National Conference on Behavioural Economics and Intelligent Decision Systems for Climate Change and Sustainable Development
Journal of Advanced Research in Embedded System
Volume 13, Issue 1&2 - 2026

Research Article

# Measuring ENSO Predictability Sources Using Machine Learning Techniques

Dibyadarshini Maharatha[1], Prashant kumar Chauhan[2]

[1]Research Scholar,[2]Associate Professor, Department of Applied Science, National Institute of Technology Delhi, New Delhi, India

**DOI:** https://doi.org/10.24321/2395.3802.202608

## INFO

**Corresponding Author:**
Prashant kumar Chauhan, Department of Applied Science, National Institute of Technology Delhi, New Delhi, India
**E-mail Id:**
prashantkumar@nitdelhi.ac.in
**Orcid Id:**
http://orcid.org/0000-0001-8480-7490

**How to cite this article:**
Maharatha D, Chauhan P K. Measuring ENSO Predictability Sources Using Machine Learning Techniques. *J Adv Res Embed Sys* 2026; 13(1&2): 68-72.

Date of Submission: 2025-10-04
Date of Acceptance: 2025-10-28

## ABSTRACT

Important global climate anomalies, such as the El Niño-Southern Oscillation (ENSO), can impact socio-economic systems, water resources, and agriculture. Strategies for disaster preparedness and climate adaptation rely on accurate ENSO event predictions. Here, we use machine learning techniques to investigate potential causes of ENSO prediction. The ENSO indices (Niño 3.4 and Niño 4) were modelled using historical records of sea surface temperature (SST), sea level pressure (SLP), and subsurface ocean temperatures. To unravel the data's patterns, including its non-linear linkages and temporal dependencies, we employed Random Forest (RF), Gradient Boosting (GB), and Long Short-Term Memory (LSTM) networks. To evaluate model performance, we used RMSE, MAE, and correlation coefficients for performance metrics, as well as feature importance metrics and seasonal analysis to evaluate phase-dependent predictability (i.e., modelling winter ENSO and summer ENSO). The results indicate that the LSTM model yields performance levels superior to the tree-based models and predicts the highest levels of ENSO prediction accuracy; the strongest predictors were SST anomalies in the Niño 3.4 region and subsurface temperatures. In terms of seasonal predictability, we found that ENSO events during winter months were more predictable than summer months, which aligns with phase-locking behaviour. Overall, this research shows that machine learning can provide reliable understanding of ENSO dynamics and identify the important climate drivers, which together provides a step towards better forecasting and early warning capacity.

**Keywords:** Enso, El Niño–Southern Oscillation, Machine Learning, Lstm, Random Forest, Gradient Boosting, Sea Surface Temperature, Sea Level Pressure

## Introduction

Sea surface temperatures (SST) and air pressure in the equatorial Pacific undergo periodic change as part of the El Niño-Southern Oscillation (ENSO), a significant climate phenomenon.[1] Changes to precipitation, temperature extremes, agriculture, water resources, and socio-economic systems are among the significant effects of ENSO-driven events, which encompass both warm (El Niño) and cold (La Niña) phases.[2] Predicting ENSO events appropriately is important to prepare for disasters, inform agricultural practices, and adapt to climate change.[3]

While conventional forecasting strategies such as statistical models and dynamical models have shortcomings in considering coupled and non-linear ocean-atmosphere processes that underpin ENSO variability[4], recent developments in machine learning (ML) hold promise to inform us about the next generation of ML tools that allow us to analyse very large climate datasets [5], discover latent patterns, and compare accuracy improvements to historical approaches[6]. Patterns in historical ocean and atmospheric data may be used to identify dominant predictors and metrics based on their weight on predictability, leading to new insights that could potentially outperform conventional predictions.[7]

## Objectives of the Study

- To model ENSO indices using advanced machine learning techniques and evaluate their predictive skill.
- To identify and quantify the dominant oceanic and atmospheric variables contributing to ENSO predictability.
- To assess seasonal variations in ENSO predictability and understand phase-dependent forecasting accuracy.

## Literature Review

Colfescu et al. (2024)[8] analysed oceanic characteristics (SST, heat content) and atmospheric variables (near-surface zonal wind, U10) to determine the long-term predictability of ENSO using machine learning. During the late fall to late spring, they discovered that tropical SST was the most influential factor on predicting skill; however, U10 on its own showed similar forecasting ability for lead durations of 11-21 months. An atmospheric bridge in the Pacific explains the long-lead signal as a result of linked wind-SST interactions over the Indian Ocean. A preliminary link between anomalies in the western Indian Ocean's sea surface temperature and anomalies in the eastern Indian Ocean's wind speed has been suggested via linear correlation studies. The study emphasised the significance of U10 for ENSO predictions beyond one year.

