

Multi-Task Deep Learning Based Spatiotemporal Arctic Sea Ice Forecasting

Eliot Kim*, Peter Kruse[†], Skylar Lama[‡], Jamal Bourne Jr.[§], Michael Hu[¶], Sahara Ali^{||}, Yiyi Huang^{||}, Jianwu Wang^{||}
*Nelson Institute of Environmental Studies Department of Statistics, University of Wisconsin-Madison, Madison, WI, United States

Email: ejkim23@wisc.edu

[†]Department of Accounting, Business, and Economics, Juniata College, Huntingdon, PA, United States

Email: krusepi18@juniata.edu

[‡]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, United States

Email: slama@umd.edu

[§]Department of Mathematics and Computer Science, McDaniel College, Westminster, MD, United States

Email: jeb0100@mcDaniel.edu

[¶]College of Computing, Georgia Institute of Technology, Atlanta, GA, United States

Email: mhu94@gatech.edu

^{||}Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, United States

Email: {sali9,yhuang10,jianwu}@umbc.edu

Abstract—Critical natural resources and processes in the Arctic depend heavily on sea ice. Thus, accurate and timely predictions of Arctic sea ice changes is important. Arctic sea ice forecasting involves two connected tasks: predicting sea ice concentration (SIC) at each pixel and predicting overall sea ice extent (SIE). Instead of having two separate models for these two forecasting tasks, in this paper we study how to use multi-task learning techniques and leverage the connections between ice concentration and ice extent to improve accuracy for both forecasting tasks. Because of the spatiotemporal nature of the data, we designed two novel multi-task learning models based on the CNN and ConvLSTM, respectively. Further, in conjunction with multi-task models, we developed custom loss functions which train the models to ignore land pixels and optimize for both concentration and extent when making predictions. Our experiments show that multi-task models provide better accuracy for a 1-month lead time than models that predict sea ice extent and concentration separately. Our accuracies are better than or comparable to results in related state-of-the-art studies. Our best model in SIC prediction outperformed the best existing SIC prediction model in the literature with 1.78% less error, and our best model in SIE prediction outperformed the best existing SIE prediction model with 0.283 million km² less error.

Index Terms—arctic sea ice forecasting, convolutional neural network (CNN), convolutional long short-term memory (ConvLSTM) network, multi-task deep learning, custom loss function

I. INTRODUCTION

Arctic sea ice variations drive atmospheric processes, oceanic circulation, and polar ecosystems. The albedo of sea ice is much larger than that of open ocean. Thus, sea ice regulates the Earth's temperature by reflecting sunlight away from the surface [21], [38]. However, since 1981, Arctic sea

ice extent has declined at a rate of 13.1% per decade. This trend has accelerated in recent years. In September 2012, Arctic sea-ice extent (SIE) reached a record low of 3.57 million km² [37]. If current trends continue, the Arctic Ocean may be ice-free by the middle of this century based on climate model projections [5], [25], [31].

Melting of Arctic sea ice poses significant challenges for local and global communities. Decreased ice extent increases absorption of solar radiation by the ocean, which results in warmer sea surface temperatures. This in turn accelerates sea ice melting, leading to a feedback loop of warming [19]. Furthermore, reduced ice formation inhibits oceanic circulation and the transport of heat between continents [34]. On the local scale, sea ice loss threatens native species and the livelihoods of Indigenous people [9], [32]. In light of these consequences, accurate predictions of future Arctic sea ice levels are essential for planning mitigation and resilience measures for climate change.

In this paper, we address the following two main challenges in arctic sea ice forecasting. First, although there is a relationship between sea ice concentration and sea ice extent, researchers have not yet created a model that can accurately and simultaneously predict both spatial sea ice concentration (SIC) images and temporal sea ice extent (SIE) values. Additionally, remote sensing techniques often produce noisy values, making it difficult for models to discriminate between land, sea and ice pixels, resulting in increased prediction error.

The main goal of this study was to apply novel deep learning methods to improve on the best sea ice prediction accuracies in the literature with the following contributions. The software implementations of our work are open-sourced at [1].

- We propose multi-task deep learning models, namely Multi-Task CNN and ConvLSTM, to simultaneously predict both SIC and SIE. Each model is able to learn both

Peter Kruse and Eliot Kim contributed equally to this work as first authors. Skylar Lama and Jamal Bourne Jr. contributed equally to this work as second authors.

spatial information from SIC and temporal information from SIE, and thus able to improve prediction ability for both metrics. The results indicate that this multi-task method improves ice forecasts compared to deriving extent based on concentration or training two separate models for concentration and extent.

- We propose a custom loss function in our models to focus on relevant pixels by applying a land mask to each epoch's predicted SIC image. Models are thus able to optimize more efficiently on the grid cells where ice formation is possible.
- We conducted extensive experiments to evaluate our work. Results show our approaches can attain comparable or lower SIC and SIE errors than baseline methods and related studies. The results confirm the ability of deep learning methods to predict Arctic sea ice trends, and the ability of multi-task models to provide the optimal accuracy.

The rest of the paper is organized as follows. We first give an overview of numerical, statistical, and deep learning methods used in the literature to forecast Arctic sea ice in Section II. This is followed by a description of the dataset in Section III. Details about the deep learning model architectures and model training and testing are provided in the Methods section (Section IV). A discussion of our SIC and SIE prediction results and comparisons with results from related studies is in Section V. Lastly, Section VI concludes this work.

II. BACKGROUND AND RELATED WORK

A. Numerical and Statistical Methods

Recent studies have utilized numerical, statistical, and machine learning methods to predict Arctic sea ice. Using CFSv2 (Climate Forecast System, Version 2), a physics-based numerical climate model, to predict September SIE for 2005 through 2014, Collow et al. obtained an RMSE in the range of 0.55-0.65 million km² [39]. Wang et al. used CFSv2 to conduct a detailed analysis of SIE prediction ability for 1982 through 2007, and obtained RMSE values between 0.2 and 0.6 million km² [42]. These numerical studies have attained reasonable SIE predictions, but the computational resources and complex physics expertise necessary to harness dynamical models inhibit ready use of these methods.

Statistical methods, while much simpler to implement and interpret than dynamical models, show limited potential for predicting Arctic sea ice. Regression techniques have provided adequate sea ice predictions results for up to 7 month lead times, using only sea ice input [23] or additional predictors [17]. Wang et al. used vector autoregression, a multivariate time series model, to predict daily summertime sea ice at an intraseasonal timescale of 20-60 days. The study reached one month-ahead SIE prediction RMSE of approximately 0.45 million km² and one month-ahead SIC prediction RMSE of 10-15%, depending on the specific region [41]. Greater accuracy is desired for reliable long-term ice forecasts.

