An Entity of the Ocean Data Prediction System Using Machine Learning Techniques

S.Oviya[1*], Dr.T.Ananth Kumar[2], Dr.D.Jessintha[3], Dr.R.Rajmohan[4], N.Padmapriya[5] & Dr.S.Arunmozhi selvi[6]

[1,2,5]*Department of Computer Science and Engineering, IFET College of Engineering, Tamilnadu, India.*
[3]*Department of Electronics and Communication Engineering, Easwari Engineering College, Tamilnadu, India.*
[4]*School of Computer Science and Engineering, VIT-AP University, Amaravati, India.*
[6]*Principal, Holy Cross Engineering College, Tamilnadu, India.*
*Corresponding Author (S.Oviya) Email: oviyakoliyanur@gmail.com**

ABSTRACT

The tropical cyclone is one of the most powerful and destructive meteorological systems on Earth. Researchers note tropical cyclone data every few seconds, but utilizing all of the data with the appropriate accuracy values is difficult. In this system, we predict the various elements' status accuracy and loss in the ocean data set. The use of machine learning methods has developed a lot, and the prediction of the value of the ocean data follows the new enhanced term to give the status of the elements in the data. The LSTM (long short-term memory neural network excavation model) of the historical track's helpful information is more profound and more precise. Bi-LSTM goes the both forward and backward directions, and Adam optimizer, two updated machine learning techniques, are utilized to assess the status of the ocean element in the data set. It goes beyond the existing system to offer an opportunity for a different system result. The data set with a large number of values will also perform accurately. The project's ultimate objective is to give oceanographers a tool to anticipate the quality of ocean data in real-time, which can increase the precision of climate models and help with improved ocean-related decision-making.

Keywords: Machine Learning; Ocean DATA; Bi-LSTM; Adam Optimizer.

1. Introduction

With advances in technology and data processing capabilities, the amount of marine data available is increasing rapidly. This presents new opportunities for research and innovation and new challenges for data management and analysis. Effective use of marine data requires collaboration between scientists, policymakers, and stakeholders from different sectors to ensure that the information is used to benefit society and protect the ocean environment [1]. Ocean data is data collected from the oceans and other marine environments. It includes observations from the sea surface, subsurface, and bottom and data from satellites, buoys, tags, and other instruments placed in the oceans. Examples of ocean data include temperature and salinity readings, current speeds and directions, wave heights and spectra, chemical concentrations in ocean water, ocean floor topography and maps, and satellite imagery [2].

Ocean data is used in various fields, from climate research and marine biology to search and rescue operations and fishing management. It needs the proper data status accuracy and loss prediction for the best climate weather prediction as it performs according to the feasibility of the input data. The significant development of machine learning (ML) over the last few decades has expanded the use of this data-driven methodology in research, business, and industry. There has recently been an increase in interest in using ML to address data quality and precision issues.

The Bi-LSTM Neural network in particular, has been utilized with ML methods to create a functional system [3]. Accurate and the loss value gained from this system will enhance the weather prediction process. Recurrent neural network (RNN) architectures such as the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm process

sequential input. Bi-LSTM models concurrently process data in both ways, unlike conventional LSTM models, which only process data in one direction (either from the past to the future or vice versa [4]. This makes it possible for the model to include data from both historical and upcoming inputs, improving predictions. Typically, bi-LSTM models have two independent LSTM layers, one of which processes the input data forward and the other of which processes the input data backward. demonstrate the degree of comparison between the BI-LSTM and LSTM, demonstrating that the BI-LSTM performs better than the LSTM. As a result, the Bi-major LSTM's accuracy methods are applied to the intake data, and the Adam optimization is combined along with this system to maintain the epoch performance [5]. The system recognizes and displays the value concisely for the numerous data types found under the ocean. The data are divided into two primary devisors, test and train, and throughout the prediction process, they provide values for precision such as accuracy, precision, sensitivity, specificity, and F-Score for both accuracy and loss.

2. Related Works

Researchers have created a numerical, deep-learning version using real information for SST and SWH forecasting. Finally, we proposed using four evaluation criteria to examine these three prediction strategies [6]. Experimental data showed that the two techniques much improved on the statistical prediction model, but the deep learning model only slightly exceeded the various types of machine learning models. Ranran, Luo, and others examine how machine learning is used in several applications, such as ocean component prediction, source identification and localization, and tracking of deep-sea resource availability. Here is a summary of current research and particular machine-learning techniques used on ocean data [7]. It discusses some of the study's shortcomings, possible uses, and future directions.

