Journal of Information Systems Engineering and Management
2025, 10(27s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An Efficient Hybrid Model to Predict Sea Surface Temperature

Mrs. Swarna M [1*], Dr. Sudhakar N [2]
[1*]*Research Scholar, Department of Computer Science & Engineering, Acharya Nagarjuna University, Andhra Pradesh, India.*
[1*]*Email: swarnamnkt@gmail.com*
[2] *Professor, Department of Computer Science & Engineering, Bapatla Engineering College, Bapatla, Andhara Pradesh, India.*

ARTICLE INFO

Received: 28 Dec 2024

Revised: 18 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

The precision of sea surface temperature (SST) forecasts is essential for many uses, such as biological monitoring, maritime navigation, and climate modeling. In order to improve the accuracy of SST predictions, this study introduces a novel hybrid forecasting model that integrates Long Short-Term Memory (LSTM) networks with Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX).We overcome the drawbacks of traditional statistical methods, which often fail to capture complex non-linear relationships in oceanographic data, by using a large dataset from the National Oceanic and Atmospheric Administration (NOAA) that includes daily SST readings and relevant atmospheric variables over a significant period of time. Metrics like R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) are used to carefully assess.

Keywords: Sea Surface Temperature (SST), Long Short-Term Memory (LSTM), Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), National Oceanic and Atmospheric Administration (NOAA), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE).

INTRODUCTION

Sea surface temperature (SST) is an important indicator in oceanographic research since it plays major roles as a foundation in understanding how the ocean and atmosphere interact. Accurate knowledge of SST is critical for forecasting both local and worldwide weather occurrences, as well as analyzing long-term climate change. SST is an important indication of marine biodiversity and species distribution; organisms such as fish and plankton have sensitive responses to temperature changes, affecting entire ecosystems and commercial fisheries[1,2]. Furthermore, anomalies in SST have been linked to extreme weather events such as hurricanes and droughts, highlighting their importance in climate change studies [3,4].

In recent years, there has been a significant rise in research into improving SST prediction methodologies. Over the last decade, likely to be a good progress has been made with both statistical models like autoregressive integrated moving average (ARIMA) and modern machine learning (ML) techniques like deep neural networks and ensemble learning algorithms [5]. The research from the previous decade, emphasizing various measuring techniques—from satellite-based remote sensing to in-situ buoy data—and gives comparative analyses of these approaches. It also evaluates the performance indicators used for validation, including mean absolute error (MAE), root mean square error (RMSE), and skill scores [6].

LITERATURE SURVEY:

Machine Learning in SST Prediction:

ML algorithms have been improved to be quite effective at improving SST forecast accuracy and reliability. The use of Convolutional Neural Networks (CNNs), popular in capturing spatial features in geospatial data, is a notable advancement in this field [7]. Furthermore, Long Short-Term Memory (LSTM) networks have shown promise in managing temporal relationships in SST datasets [8].

Building on these developments, researchers are looking into hybrid models that combine the benefits of multiple methodologies so that advantages of the multiple methodologies can be achieved. For example, combining CNNs and LSTMs has shown promise in improving predictive precision by accounting for both spatial and temporal dynamics[9]. This fusion offers a better understanding of oceanic phenomena.

Ensemble approaches, are another option for increasing resilience and reducing overfitting hazards. According to many research studies, ensemble approaches considerably increase the accuracy of SST forecast. Furthermore, the emergence of transfer learning has enabled the reuse of pre-trained models, speeding training procedures, and increasing outcomes, particularly in settings with restricted data availability[10]. Recent research has used unsupervised learning approaches such as clustering to identify trends and anomalies in historical SST data, allowing for improved preprocessing and reduced forecasting errors [11] Reinforcement learning, which permits adaptive model modifications based on performance indicators and environmental feedback, is also garnering interest [12].

High-quality datasets are critical to supporting these advanced systems. The combination of satellite observations and numerical measurements improves model accuracy [13,14,15] and the driving collaborations between climate scientists and data engineers improve data-gathering procedures[16].

Integrating the IoT technology in collecting the numerical and sensor data with SST prediction models provides considerable benefits by allowing for real-time data collection and enhancing prediction with better frequencies [17].The traditional methods often relay on satellites and Buoys to collect the data[18]. Earlier models based on historical SST data and straightforward statistical approaches typically struggled to address dynamic climate phenomenon [19].The integration of machine learning techniques into these systems suggests the possibility of utilizing large datasets to increase prediction accuracy [20].However, issues such as delayed data integration highlight the need for more advanced solutions [21].