B. Deep Learning Methods

Deep learning models have demonstrated improved accuracies over statistical models for ice prediction. Chi et al. used two deep learning models, a multilayer perceptron and a long-short term memory model, to predict monthly 2015 Arctic SIC. Sea ice concentration for the preceding 12 months was the only input variable used for their models, but this study was able to achieve an RMSE of 8.89% at 25km x 25km resolution [7]. Kim et al. trained a convolutional neural network (CNN) as well as random forests to make one month-ahead monthly SIC predictions. Prior meteorological measurements and ice values were used as predictors. The CNN had the best performance, with an overall RMSE of 5.76% for predictions of Arctic sea ice from 2000-2017 [27]. Liu et al. compared the performance of CNN and ConvLSTM models in predicting Arctic SIC at the daily scale for 2018. The spatial domain was divided into 20 sub-grids, and the two previous days were used to predict the next day's ice concentrations. The CNN had an average RMSE of 8.058%, and the ConvLSTM had an improved 6.942% RMSE for 2018 sea ice [28]. Ali et al. compared different machine/deep learning based SIE forecasting techniques and proposed a multi-temporal ensemble model that achieved the best forecast accuracy with an RMSE of 4.11% [2]. Most recently, Andersson et al. proposed a U-Net based ensemble model for predicting sea-ice probabilities for a lead time of 6 months [4]. Their model takes in images as input and forecasts as output sea-ice concentration (SIC) maps in the form of three classes (open-water region SIC < 15%, ice-edge region 15% < SIC < 80%, and confident ice region SIC > 80%). These studies highlight the promise of deep learning models for producing accurate ice predictions at a high spatiotemporal resolution. Yet, to the best of our knowledge, there are still no studies that simultaneously forecast spatiotemporal Arctic sea ice in terms of both SIC and SIE.

III. DATASETS, PROBLEM DEFINITION AND DATA PROCESSING

A. Datasets

This study uses sea ice, atmospheric, and meteorological data from 1979 through 2020 covering the Arctic Ocean and adjacent land areas. Sea ice concentration data with 25km by 25km grid cells were accessed from the National Snow and Ice Data Center [6], [35]. The SIC dataset produces an uncertainty of about +5% in the Arctic winter when sea ice tends to reach its peak in concentration levels. During summer months, this uncertainty increases to about +15% as there are more melt ponds present which can skew data collection [6]. The uncertainty comes from either instruments or satellite retrieval algorithms. Here, for modeling purposes, the satellite-retrieved SIC data were considered to be the ground truth.

Atmospheric and meteorological variables were obtained from European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-5 global reanalysis product [14], [15]. Information on these atmospheric variables along with sea ice concentration are listed in Table I. The variables were chosen

because they demonstrated causal links with sea ice variations [22].

The inclusion of each atmospheric and meteorological variable was based on their physical impact on sea ice trends. Air temperature is the main driver of changes in sea ice, and record low sea ice extents during recent melting seasons have been associated with warmer atmospheric temperatures [36]. The inclusion of sea surface temperature and 2 meter air temperature in the dataset provide the models with information on oceanic and atmospheric heat. Studies have also shown that Arctic circulation and wind patterns have seasonally varying relationships with sea ice [10], [18]. For example, poleward winds specifically play a key role in transporting heat to the Arctic, which contributes to ice melt [3], [26], [40]. Precipitation trends are also connected to sea ice patterns. In recent years, earlier rainfalls during spring have triggered earlier snowmelt and, via feedback loops, earlier Arctic ice melt [11], [29]. The complexity of atmospheric, oceanic, and sea ice interactions is illustrated in [16], [20], which highlight the pathway by which regional differences in atmospheric pressure facilitate increased Arctic humidity, which in turn enables higher levels of longwave radiation to reach the sea surface, leading to earlier melting of sea ice. Thus, each predictor impacts Arctic sea ice through complex physical interactions in the ocean and atmosphere.

TABLE I: Input Features for CNN and ConvLSTM models

Feature	Source	Units	Range
Sea Ice Concentration	NSIDC	% per pixel	0-100
Surface Pressure	ERA5	Pa	40000-110000
10m Wind Speed	ERA5	m/s	0-40
Near-Surface Humidity	ERA5	kg/kg	0-0.1
2m Air Temperature	ERA5	K	200-350
Shortwave Radiation	ERA5	W/m ²	0-1500
Longwave Radiation	ERA5	W/m ²	0-300
Rain Rate	ERA5	mm/day	0-800
Snow Rate	ERA5	mm/day	0-200
Sea Surface Temperature	ERA5	K	200-350

All variables were averaged from a daily resolution to the monthly scale. Prior to model-specific pre-processing, the dataset had 504 images, each with 448 by 304 grid cells and 10 channels, corresponding to the 504 months in the dataset and the 10 input features detailed in Table I.

B. Problem Statement

Given n months of historical data X comprising of the above 10 atmospheric and ocean variable measurements in Arctic region for each pixel, learn a function to forecast pixel-wise sea-ice concentration Y_C , shown in Equation (1), and total sea-ice extent Y_E , shown in Equation (2), for the next month.

$$Y_C = f_C(X_{t-n}, X_{t-n+1}, \dots, X_t) \quad (1)$$

$$Y_E = f_E(X_{t-n}, X_{t-n+1}, \dots, X_t) \quad (2)$$

C. Data Preprocessing

1) *Convolutional Neural Network Data Preprocessing:* CNN models were trained on the first 407 months of the data (January 1979 - November 2012) and tested on the last 96 months (January 2013 - November 2020), with a one-month lead time. Each image in the dataset was considered to be an individual training example and was used to predict per-pixel sea ice concentrations for the next month. For example, the image corresponding to January 1979 was used to predict ice concentrations for February 1979. Thus, the training dataset learned per-pixel sea ice concentrations for February 1979 - December 2012, and the testing dataset predicted per-pixel sea ice concentrations for February 2013 - December 2020.

2) *Convolutional LSTM Data Pre-Processing:* In order to fully capture the spatio-temporal nature of our data using a convolutional LSTM model, heavy data preprocessing was necessary. The model was trained on the first 407 months of the data and tested on the last 96 months of the data. In Keras, ConvLSTM2D layers require 5 dimensional inputs of shape (*samples, timesteps, rows, columns, features*). To reshape the data, a stateless rolling window was applied to the training and testing data, creating 384 samples of 12 months each. Sample one contained months 1-12, sample two contained months 2-13, and the last sample contained months 372-384. The final shape of the training input data was 384 samples with 12 months of 448×304 pixel images, each containing 10 feature measurements at each pixel. Similarly, the final shape of the test input data was 84 samples with 12 months of 448×304 pixel images each, all containing 10 feature measurements at each pixel.

The data consisted of 384 images in the training set and 84 images in the test set. Each image contained the average sea ice concentration for the corresponding month in each pixel. The first sample of input data, consisting of the first 12 months of images, was used to predict the sea ice concentrations in the 13th month in the output data; the second sample was used to predict the SIC in the 14th month and so on. Including such a rolling window with 12-month timesteps allows the ConvLSTM to learn yearly variations and relationships in SIC.

D. Post-Processing

After model training, each SIC prediction image was post-processed to obtain a more realistic result. SIC strictly ranges from 0% to 100%, and can only have a non-zero value over ocean pixels. Thus, any non-zero SIC prediction over land pixels were set to 0, all SIC predictions over ocean pixels below 0% were set to 0%, and all SIC predictions over 100% were set to 100%.

