

Received February 7, 2020, accepted March 15, 2020, date of publication March 19, 2020, date of current version March 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982035

Classifying El Niño-Southern Oscillation Combining Network Science and Machine Learning

MATHEUS A. DE CASTRO SANTOS¹, DIDIER A. VEGA-OLIVEROS^{2,3},
LIANG ZHAO², (Senior Member, IEEE), AND LILIAN BERTON¹

¹Institute of Science and Technology, Federal University of São Paulo, São José dos Campos 12247-014, Brazil

²Faculty of Philosophy, Sciences, and Letters at Ribeirão Preto (FFCLRP), University of São Paulo, Ribeirão Preto 14040-901, Brazil

³School of Informatics, Computing and Engineering, Indiana University Bloomington, Bloomington, IN 47408, USA

Corresponding author: Lilian Berton (lberton@unifesp.br)

This work was supported in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grant 2018/04029-7, Grant 2018/01722-3, Grant 2018/24260-5, Grant 2016/23698-1, and Grant 2015/50122-0, and in part by the DFG-GRTK under Grant 1740/2.

ABSTRACT Machine learning and complex network theory have emerged as crucial tools to extract meaningful information from big data, especially those related to complex systems. In this work, we aim to combine them to analyze El Niño Southern Oscillation (ENSO) phases. This non-linear phenomenon consists of anomalous (de)increase of temperature at the tropical Pacific Ocean, which has irregular occurrence and causes climatic variability worldwide. We construct temporal Climate Networks from the Surface Air Temperature time-series and calculate network metrics to characterize the warm and cold ENSO episodes. The metrics are used as topological features for classification. We employ ten classifiers and achieved 80% AUC ROC when predicting the intensity of Strong/ Weak El Niño and Strong/ Weak La Niña for the next season. The complex network represents the relationship among different regions of the planet and machine learning creates models to classify the different classes of ENSO. This work opens new paths of research by integrating network science and machine learning to analyze complex data like global climate systems.

INDEX TERMS Classification algorithms, climate networks, complex network, EL Niño intensity, machine learning, network sciences.

I. INTRODUCTION

The study of climate systems is of paramount relevance for understanding the impact on local and global economics [1], health and social effects [2], [3], and disaster risk reduction and vulnerability [4] on society. One traditional approach for studying climatic systems is by the development of climate models, which require data from diverse sources, such as atmosphere measures, land use, carbon release, among many others [5]. However, gathering and compiling these data represents a problem for some regions, they are time-expensive and consuming, and specialized knowledge and expertise is required to build, setup, or interpret the results [5]. On the other hand, since the notable increase in computational power, Machine Learning (ML) methods have shown

effective as a support in several tasks [4], [6]–[11], like classification, regressions and estimation problems.

In particular, the El Niño Southern Oscillation (ENSO) is a non-regular and complex anomaly fluctuation across the central and east-central equatorial Pacific Ocean [1], [2], which notably affect human life by producing extreme precipitations or droughts around the world [2], [12], [13]. The main characteristic is a variability out of expected sea surface temperature (SST) in this region. The positive phase happens when the average SST is warmer than expected in the ENSO region, which is called El Niño. The opposite phase, colder than average SST, is known as La Niña. This cycle fluctuation is responsible for changes in precipitation and temperature around the world, known as teleconnections [13]–[15].

Some studies have reported statistical and ML approaches to understand the ENSO phenomenon [13]–[15]. Most of the works concentrate on the long-term forecasting/regression problem [15]–[19], which is a hard and challenging problem

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao¹.

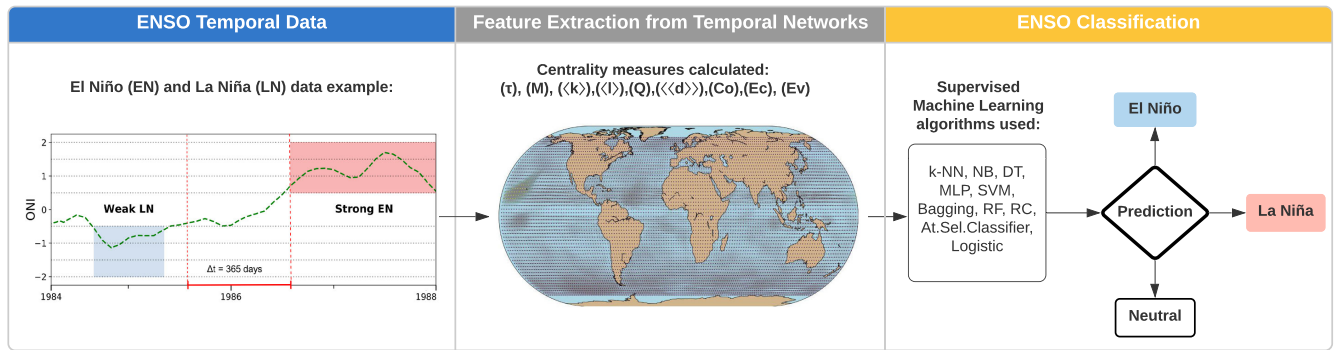


FIGURE 1. Schematic of the proposed method. First, we show an excerpt of the ONI time-series, which is the prediction goal for regression tasks. The inside labels represent the intensity level of each episode. Second, we divide by months the spatiotemporal Surface Air Temperature Anomaly dataset into overlapping windows of $\Delta t = 365$ days. For each sliding window, we calculate the Pearson correlation among all the time-series (points in the globe); then, we establish connections between nodes when the correlation values are significant generating temporal climate networks. For each network, we calculate nine centrality measures that work as features for Machine Learning algorithms. Finally, we perform a classification task analyzing the El Niño episodes and intensities (labels) using 10 (traditional and ensemble) classifiers.

given the non-linear and chaotic behavior of the El Niño and La Niña phases. Despite the efforts, previous works disregarded the inclusion of topological information in the methods, as well as spatiotemporal similarity patterns. Moreover, it is important to understand first the micro and macro characteristics of the phenomenon in the short term (time) and local areas (space). For this reason, we focus on the classification task of the ENSO occurrences to better characterize the behavior of the system, and the proposal of suitable topological descriptors, or attributes, that can be employed in further predictions models.

For structural features of the systems, Network Sciences (NS) brings a powerful toolbox for representing and characterizing spatiotemporal datasets [7], [8]. For example, to analyze macro and intermediate-scale interactions, or local transitive dynamics like the clustering coefficient or motifs patterns in the network [8]. The topological characterization of the complex system has shown improvements in analyzing the ENSO region and other climate systems [20], [21], due to the precise network construction of the dataset in the attribute space [22], [23]. Therefore, the straightforward approach is to combine NS with ML [7], [8], in which the former provides significant inputs to the algorithms. These enriched descriptors of the interactions among components can enhance the results in the case of Climate Networks (CN) [21] and the classification task of diverse domains [8].

Here, we present an approach to classify patterns in spatiotemporal grid data by considering structural properties from functional networks and applying ML algorithms. A method is proposed to classify the previous conditions around the world of ENSO states, combining network-based features and ML algorithms. The idea is to transform the daily near-Surface Air Temperature (SAT) time-series around the globe into a temporal functional network, in such a way that strongly correlated time-series are represented by links [18], [21]. Then, several structural measures are extracted to characterize the ENSO dynamic, which are used as attributes in ten ML algorithms. Therefore, the algorithms learn a model to recognize the micro and macro patterns for

the intensity of the next season in terms of an El Niño or La Niña episodes. The topological properties extracted from the network, which was constructed from the spatiotemporal dataset, brings relevant information that single ML algorithms can not detect from systems with high complexity [8].

We test our method in different experimental setups: a binary problem (El Niño and non-El Niño occurrences); and a multi-class problem (strong El Niño, weak El Niño, strong La Niña, weak La Niña, and regular year). In all the cases, the results indicate that the combined classification method correctly predicted the class of the ENSO episode with an AUC ROC equal to or higher than 79%, i.e., nearly 4 out of 5 times, by using the Random Forest classifier. To the best of our knowledge, this is the first work that employs classification algorithms and NS to analyze the ENSO dynamic. More important, as an initial endeavor in this direction, we pave the way towards spatiotemporal classification problems together with NS, which can be adapted according to the particularities of the domain, e.g., sliding windows, networks measures, ML algorithms, and techniques.

The paper is organized as follows. In Section II, we present some of the works related to this proposal. In Section III, we present the motivation of the research problem. In Section IV, we give some basic concepts about the El Niño-Southern Oscillation and the Oceanic Niño Index (ONI) and describe the NS and ML methods adopted in this work. In Section V, we illustrate our proposed methodology using SAT; we describe the classification problems and the statistical indicators used to evaluate the performance of the classifiers. In Section VI, we have the experimental results about the characterization of the network and the classification of the ENSO episodes. Finally, in Section VII, we conclude with some final remarks, discussions, and suggestions.

