Earth Science Informatics (2026) 19:72
https://doi.org/10.1007/s12145-026-02125-7

RESEARCH

Long-term prediction of sea ice concentration with convolutional long short-term memory

Özgür Boran AK[1] · Ertan KURUÖZ[1] · Ecenur ÖZDEMİR[1] · Ayça AK[2]

Received: 12 January 2026 / Accepted: 3 April 2026
© The Author(s) 2026

Abstract

Estimation of sea ice concentration (SIC) is not only important for the Earth's temperature and water balance, but also for the protection of human activities such as ship navigation, oil and gas exploration, and fishing in polar regions. The decrease in SIC directly affects global climate change. Knowing in advance how sea ice will change over time is important in terms of taking the necessary precautions and preventing undesirable situations. In this study, a convolutional long short-term memory (ConvLSTM) model is designed to predict SIC, surface albedo (SAL), and fractional cloud cover (CF) of the South and North Poles for different seasons and months in 2030 and 2040. The model is trained using 46 years of satellite observation data and generates 16-year predictions of SIC, SAL, and CF using a sliding-window strategy. The results indicate a substantial contraction of sea ice concentration in both hemispheres toward 2030 and 2040, consistent with observed historical trends. The proposed framework utilizes artificial intelligence and perceptual metrics to learn spatiotemporal patterns in SIC, SAL, and CF. Therefore, produced outputs should be interpreted as a data-driven scenario extension of satellite observations rather than a deterministic climate prediction. The best model achieved a Structural Similarity Index Measure (SSIM) of 0.9682, a Deep Image Structure and Texture Similarity (DISTS) score of 0.0662, and a Mean Absolute Error (MAE) of 0.0426 in the validation phase for single-year prediction. When predicting the year 2024 using data from 1986 to 2008, the model achieved an SSIM of 0.9239, a DISTS score of 0.1087, and an MAE of 0.0623.

Keywords ConvLSTM · Sea Ice · Climate Change · Artificial Intelligence

Introduction

Global warming is characterized by long-term changes in the average climate conditions that shape Earth's local, regional, and global environments. These changes are associated with a wide range of observable impacts on natural and human systems (NASA 2023). Global warming is a serious environmental problem that can lead to many disasters in the future, if necessary measures are not taken (Koçoğlu and Gökalp 2021). Global warming causes the sea level to rise, large amounts of water masses to evaporate and mix into the atmosphere, and strong winds to occur. This brings heavy rains, storms, and tsunamis. One of these disasters is the increasingly rapid melting of snow and decline of Arctic Sea ice concentration (SIC) (Akın 2013). Arctic ice melting rapidly due to global warming irreversibly disrupts the balance of the ecosystem and changes waterways (Sağlam et al. 2008). At the same time, sea ice affects the voyage of ships, and accurate estimation of sea ice is therefore the basis for safe navigation (Liu et al. 2021). Surface albedo

Communicated by Hassan Babaie

✉ Ayça AK
aycaak@marmara.edu.tr

Özgür Boran AK
ozgurboranak@gmail.com

Ertan KURUÖZ
ertankuruoz@hotmail.com

Ecenur ÖZDEMİR
0ecenur40@gmail.com

[1] Şaban Teoman Duralı Science and Art Centre, İstanbul, Kartal, Türkiye

[2] Marmara University Vocational School of Technical Sciences, İstanbulKartal, 34865, Türkiye

Published online: 18 April 2026

(SAL) and fractional cloud cover (CF) also play an important role in climate change. SAL is the rate of reflected sunlight of total solar radiation. As absorbing generates heat, higher albedo means less warming. Clouds also reflect light before it reaches the ground, causing temperatures to drop.

The problems caused by climate change on Earth and at the poles are being analyzed by many researchers. Scientists aim to prevent this situation by using various methods. The latest technological and scientific developments are needed to understand the problem, make the necessary measurements and predict the effects (Doğan and Tüzer 2011). Climate models are widely used to investigate and simulate the potential impacts of climate change (Çeçen and Güvenç 2022). These models aim to represent the interactions among different components of the climate system by using measurable and observable physical variables (General Directorate of Meteorology 2015).

Predicting SIC over longer lead times remains challenging due to the complex interactions between oceanic, atmospheric, and cryospheric processes. Recent years have witnessed an increasing number of studies leveraging data-oriented techniques with many studies employing machine learning and deep learning methodologies. In Choi et al., the authors develop a short-term forecasting model for Arctic SIC using a recurrent neural network architecture (2019). Specifically, they employ a Gated Recurrent Unit encoder to capture temporal dependencies in SIC data and a feedforward neural network as the decoder to generate the forecasts. In Zheng et al.'s study, a long short-term memory (LSTM) network was employed to capture and encode the temporal structure of the time series, and the resulting representations were subsequently decoded through a deep neural network (Zheng et al. 2022). Kim et al. proposed a 1-month SIC prediction method with eight predictors using Convolutional Neural Networks (CNN) (2020). Liu et al. proposed a daily SIC prediction model based on the ConvLSTM algorithm and found that the single prediction effect of ConvLSTM was better than CNNs (2021).

Ali et al. trained an LSTM model with National Snow and Ice Data Center (NSIDC) and European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5) data and found that it performed better than similar deep learning models (2021). In this study, no visualizations were created. When the LSTM model was compared with a traditional statistical model, it was found that LSTM performed well for 1-month SIC forecasting with an average monthly error of less than 9%. Le et al. used an LSTM neural network for flood forecasting and achieved high accuracy (2019). Handcock and Raphael mathematically and statistically decomposed the temporal change in sea ice and showed that the abnormal deterioration in sea ice is related to phase change (2020). Wang et al. aimed to increase the accuracy by combining Chi-square Test/Clustering (CT) and LSTM models for air quality prediction (2021). Liu et al. compared ConvLSTM and CNN models in predicting monthly changes in SIC and found that the prediction of ConvLSTM is better than that of CNN (2021). Wei et al. achieved more successful results than traditional LSTM models by utilizing different temporal intervals in the training process of Artificial Intelligence (2022). In this study, only SIC was predicted. Madaeni et al. examined CNN and LSTM models to predict the probability of glacier compression and found that these models give the most efficient results when used together (2022). Furthermore, their methodology lacked comprehensive spatial visualizations. Han et al. combined the Improved Sparrow Search Algorithm CNN-LSTM models to predict ice cover thickness and found that this model is much better than sparrow search algorithm and CNN models (2023).

In 2024, Sureja assessed the time series forecasting for five regions in Antarctica and the Arctic by using a combination of ConvLSTM and bidirectional LSTM models, and compared the results with those from a standalone LSTM model (Sureja 2024). In Dong et al.'s study, using satellite-derived SIC records from 1989 to 2016, a ConvLSTM deep neural network was designed to predict Antarctic Sea ice dynamics for the subsequent 60 days (2024). The model demonstrated reliable performance in forecasting the day-to-day spatial distribution of Antarctic SIC for nearly 30 days. Between 2018 and 2022, the best forecasting ability was found in the Australian autumn and winter. The ConvLSTM deep neural network also exhibits strong real-time forecasting performance during February and September, periods characterized by declining Antarctic Sea ice. Zhang et al. proposed an innovative encoder-decoder pyramid extended ConvLSTM network. Their experimental results show that the distribution of SIC predicted by their proposed model is in high agreement with the ground-based model (Zhang et al., 2024). In the paper by Li et al., a ConvLSTM model was used to iteratively forecast monthly series of 6 or 12 months of Arctic SIC reanalysis data (2021). They proved the effect of different seasons on the iterative prediction accuracy of the model.

Numerical forecasting models for SIC often rely on large multi-source datasets and focus primarily on short-term predictions. In contrast, this study develops a ConvLSTM-based spatiotemporal prediction framework designed for long-horizon prediction while maintaining computational efficiency through internal spatial down-sampling and reconstruction mechanisms. The model is trained on a 46-year satellite record and jointly predicts SIC, SAL, and CF. In this way, computational cost is reduced by predicting three variables simultaneously; meanwhile, a three-channel output is ideal for perceptual image-quality metrics. Moreover, instead of limiting the analysis to short-term forecasts,

the proposed framework generates scenario extensions for the years 2030 and 2040. The main contributions of this study can be summarized as follows:

1. A long-term spatiotemporal prediction framework based on ConvLSTM is proposed to predict SIC, SAL, and CF simultaneously for both Arctic and Antarctic regions.
2. Unlike most existing studies focusing on short-term (daily or monthly) forecasting, this work extends the prediction horizon to 2030 and 2040 using a recursive multi-step iteration strategy trained on a 46-year satellite record (1979–2024).
3. A computationally efficient ConvLSTM architecture is designed by incorporating spatial down sampling and transposed convolution layers, enabling long-sequence learning with reduced memory and GPU usage.
4. The study integrates perceptual image-quality metrics (Structural Similarity Index Measure (SSIM), Deep Image Structure and Texture Similarity (DISTS), and Optical Flow Error (OFE)) alongside conventional Mean Absolute Error (MAE) and Mean Squared Error (MSE), providing a more comprehensive evaluation of spatiotemporal prediction performance.

The remainder of this paper is organized as follows. Section II introduces the proposed ConvLSTM-based prediction framework and describes the methodological components of the model. Section III explains the dataset and preprocessing steps used for training and evaluation. Section IV presents the experimental results and discusses the performance of the proposed approach using multiple evaluation metrics. Finally, Section V summarizes the main findings and outlines potential directions for future work.

