IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 September 2024, accepted 19 October 2024, date of publication 28 October 2024, date of current version 5 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3486914

RESEARCH ARTICLE

Multi-Dilated Convolutional LSTM With U-Net for Global Sea Surface Temperature Forecasting

MANVENDRA JANMAIJAYA[1], (Member, IEEE), AMIT RAUNIYAR[2], (Member, IEEE), AMIT K. SHUKLA[3], (Senior Member, IEEE), SANDEEP KUMAR[1], REM COLLIER[2], (Member, IEEE), AND PRANAB K. MUHURI[1], (Senior Member, IEEE)

[1]Department of Computer Science, South Asian University, New Delhi 110068, India
[2]School of Computer Science, University College Dublin, Dublin 4, D04 V1W8 Ireland
[3]School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland

Corresponding author: Amit Rauniyar (amit.rauniyar@ucd.ie)

ABSTRACT Accurate forecasting of sea surface temperature (SST) is pivotal for a wide range of applications ranging from climate modelling to marine ecosystem management. This study introduces an ingenious multi-dilated model that employs dilated Convolutional Long Short-Term Memory (ConvLSTM) networks alongside a U-Net architecture which aims to enhance the precision of monthly SST predictions at lead periods of 3, 6 and 12 months. Our approach innovatively combines dilated convolutions within ConvLSTM with the segmentation capabilities of U-Net to adeptly capture the complex Spatio-temporal dynamics as well as relevant attributes of SST. The study utilizes a high-resolution MPI-ESM1-2-HR SST dataset for a series of rigorous tests, achieving a Mean Square Error of 0.01, indicating a high accuracy level in SST predictions. The model at a lead time of 12 months also showed a Sea Surface Microwave Index of 0.036, illustrating its effectiveness in reflecting SST's microwave emissivity characteristics, and an Earth Mover's Distance score of 0.97, highlighting its ability to closely match the predicted SST distribution with the actual one. Furthermore, a cosine similarity score of 0.99 suggests a significant alignment between the predicted and actual SST patterns. By merging multi-dilated convolutions with segmentation, our model tackles the intricate challenge of simultaneously capturing the spatial and temporal dependencies of SST, setting a new standard for forecasting accuracy in the field.

INDEX TERMS Multi-dilation, segmentation, U-Net, spatio-temporal, sea surface temperature, CNN.

I. INTRODUCTION

Sea Surface Temperature (SST) is an essential climate variable critical to our understanding of the climate system [1]. SST, at the interface of atmosphere and ocean, is instrumental in comprehending, quantifying as well as predicting the complex interactions, i.e., the flux of heat, gases, moisture, and momentum between oceanic and atmospheric realms [2], [3]. It has been estimated that oceans absorb more than 80 percent of the heat added to the global climate system, besides assimilating significant anthropogenic CO_2 emissions [4]. Unsurprisingly, SST variability is much inflated due to the unprecedented rate of anthropogenic climate change, besides consisting of a significant natural variability component [5]. This SST variability, both spatial and temporal, influences atmospheric circulation, thereby, affecting precipitation and temperature patterns, locally as well as remotely [6], occurrence of extreme weather events [5], [7] and marine ecology [8]. The accurate prediction of SST variability at different timescales is of critical importance for attributing past climate as well as improving future climate change monitoring, assessment, and prediction [9], [10], [11] as well as preserving marine ecosystems. Therefore, SST has immense socio-economic-ecological ramifications inter-regional influences, and therefore an efficient model for the global forecast of SST which accounts for its linear and

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

non-linear dependencies for modelling the wholesome distribution of SST is of critical importance.

The SST prediction efforts are broadly categorized into physics-based state-of-the-art numerical models, data-driven traditional statistical and machine learning (ML) methods and a combination of these two approaches [12]. The availability of a rich archive (almost 2 centuries) of SST data consisting of in-situ measurements (ships, buoys, etc.), satellite observations as well as blended data products [13] has enabled the use of ML algorithms for prediction of SST. In comparison to physics-based numerical models, data driven forecasting method are computationally efficient and can automatically learn the spatial relationships in data and require no or limited oceanic and atmospheric domain knowledge [58], [59]. Various ML algorithms such as Artificial Neural Network (ANN) [14], [15], [16], Support Vector Machine (SVM) [17], Multi-Layer Perceptron (MLP) Neural Network (NN) [18], Wavelet NN [19] have been applied to the problem of SST prediction. However, it was the use of Recurrent Neural Network (RNN) algorithms, particularly Long Short-Term Memory (LSTM) [20], that pioneered the extensive use of ML and Deep Learning (DL) methods for SST prediction [21], [22], [23], [24]. Zhang et al. [25] employed LSTM, SVM and MLP for SST prediction over China seas on a time scale ranging from days (1-3) to weeks and months. Gated Recurrent Unit (GRU), an efficient LSTM variant, has been used to predict SST in Bohai Sea with high accuracy [25]. Xie et al. [26] employed a GRU encoder-decoder (GED) for the Bohai Sea and South China Sea SST prediction. Xiao et al. [27] combined LSTM with AdaBoost to predict SST over short and medium timescales. Jahanbakht et al. [28] used an ensemble of two stacked Deep NNs for SST prediction. Most SST prediction studies primarily focus on temporal data and fail to incorporate spatial information [29], [30], [31], [32]. However, SST prediction is broadly a spatiotemporal sequence prediction task. A Deep GRU and Convolutional Neural Network (CNN) were integrated into a novel architecture DGCnetwork for SST prediction, using both spatial and temporal data [31]. Yang et al. [33] integrates a CNN with a fully connected LSTM network to put forth a CFCC-LSTM model to learn temporal and spatial information from SST. Hou et al. [34] created an encoder-decoder model (MIMO) for multi-scale SST prediction. The model extracts Spatio-temporal features from SST data at multiple scales and fuses features via cross-scale fusion (CSF). Zhang et al. [25] introduced a multilayer Convolutional LSTM (M-ConvLSTM) model designed for predicting 3D ocean temperature that can gather ocean temperature from horizontal and vertical directions for different depth layers.

SST impact different parts of the Earth's system in varying degrees. Therefore, previous research on SST prediction using spatial information focused on specific regions such as China Seas [22], [24], [27], coastal areas of Korea [21], [23], Red Sea [15], [35], Indian Ocean [15], [32], Black Sea [36], Atlantic Ocean [37], Pacific Ocean [25], Mediterranean Sea [38], etc. Comparatively, there are fewer studies predicting global SST. Nevertheless, it is of critical importance to study global SST variability in comprehensive climate change assessment [30]. In this context, this paper attempts to harness the capabilities of DL models to effectively learn spatial SST variability by analyzing complex patterns in data across different geographical locations by introducing a novel hybrid approach based on multi-dilation based ConvLSTM networks. It is further combined with the U-Net architecture, trained on high-resolution data to effectively integrate multi-scale spatial features, enabling precise localization and understanding of thermal variations across the ocean's surface.

In this context, this paper introduces a novel hybrid approach based on multi-dilation ConvLSTM networks, combined with the U-Net architecture, trained on high-resolution data to enhance the global SST forecasting capabilities. ConvLSTM has shown remarkable capabilities for time-series sequential data, while dilation operations, characterized by their unique ability to expand the receptive field without losing resolution or increasing the number of parameters, are particularly beneficial for geospatial datasets like SST. They enable the model to capture broader contextual information from the input data, enabling the identification of large-scale oceanic patterns that are crucial for accurate SST prediction. On the other hand, the U-Net architecture, originally designed for biomedical image segmentation, is proficient at handling spatial data with its deep, symmetric structure that includes a contracting path to capture context and an expansive path to enable precise localization. This structure is inherently suited for geospatial analysis, as it allows for the detailed examination of SST patterns at various scales, which ensures both global trends and local anomalies are accurately captured. Together these two architectures can effectively capture both the patterns in SST data and produce more accurate results.

In summary, the contributions of this study can be summarised as follows:

1. The introduction of a multi-dilation ConvLSTM with U-Net architecture for accurate SST forecasting. This hybrid approach utilizes the capabilities of ConvLSTM for sequential data analysis and spatial awareness of U-Net to achieve accurate and thorough SST forecasting.
2. Multi-dilation in convolutions allows for the extraction of information at various scales, enhancing the model's ability to capture both fine-grained details and broader patterns, thereby improving forecasting capabilities.
3. Inclusion of temperature data from multiple ocean depths to provide a comprehensive perspective of SST forecasting. This ensures that the effects of different layers of the ocean are considered.
4. The proposed Multi-dilated ConvLSTM proved to be improving long-horizon forecast after rigorous experiments.

TABLE 1. Comparative overview of proposed and existing related studies.

Algorithm	Data	Time Resolution	Variables Used	Forecast Window	Dilation	U-Net	Data Spatial resolution	Training time period	Spatial: localized/global
D2CL[41]	OISST, NOAA	Daily	SST	27 days	Yes	No	0.25°x0.25°	NA	Localized
LSTM[[39]	Geospatial Agency of Indonesia	Daily	SST	60 days	No	No	NA	NA	Localized
Spatial 2D CNN[42]	ECMWF	NA	SST	30 days	No	No	0.25°x0.25°	NA	Localized
LSTM[20]	NOAA	Daily	SST	NA	No	No	0.25°x0.25°	Sept. 1981 - Aug. 2012	NA
LSTM [40]	NOAA	Weekly	SST	NA	No	No	1° x 1°	NA	Global/Localized
Adaboost-Ensemble LSTM [27]	NOAA	Daily	SST	10 Days	No	No	0.25°x0.25°	NA	Localized
DGRU and CNN [31]	OISST, NOAA	Daily	SST	2 days – 1 month	No	No	1° x 1°	NA	Localized
Proposed Approach	**CMIP6**	**Monthly**	**SST, T300, T700**	**3-12 months**	**Yes**	**Yes**	**0.25°x0.25°**	**1851-2000**	**Global**

5. The proposed approach notably achieves enhanced forecasts without the need for a substantial increase in the number of trainable parameters. This guarantees that the approach is computationally manageable.
6. The suitability of other convolutional approaches like 2D and 3D CNNs along with their dilated models were also examined for capturing the spatial and temporal features of SST.

