IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 14, 2022, accepted April 5, 2022, date of publication April 13, 2022, date of current version April 20, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3167176

Deep Learning Models to Predict Sea Surface Temperature in Tohoku Region

KALPESH R. PATIL[1] AND MASAAKI IIYAMA[2], (Member, IEEE)
[1]Application Laboratory, VAiG, JAMSTEC, Yokohama, Kanagawa 236-0001, Japan
[2]Faculty of Data Science, Shiga University, Hikone, Shiga 522-8522, Japan

Corresponding author: Kalpesh R. Patil (patilkalpeshr@gmail.com)

This work was supported by the Japan Science and Technology Agency (JST) CREST, Japan, under Grant JPMJCR19F1.

ABSTRACT The prediction of sea surface temperature (SST) is a challenging task, especially for regions with high SST variability. Such predictions are either achieved by physics-based models, which often yield poor predictions and are computationally intensive, or by using data-driven methods, which are skillful and computationally less intensive. However, recent machine learning studies exploring SST prediction have not included the 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 day-ahead SST. The proposed models include a deep multilayer perceptron (deep MLP), long short-term memory network (LSTM) and spatial 2-dimensional convolutional neural network (spatial 2D CNN). We explore the potential of the proposed DL models for day-ahead SST prediction across different locations in the Tohoku region (Japan's east coast), including in situ validation. Evaluation of these DL models' prediction skills suggests that the spatial 2D CNN's are highly skillful at coastal locations, whereas at offshore locations, equal prediction skills were noted from the deep MLP and LSTM. We further attempted to improve the spatial 2D CNN by including past SST features, and such improvisation showed 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. In addition, we extracted features from the spatial 2D CNN to understand the reason underlying such improved skills, and we noted that the proposed DL model successfully captured the major meteorological and oceanic features governing SST variability. This led us to conclude that the proposed DL models are capable of producing highly reliable SST predictions, and may be equally applicable to other study regions.

INDEX TERMS Deep multi-layer 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 the air-sea flux interactions between oceans and the atmosphere. SST variability strongly controls the local climate (precipitation changes and marine heatwaves) and marine ecosystems (potential fishing zones and eddies); thus, its prediction will aid in understanding the probable effects due to its variability in advance. Methods applied to the prediction of SST can be broadly classified into two major categories. First, the solution of physics-based energy, momentum, and flux equations was used, followed by data-driven techniques. The major drawbacks of physics-based methods are the complex assumptions involved in formulating the governing equations, the need for accurate forcing, and high computational dependencies [1], [2]. In addition, the accuracy of physics-based methods largely varies over a large 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 previous studies have reported the use of simple multi-layered perceptron (MLP) models for SST prediction in various parts of the global ocean. This includes multi-linear

The associate editor coordinating the review of this manuscript and approving it for publication was Huiyu Zhou.

regression at a few locations over the Indian Ocean [3], MLP in the Indian Ocean [3], [4] and tropical Pacific Ocean [5], support vector machines/regressions (SVM/R) in the northeast Pacific [6], and tropical Atlantic Ocean [7]. Such past studies have considered features only from past SST at a target grid, assuming that future SST values can be predicted from their own past. Very few past studies have experimented with a cause-effect approach in SST prediction, including MLP in 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 owing to the non-inclusion of features from nearby grids around the target location. Aparna *et al.* attempted to include features from surrounding grids, but only from past SST. Because meteorological parameters were missing from the features, this study also exhibited high errors [10]. In recent past, many attempts using DL models can be noted for SST predictions. This includes the LSTM network in the Bohai Sea [11], Yellow Sea [12], Korea Sea [13], gated recurrent unit (GRU) in the Bohai Sea [14], [15] and Yellow Sea [16], and CNN in the Pacific Ocean [15], [17].

Various past attempts have been made either by simple machine learning models or by lacking causal parameters in features. Therefore, high errors were noted for SST predictions in those attempts; further, this raises questions on the applicability of such methods for longer lead times. In addition, prediction errors were observed to be biased towards warm SST in many of these past studies [14], [15], [17].

Therefore, to address this significant research gap, we propose DL models that consider both meteorological and oceanic features to predict SST at longer lead times in one of the high SST variability regions in the northwest Pacific Ocean near Japan's east coast, 'Tohoku'.

