Journal of Advances in Management, Engineering and Science (JAMES), July 2026; 1(3): 10-22

Research Article

Ensemble Learning Model for El Niño Prediction and Global Warming Risk Assessment

Venkata Krishna Azith Teja Ganti, Sr Data Support Engineer, Microsoft Corporation, Charlotte NC, Charlotte NC, USA. azith.ganti.microsoft@gmail.com

Received: 11th May 2026; Accepted: 18th June 2026; Published: 9th July 2026

https://doi.org/10.65470/james.v1i03.32

Abstract – *Climate change and the growing prevalence of extreme weather events have made El Niño prediction more precise and the evaluation of risks from global warming more dependable. With significant influence of El Niño on global temperature, rainfall distribution, droughts, floods and ecosystem stability, early prediction is crucial for effective climate adaptation and disaster preparedness. In this study, a novel Hybrid Ensemble Learning Model for El Niño Prediction and Global Warming Risk Assessment is presented by using state-of-the-art climate data preparation, feature engineering, adaptive feature selection, climate similarity clustering and an adaptive weighted stacking ensemble of Random Forest, Extra Trees, LightGBM and CatBoost. This framework converts raw oceanic and atmospheric observations into informative climate indicators that improve the learning capacity of the predictive models while removing redundant information. Multiple ensemble learners' outputs are intelligently combined to improve prediction robustness, reduce over-fitting, and increase generalisation ability across different climate conditions. To improve the model transparency, SHAP-based explainable artificial intelligence is integrated to determine the contribution of the single climate variables towards the prediction outcomes, supporting the scientific interpretable decision-making. Furthermore, a Global Warming Risk Index is developed to assess future climate vulnerability and to provide useful information for environmental monitoring and early warning systems. The experimental evaluation reveals that the proposed hybrid framework outperforms the conventional machine learning approaches in terms of accuracy, precision, recall, F1-score, and error reduction, while it is robust under different climate scenarios. The proposed model offers a reliable, scalable, and interpretable solution for long-term climate forecasting, disaster risk reduction, environmental policy planning, and sustainable climate management, demonstrating the increasing potential of intelligent ensemble learning techniques to address complex global climate challenges.*

Keywords—*El Nino (ENSO) data****,*** *Boosting Algorithms, Decision trees, Machine Learning, SHAP Model, Global Warming, Environmental changes, Climatic variance.*

INTRODUCTION

In the Twenty-first century the environmental change i.e. the climate change has become one of the most affected scenarios by which the human life pattern is also affecting, so this change should be noted and try to resolve this before becoming a big threat and cause the people to death. The rising greenhouse gas emission, increase in the global temperature, change in precipitation, all these factors lead to the creation of a model that predicts the prospects and the changes by analysing the current and past data. El Nino is one of the natural climate phenomena that occurs when the surface waters of the eastern and central tropical Pacific Ocean become more warmer than usual, this changes the global weather patterns, that leads to the droughts, floods, heat waves and drastic changes in the climate. The non-linear relationship between the atmospheric and the environment variables makes the difference and cause the threats. In the recent years the Artificial Intelligence and Machine learning has enabled the solutions for the critical issues also by analysing the large data. But for the current criteria and the situation prediction of the threat using the Single model is not enough since that cannot handle the entire process efficiently, So Hybrid ensemble learning model has become one of the widest used method to find the solutions, which first learns about the data correctly and then calculates the equation as required and proceed for the further steps and in each step the double check of the methods is adopted so that the after each step this filter will help to eradicate the minute mistakes too, this enhances the models capability to solve the most toughest difficulties faced in this digital era. Motivated by these challenges this study proposes the Hybrid Ensemble Learning Model for El Nino Prediction and the Global warming Risk Assessment that integrates the ensemble learning techniques and the climate changes. The proposed system main aim is to improve the prediction accuracy, supporting the environment and the climate change criteria, that helps for the whole era that includes the government, social activist, common people and the institutions. The first and foremost step that is to be taken is to change the methodology used by people like the way of behaving with environment makes the difference that indirectly will help to resolve the issue (1). The Global warming is the main cause for this kind of changes in the environment, and these patterns will differ according to the regions and the continents (2). The constructive climate changes will effect the people as more disturbed and this may affect the future also but if that is taken as the topic into the discussion then it can be resolved since the different solutions can be given in different prospects (3). The climate situation and the environment changes should be known to all so that the awareness and the taking a step towards a better approach will happen so this will happen when the social media focuses on this area and lets the public and the government know what the future of the environment is (4). The El Nino Southern oscillations (ENSO) is one of the major ocean atmosphere climate phenomena that influences the global weather patterns, that includes the floods, droughts, and tropical storm. So, the prediction of the ENSO data and the prediction of the ENSO is essential and must for protecting the environment (5). The heat waves are the major issues and the increasing frequency and the intensity of them under the global warming leads to variations in the climate and the temperature so to avid such major issues a model that integrates the AI and ML should be integrated to find the better solution (6). The Super El Nino events on the climate have the diverse effect by which the Climate regime Shifts (CRS) are differentiated and lead to the severe destruction so this should be taken care (7). The North Pacific Oscillation (NSO) is a key driver of the El Nino events that plays a main role by finding the best solution to monitor and control the events so by implementing the NSO in the methodology can help to get the better prediction (8) The Artificial intelligence is implemented in the Early Warning System (EWS) in the environmental criteria so that the future threats can be known and take the precautionary measures (9). The event-based surveillance can be implemented so that at once the system no need to scan the whole data, it can do events by events so that the whole model performance, accuracy and the final prediction increases and system will be in the constant mode (10). The Seasonal forecasting should be implemented so that the changes in the seasons will not affect the prediction of the environment and this also enhances the stability and the performance of the whole methodology (11).

