This study evaluates the performance of nine machine learning algorithms for Sea Surface Temperature (SST) prediction, concluding that the Bi-LSTM model achieves the best performance. However, while the comparison is comprehensive, the manuscript in its current form lacks sufficient innovation and contains several methodological and presentation weaknesses. I do not believe the manuscript is suitable for publication based on the following concerns.

1. Innovation and Novelty

The primary concern regarding this manuscript is the limited scientific innovation.

• Redundancy of Research: The application of Bi-LSTM for SST forecasting has been extensively studied in recent years. This manuscript appears to be a repetition of existing work without providing new insights or methodological breakthroughs. Specifically, several published works have already established the efficacy of Bi-LSTM and deep learning in this domain, such as:

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., ... & Gong, J. (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. Environmental Modelling & Software, 120, 104502.

Zrira, N., Kamal-Idrissi, A., Farssi, R., & Khan, H. A. (2024). Time series prediction of sea surface temperature based on BiLSTM model with attention mechanism. Journal of Sea Research, 198, 102472.

Xu, T., Zhou, Z., Li, Y., Wang, C., Liu, Y., & Rong, T. (2023). Short-term prediction of global sea surface temperature using deep learning networks. Journal of Marine Science and Engineering, 11(7), 1352.

The authors claim that 'Bi-LSTM is a powerful tool for improving the accuracy of SST bias correction in the TCWAGFS model.' However, the study does not actually implement or demonstrate a bias correction framework. It merely compares the errors between Bi-LSTM and TCWAGFS. The title and the stated objectives are therefore misleading, as the actual work does not reflect the promised 'bias correction' application.

2. Selection of Research Areas

The authors selected five major areas but failed to provide the specific geographic coordinates (latitude/longitude degrees) for these areas. What's more, the 'major' areas do not necessarily equate to regions with the highest prediction errors. The authors should clarify the logic behind this selection. The results show that Bi-LSTM provides the most significant improvement in SST bias prediction in the Southern Ocean, which was not emphasized as a primary focus area. This discrepancy between the selection criteria and the results needs to be addressed.

3. Visualization and Presentation

The figures is repetitive and lacks diversity. Almost all figures in the manuscript utilize the same chart type. This makes the manuscript feel monotonous and hinders readability. The authors should employ a variety of figures to better illustrate their findings. For instance, time-series line graphs could be used to show SST prediction performance over different lead times in the key study regions, which would provide more intuitive insights than static error comparisons.