

# How neural networks learn to classify chaotic time series

Alessandro Corbetta<sup>†</sup>

Department of Applied Physics,

Eindhoven University of Technology, 5612AK Eindhoven, The Netherlands.

a.corbetta@tue.nl

Thomas Geert de Jong<sup>†</sup>

Faculty of Mathematics and Physics, Institute of Science and Engineering,

Kanazawa University, 920-1192 Kanazawa, Japan.

tgdejong@se.kanazawa-u.ac.jp

**Abstract.** We tackle the outstanding issue of analyzing the inner workings of neural networks trained to classify regular-versus-chaotic time series. This setting, well-studied in dynamical systems, enables thorough formal analyses. We focus specifically on a family of networks dubbed Large Kernel Convolutional Neural Networks (LKCNN), recently introduced by Boullé et al. (2021). These non-recursive networks have been shown to outperform other established architectures (e.g. residual networks, shallow neural networks and fully convolutional networks) at this classification task. Furthermore, they outperform “manual” classification approaches based on direct reconstruction of the Lyapunov exponent. We find that LKCNNs use qualitative properties of the input sequence. We show that LKCNN models trained from random weight initialization, end in two most common performance groups: one with relatively low performance (0.72 average classification accuracy) and one with high classification performance (0.94 average classification accuracy). Notably, the models in the low performance class display periodic activations that are qualitatively similar to those exhibited by LKCNN with random weights. This could give very general criteria for identifying, *a priori*, trained weights that yield poor accuracy.

**Keywords:** Dynamical systems, chaos, deep learning, convolutional networks, time series, classification, Savitsky-Golay.

Neural networks are increasingly employed to model, analyze and control non-linear dynamical systems ranging from physics to biology. Owing to their universal approximation capabilities, they regularly outperform state-of-the-art model-driven methods in terms of accuracy, computational speed, and/or control capabilities. On the other hand, neural networks are very often taken as black boxes whose explainability is challenged, among others, by huge amounts of trainable parameters. In this paper, we analyze how neural networks successfully manage the longstanding challenge of classifying signals in chaotic or regular. We consider a neural network with an architecture which lends itself well for analysis, Large Kernel Convolutional Neural Networks (LKCNN). We have shown that to classify signals with high accuracy, LKCNNs use qualitative properties of the input

---

<sup>†</sup>Joint first author

sequence. This enables them to outperform classical methods. LKCNNs higher classification accuracy strongly emerges as we consider regular signals which are almost chaotic. We also investigated the emerging connection between input periodicity and periodicity within the network layers. We have shown this aspect to be paramount for performance. This could give new baseline requirements during neural network training.

## 1 Introduction

During the last decade there has been a strong acceleration in the adoption of machine learning, typically through artificial neural networks, to model, analyze, and control a broad spectrum of dynamical systems [BK22]. Neural networks are nonlinear parametric models satisfying universal approximation principles [KHW89]. Usually, they are trained in data-driven contexts to fit pre-annotated data. Beside boosting well-known super-human performance in contexts as automated vision [WBGL15], natural-language processing [SCGH05], and long-term strategy games [SHM<sup>+</sup>16], they are gaining substantial momentum in fundamental research in connection with non-linear/chaotic dynamics. In these contexts they are regularly surpassing “traditional” state-of-the-art model-driven approaches [TKL<sup>+</sup>20].

Astounding are the achievements connected to three-body problem long-term forecasts [BFBZ20], weather modeling [KMA<sup>+</sup>21], fluid turbulence closures [DIX19, OCRT22], measurements [CMBT21], and control [RJK<sup>+</sup>19]. Machine learning methodologies allowed progress in scientific context-agnostic issues, e.g., model-free chaotic dynamics predictions [PHG<sup>+</sup>18], as well as the problem dealt with in this paper: the classification of time series between chaotic and regular. Dynamics that can bifurcate from regular to chaotic and vice versa are present in every scientific discipline: astronomy, biology, meteorology, physics etc. [PJSF04, Str18, Cel10, Lor63, KKK98, GH13, BT11]. The issue of their classification is thus longstanding (e.g [LY03, KPDV92, PRK92]), and has been recently addressed via new data-driven machine learning methods [CGRFV22, HMS21, WLTMK22, AUK22, BLMC<sup>+</sup>23, ZGS<sup>+</sup>23, WAB10, PLDJ21, MOG97]. Among these, the recent works by Bouillé et al. [BDNS20] showed that convolutional neural networks, specifically with large kernels, can be strikingly successful at the task. In general, convolutional networks [GBC16] have proven suitable for time series classification [WYO17] and inference [CMBT21]. Building on this approach, Bouillé et al. proposed to consider convolutional neural networks with large kernels; this means that the kernel size of the convolutional layers is large in relation to the input sequence length. Bouillé et al. have shown that LKCNNs boast highest performance in comparison with other established machine learning approaches such as residual, fully convolutional, multilayer perceptrons and shallow networks [Wan19, WYO17]. Besides, they manage generalization properties to data outside the training set. It is worth

mentioning that, in general, neural networks are (ubiquitously) employed as data-driven black boxes. Especially as the network size grows, straightforward insights into their internal mechanics becomes impossible [BLH21]. However, identifying the features that that network leverage on and that are key for performance, could become a crucial component towards new discoveries.

This creates the case for the analysis of this paper. We consider how Large Kernel Neural Networks successfully manage the longstanding challenge of classifying discrete-time series produced by dynamical systems, identifying regular vs. chaotic motions. We provide a two-fold contribution: experiments that show that LKCNNs can identify qualitative properties of the dynamics, currently missing, and a formal analysis of the inner workings of LKCNNs. This is possible because of two key factors: the relative technical simplicity of LKCNNs and the vast understanding on chaotic maps.

We take the approach of identifying chaos through sensitive dependence on initial conditions by computing the Lyapunov exponent which measures the average rate of exponential departures of small perturbations [BGG80, PC12]. Formally, we consider the following classification problem: Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a smooth map recursively generating the bounded time series  $\{x_1, x_2, \dots\}$  as

$$x_{n+1} = f(x_n), \quad (1)$$

for a given initial condition  $x_0$ . We aim at classifying finite-length time sequences between *chaotic* or *regular* without knowledge of the analytical expression of  $f$ . In other terms, given finite sequences

$$x^\ell = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N,$$

we target accurate classifiers  $G$ , such that

$$\begin{aligned} G : \mathbb{R}^N &\rightarrow L := \{\text{chaotic, regular}\}, \\ x^\ell &\mapsto \ell \in L. \end{aligned} \quad (2)$$

Analogously to [BDNS20], we shall consider  $N = 500$ . Note that if the function  $f$  and its derivative are known analytically, the classification labels in  $L$  can be immediately determined by estimating the Lyapunov exponents of  $f$ . More specifically, the Lyapunov exponent,  $\lambda$ , corresponding to  $\{x_n\}$  generated by  $f$  is defined as

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log(|f'(x_i)|), \quad (3)$$

if the limit exists. Bounded sequences are chaotic when  $\lambda > 0$ , as attractors with  $\lambda > 0$  can only be bounded if a type of folding process merges widely separated trajectories (e.g. [PC12, WSSV85]). Conversely, when only data in the form of sequences  $(x^\ell)$  is available, direct approaches which reconstruct the Lyapunov exponent by estimating  $f'$  could be used. These approaches, however, require

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

a sufficiently large  $N$  (e.g. [VK86, BBA90]). Our experiments show that our choice  $N = 500$  is large enough to capture the dynamics and short enough that direct approaches are error prone.

Our analysis of the inner workings of LKCNNs leverages on well-known 1D maps: the Logistic map and the Sine-circle map. Additionally, we compare LKCNNs performance with a “traditional” classifier based on an estimate of  $\lambda$  via a local polynomial reconstruction of  $f'$  using Savitsky-Golay filtering [SG64]. We restrict to a data set with trajectories that do not exhibit periodic behavior over the input length, we refer to these trajectories as non-periodic. Non-periodic trajectories with Lyapunov exponents close to zero are the most challenging to classify: small perturbations can lead in fact to misclassification. For these, the absolute derivative along the orbits typically oscillate around 1. Overall, we show that:

- LKCNNs can identify qualitative properties of the dynamics of the input sequence as they outperform direct reconstruction of the Lyapunov exponent from the time series. Our experiment is performed on a restricted data set with only non-periodic sequences which are hard to classify as the average separation between consecutive time-steps is almost constant. Additionally, for these sequences the LKCNN cannot classify regular sequences by determining if they are repeating.
- LKCNNs map periodic inputs to periodic activation with exception of the dense layers. We can capture this mapping in a two dimensional matrix which we refer to as the *period matrix*. Untrained LKCNNs, meaning random weight initialized networks, have generically the same period matrix. The period matrix is a property of the architecture. We prove this property in a limit condition.
- Grouping trained LKCNNs for different weight initializations by period matrices, we observe a single period matrix which correlates with low performance models. This period matrix is equal to the period matrix of generic random weight initialization.