The paper is organized as follows: Section II discusses the existing and related approaches available for SST forecasting using deep learning methods and provides a comparative analysis with our proposed approach. Section III provides the discussion about the problem formulation. The details about the design of the proposed solution approaches along with their elements and performance metrics for comparative study are provided in Section IV. Section V presents the discussion about experimental setups can comparative analysis of obtained results. Lastly, the overall study is concluded in Section VI with direction towards future work.

II. RELATED WORKS

As already stated in the above section, several methodologies have been developed for SST prediction. In this section, we provide a detailed comparative analysis of our proposed approach with some latest and related approaches used for SST forecasting using deep learning approaches.

Hou et al. [41] proposed a novel approach of dense dilated convolutional LSTM (D2CL) which utilized dilation and daily SST data at 0.25° × 0.25° resolution for a 28-days forecast window. However, the D2CL model lacked U-net architecture and focused only on localized forecasting. Similarly, spatial 2D CNN used by Minnett et al. [3] also predicted at a localized scale with 0.25° × 0.25° dimension resolution, but with a 30-day window and no dilation or U-Net. Authors of [20], [39], and [40] utilized the standard LSTM model for the SSAT variable. Winona and Adytia [39] measured daily time resolution for localized forecasting for a forecasting window of 60 days, Patil et al. [19] also considered daily time resolution with the training time period from September 1981 to August 2012, while Saxena et al. [40] did the longer weekly time resolution with both globalized or localized forecasting. Xiao et al. [27] proposed an Adaboost-Ensemble LSTM which uses daily data at 0.25° × 0.25° resolution for 10-day forecasts but remains localized. Xu et al. [30] utilized a Deep Gated Recurrent Unit (DGRU) combined with CNN and considered daily SST data for localized forecasting with a flexible window of 2 days to 1 month.

Among all of the above studies, a wide range of data sets are used for experimentation including, Optimum Interpolation SST (OISST), National Oceanic and Atmospheric Administration (NOAA), European Centre for Medium-Range Weather Forecasts (ECMWF). These data sets vary in the physical model that has been used. The varied range of temporal and spatial resolutions is another important differentiating factor in these types of datasets.

In conclusion, almost all the discussed approaches focused on localized SST forecasting with resolution ranging from 0.25° to 1°. The time resolution is either daily or in one case weekly, with forecast windows ranging from 2 days to 60 days. The technique to capture long-range dependencies i.e., dilation is not widely adopted in the literature, and all the studies consider only SST as the forecasting variable. In contrast to the above approaches, our proposed approach introduces a novel globalized SST forecasting method utilizing Conv. LSTM with multi-dilation and U-net architecture. Additionally, we use the dataset Coupled Model Intercomparison Project Phase 6 (CMIP6), and our model incorporates multiple atmospheric variables such as SST, T300, and T700. Also, we experimented with a flexible time window ranging from 3 to 12 months. Therefore, with a broader range of variables and global domain, we aim to provide a more comprehensive and inclusive SST forecasting. Table 1 shows a comparison of our proposed approach with contemporary methods and consists of wide attributes viz., data, time resolution, variables, forecast window, dilations and U-Net used, data resolution, training time period, and localized or global prediction.

III. PROBLEM FORMULATION

This section discussed the problem formulation of SST. Let the global SST values corresponding to the grid of earth be denoted by a $I \times J$ grid, where I and J are latitudes and longitudes. Let K be the number of channels in conjunction with the number of layers or variables being considered, for a certain Grid reference system G_i to represent the state at a certain time T_t and each grid point represents the observed SST and potential temperature at depths 300m and 700m at that particular point. The periodic observation can be represented as a sequence of tensors given as $G_1, G_2, G_3, \ldots\ldots G_t$. Given this data, the SST forecasting problem can be defined as:

$$G_{t+l} = \underset{G_{+l}}{\operatorname{argmax}}\, p\left(G_{t+l} \mid G_{t-hs+1}, G_{t-hs+2} \ldots . G_{t-hs+hs}\right) \tag{1}$$

where, l is the lead time, hs is the history size, and p is the probability. The lead time is defined as the count of months that occur between the most recent month with the observed data utilized for forecasting and the target month. History size is defined as the number of months used as input for the forecast. For example, for a given year, as given in Eq. (1), for a lead time of three months, for a target month "December" and a history size of three months, the model uses SST values for months June, July and August to forecast the SST in December. The horizon of the forecast in the future is determined by the lead time and is determined as the difference between the last observed month and the predicted month.

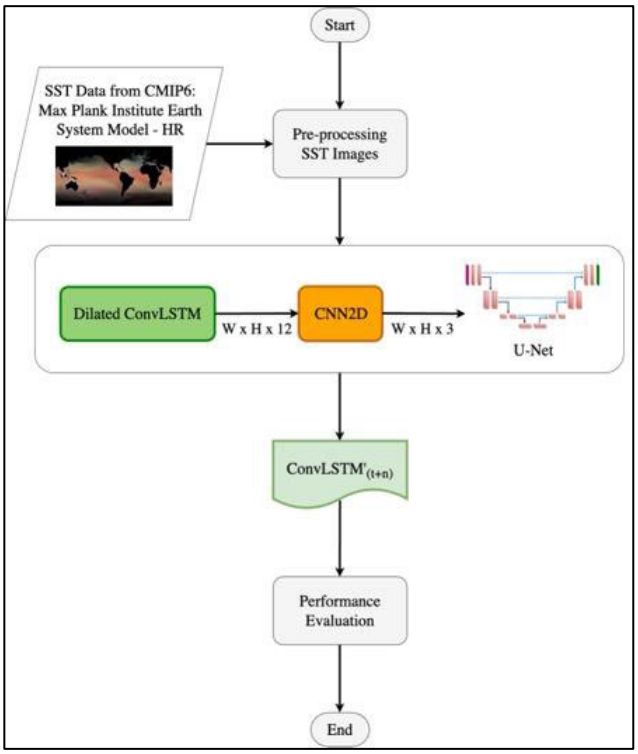

FIGURE 1. Flowchart demonstrating the working of the proposed approach.

IV. PROPOSED APPROACH

In this section, the methodology employed for SST forecasting and the proposed approach is presented. The flow chart of the approach is shown in Fig. 1.

A. DATA DESCRIPTION AND COLLECTION

The study is based on hindcast climatic simulations from the Max Plank Institute Earth System Model's high-resolution version (MPI-ESM1-2-HR) [43] provided by the climate modelling centre at Max Plank Institute of Meteorology, Germany that participated in Coupled Model Intercomparison Project (CMIP) phase 6. This dataset consists data of very high spatial resolutions which is pivotal for capturing even fine variations. Moreover, the dataset comprises the data over longer periods of time with a varied range of atmospheric parameters. The model output used is based on SST historical simulations spanning the period from 1850 to 2014 (past ∼1.5 centuries) [43]. For any experiment, every simulation of an ensemble of runs by a model is uniquely identified through realization (initial conditions), initialization (procedure, i.e., algorithms used), physics (model parameterizations), and forcing index (ripf indices). The study is based on single ensemble member of the model i.e., r1i1p1f1. The present study is based on the model's native tripolar grid with a spatial resolution of ∼ 0.4° and a monthly temporal resolution. Besides SST, the study uses potential temperature integrated

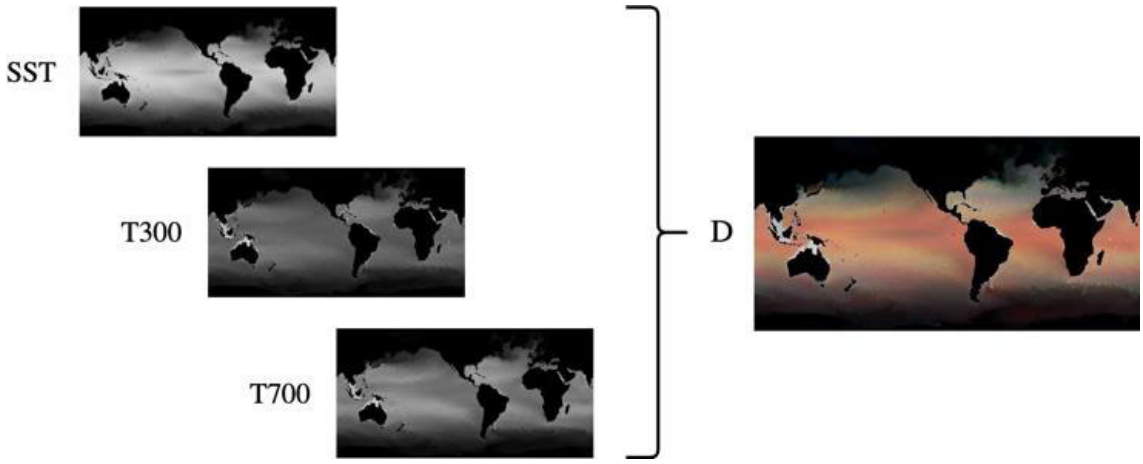

FIGURE 2. Final Image Input (D) consists of 3 stacked images (A) SST (B) Potential Temperature averaged at 300 m below sea level and (C) Potential Temperature averaged at 700 m below sea level.

at ocean depth of 0-300 meters (T300) and 0-700 (T700) meters for SST prediction.