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

II. STUDY LOCATIONS

This study focuses on several locations in the Tohoku region (38°N–40°N, 141°E-144°E, Fig. 1), which hosts a major mixture of the 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 SST prediction. In addition, the Tohoku region is the most habitable region for major fish species. Hence, SST predictions with a lower computational complexity are highly appreciated in the Tohoku region. Various study locations have been considered, including coastal waters and offshore regions. In addition, these study locations are at the fronts of major spatiotemporal SST variability (right panel of Fig. 1). We also developed the proposed DL model at the Kuroshio Extension Observatory (KEO) buoy location (32.3°N, 144.6°E) to validate it against in-situ SST. The in-situ SST data at the KEO buoy are available through.*[1]

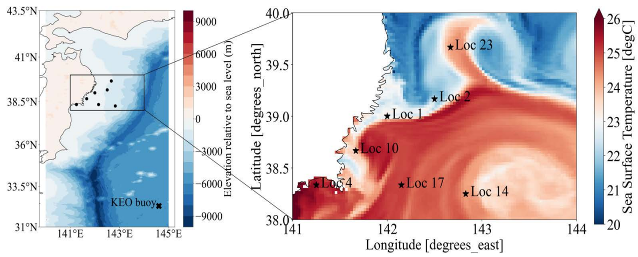

FIGURE 1. Various study locations in Tohoku region. Loc 1, 4, and 10 are in coastal region, whereas rest of them are in offshore region. Loc 1-23 the DL models were validated against target SST, whereas at KEO buoy the validation was performed against in-situ SST.

III. DATA

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

A. TRAINING DATA

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

ERA5 data provide several meteorological parameters, of which important parameters are extracted over the study locations. These includes the 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 temperature show strong coupling with SST [20], solar radiation and sensible heat flux govern the major heat source [21], [22], and wind speed and total cloud cover govern the cooling [22], without considering that any of these variables could lead to poor skills [8], [9]. ERA5 data are sufficiently accurate over the study area to develop DL models [23]. The spatial resolution of the ERA5 data is 0.25° × 0.25°, and it is available on an hourly timescale. ERA5 data were downloaded using the ECMWF Climate Data Store API (CDSAPI) [24]. Meteorological variables were extracted on

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

the grids closest to the study locations; thus, DL models were trained at 0.25° × 0.25° to account for the wider spatial domain.

B. TARGET DATA

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

IV. PROPOSED DEEP LEARNING MODELS

The proposed DL models were developed at various locations shown in Fig. 1 over the Tohoku region. Features of DL models were governed by past instances of ERA5 reanalysis data from 2m-air temperature, 2m-dew point temperature, solar radiation, total cloud cover, surface sensible heat flux, wind speeds (meridional and zonal), and ocean currents from JCOPE2. Features were moving-averaged over a window of 2 to 4 h to reduce the noise in the features. Training features were extracted at the grid closest to the study locations and surrounding areas, whereas the target SST was interpolated at study locations using bi-linear interpolation. The duration of the training and validation data was solely based on the availability of the target SST. The total duration of the target SST data was available from 2018-08-01 to 2018-12-31 on an hourly timescale. In total, 3672 h (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 there was no overlap between the training, validation, and testing datasets. The testing set has approximately 730 h (30 d), which is sufficient to consider the diurnal variability in SST [9], [15]. DL models were formulated based on the number of grids considered for the features (see Fig. 2 for details). The Deep MLP and LSTM consider features only from the target grid, whereas the spatial 2D CNN collects 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, wind speeds (meridional and zonal), and ocean currents constitute the features of the DL models. SST values alone may not be sufficient to capture high SST variability, as they are governed more by other causal variables [2], [8], [9]. The proposed DL models were targeted for SST prediction at a lead-time of 24-hours. Details of the individual models are explained in the following sections.

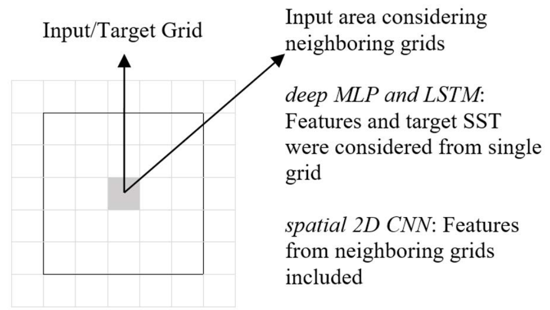

FIGURE 2. Input area/grid used for formulation of DL models using various meteorological parameters and ocean currents.

A. DEEP MLP

A deep MLP was 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 the target grid surrounding the study location were extracted as features for DL models. These features were moving averaged to reduce their noise, and the target SST data of future time steps were considered for training. All the features were converted into a single column vector (as shown in Fig. 3) to form an input layer. This input layer is connected to several deep hidden layers.