This last insight might also be useful for addressing more general models and settings.

The remainder of this paper is organized as follows. In Section 2, we present the dynamical systems for the classification problem, data sets and performance metrics. In Section 3, we present a method to directly reconstruct the Lyapunov exponent from the times series with Savitsky-Golay polynomial filtering. In Section 4, we revise LKCNN, present the architecture, explain the choice of hyperparameters and data sets. In Section 5, we present the main results related to Lyapunov reconstruction with Savitsky-Golay and periodic activation. A final discussion closes the paper. Table 1 can be used as a notation guideline throughout this paper.

Notation	Definition
$f$	one dimensional smooth map
$\{x_i\}$	one dimensional sequence
$k$ -periodic $\{x_i\}$	sequence satisfying $x_i = x_{i+k}$ with $x_1, x_2, \dots, x_k$ distinct
$\mathbb{Z}, \mathbb{N}, \mathbb{R}$	integers, natural numbers, real numbers
$\lambda$	Lyapunov exponent
$\mu$	mean
$\sigma$	standard deviation
$s k$	$k \in \mathbb{N}$ is divisible by $s \in \mathbb{N}$
$s \nmid k$	$k \in \mathbb{N}$ is not divisible by $s \in \mathbb{N}$
$T_{[.]}$	# true predictions classified as [.]
$F_{[.]}$	# false predictions classified as [.]

Table 1: Summary of notation

## 2 Time series, data sets, performance metrics

### 2.1 Time series: the Logistic and Sine-circle map

As in [BDNS20], we analyze LKCNN considering data sets built using two well-known maps: the Logistic map and the Sine-circle map. We briefly review some crucial features of the time series that these maps generate. The Logistic map [Rob76] is given by

$$f(x) = \alpha x(1 - x), \quad \alpha \in [0, 4], \quad x_0 = 0.5, \quad (4)$$

where  $\alpha$  is the bifurcation parameter. Observe that  $x_n \in [0, 1]$  for all  $\alpha$ . The bifurcation diagram corresponding to long-term evolution of the orbits is given in Figure 1a. The bifurcation diagram exhibits orbits doubling in period with geometric rate. After the period-doubling accumulation point,  $\alpha \approx 3.56995$ , chaos onsets and the map exhibits alternation of regular and chaotic behavior [Fei78].

The second map considered is the Sine-circle map [H<sup>+</sup>00, Boy86, Her79]:

$$f(\theta) = \theta + \Omega - \frac{\beta}{2\pi} \sin 2\pi\theta \mod 1, \quad \theta_0 = 0.5. \quad (5)$$

The Sine-circle map also exhibits a transition from regular to chaotic dynamics as the bifurcation parameter  $\beta$  changes. For  $\beta < 1$  only periodic and quasi-periodic trajectories occur since the Sine-circle map is invertible. Recall that quasi-periodic motion is characterized by non-chaotic and non-periodic dynamics, for a classical example take  $\Omega$  irrational and  $\beta = 0$  in (5). Invertibility implies that the folding dynamics which characterizes chaotic dynamics cannot occur [H<sup>+</sup>00]. As in [BDNS20] we consider  $\Omega = 0.606661$ ,  $\beta \in [0, 5]$ . We note that this choice of  $\Omega$  is sufficiently irrational that for  $\beta = 0$  the sequence is quasi-periodic from a numerical perspective. For this particular  $\Omega$  value, chaotic dynamics occur right after  $\beta = 1$  [H<sup>+</sup>00]. The bifurcation diagram corresponding to long-term evolution of the orbits is given in Figure 1b.

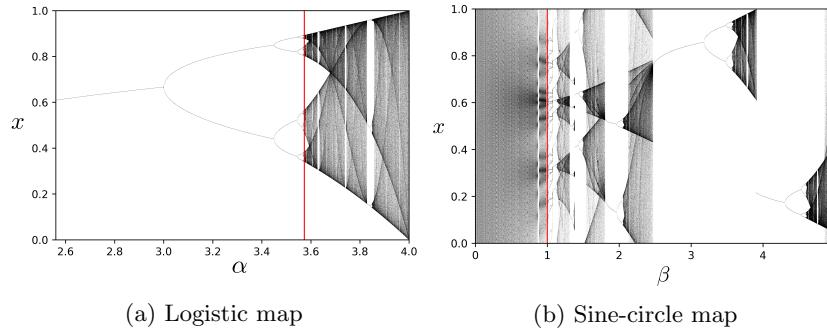


Figure 1: Bifurcation diagrams: In (a) the parameter corresponding to the onset of chaos is indicated by a red line. In (b) the parameter corresponding to the transition from invertible to non-invertible is indicated by the red line. Invertible maps do not exhibit chaotic dynamics.

## 2.2 Data sets

Given a discrete parameter set we consider trajectories corresponding to long-term evolution of the time series generated by (4) and (5). We use trajectories corresponding to equi-spaced parameters on a log scale for the Logistic data set as chaotic and regular trajectories are better balanced as a consequence of the geometric doubling over the parameter space. In [BDNS20], the trajectories corresponding to equi-spaced parameters on a linear scale are considered for the Logistic map. This data set is biased towards regular trajectories, see Table 2. We observe that periods of periodic orbits are much more uniformly distributed in the log-scale than in the linear-scale, Figure 2. Moreover, periodic orbits with low orbit periods are over-represented in the linear-scale.

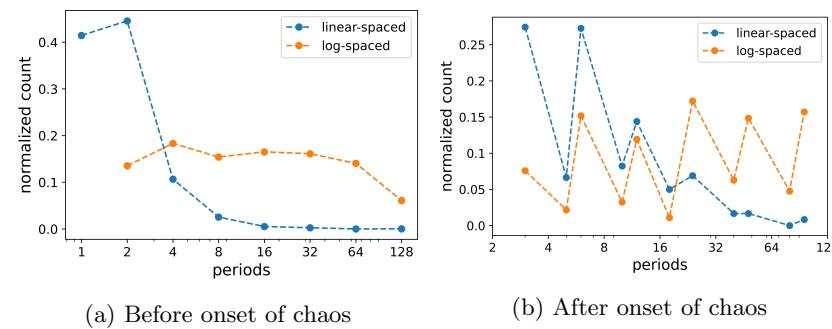


Figure 2: Logistic map period distributions for equi-spaced parameters on a linear- and log-scale: We consider periodic orbits before and after the onset of chaos. The period distribution for the log-scale is more uniform than the linear-scale.

Data sets	Parameter scale	Chaotic	Regular	
			Periodic	Non-periodic
Logistic	linear	27%	72 %	1 %
Logistic	log	46%	45 %	9 %
Sine-circle	linear	41%	49%	10 %

Table 2: Subdivision data sets: Considering equi-spaced parameters for the Logistic map on a log-scale yields a trajectory data set which is more balanced in comparison to the linear-scale. For the log-scale the parameters are logarithmically spaced at the onset of chaos. We note that regular non-periodic orbits for the Logistic map typically correspond to orbits which accumulate around orbits of low period whereas regular non-periodic orbits for the Sine-circle map can exhibit quasi-periodic dynamics which covers a large domain of the phase space.

For the Sine-circle map we consider trajectories corresponding to equi-spaced parameters on a linear scale. The regular trajectories can exhibit periodic and non-periodic behavior. The chaotic trajectories are all non-periodic. We note that the sign of the short time Lyapunov exponent is a good predictor of the sign of the converged Lyapunov exponent if the absolute Lyapunov exponent is sufficiently large. The Sine-circle data set is well-balanced between chaotic and regular trajectories, see Table 2. From Table 2 we observe that approximately 60% of all trajectories of the Sine-circle map are regular. We can heuristically derive this. For  $\beta \in (0, 1)$  the trajectory is periodic or quasi-periodic [H<sup>+00</sup>]. Consider  $U \subset \mathbb{R}$  such that for all  $\beta \in U$  the trajectory has either period 1, period 2 or period 4. From Figure 1b we observe that  $U$  has approximately length 2.

