The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 44−51

Sea State Prediction for Naval Operations Using Machine Learning Models and ERA5 Reanalysis Data

P. Prasad[1], D. V. L. Priyanka[2], B. Suguna[3*]

[1,2,3]Department of MCA, NSRIT, Visakhapatnam, Andhra Pradesh, India

Corresponding Author *: sugunaboddu6@gmail.com

Abstract

Accurate sea state prediction plays a vital role in supporting naval operations, maritime transportation, and marine safety. Reliable information about ocean conditions helps improve navigation, operational planning, and decision-making in marine environments. This study presents a machine learning-based framework for predicting sea state in the Bay of Bengal using the ERA5 reanalysis dataset. The prediction model was developed using key environmental variables, such as wind speed, sea surface temperature, surface pressure, wave period, and wave direction, as these factors have a direct influence on ocean wave behaviour. To identify the most suitable prediction model, four machine learning algorithms—Random Forest, XGBoost, LightGBM, and TabNet—were trained and compared. Their performance was assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). Among the evaluated models, Random Forest produced the most accurate predictions, achieving an R^2 score of 0.9818. These findings indicate that machine learning provides reliable support for sea state forecasting, contributing to better planning and safer decision-making in naval and maritime applications.

Keywords: Sea State Prediction, Machine Learning, Random Forest, ERA5 Reanalysis Data, Naval Operations, Bay of Bengal.

1. Introduction

Sea state prediction is essential for naval operations, maritime transportation, and marine safety. Ocean conditions such as wind speed, wave height, sea surface temperature, and atmospheric pressure directly influence navigation and operational planning. Accurate prediction of these conditions helps improve decision-making and reduces risks during maritime activities. In this study, a machine learning-based framework is proposed to predict sea state conditions in the Bay of Bengal using the ERA5 reanalysis dataset. Environmental parameters, including wind speed, sea surface temperature, surface pressure, wave period, and wave direction, were used as input features to train four machine learning models: Random Forest, XGBoost, LightGBM, and TabNet. The objective was to identify the model that provides the most accurate sea state prediction for supporting reliable maritime and naval operations [1,18,19].

2. Literature Survey

Sea state prediction has traditionally relied on numerical wave forecasting models such as **SWAN** and **WAVEWATCH III**, which simulate ocean wave behaviour using atmospheric and oceanographic parameters. These models have been widely used for maritime navigation and offshore operations because they provide reliable wave forecasts. However, they require high computational resources and detailed physical modelling, making them less suitable for real-time prediction. Recent studies have also combined **ERA5 reanalysis data** with numerical models to improve forecasting accuracy [19,20].

Recent research has shown that **machine learning and deep learning** techniques can effectively improve sea state prediction by learning patterns from historical oceanographic data. Algorithms such as **Random Forest, XGBoost, ANN, LSTM, GRU, ConvLSTM, and Transformer-based models** have achieved promising results in predicting significant wave height and other ocean conditions. However, prediction performance varies depending on the dataset and geographical region. Therefore, this study evaluates **Random Forest, XGBoost, LightGBM, and TabNet** using the **ERA5 reanalysis dataset** to identify the most suitable model for sea state prediction in the **Bay of Bengal** [1,18].

3. Methodology

The proposed methodology consists of five main stages: data collection, data preprocessing, feature engineering, model development, and performance evaluation. ERA5 reanalysis data for the Bay of Bengal was used as the primary dataset because it provides comprehensive atmospheric and oceanographic information [19,20]. The collected data was preprocessed to remove missing values, correct inconsistencies, and improve data quality. Feature engineering was performed by calculating wind speed from the zonal (u10) and meridional (v10) wind components. The processed dataset was used to train four machine learning models: Random Forest, XGBoost, LightGBM, and TabNet. Their performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). Finally, the Random

Forest model, which achieved the highest prediction accuracy, was selected as the final prediction model.

