RESEARCH ARTICLE | JULY 19 2024

Recurrence microstates for machine learning classification

G. S. Spezzatto ; J. V. V. Flauzino ; G. Corso ; B. R. R. Boaretto ; E. E. N. Macau ; T. L. Prado ; S. R. Lopes

Chaos 34, 073140 (2024)
https://doi.org/10.1063/5.0203801

Recurrence microstates for machine learning classification

Cite as: Chaos 34, 073140 (2024); doi: 10.1063/5.0203801
Submitted: 16 February 2024 · Accepted: 6 July 2024 ·
Published Online: 19 July 2024

G. S. Spezzatto,[1] J. V. V. Flauzino,[1] G. Corso,[2] B. R. R. Boaretto,[3] E. E. N. Macau,[3] T. L. Prado,[1,4,5,a)] and S. R. Lopes[1,5,6,b)]

AFFILIATIONS

[1]Department of Physics, Federal University of Paraná, 81531-980 Curitiba, Brazil
[2]Biophysics and Pharmacology Department, Federal University of Rio Grande do Norte, 59078-900 Natal, Rio Grande do Norte, Brazil
[3]Institute of Science and Technology, Federal University of São Paulo, 12231-280 São José dos Campos, São Paulo, Brazil
[4]Department of Physics, University Rey Juan Carlos, Móstoles, 28933 Madrid, Spain
[5]Interdisciplinary Center for Science, Technology and Innovation CICTI, Federal University of Paraná, 81531-980 Curitiba, Brazil
[6]Potsdam Institute for Climate Impact Research—Telegraphenberg, 14473 Potsdam, Germany

[a)]**Electronic mail:** thiago@fisica.ufpr.br
[b)]**Author to whom correspondence should be addressed:** lopes@fisica.ufpr.br

ABSTRACT

Recurrence microstates are obtained from the cross recurrence of two sequences of values embedded in a time series, being the generalization of the concept of recurrence of a given state in phase space. The probability of occurrence of each microstate constitutes a recurrence quantifier. The set of probabilities of all microstates are capable of detecting even small changes in the data pattern. This creates an ideal tool for generating features in machine learning algorithms. Thanks to the sensitivity of the set of probabilities of occurrence of microstates, it can be used to feed a deep neural network, namely, a microstate multi-layer perceptron (MMLP) to classify parameters of chaotic systems. Additionally, we show that with more microstates, the accuracy of the MMLP increases, showing that the increasing size and number of microstates insert new and independent information into the analysis. We also explore potential applications of the proposed method when adapted to different contexts.

Many physical systems exhibit similar dynamics when subjected to small parametric variations, and they are said to be robust to these parameter changes, which may represent a challenge to assess parameter values based on observed data. In such cases, it is crucial to employ sensitive methods to assess parameter values based solely on the system's observed data. Recurrence analysis becomes explicit many data properties, such as being chaotic, random. This analysis can be performed on linear and nonlinear systems. The information obtained by using recurrence analysis can be used to detect changes, classify, and even forecast the dynamic evolution. We propose a novel approach that integrates recurrence analysis with supervised machine learning techniques. This method involves utilizing recurrence quantifiers: microstate probabilities, entropy, and the threshold, associated with a multi-layer perceptron to evaluate parameters governing the dynamics of systems. We show the effectiveness of the method in determining parameters of chaotic maps, using the generalized Bernoulli, the so-called β-x, mod 1, logistic, and Hénon maps. The microstate multi-layer perceptron (MMLP) distinguishes minor parameter changes. Our findings indicate that, for machine learning classification, employing higher values of the recurrence plot threshold ϵ associated with the system's average maximum entropy yields superior results compared to smaller thresholds commonly recommended in the literature.

I. INTRODUCTION

The problem of optimal data acquisition involves capturing the maximum amount of information about a physical system using the smallest amount of data possible. This process is crucial for

obtaining information and developing models and, consequently, improving our ability to predict events.[1,2]

Numerous approaches are available to analyze the fundamental traits of data, encompassing aspects, such as its deterministic or stochastic nature.[2,3] The search for these characteristics is essential so that the information contained therein can be correctly processed. For example, it does not make sense to evaluate the Lyapunov exponent[4] for a stochastic time series. To study these properties, many methods have been developed. In general, the data are transformed into a smaller set of numbers, the so-called data quantifiers, making visible data properties, e.g., symbolic dynamics,[5–7] machine learning algorithms,[8] such as neural networks[9,10] and recurrence methods.[3,11–14]

Here, we use recurrence quantifiers to analyze the time series. The simplest example of recurrence is the case of periodic signals, exhibiting trivial recurrences. On the other hand, chaotic systems, despite being deterministic and stationary (considering trajectories embedded in an attractor or energy-constant-surface), lack a fully recurrent trajectory,[4] but still preserve the ability to recur in the phase space. Moreover, there are the cases of stochastic systems, governed by probabilistic events, which never can be exactly repeated.[1,2] To deal with this situation, a tolerance recurrence radius was originally inserted in the recurrence theorem by Poicaré,[15] adding an uncertainty to trajectories re-occur. Several mathematical and numerical tools have been developed based on this idea.[13] We may cite, among others, the close return plots,[13,16] recurrence time statistics,[17,18] and recurrence plots (RPs).[11,13]

RP is constructed by a very simple mathematical formulation, being a qualitative visual representation of the dynamics.[11] Typically, the RP is a visual representation of the square recurrence matrix populated with "ones" (meaning a recurrent pair of states) and "zero" (non-recurrent). If two dynamical states (i, j) (line and column indexes of the recurrence matrix) in the phase space closely resemble each other, within a specified threshold, the "one" is converted into a black dot in the RP matrix; otherwise, the matrix pixel is white. This creates many possible sequences and groups of black pixels in different levels over the RP. This representation provides a visual perspective of the underlying dynamics of the systems. To complement the qualitative view offered by the RP, a complementary technique was developed to quantify these sequences and groups of pixels, the recurrence quantification analysis (RQAs[13,19,20]).

