Journal of Sea Research 200 (2024) 102515

Contents lists available at ScienceDirect

Journal of Sea Research

journal homepage: www.elsevier.com/locate/seares

STA-SST: Spatio-temporal time series prediction of Moroccan Sea surface temperature

Isam Elafi [a], Nabila Zrira [b], Assia Kamal-Idrissi [c], Haris Ahmad Khan [d,e,*], Aziz Ettouhami [a]

[a] Laboratory of Conception and Systems - Mohammed V University in Rabat, Rabat, Morocco
[b] ADOS Team, LISTD Laboratory, National Superior School of Mines Rabat, Rabat, Morocco
[c] Ai Movement, Center of Artificial Intelligence, Mohammed VI Polytechnic University, Rabat, Morocco
[d] Agricultural Biosystems Engineering, Wageningen University & Research, Wageningen, the Netherlands
[e] Data Science, Crop Protection Development, Syngenta, the Netherlands

ARTICLE INFO

Keywords:
BiLSTM
Sea surface temperature (SST)
Forecasting
Marine data
Morocco
Convolution
Attention

ABSTRACT

Global Sea Surface Temperature (SST) trends have garnered significant attention in several ocean-related domains, including global warming, marine biodiversity, and environmental protection. This involves having an accurate and efficient forecast of future SST to ensure early detection and response in time to these events. Deep learning algorithms have become popular in SST prediction recently, although directly obtaining optimal prediction results from historical observation data is not simple. In this paper, we propose STA-SST, a new deep learning approach for forecasting SST, by combining the temporal dependencies of SST using the Bidirectional Long Short-Term Memory (BiLSTM) model, spatial features extracted from the convolution layer, and relevant information from the attention mechanism. To assess how well the Attention-BiLSTM with convolution layer predicts SST, we conducted a case study in the Moroccan Sea, concentrating on five different areas. The proposed model was compared against alternative forecasting models, including LSTM, XGBoost, Support Vector Regression (SVR), and Random Forest (RF). The experimental results show that STA-STT produces noticeably the best prediction results and is a solid choice for field implementation.

1. Introduction

In situ observations, such as marine stations, volunteer observation ships, and buoys, along with satellite instrumentation, have yielded invaluable studies of the physical properties of the ocean surface over the last few years. These findings are now essential for comprehending ocean dynamics, even with their disparate spatial and temporal sampling and measurement methods. Notably, our understanding of medium- to large-scale ocean variability (about 100–200 km) has been improved by using altimetry data.

For both short-term and long-term forecasts, an accurate SST prediction is crucial. These forecasts support decision-making for tasks like the distribution of fishing resources and the preservation of the marine environment. As a result, SST prediction is essential for many applications, such as ship route planning (Qian et al., 2023), weather forecasting (Ahn and Dowd, 2023), tracking marine animals (Horton et al., 2024), and marine fisheries (Mediodia et al., 2023; Khaled et al., 2023).

SST prediction has been the subject of intense investigation in recent literature. This topic involves applying prediction models with various time horizons to generate future SST values in advance. SST data, however, is frequently lacking and organized as a time series with intricate or pointless coupling processes and spatio-temporal correlations. These characteristics present difficulties for conventional methods of data analysis.

Traditionally, SST prediction relies on kinetic and thermodynamic equations to model the parameters of the physical environment. In recent research, scientists have developed methods for predicting SST based on data-driven statistical methods. As these techniques autonomously learn patterns from historical data and leverage this knowledge for forecasting future SST values, they demand substantial data volumes while relying less on specific insights into oceanic or atmospheric conditions. This class encompasses artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) models, known for their demonstrated accuracy and robustness. Support Vector Machines (SVM) and XGBoost are two ML models that have demonstrated significant ability and enormous potential for weather and ocean parameter

* Corresponding author at: Agricultural Biosystems Engineering, Wageningen University & Research, Wageningen, the Netherlands.
E-mail address: haris.khan@wur.nl (H.A. Khan).

https://doi.org/10.1016/j.seares.2024.102515
Received 5 March 2024; Received in revised form 15 May 2024; Accepted 18 June 2024
Available online 27 June 2024
1385-1101/

prediction.

Recently, DL models have gained a lot of traction because of their capacity to extract features from hidden data and represent intricate relationships. They are extensively employed in a variety of applications, including time-series prediction. The DL models used for SST prediction can be divided into three categories. The first set groups the temporal methods. This refers to techniques that take into account the temporal dependencies and patterns in the time series data. SST data typically involves observations collected over time, and the goal is to predict future SST values based on historical data.

Temporal methods in deep learning for SST prediction often include Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs). These types of neural networks are designed to capture sequential dependencies in time-series data. They have memory mechanisms that allow them to learn and remember patterns over different time steps, making them suitable for modeling temporal relationships. RNNs are a type of neural network architecture that has connections feeding back on themselves, allowing them to maintain a memory of previous inputs. LSTM is a specialized type of RNN designed to address the vanishing gradient problem, which can occur in traditional RNNs when learning long-term dependencies. Zhang et al. (2020) introduced a prediction model for medium- and long-term SST based on the Gated Recurrent Unit (GRU) neural network algorithm. The model leverages the GRU layer to capture the time regularity of SST and produces predicted outcomes through the fully connected layer. The model's practicality and stability are assessed using SST data at various time scales, specifically on a monthly and quarterly basis. Lin and Zhong (2021) proposed an enhanced SST prediction model called Multichannel LSTM (MC-LSTM). In the initial step, they applied a wavelet transform to decompose time-sequence data into distinct sequences representing trend, period, and disturbance. Subsequently, multiple LSTM channels were utilized to concurrently train these decomposed sequences, followed by the aggregation of their prediction results to produce a comprehensive prediction outcome.

The second category concerns spatial methods that take into account the spatial relationships and patterns in the data. Unlike temporal methods, which focus on capturing dependencies over time, spatial methods focus on capturing dependencies across different locations or regions. When dealing with SST prediction, the spatial distribution of temperatures across oceans is essential. DL models that incorporate spatial information aim to understand how SST values vary across different geographical locations and how these variations impact each other.

Convolutional Neural Networks (CNN) are widely used for spatial analysis in computer vision tasks (Moussaid et al., 2023; Belharar and Zrira, 2022; Mouhcine et al., 2024), but they can also be applied to spatial patterns in SST data. These networks use convolutional layers to learn spatial hierarchies and feature representations from input data, making them effective for capturing spatial dependencies. Another technique is the use of Multi-Layer Perceptron (MLP), which is characterized by layers of interconnected nodes and proves versatile for capturing intricate patterns and non-linear relationships within SST data (Wolff et al., 2020).

In the last category, we find spatio-temporal methods. This kind of technique combines both spatial and temporal information to model how SST values evolve over both space and time. These methods are designed to capture the complex interactions and dependencies that exist in both the spatial and temporal dimensions of the data.

Combining LSTMs with spatial components, such as CNNs, allows the model to capture both spatial and temporal dependencies. This hybrid approach is effective for understanding how SST values vary across different locations and how those variations change over time, as claimed by Xiao et al. (2019). In the work (Fei et al., 2022), the authors have adapted the ConvLSTM, initially proposed for precipitation prediction (Shi et al., 2015), to SST prediction. They applied the ConvLSTM to discern the spatio-temporal interconnections within the data, while an attention model was utilized to allocate weights to historical information. The attention mechanism refines the ConvLSTM output by dynamically learning suitable weights at each step. This dual approach ensures precise correction of SST values, contributing to high-precision predictions. Qiao et al. (2023) introduced a learning strategy that merges the Predictive Recurrent Neural Network (PredRNN) architecture with an attention mechanism to enhance the accuracy of SST field predictions. Zhang et al. (2021) proposed a model called the Memory Graph Convolutional Network (MGCN). Comprising two memory layers, one graph layer, and one output layer, the MGCN is designed to capture various aspects of SST dynamics. The memory layers employ temporal convolution units and gate linear units to capture temporal changes in SST. Simultaneously, the graph layer encodes spatial changes by leveraging characteristics derived from the graph Laplacian. The output layer synthesizes information from the preceding layers to generate accurate SST predictions. Other papers used a graph neural network (GNN) by representing the spatial locations with nodes and the spatial relationships with edges (Kim et al., 2023). The temporal dimension is then incorporated to model how these relationships evolve over time. Sun et al. (2021) introduced the Time-Series Graph Network (TSGN). This model adeptly combines the capture of graph-based spatial correlation with temporal dynamics. TSGN employs an LSTM network for feature aggregation in time series data and establishes a GNN model to successfully fulfill the SST prediction task.