The study uses historical global SST data and sub-surface potential temperature data (lagged data) to forecast global SST in the future. The DL models are trained using three (SST, T300, T700) 404 × 808 (latitude × longitude) images over monthly data from 1851-1990 (1680 months). The three layers of the ocean are represented as three channels of an image. Fig. 2 shows the three layers of the ocean and the combined image. Further, training was tested and validated on the data from 2000-2014. To remove any ocean memory impact, a ten-year gap is taken between the training and testing period.

B. DATA PRE-PROCESSING

In the data pre-processing stage, for the effectiveness of the U-Net segmentation model, we initiated the workflow by regridding the MPI-ESM1-2-HR climate model outputs, originally in the netCDF format with spatial resolution of 404 by 808 grid points, to a refined grid of 448 by 896 using the Climate Data Operators (CDO) [44] tool. This particular resizing ensures that each downscaling step in the U-Net halves the input dimension, thereby maintaining the model's structural integrity without encountering incompatible tensor dimensions. Following the regridding process, normalization was applied to the dataset, standardizing the range of independent variables or features of data. In the context of our climate data, normalization is a critical step as it ensures the consistency of numerical ranges between variables, thus facilitating a faster and more stable convergence during the model training phase. This approach also aids in improving the gradient flow through the network, which is essential for the learning process of deep learning architectures, particularly those dealing with heterogeneous datasets like climate model outputs. Therefore, in this study, we have used the StandardScaler function from the scikit-learn library. This approach standardizes by subtracting the mean, thereby moving towards a unit variance [60], [61].

C. MULTI-DILATION ConvLSTM WITH U-Net (MDCU-Net)

The proposed model with multi-dilated ConvLSTM and U-Net (termed as MDCU-Net) is discussed in this section. Initially, the different blocks of the proposed model are discussed, followed by the proposed model. The proposed model is shown in Fig. 4 and the corresponding algorithm is compiled as Algorithm 1.

Algorithm 1 MDCU-Net: Multi-View Convolutional LSTM-Based UNET for SST Forecast

Input: $G_i, G_{i+1}, G_{i+2}, G_{i+3}, \ldots\ldots\ldots\ldots G_{i+hs}$

Output: Forecasted SST, MSE, SSIM, EMD, Cosine Similarity

```
Approach MDCU-Net() {
Data collection and pre-processing (tos_i, T300_i, T700_i)
  Regrid (tos_i, T300_i, T700_i)
    Normalize(tos_i, T300_i, T700_i)
    tos_i, T300_i, T700_i → G_i
 for (hs)
     for Lead
       for dilation_rate
          d_dilation_rate →ConvLSTM(G_i, hs, Lead,
        dilation_rate)
            Concat (∅, d_dilation_rate)
        STMD_Gi → CNN(∅)
       SST_i+Lead ← U-Net (STMD_Gi)
       MSE(SST_i+Lead, G_i+lead[:][:][0] )
       SSIM(SST_i+Lead, G_i+lead[:][:][0])
       EMD(SST_i+Lead, G_i+lead[:][:][0])
       Cosine_Similarity(SST_i+Lead, G_i+lead[:][:][0]) }
```

1) CONVOLUTIONAL LSTM

Convolutional Dilation is a variation of traditional CNN used in semantic segmentation which has garnered considerable

attention lately [45], [46]. The presence of a pooling layer compromises network accuracy and layer deepening increases computational cost manifold in conventional CNN. A dilated convolution optimizes the performance of CNN by replacing the convolutional kernel with a dilated kernel which expands the receptive field allowing more information retrieval with the same number of parameters without an increase in computational and memory cost. A dilated convolutional kernel is formed by introducing gaps or holes (with 0 weight) between the values in the convolutional kernel as shown in Fig 3, for a kernel of size 3. The number of gaps is referred as ''dilation rate''. Dilation increases the receptor field according to the equation.

$$k_r = k + (k-1) \times (r-1) \tag{2}$$

where, k is the size of the filter and r is the dilation rate.

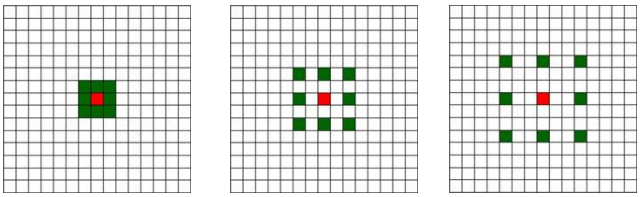

FIGURE 3. Dilated convolution kernels with multi-dilation rates of (a) 1, (b) 2, and (c) 3.

Dilation, without any loss in resolution supports exponential expansion of the receptive field. The study models geospatial data, therefore, the receptor field here represents the actual area being considered on the earth grid. This data is complex multi-dimensional data that incorporates multi-scale spatial and temporal dependencies. Yu and Koltun [45] make the case for the application of dilated convolutions to multi-scale contextual data for dense prediction tasks. The convolution operator in a conventional ConvLSTM is replaced by a dilated convolution operator. This operation is given by the Eq. (3):

$$Y = X \widetilde{\circledast} W_{k_r} \tag{3}$$

where Y is the output, X is the input, $\widetilde{\circledast}$ is the dilated convolution operator, and W_{k_r} is the kernel. Thus, in the ConvLSTM equations, the convolution operator can be replaced by the dilated convolution operator.

$$i_t = \sigma\left(W_{xi} \circledast \widetilde{X}_t + W_{hi} \circledast \widetilde{H_{t-1}} + W_{ci} \diamond C_{t-1} + b_i\right) \tag{4}$$

$$f_t = \sigma\left(W_{xf} \widetilde{\circledast} X_t + W_{hf} \widetilde{\circledast} H_{t-1} + W_{cf} \diamond C_{t-1} + b_f\right) \tag{5}$$

$$\widetilde{C}_t = tanh\left(W_{xc} \widetilde{\circledast} X_t + W_{hc} \widetilde{\circledast} H_{t-1} + b_c\right) \tag{6}$$

$$C_t = f_t \diamond C_{t-1} + i_t \diamond \widetilde{C}_t \tag{7}$$

$$o_t = \sigma\left(W_{xo} \widetilde{\circledast} X_t + W_{ho} \widetilde{\circledast} H_{t-1} + W_{co} \diamond C_t + b_o\right) \tag{8}$$

$$H_t = o_t \diamond tanh\left(C_t\right) \tag{9}$$

2) MULTI-DILATION FEATURE EXTRACTION

Multi-dilation is an advanced variation of dilation in CNNs. Here, multiple dilation rates are employed in the same network. This enables capturing features at various scales, thereby learning representations at different granularities. Utilizing various dilation rates in convolutional networks facilitates the extraction of information across multiple levels. This approach improves the model's capacity to discern both intricate details and wider trends. In this proposed approach, we are using four ConvLSTM layers with a kernel size of 3. Figure 5 shows the convolutional kernel with different dilation rates. The layers use dilation rates ranging from 1 to 3. Therefore, going by Eq. (2), the receptor area of each of the layers is defined as follows:

$$Layer_1 = (3 * 3)$$
$$Layer_2 = (5 * 5)$$
$$Layer_3 = (7 * 7)$$

Further, the outputs of all these layers are concatenated together. One thing worth mentioning here is that, even after concatenation, the number of parameters remains the same as the kernel size remains fixed. Multi-dilation allows the network to capture features at multiple scales simultaneously. Each dilated ConvLSTM layer, with its respective receptive field, captures patterns and dependencies over time and across different scales within the SST, T300, and T700 data. The outputs from these layers, which encapsulate temporal dynamics at varying spatial scales due to the dilation, are concatenated. The concatenation of feature maps with different receptive fields is not just a simple sum; it combines the distinctive perspectives into a unified representation that can be described by the combined extent of their individual receptive fields. The produced feature map encapsulates the information required for the subsequent forecasting task. From Eq. (8), it can be derived that

$$o_t = concat\left(o_{t,1}, o_{t,2}, o_{t,3}, o_{t,4}\right) \tag{10}$$

where $o_{t,r}$ denotes the output of the ConvLSTM cell at time t and a dilation rate r of the convolutional operator. Similarly, the hidden output can be written as,

$$H_t = concat\left(H_{t,1}, H_{t,2}, H_{t,3}, H_{t,4}\right) \tag{11}$$

Therefore, the total output of the dilation rates can be given by the equation

$$Multi - view\_features = concat(ConvLSTM_{D1} + ConvLSTM_{D2} + ConvLSTM_{D3} + ConvLSTM_{D4}) \tag{12}$$

Here, $ConvLSTM_{Dn}$ represents the output from a ConvLSTM layer with a dilation rate n, and $Concat$ denotes the concatenation operation. This output is a multi-channel feature map where each set of channels corresponds to the features extracted at a particular dilation rate. The output of this concatenation is then fed into a Conv2D layer that integrates these multi-scale temporal-spatial features through convolutional processing. For a given position (x, y)