To analyze the data, we employ a RQA[13,19] that account on the RP space, the diversity of small square structures (called recurrence microstates)[14] to quantify the complexity of the underlying system. The probabilities of each possible microstate are then used to compute an information entropy (a Shannon entropy)[21] following a maximization entropy protocol.[22,23] A direct consequence of this protocol is the automatic computation of the recurrence threshold proposed by Poincaré to characterize recurrences of finite time signals.[23,24]

Here, we integrate the probabilities of occurrence of each possible microstate and machine learning algorithms (ML). The ML concept has evolved rapidly and now plays a vital role in data analysis, offering automated solutions in a wide range of applications.[25,26] Simplest ML-methods are based on the perceptron models, which is a kind of feed-forward mathematical fitting to some known rule. The perceptron models were largely improved in more advanced multi-layer networks with supervised learning, the addition of bias, better-suited activation functions, and learning ratio changes. Other kinds of ML algorithms were proposed as the convolution neuronal network (CNN) or the recurrent neuronal network (RNN), with supervision, partially supervised, or non-supervised learning.[27] These algorithms and several others as the reservoir network or the generative transformer text (GPT) were also latter proposed for classification problems, regression, natural language reading, optimization, and image frame ratio improvement in videos.[25,28]

The intersection of machine learning and recurrence analysis has been discussed as a promising trend.[29] The ideas have initially focused on the classification and regression with the support vector machine (SVM). After that start, it has incorporated other methods, such as the random forest (RF)[30] and convolutional neural networks (CNNs).[31] Most of these methods operate in the preprocessed recurrence quantifier data,[32] but applications directly into the recurrence plot using CNN may also be found.[31] Another class of methods known as reservoir computing was proposed to use recurrence methods to do prediction of future dynamical states without the need for dynamical equations.[33]

We propose here a new way to evaluate multi-dimensional data doing a supervised classification based on the probability patterns of recurrence microstates occurring in a recurrence plot. We focus on the classification of chaotic systems parameters, which can be a challenging task for systems when these parameters are almost the same. The classification is done based only in data. This new method can open new possibilities for the intersection of both areas for classification algorithms.

We show that the use of sets of recurrence microstate probabilities may be used as inputs into a machine learning algorithm to classify similar time series. The technique can improve its accuracy simply increasing the number of analyzed microstate probabilities, suggesting that each set of probabilities of occurrence of microstates (for distinct microstate sizes N) carries new and independent information.

Figure 1 illustrates a schematic representation of the methodology. In Step 1, data must be collected. It may came from distinct sources, such as simulations of dynamical systems or experimental data. These data lead to Step 2, where the data are organized into a set of W independent datasets, each characterized by distinct labels (in this study, these labels represent different parameters of chaotic systems). Two groups of data are created: one for training the machine learning (ML) algorithm and another for testing its performance. Moving to Step 3, all data groups undergo processing to compute a set of features for each of the W datasets. These features include the probability of each microstate, as well as the maximum entropy and optimal ε. The resulting set of features, representing the information extracted from the datasets, serves as input for training and testing the multi-layer perceptron (MLP) with microstate information to classify the chaotic parameters of each dataset in Step 4.

This paper is structured as follows: In Sec. II, we delve into recurrence techniques and the maximum entropy principle. Additionally, we explore stochastic and chaotic data, along with various machine learning algorithms for data classification. Moving on to Sec. III, we showcase the results obtained from applying the machine

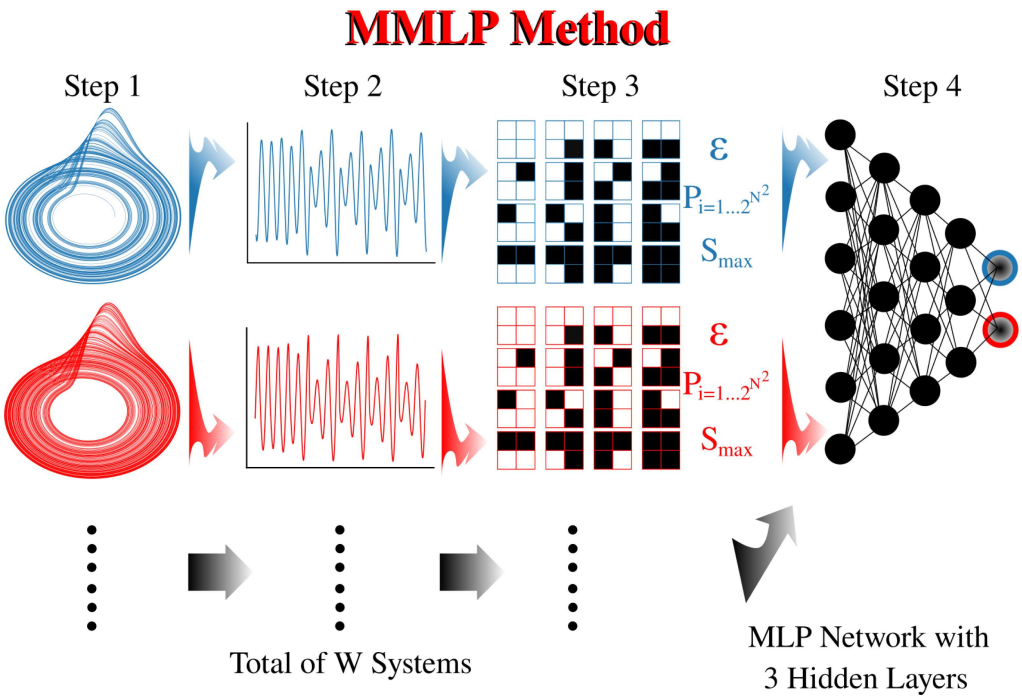

FIG. 1. Concept of the proposed MMLP (microstate multi-layer perceptron) operates through four key steps. In Step 1, data are either generated or captured from a source, which can be continuous or discrete. Step 2 involves dividing the data into two sets of W time series: one for training and the other for testing. In Step 3, both sets undergo analysis using recurrence microstates, resulting in the generation of sets of $2^{N^2}+2$ features for each time series. These features include probabilities of each recurrence microstate, recurrence entropy, and threshold. Finally, Step 4 entails the implementation of a machine learning (ML) algorithm. In the case of the MMLP discussed in this work, the ML algorithm consists of a multi-layer perceptron with an input layer, three hidden layers, and an output classification layer.

learning algorithm to recover parameters based on time series data from deterministic dynamical systems. We evaluate the method's accuracy across different lengths of data sequences. Finally, in the concluding section, we outline our key findings and discuss potential avenues for future research.