Moreover, Bidirectional LSTM (BiLSTM) is a type of RNN architecture that consists of two LSTM layers: one processing the input sequence from left to right (forward LSTM) and the other processing the input sequence from right to left (backward LSTM). When using BiLSTM for prediction tasks, such as time series forecasting or sequence prediction, the model takes advantage of its ability to capture both past and future context to make more accurate predictions.

More recently, Zrira et al. (2024) proposed a new approach to SST prediction based on the BiLSTM model with an attention mechanism. The approach achieved significantly superior prediction outcomes in the Moroccan Sea. In this paper, we extend the work of (Zrira et al., 2024) to also explore representations of spatial patterns and structures learned by the network through the convolutional layer. Analyzing and understanding these spatio-temporal features in SST is essential for various applications, including weather forecasting, climate modeling, and assessing the health of marine ecosystems.

The subsequent sections of this paper are structured as follows. In Section 2, we provide a theoretical background on deep learning concepts. Section 3 presents the proposed BiLSTM with an attention method and a convolution layer for SST prediction. Section 4 details the study area. Section 5 depicts the experimental setting. Section 6 showcases the experimental results. Lastly, Section 7 serves as the conclusion of this paper.

2. Theoretical background

2.1. Long short-term memory

LSTM is a variant of RNN equipped with multiple units. These units, introduced by German researchers Hochreiter and Schmidhuber in 1997 (Hochreiter and Schmidhuber, 1997), were designed to overcome the challenge of the vanishing gradient problem. These networks have demonstrated exceptional efficacy across diverse problem domains and are extensively applied in various applications. Comprising memory blocks housing self-connected memory cells, LSTMs keep the temporal information of the network. LSTMs is composed of several units named "gates", which enable the storing, modification, and retrieval of information within cells. Reading, writing, and resetting gates are controlled by each cell independently. Despite their binary operation—open or closed—these gates utilize element-wise multiplication by the sigmoid function, constraining values between 0 and 1 for differentiability and seamless integration with the backpropagation algorithm. Functioning as information filters, the gates selectively permit or block signals based on relevance and importance, modulated by adjustable weights refined during the learning of the recurrent network, thereby influencing the input and hidden states of the LSTM. Here's a description of this process with equations:

Memory Cell: The fundamental element of an LSTM is the memory cell, denoted as C_t, which serves as a repository for preserving information over prolonged periods. This cell incorporates a self-connection, enabling it to retain its state over time. Mathematically, the memory cell is expressed as:

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

where, f_t is the forget gate output's, i_t is the the input gate output's, and $\widetilde{C}_t$ is the new candidate value.

Forget Gate: The forget gate determines what information from the previous state should be discarded. It is computed using the sigmoid activation function and is given by:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$

where, W_f is the weight matrix for the forget gate, σ is the sigmoid activation function, h_{t-1} is the previous hidden state, x_t is the input at time t, and b_f is the bias term.

Input Gate: The input gate decides which new information should be stored in the memory cell. It is calculated as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

where, W_i is the weight matrix for the input gate, and b_i is the bias term.

Candidate Value: A new candidate value is generated to be added to the memory cell. This is achieved through the hyperbolic tangent activation function:

$$\widetilde{c}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$

where, W_C is the weight matrix for the candidate value, and b_C is the bias term.

Output Gate: The output gate determines the next hidden state based on the current input and the memory cell. It is expressed as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

where, W_o is the weight matrix for the output gate, and b_o is the bias term.

Hidden State: The final hidden state, h_t, is calculated by combining the output gate output (o_t) with the hyperbolic tangent of the memory cell (C_t):

$$h_t = o_t \cdot tanh(C_t) \tag{6}$$

As depicted in Fig. 1, each step of the LSTM involves intricate computations using gates and memory cells, allowing the network to selectively retain, forget, and update information over sequential data, thereby addressing challenges associated with vanishing gradients and long-term dependencies.

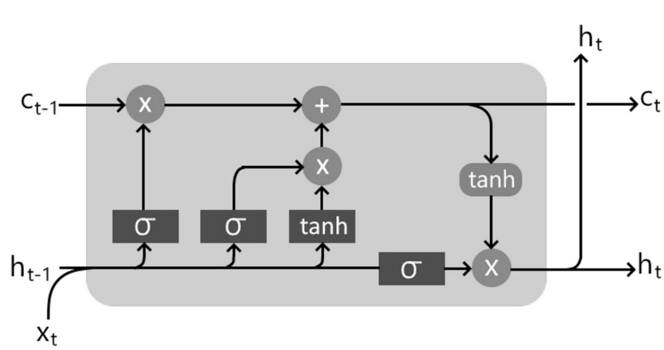

Fig. 1. LSTM architecture.

2.2. Bidirectional long short-term memory

Bidirectional Long Short-Term Memory (Graves and Schmidhuber, 2005) extends the traditional LSTM architecture by handling input data in both forward and backward directions. Unlike a standard LSTM that processes input sequences sequentially from start to end, BisLSTM introduces a layer that processes the sequence in reverse. This bidirectional processing helps capture information from both past and future contexts, enhancing the network's ability to understand and learn dependencies in both directions.

In a BiLSTM network, the hidden state at each time step is determined by considering both the preceding and succeeding elements in the sequence. This bidirectional approach is particularly useful in tasks where context from both directions is crucial. Mathematically, the forward and backward computations in a BiLSTM can be represented as follows:

In the forward pass, the input sequence is processed from the beginning to the end, similar to a conventional LSTM. The hidden state at each time step $\overleftarrow{h_t}$ is computed using the input $\overleftarrow{x_t}$ and the preceding hidden state $\overleftarrow{h_{t-1}}$:

$$\overleftarrow{h_t} = \text{LSTM}\left(\overleftarrow{x_t}, \overleftarrow{h_{t-1}}\right) \tag{7}$$

Conversely, in the backward pass, the input sequence is processed in reverse, allowing the network to consider information from future time steps. The hidden state at each time step $\overrightarrow{h_t}$ is computed using the input in the reversed sequence $\overrightarrow{x_t}$ and the succeeding hidden state $\overrightarrow{h_{t+1}}$:

$$\overrightarrow{h_t} = \text{LSTM}\left(\overrightarrow{x_t}, \overrightarrow{h_{t-1}}\right) \tag{8}$$

The final hidden state h_t is obtained by concatenating the forward and backward hidden states:

$$h_t = \left[\overleftarrow{h_t}, \overrightarrow{h_t}\right] \tag{9}$$

This bidirectional processing of information empowers Bi-LSTM networks to capture dependencies in both directions, making them particularly effective for tasks that demand a comprehensive understanding of sequential data. Fig. 2 shows the BiLSTM architecture.

2.3. Attention mechanism

The attention mechanism (Niu et al., 2021) serves as a crucial element in neural network architectures designed for sequence-to-sequence tasks, such as natural language processing and machine translation. Its functionality lies in enabling the model to prioritize specific segments of the input sequence during predictions, assigning varying degrees of significance to different elements. Particularly in the context of processing sequential data, the attention mechanism significantly enhances the network's capacity to capture long-range dependencies, thereby improving its effectiveness in tasks where certain portions of the input sequence hold more relevance than others.

