Deep learning models to predict sea surface temperature in Tohoku region

Kalpesh Patil¹, Masaaki Iiyama²

¹Application Laboratory, VAiG, JAMSTEC, Yokohama, Kanagawa, 236-0001 Japan
²Faculty of Data science, Shiga University, Hikone, Shiga, 522-8522 Japan

Corresponding author: First A. Author (e-mail: patilkalpeshr@gmail.com).

“This work was supported by JST CREST Grant Number JPMJCR19F1, Japan.”

ABSTRACT Prediction of sea surface temperature (SST) is a very challenging task, especially for the regions of high SST variability. Such predictions are either achieved by physics-based models, which often yield poor predictions and computationally intensive, or using data-driven methods, which are skillful and computationally less intensive. However, recent past machine learning studies exploring SST prediction do not include important meteorological parameters governing SST variability. Therefore, in this study we propose various of deep learning (DL) models trained using past meteorological features to predict a day ahead SST. Proposed models include a deep multilayer perceptron (deep MLP), Long short-term memory networks (LSTM) and spatial 2-dimensional convolutional neural network (spatial 2D CNN). We explore the potential of the proposed DL models at a day ahead SST predictions across different locations in Tohoku region (Japan’s east coast) including the in-situ validation. Evaluation of these DL models’ prediction skills suggests that the spatial 2D CNNs are highly skillful at coastal locations, whereas at offshore location's equal predictions skills were noted from deep MLP and LSTM. We further attempted to improve the spatial 2D CNN by including past SST features and such improvisation shown very low errors ranging from 0.35°C to 0.75°C and high correlation skill from 0.64 to 0.96. These improved skills were also compared with persistent model (PM) skills using RMSE and Correlation (RC) phase diagram, where we found that improved skills are consistently better than PM skills. Apart from that, we also extracted features from spatial 2D CNN to understand the reason underlying for such improved skills, and we noted that proposed DL model successfully captures the major meteorological and oceanic feature governing SST variability. This led us to conclude that, proposed DL models are capable of producing highly reliable SST predictions and may be equally applicable for other study regions.

INDEX TERMS Deep multilayer perceptron (Deep MLP), spatial 2-dimensional convolutional neural network (spatial 2D CNN), Long short-term memory networks (LSTM), deep learning for SST prediction, Meteorologic parameter forced SST prediction

I. INTRODUCTION

Sea surface temperature (SST) is a crucial factor responsible for air-sea flux interactions between oceans and atmosphere. SST variability strongly controls the local climate (precipitation changes, marine heatwaves) and marine ecosystem (potential fishing zones, eddies); thus, its prediction will aid us to understand the probable effects due to its variability in advance. Methods applied for prediction of SST can be broadly classified into two major categories. First, is using the solution of physics-based energy, momentum, and flux equations, and second is, using data-driven techniques. Major drawbacks of physics-based methods are, the complex assumptions involved in formulating governing equations, the need for accurate forcings, and high computational dependencies [1-2]. Apart from that, physics-based methods’ accuracy largely varies over the big spatial domain and may yield poor predictions [1]. On the other hand, data-driven techniques are recently becoming very popular as an alternative in the field of SST prediction due to their less assumptive, low computational nature and higher predictive skills.

Several past studies were reported elaborating the use of simple multi-layered perceptron (MLP) models for SST prediction at various parts in the global ocean. This
includes multi-linear regression at a few locations over Indian Ocean [3], MLP at Indian Ocean [3-4] and tropical Pacific Ocean [5], support vector machines/regressions (SVM/R) at northeast Pacific [6] and tropical Atlantic Ocean [7]. Such past studies considered features only from past SST at a target grid, assuming that future SST values can be predicted from its own past. Very few past studies had experimented cause-effect approach in SST prediction; this includes MLP at the western Mediterranean Sea [8] and random forest (RF) at selected locations over the global Ocean, including the eastern coast of Japan [9]. Such cause-effect approaches were implemented with simple machine learning models and thus exhibited high errors due to non-inclusion of features from nearby grids around the target location. Aparna et al. did attempt to include the features from surrounding grids, but only from past SST. Since the meteorological parameters were missing from the features; such study also exhibited high errors [10]. In recent past, many attempts using DL models can be noted for SST predictions. This includes LSTM network at Bohai Sea [11], Yellow Sea [12] and Korea Sea [13], gated recurrent unit (GRU) at Bohai Sea [14-15] and Yellow Sea [16], CNN at Pacific Ocean [15][17].