### 2.3 Performance metrics

The performance is assessed using the classification accuracy which is defined as

$$\text{Accuracy} := \frac{T_{\text{regular}} + T_{\text{chaotic}}}{T_{\text{regular}} + T_{\text{chaotic}} + F_{\text{regular}} + F_{\text{chaotic}}},$$

where  $T_{\text{regular}}$  = true regular predictions,  $T_{\text{chaotic}}$  = true chaotic predictions,  $F_{\text{regular}}$  = false regular predictions and  $F_{\text{chaotic}}$  = false chaotic predictions. Our test sets are well-balanced hence we do not need to consider alternatives [OD08, BOSB10]. We will also consider a precision measure for chaotic and regular orbits:

$$P_{\text{chaotic}} := \frac{T_{\text{chaotic}}}{T_{\text{chaotic}} + F_{\text{chaotic}}}, \quad P_{\text{regular}} := \frac{T_{\text{regular}}}{T_{\text{regular}} + F_{\text{regular}}}. \quad (6)$$

### 3 Direct reconstruction of Lyapunov exponents via Savitsky-Golay polynomial filtering

We can reconstruct the Lyapunov exponent directly from the times series by approximating derivatives. Here, we consider Savitsky-Golay as the fluctuations occurring for Lyapunov exponent close to zero are smoothed out which makes it outperform the classical method [WSSV85, RCDL93].

Savitsky-Golay is used to obtain the graph of  $f$ . The coefficients of the Savitsky-Golay approximation can be used to compute the derivative of  $f$ . Standard Savitsky-Golay works on an equidistant grid [SG64] but the extension to a non-equidistant grid is straightforward. We will briefly outline the method. The approximation of the graph of  $f$  is described by  $(x_i, x_{i+1})$ . We sort these points with respect to  $x_i$ . Denote the reordered points by  $(\bar{x}_i, y_i)$ . Consider  $k$  consecutive points  $(\bar{x}_i, y_i)$  through which we interpolate an  $m$ -degree polynomial with coefficients  $\mathbf{c} \in \mathbb{R}^m$ . For convenience we will consider  $i = 1, \dots, k$ . Take  $\kappa = (k-1)/2$ . Define the non-equidistant design matrix by  $A_{ij} = (\bar{x}_i - \bar{x}_\kappa)^j$ . Define  $\mathbf{y} = (y_1, \dots, y_k)^T$ . The set of equations that need to be solved are

$$A\mathbf{c} = \mathbf{y}.$$

From  $\mathbf{c}$  we can obtain the derivatives at  $x_i$  by straightforward calculus. Finally, we log-average over these derivatives to obtain an estimate for the Lyapunov exponent.

The graph of the attractor will generally consist of components which are disconnected from each other. So we perform Savitsky-Golay polynomial filtering on each component separately. Finally, we note that this procedure does not directly extend to higher dimensions.

### 4 Large Kernel Convolutional Neural Networks (LKCNN)

We review here the core features of Large Kernel Convolutional Neural Networks (LKCNN) as introduced in [BDNS20]. A Large Kernel Convolutional Neural Network has convolutional layers with kernel size large in relation to its feature channels. Conversely, standard convolutional networks usually consider a large number of feature channels in relation to the kernel size [IFFW<sup>+</sup>19, WYO17]. In this paper, as in [BDNS20], we use LKCNN for binary classification of sequences.

#### 4.1 Architecture

We report in Figure 3 the LKCNN architecture we employ to tackle (2). Specifically, we stack the following layers in a feed-forward fashion:

- two convolutional layers (kernel size: 100, large in relation to the signal length; stride 2; relu activation);

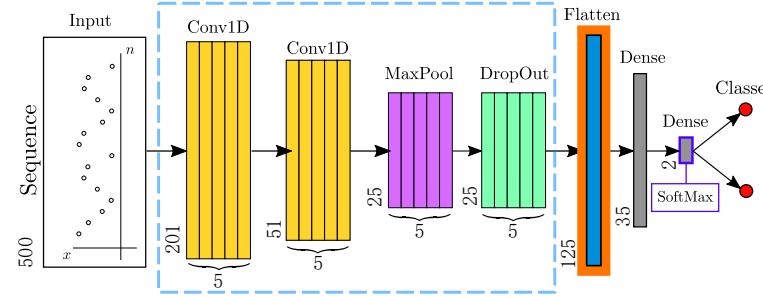


Figure 3: LKCNN architecture: The input time series  $x$  is classified as chaotic or regular. LKCNN have a typical CNN architecture but the convolutional layers have a large kernel size. The two Conv1D each have 5 filters with kernel size 100. The loss function used is cross-entropy. The internal analysis of the network will focus on the effect of the input on the activation at the flatten layer, highlighted in orange. Specifically, we will see how the activation of the layers in the blue dashed rectangle is mapped to an activation at the flatten layer.

- maxpooling layer (pool size 2);
- dropout regularization layer (rate 0.5);
- dense layer (sigmoid activation);
- final dense layer with softmax activation over the two classes in  $L$ . In other words, the network outputs a probability vector  $(p_{\text{regular}}, p_{\text{chaotic}}) = (p_{\text{regular}}, 1 - p_{\text{regular}})$ .

We consider an input sequence as chaotic if  $p_{\text{chaotic}} > 0.5$  and regular otherwise. We employ cross-entropy as training loss.

The overall architecture is the same as in [BDNS20] but with different hyperparameters. In fact, we use exclusively the Lyapunov exponent (3) to assign the classification labels on the training set. In particular, we estimate the Lyapunov exponent employing orbits much longer than  $N = 500$  steps. Our data sets are then built by restricting these long orbits to chunks  $N$ -steps long. On the opposite, [BDNS20] considers a ground truth labeling criterion based on a combination of Lyapunov exponent and Shannon entropy.

In Appendix B, we provide exhaustive analysis of the network performance considering various strides, activation functions and size of the first dense layers. As in [BDNS20], we observe that increasing the number of convolutional layers does not improve the performance. This suggests that sequence information is maximally condensed by 2 convolutional layers.

## 5 Uncovering the hidden workings of LKCNNs

### 5.1 LKCNNs can classify non-periodic orbits and outperforms Lyapunov reconstruction

We subdivide the input sequences corresponding to regular orbits in periodic and non-periodic. Here, periodic and non-periodic are identified by repeating and not repeating over the sequence input length, respectively. As in [BDNS20] we train the LKCNN on the Logistic trajectories, yet for equi-spaced parameters on a log-scale, and test it on the Sine-circle data set, see Table 3 for the performance results and Table 5 for the details on the data sets. The training accuracy was set to 0.975. We observe low performance on the regular non-periodic Sine-circle subset to the extend that the performance would be similar if we predict that a repeating sequence is periodic and a non-repeating sequence is chaotic. Poor performance is to be expected since the non-periodic subset of the Logistic and Sine-circle map are qualitatively different.

If we train the network on the Sine-circle data set we observe that the accuracy on the training set will not exceed 0.8. If we remove the periodic trajectories from the data set we observe accuracy exceeding 0.8, see Table 3. Hence, this is a labeling problem where the regular trajectories need to be subdivided into periodic and non-periodic trajectories to distinguish them from chaotic trajectories.

Training set	Test set	Accuracy		
		Chaotic $\mu \pm \sigma$		Regular $\mu \pm \sigma$
		Periodic	Non-periodic	
Logistic	Logistic	0.98 $\pm$ 0.0075	0.98 $\pm$ 0.0042	0.94 $\pm$ 0.014
Logistic	Sine-circle	0.88 $\pm$ 0.090	0.93 $\pm$ 0.052	0.045 $\pm$ 0.01

Table 3: **LKCNN trained on Logistic map can classify with high accuracy periodic orbits of the Sine-circle map but cannot accurately classify regular non-periodic orbits of the Sine-circle map.** We observe that training on the Logistic set generalizes well to Sine-circle data on the chaotic and periodic data subset of the Sine-circle map but performs poorly on the regular periodic set of the Sine-circle map. The latter is to be expected since the non-periodic trajectories are qualitatively different.

In Table 4, we investigate the performance on non-periodic data for LKCNN, average Lyapunov exponent over the input sequence which we refer to as short time Lyapunov exponent and direct reconstruction of Lyapunov exponent with Savitsky-Golay-polynomial filtering, Section 3. For all models in Table 4 the same test set is considered, see Table 5.