The attention mechanism works by assigning weights to different elements of the input sequence, indicating their importance in the context of the current prediction. These weights are dynamically adjusted during the training process based on the interaction between the model's current state and the elements of the sequence. Mathematically, let's consider a simple attention mechanism. Given an input sequence $X = \{x_1, x_2, \ldots, x_n\}$ and a hidden state h_t of the model at a particular time step, the attention weights α are calculated as follows:

The final hidden state h_t is obtained by concatenating the forward and backward hidden states:

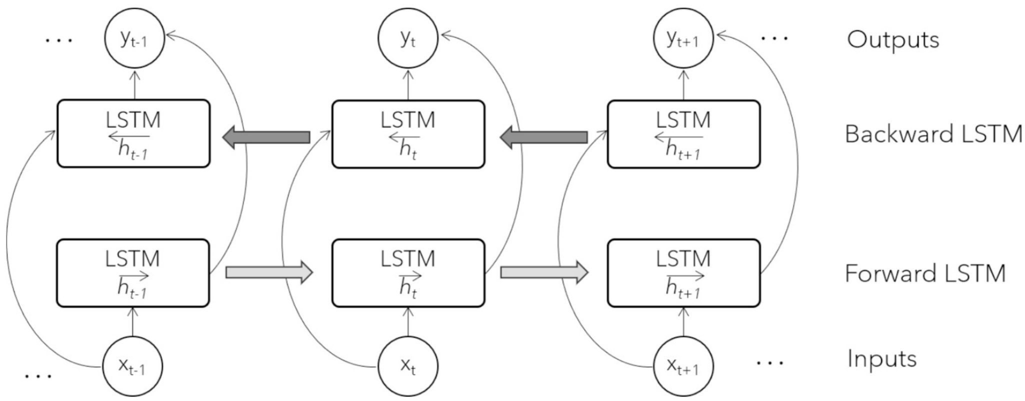

Fig. 2. BiLSTM architecture.

$$\alpha_i = \frac{exp(e_i)}{\sum_{j=1}^{n} exp(e_j)} \tag{10}$$

where e_i is a score associated with the importance of the i-th element. These scores are often computed using a learned function, for example, a feedforward neural network. α_i represents the attention weight for the i-th element, and it is normalized using the softmax function.

The context vector c_t is then calculated as a weighted sum of the input elements:

$$c_t = \sum_{i=1}^{n} \alpha_i \cdot x_i \tag{11}$$

The context vector c_t is used in combination with the current hidden state h_t to make predictions or update the model's internal state.

The attention mechanism thus allows the model to selectively focus on different parts of the input sequence, adapting its attention weights dynamically during training to improve its understanding of the data and enhance performance on complex tasks.

2.4. Convolutional 1D layer

A Convolutional 1D layer, often denoted as Conv1D, is a type of layer used in deep learning models, specifically in the domain of sequence data or one-dimensional data. It is part of convolutional neural networks and is designed to capture patterns, features, and relationships within one-dimensional sequences.

In a Conv1D layer, a set of filters, also known as convolutional kernels, slide or convolve across the input sequence. These filters learn to recognize local patterns, capturing information about the relationships between neighboring elements in the sequence. The convolution operation involves element-wise multiplication and summation, producing a set of feature maps that highlight relevant patterns in the data.

2.5. Prediction measures

To assess and choose the most effective models for time series prediction, we utilize commonly employed evaluation metrics. These metrics are calculated based on the following equations, taking into account x as the actual observation, N as all observations in the testing set, and y as the prediction:

The Mean Absolute Error (MAE) is a measure that signifies the average of absolute errors between the predictions and actual observations for a given set of data points.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i| \tag{12}$$

The Mean Absolute Percentage Error (MAPE) is an indicator that denotes the average of the absolute percentage errors.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_i - y_i}{x_i}\right| \tag{13}$$

The Root Mean Squared Error (RMSE) compute, as indicated by his name, the square root of the mean squared error. This measure is widely employed in the assessment of numerical predictions.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|^2} \tag{14}$$

The R-squared (R^2), also known as the coefficient of determination, offers more comprehensive information compared to RMSE, MAE, and MAPE metrics. R^2 represents the portion of the variability in a dependent variable that is explained by an independent variable.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}|x_i - y_i|^2}{\sum_{i=1}^{N}|x_i - \overline{x_i}|^2} \tag{15}$$

where $\overline{x_i}$ represents the mean of x observations.

3. Proposed model STA-SST

In this paper, we propose the STA-SST model to solve the problem of SST prediction. The whole approach is depicted in Fig. 3, which consists of three main phases: extracting bidirectional temporal features, extracting spatial correlation of SST data, and finally capturing important information. The main idea is to use the BiLSTM architecture to capture dependencies across the entire SST sequence. This allows the model to better understand the context surrounding each data point in the sequence. In time series analysis, having access to future information can be especially useful for making predictions, as it allows the model to incorporate information about upcoming trends or patterns. BiLSTMs can be easily integrated into various architectures and combined with other types of layers, such as convolutional layers for extracting spatial information or attention mechanisms for focusing on specific parts of the sequence. By incorporating attention mechanisms, models can effectively leverage spatio-temporal information to improve performance on tasks involving SST data. Moreover, dropout and batch normalization layers are added to the STA-SST to improve the training and generalization performance of our model, including those applied to time series data with BiLSTM networks. Dropout helps prevent overfitting by randomly dropping neurons during training, while batch normalization stabilizes training and accelerates convergence by normalizing the inputs of each layer. Integrating these techniques into BiLSTM architectures can lead to more robust and effective models for analyzing SST data. Finally, a dense layer, also known as a fully connected layer, is used to connect the input from every neuron in the previous attention layer, which is combined with a set of weights and then passed through

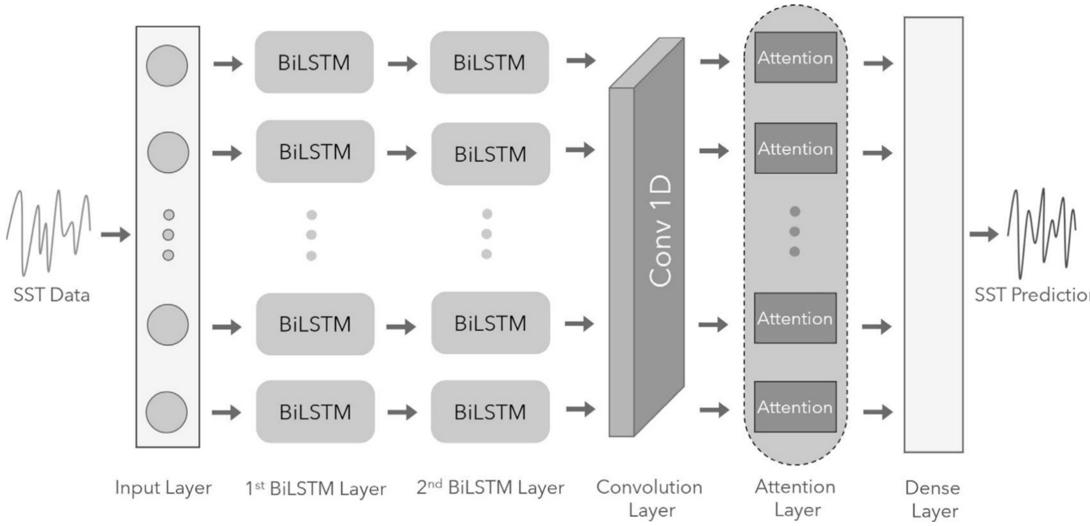

Fig. 3. Our proposed model STA-SST.

an activation function to produce the output of each neuron in the dense layer. Table 1 summarizes the output shape of each layer in the STA-SST model.

4. Dataset

The methodology involves two principal phases: data preparation and network training. Initially, a study region is chosen for predicting SST values, and the time series data is subjected to potential issues such as missing values or outliers. To address these challenges, a pre-processing step is employed. Additionally, normalization of the time series data is carried out to enhance training stability and the overall performance of the STA-SST model. Following this, the processed data is divided into two sets: training and testing sets.

4.1. Study sites

Morocco, situated in the northwestern part of Africa, extends its seafront to a part of the Mediterranean Sea to the north and the Atlantic Ocean to the west. This study focuses on predicting Sea Surface Temperature (SST) values in five distinct Moroccan cities: Nador, Martil, Tangier, Rabat, and Taghazout. The selection of these cities, depicted in Fig. 4, is grounded in their proximity to the sea.