Since such various past attempts were formulated either by simple machine learning models or lacking causal parameters in features. Therefore, high errors were noted for SST predictions in those attempts, further it raises question on the applicability of such methods for longer leadtimes. Apart from that, prediction errors were observed to be biased towards warm SST in many these past studies [14-15][17].

Therefore, to address this significant research gap, we propose DL models, which considers both meteorological and oceanic features to predict SST at longer leadtimes at one of the high SST variability regions in the north-west Pacific Ocean near Japan’s east coast, ‘Tohoku’.

The rest of the paper is arranged in five sections. Section II highlights the various study locations inside Tohoku region, Section III describes datasets used for training and validation of DL models, Section IV discusses the formulation of proposed DL models and their parameter optimization, Section V elaborates results from proposed DL models and further their improvisation. In Section VI, brief discussion regarding the extracted features from CNN, error distribution and validation with in-situ SST is presented and, in last Section VII, conclusions and future work of the proposed study are mentioned.

II. STUDY LOCATIONS

This study is focused on several locations in Tohoku region (38°N - 40°N, 141°E-144°E, Fig. 1), which hosts a major mixture of Oyashio and Kuroshio currents and thus exhibits high spatio-temporal SST variability [18]. Therefore, it is a highly suitable region for evaluating the efficacy of the proposed DL models for the predictability of SST. Apart from that, the Tohoku region hosts most habitable region for major fish species. Hence SST predictions with lesser computational complexity are highly appreciated in Tohoku region. Various study locations were considered in coastal waters and offshore region. Also, these study locations are at the fronts of major spatio-temporal SST variability (right panel of Fig. 1). We also developed the proposed DL model at Kuroshio Extension Observatory (KEO) buoy location (32.3°N, 144.6°E) to validate them against in-situ SST. The in-situ SST data at KEO buoy is available through[1].

III. DATA

Two different datasets were used for the development of DL models. The first is training data mainly utilized as features to DL models, and the other is target data used for supervised training of the DL models.

A. TRAINING DATA

Training data is obtained from reanalysis data, an amalgamation from various sources refined by an optimal interpolation method; thus, reanalysis data are very close to the observed data. This study makes use of European Centre for Medium Range Weather Forecast’s (ECMWF) fifth-generation atmospheric reanalysis products (ERA5) of global climate [19] as features for training DL models.

ERA5 data provides several meteorological parameters, out of which important ones are extracted over the study locations. These includes 2m-air temperature, 2m-dew point temperature, solar radiation, total cloud cover, surface sensible heat flux, and wind speeds (meridional and zonal). These meteorological variables are essential to account for the rapid variability in the SST, e.g., 2m air-temp and dew point temp shows strong coupling with SST [20], solar radiation and sensible heat flux governs the major heat source [21-22], wind speed and total cloud cover governs the cooling [22], without considering any these variables could lead to the poor skills [8-9]. ERA5 data is accurate enough over the study area to develop the DL models [23]. Spatial resolution of ERA5 data is 0.25° x 0.25°, and it is available at an hourly timescale. ERA5 data was

*1: https://www.pmel.noaa.gov/ocs/data/disdel/

FIGURE. 1: Various study locations in Tohoku region. Loc 1, 4, and 10 are in coastal region, whereas rest of them are in offshore region. Loc 1-23 the DL models were validated against target SST, whereas at KEO buoy the validation was performed against in-situ SST.
downloaded using the ECMWF Climate Data Store API (CDSAPI) [24]. Meteorological variables were extracted on the grids closest to the study locations, thus the training of DL models is performed on 0.25° x 0.25° to account for the wider spatial domain.