4. Experimental Setup and Implementation

The experimental setup was designed to implement the proposed sea state prediction framework using Python and widely used machine learning libraries. The implementation process involved data preprocessing, feature engineering, model training, performance evaluation, and model comparison. The following steps were carried out:

4.1 Data Collection: ERA5 reanalysis data for the Bay of Bengal was collected, containing oceanographic and atmospheric parameters such as wind components, sea surface temperature, surface pressure, significant wave height, mean wave period, and mean wave direction.

4.2 Data Preprocessing: The collected dataset was cleaned by handling missing values, removing duplicate records, and correcting inconsistencies. The processed dataset was then prepared for machine learning model development.

4.3 Feature Engineering: Wind speed was calculated from the zonal (u10) and meridional (v10) wind components. The final feature set included sea surface temperature, surface pressure, significant wave height, mean wave period, mean wave direction, and wind speed.

4.4 Model Training and Evaluation:

- The dataset was divided into training (80%) and testing (20%) sets.
- Four machine learning models—Random Forest, XGBoost, LightGBM, and TabNet—were trained using the prepared dataset.
- Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2).

4.5 Model Comparison: The performance of all four machine learning models was compared based on the evaluation metrics. Random Forest achieved the highest prediction accuracy and was selected as the final model for the proposed sea state prediction system.

4.6 Prediction System: The selected Random Forest model was integrated into a Flask-based prediction system, where users can select a location and obtain sea state predictions through a simple web interface.

5. Result Analysis

The performance of the developed sea state prediction system was evaluated using four machine learning models: Random Forest, XGBoost, LightGBM, and TabNet. The models were trained and tested with ERA5 reanalysis data from the Bay of Bengal. Their prediction accuracy was

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 44–51

measured using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2).Among the evaluated models, Random Forest delivered the best overall performance by achieving the lowest error values and the highest R^2 score of **0.9817**. Although XGBoost, LightGBM, and TabNet also produced reliable predictions, their performance was slightly lower than that of Random Forest. These results show that the proposed machine learning framework can accurately predict sea state conditions and provide useful support for naval operations and other maritime applications

Table 1: Performance Comparison

Model	MAE	RMSE	R^2 Score
Random Forest	0.062351	0.097220	0.981762
XGBoost	0.116462	0.155389	0.953408
LightGBM	0.120442	0.159811	0.950718
TabNet	0.113560	0.152704	0.955004

The performance comparison indicates that the Random Forest model achieved the best results with the lowest MAE (0.062351), lowest RMSE (0.097220), and highest R^2 score (0.981762). Compared to XGBoost, LightGBM, and TabNet, Random Forest provided superior prediction accuracy, making it the most suitable model for the proposed sea state prediction framework.