Nador is a coastal city located in the northeastern part of Morocco, known for its strategic position along the Mediterranean Sea. With its picturesque coastal landscapes, including the stunning Mar Chica lagoon, Nador attracts those seeking scenic beauty and recreational activities. Martil, situated in northern Morocco and nestled along the shores of the Mediterranean Sea, is renowned for its pristine beaches and crystal-clear waters. The city's coastal context not only contributes to its aesthetic allure but also plays a crucial role in supporting activities like fishing and water sports, adding vitality to the local economy. Martil, with its blend of natural beauty and cultural richness, stands as a coastal *gem* on Morocco's northern shores. Tangier, resting on the northern coast of Morocco, overlooks the strategic Strait of Gibraltar. As one gazes upon the Atlantic Ocean and the Mediterranean Sea merging at this maritime crossroads, Tangier unveils its unique maritime identity. Rabat, Morocco's capital city, is situated along the Atlantic Ocean, its maritime context is defined by the scenic Bouregreg River estuary, which flows into the ocean, shaping the city's coastal landscape. The city's waterfront, marked by the Oudaya Gardens and the Hassan Tower, offers stunning views of the Atlantic. Taghazout is a village along the southwest coast of Morocco. Situated within the Agadir province, it is located 20 km away from the city of Agadir. Nestled between the Atlantic Ocean and the rugged cliffs of the Atlas Mountains, Taghazout is renowned for its world-class surfing spots and the fishing boats that dot the horizon.

The sea surface temperature measurements utilized in this study are sourced from the National Oceanic and Atmospheric Administration (NOAA), covering the period from January 1, 2012, to July 9, 2023. Table 2 shows key statistics related to the SST data, encompassing site coordinates, mean and standard deviation measures, and minimum and maximum temperatures. The selection of these five cities is informed by statistically similar characteristics across most cities. In the realm of deep learning, utilizing data with comparable statistics proves advantageous for constructing models with robust generalization. Models commonly demonstrate improved performance when utilized on the novel, unseen data that conforms to a similar distribution. Furthermore, ensuring the alignment of statistical properties during the division of data into training and testing sets contributes to the creation of dependable models and simplifies accurate performance evaluation.

Table 1
STA-STT architecture.

Layer (type)	Output Shape
Bidirectional	(3593, 60, 256)
Dropout	(3593, 60, 256)
Bidirectional	(3593, 60, 128)
Dropout	(3593, 60, 128)
Conv1D	(3593, 60, 64)
Attention	(3593, 60, 64)
BatchNormalization	(3593, 60, 64)
Flatten	(3593, 3840)
Dense	(3593, 1)

4.2. Data preprocessing

Creating an accurate predictive model for time series necessitates having comprehensive data without any missing or outlier values. Therefore, it is necessary to execute a relevant preprocessing phase before deploying the model. In this work, we use standard preprocessing techniques to improve data quality, incorporating conventional imputation to handle missing values. Furthermore, Normalizing time series data is indispensable owing to its diverse measurement scales. The methods based on deep learning aim to establish a correspondence between input data and output across all samples within the training set. In fact, model Weights are initially set randomly and subsequently updated through optimization approaches such as Root Mean Square Propaga-

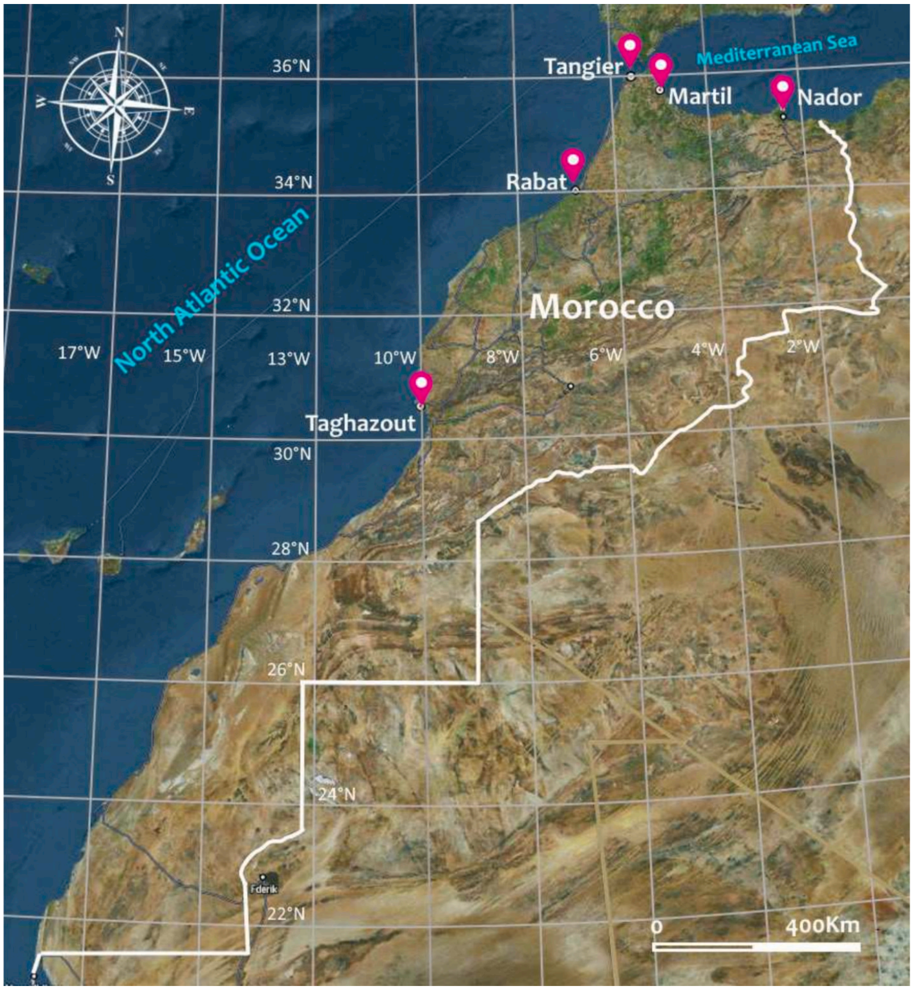

Fig. 4. Moroccan map with chosen cities localization.

Table 2
SST statistics.

City	Coordinates	Mean	Max	Min	STD
Tangier	35.76727,-5.79975	18.329208	24.19	14.62	2.226597
Taghazout	30.54500,-9.70863	18.423433	23.82	14.5	1.79698
Rabat	34.01325,-6.83255	19.582309	24.7	15.17	2.367962
Martil	35.61666,-5.26666	18.537168	25.1	14.41	2.516966
Nador	35.16813,-2.93352	19.684396	28.55	13.78	3.774163

tion (RMSProp). Failure to appropriately scale time series data for prediction can result in an erratic learning process. To address this, we employ a normalization method based on the minimum and maximum values of the data, transforming time series information within the range (0,1). Eq. 16 illustrates the scaling equation:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{16}$$

where x represents the initial value, x_{min} and x_{max} are, respectively, the minimum and the maximum values.

4.3. Dataset partitioning

The temporal data employed in this investigation spans approximately 12 years, precisely from January 1, 2012, to July 9, 2023, covering a total of 4207 days. The time series data is partitioned into two subsets: training and testing. The data allocated for training extends from January 1, 2012, to December 31, 2022 (totaling 4018 days). The remaining dataset, comprising 189 days, is set aside to assess the model's performance during testing. Fig. 5 demonstrates a visual representation of the data distribution for each region in the study area. It delineates the training and testing datasets.

5. Experimental setting

To gauge the efficiency of the spatio-temporal attention method proposed, we conducted a benchmark comparison against four traditional methods. This benchmark comprises two conventional machine learning models, namely Random Forest (RF) and Support Vector Regression (SVR). Additionally, we incorporate a gradient-boosting technique known as Extreme Gradient Boosting (XGBoost). The benchmark also features a deep learning method explicitly designed for time series data such as LSTM. We execute all experiments with the same configurations to guarantee an unbiased comparison. Using the same testing set and experimental setting allows for a direct comparison of results, facilitating the evaluation of the STA-SST method's performance.