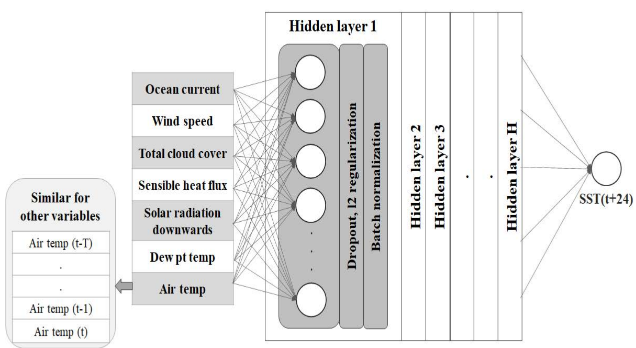

FIGURE 3. Deep MLP architecture proposed based on meteorological parameters and ocean currents as features for SST prediction. Features were considered from single target grid and stacked together in a single column vector.

Each deep hidden layer consists of several hidden neurons, a dropout ratio, regularization, and batch normalization layers. Drop-out and regularization layers were added to avoid overfitting during training. The final deep hidden layer is connected to an output layer consisting of a single neuron.

The output layer compares the predicted SST from the deep MLP with the target SST, and adjusts the internal parameters with respect to the observed error. The training procedure for the deep MLP was repeated for several iterations until the desired error was achieved. For each iteration, different hyperparameters were used to obtain the generalization ability.

B. SPATIAL 2D CNN

The spatial 2D CNN model exhibits a major difference in feature gathering. A spatial 2D CNN gathers features from the closest grid to the study location, as well as from its neighbors. The features of spatial 2D CNN are a three-dimensional matrix with the size of ‘latitudes times longitude times past_hourly timesteps of 2m-air temperature, 2m-dew point temperature, solar radiation, total cloud cover, surface sensible heat flux, wind speeds (meridional and zonal), and ocean currents. A spatial 2D CNN is thus supposed to extract 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 attempts to extract the features from the previous convolutional layer. Each convolutional layer is further connected to the *dropout layer* with regularization to avoid overfitting. After several convolutional layers, feature maps were extracted in a single column vector, which was further connected to several hidden layers, as in the deep MLP. The internal parameters of the spatial 2D CNN were adjusted with respect to the error between the spatial 2D CNN output and target data. Fig. 4 shows the architecture of the proposed spatial 2D CNN model.

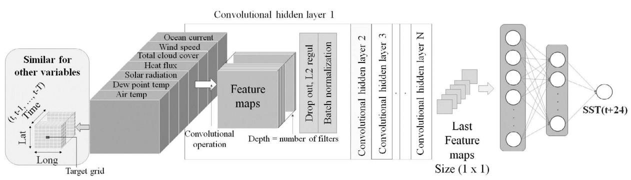

FIGURE 4. **Proposed spatial 2D CNN architecture using spatio-temporal features around target grid from meteorological parameters and ocean currents as features for prediction of SST.**

C. LONG SHORT-TERM MEMORY NETWORKS

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

D. DL MODEL PARAMETER OPTIMIZATION

There are two types of parameters in the DL models to be optimized. The first is *internal parameters* that can be adjusted

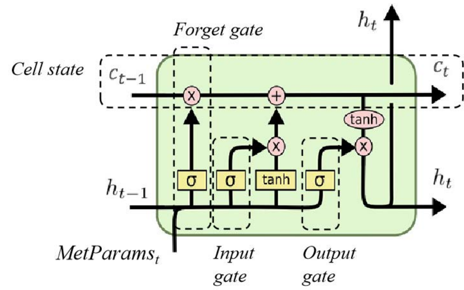

FIGURE 5. **Proposed recurrent neural network model for SST prediction trained using past meteorological features at each study location (as in Fig. 1) based on LSTM [26].**

with respect to the observed error between the DL model output and the target SST. A mean square error based on *adam* optimizer was considered to adjust the internal parameters of the DL model. The second is external or popularly known as hyper-parameters that cannot be adjusted by the observed error, but rather fixed by a trial-and-error approach or with the help of a sophisticated optimization algorithm. In the current study, the popular but computationally less intensive and equally effective *random search algorithm* is used to optimize the hyperparameters, as the performance of the DL model is very sensitive to them. Various hyper-parameters include past time-steps of causal features (2–16 h), moving average window for features (2–4 h), number of hidden/convolutional/LSTM layers (2 to 5), number units in each hidden/convolutional/LSTM layer (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 (0.15–0.4), and L2 regularization (1e-5 to 1e-2).

V. RESULTS

The results from various proposed DL model experiments during the testing period of 730 h (30 d) at a lead-time of 24-hours. The performance of the DL model was evaluated using the root mean square error (RMSE) and the correlation skill (corr. skill) calculated between the DL model output and the target SST or in-situ SST. Lower RMSE and higher correlation. skill values are better, and vice versa.