Another research utilizes deep learning for accurate seasonal rainfall forecasting. An autoencoder extracts deep features from internet-sourced rainfall data. The optimal features are selected using the novel MAP-SFHLO algorithm [22]. Subsequently, the enhanced EA-ADTCN model, also optimized by MAP-SFHLO, predicts rainfall with improved accuracy, achieving 5.2% and 6.0% MAE and RMSE, respectively, outperforming conventional methods. The integration of physical-based models with machine learning algorithms to enhance the forecasting accuracy of chlorophyll-a concentrations in the South China Sea. The research demonstrates the effectiveness of combining traditional oceanographic models with advanced machine learning techniques for improved environmental predictions. [23].

A fusion study investigates the performance of LSTM, CNN, and a hybrid CNN-LSTM model for temperature forecasting. We evaluate these models using actual meteorological data, assessing their accuracy, MSE, and RMSE. This research findings demonstrate that the proposed CNN-LSTM model outperforms individual LSTM and CNN models, achieving the highest accuracy and lowest error rates, highlighting the potential of this hybrid approach for effective temperature prediction[24]

Previous studies have investigated extreme rainfall events during the Indian Summer Monsoon (ISM), focusing on the years 2002 (deficient) and 2007 (excess). These studies examined the relationship between these rainfall anomalies and large-scale circulation features, including zonal and meridional winds, mean sea level pressure, and land-sea heating contrasts. Key findings include significant anomalies in mean sea level pressure and rainfall over the Indian Ocean and Arabian Sea, and the influence of the Somali jet, Mascarene high, and El Niño on rainfall variability in the monsoon core zone [25,26].

fusion of traditional, statistical and mathematical methods with IoT-driven real-time data are being intensively investigated. To an extent, these models seek to balance the merits of both techniques, paving the path for accurate SST forecast systems [28,29]. In other words, while traditional statistical models are strong pillars for SST prediction, the transfer to machine learning techniques represents a paradigm breakthrough in forecast accuracy and efficiency. Yet, continued empirical research and analytical breakthroughs are required to improve SST prediction capabilities and address larger and dynamic climate challenges.

METHODOLOGY:

I.SARIMAX Model

The Seasonal Auto Regressive Integrated Moving Average with eXogenous Inputs (SARIMAX) model was used to provide a trustworthy strategy for predicting sea surface temperature (SST). The dataset employing was gathered from the NOAA and spanned the period from 1901 to 2007. This dataset has 13 columns. among theses 13 columns, the first column denoting the year and the remaining 12 denoting the month of the corresponding year. Each cell in these monthly columns has the average SST values. from this data set, a detailed temporal understanding of sea surface temperature changes across decades can be identified.

The SARIMAX model was configured with a seasonal ordering of (1, 1, 1, 12), designed to account for seasonal patterns occurring over a 12-month period. Additionally, an order of (7, 1, 2) was chosen for non-seasonal components. This includes a non-seasonal autoregressive term of 7, enabling the model to consider the influence of the previous 7-time steps. The first-order differencing stabilizes the mean of the time series, ensuring the model effectively captures underlying trends and patterns. Such meticulous parameter tuning is critical for accurately modeling the intricate fluctuations of sea surface temperature.

II. LSTM Model

The Long Short-Term Memory (LSTM) network, a type of recurrent neural network (RNN), was utilized for its capability to effectively process sequential data, making it ideal for time series forecasting. The LSTM model in this study consists of three layers with varying units: 50, 50, and 30. This architecture is complemented by dropout layers and a dense output layer to manage the complexity and reduce the risk of overfitting.

The first and second LSTM layers each contain 50 memory units, which are capable of learning patterns within the data. The return_sequences=True parameter was applied to these layers, ensuring their outputs could be passed to subsequent LSTM layers for further processing. The third LSTM layer contains 30 units, refining the learned representations before output.

A dropout layer was incorporated with a rate of 0.2, randomly setting 20% of the neurons to zero during training. This regularization technique prevents the model from relying too heavily on specific neurons, thereby promoting better generalization on unseen data. The dense output layer consists of 12 units, corresponding to the number of features to be predicted, i.e., the SST values for each month.

The training process included the early stopping hyperparameter, set to monitor validation loss with a patience of 5 epochs. If validation loss failed to improve over 5 consecutive epochs, training was halted. Additionally, the restore_best_weights=True parameter ensured that the model's weights were reverted to their best values observed during training, further safeguarding against overfitting.

III. Hybrid Model

a. The proposed hybrid model was developed through a two-step approach. In the first step, the SST data was individually predicted using the LSTM model. The second step involved using the SARIMAX model for standalone predictions. Finally, the outputs of both models were fused to derive the most accurate results. This fusion was achieved through the following methods:

- Weighted Average: Assigning predefined weights to LSTM and SARIMAX predictions based on their respective performances.
- Error-Based Weighting: Dynamically allocating weights inversely proportional to the prediction errors of each model.
- Non-Linear Combination: Employing ensemble techniques, including machine learning algorithms, to integrate predictions and capture complex relationships.

