Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:42
https://doi.org/10.28919/cmbn/8371
ISSN: 2052-2541

AN INVESTIGATION INTO IMPROVING EL NIÑO-SOUTHERN OSCILLATION PREDICTION BASED ON TEMPORAL TRANSFORMER ARCHITECTURE

RENALDY FREDYAN*, KARLI EKA SETIAWAN

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Abstract: Precisely forecasting the EI Nino-Southern Oscillation (ENSO) holds significant importance in predicting seasonal climate. The Recurrent Neural Network (RNN) has been demonstrated to be the most efficacious approach for ENSO prediction. The localized nature of the recurrent neuron poses a challenge in capturing distant antecedents of ENSO. The transformer architecture has been utilized in the domain of natural language processing (NLP) for a significant period of time due to its capacity to attend to global features. This study presents the introduction of the ENSO transformer with recurrent neuron to the ENSO research community. The current investigation demonstrates the efficacy of the ENSO Transformer model in accurately predicting the upcoming monthly mean Nino index. As the lead time increased, a temporal progression was observed in the activation map values. The study's results indicate that various climatic precursors of ENSO events have a significant impact, and each of them exhibits distinct temporal patterns. This suggests that the transformer with recurrent neuron model could be a useful tool for diagnosis. The present research suggests employing the ENSO Transformer RNN in tandem with the variant-based deep learning approach to achieve short-term prediction. The present investigation utilizes a comprehensive dataset that covers the

*Corresponding author
E-mail address: renaldy.fredyan@binus.ac.id
Received November 30, 2023

complete Nino region spanning from 1950 to 2019. In contrast to other frequently employed forecasting models, the model we put forth demonstrated superior performance in benchmark evaluations and exhibited greater accuracy in reproducing the variations in predictive precision.

Keywords: ENSO prediction; transformer; LSTM; GRU; deep learning.

2020 AMS Subject Classification: 68T05, 97R40, 97R30.

1. INTRODUCTION

The implementation of calendar months is a feasible strategy for distinguishing the behavior of El Nino-Southern Oscillation (ENSO) phenomena. ENSO occurrences usually initiate in the northern hemisphere spring, undergo a swift intensification in the summer and fall seasons, and reach their peak magnitude in the winter [1]. The phenomenon known as "seasonal phase-locking" of ENSO is ascribed to the yearly fluctuations in the atmospheric reaction to a specific the fundamental conditions of the oceans and sea surface temperature (SST) [2].

The primary determinant of the seasonal variability in the precision of ENSO prediction is the phenomenon of seasonal ENSO periodic. When El Niño-Southern Oscillation (ENSO) events reach maturity during the boreal winter and are predicted by a combination of dynamical and statistical models, the accuracy of forecasting generally remains high. Conversely, it experiences a swift decline in the boreal spring season as the ENSO phenomenon leads to escalate [3]. The term "spring predictability barrier" is a widely used nomenclature to refer to the characteristic feature of ENSO predictions [4].

The seasonal variability of the ENSO precursor holds substantial importance in determining the accuracy of ENSO prediction. The Indian Ocean Dipole (IOD) is a widely researched phenomenon that serves as a leading indicator of the El Niño-Southern Oscillation (ENSO). Its highest intensity is observed during the autumn season in the Northern Hemisphere, with minimal impact during other seasons [5]. Hence, it can be deduced that the Indian Ocean Dipole (IOD) is a highly dependable parameter for making predictions that initiate in the autumn season of the Northern Hemisphere [6], and North Tropical Atlantic SST [7], demonstrate notable seasonal fluctuations in their magnitudes, which restrict their predictive efficacy to specific seasons. To put

it in other words, a precursor of the El Niño-Southern Oscillation (ENSO) that signals an upcoming season may not necessarily be the most effective precursor for the following seasons.

In their study, Ham et al. constructed an ENSO forecast model, named H19, that utilized deep learning techniques. The model was customized for each target season and forecast lead month to accommodate the seasonality of the predictors. This information was reported in reference [8]. The production of ENSO forecasts for all target seasons, which includes 23 lead months and 12 target seasons, necessitates the use of 276 H19 models. Subsequent studies utilized deep learning models such as variant of the Convolutional Neural Network (CNN) [9], LSTM [10], and the combination of the two that becomes a Convolutional Long Short-Term Memory Network (ConvLSTM) [11].The present deep learning models were found to be insufficient in reproducing the diverse seasonal characteristics of the ENSO, as they were created separately for each input season.

Furthermore, the temporal sequence of predictions exhibits a diminishing degree of coherence over time due to the utilization of distinct H19 models for each lead month. The H19 model postulates a monthly variation in the ENSO index that lacks empirical evidence. The most effective solution to this issue is to develop a comprehensive H19 model that encompasses all the relevant seasons and projected lead months. The act of consolidating monthly training samples into a unified dataset significantly reduces the predictive ability for ENSO events.

The current investigation presents a new methodology that combines the Empirical Mode Decomposition (EMD) technique with the convolutional Long Short-Term Memory (LSTM) Encoder-Decoder model, resulting in a hybrid approach. The proposed methodology exhibits the capacity to provide accurate predictions of Oceanic Niño Index (ONI) values pertaining to El Niño for a duration of 12 months. Furthermore, it has the capability to forecast the onset of El Nino phenomena with a 12-month lead time. The objective of incorporating Empirical Mode Decomposition (EMD) is to streamline the process of breaking down past Oceanic Niño Index (ONI) data points into a collection of Intrinsic Mode Functions (IMFs) and a remaining residual component. Efficient training of sequence prediction using the convolutional LSTM Encoder-Decoder model is made possible by the local stationarity exhibited by each decomposed

component. The Convolutional LSTM Encoder-Decoder architecture comprises distinct submodels for encoding and decoding, specifically the encoder and decoder submodels. The submodel responsible for encoding integrates Long Short-Term Memory (LSTM) units with convolutional reading. The Convolutional Long Short-Term Memory (LSTM) model demonstrates proficiency in assimilating information from input that changes over time and consolidating sequential data into an underlying representation. The decoder submodel utilizes a series of stacked LSTM layers to decode the latent state and produce subsequent sequences. In the end, the sequential data generated by individual components are reconstructed to produce the anticipated results. The proposed approach is evaluated through the application of several metrics, including the Absolute Error (AE), Mean Absolute Percent Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The results of the analysis indicate that the proposed approach is superior to the existing models cited in academic literature for forecasting ONI and El Nino events. The approach put forth exhibits precise forecasting of El Nino events during the periods of 2009-2010, 2015-2016, and 2018-2019, with a lead time of 12 months, within the time span of 2008 to 2019.

The subsequent sections are structured in the following manner. The second section of the document elucidates the fundamental principles underlying the Transformer with recurrent architecture. The third section elucidates the methodologies employed for scrutinizing and predicting time series information utilizing the suggested approach. The fourth section of the document delineates the outcomes of the experiment, provides a comparison of the results, and offers an evaluation of the findings. The findings of the investigation are expounded upon in Section 5.

2. RELATED WORKS

Over the past few years, a number of machine learning approaches have surfaced as substitutes for traditional analysis and prediction methods in addressing concerns related to time series. As per the citation provided in reference [12], non-linear models are commonly afflicted by the issue of complexity and over-parameterization. According to the authors cited in reference [13], a feed-

forward neural network that incorporates a solitary hidden layer can function as a flexible substitute for linear models that possess predetermined specifications, particularly for prediction timeframes that surpass a solitary step into the future. The techniques mentioned above have been widely employed in the discipline of oceanography to provide forecasts regarding oceanic variables, such as ignificant events like El Nino, waves, sea surface temperature (SST), monsoon models, and sea level, [14]. This study emphasizes the effectiveness of Self-Organizing Map (SOM), Genetic Programming (GP), and Artificial Neural Network (ANN) in predicting sea level. The investigation described in citation [15] utilizes a Self-Organizing Map (SOM) neural network to forecast the water level in Hamburg. The outcomes derived from this methodology are juxtaposed with those of six alternative models, alongside empirical observations. The GA model employed in reference [16] underwent training and validation protocols using tide gauge measurements. The application of the GA algorithm is utilized to forecast variations in sea level, with a forecast horizon of three time intervals ahead, over a specified time span of 12 hours, 24 hours, 5 days, and 10 days. The methodology employed involves conducting a comparative analysis of the results obtained from an Artificial Neural Network (ANN) model and an Emotional Artificial Neural Network (EANN) model.

