Hybrid Deep Learning Models For Forecasting Harmful Algal Blooms (HABs)

[1] Amit Kumar Pandey Dr. Santosh Kumar Singh [2] Sakina Sayyad [3]Aditya Yadav [4]
[1]Guide [2]Head of Dept. [3,4]PG Student
[1,2,3,4]Department of Information Technology
[1,2,3,4]Thakur College of Science and Commerce (East), Mumbai, India
[1]amitpandey8089@gmail.com, [2]sksingh@tcsc.edu.in,
[3]sakinasayyad100@gmail.com ,[4]adi210903@gmail.com

***Abstract*- This paper designs and analyzes hybrid deep learning methods of forecasting HABs along California and Southern Oregon coastlines between the month of April and July 2025. They are three models that are operated: a CNN-GRU to predict chlorophyll-a; a multi-variable CNN-GRU to identify the risk of the HAB; and a Bloomformer-CNN-BiLSTM to predict the toxin level. These are models that involve biological and environmental information on water temperature, salinity, and chlorophyll-a levels. Findings indicate that multi-variable and transformer-enhanced models provide more reliable forecasts that can be used to make early warning and coastal management. Bloomformer-CNN-BiLSTM model is the most representative of the capacity-level and accuracy with toxin-related parameters of the system levels of Pseudo-nitzschia and domoic acid. This can establish a base on hybrid deep learning techniques in predicting HABs, informative in establishing the financial, health, and environmental dangers of algal blooms.**

Keywords: Chlorophyll-a, Forecasting, Deep Learning, Bloomformer, Toxic Level, GRU and LSTM

I. INTRODUCTION

HABs refer to events when marine-damaging algal species multiply rapidly and discharge toxins, causing health hazards to humans and ecological life and contaminating sea food. The bloom typically occurs along the Oregon coast southward and occurs on the California coast. Algal blooms are triggered by varying conditions of temperature, salinity, or nitrogen in the ocean. The most detrimental Pseudo-nitzschia blooms are those that produce domoic acid, a strong neurotoxin that bioaccumulates in fish and shellfish and poses a threat to human and animal health. Hence, it is necessary to monitor and forecast HABs for the security of coastal communities, aquaculture, and fisheries.

In recent years, it has become a stronger forecasting tool for environmental phenomena specifically because of its ability to extract patterns from big datasets. Unlike classical statistical models, deep learning can seize short-term fluctuations and long-term relationships in biological and environmental data. Convolutional neural networks and recurrent networks such as GRU or BiLSTM, along with advanced transformer-based architectures like Bloomformer, serve as examples of hybrid models combining multiple deep learning architectures to impart superior accuracy and robustness. Using such architectures, the study aims at forecating hazardous levels associated with HABs, in particular particulate domoic acid, cellular domoic acid, and pseudo-nitzschia abundance, for enhanced early-warning and sustainable coastal management

A. Objective

This study is aimed at designing and testing hybrid deep learning between April and July 2025 based on biological and environmental data, obtained through various sources. Models forecast concentrations of harmful algal blooms (HABs) and associated toxins in the shores of California and Southern Oregon. A CNN-GRU model shall be adopted in predicting chlorophyll-a concentrations, which will provide early warning of the bloom development especially through determination of temporal trends and spatial patterns. The same CNNGRU approach of estimating bloom occurrence and magnitude will be utilized to score water temperature, salinity, and chlorophyll-a with the goal of predicting the risks of HABs. The pseudo-nitzschia density and the level of domoic acid would be estimated when used together with a Bloomformer over the remaining four months. RMSE, MAE, and R2 will also be used to evaluate the performance of the model as compared to traditional and individual deep learning codes. The findings seek to mitigate the financial, health, and environmental risks that are related to HABs through proper predictions to support aquaculture, public health, and sustainable coastal management.

B. Scope

The proposed research will focus on modeling harmful algal bloom (HABs) and toxins off the coast of Southern Oregon and California and will design and test hybrid deep learning model. This is comprised of the biological and environmental indicators found in the different sources such as salinity, temperature of water, chlorophyll-a and Pseudo-nitzschia abundance to generate reliable short- and medium-term forecasts. Early warning systems, supporting sustainable coastal management, safeguarding human health, and supporting fisheries and aquaculture to lessen the environmental and economic impacts of events caused by HABs are the objectives of the project.