B. TARGET DATA

Target SST choice is very crucial in the development of DL models. Target SST should be of very high resolution in spatial and temporal scale; apart from that, it should be very close to the observations. With such criteria, the choice of target data was limited over the study area. The satellite data from Himawari was a possible choice, but it had major temporal and spatial gaps as the product was available for L3 level. Another choice was the output of a numerical model from Japan Coastal Ocean Predictability Experiment ver2 (JCOPE2) targeted over the entire North-western Pacific, including the proposed study area. JCOPE2 is an operational nowcast/forecast system that assimilates the Himawari-8 satellite SST data into JCOPE using a three-dimensional variational scheme [25]. The resolution of the target data is 0.02° x 0.03° on a spatial scale, and it is available at an hourly timescale. JCOPE2 data is highly in agreement with ERA-5 reanalysis SST data at selected study locations and therefore meets the required criteria as a target SST [25]. Although ERA-5 SST data was available due to its low-resolution, JCOPE2 SST data is preferred as target data.

IV. PROPOSED DEEP LEARNING MODELS

Proposed DL models were developed at various locations shown in Fig. 1 over the Tohoku region. Features for DL models were governed from the past instances of ERA5 reanalysis data from 2m-air temperature, 2m-dew point temperature, solar radiation, total cloud cover, surface sensible heat flux and wind speeds (meridional and zonal) and ocean currents from JCOPE2. Features were moving averaged with a window of 2 to 4 hrs to reduce the noise in features. Training features were extracted at grid closest to the study locations and surrounding to it, whereas and target SST were interpolated at study locations using bi-linear interpolation. The duration of the data for training and validation is solely based on the availability of target SST. The total duration of the target SST was available from 2018-08-01 to 2018-12-31 at hourly timescale. In total, 3672 hours (153 days) were available, out of which 75% of data was used for training the DL models, 5% was used for validation of DL model parameters, and the remaining 20% was used for testing; and we note that there is no overlap between training, validation, and testing dataset. The testing set has an approximately 730 hours (30 days) which is sufficient to consider the diurnal variability in SST [9][15]. DL models were formulated based on the number of grids considered for features (refer to Fig. 2 for details). Deep MLP and LSTM, considers features only from target grid whereas spatial 2D CNN collect features from neighboring grids surrounding the target grid. Past hourly instances of meteorological variables, 2m-air temperature, 2m-dew point temperature, solar radiation, total cloud cover, surface sensible heat flux and wind speeds (meridional and zonal), and ocean currents constitutes the features to the DL models. Only SST values alone may not be sufficient to capture high SST variability as it is governed more by other causal variables [2][8-9]. Proposed DL models were targeted for SST prediction at leadtime of 24-hours. The details of individual models are explained in further sections.

FIGURE 2: Input area/grid used for formulation of DL models using various meteorological parameters and ocean currents.

A. DEEP MLP

Deep MLP is designed to receive features from a single target grid, as shown in Fig 2. Various meteorological features (2m-air temperature, 2m-dew point temperature, solar radiation, total cloud cover, surface sensible heat flux and wind speeds) and ocean currents at target grid surrounding the study location were extracted as features to DL models. These features were moving averaged to reduce the noise in them, and target SST data of future time step was considered for training. All features were converted in a single column vector (as mentioned in Fig. 3) to form an Input layer. This input layer is further connected to several deep hidden layers.

FIGURE 3: Deep MLP architecture proposed based on meteorological parameters and ocean currents as features for SST prediction. Features were considered from single target grid and stacked together in a single column vector.
Each deep hidden layer consists of several hidden neurons, dropout ratio, regularization, and batch normalization layers. Drop-out and regularization layers were added to avoid over-fitting during training. The final deep hidden layer is connected to an output layer which consists of a single neuron. The output layer compares the predicted SST from deep MLP with target SST and further adjusts the internal parameters with respect to observed error. The training procedure of deep MLP is repeated for several iterations until the desired error is achieved. For each iteration, different hyper-parameters were tried to gain the generalization ability.

B. SPATIAL 2D CNN

The spatial 2D CNN model has a major difference in feature gathering. Spatial 2D CNN gather features from the closest grid to study location as well as from its neighbors. The features of spatial 2D CNN are a three-dimensional matrix with the size of ‘latitudes longitude times past_hourly timesteps’ of 2m-air temperature, 2m-dew point temperature, solar radiation, total cloud cover, surface sensible heat flux, and wind speeds (meridional and zonal) and ocean currents. Spatial 2D CNN is thus supposed to extract the important features from causal variables governing SST variability. A three-dimensional feature matrix forms an input layer for the spatial 2D CNN model, which is further connected to several convolutional hidden layers. Each hidden convolutional layer tries to extract the features from the previous convolutional layer. Each convolutional layer is further connected to the dropout layer with regularization to avoid over-fitting. After several convolutional layers, feature maps were extracted in a single column vector, this vector is further connected to several hidden layers as in deep MLP. Internal parameters of spatial 2D CNN were adjusted with respect to the error between spatial 2D CNN output and target data. Fig. 4 elaborates the architecture of the proposed spatial 2D CNN model.