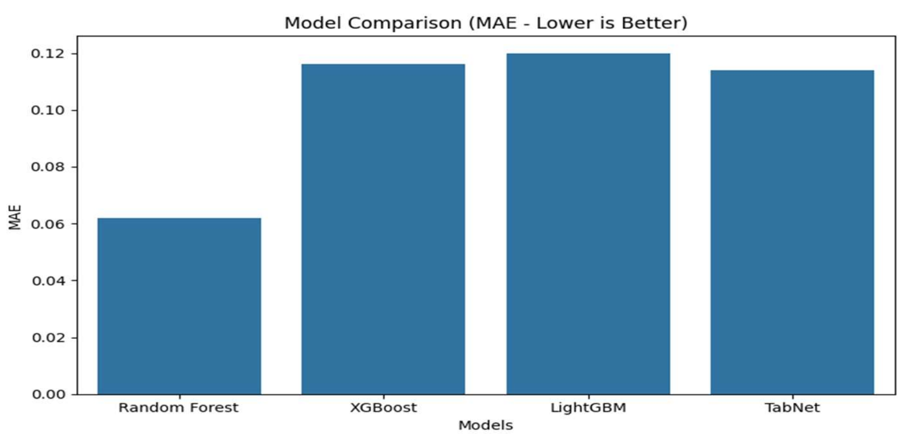

Figure 1. MAE Comparison Graph

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 44−51

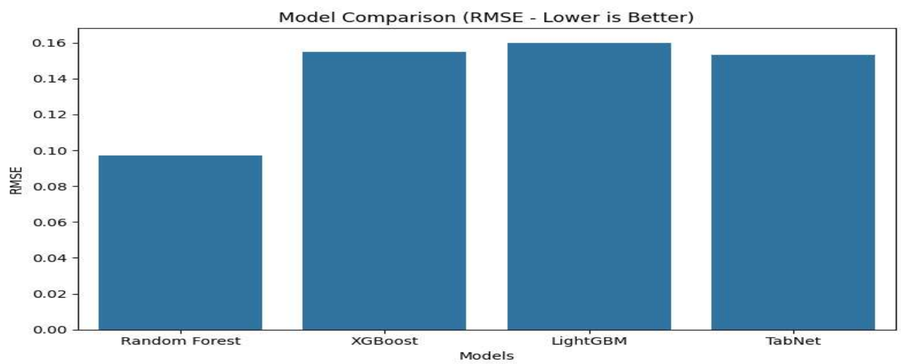

Figure 2. RMSE Comparison of Machine Learning Models

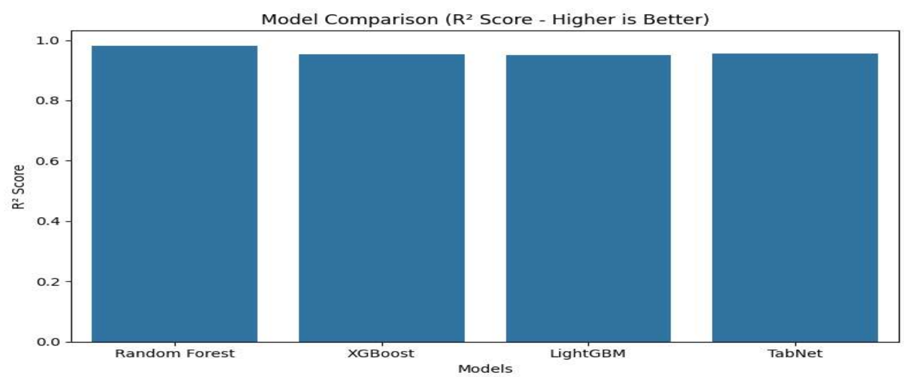

Figure 3. R² Score Comparison Graph

Figure 4: Train vs Test R² Comparison

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 44-51

6. Conclusion

This research developed a machine learning-based framework for predicting sea state conditions in the Bay of Bengal using the ERA5 reanalysis dataset. Four machine learning models—Random Forest, XGBoost, LightGBM, and TabNet—were trained and evaluated to identify the most suitable approach for sea state prediction. Among the evaluated models, Random Forest achieved the highest prediction accuracy and the lowest error values, making it the best-performing model in this study. The results show that machine learning techniques can effectively analyse oceanographic and atmospheric data to produce reliable sea state predictions. The developed framework has the potential to support naval operations, maritime transportation, and marine safety by providing accurate information for operational planning and decision-making. Overall, this work demonstrates that machine learning offers a practical and reliable approach for improving sea state forecasting in the Bay of Bengal.

References

1. Song, T., Han, R., Meng, F., Wang, J., Wei, W., & Peng, S. (2022). A significant wave height prediction method based on deep learning combining the correlation between wind and wind waves. Frontiers in Marine Science, 9, Article 983007. https://doi.org/10.3389/fmars.2022.983007

2. Hao, P., Li, S., Yu, C., & Wu, G. (2022). A prediction model of significant wave height in the South China Sea based on attention mechanism. Frontiers in Marine Science, 9, Article 895212. https://doi.org/10.3389/fmars.2022.895212

3. Hao, P., Li, S., & Gao, Y. (2022). Significant wave height prediction based on deep learning in the South China Sea. Frontiers in Marine Science, 9, Article 1113788. https://doi.org/10.3389/fmars.2022.1113788

4. Hou, B., Fu, H., Li, X., Song, T., & Zhang, Z. (2024). Predicting significant wave height in the South China Sea using the SAC-ConvLSTM model. Frontiers in Marine Science, 11, Article 1424714. https://doi.org/10.3389/fmars.2024.1424714