Methodology

This section presents the methodological framework used to model the prediction of SIC, SAL, and CF. The proposed approach employs a ConvLSTM-based architecture to capture spatial patterns and temporal evolution in long-term satellite observations. To enhance computational efficiency and prediction performance, several architectural design choices are incorporated into the model. The subsequent subsections describe the model structure, dataset preparation procedures, and the training configuration used in the experimental study.

Long Short-Term Memory (LSTM)

LSTM is a deep learning algorithm that makes predictions about time-varying sequential data, such as graphics or video. The inadequacy of the Recurrent Neural Network (RNN) algorithm in long sequences necessitated the development of LSTM. This algorithm can generate another sequential data from one sequential data, as well as make predictions by reducing the sequence to a single data. Figure 1 presents the architecture of an LSTM cell. In this diagram, σ denotes the sigmoid activation function, which is responsible for gating mechanisms within the cell, while *b* represents the bias term added to each gate's linear transformation to improve learning flexibility.

LSTM algorithm consists of unit cells. These cells take the first element of the sequence as input and produce output. This output and the next element of the sequence become the input of the next cell and so on through the entire sequence. In the LSTM algorithm, each cell produces two outputs: the hidden state (h) and the cell state

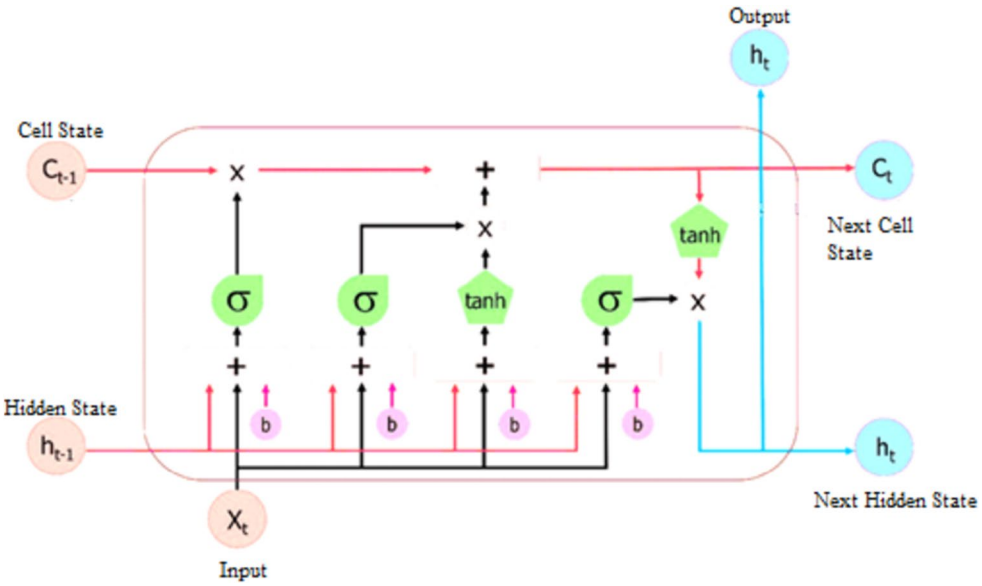


Fig. 1 LSTM cell structure (Le et al. 2019)

(C). Information processing within each LSTM module is modulated by three specialized gates: forget, input, and output, which collectively determine the long-term memory persistence of the network. In the forget gate, the corresponding element of the queue and the latent state output of the previous cell are multiplied by their weights. The results and the bias value are summed and passed through sigmoid activation:

$$f_t = \sigma\left(W_f\left[h_{t-1}, X_t\right] + b_f\right) \tag{1}$$

The parameters W_f and b_f refer to the weight matrix and bias term that control the behaviour of the forget gate. The result is multiplied by the cell state of the previous cell. During the gating operation, element-wise multiplication allows the network to suppress less relevant information by assigning values close to zero. At the input gate, the hidden state of the previous cell is multiplied by the corresponding element of the queue and the weights of the hidden state output of the previous cell. In other words, in the subsequent step, the cell state is not only updated but also stabilized by incorporating and retaining the information from the new input (X_t). This step consists of sigmoid and, secondly, *tanh* processes:

$$N_t = tanh\left(W_n\left[h_{t-1}, X_t\right] + b_n\right) \tag{2}$$

The new cell state is updated through the product of two activation-derived values, which is then summed with the prior memory to establish the current C_t (Antoniadis 2024; Le et al., 2019).

$$C_t = C_{t-1} f_t + N_t i_t \tag{3}$$

where C_t and C_{t-1} *1* denote the cell states corresponding to the time steps t and $t-1$, respectively, b is the bias, and W is the weight matrices (Le et al., 2019). The input gate results are multiplied by the candidate cell state, and the product is summed with the cell state updated by the forget gate to obtain the final cell state output. In the output gate, the related element and the hidden state output of the previous cell are multiplied by their weights. The results and the equilibrium value are summed and sigmoidally activated. Hyperbolic tangent activation is applied to the cell state obtained with the input gate and the latent state output of the cell is obtained by multiplying these two values. By using the output of the last cell, the sequence can be reduced to a single element or the number of elements can be maintained by using the output of each cell.

Convolutional LSTM

CNN is widely used in image processing and classification problems. Deep learning is performed with images in the convolution layer. In the convolution process, a filter, which is a weight array of a certain size and has the same number of channels as the input, is shifted by stride each time and the numbers in the places of the input corresponding to the filter are multiplied by the numbers in the filter, the results and the bias are summed to form a single number corresponding to the position of the filter in the output. Each filter used forms one channel of the output. The output has as many channels as the number of filters used. A stride value greater than 1 ensures that the size of the output is smaller than the input. The kernels (*f*) (or filters) used in the convolutional layer are defined as having a size of $H \times W \times D^l$. Each kernel has a fixed spatial size $H \times W$, which determines how much of the input's spatial area it looks at during convolution, and it spans the full depth of the input, hence the D^l. This means that each filter operates across all channels of the input but only over a small patch spatially. Following this, the convolution can be formulated as follows:

$$y_{i,j,d}^{l+1} = \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} \sum_{k=0}^{D^l-1} f_{m,n,k}^{(d)} . x_{i+m,j+n,k}^{l} \tag{4}$$

In this expression, $y_{i,j,d}^{l+1}$ denotes the output activation at spatial position (i, j) and output channel d in layer $l+1$. The value is computed by sliding a learnable convolutional kernel f^d over the input feature map x^l from the previous layer. The kernel has spatial dimensions $H \times W$ and spans the full depth of the input, denoted as D^l. The indices m, n, and k iterate over the height, width, and depth of the kernel, respectively.

The pooling layer works separately on x^l channels. It partitions each channel into smaller regions of size $H \times W$. After that, a pooling function reduces each region to a single value, thereby lowering the overall dimensionality. Average pooling can be expressed as follows (Sureja 2024):

$$y_{i,j,d}^{l+1} = \underset{0 \le m < H, 0 \le n < W}{\text{mean}} x_{i.H+m,j.W+n,d}^{l} \tag{5}$$

ConvLSTM uses a CNN instead of a one-dimensional neural network in cell gates. In this way, processing can be done with videos instead of one-dimensional number sequences. Transposed convolution calculates the input of a convolution layer with given parameters, which gives the input as output. In this way, image resolution can be increased with

a stride value greater than 1. Transposed convolution is frequently used in generative artificial intelligence and resolution enhancement applications.

In this study, MAE, MSE, SSIM, DISTS, and OFE were used to assess prediction skill. MAE provides a robust assessment of model accuracy by averaging the absolute differences between the estimated variables and the actual measured data. MAE is calculated from the values in the range of [0, 1] as follows (Zhang et al., 2024):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|True_i - Predict_i| \tag{6}$$

where N represents the total number of samples.

MSE provides a second-order measure of accuracy by averaging the squares of the differences between the predicted trajectories and the empirical observations. Compared to MAE, this metric penalizes larger errors more strongly. The formula for MSE is written below:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(True_i - Predict_i)^2 \tag{7}$$

SSIM quantifies the structural integrity of the output images by modelling the interaction between luminance, contrast, and spatial structure. It serves as a more representative measure of visual quality for complex geophysical datasets compared to standard error-based metrics. The SSIM metric operates by evaluating the luminance, contrast, and structural integrity between a reference image X and a target image Y. To quantify the fidelity at each spatial coordinate, the respective comparison functions are defined as (Peng et al. 2020):

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{8}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{9}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{10}$$

where μ_x and μ_y denote the local means, σ_x and σ_y represent the standard deviations and σ_{xy} represent cross-covariance of image X and image Y, respectively. To ensure numerical stability, the coefficients C_1, C_2, and C_3 acts as stabilization constants. The underlying calculations for μ_x, σ_x, and σ_{xy} are defined by:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N}x_i \tag{11}$$

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{\frac{1}{2}} \tag{12}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \tag{13}$$

i is the index of all points in a local area, and N is the total number of points in the same local area. Finally, the SSIM index combines three sub-functions and has the following form:

$$\mathrm{SSIM(x,y)} = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{14}$$

Setting $\alpha=\beta=\gamma=1$ and $C_3=C_2/2$, the final form of the SSIM index is as follows:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_x^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{15}$$

DISTS is a comprehensive full-reference metric that effectively bridges the gap between mathematical error calculation and human visual assessment. This metric excels in balancing global structure and local texture characteristics, which is critical for verifying the realistic representation of sea ice concentration patterns. DISTS is designed to capture 'visual textures,' represented by spatially uniform areas with iterative components, ensuring that the model's output retains its original textural characteristics without significant distortion. DISTS evaluates similarity using deep neural network features and is designed to better capture perceptual differences between images (Li et al. 2025).