II. RELATED WORK

In the case of the El Niño phenomenon, some works focused on understanding the effects of the ENSO phases around the world [13], [14], [24] and possible connections with other climate or different phenomena [1]–[3], [12], [24], [25]. In terms

of ML applications, several works seek to predict or forecast long-term ENSO episodes [15]–[19], [26], [27]. They select a group of spatiotemporal datasets and information to feed the forecasting model and perform some regression analyses. The goal is to minimize forecasting errors given a specific indicator of the problem [15], like the Oceanic Niño Index (ONI) shown in Fig. 1. This is a tough problem, in which the forecasting models tend to follow the previous tendency of the index [16], with not significant differences with a random model [27], and the errors gradually increase for longer lead times [15], [16]. Besides, they can not completely model the non-linear characteristic and tendency changes given the high sensitivity to small perturbations of the ENSO dynamic [28]–[31].

Other works have exploited clustering techniques. For example, in [13], the authors proposed to cluster the two types of El Niño episodes based on the similarity between the produced networks by the cross-correlations between 18 specific areas. This way, they produced a new network of 11 nodes (different El Niño events) with the similarity connections and a tentative community division. In [32], the authors employed a probabilistic clustering technique to describe tropical cyclone tracks in the eastern North Pacific. They analyzed the clusters in terms of many factors, such as location, intensity, and their relationships with ENSO. Authors in [33] proposed an index based on the transitivity measure and the average node strength of the El Niño years. They were able to discriminate, without any visual inspection, between Central Pacific (CP) or Eastern Pacific (EP) El Niño events. These results support our hypothesis that CN can provide high-level features for understanding complex phenomena. Using ML algorithms combined with network features can lead to automatic methods for mining spatiotemporal time-series. However, there are two issues limiting clustering approaches: 1) the choice of the number of clusters to be used is not uniquely determined; and 2) there is no precise metric for cluster validation [6].

Recently, some works have used deep learning for ENSO forecasting, such as [27] that employed a deep-learning approach that produces ENSO forecasts for long-time periods. They used a transfer learning to train a CNN on historical and reanalysis data. The proposed approach predicted types of El Niño events 12 months in advance and achieved a hit rate of 66.7% in the validation period (1984-2017). In [34], the authors compared long-short-term-memory models (LSTMs) to linear regression models (LR). They combined SST and zonal winds as predictors data. LSTM exhibited some advantage over LR in terms of the correlation coefficient using daily data; the results are presented via box-plots and graphics and do not have a precise value. In [35], the authors extracted network metrics from CN with LSTMs to forecast ENSO. A 6-month lead prediction measure of Root mean squared error (RMSE) = 0.8897 and Mean absolute error (MAE) = 0.6376 are presented. Both the MAE and RMSE can be used together to diagnose the variation in the errors in a set of forecasts, and the values can range from 0 to

∞ ; however, lower values are better. In this case, the results are not convincing.

Several studies have successfully analyzed their domain problem by adopting the NS framework. Transforming the dataset into a network has allowed finding many interesting patterns and results in different areas, like text-mining [36], health sciences [37], stock markets, among many others [38]. In the case of Earth sciences, the network representation has been largely used for analyzing global climate effects and teleconnections [20], [25], wild-fires events [39], anomalies in annual hurricanes events [40], seismic events [41], and continental moisture recycling process [42].

III. RESEARCH PROBLEM

Instead of performing a forecasting or clustering task to predict some ENSO descriptor, here, we aim to classify and understand the intensity pattern of the El Niño or La Niña episode, as depicted in the inner labels of Fig. 1. In this classification problem, the goal is to recognize the previous climate patterns around the world that can trigger a strong/weak ENSO condition. For example, Fig. 1 shows the complete El Niño episode between 1986 to 1988.¹ We can consider the temperature anomalies of the previous 365 days, construct the corresponding CN, and recognize the topological patterns associated with the activation and intensity of the ENSO events by some ML algorithms.

In regression models, a numerical quantity is predicted, whereas, in classification tasks, a category (one or more labels) is predicted. Clustering is part of unsupervised learning, i.e., it does not use label data to train a model. The proposal and novelty of this work are to employ classification for ENSO analysis using high-level topological features from the temporal networks. It is possible to construct networks from any dataset [8], [21], [36], [39], to generate elaborated features that feed a classification algorithm. More topological features allow more possibilities to be exploited than using only the time-series values to perform the predictions [8]. Moreover, some classification algorithms are not dependent on big data to obtain good results, while deep learning approaches need a vast amount of instances to achieve proper training and good results, which is a limitation in the case of historical climate data.

IV. BACKGROUND REVIEW

A. THE OCEANIC Niño INDEX

In the central and east-central equatorial Pacific Ocean, El Niño and La Niña are opposite phases of the El Niño-Southern Oscillation (ENSO) region, which is one of the most influential climate patterns in the Earth. This irregular oscillation produces significant changes in the atmospheric precipitation system and jet streams, influencing the temperature and precipitation around the world [13], [14]. El Niño is the phase when the Sea Surface Temperature (SST) in the ENSO region exhibits positive anomalous temperatures,

¹which is strong by reached a maximum value above 1.5 in ONI [43]

i.e., warmer than average conditions during the same period. Opposite, La Niña happens when the SST in the ENSO region is colder than average conditions; i.e., it presents negative anomalous temperatures.

The ONI is one of the main index used to monitor the occurrence of El Niño and La Niña phenomena. We used the public updated NOAA’s ONI compiled data according to the Extended Reconstructed SST version 5 (ERSSTv5) filter [44], from the Climate Prediction Center (CPC) covering the years 1950 to July 2019. The index is defined as the average of monthly SST anomalies in the Niño 3.4 region (Fig. 2). The monthly anomalies are calculated by subtracting the average temperature in the same month over the past 30 years. Then, it is calculated the running three-months average [44]. The conditions of El Niño occur when the ONI anomaly values are greater than or equal to $+0.5^{\circ}\text{K}$, indicating that the El Niño 3.4 region is warmer than usual. La Niña conditions are presented when observed values less than or equal to -0.5°K on the ONI, indicating that the same Pacific region is colder than usual [33], [44].

Episodes of El Niño or La Niña are defined active only if five or more consecutive overlapping 3-month running intervals happen during the year. This is the specification that CPC uses to define El Niño or La Niña episodes [44]. Here, we use the ENSO category of intensity generalizing to weak or strong events [43]. When the average of the three months anomalies over the continuous El Niño episode of the year are larger than 1.0, the event is categorized as Strong El Niño (SEL). Otherwise, it is a Weak El Niño (WEL) episode. If two successive events started during the same year, we adopted the one that was larger and more intense. Similarly, a Strong La Niña (SLA) episode is defined for the average of three months anomaly values lower than -1.0 . Otherwise, it is considered as a Weak La Niña event. In this way, we present in Table 1 all the years, from July 1950 to July 2019, categorized according to strong or weak ENSO episodes.

It is important to notice that there can exist a variety of El Niño intensity categories in the literature [24], which can be used as labels in the classification task. However, here we are not proposing a new ENSO categorization, but how classification algorithms can predict the adopted category for the ENSO episodes.

B. CLIMATE NETWORKS REVISITED

We start defining a network $G = \{V, E\}$, which consists of a set of $N = |V|$ nodes that represent spatial points where a piece of information or signal over time is collected. The set of $M = |E|$ links show the nodes that are connected between each other, where $E \subseteq \{(i, j) \mid i, j \in V\}$. The links describe the similarity level between the nodes. In the case of unweighted networks, the links have binary weights. The mathematical entity that describes the network is the adjacency matrix $A = A_{ij}$, where the elements A_{ij} are equal to unity if i and j are connected, and zero otherwise.

The dataset is not always naturally represented as a network, but rather by spatiotemporal events, or time series.

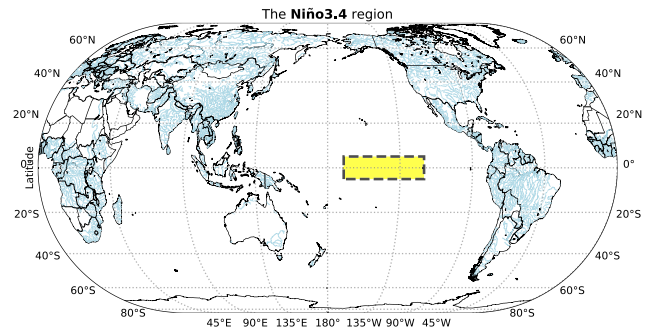


FIGURE 2. Region of the El Niño3.4 considered for the ONI index, delimited from latitude 5°N to 5°S and from longitude 170°W to 120°W .

TABLE 1. Adopted binary categorization of the years according to the El Niño 3.4 ONI anomaly intensity [43]. We consider the Weak episodes as the class (Weak) and all Moderate, Strong, and Very Strong episodes as the class (Strong). Italic years are originally Moderate and bold years are Very strong episodes.