A machine-learning-based prediction approach are provided and validated using simulated data from the two-dimensionally complicated Ginzburg-Landau equation and real wind speed data from the North Atlantic Ocean [8]. Discussed are the trade-offs between forecast accuracy, spatial resolution, and prediction horizon, as well as how catastrophic event incidence that is geographically skewed impacts forecast accuracy. The deep learning architecture allows for predicting catastrophic events in the actual world [9]. Thomas Bolton et al.'s example shows how data-driven approaches may be used to predict small-scale and large-scale events while upholding physical laws, even when data are restricted to a specific area or external force. This work shows that ocean eddy sequences that are used in coarse-resolution climate models may be successfully constructed [10]. In addition to Partha Pratim Sarkar, Five locations across India were used to provide location-specific SST forecasts for three different time frames by mixing deep learning neural network estimators with numerical estimators. (daily, weekly, and monthly) [11]. Ordinary neural networks (NNs) were first used to make predictions, and deep learning networks were then used. The deep learning long short-term memory (LSTM) neural network was used further to improve the results of the NNs at all times and at all the chosen locations [12]. The outcomes of the NNs significantly outperformed those of the numerical forecasts. Several statistical tests showed that the model functioned well, with correlation values near 1.0 and decreased error. Felipe C.Minuzzia and colleagues [13] With the best accuracy approaching Ninety-five compared to reexamination data and 87% when compared to buoy data, it demonstrates that a method based on data may be employed in place of the computationally costly physical

models. Since there are several applications for wind speed prediction, it might not be easy to categorize the duration range of forecast horizons in these systems [14]. Reaction times, minutes, seconds, or fractions of minutes are one of the main issues with wind speed prediction, yet grid-based factory scheduling and market response require more time frames. Depending on the energy market, different time frames are required for forecasts [15]. An answer is needed in a matter of minutes for the real-time energy market, while The needs of the day-ahead energy market divination are up to 24 hours in advance since it involves data as long as energy deal the backing day. The borders between each of these periods may also need separate estimated time scales. For instance, when employing options like profitable cargo ship and charge raise/down, scheduling between 30 minutes and 6 hours in advance is required. We concentrate on predicting short-term wind speed in our work.

The Argo project, a cooperation of more than 30 entities from all continents, has collected more than 1.5 million salinity, pH climate, and salinity characteristics using Argo floats worldwide. Thanks to the Argo project, extreme weather conditions and ocean processes like El Nino may now be forecast [16]. Reference examined the accuracy of temperature and heat storage estimations from the Argo profiling float dataset. The mixed-layer depth in the Southern Ocean was calculated using the temperature, salinity, and pressure profiles from the Argo floats. In Reference, the temperature profiles of the Argo were handled using the unsupervised categorization approach. It was discovered how spatially varied the installment fields were in the upper 1400 metres of water using the Argo dataset. Traditional statistical methods are no longer acceptable for controlling this information collection due to recent growth in marine data [17]. Therefore, techniques for machine learning should be used to analyse the Argo dataset. Argo data were examined using an approach to machine learning to categorise different types of moline. The KNN regression technique anticipated ocean salinity and temperature [18]. Utilizing the Argo dataset, the SVR approach was used to study and project the lateral thermocline boundary.

Studies that focus on ocean temperature, salinity, and QC methods have also recently been carried out. The findings indicated that by utilizing 28 full-depth hydrographic sections, the median thermotropic levels of the sea rose globally. between 0.11 and 0.100 mm per year throughout were identified, and the thermocline's seasonal variation characteristics were also looked at [19] Reference developed an the year 1990 and the present. Using information on surface wind velocity and the temperature of seawater collected over a 51-year period, the reaction of the set depth to the El collection-Southern Oscillation was examined [20]. Using water temperature profiles from the Southern Chinese's uppermost set limits and the boundaries of the Chinese Ocean automated QC method to remove [21] anomalous values from the profiles by objective mapping. Regarding a fictitious autonomous For an ongoing flow of Argo information the QC system, Reference used a machine learning technique for delayed-mode quality assurance of Argo profiles. A deep learning framework has been made available to handle data from spatial ocean sensing and anticipate thermoclines [22]. Processing Argo data is more effective using this method; It is challenging to quickly and precisely estimate the ocean's temperature and salinity due to the dearth of research explicitly addressing this topic [23]. This paper aims to create a deep Bi-LSTM learning strategy for Argo that will properly predict ocean temperature and salinity as a backup plan.

