Sea Surface Temperature Forecasting with Ensemble of Stacked Deep Neural Networks

Mohammad Jahanbakht, *Student Member, IEEE*, Wei Xiang, *Senior Member, IEEE*, and Mostafa Rahimi Azghadi, *Senior Member, IEEE*

***Abstract*—Oceanic temperature has a great impact on global climate and worldwide ecosystems, as its anomalies have been shown to have a direct impact on atmospheric anomalies. The major parameter for measuring the thermal energy of oceans is the Sea Surface Temperature (SST). SST prediction plays an essential role in climatology and ocean-related studies. However, SST prediction is challenging due to the involvement of complex and nonlinear sea thermodynamic factors. To address this challenge, we design a novel ensemble of two stacked deep neural networks that uses air temperature, in addition to water temperature, to improve the SST prediction accuracy. To train our model and compare its accuracy with the state-of-the-art, we employ two well-known datasets from the national oceanic and atmospheric administration as well as the international Argo project. Using deep neural networks, our proposed method is capable of automatically extracting required features from the input timeseries and utilizing them internally to provide a highly accurate SST prediction that outperforms state-of-the-are models.**

***Index Terms*—Sea surface temperature, deep neural networks**

I. Introduction

SEASONAL weather forecasting and SST prediction has attracted increasing attention in scientific literature [1], [2]. Because oceans cover approximately three quarters of the surface of our planet, accurate SST prediction can provide noticeable benefits to many environmental-related studies and applications. To address the SST forecasting problem, various prediction algorithms have been introduced in the literature [3], [4], [5]. These diverse predictive models can be categorized into: (i) physics-based numerical models, (ii) classic statistical methods, (iii) traditional neural networks, and (iv) deep neural networks.

Physics-based numerical models use complex kinetic and thermodynamic equations along with exciting parameters and boundary conditions. General circulation models [3] and regional ocean modeling system [6] are commonly used methods in this category. The second category of the SST predictive models, uses classic statistical methods to create a mathematical model that embodies numerical relationship between one or more random variables. Markov model [4] and linear regression [5] are two repeatedly reported statistical methods. Traditional (a.k.a shallow) neural networks including Support Vector Regression (SVR) [1] and wavelet neural networks [6], are the third type of methods used for SST forecasting.

The fourth category of SST predication models are Deep Neural Networks (DNN), which incorporate multiple hidden layers to extract data features and to automatically learn SST variation rules. These methods are very popular due to easy access to big and up-to-date collections of in-situ and remotely sensed SST data collected by various organizations and made publicly available. Some well-established models in this category include Fully-Connected Long Short-Term Memory (FC-LSTM) [1], Gated Recurrent Unit Encoder–Decoder (GED) [1], and Convolutional LSTM [2], which are shown to have higher accuracy compared to the first three categories [1].

The SST forecasting DNN models can be divided into timeseries and next-frame predictors. While a timeseries predictor works with spatial averaged SST [2], next-frame predictors use SST distribution matrix in an area [7]. From a functionality perspective, a next-frame predictor results in a higher mean squared error, compared to its timeseries counterparts. This is simply because the next-frame predictor needs to predict a 2D matrix, while timeseries predictors predict a single value. Consequently, various practical applications need to choose between next-frame or timeseries predictors, depending on the trade-off between spatial coverage or prediction accuracy.

Here, we propose a new SST timeseries prediction model, which consists of two stacked DNNs. While previous DNN models have only used water temperature as their input [1], [2], we also use readily available air temperature data for improving the efficiency of our model. Two correlated air and water temperature variables are fed separately to two stacked DNNs. These two networks then form an ensemble to create our highly accurate model. Simulation results show that our model outperforms the state-of-the-art, even without using the air temperature data. However, the accuracy is enhanced by introducing an ensemble node, which merges the SST and air temperature timeseries into the final prediction output.

II. Background and Problem Formulation

Various factors can affect SST variations. These include solar radiation, surrounding air temperature, heat exchange with atmosphere, wind speed at sea surface, evaporation, ocean internal processes, etc. [1]. Due to the diverse nature

This work is funded by the Australian Government Research Training Program Scholarship.