5. Zhang, Y., Wang, X., & Chen, L. (2025). Machine learning-based short-term forecasting of significant wave height during typhoons using SWAN data: A case study in the Pearl River Estuary. Journal of Marine Science and Engineering, 13(9), Article 1612. https://doi.org/10.3390/jmse13091612

6. Kang, X., Song, H., Zhang, Z., Yin, X., & Gu, J. (2024). A transformer-based method for correcting significant wave height numerical forecasting errors. Frontiers in Marine Science, 11, Article 1374902. https://doi.org/10.3389/fmars.2024.1374902

7. Zhang, Y., Wang, H., Wu, B., Sun, J., Fan, M., Dai, S., Yang, H., & Xu, M. (2025). Automated machine learning-based significant wave height prediction for marine operations. Journal of Marine Science and Engineering, 13(8), 1476. https://doi.org/10.3390/jmse13081476

8. Feng, Z., Hu, P., Li, S., & Mo, D. (2022). Prediction of significant wave height in offshore China based on the machine learning method. Journal of Marine Science and Engineering, 10(6), 836. https://doi.org/10.3390/jmse10060836

9. Han, L., Ji, Q., Jia, X., Liu, Y., Han, G., & Lin, X. (2022). Significant wave height prediction in the South China Sea based on the ConvLSTM algorithm. Journal of Marine Science and Engineering, 10(11), 1683. https://doi.org/10.3390/jmse10111683

10. Shi, J., et al. (2023). A machine-learning approach based on attention mechanism for significant wave height forecasting. Journal of Marine Science and Engineering, 11(9), 1821. https://doi.org/10.3390/jmse11091821

11. Han, Y., Li, X., & Wang, Z. (2024). A significant wave height prediction method based on improved temporal convolution network with attention mechanism. Electronics, 13(24), 4879. https://doi.org/10.3390/electronics13244879

12. Ma, J., Song, N., Nie, J., Ye, M., Liu, X., Yuan, Y., ... & Wei, Z. (2025). MPST: Effective significant wave height prediction method based on multiscale physical space-time-frequency fusion. Intelligent Marine Technology and Systems, 3, Article 27. https://doi.org/10.1007/s44295-025-00080-5

13. Ouyang, Z., Wang, Y., & Li, X. (2024). Significant wave height forecasting based on EMD-TimesNet. Journal of Marine Science and Engineering, 12(4), Article 536. https://doi.org/10.3390/jmse12040536

14. Wang, X., Wang, J., Lu, W., Dong, C., Qin, H., & Jiang, H. (2025). Data-driven rolling model for global wave height. Geoscientific Model Development, 18, 5101–5114. https://doi.org/10.5194/gmd-18-5101-2025

15. Wu, L., An, Y., Qin, P., & Hu, H.-S. (2025). Prediction of significant wave height based on feature decomposition and enhancement. Expert Systems with Applications. Advance online publication. https://doi.org/10.1016/j.eswa.2025.127255

16. Fu, Y., Ying, F., Huang, L., & Liu, Y. (2023). Multi-step-ahead significant wave height prediction using a hybrid model based on an innovative two-layer decomposition framework and LSTM. Renewable Energy, 203, 455–472. https://doi.org/10.1016/j.renene.2022.12.079

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 44–51

17. Hakimi, A., & Farjami, H. (2025). Enhanced significant wave height prediction in the Southern Ocean using an ANFIS model optimized with subtractive clustering. Scientific Reports, 15, 35072. https://doi.org/10.1038/s41598-025-18140-5

18. Wang, X., Jiang, H., et al. (2024). Physics-guided deep learning for skillful wind-wave modeling. Science Advances, 10(49), eadr3559. https://doi.org/10.1126/sciadv.adr3559

19. Bellotti, G., et al. (2021). Regional downscaling of Copernicus ERA5 wave data for coastal applications using SWAN. Water, 13(6), 859. https://doi.org/10.3390/w13060859

20. Yao, R., Li, Y., & Wang, H. (2025). Numeric modeling of sea surface waves using WAVEWATCH III and ERA5 forcing. Journal of Marine Science and Engineering, 13(10), 1450. https://doi.org/10.3390/jmse13101450