Dijkstra et al. (2019)[9] assessed progress toward forecasting El Niño events in the tropical Pacific, focusing on machine learning techniques derived from artificial neural networks. It was indicated that typical statistical and dynamical models experienced substantial skill loss when forecasting for lead times greater than six months. In contrast, initial machine learning approaches showed superior forecasting skill for lead times longer than 12 months. In addition, the article discussed ways to select the most effective forecasting characteristics, particularly for attributes derived from complex networks, and offered a critical viewpoint regarding potential future directions to improve ENSO forecasting.

Lima et al. (2015)[10] studied how well statistical and dynamical models could predict ENSO, and they all failed miserably when it came to making predictions more than six months out. They unveiled a state-of-the-art ENSO forecast model that was created at UNB/CWC using regularised least squares regression and a nonlinear dimensionality reduction technique. For longer lead times in particular, their model outperformed other ENSO forecast models. Contrary to numerous other models and in line with observations, the UNB/CWC model did not anticipate a large El Niño event in 2014, according to an examination of that event. In particular, with longer lead times, this work demonstrated that nonlinear statistical approaches have the ability to improve the accuracy of ENSO predictions.

## Research Methodology

This study aims to identify the sources of ENSO predictability using machine learning techniques.

## Research Design

This study adopts a quantitative, analytical research approach that utilises historical datasets on the ocean and atmosphere to develop ENSO indices. This research combines comparative modelling with feature attribution and seasonal analysis to characterise both predictive skill and dominant predictors.

## Data Sources

This study uses reliable datasets from 1980 to 2023 detailing key oceanic and atmospheric variables that influence ENSO. Sea surface temperature (SST) data is from NOAA OISST, while sea level pressure (SLP) is obtained from ERA5 reanalysis. Subsurface ocean temperatures at 50 m and 100 m depth come from Argo floats and ORAS5. SST anomalies from Niño 3.4 and Niño 4 indices serve as target variables. Together, these datasets are a complete representation of the coupled ocean-atmosphere processes that underlie ENSO.

## Data Preprocessing

The climate datasets are initially prepared for compatibility and robustness of machine learning analysis. For missing values, we use the k-nearest neighbours (KNN) method, which preserves local time correlations and local spatial correlations. All variables are scaled to zero mean and unit variance to avoid potential bias, and seasonal trends are adjusted out to focus only on interannual variations that contribute to ENSO. Finally, all datasets are aligned on a monthly scale to temporally synchronise predictors to ENSO indices.

## Machine Learning Models

For Random Forest (RF), 200 estimators were used with maximum depth determined through grid search. Gradient Boosting (GB) employed a learning rate of 0.05 with 300 estimators and a subsample ratio of 0.8 to prevent overfitting. The Long Short-Term Memory (LSTM) network consisted of two hidden layers with 64 and 32 units, ReLU activation for intermediate layers, and a linear output layer.

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202608

Dropout regularisation (rate = 0.2) was applied to mitigate overfitting. The model was trained for 100 epochs using the Adam optimiser with a learning rate of 0.001 and a batch size of 32.

- **Random Forest (RF):** Conducts non-linear regression and evaluates feature importance.
- **Gradient Boosting (GB):** Captures complex interactions and increases prediction accuracy.
- **Long Short-Term Memory (LSTM):** Represents sequential data with respect to time dependence.

The training process consisted of 80% of data used for training and 20% for testing; hyperparameter tuning was performed through grid search and cross validation. Additionally, temporal cross-validation was implemented using a sliding window approach. Models were trained on earlier segments of the time series and validated on subsequent periods to ensure realistic temporal forecasting conditions and prevent data leakage across time.

## Evaluation Metrics

Model performance is evaluated using RMSE and MAE to quantify prediction errors and the correlation coefficient (R) to assess the agreement between observed and predicted ENSO indices. In the RF and GB models, feature importance scores specify the contribution of each predictor to the overall model. The models are also explored seasonally for DJF, MAM, JJA, and SON, analysing associations between phase-dependent behaviour and ENSO predictability.