5.1. Experiment configuration

The experimental study was carried out in Google Colab, employing the Attention frameworks, Keras, and TensorFlow. Both training and testing phases are executed within the Python 3.5 programming environment on GPUs T4 and P100.

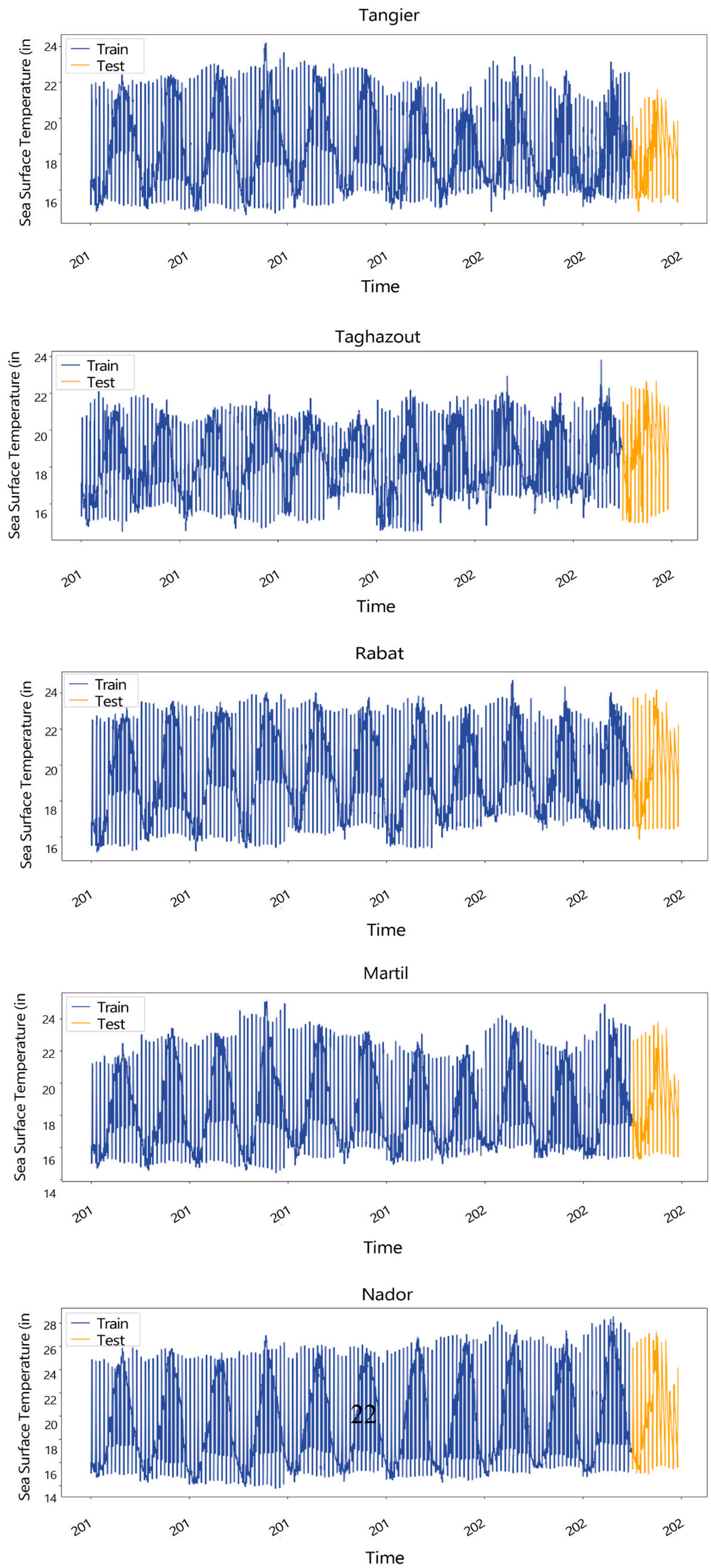

Fig. 5. Training and testing data partitioning.

The model combining Attention and BiLSTM underwent training for around 200 epochs, incorporating a stopping mechanism to mitigate overfitting. Model optimization utilizes the Adam technique with a mini-batch size of 32, and 0.001 as the starting learning rate.

Table 3 outlines the hyperparameters employed during the training phase of deep learning methods. These hyperparameters encompass the batch size, the count of LSTM units, the learning rate, and the dropout rate. The selection of these hyperparameters was driven by a process of experimentation and fine-tuning, strategically chosen to optimize the performance of both the LSTM and STA-STT.

Table 4 offers an overview of the optimal parameters employed in training the SVR, XGBoost, and RF models. The used parameters are discerned through a rigorous experimentation study and strategically selected to attain the best achievement score in each model.

5.2. Time step

In the field of RNN, the concept of "time step" refers to the count of previous observations or time intervals considered when predicting the next step within a sequence. This parameter holds significance as a hyperparameter and may vary depending on the inherent patterns and relationships within the data. To identify the optimal time step, we conducted a series of experiments, evaluating various samples of time steps to ascertain the most effective one. The assessment involved calculating the metrics explained previously. These experiments were specifically applied to Taghazout data, and the results are summarized in Table 5. The highest R^2 value is considered optimal, while the lowest values for MAPE, MAE, and RMSE indicate superior performance. Our experiments revealed that employing a 60-time step configuration yielded the best results. This choice aligns with the ability of two months to capture changes associated with seasonal variations in Sea Surface Temperature. Monitoring SST changes over this time step facilitates a comprehensive understanding of temperature fluctuations during seasonal transitions, which is important for deciphering SST behaviors. Therefore, we choose 60 as the optimal time step for this experiment.

6. Experimental results

The experimental study is divided into two phases. Initially, we assess the effectiveness of employing spatio-temporal methods coupled with attention mechanisms in contrast to conventional methods. Subsequently, in the second phase, we conduct a comparative analysis of our approach against various existing methods arising from diverse theoretical backgrounds.

6.1. Ablation study

To validate the effectiveness of our proposed approach, we explore various scenarios:

- Spatial: Conv1D only;

Table 3
Hyperparameters preserved throughout the STA-SST training process.

Hyperparameter	Value
Input_shape	(3593, 60, 1)
BiLSTM_units	256–128
Kernels	64
Kernel_initializer	Glorot uniform
Batch_size	32
Learning_rate	0.0001
Optimizer	ADAM
Epochs	200
Loss_function	MSE
Dropout_rate	0.2

Table 4
Parameters using in XGBoost, RF, and SVR training phase.

Model	Parameter	Value
RF	n_estimators	100
	criterion	absolute_error
XGBoost	max_depth	3
	learning_rate	0.1
	n_estimators	100
SVR	degree	1
	kernel	linear

Table 5
The predicting performance on Taghazout for various time step samples.

Days	MAE	MAPE	RMSE	R^2
1	0.847029	0.046929	0.931680	0.783537
7	0.393499	0.020563	0.528211	0.930423
30	0.395406	0.021217	0.511922	0.934648
60	**0.331507**	**0.017621**	**0.444958**	**0.950627**
90	1.077337	0.058461	1.153988	0.667912

- Spatial and Attention: Conv1D and attention;
- Temporal: BiLSTM only;
- Temporal and Attetion: BiLSTM and attention;
- Spatio-temporal: Conv1D and BiLSTM;
- Spatio-temporal and Attention: Conv1D, BiLSTM, and attention.

From the results shown in Table 6 and Fig. 6, we can deduce that incorporating spatial information yields significant advantages. Indeed, by integrating geographical data and spatial characteristics, predictive models gain a more comprehensive understanding of the intricate relationships between temperature variations and geographic locations. This inclusion enables the model to discern localized patterns. Integrating an attention mechanism into models confers notable advantages. The attention mechanism allows the model to dynamically focus on specific regions or patterns within the SST data, assigning varying degrees of importance to different temporal and spatial patterns. This adaptability is particularly advantageous in scenarios where certain features may have a more significant impact on SST variations. The integration of spatial and temporal information allows the model to capture both the geographical influences on SST. The attention mechanism further refines this process by dynamically allocating weights to specific spatial and temporal components, emphasizing their significance during prediction. This dual approach ensures that the model not only considers the broader context of SST dynamics but also zooms in on crucial details. (See Fig. 7.)