M. Jahanbakh and M. Rahimi Azghadi are with the College of Science and Engineering, James Cook University, Queensland, Australia (e-mail: mohammad.jahanbakht@my.jcu.edu.au; mostafa.rahimiazghadi@jcu.edu.au).

Wei Xiang is with the School of Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia (e-mail: w.xiang@latrobe.edu.au).

TABLE I
TEMPOROSPATIAL COORDINATES OF THE SEA AREAS UNDER STUDY AND THEIR CORRESPONDING LAND WEATHER STATIONS

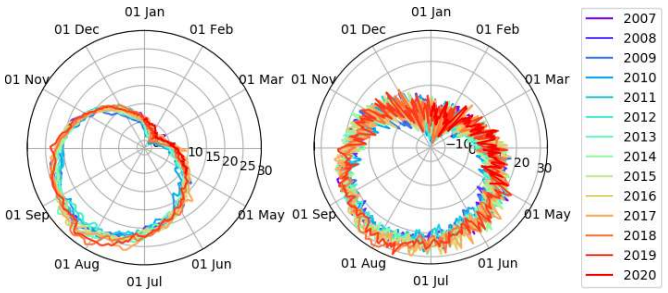
Location Name	Type of Timeseries	Data Source	Spatial Coordinates	Sampling Duration
Bohai Sea	Endogenous Averaged SST	NOAA PSL	117.5- 121.5E, 36.5- 40.5N	1998-2020
Dalian City	Exogenous Air Temp	NOAA NCEI	121.633E, 38.9N	1998-2020
South China Sea	Endogenous Averaged SST	NOAA PSL	112.5- 119.5E, 6.5- 21.5N	1998-2020
Dagupan City	Exogenous Air Temp	NOAA NCEI	120.35E, 16.083N	1998-2020
North Pacific Ocean	Endogenous Averaged SST	Argo	130- 190E, 10- 50N	2004-2019
Amami Island	Exogenous Air Temp	NOAA NCEI	129.5E, 28.383N	2004-2019

of these factors, accessing them in a desired location is not always possible. Among these factors, the air temperature is a common observation that can be found in almost any weather station worldwide. Therefore, the air temperature can be easily obtained to be used in conjunction with historical SST to devise a better SST prediction model. If the air temperature is not available for a desired point, it can be obtained from an iso-latitude station to act as an exogenous variable.

In order to perform precise SST forecasting, here we propose to utilize SST and air temperature, simultaneously. The SST timeseries are extracted from the National Oceanic and Atmospheric Administration (NOAA) Physical Science Laboratory (PSL) and the Argo data sources, while the air temperature timeseries come from NOAA National Centers for Environmental Information (NCEI) [8]. Geolocations of these timeseries are presented in Table I. The SST values in this table are area averaged over the corresponding spatial coordinates. The choice of these referenced coordinates are in accordance with other published works [1], [2], which enable us to compare the accuracy of our approach with literature.

The bilateral relationship between the SST and air temperature has already been studied in climatological research. More specifically, SST has been used as a reliable predictor of weather anomalies [9], while air temperature is used to assist in predicting SST [10].

A close inspection of Fig. 1 reveals the close relationship between SST and air temperature factors in two typical geolocations of our datasets (i.e., Bohai Sea and Dalian City). However, the air temperature usually has higher dynamics compared to the SST, which results in a lower statistical correlation between them. To address this problem, a moving average window can be slid over the air temperature data sequence. This window acts as a low-pass filter (LPF), which smooths the high dynamics of data. This smoothing process will increase the accuracy of our proposed SST forecasting algorithm. To elaborate, by applying an LPF to the air temperature, the Pearson correlation coefficient increases to over 0.8 for all the geolocations in Table I. This indicates the existence of a high linear relationship between the smoothed



(a) SST at Bohai Sea (b) Air Temp in Dalian City

Fig. 1. Polar plot of (a) the area-averaged SST at Bohai Sea and (b) the air temperature at its iso-latitude Dalian City, both from the NOAA data source.

air temperature and SST.