RELATED WORKS

The Extreme Gradient Boost (XGBoost) is an advanced ensemble learning, machine learning method where the several decision trees combine and achieve the best accuracy (12). The stability-based machine learning framework is implemented so that the model will work for the longer time and supports the performance by removing the minor issues (13). The usage of the machine learning approach always gives the best prediction since in this method the data is trained regularly and then tested and finally implemented by which the errors and the unwanted inbuilt chains will be removed and the format will support for the long term also (14). The implementation of the Conventional Neural Network and the Long Short-Term Memory can predict the climate driven streamflow prediction that gives the better approach, but this should be continuously monitored (15). The changes in the sand dunes are detected by which the climate changes can also be identified using the hybrid deep learning models so that the clear picture of the next state will be known (16). To manage the complex environment the method should be implemented correctly so first to monitor and maintain the data the machine learning is implemented and to explore and to perform the logics the Artificial Intelligence is employed so that the big risks can be easily solved (17). The Sea surface temperature (SST) and the Ocean Heat Content (OHC) will describe the variations in the temperature as well as the environmental factors that very accordingly by which the more changes can be seen to avoid these the two factors should be properly maintained (18). The Empirical Orthogonal Function (EOF) is applied to the global sea surface temperature data and found the variability in the ENSO, PDO (Pacific Decadal Oscillation), AMO (Atlantic Multidecadal Oscillation). So, to monitor them regularly the better model is required that is proposed in our model (19). The tropical Atlantic Sea Surface Temperature variability and the ENSO relationship is discussed and evaluated where the ocean atmospheric condition should be maintained properly and evaluated that the better AI enabled model is required to continuously monitor and keep in constant (20). For any of the digital world maintenance the data is maintained correctly so that no further issues or the problem of tracking appears so here they introduced the large database where the entire current and past data of the environmental factors updates can be found and modified but this has some limitations to use (21). The Global Climate Models (GCM) and the Transient

Climate Response (TCR) are introduced to note the uncertainty in all factors that influence the environmental changes that resulted the comparison between the older and newer model of predicting the most reliable one and shows the better results also (22). The future risk of the Compound dry hot events (CDHEs) are discussed by involving the climate hazards and the current criteria that leads to the global warming and this gives the conclusion about the future risks of increase in global warming due to several factors that affect the changes even in the ENSO oscillations and the environment too (23). The location specific deep learning method with the integration of the Sea Surface Temperature is built so that the several places can be identified easily and concluded with less latency, but the AI integration can help them to solve the logics and the predictions more accurately when compared to the results obtained (24). The Hybrid Deep Gated Recurrent Unit Convolutional Neural Network (DGCNetwork) for the SST forecasting is employed that captures the spatial and the temporal climate features that enables the improvement in the SST levels and the whole model, but the integration of the Machine learning is not employed (25). The combined deep learning method integrates the neural networks and the climate models to enhance the ENSO data and the prediction of the total model by increasing the performance of each step (26). The coupled ocean models are proposed that contains the Tropical Ocean Global Atmosphere (TOGA) program which increase the prediction of the model, that eradicates the current challenges of the ENSO but lacks in the latency that will change accordingly in several changes (27). The Climate Forecast System (CFS) is implemented that predicts using all the required parameters and integrates the AI and logics for the Deep learning this gives the better prediction compared to the other models (28). The linear Markov model is introduced that is based on the Empirical Orthogonal Functions to improve the ENSO prediction and analyse the forecasting correctly without missing the minute information too (29).Using the linear and quadratic regression the model created the stochastic invers model that improve the ENSO and the SST uses the El Nino and La Nina for the further variations and gives the good prediction (30). Many improvements and innovations are done for the El Nino and global warming reduction but somehow they lacks in one or the other way and no ensemble learning is introduced in which only one model performs all the actions and gives the good prediction in the real time so the proposed model does that and gives the unique and best prediction method.

METHODOLOGY

The Proposed Hybrid Ensemble Learning Model for El Niño Prediction and Global Warming Risk Assessment is introduced to increase the robustness, stability, and the interpretability of the climate forecasting by increasing the prediction of the whole model. First the dataset is collected from the open source Kaggle, in real time implementations the data will be coming from real time devices that is connected, then this data is pre-processed using the normalization and the other techniques which enables to remove the duplicates, missing values and the unusual data that is brought into the standard format. Then the second step is climate feature engineering where the required data i.e. the informative data like the temperature anomalies, moving averages, seasonal trends and the ocean – atmospheric reactions are retrieved so that all other factors are removed from the data that increases the perfectness and the stability of the model. To reduce the computational efficiency the feature selection method is implemented where the most influential factors are only selected for the further steps then the processed data are grouped using the climate similarity clustering that groups the historical and the current data into the Strong El Nino, Moderate El Nino, Neutral and weak La Nina and Strong La Nina so that in the final prediction the data is classified correctly and get the result. In the traditional methods the non-linear relationships are resolved using only one model that are not supporting the heavy models but in the proposed model four ensemble learners are introduced that will handle all the stuffs correctly that are namely the Random Forest, Extra trees, LightGBM, and the CatBoost then the predictions are forwarded to the Adaptive weighted stacking Ensemble where the bias and the variance are minimized and combines all the models prediction then the occurrence and the intensity of the El Nino events are computed and the Global Warming Risk Index is introduced, where the combined analysis of the ENSO, SST, atmospheric variables, greenhouse gas trends are considered. The SHAP (Shapely Additive Explanations) is incorporated to increase the transparency and the scientific interpretability. Finally, the Early warning and decision support step gives the timely alerts and the actionable insights for the Government or disaster management organizations and the forecasting team by integrating the El Nino predictions and the global warming risk levels that gives the best prediction and immediate alerts which helps to find the threats easier and take the precautionary measures.

A. Data Collection:

The First step of the model is collection of the data for this model the data is collected from the Internationally recognized organizations that helps to perform the better prediction and find the toughest methods solution, so the data is collected from the NOAA (National Oceanic Atmospheric Administration), WMO (World Meteorological organization) this includes the Sea surface Temperature (SST), Oceanic Nino Index (ONI), Southern Oscillation Index (SOI), Ocean Heat Content (OHC), rainfall, Air temperature, Carbon dioxide (CO2), Greenhouse gas concentration, wind speed, Relative Humidity that are represented as the dataset as shown in Eq (1) where the $X_1, X_2 \ldots$ represents the different parameters.