6.2. Comparative study

It's essential to note that we employ four evaluation metrics to assess and compare the performance of the competing methods in this study: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R2.

Tables 8, 10, 9,7, and 11 offer a comprehensive assessment of various models' performance across diverse regions in the test dataset. Notably,

Table 6
Ablation study on Taghazout.

Models	MAE	MAPE	RMSE	R^2
Spatial	0.329419	0.017522	0.434195	0.952987
Spatial and Attention	0.317757	0.017192	0.423555	0.955263
Temporal	0.364238	0.019240	0.485684	0.941176
Temporal and Attention	0.319338	0.017179	0.424272	0.955111
Spatio-Temporal	0.332824	0.017835	0.430741	0.953732
Spatio-Temporal and Attention	**0.314188**	**0.016826**	**0.420749**	**0.955854**

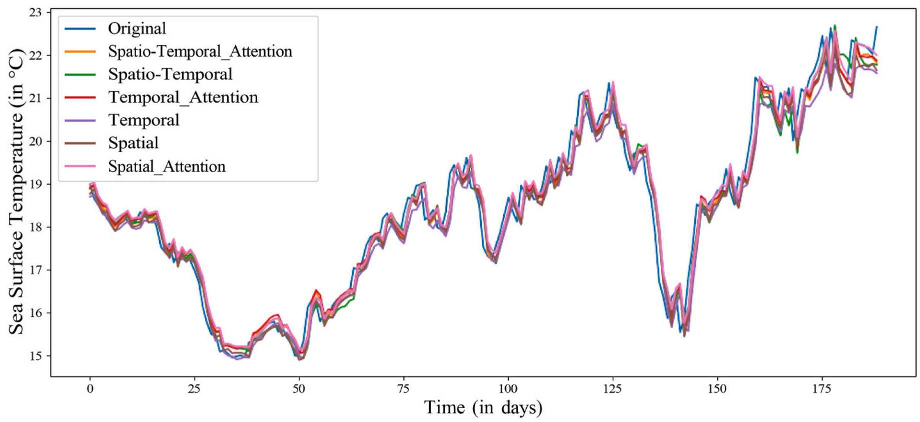

Fig. 6. The ablation study in Taghazout.

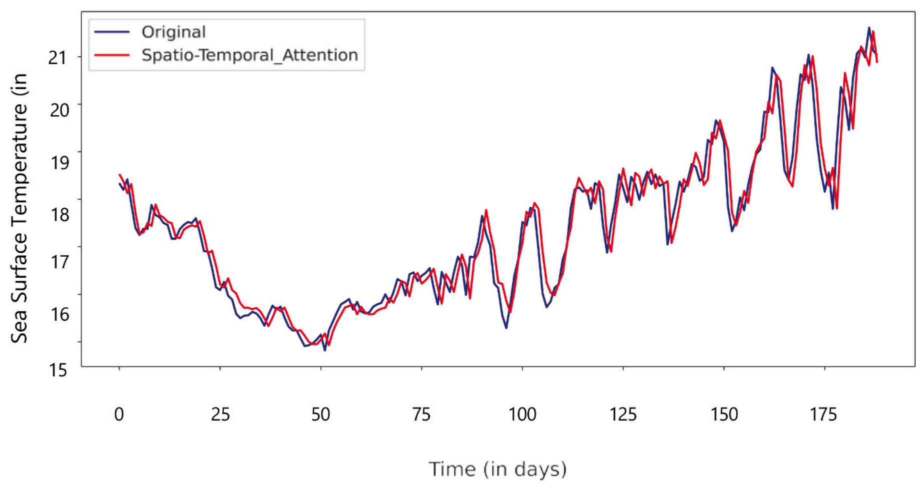

Fig. 7. The forecasting of SST in Tangier.

Table 7
The forecasting precision in Rabat.

Model	MAE	MAPE	RMSE	R^2
XGBoost	0.200578	0.010469	0.268218	0.985756
RF	0.208833	0.010906	0.281054	0.984580
SVR	0.196725	0.010205	0.270641	0.985601
LSTM	0.235194	0.012324	0.306578	0.981053
STA	**0.199933**	**0.010441**	**0.264873**	**0.985857**

the RF model consistently exhibits comparatively inferior performance, displaying elevated error values and a reduced determination coefficient. In contrast, the STA-SST model consistently achieves the highest R^2 across all Moroccan test regions, underscoring its capability to fit SST observations. The Nador region particularly excels with an R^2 of 0.989034, indicative of remarkably accurate sea temperature predictions. Despite the computational intensity and time investment in training deep learning models, their efficiency becomes evident when it's employed for inference.

Figs. 8, 9, 10,11, and 12 present a visual representation of the forecasted SST values for the Nador, Martil, Tangier, Rabat, and Taghazout regions. These graphs serve as evidence of the STA-SST's ability to predict SST values, closely aligning with the real observations. The proposed model's predictive performance is notably robust, demonstrating remarkable accuracy across diverse regions within the test dataset. Indeed, the plots showcase the close correspondence between the forecasted and actual SST values, underscoring the model's reliability. This alignment across different regions further emphasizes the generalizability and effectiveness of the proposed STA-SST approach.

The visual insights provided by these figures strengthen our confidence in the model's ability to capture intricate SST patterns across various geographical locations, validating its utility for comprehensive sea temperature predictions.

Table 8
The forecasting precision in Nador.

Model	MAE	MAPE	RMSE	R^2
XGBoost	0.250927	0.012165	0.362656	0.988475
RF	0.261508	0.012636	0.374882	0.987684
SVR	0.252743	0.012187	0.360607	0.988604
LSTM	0.258472	0.012648	0.358683	0.988578
STA	**0.248751**	**0.012214**	**0.351461**	**0.989034**

Table 9
The forecasting precision in Tangier.

Model	MAE	MAPE	RMSE	R^2
XGBoost	0.317871	0.017927	0.425660	0.930201
RF	0.339133	0.019120	0.450237	0.921908
SVR	0.315809	0.017772	0.422599	0.931201
LSTM	0.315732	0.017825	0.412877	0.932604
STA	**0.313230**	**0.017727**	**0.411709**	**0.932985**

Table 10
The forecasting precision in Martil.

Model	MAE	MAPE	RMSE	R^2
XGBoost	0.246239	0.013307	0.338883	0.973984
RF	0.250376	0.013589	0.342840	0.973373
SVR	0.244631	0.013240	0.339299	0.973920
LSTM	0.235010	0.012794	0.324478	0.974988
STA	**0.238833**	**0.012995**	**0.323111**	**0.975595**

Table 11
The forecasting precision in Taghazout.

Model	MAE	MAPE	RMSE	R^2
XGBoost	0.323082	0.017242	0.431248	0.955063
RF	0.320332	0.017114	0.444946	0.952163
SVR	0.314611	0.016826	0.424675	0.946423
LSTM	0.318463	0.017028	0.425976	0.954750
STA	**0.314188**	**0.016826**	**0.420749**	**0.955854**

6.3. Discussion

The integration of the attention mechanism within the model stands out as a factor contributing to enhanced predictions, granting the model the ability to selectively focus on crucial features. In comparison to ordinary models, the STA-SST model emerges as an exceptional solution for the accurate prediction of SST values. The comprehensive evaluation, encompassing experimental outcomes, visualizations, and comparative analyses, collectively underscores the effectiveness of the proposed STA-SST method across diverse regions in Morocco. In fact, the experimental results affirm the reliability and efficacy of the proposed model in delivering precise SST predictions for varied geographical regions.

When integrated into a larger neural network model, the Conv1D layer, along with other components like recurrent layers and attention mechanisms, contribute to the model's capacity to understand and forecast the complex dynamics of SST. Indeed, the use of spatio-temporal methods, particularly the incorporation of the attention mechanism and Conv1D in our model, has showcased a significant enhancement in prediction when compared to conventional approaches. The inherent strength of spatio-temporal methods lies in their ability to concurrently capture both the spatial and temporal intricacies inherent in SST dynamics. Unlike singularly spatial or temporal methods, the combination of spatio-temporal modeling facilitates a more holistic understanding of how SST values evolve both across different locations and over time. The attention mechanism acts as a discerning lens, dynamically focusing on crucial spatial and temporal elements during the prediction process. This dual perspective not only enhances the model's precision but also enables it to discern complex patterns and dependencies that would elude traditional spatial or temporal methods.