IV. METHODS

A. Statistical Models

Vector Autoregression (VAR) is a family of multivariate time series models. Each variable is used as an autoregressive predictor for every other variable in the input data. A VAR model with lag 10, selected using the Bayesian Information Criterion (BIC), was used as a baseline for comparison with

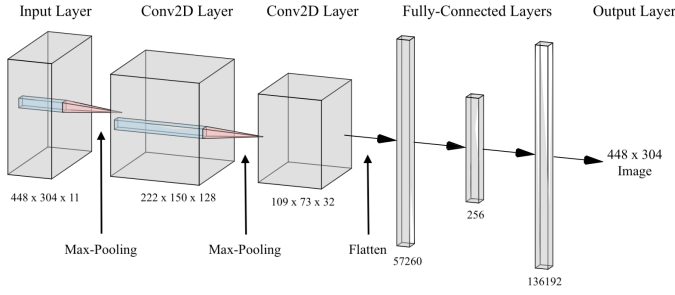


Fig. 1: CNN model architecture.

deep learning model results. As a temporal model, VAR was only trained to predict SIE with a one-month lead time, taking daily inputs from the 10 variables described in Table I.

B. Deep Learning Models

1) *Convolutional Neural Network*: The first deep learning technique this study implements is a Convolutional Neural Network (CNN). CNNs are a type of deep learning models particularly suited for images, speech, and audio signals. Thus, we chose to implement it to process our per-pixel data, which is in the form of images. The CNN input data for this study is a three-dimensional array with the dimensions $height \times width \times channel$ (448, 304, 10).

We created a CNN model in order to provide a baseline for comparison with other, more complex deep learning models. Figure 1 shows the CNN model architecture, which consists of three types of layers: convolutional layers, pooling layers, and fully connected layers [12]. The CNN we implemented features 9 layers. The first part of the model is an alternating sequence of 3 convolutional layers and 2 max-pooling layers. The convolutional layers contain 128, 32, and 8 filters respectively, all with kernels of size 5×5 . The max-pooling layers have kernels of size 2×2 .

Once the data was fed through the convolutional and max-pooling layers, it was then flattened and propagated through two fully-connected layers. The first fully-connected layer contained 256 neurons, while the next layer contained $448 \times 304 = 136192$ neurons, corresponding to the dimensions of the output images. The output vector is then reshaped into a matrix ($I \times J$) of 448 by 304 grid cells, representing a monthly SIC image prediction. The number corresponding to each pixel in the output image represents the sea ice concentration percentage for that pixel.

$$Y_E = \sum_I \sum_J \begin{cases} A[i, j], & \text{if } Y_C[i, j] \geq 15\% \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The monthly SIE value was derived from SIC using Equation (3). $A[i, j]$ is the area of the pixel at $[i, j]$. If the predicted SIC value of a pixel is greater than or equal to 15%, the pixel's value is set equal to 1; if the predicted SIC value is less than 15%, the value is set equal to 0. Each pixel is multiplied by its corresponding area, and the resultant matrix is summed to

produce the predicted SIE. This calculation is also used by the National Snow and Ice Data Center [33].

The model is optimized with the Adamax optimizer, and it utilizes a custom masked loss function, detailed in section IV-C1. All layers in the model use ReLU activation with the exception of the output layer, which uses linear activation. Early stopping with a patience of 10 was used to reduce training time.

2) *Long Short-Term Memory Network*: This study also implements a Long Short-Term Memory Network (LSTM) in order to obtain a baseline for comparison with convolutional and multi-task deep learning models. LSTM is a type of Recurrent Neural Network (RNN) that is used in analyzing time-series data with proficiency in forecasting longer lead times. An LSTM network comprises a special memory gate that gathers and processes input from previous time steps to influence the output value of successive time steps. Since our data possess spatial and temporal properties, the use of an LSTM was suitable to be considered for baseline experiments. The architecture of our baseline LSTM comprises two many-to-one LSTM layers, one dropout layer and three fully-connected layers. As a temporal model, we have implemented the LSTM to only forecast monthly SIE with a one-month lead time, taking daily inputs from the 10 variables described in Table I. The input dimensions for the model were $N \times F \times T$. Here, N is the number of samples, F is the number of features, that is 11, and T is the timestep, that was kept 1. The training and test split was same as other models to keep fair comparison. The model was trained on 500 epochs using early stopping method, the optimizer used was 'Adam', whereas the batch size was kept 12.

$$Y_{t+1} = f(X_{t-n}, X_{t-n+1}, \dots, X_t, Y_t) \quad (4)$$

Equation (4) represents the functionality of our baseline LSTM model. The outcome of the model is Y_{t+1} , that is the predicted sea-ice extent for next month $t + 1$, given the atmospheric data and sea-ice extent for previous month t .

3) *Convolutional Long Short Term Memory Network*: ConvLSTM architecture combines the spatial recognition capabilities of convolutional neural networks and the temporal modeling capabilities of long short-term memory models (LSTM) to produce an output which takes both spatial and temporal patterns into account. LSTMs use matrix multiplication on each gate in an LSTM cell; ConvLSTMs replace this matrix multiplication with convolutions, allowing the model to capture underlying spatial features in multi-dimensional data [13].

We created a ConvLSTM model to give us a baseline for comparison with our Multi-Task models. Figure 2 shows the ConvLSTM model architecture. Our ConvLSTM model had 9 layers. The first layer is a ConvLSTM layer with 16 filters of size 5×5 . Following this layer is a set of two alternating max-pooling and convolutional layers. Both max-pooling layers have a 2×2 kernel. The first convolutional layer features 128

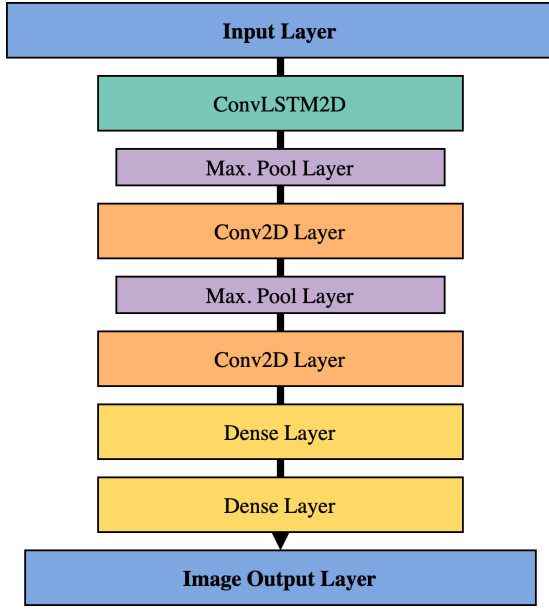


Fig. 2: ConvLSTM model architecture.