Model	$P_{\text{chaotic}}$	$P_{\text{regular}} \text{ (non-periodic)}$
LKCNN	1.0	1.0
Short-time Lyapunov exponent input sequence	1.0	0.99
Direct reconstruction Lyapunov exponent with Savitsky-Golay	1.0	0.93

Table 4: **LKCNN trained on only non-periodic orbits can distinguish regular non-periodic orbits from chaotic orbits.** We performed a precision comparison for models trained and tested on non-periodic Sine-circle orbits (cf. Equation (6)). We observe that the short-time Lyapunov exponent which is the Lyapunov exponent computed over the input sequence very accurately predicts the long time Lyapunov exponent. Hence, we can compare LKCNN prediction to prediction based on direct reconstruction of Lyapunov exponents via Savitsky-Golay polynomial filtering. The LKCNN outperforms the reconstruction.

Table ref.	Data set	# trajectories	trajectories per parameter
Table 3	Logistic	32 000	10
	Sine-circle	40 000	20
Table 4	Sine-circle	30 000	3

Table 5: Details training and testing data sets: Training to test ratio is 2:1. The Sine-circle data set corresponding to Table 3 is only used for testing.

Short time Lyapunov exponent is obtained with knowledge of  $f$  and direct reconstruction with Savitsky-Golay is obtained without knowledge of  $f$ . As the short time Lyapunov exponent is nearly perfect it implies that if the direct reconstruction can perfectly determine the derivatives then it can nearly perfectly classify the sequence. Hence, we shall compare the performance of the direct reconstruction to LKCNN. We observe that LKCNN performs significantly better. To add more rigor, the data set contains sequences where the Lyapunov exponent has converged to  $k$ -decimal precision which are then rounded to  $k$ -decimals to determine the label. This methodology applies to the short time Lyapunov exponent and the direct reconstruction of Lyapunov exponent in Table 4. The results in Table 4 are for  $k = 4$  but these results persist, see Appendix D.

## 5.2 Mapping periodic input to periodic activation is generic for random weight initialized untrained LKCNNs

We consider the LKCNN trained on the Logistic data set with parameters on a log-scale, Section 2.2. The activations for the first convolutional layer to the flatten layer preserves non-periodic and periodic structures. For a chaotic sequence the activations are non-periodic and for a periodic sequence the activations are

periodic, see Figure 4. We first present a rigorous result on the periodicity of a convolutional layer with stride  $s$  for periodic inputs and then extend this result to LKCNN.

### 5.2.1 Periodic activations of convolutional layers

Borrowing from equivariant theory for convolutional networks [CW16], we show that in an idealized setting where the network's input sequence has infinite length the convolutional filter applied to a  $k$ -periodic sequence yields an activation which is  $k$ -periodic. Furthermore, for stride  $s$  the period becomes  $k/s$  if  $k$  is divisible by  $s$ .

Denote by  $F : \mathbb{Z} \rightarrow \mathbb{R}$  the sequence feature map which is the map associated to a sequence  $\{x_i\}$  such that  $F(i) = x_i$ . Denote the convolutional filters by  $\phi : \mathbb{Z} \rightarrow \mathbb{R}$ . The convolutional filters also take as input the sequence indexes. A map  $F$  is  $k$ -periodic if  $F(z+k) = F(z)$  for all  $z \in \mathbb{Z}$ . Convolution of a sequence feature map  $F$  by the filter  $\phi$  is defined by

$$[F * \phi](z) := \sum_{y \in \mathbb{Z}} F(y)\phi(z-y).$$

Define  $S : \mathbb{Z} \rightarrow \mathbb{Z}$  by  $S(z) = sz$  with  $s \in \mathbb{N}$ . Then a convolution with stride  $s$  is given by  $[F * \phi] \circ S$ . Denote  $k \in \mathbb{N}$  divisible by  $s$  as  $s|k$ .

**Lemma 7.** *Let  $F$  be  $k$ -periodic. If  $s|k$  then the period of  $[F * \phi] \circ S$  is  $k/s$ . If  $s \nmid k$  then the period of  $[F * \phi] \circ S$  is the product of all the terms in the prime factorization of  $k$  that are not contained in  $s$ .*

Observe that if  $F$  is  $k$ -periodic and  $s = 1$  then the period of  $[F * \phi] \circ S$  is  $k$ . Note that  $[F * \phi]$  can have period less than  $k$  for suitably chosen  $\phi$ . For example, if  $\phi$  is the zero function then  $[F * \phi]$  has period 1. For this reason the definition of  $k$ -periodic for  $F$  does not require distinctness of  $F(1), F(2), \dots, F(k)$ .

*Proof.* Take  $\hat{k} \in \mathbb{N}$ . We define  $\hat{y} = y - s\hat{k}$ . We can write

$$F(y)\phi(s(z + \hat{k}) - y) = F(\hat{y} - s\hat{k})\phi(sz - \hat{y}).$$

Using the above we can write

$$[F * \phi]S(z + \hat{k}) = \sum_{\hat{y} \in \mathbb{Z}} F(\hat{y} - s\hat{k})\phi(sz - \hat{y}).$$

Hence, if  $s|k$  then we obtain that  $[F * \phi]S(z + \hat{k}) = [F * \phi]S(z)$  for  $\hat{k} = k/s$  and if  $s \nmid k$  then we obtain that  $[F * \phi]S(z + \hat{k}) = [F * \phi]S(z)$  for  $\hat{k}$  equal to the product of all the terms in the prime factorization of  $k$  that are not contained in  $s$ .  $\square$

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
 PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

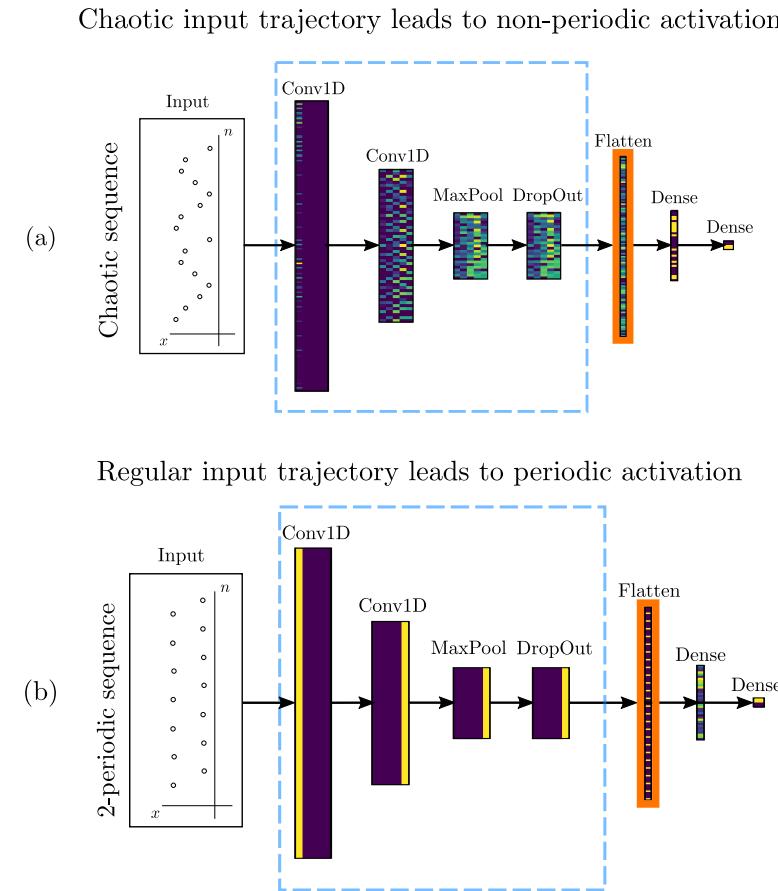


Figure 4: Activations LKCNN: We visualize the activation for a correctly classified chaotic and periodic trajectory. From the first convolutional layer to the flatten layer the non-periodic, (a), or periodic structure, (b), is preserved. The last two layers are dense and do not preserve any periodic or non-periodic structure.

### 5.2.2 Periodic activations of LKCNN

The result from Section 5.2.1 assumes an infinitely large network. Practically, the periods that can be captured by the network depend on the size of the network. Observe that the largest activation period that the first convolutional layer can capture is 100 since it has size  $201 \times 5$ , see Figure 4. Hence, since we have stride 2 the largest period the network can capture in terms of the input  $x$  is 200.

Denote by  $k$  the period of the activation for Maxpooling and  $p$  the pool size. Then if  $p|k$  the output period is  $k/p$  and if  $p \nmid k$  then the output period is  $k$ . Here, we have  $p = 2$ . Again, in a practical setting the size of the layer restricts the periods that can be captured. Here, the output activation period can be at most 12.