Due to its potential impact on various activities, SST includes activities, aquaculture, maritime the natural world, and weather forecasting. is an important aspect to consider when dealing with maritime environments. As a result,

forecasting the SST over short and long durations is a current issue that has lately caught researchers' attention [24]. SST has been estimated for the Tropical Atlantic region using a support vector machine trilogy (SVM)-based forecast approach. The dataset employed in this study was gathered from two PIRATA buoys that are located, respectively, at 8°N and 10°S, and it serves as the SVM model's initial data intake. The approach employed by the authors expands upon that. Most of the time, it is feasible to forecast both the long-term (a few weeks) and the short-term (a few days) ocean surface temperature. This issue may be framed as a timeline recurrence issue. Given this, it is possible to predict the SST using techniques similar to long-term memory [25]. (LSTM). This work initially used an LSTM layer to replicate the time series. A fully connected layer uses the output produced by the LSTM layer to forecast SST. For the series of coastal waters, the authors recommended sea surface temperatures.

A strategy for predicting day sea temperature periods was developed using Eastern China Gulf case study and forty-six years' worth of information from sensor series of time. This method made use of the historical satellite anomaly instead of determining the ocean's temperature on the surface [26]. This machine learning system utilizes a large short-term memory in conjunction with an AdaBoost batch learning model to increase accuracy and provide trustworthy temperature forecasts. (LSTM). On the datasets for the East China Sea and the Yellow Sea, another Deep Gated Recurrent Unit merger was used. The deeply concealed time components and A location-based components of the SST data are extracted using the +e deep GRU and the convolutional layers, respectively. This method proposes a hybrid strategy that combines data-driven and numerical techniques, and it effectively attained a 98 percent accuracy rate. When used in a localized investigation, the numerical estimate of the sea surface shows significant variations and has decreased dependability for Permanent forecast. It mitigates these undesirable effects. This study forecasted daily, weekly, and monthly at several places in India using deep-learning neural networks and statistical estimators [27]. Once conventional neural networks for prediction have been deployed, the LSTM is used throughout all periods. Compared to linear approaches, the +e LSTM is more sensitive to gap lengths and has a more extensive data space. The linear model (ARIMAX), which was used as a comparison, was shown to be unable to manage wide and variable temporal horizons effectively.

The authors recommended creating a prediction model to forecast SST throughout the whole China Sea. They profited from information gathered over a year. SST prediction, they proposed a deep learning model that makes use of the LSTM architecture [28]. They suggested dividing the data into SST anomalies and SST averages for their investigation. They trained the appropriate LSTM model using each data split. Additionally, they advocated the use of independent characteristic map (SOM) neural networks to categorize various subregions since the accuracy of the SST forecast is improved by the results of these classifier models. The Gulf of Mexico, Korea, and the UK, respectively, were the sources of the marine data used by the writers. Each of the four daily (i.e., every six hours) datasets is provided by the 13 sites dispersed over these three zones. The suggested model forecasts the day SWH at each place at midnight [29]. The Support Vector Regression (SVR) and extreme learning machines (ELM) models were contrasted with the authors' two suggested models. The outcomes show that the suggested models drastically underperform ELM and SVR. The suggested models fared better than the conventional ELM and SVR.

For various coastal engineering applications, wavelength forecasting is essential. The authors have used Multilayer Perceptrons with assistance vector machines (SVR) to forecast substantial wave height [30]. The scientists

anticipated wave height values in Lake Michigan using a data collection and used that data set to do this. Wave height predictions were made by Shamshirband et al., using information from two separate Persian Gulf sites and wind data. ML-based models, including synthesis neural networks, advanced learning machines (ELM), and SVR, are used in the research to predict wave height. Conducting Waves Nearshore (SWAN) is a numerical model that simulates waves near shore.

