



ENSO dataset & comparison of deep learning models for ENSO forecasting

Shabana Mir¹ · Masood Ahmad Arbab² · Sadaqat ur Rehman³

Received: 29 October 2023 / Accepted: 22 March 2024
© Crown 2024

Abstract

Forecasting the El Niño-Southern Oscillation (ENSO) is a challenging task in climatology. It is one of the main factors responsible for the Earth's interannual climatic fluctuation and can result in many climatic anomalies. The impacts include natural disasters (floods, droughts), low & high agriculture yields, price fluctuation, energy demand, availability of water resources, animal movement, and many more. This study presents a comprehensive ENSO dataset containing standard indicators and other relevant data to facilitate ENSO analysis and forecasting. To ensure the dataset's validity and reliability, we performed extensive data analysis and trained four basic deep models for time series forecasting (i.e. CNN, RNN, LSTM, and hybrids). The data analysis confirmed the accuracy and suitability of the dataset for ENSO forecasting. The LSTM model achieved the best fit to the data, leading to superior performance in forecasting ENSO events.

Keywords El Niño-Southern Oscillation (ENSO) · El Niño La Nina prediction · Climate changes · Time series analysis · Time series forecasting · Deep learning · Neural networks

Introduction

ENSO is an irregular periodic fluctuation (i.e. every 2–7 years) in wind and sea surface temperature (SST) in the central and eastern tropical Pacific Ocean. It has a significant impact on global climate patterns because it can change atmospheric circulation worldwide, which affects temperature and precipitation. Agriculture yields, commodity supply chains, energy demand, water resources, animal migration,

etc. are all impacted. The ENSO cycle has three phases: El Niño, La Niña, and Neutral (L'Heureux 2014).

- El Niño / warm phase: the unusual warm ocean temperatures in the eastern Pacific
- La Niña / cool phase: the unusual cool ocean temperatures in the eastern Pacific
- Neutral: neither El Niño nor La Niña

ENSO indicators include Oceanic Niño Index (ONI), Multivariate ENSO Index (MEI.v2), Southern Oscillation Index (SOI), Trans-Niño Index (TNI), Pacific-North American (PNA) Index, Outgoing Longwave Radiation (OLR), TAO/Triton Data Display, and sea surface temperature (SST) indices which are Niño 1+2, Niño 3, Niño 3.4, Niño 4. According to the National Oceanic and Atmospheric Administration (NOAA), the ENSO indicators are the ONI, SOI, OLR, MEI.v2, TAO/Triton Data Display, and SST anomalies (Barnston 2015; NOAA n.d.).

Linear approaches have historically dominated time series forecasting since they can solve many straightforward forecasting problems. To address the shortcomings of linear methods, nonlinear approaches have been applied to climate studies. Deep learning has been used to improve weather forecasts and predict climate variability because it is more effective at identifying weather features (Gupta 2019; Shin et al. 2022).

Communicated by: H. Babaie

✉ Sadaqat ur Rehman
s.rehman15@salford.ac.uk

Shabana Mir
smir@uetpeshawar.edu.pk

Masood Ahmad Arbab
arbabmasood@uetpeshawar.edu.pk

¹ Department of CS & IT, University of Engineering & Technology, Peshawar, Pakistan

² Department of Computer System Engineering, University of Engineering & Technology, Peshawar, Pakistan

³ School of Sciences Engineering and Environment University of Salford, Salford, UK

ENSO exhibits seasonal patterns and is a time series forecasting problem (NOAA n.d.). Deep learning holds significant potential for forecasting time series, especially learning about temporal dependencies and addressing temporal patterns like trends and seasonality. Deep neural networks support multiple inputs and outputs and can learn arbitrarily complicated mappings. Deep learning techniques for forecasting time series include convolutional neural network (CNN), recurrent neural network (RNN) such as simple RNN and long short-term memory (LSTM), and hybrids (CNN-LSTM, ConvLSTM) (Brownlee n.d.).