OFE can be defined as the average number of pixels in which details are misplaced in predicted images. To calculate OFE, firstly, the optical flows between the true previous year and next year images, and between the true previous year and predictions are computed with the Recurrent All-Pairs Field Transforms (RAFT) small model. The flows contain two channels: the horizontal and vertical movement of each pixel. Each channel of the flows is subtracted to get the horizontal and vertical error matrix. The total error matrix is calculated with Pythagoras' theorem. The final OFE is the mean of this matrix.

RAFT integrates a per-pixel feature encoder with a 4D correlation layer and a recurrent update module to iteratively solve for optical flow. This architecture mimics iterative optimization by retrieving values from pooled correlation volumes to refine a zero-initialized flow field. By replacing traditional handcrafted features with deep-learned representations, RAFT enhances the model's ability to capture complex motion dynamics (Teed and Deng, 2020).

The models have two ConvLSTM layers at the entry of the network. The stride of these layers is 2. In this way, the

network is accelerated by reducing the image size. The former has 64, the latter has 128 channels. They are followed by four ConvLSTM and 3D convolution layers. These layers have 128 channels. The frame numbers are perceived as the third dimension by 3D convolution layers. Stride of these layers is 1. Among these layers, there are residual connections or attention blocks. The convolution kernel size in the model is 4 × 4. This study compares the performance of these two architectures as well as Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (Adam) optimizers. After that, two transpose 3D convolution layers are added. The former has 64, the latter has 3 channels for output. The stride of these layers is (1,2,2). In this way, the image size is restored without increasing the number of frames. The computational efficiency of the proposed models was evaluated during the inference phase. Both the Residual and Attention-based architectures demonstrate high operational efficiency, requiring only 0.32 s on an RTX 5080 GPU to generate a full-year prediction. In terms of hardware resource allocation, the models maintain a stable memory footprint of 4.8 GB RAM, making them suitable for deployment on standard workstation GPUs. This neural network was trained for 100 to 600 epochs, depending on the progress of the network. Three loss algorithms are used to train the networks: MAE, SSIM, and DISTS. MSE and OFE are used to monitor the performance of models. The models are shown in Fig. 2. Images are rescaled to [0, 1] during preprocessing. Details of the designed LSTM model are given in Table 1.

Table 1 Overview of the LSTM model

Objects	Details
Training hyperparameters	SGD default parameters Adam default parameters Epochs: 100–600
Input Variable	SIC, SAL, CF
Prediction goal	Predictions for 2030 and 2040
Data year range	1979 to 2024
Number of images	Total 3312 images
Images size	200 × 200

Data

SIC, SAL, and CF images were used in this study. SIC images were downloaded from the European Organization for Meteorological Satellites (EUMETSAT) Ocean and Sea Ice (OSI), Satellite Application Facility (SAF). SAL and CF images are downloaded from EUMETSAT Climate Monitoring (CM) SAF. Data are downloaded at 1-month intervals from 1979 to 2024. Monthly snapshot data were used for

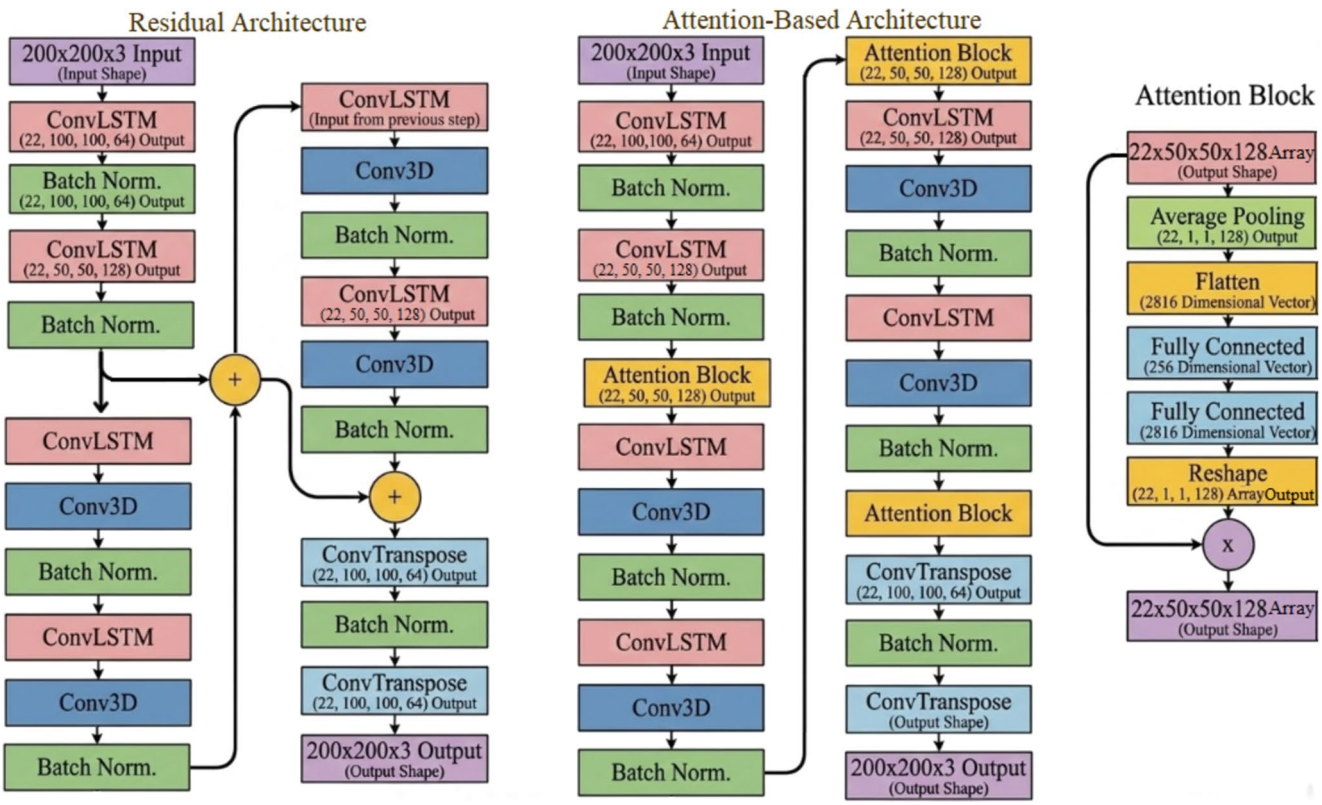


Fig. 2 Designed ConvLSTM neural network structure

SIC, while monthly mean values were used for SAL and CF. Both daily and monthly mean data are available in CM SAF, and the monthly mean data are preferred as they provide a more stable representation of their temporal behavior, reducing short-term fluctuations and noise. On the other hand, only snapshot data is available in OSI SAF, hence snapshot data is used for SIC. Figure 3 illustrates the study area.

In order to compute SIC, firstly, level 1 microwave observations are calibrated to adjust for land propagation effects, ensuring a more precise representation of the geophysical variables across the study area. The first-guess SIC is established via the NASA Team algorithm using static tie-points, followed by an atmospheric correction process driven by radiative transfer and numerical weather data. The final OSI SAF SIC values are calculated using 30-day average tie-points, ensuring that atmospheric noise is minimized and the total error at each spatial coordinate within the satellite swath is recorded (Eumetsat OSI SAF 2023).

The SAL is the integral of the light reflected from a single direction towards all viewing directions in the bow and azimuth planes. The spectral dependence of the surface reflectance is ignored here. Full (black sky) albedo is a broadband surface reflectance. For the Advanced Very High-Resolution Radiometer (AVHRR), it is defined in the wavelength range 0.25–2.5 μm. It is obtained by integrating the spectrally directional hemispherical reflection over the investigated waveband. AVHRR channels 1 and 2 (0.58–0.68 μm and 0.725–1 μm) are used as radiation sources for SAL product generation. The data conditioning stage of satellite data for use with SAL is performed by the Polar Processing System software package. The package converts the observed satellite irradiances into top-of-atmosphere reflections for SAL and performs critical cloud detection (Eumetsat CM SAF 2023).

The CF is directly derived from the results of a cloud detection method. Specifically, it is provided as part of the level-2b binary cloud mask (sampled daily on a global 0.05° grid) and the level-3 average cloud fraction product (available on a global 0.25° latitude/longitude grid). The cloud fraction is defined as the ratio of cloudy pixels to the total number of analyzed pixels within each grid cell, and it is expressed as a percentage. The probabilistic cloud mask, referred to as CMAPROB, serves as the foundation for generating the binary cloud mask and is included in the Level-2b product. This percentage-based representation allows users to flexibly select the reliability threshold of the cloud mask by interpreting cloud fractions on a scale from 0 to 100% (Eumetsat CM SAF 2023).

Signal processing

The software used in this study was developed using the Python programming language. Additionally, the OpenCV, NumPy, and Matplotlib libraries were utilized for data processing, management, and visualization. Keras and PyTorch libraries were used for network training.

During the pre-processing stage, the obtained images were converted to grayscale, reducing the channel dimensionality from three to one while preserving the essential spatial information. The size of the images was reduced to 200×200 due to the difficulty of training at high resolution. Different types of images were added one after the other to obtain $200 \times 200 \times 3$ arrays. The data is divided into two pieces: from 1979 to 2001 and from 2002 to 2024.