El Niño	La Niña
Weak	
1952-53, 1953-54, 1958-59, 1969-70, 1976-77, 1977-78, 1979-80, 2004-05, 2006-07, 2014-15, 2018-19	1954-55, 1964-65, 1971-72, 1974-75, 1983-84, 1984-85, 2000-01, 2005-06, 2008-09, 2016-17, 2017-18
Strong	
<i>1951-52, 1954-55, 1957-58, 1963-64, 1965-66, 1968-69, 1972-73, 1982-83, 1986-87, 1987-88, 1991-92, 1994-95, 1997-98, 2002-03, 2009-10, 2015-16</i>	<i>1955-56, 1970-71, 1973-74, 1975-76, 1988-89, 1995-96, 1998-99, 1999-00, 2007-08, 2010-11, 2011-12</i>

In these cases, a suitable network construction method must be selected [21]–[23], [39]. Here, we adopted the functional or correlation networks as the approach for constructing the CN [21]. The linking process considers the grid cells or spatial points that are significant and high correlated with each other. Pearson-based correlations are the standard functions employed for assessing the similarity between time-series [18], [20], [21], [38] but many other functions can be appropriated as well.

After the construction process, several topological measures can be employed to characterize the CN. We briefly describe the measures used here. However, interested readers can find the full concepts and definitions in [23], [45]. The primary descriptors are related to the number of links M and the average degree $\langle k \rangle = 2M/N$, which is the density of links per node. From network connectivity, the triangles – or cycles of order three, are the cases in which three nodes are mutually connected. We quantify the proportion of triangles τ on the network employing the transitivity measure [45].

Regarding the adjacency matrix, it is possible to identify the influence of nodes given the spectral properties of the matrix. The Eigenvector of the highest eigenvalue of A describes the importance of the nodes related to the location and connections. We consider Ev as the average Eigenvector centrality of the nodes. On the other hand, the network can also be decomposed into cores by using the k-shell index.

Nodes with higher k-shell values are central in the network. The coreness of the network is the average k-shell values among all the nodes.

In terms of path lengths, the shortest-path length between two different nodes $\ell_{i,j}$ is the minimum number of edges in a path that connects both i and j nodes. Then, the average shortest-path $\langle \ell \rangle = \langle \ell_{i,j} \rangle$ is the average number of edges crossed along all the shortest-paths between all the nodes of the network. In the case that there is no path between i and j , we assume $\ell_{i,j} = N$. The Eccentricity of a node i is its largest shortest-path length over all other nodes, i.e., $\max(\ell_{i,v})$. The eccentricity of the network is then the average eccentricity of the nodes, $Ec = \langle \max(\ell_{i,v}) \rangle$.

In terms of distances, the normalized distance between two different nodes is $\hat{d}_{i,j} = D_{i,j}/D_{\max}$, i.e., the geodesic distance between the spatial location points of i and j divided by half the circumference of the Earth D_{\max} , which is the maximum possible distance on the planet. The local link distance $d_i = \langle \hat{d}_{i,j} | A_{ij} > 0 \rangle$ is the average of the normalized distance of all the network neighbors of i . Therefore, the Global Average Link Distance $\langle \langle d \rangle \rangle$ [21] is the average local link distances of the nodes, which is $\langle \langle d \rangle \rangle = \langle d_i \rangle$.

Last, the modularity indicates how well modular divided into communities is the network, i.e., networks with higher modularity value are well divided in communities with a larger proportion of links intra-community than inter-community. We consider here the Q modularity [45].

C. MACHINE LEARNING REVISITED

Machine learning (ML) is an area of Artificial Intelligence (AI) whose objective is the development of computational techniques on learning as well as the construction of systems capable of acquiring knowledge automatically [46]. Algorithms in ML can be classified into predictive and descriptive tasks [6]. The first task aims to describe the information encoded in the data, by finding patterns and without providing any forecast. Descriptive algorithms mostly rely on clustering or association rules techniques. On the other hand, predictive tasks aim to predict a target feature, based on the information from other variables and a set of labeled data. Predictive algorithms mostly rely on classification and regression problems.

Here, we focus on the classification problem and employed predictive algorithms. The classification problem consists of acquiring data samples and decide which of the M classes they belong to, based on training using examples from each class. The classification problem is discrete, so each example belongs precisely to one class within a set of classes $C = \{c_1, \dots, c_m\}$. The formal definition of the classification problem is: given a set $D = \{(x_i, f(x_i)), i = 1, \dots, n\}$, where f represents an unknown function from which the machine learning algorithm learns a f' approximation of the f function. Using the f' function it is possible to estimate the value of f for new values of x , so that in the case of classification we have $y_i = f(x_i) \in c_1, \dots, c_m$.

In the following, we briefly present the ten classifications algorithms (classifiers) selected in this work, and more details can be found in [6], [46]. We use the implementation from Weka² and Scikit-learn.³ Since there is no single algorithm that obtains the best performance in all problems — no free lunch assumption, it is important to evaluate different classifiers in order to find the most appropriate for the ENSO classification with CNs.

- *k-Nearest Neighbors (k-NN)*: classifies an example by the vote of its neighbors, assigning the most common class among its k nearest neighbors (given some distance measure like Euclidean). We used k equal to three.
- *Naive Bayes*: is a probabilistic classifier based on Bayes' theorem with strong independence assumptions between the features. It assigns to a instance x probabilities for each of possible class C_m as $P(C_m|x) = p(C_m)p(x|C_m)/p(x)$.
- *Decision Tree*: generates a structure similar to a diagram in which each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents the target variable (class label). There are many specific decision-tree algorithms; here, we used the J48 algorithm.
- *Multilayer Perceptron (MLP)*: is a feedforward artificial neural network that uses a mathematical model for information processing. MLP uses backpropagation for calculating the weights updates in the network until it can perform the task for which it is being trained. Besides the input and output layer, the number of hidden layers considered here was $(\text{attributes} + \text{classes})/2$.
- *Support-Vector Machines (SVMs)*: constructs a set of hyperplanes in a high-dimensional space that maximizes the distance from each example to the nearest data point on each side. It can efficiently perform linear and non-linear classification using a kernel trick that implicitly maps the inputs into high-dimensional feature spaces. The kernel used here was PolyKernel and John Platt's sequential minimal optimization algorithm.
- *Bootstrap Aggregation (Bagging)*: is based on Bootstrap, a statistical estimation technique which performs some statistical measure like mean from multiple random samples with replacement from the dataset. Different machine learning models are trained from several random samples of the training data. The results are averaged to make a better prediction. The used classifier was the decision tree (CART).
- *Random Forest*: is an improvement of Bagging since, during the tree creation, splits points can be selected from a random subset of the attributes and not by the traditional greedy strategy that selects the best split point at each step.
- *Random Committee*: generates several random base classifiers by using a different random number seed.

²Weka: <https://www.cs.waikato.ac.nz/ml/weka/>

³<https://scikit-learn.org/stable/>

The final prediction is an average of the predictions generated by the individual base classifiers. The used classifier was Random Tree.

- *AttributeSelectedClassifier*: reduces the dimensionality of training and test data by attribute selection before passing the data to a classifier. The classifier used was Decision Tree J48.
- *MultiClass*: is a meta classifier for handling multi-class datasets with 2-class classifiers. The classifier used was Logistic.

V. MATERIAL AND METHODS

A. DATA

We used the global daily near-surface (1000 hPa) Air Temperature data (SAT), from the National Oceanic and Atmospheric Administration (NOAA) and the Reanalysis I project of the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) [47]. SAT has been used by several authors to study and analyze the ENSO phenomenon [26], [33], [35]. We opt for using this dataset instead of the SST due to the inclusion of the land areas, which is extra information that we consider important for the analyses. Besides, the inclusion of global data is relevant given that the El Niño Oscillation is not an isolated phenomenon in the central pacific, but connected with many other climate regions around the world as a complex system.

The dataset covers the period from 1948 to 2018, with the information of the daily temperature of spaced points of $2.5^\circ \times 2.5^\circ$ degrees of latitude and longitude resulting in a total of 10512 points around the globe, i.e., land and sea areas. We removed the points located in the poles, considering only the geographical data located between latitudes 66.5°N and 66.5°S . This strategy is because the high density of points located in the poles negatively influenced the analysis.

After that, we calculate the long-term average temperature centered on each of the 365 calendar days, to remove the seasonality in the daily temperature on each spatial point. In other words, the 365-day climatology is the average over 70 samples of the same day among all the years (1948 – 2018). For each calendar day, we subtract its corresponding long-term average temperature, obtaining the SAT anomalies time-series of the grid-points. We removed the data corresponding to February 29th in the case of leap-years. As a result, the SAT anomalies (SATa) are the raw data used here for constructing the CNs.

B. TEMPORAL CN AND ML ANALYSIS

Initially, we divide the spatiotemporal SATa dataset into overlapping windows of Δt days for all the spatial points. Given that the complete climatic seasonal cycle of the planet is one year (four seasons) and the interannual nature of the ENSO, we set $\Delta t = 365$ days [26]. The sliding window is moved every month, in which the reference is the last month in the interval, e.g., (July 1, 1950 – Δt , July 1, 1950), (August 1, 1950 – Δt , August 1, 1950), . . . , ($m\text{-yy} - \Delta t$, $m\text{-yy}$) and so on.