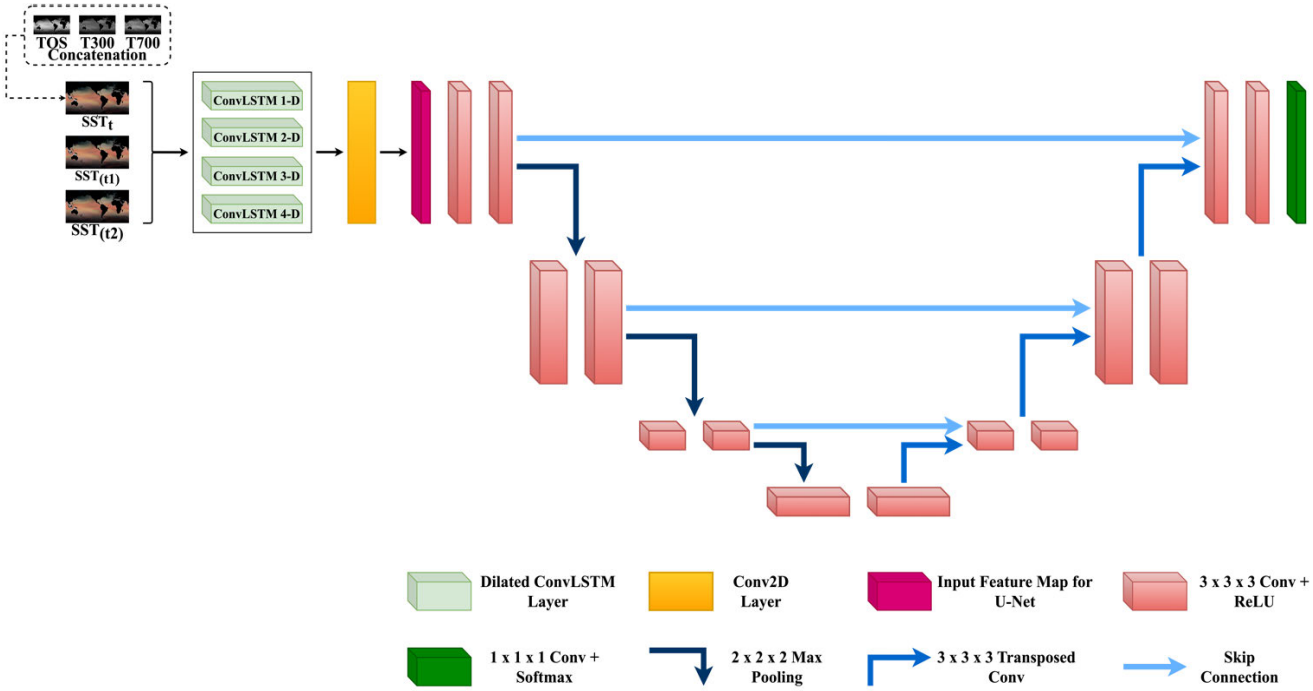

FIGURE 4. Proposed multi-dilation-based ConvLSTM with U-Net.

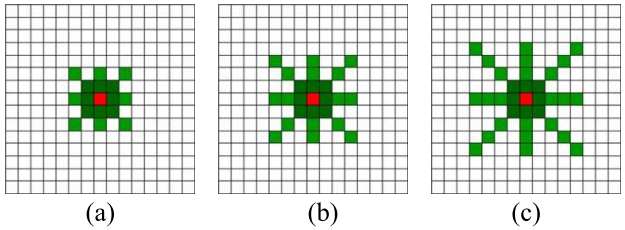

FIGURE 5. Dilated convolution kernels with multi-dilation rates of (a) (1,2), (b) (1, 2, 3), and (c) (1, 2, 3, 4).

in this multi-channel feature map F, convolution operation is defined as:

$$F(x, y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} I(x+i, y+i) .K(i+a, j+b) \quad (13)$$

where, i and j iterate over the kernel dimensions, $K(i+a, j+b)$ refers to the value at the position $(i+a, j+b)$ in the kernel K. The values a and b are determined by the kernel's size. For a $(2a+1) \times (2b+1)$ kernel, a and b define the extent to which the kernel extends from its centre. The convolutional layer extracts the spatio-temporal multi-dilated (STMD) features from the ConvLSTM using Eqs. (4)-(9).

3) U-Net

U-Net is a modified CNN whose architecture mimics the alphabet 'U', hence the name. Originally applied for image segmentation to sparse biomedical data [47], the model is being introduced to diverse prediction problems [48], [49], including the prediction of SST [50]. A U-Net model basically consists of 2 paths or subnetworks. A contracting path (variably called the encoder, analysis, downsampling path) is similar to an archetypical CNN that uses convolution, pooling and non-linear activation functions (ReLu) to classify/identify features at mesoscale and synoptic scales [51]. On the other hand, the expanding path (also called the decoder, synthesis, upsampling path) offers the U-Net its uniqueness by allowing localisation, i.e. pixel-by-pixel class labelling. Here, the feature from a contracting layer is concatenated with a corresponding upsampled layer (using a transposed convolution layer) in the expanding path, followed by 2 successive convolutions and ReLu activation. Finally, a 1×1 convolution is employed to get the desired number of channels, generating a precise segmented image. U-Net lacks any fully connected layers and uses only those pixels for which full context is available in the input data, allowing segmentation of large images by the overlap-tile method [47], [52].

The concatenated output from Eq. (10), now encoded with rich temporal-spatial information using Eq. (11), serves as the input to a U-Net architecture. The U-Net, with its encoding-decoding structure, refines the feature map through downsampling and upsampling paths, incorporating skip connections to preserve spatial information. Ultimately, a 1×1 convolution and a SoftMax activation function generate the final forecasted SST, mapping the complex input relationships to a predictive output.

TABLE 2. Comparative study of results of six models for history size 3 and 6 combination with lead 3.

	History Size =3				History size=6				History size=12			
	MSE	SSIM	EMD	Cosine Similarity	MSE	SSIM	EMD	Cosine Similarity	MSE	SSIM	EMD	Cosine Similarity
CNN2D	0.05261	0.90183	0.08059	0.99112	0.34175	0.43111	0.84243	0.67092	0.40174	0.23199	0.90137	0.62765
CNN2D'	0.01436	0.96371	0.05844	0.99900	0.29107	0.49710	0.51873	0.75132	0.28582	0.46359	0.44408	0.74234
CNN3D	0.01416	0.96695	0.04987	0.99942	0.28291	0.46678	0.25036	0.72892	0.34321	0.54299	0.40965	0.77240
CNN3D'	0.01311	0.9695019,	0.04238	0.99950	0.27882	0.49941	0.46894	0.75707	0.29977	0.70827	0.40286	0.79269
ConvLSTM	0.01294	0.97067	0.05209	0.99951	0.11027	0.81003	0.22462	0.97202	0.15631	0.75281	0.30239	0.91289
MDUNet	**0.01267**	**0.97205**	**0.04658**	**0.99954**	**0.08260**	**0.87343**	**0.24613**	**0.98513**	**0.11204**	**0.81204**	**0.22029**	**0.92298**

D. OTHER COMPARISON MODELS

The above-proposed model is compared with other convolutional approaches in conjunction with the U-Net architecture to benchmark and evaluate our proposed multi-dilation ConvLSTM framework. These comparison models include both simple convolutional models and dilated convolutional models, providing a comprehensive analysis of their effectiveness in SST forecasting.

Simple Convolutional Models:

i. *Convolutional 2D (CNN2D):* It is a straightforward model that uses 2D convolutional layers for feature extraction. It operates on the SST data without considering the temporal dependencies between consecutive timesteps. The architecture of this model uses two convolutional 2D layers with 12 and 3 size filters respectively followed by U-Net. This model considers the input as one image with 9 channels.

ii. *Convolutional 3D (CNN3D):* It extends the 2D convolutional approach to incorporate the temporal dimension. This model employs 3D convolutional layers, which have the capability to capture both spatial and temporal features. The architecture of this model has two convolutional 3D layers with filter sizes 12 and 3 respectively along with stride rate 3. In this model, each month is considered as a single 3D image.

iii. *Convolutional LSTM (ConvLSTM):* It integrates the LSTM units with the convolutional layers extending the LSTM's capability to process spatial information by incorporating convolutional operations. The model architecture uses two ConvLSTM layers with filter sizes 12 and 3, respectively, considering each month's data as an image with temporal relations with the previous months.

Since the filter sizes are kept constant throughout the above three models, the number of trainable parameters is also the same for all three models. This has been done for a fair comparison when evaluating the different models.

Dilated Convolutional Models:

i. *Dilated Convolutional 2D with U-Net (also denoted as CNN2D'):* Continuing from the proposed model, Convolutional 2D with dilation has been used. This modification of dilation enables the model to capture the spatial pattern with a large receptive field while maintaining computational efficiency. This approach consists of four dilated convolutional 2D layers with dilation rates ranging from 1 to 4. The base filter size of each of the layers is (3×3), however with dilation the dilated filter sizes are (3×3), (5×5), (7×7) and (9×9). This allows the model to capture the spatial features at varied receptive fields. The output of the 4 layers is concatenated and then fed to a convolutional 2D layer with a filter size of 3, followed by the U-Net. This approach also considers the input as one image with 9 filters.

ii. *Dilated Convolutional 3D with U-Net (also denoted as CNN3D'):* Extending the Dilated Convolutional 2D model, the dilated CNN3D model applies the dilation rates to the 3D convolutional layers. This in turn allows the model to capture spatio-temporal patterns with varying scales. Four CNN3D layers with dilation rates ranging from 1 to 4 with their outputs concatenated are fed as input to a convolutional 3D layer. In this model, each month is considered as one 3D image. The U-Net model is further used to forecast the SST in the future.