C. LONG SHORT-TERM MEMORY NETWORKS

We also developed long short-term memory (LSTM) networks which is a recent form of recurrent neural networks [26]. LSTM are the improved recurrent neural networks enhanced by additional components of cell state, forget gate, input gate and output gate (Fig. 5). We tested the LSTM model prediction skills at each location from the past meteorological features. Similar measures as seen deep MLP case, were taken to avoid over-fitting, by adding the drop out, batch-normalization layers and weight regularization.

D. DL MODEL PARAMETER OPTIMIZATION

There are two types of parameters in DL models to be optimized. First are internal parameters which can be adjusted with respect to the observed error between DL model output and target SST. To adjust the internal parameters of the DL model, a mean square error based on adam optimizer is considered. Second is, external or popularly known as hyper-parameters which cannot be adjusted by observed error, but rather fixed by trial-and-error approach or with the help of some sophisticated optimization algorithm. In the current study, the popular but computationally less intensive and equally effective random search algorithm is used to optimize the hyper-parameters, as the performance of the DL model is very sensitive towards them. Various hyper-parameters include past timesteps of causal features (2 to 16 hrs), moving average window for features (2 to 4 hrs), number of hidden/convolutional/LSTM layers (2 to 5), number units in each hidden/convolutional/LSTM layers (50 to 300), number of neighboring grids in spatial 2D CNN (2 to 13), filter size in spatial 2D CNN (1 to 3), dropout ratio (0.15 to 0.4) and L2 regularization (1e-5 to 1e-2).

V. RESULTS

Results from various proposed DL model experiments are discussed in this section during testing period of 730 hours (30 days) at a leadtime of 24-hours. The performance of the DL model was evaluated using root mean square error (RMSE) and correlation skill (corr. skill) calculated between DL model output and target SST or in-situ SST. Lower RMSE and higher corr. skills values are better and vice-versa.

A. Comparison of proposed DL models

To understand the effectiveness of gathering the features from neighboring grids, the performance of spatial 2D CNN was compared with deep MLP and Long-short term memory (LSTM) networks at each study location. Fig. 6 shows such comparison of DL model predictions against target SST at 24-hour leadtime.
The errors in SST predictions at 24-hour leadtime with deep MLP were observed to be between 1.0°C to 2.5°C, and the same from the LSTM models were found between 1.0°C to 2.25°C. The poor skills observed in deep MLP are due to few features and those from LSTM models are due to few parameters and inability to extract useful pattern from short length of past variables compared to spatial 2D. Such poor skills of LSTM in SST predictions were also noted at few locations in Japan’s east coast [9]. The above-mentioned errors are quite high due to the non-inclusion of features from the neighboring grid, but comparable to some of the past studies which considered similar approach as noted in [3-4][9]. However, errors from spatial 2D CNN models were limited to 1.0°C to 1.5°C, such higher skill was due to the ability of spatial 2D CNN model’s in handling the spatio-temporal features from neighboring grids [17-18]. From Fig. 6, it is evident that spatial 2D CNN is experiencing lesser errors than deep MLP and LSTM at each location. The errors, at the offshore locations (Loc 17, 23, and 14), were marginally lesser (except Loc 2) and significantly lesser at coastal locations (Loc 4, 10, 1). The lower RMSE at coastal locations was due the inclusion of past spatio-temporal causal features from neighboring grids (especially from wind speed and ocean current), which are largely responsible for high variability in SST.

Even though the errors in spatial 2D CNN were lesser compared to deep MLP and LSTM, the range of 1°C to 1.5°C was still higher considering the usability of such predictions for various applications such as fisheries and marine-ecosystem management. Therefore, we further attempt to improvise the spatial 2D CNN models to reduce these errors.