The contributions of this study include:

- A comprehensive ENSO dataset containing standardized monthly climate data spanning 74 years, from 1950 to 2023 (Mir 2022a)
- Data analysis to validate the dataset (Mir 2022b)
- Comparison of deep models to verify the dataset and assess how well they fit the data (Mir 2023)

Related work

The deep models used in ENSO analysis and prediction mainly include CNN, RNN, and hybrids, as shown in Table 1.

Shin et al. (Shin et al. 2022) used CNN to analyze ENSO patterns from climate model simulations. The model accurately forecasted ENSO for a 9-month lead with high correlation (~ 0.82). Ham et al. (2019) trained a convolution network using transfer learning on historical simulations and reanalysis data for multi-year ENSO predictions. The model successfully forecasted ENSO 12 months ahead, with an accuracy of 66.7%. Cao et al. (Cao et al. 2020) studied ENSO prediction based on LSTM using meteorological time-series data. They also showed that the quality of the dataset is important for good prediction results. Ha et al. (2021) used LSTM, ConvLSTM-LSTM, and ConvLSTM-GRU networks, to predict the monthly streamflow of the Yangtze River. They proved that there is an implicit relationship between ENSO and streamflow data that can be learned by the neural network. Mahesh et al. (2019) trained a convolutional and recurrent neural networks model on physical simulations for predicting monthly ENSO temperatures. Wang et al. (2023) reviewed how

deep learning has been used to predict ENSO in the past 10 years and provided a summary of the most influential papers on this topic. They discussed the potential of deep learning for ENSO prediction and the challenges that need to be addressed.

Methodology

ENSO data since 1950 was collected from reliable sources such as NOAA (NOAA n.d.), NCEI (NCEI n.d.), NCAR (NCAR n.d.), and other official weather sites (Null 2023; National Centers for Environmental Information NOAA n.d.). Data prior to 1950 is available for some indicators like Nino 3.4 SST, but for the primary and most recent indicators such as MEI.v2 and OLR, the data is unavailable. The dataset was prepared and formatted using pandas and Excel.

A range of data analysis and visualization techniques were utilized to validate the dataset by examining the following aspects: seasonal patterns in the ENSO, the strength of ENSO phases over time, the relationship between ENSO and ONI, and the correlation between different indicators. ONI is widely accepted and used by NOAA as the standard for classifying ENSO events. The correlation analysis was specifically conducted to explore the relationship between ONI and other ENSO indicators to identify potential predictor variables for ONI prediction.

During preprocessing, several operations were performed to prepare the data for training. The data was transformed into supervised learning format to get input and output sequences. This is necessary as time series forecasting problems need to be reframed as supervised learning problems. The preceding 12 months' time steps were employed to forecast the subsequent 3 months' time steps. The data was divided into training, validation, and test sets in an 80:10:10 ratio. Separate scalers were employed to normalize the input and output variables, ensuring restoration of the scaling to obtain the actual output values. To assess the suitability of the dataset and predict ONI, four deep learning models were trained: CNN, Simple RNN, LSTM, and CNN-LSTM. The data was also reshaped to adhere to the input shape specifications of the model.

Table 1 Deep learning models and data used for forecasting ENSO

References	Methods	Data
Shin et al. 2022)	CNN	Climate model data
Ham et al. 2019)	CNN + Transfer Learning	Historical simulations & reanalysis data
Cao et al. 2020	LSTM	Meteorological time series data
Ha et al. 2021)	LSTM, ConvLSTM	Streamflow and ENSO data
Mahesh et al. 2019)	CNN-LSTM	Physical simulations & observational data

To evaluate the performance and accuracy of each model in fitting the data and making predictions, several evaluation metrics such as root-mean-square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and r-squared (R^2) were used. These metrics provide a comprehensive assessment of the model's accuracy and prediction quality.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{1}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{3}$$

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{4}$$

(n sample size, y_i actual value, \hat{y}_i predicted value, \bar{y} mean)

Figure 1 summarizes the entire process discussed above from data collection to prediction, showing the basic steps involved in training a machine learning model.

