

Impact of Upper-Level Atmospheric Data on the Accuracy of Predicting Different ENSO Phases Using Hybrid Machine Learning Models in the Southern Provinces of Iran

Mahdi Malekian¹, Ali Akbar Sabziparvar^{2*}

¹ Ph.D. Candidate of Agricultural Meteorology, Department of Water Science Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran

^{2*} Professor of Meteorology, Department of Water Science Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran

(Received: 01 May 2025, Accepted: 23 August 2025)

Summary

The El Niño-Southern Oscillation (ENSO), as one of the world's most important Teleconnection phenomena, has widespread impacts on meteorological patterns and extreme events in various regions, including southern Iran. This study aims to enhance the predictive capabilities for ENSO phases, leveraging advanced machine learning techniques to address climate variability in vulnerable regions. This research investigates the impact of upper-air Parameters/variables on the prediction accuracy of ENSO phases using a hybrid ConvLSTM2D machine learning model in the southern provinces of Iran. The results indicate the key role of upper-level atmospheric data in improving the prediction of ENSO phases. Meteorological data including geopotential height, temperature, relative humidity, specific humidity, wind speed, and wind direction were collected from the ERA5 (ECMWF Reanalysis v5) reanalysis dataset at four pressure levels (1000, 850, 700, and 500 hPa) for the period 1994 to 2023. These atmospheric variables were meticulously processed to capture intricate Spatio-Temporal relationships critical for accurate climate forecasting. The proposed model was designed by combining ConvLSTM2D, LSTM, and Dense layers to extract complex Spatio-Temporal features, and the outcomes were evaluated using metrics such as R^2 , MAE, and MSE. Results showed that inclusion of upper-air data (500 and 700 hPa), compared to near-surface levels, could significantly increase prediction accuracy. The highest model performance was noted at the 500 hPa level with a three month lead time ($R^2 = 0.93$, MAE = 0.14). This also shows the 500 hPa level of the atmosphere does not change its behavior, which in turns suggests why there is a delay during change in the atmosphere. This finding underscores the importance of mid-tropospheric dynamics in long-term climate predictions. Feature Importance analysis also showed the key role of dynamic variables (geopotential height and horizontal wind components) and thermodynamic variables (temperature and specific humidity) in the mid-level atmosphere. Furthermore, adding the suggest hybrid model in addressing non-linear relationships also makes it more applicable to complex climate systems. This research emphasizes the necessity of using a wide range of atmospheric data and optimizing the time horizon in climate predictions. The results can contribute to improving water resource management and reducing atmospheric hazards in southern Iran, although challenges such as model interpretability require further research. Future studies could explore integrating multi-source datasets to further refine predictive accuracy. In order to enhance prediction accuracy and improve performance in practical applications, it is recommended that hybrid models which incorporate Attention Mechanisms enabling focusing on salient features be developed.

Keywords: Forecast, Teleconnection, Artificial Intelligence, Hybrid Models, Feature Importance