II. METHODS

A. Recurrence plots and recurrence microstates

The concept of RP was originally introduced by Eckmann *et al.*[11] in 1987. To generate the recurrence matrices, it is essential to define a recurrence threshold (ε), which specifies the maximum permissible distance between two dynamic states within the phase space to be considered recurrent to each other. Formally, each element of the recurrence matrix is expressed as $\mathbf{R}_{i,j}(\varepsilon) = \theta(\varepsilon - ||\mathbf{x}_i - \mathbf{x}_j||)$, where $||\cdot||$ represents an appropriate metric and θ denotes the Heaviside step function. For a given sequence of data points $x_1, x_2, \ldots, x_K$, the recurrence matrix elements can be simplified to

$$\mathbf{R}_{i,j}(\varepsilon) = \theta(\varepsilon - |x_i - x_j|). \tag{1}$$

Many authors have developed methods to define automatically the recurrence threshold ε. Some argue that an appropriate choice should be no more than 10% of the phase space diameter,[12,34] while others propose some ratio from the recurrence ratio quantifier[19] or some factor of the underlying data standard deviation.[35] There are even some alternative methods that evaluate directly on the distance matrix and, therefore, does not need a recurrence threshold.[36] Here, we set ε using the maximum entropy principle proposed by Jaynes[22] and applied to recurrence microstates entropy.[23]

In 1992, Zbilut and Webber, Jr.*et al.*[12] introduced a set of mathematical tools designed to quantify the dynamical patterns evident in recurrence plots (RPs). These tools, along with numerous others developed over time,[13] are collectively referred to as recurrence quantification analyses (RQAs). Generally, these quantifiers evaluate several possible patterns, such as the amount of recurrent diagonals or vertical/horizontal lines and also compute probabilities, such as the recurrence density (known as the recurrence rate).

An innovative approach for structuring square patterns from the RP, namely, *recurrence microstates*, was introduced in 2018 by Corso *et al.*[14] These microstates are binary matrices obtained from two sequences of N consecutive data points embedded within the dataset. For instance, if $N = 2$, we have the following microstate:

$$\begin{bmatrix} \mathrm{R}_{i,j} & \mathrm{R}_{i,j+1} \\ \mathrm{R}_{i+1,j} & \mathrm{R}_{i+1,j+1} \end{bmatrix}, \tag{2}$$

which has 16 (2^{N^2}) distinct binary possible microstates. Let K denote the number of data points, and considering that $K \gg N$, one may

extract randomly M microstate samples from the data sequence. The number M must be large enough so that the results are not a function of M itself. In all cases analyzed here, 10% of the total possible microstates embedded in the RP are used. This is one of the advantages of this technique. We do not need to calculate all microstates, which would demand more computational time, but only a fraction of all possible microstates. There is a more extensive discussion about this point in Refs. 23 and 37. In general, from a statistical point of view, typically 10% of all possible microstates are enough for the set of probabilities to be stationary.

The probability of a possible recurrence microstate occurring is then calculated as $P_i = n_i/M$, where n_i is the number of occurrence of the microstate i. An important RQA is the recurrence entropy. Originally, it was proposed based on the recurrence probabilities of RP diagonal structures.[13] Later, other entropy measures have been defined, including the analysis of block (rectangular) structures based on recurrence points that emerge on RP obtained from categorical time series,[38] entropy of vertical non-recurrence lines,[39,40] and also an alternative procedure to compute entropy based on the non-recurrent diagonal statistics.[41] In 2018, Corso *et al.*[14] proposed an approach leading to a new Shannon information entropy computed as

$$S = -\sum_i P_i \ln P_i. \tag{3}$$

In this approach, the probabilities P_i are computed over the occurrence of each microstate, and S is defined as a recurrence microstate entropy. Observing that S vanishes in the limits of null or maximum ε, it may be computed associated with the maximum entropy concept.[22,23] This approach automatically finds an optimal ε, namely, the one that results in a maximum for S. This choice also evokes maximum diversity of structures in the RP and will be explored in this work as an optimizer for recurrence-based machine learning algorithms.

B. Dynamical systems data

Many natural phenomena can originate in chaotic systems. Classic examples include meteorological systems,[42,43] heartbeat mechanisms,[44] neuronal electrical behavior,[45,46] among others. Therefore, a reliable method for detecting system parameters or parameter changes (nonstationary behavior) based only on available data is essential. Let us consider three examples of paradigmatic chaotic dynamical systems,[4] the generalized version of the Bernoulli map, called here βx mod(1) map, the logistic map, and the two-dimensional map, known as the Hénon map. We show that MMLPs are capable of distinguishing parameters based solely on data generated by these systems.

For each map, 128 time series of 1000 data points are generated. The time series have distinct values of the control parameter, which are uniformly distributed in the variation range specific for each map. Half of the data is used for training, and half is used to test the ability of the MMLP to distinguish the parameter used in the time series, a 64 + 64 scheme.

The βx mod (1) map is defined as

$$x_{n+1} = \beta x_n \mod(1), \tag{4}$$

where $n = 1, 2, 3, \ldots, K-1$. The sequence of values $0 \leq x_n \leq 1$ generates a chaotic discrete trajectory for $\beta > 1$.[4] Here, we generate trajectories for a uniformly distributed values of β, $1.99 \leq \beta \leq 4.99$.

The second dynamical system is a uni-dimensional system originally obtained to study population dynamics. The logistic map is given by[4]

$$x_{n+1} = \rho x_n(1 - x_n), \tag{5}$$

where the control parameter ρ distinguishes different dynamical behaviors. The logistic system may display a cascade of bifurcations that ultimately lead to a chaotic regime. Here, we take parameters uniformly distributed inside the interval $3.95 \leq \rho \leq 4.00$ using only chaotic trajectories (we test and discard eventual regular trajectories).

As a third example, we consider the bi-dimensional Hénon map defined as[4]

$$x_{n+1} = 1 - ax_n^2 + y_n, \tag{6}$$

$$y_{n+1} = bx_n. \tag{7}$$

The map has two parameters (a and b). As traditionally used,[4] we take the system with a fixed $b = 0.3$ parameter, and the parameter a is varied uniformly in the interval $1.1 \leq a \leq 1.2$, being a responsible for changes in the dynamics, including values that exhibits chaotic properties. We also exclude initial conditions that lead to a divergent trajectory for the Hénon map.

Figure 2 shows details about the transformation of the data performed by the method. It illustrates the transformation of time series generated from the βx map, as described in Eq. (4), with varying β values, into a reduced recurrence space. This space comprises 16 coordinates, determined by the probabilities of a microstate occurrence with a size of $N = 2$, alongside the recurrence maximum entropy and the threshold values. Note that each time series results in a horizontal set of 18 values (features) in Fig. 2 [panels (a)–(c)]. In this new space, the set of values has subtly different coordinates, distinct enough for the MMLP to be able to correctly quantify the β value used to generate e data. To better observe the distinctness among elements of the sets we have magnified in Fig. 2(d), six values of β grouped in three sets of two near values. Note that very different β-values lead to visually different fingerprints. However, for close values of the parameter, classification starts to become more difficult. In these cases, two paths can be followed: The first is to increase the size of the time series, which naturally increases the ability to distinguish. There may be cases where this increase in the size of the data may not be possible, in which case an increase in the size of the microstate is necessary.

Figure 3 shows the same case presented in Fig. 2(d), but using microstates of size $N = 3$. Time series of $K = 1000$ data points are used. In this case, we have 512 features plus the recurrence entropy and the recurrence threshold. Note that the larger the size of the microstates, the better the visual ability to distinguish between states. Following this procedure, we can now generalize our results by classifying these visual patterns or, in other words, these fingerprints of each parameter value using a machine learning procedure.