This data is taken from the open source Kaggle (31), to get the good prediction.

$$D = \{X_1, X_2, \dots \dots \dots, X_n\} \quad (1)$$

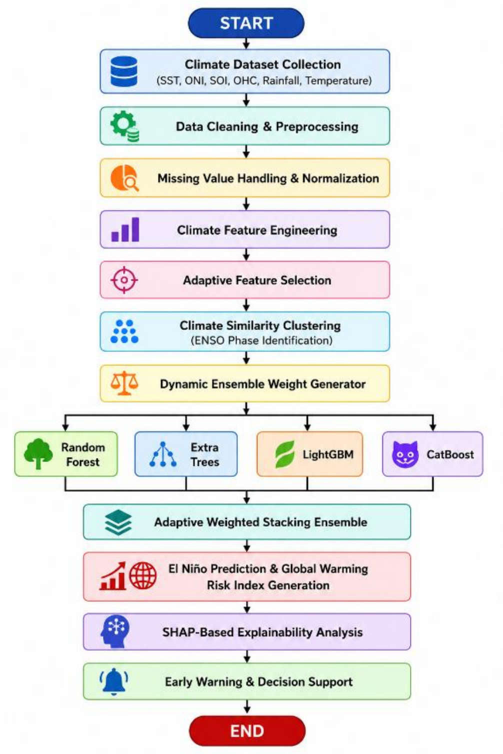

Fig 1. Flowchart of the Proposed Hybrid model

B. Cleaning of the data and the preprocessing:

The Next step is the raw data preprocessing by which the data will be coming to the stable format and this can be used in the further steps, so this data contains the inconsistency, duplicates, missing values, empty values, outliers that are removed which is generally represented by the Eq (2) as follows:

$$d_p = f(d) \quad (2)$$

Where the d represents the raw climate dataset, d_p is the Processed data, and f(d) is denoted as the cleaning and preprocessing function.

1. Normalisation and handling the Missing values:

Because of the failures of the machines or the tools used and due to the latency time or the natural calamities affects the data and gives rise to the missing values so if this data is removed or delated then the real information or the main data can be lost which cannot be recovered after so the missing values are estimated using the techniques of the interpolation, that helps to get the good results. The linear interpolation is calculated as in Eq (3) between the two neighbouring values as follows:

$$X_m = O_1 + \frac{(T_1 - T_2)}{(T_2 - T_1)}(O_2 - O_1) \quad (3)$$

Where the $O_1$is the Previous Observation, O_2 is the Next Observation, T is the current time T_1, T_2 are the neighbouring time stamps. After the missing values estimation the data is still not prepared for the next step, the Normalization is performed now to transform all the values into a common numerical scale. The direct comparison for the machine learning in further steps becomes difficult because one data is in the degree Celsius format and the other data in ppm so all should be normalized using the Eq (4) as follows:

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Where the x is the original value, x_{min} is the minimum feature value, x_{max} is the maximum feature value and the x_n is the normalized value. This Normalization helps to prevent the numerical imbalance among the climatic variables and accelerates the convergence during the model training.

C. Climate Feature Engineering:

The climate dataset is transformed into a more informative data. The Raw climatic parameters are not enough to calculate and get the results that are responsible for the formation El Nino and the global warming, so climate indicators are formed using the combination of the temporal, oceanic and the atmospheric parameters. These enables the model to recognize the hidden relationship and the oscillations and the long-term climate trends. The Proposed model defines the meaningful attributes that includes the Ocean Heat Content Trend (OHCT), Rainfall Anomaly (RA), Temperature Anomaly (TA), CO_2 Growth Rate (CGR), Wind Speed Trend (WST), and the lag features. The engineered feature vector is represented as given in the Eq (5) as follows:

$$F = \{f_1, f_2, \dots \dots \dots \dots, f_n\} \quad (5)$$

Where the F represents the engineer feature set, f_i is the individual engineered feature, m is the total number of generated features. The moving average is calculated as given in Eq (6) as below:

$$Ma_t = \frac{1}{k} \sum_{i=t-k+1}^{t} x_i \quad (6)$$

Where the Ma_t is the moving average at time t, k is the window size, x_i is the climate observation. The temperature anomaly is calculated as given in Eq (7) as follows, and the SST anomaly is calculated as in Eq (8) as follows:

$$T_a = T_{current} - T_{average} \quad (7)$$

Where the T_a represents the temperature anomaly and the $T_{current}$ is the observed temperature, $T_{average}$ is the long-time climatological value.

$$SST_A = SST_O - SST_C \quad (8)$$

Where the SST_A is the Sea Surface Temperature Anomaly, SST_O is the current SST, SST_C is the long-term average of the SST. These engineered variables provide the richer climate information and improve the prediction accuracy.

D. Adaptive Feature Selection:

Although the feature engineering gives the parameters, but those features are not always contributed equally to the prediction. Some may contain the redundancy, and some may contain the noise that increases the computational activity, and overfitting. So, the Adaptive Feature Selection (AFS) is used to automatically identify the informative features and train it. The traditional methods use the static ranking creation, but this method relies on the performance i.e. it sees the prediction and the stability towards the model and select those features only, the features with the high importance values are only moved to the next steps all other weak or the redundant features are discarded in this method. The importance of each parameter or the feature is calculated using the Eq (9) as follows:

$$S(f_i) = \frac{1}{n}\sum_{j=1}^{N} Imp_{ij} \tag{9}$$

Where the $S(f_i)$ is the importance score of the feature i the Imp_{ij} is the importance assigned during the iteration of the j and N is the number of the evaluation metrices. The optimal feature subject is represented as given in Eq (10) as given below:

$$F = \{f_i | I(f_i) \geq \theta\} \tag{10}$$

Where F is the Selected feature subset, and θ is the feature selection threshold. The features that exceed the threshold are eliminated and the features which do not exceed the threshold are sent to the next step. The Advantages of using the Adaptive Feature Selection is as follows:

- This removes the irrelevant climatic variables.
- It reduces the computational effect and the overfitting.
- It improves the prediction and makes the model as stable one.
- It also enhances the ensemble learning performance.

E. Climate similarity clustering:

This step is also considered as the ENSO phase identification, the climate behaviour changes accordingly across the different periods during the natural oscillations, and the condition of the atmosphere too. Training the prediction model on the present and static data may not support for the future so the proposed methodology introduces the climate similarity clustering stage that groups the observations with similar climatic behaviour before the ensemble learning. This clustering process organizes into the homogeneous data of the climate records to the different ENSO phase that includes the Strong El Nino, weak El Nino, Strong La Nina, Weak La Nina and the neutral conditions. Training of the separate climate patterns allows the hybrid method to analyse the specialized relationship instead of treating all the values as same. The Euclidean distance is given in the Eq (11) as follows:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - y_i)^2} \tag{11}$$

F. Dynamic Ensemble weight Generator:

The traditional methods assigns the equal weights to all the features that are available in the real time then proceed to the next step, but due to this the prediction cannot be successful since the model will not be stable, so in the proposed methodology the weights are assigned dynamically according to the current climate conditions, and also the model performance, so finally the models i.e. features that produces the higher prediction will be considered into the final model without any confusion. The dynamic weight is represented as given in Eq (12) as follows:

$$W_i = \frac{Accuracy_i}{\sum_{j=1}^{m} Accuracy_j} \tag{12}$$

Where the $\sum W_i = 1$ follows for the whole model.