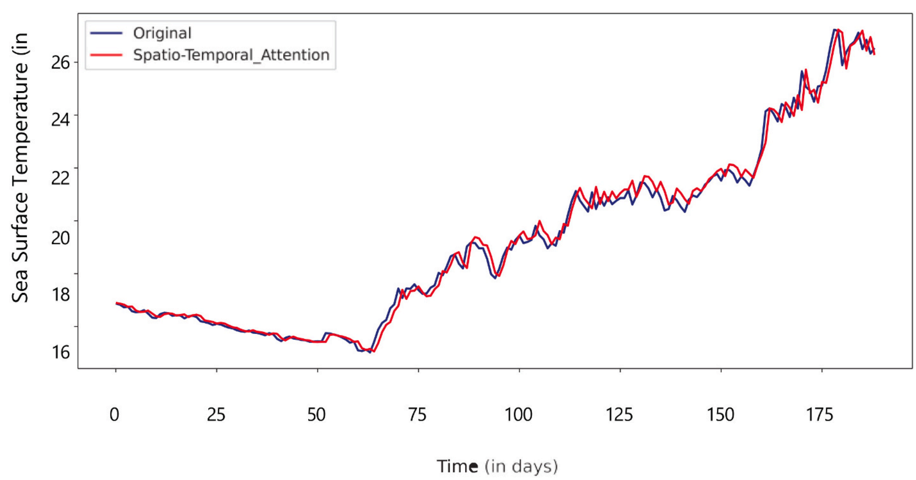

Fig. 8. The forecasting of SST in Nador.

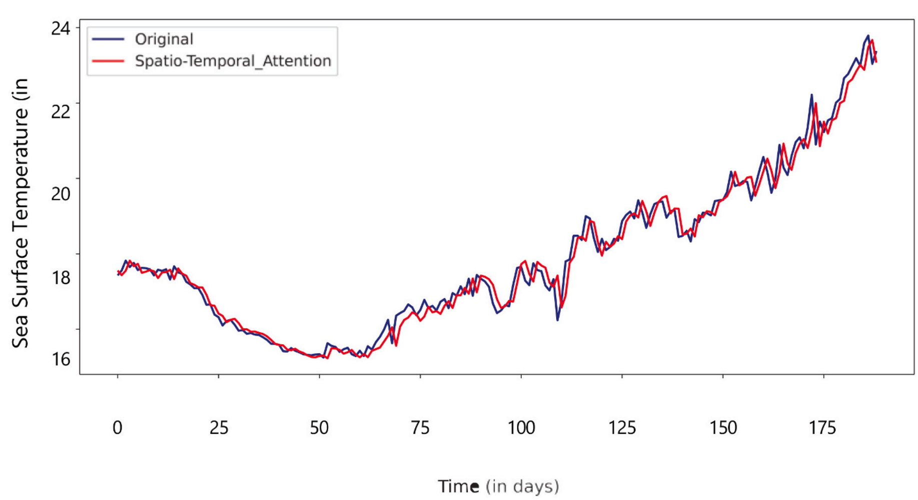

Fig. 9. The forecasting of SST in Martil.

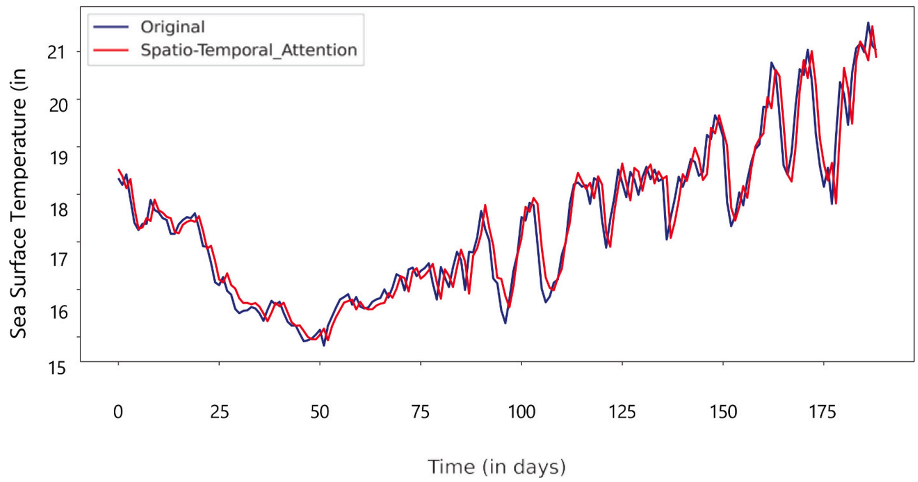

Fig. 10. The forecasting of SST in Tangier.

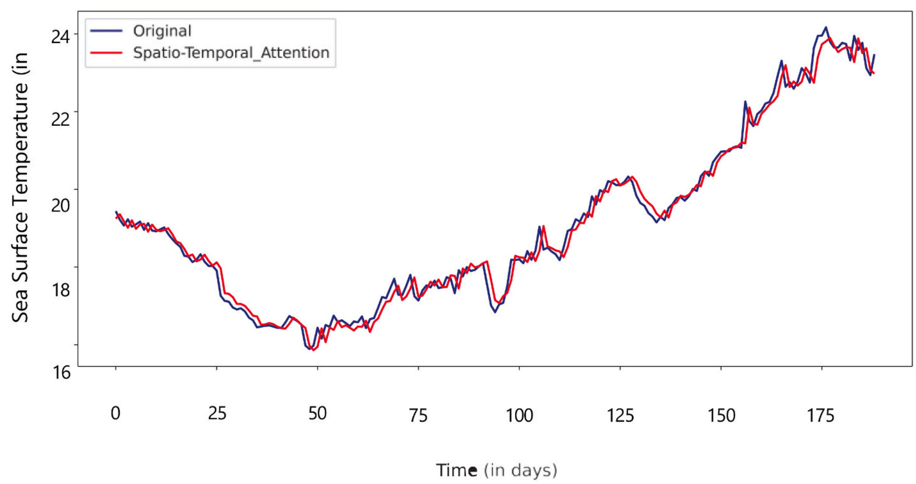

Fig. 11. The forecasting of SST in Rabat.

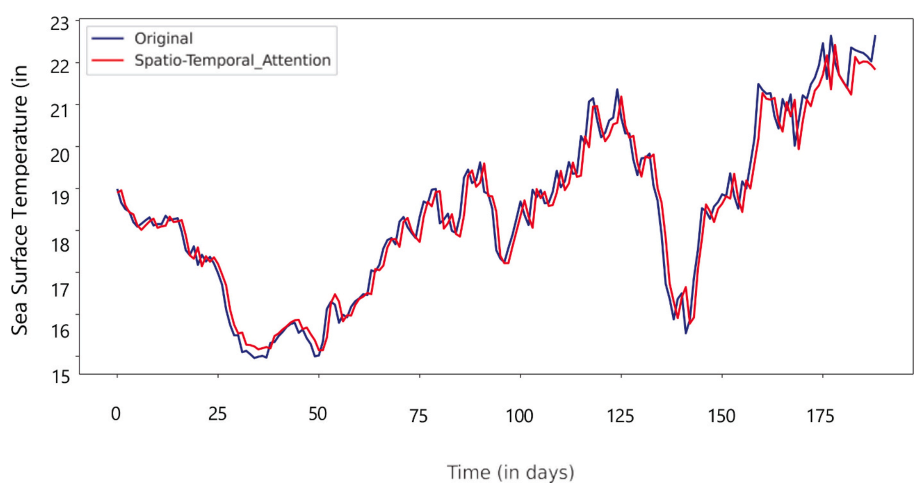

Fig. 12. The forecasting of SST in Taghazout.