A. COMPARISON OF PROPOSED DL MODELS

To understand the effectiveness of gathering 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 a comparison of the DL model predictions against the target SST at 24-hour lead-time.

The errors in SST predictions at 24-hour lead-time with deep MLP were observed to be between 1.0°C and 2.5°C, and those from the LSTM models were found between 1.0°C and 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 the inability to extract useful patterns from short lengths of past variables compared to spatial 2D.

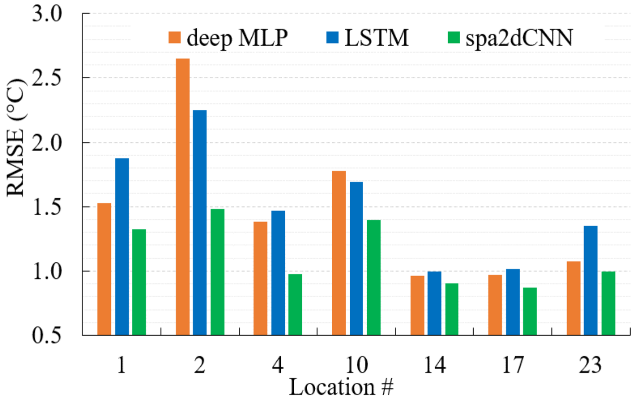

FIGURE 6. RMSE comparison between deep MLP, LSTM and spatial 2D CNN based on RDL models output and target SST. The location numbers in this figure correspond to those in Fig. 1.

Such poor skills of LSTM in SST predictions have also been noted at a few locations on the east coast of Japan [9]. The above-mentioned errors are quite high owing to the non-inclusion of features from the neighboring grid, but comparable to some of the previous studies that considered a 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 the spatial 2D CNN experienced fewer errors than the deep MLP and LSTM at each location. The errors at the offshore locations (Locs 17, 23, and 14) were marginally lower (except Loc 2) and significantly lesser at coastal locations (Locs 4, 10, and 1). The lower RMSE at coastal locations was due to the inclusion of past spatiotemporal causal features from neighboring grids (especially from wind speed and ocean current), which are largely responsible for the high variability in SST.

Even though the errors in spatial 2D CNN were less than those in 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 attempted to improve spatial 2D CNN models to reduce these errors.

B. IMPROVISED SPATIAL 2D CNN

To improve the earlier skillful spatial 2D CNN's, we propose adding 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 in the data section) but also from past SST values, and overcomes the crucial shortcomings of several past studies.

The performance of the improvised spatial 2D CNN is depicted in Fig. 7 and Fig. 8 at each location, using a time-series plot of the target SST against the predictions from the improved spatial 2D CNN at 24-hour lead-time. Improvised spatial 2D CNN substantially improved the errors in SST prediction compared with its earlier range (Fig. 6). The improvised spatial 2D CNN model now observes an RMSE in the range of 0.35°C to 0.75°C, and correlation skills from 0.64 0.96. Such skills are well within the acceptable limits of usability considering various applications, such as locating potential fishing zones, eddy detection, and cyclone progression [1], [2], [5], [17], which were 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].

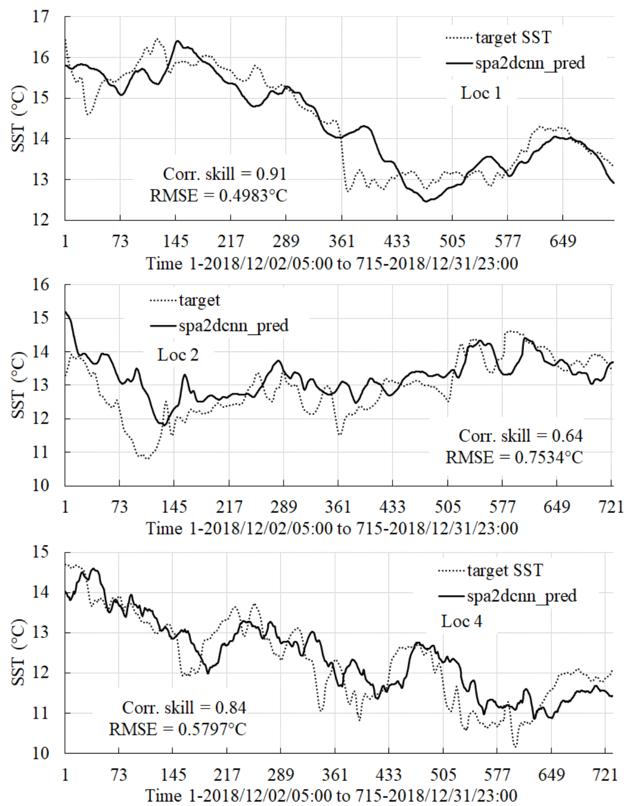

FIGURE 7. Improvised spatial 2D CNN SST prediction comparison against target SST for 24-hour lead-time at Loc 1, 2 and 4. Corresponding correlation skill and RMSE values are mentioned for each location. Improvised models output captures SST variability very well with no bias towards high SST.