G. Hybrid Ensemble Learning:

This is the core step of the whole proposed ensemble-based model, so this method uses the most powerful four ensemble learning algorithms namely Random Fores (RF), Extra Trees (ET), Light Gradient Boosting Machine (LightGBM), and the CatBoost. When the topic comes about the climate usually changes and the variability will be seen from seconds to seconds, these are considered as the non-linear problems that is caused due to the oceanic variables, difference in the temporal and spatial scales. The usage of the single Machine Learning algorithm will fail to give the better prediction and stable model so that in the proposed model the different and best algorithms are implemented by combining them i.e. training each one separately and finally combining for estimation that enhances the whole model's ability to predict and the latency also decreases.

Now the output obtained from the previous step is taken as input to this step by which the process moves forward where each algorithm learns the data independently i.e. the relationship between the climate variables and the El Nino occurrence. Then their predictions are later combined by the Adaptive weighted Stacking Ensemble to get the final prediction. This strategy always performs better in the real time also that increases the prediction, stability and the robustness and easier way of dealing with the dynamic large data. First the Input vector is represented as shown in the Eq (13) as follows:

$$X_f = \{x_1, x_2, x_3, \dots \dots, x_n\} \tag{13}$$

Where the X_f represents the climate feature vector, x is the individual climate factor, and the n is the total number of the selected features. Each ensemble model generates an independent prediction that is represented as shown in the Eq (14) as follows:

$$P = P_{rf} + P_{et} + P_{Lgbm} + P_{cb} \quad (14)$$

Where the P is the Total Prediction, P_{rf} represents the Random forest prediction, P_{et}, P_{lgbm}, P_{cb} are the predictions of the extra tress, LightGBM, CatBoost respectively.

1. Random Forest:

The Random Forest is the bagging-based learning algorithm that selects randomly the subsets of the training data and then constructs the multiple decision trees, in this every decision tree independently predicts the occurrence of the El Nino based on the processed climate variables. The final prediction is obtained from the average of the outputs of all decision trees; this reduces the model variance and increases the stability. This is used because this can deal with the non-linear relationships and it also has the noise resistance ability that will always be a positive sign to use in the climate prediction and analysis that gives the best results compared to other methods. The Random Forest is defined as in the Eq (15) as follows:

$$P_{rf} = \frac{1}{N}\sum_{i=1}^{N} t_i(x) \quad (15)$$

Where the P_{rf} is the Final prediction of the Random forest, $t_i(x)$ is the prediction generated by the i^{th} decision tree, and the N is the total number of trees. The impurity at each node is reduced using the Eq (16) as follows:

$$G = 1 - \sum_{i=1}^{c} P_i^2 \quad (16)$$

Where the G is denoted as the Gini impurity, P_i is the probability of the class i, c is the number of classes. A lower Gini index will always represent the better split of the climate data. The reasons to use the random forest is that this handles the non-linear climate relationships, reduces the overfitting through the bagging method, performs better at the high dimensional data, and provides the feature importance score.

2. Extra Trees:

The Extremely Randomized trees also referred as the Extra trees, is one of the tree-based ensembles learning method or algorithm that introduces the great randomness during the tree construction. Here each extra tree selects the split thresholds randomly, this increases the model diversity, decreases the variance and improves the generalization. The uncertainties and the variables can be removed using this method and the randomization of this model enables the model to collect the huge, diversified data and their performance. The prediction of the Extra trees is calculated using the Eq (17) as follows:

$$P_{et} = \frac{1}{m}\sum_{j=1}^{m} t_j(x) \quad (17)$$

Where the $P_{et}$is the prediction of the Extra tree, m is the total number of the trees, $t_j(x)$ is the prediction of the j^{th}randomized tree. Here the node is splitted according to the random feature selection, randomly generated split thresholds, reducing the correlation among the individual trees.

3. Light Gradient Boosting Machine:

As the name says this is one of the boosting methods that is based on the ensemble learning model algorithm. This is made only for the increase the efficiency of the prediction of the large dataset that too dynamic. This is one of the unique because it builds the decision tree in a sequential format like one after the other where each new tree attempts to correct the prediction error of the previous one. This mainly involves the leaf growth strategy so that o data is missed and replaced by the unused data, so this usually achieves the higher prediction with more stability. The prediction update Equation is given in the Eq (18) as follows:

$$F_m(x) = F_{m-1}(x) + \eta r_m(x) \quad (18)$$

Where the $F_m(x)$ is the prediction after the m iteration, $F_{m-1}(x)$ is the prediction from the previous trees, $r_m(x)$ is the new regression tree and the η represents the Learning rate. The objective function is minimized as follows given in Eq (19):

$$Object = \sum_{i=1}^{n} L(y_i, \hat{y_i}) + \Omega(h) \quad (19)$$

Where the $L(y_i, \hat{y_i})$ is the Loss function, the regularization term is represented as $\Omega(h)$, and the n is the number of the training samples. This always improves the generalization by removing the overfitting.

4. CatBoost:

This is one of the advanced gradient boosting methods, that minimizes the prediction bias through the ordered boosting and the symmetric decision trees. This will be more helpful when the data contains the mixed feature distributions and the complex nonlinear relationships. This complements the other ensemble learners by knowing the interactions among the climatic variables as said in the first step like the Sea surface Temperature, Ocean Hot content, greenhouse gas, concentration, and the atmospheric pressure. The catBoost prediction is represented as given in Eq (20) as follows:

$$P_{cb} = \sum_{k=1}^{k} \alpha_k h_k(x) \quad (20)$$

Where the P_{cb} is the prediction of the CatBoost, $h_k(x)$ is the prediction of the k^{th} boosting tree, α_k is the weight assigned to the tree, k is the number of the boosting iterations. The minimization of the squared error loss is given in Eq (21) as below:

$$L = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 \quad (21)$$

Where the L is represented as the Loss function, y_i is the actual value, $\hat{y_i}$ is the predicted value, and n is the number of the observations. During each iteration the CatBoost minimizes the loss and reduces the overfitting also.