For example, when talking about the CN of December 2019 ($m\text{-yy}$), we are referring the time-series of SATa in the interval (12-2018, 12-2019), i.e., the CN of the previous year, and inferring the class of the ENSO episode for the next three months according to the topological properties of the CN in ($m\text{-yy}$).

The temporal CN (\mathcal{G}) is the ensemble of CNs from all the months in the dataset. The \mathcal{G} describes the ordered sequence of networks at the different intervals [48], [49], i.e., $\mathcal{G} = \{G_1^{(7-50)}, G_2^{(8-50)}, \dots, G_l^{(m\text{-yy})}\}$, with l the number of layers or sliding windows and the superscript ($m\text{-yy}$) is the month of reference. This temporal CN contains not only the similarity between nodes but also the global evolution by months of the climatic temperature system.

For each sliding window $G_x^{(m\text{-yy})}$, we calculate the Pearson correlation among all the time-series, producing the similarity matrix between the spatial points or nodes. From the similarity or correlation matrix, we establish connections between nodes when the absolute correlation values are higher than 0.65, based on previously reported works [33], [50]. Therefore, this is the adjacency matrix that represents the CN of the SATa from the previous year until the month of reference ($m\text{-yy}$). Each $G_x^{(m\text{-yy})}$ network contains the climatic dynamic and similarity of the nodes around the globe, in which the topological features can describe the formation and evolution of the ENSO phenomenon [7], [8], [21], [23]. Here, we aim to understand the El Niño and La Niña evolution by each month with the information of the CN constructed with the SATa of the previous year.

To conduct the classification experiments, we calculate several structural measurements for each of the CNs in \mathcal{G} and put the new information into a new dataset of attributes. The adopted measures are the transitivity (τ), the number of links (M), average degree ($\langle k \rangle$), average shortest-path length ($\langle \ell \rangle$), modularity (Q), global average link distance ($\langle \langle d \rangle \rangle$), and the averages of Coreness (Co), Eccentricity (Ec), and Eigenvector (Ev) centralities. Additionally, each of the 822 $G_x^{(m\text{-yy})}$ CNs has its ONI value, which correspond to the next three month of the reference, i.e., ($m + 1\text{-yy}$), ($m + 2\text{-yy}$), ($m + 3\text{-yy}$). Hence, we consider the following three experimental setups:

- **Dataset1**, two classes with El Niño (EN) and no El Niño occurrences or anything else (AE);
- **Dataset2**, three classes with El Niño (EN), La Niña (LN) occurrences, and regular years (RY) that means neither El Niño nor La Niña occurred;
- **Dataset3**, five classes with strong El Niño (SEN), weak El Niño (WEN), strong La Niña (SLN), weak La Niña (WLN), and regular year (RY).

Thus, we explore one binary and two multi-class classification problems. Instances of an ENSO episode, as explained in Section IV-A, receive the average of the ONI values of the episode they pertain. Afterward, we label the instances according to the intensity of the El Niño or La Niña occurrence they had a place. Information about the number of instances, classes, and distributions are reported in Table 2.

TABLE 2. Information of the experimental setup and classes.

	# Classes	# Samples in each class
Dataset1	2	EN (234), AE (588)
Dataset2	3	EN (234), LN (227), RY (361)
Dataset3	5	SEN (137), SLN (84), RY (361), WEN (97), WLN (143)

The classifiers presented in Section IV-C were used to perform the classification task in the three proposed datasets.

C. STATISTICAL EVALUATION

To evaluate the performance of the classifiers, the original dataset was divided into the training set and the test set. In the training set, we used the K-Fold Cross Validation method. The data set was divided into $k = 10$ partitions so that $k - 1$ groups are used to train the model, and the remaining partition is used for evaluation. This process is performed k times by alternating the evaluation partition, and, at the end of the iterations, we obtain the performance of the model with the test set [6].

The confusion matrix (Fig. 3) is the final result of the classifier when applying the test set. This is a square matrix with a size equal to the number of target labels or classes. The columns represent the true classes the algorithm is trying to predict, and the rows represent the classes predicted by the classifier. Each element in the matrix matches the number of samples of class column- j that were classified as belonging to class row- i [6]. The results of the confusion matrix are essential because several performance metrics of the classifier are calculated according to the elements of the matrix [6].

In Fig. 3, we show the metrics used in this work: **Accuracy** is the ratio between the number of correct predictions over the total number of samples in the test. **Recall** is the ratio between the total correct predicted instances of a class over the total number of instances of that class, i.e., relevant instances that were correctly retrieved. **Precision** or positive predicted value is the fraction of correctly predicted samples of a class over the total number of predicted samples of the class. **F1-Score** is the harmonic mean between the precision and recall measures. **AUC** (Area Under The Curve) **ROC** (Receiver Operating Characteristics) curve is an evaluation metric for the classification model's performance. ROC is a probability curve that plots the TP rate = $(TP / (TP + FN))$ against the FP rate = $(FP / (FP + TN))$ at various threshold settings. AUC measures the entire two-dimensional area underneath the entire ROC curve and represents a measure of separability. An excellent model has AUC near to 1, which means it has a good measure of separability (i.e., if we plot the positive and negative curves they do not overlap), and an AUC near to 0 means the model has the worst measure of separability (the model is predicting negative class as a positive class and vice versa). If AUC is 0.5, it means the model has no class separation capacity (i.e., the distributions overlapping following a random selection).

D. FEATURE SELECTION

Feature selection consists of choosing a small subset of features that ideally is sufficient to describe the target value [51]. The main challenge is to perform a trade-off and not select too many or too few features. In this paper, we employed two classical methods namely Information gain and Gain ratio to rank the network-based features/attributes from the dataset.

The Information gain is calculated for each attribute as follows: given an attribute A that splits the set S into subsets S_i , the average entropy is calculated and its sum is compared to the entropy of the original set S . The attribute that maximizes the difference expressed in Equation 1 is selected.

$$Gain(S, A) = E(S) - \sum_i \frac{S_i}{S} \cdot E(S_i) \quad (1)$$

Information gain is biased towards choosing attributes with a large number of values. Gain ratio modifies Information gain to reduce its bias considering the number and size of branches when choosing an attribute. It considers the intrinsic information that is the entropy of the distribution of instances into branches.

$$Gain - ratio(S, A) = \frac{Gain(S, A)}{IntI(S, A)} \quad (2)$$

VI. EXPERIMENTAL RESULTS

In this section, we report the results of the proposed approach for analyzing and classifying the ENSO phenomena using NS and ML.

A. SPATIOTEMPORAL NETWORK CHARACTERIZATION AND EVOLUTION

Initially, we show, on a global scale, the evolution of the topological properties of the SATa temporal CNs. Fig. 4 depicts the variation in the measurements between 1950 and 2019, along with the periods of El Niño or La Niña occurrences. So, we have detailed how the spatiotemporal properties of the temporal networks evolve, in which the blue/red background colors indicate La Niña or El Niño episodes, respectively. Most of the measures present some notable perturbations when occurring El Niño or La Niña phases. In other words, given CNs of the previous twelve months, there exist some characteristic patterns in agreement to the next ENSO phase.

In general, τ , $\langle k \rangle$, $\langle \langle d \rangle \rangle$, Co , and Ev measures present a similar peak behavior. In this scenario, we have the exact behavior between the number of links M and the average degree $\langle k \rangle$, due to the size of the networks is always the same. On the other hand, the modularity has an opposite or inverse behavior that the measures mentioned above. Lower Q values mean that the network is less modular divided into communities or sub-components. Larger $\langle \langle d \rangle \rangle$ means that the nodes have more long-range connections, which is the principle of teleconnections. Larger $\langle k \rangle$ together with larger Co means that the network has a larger number of hubs, which are densely connected in specific clusters. This is in accordance with the high Ev values, evidencing a network with a high

		True condition	
		True Positive (TP)	False Positive (FP)
Predicted condition	True Positive (TP)	True Positive (TP)	False Positive (FP)
	False Negative (FN)	False Negative (FN)	True Negative (TN)

FIGURE 3. Example of a confusion matrix for a disjoint two-class problem (left), and Accuracy, Precision, and Recall measures calculated from the matrix (right).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)},$$

$$Precision = \frac{TP}{(TP + FP)}, \text{ Recall} = \frac{TP}{(TP + FN)},$$

$$F_1\text{-score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

diffusion rate when propagating a signal. Only Ec and $\langle \ell \rangle$ show different behaviors than the other measures.