The researchers duplicated their prior methodology at an alternate location situated in the Cocos (Keeling) Islands, which are positioned in the Indian Ocean [17]. The Regional Neural Network for Water Level (RNN WL), a feed-forward neural network (FFNN), was utilized in a study conducted by [18]. The training data for the RNN WL was obtained from a remote station located on Long Island, specifically on the South Shore of New York. The objective of the aforementioned training was to forecast the water levels in a coastal inlet for an extended duration. In a previous investigation [19] the authors employed diverse methodologies, including multilinear regression (MLR), feed-forward backpropagation (FFBP), radial basis function (RBF), and generalized regression neural network (GRNN), to approximate the daily average sea level elevations. The aforementioned methods were utilized to conduct a least squares estimation on the model of sea level. The results of the statistical analysis indicated that the employment of neural

network-based techniques yielded higher levels of accuracy in comparison to the conventional least squares method in the determination of daily mean sea levels. The methodology under consideration entails the modification of connection weights within a feedforward neural network (FFNN) to tackle intricacy and predict river levels with a lead time of up to five hours. In their study, Pashova and Popova conducted a comparison of various artificial neural network (ANN) models with the aim of predicting sea levels [20]. The research assessed various neural network techniques, including multilayer feed-forward (FF), cascade-feed-forward (CFF), feed-forward time delay (FFTD), RBF, and generalized regression (GR), as well as multiple linear regression (MLR). The utilization of Artificial Neural Networks (ANNs) has been discovered to surpass conventional harmonic analysis methodologies in effectively capturing temporal interdependencies in time series data [21]. In a previous study, the authors cited as reference [22] employed hourly time series data of atmospheric pressure, wind, and harmonically calculated tides as predictors, and the observed tides as the response variable, to train an artificial neural network (ANN). The implementation of the network yielded a 50% decrease in the margin of error. The present study utilized two distinct data-driven approaches, namely the adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN), to estimate hourly sea level.

The study employed the MLR methodology to ascertain the most favorable amalgamations of hourly sea level inputs, predicated on lag times. The performance of the Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Auto-Regressive Moving Average (ARMA) models were compared, revealing that the ANN and ANFIS models demonstrated comparable or superior performance. The scholars utilized the Extreme Learning Machine (ELM) methodology, which is grounded on Artificial Neural Networks (ANN), to predict alterations in sea level in Chiayi, Taiwan [23]. Additional methodologies employed encompassed the Relevance Vector Machine (RVM), Support Vector Machine (SVM), and Radial Basis Function (RBF) models. The findings indicate that the ELM and RVM models exhibited superior performance compared to the other methodologies. The article cited as [24] outlines several techniques for handling missing values in time series data. The predominant approach involves utilizing the

Empirical Orthogonal Function (EOF) reconstruction method in conjunction with either a prediction network (FCnet) or a comparable backcast network (BCnet). The study employed a neural network architecture to forecast regional mean sea level anomalies (MSLA) by leveraging a particular set of tide gauges following the restoration of the time series. Reference [25] presented a methodology for predicting tide levels in the event of typhoon-induced storm surges. The present methodology employs typhoon parameters as input and utilizes a cubic B-spline curve with a knot insertion mechanism in conjunction with the forecasts. Although conventional artificial neural network architectures are purported to be capable of conducting time series analysis and prediction, comprehending long-term correlations within an extended time series remains a difficult task.

The convolutional neural networks (CNNs) possess the ability to process multidimensional data and capture spatial correlations, while the recurrent neural networks (RNNs) are intrinsically suitable for managing time series data owing to their efficiency in capturing long-term spatial and/or temporal relationships. The amalgamation of Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) architectures results in a composite model that exhibits proficiency in processing spatio-temporal data. The integration of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) has been suggested as a method for exploring the spatial and temporal progression of sea level and predicting interannual sea level anomalies [26]. The Long Short-Term Memory (LSTM) model effectively addresses numerical obstacles that arise from the vanishing or exploding gradient problem during the training phase. Moreover, it eliminates the need for explicitly choosing a primary time for the sliding window approach as discussed in reference [27], which concerns the number of time steps for moving the sliding window in the time series. Long Short-Term Memory (LSTM) models have exhibited effectiveness in diverse investigations focused on forecasting occurrences in time series data distinguished by significant intervals and lags. In a particular investigation [28], a speaker verification model that is not reliant on specific texts was created by combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Likewise, a separate investigation [29] utilized a Long Short-Term Memory (LSTM) framework incorporating an

architectural refinement strategy referred to as Enhanced Forget Gate (EFG) for the purpose of forecasting wind power. The aforementioned methodology yielded a significant enhancement of 18.3% in contrast to alternative prognostication techniques.

3. Study Area and Dataset

The ENSO phenomenon is a notable climatic variation that occurs on an annual basis, with its source being traced to the tropical Pacific Ocean. Its effects are far-reaching and have a significant negative impact on the global climate system. A cyclical pattern of events is observed in the expansive Pacific Ocean, spanning from the shores of Peru and Ecuador to its central region in proximity to the International Date Line. These events are characterized by alternating El Nino and La Nina occurrences. The area known as Nino3.4, which is defined by the geographical coordinates of 5°S-5°N and 170°W-120°W as illustrated in Figure 1, displays a sequence of five consecutive 3-month moving averages of sea surface temperature (SST) anomalies that exceed or fall below the threshold of + 0.5 °C or 0.5 °C, respectively. The Oceanic Nino Index (ONI) is utilized as a standardized metric for the purpose of monitoring the El Nino-Southern Oscillation (ENSO) phenomenon [30].

The Oceanic Niño Index (ONI) is derived from the anomalous sea surface temperature (SST) values relative to the long-term mean in predetermined areas of the tropical Pacific. The Oceanic Niño Index (ONI) concentrates its attention on the Niño 3.4 area, which encompasses a geographical range of roughly 5°N to 5°S latitude and 120°W to 170°W longitude. The anomalies in sea surface temperature (SST) within this particular area are utilized for the purpose of ascertaining the magnitude and duration of El Niño or La Niña occurrences.

A positive ONI value is indicative of the existence of El Niño conditions, which implies that the sea surface temperatures in the Niño 3.4 region are higher than the average. The conditions that are warmer than the norm usually have an impact on weather patterns worldwide, resulting in alterations in the distribution of temperature, atmospheric circulation, and precipitation.

On the other hand, a negative Oceanic Niño Index (ONI) value indicates the presence of La Niña conditions, which are distinguished by lower-than-average sea surface temperatures in the

Niño 3.4 area. La Niña phenomena have distinct implications on worldwide weather patterns, frequently resulting in cooler and wetter climatic conditions in certain areas.