3. Problem Statement

This research aims to develop a machine-learning model for predicting the status of ocean data, such as temperature, salinity, currents, and sea level. The prediction model should use historical ocean data collected from various sources, such as Atlantic Ocean data and pacific ocean data in a certain period, and be able to provide accurate predictions of future ocean data status. Predicting ocean data status is essential for various industries, such as fisheries, shipping, and tourism, as it can help anticipate potential risks and optimize operations. However, predicting ocean data status can be challenging due to the complexity and variability of the ocean environment. Therefore, the machine learning model we use here is Bi-LSTM, designed to handle a large amount of data and make the process fast and simple. The model should also adapt to changing environmental conditions and be scalable to handle large amounts of data. The success of this project would result in a valuable tool for ocean-related industries to improve their operations and decision-making processes, ultimately leading to more sustainable and efficient use of ocean resources.

4. Conclusion

Forecasting the status of ocean data and the accuracy of ocean data is crucial and necessitates considerable thought and attention to detail. For a wide range of applications, such as climate modeling, studies of marine ecology, and economical operations like shipping and fishing, accurate and trustworthy ocean data is crucial. Our proposed model Bi-LSTM work performs well in predicting the status of elements, accuracy, and loss in data, which gives accuracy, Precision, Sensitivity, Specificity, and F-Score the result shown for both the accuracy and the loss in the data. Were the output driven in graphic representation, which shows a variation in output? This fast prediction method is well than other methods Models that forecast oceanic losses from natural disasters like hurricanes, tsunamis, and sea level rise may be created by scientists using precise data. These models can aid decision-makers and stakeholders in mitigating risks and minimizing the effects of catastrophic disasters on coastal communities and economies.

References

[1] Wang, Yuchao, Hui Wang, Bin Zhou, and Huixuan Fu (2021). Multi-dimensional prediction method based on Bi-LSTMC for ship roll. Ocean Engineering, 242: 110106.

[2] Wei, Yunyu, Zezong Chen, Chen Zhao, Yuanhui Tu, Xi Chen, and Rui Yang (2021). A BiLSTM hybrid model for ship roll multi-step forecasting based on decomposition and hyperparameter optimization. Ocean Engineering 242: 110138.

[3] Zhao, Jiansen, Zhongwei Yan, Xinqiang Chen, Bing Han, Shubo Wu, and Ranxuan Ke (2022). k-GCN-LSTM: A k-hop Graph Convolutional Network and Long–Short-Term Memory for ship speed prediction. Physica A: Statistical Mechanics and its Applications, 606: 128107.

[4] Rajmohan, R., T. Ananth Kumar, M. Pavithra, and S. G. Sandhya (2020). Blockchain: Next-generation technology for industry 4.0. In Blockchain Technology, Pages 177-198.

[5] Gao, Song, Juan Huang, Yaru Li, Guiyan Liu, Fan Bi, and Zhipeng Bai (2021). A forecasting model for wave heights based on a long short-term memory neural network. Acta Oceanologica Sinica, 40(1): 62-69.

[6] Abduljabbar, Rusul L., Hussein Dia, and Pei-Wei Tsai (2021). Unidirectional and bidirectional LSTM models for short-term traffic prediction. Journal of Advanced Transportation, Pages 1-16.

[7] Hao, Xueli, Ying Liu, Lili Pei, Wei Li, and Yaohui Du (2022). Atmospheric Temperature Prediction Based on a BiLSTM-Attention Model. Symmetry, 14(11): 2470.

[8] Adytia, Didit, Deni Saepudin, Sri Redjeki Pudjaprasetya, Semeidi Husrin, and Ardhasena Sopaheluwakan (2022). A deep learning approach for wave forecasting based on a spatially correlated wind feature, with a case study in the Java Sea, Indonesia. Fluids, 7(1): 39.

[9] Usharani, S., P. Manju Bala, R. Rajmohan, T. Ananth Kumar, and M. Pavithra (2022). Data Visualization for Healthcare. In Intelligent Interactive Multimedia Systems for E-Healthcare Applications, Pages 3-32.

[10] Jin, Emilia K., James L. Kinter, Bin Wang, C-K. Park, I-S. Kang, B. P. Kirtman, J-S. Kug et al. (2008). Current status of ENSO prediction skill in coupled ocean–atmosphere models. Climate Dynamics, 31: 647-664.

[11] Fujii, Yosuke, Elisabeth Rémy, Hao Zuo, Peter Oke, George Halliwell, Florent Gasparin, Mounir Benkiran et al. (2019). Observing system evaluation based on ocean data assimilation and prediction systems: On-going challenges and a future vision for designing and supporting ocean observational networks. Frontiers in Marine Science, 6: 417.