The lowest error in the improvised spatial 2D CNN model was observed at Loc 17, and the highest at Loc 2. Loc 2 showed a sudden increase in depth compared to nearby locations, due to which the ocean dynamics exhibited rapid changes and hence showed a higher error. Fig. 7 and Fig. 8 show that improvised spatial 2D CNN models capture rapid changes effectively and accurately. This suggests that improvised spatial 2D CNN models are highly capable of providing skillful long-lead-time SST predictions.

VI. DISCUSSION

In the proposed study, long lead time SST predictions were attempted using various DL models. Several previous studies have attempted SST prediction based only on a time-series approach [3], [4], [15]. The major limitation of such

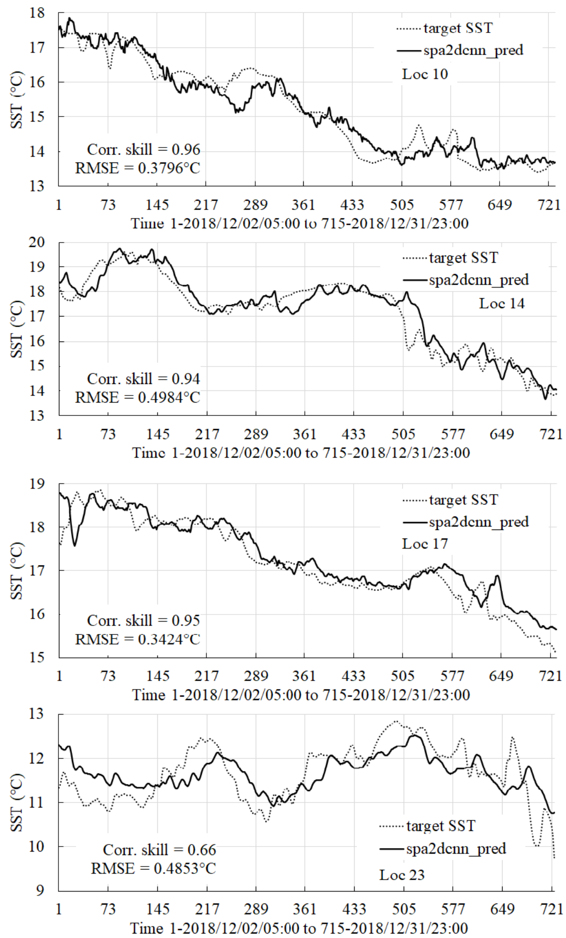

FIGURE 8. Same as Fig.7 but for Loc 10, 14, 17 and 23.

an approach is its inability to capture rapid changes in SST variability, especially at coastal locations. Therefore, to overcome these limitations, the proposed study combined spatiotemporal features from past meteorological data along with oceanic parameters. Various DL models have been developed for this purpose, in which deep MLP and LSTM consider features from a single target grid, and spatial 2D CNN models gather the features from neighboring grids. We further analyze the skills from the proposed models and discuss their validation with in-situ SST at the KEO buoy, comparison with persistent models (PM), extraction of trained feature maps from the first convolution layer, and error distribution at each location in further subsections.

A. VALIDATION OF IMPROVED MODELS WITH IN-SITU SST

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

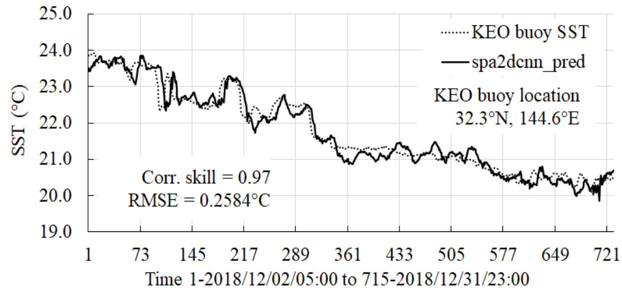

FIGURE 9. In-situ validation of improvised spatial 2D CNN model prediction at KEO buoy at 24-hr lead-time, exhibiting very high prediction skills.

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 the deep MLP and LSTM, the spatial 2D CNN models outperformed significantly at coastal locations and marginally at offshore locations (Fig. 6). Even though such errors were comparable with those of earlier studies, this margin was high considering the applications of such SST predictions. To further enhance the performance of the proposed DL models, past spatiotemporal SST values were added along with meteorological features and ocean currents.