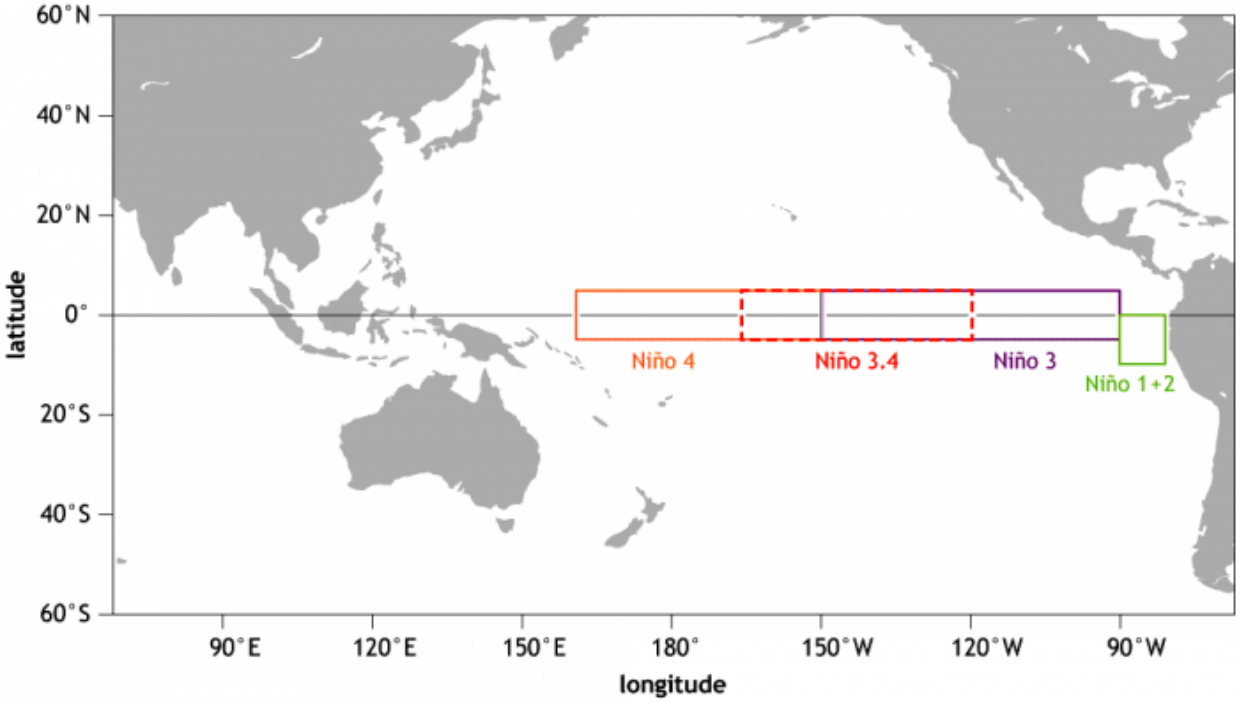

FIGURE 1. The El Nino area in the Pacific Ocean is divided into several regions along the equator starting from the Asian continent to the American continent.

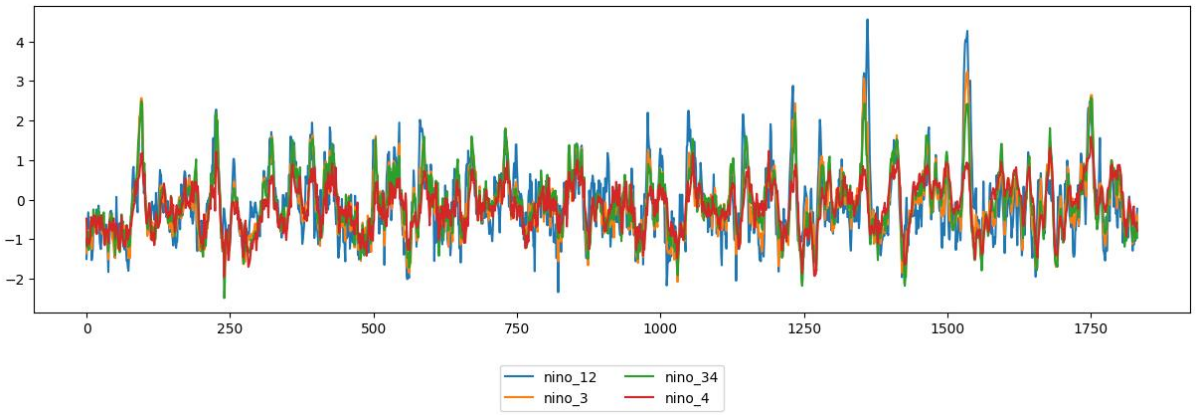

FIGURE 2. Graph can represent the length of the data in a time series where changes in sea surface temperature can be observed and form a temporal pattern.

The present study selected the Nino 3.4 region, which extends from 5S to 5N latitude and 170W to 120W longitude, as the primary site for investigating El Nino events through SST analysis, owing to its frequent usage in previous research. A variety of quantitative climatic measures can

be employed to characterize the incidence of El Nino and La Nina phenomena, utilizing the sea surface temperature (SST) data obtained from the Nino 3.4 region. The Nino 3.4 index and the Oceanic Nino Index (ONI) are commonly employed indices for the purpose of monitoring El Nino and La Nina phenomena. The computation of these indices is predicated on the anomalies in sea surface temperature (SST) within the Nino 3.4 region. The Oceanic Niño Index (ONI) is utilized by the National Oceanic and Atmospheric Administration (NOAA) as the principal metric to detect the occurrence of El Niño and La Niña events. The Oceanic Niño Index (ONI) is computed by utilizing the Extended Reconstructed Sea Surface Temperature (ERSST.v5) anomalies within the Nino 3.4 region, with a reference period of 30 years. As per the National Oceanic and Atmospheric Administration (NOAA), the Oceanic Niño Index (ONI) exceeding the +0.5 °C threshold for a consistent duration of five months is suggestive of an El Niño year in the subsequent year. Similarly, La Nina phenomena are characterized by a consecutive series of five three-month intervals during which the temperature remains at or below 0.5 °C. In addition, according to the Oceanic Niño Index (ONI), El Niño events can be classified into four tiers based on their intensity levels: weak (0.5-0.9 C), moderate (1.0-1.4 C), strong (1.5-1.9 C), and very strong (2.0 C). On the other hand, the magnitude of La Niña occurrences is categorized based on the corresponding negative numerical values.

The National Oceanic and Atmospheric Administration (NOAA) website offers access to archived data on the Oceanic Niño Index (ONI). The Oceanic Niño Index (ONI) is a metric utilized to observe and forecast El Niño and La Niña occurrences by gauging the deviations in sea surface temperature in the equatorial Pacific Ocean. The data may be retrieved from the website https://origin.cpc.ncep.noaa.gov. The present study employs a dataset consisting of ONI data that covers a time frame of 70 years, starting from January 1950 and ending in December 2019. The dataset consists of 840 data points that were gathered on a monthly basis. Each data point's date represents the third month of the corresponding three-month interval. The Oceanic Niño Index (ONI) is computed by averaging the sea surface temperature (SST) anomaly measurements collected during a consecutive three-month interval. For instance, the JFM timeframe is utilized to

indicate the month of February, while the ONI measure for February is derived by averaging the ONI readings that cover the months of January through March.

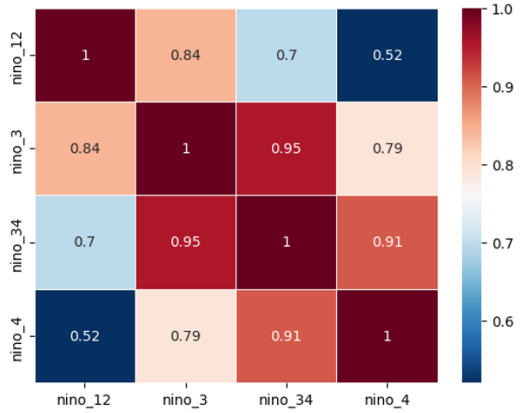

FIGURE 3. Pearson correlation calculations on the data to show the relationship between data.

As shown in Figure 3, in the aforementioned research by Jebli et al. [31] utilized the Pearson Correlation Coefficient (PCC) to quantify the level of dependence or association among variables, specifically in relation to El Nino regions, as illustrated in Figure 3. The PCC was employed to examine the correlation between different El Nino regions through their comparison. The objective of our study was to identify the interconnections or links among the aforementioned zones, with the aim of obtaining valuable information that could be utilized for analysis and prediction purposes.

The statistical measurement of the linear correlation between two variables is accomplished through the use of the Pearson correlation coefficient (PCC). The numerical value of correlation coefficient ranges from -1 to 1, where a value of -1 indicates a strong negative correlation, a value of 0 indicates no correlation, and a value of 1 indicates a strong positive correlation. In this study, the Pearson correlation coefficient (PCC) was employed to assess the magnitude and direction of the associations among different El Nino zones. This valuable information can aid in the understanding of the dynamics and interactions between these geographical areas. Our objective was to enhance their understanding of the complex dynamics of El Nino events through the utilization of the Pearson Correlation Coefficient (PCC) to examine the interrelationships among El Nino regions.

4. PROPOSED METHOD

4.1 Preprocessing Data

The dataset was used in this research was relatively good because there was no missing value, so that there was no need to perform data cleaning. This study split the dataset into two parts: train and test data, where 80 percent of the dataset was used for the train process and 20 percent was used for the test process, as shown in Figures 4, 5, 6, and 7. This research also split the training data into two parts, with 80 percent of the training data allocated for training the models and the rest for validating the models.

