmonics of the climatological seasonal cycle. Each input variable was then normalized by its own domain-averaged standard deviation before being used as input to the CNN$_{OBS}$. The RMM indices, the output variable for the CNN$_{OBS}$, are obtained as the principal components of the leading two combined empirical orthogonal function (CEOFs) of meridionally averaged (15°S-15°N) OLR, U250, and U850 anomalies[53]. As in Wheeler & Hendon[53], the previous 120-day mean was removed from the three variables before obtaining the CEOFs and their principal components.

For the observational data, we used NOAA interpolated OLR[54], ERA-5 reanalysis[55] U200, U850, PW, and TS from 1 January 1979 to 31 December 2018. The horizontal resolution and pre-processing methods used were the same as those used for CESM2.

Convolutional neural network (CNN) and transfer learning

CNN[56] is specialized in learning the relationship between input and output by extracting features from input images. With maps of five variables as the input and two numbers as the output, we constructed a simple CNN structure that consists of three convolutional layers of 30 2 × 2 convolution filters, one 50-unit hidden layer, and one output layer with 2 units (Supplementary Fig. 8). We used the rectified linear unit, the mean square error, and Adam as the activation function, the loss function, and the optimizer, respectively. The learning rate is 0.005 for CNN$_{CESM2}$ and 0.0005 for CNN$_{OBS}$. To set the number of epochs, we applied early stopping.

Firstly, we trained a CNN with a 1200-year-long CESM2 simulation (CNN$_{CESM2}$). To use all available data for pre-training, we did not conduct cross-validation on the CESM2 simulation and utilized 75% of the entire CESM2 dataset for training and 25% for validation. In this study, the CNN$_{CESM2}$ is independently trained for each forecast lead time (1 to 30 days). Furthermore, we trained 20 different networks, each initialized

with different weights, to take into account the sensitivity of the CNN_{CESM2} to weight initialization.

The prediction skill of CNN_{CESM2} increases as more training data are used (Supplementary Fig. 9), and more than 400 years of training data are required for the model to yield the skill score greater than 0.6 at the 15-day forecast lead time. Interestingly, the forecast skill is not completely saturated even with more than 700 years of training data (corresponding to 100% in the x-axis in Supplementary Fig. 9), suggesting that using longer simulation data would be beneficial to further improve MJO prediction.

Second, we utilize transfer learning[32] to fine-tune the weights of all layers of the pre-trained model (i.e., CNN_{CESM2}) to fit the observation data. We use transfer learning to simultaneously acquire 1) the advantages of CESM2 with vast amounts of data and 2) the benefits of using observational (and reanalysis) data. For cross-validation, we split the observation dataset into three groups with 3:3:4 ratios and performed three rounds of transfer learning with one of the three groups as a test dataset and the remaining as training and validation datasets. As a result, we obtained a total of 60 independent CNNs (20 networks × 3 datasets) for each forecast lead time. All results in this study are the average of all networks, which is referred to as the CNN_{OBS}.

Signed-contribution (SC) map

Shin et al.[21] proposed an XAI method called the contribution map, which quantitatively estimates the relative influence of each input variable at each grid point on the output. For every input value, its influence on the prediction is measured as the difference (D in Eq. (1)) between the original prediction and the prediction after zeroing the input variable:

$$D_{i,j} = f(X_j) - f(X_j^i),$$ (1)

where f is an ML prediction model, X_j is the full input data array for the jth sample, and X_j^i is the modified array with the ith element replaced with zero. The overall 'contribution' of the ith element to the prediction results is then quantified as the root-mean-squared D:

$$C_i = \sqrt{\overline{D_{i,j}^2}},$$ (2)

where ⁻ indicates the average over the entire sample. In the Contribution map, C_i is displayed as a function of longitude and latitude for each input variable (e.g., Shin et al.[21]). While providing useful insights into how an ML model makes a specific prediction, the original Contribution map measures only the magnitude of the influence, lacking sign information. To overcome this limitation, we developed the signed-contribution (SC) map, which provides the sign of the relationship between the input and output together with the magnitude of the importance. To produce an SC map, we first obtain two conditional averages of $D_{i,j}$ by using only those with sufficiently high (i.e., greater than 1.0 standard deviation) and low (i.e., lower than −1.0 standard deviation) values of the ith element in the input data:

$$D_i^{high} = \overline{D_{i,j}}^{X(i)_j > SD_i}; D_i^{low} = \overline{D_{i,j}}^{X(i)_j < -SD_i}$$ (3)

where $X(i)_j$ is the ith element of the input data for the jth sample and SD_i is the standard deviation of the ith input element. The 'signed-contribution' is obtained as the difference between the two conditional averages:

$$SC_i = D_i^{high} - D_i^{low}.$$ (4)

The sign and magnitude of SC_i reflect the direction and amplitude of the output changes associated with removing the effect of the ith input element, respectively. As in the contribution map, SC_i is displayed as a function of longitude and latitude for

each input variable in the SC map (e.g., Figs. 2 and 3). In this paper, we show the average of SC maps for all individual networks, initialized with different weights.