The flatten layer will have periodic activation if the dropout layer also has periodic activation. Furthermore, if the activation matrix at the dropout layer is non-constant then the periodicity is increased by a multiplicative factor of 5 since the number of columns of the activation matrix is 5 which is prime. Our experiments indicate that activation at the flatten layer can vary if we vary over the period of  $x$ , Figure 5 and Appendix A.

The activation periodicity is lost in the last two dense layers. Hence, to study network periodicity we will focus on the periodicity at the flatten layer in relation to the periodicity of the sequence.

Combining the results from this section with Lemma 7 we obtain the following heuristic statement: *for an untrained LKCNN we expect that for  $k \leq 96$  the period at the flatten layer is  $5k/2^i$  where  $i$  is the largest  $i \leq 3$  for which  $2^i|k$ .*

### 5.3 Activation periodicity influences performance

Generally, a sufficiently small periodic input implies a periodic activation at the flatten layer. We refer to the latter as the network period. For example, the 2 period orbit in Figure 5 has network period 5. We experimentally observe that orbits from the data set with the same period map into a single network period if the orbit period is sufficiently small. Consequently, we represent this mapping by a binary matrix which identifies orbit periods to network periods, see Figure 6. We note that we could formulate these results in a non-binary setting where the orbit period is not uniquely mapped to a single network period. However, similar results still hold, see Appendix C. We also note that orbits from outside the data set can inconsistently map into a single network period even for orbits with small periods.

The binary matrix in Figure 6 is called the period matrix. Period matrices can differ per trained network weights. We consider 250 random weight initializations trained over 1000 epochs with patience 50. Here the Logistic data set with equi-spaced parameters on a log-scale is used. We then classify the trained weights using the period matrices. The largest class consist of 32% of trained network weights and has period matrix as depicted in 7a which will be referred to as period matrix  $A$ . Single classes in the complement of  $A$  make up at most

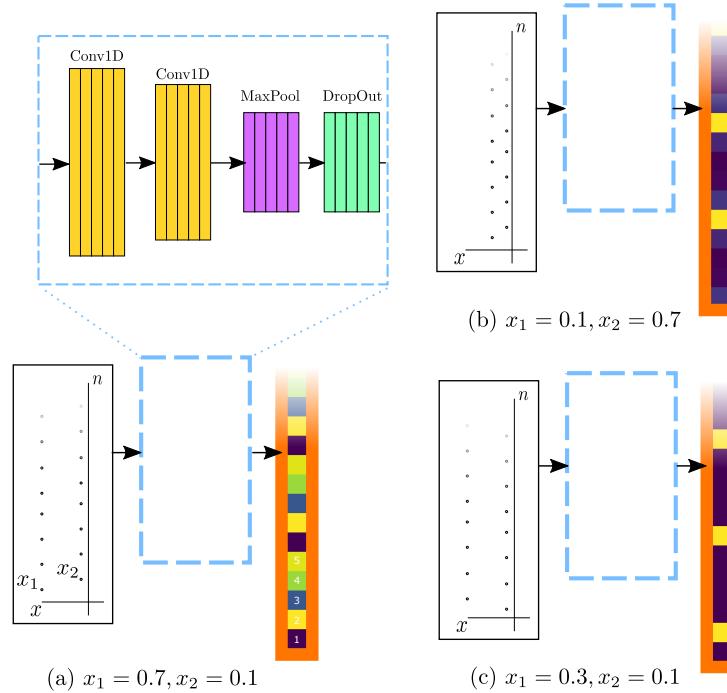


Figure 5: Activation at flatten layer: We consider the first 15 nodes at the flatten layer for period 2 trajectories given by  $x_1, x_2$ . We observe that the periodicity is invariant for all three cases. Reversing the order alters the activation (a)(b). Position of the maximum activation, yellow, and minimal activation, purple, can vary (a)(b)(c).

4% of all trained networks. We consider these classes to be non-significant. We note that the great variety in period matrices is not in contradiction with Lemma 7 since the Lemma only gives an upper bound on the activation period of the convolutional layers.

We observe that 56% of class  $A$  converges to a local minimum which has accuracy between 55-60%, Figure 7b. For the network weights in the complement of class  $A$  we have that Q1 and Q3 are at 0.94 and 0.98 accuracy, respectively. The average accuracy of class  $A$  is 0.72 whereas the average accuracy of its complement is 0.94. Hence, network weights of class  $A$  have a negative effect on the overall performance.

The period matrix  $A$  exactly corresponds to the theoretically determined periods for untrained networks, Section 5.2.2. For other period matrices the weights reduce the network period. This appears to generally have a positive effect on the performance. Heuristically, we can argue that there is no benefit to preserve a property of the untrained network which is imposed on the network independent of data properties. Therefore, network weights with period

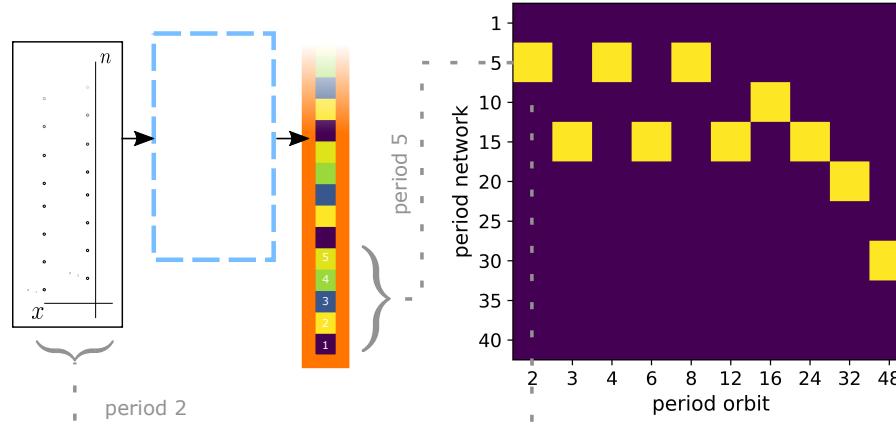


Figure 6: Period matrix: On the left a period 2 orbit with flatten layer period 5 is depicted. For each periodic orbit we identify the orbit's period to a period at the flatten layer. The period at the flatten layer is referred to as network period. Orbits with the same period do not uniquely map to a network period. However, in period range 2 to 48 the majority of period orbits map to a single network period. We associate a binary matrix to this identification. This period matrix can differ per trained network weight initialization. However, we can identify classes which have the same period matrix

matrix  $A$  underperform in relation to its complement.

## 6 Discussion

In this work we consider the issue of classifying time sequences, discriminating regular from chaotic dynamics. To this purpose we analyze a state-of-the-art approach based on Large Kernel Convolutional Neural Networks (LKCNNs), and compare it with a “traditional” quantitative approach that reconstructs the Lyapunov exponent. On these bases, we show that LKCNNs can identify qualitative properties of the dynamics which are lost to quantitative methods. LKCNNs have a simple structure: few convolutional layers with peculiarly large filters connected to a dense layer. This structure simplifies numerical experiments and allowed us to identify relevant network features key for performance. Specifically, LKCNNs do not determine whether a sequence is chaotic or regular by checking if the sequence repeats as LKCNN still have high performance on a data set where all regular orbits are non-periodic. Hence, LKCNNs have a qualitative understanding of the dynamics of the input sequence. This enables them to outperform direct Lyapunov exponent reconstruction methods. Here, we consider a reconstruction approach based on Savitsky-Golay polynomial filtering. LKCNNs higher classification accuracy strongly emerges as we consider sequences whose Lyapunov exponent is close to zero: the hardest to evaluate as

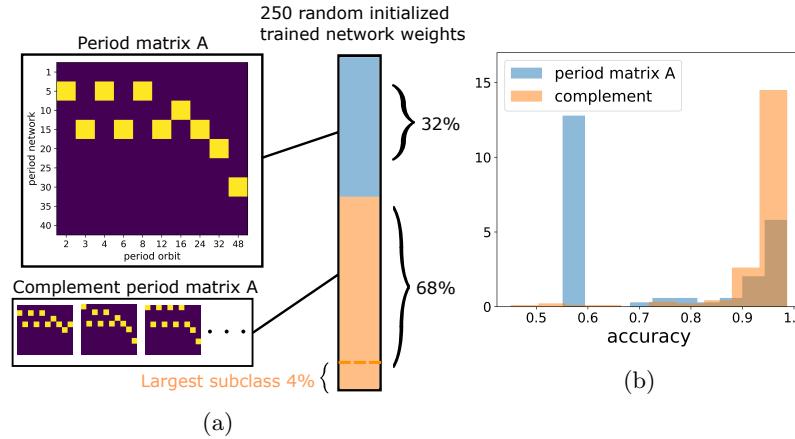


Figure 7: Network weights classified by period matrix: We investigated the period matrices associated to 250 random initialized trained network weights. In (a) we observe that the most common period matrix is period matrix A to which 32% of network weights converge. The second largest period matrix class is 4%. We compare the performance between network weights with period matrix A and network weights in their complement. In (b) we plot the density functions for the accuracy of both classes. We observe that network weights with period matrix A have on average lower performance.

the average separation between consecutive time-steps is almost constant. We investigated the emerging connection between input periodicity and periodicity in the activation of the non-dense layers. We have shown this aspect to be paramount for performance. For this analysis, we have introduced the notion of a period matrix which is a two dimensional binary matrix which represents the mapping between periodic inputs and the network's activation excluding the dense layers.