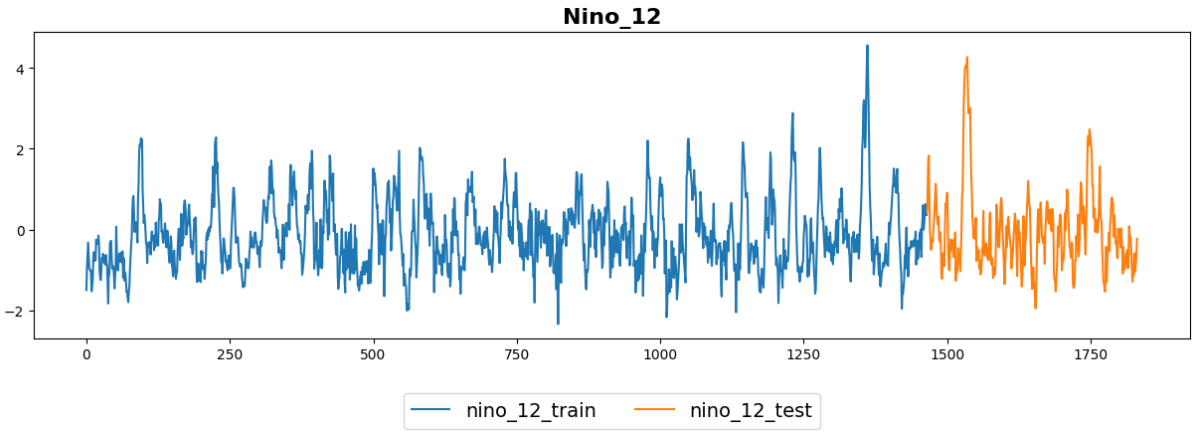

FIGURE 4. Split Nino 12 anomaly data.

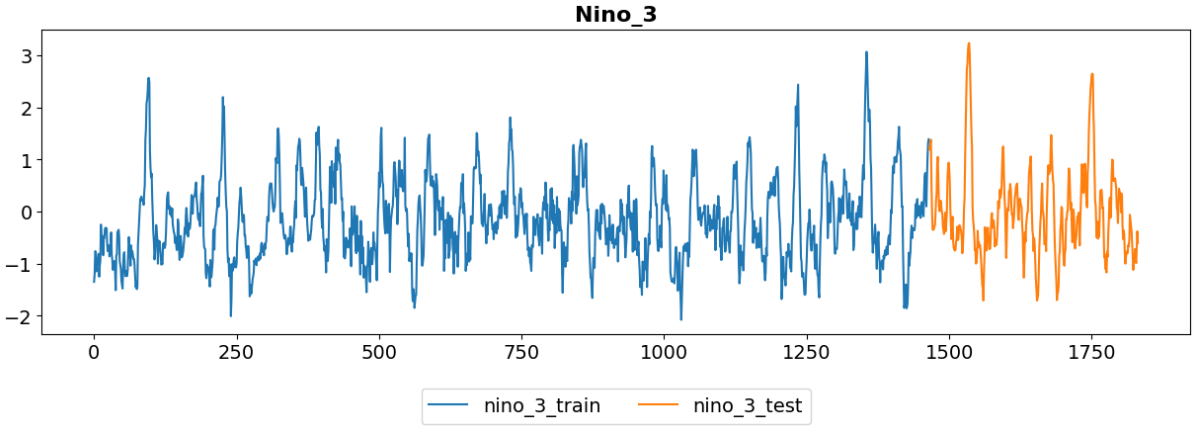

FIGURE 5. Split Nino 3 anomaly data.

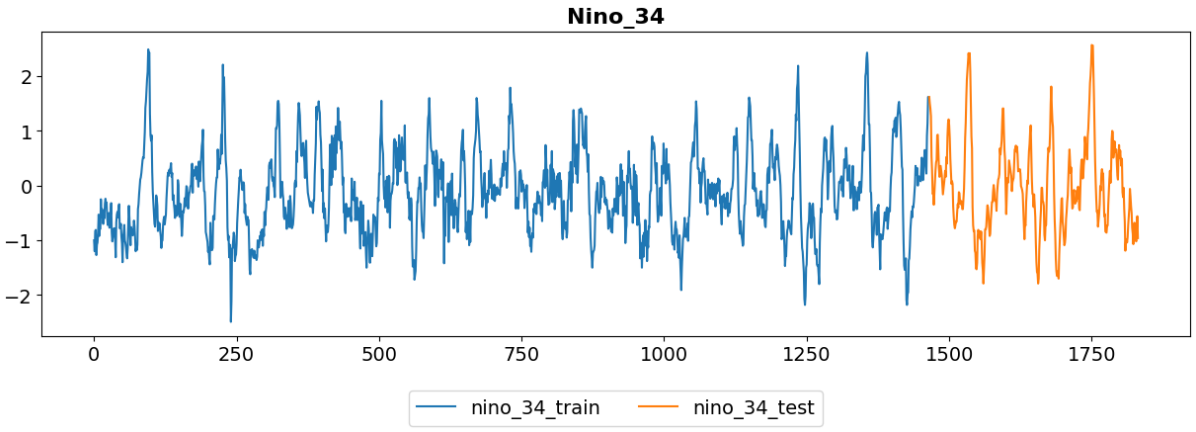

FIGURE 6. Split Nino 34 anomaly data.

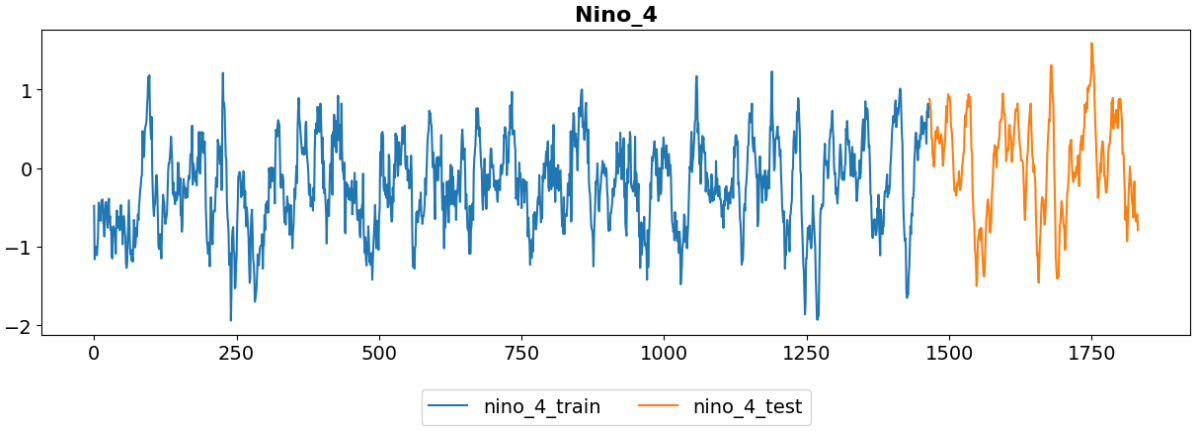

FIGURE 7. Split Nino 4 anomaly data.

This research did not implement any normalization or standardization because the data was relatively in small range$(-3,5)$. This research designed all predictive models that can predict data one month in the future based on previous twenty-four month as in figure 8.

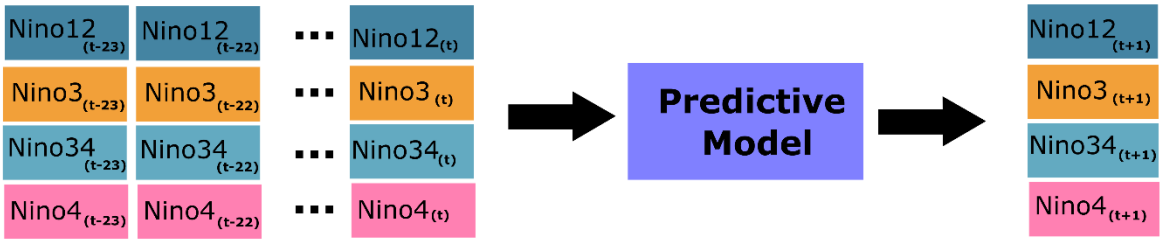

FIGURE 8. Prediction scenario with models.