## Analytical Approach

Using a comparative modelling strategy, the study evaluates the predictive skill of RF, GB, and LSTM models, including the effects of temporal dependencies and nonlinearities in predicting ENSO. The study uses feature importance metrics to quantify the influences of the sea surface temperature (SST), subsurface temperatures, sea level pressure (SLP), and wind anomalies (as well as other variables) on the ENSO indices. Seasonal performance is also investigated to show seasons of generally higher or lower predictability, providing information about phase locking and stochastic variability. The approach allows robust quantification of ENSO predictability while simultaneously identifying relevant climate drivers.

## Results And Discussion

The comparative performance shows the LSTM's capability to handle sequential dependencies and non-linearities inherent in ENSO dynamics. The superior correlation and reduced error metrics confirm that recurrent neural architectures can learn temporal memory patterns more effectively than tree-based methods. This is consistent with prior ENSO prediction studies that emphasise temporal coherence as a critical predictor of skill.

This section presents how the performance of the machine learning models predicts ENSO indices, explores the relative influence of ocean and atmospheric predictor variables, and examines seasonal variability with respect to predictive skill to provide a holistic understanding of predictability associated with ENSO.

## Model Performance

Table 1 summarizes the outcomes of the three machine learning models' predicted performance on the Niño 3.4 index.

The LSTM model produced the lowest values of RMSE (0.59°C) and MAE (0.46°C) as well as the greatest correlation (R = 0.91), suggesting improved predictive skill over tree-based models. Gradient boosting performed marginally better than random forest, most likely because the model was able to account for complex non-linear interactions among predictors. These results underscore the value of representing temporal dependencies with respect to ENSO dynamics, as LSTM models are proposed to better incorporate sequential data.

## Feature Importance Analysis

To assess the strongest predictors of ENSO, feature importance assesses from RF and GB models. Fig. 1 lists the 10 most important predictors.

SST anomalies in the Niño 3.4 area have been consistently the most crucial predictor across both models, illustrating the fundamental role of central Pacific Ocean temperatures in driving ENSO events. Furthermore, subsurface temperatures (at both 50 m and 100 m depths) were also deemed essential predictors, indicating that heat content below the surface has a strong impact on the evolution of ENSO. In addition, significant contributions were also seen from various atmospheric variables (specifically SLT and wind anomalies), confirming the associated ocean-atmosphere nature of processes involved in ENSO dynamics. Taken together, oceanic and atmospheric predictors strongly indicate that proper ENSO forecasts should include multiple climate-related predictors rather than be limited to SSTs.

**Table 1.Model Performance Metrics For Enso Prediction**

| Model | RMSE (°C) | MAE (°C) | R (Correlation) |
|---|---|---|---|
| Random Forest | 0.68 | 0.52 | 0.87 |
| Gradient Boosting | 0.62 | 0.49 | 0.89 |
| LSTM | 0.59 | 0.46 | 0.91 |

**ISSN: 2395-3802**
**DOI: https://doi.org/10.24321/2395.3802.202608**

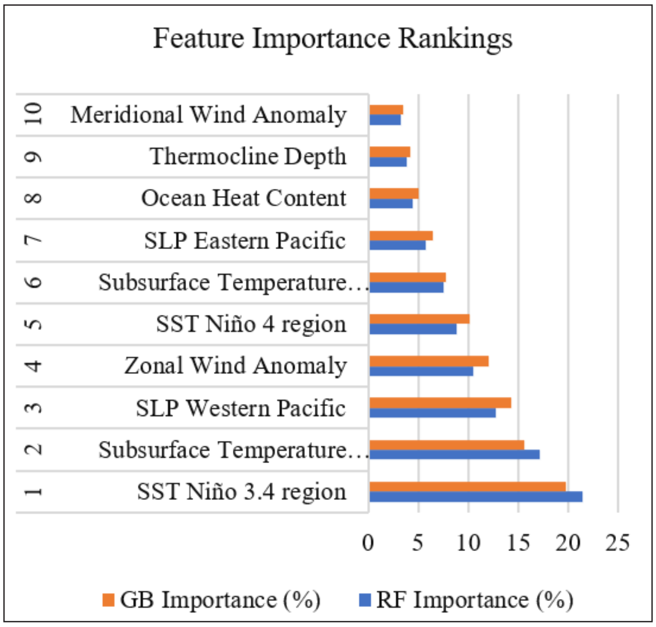


**Figure 1.Feature Importance Rankings**

## Limitation and Future Work

Although the models perform well, the study is limited by the spatial resolution of datasets and potential errors in reanalysis data. Future work may include hybrid models that integrate physical constraints or use attention-based architectures (e.g., Transformers) for better interpretability and long-lead prediction. Additionally, incorporating ocean reanalysis ensembles may improve robustness.