We considered generic untrained networks first. For these, we have determined theoretically the connection between the input period and the periodicity of the activation of the convolutional layers. Effectively, this yields a period mapping-type relation. This mapping is represented by the period matrix. We showed that for generic networks this period matrix is unique. The weights minimizing the loss function are meager within the weight space. Consequently, this period mapping needs not be preserved by training. Indeed, we observed that trained network weights can lead to a variety of period mappings. Nevertheless, a significant percentage of trained networks have period matrix equal to that of generic untrained network weights. These network weights underperform. In other terms, we numerically verified how the period matrix is a feature correlating with performance. Heuristically, there is no benefit to preserve a property of the untrained network if it has no relation to the data. If a property of the data is reflected on the network we would expect an increase in performance.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

Models with high performance have period matrices featuring network periods which are lower compared to the case of a generic untrained network. Additionally, high performance is not a property of a singular period matrix. Two aspects remain outstanding: how the network training yields period reduction, and whether properties of the periodicity matrix correlating with high classification performance can be identified independently on the periodicity matrix of the untrained network. Possibly, the absence of clear connections indicates that restrictions on the period matrix adversely affect performance.

Finally, here we focused on discrete dynamical systems. It would be relevant to formulate suitable experiments in a continuous setting. We note that our approach leverages on periodicity of trajectories. Hence, it could be applied to other discrete dynamical systems or settings where it is possible to construct Poincaré maps.

This paper illustrates how dynamical system problems can be solved using neural networks and how neural network problems can be understood using dynamical systems. Thanks to a neural network approach we could achieve higher performance than traditional methods at classifying chaotic from regular time series. On the other hand, an approach hinged on dynamical systems analysis has been key to understand the inner workings of the network.

**Acknowledgments:** During this research Thomas de Jong was also affiliated to University of Groningen and Xiamen University. Many thanks to Alef Sterk for his helpful comments and literature recommendations. Also, many thanks to Klaas Huizenga for providing hardware during Thomas de Jong's stay in Groningen. This research was partially supported by JST CREST grant number JPMJCR2014.

## Appendix

### A Generalizations of 2-periods

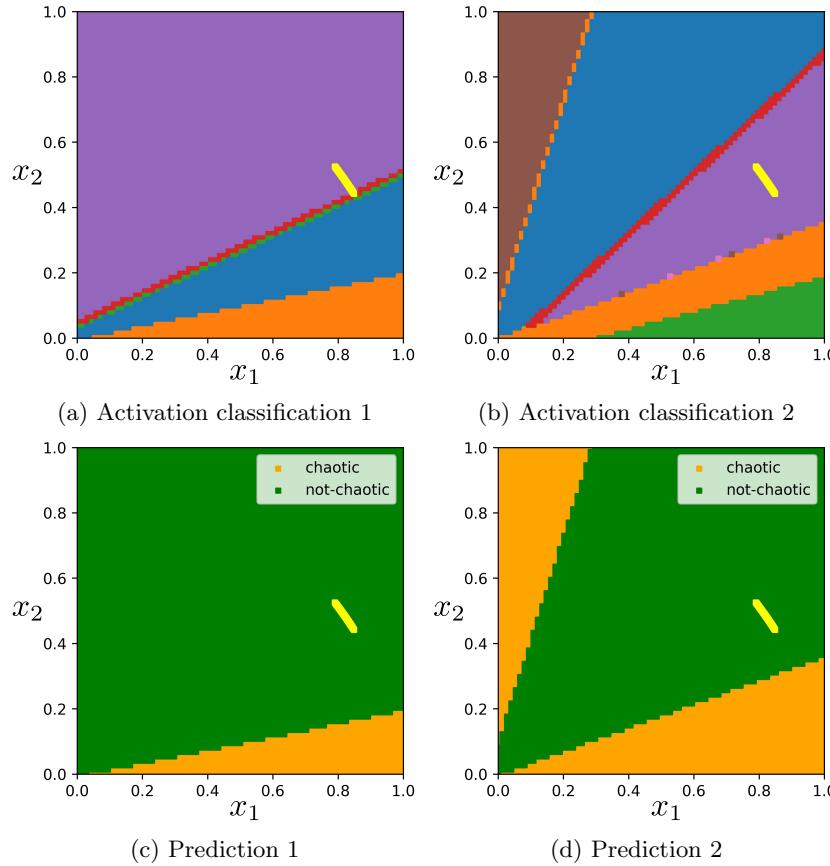


Figure 8: Period 2 orbits activation classification with predictions. The yellow line corresponds to period 2 orbits in the Logistic data set. The other colors in (a)(b) correspond to period 2 orbits outside the data set. These are grouped by activation at the flatten layer where points with the same color have zero activation at the same position in the flatten layer.

Using the method from Section 5.2 we investigate the network activation at the flatten layer for all possible trajectories with period 2. We denote the period 2 input sequence by  $x = [x_1, x_2, x_1, x_2, \dots]$ . Hence, can represent  $x$  as a point in  $\mathbb{R}^2$ . In Figure 8 the yellow line corresponds to period 2 trajectories from the Logistic data set. For period 2 orbits the activation at the flatten later can be represented by  $y = [y_1, y_2, y_3, y_4, y_5, y_1, y_2, \dots]$ , Section 5.2.2. We classify the

activation by the position of zero elements. Hence, activation  $[1, 0, 0, 0, 0, \dots]$  is in the same class as  $[0.1, 0, 0, 0, 0, \dots]$ . In Figure 8ab we display the domains corresponding to activation classes of  $x$  for two different weight initializations of LKCNN and in Figure 8cd we visualized the corresponding LKCNN prediction of  $x$ .

We observe that in Figure 8 the domains of the generalized period 2 trajectories for the activation classes and prediction vary based on the initialization. In Figure 8ab the activation classification is not preserved under changing the order of  $x_1, x_2$ . This order preservation is also not satisfied for the prediction, see Figure 8cd. As expected the prediction is incorrect if the prediction is applied to period 2 inputs far away from the trained data. In Figure 8b there are 9 domains (excluding the yellow data domain). This means that there are different permutations of zero activation with the same number of zeros since the period at the flatten layer is at most 5.

The domains in Figure 8ab are all bounded by linear functions of the form  $x_2 = ax_1 + b$  with  $a > 0$ . In addition, in Figure 8a a perturbation of a period 2 input sequence in the data set can lead to different activation classes. To explore this a bit further we can investigate which  $x$  have activation close to a period 2 orbit from the data set, see Figure 9. We observe again a domain bounded by linear functions.

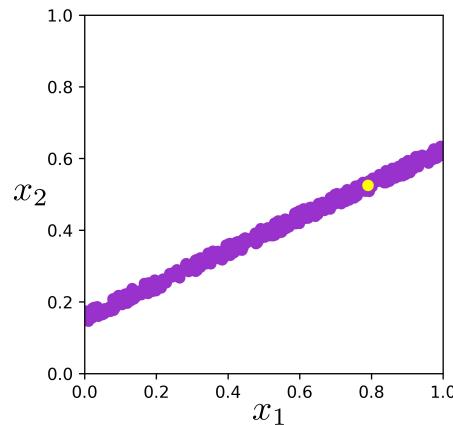


Figure 9: Sub-domain of Figure 8a: The purple domain corresponds to a sub-domain of the purple domain in Figure 8a. The yellow point corresponds to a period 2 orbit from the data set. The purple domain has activation close to the yellow data point in the 2-norm.

## B Architecture optimization

Our experiments show that the performance can be improved by adjusting the following two hyperparameters:

- Stride in the two convolutional layers,
- Number of nodes in the fully connected dense layer after the flatten layer.