4.2 Nino Transformer with Recurrent Neuron

To mitigate the issues associated with the transformer, we incorporated a specialized recurrent neuron into its architecture, specifically tailored for the purpose of predicting ENSO. The proposed model integrates Recurrent Neural Network (RNN) and Transformer architectures, as illustrated in Figure 10. Utilizing a method of progressive abstraction from superficial to profound, Recurrent Neural Networks (RNNs), a hierarchical approach to data representation, extract characteristics imbued with high-level semantic information. The lower-level feature representation serves as a fundamental basis for the higher-level feature representation. The enhanced transformer model depicted in Figure 9 was derived by means of a comparative analysis with the transformer baseline.

Extracting the underlying properties, such as critical locations, lines, and Nino zones, is deemed more advantageous. One approach to achieve this is through the utilization of a common neuron for feature extraction, resulting in a potential reduction in the necessary parameters to prevent redundant processing. However, the incorporation of encoder operation can furnish the network with a specific degree of translation invariance. The transformer employs a self-attentive mechanism to obtain comprehensive contextual information and establish distant relationships with the target, thereby enhancing its efficacy in governing high-level features such as connecting Nino regions with other regions. In order to achieve a harmonious integration of benefits, a combination of Recurrent Neural Networks (RNN) and Transformers is employed, with the former being utilized for the foundational features and the latter for the more advanced attributes. The subsequent passage delineates the complete computational process of the model.

$$
\begin{aligned}
a_0 &= \left[X_{pre}, PD(DS(x))\right] + X_{pos}, X_{pos} \in R^{(N+1)\times C_2} \\
a_\ell &= MHA\left(LN(a_{1-1})\right) + a_{1-1}, \ell = 1, \dots, L \\
a_l &= MLP\left(LN(a_l)\right) + a_l, \ell = 1, \dots, L \\
y &= FC(a_L^0)
\end{aligned}
\tag{1}
$$

To produce the feature map $f_1 \in R^{C\times H\times W}$, downsampling (DS) is first performed on the input data $x \in R^{C\times H\times W}$, the dataset comprises of anomalies in sea surface temperature (SST) and heat content for a period of three consecutive months. The feature map f_1 is divided into N feature maps with the same dimension, $f_2 \in R^{C_2\times p1\times p2}$, and then these vectors are turned into 1-D vectors with C_2 as their dimension. After patch division (PD), the semantic feature data's dimension is (N, C_2).

It is noteworthy that the learnable embedding token $a_0^0 = X_{pre}$is pre-set on the embedded sequence. Incorporating position embedding coding is necessary to obtain position features X_{pos}. The absence of positional information in the input data is deemed critical for meteorological data. The semantic and location features are provided as inputs to the transformer structure. The transformer architecture comprises of Multilayer Perceptron (MLP) and Multi Head Attention (MHA) blocks. Layer normalization (LN) is applied prior to each block, and a residual connection is applied after each block. The outcome of the prediction, denoted as y, is obtained through a series of coding layers, whereby the embedding feature of the final layer is propagated into the fully connected (FC) layer.

$$[q, k, v] = z * U_{qkv}, U_{qk}, \in R^{C_2\times 3C_h}$$

$$S(z) = sofmax\left(\frac{qk^T}{\sqrt{C_h}}\right) * v \tag{2}$$

$$MHA(z) = [S_1(z), \dots, S_k(z)] * U_{mha}, U_{mha} \in R^{hC_h\times C_2}$$

A common transformer architecture building component is self-attention (S) [32]. For each element in the input sequence $z \in R^{(N+1)\times C_2}$, we calculate a weighted total of all values v.The attention weights are determined by how closely two items in the sequence and their corresponding query, q and key, k, match up pairwise. In MHA, an extension of S, we perform h self-attentive processes (also known as "heads") concurrently and forecast their results. It is typical to set C_h to $\frac{C_2}{h}$ in order to maintain the computation and the amount of parameters while adjusting h.

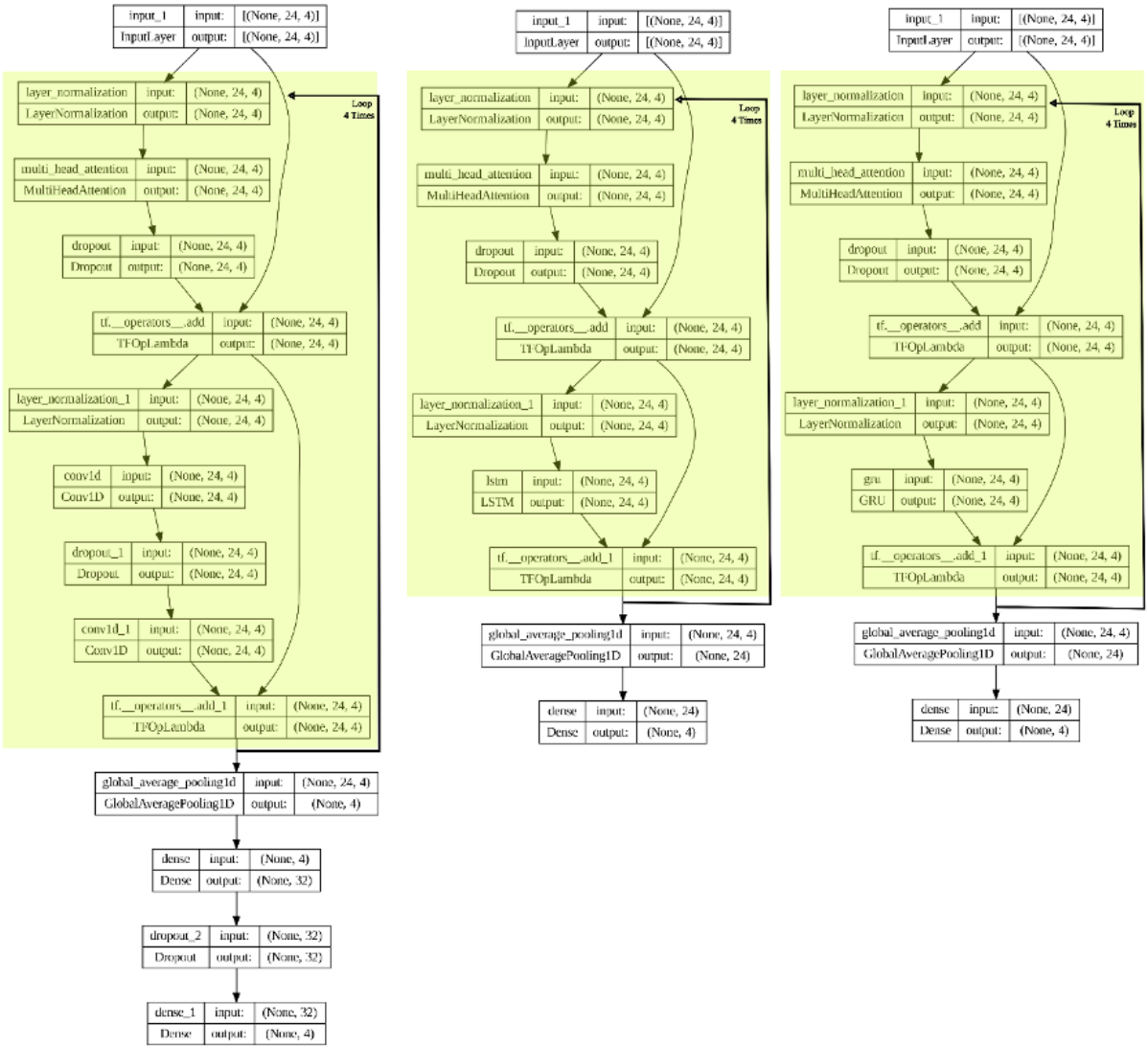

FIGURE 9. Transformer Baseline Model (left), Transformer LSTM Model (middle), and Transformer GRU Model(right)