In this paper, the SST forecasting problem is represented by function $F_1(SST)$, which returns a prediction value. Similarly, predicting SST from historical air temperature values can be formulated as function $F_2(Air)$, which returns another SST prediction value. The predictive outputs of these two functions can then be merged to build an enhanced SST predictor, i.e.,

$$F^{+}(SST, Air) = w_{SST}\ F_1(SST) + w_{Air}\ F_2(Air)\,, \quad (1)$$

where w_{SST} and w_{Air} coefficients adjust the relative contributions of $F_1(SST)$ and $F_2(Air)$ in the final prediction.

It is worth mentioning that the SST observations for the North Pacific Ocean in the Argo data source, along with its corresponding air temperature reads (i.e., the air temperature from the Amami Island) are monthly averaged throughout this paper. Consequently, any further smoothing is unnecessary for this geolocation.

III. NETWORK ARCHITECTURE

The high-level block diagram of the proposed voting ensemble of stacked DNNs is shown in Fig. 2(a). This model consists of two separate stacked DNN branches that are trained with different datasets. The first stage of the top branch ($F_1(SST)$) takes the SST timeseries as input and performs general preprocessings e.g. outlier detection and missed data interpolation. It then passes the clean data to the seasonality decomposition block. This block decomposes the year seasonality from its raw input timeseries, outputting trend and residual.

Trend is a straight line that matches as closely as possible to the original timeseries. It can be easily found by linear regression. Year seasonality on the other hand, is the repeating one-year-long cycle in data. It can be calculated by yearly averaging SST, after subtraction of the trend line. Calculation of both the year seasonality and trend must be carried out with the training dataset only, leaving the test dataset completely unseen to the system. Finally, the residual is the remaining random variation in the SST timeseries, which has not been taken into account in the trend and seasonality [11]. The sum of the trend and residual outputs are then fed to the core processing block, which has an stacked architecture [12] of an LSTM and a Multilayer Perceptron (MLP) network. This block is illustrated in Fig. 2(b). It consists of two cascaded LSTM layers and one fully-connected MLP layer.

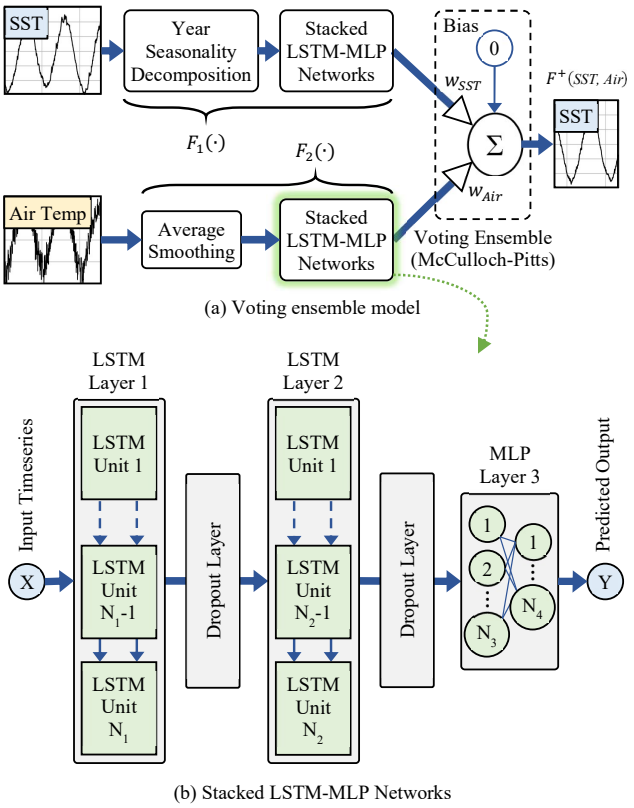


Fig. 2. Block diagram of (a) the voting ensemble model for SST forecasting, which consists of two (b) stacked LSTM-MLP deep neural networks.