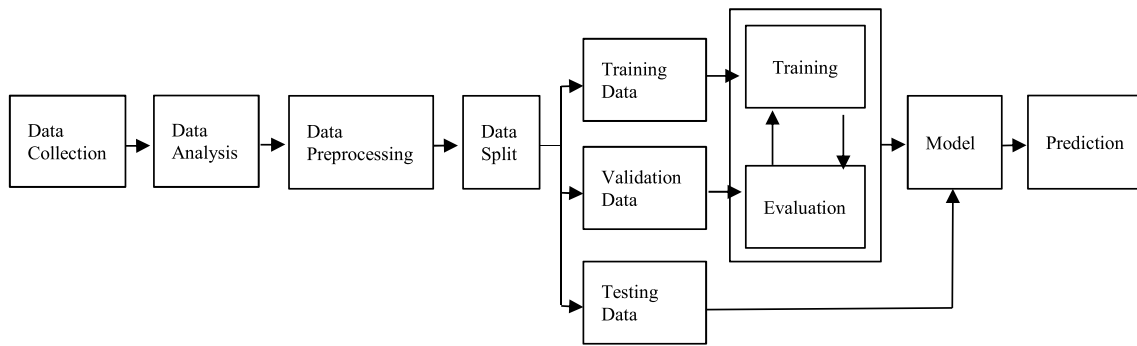


Fig. 1 The basic process of training and evaluating a deep learning model

Table 2 ENSO dataset: indicator columns details

Column	Description (unit)	Column	Description (unit)
ONI	Oceanic Nino Index (°C)	SOI	Southern Oscillation Index
OLR	Outgoing Longwave Radiation (W/m ²)	MEI.v2	Multivariate ENSO Index Version 2
TNI	Trans-Nino Index	PNA	Pacific-North American Index
Nino 1 + 2 SST	Nino 1 + 2 region Sea Surface Temperatures (°C)	Nino 1 + 2 SST Anomalies	Nino 1 + 2 region Sea Surface Temperature Anomalies (°C)
Nino 3 SST	Nino 3 region Sea Surface Temperatures (°C)	Nino 3 SST Anomalies	Nino 3 region Sea Surface Temperature Anomalies (°C)
Nino 3.4 SST	Nino 3.4 region Sea Surface Temperatures (°C)	Nino 3.4 SST Anomalies	Nino 3.4 region Sea Surface Temperature Anomalies (°C)
Nino 4 SST	Nino 4 region Sea Surface Temperatures (°C)	Nino 4 SST Anomalies	Nino 4 region Sea Surface Temperature Anomalies (°C)

Table 3 ENSO dataset: other columns details

Column	Description (unit)	Column	Description (unit)
Date	date (yyyy/mm/dd)	Year	year (yyyy) e.g. 1950
Month	month (MMM) e.g. JAN	Season (2-month)	two month season e.g. DJ
Season (3-month)	three month season e.g. DJF	Season (12-month)	twelve month season e.g. 1950–1951
Global Temperature Anomalies	monthly global temperature anomalies (°C)	ENSO Phase-Intensity	yearly ENSO phase & intensity e.g. SL (strong La Nina)