A significant improvement in the RMSE was noted after the addition of past spatiotemporal SST values as additional features. The errors from the improved spatial 2D CNN's were reduced from 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 introduced an RMSE vs. corr. skill based two-dimensional space diagram (RC). In the RC diagram, the skills are plotted for each location. Along the x-axis, the higher the correlation, the point moves towards the right; similarly, along the y-axis, the point moves towards the bottom. Therefore, in general, an improvement in the RC diagram can be noted with points moving towards the bottom-right corner. In Fig. 10, the RC diagram between the improvised model and PM skills is depicted, and the movement of skills from the improvised models in the bottom-right corner suggests the superior skills of improvised spatial 2D-CNN compared with PM, except at Loc 2.

C. FEATURE MAPS FROM IMPROVED SPATIAL 2D CNN

Feature maps from CNN's are essential tools for analyzing the connection between the training features and target data (Fig. 4). These feature maps are the outputs of the intermediate convolutional layers. This helps to understand how the training features are transformed into convolution layers to output the desired predictions, in this case, the hourly SST at a particular location. For this purpose, we analyzed the feature maps for Loc 23 and compared them with training features.

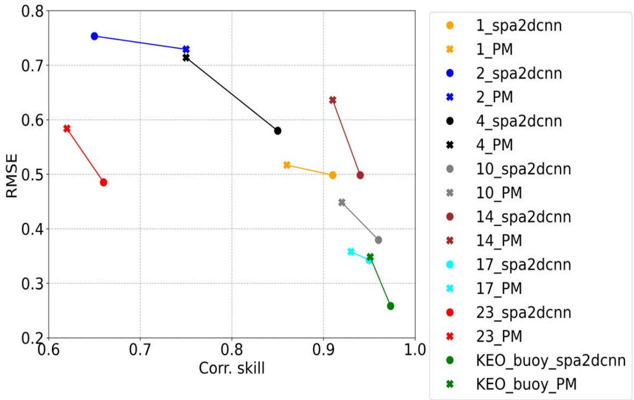

FIGURE 10. **Comparison of improvised spatial 2D CNN with PM skills at each location. The movement of points towards the bottom right corner denotes the skill improvement. Improvised models show significantly better skills than PM at each location except at Loc 2.**

In Fig. 11a, the training features from past meteorological and oceanic features used for the prediction of the 108th SST data point during the testing period are shown. The trained features extracted from the first convolutional layer are shown in Fig. 11b. The different training features were merged into common and separate features. For example, the meridional wind (v-wind) component (Fig. 11a) was captured by filter number 2 in Fig. 11b. Similarly, zonal wind (u-wind) is captured in filter numbers 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 mixing (Fig. 1); therefore, among the various training features, SST, zonal and meridional wind, and meridional currents are found to dominate over the other. Thus, CNN can efficiently extract the required spatial meteorological and oceanic features and merge them to pass on to further convolutional layers.

D. EFFECTIVENESS IN CAPTURING LOW SST

Sudden cooling of ocean water was noted at a few study locations, for example, Locations 1, 10, 14, and 17. This was due to the rapid mixing of currents and changes in 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 is more persistent than cold SST. The cold SST arises owing to the sudden change in meteorological and physical factors; 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 the fishing zone or maintain a low-pressure zone during cyclones. Hence, it is very important that the prediction of such a low SST be as accurate as the total study period. With the improvised spatial 2D CNN, we noted a similar accuracy in the low SST in comparison with the error observed during the complete testing period (Fig. 7 and 8).