The two cascaded LSTM layers are referred to in parlance as *encoder-decoder* or *Seq2Seq*[13]. The first LSTM layer translates the dynamicity of the input sequence into a higher dimensional representation, whilst the second LSTM layer extracts useful features to feed to the next MLP layer. The MLP layer then combines all the automatically discovered features in the data, into its predictive output. Additionally, two dropout layers are placed between each consecutive DNN pairs to prevent the model from overfitting.

The second DNN branch in Fig. 2(a) (i.e., $F_2(Air)$) takes the air temperature as input and performs the same preprocessing as that on the SST. The next block smooths the given high-dynamic data with a moving average window, which acts like an LPF to increase the correlation between the air temperature and SST. Similar to the first branch, the smoothed data is then fed to the core stacked processing block to produce the final outcome of the second branch.

Later in Section IV, we will show the prediction efficiency of $F_1(SST)$ as a standalone branch. However, to further improve the prediction accuracy of our model, we implement (1) by placing an additional block to combine the outputs of $F_1(SST)$ and $F_2(Air)$, at the end of the network architecture in Fig. 2(a). This block takes $F_1(SST)$ and $F_2(Air)$ as the inputs to a McCulloch-Pitts neuron with zero biasing and with a linear activation function. This single neuron is referred to in parlance as the *voting ensemble* [14]. As shown in (1), the main objective of this block is to merge the independently predicted results to form the final output of the network, i.e., $F^+(SST, Air)$. Nevertheless, the weights of this single neuron would be optimized by backpropagation through gradient descent algorithm to produce the ultimate outcome.

The proposed voting ensemble model in Fig. 2 includes some hyper-parameters that require optimization. The first one is M_1, which represents the number of data elements inside the input timeseries to the first LSTM layer in Fig. 2(b). This parameter can be defined as,

$$M_1 = \frac{Input\ Horizon}{Sampling\ Period}, \tag{2}$$

where the *Input Horizon* refers to the length of time, in which we look back in our historical data to provide future predictions. Additionally, N_1 and N_2 are the number of LSTM units in the first and second layers of the stacked LSTM-MLP network shown in Fig. 2(b). These two parameters should be *adequately* large. Here *'adequately'* means, firstly,

$$N_2 > N_1 > M_1\,. \tag{3}$$

Secondly, N_1 and N_2 should be large enough so that the dropped-out neurons in Fig. 2(b) do not adversely affect the network performance.

The output of the second LSTM layer is fed into the first and second dense layers of the following MLP network with N_3 and N_4 neurons, respectively. It is worth noting that very large or small values of N_i will respectively result in overfitting and underfitting, which consequently reduces the overall accuracy. To achieve the best performance, we performed a hyperparameter optimization process that resulted in $N_i\ \forall i \in \{1, 2, 3\}$ values with 200, 300, and 100 neurons. Finally, the number of output neurons from the MLP layer (N_4) should be equivalent to the number of elements in the predicted output timeseries. This parameter can be defined as the ratio of the *forecasting horizon* to the sampling period, where the *forecasting horizon* simply indicates the length of time of the prospective forecasting period.

IV. Results and Discussions

In this section, we introduce and discuss the utilized air and water temperature data sources. We also evaluate the efficacy of our proposed model compared to literature.

A. Description of Data Sources

The conducted experiments in this paper are carried out using two major data sources. The first data source is NOAA, which is a USA national scientific agency. NOAA focuses and monitors the conditions of both the oceans and atmosphere, facilitated by its various centers and laboratories. The NOAA PSL contains SST timeseries from September 1981 to present. This data covers the global oceans in daily sampling period, with 0.25 spatial resolution. In addition, the NOAA NCEI has a global information system tool, which enlists global weather stations. For each station, daily sampled atmospheric measurements such as air temperature, precipitation, wind speed, etc. are publicly available [15].