II. PROBLEM STATEMENT

HABs (Harmful Algal Blooms) are known to be a significant menace to the marine environment, human health, and the economy of coastal Californian and southern Oregon. Conventional forecasting thrives have a hard time defining the occurrence of a bloom accurately as the environment and biological parameters vary. Therefore, there is a necessity in the sophisticated data-driven framework that would predict the level of chlorophyll-a, toxin production, and risks of the bloom with the help of the deep learning and machine learning models.

III. LITERATURE REVIEW

Silvia Martin-Suazo, Jesús Moron-Lopez, Stanislav Vakaruk, Amit Karamchandani, Juan Antonio Pascual Aguilar, Alberto Mozo, Sandra Gomez-Canaval, Meritxell Vinyals, and Juan Manuel Ortiz published an article named "Deep learning methods to multi-horizon long-term predictions of harm of algal blooms" in Knowledge-Based systems (Elsevier) in 2024. The study aimed to improve the prediction of HAB by comparing the N-BEATS algorithm to LSTM, CNN, and CNN-LSTM to predict the chlorophyll-a (Chl-a) level in the As Conchas reservoir (NW Spain). WHO HAB alert criteria and 3 year high-frequency buoy data were used to evaluate the models on a one-day and one-week permanent basis. These findings indicated that N-BEATS outperformed other models, especially when the predictions were 7 weeks, and the MAE and the F1-score was high and the generalization on buoy-to-buoy was high. It is curious to observe that the influence of exogenous parameters such as temperature, pH and conductivity did not significantly bias the accuracy, and apparently, simple Chl-a-only models are sufficient. The paper focuses on the manner of enhancing water quality management and protecting the health of the populace by the N-BEATS due to computational efficiency, mobility across monitoring locations, and the possibility of real-time early warning systems of HABs. [1].

One of the book chapters, "Harmful Algal Blooms in Indian Waters", by Vaisakh G. (ICAR–CMFRI, Calicut Regional Station), circa 2020 or later, provides an overview of the reasons, effects, toxins, and control of Harmful Algal Blooms (HABs) in India. The causes of HABs are nutrient enrichment, coastal waters pollution, climate change, ballast water exchange, and aquaculture practices that have brought about toxin production (PSP, DSP, ASP, CFP, NSP), which have caused significant ecological, economical, and health effects. There have been growing rates of HABs in India whereby there have been huge incidences in Kerala, Karnataka and Tamil Nadu coasts affecting fisheries, aquaculture, tourism as well as the health of the people. The discussion on remote sensing in operation (i.e. bio-optical algorithm, Red tide index and bloom index by INCOIS) and monitoring schemes to detect HABs at the initial stages. Timely intervention proved to be significant in the chapter by utilizing the concept of sustainable practices, phycotoxin monitoring, and environmental control to reduce the risk. It gives a summary of the Indian waters in relation to various outlooks of the monitoring and management of HABs in reef of marine ecosystem and human wellbeing.[2].

The framework of the "Deep learning approach to simulating harmful algal blooms of the ocean numerical modeling" is placed in the background of the article by Baek et al. (2021).The paper indicates that algal bloom is simulated through the lenses of harmful bloom. It presents a new outlook of deep-learning model coupled with ocean model to simulate Alexandrium catenella HABs. With the CNN-based classification and regression (GoogleNet and ResNet-101) method, the research was capable of attaining the highest accuracy at bloom initiation of 96.83 percent and prediction of bloom density with a RMSE of 1.20 logcells/L. The data of EFDC model (2017-2019) were analyzed using Grad-CAM to interpret physical, chemical, and biological variables and it was evident that salinity, temperature, and NH4-N were the most influential variables with a lookback period of 530 days. The models had the capability to give the distribution of blooms as observed and their efficacy deteriorated with a protrusion in forecast horizons and an insufficiency of PO4P and NH4N information restricted them. This paper provides that AI-based HAB modeling could be employed towards superior tools of interpretation, prediction, controlling, and suppression, particularly in economic terms and health to the community. [3].

The article by Feng Zhang et al. (2020) has discussed the prospect of predicting algal blooms using a "Deep-Learning-Based Approach for Prediction of Algal Blooms". It was applied to the East China coastal waters (Zhejiang Province in 2008-2012), whereas the five-layer DBN (structure: 12-11-11-11-11) was much more successful in the prediction of the algal bloom events than the standard Backpropagation Neural Network (BP) (RMSE = 0.0475) was. This model is extremely generalizable and can be used to predict the crest values of algal cell density; it is also highly stable on the training and test sets. Even though it acknowledged the challenges of coming up with an optimized design of DBN architecture, this research paper consolidates the thought that deep-learning-based applications are more effective in identifying nonlinear correlations between physical, chemical, and biological variables. In general, this article gives relevance to DBNs as the means of environmental protection, algal bloom prediction, and disaster management.[4].