H. Adaptive Weighted Stacking Ensemble:

After getting the output from the four algorithms those outputs are fed as inputs to this step for combining using the Adaptive Weighted Stacking Ensemble (AWSE), the previous and traditional methods gives the same importance values for all but this will be not useful, so the proposed adaptive stacking integrates the prediction according to the dynamically generated weights that is calculated using the Eq (12), the model that perform better performance contribute more significantly to the final El Nino prediction. The stacking level combines the two levels, the first one is that contains the four ensemble learners namely Random forest, Extra trees, XGBoost, and CatBoost that produces independent prediction, the second level is the meta learner stage that aggregates the predictions and considering the average as the final prediction. The final ensemble prediction is given by Eq (22) as follows:

$$\hat{F} = \sum_{i=1}^{4} w_i \, p_i \quad (22)$$

Where the $\hat{F}$ is the Final Ensemble Prediction, w_i is the adaptive weight of the i^{th} model, p_i is the prediction of the i_{th} base learners, that subjects to $\sum_{i=1}^{4} w_i = 1$. This adaptive feature increases the strengths of the multiple ensemble algorithms while minimizing the weakness of the individual models. This gives the model stability to identify the complex El Nino patterns under the variable conditions of the climate and the temperature and forms the base for the Global warming Risk step. The advantages of using this step are as follows: This integrates the strengths of the multiple ensemble learners, reduces the individual bias, increases the prediction and accuracy, improves the generalization and adaptability to the climate scenarios.

I. El Nino Prediction and the Global warming Risk index Generation:

After the adaptive stack method, the final prediction is done, to do this the framework calculates the El Nino prediction and the Global Warming Risk Generation (GWRG), by integrating the climate indicators or the features like the SST, OHC, Greenhouse gas, rainfall anomalies, temperature, anomalies, this integration a better stable format of the output. The GWRG is formulated as the weighted combination of the normalized climatic indicators, where each variable contributes according to the relative influence on the climate changes. The mathematical expression is given in the Eq (23) as follows:

$$GWRI = \sum_{i=1}^{n} \alpha_i \, f_i \quad (23)$$

Where the f_i is the Normalized climate feature, α_i is the importance weight of the feature i and n is the total number of the selected climate features. The calculation is not only enough to identify the final prediction, so the ranges are there, so the calculated GWRI values are categorized into different climate risk levels as follows:

Table I: Risk level classification based on the GWRI

GWRI Range	Risk Level
0.00-0.25	Low Risk
0.26-0.50	Medium Risk
0.51-0.75	High Risk
0.76-1.00	Extreme Risk

The Table I represents the comparison values, this enables the governments, climate scientist, forecasting team, social activists, disaster management agencies to quickly identify the danger regions and alert them.

J. SHAP Based Explainability analysis:

The machine learning based algorithms or the methods always achieve the better prediction but operates as the black box i.e. difficult to understand and move forward, to improve the interpretability and the transparency, the proposed model integrates the SHapely Additive exPlanations (SHAP), this gives the contribution of each feature by assigning the importance value based on the cooperative game theory. For every feature it calculates the individual contribution using the Eq (24) as follows:

$$\emptyset_i = \sum_{S \subseteq F\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (24)$$

Where the $\emptyset_i$ represents the SHAP value of feature i, F is the complete feature set, S is the subset of the features, f(s) is the Prediction function. A Positive SHAP value indicates that the feature increases the Predicted El Nino Probability or the climate risk, whereas the negative SHAP value decreases the prediction.

K. Early warning and the Decision support System:

This is the final and the last stage of the whole model that converts the Predicted El Nino events and the Global Warming Risk Indicator to an Intelligent Early warning Decision Support System (EWDSS). The main aim to provide the timely alerts and provide a safety measure to the whole world by analysing the entire data and finding the solution and giving the alerts. This is done as follows, by considering the El Nino values and the weights, and SHAP values the system automatically groups the data into the pre-defined warning levels. These indicates the severity of the expected climatic event, and the recommended appropriate strategies. The warning function is given in the Eq (25) as follows:

$$Warning = f(P_{ENSO}, GWRI) \quad (25)$$

Where the P_{ENSO} represents the Predicted El Nino probability, GWRI is the global Warming Risk Index.

Table II Warning categorization representation

Predicted Probability	Risk Level	Recommended Action
<0.30	Normal	Continue monitoring
0.30-0.50	Advisory	Precautionary measures
0.50-0.75	Warning	Activate mitigation plans
>0.75	Emergency	Immediate response

The Table II represents the warning categorization that is used and helpful for the future prediction and emergency calls also by which the entire world can be profitable and use this methodology for the whole lifetime without any loss.

L. Performance Evaluation metrics:

The Performance of the whole model is important when it comes to the final steps, by analysing the different ratios that got from the methodology, since the method gives the El Nino event classification and the Global warming Risk Prediction, both the classification and the regression methods are used. These helps the model to correctly predict the results and the climate forecasting. The following metrices are used for the evaluation of the whole model.

1. Mean Absolute Percentage Error:

This measures the percentage error between the actual and the predicted values, the lower value indicates the better forecasting accuracy. The MAPE is given in the Eq (26) as follows:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}|y_i - \bar{y}_i| \quad (26)$$

2. Symmetric Mean Absolute Percentage Error:

This provides the balanced percentage error that is less affected by the large or small values. It is widely used in the climate and weather forecasting. The Equation is given in Eq (27) as follows:

$$sMAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \bar{y}_i|}{\frac{|y_i - \bar{y}_i|}{2}} \quad (27)$$

3. Nash-Sutcliffe Efficiency:

The one of the most popular evaluation metrics in the climate prediction and the hydrology, the NSE value close to 1 indicates excellent predictive performance it is represented as given in the Eq (28) as follows:

$$NSE = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma\ (y_i - \bar{y}_i)^2} \quad (28)$$

Where the y_i is the Actual value, $\hat{y}_i$ is the Predicted value, $\bar{y}$ is the Mean observed value.

4. Pearson Correlation coefficient:

This measures the strength of the linear relationship between the observed and the predicted value. The value close to 1 indicates the strong positive agreement. This is given in Eq (29) as follows:

$$R = \frac{\Sigma(y_i - \bar{y})(\hat{y}_i - \hat{y})}{\sqrt{\Sigma(y_i - \bar{y})^2}\ \sqrt{\Sigma(\hat{y}_i - \hat{y})^2}} \quad (29)$$

5. Explained variance score:

This measures the how much variation in the data is explained by the model, a higher EVS means the model explains the more of the variability in the observed data. This is given in the Eq (30) as follows:

$$EVS = 1 - \frac{var(y - \hat{y})}{Var(y)} \quad (30)$$

RESULT AND DISCUSSION

The Experimental results shows that the proposed hybrid model provides the more accurate values than the existing machine learning techniques, and in one model it combines two strategies of finding the ENSO and the global warming risk prediction, this uses the several methods to analyse the data and the methods, in the below the different graphs are shown that gives the results of the whole model and the comparison graph gives the better understanding of the base model.