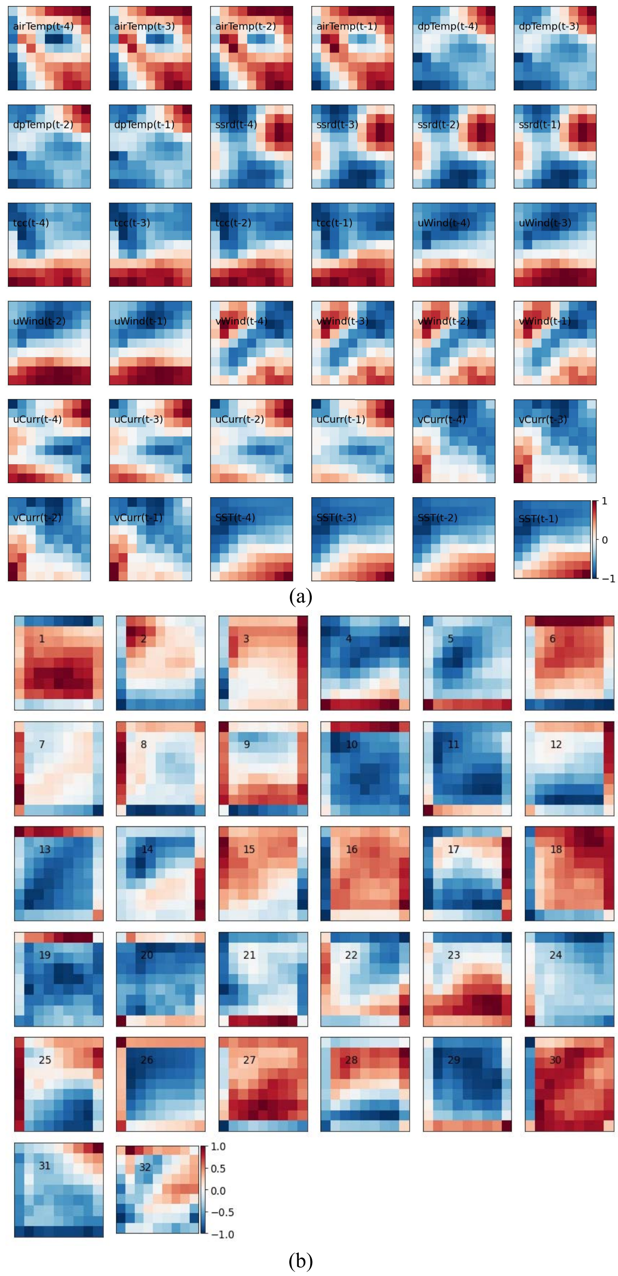

FIGURE 11. **(a) Training features from past four hours surrounding Loc 23 for 108th data-point during testing period (Ref. Fig. 8). Center pixel in the above figure corresponds to the Loc 23, and the x-axis and y-axis represents the latitudes and longitudes, resp. Training features are normalized within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

E. ERROR DISTRIBUTION

To understand the nature of the bias in improvised spatial 2D CNN predictions, we further investigated the error distribution between the target SST and improvised spatial

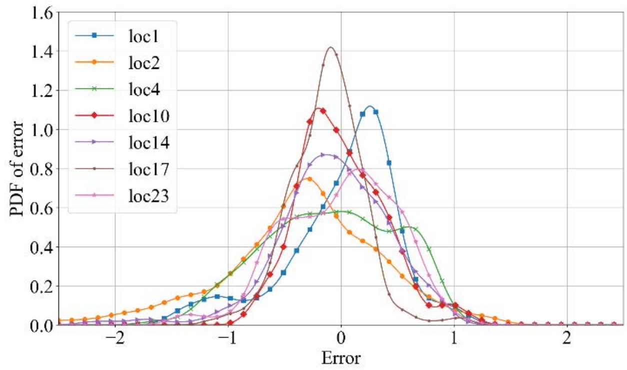

FIGURE 12. **Probability distribution function of error at each location. Large number of errors are concentrated between −0.5° to 0.5° C, suggesting that high errors are less likely to occur.**

2D CNN predictions. Fig. 12 shows the probability distribution function of the error 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 that CNN predictions do not exhibit large errors, and hence, are more reliable.

VII. CONCLUSION

- Among the proposed deep learning models, the deep MLP and LSTM showed high errors at coastal locations, whereas equal prediction skills were noted at offshore locations.
- The inclusion of more spatio-temporal features from the surrounding grids significantly reduced the errors in the spatial 2D CNN, especially at coastal locations.
- To further improve the spatial 2D CNN, spatio-temporal past SST features were added, which helped them surpass the PM prediction skills (Fig. 10).
- Various important features pertaining to the inputs surrounding the study area were effectively noted by the improved spatial 2D CNN, as shown in Fig. 11a and Fig. 11b. This suggests that the applicability of the proposed improvised models in other study areas will result in similar skills in SST prediction, exhibiting high variability in SST.
- The peculiar feature of improvised spatial 2D CNN is the non-biasedness of prediction skills towards warm SST. Such important observations were missing in past studies on SST prediction [14], [15], [17].
- The future work of the current study is to consider high-resolution input features and more oceanic parameters (e.g., sub-surface temp) in a large spatial domain instead of a few locations. Because the proposed method is highly adaptive and efficiently captures the high SST variability, it is possible to cluster similar variability SST locations and collectively develop a DL model for each cluster. This significantly reduces the computational cost during training when the proposed DL models are applied over a wide spatial domain.
- When applied to the spatial domain, the proposed framework will be highly useful for estimating potential fishing zones [27], probable movements of the progressing cyclone [28], knowing expected heatwaves, and detecting 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,'' *J. Geophys. Res., Oceans*, vol. 104, no. C5, pp. 11337–11358, May 1999.