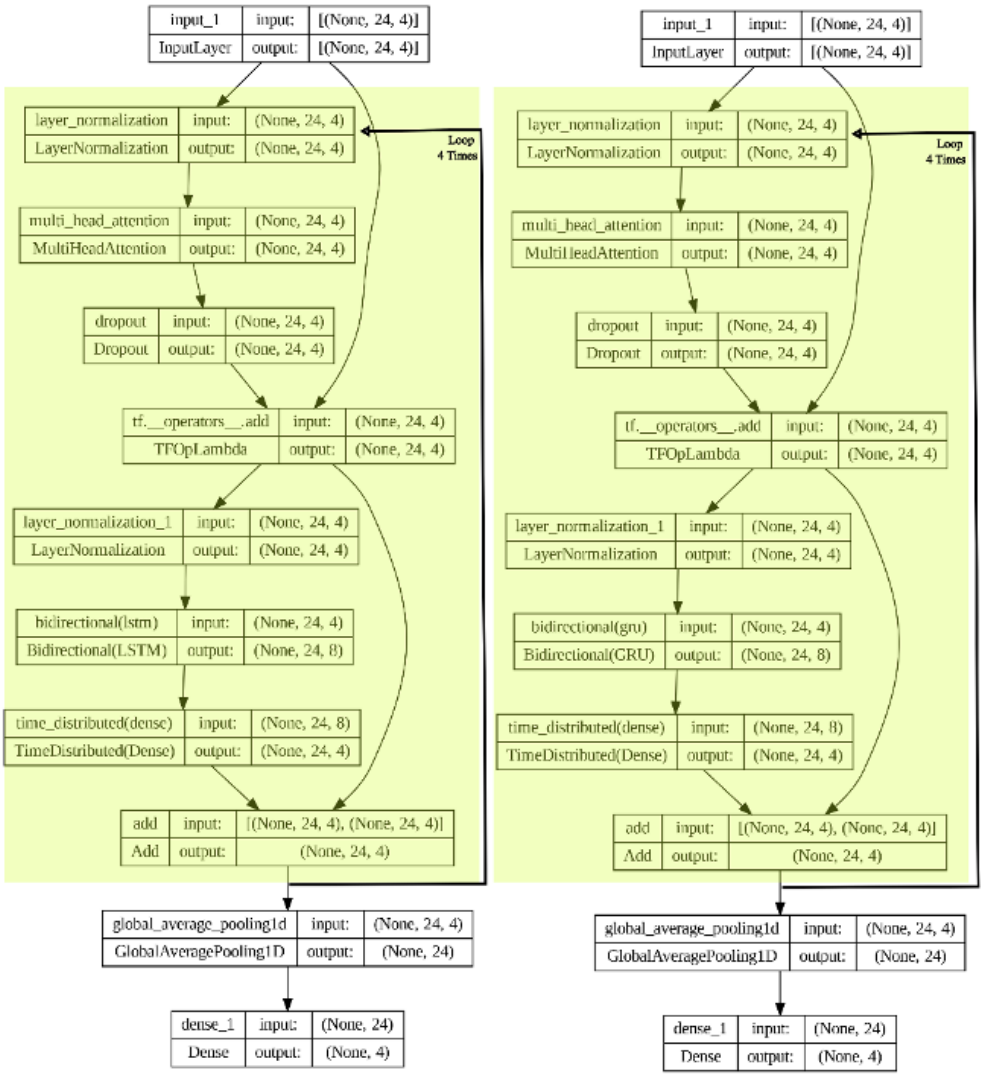

FIGURE 10. Transformer BiLSTM Model (left), and Transformer BiGRU Model(right)

In this research, the transformer baseline model for time series was built based on Keras documentation as shown in Figure 4, while the other proposed transformer model was modified from the baseline model by implementing RNN layers such as LSTM, BiLSTM, GRU, and BiGRU and making some model adjustments as in Figure 4 and 5.

To ensure a fair comparison, this research implemented the same settings for each of the compared models, such as the number of units in each RNN layer, such as LSTM, Bi-LSTM, GRU, and Bi-GRU, being set to four, the learning rate was set to with the Adam optimizer, and the batch size and the number of epochs were 256 and 100, respectively. As can be seen in Figures 4 and 5,

the green area was the transformer block. That number of transformer blocks was set to four, so that the models would perform a four-time loop.

4.3 Evaluation Metrics

To Evaluate the performance of predictions that is so crucial during Nino events, the primary accuracy measure for a deterministic forecast is the root-mean-square error (RMSE) in equation (8):

$$RMSE = \sqrt{\frac{\sum(y_i - q_i)^2}{n}} \tag{3}$$

Where i = 1, 2, ..., 24, n is the length of samples, G and Y denote the observation and prediction values, $\overline{G}$ and $\overline{Y}$ are observation mean value and prediction mean value, respectively. MAE can be formulated as [33] in equation (9).

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_i - q_i| \tag{4}$$

5. RESULT AND DISCUSSION

Five models have been built to forecast El Nino in Pacific Ocean leading by 12 months' time step to predict one month. Those models are:

- Transformer baseline: Transformer baseline refers to a basic or standard implementation of the transformer architecture in natural language processing tasks.
- Transformer LSTM: a hybrid model that combines elements of both the transformer architecture and the LSTM (Long Short-Term Memory) model.
- Transformer GRU: the transformer architecture and the GRU (Gated Recurrent Unit) model are combined to create a hybrid model.
- Transformer BiLSTM: Similarly with transformer LSTM but we utilize LSTM as a back propagation algorithm.
- Transformer BiGRU: Combination of Transformer and GRU with back propagation use GRU.

TABLE 1. Testing Results for each model.

	Metrics	Transformer Baseline	Transformer LSTM	Transformer GRU	Transformer BiLSTM	Transformer BiGRU
NINO 12	MAE	0.695619	0.709542	0.842349	0.688110	0.698858
	RMSE	1.020399	1.066383	1.128625	0.988143	0.978457
NINO 3	MAE	0.624594	0.705692	0.592944	0.641332	0.718975
	RMSE	0.882535	0.986601	0.835614	0.885487	0.957922
NINO 34	MAE	0.657151	0.725354	0.657338	0.657001	0.646685
	RMSE	0.856256	0.947293	0.837290	0.854008	0.843343
NINO 4	MAE	0.573183	0.725275	0.563845	0.526472	0.702675
	RMSE	0.681316	0.865769	0.678396	0.637363	0.859172

All models were evaluated with the test dataset (orange-colored data in Figures 4, 5, 6, and 7). Table 1 is the evaluation summarization of every compared model, where the best result is colored in green. From the table, it can be seen that transformer baseline and transformer LSTM did not perform superiorly in this research comparison in predicting Nino data. Transformer GRU, Transformer BiLSTM, and Transformer BiGRU were very competitive.

All the deep learning models built in this research can predict all four NINO data sets, such as Nino 12, Nino 3, Nino 34, and Nino 4, in a single time process, as depicted in Figure 8. Based on testing results, the best model for predicting Nino 3 data was Transformer GRU, and the best model for predicting Nino 4 data was Transformer BiLSTM. Meanwhile, Transformer BiLSTM and Transformer BiGRU were competing to be the best at predicting Nino 12 data, and Transformer GRU and Transformer BiGRU were competing against each other to be the best at predicting Nino 34 data. Table 1 was to show the competition between each model in predicting Nino data. It seems that the transformer LSTM seems to be worse than the transformer baseline, but the transformer LSTM was actually better than the baseline model in the next analysis in the next discussion based on Figures 11 and 12.