E. PERFORMANCE EVALUATION METRICS

1) MEAN SQUARED ERROR (MSE)

MSE is used to quantify the average squared difference between the predicted SST values and the corresponding ground truth values across all samples. It is calculated using the equation:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2 \tag{14}$$

where, N is the number of samples, y_i is the actual SST values for the sample i, $\widehat{y_i}$ is the predicted SST for the sample

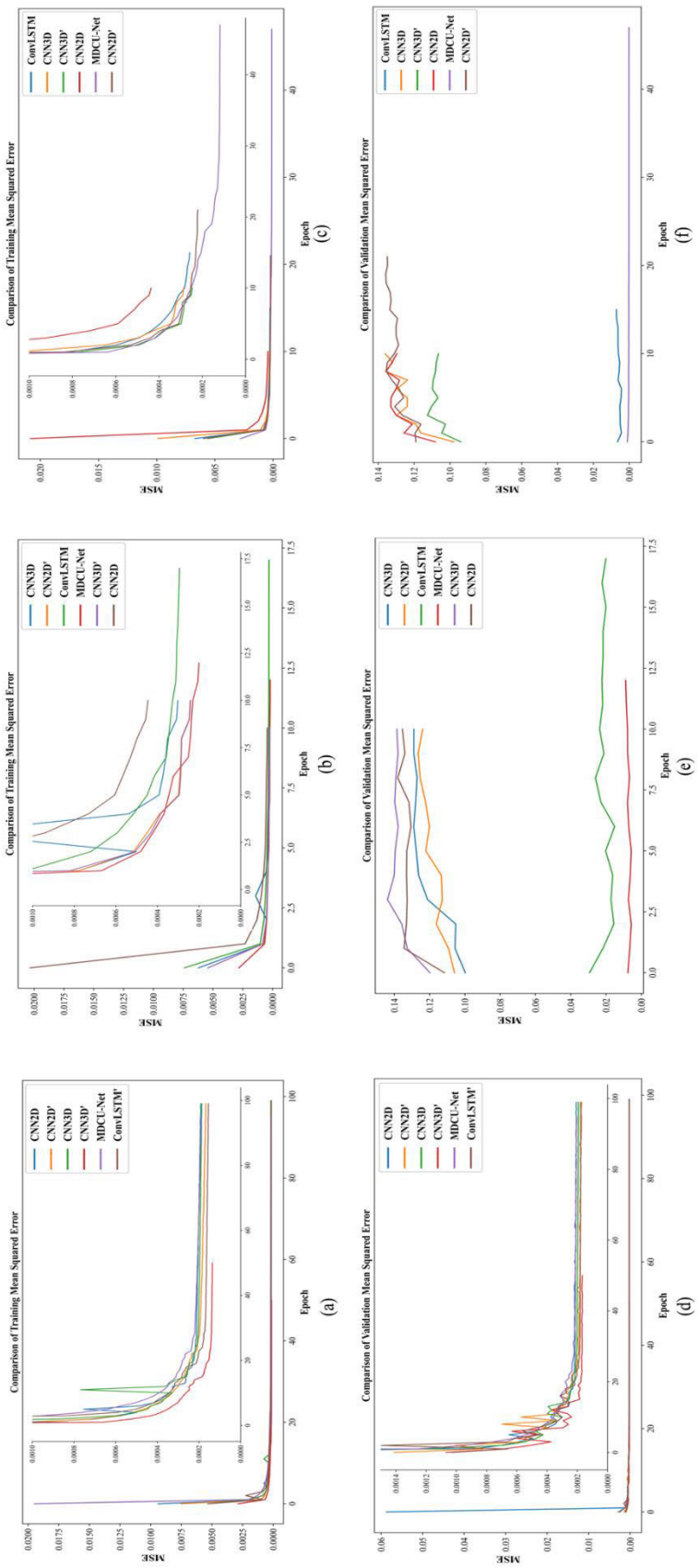

FIGURE 6. (a)-(c) Training mean squared error for lead 3, lead 6, and lead 12, respectively, (d)-(e) Validation mean squared error for lead 3, lead 6, and lead 12, respectively.

i. A lower MSE indicates a better agreement between the predicted and actual SST values.

2) STRUCTURAL SIMILARITY INDEX (SSIM)

SSIM [53] measures the similarity between two images by considering their luminance, contrast, and structure. It's computed using the following formula:

$$SSIM\ (x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{15}$$

where, x and y are the input images (in our case, predicted and ground truth SST images), μ_x and μ_y are the means of x and y, σ_x^2 and σ_y^2 are the variances of x and y, σ_{xy} is the covariance of x and y, and c1 and c2 are constants to stabilize the division with a weak denominator. SSIM ranges from -1 to 1, where 1 indicates perfect similarity between the images.

3) EARTH MOVERS DISTANCE (EMD)

EMD [54] measures the dissimilarity between two probability distributions by computing the minimum "work" required to transform one distribution into the other. It's evaluated using the following formula:

$$EMD\ (P, Q) = min \sum_{i=1}^{n} \sum_{j=1}^{m} f_{ij}.d_{ij} \tag{16}$$

where, P and Q are the probability distributions of SST values in the predicted and ground truth images, f_{ij} represents the amount of mass to be moved from the bin i of distribution P to bin j of distribution Q. d_{ij} is the distance between bin i and j (Euclidian distance). The minimum is taken over all possible assignments of mass. EMD measures the dissimilarity between the distributions, with lower values indicating greater similarity.

V. EXPERIMENTAL RESULTS

This section discusses in detail about the dataset used for experiments, training and experimental settings and a comparative study of results obtained using all developed techniques.

A. TRAINING AND EXPERIMENTAL SETUP

In the training setup of the MDCU-Net, various parameters have been carefully configured for optimal training results. To prevent overfitting of the model early stopping has been used with the patience parameter set to 15. It allows the training process to stop when the model's performance on the validation set does not improve. A learning rate scheduler, ReduceLROnPlateau is used as an adaptive learning rate (LR) strategy which adjusts the learning rate dynamically during the training. The parameters for this scheduler were patience $= 4$ and factor $= 0.3$ i.e., when the validation loss does not decrease for 4 epochs, the learning rate is reduced by a factor of 0.3. The minimum learning rate is set to 0.00001, which ensures that the learning rate does not drop below this level.

TABLE 3. Comparative study of results of six models for history size 3 combination with Lead 3, 6, and 12.

	MSE	SSIM	EMD	Cosine Similarity
LEAD 3				
CNN2D	0.05261	0.90183	0.08059	0.99112
CNN2D'	0.01436	0.96371	0.05844	0.99900
CNN3D	0.01416	0.96695	0.04987	0.99942
CNN3D'	0.01311	0.96950	0.04238	0.99950
ConvLSTM	0.01294	0.97067	0.05209	0.99951
MDCU-Net	0.01267	0.97205	0.04658	0.99954
LEAD 6				
	MSE	**SSIM**	**EMD**	**Cosine Similarity**
CNN2D	0.28192	0.43497	0.30076	0.73076
CNN2D'	0.29913	0.50639	0.34471	0.69675
CNN3D	0.33614	0.48332	0.84920	0.67167
CNN3D'	0.28072	0.52922	0.64749	0.75433
ConvLSTM	0.13162	0.77363	0.27330	0.95557
MDCU-Net	0.08247	0.86181	0.27319	0.98363
LEAD 12				
	MSE	**SSIM**	**EMD**	**Cosine Similarity**
CNN2D	0.31819	0.48311	0.42967	0.66485
CNN2D'	0.29754	0.44708	0.41403	0.70637
CNN3D	0.26217	0.59511	0.65126	0.77846
CNN3D'	0.27401	0.54290	0.59420	0.77997
ConvLSTM	0.08572	0.84207	0.09972	0.97818
MDCU-Net	0.01224	0.97343	0.03621	0.99957

The models were trained for 200 epochs using the Adam optimizer utilizing stochastic gradient descent and backpropagation. The loss function used was binary cross-entropy. It is a commonly used loss function that measures the dissimilarity between the predicted and actual values. The model was trained on a system with an AMD Ryzen 7 3700x processor with 32 GB RAM and an NVIDIA RTX 2080ti GPU with 11 GB VRAM. The deep learning frameworks used for the implementation of the above models are TensorFlow and Keras.

B. RESULTS ANALYSIS

The experimental results have been obtained using six different developed techniques for the problem studied. The techniques include three non-dilated models namely: CNN2D, CNN2D, and ConvLSTM and the other three are their dilated counterparts termed as, CNN2D', CNN3D' and the proposed MDCU-Net. Each of these techniques was executed for Lead 3, Lead 6 and Lead 12 combined with the history size of 3, 6 and 12. In order to evaluate the performance of developed techniques, we conducted four popular metrics: MSE, SSIM, EDM and Cosine similarity [56], [57] over the obtained experimental results.

We first conducted experiments with history size 3, 6 and 12 with Lead 3 to observe the behaviour of the approaches. The history sizes are in the ratio of 3 because, which is set to evaluate the seasonality parameter as SST exhibits a seasonal

cycle [55]. It can be witnessed from Table 2 that the lower the length of history size better is the performance of approaches. This is because the larger history size bears complex patterns and variations which makes it complex for techniques to capture it. However, it can also be observed that MDCU-Net is able to capture these variabilities in a better way compared to others but still, it has not been able to provide a better forecast. Therefore, for further exhaustive experiments, we set history size to 3 with combinations of Lead 3, 6, and 12 to avoid the effect of seasonality patterns. The training and validation MSE are given in the Fig. 6. The proposed MDCU-Net performs the best among the other comparison algorithms. Since early stopping was used the MDCU-Net model converged earlier than all the other algorithms. Table 3 provides comparative results of all six techniques for the combination of history size 3 with Lead 3, 6 and 12 which depicts short-term forecast, medium-term forecast and long-term forecast, respectively.

Observations Per Lead Time:

Lead 3 (Short-Term Forecast): The MSE values resulting from the dilated models have lower values which signifies better performance in terms of average squared differences from the actual SST values. And this is the lowest for our proposed MDCU-Net with a value of 0.01267. In terms of the SSIM evaluation, all models perform very well with scores above 0.9, indicating high structural similarity. The dilated models, however, slightly outperformed their non-dilated counterparts, which suggests that dilation may help capture the spatial structure in SST data efficiently. Again, our proposed model resulted in a better SSIM value of 0.97205. Similar is the case with the evaluation of EMD values, which needs to be lower and MDCU-Net resulted in an EMB value of 0.04658. The cosine similarity signifying the alignment of the directionality of the forecasts with the actual data, is extremely high (above 0.99) for all models, and even in that the proposed model even though marginally, performed better with the value of 0.99953.