We observe an intriguing phenomenon between 1993 and 1994 (Fig. 4), where there is a peak or perturbation detected by many of the measures. However, there were none ONI conditions during that period. We found that this anomalous period was reported and studied in other works [28]–[31]. The cause of the phenomenon is not clear yet, varying from: the connection to some climatic adjustment occurring in a recurrent period of tens or hundreds of years [28], [29]; a stronger exogenous effect caused by climate changes [30]; or the influence of strong volcanic eruptions in the global temperature [31]. Given the anomalous condition of the global climate system in this period, and that investigating this topic is out-of-scope here, we decided to remove the nine months corresponding to this time frame in the classification analyses. This way, we avoid the outlier data from extrinsic phenomena undermining the classification task.

We also show, as examples, the local Eigenvector centrality of the global SATa temporal networks, for a Regular (Fig. 5a), and Strong El Niño (Fig. 5b) and La Niña (Fig. 5c) years. Recalling that for each map, we construct the CN for the interval of the previous 365 days until the month of reference ($mm\text{-}yy$). Then, the nodes around the world are connected if their SATa time-series are high correlated according to the Pearson coefficient, as explained in Section V-B. After obtaining the CNs, we calculated the Eigenvector centrality and plotted the value of the nodes in Fig. 5.

We observe that for the (12-1997) and (12-1998) CNs (Fig. 5(b) and (c), respectively), the ENSO region in the eastern tropical Pacific has the most influential nodes in the CNs. Given the network construction on top of higher correlated SATa points around the world, nodes with the highest Eigenvector centrality are those highly influencing the temperature anomalies in the CN. More interesting, in Fig. 5a, we can observe the influential nodes located in the South Pacific Ocean. These high central points are located in a region known as the South Pacific Oscillation (SPO) [52]–[54], between $35^\circ\text{N} - 60^\circ\text{S}$ and $170^\circ\text{W} - 90^\circ\text{W}$. Recently, some works found pieces of evidence that the intensity of the cold or warm ENSO episodes is likely related to patterns of variation in the sea level pressure and temperature in the SPO region [52], [53], in the previous months. Thus, the CN previous the strong El Niño year is showing a cluster of influential

nodes in the SPO region. The results in Fig. 5 show that the proposed temporal network approach and the topological characterization of the CN can capture the complex dynamics and relations between key regions around the world.

B. DATA CLASSIFICATION

The SATa dataset, described in section V-A, was employed to construct the temporal CNs, as explained in Section V-B. Nine network measures were calculated and used as features in the classification algorithms. We consider ten algorithms described in Section IV-C: k-NN, Naive Bayes, Decision Tree, Multilayer Perceptron (MLP), Support-Vector Machines (SVM), Bootstrap Aggregation (Bagging), Random Forest, Random Committee, Attribute Selected Classifier (Att.Sel.Classifier) and MultiClass. We present the classification results regarding the three datasets configurations. The classifiers evaluation was considering the Accuracy, Precision, Recall, F1-Score, and AUC ROC, from 10-fold cross-validation, which is a re-sampling procedure used to avoid overfitting in ML models [46].

C. RESULTS FOR DATASET1

Dataset1 consists of two classes with El Niño (EN) and no El Niño occurrences/anything else (AE). Classification results are shown in Table 3. Although this dataset represents a binary classification, we have a very imbalanced class distribution where only 28.5% of the data represents El Niño occurrences. All classifiers achieved Accuracy around 70% and F1-score between 59% and 76%. The best results were reached by Random Forest (0.78 of Accuracy, 0.76 of F1-score, and 0.79 of AUC ROC).

D. RESULTS FOR DATASET2

Dataset2 consists of three classes with El Niño (EN), La Niña (LN) occurrences, and regular years (RY), which means no El Niño neither La Niña occurrence. Classification results are shown in Table 4. This dataset is a bit more difficult than Dataset1 because it represents a multiclass problem (three classes) and it is also imbalanced with 28.5% of the data representing EN and 27.6% of the data representing LN, while the remaining represent RY. For this reason, Accuracy decreased to 48%-58%. Again the best results were reached by Random Forest (0.58 of Accuracy, 0.57 of F1-score and 0.79 of AUC ROC).

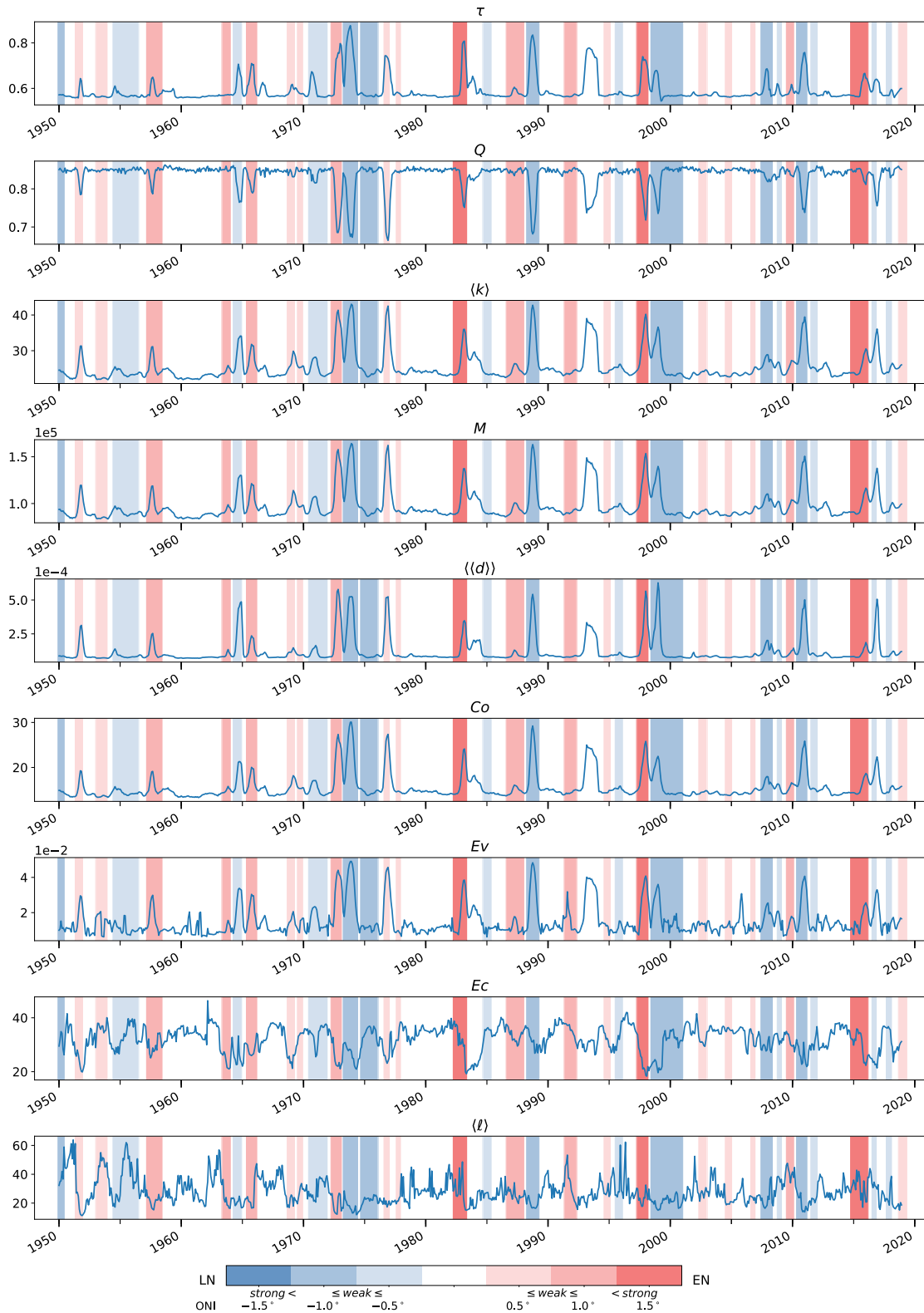


FIGURE 4. Characterization of the temporal networks according to the adopted topological measures. In the background, red colors (right side in the legend) show the months that occur El Niño events, and blue colors (left side in the legend) are the corresponding La Niña events. The measures, from top to bottom, are the transitivity (τ), modularity (Q), average degree ($\langle k \rangle$), the number of links (M), global average link distance ($\langle\langle d \rangle\rangle$), and the averages of Coreness (Co), Eigenvector (Ev) centralities, Eccentricity (Ec), and shortest-path length ($\langle \ell \rangle$).