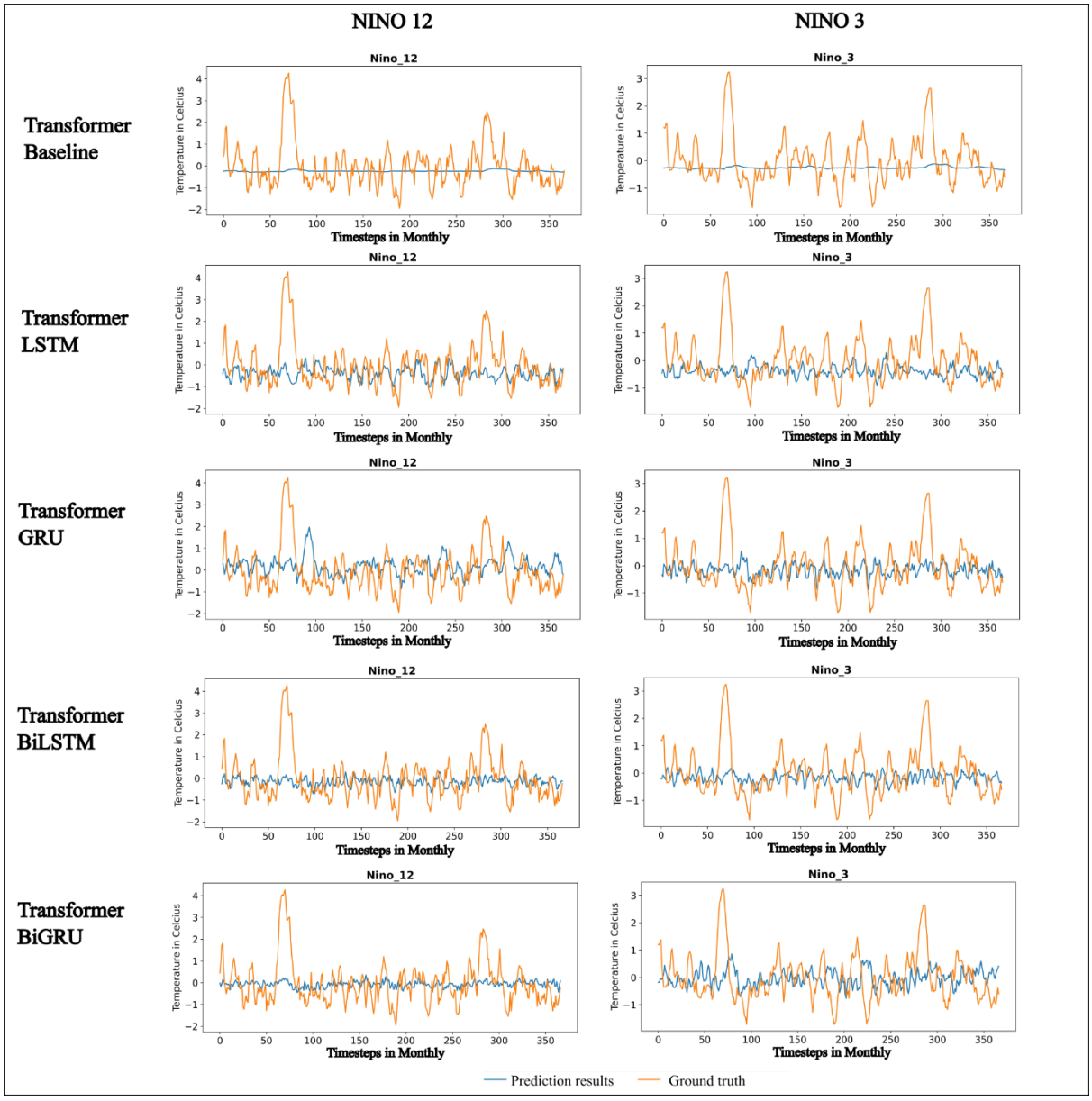

FIGURE 11. Nino 12 and Nino 3 prediction results visualization for each model with testing data.

A quick glance at Figures 11 and 12, it seems that Nino 12 and Nino 3 ground truth data have similar patterns, as do Nino 34 and Nino 4. It was because Nino 12 and Nino 3 have an extremely high correlation of 0.84 in Pearson correlation, whereas Nino 34 and Nino 4 have an extremely high correlation of 0.91 in Pearson correlation.

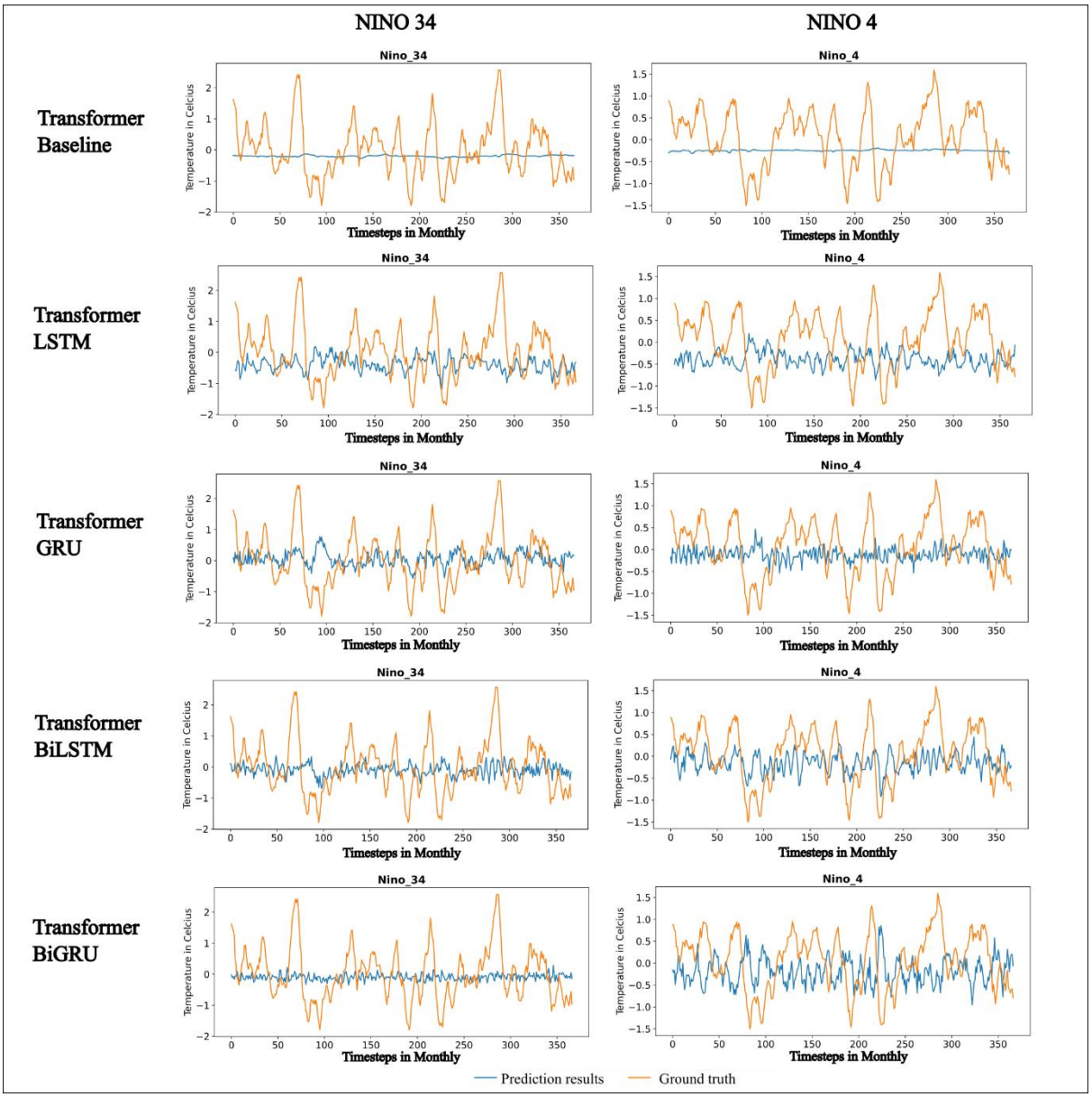

FIGURE 12. Nino 34 and Nino 4 prediction results visualization for each model with testing data.

To see how good our models were at predicting Nino data, this research visualized the test prediction results in Figures 11 and 12, where the orange line is the ground truth data and the blue line is the prediction result. In our opinion, the transformer baseline model seems to have failed to predict the Nino data, where its prediction result looks like a flat line, while the other line fluctuated while trying to predict the future data.

It seems that the implementation of RNN layers such as LSTM, BiLSTM, GRU, and BiGRU in transformer models gives significant improvement, so for future research there is still a chance to modify the transformer model to be a robust model for forecasting time series data. This research faced the obstacle that the number of neuron units in the RNN layer cannot be changed other than to only four units. In our opinion, the alteration in unit number in the RNN layer may bring significant improvement in the prediction of time series data because the RNN layer is very suitable for sequence data.

Beside the fact that modification of deep learning models can improve the prediction, there is a belief that time series forecasting is a complex task, so involving new variables such as atmospheric data such as wind pattern and atmospheric pressure, oceanic variables such as sea-level height level, ocean current, and subsurface ocean temperature, and precipitation pattern data may improve the prediction results. The collaboration with oceanographers is needed for our future research with the intention of learning about our earth and saving it from the damage that will harm future humanity.