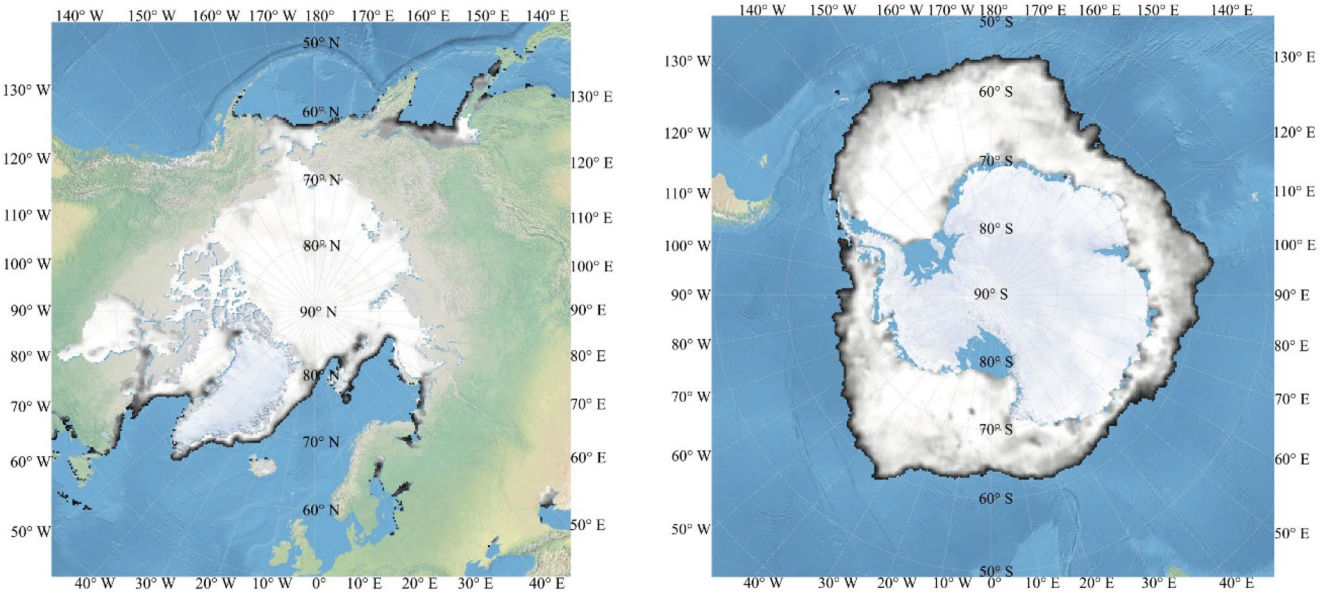


Fig. 3 The study area

The data of June for the years 1979 to 2001 and the data of January for the years 2002 to 2024 are separated for validation. For the North and South Poles, 48 arrays with $23 \times 200 \times 200 \times 3$-dimensions were obtained by combining the images belonging to the same month in different years. These arrays consist of images of one month with one year interval for a period of 23 years. The ConvLSTM network was trained and validated using these arrays. The model uses 22 years of data to predict the next year. For example, the model predicts 2003 to 2025 data using the input of 2002 to 2024. All training and inference procedures strictly preserve chronological order, and no future observations are used as inputs during model training or validation. The model is trained to generate the data of 2003 to 2024 from 2002 to 2023 and 1980 to 2001 from 1979 to 2000. The generated data from 2003 to 2024 and 1980 to 2001 is only for measuring the performance of the model during training and validation. In inference, the predicted 2025 data are added to the 2002 to 2024 sequence, while 2002 is removed to get the 2003 to 2025 data. This process is continued until 2040.

The input length $L = 22$ years was selected as a compromise between capturing decadal-scale variability and maintaining a sufficient number of training samples under the fixed 1979–2024 record. Shorter windows degraded the representation of low-frequency variability, while longer windows reduced the number of effective training pairs and increased training instability. For each hemisphere and month-specific sequence, input–target pairs are constructed using a causal setup, where an input window of $L = 22$ years is mapped to the subsequent year. During training, all targets belong to years after their corresponding input windows, and no future frames are included in the inputs. Validation is performed on held-out month–year sequences that are never used for weight updates. For multi-step inference (up to 2040), recursive prediction is applied. The model first predicts year $t+1$ from $[t-L+1,\ldots,t]$; then the prediction x^{t+1} is appended to the window, and the oldest year is dropped to predict $t+2$. This sliding window approach is shown in Fig. 4.

Results

In this study, several models are developed to predict future values based on different historical time windows. Each model uses a fixed number of past years as input and generates predictions for one or more future years, aiming to improve both short and long-term predictive accuracy. The following section summarizes the models and their training strategies.

The ConvLSTM neural network was trained with SGD and Adam optimizers. Default parameters of optimizers are preserved. Batch size is configured as 1. Early-stopping is involved during training to avoid unnecessary training. The study was done for model architectures with residual or attention and with loss functions MAE, SSIM and DISTS. 1-SSIM is used as the loss function to be minimized. The results are presented in Table 2. Attention-based models trained with the Adam optimizers demonstrated better performance compared to the other models. Adam-residual, SGD-attention, and SGD-residual models behaved similarly. Since SSIM captures perceptual differences in structure, luminance, and contrast rather than pixel-level error or textures, it performed better at teaching features such as sea ice edge, land, and distribution of albedo. Metrics of the cross-validated attention model trained with Adam and SSIM for 200 epochs are shown in Table 3. Estimated metric values for 2024, based on data from 1986 to 2008, are presented in Table 4.

The learning curves for various evaluation metrics, including SSIM and OFE, obtained with the Adam optimizer and SSIM-based loss weighting, are presented in Fig. 5.

Figure 6 illustrates the variation in metrics relative to the number of lead years when predicting the year 2024 using data spanning from 1986 to 2008.

Figure 7 presents the comparison between observed 2024 images and the predictions generated using 1986–2008 input sequences. The results allow the performance of different model configurations to be visually evaluated.

The predicted images are slightly blurred. However, the network demonstrates the ability to predict the overall progress of SIC, SAL, and CF. The predictions of SAL and CF are more erroneous than sea ice predictions. The reason why generated SIC images are clearer and more accurate than SAL and CF images is that SAL and CF images are only available for day-time while SIC observations are made all year. The overall prediction errors for the future steps remain relatively small. As expected, the first-step predictions are the closest to the observed

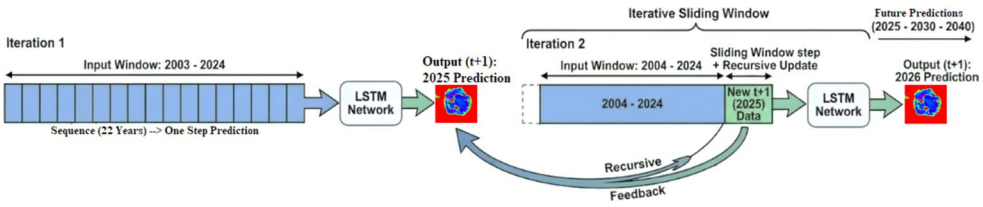


Fig. 4 Sliding window used in the model

Table 2 Performance of models

Archit.	Opt.	Loss Fn	Train MAE	Train MSE	Train OFE	Train DISTS	Train SSIM	Valid MAE	Valid MSE	Valid OFE	Valid DISTS	Valid SSIM	Eph.
Attention	Adam	SSIM	0.0428	0.0058	0.4096	0.0579	0.9712	0.0426	0.0061	0.4202	0.0662	0.9682	199
Residual	Adam	SSIM	0.0446	0.0047	0.2965	0.0432	0.9791	0.0443	0.0046	0.2756	0.0464	0.9788	96
Attention	SGD	SSIM	0.0693	0.0134	0.7603	0.1066	0.9376	0.0671	0.0127	0.8377	0.1121	0.9391	186
Residual	SGD	SSIM	0.0563	0.0089	0.5099	0.0784	0.9570	0.0535	0.0081	0.5236	0.0807	0.9593	200
Attention	Adam	DISTS	0.0596	0.0097	0.5106	0.0851	0.9496	0.0574	0.0095	0.5397	0.0893	0.9489	100
Residual	Adam	DISTS	0.0348	**0.0026**	**0.2364**	**0.0337**	**0.9863**	0.0339	**0.0025**	**0.2436**	**0.0358**	**0.9860**	96
Attention	SGD	DISTS	0.0816	0.0173	0.9223	0.1251	0.9194	0.0781	0.0161	1.1429	0.1297	0.9226	600
Residual	SGD	DISTS	0.0727	0.0141	0.7081	0.1065	0.9315	0.0686	0.0127	0.7378	0.1099	0.9355	600
Attention	Adam	MAE	0.0553	0.0105	0.4689	0.0927	0.9402	0.054	0.0108	0.4969	0.1070	0.9379	192
Residual	Adam	MAE	**0.0313**	0.0036	0.3330	0.0427	0.9800	**0.0308**	0.0035	0.3295	0.0445	0.9797	96
Attention	SGD	MAE	0.0690	0.0174	0.9216	0.1439	0.9163	0.0647	0.0158	1.1229	0.1461	0.9209	186
Residual	SGD	MAE	0.0563	0.0125	0.6511	0.1130	0.9402	0.0522	0.1120	0.6385	0.1185	0.9343	199

image. As predictions move further in the future, the differences with the observed images slightly increase. The selection of the proper model is critical for developing an accurate prediction model. The attention model trained with Adam and SSIM generated images with better quality.

Figure 8 shows the observed and predicted SIC maps for January, May, and September in the Northern Hemisphere for the years 2030 and 2040. The predictions reveal seasonal variations in sea ice distribution across the selected months. In this color scheme, iceless sea is shown as red, lands are shown green, sea ice with high concentration is shown blue, and sea ice with low concentration is shown green. The predictions generated by the proposed ConvLSTM model indicate a significant contraction trend in SIC across both Arctic and Antarctic regions over the next two decades. Analysis of the Northern Hemisphere predictions in Fig. 8 reveals a pronounced retreat of sea ice edges toward the central Arctic, particularly during September minima. The transformation of the high-concentration core ice mass (indicated by blue regions) into a more fragmented structure aligns with "ice-free summer" scenarios documented in recent literature (e.g., Kim et al. 2020).