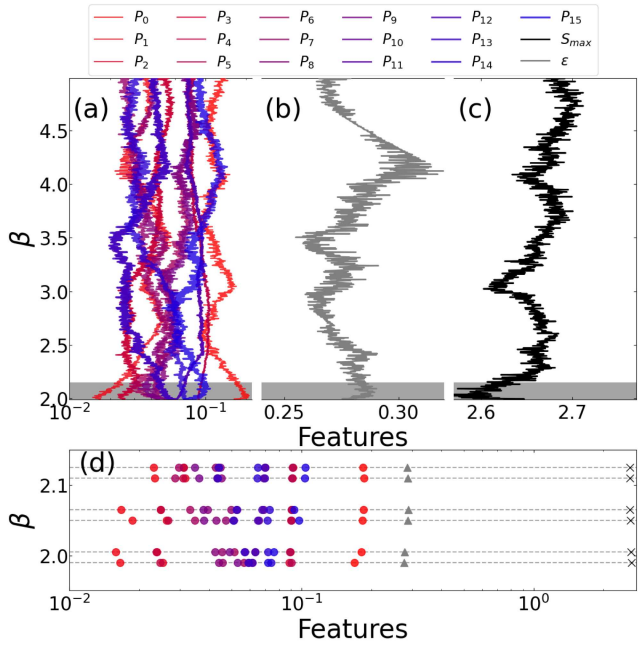

FIG. 2. All 18 ($N = 2$) features are composed of 16 recurrence microstate probabilities: panel (a), maximum recurrence entropy; panel (b) and the recurrence threshold, and panel (c) plotted as a function of β values. The 18 features configure the discrete space where each β-value time series of the βx map is represented. Time series size is $K = 1000$. Observe that each horizontal line in panels (a)–(c) depicts a distinct set of values (features). To better observe this distinctness, panel (d) displays that magnifications of the gray are of panels (a), (b), and (c) for six values of the β parameter. The sampled values are grouped in three sets of two near values of β. As may be observed, even small differences in the β values are translated into very distinct sets of features, turning them in ideal distinction quantifiers.

III. SYSTEM CLASSIFICATION

A. The βx map classification with MMLP

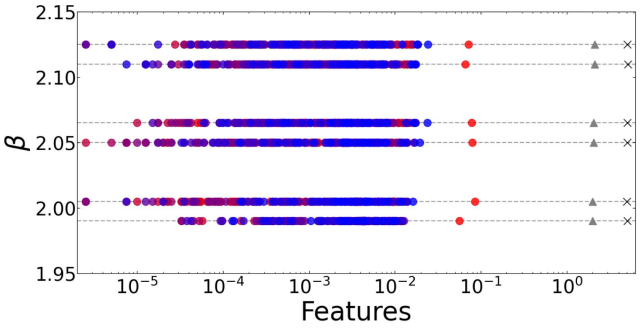

FIG. 3. All 512 ($N = 3$) recurrence microstate probabilities (features) plus recurrence entropy (up-triangle) and recurrence threshold ("x"). Again configuring the discrete space where each β-value time series of the βx map is represented. Time series size of $K = 1000$ is used. Observe that each line depicts a distinct set of values (features) as detailed in the caption of Fig. 2.

Figure 4 shows our results applied to the distinction of β values based on the dynamics of βx map, Eq. (4). Three panels for different microstate sizes N are shown: (a) $N = 2$, (b) $N = 3$, and (c) $N = 4$. Along the x axis, it is varied the time series size $500 \leq K \leq 4000$, and along the y axis, it is plotted the number of possible β values (labels) in the interval $1.99 \leq \beta \leq 4.99$, from 16, until 512, the largest resolution we have simulated. Considering W as the total number of time series analyzed, it is important to mention that the ratio $W/(\text{Number of Labels})$ is maintained constant at 64. This means that 64 training series and 64 test series will always be used for each label. In color coded, it is plotted the accuracy of the MMLP algorithm going from zero (the algorithm does not correctly predict the β parameter for none of the labels) to one (the algorithm predicts exactly the β parameter for all labels). Here, accuracy is defined as the ratio of correctly predicted parameters to the total number of parameters across all W time series.

The first important result we observed in Fig. 4 is that in general, the larger the time series size K or microstate sizes N, the better the accuracy. The number of possible β values (labels) increases the similarity from one data set (64 time series) to another once the differences from one β value to another become smaller. In these cases, a larger number of labels reduces the accuracy. It is noteworthy to consider that more important than the number of labels is the difference imposed by these labels into distinct time series. In the βx map example, the beta value defines the level of chaos of the map, measured by the Lyapunov exponent, $\lambda = \ln \beta$. The difference from one β_i value to the following β_{i+1} for the case of 512 labels for the βx map is 0.006, then two consecutive values of β as supposed, shows us that the method is capable of distinguishing very small differences in the degree of chaoticity of the system. We note in (c) that even for 512 Labels, the accuracy is close to 0.7 for $K = 4000$ time series sizes. We also note that the results considerably improve when we compare $N = 3$ and $N = 4$ [see the substantial increase in the red-orange areas in panel (c)], showing that increasing the size of the microstate is an important tool in detecting small changes in the behavior of time series.

B. The logistic and Hénon maps classification with MMLP

In order to show the robustness of the findings depicted in Fig. 4 across different dynamical systems, Fig. 5 presents similar results also for the logistic and Hénon maps. In both cases, the recurrence thresholds are computed using the maximum entropy criterion while varying the number of labels from 16 to 512. Two time series sizes are considered: 500 data points in dashed lines and 4000 data points in solid lines. Furthermore, two microstate sizes are used: $N = 2$ blue lines and $N = 3$ green lines. In general, larger microstates lead to better results, while increasing the number of possible labels makes it difficult to correctly predict their values. We also observe that the larger the time series, the better the results. Again, it can be seen that increasing the size of the recurrence microstate is much more effective in obtaining better accuracy than increasing the size of the time series. Note that an increase of almost eight times in the size of the time series is necessary to obtain accuracy similar to those obtained simply by increasing the size of the microstate from $N = 2$ to $N = 3$ (comparing a green dashed line