[2] P. C. Chu, C.-L. Fang, and C. S. Kim, ''Japan/East sea model predictability,'' *Continental Shelf Res.*, vol. 25, no. 17, pp. 2107–2121, Nov. 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 J. Mar. Sci.*, 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,'' *J. Atmos. Ocean Technol.*, vol. 35, no. 7, pp. 1441–1455, Jul. 2018.

[5] 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.

[6] Q.-J. Li, Y. Zhao, H.-L. Liao, and J.-K. Li, ''Effective forecast of northeast Pacific sea surface temperature based on a complementary ensemble empirical mode decomposition–support vector machine method,'' *Atmos. Ocean. Sci. Lett.*, vol. 10, no. 3, pp. 261–267, May 2017.

[7] 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.

[8] E. Garcia-Gorriz and J. Garcia-Sanchez, ''Prediction of sea surface temperatures in the western Mediterranean Sea by neural networks using satellite observations,'' *Geophys. Res. Lett.*, vol. 34, no. 11, 2007, Art. no. L11603.

[9] 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.

[10] 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, 2018.

[11] 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.

[12] 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.

[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 Sens.*, vol. 12, no. 21, p. 3654, Nov. 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,'' *J. Mar. Sci. Eng.*, vol. 8, no. 4, p. 249, Apr. 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 Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 740–744, May 2020.

[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,'' *Math. Problems Eng.*, vol. 2020, pp. 1–9, May 2020.

[17] G. Zheng, X. Li, R.-H. Zhang, and B. Liu, ''Purely satellite data–driven deep learning forecast of complicated tropical instability waves,'' *Sci. Adv.*, vol. 6, no. 29, Jul. 2020, Art. no. eaba1482.

[18] T. Kono, M. Foreman, P. Chandler, and M. Kashiwai, ''Coastal Oyashio south of Hokkaido, Japan,'' *J. Phys. Oceanogr.*, vol. 34, no. 7, pp. 1477–1494, Jul. 2004.

[19] L. Bengtsson, ''Medium-range forecasting—The experience of ECMWF,'' *Bull. Amer. Meteorolog. Soc.*, vol. 66, no. 9, pp. 1133–1146, Sep. 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," *J. Korean Soc. Mar. Environ. 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 J. Mar. Sci.*, vol. 23, pp. 8–13, Mar. 1994.

[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," *J. Geophys. Res.*, vol. 112, no. C7, 2007, Art. no. C07015.

[23] H. Hersbach, B. Bell, P. Berrisford, A. J. M. S. Horányi, J. M. Sabater, J. Nicolas, R. Radu, D. Schepers, A. Simmons, C. Soci, and D. Dee, "Global reanalysis: Goodbye ERA-Interim, hello ERA5," *ECMWF Newslett.*, vol. 159, pp. 17–24, Jun. 2019.

[24] *ERA5: Fifth Generation of ECMWF Atmospheric Reanalysis of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS)*. Accessed: Jun. 20, 2019. [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview

[25] Y. Miyazawa, S. M. Varlamov, T. Miyama, X. Guo, T. Hihara, K. Kiyomatsu, M. Kachi, Y. Kurihara, and H. Murakami, "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 Dyn.*, vol. 67, no. 6, pp. 713–728, Jun. 2017.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, 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 D.-C. Liu, "Sea surface temperature fronts affect distribution of Pacific saury (Cololabis saira) in the northwestern Pacific ocean," *Deep Sea Res. II, Topical Stud. Oceanogr.*, vol. 107, pp. 15–21, Sep. 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," *Meteorolog. Appl.*, vol. 26, no. 1, pp. 130–139, Jan. 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 Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7224–7234, Sep. 2021.

KALPESH R. PATIL received the B.Tech. degree in civil engineering from the College of Engineering Pune, Maharashtra, India, in 2007, and the M.Tech. degree in water resources engineering and the Ph.D. degree in ocean sciences from IIT Bombay, Maharashtra, in 2012 and 2019, respectively. From 2008 to 2010, he was a Senior Software Developer of mobile application development at Mahindra Satyam (formerly, Satyam Computer Services Ltd.), Bangalore, India. From 2013 to 2014, he was a Research Assistant at the Civil Engineering Department, Institute of Technology, Bombay. Currently, he is working as a Young Research Fellow with the Application Laboratory, Japan Agency for Marine Earth-Science and Technology, Yokohama, Japan. 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 (Member, IEEE) 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 with the Faculty of Data Science, Shiga University. His research interests include computer vision, pattern recognition, machine learning, and remote sensing for fishery.