SIC observation and prediction images for the Southern Hemisphere in July, November, and March are given in Fig. 9. Figure 10 illustrates the predicted SAL distributions for different months, providing insight into the spatial variability of salinity patterns. Low to high albedo is displayed as blue, green, brown, yellow, and white colors, respectively. SAL predictions are consistent with the reduction in SIC, indicating a diminished reflective capacity of the surface, subsequently increasing solar radiation absorption by the ocean.

SAL observation and prediction images for the Southern Hemisphere in February and October are given in Fig. 11.

The model suggests critical SAL declines along Arctic coastlines by 2040, and the convergence of decay areas of SIC and SAL, suggests that the "ice-albedo feedback" mechanism may has been implicitly reflected by the deep learning architecture.

Figure 12 presents the predicted CF patterns for the selected months in Northern Hemisphere, allowing the spatial distribution of CF to be analyzed. Clear and cloudy sky is shown as blue and white, respectively.

CF observation and prediction images for the Southern Hemisphere and February, October are given in Fig. 13. While CF predictions exhibit a slight downward trend, indicating shifts in atmospheric moisture transfer and radiative forcing.

Figure 14 summarizes the mean spatial distributions of SIC, SAL, and CF derived from the predicted images.

Table 3 The metrics of the cross-validated attention model trained with Adam and SSIM for 200 epochs

Train MAE	Train MSE	Train OFE	Train DISTS	Train SSIM	Valid. MAE	Valid. MSE	Valid. OFE	Valid. DISTS	Valid. SSIM
0.0489	0.0075	0.4144	0.0673	0.9619	0.052	0.0087	0.4297	0.0725	0.9553

Table 4 Metric values when 2024 is predicted from the data from 1986 to 2008

Architecture	Optimizer	Loss Fn.	MAE	MSE	OFE	DISTS	SSIM
Attention	Adam	SSIM	**0.0623**	**0.0157**	**0.7385**	**0.1087**	**0.9239**
Residual	Adam	SSIM	0.1106	0.0348	1.8827	0.1518	0.8546
Attention	SGD	SSIM	0.0822	0.0175	1.0244	0.1465	0.9126
Residual	SGD	SSIM	0.0972	0.0242	1.1423	0.1604	0.8889
Attention	Adam	DISTS	0.0803	0.0215	0.8349	0.1410	0.9005
Residual	Adam	DISTS	0.1214	0.0335	1.3145	0.1476	0.8572
Attention	SGD	DISTS	0.0982	0.0226	1.2259	0.1708	0.8897
Residual	SGD	DISTS	0.1029	0.0225	1.6275	0.1687	0.8803
Attention	Adam	MAE	0.0816	0.0287	0.9467	0.1612	0.8632
Residual	Adam	MAE	0.1132	0.0369	1.0655	0.2201	0.8030
Attention	SGD	MAE	0.0846	0.0228	1.4746	0.1807	0.8882
Residual	SGD	MAE	0.0824	0.0226	1.4487	0.1837	0.8911

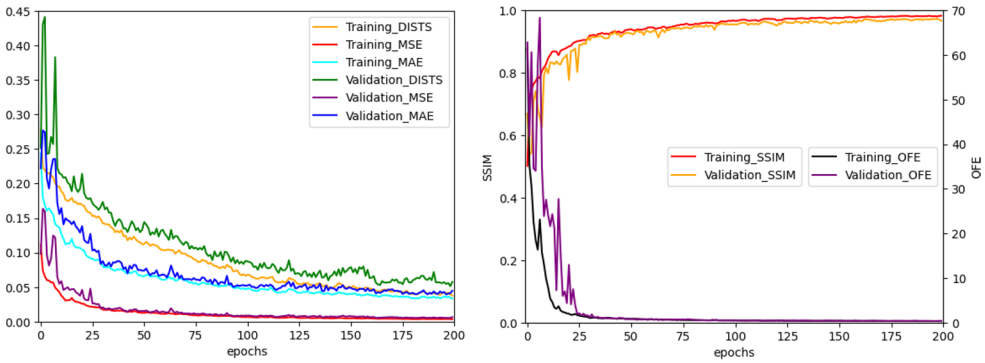


Fig. 5 Learning curves for error and similarity metrics

(a) DISTS, MSE, and MAE loss convergence (b) SSIM and OFE performance during training

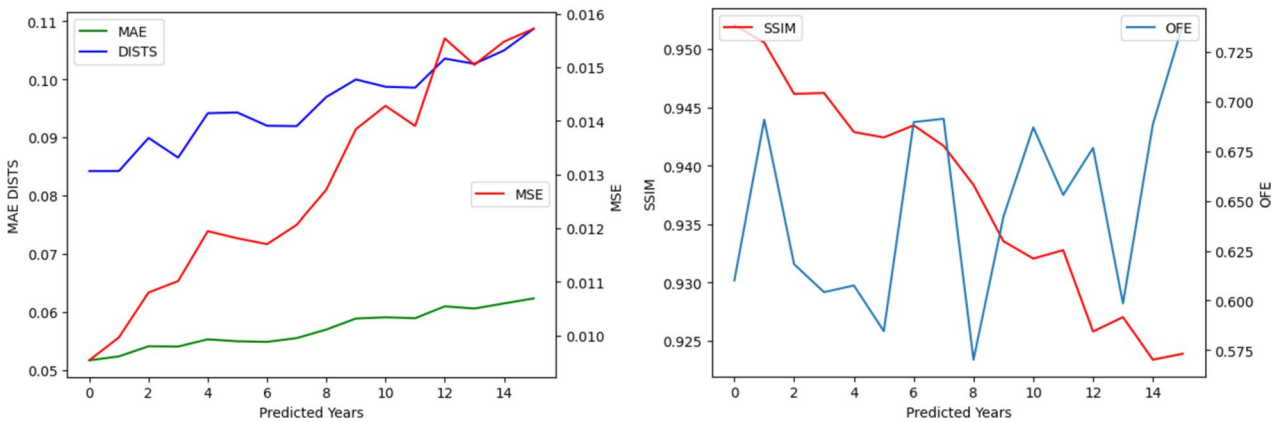
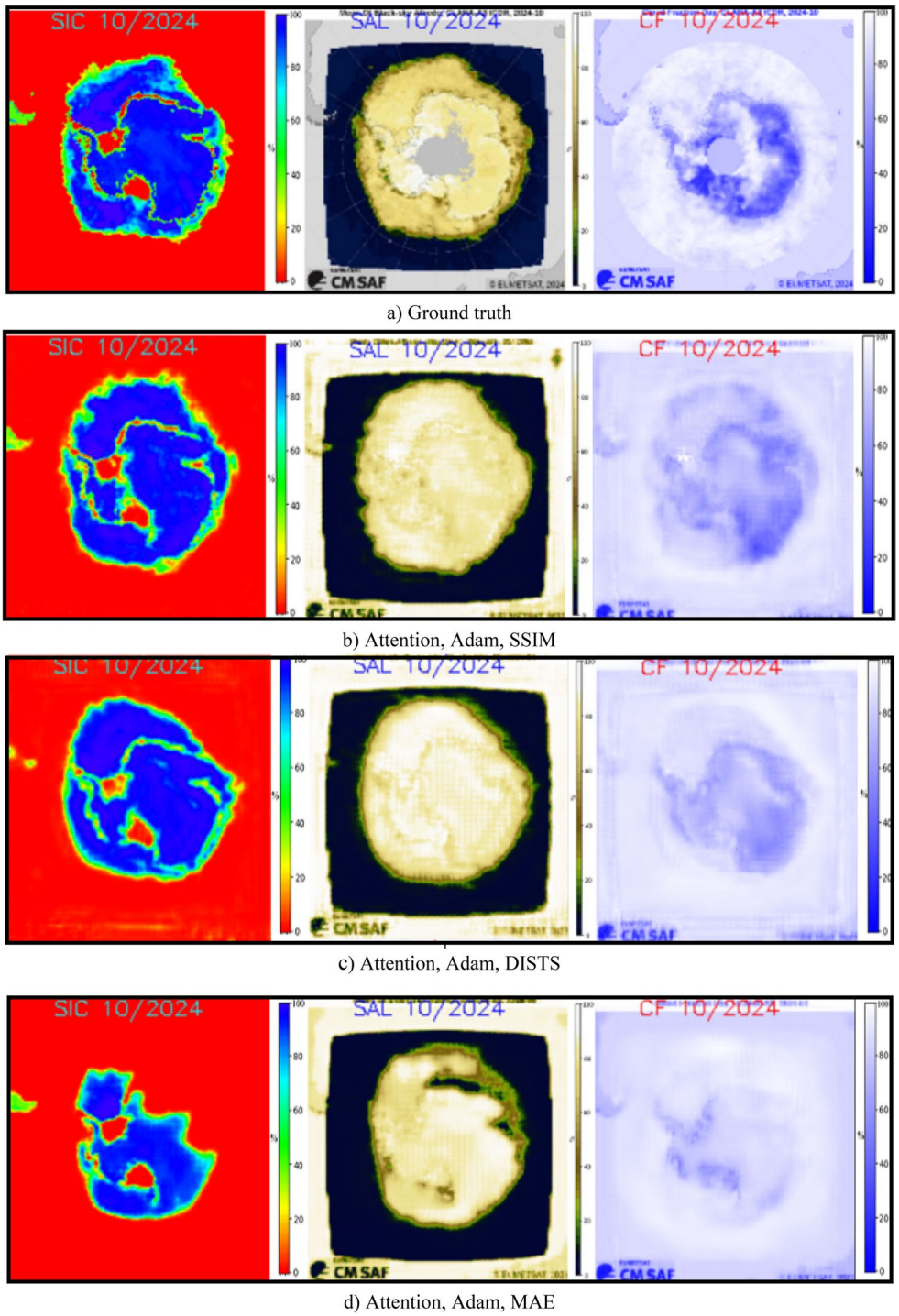