The second data source used in our experiments is Argo, which is dedicated to oceanographic research and is collected

TABLE II
MEAN SQUARED ERROR OF AREA-AVERAGED SST FORECASTING AT THE BOHAI SEA, COMPARED WITH THE DIFFERENT SCHEMES USED IN [1]

Model Name	Daily Mean			Weekly Mean		Monthly Mean
	1 Day	3 Days	7 Days	1 Week	3 Weeks	1 Month
GED [1]	**0.166**	0.415	0.742	0.350	**0.514**	0.581
FC-LSTM [1]	0.170	0.424	0.787	0.382	0.592	0.687
SVR [1]	0.472	0.692	1.005	0.578	0.627	0.711
Our $F_1(SST)$	0.166	**0.322**	**0.514**	**0.310**	0.696	0.272
Our $F_2(Air)$	2.289	2.829	1.633	0.835	1.135	**0.242**
Our $F^+(SST, Air)$	**0.157**	**0.318**	**0.508**	**0.294**	0.696	**0.194**

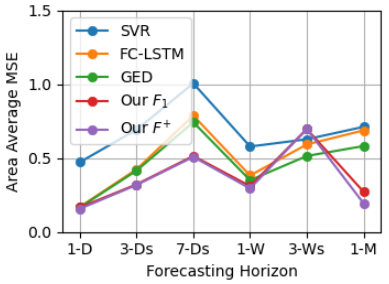


TABLE III
MEAN SQUARED ERROR OF AREA-AVERAGED SST FORECASTING AT THE SOUTH CHINA SEA, COMPARED WITH THE DIFFERENT SCHEMES USED IN [1]

Model Name	Daily Mean			Weekly Mean		Monthly Mean
	1 Day	3 Days	7 Days	1 Week	3 Weeks	1 Month
GED [1]	0.063	0.125	0.211	0.162	0.267	0.207
FC-LSTM [1]	0.061	0.140	0.218	0.168	0.285	0.343
SVR [1]	0.095	0.157	0.242	0.214	0.285	0.212
Our $F_1(SST)$	**0.055**	**0.084**	**0.131**	**0.078**	**0.135**	**0.104**
Our $F_2(Air)$	0.692	0.738	0.802	0.676	0.812	0.874
Our $F^+(SST, Air)$	**0.055**	**0.084**	**0.131**	**0.077**	**0.135**	**0.099**

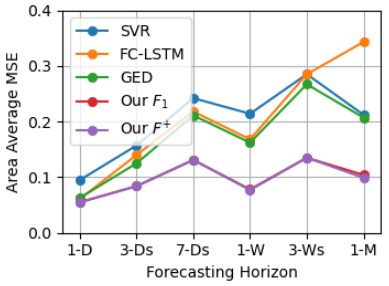
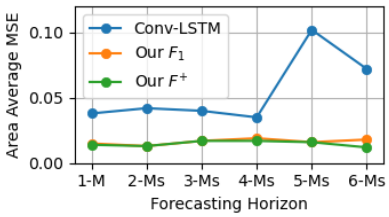


TABLE IV
MEAN SQUARED ERROR OF AREA-AVERAGED SST FORECASTING AT THE NORTH PACIFIC OCEAN, COMPARED WITH THE PROPOSED MODEL IN [2]

Model Name	Monthly Mean					
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months
Convolutional LSTM [2]	0.038	0.042	0.040	0.035	0.102	0.072
Our $F_1(SST)$	**0.015**	**0.013**	**0.017**	**0.019**	**0.016**	**0.018**
Our $F_2(Air)$	0.283	0.456	0.294	0.288	0.440	0.488
Our $F^+(SST, Air)$	**0.014**	**0.013**	**0.017**	**0.017**	**0.016**	**0.012**



and made publicly available through an international program since the early 2000s. The floating buoys of Argo record temperature, salinity, oceanic currents, bio-optical properties, etc. In contrast to NOAA, Argo does not measure atmospheric parameters, and it has as low as 1 spatial resolution [8].