Bivariate-mean-square-difference (BMSD)

We conducted a single-variable zeroing test using our CNN_{OBS} to evaluate the extent to which the RMM indices themselves are altered by removing the effect of individual input variables (Fig. 4). To do this, we computed the BMSD for each input variable ($BMSD_k$ in Eq. (5)), which is the mean-square-differences between the original predictions and the single-variable zeroing prediction for both RMM1 and 2 indices, as follows:

$$BMSD_k = \frac{1}{N} \sum_{j=1}^{N} \left(\left[f_1(X_j) - f_1(X_j^k) \right]^2 + \left[f_2(X_j) - f_2(X_j^k) \right]^2 \right)$$ (5)

where f is an ML prediction model, the subscripts 1 and 2 indicate the prediction value of RMM1 and RMM2 indices, respectively. X_j is the full input data array for the jth sample (N=total number of samples), and X_j^k denotes the modified array with all k^{th} variables replaced by zero. The total BMSD (Eq. (6)) is the sum of the BMSDs of five input variables:

$$total\ BMSD = \sum_{k=1}^{5} BMSD_k.$$ (6)

To quantify the contribution of each input variable, we calculated the ratio of each variable's BMSD ($BMSD_k ratio$ in Eq. (7)) to the total BMSD as follows:

$$BMSD_k ratio = BMSD_k / total\ BMSD.$$ (7)

In this paper, we show the average of BMSDs for all individual networks, initialized with different weights.

DATA AVAILABILITY
The training and test datasets for the CNN_{CESM2} and the CNN_{OBS} used in this study are available at https://doi.org/10.6084/m9.figshare.21972653.v1. The NOAA Interpolated Outgoing Longwave Radiation (OLR) reanalysis data were downloaded from https://psl.noaa.gov/data/gridded/data.olrcdr.interp.html. The ERA5 hourly data on pressure levels are available from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form.

CODE AVAILABILITY
The codes used to produce the results except for training, test, and SC maps are available from the corresponding author upon reasonable request. The codes related to the CNN_{CESM2} and the CNN_{OBS} are available at https://doi.org/10.6084/m9.figshare.21972653.v1.

REFERENCES
1. Madden, R. A. & Julian, P. R. Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *J. Atmos. Sci.* **28**, 702–708 (1971).
2. Madden, R. A. & Julian, P. R. Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period. *J. Atmos. Sci.* **29**, 1109–1123 (1972).
3. Domeisen, D. I. V. et al. Advances in the Subseasonal Prediction of Extreme Events: Relevant Case Studies across the Globe. *B Am. Meteorol. Soc.* **103**, E1473–E1501 (2022).
4. Schreck, C. J. Global Survey of the MJO and Extreme Precipitation. *Geophys. Res. Lett.* **48**, (2021).
5. Zhang, C. Madden–Julian Oscillation: Bridging Weather and Climate. *B Am. Meteorol. Soc.* **94**, 130405130926004 (2013).
6. Kim, H., Vitart, F. & Waliser, D. E. Prediction of the Madden-Julian Oscillation: A Review Prediction of the Madden-Julian Oscillation: A Review. *J. Clim.* **31**, 9425–9443 (2018).

7. Kim, H., Janiga, M. A. & Pegion, K. MJO Propagation Processes and Mean Biases in the SubX and S2S Reforecasts. *J. Geophys Res Atmosph.* **124**, 9314–9331 (2019).

8. Lim, Y., Son, S.-W. & Kim, D. MJO prediction skill of the subseasonal-to-seasonal prediction models. *J. Clim.* **31**, 4075–4094 (2018).

9. Kim, H.-M. & Kang, I.-S. The impact of ocean–atmosphere coupling on the pre-dictability of boreal summer intraseasonal oscillation. *Clim. Dynam* **31**, 859 (2008).