Lead 6 (Medium-Term Forecast): For this Lead, there's a notable improvement in MSE for all models as compared to Lead 3. Among all, the MDCU-Net model shows the best performance with a value of 0.08247. In contrast to this, the scores for SSIM have decreased compared to the previous Lead 3, particularly for CNN2D models. The MDCU-Net still maintains a desired high SSIM of 0.86181. For EMD, MDCU-Net shows minimal improvement, indicating a relatively better forecast in terms of distribution. The cosine similarity for this Lead is much more improved percentage-wise for the proposed approach. However, this measure shows poor performance for the CNN2D' as compared to their non-dilated version.

Lead 12 (Long-Term Forecast): Our proposed approach has shown remarkable performance for this Lead in all the evaluation metrics. The MSE, SSIM, EMD, and cosine similarity are 0.01224, 0.97343, 0.03621, and 0.99957, respectively which is significantly better than the other models of CNN2D and CNN3D, their corresponding non-dilated versions.

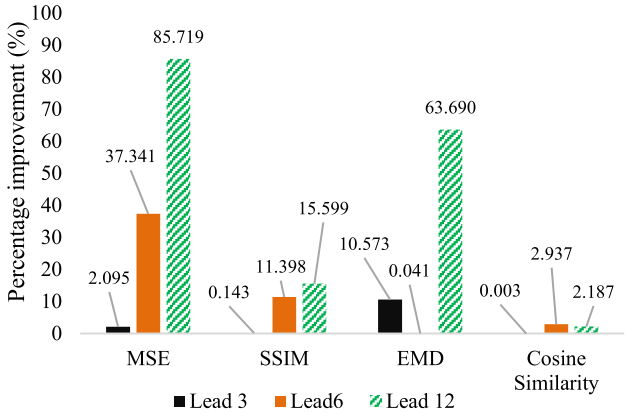

FIGURE 7. Percentage improvement of different metrics of ConvLSTM and MDCU-Net for Lead 3, 6, and 12.

In conclusion of the Table 3 analysis, we observed that the dilated versions of the models tend to significantly outperform the regular ones, particularly in MSE, SSIM, and EMD metrics, which suggests that dilation contributes positively to capturing the SST patterns accurately. Moreover, ConvLSTM models showed better performance across all metrics including cosine similarity, especially in the medium to long-term forecasts, which indicates their effectiveness in capturing both the spatial and temporal dynamics of SST data. However, it has to be noted here that the cosine similarity remains consistently high, which doesn't actually reflect the increase in MSE and EMD or the decrease in SSIM with longer lead times. This suggests that while forecasts might align directionally, they can still be quite different in terms of magnitude and structure, which is critical for SST forecasting. Therefore, cosine similarly has its limitations, and our consideration of several crucial metrics provides an overly optimistic and holistic view of the forecast quality. Also, despite the dominance of the annual cycle, it has been argued [55] that the SST annual cycle can also change significantly over time. This is particularly important in the context of climate change, where shifts in SST patterns can indicate broader environmental shifts. Furthermore, testing our model's ability to forecast over a 12-month period, despite the inherent challenges, is critical for validating and refining its predictive capabilities. It allows us to understand the limitations of the model's utility in practical scenarios and to make necessary adjustments to improve the accuracy.

It can be clearly observed that the dilated variants of techniques generate better metrics values compared to their non-dilated counterparts for every Lead size. This states that the incorporation of the dilation method enhances the capability of the model to learn from spatial relationships within the data and capture the SST patterns accurately. Moreover, the performance metrics value for Lead 3 are far better compared to Lead 6 and Lead 12. This suggests that the short-term forecast provides better accuracy, capturing the spatial structure in SST and achieving SST distributions closer to the actual SST distributions. After the analysing the results in terms

of technique comparison, the non-dilated ConvLSTM models outperform CNN2D and CNN3D in every measurement. It also performs well for forecasting efficiently not only for Lead 3 but also for Lead 6 and Lead 12. This establishes that ConvLSTM is also able to make better predictions, especially for medium and long-term forecasts.

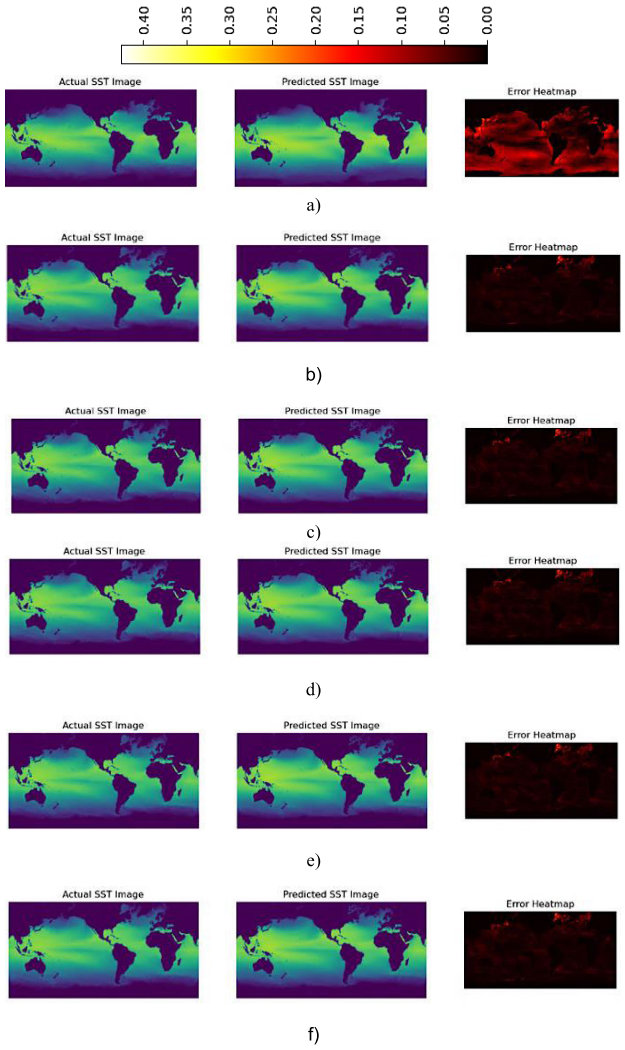

FIGURE 8. Prediction vs the actual SST Image vs error heat map in Lead 3 for (a) CNN2D, (b) CNN2D', (c) CNN3D, (d) CNN3D', (e) ConvLSTM, and (f) MDCU-Net.

On the other hand, dilated MDCU-Net outperforms all other techniques for every combination of Lead size. Because of the larger receptive field is able to model the SST in a better way. Therefore, MSE for CNN2D' is better compared non-dilated variant. In the case of CNN2D' and 3D', the CNN3D' model uses a 3-dimensional filter to perform convolutions which helps to capture the non-linear relationship 3 variables (TOS, T_300, T_700). However, the difference in MSE between CNN3D and CNN3D' is not as significant as in the case of CNN2D as this can be attributed to the 3-dimensional nature of the kernel. Figure 7 provides the bar plot for percentage improvement in performance metrics of ConvLSTM and MDCU-Net for Lead 3, 6, and 12. Further, figure 8 provides a graphical representation to demonstrate the difference between actual SST vs predicted SST image along with the corresponding error heatmap generated using the results obtained from CNN2D, CNN2D', CNN3D, CNN3D', ConvLSTM, and MDCU-Net, respectively. The results are for the month of January 2012. It can be easily seen that the dilated variants bear less error compared to their non-dilated counterparts. Moreover, ConvLSTM-based techniques have very minor errors.

VI. DISCUSSION, CONCLUSION, AND FUTURE WORKS

Forecasting SST with precision is critical for advancing our understanding and management of various oceanic and climatic phenomena, including weather patterns, marine ecosystems, and climate change impacts. In this study, we introduce a novel data-driven approach which combines multi-dilated convolutions with ConvLSTM networks and U-Net architecture. This novel architecture is designed to enhance the accuracy of monthly SST predictions over lead time of 3, 6 and 12 months. Our approach uniquely integrates dilated convolutions within ConvLSTM to capture the intricate spatio-temporal dynamics of SST. These dilated convolutions increase the receptive field, enabling the model to capture broader spatial patterns like oceanic currents and temperature gradients without losing local detail. Additionally, the framework allows for better long-range spatial understanding with fewer parameters, thus improving computational efficiency. The integration of segmentation strategy through U-Net effectively isolates and analyses pertinent features within the SST data, which ensures enhanced precision. Further, this architecture, with its encoder-decoder structure, efficiently retains fine spatial details through skip connections, which are crucial for preserving the resolution of SST features during forecasting. Utilizing the high-resolution MPI-ESM1-2-HR dataset, we conducted comprehensive experiments to evaluate the performance of our model across various forecast leads and historical data sizes. The experimental results demonstrate the model's superior forecasting capabilities, achieving a MSE of 0.08, which signifies a high degree of accuracy in predicting the actual SST values. Furthermore, the model exhibited a SSMI of 0.21, indicating its effectiveness in capturing the SST's microwave emissivity characteristics. EMD score of 0.99 further underscores the model's proficiency in matching the predicted SST distribution with the actual distribution, while a cosine similarity score of 0.67 reflects a favourable alignment between the predicted and actual SST directional patterns.

By successfully integrating multi-dilated convolutions with a segmentation approach, our model addresses the critical challenge of concurrently capturing the SST's complex spatial and temporal dependencies, inherent in such type of data. This research not only enhances the state-of- the-art in SST forecasting within the domain of data-driven statistical

methods but also offers valuable insights for the development of advanced predictive models in oceanography and climatology. The promising results advocate for the potential application of our model in various related domains, including weather forecasting, marine resource management, and climate change mitigation strategies.