All hyperparameter variations exhibited instability when trained using data set corresponding to parameters on a log-scale. We resolved this by setting the learning rate to 0.000388.

The experiments in Figure 10 concern the accuracy over the validation set of 50 models for 200 epochs per architecture. Model 0 is the LCKNN architecture used in [BDNS20]. Since most of the experiments concerns models trained with high accuracy we are interested in selecting the hyperparameters which can consistently train models with high accuracy. We selected model 1 since the accuracy range over Q2 to Q3 is the highest.

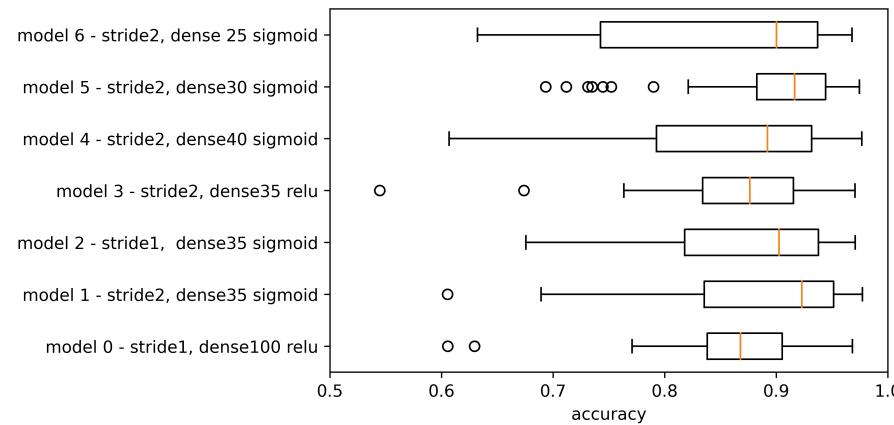


Figure 10: Validation accuracy boxplots for different hyperparameters.

## C Period matrices

### C.1 Period matrix classes

We recall that period matrix class A, Figure 7a makes up 32% of the 250 trained models. In Figure 11 we visualize the 2nd, 3rd and 4th largest classes.

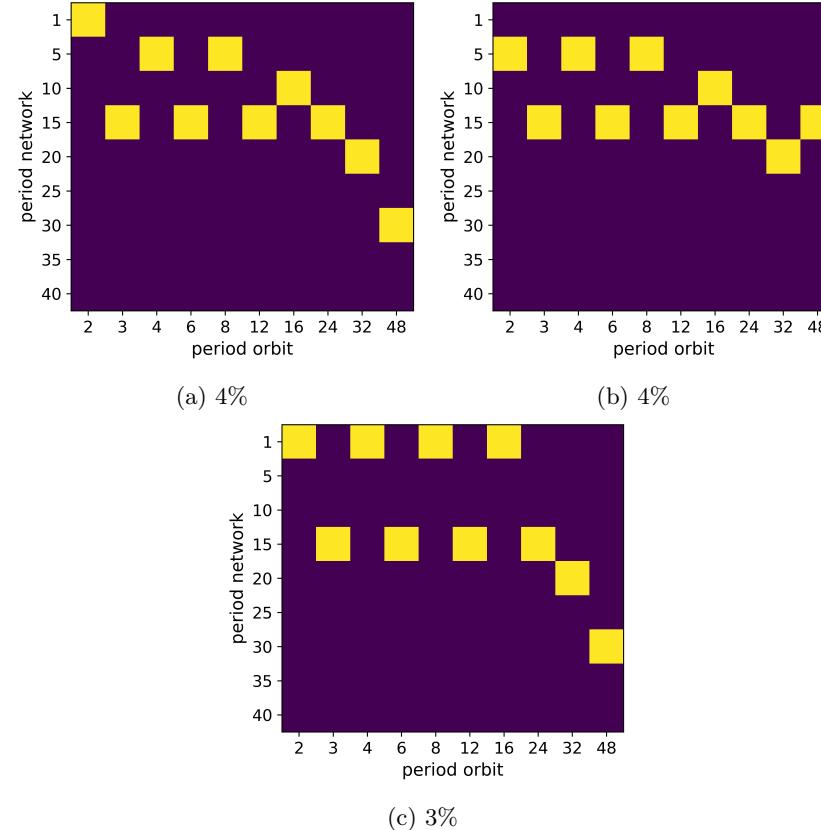


Figure 11: Period matrices: The percentage indicates the size of the class with respect to all the models

### C.2 Non-binary period matrices

Recall that for low orbit periods we generally obtain a unique period of the network. Most high orbit periods map onto more than one network period. We normalize the period network over each period orbit by dividing by the total orbits of a specific period and then put the values in a matrix as before, see Figure 12.

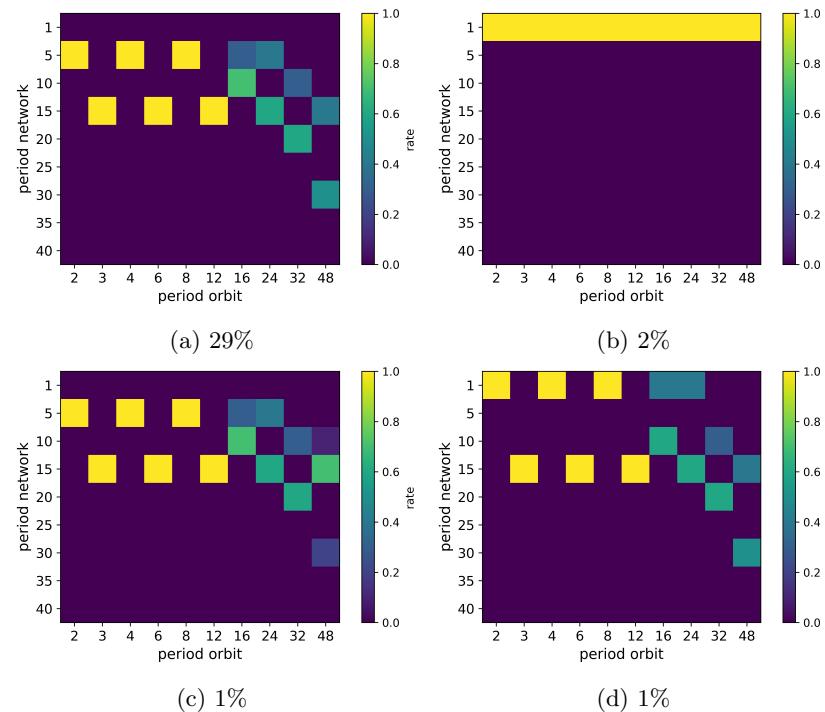


Figure 12: Non-binary period matrices: The percentage indicates the size of the class with respect to all the models

In Figure 12 we visualized the 4 most common classes. Figure 12a is the major class making up Figure 7a. We also note that the max over each column of Figure 12c gives Figure 11b. If we take the max over each column of Figure 12b or Figure 12d we do not get Figure 11a or Figure 11c.

## D Performance models non-periodic Sine-circle data

In Table 4 of Section 5.1 we presented the performance of LKCNN, Savitsky-Golay reconstruction of the Lyapunov exponent and short-time Lyapunov exponent on classifying chaos for non-periodic trajectories of the Sine-circle map. For non-periodic trajectories the performance depends strongly on the decimal precision of the Lyapunov exponent. Here we consider sequences which have a Lyapunov exponent that converges in  $k$ -decimals and classify the sequence as chaotic using the first  $k$ -decimals. In Figure 13 we consider the performance as function of the convergence of the first  $k$ -decimals of the Lyapunov exponent.