B. Prediction Accuracy and Comparison

The prediction results of our voting ensemble model are compared with other published works in Tables II, III, and IV. The last three rows in these tables correspond to the three distinct nodes in our proposed model architecture, namely $F_1(SST)$, which is the output of the top branch in Fig. 2(a) (SST data only); $F_2(Air)$, which is the output of the bottom branch in Fig. 2(a) (air temperature only); and $F^+(SST, Air)$ that represents the output of our voting ensemble model in Fig. 2(a) (both SST and air temperature data).

Similar to [1], our models in Tables II and III are separately trained with the daily mean, weekly mean, and monthly mean data. Therefore, the Mean Squared Error (MSE) values for 7 days in the daily mean category is different from the MSE values for one week in the weekly mean category.

As can be seen from Tables II to IV, the prediction capability of the standalone first branch in Fig. 2(a) (i.e., $F_1(SST)$) is very strong. Among all the comparisons made, this single branch provides better SST forecasting in 16 cases out of the total 18 compared to literature. On average, our $F_1(SST)$ offers 15% and 39% better MSEs, compared to GED [1] in Tables II and III, respectively. It also provides 65% better MSE, compared to convolutional LSTM [2] in Table IV. Note that, GED is the best performing model presented in [1].

After combining the $F_1(SST)$ and $F_2(Air)$ branches using a McCulloch-Pitts neuron to build a voting ensemble, our model outperforms $F_1(SST)$ results and all previously reported works in 17 out of the 18 comparisons. To summarize, on average our proposed $F^+(SST, Air)$ ensemble model provides 19% and 40% better MSEs, compared to GED [1] in Tables II and III, respectively. It also provides 68% better MSE, compared to convolutional LSTM [2] in Table IV.

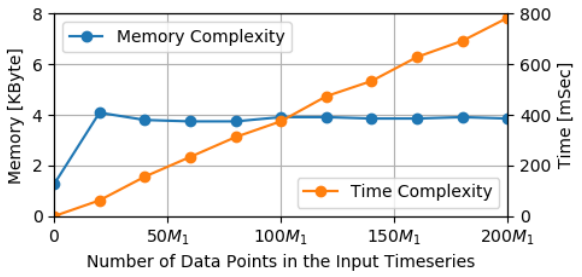


Fig. 3. Time and memory complexities with respect to the number of data points for one-day $F^+(SST, Air)$ forecasting in Bohai Sea.

These results show that our model outperforms previous works, when only using SST data in most of the cases, while it achieves a slight accuracy improvement, if an ensemble is used to include some readily available air temperature data.

To further analyze the efficiency of our proposed model, we used the big-O approach to evaluate our model's demand when changing the input size [16]. The analysis of both memory and time complexities for one-day SST prediction at the Bohai Sea are illustrated in Fig. 3. These plots quantify the growth in required computational resources, against the number of data points M_1 in the input timeseries in (2). The results are obtained during inference, where the weights and biases are fixed. The plots reveal an $O(n)$ linear growth for simulation time, and an $O(1)$ constant growth for the memory demand. These suggest our model is efficient.

When comparing our model to the state-of-the-art for SST prediction, it offers the following advantages.

- To the best of our knowledge, the proposed simple and efficient ensemble of stacked DNNs for SST prediction is unique and has not been reported previously in literature.
- Creatively incorporating the air temperature data from close-by iso-latitude weather stations makes the model more accurate and versatile. As a result, our model outperforms the previous works by 19% to 68% better prediction accuracy by adding readily available air temperature data. This improvement is more prominent in some geolocations than others. For instance, the accuracy improvement for the Bohai Sea increases from 15% to 19% when including air data.
- The innovative injection of air temperature is beneficial to our model not only for accuracy enhancement, but also in reliability improvement. Our model is more resilient to missing SST values and outliers, as the air temperature data is present. By inspecting Tables II, III, and IV, it is obvious that we can predict the SST, using the air temperature only. This prediction has less than $\pm 1C$ error in most of the cases. That is, we can use the second $F_2(Air)$ branch as a missing SST value estimator.