### B. IMPROVISED SPATIAL 2D CNN

To improvise the earlier seen skillful spatial 2D CNN’s, we propose to add past SST values as an additional feature from neighboring grids. Thus, the improvised spatial 2D CNN not only takes features from important causal parameters (meteorological and ocean currents, as mentioned data section) but also from past SST values and overcomes the crucial shortcomings of several past studies.

The performance of improvised spatial 2D CNN is depicted in Fig. 7 and Fig. 8 at each location, using a time-series plot of target SST against the predictions from improved spatial 2D CNN at 24-hour leadtime. Improvised spatial 2D CNN substantially improve the errors in SST prediction than its earlier range (Fig. 6). The improvised spatial 2D CNN model now observes RMSE in the range of 0.35°C to 0.75°C, and correlation skills from 0.64 to 0.96. Such skills are well within the acceptable limits of usability considering from the perspective of various applications, such as locating potential fishing zones, eddy detection and cyclone progression [1-2][5][17] and such skills were further found to be significantly better in comparison with similar past works where RMSE values were observed to be higher than 1.0°C [9-10][17].

![FIGURE 6: RMSE comparison between deep MLP, LSTM and spatial 2D CNN based on RDL models output and target SST. The location numbers in this figure correspond to those in Fig. 1.](image)

![FIGURE 7: Improvised spatial 2D CNN SST prediction comparison against target SST for 24-hour leadtime at Loc 1, 2 and 4. Corresponding correlation skill and RMSE values are mentioned for each location. Improvised models output captures SST variability very well with no bias towards high SST.](image)
do capture the rapid changes effectively and accurately. This suggests that improvised spatial 2D CNN models are highly capable of providing skillful long leadtime SST predictions.

VI. DISCUSSION

In the proposed study, long lead time SST predictions were attempted using various DL models. Several past studies attempted SST prediction based only on a time-series approach [3-4, 15]. The major limitation of such approach is the inability to capture the rapid changes in the SST variability, especially at coastal locations. Therefore, to overcome these limitations, proposed study combines the spatio-temporal features from past meteorological along with the oceanic parameters. Various DL models were developed for this purpose, in which deep MLP, LSTM considered features from single target grid and spatial 2D CNN models gathered the features from neighboring grids. We further analyze the skills from the proposed models and discuss their validation with in-situ SST at KEO buoy, comparison with persistent models (PM), extraction of trained feature maps from first convolution layer and error distribution at each location in further sub sections.

A. VALIDATION OF IMPROVED MODELS WITH IN-SITU SST

The improvised spatial 2D CNN model was also developed at KEO buoy location to validate its predictions against the in-situ SST data. The improvised spatial 2D CNN at KEO buoy was trained from past meteorological and JCOPE2 SST features, whereas in-situ SST was kept aside for validation purpose. In this validation, we found that RMSE from improvised models was very low around 0.26°C, and high correlation skill of 0.97 (Fig. 9). Such superior skills strongly suggest that improvised spatial 2D CNN models can capture the ocean dynamics at various locations.

B. IMPROVED SKILLS IN SPATIAL 2D CNN AND COMPARISON WITH PM

In comparison to deep MLP and LSTM, spatial 2D CNN models outperformed significantly at coastal locations and marginally at offshore locations (Fig 6). Even though such errors were comparable with earlier studies, this margin was high in consideration of the applications of such SST predictions. To further enhance the performance of the proposed DL models, past spatio-temporal SST values were added along with meteorological features and ocean currents. A significant improvement in the RMSE was noted after addition of past spatio-temporal SST values as additional features. The errors from improved spatial 2D CNNs were reduced from a range of 1.0°C-1.5°C to 0.35°C-0.75°C. To understand the comparison of predictive skills from improvised spatial 2D CNN against PM skills, we introduce a RMSE vs corr. skill based two-dimensional space diagram (RC). In RC diagram the skills were plotted at each location. Along x-axis higher the correlation, point moves towards right, similarly along y-axis lower the RMSE, point move towards bottom. So, in general, the improvement in RC diagram can be noted with points moving towards bottom-right corner. In Fig. 10, RC diagram between improvised model and PM skills is depicted, the movement of skills from the improvised models in the bottom-right corner suggest the superior skills of improvised spatial 2D-CNN’s than PM except at Loc 2.
FIGURE 10: Comparison of improvised spatial 2D CNN with PM skills at each location. The movement of points towards the bottom right corner denotes the skill improvement. Improvised models show significantly better skills than PM at each location except at Loc 2.