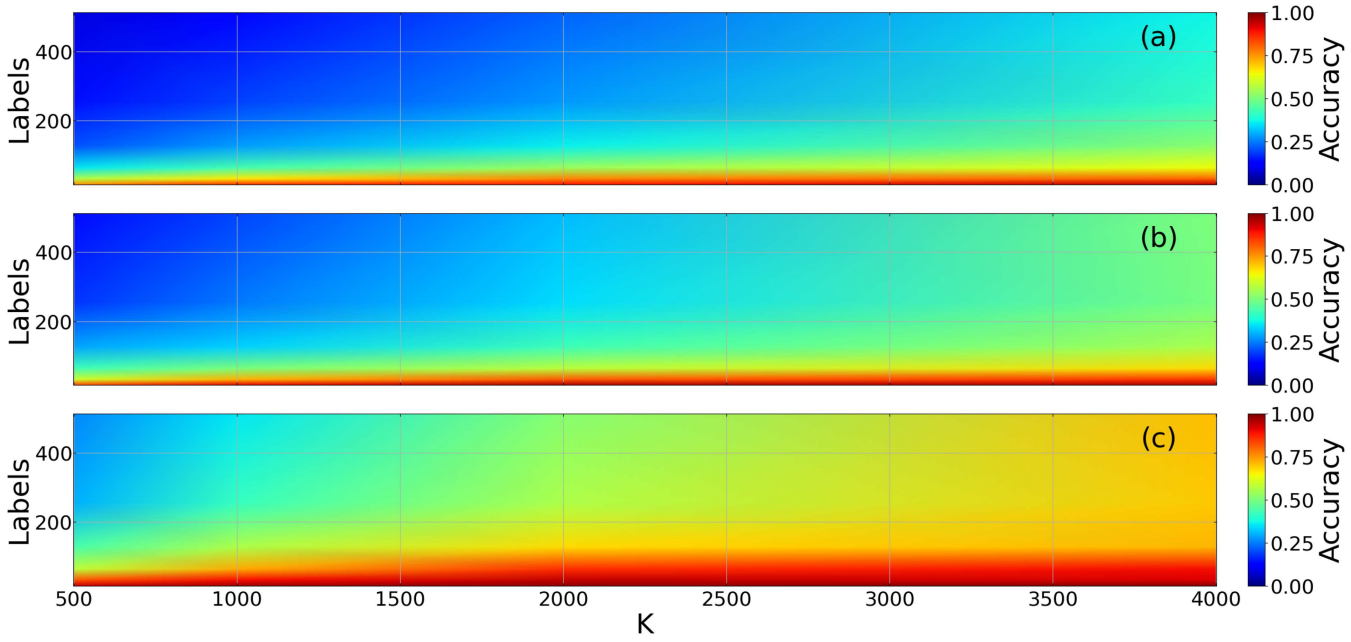

FIG. 4. Color coded accuracy as a function of the time series size and the number of possible β parameter of the βx map in the parameter interval $1.99 < \beta < 4.99$. Microstate sizes of (a) $N = 2$, (b) $N = 3$, and (c) $N = 4$ are used. Along the x axis, it is varied the time series size, and along the y axis, it is plotted the number of possible β values (Labels) in the interval $1.99 \leq \beta \leq 4.99$, from 16, until 512. Labels characterize uniquely each β in this interval. The larger time series size K or larger microstate sizes N, the better the accuracy, but increasing the size of microstates is more effective in improving accuracy.

with a blue solid line). Also, note the small dispersion of the values observed most of the time. In fact, many bars and scatter are not visible, suggesting that even for a small number of time series, the results are reliable.

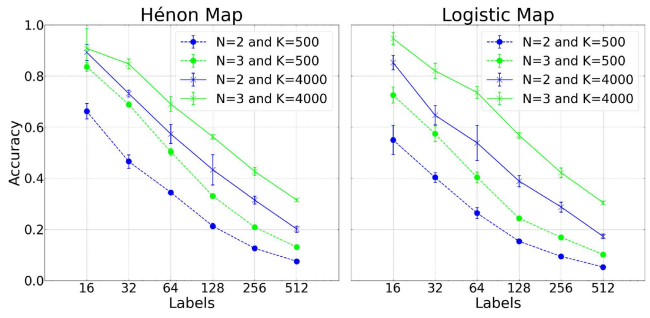

FIG. 5. Analysis of the MMLP accuracy in the supervised learning of a chaotic dynamical system for optimized ε. We consider 16 until 512 equally spaced labels. For the logistic map $3.95 \leq \rho \leq 4.00$, for the Hénon map, $1.1 \leq a \leq 1.2$. Considering time series sizes of $K = 500$, 1000, and 4000 data points. In general, accuracy decreases with the increase in the number of possible labels, regardless of the map treated. Scatter bars are due to five simulations of each condition.

C. Confusion matrices analyses

To complete these analyses, we show in Fig. 6 confusion matrices for two cases. A first case with 16 different parameter values (labels) is used, panels (a)–(c), and an extreme case with 512 values

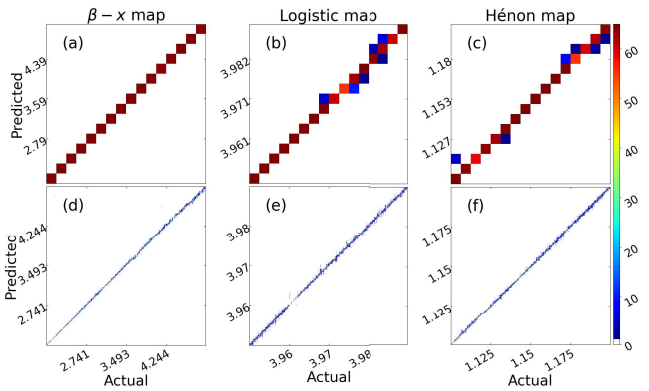

FIG. 6. Confusion matrices for βx map [panels (a) and (d)], the logistic map [panels (b) and (e)], and the Hénon map [panels (c) and (f)]. In panels (a)–(c), 16 possible labels are considered in the respective parameter intervals. Panels (d)–(f) result from 512 possible labels. In all cases, the simulations were conducted with $K = 4000$, $N = 3$, and 64 different initial conditions are simulated for each label. Observe that the predictions are mostly correct, exactly or very near the correct values in all panels. Predictions very far from the simulated real parameter are rare.

is used, panels (d)–(f). In both cases, training and testing are performed with $64 + 64$ time series of size $K = 4000$ for each parameter value. The best results clearly occur for the βx map and 16 labels; see Fig. 6(a). Moreover, in Fig. 6(d), 512 labels are used, but the MMLP prediction spreads slightly from the main diagonal. In this case, the MMLP may make small prediction mistakes, but the predicted values are close to the actual ones. Similar results are observed for the logistic and Hénon maps; see Fig. 6 panels (b) and (e) for the logistic map and for (c) and (f) for the Hénon map. In both cases, the ranges assumed for a parameter variation are much smaller when