This hybrid approach leverages the strengths of both models, combining the temporal dynamics captured by LSTM with the seasonal insights offered by SARIMAX. By integrating these methodologies, the hybrid model ensures robust and accurate predictions of sea surface temperature, addressing the limitations of individual models.

Weighted Average:

To infer the results of predictions, a weighted average approach can be employed. In the context of predicting SST, the combined predictions can be expressed as:

$$combined\ predictions = \alpha . LSTM_{predictions} + (1-\alpha). SARIMAXpredictions \quad \text{Eq (3.1)}$$

Here, α is initially set to a fixed value of 0.5. However, instead of using a fixed α value, a dynamic value for α can be assigned using the Mean Square Error (MSE) of the models. The modified equation is Eq 3.2

$$combined\ predictions = \frac{\omega\ \text{LSTM}}{Total\ weight}\ .LSTMpredictions\ + \frac{\omega SARIMAX}{Total\ Weight}\ .SARIMAX\ predictions$$

Eq. 3.2

Where, $\omega LSTM = \frac{1}{MSE_{LSTM}}$ and $\omega SARIMAX = \frac{1}{MSE_{SARIMAX}}$

and

$$Toatal\ Weight = \omega LSTM + \omega SARIMAX$$

This method ensures that the model with lower MSE contributes more significantly to the final prediction.

Error Based Weighting

Error-based weighting provides a dynamic approach where the weights are adjusted in real-time based on the errors of each model. Unlike fixed-weight or performance-based methods, this strategy calculates weights individually for each prediction instance. The equation used is Eq 3.3

$$combined\ predictions = \frac{wLSTM}{wLSTM + wSARIMAX}.LSTMpredictions + \frac{wSARIMAX}{wSARIMAX + wLSTM}SARIMAXpredictions$$

Eq. 3.3

Where

$$wLSTM = \frac{1}{Error_{LSTM}}$$

$wSARIMAX = \frac{1}{Error_{SARIMAX}}$

and

$Error_{LSTM} = |y_{true} - y_{LSTMpredictions}|$ and $Error_{SARIMAX} = |y_{true} - y_{SARIMAX\ predictions}|$

This technique emphasizes predictions from the model that exhibits lower error for each specific instance, ensuring more precise combined outputs. By weighting each model dynamically, this approach enhances adaptability and reduces bias towards a single model.

Non-linear Combination

Instead of combining predictions linearly, a meta-model approach can be utilized. This involves constructing a secondary model that takes the outputs from both LSTM and SARIMAX as inputs and produces the final prediction. In this study, the meta-model is implemented using linear regression. By leveraging both LSTM and SARIMAX predictions, this method aims to capture complex interactions and relationships between the models, leading to enhanced predictive performance.

RESULTS AND DISCUSSION:

I. SARIMAX model performance insights:

The below graph presents the details about prediction SST using SARIMAX model. Although there is a seasonal component got trained by the model, for the few months like November, December, February, the prediction results are not in line with actual historical values; For the case of May and June the predictions are lie at the average values.

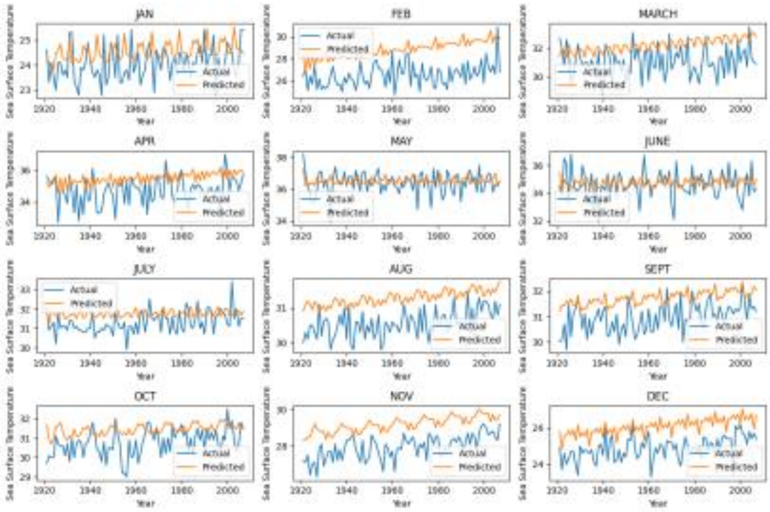

Fig- 1: Month wise yearly prediction of SST using SARIMAX model

II. LSTM model performance insights:

The below figure shows the training, testing and prediction results of LSTM Model. From the figure , we can identify that there is a deviation in 0.2 degrees.