C. FEATURE MAPS FROM IMPROVED SPATIAL 2D CNN
Feature maps from CNNs are an essential tool used to analyze the connection between the training features and target data (Fig. 4). These feature maps are the outputs of intermediate convolutional layers. It helps to understand how the training features are transformed in convolution layers to output the desired predictions, in this case hourly SST at a particular location. For this purpose, we analyzed the feature maps for Loc 23 and compared them with the training features.

FIGURE 11a: Training features from past four hours surrounding Loc 23 for 108th data-point during testing period (Ref. Fig. 8). Center pixel in the above figure corresponds to the Loc 23, and the x-axis and y-axis represents the latitudes and longitudes, resp. Training features are normalized within -1 to +1 range, and darker shades represent strong features.

In Fig. 11a, the training features form past meteorological and oceanic features used for prediction of 108th SST data-point are shown during the testing period. The extracted trained features from the first convolutional layer are shown in Fig. 11b. Different training features were merged into a common or separate features. For example, meridional wind (v-wind) component (Fig. 11a) is captured in the filter number 2 in Fig. 11b. Similarly, zonal wind (u-wind) is seen to be captured in filter number 4 and 5, SST in filter number 23, meridional current (v-curr) in filter number 24 and dew point temperature in filter number 25. At Loc 23, where Oyashio cold currents and wind majorly govern the mixing (Fig. 1), and therefore among the various training features SST, zonal and meridional wind, and meridional current are found to be dominating over other. Thus, CNN is seen to be efficiently extracting the required spatial meteorological and oceanic features and merging them to pass on to the further convolutional layers.

FIGURE 11b: Feature maps corresponding to Fig. 11a from first convolutional layer for the improved spatial 2D CNN model for predicting 24-ahead SST at Loc23. Extracted feature maps resemble the important training features, e.g. Filt #2 – vWind, #4,5 – uWind, #23 – SST, #24 – vCurr and #25-DewPtTemp.

D. EFFECTIVENESS IN CAPTURING LOW SST
Sudden cooling of ocean water is noted at few study locations, e.g., Loc 1, 10, 14, and 17. This is due to the rapid mixing of currents and change in the bathymetry. Usually, physics-driven models and many past DL methods exhibit a high bias for cold SST and low bias for warm SST, because warm SST has more persistence than cold SST. The cold SST arises due to the sudden change in the meteorological and physical factors and hence it does not persist for long periods. It is highly beneficial to capture such rapid variability in SST, as it could govern the potential location of fishing zone or keep persisting the low-pressure zone during cyclones. Hence it is very important that prediction of such low SST should be as accurate as the total study period. With improvised spatial 2D CNN we note the similar accuracy in the low SST in comparison with the error observed during complete testing period (Fig. 7 and 8).
E. ERROR DISTRIBUTION

To understand the nature of bias in improvised spatial 2D CNN predictions, we further investigated the error distribution between target SST and improvised spatial 2D CNN predictions. In Fig. 12 the probability distribution function of error is shown for each location. It can be noted that, large number of errors do concentrate at the center within a range of -0.5°C to 0.5°C. This suggests, CNN predictions do not exhibit large errors and hence are more reliable.

![Figure 12: Probability distribution function of error at each location.](image)