Fig. 6 Plots of metrics according to the number of predicted years



a) Ground truth

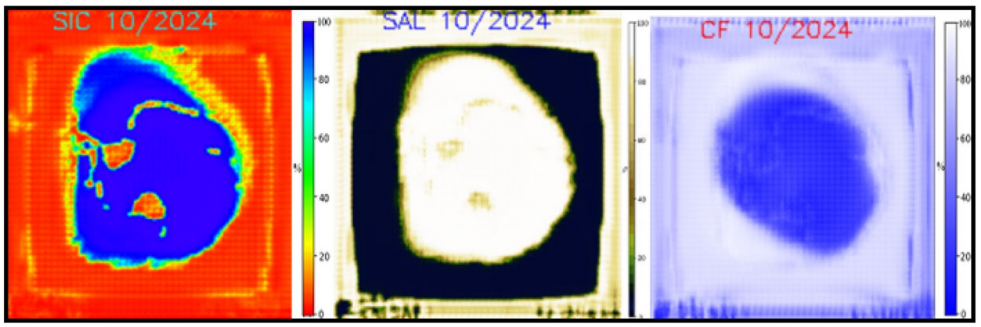
b) Attention, Adam, SSIM

c) Attention, Adam, DISTS

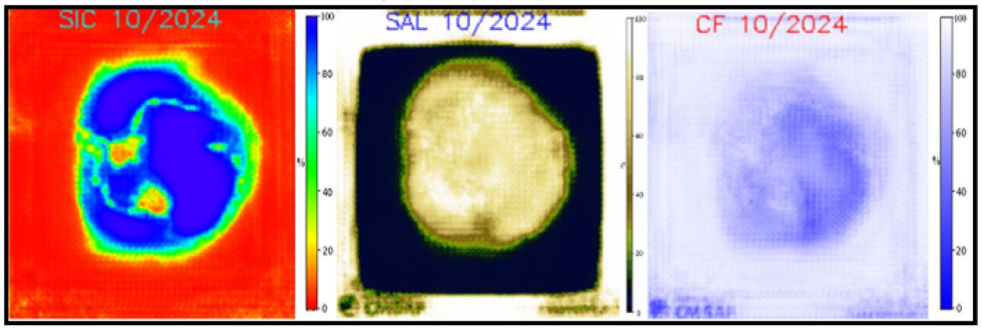
d) Attention, Adam, MAE

Fig. 7 Comparison of observed 2024 images and model predictions using 1986–2008 input sequences

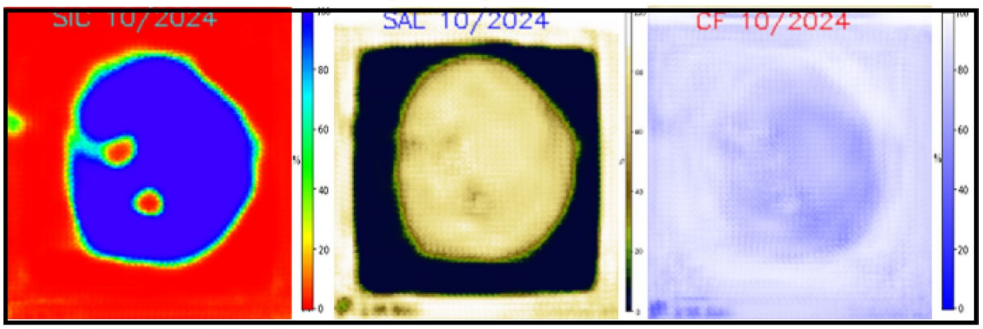
In both Hemispheres (Fig. 14), a persistent downward trend in SIC means is observed. For instance, the mean of the images decreased from 0.3069 to 0.3011 in the Northern Hemisphere and from 0.2007 to 0.1860 in the Southern Hemisphere in the period from 2024 to 2040. A less significant decay trend in SAL is observed in both hemispheres. The fluctuations continued in CF in both hemispheres, while a dwindling is present in the Southern



e) Residual, Adam, DISTS



f) Residual, SGD, SSIM



g) Attention, SGD, MAE

Fig. 7 (continued)

Hemisphere. This can be attributed to the fact that CF is characterized by high-frequency stochastic atmospheric fluctuations. The model successfully captured the historical inter-annual variability and predicted a continued decline into the 2030–2040 decade. This suggests that the ConvLSTM architecture has effectively learned recurring spatiotemporal patterns associated with long-term variability in the historical satellite record (1979–2024). While SIC predictions show high stability, the CF and SAL time series exhibit higher volatility (Fig. 14b and d). This discrepancy may be attributed to data constraints of daytime-only satellite observations for SAL and CF, as noted in the methodology. The fluctuations in CF and SAL before 2000 are due to missing places in images owing to early-generation satellite technology, as can be seen in Figs. 11, 12 and 13.

The concurrent decline in SIC and SAL is consistent with the ice–albedo feedback mechanism, whereby reduced ice cover lowers surface reflectivity and enhances solar energy absorption by the ocean. The model's ability to reproduce