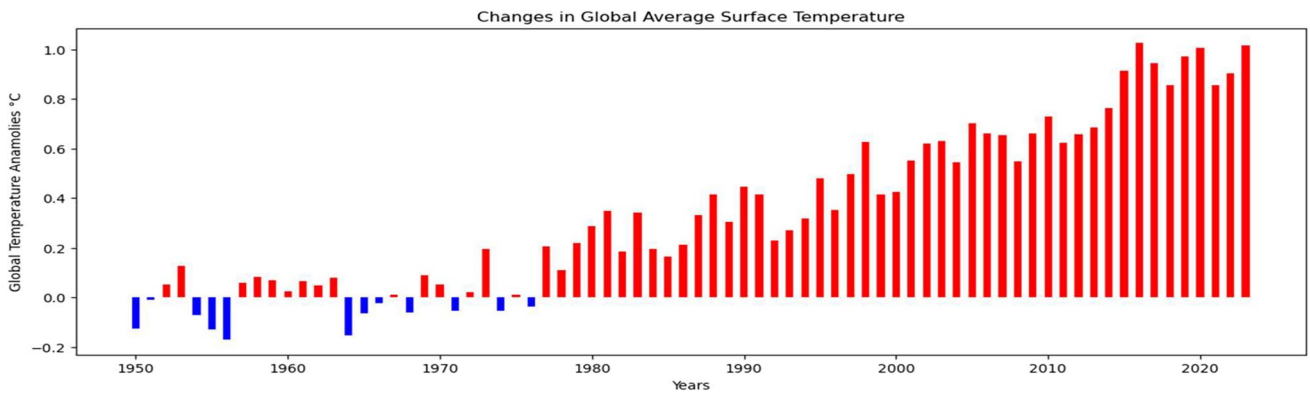


Fig. 2 Changes in global average surface temperature from 1950–2023. Blue bars indicate cooler-than-average years; red bars show warmer-than-average years

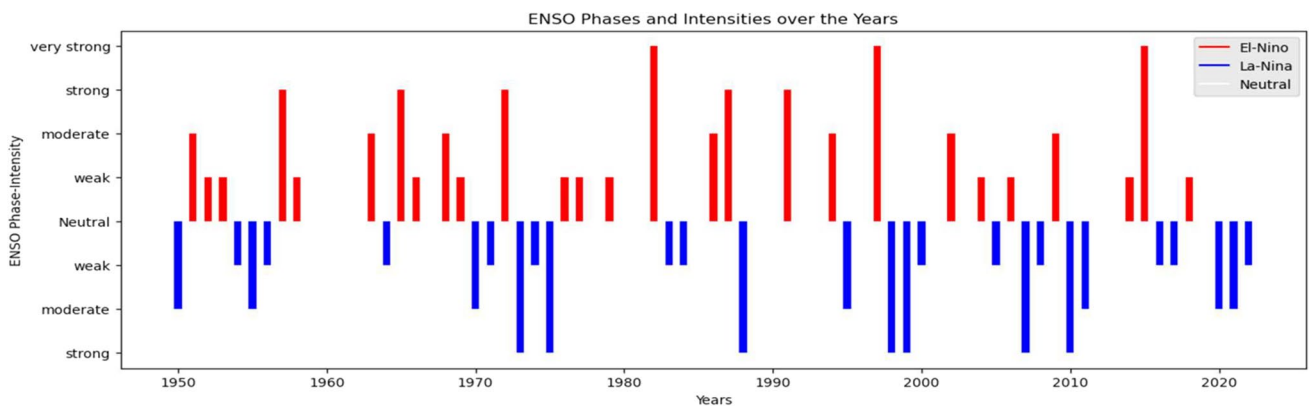


Fig. 3 ENSO events and their strength since 1950. The positive bars in red correspond to El Niño events while the negative bars in blue correspond to La Niña events