The paper "Algal Blooms Forecasting with Hybrid Deep Learning Models from Satellite Data in the Zhoushan Fishery " by Wenxiang Ding and Changlin Li (2024) study on the algal bloom and chlorophyll-a (Chl).The hybrid product performed better than each of the individual products, which had the highest R, lowest RMSE, and very good bloom prediction metrics (POD, POCR, HSS). Unlike in the individual model, the hybrid product might have precisely predicted the 24-28 June, 2022 bloom event as Chl concentration overruled the prediction in the high-Chl area, and the meridional wind and current overruled the prediction in the medium/low Chl area. The hybrid had most of its best performances during summer when the blooms occurred on a seasonal basis.Processes which develop in complex sequential patterns not known to the hybrid were totally disregarded as also were the absence of such essential variables as dissolved oxygen and nutrients. In conclusion, the research offers yet another effective AI-powered solution to fishery management and control and sustainable use of resources at Zhoushan Archipelago. [5].

The paper from 2018 by Sangmok Lee and Donghyun Lee titled "Improved Prediction of Harmful Algal Blooms in Four Major South Korea's Rivers Using Deep Learning Models" proposed and statistically compared deep learning models (MLP, RNN, LSTM) and Ordinary Least Squares (OLS) for predicting chlorophyll-a (Chl-a) as a proxy for algal blooms. With weekly water quality and hydrological data from 16 dammed pools across four rivers, the study found that all deep learning models performed better than OLS, with the LSTM working best because it was able to take into temporal dependencies and disturbed data loss. Temperature, pH, DO, BOD, and COD were some of the most important variables positively correlated with Chl-a, while water level, pondage, and cyanobacteria were negatively correlated. Being one-week-holder of data for parameterization of the LSTM despite the missing cyanobacteria data, the study aside proved the use of LSTM as the best for predicting short-term HABs, which is the first application of deep learning methods towards environmental management in Korean rivers.[6].

The article under discussion is a research article by the title of “High-throughput phytoplankton monitoring and screening of harmful and bloom-forming algae in coastal waters with updated functional screening database”, which was published in 2026 in the journal Marine Pollution Bulletin (Elsevier). This paper has utilized high throughput DNA metabarcoding (18S V4 rRNA) to characterize phytoplankton communities in the coastal waters of Hong Kong and has created an updated Harmful Algae Database containing 469 species recognized to produce toxins or cause bloody algae blooms. Among the new damaging algae revealed in the findings, there are Karlodinium veneficum and Cyclotella choctawhatcheeana the distribution of which is highly dependent on environmental parameters like salinity, pH, and nitrogen concentrations. The combination of the molecular methods with the conventional monitoring enhances the detection of harmful algal bloom so early and ecological risk assessment, which will be of great value in water quality management and management of coastal ecosystems globally.[7].

A paper by Aditya R. Nayak, Srinivas Kolluru, Aloke Kumar, and Punyasloke Bhadury, 2025 article named “Revisiting Harmful Algal Blooms in India through a Global Lens: An Integrated Framework of Enhanced Research and Monitoring”, is a comprehensive perception of the research and monitoring of HAB in India as a global program. The report indicates that, 218 HAB events in Indian waters were recorded as of October 2022 with 88 percent of these events recorded since 2000 with dinoflagellates and cyanobacteria recorded together 65 percent of them, and 1/3 all along the coasts of Kerala, Karnataka and Tamil Nadu respectively. The drivers mentioned in this section are nutrients, temperature, salinity and biogeochemistry powered by the monsoon winds yet there exist glaring gaps in documentation, integrated monitoring networks and epidemiological studies. It proposes a four-pronged model which includes new technologies (satellites, imaging sensors, etc.), combined observations networks, clinical health research, and a public interface/citizen science. The review, nevertheless, has pointed at the limitations of satellite reliability, cost of imaging, and a U.S.-biased bias in the program, but it emphasizes the importance of the development of HAB monitoring in India to blue economy, and populations of coastal areas and worldwide SDG-14.[8].