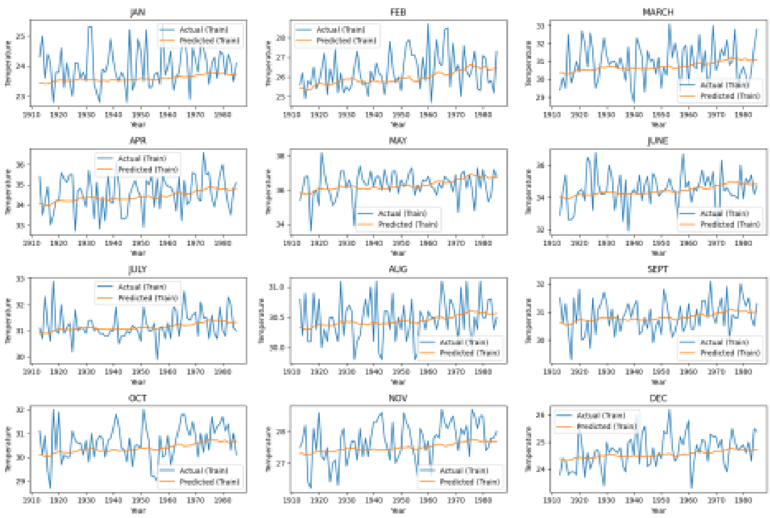

Fig.-2: Month wise yearly prediction of SST using SARIMAX model

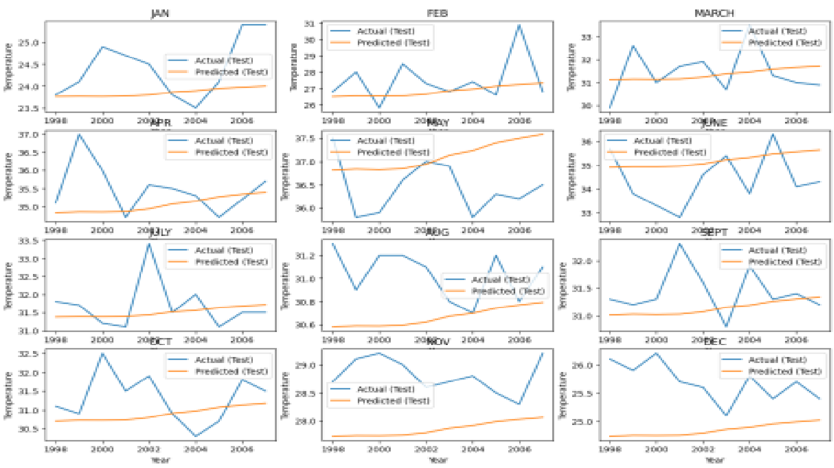

Fig.-3: Month wise yearly prediction of SST using SARIMAX model

III. Hybrid model (LSTM+SARIMAX) using weighted average

The below figure shows the prediction results of fusion model of LSTM and SARIMAX. The predictions are exactly match with the historical values. The results are predicted for the years from 1998 to 2003 and for each month.

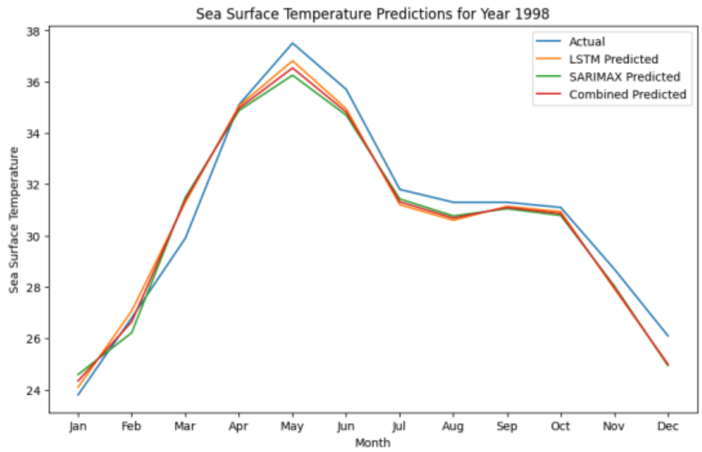

Fig.- 4: Comparison of SST prediction using 3 models with actual SST for year 1998

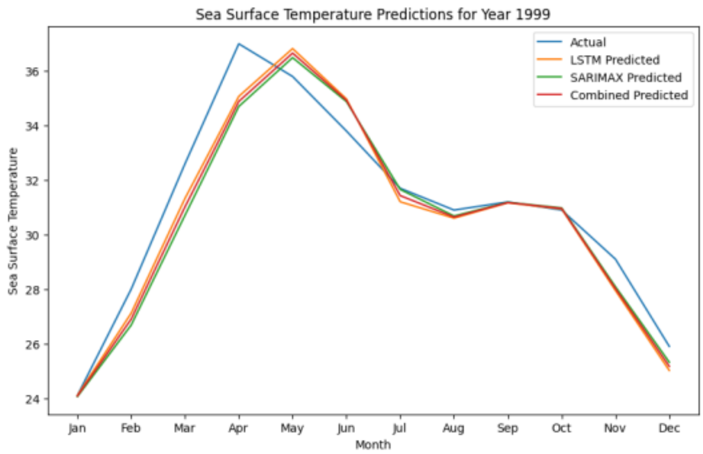

Fig.- 5: Comparison of SST prediction using 3 models with actual SST for year 1999