VII. CONCLUSIONS

- Among the proposed deep learning models, Deep MLP and LSTM showed high error at coastal locations, whereas equal prediction skills from them were noted at offshore locations.
- Inclusion of more spatio-temporal features from surrounding grids reduced the errors in spatial 2D CNN significantly, especially at coastal locations.
- To further improvise the spatial 2D CNN, additional spatio-temporal past SST features were added, which helped them to surpass the PM prediction skills (Fig. 10).
- Various important features pertaining to inputs surrounding the study area were effectively noted by improved spatial 2D CNN as seen in Fig. 11a and Fig. 11b. This suggests the applicability of the proposed improvised models at other study areas will experience similar skills in the SST prediction exhibiting high variability in SST.
- Peculiar feature of improvised spatial 2D CNN is the non-biasedness of prediction skills towards the warm SST. Such important observation was found to be missing in past studies of SST prediction [14-15][17].
- The future work of the current study is to consider the higher-resolution input features and more oceanic parameters (e.g., sub-surface temp) at large spatial domain instead of few locations. Since the proposed method is highly adaptive and efficiently seen to capture the high SST variability, it is possible to cluster the similar variability SST locations and collectively develop DL model for each cluster. This will significantly reduce the computation cost during training when proposed DL models are considered to apply over wide spatial domain.
- The proposed framework when applied on spatial domain will be highly useful for estimating potential fishing zones [27], probable movements of the progressing cyclone [28], knowing expected heatwaves and detection of eddies [29].

REFERENCES

[1] R. F. Milliff, W. G. Large, J. Morzel, G. Danabasoglu and T. M. Chin, "Ocean general circulation model sensitivity to forcing from scatterometer winds," Journal of Geophysical Research: Oceans, vol. 104, no. 5, pp. 11337 - 11358, 1999.
[2] P. Chu, C. Fang and C. Kim, "Japan/East Sea model predictability," Continental Shelf Research, vol. 25, no. 17, pp. 2107-2121, 2005.
[3] 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 Journal of Marine Sciences, vol. 35, no. 3, pp. 210-220, 2006.
[4] K. Patil and M. C. Deo, "Basin-Scale Prediction of Sea Surface Temperature with Artificial Neural Networks," Journal of Atmospheric and Oceanic Technology, vol. 35, no. 7, pp. 1441-1455, 2018.
[5] A. Wu, W. W. Hsieh and B. Tang, "Neural network forecasts of the tropical Pacific sea surface temperatures," Neural Networks, vol. 19, no. 2, pp. 145-154, 2006.
[6] Q. Li, Y. Zhao, H. Liao and J. Li, "Effective forecast of Northeast Pacific sea surface temperature based on a complementary ensemble empirical mode decomposition–support vector machine method," Atmospheric and Oceanic Science Letters, vol. 10, no. 3, pp. 261-267, 2017.
[7] I. D. Lins, M. Araujao, M. das C. Moura, M. A. Silva, and E. L. Drogueit, “Prediction of sea surface temperature in the tropical Atlantic by support vector machines,” Computational Statistics and Data Analysis, vol. 61, no. May, pp. 187–198, 2013.
[8] E. Garcia-Gorriz and J. Garcia-Sanchez, “Prediction of sea surface temperatures in the western Mediterranean Sea by neural networks using satellite observations,” Geophysical Research Letters, vol. 34, no. 11, Article ID L11603, 6 pages, 2007.
[9] S. Wolf, F. O'Donncha, and B. Chen, “Statistical and machine learning ensemble modelling to forecast sea surface temperature,” Journal of Marine Systems, vol. 208, no. Aug, p. 103347, 2020.
[10] S. G. Aparna, S. D'Souza, and N. B. Arjun, "Prediction of daily sea surface temperature using artificial neural networks," International Journal of Remote Sensing, vol. 39, no. 12, pp. 4214–4231, Mar. 2018.
[11] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, “Prediction of sea surface temperature using long short-term memory,” IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, pp. 1745–1749, Oct. 2017.
[12] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, “A CFCC-LSTM Model for Sea Surface Temperature Prediction,” IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 2, pp. 207–211, Feb. 2018.
[13] M. Kim, H. Yang and J. Kim, “Sea Surface Temperature and High Water Temperature Occurrence Prediction Using a Long Short-Term Memory Model,” Remote sensing, vol. 12, no. 21, p. 3654, 2020.
[14] 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," Journal of Marine Science and Engineering, vol. 8, no. 4, p. 249, 2020.
[15] 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 Geoscience and Remote Sensing Letters, vol. 17, no. 5, pp. 740-744, 2019.
[16] 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,” Mathematical Problems in Engineering, vol. 2020, Article ID 6387173, p. 9, 2020.
[17] G. Zheng, X. Li, R. H. Zhang and B. Liu, "Purely satellite data-driven deep learning forecast of complicated tropical instability waves," Science advances, vol. 6, no. 29, p. eaba1482, 2020.