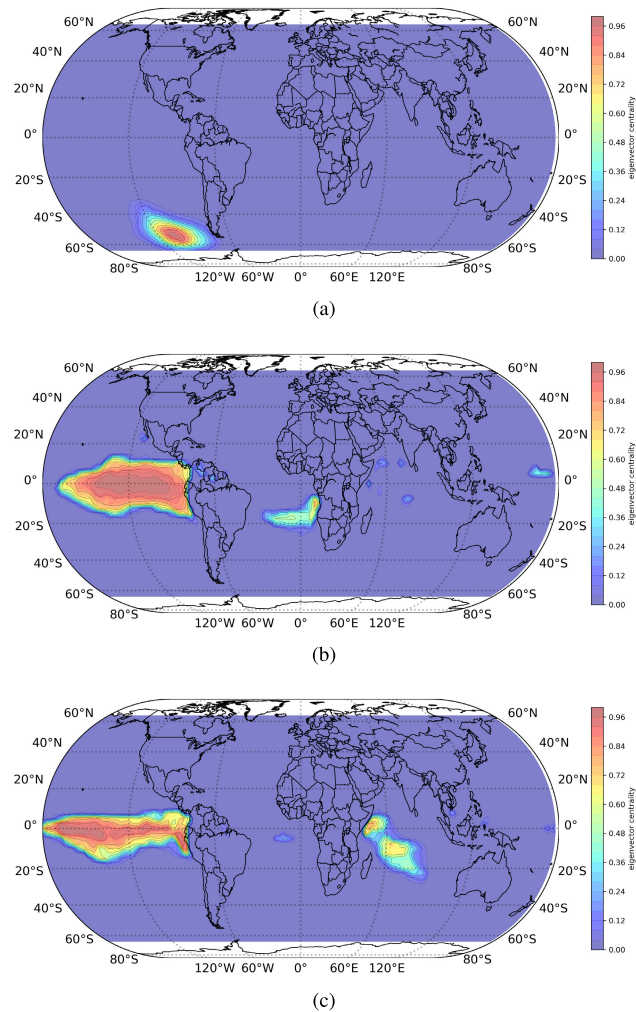


FIGURE 5. Local characterization by the Eigenvector centrality for three cases of temporal networks: (a) (12-1996), representing a CN with no El Niño or La Niña episode; (b) (12-1997), representing a CN of El Niño episode; (c) (12-1998), representing a CN of La Niña episode. Each CN is constructing using the SATa of the previous 365 days according to the month of reference.

TABLE 3. Classification results for the two cases: El Niño (EN) or anything else (AE). In bold are the best values.

Method	Accuracy	F1-score	Precision	Recall	AUC ROC
Decision Tree	0.71	0.63	0.66	0.71	0.59
Random Forest	0.78	0.76	0.77	0.78	0.79
MLP	0.74	0.68	0.72	0.74	0.68
SVM	0.71	0.59	0.51	0.71	0.50
Naive Bayes	0.70	0.66	0.65	0.70	0.60
Bagging	0.76	0.72	0.74	0.76	0.75
Att.Sel.Classifier	0.72	0.61	0.80	0.72	0.57
Rand. Committee	0.75	0.74	0.73	0.75	0.76
KNN-3	0.73	0.73	0.72	0.74	0.70
MultiClass	0.72	0.65	0.69	0.72	0.63

E. RESULTS FOR DATASETS

Dataset3 consists of five classes with strong El Niño (SEN), weak El Niño (WEN), strong La Niña (SLN), weak La Niña (WLN), and regular year (RY). Classification results are shown in Table 5. This dataset is more difficult than

TABLE 4. Classification results for the three cases: El Niño (EN), La Niña (LN), and Regular year (RY). In bold are the best values.

Method	Accuracy	F1-score	Precision	Recall	AUC ROC
Decision Tree	0.49	0.48	0.48	0.49	0.65
Random Forest	0.58	0.57	0.58	0.58	0.79
MLP	0.51	0.47	0.52	0.51	0.67
SVM	0.50	0.41	0.53	0.50	0.57
Naive Bayes	0.48	0.40	0.43	0.48	0.62
Bagging	0.54	0.53	0.53	0.54	0.74
Att.Sel.Classifier	0.50	0.47	0.47	0.49	0.64
Rand. Committee	0.58	0.58	0.58	0.58	0.76
KNN-3	0.57	0.56	0.57	0.57	0.72
MultiClass	0.51	0.46	0.50	0.51	0.65

TABLE 5. Classification results for the five cases: Weak El Niño (WEN), Strong El Niño (SEN), Weak La Niña (WLN), Strong La Niña (SLN), and Regular Year (RY). In bold are the best values.

Method	Accuracy	F1-score	Precision	Recall	AUC ROC
Decision Tree	0.45	0.45	0.44	0.45	0.66
Random Forest	0.55	0.53	0.54	0.55	0.80
MLP	0.48	0.39	0.43	0.48	0.69
SVM	0.47	0.34	0.32	0.47	0.57
Naive Bayes	0.46	0.36	0.33	0.46	0.63
Bagging	0.53	0.49	0.50	0.53	0.76
Att.Sel.Classifier	0.47	0.46	0.46	0.48	0.68
Rand. Committee	0.53	0.51	0.50	0.53	0.77
KNN-3	0.50	0.48	0.48	0.50	0.72
MultiClass	0.47	0.37	0.40	0.47	0.67

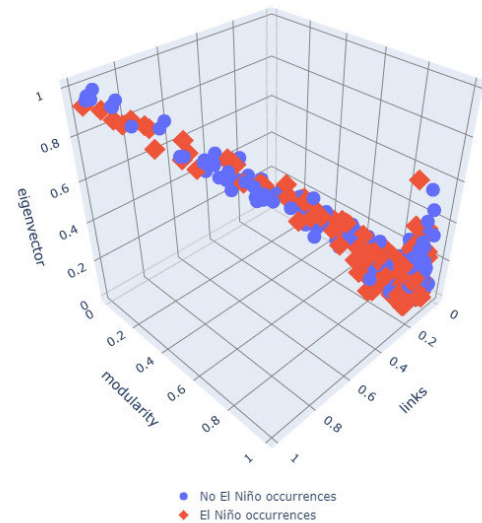


FIGURE 6. Attribute space distribution of the samples according to topological features in the case of Dataset1 (two classes).

Dataset1 and Dataset2 because it represents a multiclass problem (five classes) and is also imbalanced with 16.6% of the data representing SEN, 10.2% of the data representing SLN, 11.8% of the data representing WEN, 17.4% representing WLN and 43.9% of the data are RYs. For this reason, Accuracy decreased to 47%-55%. Again the best results were reached by Random Forest (0.55 of Accuracy, 0.53 of F1-score and 0.80 of AUC ROC).

TABLE 6. Pair-wise spearman correlation between the topological features among the temporal climate networks.

MEASURES	τ	$\langle k \rangle$	$\langle \ell \rangle$	Q	$\langle \langle d \rangle \rangle$	E_c	E_v	C_o	M
Transitivity	1.00								
Average degree	0.88	1.00							
Average path length	-0.37	-0.42	1.00						
Modularity	-0.49	-0.54	0.22	1.00					
Global average link distance	0.88	0.98	-0.43	-0.53	1.00				
Eccentricity	-0.61	-0.72	0.48	0.35	-0.75	1.00			
Eigenvector	0.50	0.59	-0.30	-0.60	0.59	-0.44	1.00		
Coreness	0.89	1.00	-0.43	-0.53	0.98	-0.73	0.58	1.00	
Links	0.88	1.00	-0.42	-0.54	0.98	-0.72	0.59	1.00	1.00

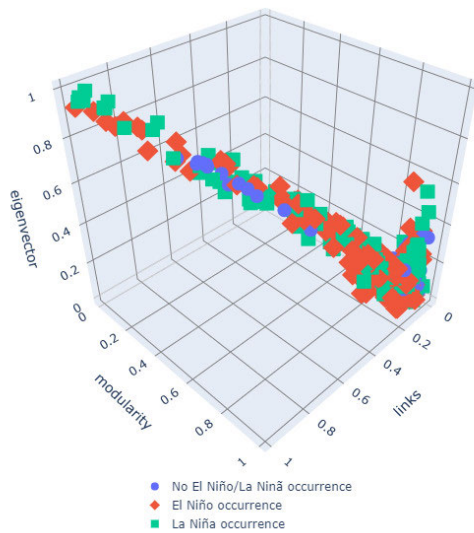


FIGURE 7. Attribute space distribution of the samples according to topological features in the case of Dataset2 (three classes).

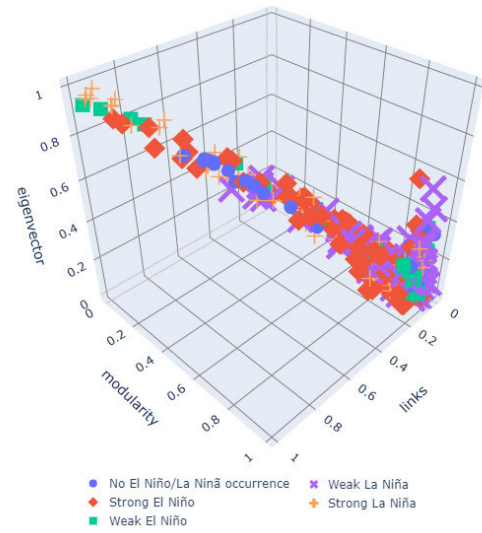


FIGURE 8. Attribute space distribution of the samples according to topological features in the case of Dataset3 (five classes).