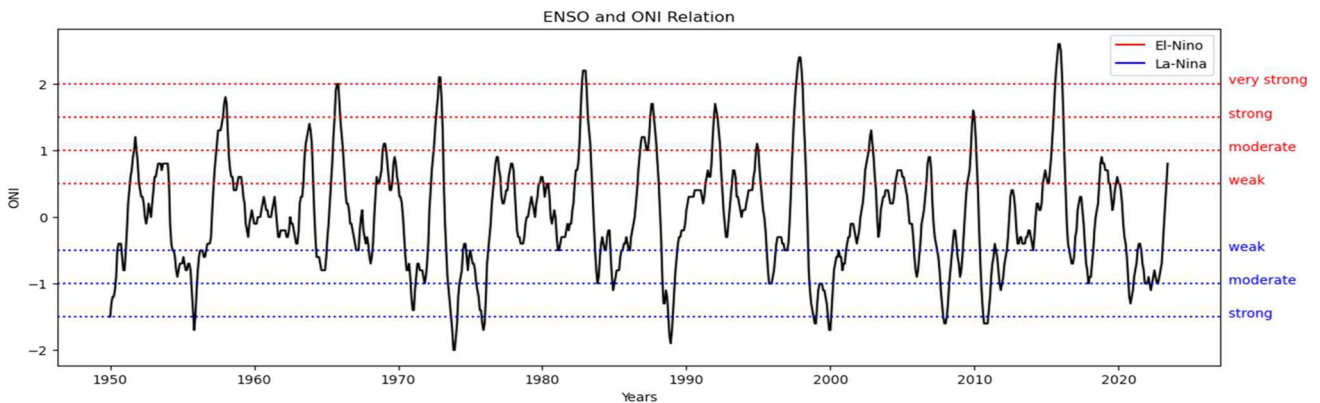


Fig. 4 Effect of ONI on defining ENSO event type and intensity. The positive peaks correspond to El Niño events while the negative peaks correspond to La Niña events

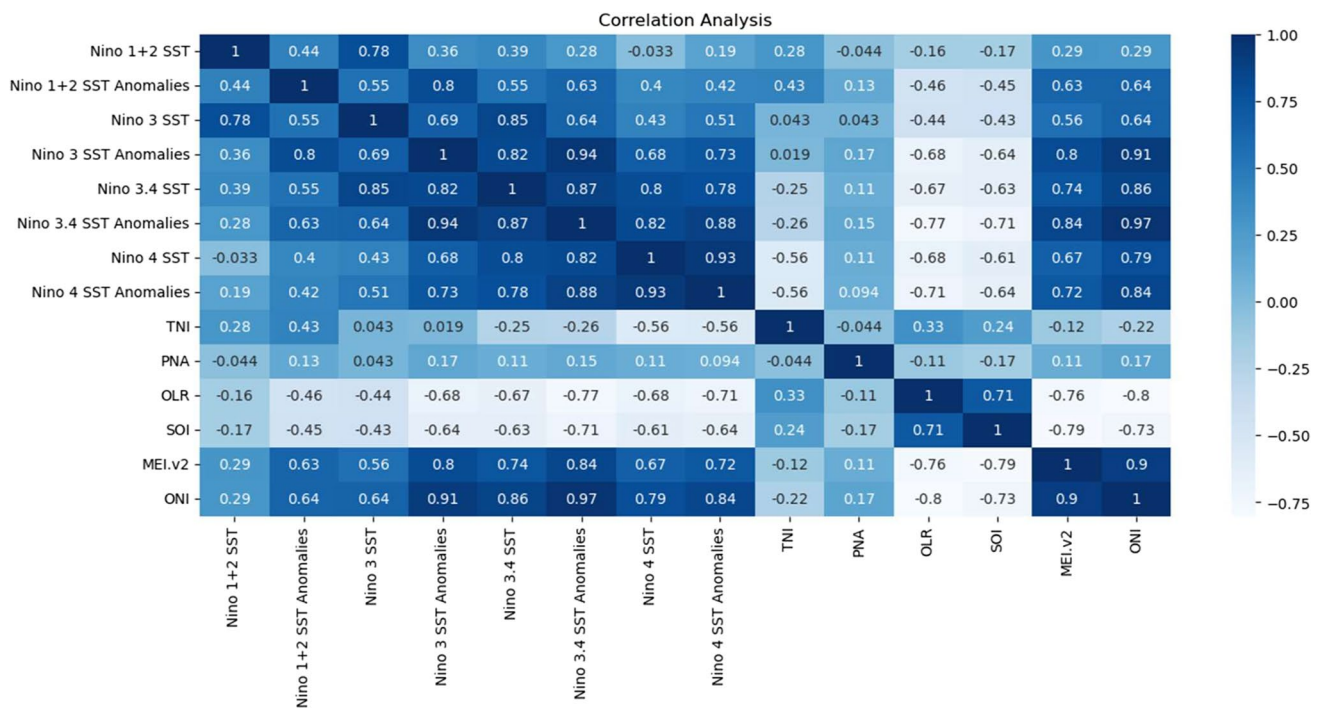


Fig. 5 Relation of ONI with other indicators. A value above +0.5 indicates a strong positive (direct) relation and below -0.5 indicates a strong negative (inverse) relation

Results

A comprehensive ENSO dataset spanning 1950–2023 was created, containing standardized monthly data. Tables 2 and 3 provides an overview of the dataset's content.