[18] T. Kono, M. Foreman, P. Chandler and M. Kashiwai, "Coastal Ocean Observations—A Japanese Experience of the Global Ocean Observing System," Journal of physical oceanography, vol. 34, no. 7, pp. 1477-1494, 2004.

[19] L. Bengtsson, "Medium-Range Forecasting—The Experience of ECMWF," Bulletin of the American Meteorological Society, vol. 66, no. 9, pp. 1133–1146, 1985.

[20] S. M. Jang, S. S. Kim, Y. C. Choi and S. G. Kim. “A study of correlations between air-temperature of Jeju and SST around Jeju Island”. Journal of the Korean Society for Marine Environment & Energy, vol. 9, no. 1, pp. 55-62, 2006.

[21] M. R. Kumar and A. A. Fernandes. “Variability of surface meteorological parameters over the Arabian sea”. Indian Journal of Marine Sciences, Vo. 23, Mar 1994, pp. 8-13.

[22] H. Qin, H. Kawamura, and Y. Kawai. “Detection of hot event in the equatorial Indo-Pacific warm pool using advanced satellite sea surface temperature, solar radiation, and wind speed”. Journal of Geophysical Research: Oceans, vol. 112, no. (C7), 2007.

[23] H. Hersbach et al, “Global reanalysis: goodbye ERA-Interim, hello ERA5,” ECMWF Newsletter, vol. 159, Meteorology, pp. 17-24, Apr-2019.

[24] ERA5: Fifth generation of ECMWF atmospheric reanalysis of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-re-levels?tab=overview), Accessed on 2019-Jun-20.

[25] Y. Miyazawa, et al, “Assimilation of high-resolution sea surface temperature data into an operational nowcast/forecast system around Japan using a multi-scale three-dimensional variational scheme,” Ocean Dynamics, vol. 67, no. 6, pp. 713–728, 2017.

[26] S. Hochreiter and J. Schmidhuber. “Long short-term memory”. Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[27] C. T. Tseng, C. L. Sun, I. M. Belkin, S. Z. Yeh, C. L. Kuo, and C. D. Liu, “Sea-surface temperature affect distribution of Pacific Saury (Cololabis saira) in the Northwestern Pacific Ocean”. Deep Sea Research Part II: Tropical Studies in Oceanography, vol. 107, pp. 15-21, 2014.

[28] D. Rai, S. Pattnaik, P. V. Rajesh and V. Hazra, “Impact of high resolution sea surface temperature on tropical cyclone characteristics over the Bay of Bengal using model simulations.” Meteorological Applications, vol. 26, no. 1, pp. 130-139, 2019.

[29] X. Sun, M. Zhang, J. Dong, R. Lguensat, Y. Yang, and X. Lu, "A Deep Framework for Eddy Detection and Tracking From Satellite Sea Surface Height Data." IEEE Transactions on Geoscience and Remote Sensing, 2020.

KALPESH R. PATIL received his B.Tech. degree in Civil Engineering from College of Engineering Pune, Maharashtra, India, in 2007; the M. Tech. degree in Water resources engineering from Indian Institute of Technology, Bombay, Maharashtra, India in 2012, and Ph.D. degree in Ocean Sciences from Indian Institute of Technology, Bombay, Maharashtra, India in 2019. Currently he is working as a young research fellow at Application laboratory, Japan Agency for Marine Earth-Science and Technology, Yokohama, Japan. From 2008 to 2010, he was a senior software developer for mobile application development at Mahindra Satyam (formerly Satyam Computer Scr. Ltd.), Bangalore, India. From 2013 to 2014 he was a research assistant with the Civil engineering department of Institute of Technology, Bombay, Maharashtra, India. His research interests include the analysis and predictability of important climatic and ocean index oscillations, development of artificial intelligence methods for analysis of triggering mechanism of cyclones, precipitation and various important weather phenomena, and prediction of ocean state parameters using state-of-the art deep learning methods.

MASAAKI IIYAMA Masaaki Iiyama received the B.S. degree in engineering informatics, and the M.S. and Ph.D degrees in informatics from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2006, respectively. Currently, he is a professor at Faculty of Data Science, Shiga University. His research interests include computer vision, pattern recognition, machine learning and remote sensing for fishery.