Data driven ML and DL models are increasingly used to understand extremely voluminous and complex earth data, comprehend non-linear relationships and build computationally efficient prediction models. Having said that, these data driven methods do face challenges, which includes the inability to incorporate physical laws that govern the earth system. The proposed MDCU-Net is a data-driven model which only focus on the oceanic variables, which is posed to complement the dynamical models or evolve into physics-informed models for modelling climate efficiently. However, the interactions between atmospheric and oceanic variables could provide suitable intertwined environmental features for better and longer forecast leads. In addition, it could also provide the ability to work in isolation with seasonality. Therefore, future work shall focus on refining our models performance by combining atmospheric and oceanic variables. Also, the accuracy of the model can be tested in forecasting various phenomena based on SST, for instance, ENSO, IOD, etc.

REFERENCES

[1] V. Banzon, T. M. Smith, T. M. Chin, C. Liu, and W. Hankins, "A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies," *Earth Syst. Sci. Data*, vol. 8, no. 1, pp. 165–176, Apr. 2016.

[2] A. Bentamy et al., "Review and assessment of latent and sensible heat flux accuracy over the global oceans," *Remote Sens. Environ.*, vol. 201, pp. 196–218, Nov. 2017.

[3] P. J. Minnett, A. Alvera-Azcárate, T. M. Chin, G. K. Corlett, C. L. Gentemann, I. Karagali, X. Li, A. Marsouin, S. Marullo, E. Maturi, R. Santoleri, S. Saux Picart, M. Steele, and J. Vazquez-Cuervo, "Half a century of satellite remote sensing of sea-surface temperature," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111366.

[4] E. S. Poloczanska, C. J. Brown, W. J. Sydeman, W. Kiessling, D. S. Schoeman, P. J. Moore, K. Brander, J. F. Bruno, L. B. Buckley, M. T. Burrows, C. M. Duarte, B. S. Halpern, J. Holding, C. V. Kappel, M. I. O'Connor, J. M. Pandolfi, C. Parmesan, F. Schwing, S. A. Thompson, and A. J. Richardson, "Global imprint of climate change on marine life," *Nature Climate Change*, vol. 3, no. 10, pp. 919–925, Oct. 2013.

[5] K. E. Trenberth, J. T. Fasullo, and T. G. Shepherd, "Attribution of climate extreme events," *Nature Climate Change*, vol. 5, no. 8, pp. 725–730, Aug. 2015.

[6] R. Ruela, M. C. Sousa, M. deCastro, and J. M. Dias, "Global and regional evolution of sea surface temperature under climate change," *Global Planet. Change*, vol. 190, Jul. 2020, Art. no. 103190.

[7] W. A. Atiah, G. M. Tsidu, L. K. Amekudzi, and C. Yorke, "Trends and interannual variability of extreme rainfall indices over Ghana, West Africa," *Theor. Appl. Climatol.*, vol. 140, nos. 3–4, pp. 1393–1407, May 2020, doi: 10.1007/s00704-020-03114-6.

[8] G. Ottersen, S. Kim, G. Huse, J. J. Polovina, and N. C. Stenseth, "Major pathways by which climate may force marine fish populations," *J. Mar. Syst.*, vol. 79, nos. 3–4, pp. 343–360, Feb. 2010.

[9] C. Deser, M. A. Alexander, S.-P. Xie, and A. S. Phillips, "Sea surface temperature variability: Patterns and mechanisms," *Annu. Rev. Mar. Sci.*, vol. 2, no. 1, pp. 115–143, Jan. 2010, doi: 10.1146/annurev-marine-120408-151453.

[10] B. Jha, Z.-Z. Hu, and A. Kumar, "SST and ENSO variability and change simulated in historical experiments of CMIP5 models," *Climate Dyn.*, vol. 42, nos. 7–8, pp. 2113–2124, Apr. 2014, doi: 10.1007/s00382-013-1803-z.

[11] T. Laepple and P. Huybers, "Ocean surface temperature variability: Large model–data differences at decadal and longer periods," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 47, pp. 16682–16687, Nov. 2014, doi: 10.1073/pnas.1412077111.

[12] M. Haghbin, A. Sharafati, D. Motta, N. Al-Ansari, and M. H. M. Noghani, "Applications of soft computing models for predicting sea surface temperature: A comprehensive review and assessment," *Prog. Earth Planet. Sci.*, vol. 8, no. 1, p. 4, Dec. 2021, doi: 10.1186/s40645-020-00400-9.

[13] B. Huang, C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang, "Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1," *J. Climate*, vol. 34, no. 8, pp. 2923–2939, Apr. 2021.

[14] K. C. Tripathi, I. M. L. Das, and A. K. Sahai, "Predictability of sea surface temperature anomalies in the Indian Ocean using artificial neural networks," *Indian J. Mar. Sci.*, vol. 35, no. 3, pp. 210–220, Sep. 2006. Accessed: Apr. 8, 2024. [Online]. Available: https://nopr.niscpr.res.in/handle/123456789/1518

[15] K. Patil and M. C. Deo, "Basin-scale prediction of sea surface temperature with artificial neural networks," *J. Atmos. Ocean. Technol.*, vol. 35, no. 7, pp. 1441–1455, Jul. 2018.

[16] S. G. Aparna, S. D'Souza, and N. B. Arjun, "Prediction of daily sea surface temperature using artificial neural networks," *Int. J. Remote Sens.*, vol. 39, no. 12, pp. 4214–4231, Jun. 2018, doi: 10.1080/01431161.2018.1454623.

[17] I. D. Lins, M. Araujo, M. D. C. Moura, M. A. Silva, and E. L. Droguett, "Prediction of sea surface temperature in the tropical Atlantic by support vector machines," *Comput. Statist. Data Anal.*, vol. 61, pp. 187–198, May 2013.

[18] A. Wu, W. W. Hsieh, and B. Tang, "Neural network forecasts of the tropical Pacific sea surface temperatures," *Neural Netw.*, vol. 19, no. 2, pp. 145–154, Mar. 2006.

[19] K. Patil, M. C. Deo, and M. Ravichandran, "Prediction of sea surface temperature by combining numerical and neural techniques," *J. Atmos. Ocean. Technol.*, vol. 33, no. 8, pp. 1715–1726, Aug. 2016.

[20] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1745–1749, Oct. 2017.

[21] H.-M. Choi, M.-K. Kim, and H. Yang, "Abnormally high water temperature prediction using LSTM deep learning model," *J. Intell. Fuzzy Syst.*, vol. 40, no. 4, pp. 8013–8020, Apr. 2021.

[22] X. Jia, Q. Ji, L. Han, Y. Liu, G. Han, and X. Lin, "Prediction of sea surface temperature in the East China Sea based on LSTM neural network," *Remote Sens.*, vol. 14, no. 14, p. 3300, Jul. 2022.

[23] M. Kim, H. Yang, and J. Kim, "Sea surface temperature and high water temperature occurrence prediction using a long short-term memory model," *Remote Sens.*, vol. 12, no. 21, p. 3654, Nov. 2020.

[24] L. Wei, L. Guan, L. Qu, and D. Guo, "Prediction of sea surface temperature in the China seas based on long short-term memory neural networks," *Remote Sens.*, vol. 12, no. 17, p. 2697, Aug. 2020.

[25] Z. Zhang, X. Pan, T. Jiang, B. Sui, C. Liu, and W. Sun, "Monthly and quarterly sea surface temperature prediction based on gated recurrent unit neural network," *J. Mar. Sci. Eng.*, vol. 8, no. 4, p. 249, Apr. 2020.

[26] J. Xie, J. Zhang, J. Yu, and L. Xu, "An adaptive scale sea surface temperature predicting method based on deep learning with attention mechanism," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 740–744, May 2020.

[27] C. Xiao, N. Chen, C. Hu, K. Wang, J. Gong, and Z. Chen, "Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111358.

[28] M. Jahanbakht, W. Xiang, and M. R. Azghadi, "Sea surface temperature forecasting with ensemble of stacked deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[29] S. Kartal, "Assessment of the spatiotemporal prediction capabilities of machine learning algorithms on sea surface temperature data: A comprehensive study," *Eng. Appl. Artif. Intell.*, vol. 118, Feb. 2023, Art. no. 105675.

[30] T. Xu, Z. Zhou, Y. Li, C. Wang, Y. Liu, and T. Rong, "Short-term prediction of global sea surface temperature using deep learning networks," *J. Mar. Sci. Eng.*, vol. 11, no. 7, p. 1352, Jul. 2023.

[31] X. Yu, S. Shi, L. Xu, Y. Liu, Q. Miao, and M. Sun, "A novel method for sea surface temperature prediction based on deep learning," *Math. Problems Eng.*, vol. 2020, pp. 1–9, May 2020.