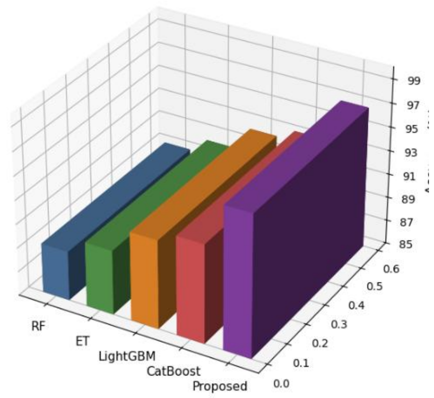

Fig. 2 Overall Performance of the proposed model

The Figure 2 shows the 3D model graph for the proposed model by which the performance can be known, and this model gives the highest prediction model.

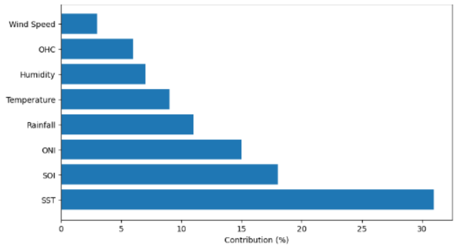

Fig 3. Climate feature condition of the model

The graph 3 shows the feature importance chart that highlights the influence of the different climate variables for the El Nino Prediction.

TABLE III PERFORMANCE MODEL (%)

MODEL	MAPE	sMAPE	NSE	Pearson
Random Forest	84.52	84.81	84.76	85.12
Extra Trees	86.13	86.45	86.28	86.74
LightGBM	88.34	88.72	89.02	89.02
CatBoost	90.18	90.44	90.82	90.08
Proposed Hybrid model	92.87	92.33	93.24	93.02

The Table III represents the proposed models and the independent single model's performance in which the better one is given by the combined model in the real time scenario.

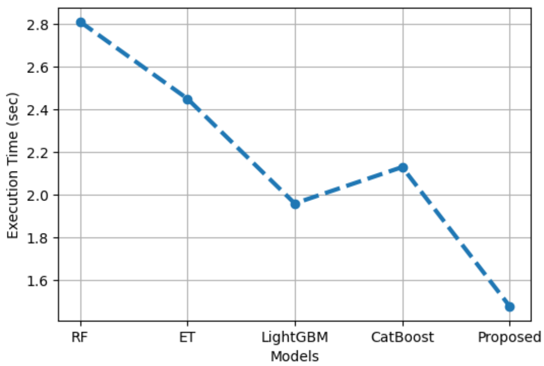

Fig 4. The Execution time for the whole model

Figure 4 shows the execution time that is noted from the scratch before data collection the time is started then the timer ends at the final prediction that gives the performance.

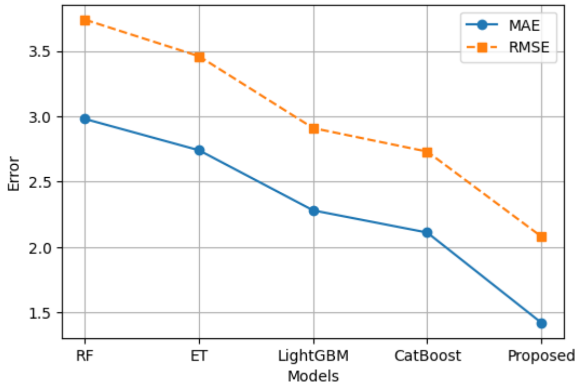

Fig 5. MAE and the RMSE Comparison graph

Figure 5 gives the Mean Absolute Error, Root Mean Squared Error of different machine learning models. This records the lowest values indicating the higher prediction accuracy.

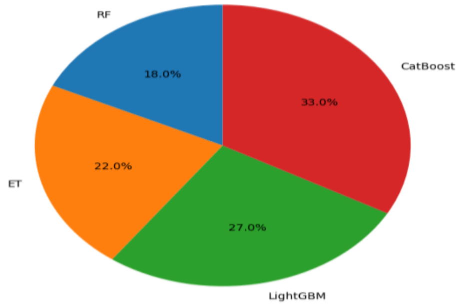

Fig. 5 Adaptive Ensemble weight graph

The Figure 5 shows the pie chart that illustrates the contribution of the ensemble model to the final prediction process. The proposed weighting strategy assigns the higher importance.

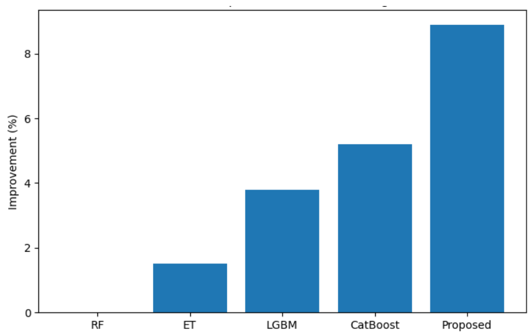

Fig. 6 Performance over the Existing models

The Figure 6 shows the performance increasing over the existing traditional models that gives the best performance compare to the other models.

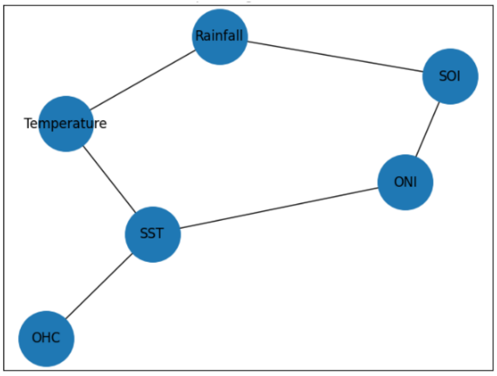

Fig.7 Relationship between the climatic variables

The Figure 7shows the relationship between the climatic variables that are displayed above, this is shown since the relationships matter when the final prediction is done and while making the decision.

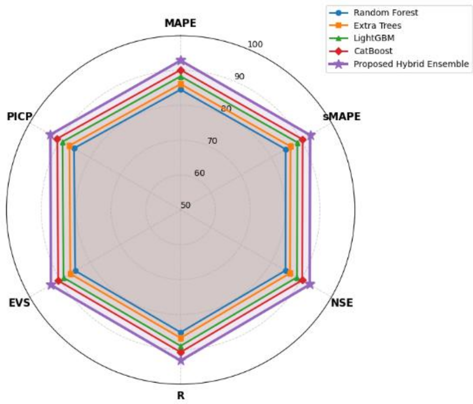

Fig. 8 The Radar chart showing whole model performance

The Figure 8 gives the radar chart by analysing the performance of the different single models and the proposed hybrid model gives the better one.