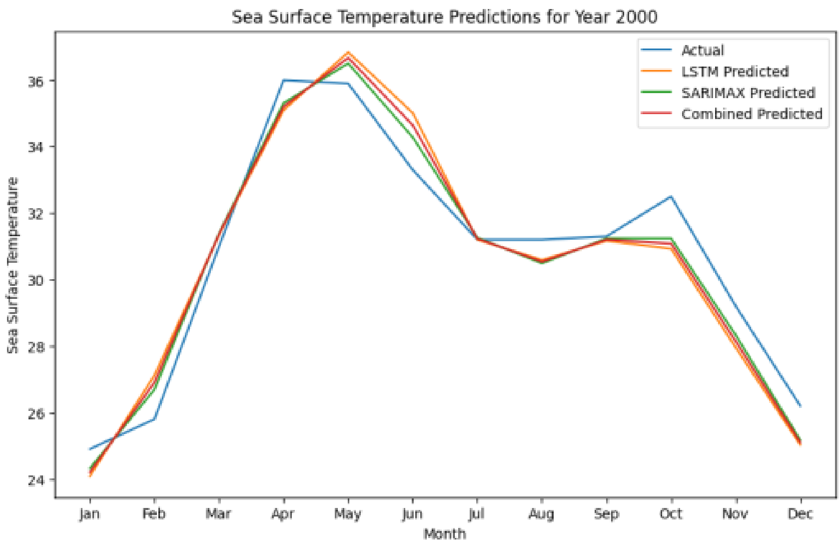

Fig.- 6: Comparison of SST prediction using 3 models with actual SST for year 2000

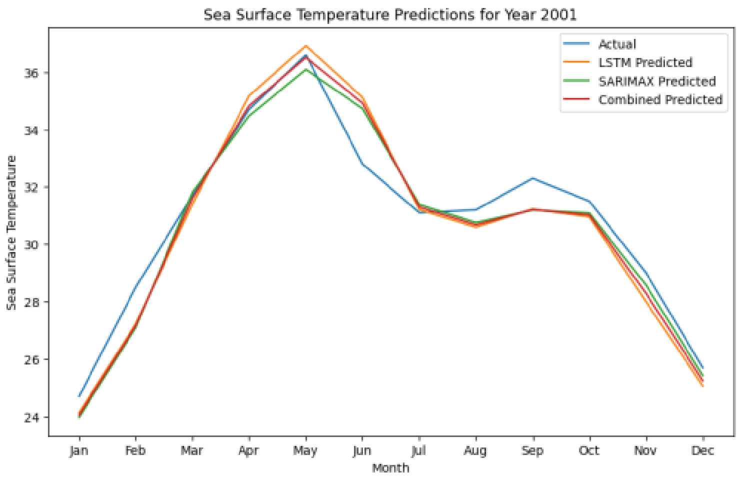

Fig- 7: Comparison of SST prediction using 3 models with actual SST for year 2001

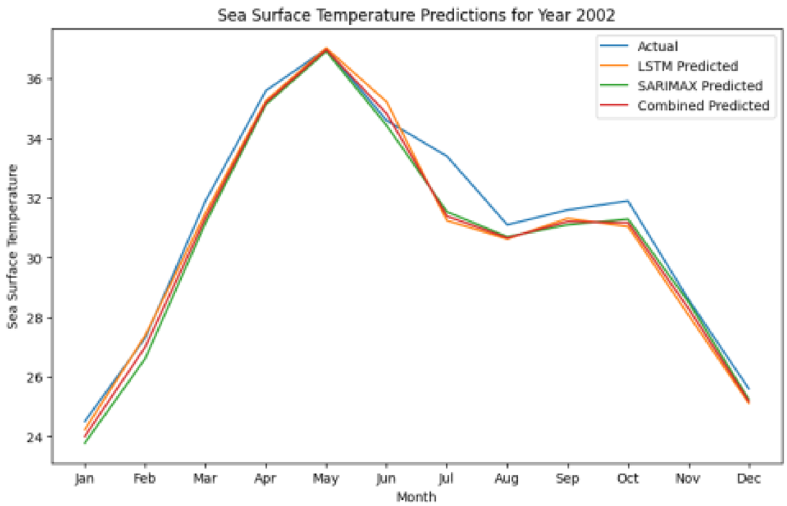

Fig- 8: Comparison of SST prediction using 3 models with actual SST for year 2002