F. DISCUSSION

The accuracy degraded as the number of classes increased, which is expected since it becomes more difficult to separate the data in the attribute space. The before can be seen in Figs. 6, 7 and 8. Multiclass and multivariate classification problems are a challenge and relevant in recent machine learning studies. Here, we employ different techniques (traditional classifiers and ensemble), and all of them present a similar performance. Ensemble algorithms achieved a bit higher results, especially Random Forest, than single classifiers. They employ a set of classifiers and combine the predictions. Random Forest, for example, uses a multitude of decision trees. The trees are created by a random subset of the attributes to reduce the correlation between estimators in the ensemble by training them on random samples of features and avoid bias on features that appear highly descriptive in the training set. So, this model can deal better with high dimensional data and avoid overfitting.

A high number of features, typically called multivariate, in some cases, can difficult the class separation capabilities. We employ two feature selection techniques – Information Gain and Gain Ratio – to rank the best attributes from the nine network measures in the classification: transitivity,

modularity, average degree, the number of links, global average link distance, and the averages of Coreness, Eigenvector centralities, Eccentricity, and shortest-path length. The best-ranked attributes were Eigenvector, Modularity, and number of links. We also calculated the pair-wise Spearman coefficient between all the measures (Table 6) to check the information correlation of the features. We confirm that the best-ranked measures by the feature selection algorithm also have a lower correlation between each other. This way, the classifiers could be trained using only these three features leading to short running time and similar results.

Networks are a useful tool to represent complex systems by modeling interactions among the elements from the data. When analyzing the problem with this representation, we can find patterns about the system and better understand its behavior. Moreover, networks help us understand how microscopic interactions lead to macroscopic regularities. Here, we combine NS with ML to study the ENSO phenomenon. ML builds models to learn patterns/behavior without much domain knowledge, whereas NS aims to find domain knowledge by characterizing the system and its interactions. One challenge is representing temporal connectivity patterns, and this work contributes in this direction.

VII. CONCLUSION

This work presents an approach combining complex network measures and machine learning techniques to classify intensities of El Niño episodes. We model the spatiotemporal dataset as temporal climate networks and extracted nine topological measures. The measures were employed as features in binary and multi-class classification problems. Random Forest was always the best classifier. In the binary classification, the best result was 79% of AUC ROC and 76% of F1-score; in the three-class dataset, the best result was 79% of AUC ROC and 58% of F1-score; whereas in five-class the best result was 80% of AUC ROC and 53% of F1-score.

The proposed method is an objective and automatic classification approach that not requires visual inspection and depends only on a few attributes. The attributes are the topological properties of the evolving climate networks constructed from the publicly available global near-surface (1000 hPa) air temperature and the adopted ENSO labels (categories). Given the computational cost lower than deep learning or EOF-based methods and no limitations about dataset size, our approach is a potential tool to analyze any other complex spatiotemporal phenomena related to time-series datasets.

We conclude that both complex networks and machine learning are efficient tools for analyzing vast volumes of data collected by satellites and sensors. Besides, climate change research can help society cope with climate extremes and minimize their impacts. Network science is expanding and is an excellent tool for interdisciplinary applications. Here, network science meets machine learning, and enhanced results can be achieved for the classification of complex phenomena like ENSO.

REFERENCES

- [1] P. Cashin, K. Mohaddes, and M. Raissi, "Fair weather or foul? The macroeconomic effects of El Niño," *J. Int. Econ.*, vol. 106, pp. 37–54, Nov. 2017.
- [2] V. B. Rao, K. Maneesha, P. Sravya, S. H. Franchito, H. Dasari, and M. A. Gan, "Future increase in extreme EL Niño events under greenhouse warming increases Zika virus incidence in south America," *NPJ Climate Atmos. Sci.*, vol. 2, no. 1, p. 4, Dec. 2019.
- [3] S. M. Hsiang, K. C. Meng, and M. A. Cane, "Civil conflicts are associated with the global climate," *Nature*, vol. 476, no. 7361, pp. 438–441, Aug. 2011.
- [4] L. Santos, L. Londe, T. Carvalho, D. Menasché, and D. Vega-Oliveros, *About Interfaces Between Machine Learning, Complex Networks, Survivability Analysis, and Disaster Risk Reduction*. Cham, Switzerland: Springer, 2019, pp. 185–215.
- [5] L. N. Ferreira, D. A. Vega-Oliveros, L. Zhao, M. F. Cardoso, and E. E. N. Macau, "Global fire season severity analysis and forecasting," *Comput. Geosci.*, vol. 134, Jan. 2020, Art. no. 104339.
- [6] S. Marsland, *Mach. Learning: An Algorithmic Perspective*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2014.
- [7] T. Christiano Silva and L. Zhao, *Machine Learning in Complex Networks*. Cham, Switzerland: Springer, 2016.
- [8] M. Zanin, D. Papo, P. A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik, and S. Boccaletti, "Combining complex networks and data mining: Why and how," *Phys. Rep.*, vol. 635, pp. 1–44, May 2016.
- [9] D. A. Vega-Oliveros, L. Zhao, and L. Berton, "Evaluating link prediction by diffusion processes in dynamic networks," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 10833.
- [10] L. Berton, T. de Paulo Falciros, A. Valejo, J. Valverde-Rebaza, and A. de Andrade Lopes, "RGCLI: Robust graph that considers labeled instances for semi-supervised learning," *Neurocomputing*, vol. 226, pp. 238–248, Feb. 2017.
- [11] L. Berton, D. A. Vega-Oliveros, J. Valverde-Rebaza, A. T. da Silva, and A. D. A. Lopes, "Network sampling based on centrality measures for relational classification," in *Information Management and Big Data*, J. A. Lossio-Ventura and H. Alatrística-Salas, Eds. Cham, Switzerland: Springer, 2017, pp. 43–56.
- [12] J. Rao, C. I. Garfinkel, and R. Ren, "Modulation of the northern winter stratospheric El Niño-southern oscillation teleconnection by the PDO," *J. Climate*, vol. 32, no. 18, pp. 5761–5783, Sep. 2019.
- [13] J. Fan, J. Meng, Y. Ashkenazy, S. Havlin, and H. J. Schellnhuber, "Network analysis reveals strongly localized impacts of EL Niño," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 29, pp. 7543–7548, Jul. 2017.
- [14] K. Yamasaki, A. Gozolchiani, and S. Havlin, "Climate networks around the globe are significantly affected by EL Niño," *Phys. Rev. Lett.*, vol. 100, no. 22, Jun. 2008, Art. no. 228501.
- [15] A. G. Barnston, M. K. Tippett, M. Ranganathan, and M. L. L'Heureux, "Deterministic skill of ENSO predictions from the north american multimodel ensemble," *Climate Dyn.*, vol. 53, no. 12, pp. 7215–7234, Dec. 2019.
- [16] A. G. Barnston, M. K. Tippett, M. L. L'Heureux, S. Li, and D. G. DeWitt, "Skill of real-time seasonal ENSO model predictions during 2002–11 is our capability increasing?" *Bull. Amer. Meteorol. Soc.*, vol. 93, no. 5, pp. 631–651, 2012.
- [17] J. Ludescher, A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber, "Improved EL Niño forecasting by cooperativity detection," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 29, pp. 11742–11745, Jul. 2013.
- [18] J. Meng, J. Fan, Y. Ashkenazy, A. Bunde, and S. Havlin, "Forecasting the magnitude and onset of EL Niño based on climate network," *New J. Phys.*, vol. 20, no. 4, 2018, Art. no. 043036.
- [19] M. Li, S. J. Koopman, R. Lit, and D. Petrova, "Long-term forecasting of EL Niño events via dynamic factor simulations," *J. Econometrics*, vol. 214, no. 1, pp. 46–66, Jan. 2020.
- [20] F. Falasca, A. Bracco, A. Nenes, and I. Fountalis, "Dimensionality reduction and network inference for climate data using δ -MAPS: Application to the CESM large ensemble sea surface temperature," *J. Adv. Model. Earth Syst.*, vol. 11, no. 6, pp. 1479–1515, Jun. 2019.
- [21] R. V. Donner, M. Wiedermann, and J. F. Donges, *Complex Network Techniques for Climatological Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2017, pp. 159–183.
- [22] L. Berton, A. de Andrade Lopes, and D. A. Vega-Oliveros, "A comparison of graph construction methods for semi-supervised learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.
- [23] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, "Complex network approaches to nonlinear time series analysis," *Phys. Rep.*, vol. 787, pp. 1–97, Jan. 2019.
- [24] J. Rao and R. Ren, "Parallel comparison of the 1982/83, 1997/98 and 2015/16 super el Niños and their effects on the extratropical stratosphere," *Adv. Atmos. Sci.*, vol. 34, no. 9, pp. 1121–1133, Sep. 2017.
- [25] N. Boers, B. Goswami, A. Rheinwalt, B. Bookhagen, B. Hoskins, and J. Kurths, "Complex networks reveal global pattern of extreme-rainfall teleconnections," *Nature*, vol. 566, no. 7744, pp. 373–377, Feb. 2019.
- [26] J. Meng, J. Fan, J. Ludescher, A. Agarwal, X. Chen, A. Bunde, J. Kurths, and H. J. Schellnhuber, "Complexity-based approach for EL Niño magnitude forecasting before the spring predictability barrier," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 1, pp. 177–183, Jan. 2020.
- [27] Y.-G. Ham, J.-H. Kim, and J.-J. Luo, "Deep learning for multi-year ENSO forecasts," *Nature*, vol. 573, no. 7775, pp. 568–572, Sep. 2019.
- [28] R. J. Allan and R. D. D'Arrigo, "'Persistent' ENSO sequences: How unusual was the 1990–1995 EL Niño?" *Holocene*, vol. 9, no. 1, pp. 101–118, 1999.
- [29] M. Latif, R. Kleeman, and C. Eckert, "Greenhouse warming, decadal variability, or EL Niño? An attempt to understand the anomalous 1990s," *J. Climate*, vol. 10, no. 9, pp. 2221–2239, Sep. 1997.
- [30] K. E. Trenberth and T. J. Hoar, "The 1990–1995 EL Niño-southern oscillation event: Longest on record," *Geophys. Res. Lett.*, vol. 23, no. 1, pp. 57–60, Jan. 1996.
- [31] E. Predybaylo, G. L. Stenchikov, A. T. Wittenberg, and F. Zeng, "Impacts of a pinatubo-size volcanic eruption on ENSO," *J. Geophys. Res., Atmos.*, vol. 122, no. 2, pp. 925–947, Jan. 2017.