In the paper “Hybrid Machine Learning Techniques in the Management of Harmful Algal Blooms Impact” by Andres Molares-Ulloa, Daniel Rivero, Jesuis Gil Ruiz, Enrique Fernandez-Blanco and Luis de la Fuente-Valentin (2023), the authors explore the use of hybrid ML methods (BAGNET, SVM-KNN) to make a certain prediction of HAB influence on mollusc aquaculture in Galicia, Spain (200419). By the environmental and biological data of chlorophyll, phytoplankton, nutrients, temperature, salinity, oxygen and upwelling index, the models categorized the production areas as open/closed with respect to shellfish toxicity. The findings indicated that BAGNET recorded the best recall of 93.41, and it was always better than the recall of Random Forest, ANN, kNN, SVM, XGBoost, and Naive Bayes and showed good generalization capacity across all the estuaries. In this study, although the datasets are irregular and unbalanced and there is no standardized public dataset to prevent comparison, the hybrid ML models have proven effective and dependable. This paper is unique in that it presents the concept of predicting the status of activities using toxicity, hence underlining the concept of public health and aquaculture safety, and aiding the development of common benchmark datasets to be used in future research in HAB prediction. [9].

The paper “A Remote Sensing and Machine Learning-Based Approach to Forecast the Onset of Harmful Algal Bloom” by Izabi, Sultan, El Kadiri, Ghannadi, and Abdelmohsen (2021) compared and contrasted the XGBoost random forest and SVM models to forecast the occurrence of Karenia brevis bloom in Florida. These parameters are collected on a daily basis using satellite data MODIS ocean color products (2000-2020) e.g.: chlorophyll-a, SST and euphotic depth and fluorescence line height, and can be used to predict the onset of bloom in advance of up to 8 days. XGBoost that incorporated euphotic depth, SST and chlorophyll-a as the key parameters were the best with an accuracy of 96, k = 0.93, F1 = 0.97 and AUC = 0.98 as compared to RF and SVM. The method addresses two key problems of other studies: there is a lack of cloud cover data, variable lag time, but it remains limited by small datasets of cloud free and region specificity. In this regard, this research will provide an effective model to launch a timely alert, thus securing improved tracking and management of HABs and safeguarding human wellbeing.[10].

IV. Methodology

A. Data Collection and Preprocessing

The data was collected in 2025 during the months of April, May, June, and July. The Environmental Research Division Data Access Program (ERDDAP) website is where we obtained the data in a CSV file.

B. Data description

(Figure 1) It contains Key biological and environmental characteristics are included in the data, including salinity, water temperature, chlorophyll-a concentration (chla_filled), pseudo-Nitzschia, cellular domoic acid, particulate domoic acid, and R486 and R551 fitted values (r486_filled, r551_filled). The following is a simple and short description of the dataset.

Parameters	Description of dataset
pseudo_nitzschia	Pseudo-nitzschia algae concentration exceeding 10,000 cells/liter. Practically, higher the chance, higher the risk of the presence of harmful algae.
particulate_domoic	Probability of particulate domoic acid concentration exceeding 500 ng per liter. This poses the threat of toxin presence in the water.
cellular_domoic	Probability of cellular domoic acid concentration above 10 picograms per cell. Measures toxin levels inside the algae cells.
chla_filled	Chlorophyll-a concentration (mg/m³) calculated from NOAA VIIRS satellite, gap-filled by DINEOF to account for missing data. It denotes the abundance of algae.
r486_filled	Satellite reflectance at 486nm gap-filled by DINEOF to detect algae or alterations in water color.
r551_filled	Satellite reflectance at 551nm, gap-filled using DINEOF for similar purposes as above: monitoring algal patterns and water properties.
salinity	Surface water salinity from WCOFS; unit is PSU (practical salinity units). Influences the growth of algae.

water_temperature	Surface water temperature from WCOFS; in °C. Influences algae growth and bloom development.

Fig.1. Data Description

C. Data pre-processing

(Figure 2) The heatmap displays the Pearson correlation coefficients between various environmental and biological variables. The values range from -1 (strong negative correlation, in blue) to +1 (strong positive correlation, in red).