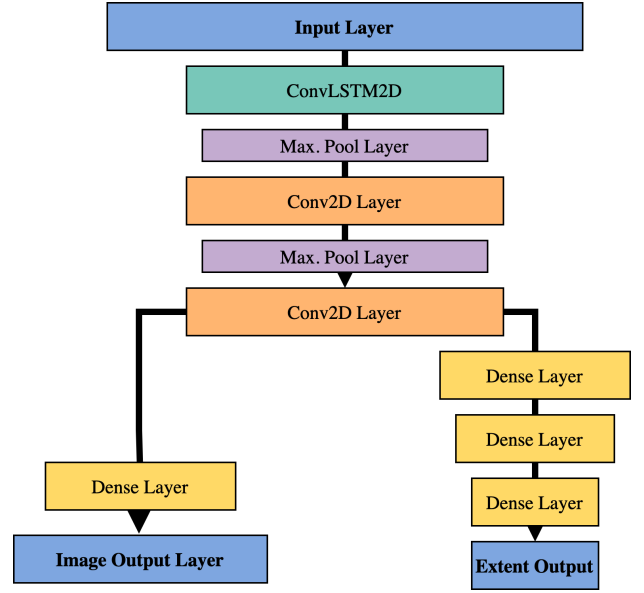


Fig. 3: Multi-Task ConvLSTM model architecture.

filters, while the second contains 32; both layers have a kernel size of 5×5 .

After propagation through the ConvLSTM and convolutional layers, the model's output is flattened and fed through three fully-connected layers. The first layer contains 256 neurons, the second contains 512 neurons, and the output layer contains $448 \times 304 = 136192$ neurons, equal to the number of pixels in each output image. The final output vector is then shaped into a 448 by 304 image, corresponding to a monthly SIC image output. The number corresponding to each pixel in the output image represents the sea ice concentration percentage for that pixel. As with the CNN models, the monthly SIE value was derived from SIC using Equation (3).

The model is optimized with the Adamax optimizer, and it utilizes a custom masked loss function, detailed in IV-C1. All layers in the model use ReLU activation with the exception of the output layer, which uses linear activation. As with the CNN, early stopping with a patience of 10 was used to reduce training time.

4) *Multi-Task Models:* Multi-task learning is a subset of machine learning where multiple tasks are learned by a shared model [8]. Using a branched architecture, we trained multi-task models to produce both monthly image predictions of SIC for each pixel and monthly sea ice extent values.

Multi-task learning was integrated into both the CNN and ConvLSTM models. Figure 3 shows the architecture for our Multi-Task ConvLSTM model, which is nearly identical to the Multi-Task CNN architecture. The models contain a shared root, through which the input data are passed, and two subsequent branches which produce the SIC image and sea ice extent outputs. The input root for the ConvLSTM comprises one ConvLSTM2D layer with 8 filters of size 5×5 and ReLU activation, followed by two alternating max pooling

and convolution layers. The only difference for the Multi-Task CNN is that the first layer is a convolutional layer; everything else that follows replicates the architecture of the Multi-Task ConvLSTM. The max pooling layers contain filters of size 4×4 , while the convolutional layers contain 128 and 32 size 5×5 filters respectively. The convolutional layers use ReLU activation. The data are then flattened and propagated through a dense layer with 256 nodes and ReLU activation.

The image branch of the architecture receives the model's root output and propagates it through a dense layer of size $448 \times 304 = 136,192$ with linear activation. The data are then reshaped into an image of size 448 rows by 304 columns. Each pixel in the image has an associated SIC value.

The extent branch also receives the model's root output; it propagates the root output vector through 4 dense layers of size 128, 32, 8, and 1 respectively, returning a single sea ice extent result for each input sample. The first 3 dense layers use a ReLU activation function, while the output layer utilizes linear activation for regression.

Two loss functions are used to optimize the models. The image branch is optimized using the custom masked loss function described in section IV-C1, while the sea ice extent branch is optimized using MSE loss. The performance of the models are evaluated using the RMSE metric for both branches.

C. Custom Loss Functions

1) *Masked Loss Function:* Neural networks use loss functions to measure the error in their predictions after each epoch. Once the error has been measured, the model optimizes the loss function using back-propagation.

To improve the accuracy of our deep learning models, a custom masked loss function was implemented in the architecture of each network, except the LSTM. A land mask was applied

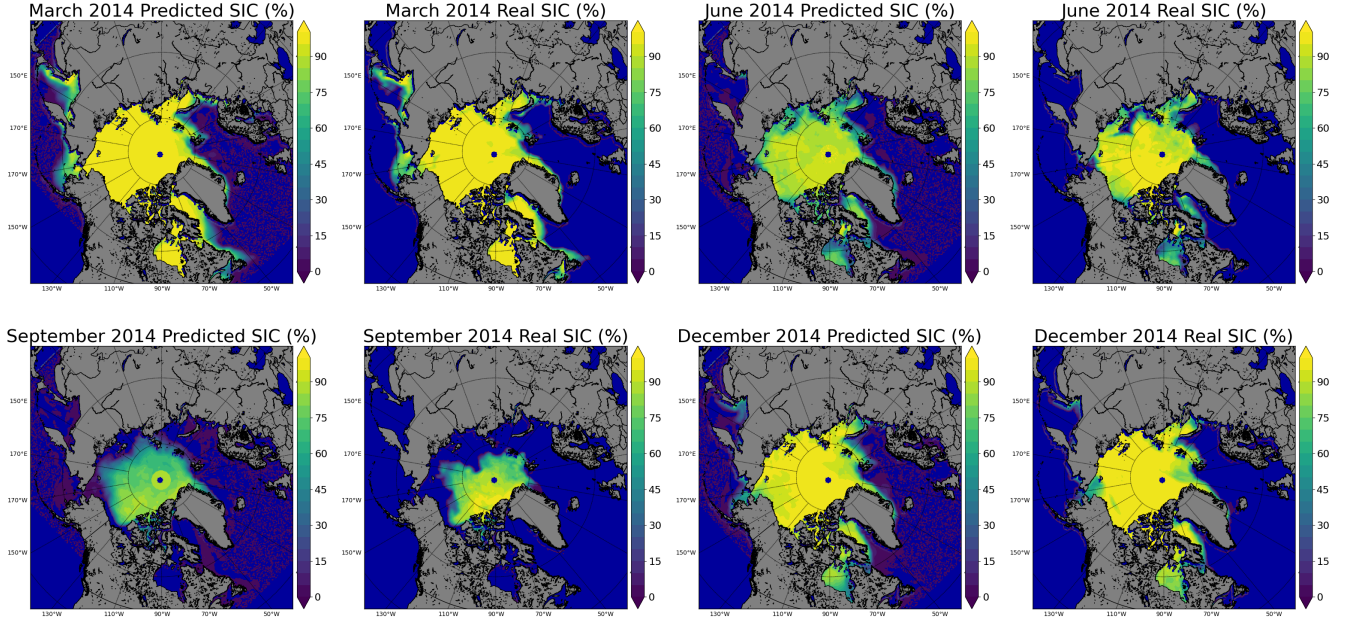


Fig. 4: Multi-Task ConvLSTM SIC predictions and true values for March, June, September, and December 2014.

to each output of the network before loss was evaluated. This mask included one image for each month in the dataset; land pixels were given a value of 0 and non-land pixels were given a value of 1. Each predicted output and its corresponding mask were multiplied elementwise. Thus, any predicted values over land became 0, and any predictions over ocean pixels were unchanged, resulting in land pixels being ignored when calculating the loss.