One may further improve our model by adding extra branches into its modular design to incorporate more exogenous factors, e.g., solar radiation or wind speed. Besides, the stacked LSTM-MLP DNN in our model can be replaced by novel Transformer networks, which may better learn long-term dependencies compared to an LSTM.

V. Conclusion

Accurate long-term SST prediction is challenging. To address this problem, we proposed a light-weight and highly-accurate new DNN structure that leverages the correlation between SST datasets and air temperature at nearby iso-latitude weather stations. We devised two stacked LSTM-MLP networks and trained them with the correlated SST and air temperature datasets. We then integrated the outputs of the two stacked networks in a voting ensemble to form a highly accurate model. We used the two well-known NOAA and Argo data sources to train and test our models. We demonstrated that our model significantly outperforms the state-of-the-art SST prediction algorithms.

References

[1] 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, Jul. 2019.

[2] K. Zhang, X. Geng, and X. Yan, "Prediction of 3-D ocean temperature by multilayer convolutional LSTM," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, Jan. 2020.

[3] G. Aparna, S. D'Souza, and B. N. Arjun, "Prediction of daily sea surface temperature using artificial neural networks," *International Journal of Remote Sensing*, vol. 39, pp. 4214–4231, Mar. 2018.

[4] N. Bounceur, I. Hoteit, and O. Knio, "A Bayesian structural time series approach for predicting Red Sea temperatures," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1996–2009, Apr. 2020.

[5] X. Meng and J. Cheng, "Estimating land and sea surface temperature from cross-calibrated Chinese Gaofen-5 thermal infrared data using split-window algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 3, pp. 509–513, Mar. 2020.

[6] K. Patil and M. Deo, "Prediction of daily sea surface temperature using efficient neural networks," *Ocean Dynamics*, vol. 67, Feb. 2017.

[7] 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.

[8] K. Iqbal, M. Zhang, S. Piao, and G. He, "Gradual but persistent quest for the ocean observation by employing multifarious sensing gadgets: a preview," in *Proceedings OCEANS*. Marseille, France: IEEE, Jun. 2019, pp. 1–7.

[9] Y. Xu *et al.*, "Contribution of SST change to multidecadal global and continental surface air temperature trends between 1910 and 2013," *Climate Dynamics*, vol. 54, no. 3, pp. 1295–1313, Feb. 2020.

[10] S. Wolff, F. O'Donncha, and B. Chen, "Statistical and machine learning ensemble modelling to forecast sea surface temperature," *Journal of Marine Systems*, vol. 208, p. 103347, Aug. 2020.

[11] K. Bandara, C. Bergmeir, and H. Hewamalage, "LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, Apr. 2020.

[12] Y. Lan, Y. Hao, K. Xia, B. Qian, and C. Li, "Stacked residual recurrent neural networks with cross-layer attention for text classification," *IEEE Access*, vol. 8, pp. 70 401–70 410, Apr. 2020.

[13] Y. Zhang, P. J. Thorburn, W. Xiang, and P. Fitch, "SSIM—a deep learning approach for recovering missing time series sensor data," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6618–6628, Aug. 2019.

[14] A. Yazdizadeh, Z. Patterson, and B. Farooq, "Ensemble convolutional neural networks for mode inference in smartphone travel survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 6, pp. 2232–2239, Jun. 2020.

[15] N. R. Nalli, A. Gambacorta, Q. Liu, C. D. Barnet, C. Tan, F. Iturbide-Sanchez, T. Reale, B. Sun, M. Wilson, L. Borg, and V. R. Morris, "Validation of atmospheric profile retrievals from the SNPP NOAA-unique combined atmospheric processing system. part 1: temperature and moisture," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 180–190, Jan. 2018.

[16] R. Vaz, V. Shah, A. Sawhney, and R. Deolekar, "Automated big-O analysis of algorithms," in *Proceedings International Conference on Nascent Technologies in Engineering (ICNTE)*. Mumbai, India: IEEE, Jan. 2017, pp. 1–6.