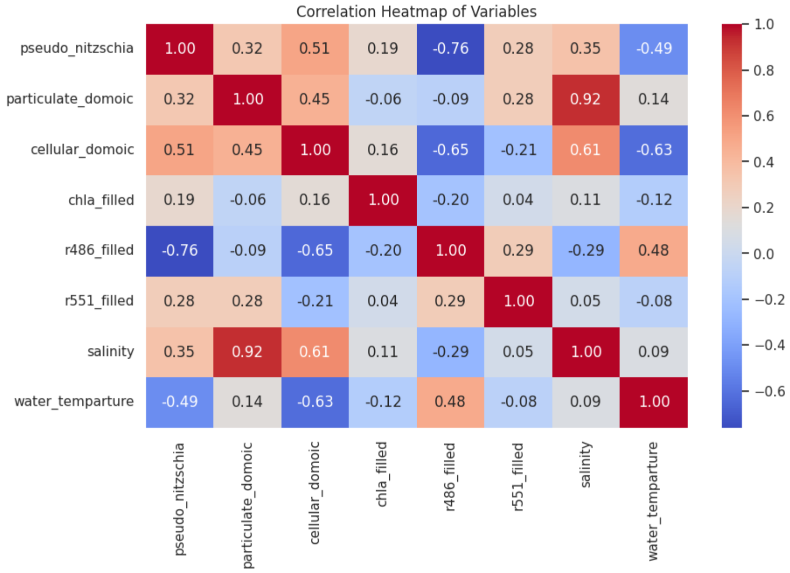

Fig. 2. Heat map

D. Model Implementation

1. Hybrid Model for Forecasting Chlorophyll-a Forecasting using (1D- CNN + GRU)

(figure 3) To forecast chlorophyll-a concentrations in the subsequent four months, you utilized a hybrid CNN+GRU model. The 1D Convolutional Neural Network (1D-CNN) in this feature extractor identifies short-term local structure in time series, i.e., abrupt jumps or drops in chlorophyll concentrations. These identified features are then input to the Gated Recurrent Unit (GRU), which identifies seasonal cycles and long-term dependencies underlying bloom dynamics. The model can be easily fitted to forecast harmful algal bloom risk indicators like Chlorophyll-a due to CNN's strength in identifying local structure and GRU's capacity to recall previous context. This helps both short-term oscillations as well as long-term bloom behaviour to be effectively picked up by this model.

2. Hybrid Deep Learning Model for Forecasting Multi- Variable HAB Risks using (1D- CNN + GRU)

(figure 4) This model uses a hybrid CNN + GRU network to assess HAB risks by forecasting the chlorophyll-a levels in the future along with salinity and water temperature. The 1D CNN layer extracts local space-time features, like short-term spikes and cycles or sudden changes in water quality. Meanwhile, the GRU remembers long-term dependencies and seasonal variations-to forecast-the time series. A Dense layer combines these features for a multi-step prediction. The data will be scaled, converted to sequences, then split into a training/testing set. The model will be trained with respect to the MSE loss and evaluation will be done with MSE, RMSE, and R-square, thereby making a 4-month forecast accompanied by a visualization of past and predicted trends.

3. Hybrid Deep Learning Model for Forecasting Toxic Level using (Bloomformer+ CNN-BiLSTM)

(figure 5) This innovative hybrid deep learning model brings together Gated Recurrent Units (GRU) and 1D Convolutional Neural Networks (1D CNN) to predict various factors influenced by harmful algal blooms, including water temperature, salinity, and chlorophyll-a levels. The 1D-CNN is particularly adept at spotting sudden changes and quick local shifts in input sequences, much like how filters can bring out key features in time-series data. Then the GRU, which is a type of recurrent neural network, captures temporal dependencies and trends over time to decide which historical data should be remembered or forgotten for future forecasts. This way, one can forecast for the next four-time steps by remembering long-term dependencies through GRU and extracting local patterns with CNN. The model then considers and compares recent past values as evidence to demonstrate the prediction of future environmental conditions.