## Temporal and Seasonal Analysis

The predictive skill of the LSTM model was evaluated for seasonal variations representing phase-dependent ENSO predictability. Table II provides a summary of the RMSE and correlation (R) for various seasons.

**Table 2.Seasonal Predictive Skill Of Lstm Model**

| Season | RMSE (°C) | R (Correlation) |
|---|---|---|
| DJF (Winter) | 0.57 | 0.92 |
| MAM (Spring) | 0.61 | 0.90 |
| JJA (Summer) | 0.62 | 0.88 |
| SON (Autumn) | 0.60 | 0.89 |

The most accurate predictions of ENSO events took place in winter (DJF) (R = 0.92), whereas the predictions made in summer (JJA) displayed the lowest predictability (R = 0.88). This finding is consistent with the well-known seasonal phenomenon of the phase-locking mechanism in ENSO events, with ENSO typically peaking in boreal winter, thus making prediction more certain. ENSO's slightly lower performance in the summer may be due to the effects of stochastic processes at those ENSO transitional phases, which likely reduce the model's predictability. These seasonal differences are important to consider for operational forecasting and early warning applications.

## Conclusion

El Niño-Southern Oscillation (ENSO) can be better understood and predicted with the help of machine learning techniques, as demonstrated in this study. Of the approaches tested, the Long Short-Term Memory (LSTM) network showed the best predicting ability, highlighting the importance of capturing ENSO's temporal features. Among the notable factors that contribute to the predictability of ENSO, according to the feature importance analysis, are sea surface temperature anomalies in the Niño 3.4 zone, subsurface ocean temperature, and atmospheric variables such as sea level pressure and wind anomalies. Seasonal evaluation showed that ENSO events are the most predictable in winter (DJF), consistent with established phase-locking tendencies of ENSO; summer events resulted in lower predictability due to transitional and stochastic, or noise, processes. Overall, the results indicate that incorporating oceanic and atmospheric variables into machine learning frameworks improves predictive ability over traditional ML-based techniques.

## References

1. Ibebuchi CC, Rainey S, Obarein OA, Silva A, Lee CC. Comparison of machine learning models in forecasting different ENSO types. Physica Scripta. 2024 Jul 30;99(8):086007.
2. Li T, Tang Y, Lian T, Hu A. Quantifying the relative contributions of the global oceans to ENSO predictability with deep learning. Geophysical Research Letters. 2024 Mar 16;51(5):e2023GL106584.
3. Liu Y, Duffy K, Dy JG, Ganguly AR. Explainable deep learning for insights in El Niño and river flows. Nature Communications. 2023 Jan 20;14(1):339.
4. Pal M, Maity R, Ratnam JV, Nonaka M, Behera SK. Long-lead prediction of ENSO modoki index using machine learning algorithms. Scientific reports. 2020 Jan 15;10(1):365.
5. Patil KR, Doi T, Jayanthi VR, Behera S. Deep learning for skillful long-lead ENSO forecasts. Frontiers in Climate. 2023 Jan 4;4:1058677.
6. Tang Y, Zhang RH, Liu T, Duan W, Yang D, Zheng F, Ren H, Lian T, Gao C, Chen D, Mu M. Progress in ENSO prediction and predictability study. National Science Review. 2018 Nov 1;5(6):826-39.
7. Wang B, Feng P, Waters C, Cleverly J, Li Liu D, Yu Q. Quantifying the impacts of pre-occurred ENSO signals on wheat yield variation using machine learning in Australia. Agricultural and Forest Meteorology. 2020 Sep 15;291:108043.
8. Colfescu I, Christensen H, Gagne DJ. A machine learning-based approach to quantify ENSO sources of

predictability. Geophysical Research Letters. 2024 Jul 16;51(13):e2023GL105194.
9. Dijkstra HA, Petersik P, Hernández-García E, López C. The application of machine learning techniques to improve El Niño prediction skill. Frontiers in Physics. 2019 Oct 10;7:153.
10. Lima CH, Lall U, Jebara T, Barnston AG. Machine learning methods for ENSO analysis and prediction. InMachine Learning and Data Mining Approaches to Climate Science: Proceedings of the 4th International Workshop on Climate Informatics 2015 Sep (pp. 13-21). Cham: Springer International Publishing.

**ISSN: 2395-3802**
**DOI: https://doi.org/10.24321/2395.3802.202608**