In order to correctly apply each month's land mask to the corresponding sample, the model accepted the land mask values and actual values as inputs. However, they were not propagated through the model; instead, they were exclusively used for the calculation purposes of the loss function.

2) *Extent Loss Function*: A separate convolutional neural network model was trained with a loss function which adds a penalty for SIE loss in addition to the land masking discussed in section IV-C1. For each SIC image prediction, the SIE value was derived using Equation (3), and the SIE RMSE was calculated using Equation (6). This SIE RMSE was added to the SIC RMSE for the loss function output value. This addition to the custom loss function enabled the CNN model to better optimize both SIC and derived SIE predictions.

V. RESULTS AND DISCUSSION

A. Models

Based on the methods described in section IV, we conducted experiments to compare results of seven total models: one statistical model, namely VAR in IV-A, and six deep learning models. The six deep learning models evaluated were a CNN with a masked loss function, an LSTM without a masked loss function, a ConvLSTM with a masked loss function, a CNN

with an extent loss function described in IV-C2, a Multi-Task CNN with a masked loss function, and a Multi-Task ConvLSTM with a masked loss function. Since the LSTM model outputs single values rather than images, the use of a masked loss function is not applicable. We also tested a CNN without a custom loss function and a Multi-Task CNN with an extent loss function, but these results are not included in this paper due to their poor performance. The CNN-based models were trained over 400 epochs, and early stopping was utilized to reduce training time. The base CNN had a batch size of 4, while the Multi-Task CNN had a batch size of 32. The ConvLSTM-based models were trained over 1000 epochs; early stopping was also used in these models. The ConvLSTM models had a batch size of 4. Our implementation code can be accessed at [the Big Data REU GitHub repository \[1\]](#).

B. Evaluation Metrics

Root-mean-squared error (RMSE) was the main metric for evaluating model predictions. Equation (5) shows the RMSE calculation for evaluating SIC results. The squared error was calculated for each predicted pixel, where $I = 448$ and $J = 304$. Here, Y_C represents real SIC and $\hat{Y}_C$ represents predicted SIC values. Equation (6) shows the RMSE calculation for evaluating SIE results. The squared error was calculated for each month, where $M = 84$. Y_E represents real or derived SIE and $\hat{Y}_E$ represents predicted SIE values.

$$RMSE_{SIC} = \sqrt{\frac{\sum_I \sum_J (Y_C[i, j] - \hat{Y}_C[i, j])^2}{N}} \quad (5)$$

TABLE II: SIC and SIE RMSE percentage values of our models. Values without * represent SIE predicted by the model; values with * represent SIE derived based on Equation (3).

Method	SIC Train RMSE	SIC Test RMSE	SIC Post-Proc RMSE	SIE Train RMSE	SIE Test RMSE
VAR	-	-	-	-	0.424
LSTM	-	-	-	0.179	0.347
CNN	11.734	12.005	7.106	-	0.631*
Extent Loss CNN	11.911	12.228	7.150	-	0.571*
ConvLSTM	10.054	11.478	8.161	0.908*	0.938*
Multi-Task CNN	13.108	13.348	7.527	0.375	0.536
Multi-Task ConvLSTM	9.989	10.807	7.155	0.335	0.303

$$RMSE_{SIE} = \sqrt{\frac{\sum_M (Y_E[m] - \hat{Y}_E[m])^2}{N}} \quad (6)$$

Normalized RMSE (NRMSE) was also utilized to improve the comparability of predictions from different models. NRMSE calculation involves dividing the RMSE value by the mean of the predicted values.

C. Model Comparison on Sea Ice Concentration Forecasting

SIC prediction results of our models are shown in columns 2-4 of Table II. Note that all RMSE values stated in the following section refer to test RMSE after post-processing, unless otherwise specified. The base CNN model resulted in the lowest SIC RMSE, 7.106%, followed by the CNN with extent loss function, with SIC RMSE of 7.150%. The Multi-Task ConvLSTM resulted in slightly less accurate predictions than the CNN models, with an SIC RMSE of 7.155%. The Multi-Task CNN had worse accuracy than the single-task CNN models, while the Multi-Task ConvLSTM improved on the single-task ConvLSTM's accuracy.

Prior to post-processing, the Multi-Task ConvLSTM had a significantly lower SIC test RMSE compared to all other models. Additionally, each ConvLSTM model outperformed all CNN models on SIC test RMSE prior to post-processing. This indicates that the ConvLSTM models better avoided predicting SIC values over land or values beyond the [0%, 100%] SIC range, possibly due to the additional temporal information.

Figure 4 shows Multi-Task ConvLSTM prediction and real SIC values for March, June, September, and December 2014. The model was able to capture the overall spatial distribution of sea ice, as seen by the consistent concentration distribution between predicted and real maps. Upon visual inspection, the model performs particularly well for March and December, winter months for which SIC is relatively high. June and September display worse model accuracy. Specifically, the model underestimates June SIC towards the Canadian Archipelago. As for September, the model underestimates the SIC in the north of Greenland and Canadian Archipelago, and overestimates SIC in the Pacific side of Arctic Ocean, particularly in the Beaufort Sea, Chukchi Sea and East Siberian Sea. In addition, the sea ice-covered area in the predicted September map is much larger than that of the real September sea ice distribution.

The visual results gathered from the spatial plots are confirmed by the time-based RMSE plots in Figures 5 and 6. SIC

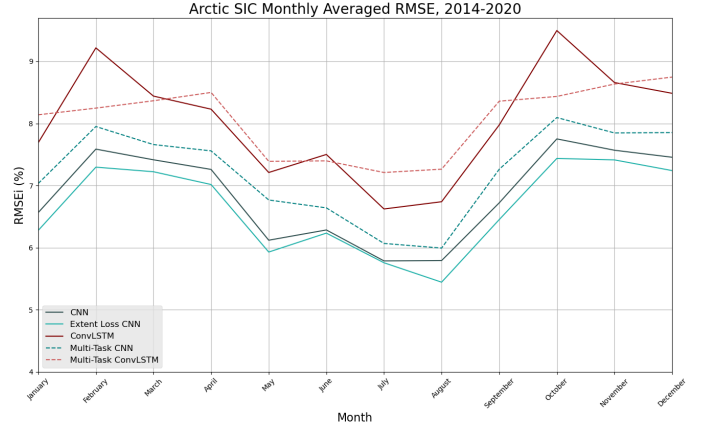


Fig. 5: SIC monthly averaged RMSE values for each model.