E. Performance Evaluation and Visualize

1. Hybrid Model for Forecasting Chlorophyll-a Forecasting using (1D- CNN + GRU)

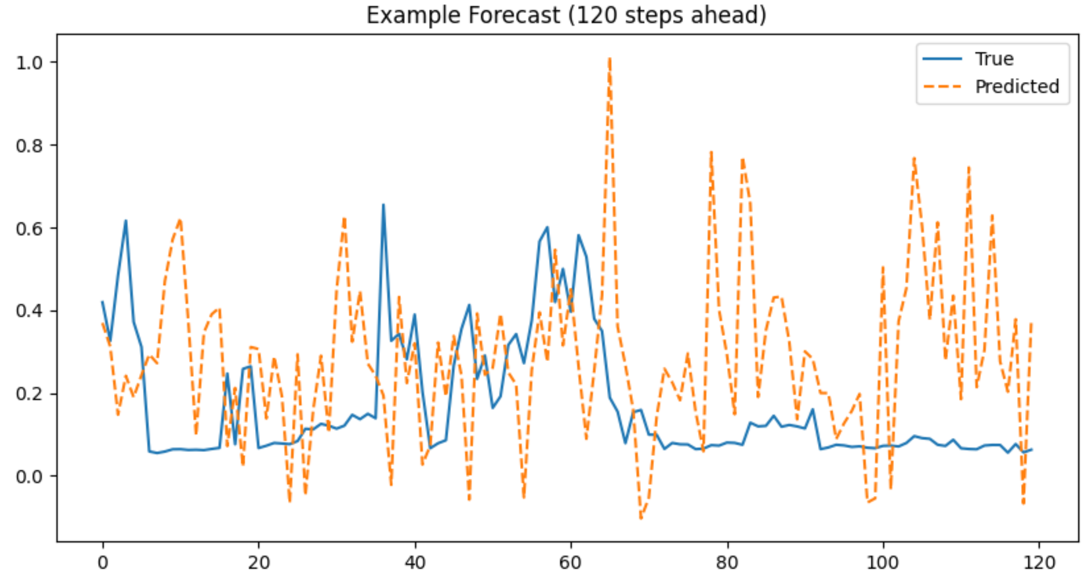

Fig.3.Hybrid Model for Forecasting Chlorophyll-a Forecasting using (1D- CNN + GRU)

2. Hybrid Deep Learning Model for Forecasting Multi- Variable HAB Risks using (1D- CNN + GRU)

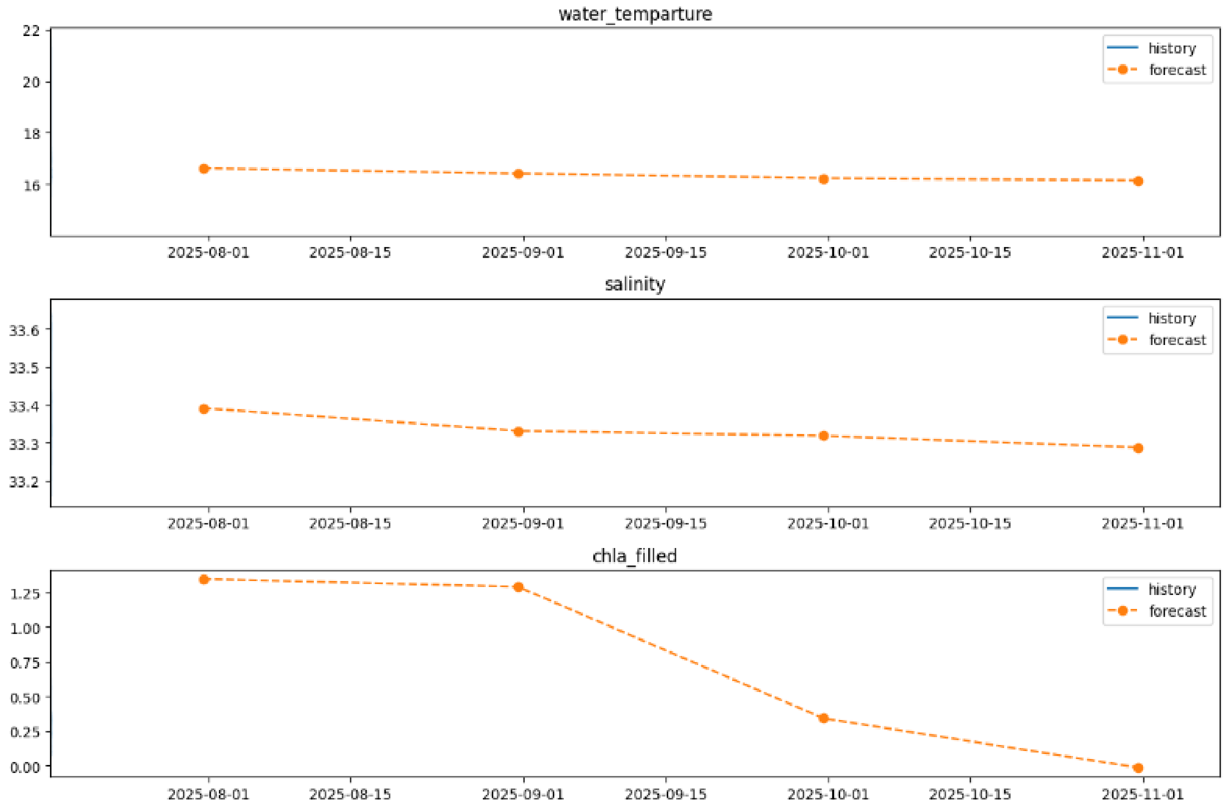

Fig.4.Hybrid Deep Learning Model for Forecasting Multi- Variable HAB Risks using (1D- CNN + GRU)