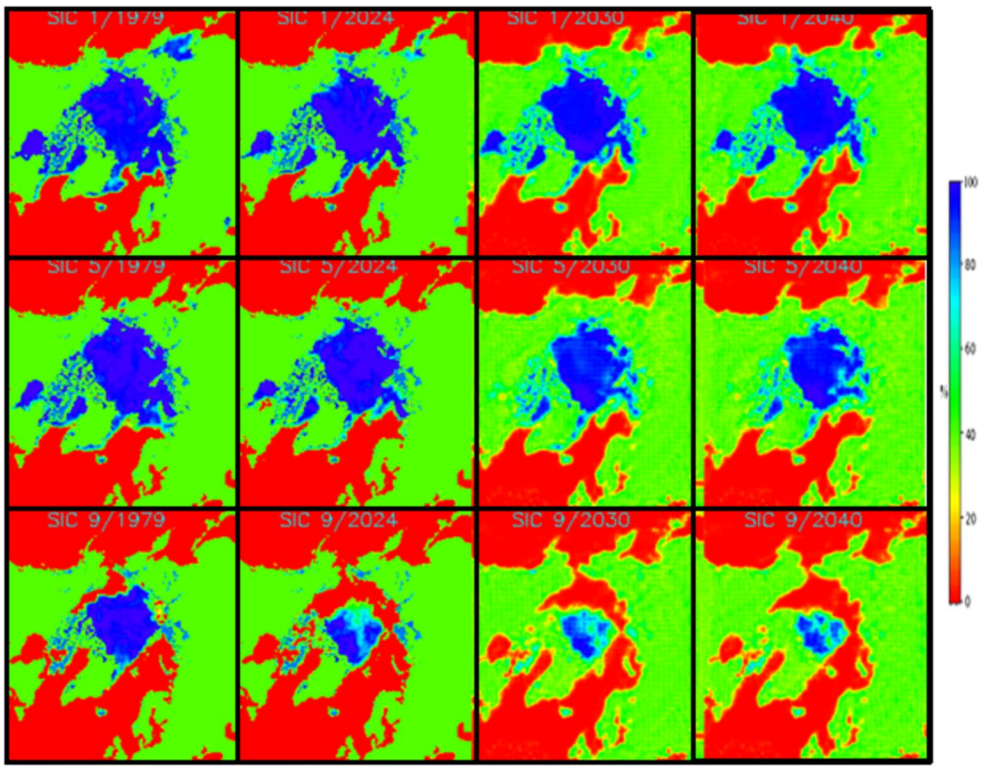


Fig. 8 Observed and predicted SIC in the Northern Hemisphere

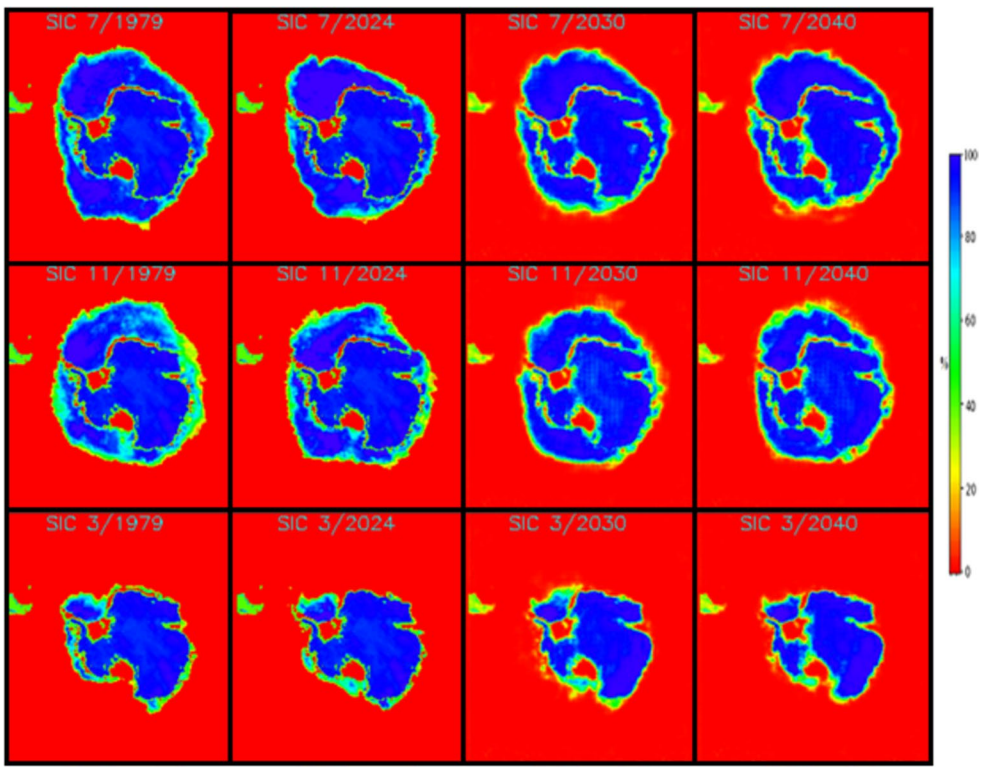


Fig. 9 Observed and predicted SIC in the Southern Hemisphere

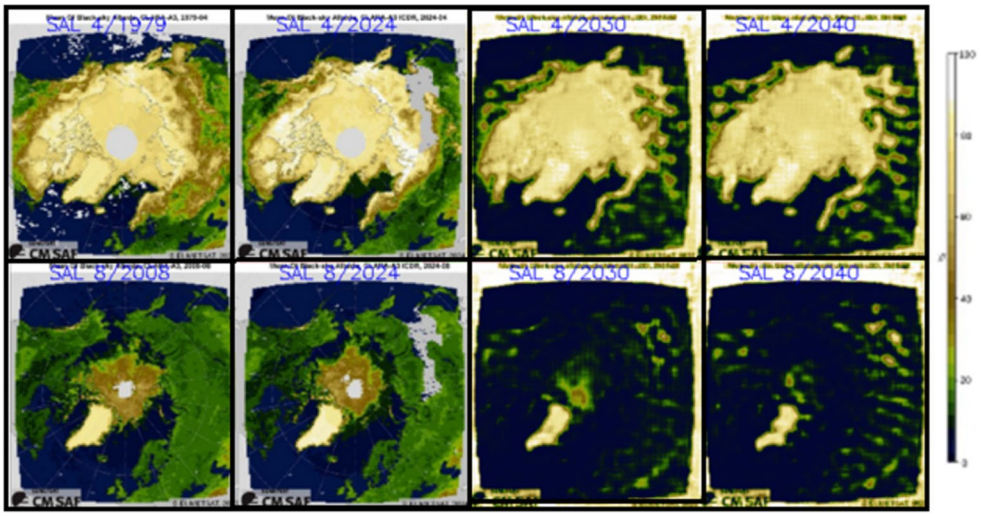


Fig. 10 Observed and predicted SAL in the Northern Hemisphere

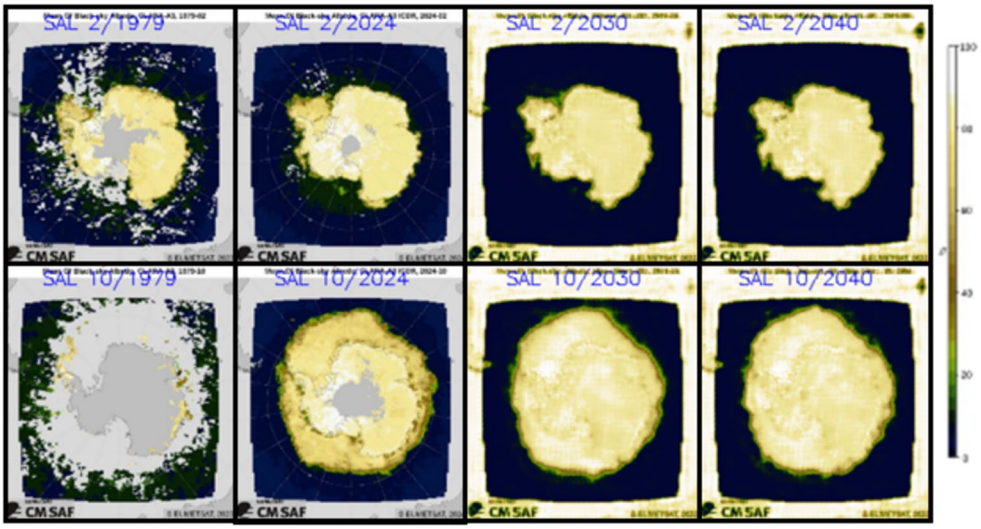


Fig. 11 Observed and predicted SAL in the Southern Hemisphere

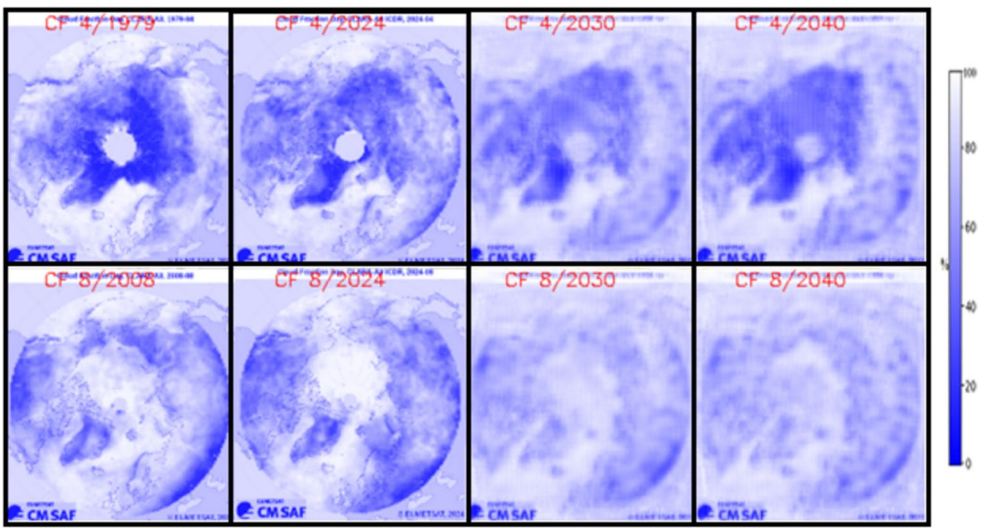


Fig. 12 Observed and predicted CF in the Northern Hemisphere

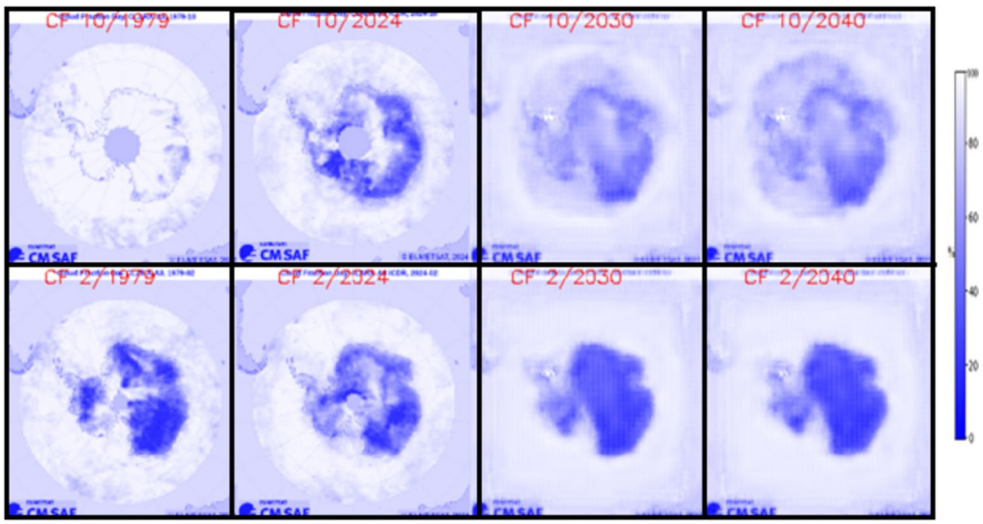


Fig. 13 Observed and predicted CF in the Southern Hemisphere

this coupled behavior suggests that it may have implicitly learned physically meaningful relationships from the historical satellite record, despite the absence of explicit physical constraints.

Conclusion

SIC is important in terms of climate change, sea level changes, and maritime shipping. Specifically, estimating regional SIC would improve planning by taking into account the unique climatic conditions and the importance of the regions surrounding the Antarctic and the Arctic. Predicting SAL and CF is also important for climate research.

In this study, a ConvLSTM-based approach was used to predict SIC, SAL, and CF using a long-term dataset covering 46 years. This methodology allows the model to capture emergent patterns in the data while maintaining computational efficiency. Different model architectures, optimizers, and loss functions were evaluated to analyze their effects on prediction performance. The results indicate that the attention-based model trained with the Adam optimizer and SSIM loss function produced visually higher-quality predictions compared to the other tested configurations. Analysis of the predicted maps suggests a decrease in SIC and slight declines in SAL and CF in both hemispheres. The model does not explicitly enforce physical conservation laws or atmospheric–oceanic energy balance constraints. Therefore, the generated long-horizon outputs should be interpreted as data-driven scenario extensions, rather than physically constrained climate projections. In addition, one limitation of this study is that SAL and CF datasets are derived from daytime satellite observations, which may not fully capture diurnal variability.

Overall, the model introduced in this study represents a notable advancement in SIC prediction, which will aid climatology research and support economic planning in polar regions. The findings indicate a substantial potential for the use of artificial intelligence methods and perceptual loss functions in SIC, SAL, and CF prediction.