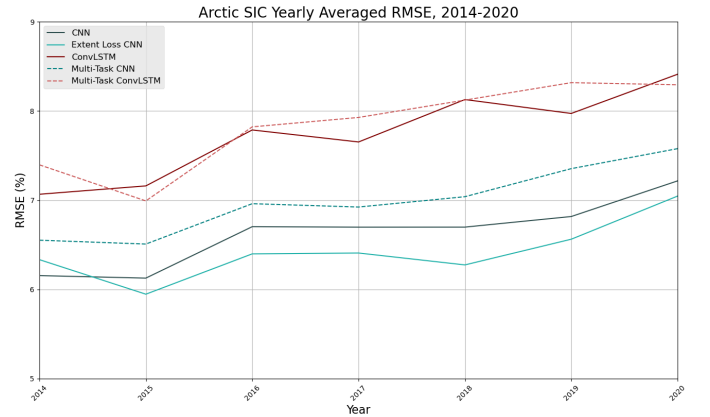


Fig. 6: SIC yearly averaged RMSE values for each model.

test RMSE for each model was averaged over each month for Figure 5. Overall, deep learning models predicted SIC with higher accuracy during months with stable changes in ice, and displayed poor accuracy during periods with high ice variability. RMSE is high for February and March as well as October through December. These periods mark the beginning of the melting and freezing seasons, respectively [30]. The models were not able to fully capture the rapidly changing atmospheric and oceanic conditions during these periods [20]. SIC test RMSE for each model was averaged over each year for Figure 6. This figure clearly shows a positive relationship between RMSE and elapsed time since the end of the model

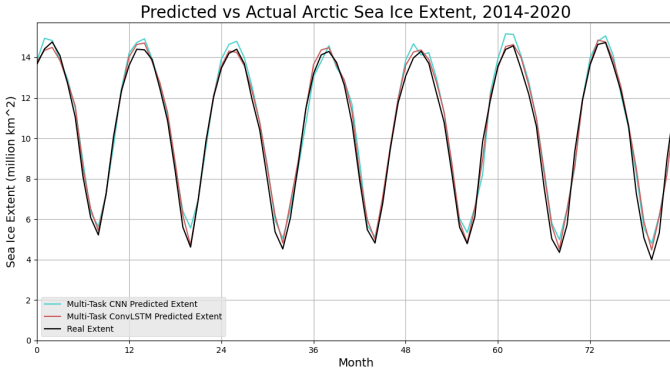


Fig. 7: Multi-task CNN derived SIE predictions and Multi-task ConvLSTM SIE predictions vs. actual SIE values in million km^2 .

training dataset. The increasing disparity between the testing and training periods due to changing climatic conditions may contribute to the trend seen in Figure 6. In order to obtain accurate predictions, models may need to be re-trained on more recent data.

Figures 5 and 6 confirm the relatively poor performance of the Multi-Task CNN model for SIC prediction. The Multi-Task CNN results in higher RMSE values than the base and extent CNNs across all months and years in the testing period. This confirms the high SIC RMSE value in Table II. The Multi-Task ConvLSTM and base ConvLSTM models have comparable performance across the time for SIC prediction. However, a clear distinction can be seen for February and October in Figure 5, for which the Multi-Task ConvLSTM model maintains a stable RMSE of around 8%, while the base ConvLSTM SIC RMSE spikes to greater than 9%.

D. Model Comparison on Sea Ice Extent Forecasting

The SIE forecasting performance of all seven models are shown in the last two columns of Table II. It indicates that our Multi-Task ConvLSTM model has the best performance with test data. Note that all RMSE values stated in this section refer to test RMSE, unless otherwise specified. The lowest SIE RMSE value was attained by the Multi-Task ConvLSTM model, at 0.303 million km^2 . Two time series models, VAR and LSTM, provided the next-best performance. The LSTM, without any convolutional layers and thus without spatial information, reached an SIE RMSE of 0.347 million km^2 . Vector Autoregression, the baseline model for SIE, reached an RMSE of 0.424 million km^2 .

Each of the three aforementioned time series models had a lower SIE RMSE than any CNN model. The best-performing CNN model was the Multi-Task CNN, with an SIE RMSE of 0.536 million km^2 , 0.233 million km^2 greater than the Multi-Task ConvLSTM SIE RMSE. The extent loss CNN had a marginally higher SIE RMSE of 0.571 million km^2 . The base CNN, which was not provided any information about ice extent in the input data or loss function, had an SIE RMSE of

0.868 million km^2 . The SIE predicted values for the base CNN and extent loss CNN were calculated from the SIC predicted values using Equation (3).

Figure 7 shows the Multi-Task CNN predicted SIE values in blue and the real SIE values in black. The general time-series trend of SIE is captured well. For the majority of months, especially during periods of steady melting or freezing, the model performance is exceptional. However, the Multi-Task CNN overestimates SIE maximums and underestimates SIE minimums. For September 2020, the Multi-Task CNN model overestimates SIE by nearly 1 million km^2 . The significant improvement in performance by the Multi-Task ConvLSTM is clearly seen in Figure 7, which also shows Multi-Task ConvLSTM predicted SIE in red and real SIE values in black. This model again accurately captures the general 3time-series trend of SIE, but the major improvement comes in predicting the SIE maxima and minima. The Multi-Task ConvLSTM is able to consistently predict SIE maxima with high accuracy, and SIE minima with only slight overestimates. Predicting sea ice during the March maximum and September minimum has traditionally been a highly difficult task [24], highlighting the particular benefits of the Multi-Task ConvLSTM.

These results indicate that 1) multi-task learning is greatly beneficial for SIE prediction, 2) time-series models perform significantly better for SIE prediction and 3) models cannot accurately predict SIE solely based on SIC.

Based on conclusion 2), it is important to note the significant underperformance in SIE prediction from one time-series model, the ConvLSTM. This underperformance is due to the fact that the ConvLSTM was predicting derived SIE values, while all other time-series models made predictions based on actual SIE values.

E. Discussion and Related Work Comparison

As shown in Table II, multi-task models offered the clearest benefits for predicting SIE. During model training, learning to optimize for SIC may have provided valuable information for minimizing SIE error. On the contrary, the current results do not indicate a benefit for SIC prediction from having models learn SIE in conjunction. As mentioned in IV-C1, uncertainty remains in the satellite-retrieved SIC dataset, and it should be taken into account along with the prediction errors. However, addressing uncertainties related to observational datasets was beyond the scope of this study.

To further compare our work with related studies, Table III presents the SIC RMSE values for related studies as well as our best-performing models. From the table, we can see different methods vary in terms of whether additional physical variables except sea ice variable are used, temporal resolution, forecast lead time and train/test data. Each model from this study in Table III outperforms each model from related studies, with the exception of 5.76% by [27]. Work [27] uses all past data starting 1988 to forecast next month's SIC, so its train datasets range from 12 to 29 years. Because it does not have static train/test data split like other studies, its result is not directly comparable. So our best model (Base CNN)

TABLE III: Comparison with related studies on SIC prediction performance.