[32] P. P. Sarkar, P. Janardhan, and P. Roy, "Prediction of sea surface temperatures using deep learning neural networks," *Social Netw. Appl. Sci.*, vol. 2, no. 8, p. 1458, Aug. 2020, doi: 10.1007/s42452-020-03239-3.
[33] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, "A CFCC-LSTM model for sea surface temperature prediction," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 207–211, Feb. 2018.
[34] S. Hou, W. Li, T. Liu, S. Zhou, J. Guan, R. Qin, and Z. Wang, "MIMO: A unified spatio-temporal model for multi-scale sea surface temperature prediction," *Remote Sens.*, vol. 14, no. 10, p. 2371, May 2022.
[35] M. M. Hittawe, S. Langodan, O. Beya, I. Hoteit, and O. Knio, "Efficient SST prediction in the Red Sea using hybrid deep learning-based approach," in *Proc. IEEE 20th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2022, pp. 107–117. Accessed: Apr. 8, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9976090/
[36] N. B. Avsar, S. Jin, and S. H. Kutoglu, "Interannual variations of sea surface temperature in the black sea," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 5617–5620. Accessed: Apr. 8, 2024.
[37] S. Wolff, F. O'Donncha, and B. Chen, "Statistical and machine learning ensemble modelling to forecast sea surface temperature," *J. Mar. Syst.*, vol. 208, Aug. 2020, Art. no. 103347.
[38] G. Bonino, G. Galimberti, S. Masina, R. McAdam, and E. Clementi, "Machine learning methods to predict sea surface temperature and marine heatwave occurrence: A case study of the Mediterranean Sea," *Ocean Sci.*, vol. 20, pp. 417–432, 2024.
[39] A. Y. Winona and D. Adytia, "Short term forecasting of sea level by using LSTM with limited historical data," in *Proc. Int. Conf. Data Sci. Appl. (ICoDSA)*, Aug. 2020, pp. 1–5. Accessed: Apr. 8, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9213025/
[40] A. Saxena, R. Chauhan, D. Chauhan, and G. Aggarwal, "Time series prediction of sea surface temperature using LSTM," in *Proc. 13th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2023, pp. 458–461. Accessed: Apr. 8, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10048855/
[41] S. Hou, W. Li, T. Liu, S. Zhou, J. Guan, R. Qin, and Z. Wang, "D2CL: A dense dilated convolutional LSTM model for sea surface temperature prediction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12514–12523, 2021.
[42] K. R. Patil and M. Iiyama, "Deep learning models to predict sea surface temperature in Tohoku region," *IEEE Access*, vol. 10, pp. 40410–40418, 2022.
[43] H. Haak et al., "MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 FAFMIP faf-all," World Data Centre Climate(WDCC) DKRZ, Germany, 2023. [Online]. Available: https://hdl.handle.net/21.14106/8183142d045807c448cce5c4c2567402b6bd5bf3
[44] U. Schulzweida, L. Kronblueh, and R. G. Budich. (2019). *CDO: Climate Data Operators*. Accessed: Apr. 8, 2024. [Online]. Available: https://earth.bsc.es/gitlab/ces/cdo/-/raw/6f73055c48a240a9e5e5b318512ff3bc5ea6b3a1/doc/cdoBrochureWeb.pdf
[45] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 636–644. Accessed: Apr. 8, 2024. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2017/html/Yu_Dilated_Residual_Networks_CVPR_2017_paper.html
[46] Y. Kudo and Y. Aoki, "Dilated convolutions for image classification and object localization," in *Proc. 15th IAPR Int. Conf. Mach. Vis. Appl. (MVA)*, May 2017, pp. 452–455. Accessed: Apr. 8, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7986898/
[47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham, Switzerland: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[48] J. G. Fernández, I. A. Abdellaoui, and S. Mehrkanoon, "Deep coastal sea elements forecasting using UNet-based models," *Knowl.-Based Syst.*, vol. 252, Sep. 2022, Art. no. 109445.
[49] Y. Ren and X. Li, "Predicting the daily sea ice concentration on a subseasonal scale of the Pan-Arctic during the melting season by a deep learning model," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4301315. Accessed: Apr. 8, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10131954/
[50] B. Xie, J. Qi, S. Yang, G. Sun, Z. Feng, B. Yin, and W. Wang, "Sea surface temperature and marine heat wave predictions in the South China Sea: A 3D U-Net deep learning model integrating multi-source data," *Atmosphere*, vol. 15, no. 1, p. 86, Jan. 2024.
[51] P. Rozas Larraondo, L. J. Renzullo, I. Inza, and J. A. Lozano, "A data-driven approach to precipitation parameterizations using convolutional encoder–decoder neural networks," 2019, *arXiv:1903.10274*.
[52] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.
[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[54] Y. Rubner, "The Earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000, doi: 10.1023/a:1026543900054.
[55] F. Yang and Z. Wu, "The phase change in the annual cycle of sea surface temperature," *NPJ Climate Atmos. Sci.*, vol. 7, no. 1, pp. 1–8, Feb. 2024.
[56] H. Lee, K. Lee, J. H. Kim, Y. Na, J. Park, J. P. Choi, and J. Y. Hwang, "Local similarity Siamese network for urban land change detection on remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4139–4149, 2021, doi: 10.1109/JSTARS.2021.3069242.
[57] G. Wang, K. Sun, and L. Tang, "No-reference DIBR-synthesized video quality assessment based on spatio-temporal texture inconsistency measurement," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Nov. 2022, pp. 1–4, doi: 10.1109/ISPACS57703.2022.10082823.
[58] F. Hussain, R.-S. Wu, and J.-X. Wang, "Comparative study of very short-term flood forecasting using physics-based numerical model and data-driven prediction model," *Natural Hazards*, vol. 107, no. 1, pp. 249–284, May 2021.
[59] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and F. Prabhat, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.
[60] A. P. Joshi, V. Kumar, and H. V. Warrior, "Modeling the sea-surface pCO_2 of the central Bay of Bengal region using machine learning algorithms," *Ocean Model.*, vol. 178, Oct. 2022, Art. no. 102094.
[61] S. Ali, Y. Huang, and J. Wang, "AI for sea ice forecasting," in *Artificial Intelligence in Earth Science*. Amsterdam, The Netherlands: Elsevier, 2023, pp. 41–58.

MANVENDRA JANMAIJAYA (Member, IEEE) received the B.Sc. degree (Hons.) in computer science from the University of Delhi, India, in 2012, the M.Sc. degree in computer science from South Asian University (SAU), New Delhi, India, in 2014, and the Master of Technology degree in computer science engineering from Gautam Buddha University, Greater Noida, India, in 2016. He is currently pursuing the Ph.D. degree with the Department of Computer Science, South Asian University. His current research interests include neural networks, machine learning, and scientometrics.

AMIT RAUNIYAR (Member, IEEE) received the bachelor's degree in computer applications from HNB Garhwal University, in 2012, and the M.Sc. and Ph.D. degrees in computer science from South Asian University (SAU), in 2015. He is currently a Postdoctoral Research Fellow with the School of Computer Science, University College Dublin. His research interests include evolutionary algorithms, multi-factorial optimization, real-time, and multi-robots systems. He has been awarded Visvesvaraya Ph.D. (Ministry of Electronics & IT, Government of India) Fellow. He is an active member of the IEEE Computational Intelligence Society, Robotics and Automation Society, and IEEE Systems, Man and Cybernetics Society.

AMIT K. SHUKLA (Senior Member, IEEE) received the master's (Hons.) and Ph.D. degrees in computer science from South Asian University, New Delhi, India. He is currently an Associate Professor with the School of Technology and Innovations, University of Vaasa, Finland. His research interests include fuzzy sets and systems, anomaly detection, industry 4.0 systems, transfer learning, and evolutionary optimization. He was a recipient of the prestigious INSPIRE Fellowship from the Department of Science and Technology, Government of India, during his Ph.D. tenure. He is also serving as an Associate Editor for *Heliyon* (Elsevier), along with actively reviewing publications from various journals, such as IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, INFORMATION SCIENCES, *Applied Soft Computing*, *Engineering Applications of Artificial Intelligence*, and *Neural Computing and Applications*.

SANDEEP KUMAR received the master's (Hons.) and Ph.D. degrees from South Asian University, New Delhi, India. His research interests include transfer learning, machine learning, and intrusion detection. He is an active member of IEEE Computational Intelligence Society. He is also an active reviewer in various journals, such as *Heliyon*, *Applied Soft Computing*, and *Engineering Applications of Artificial Intelligence*. He received the prestigious INSPIRE fellowship from the Department of Science and Technology, Government of India, during the Ph.D. degree.

REM COLLIER (Member, IEEE) is currently an Associate Professor with the UCD School of Computer Science, Dublin, Ireland, where he has been a Faculty Member, since 2004. His primary research interests include multi-agent systems and in particular agent-oriented software engineering/programming. From 1996 to 2010, he developed the Agent Factory Framework, a cohesive framework for the rapid prototyping of multi-agent systems (MAS). During that time, he was involved in a range of projects exploring its application to social robotics, virtual reality, and mobile computing. He was a Postdoctoral Researcher on the team that won the prestigious 2003 CIA Systems Innovation Award for work on the Agents Channelling ContExt Sensitive Services (ACCESS) Architecture. Since 2017, his research interests has centered around the areas of MAS, microservices, REST, and linked data, where he has proposed the multi-agent microservices (MAMS) architectural style as a novel approach to integrating agent technologies into microservice-based software architectures. He has been the Primary Supervisor of 16 Ph.D. and 12 M.Sc. theses and published over 130 academic articles. He is currently the Lead-PI of CONSUS, a €17.5M nationally funded Strategic Partnership Project focused on Digital Precision Agriculture & Crop Science and is a Principal Investigator on the €9M CAMEO Project and a Disruptive Technologies Innovation Project charged with developing Ireland Future Earth Observation Platform.

PRANAB K. MUHURI (Senior Member, IEEE) was born in Chittagong, Bangladesh. He received the Ph.D. degree in computer engineering from IT-BHU [currently, Indian Institute of Technology (BHU)], Varanasi, India, in 2005. He is currently a Professor with the Department of Computer Science, South Asian University, New Delhi, India, where he is leading the Computational Intelligence Research Group. He has published more than 100 papers in reputed journals and conferences, including IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Reliability Engineering & Systems Safety*, *Fuzzy Sets and Systems*, *Applied Soft Computing*, *Computers & Industrial Engineering*, and *Future Generation Computer Systems*. His current research interests include real-time systems, fuzzy systems, evolutionary algorithms, perceptual computing, and machine learning. He has been serving as an Editorial Board Member in journals, such as *Applied Soft Computing* and *Engineering Applications of Artificial Intelligence*.