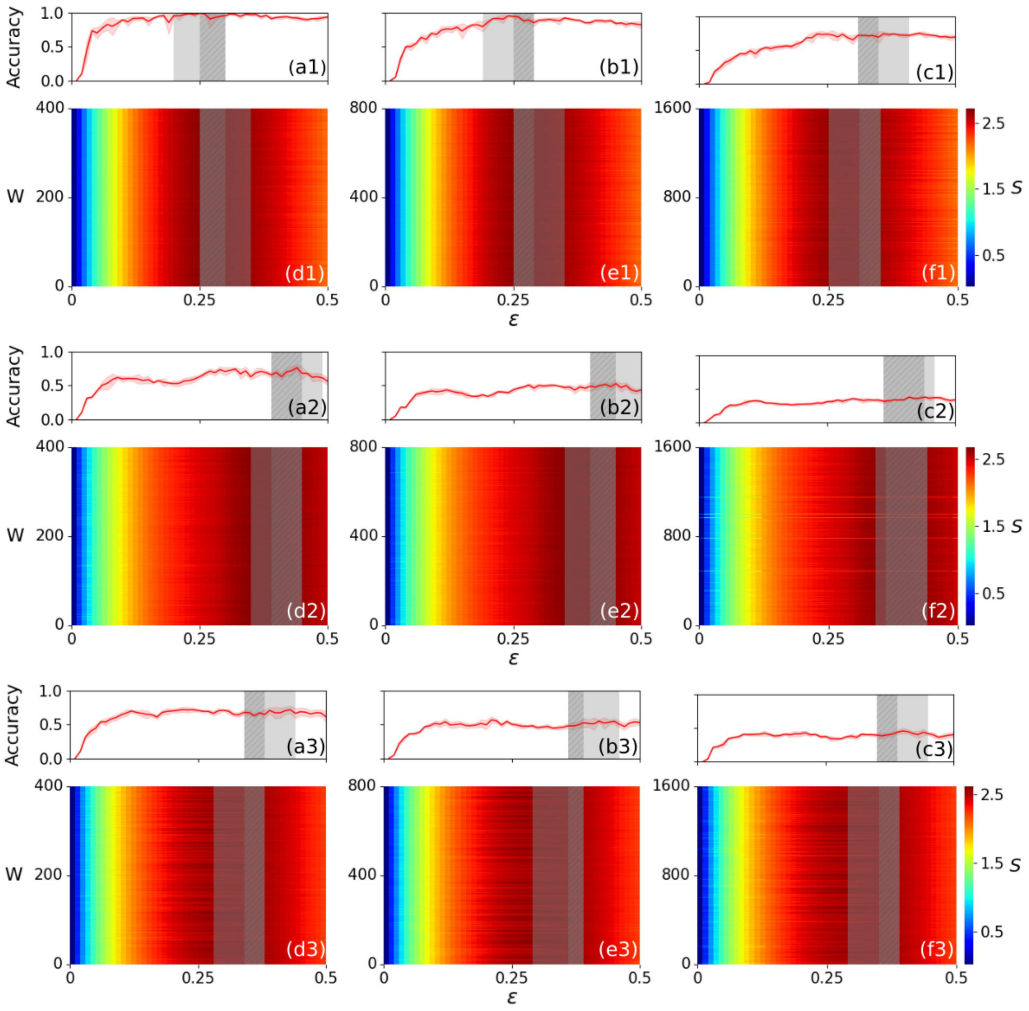

FIG. 7. Analysis of MMLP algorithm accuracy in the supervised learning of three chaotic dynamical system parameters as a function of the number of simulated equally spaced values. Considering the βx map, Eq. (4), in panel (a1), 10 values of parameter β are equally distributed in the interval $1.99 \leq \beta \leq 4.99$, in panel (b1), 20 values, and in panel (c1), 40 values. Considering the logistic maps, Eq. (5), in panel (a2), 10 values of the parameter ρ are equally distributed in the interval $3.95 \leq \rho \leq 3.99$, in panel (b2), 20 values, and in panel (c2), 40 values. Finally, for the Hénon map, Eqs. (6) and 7 in panel (a3), 10 values of parameter a are equally distributed in the interval $1.1 \leq a \leq 1.2$, in panel (b3), 20 values, and in panel (c3), 40 values. For all situations, a large interval of $0 < \varepsilon < 0.5$ values is considered. Each time series has $K = 1000$ data points. For all resolutions in the parameter spaces, despite the decrease of the accuracy, the best results obtained by MMLP occur for values of the recurrence threshold ε associated with the computation of maximum entropy S_{max}, as may be observed in the color panels (d1), (e1), and (f1) for the βx map, (d2), (e2), and (f2) for the logistic map, and (d3), (e3), and (f3) for the Hénon map. To compute the color maps showing that the best accuracy occurs near the maximum values of the recurrence entropy, we have computed 40 simulations of the maps for each value of the parameters, resulting in 400 time series for (d1), (d2), and (d3) until 1600 simulations for (f1), (f2), and (f3). All values for the maximum computed entropy fit the shadow areas in color panels (d1)–(d3), (e1)–(e3), and (f1)–(f3). Dashed areas are inside the shaded areas in all cases. On the other hand, the shadow areas in panels (a1), (b1), and (c1) give us the minimum and maximum recurrence thresholds where the averages of the maximum accuracy are obtained. The dashed shadow areas are the simultaneous overlap between the intervals where the maximum recurrence entropy and the maximum accuracy occur.

compared to the βx map, considerably increasing the difficulty in predicting exactly the parameters used. However, it is clear that the predicted values are always in the vicinity of the correct values.

IV. THE ROLE OF THE RECURRENCE THRESHOLD VALUE IN THE ACCURACY OF THE MMLP

So far, our results were presented for a specific choice of recurrence threshold parameter ε, namely, the one we get when we maximize the recurrence entropy of the time series. To show that this choice is, in fact, the one that brings the best results, Fig. 7 shows results for the accuracy of the MMLP applied to three chaotic dynamical system parameters, the parameter interval of $1.99 \leq \beta \leq 4.99$ for the βx map, Eq. (4), a narrow interval of the parameter $3.95 \leq \rho \leq 3.99$ for the logistic maps, Eq. (5), and the interval of parameter $1.1 \leq a \leq 1.2$ for the Hénon Map, Eqs. (6)–(7). Once the Hénon map is bi-dimensional, we employ the analyses using the Euclidean norm of the phase space. All panels (a) depict the mean values of the accuracy due to 10 trials, considering $W = 400$ time series and 10 equally spaced but randomly selected possible values of the parameters β, (a1), ρ, (a2), and a (a3), respectively. Shaded pale colors mean the dispersion of values. Panels (b) and (c) depict similar results but for $W = 800$ and $W = 1600$ time series and considering 20 and 40 randomly selected possible equally spaced values of β, ρ, and a, respectively. In these cases, the number of systems evaluated increases to maintain the ratio of the number of systems per label (here, it is 40 time series each one with $K = 1000$ data points for the training and another equal size set for the test).