Paper	Model	Data	Physical Variables	Temporal Resolution	Lead Time	Train Data	Test Data	RMSE	NRMSE
Liu et al. [28]	ConvLSTM	25x25 km	✓	Daily	1 day	9 years	1 year	11.2%	NA
Liu et al. [28]	CNN	25x25 km	✓	Daily	1 day	9 years	1 year	13.7%	NA
Kim et al. [27]	CNN	25x25 km	✓	Monthly	1 month	12-29 years (dynamic)	18 years	5.76%	0.1615
Chi et al. [7]	LSTM	25x25 km	X	Monthly	1 month	36 years	1 year	8.89%	NA
This work	ConvLSTM	25x25 km	✓	Monthly	1 month	33 years	7 years	8.161%	0.396
This work	Base CNN	25x25 km	✓	Monthly	1 month	33 years	7 years	7.106%	0.345
This work	Multi-Task CNN	25x25 km	✓	Monthly	1 month	33 years	7 years	7.527%	0.365
This work	Multi-Task ConvLSTM	25x25 km	✓	Monthly	1 month	33 years	7 years	7.155%	0.347

TABLE IV: Comparison with related studies on SIE prediction performance.

Paper	Model	Data	Physical Variables	Train Data	Test Data	RMSE	NRMSE
Ali et al. [2]	Attention-based Ensemble LSTM	Spatially averaged daily and monthly inputs	✓	34 years	5 years	0.586 million km ²	0.0567
This work	LSTM	Spatially averaged daily inputs	✓	33 years	7 years	0.347 million km ²	0.0305
This work	Base CNN	25x25 km monthly averaged	✓	33 years	7 years	0.631 million km ²	0.0607
This work	Multi-Task CNN	25x25 km monthly averaged	✓	33 years	7 years	0.536 million km ²	0.0516
This work	Multi-Task ConvLSTM	25x25 km monthly averaged	✓	33 years	7 years	0.303 million km²	0.0292

is 1.78% better than the best SIC prediction result, namely [7], from related work. Table III also shows our model results represent a major improvement over prior results using similar ConvLSTM and CNN based models respectively.

Similarly, Table IV displays the SIE RMSE values for our best-performing models and related work. The Multi-Task ConvLSTM, and Multi-Task CNN outperform the work by Ali et al. SIE RMSE of 0.586 million km². Our best model, namely Multi-Task ConvLSTM, has less RMSE by 0.283 million km². This confirms the value of multi-task learning, which by simultaneously optimizing for both SIC and SIE helps the model improve SIE prediction performance.

We also note that Tables III and IV show that none of the related studies can predict SIC and SIE simultaneously. While our two multi-task deep learning models can achieve both predictions with good performance.

VI. CONCLUSIONS

In this paper, we introduced two types of Multi-Task deep learning models, Multi-Task CNN and Multi-Task ConvLSTM, to simultaneously forecast spatial pixel-wise sea ice concentration (SIC) and temporal total sea ice extent (SIE) values. We considered sea ice, atmospheric, and meteorological data in image time-series form, allowing our models to learn fluctuations in data's spatial and temporal dimensions. Our models included a branched multi-task architecture and custom loss function. The models' multi-task structures allowed them to effectively learn to predict SIC images and SIE values, making the optimization process more efficient with the use of multiple branches and outputs. The use of a custom loss function forced the model to ignore land pixels when making

image predictions; therefore, the models learned to only focus on predicting sea-ice values. To assess the performance of our models, we conducted experiments on 34 years of training data and 7 years of testing data, comparing the results of the models against each other and sea ice prediction models from other literature. The results of our experiments indicated that multi-task modeling performs comparably to or better than state-of-the-art deep learning models in both SIC and SIE prediction while learning both tasks simultaneously. We found that, out of all models we developed, the Multi-Task ConvLSTM achieved the best training and testing accuracies for SIC and SIE before post-processing.

We anticipate that this work and results will be useful for developing forecasts of ice extent alongside concentration forecasts, both of which are important characteristics of Arctic ice. Additional analysis may elucidate spatiotemporal relationships between SIC and SIE. We also believe that this work highlights the merits of multi-task deep learning, and provides a valuable template for further applications.

For future research, we will mainly focus on following facets: 1) eliminate the need for model post-processing by using a scaled activation function, 2) evaluate the performance of multi-task models on lead times greater than one month, 3) perform extensive hyperparameter tests to reduce overfitting and achieve optimal model performance, 4) incorporate probabilistic modeling to capture data and model uncertainties.

ACKNOWLEDGMENTS

This work is supported by the grant "REU Site: Online Interdisciplinary Big Data Analytics in Science and Engineering" from the National Science Foundation (grant no. OAC-2050943). Co-author Ali and Wang additionally acknowledge

support by the grant “CAREER: Big Data Climate Causality Analytics” from the National Science Foundation (grant no. OAC–1942714). The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS–0821258, CNS–1228778, and OAC–1726023) and the SCREMS program (grant no. DMS–0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC).