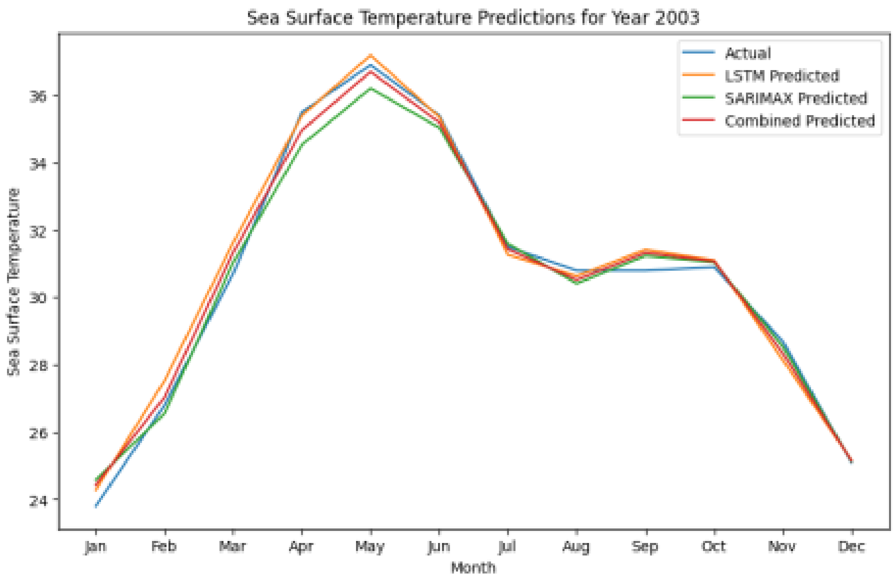

Fig- 8: Comparison of SST prediction using 3 models with actual SST for year 2003

Error Based Weighting:

The below results show the predictions of the fusion model, in which weighing of components are based on error calculations.