10. Neena, J. M., Lee, J. Y., Waliser, D., Wang, B. & Jiang, X. Predictability of the Madden–Julian Oscillation in the Intraseasonal Variability Hindcast Experiment (ISVHE)*. *J. Clim.* **27**, 4531–4543 (2014).

11. Reichler, T. & Roads, J. O. Long-Range Predictability in the Tropics. Part II: 30–60-Day Variability. *J. Clim.* **18**, 634–650 (2005).

12. Ahn, M.-S., Kim, D., Ham, Y.-G. & Park, S. Role of Maritime Continent Land Con-vection on the Mean State and MJO Propagation Role of Maritime Continent Land Convection on the Mean State and MJO Propagation. *J. Clim.* **33**, 1659–1675 (2019).

13. Jiang, X. et al. Vertical structure and physical processes of the Madden-Julian oscillation: Exploring key model physics in climate simulations. *J. Geophys Res Atmosph.***120**, 4718–4748 (2015).

14. Lin, J.-L. et al. Tropical Intraseasonal Variability in 14 IPCC AR4 Climate Models. Part I: Convective Signals. *J. Clim.* **19**, 2665–2690 (2006).

15. Slingo, J. M. et al. Intraseasonal oscillations in 15 atmospheric general circulation models: results from an AMIP diagnostic subproject. *Clim. Dynam* **12**, 325–357 (1996).

16. Jiang, X. et al. Fifty Years of Research on the Madden-Julian Oscillation: Recent Progress, Challenges, and Perspectives. *J Geophys. Res. Atmos.* **125**, 68 (2020).

17. Zhang, C., Adames, Á. F., Khouider, B., Wang, B. & Yang, D. Four Theories of the Madden-Julian Oscillation. *Rev. Geophys* **58**, e2019RG000685 (2020).

18. Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).

19. Dueben, P. D. & Bauer, P. Challenges and design choices for global weather and climate models based on machine learning. *Geosci. Model Dev. Discuss* **11**, 1–17 (2018).

20. Weyn, J. A., Durran, D. R., Caruana, R. & Cresswell-Clay, N. Sub-Seasonal Fore-casting With a Large Ensemble of Deep-Learning Weather Prediction Models. *J. Adv. Model Earth Syst.* **13**, e2021MS002502 (2021).

21. Shin, N.-Y., Ham, Y.-G., Kim, J.-H., Cho, M. & Kug, J.-S. Application of deep learning to understanding ENSO dynamics. *Artif. Intell. Earth Syst.* **1**, 1–37 (2022).

22. Mamalakis, A., Ebert-Uphoff, I. & Barnes, E. A. Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *Environ. Data Sci.* **1**, p.e8 (2022).

23. Adadi, A. & Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *Ieee Access* **6**, 52138–52160 (2018).

24. Delaunay, A. & Christensen, H. M. Interpretable Deep Learning for Probabilistic MJO Prediction. *Geophys. Res. Lett.* **49**, e2022GL098566 (2022).

25. Kim, H., Ham, Y. G., Joo, Y. S. & Son, S. W. Deep learning for bias correction of MJO prediction. *Nat. Commun.* **12**, 3087 (2021).

26. Martin, Z. K., Barnes, E. A. & Maloney, E. Using Simple, Explainable Neural Net-works to Predict the Madden-Julian Oscillation. *J. Adv. Model Earth Sy* **14**, e2021MS002774 (2022).

27. Toms, B. A., Kashinath, K., Prabhat & Yang, D. Testing the Reliability of Inter-pretable Neural Networks in Geoscience Using the Madden-Julian Oscillation. *Geosci. Model Dev. Discuss* **2020**, 1–22 (2020).

28. Wang, B., Chen, G. & Liu, F. Diversity of the Madden-Julian Oscillation. *Sci. Adv.* **5**, eaax0220 (2019).

29. Danabasoglu, G. et al. The Community Earth System Model Version 2 (CESM2). *J. Adv. Model Earth Sys.* **12**, MS001916 (2020).

30. Ahn, M. et al. MJO Propagation Across the Maritime Continent: Are CMIP6 Models Better Than CMIP5 Models? *Geophys. Res. Lett* **47**, e2020GL087250 (2020).

31. Kang, D. et al. The Role of the Mean State on MJO Simulation in CESM2 Ensemble Simulation. *Geophys. Res. Lett.* **47**, e2020GL089824 (2020).