REFERENCES

- [1] GitHub Repository for Multi-Task Deep Learning Based Spatiotemporal Arctic Sea Ice Forecasting. <https://github.com/big-data-lab-umbc/big-data-reu/tree/main/2021-projects/team-1>, 2021. Accessed: 2021-11-5.
- [2] S. Ali, Y. Huang, X. Huang, and J. Wang. Sea Ice Forecasting using Attention-based Ensemble LSTM. *arXiv e-prints*, page arXiv:2108.00853, July 2021.
- [3] R. Alkama, E. Koffi, S. Vavrus, T. Diehl, J. Francis, J. Stroeve, G. Forzieri, T. Vihma, and A. Cescatti. Wind amplifies the polar sea ice retreat. *Environmental Research Letters*, 15:1–16, 11 2020.
- [4] T. R. Andersson, J. S. Hosking, M. Pérez-Ortiz, B. Paige, A. Elliott, C. Russell, S. Law, D. C. Jones, J. Wilkinson, T. Phillips, et al. Seasonal arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, 12(1):1–12, 2021.
- [5] J. Boé, A. Hall, and X. Qu. September sea-ice cover in the arctic ocean projected to vanish by 2100. *Nature Geoscience*, 2(5):341–343, 2009.
- [6] D. Cavalieri, C. Parkinson, P. Gloersen, and H. Zwally. Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. Technical report, NASA DAAC at the National Snow and Ice Data Center, 1996.
- [7] J. Chi and H. Kim. Prediction of arctic sea ice concentration using a fully data driven deep neural network. *Remote Sensing*, December 2017.
- [8] M. Crawshaw. Multi-task learning with deep neural networks: A survey. *CoRR*, abs/2009.09796, 2020.
- [9] A. Davydov and G. Mikhaylova. Climate change and consequences in the arctic: Perception of climate change by the nenets people of vaigach island. *Global health action*, 4, 11 2011.
- [10] Q. Ding, A. Schweiger, M. L’Heureux, D. S. Battisti, S. Po-Chedley, N. C. Johnson, E. Blanchard-Wrigglesworth, K. Harnos, Q. Zhang, R. Eastman, et al. Influence of high-latitude atmospheric circulation changes on summertime arctic sea ice. *Nature Climate Change*, 7(4):289–295, 2017.
- [11] T. Dou, C. Xiao, J. Liu, W. Han, Z. Du, A. Mahoney, J. Jones, and H. Eicken. A key factor initiating surface ablation of arctic sea ice: Earlier and increasing liquid precipitation. *The Cryosphere*, 13:1233–1246, 04 2019.
- [12] I. C. Education. What are convolutional neural networks? <https://www.ibm.com/cloud/learn/convolutional-neural-networks>, October 2020.
- [13] J.-M. Y. et al. Spatial mapping of short-term solar radiation prediction incorporating geostationary satellite images coupled with deep convolutional lstm networks for south korea. *Environmental Research Letters*, 15(9), 2020.
- [14] European Centre for Medium-Range Weather Forecasts. ERA-5 global reanalysis product. <https://cds.climate.copernicus.eu/cdsapp#!/home>, 2021. Accessed: 2021-9-5.
- [15] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.
- [16] S. Horvath, J. Stroeve, B. Rajagopalan, and A. Jahn. Arctic sea ice melt onset favored by an atmospheric pressure pattern reminiscent of the north american-eurasian arctic pattern. *Climate Dynamics*, 04 2021.
- [17] S. Horvath, J. Stroeve, B. Rajagopalan, and W. Kleiber. A bayesian logistic regression for probabilistic forecasts of the minimum september arctic sea ice cover. *Earth and Space Science*, 7:e2020EA001176, 2020.
- [18] Y. Huang, Q. Ding, X. Dong, B. Xi, and I. Baxter. Summertime low clouds mediate the impact of the large-scale circulation on arctic sea ice. *Communications Earth & Environment*, 2(1):1–10, 2021.
- [19] Y. Huang, X. Dong, D. A. Bailey, M. M. Holland, B. Xi, A. K. DuVivier, J. E. Kay, L. L. Landrum, and Y. Deng. Thicker clouds and accelerated arctic sea ice decline: The atmosphere-sea ice interactions in spring. *Geophysical Research Letters*, 46(12):6980–6989, 2019.
- [20] Y. Huang, X. Dong, B. Xi, and Y. Deng. A survey of the atmospheric physical processes key to the onset of arctic sea ice melt in spring. *Climate Dynamics*, 52(7):4907–4922, 2019.
- [21] Y. Huang, X. Dong, B. Xi, E. K. Dolinar, and R. E. Stanfield. The footprints of 16 year trends of arctic springtime cloud and radiation properties on september sea ice retreat. *Journal of Geophysical Research: Atmospheres*, 122(4):2179–2193, 2017.
- [22] Y. Huang, M. Kleindessner, A. Munishkin, D. Varshney, P. Guo, and J. Wang. Benchmarking of data-driven causality discovery approaches in the interactions of arctic sea ice and atmosphere. *Frontiers in Big Data*, 4:72, 2021.
- [23] M. Ionita, K. Grosfeld, P. Scholz, R. Treffeisen, and G. Lohmann. September arctic sea ice minimum prediction - a skillful new statistical approach. *European Geosciences Union*, March 2019.
- [24] M. Ionita, K. Grosfeld, P. Scholz, R. Treffeisen, and G. Lohmann. September arctic sea ice minimum prediction- a skillful new statistical approach. *Earth System Dynamics*, 10:189–203, 03 2019.
- [25] A. Jahn, J. E. Kay, M. M. Holland, and D. M. Hall. How predictable is the timing of a summer ice-free arctic? *Geophysical Research Letters*, 43(17):9113–9120, 2016.
- [26] L. Jakobson, T. Vihma, and E. Jakobson. Relationships between sea ice concentration and wind speed over the arctic ocean during 1979–2015. *Journal of Climate*, 32, 08 2019.
- [27] Y. Kim, H. Kim, D. Han, S. Lee, and J. Im. Prediction of monthly arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. *European Geosciences Union*, March 2020.
- [28] Q. Liu, R. Zhang, Y. Wang, H. Yan, and M. Hong. Daily prediction of the arctic sea ice concentration using reanalysis data based on a convolutional lstm network. *Marine Science and Engineering*, March 2021.
- [29] A. Marcovecchio, A. Behrangi, X. Dong, B. Xi, and Y. Huang. Precipitation influence on and response to early and late arctic sea ice melt onset during melt season. *International Journal of Climatology*, n/a(n/a), 2021.
- [30] T. Markus, J. C. Stroeve, and J. Miller. Recent changes in arctic sea ice melt onset, freezeup, and melt season length. *Journal of Geophysical Research: Oceans*, 114(C12), 2009.
- [31] F. Massonnet, T. Fichefet, H. Goosse, C. M. Bitz, G. Philippon-Berthier, M. M. Holland, and P.-Y. Barriat. Constraining projections of summer arctic sea ice. *The Cryosphere*, 6(6):1383–1394, 2012.
- [32] W. Meier, G. Hovelsrud, B. van Oort, J. Key, K. Kovacs, C. Michel, C. Haas, M. Granskog, S. Gerland, D. Perovich, A. Makshtas, and J. Reist. Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. *Reviews of Geophysics*, 52:185, 09 2014.
- [33] National Snow and Ice Data Center. Data: Terminology. <https://nsidc.org/cryosphere/seaice/data/terminology.html>, Apr 2020.
- [34] National Snow and Ice Data Center. Environment : Climate. https://nsidc.org/cryosphere/seaice/environment/global_climate.html, Apr 2020.
- [35] National Snow and Ice Data Center. Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. <http://nsidc.org/data/NSIDC-0051>, 2021. Accessed: 2021-9-5.
- [36] D. Olonscheck, T. Mauritsen, and D. Notz. Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nature Geoscience*, 12, 05 2019.
- [37] C. L. Parkinson and J. C. Comiso. On the 2012 record low arctic sea ice cover: Combined impact of preconditioning and an august storm. *Geophysical Research Letters*, 40(7):1356–1361, 2013.
- [38] A. Riihela, T. Manninen, and V. Laine. Observed changes in the albedo of the arctic sea-ice zone for the period 1982–2009. *Nature Climate Change*, 3(10):895–898, 2013.
- [39] J. Stroeve, E. Blanchard-Wrigglesworth, V. Guemas, S. Howell, F. Massonnet, and S. Tietsche. Improving predictions of arctic sea ice extent. *Eos*, 96, 06 2015.
- [40] J. Wang, J. Zhang, E. Watanabe, M. Ikeda, K. Mizobata, J. E. Walsh, X. Bai, and B. Wu. Is the dipole anomaly a major driver to record lows in arctic summer sea ice extent? *Geophysical Research Letters*, 36(5), 2009.

- [41] L. Wang, X. Yuan, M. Ting, and C. Li. Predicting summer arctic sea ice concentration intraseasonal variability using a vector autoregressive model. *American Meteorological Society*, pages 1529–1543, February 2017.
- [42] W. Wang, M. Chen, and A. Kumar. Seasonal prediction of arctic sea ice extent from a coupled dynamical forecast system. *Monthly Weather Review*, 141:1375–1394, 04 2013.