It is clear that the maximum accuracy is obtained in a range of recurrence threshold values, shadow areas in panels (a1), (b1), and (c1). On the other hand, the maximum entropy is system dependent, each map simulation, even for a fixed parameter value will have small deviations in its maximum entropy. The only possible way to have no deviation in the computed values is the case of very large (infinity) time series. All values for the maximum computed entropy fit the shadow areas in Fig. 7 [color panels (d)–(f)]. The dashed shadow areas are the simultaneous overlap between the intervals where the maximum recurrence entropy and the maximum accuracy occur.

A second point is the small threshold dispersion, showing a robust and precise accuracy of the MMLP as may be observed in Fig. 7. Again, the larger the number of possible values of the parameter, the smaller the accuracy, reflecting the difficulty to distinguish close values of the parameter. To show that the better accuracy interval is associated with those for which the microstate entropy is maximum, panels (d)–(f) show the microstate entropy values as a function of the recurrence threshold for the same situation explored in panels (a)–(c) for all three maps. As noted, the ranges of recurrence threshold values for which the microstate entropy is maximum are clearly related to the accuracy maxima. For the Hénon map parameters, the accuracy of the method also decreases. It is observed that accuracy is also dependent on the dynamic system evaluated. In general, the accuracy increases with longer time series. We have 100% (little less for the Hénon map) accuracy for 10 labels when the ε is correctly chosen, but quite low accuracy for wrong ε choice, which reinforces the importance of an adequate choice of the recurrence threshold. Similar behavior occurs for a larger number of labels, but the accuracy is decreased since the difference from two successive ρ or a parameter values can be as small as 0.01.

V. DISCUSSIONS AND CONCLUSIONS

In this work, we presented a novel recurrence-based machine learning protocol, called a microstate multi-layer perceptron (MMLP). The idea is based on the transformation from the collected data space into the recurrence space. In this case, we define a new space composed of all possible probabilities of occurrence of recurrence microstates plus recurrence entropy and threshold. Each component of this new space works as features of the machine learning protocol. The new space shows that it is possible to distinguish time series of subtly different parameters, proving to be an excellent classifier of parameters of dynamic systems. We present results for three well known dynamical systems: βx, logistic, and Hénon maps. It is seen that in this high order dimensional space, each time series has an almost uniquely pattern.

We have shown that in the supervised learning, the size of the analyzed time series positively impacts the accuracy. Another approach to achieve higher accuracy can be obtained by increasing the size of the recurrence microstates. In this case, it is a remarkable result that the algorithm presents a higher accuracy without the need for new information by just changing the only truly free parameter of the methodology (the microstate size). In general, the larger the microstate size, the better the information the analysis brings about the system. It is also true that all the information obtained by a microstate size is contained in larger ones; that is, part of the information added with the increase in the size of the microstates is not new. In this sense, we may say that independent (new) information is added using larger microstate sizes. It is also important to mention that even for quite a large number of different parameters of a chaotic system, the algorithm still keeps an acceptable accuracy even for a small number of features (the size of the recurrence space).

Our results present a clear new path in machine learning recurrence analysis. The reduction of data to a more concise set of features opens doors to new scientific problems in big data. Furthermore, achieving high accuracy with small datasets encourages the advancement of recurrence tools, aiming to enhance the uniqueness within the recurrence space.

Finally, we should mention that there is no consensus in the literature about an optimal recurrence threshold ϵ. Here, we show that the MMLP has better accuracy around the maximum recurrence microstate entropy which in this case is an optimal choice. It is noteworthy that the recurrence classification algorithms need a higher threshold than generally found in the literature.[13] In many situations, other parameters specific to recurrence analysis, such as the size of the minimum diagonal, can still coexist with large thresholds as long as they are treated appropriately,[24,47] making the choice of parameters an automated task.

ACKNOWLEDGMENTS

This work was supported by the following Brazilian research agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Grant Nos. 308 441/2021-4, 307 907/2019-8,

407 072/2022-5, 305 189/2022-0, 408 254/2022-0, and 300 064/2023-3), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Grant Nos. 88 887.898 929/2023-00, 88 887.898 924/2023-00, and 88 881).895 032/2023-01, and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Proc. 2018/03211-6 and 2021/09839-0).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

G. S. Spezzatto: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Visualization (equal). **J. V. V. Flauzino:** Data curation (equal); Formal analysis (equal); Methodology (equal); Visualization (equal). **G. Corso:** Formal analysis (equal); Investigation (equal); Writing – review & editing (equal). **B. R. R. Boaretto:** Investigation (equal); Methodology (equal). **E. E. N. Macau:** Funding acquisition (equal); Methodology (equal); Writing – review & editing (supporting). **T. L. Prado:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Visualization (equal); Writing – review & editing (equal). **S. R. Lopes:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

All the details of the code to compute recurrence microstate entropy and optimum recurrence threshold is available in Ref. 48. About the machine learning algorithm, the method uses 128 neurons per layer, with reduction of the total number of neurons of 10% in each one of the three hidden layers and the activation function is the SELU (Scaled Exponential Linear Unit). The output layer uses a softmax function. The optimizer is the Adam method with sparse categorical cross-entropy as a loss function. This part of the code uses the TensorFlow and Keras packages in the Python language.