6. CONCLUSION

This research result show that the implementation of RNN layers in transformer model can improve the prediction results of Nino data, where the transformer GRU, Bi-LSTM, and Bi-GRU were compete each other overcome the transformer baseline. From testing prediction result visualization, it seems that the baseline model failed to predict timeseries data with Nino data, meanwhile the modified transformer with RNN layer can predict better.

Based on this research challenge in modifying the RNN layer in the transformer architecture model, the changing of neuron units in the RNN layer that is suitable for the Nino dataset may still have more potential to be investigated further for future research. Involving new variables that have a correlation value with NINO data may help improve models for predicting precisely in future research.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] A. Timmermann, S.I. An, J.S. Kug, et al. El Niño–Southern Oscillation complexity, Nature. 559 (2018), 535–545. https://doi.org/10.1038/s41586-018-0252-6.

[2] Y.G. Ham, J.H. Kim, E.S. Kim, et al. Unified deep learning model for El Niño/Southern Oscillation forecasts by incorporating seasonality in climate data, Sci. Bull. 66 (2021), 1358–1366. https://doi.org/10.1016/j.scib.2021.03.009.

[3] A.F.Z. Levine, M.J. McPhaden, The annual cycle in ENSO growth rate as a cause of the spring predictability barrier, Geophys. Res. Lett. 42 (2015), 5034–5041. https://doi.org/10.1002/2015gl064309.

[4] A.G. Barnston, M.K. Tippett, M. Ranganathan, et al. Deterministic skill of ENSO predictions from the North American multimodel ensemble, Clim. Dyn. 53 (2017) 7215-7234. https://doi.org/10.1007/s00382-017-3603-3.

[5] N.H. Saji, B.N. Goswami, P.N. Vinayachandran, et al. A dipole mode in the tropical Indian Ocean, Nature. 401 (1999), 360–363. https://doi.org/10.1038/43854.

[6] H. Dayan, J. Vialard, T. Izumo, et al. Does sea surface temperature outside the tropical Pacific contribute to enhanced ENSO predictability?, Clim. Dyn. 43 (2013), 1311-1325. https://doi.org/10.1007/s00382-013-1946-y.

[7] Y. Ham, J. Kug, J. Park, Two distinct roles of Atlantic SSTs in ENSO variability: North Tropical Atlantic SST and Atlantic Niño, Geophys. Res. Lett. 40 (2013), 4012–4017. https://doi.org/10.1002/grl.50729.

[8] Y.G. Ham, J.-H. Kim, J.-J. Luo, Deep learning for multi-year ENSO forecasts, Nature. 573 (2019), 568–572. https://doi.org/10.1038/s41586-019-1559-7.

[9] J. Yan, L. Mu, L. Wang, et al. Temporal convolutional networks for the advance prediction of ENSO, Sci. Rep. 10 (2020), 8055. https://doi.org/10.1038/s41598-020-65070-5.

[10] C. Broni-Bedaiko, F.A. Katsriku, T. Unemi, et al. El Niño-Southern Oscillation forecasting using complex networks analysis of LSTM neural networks, Artif. Life Robotics. 24 (2019), 445–451. https://doi.org/10.1007/s10015-019-00540-2.

[11] A. Mahesh, M. Evans, G. Jain, et al. Forecasting El Niño with convolutional and recurrent neural networks, in: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, pp. 1–5, 2019.

[12] J.G. De Gooijer, R.J. Hyndman, 25 years of time series forecasting, Int. J. Forecast. 22 (2006), 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001.

[13] N.R. Swanson, H. White, Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models, Int. J. Forecast. 13 (1997), 439–461. https://doi.org/10.1016/s0169-2070(97)00030-7.

[14] A.E. Thessen, Adoption of machine learning techniques in ecology and earth science, One Ecosystem, 1 (2016), e8621.

[15] D. Zhao, Y. Zeng, Q. Wu, et al. Source discrimination of mine gushing water using self-organizing feature maps: a case study in ningtiaota coal mine, Shaanxi, China, Sustainability. 14 (2022), 6551. https://doi.org/10.3390/su14116551.

[16] A. Molajou, V. Nourani, A. Afshar, et al. Optimal design and feature selection by genetic algorithm for emotional artificial neural network (EANN) in rainfall-runoff modeling, Water Resources Manage. 35 (2021), 2369–2384. https://doi.org/10.1007/s11269-021-02818-2.

[17] M.A. Ghorbani, O. Makarynskyy, J. Shiri, et al. Genetic programming for sea level predictions in an island environment, Int. J. Ocean Clim. Syst. 1 (2010), 27–35. https://doi.org/10.1260/1759-3131.1.1.27.

[18] W. Huang, C. Murray, N. Kraus, et al. Development of a regional neural network for coastal water level predictions, Ocean Eng. 30 (2003), 2275–2295. https://doi.org/10.1016/s0029-8018(03)00083-0.

[19] G. Accarino, M. Chiarelli, S. Fiore, et al. A multi-model architecture based on Long Short-Term Memory neural networks for multi-step sea level forecasting, Future Gen. Computer Syst. 124 (2021), 1–9. https://doi.org/10.1016/j.future.2021.05.008.

[20] L. Pashova, S. Popova, Daily sea level forecast at tide gauge Burgas, Bulgaria using artificial neural networks, J. Sea Res. 66 (2011), 154–161. https://doi.org/10.1016/j.seares.2011.05.012.

[21] R. Fredyan, G.P. Kusuma, Spatiotemporal convolutional LSTM with attention mechanism for monthly rainfall prediction, Commun. Math. Biol. Neurosci. 2022 (2022), 118. https://doi.org/10.28919/cmbn/7761.

[22] B.V. Primo de Siqueira, A. de M. Paiva, Using neural network to improve sea level prediction along the southeastern Brazilian coast, Ocean Model. 168 (2021), 101898. https://doi.org/10.1016/j.ocemod.2021.101898.

[23] M. Imani, H.C. Kao, W.H. Lan, et al. Daily sea level prediction at Chiayi coast, Taiwan using extreme learning machine and relevance vector machine, Glob. Planet. Change. 161 (2018), 211–221. https://doi.org/10.1016/j.gloplacha.2017.12.018.

[24] M. Wenzel, J. Schröter, Reconstruction of regional mean sea level anomalies from tide gauges using neural networks, J. Geophys. Res. 115 (2010), C08013. https://doi.org/10.1029/2009jc005630.

[25] W. Wang, H. Yuan, A tidal level prediction approach based on BP neural network and cubic B-spline curve with knot insertion algorithm, Math. Probl. Eng. 2018 (2018), 9835079. https://doi.org/10.1155/2018/9835079.

[26] A. Braakmann-Folgmann, R. Roscher, S. Wenzel, et al. Sea level anomaly prediction using recurrent neural networks, preprint, (2017). https://doi.org/10.48550/arXiv.1710.07099.

[27] O. Makarynskyy, D. Makarynska, M. Kuhn, et al. Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia, Estuar. Coast. Shelf Sci. 61 (2004), 351–360. https://doi.org/10.1016/j.ecss.2004.06.004.

[28] Z. Zhao, H. Duan, G. Min, et al. A lighten CNN-LSTM model for speaker verification on embedded devices, Future Gen. Computer Syst. 100 (2019), 751–758. https://doi.org/10.1016/j.future.2019.05.057.

[29] R. Yu, J. Gao, M. Yu, et al. LSTM-EFG for wind power forecasting based on sequential correlation features, Future Gen. Computer Syst. 93 (2019), 33–42. https://doi.org/10.1016/j.future.2018.09.054.

[30] H.A. Dijkstra, The ENSO phenomenon: theory and mechanisms, Adv. Geosci. 6 (2006), 3–15. https://doi.org/10.5194/adgeo-6-3-2006.

[31] I. Jebli, F.-Z. Belouadha, M.I. Kabbaj, et al. Prediction of solar energy guided by pearson correlation using machine learning, Energy. 224 (2021), 120109. https://doi.org/10.1016/j.energy.2021.120109.

[32] V.B. Rao, K. Maneesha, P. Sravya, et al. Future increase in extreme El Nino events under greenhouse warming increases Zika virus incidence in South America, npj Clim. Atmosph. Sci. 2 (2019), 4. https://doi.org/10.1038/s41612-019-0061-0.

[33] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, Geosci. Model Develop. 7 (2014), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.