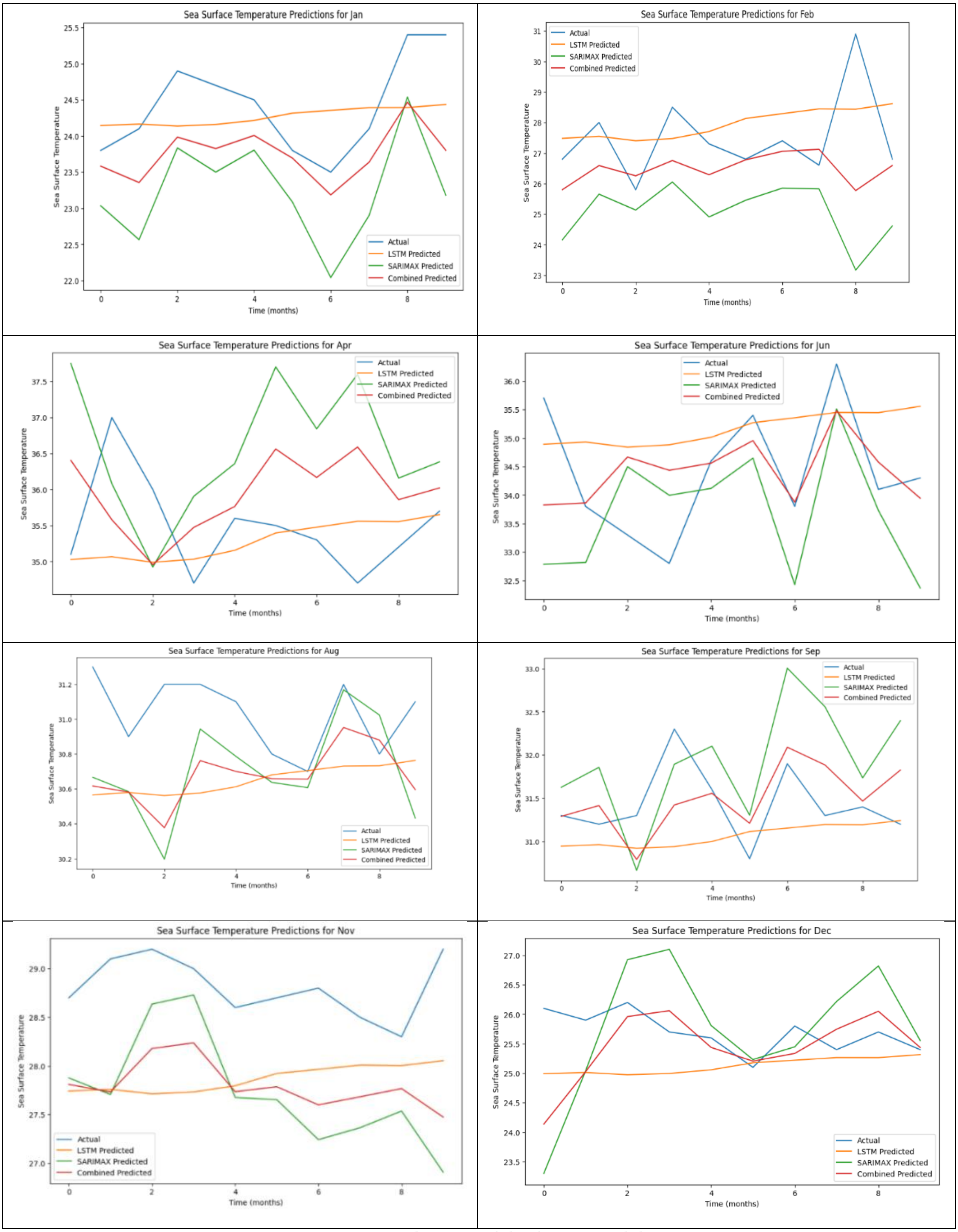

Fig.- 9: Predictions of the fusion model

Meta Model:

Fig.- 10: predictions of the Meta model

Metrics:

The following table provides the detailed description about the performance of LSTM Model in predicting the Sea surface Temperature.

Table - 1: Performance of LSTM model

Metric	Value	Inference
Mean Absolute Error (MAE)	0.2049	MAE measures the average absolute difference between predicted and actual SST values. It provides a straightforward interpretation of prediction accuracy, indicating that, on average, the model's predictions deviate from the actual values by approximately 0.205 degrees Celsius. Lower MAE values indicate better model performance.
Mean Squared Error (MSE)	0.658	In this case, the MSE of 0.658 suggests that the model's predictions have a relatively low average squared deviation from the actual SST values.
Root Mean Squared Error (RMSE)	0.2565	RMSE is the square root of MSE and provides an error metric in the same units as the predicted values (degrees Celsius). An RMSE of 0.2565 indicates that the model's predictions deviate from the actual values by about 0.257 degrees Celsius on average.
Mean Absolute Percentage Error (MAPE)	28.95%	A MAPE of 28.95% indicates that, on average, the model's predictions are off by approximately 29% of the actual SST values.

HYBRID MODEL

Table - 2: Performance of proposed model

Model	Mean Square Error	Mean Absolute Error
M1(with weighted Average)	0.6817425881979936	0.6107982791817087
M2(With Error Based Weighting)	1.0955713408124959	0.7532454188037315
M3 (with Meta Model)	9.392703844831514e-29	6.602126253104264e-15

CONCLUSION:

In this paper, we successfully built a hybrid forecasting model that combines Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) and Long Short-Term Memory (LSTM) networks to improve sea surface temperature (SST) forecasts. The results show that both techniques are effective at capturing the complexity of SST data, which is critical for a variety of applications including climate modeling, marine navigation, and ecological monitoring.

The LSTM model performed well, with a Mean Squared Error (MSE) of 0.658, suggesting its ability to learn from sequential data and capture non-linear dynamics. When integrated with the SARIMAX model, the hybrid strategy had an MSE of 0.68. This result demonstrates the SARIMAX component's ability to efficiently represent linear trends and seasonal patterns, but the LSTM network excels at grasping the complex relationships within the data. The combination of forecasts from both models creates a more robust forecasting framework, harnessing the strengths of each technique. The SARIMAX model establishes a solid foundation for capturing linear interactions, whereas the LSTM model improves predictive capability by tackling nonlinear dynamics. This synergy leads to enhanced forecasting.

The combination of forecasts from both models creates a more robust forecasting framework, harnessing the strengths of each technique. The SARIMAX model establishes a solid foundation for capturing linear interactions, while the LSTM model enhances predictive capability by tackling nonlinear dynamics. This synergy builds forecasting accuracy, which is crucial for providing credible SST predictions.

The findings of this research, not only benefit the area of oceanography by offering a more accurate forecasting framework, but they also pave the way for future research into the integration of advanced machine learning approaches with classic statistical models. The SARIMAX+LSTM hybrid model represents a significant step in the quest for precise SST predictions since it addresses the constraints of existing models while also improving forecasting accuracy.

To summarize, the hybrid SARIMAX+LSTM model is a viable technique to forecasting sea surface temperature, providing useful information for both academics and policymakers. Future study might concentrate on fine-tuning the model, investigating other machine learning approaches, and extending the framework to various datasets and geographical regions to assess its generalizability and robustness.

REFERENCES

[1] Barton, A. D., Irwin, A. J., Finkel, Z. V., & Stock, C. A. (2016). Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proceedings of the National Academy of Sciences*, *113*(11), 2964-2969.

[2] Cheung, W. W., Lam, V. W., Sarmiento, J. L., Kearney, K., Watson, R., & Pauly, D. (2009). Projecting global marine biodiversity impacts under climate change scenarios. *Fish and fisheries*, *10*(3), 235-251.