32. Bozinovski, S. Reminder of the First Paper on Transfer Learning in Neural Net-works, 1976. *Informatica* **44**, 3 (2020).

33. Rashid, H. A. & Hirst, A. C. Investigating the mechanisms of seasonal ENSO phase locking bias in the ACCESS coupled model. *Clim. Dynam* **46**, 1075–1090 (2015).

34. Vitart, F. et al. The Sub-seasonal to Seasonal Prediction (S2S) Project Database. *B Am. Meteorol. Soc.* **98**, 163–173 (2016).

35. Kang, D., Kim, D., Rushley, S. & Maloney, E. Seasonal Locking of the MJO's Southward Detour of the Maritime Continent: The Role of the Australian Mon-soon. *J. Clim.* **35**, 4553–4568 (2022).

36. Miyakawa, T. et al. Madden–Julian Oscillation prediction skill of a new-generation global model demonstrated using a supercomputer. *Nat. Commun.* **5**, 3769 (2014).

37. Xiang, B. et al. The 3–4-Week MJO Prediction Skill in a GFDL Coupled Model. *J. Clim.* **28**, 5351–5364 (2015).

38. Simonyan, K., Vedaldi, A. & Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *Arxiv* (2013).

39. Bretherton, C. S., Peters, M. E. & Back, L. E. Relationships between Water Vapor Path and Precipitation over the Tropical Oceans. *J. Clim.* **17**, 1517–1528 (2004).

40. Rushley, S. S., Kim, D., Bretherton, C. S. & Ahn, M. -S. Reexamining the Nonlinear Moisture-Precipitation Relationship Over the Tropical Oceans. *Geophys Res Lett.* **45**, 1133–1140 (2018).

41. Zhou, L. & Murtugudde, R. Oceanic Impacts on MJOs Detouring near the Maritime Continent Oceanic Impacts on MJOs Detouring near the Maritime Continent. *J. Clim.* **33**, 2371–2388 (2020).

42. Sobel, A. H. & Kim, D. The MJO-Kelvin wave transition. *Geophys. Res. Lett.* **39**, L20 808 (2012).

43. Powell, S. W. Successive MJO propagation in MERRA-2 reanalysis: Successive MJO Circumnavigation. *Geophys Res Lett.* **44**, 5178–5186 (2017).

44. Powell, S. W. & Houze, R. A. Effect of dry large-scale vertical motions on initial MJO convective onset. *J. Geophys Res Atmospheres* **120**, 4783–4805 (2015).

45. Rashid, H. A., Hendon, H. H., Wheeler, M. C. & Alves, O. Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. *Clim. Dynam* **36**, 649–661 (2011).

46. Richter, J. H. et al. Subseasonal Earth System Prediction with CESM2. *Weather Forecast* **37**, 797–815 (2022).

47. Adames, Á. F. & Kim, D. The MJO as a Dispersive, Convectively Coupled Moisture Wave: Theory and Observations. *J. Atmos. Sci.* **73**, 913–941 (2016).

48. Sobel, A. & Maloney, E. Moisture Modes and the Eastward Propagation of the MJO. *J. Atmos. Sci.* **70**, 187–192 (2013).

49. Maloney, E. D. The Moist Static Energy Budget of a Composite Tropical Intra-seasonal Oscillation in a Climate Model. *J. Clim.* **22**, 711–729 (2009).

50. Raymond, D. J. A New Model of the Madden–Julian Oscillation. *J. Atmos. Sci.* **58**, 2807–2819 (2001).

51. Pathak, J. et al. FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. *Arxiv* (2022).

52. Weyn, J. A., Durran, D. R. & Caruana, R. Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data. *J. Adv. Model Earth Sy* **11**, 2680–2693 (2019).

53. Wheeler, M. C. & Hendon, H. H. An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Mon. Weather Rev.* **132**, 1917–1932 (2004).

54. Liebmann, B. & Smith, C. A. Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset. *Bull. Am. Meteorol. Soc.* **77**, 1275–1277 (1996).

55. Hersbach, H. et al. The ERA5 global reanalysis. *Q J. R. Meteor Soc.* **146**, 1999–2049 (2020).

56. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

AUTHOR CONTRIBUTIONS

N.-Y.S., D.K., and J.-S.K. designed the research. N.-Y.S., D.K., and H.K. compiled the data. N.-Y.S. conducted analyses and prepared the figures. N.-Y.S., D.K., and J.-S.K. wrote the first draft of the manuscript, and all authors contributed in the writing of the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

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