3. **Hybrid Deep Learning Model for Forecasting Toxic Level using (Bloomformer+ CNN-BiLSTM)**

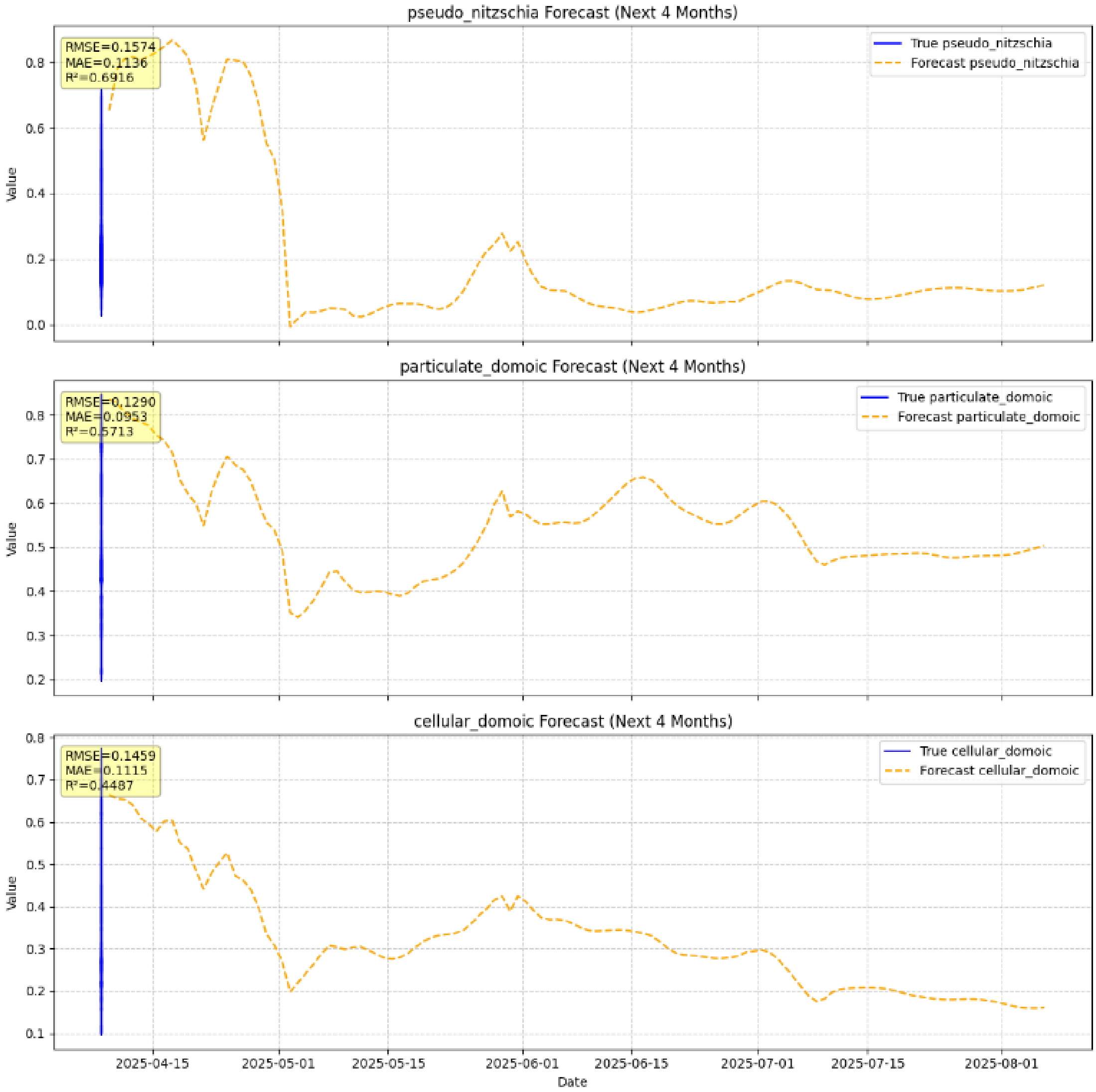

Fig.5.Hybrid Deep Learning Model for Forecasting Toxic Level using (Bloomformer+ CNN-BiLSTM)

I. Result

Hybrid Deep Learning Model	**Parameters**	**RMSE**	**MSE**	**R^2**
1D- CNN + GRU for forecasting chlorophyll-a	chla_filled	1.5987	2.5557	0.0131
1D- CNN + GRU for forecasting Multi-variable HAB Risks	water_temparture	0.0676	0.0046	0.6533
	salinity	0.0899	0.0081	0.3002
	chla_filled	0.0020	0.0000	0.0885
Bloomformer+ CNN-BiLSTM for forecasting Toxic Level	pseudo_nitzschia	0.1574	0.1136	0.6916
	particulate_domoic	0.1290	0.0953	0.5713
	cellular_domoic	0.1459	0.1115	0.4487

Fig.6. Result

V. CONCLUSION

This paper examined the use of three hybrid deep learning systems to predict harmful algal bloom (HAB) warnings and toxicity along the Southern Oregon and California coast. The study was meant to arrive at sound prediction techniques that could be efficiently applied as early warning mechanisms and management at the coastline.