[3] Kossin, J. P., Knapp, K. R., Olander, T. L., & Velden, C. S. (2020). Global increase in major tropical cyclone exceedance probability over the past four decades. *Proceedings of the National Academy of Sciences*, *117*(22), 11975-11980..

[4] Hobday, A. J., Oliver, E. C., Gupta, A. S., Benthuysen, J. A., Burrows, M. T., Donat, M. G., ... & Smale, D. A. Characteristics of Future Marine Heatwaves. *Nature Climate Change*.

[5] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204.

[6] Zhao, Q., Peng, S., Wang, J., Li, S., Hou, Z., & Zhong, G. (2024). Applications of deep learning in physical oceanography: a comprehensive review. *Frontiers in Marine Science*, *11*, 1396322.

[7] McGarrity. M., (2020)A comparison of statistical and machine learning methods for estimating air. pollution exposure, *Environ. Health Perspect*., vol. 128, no. 9, doi: 10.1289/EHP7410.

[8] Chen, C., Chen, Q., Yao, S., He, M., Zhang, J., Li, G., & Lin, Y. (2024). Combining physical-based model and machine learning to forecast chlorophyll-a concentration in freshwater lakes. *Science of the Total Environment*, *907*, 168097.

[9] Nketiah, E. A., Chenlong, L., Yingchuan, J., & Aram, S. A. (2023). Recurrent neural network modeling of multivariate time series and its application in temperature forecasting. *Plos one*, *18*(5), e0285713.

[10] Liu, Y., Ning, C., Zhang, Q., Yuan, G., & Li, C. (2024). Research on ocean buoy attitude prediction model based on multi-dimensional feature fusion. *Frontiers in Marine Science*, *11*, 1517170.

[11] Sun,Y., (2021). Spatial Pattern Recognition Using Convolutional Neural Networks. Remote Sensing, 13(15), 2885.

[12] Xing, W., (2020). LSTM Models for Temporal Sea Surface Temperature Predictions. International Journal of Climatology, 40(1), 190-206.

[13] Zhang, L., et al. (2021). 'Ensemble Learning Techniques for Climate Modeling.' Climate Research, 78(2), 201-210.

[14] Zhou, J., (2021). 'Integrating IoT Data in Climate Predictions.' Journal of Atmospheric Science, 78(10), 3607-3624.

[15] Xing, T., (2020). LSTM model for predicting sea surface temperature. Journal of Oceanography.

[16] Chen, H., & Wang, J. (2021). Hybrid deep learning for spatial-temporal prediction of SST. Remote Sensing.

[17] Zhang, Y., (2022). Ensemble modelling for improved sea temperature forecasts. Advances in Meteorology.

[18] Li, D., (2023). Leveraging transfer learning for SST prediction models. Environmental Science & Technology.

[19] Peterson, A., (2020). Anomaly detection in sea surface temperature data via clustering. IEEE Transactions on Geoscience and Remote Sensing.

[20] Kumar, V., & Jain, S. (2023). Reinforcement learning techniques in ocean prediction systems. Ocean Engineering.

[21] Smith, L., et al. (2021). Integration of satellite and in-situ data for ocean forecasts. Journal of Climate Research.

[22] Thompson, R., (2023). Model interpretability in climate predictions: Challenges and solutions. Climate Dynamics.

[23] Johnson, M., (2021). Machine learning in earth sciences: Approaches and future directions. Earth-Science Reviews.

[24] Nishant Nilkanth Pachpor., B. Suresh Kumar., Prakash S. Prasad & Salim G., Shaikh, (2024)“Intelligent rainfall forecasting model: heuristic assisted adaptive deep temporal convolutional network with optimal feature selection”, International Journal of Intelligent Information and Database Systems Vol. 16, No. 4,.

[25] Swarna, M., Sudhakar, N., & Vadaparthi, N. (2021). An effective tropical cyclone intensity estimation model using Convolutional Neural Networks. *MAUSAM*, *72*(2), 281-290.

[26] Swarna M, N Sudhakar., & Nagesh Vadaparthi, (2024), "Recent Advances in Tropical Cyclone Forecasting", Journal of Electrical Systems (JES), 20-9s. :1956-1966.

[27] Swarna M, N Sudhakar., & Nagesh Vadaparthi,(2025) "Sea surface Temperature prediction using ARIMA model", Springer Conference, , In Press.

[28] Ghulam G. Zahid, Nuzhat Fatima, Aparna Sinha, Ajhar Hussain & Firoz Ahmad, (2024) Anomalous behaviour of Indian summer monsoon rainfall due to the change in some of the monsoon semi-permanent features in surplus and deficit years, International Journal of Hydrology Science and Technology Vol. 18, No. 2.