El Nino events contribute to a slight increase in global surface temperature, while La Nina events tend to have a cooling effect. Global temperature anomalies column is added in the dataset to provide data for studying the complex relationship between ENSO and global temperature. The data analysis (Fig. 2) shows a consistent rise in global temperature, thus validating the dataset.

The primary indicators for ENSO are ONI and MEI.v2. ONI, the preferred indicator by NOAA, calculates the 3-month average SST anomaly in the Nino 3.4 region. El Nino events are identified by SST anomalies at or above +0.5 °C, La Nina events by anomalies at or below -0.5 °C and Neutral by anomalies between -0.5 °C and +0.5 °C. The threshold is further categorized as Weak (0.5 to 0.9), Moderate (1.0 to 1.4), Strong (1.5 to 1.9), and Very Strong (≥ 2.0) events (Null 2023). Visualization of ONI and ENSO phase-intensity data (Figs. 3 and 4) confirms the correlation between ENSO and ONI, validating the above statements and ensuring the dataset's accuracy. The pattern in Figs. 3 and 4 also confirms the seasonality in ENSO.

The correlation analysis results (Fig. 5) indicate that ONI has a significant positive correlation with Nino 3.4 SST anomalies and MEI.v2, while it has a strong negative

Table 4 Performance comparison of the deep models

Model	Loss	MAE	MAPE	RMSE	R ²
CNN	0.009	0.075	21.349	0.097	0.610
RNN	0.004	0.051	14.559	0.064	0.826
LSTM	0.003	0.042	11.858	0.055	0.873
CNN-LSTM	0.007	0.064	17.284	0.085	0.689

correlation with OLR and SOI. Therefore, these indicators were selected as predictors for ONI prediction. The correlation analysis also validates the dataset.

Table 4 presents the performance comparison of the deep models in forecasting ENSO using test data. The results show that the LSTM model outperformed other models in terms of accuracy with a root mean square error of 0.05, and better fitting the dataset with an r-squared value of 0.87.

Conclusion

In this study, we generated a comprehensive ENSO dataset, spanning 74 years (1950–2023), by compiling standardized monthly climate data from official sources. Through data analysis, we explored the seasonality in ENSO, the relationship between ENSO and ONI, and other trends and correlations within the ENSO data. We

also trained and compared four deep-learning networks (CNN, RNN, LSTM, and CNN-LSTM) to forecast ENSO. The LSTM network achieved superior performance and demonstrated a good fit with the data. This aligns with its ability to capture long-term dependencies and patterns in time series data, crucial for accurately predicting periodic phenomena like ENSO with its 3–7 year El Niño/La Niña cycle. Both data analysis and the prediction results validate the dataset. As a valuable resource for future research, the dataset (Mir 2022a) and code (Mir 2022b, 2023) are publicly available on Kaggle, with regular updates planned. This paves the way for advancements in ENSO analysis and forecasting, leading to enhanced global preparedness for its wide-ranging impacts.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12145-024-01295-6>.