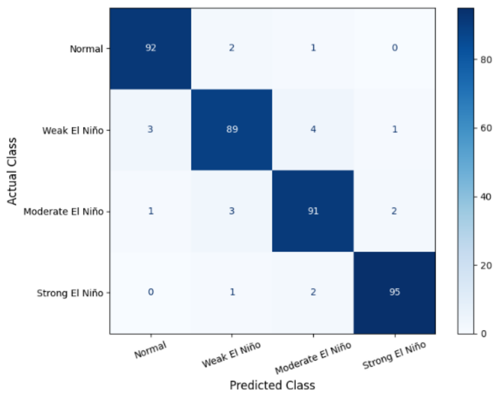

Fig. 9 Confusion matrix

The Confusion matrix shown in the Figure 9 gives the performance of the classification of the El Nino data and the proposed model correctly gives the matrix that means it identifies and classifies the data correctly.

CONCLUSION

Climate change has become one of the most important worldwide problems of the twenty first century affecting weather patterns, ecosystems, economies, and human lifestyles around the world. Among the climatic phenomena that influence the variability of the global climate, the El Niño phenomenon is especially important due to its wide impact on temperature, rainfall, droughts, floods, and extreme weather events. Hence, accurate El Nino prediction and timely global warming risk assessment are vital to improve climate resilience and facilitate evidence-based decision-making. While traditional forecasting systems have contributed to climate science, they are limited in capturing highly nonlinear interactions between oceanic, atmospheric and environmental variables. This constraint highlights the need for intelligent data-driven methods that can learn complex climate correlations and have high confidence in predictions. To address this problem, the proposed Hybrid Ensemble Learning Model combines several ensemble learning techniques into a single predictive model for climate prediction. Instead of relying on the power of a single algorithm, the proposed framework combines different learning algorithms to improve the predicted consistency and robustness in different climate scenarios. We propose a complete analytical pipeline for meaningful knowledge extraction from large scale climate datasets. The pipeline is based on advanced feature engineering, adaptive feature selection, climate similarity clustering, ensemble learning, explainable artificial intelligence and risk assessment. This multi-stage approach allows the framework to capture subtle environmental patterns, long term climatic connections and hidden interactions that can be missed by traditional statistical approaches.

One of the main contributions of this work is the emphasis on intelligent representation of climatic features. Instead of relying only on raw measurements the framework develops meaningful climatic indicators to better characterise ocean-atmosphere interactions involved in El-Nino evolution and global warming. These engineered features improve the quality of information provided to the learning algorithms, resulting in better detection of climate anomalies and long-term environmental patterns. Moreover, adaptive feature selection removes redundant information and guides the learning process to focus on the most important climate variables, thereby improving computational efficiency without compromising the quality of predictions. The hybrid ensemble architecture proposed can further improve the forecasting performance by combining the complementary characteristics of multiple ensemble learning methods. Different models interpret climate data differently. An average of predictions from many machine learning algorithms can reduce bias and variation of individual models and improve overall generalisation ability. The adaptive weighting method guarantees that the models that better predict ability under different environmental conditions receive higher weights. So, this prediction is more robust, more dependable, and more resistant to the uncertainties commonly present in long-term climate datasets in the end. This learning partnership enhances our ability to predict climate, particularly in the case of highly dynamic environmental systems that are affected by multiple interacting factors. This study not only emphasises the predictive accuracy, but also model transparency and practical usability. Explainable artificial intelligence makes it possible to interpret the influence of specific climate variables on the results of prediction. This kind of interpretability increases confidence in the model for scientists and allows researchers and policy makers to understand the environmental factors of El Niño genesis and the risk of global warming. Unlike a black-box prediction system, the proposed framework provides valuable explanations for enabling scientific validation and informed climate-related decision-making.

This study proposes a comprehensive and intelligent framework for El Nino prediction and global warming risk assessment by integrating advanced ensemble learning with climate-specific feature engineering, adaptive model fusion, explainable artificial intelligence, and climate risk analysis. The proposed methodology enhances the forecast reliability and provides interpretable and intelligible insights that can be exploited for the development of scientific research, environmental management, and sustainable policy development. The proposed paradigm helps societies to be more resilient and able to adapt to future climate variability by identifying earlier climatic anomalies and by better understanding the risks of global warming. The findings of this study demonstrate the promising potentials of hybrid artificial intelligence methods to improve the climate forecast research and provide a solid platform for future studies in intelligent environmental monitoring and global climate risk management.

Acknowledgement

The author would like to appreciate the effort of the editors and reviewers.

Author Contributions

All authors are equally contributed.

Conflict of Interests

The authors declare that they have no conflicts of interest.

Ethics Approval

There are no human subjects in this article and informed consent is not applicable.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not- for-profit sectors.

REFERENCES

1. Zaelke James Cameron D, Cameron J. Issue 2 Article 4 1990 "Global Warming and Climate Change-An Overview of the International Legal Process. American University International Law Review [Internet]. 1990. Available from: http://digitalcommons.wcl.american.edu/auilr
2. Gan Q, Leung JCH, Dong W, Yang Y, Zhao D, Li Y, et al. Beyond global mean temperature: increasing asymmetry of global warming in past and future climate change. NPJ Clim Atmos Sci. 2026 Dec 1;9(1). doi:10.1038/s41612-026-01411-x
3. Said N, Wölfl V. Impact of Constructive Narratives About Climate Change on Learned Helplessness and Motivation to Engage in Climate Action. Environ Behav. 2025 Jan 1;57(1–2):75–117. doi:10.1177/00139165251315576
4. Lineman M, Do Y, Kim JY, Joo GJ. Talking about climate change and global warming. PLoS One. 2015 Sep 29;10(9). doi:10.1371/journal.pone.0138996 PubMed PMID: 26418127.
5. Guilyardi E, Wittenberg A, Fedorov A, Collins M, Wang C, Capotondi A, et al. Understanding El Niño in ocean-atmosphere general circulation models: Progress and challenges. Bulletin of the American Meteorological Society. 2009. p. 325–40. doi:10.1175/2008BAMS2387.1
6. Vialard J, Jin FF, McPhaden MJ, Fedorov A, Cai W, An SI, et al. The El Niño Southern Oscillation (ENSO) Recharge Oscillator Conceptual Model: Achievements and Future Prospects. Reviews of Geophysics. John Wiley and Sons Inc; 2025. doi:10.1029/2024RG000843
7. De Castro Santos MA, Vega-Oliveros DA, Zhao L, Berton L. Classifying El niño-southern oscillation combining network science and machine learning. IEEE Access. 2020;8:55711–23. doi:10.1109/ACCESS.2020.2982035
8. Ding R, Tseng Y -H, Di Lorenzo E, Shi L, Li J, Yu JY, et al. Multi-year El Niño events tied to the North Pacific Oscillation. Nat Commun. 2022 Dec

1;13(1). doi:10.1038/s41467-022-31516-9 PubMed PMID: 35790767.