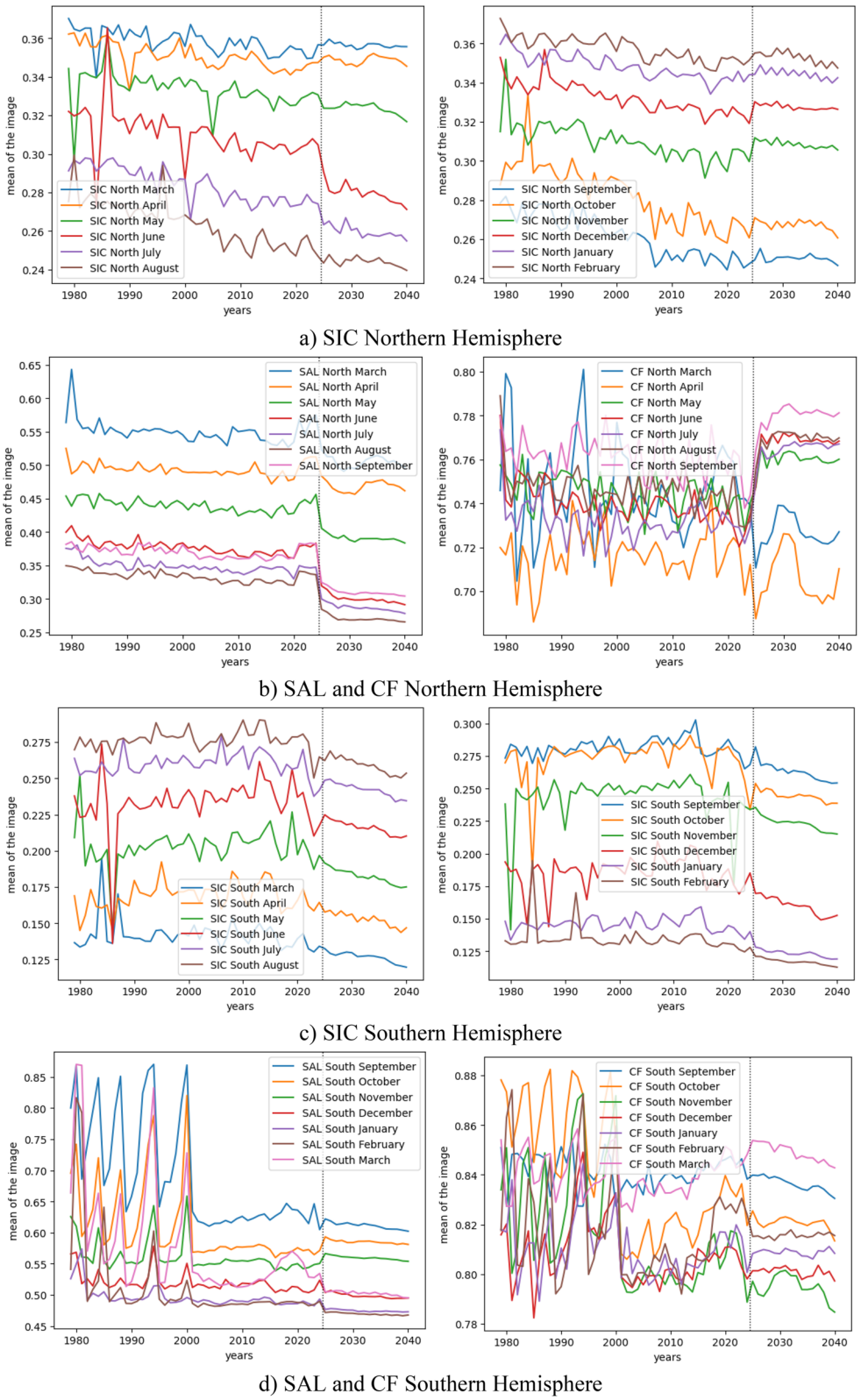


Fig. 14 Plots of the mean of the images

Author Contributions Ö. B. K: Writing – review & editing, writing – original draft, visualization, software, resources, methodology, data curation, conceptualization. E. K: Project administration, supervision. E. Ö.: Investigation, resources, writing. A. A.: Writing – review & editing.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data Availability Data sharing is not applicable to this article as no new datasets were created during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Akın G (2013) One of the main problems of our century is the melting of glaciers. Anthropology 25:9–27

Ali S, Huang Y, Huang X, Wang J (2021) Sea ice forecasting using attention-based ensemble LSTM. Atmospheric and Oceanic Physics. arXiv:2108.00853. https://doi.org/10.48550/arXiv.2108.00853

Antoniadis P (2024) Activation Functions: Sigmoid vs Tanh. https://www.baeldung.com/cs/sigmoid-vs-tanh-functions

Çeçen Z, Güvenç F (2022) Evaluation of climate change and global warming in terms of public health. SDU J Health Manage 4(1):14–25

Choi M, De Silva LWA, Yamaguchi H (2019) Artificial neural network for the short-term prediction of arctic sea ice concentration. Remote Sens 11(9):1071. https://doi.org/10.3390/rs11091071

Doğan S, Tüzer M (2011) Combating global climate change: general approaches and international efforts. Istanbul Journal of Sociological Studies 44:157–194. https://izlik.org/JA48KE93LE

Dong X, Yang Q, Nie Y, Zampieri L, Wang J, Liu J, Chen D (2024) Antarctic sea ice prediction with a convolutional long short-term memory network. Ocean Model 190:102386. https://doi.org/10.1016/j.ocemod.2024.102386

EUMETSAT CM SAF (2023) https://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=CLARA_AVHRR_V003

EUMETSAT OSI SAF (2023) https://osi-saf.eumetsat.int

General Directorate of Meteorology (2015) Global Climate Models and Global Climate Projections https://www.mgm.gov.tr

Han Z, Lv H, Liang Z, Yi J (2023) Transmission line icing thickness prediction model based on ISSA-CNN-LSTM. J Phys Conf Ser 2588(1):012020. https://doi.org/10.1088/1742-6596/2588/1/012020

Handcock MS, Raphael MN (2020) Modelling the annual cycle of daily antarctic sea ice extent. Cryosphere 14(7):2159–2172. https://doi.org/10.5194/tc14-2159-2020

Kim YJ, Kim H, Han D, Lee S, Im J (2020) Prediction of monthly arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. Cryosphere 14:1083–1104. https://doi.org/10.5194/tc-14-1083-2020

Koçoğlu E, Gökalp L (2021) Analysis of postgraduate theses on global warming in Turkey: a meta-synthesis study. J East Geogr 26(46):129–142. https://doi.org/10.17295/ataunidcd.927150

Le XH, Ho HV, Lee G, Jung S (2019) Application of long short-term memory (LSTM) neural network for flood forecasting. Water 11(7):1387. https://doi.org/10.3390/w11071387

Li P, Leng H, Ma B, He M (2021) Monthly arctic sea ice concentration prediction based on the ConvLSTM model. Int Conf Comput Inform Sci Artif Intell. https://doi.org/10.1109/CISAI54367.2021.00038

Li Y, Han H, Han G, Jian P (2025) Physics statistical descriptor-informed deep image structure and texture similarity metric as a generative adversarial network optimization criterion for three-dimensional gray-scale core reconstruction. Appl Sci 15(16):8886. https://doi.org/10.3390/app15168886

Liu Q, Zhang R, Wang Y, Yan H, Hong M (2021) Daily prediction of the arctic sea ice concentration using reanalysis data based on a convolutional LSTM network. J Mar Sci Eng 9(3):330. https://doi.org/10.3390/jmse9030330

Madaeni F, Chokmani K, Lhissou R, Homayouni S, Gauthier Y, Tolszczuk-Leclerc S (2022) Convolutional neural network and long short-term memory models for ice-jam predictions. Cryosphere 16(4):1447–1468. https://doi.org/10.5194/tc-16-1447-2022

NASA (2023) What is climate change? https://climate.nasa.gov/what-is-climate-change/

Peng J, Shi C, Laugeman E, Hu W, Zhang Z, Mutic S, Cai B (2020) Implementation of the structural SIMilarity (SSIM) index as a quantitative evaluation tool for dose distribution error detection. Med Phys 47(4):1907–1919. https://doi.org/10.1002/mp.14010

Sağlam NE, Düzgüneş E, Balık İ (2008) Global warming and climate change. Ege J Fisheries Aquat Sci 25(1):89–94. https://izlik.org/JA85GX83HJ

Sureja S (2024) Arctic and Antarctic Sea ice extent forecasting using hybrid LSTM techniques (2024) PREPRINT (Version 1) available at Research Square. https://doi.org/10.21203/rs.3.rs-4316516/v1

Teed Z, Deng J (2020) RAFT: Recurrent all-pairs field transforms for optical flow. ArXiv. https://doi.org/10.1007/978-3-030-58536-5_24. abs/2003.12039

Wang J, Li J, Wang X, Wang J, Huang M (2021) Air quality prediction using CT-LSTM. Neural Comput Appl 33:4779–4792. https://doi.org/10.1007/s00521-020-05535-w

Wei J, Hang R, Luo JJ (2022) Prediction of Pan-Arctic Sea ice using attention-based LSTM neural networks. Front Mar Sci 9:860403. https://doi.org/10.3389/fmars.2022.860403

Zhang D, Wang C, Huang B, Ren J, Zhao J, Hou G (2024) PDED-ConvLSTM: pyramid dilated deeper encoder–decoder Convolutional LSTM for Arctic Sea ice concentration prediction. Appl Sci 14(8):3278. https://doi.org/10.3390/app14083278

Zheng Q, Li W, Shao Q, Han G, Wang X (2022) A mid- and long-term Arctic Sea Ice concentration prediction model based on deep learning technology. Remote Sens 14(12):2889. https://doi.org/10.3390/rs14122889

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