The results from the experiments and the comparative analysis highlight the effectiveness of the STA-SST approach. These findings emphasize the proposed model as a robust alternative to traditional spatial and temporal methods.

Moreover, the STA-SST operates autonomously, irrespective of the specific seas chosen for testing. It relies only on the input data provided to the model for predicting SST. The proposed model transcends geographical constraints, offering a versatile framework applicable across diverse marine environments. In other words, the model does not need to know the specific characteristics of each sea or marine region to make accurate predictions. Consequently, it delivers robust and accurate SST forecasts across various marine ecosystems without necessitating extensive model adaptations. This adaptability underscores its potential for widespread deployment in oceanographic research and environmental monitoring endeavors. Moroccan seas serve as a mere illustration of the model's application.

While the STA-SST model exhibits significant strengths in capturing intricate patterns in SST variations, it is not without limitations. One limitation is the potential computational complexity associated with attention mechanisms, making the training process resource-intensive. Additionally, interpreting the attention weights and understanding the specific features the model focuses on can be challenging, introducing an element of opacity to the decision-making process. Furthermore, the effectiveness of these models may be influenced by the availability and quality of data, particularly in regions with sparse or incomplete observations. Striking a balance between model complexity and interpretability remains a challenge, as more complex models may achieve higher accuracy at the cost of increased complexity. Finally, climate change introduces dynamic shifts in temperature patterns, sea currents, and atmospheric conditions, posing challenges for models trained on historical data. The increased frequency and intensity of extreme weather events may result in non-linear changes that models may struggle to capture.

7. Conclusion

This paper introduces a novel approach called STA-SST to sea surface temperature prediction by leveraging deep learning models. Our approach integrates the bidirectional long short-term memory model, coupled with an attention mechanism and a convolution layer to incorporate spatial information. The attention mechanism significantly contributes to improving forecasting performances by grasping intricate associations among historical, spatial, and forthcoming SST values. Thorough experimentation provides evidence that our proposed model outperforms current SST prediction methods.

In our coming studies, we intend to investigate the utilization of the neural prophet model with Moroccan SST data and a diverse information set. The implementation of the neural prophet model aims to increase our comprehension of SST prediction. Such efforts advance marine research, enabling more accurate forecasting across diverse maritime areas.

CRediT authorship contribution statement

Isam Elafi: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Nabila Zrira:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Formal analysis, Conceptualization. **Assia Kamal-Idrissi:** Visualization, Validation, Software, Formal analysis, Data curation. **Haris Ahmad Khan:** Writing – review & editing, Writing – original draft, Methodology. **Aziz Ettouhami:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The authors express gratitude to the staff of the website Seatemperature for supplying observation data on the marine environment in Morocco https://seatemperature.info.

References

Ahn, Spencer, Dowd, James, 2023. On uncertainty in extreme weather forecasts in a changing climate: august 2022 floods in South Korea. J. Stud. Res. 12 (1).

Belharar, Fatima Zahra, Zrira, Nabila, 2022. Deepretino: Ophthalmic disease classification from retinal images using deep learning. In: 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). IEEE, pp. 392–399.

Fei, Tonghan, Huang, Binghu, Wang, Xiang, Zhu, Junxing, Chen, Yan, Wang, Huizan, Zhang, Weimin, 2022. A hybrid deep learning model for the bias correction of sst numerical forecast products using satellite data. Remote Sens. 14 (6), 1339.

Graves, Alex, Schmidhuber, Jürgen, 2005. Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Netw. 18 (5–6), 602–610.

Hochreiter, Sepp, Schmidhuber, Jürgen, 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Horton, Thomas W., Birch, Samantha, Block, Barbara A., Hawkes, Lucy A., van der Kooij, Jeroen, Witt, Matthew J., Righton, David, 2024. Maximising the value of transmitted data from psats tracking marine fish: a case study on Atlantic bluefin tuna. Animal Biotelemet. 12 (1), 2.

Khaled, Mostafa A., Al-Jamali, Fahad H., Said, Rashad E.M., Mohammad, Ashraf S., Ahmed, Mahmoud H., Ahmed, Hamdy Omar, 2023. Analysis of sst and chl-a data from modis-aqua in the major egyptian fishing zones of the red sea. Egypt. J. Aquat. Res. 49 (4), 520–529.

Kim, Jinah, Kim, Taekyung, Ryu, Joon-Gyu, Kim, Jaeil, 2023. Spatiotemporal graph neural network for multivariate multi-step ahead time-series forecasting of sea temperature. Eng. Appl. Artif. Intell. 126, 106854.

Lin, Yidi, Zhong, Guoqiang, 2021. A multi-channel lstm model for sea surface temperature prediction. In: Journal of Physics: Conference Series, vol. 1880. IOP Publishing, p. 012029.

Mediodia, Hanny John, Kahui, Viktoria, Noy, Ilan, 2023. Sea surface temperature and tuna catch in the eastern pacific ocean under climate change. Mar. Resour. Econ. 38 (4), 329–351.

Mouhcine, Kaoutar, Zrira, Nabila, Elafi, Issam, Benmiloud, Ibtissam, Khan, Haris Ahmad, 2024. Steff: Spatio-temporal efficientnet for dynamic texture classification in outdoor scenes. Heliyon. https://doi.org/10.1016/j.heliyon.2024.e25360. ISSN 2405-8440.

Moussaid, Abdelghani, Zrira, Nabila, Benmiloud, Ibtissam, Farahat, Zineb, Karmoun, Youssef, Benzidia, Yasmine, Mouline, Soumaya, El Abdi, Bahia, Bourkadi, Jamal Eddine, Ngote, Nabil, 2023. On the implementation of a post-pandemic deep learning algorithm based on a hybrid ct-scan/x-ray images classification applied to pneumonia categories. In: Healthcare, vol. 11. MDPI, p. 662.

Niu, Zhaoyang, Zhong, Guoqiang, Hui, Yu., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62.

Qian, Longxia, Li, Hanlin, Hong, Mei, Qi, Yuxiang, Guo, Zilong, 2023. Avoiding sudden maritime risk: a new variable speed route planning model by integrating spatio-temporal dimension. Ocean Eng. 288, 115950.

Qiao, Baiyou, Zhongqiang, Wu, Ma, Ling, Zhou, Yicheng, Sun, Yunjiao, 2023. Effective ensemble learning approach for sst field prediction using attention-based predrnn. Front. Comp. Sci. 17 (1), 171601.

Shi, Xingjian, Chen, Zhourong, Wang, Hao, Yeung, Dit-Yan, Wong, Wai-Kin, Woo, Wang-chun, 2015. Convolutional lstm network: a machine learning approach for precipitation nowcasting. Adv. Neural Inf. Proces. Syst. 28.

Sun, Yongjiao, Yao, Xin, Bi, Xin, Huang, Xuechun, Zhao, Xiangguo, Qiao, Baiyou, 2021. Time-series graph network for sea surface temperature prediction. Big Data Res. 25, 100237.

Wolff, Stefan, O'Donncha, Fearghal, Chen, Bei, 2020. Statistical and machine learning ensemble modelling to forecast sea surface temperature. J. Mar. Syst. 208, 103347.

Xiao, Changjiang, Chen, Nengcheng, Chuli, Hu, Wang, Ke, Zewei, Xu, Cai, Yaping, Lei, Xu, Chen, Zeqiang, Gong, Jianya, 2019. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. Environ. Model Softw. 120, 104502.

Zhang, Xiaoyu, Li, Yongqing, Frery, Alejandro C., Ren, Peng, 2021. Sea surface temperature prediction with memory graph convolutional networks. IEEE Geosci. Remote Sens. Lett. 19, 1–5.

Zhang, Zhen, Pan, Xinliang, Jiang, Tao, Sui, Baikai, Liu, Chenxi, Sun, Weifu, 2020. Monthly and quarterly sea surface temperature prediction based on gated recurrent unit neural network. J. Marine Sci. Eng. 8 (4), 249.

Zrira, Nabila, Kamal-Idrissi, Assia, Farssi, Rahma, Khan, Haris Ahmad, 2024. Time series prediction of sea surface temperature based on bilstm model with attention mechanism. J. Sea Res. 102472.