The 1D-CNN + GRU model of chlorophyll-a prediction was not very promising with a small R 2 of 0.0131 indicating that the model might not be sufficient to predict the dynamics of the HABs. Nevertheless, with the extension of the model to multi-variable water temperature, salinity and chlorophyll-a, the model showed better performance especially with water temperature (R 2 =.6533). This emphasizes the need to pay attention to various environmental drivers during prediction of HAB.

Bloomformer + CNN-BiLSTM hybrid model outperformed as it gave the best results when it was used to predict parameters related to the toxin. It was very predictive of Pseudo-nitzschia (R^2 = 0.6916), particulate domoic acid (R^2 = 0.5713) and cellular domoic acid (R^2 = 0.4487). Such findings imply that architecture with transformers, CNN and BiLSTM layers are better in capturing both local features and long run features in complex ecological data.

Altogether, the results prove that hybrid deep learning solutions have a great potential in HAB forecasting. The multi-variable models and the ones which consider the use of the transformer-based architecture has more trustworthy predictions in early warning and coastal management. These high-technology forecasting services are likely to help mitigate the financial, health, and environmental risks caused by HABs through the promotion of aquaculture, personal health, and sustainable management of the coast. The future studies must aim at improving these hybrid models, adding more environmental variables, and increasing the forecast horizon to enhance the accuracy and reliability levels of HAB predictions.

VI. REFERENCES

[1]. Martín-Suazo S, Morón-López J, Vakaruk S, Karamchandani A, Aguilar JA, Mozo A, Gómez-Canaval S, Vinyals M, Ortiz JM. Deep learning methods for multi-horizon long-term forecasting of Harmful Algal Blooms. Knowledge-Based Systems. 2024 Oct 9;301:112279.
[2]. D’Silva MS, Anil AC, Naik RK, D’Costa PM. Algal blooms: a perspective from the coasts of India. Natural hazards. 2012 Sep;63(2):1225-53.
[3]. Baek SS, Pyo J, Kwon YS, Chun SJ, Baek SH, Ahn CY, Oh HM, Kim YO, Cho KH. Deep learning for simulating harmful algal blooms using ocean numerical model. Frontiers in Marine Science. 2021 Oct 12;8:729954.
[4]. Zhang F, Wang Y, Cao M, Sun X, Du Z, Liu R, Ye X. Deep-learning-based approach for prediction of algal blooms. Sustainability. 2016 Oct 21;8(10):1060.
[5]. Ding W, Li C. Algal blooms forecasting with hybrid deep learning models from satellite data in the Zhoushan fishery. Ecological Informatics. 2024 Sep 1;82:102664.
[6]. Lee S, Lee D. Improved prediction of harmful algal blooms in four Major South Korea’s Rivers using deep learning models. International journal of environmental research and public health. 2018 Jul;15(7):1322.
[7]. Lo LS, Liu H, Qiu JW, Tsang PK, Dai Z, Cheng J. High-throughput phytoplankton monitoring and screening of harmful and bloom-forming algae in coastal waters with updated functional screening database. Marine Pollution Bulletin. 2026 Jan 1;222:118675.
[8]. Nayak AR, Kolluru S, Kumar A, Bhadury P. Revisiting harmful algal blooms in India through a global lens: An integrated framework for enhanced research and monitoring. iScience. 2025 Feb 21;28(2).
[9]. Molares-Ulloa A, Rivero D, Ruiz JG, Fernandez-Blanco E, De-la-Fuente-Valentín L. Hybrid machine learning techniques in the management of harmful algal blooms impact. Computers and Electronics in Agriculture. 2023 Aug 1;211:107988.
[10]. Izadi M, Sultan M, Kadiri RE, Ghannadi A, Abdelmohsen K. A remote sensing and machine learning-based approach to forecast the onset of harmful algal bloom. Remote Sensing. 2021 Sep 27;13(19):3863.