9. Reichstein M, Benson V, Blunk J, Camps-Valls G, Creutzig F, Fearnley CJ, et al. Early warning of complex climate risk with integrated artificial intelligence. Nature Communications. 2025 Dec 1;16(1). doi:10.1038/s41467-025-57640-w PubMed PMID: 40089483.
10. Rees E, Ng V, Gachon P, Mawudeku A, McKenney D, Pedlar J, et al. Risk assessment strategies for early detection and prediction of infectious disease outbreaks associated with climate change. Canada Communicable Disease Report. 2019 May 2;45(5):119–26. doi:10.14745/ccdr.v45i05a02 PubMed PMID: 31285702.
11. Brunet G, Parsons DB, Ivanov D, Lee B, Bauer P, Bernier NB, et al. Advancing Weather and Climate Forecasting for Our Changing World. Bull Am Meteorol Soc. 2023;104(4):E909–27. doi:10.1175/BAMS-D-21-0262.1
12. Bhusal S, Brake N, Nur AS, Feizbahr M, Hariri Asli H, Kandel M. An Integrated GIS and Explainable AI Framework for Climate-Resilient Municipal Pavement Management: Quantifying the Influence of Maintenance, Hydrological, and Environmental Factors on Pavement Condition Index (PCI). Sustainability. 2026 Jun 26;18(13):6510. doi:10.3390/su18136510
13. Gao Y, Zhao D, Liang B, Yang X, Xue X. Remote Sensing Monitoring of Soil Salinization Based on Bootstrap-Boruta Feature Stability Assessment: A Case Study in Minqin Lake Region. Remote Sens (Basel). 2026 Jan 12;18(2):245. doi:10.3390/rs18020245
14. Farhadi S, Tatullo S, Ferrian F. Comparative analysis of ensemble learning techniques for enhanced fatigue life prediction. Sci Rep. 2025 Dec 1;15(1). doi:10.1038/s41598-024-79476-y PubMed PMID: 40169662.
15. Shahariar S, Zobeyer H, Jahan N. Deep learning-based prediction of flood hazard and future streamflow changes in the Brahmaputra River Basin under CMIP6 climate change scenarios. J Hydrol Reg Stud. 2026 Jun 1;65. doi: 10.1016/j.ejrh.2026.103357
16. Lu H, Mokarram M. Fusion of deep neural networks and infrared wave analysis for climate pattern prediction and shifting sand hazard Assessment. International Journal of Applied Earth Observation and Geoinformation. 2025 Jul 1;141. doi: 10.1016/j.jag.2025.104649
17. Khan MI, Asfand F, Asif M, Al-Ansari T, Farooq M, Kurniawan TA, et al. Energy system resilience as an all-hazards system-of-systems property under climate and security challenges. Applied Energy. Elsevier Ltd; 2026. doi: 10.1016/j.apenergy.2026.128033
18. Cheng L, Abraham J, Trenberth KE, Boyer T, Mann ME, Zhu J, et al. New Record Ocean Temperatures and Related Climate Indicators in 2023. Adv Atmos Sci. 2024 Jun 1;41(6):1068–82. doi:10.1007/s00376-024-3378-5
19. Messié M, Chavez F. Global modes of sea surface temperature variability in relation to regional climate indices. J Clim. 2011 Aug;24(16):4314–31. doi:10.1175/2011JCLI3941.1
20. Enfield DB, Mayer DA. Tropical atlantic sea surface temperature variability and its relation to El Niño-Southern Oscillation. J Geophys Res Oceans. 1997 Jan 15;102(1):929–45. doi:10.1029/96jc03296
21. Ishii M, Mori N. d4PDF: large-ensemble and high-resolution climate simulations for global warming risk assessment. Progress in Earth and Planetary Science. Springer Science and Business Media Deutschland GmbH; 2020. doi:10.1186/s40645-020-00367-7
22. Scafetta N. Impacts and risks of “realistic” global warming projections for the 21st century. Geoscience Frontiers. 2024 Mar 1;15(2). doi: 10.1016/j.gsf.2023.101774
23. Zhang Y, Hao Z, Jiang Y, Singh VP. Global warming increases risk from compound dry-hot events to human and agricultural systems. International Journal of Climatology. 2023 Nov 30;43(14):6706–19. doi:10.1002/joc.8229
24. Sarkar PP, Janardhan P, Roy P. Prediction of sea surface temperatures using deep learning neural networks. SN Appl Sci. 2020 Aug 1;2(8). doi:10.1007/s42452-020-03239-3
25. Yu X, Shi S, Xu L, Liu Y, Miao Q, Sun M. A Novel Method for Sea Surface Temperature Prediction Based on Deep Learning. Math Probl Eng. 2020;2020. doi:10.1155/2020/6387173
26. Chen Y, Jin Y, Liu Z, Shen X, Chen X, Lin X, et al. Combined dynamical-deep learning ENSO forecasts. Nature Communications. 2025 Dec 1;16(1). doi:10.1038/s41467-025-59173-8 PubMed PMID: 40274886.
27. Latif M, Anderson D, Barnett T, Cane M, Kleeman R, Leetmaa A, et al. A review of the predictability and prediction of ENSO. J Geophys Res Oceans. 1998 Jun 29;103(C7):14375–93. doi:10.1029/97JC03413
28. Kirtman BP, Min D. Multimodel ensemble ENSO prediction with CCSM and CFS. Mon Weather Rev. 2009;137(9):2908–30. doi:10.1175/2009MWR2672.1
29. XUE Y, CANE MA, ZEBIAK SE, BLUMENTHAL MB. On the prediction of ENSO: a study with a low-order Markov model. Tellus A. 1994;46(4):512–28. doi:10.1034/j.1600-0870.1994. 00013.x
30. Kondrashov D, Kravtsov S, Robertson AW, Ghil M. A Hierarchy of Data-Based ENSO Models. 2005.
31. El Nino-Southern Oscillation (ENSO) Data [Internet]. [cited 2026 Jul 12]. Available from:

https://www.kaggle.com/datasets/shabanamir/enso
-data