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1]W. Van Der Aalst and W. van der Aalst, *Data Science in Action* (Springer, 2016).
[2]J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (John Wiley & Sons, 2011).
[3]T. L. Prado, B. R. R. Boaretto, G. Corso, G. Z. dos Santos Lima, J. Kurths, and S. R. Lopes, "A direct method to detect deterministic and stochastic properties of data," New J. Phys. **24**, 033027 (2022).
[4]E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, 2002).
[5]R. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes—An application of symbolic dynamics to information theory," IEEE Trans. Inf. Theory **29**, 5–22 (1983).
[6]C. Masoller, Y. Hong, S. Ayad, F. Gustave, S. Barland, A. J. Pons, S. Gómez, and A. Arenas, "Quantifying sudden changes in dynamical systems using symbolic networks," New J. Phys. **17**, 023068 (2015).
[7]X. Huang, H. L. Shang, and D. Pitt, "Permutation entropy and its variants for measuring temporal dependence," Aust. N.Z. J. Stat. **64**, 442–477 (2022).
[8]S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," in *Emerging Artificial Intelligence Applications in Computer Engineering*, edited by I. Maglogiannis, K. Karpouzis, M. Wallace, and J. Soldatos (IOS Press, Amsterdam, 2007), pp. 3–24.
[9]D. E. Rumelhart and J. L. McClelland, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations* (MIT press, 1987), pp. 318–362.
[10]G. P. Zhang, "Neural networks for classification: A survey," IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. **30**, 451–462 (2000).
[11]J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," Europhys. Lett. **4**, 973–977 (1987).
[12]J. P. Zbilut and C. L. Webber, Jr., "Embeddings and delays as derived from quantification of recurrence plots," Phys. Lett. A **171**, 199–203 (1992).
[13]N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," Phys. Rep. **438**, 237–329 (2007).
[14]G. Corso, T. L. Prado, G. Z. d. S. Lima, J. Kurths, and S. R. Lopes, "Quantifying entropy using recurrence matrix microstates," Chaos **28**, 083108 (2018).
[15]H. Poincaré, "Sur le problème des trois corps et les équations de la dynamique," Acta Math. **13**, 1–270 (1890).
[16]G. M. Mindlin and R. Gilmore, "Topological analysis and synthesis of chaotic time series," Phys. D **58**, 229–242 (1992).
[17]R. M. da Silva, M. W. Beims, and C. Manchein, "Recurrence-time statistics in non-Hamiltonian volume-preserving maps and flows," Phys. Rev. E **92**, 022921 (2015).
[18]C. V. Abud and R. E. De Carvalho, "Multifractality, stickiness, and recurrence-time statistics," Phys. Rev. E **88**, 042922 (2013).
[19]J. P. Zbilut, J.-M. Zaldivar-Comenges, and F. Strozzi, "Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data," Phys. Lett. A **297**, 173–181 (2002).
[20]D. G. d. B. Costa, B. M. d. F. Reis, Y. Zou, M. G. Quiles, and E. E. N. Macau, "Recurrence density enhanced complex networks for nonlinear time series analysis," Int. J. Bifurc. Chaos **28**, 1850008 (2018).
[21]C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); ce shannon, Bell System Tech. J **27**, 218 (1948)
[22]E. T. Jaynes, "Information theory and statistical mechanics," Phys. Rev. **106**, 620 (1957).
[23]T. L. Prado, G. Corso, G. Z. dos Santos Lima, R. C. Budzinski, B. R. R. Boaretto, F. A. S. Ferrari, E. E. N. Macau, and S. R. Lopes, "Maximum entropy principle in recurrence plot analysis on stochastic and chaotic systems," Chaos **30**, 043123 (2020).
[24]F. E. L. da Cruz, S. R. Lopes, and T. de Lima Prado, "How to compute the minimum diagonal length of recurrence quantifiers to optimize their sensitivity to deterministic and stochastic properties," Chaos, Solitons Fractals **173**, 113747 (2023).
[25]F. Chollet, *Deep Learning with Python* (Manning Publications Co., 2017).
[26]M. Kubat, *An Introduction to Machine Learning* (Springer, 2017), Vol. 2.
[27]P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to machine learning for physicists," Phys. Rep. **810**, 1–124 (2019).
[28]A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), Vol. 30.
[29]N. Marwan and K. H. Kraemer, "Trends in recurrence analysis of dynamical systems," Eur. Phys. J. Spec. Top. **232**, 5–27 (2023).
[30]S. V. Begum and M. P. Rani, "Classification of gait dynamics in neurodegenerative disease patients using machine learning techniques," Inter. J. Sci. Tech. Res. **9**, 6250–6254 (2020).
[31]Y.-X. Yang, Z.-K. Gao, X.-M. Wang, Y.-L. Li, J.-W. Han, N. Marwan, and J. Kurths, "A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG," Chaos **28**, 085724 (2018).
[32]I. Ghosh, R. K. Jana, and M. K. Sanyal, "Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear

dynamics, econometric models and machine learning algorithms," Appl. Soft Comput. **82**, 105553 (2019).
[33]T. Weng, H. Yang, J. Zhang, and M. Small, "Modeling chaotic systems: Dynamical equations vs machine learning approach," Commun. Nonlinear Sci. Numer. Simul. **114**, 106452 (2022).
[34]M. Koebbe and G. Mayer-Kress, "Use of recurrence plots in the analysis of time-series data," in *Santa Fé Institute Studies in the Science of Complexity—Proceedings*, Vol. 12 (Citeseer, 1992), pp. 361–361.
[35]M. Thiel, M. C. Romano, J. Kurths, R. Meucci, E. Allaria, and F. T. Arecchi, "Influence of observational noise on the recurrence quantification analysis," Phys. D **171**, 138–152 (2002).
[36]M. Hu, X. Feng, S. Zhou, and W. Xu, "Anomaly detection in time series data based on unthresholded recurrence plots," in *International Conference on Applications and Techniques in Cyber Security and Intelligence ATCI 2018: Applications and Techniques in Cyber Security and Intelligence* (Springer, 2019), pp. 477–484.
[37]L. B. Froguel, T. de Lima Prado, G. Corso, G. Z. dos Santos Lima, and S. R. Lopes, "Efficient computation of recurrence quantification analysis via microstates," Appl. Math. Comput. **428**, 127175 (2022).
[38]G. Leonardi, "A method for the computation of entropy in the recurrence quantification analysis of categorical time series," Phys. A **512**, 824–836 (2018).
[39]E. J. Ngamga, A. Nandi, R. Ramaswamy, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence analysis of strange nonchaotic dynamics," Phys. Rev. E **75**, 036222 (2007).
[40]E. J. Ngamga, D. V. Senthilkumar, and J. Kurths, "Dynamics between order and chaos revisited," Eur. Phys. J. Spec. Top. **191**, 15–27 (2010).
[41]C. Letellier, "Estimating the Shannon entropy: Recurrence plots versus symbolic dynamics," Phys. Rev. Lett. **96**, 254102 (2006).
[42]E. Lorenz, "The predictability of hydrodynamic flow," Trans. New York Acad. Sci. **25**, 409–432 (1963b).
[43]E. Lorenz, *The Essence of Chaos* (University of Washington Press, 1993), p. 227.
[44]A. L. Goldberger, "Is the normal heartbeat chaotic or homeostatic?," Physiology **6**, 87–91 (1991).
[45]V. Makarenko and R. Llinás, "Experimentally determined chaotic phase synchronization in a neuronal system," Proc. Natl. Acad. Sci. U.S.A. **95**, 15747–15752 (1998).
[46]M. Nolte, M. W. Reimann, J. G. King, H. Markram, and E. B. Muller, "Cortical reliability amid noise and chaos," Nat. Commun. **10**, 3792 (2019).
[47]F. E. L. da Cruz, J. V. V. Flauzino, S. R. Lopes, and T. de Lima Prado, "Analytical results in calculating the entropy of recurrence microstates," Phys. A **638**, 129629 (2024).
[48]https://github.com/AlgosL/maxEntropy/blob/main/Entropy.jl.