[12] Stammer, Detlef, M. Balmaseda, P. Heimbach, A. Köhl, and A. Weaver (2016). Ocean data assimilation in support of climate applications: status and perspectives. Annual Review of marine science, 8: 491-518.

[13] Ajagbe, Sunday Adeola, Joseph Bamidele Awotunde, Ademola Olusola Adesina, Philip Achimugu, and T. Ananth Kumar (2022). Internet of Medical Things (IoMT): Applications, Challenges, and Prospects in a Data-Driven Technology. Intelligent Healthcare: Infrastructure, Algorithms and Management, Pages 299-319.

[14] Moore, Andrew M., Matthew J. Martin, Santha Akella, Hernan G. Arango, Magdalena Balmaseda, Laurent Bertino, Stefano Ciavatta et al. (2019). Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: A more complete picture of the state of the ocean. Front. in Marine Science. 6: 90.

[15] Gao, Da-wei, Yong-sheng Zhu, Jin-fen Zhang, Yan-kang He, Ke Yan, and Bo-ran Yan (2021). A novel MP-LSTM method for ship trajectory prediction based on AIS data. Ocean Engineering, 228: 108956.

[16] Wang, Bin, June-Yi Lee, In-Sik Kang, J. Shukla, C-K. Park, A. Kumar, J. Schemm et al. (2009). Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Climate Dynamics, 33: 93-117.

[17] Ko, Dong S., Paul J. Martin, Clark D. Rowley, and Ruth H. Preller (2008). A real-time coastal ocean prediction experiment for MREA04. Journal of Marine Systems, 69(1-2): 17-28.

[18] Liu, Ying, Pei-Wei Lin, Xiaorong Lai, Gang-Len Chang, and Alvin Marquess (2006). Developments and applications of simulation-based online travel time prediction system: traveling to Ocean City, Maryland. Transportation Research Record 1959, 1: 92-104.

[19] Rajeevan, M. (2001). Prediction of Indian summer monsoon: Status, problems and prospects. Current Science, Pages 1451-1457.

[20] Wang, Jun, Yuhan Guo, and Yiyang Wang (2022). A sequential random forest for short-term vessel speed prediction. Ocean Engineering, 248: 110691.

[21] John, A., Shubham Kumar Singh, M. Adimoolam, and T. Ananth Kumar (2021). Dynamic sorting and average skyline method for query processing in spatial-temporal data. International Journal of Data Science, 6(1): 1-18.

[22] Key, Robert M., Are Olsen, Steven van Heuven, Siv K. Lauvset, Anton Velo, Xiaohua Lin, Carsten Schirnick et al. (2015). Global ocean data analysis project, version 2 (GLODAPv2). Ornl/Cdiac-162, Ndp-093.

[23] Key, Robert M., Are Olsen, Steven van Heuven, Siv K. Lauvset, Anton Velo, Xiaohua Lin, Carsten Schirnick et al. (2015). Global ocean data analysis project, version 2 (GLODAPv2). Ornl/Cdiac-162, Ndp-093.

[24] Fox, D. N., W. J. Teague, C. N. Barron, M. R. Carnes, and C. M. Lee (2002). The modular ocean data assimilation system (MODAS). Journal of Atmospheric and Oceanic Technology, 19(2): 240-252.

[25] Behringer, David, and Yan Xue (2004). Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. Symp. on int. observing and assimilation systems for the atmosphere, oceans, and land surface.

[26] Zhang, Zijun (2018). Improved adam optimizer for deep neural networks. In 2018 IEEE/ACM international symposium on quality of Service (IWQoS), Pages 1-2.

[27] Bock, Sebastian, Josef Goppold, and Martin Weiß (2018). An improvement of the convergence proof of the ADAM-Optimizer. ArXiv preprint arXiv:1804.10587.

[28] Bock, Sebastian, and Martin Weiß (2019). A proof of local convergence for the Adam optimizer. In 2019 joint international conference on neural networks (IJCNN), Pages 1-8.

[29] Chandriah, Kiran Kumar, and Raghavendra V. Naraganahalli (2021). RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimedia Tools and Applications, 80(17): 26145-26159.

[30] Khaire, Utkarsh Mahadeo, and R. Dhanalakshmi (2020). High-dimensional microarray dataset classification using an improved Adam optimizer (iAdam). Journal of Ambient Intelligence and Humanized Computing, 11(11): 5187-5204.