- [32] S. J. Camargo, A. W. Robertson, A. G. Barnston, and M. Ghil, "Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects," *Geochem., Geophys., Geosyst.*, vol. 9, no. 6, pp. 1–23, Jun. 2008, Paper no. Q06V05.
- [33] M. Wiedermann, A. Radebach, J. F. Donges, J. Kurths, and R. V. Donner, "A climate network-based index to discriminate different types of El Niño and La Niña," *Geophys. Res. Lett.*, vol. 43, pp. 7176–7185, 2016.
- [34] A. Huang, B. Vega-Westhoff, and R. L. Sriver, "Analyzing EL Niño-southern oscillation predictability using Long-Short-Term-Memory models," *Earth Space Sci.*, vol. 6, no. 2, pp. 212–221, Feb. 2019.
- [35] C. Broni-Bedaiko, F. A. Katsriku, T. Unemi, M. Atsumi, J.-D. Abdulai, N. Shinomiya, and E. Owusu, "El Niño-southern oscillation forecasting using complex networks analysis of LSTM neural networks," *Artif. Life Robot.*, vol. 24, no. 4, pp. 445–451, Dec. 2019.
- [36] D. A. Vega-Oliveros, P. S. Gomes, E. E. Milios, and L. Berton, "A multi-centrality index for graph-based keyword extraction," *Inf. Process. Manage.*, vol. 56, no. 6, Nov. 2019, Art. no. 102063.
- [37] Y. Zou, Z. Zhao, D. Yin, M. Fan, M. Small, Z. Liu, C. C. Hilgetag, and J. Kurths, "Brain anomaly networks uncover heterogeneous functional reorganization patterns after stroke," *NeuroImage, Clin.*, vol. 20, pp. 523–530, Jan. 2018.
- [38] S. Bialonski, M. Wendler, and K. Lehnertz, "Unraveling spurious properties of interaction networks with tailored random networks," *PLoS ONE*, vol. 6, no. 8, 2011, Art. no. e22826.
- [39] D. A. Vega-Oliveros, M. Cotacallapa, L. N. Ferreira, M. G. Quiles, L. Zhao, E. E. N. Macau, and M. F. Cardoso, "From spatio-temporal data to chronological networks: An application to wildfire analysis," in *Proc. 34th ACM/SIGAPP Symp. Appl. Comput.*, New York, NY, USA: ACM, 2019, pp. 675–682.
- [40] J. B. Elsner, T. H. Jagger, and E. A. Fogarty, "Visibility network of united states hurricanes," *Geophys. Res. Lett.*, vol. 36, no. 16, 2009, Art. no. L16702.
- [41] D. Ferreira, J. Ribeiro, A. Papa, and R. Menezes, "Towards evidence of long-range correlations in shallow seismic activities," *Euro Phys. Lett.*, vol. 121, no. 5, Mar. 2018, Art. no. 58003.
- [42] D. C. Zemp, M. Wiedermann, J. Kurths, A. Rammig, and J. F. Donges, "Node-weighted measures for complex networks with directed and weighted edges for studying continental moisture recycling," *Euro Phys. Lett.*, vol. 107, no. 5, Sep. 2014, Art. no. 58005.
- [43] J. Null. (2020). *El Niño and La Niña Years and Intensities Based on Oceanic Niño Index (ONI)*. Accessed: Oct. 3, 2019. [Online]. Available: <https://ggweather.com/ens/oni.htm>
- [44] B. Huang, P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, and H.-M. Zhang, "Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons," *J. Climate*, vol. 30, no. 20, pp. 8179–8205, Oct. 2017.
- [45] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford Univ. Press, 2010.
- [46] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.
- [47] E. Kalnay et al., "The NCEP/NCAR 40-year reanalysis project," *Bull. Amer. Meteorol. Soc.*, vol. 77, no. 3, pp. 437–471, Mar. 1996.
- [48] P. Holme and J. Saramäki, "Temporal networks," *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, 2012.
- [49] A. Li, S. P. Cornelius, Y.-Y. Liu, L. Wang, and A.-L. Barabási, "The fundamental advantages of temporal networks," *Science*, vol. 358, no. 6366, pp. 1042–1046, Nov. 2017.
- [50] F. J. Donges, Y. Zou, N. Marwan, and J. Kurths, "Complex networks in climate dynamics," *Eur. Phys. J. Special Topics*, vol. 174, no. 1, pp. 157–179, 2009.
- [51] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.
- [52] Y. You and J. C. Furtado, "The role of south pacific atmospheric variability in the development of different types of ENSO," *Geophys. Res. Lett.*, vol. 44, no. 14, pp. 7438–7446, Jul. 2017.
- [53] S. M. Larson, K. V. Pegion, and B. P. Kirtman, "The south pacific meridional mode as a thermally driven source of ENSO amplitude modulation and uncertainty," *J. Climate*, vol. 31, no. 13, pp. 5127–5145, Jul. 2018.
- [54] D. J. Amaya, Y. Kosaka, W. Zhou, Y. Zhang, S.-P. Xie, and A. J. Miller, "The north pacific pacemaker effect on historical ENSO and its mechanisms," *J. Climate*, vol. 32, no. 22, pp. 7643–7661, Nov. 2019.



MATHEUS A. DE CASTRO SANTOS is currently pursuing the bachelor's degree in computer engineering with the Institute of Science and Technology, Federal University of São Paulo, Brazil. His research interests are on machine learning, complex networks, and climate analysis.



DIDIER A. VEGA-OLIVEROS received the B.S. degree in computer engineering from the National University of Colombia, Bogotá, Colombia, in 2008, and the M.S. and Ph.D. degrees in computer sciences from the University of São Paulo (USP), São Carlos, Brazil, in 2011 and 2017, respectively. Since 2017, he has been a Postdoctoral Fellow with the Department of Computing and Mathematics, Faculty of Philosophy, Sciences, and Letters at Ribeirão Preto, USP, São Paulo, Brazil, and a Research Collaborator with the National Institute for Space Research (INPE), São Paulo. He is currently a Postdoctoral Visitor with the Center for Complex Networks and Systems Research (CNETS), Indiana University Bloomington, Bloomington, IN, USA. His research deals with network sciences with spatiotemporal data mining and applications, in particular, developing algorithms applied to climate systems, and social networks analysis.



LIANG ZHAO (Senior Member, IEEE) received the B.S. degree from Wuhan University, Wuhan, China, in 1988, and the M.Sc. and Ph.D. degrees from the Aeronautic Institute of Technology, São José dos Campos, Brazil, in 1996 and 1998, respectively, all in computer science. He joined, as a Faculty Member with the Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP), in 2000. From 2003 to 2004, he was a Visiting Researcher with the Department of Mathematics, Arizona State University, USA. He is currently a Full Professor with the Department of Computing and Mathematics (DCM), Faculty of Philosophy, Sciences, and Letters at Ribeirão Preto (FFCLRP), USP. His current research interests include machine learning, complex networks, artificial neural networks, and pattern recognition. He was a recipient of the Brazilian Research Productivity Fellowship. He was an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2009 to 2012. He is also an Associate Editor of the *Neural Networks*.



LILIAN BERTON received the B.S. degree in computer science and the Licenciatura degree in mathematics from Midwest University, Brazil, in 2007 and 2008, respectively, and the M.S. and Ph.D. degrees in computer science from the University of São Paulo, in 2009 and 2016, respectively. She is currently an Assistant Professor with the Institute of Science and Technology, Federal University of São Paulo, Brazil. Her research interests are on machine learning, graph-based methods, complex networks, and data mining.