Acknowledgements We are grateful to all those people who helped us during our research and data collection. We express our courtesy to the dataset websites from where we collected the standardized monthly climate data from 1950 to 2023.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, training and testing the different forecasting algorithms were performed by Shabana Mir, Masood Ahmad Arbab and Sadaqat ur Rehman. Writing original draft preparation, Shabana Mir. Writing, review and editing by Sadaqat ur Rehman and Masood Ahmad Arbab. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The dataset generated and used during the study is available on Kaggle (Mir 2022a), <https://www.kaggle.com/datasets/shabanamir/enso-data>

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barnston A (2015) Why are there so many ENSO indexes, instead of just one? ENSO Blog, [Online]. Available: <https://www.climate.gov/news-features/blogs/enso/why-are-there-so-many-enso-indices-instead-just-one>
- Brownlee J (n.d.) Deep Learning for Time Series Forecasting. Machine Learning Mastery, [Online]. Available <https://machinelearningmastery.com/deep-learning-for-time-series-forecasting>
- Cao X, Guo Y, Liu B, Peng K, Wang G, Gao M (2020) ENSO prediction based on Long Short-Term Memory (LSTM). IOP conference series: materials science and engineering. <https://doi.org/10.1088/1757-899X/799/1/012035>
- Gupta M (2019) Linear Forecasting Models for Univariate Time Series Prediction, Medium, [Online]. Available <https://medium.com/data-science-in-your-pocket/linear-forecasting-models-for-univariate-time-series-prediction-9bff14c2b3b3>
- Ha S, Liu D, Mu L (2021) Prediction of Yangtze River streamflow based on deep learning neural network with El Niño-southern oscillation. Sci Rep. <https://doi.org/10.1038/s41598-021-90964-3>
- Ham YG, Kim JH, Luo JJ (2019) Deep learning for multi-year ENSO forecasts. Nature 573. <https://doi.org/10.1038/s41586-019-1559-7>
- L'Heureux M (2014) What is the El Niño–Southern Oscillation (ENSO) in a nutshell? ENSO Blog, [Online]. Available: <https://www.climate.gov/comment/28>
- Mahesh A, Evans M, Jain G, Castillo M, Lima A, Lunghino B, Gupta H, Gaitan C, Hunt J, Tavasoli O, Brown P, Balaji V (2019) Forecasting El Niño with convolutional and recurrent neural networks. NeurIPS 2019 Workshop on Tackling Climate Change with Machine Learning, [Online]. Available <https://www.climatechange.ai/papers/neurips2019/40>
- Mir S (2022a) El Niño-Southern Oscillation (ENSO) Data. Kaggle, [Online]. Available <https://www.kaggle.com/datasets/shabanamir/enso-data>
- Mir S (2022b) Forecasting El Niño-Southern Oscillation (ENSO). Kaggle, [Online]. Available www.kaggle.com/code/shabanamir/forecasting-el-nino-southern-oscillation-enso
- Mir S (2023) Comparing Deep Learning Models for ENSO Forecast. Kaggle, [Online]. Available www.kaggle.com/code/shabanamir/comparing-deep-learning-models-for-enso-forecast
- National Centers for Environmental Information NOAA (n.d.) Global Time Series, Climate at a Glance, [Online]. Available <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/global/time-series>
- NCAR - National Center for Atmospheric Research (n.d.) Climate Data Guide, [Online]. Available: <https://climatedataguide.ucar.edu/climate-data>
- NCEI - National Centers for Environmental Information (n.d.) Data Access - NCEI's archive of global coastal, oceanographic, geophysical, climate, and historical weather data, [Online]. Available: <https://www.ncei.noaa.gov/access/search/index>
- NOAA - National Oceanic and Atmospheric Administration (n.d.) El Niño-Southern Oscillation (ENSO), Climate Monitoring, [Online]. Available: <https://www.ncei.noaa.gov/access/monitoring/enso/>
- Null J (2023) CCM, El Niño and La Niña Years and Intensities, Golden Gate Weather Services, [Online]. Available <https://ggweather.com/enso/oni.htm>
- Shin NY, Ham YG, Kim JH, Cho M, Kug JS (2022) Application of deep learning to understanding ENSO dynamics. Artificial Intelligence for the Earth Systems. <https://doi.org/10.1175/AIES-D-21-0011.1>
- Wang GG, Cheng H, Zhang Y, Yu H (2023) ENSO analysis and prediction using deep learning: a review. Neurocomputing. <https://doi.org/10.1016/j.neucom.2022.11.078>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.