The three models have near perfect performance when we only consider 2-decimal convergence or restrict to the chaotic subset. We note that the qualita-

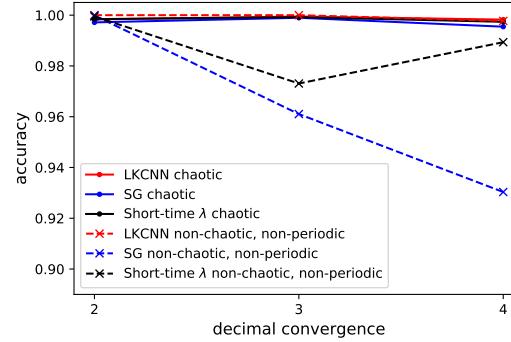


Figure 13: Performance models on non-periodic Sine-circle data set as function of decimal convergence of Lyapunov exponent: We consider the performance of LKCNN, Savitsky-Golay reconstruction of the Lyapunov exponent (SG) and short-time  $\lambda$  (averaged Lyapunov exponent over the input sequence). The  $x$ -axis corresponds to a subset where the first  $k$ -decimals of the Lyapunov exponent converge. Hence, the classification will also be determined by the first  $k$ -decimals. All models have near perfect performance when we only consider 2-decimal convergence or restrict to the chaotic subset.

tive observations of Table 4 also apply if we consider convergence in 3-decimals. Evaluating the Lyapunov exponent over regular non-periodic sequences we typically observe fluctuations around zero which makes the classification task more difficult which is clearly reflected when we compare the performance for decimal precision greater than 2. Observe that performance of short-time  $\lambda$  is non-monotone over the regular non-periodic subset. The sequences corresponding to  $k$ -decimal convergence are a subset of the sequences corresponding to  $(k+1)$ -decimal convergence. This property is insufficient to conclude anything about monotonicity in the performance results. These subsets do decrease in size with increasing  $k$ . For  $k=5$  the resulting subset is so small that the results are not reliable from a generalization perspective.

## E Regular non-periodic Logistic map trajectories

If we train the LKCNN on the Sine-circle data set from Section 2.2 we obtain poor performance on the regular non-periodic set, see Table 3. This begs the question why we obtain such good performance on the Logistic set for the regular non-periodic subset. It turns out that the majority of this subset is in a sense close to periodic trajectories of low period. We divide  $x$  in consecutive chunks of length  $k$  by defining

$$y_k^m(x) := (x_{1+k(m-1)}, \dots, x_{1+km}).$$

We consider the error given by minimizing over  $k$  the average difference between consecutive  $y_k^m(x)$ -chunks:

$$\text{period error}(x) := \min_{k \leq \frac{n}{2}} \frac{1}{\lfloor n/k \rfloor} \sum_{m=1}^{\lfloor n/k \rfloor} |y_k^{m-1}(x) - y_k^m(x)|,$$

$$K(x) := \min_{k \leq \frac{n}{2}} \operatorname{argmin} \frac{1}{\lfloor n/k \rfloor} \sum_{m=1}^{\lfloor n/k \rfloor} |y_k^{m-1}(x) - y_k^m(x)|.$$

The period error measures how close  $x \in \mathbb{R}^n$  is to a periodic orbit. More specifically, we have that if  $\{\tilde{x}_i\}$  is a  $k$ -periodic sequence and  $x = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$  with  $n \geq 2k$  then  $\text{period error}(x) = 0$  and  $K(x) = k$ .

Note that it would be computationally unfeasible to compare all chunks. However, for the computation we consider the minimum of the period error for the original and reversed sequence.

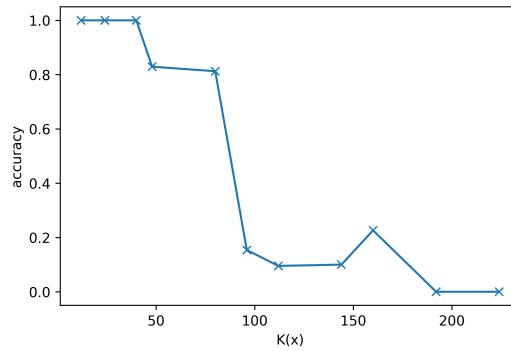


Figure 14: Accuracy non-periodic trajectories Logistic data: As the approximate period  $K(x)$  is increased the accuracy decreases. This gives evidence that the network can only classify these trajectories correctly if they are close to trajectories of sufficiently small period. An exception here are periods which are a power of 2. These periods have been excluded in the graph since the network has accuracy close to 1 on these subsets. This is to be expected since the majority of periodic trajectories have period  $2^k$

For  $\text{period error}(x) < 10^{-5}$  we consider the accuracy of non-periodic orbits which are labeled as regular. If  $K(x) = 2^k$  for  $k < 8$  the network scores high accuracy since the majority of periodic orbits in the data have period  $2^k$ , see Figure 15.

In Figure 14 we exclude  $K(x) = 2^k$ . We observe that only for non-periodic trajectories with small  $K(x)$  we obtain high-accuracy.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
 PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

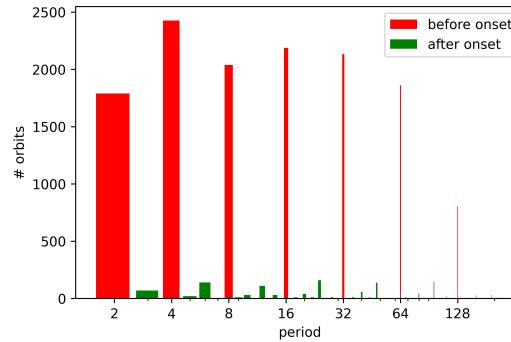


Figure 15: Period distribution for Logistic data set corresponding to log-scaled parameters.

## References

- [AUK22] B. Arıcıoğlu, S. Uzun, and S. Kaçar. Deep learning based classification of time series of chen and rössler chaotic systems over their graphic images. *Physica D: Nonlinear Phenomena*, 435:133306, 2022.
- [BBA90] P. Bryant, R. Brown, and H.D.I. Abarbanel. Lyapunov exponents from observed time series. *Physical Review Letters*, 65(13):1523, 1990.
- [BDNS20] N. Bouillé, V. Dallas, Y. Nakatsukasa, and D. Samaddar. Classification of chaotic time series with deep learning. *Physica D: Nonlinear Phenomena*, 403:132261, 2020.
- [BFBZ20] P.G. Breen, C.N. Foley, T. Boekholt, and S.P. Zwart. Newton versus the machine: solving the chaotic three-body problem using deep neural networks. *Monthly Notices of the Royal Astronomical Society*, 494(2):2465–2470, 2020.
- [BGG80] G. Benettin, L. Galgani, A. Giorgilli, and J.M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Mécanique*, 15(1):9–20, 1980.
- [BK22] S.L. Brunton and J.N. Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.
- [BLH21] Y. Bengio, Y. Lecun, and G. Hinton. Deep learning for ai. *Communications of the ACM*, 64(7):58–65, 2021.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
 PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

- [BLMC<sup>+</sup>23] R. Barrio, Á. Lozano, A. Mayora-Cebollero, C. Mayora-Cebollero, A. Miguel, A. Ortega, S. Serrano, and R. Vigara. Deep learning for chaos detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(7), 2023.
- [BOSB10] K.H. Brodersen, C.S. Ong, K.E. Stephan, and J.M. Buhmann. The balanced accuracy and its posterior distribution. In *2010 20th International conference on pattern recognition*, pages 3121–3124. IEEE, 2010.
- [Boy86] P.L. Boyland. Bifurcations of circle maps: Arnol'd tongues, bistability and rotation intervals. *Communications in Mathematical Physics*, 106(3):353–381, 1986.
- [BT11] H.W. Broer and F. Takens. *Dynamical systems and chaos*, volume 172. Springer, 2011.
- [Cel10] A. Celletti. *Stability and chaos in celestial mechanics*. Springer Science & Business Media, 2010.
- [CGRFV22] A. Celletti, C. Gales, V. Rodriguez-Fernandez, and M. Vasile. Classification of regular and chaotic motions in hamiltonian systems with deep learning. *Scientific Reports*, 12(1):1–12, 2022.
- [CMBT21] A. Corbetta, V. Menkovski, R. Benzi, and F. Toschi. Deep learning velocity signals allow quantifying turbulence intensity. *Science Advances*, 7(12):eaba7281, 2021.
- [CW16] T. Cohen and M. Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.
- [DIX19] K. Duraisamy, G. Iaccarino, and H. Xiao. Turbulence modeling in the age of data. *Annual review of fluid mechanics*, 51:357–377, 2019.
- [Fei78] M.J. Feigenbaum. Quantitative universality for a class of nonlinear transformations. *Journal of statistical physics*, 19(1):25–52, 1978.
- [GBC16] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.
- [GH13] J. Guckenheimer and P. Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42. Springer Science & Business Media, 2013.
- [H<sup>+</sup>00] R.C. Hilborn et al. *Chaos and nonlinear dynamics: an introduction for scientists and engineers*. Oxford University Press on Demand, 2000.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
 PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

- [Her79] M.R. Herman. Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Publications Mathématiques de l'IHÉS*, 49:5–233, 1979.
- [HMS21] S. Hassona, W. Marszalek, and J. Sadecki. Time series classification and creation of 2d bifurcation diagrams in nonlinear dynamical systems using supervised machine learning methods. *Applied Soft Computing*, 113:107874, 2021.
- [IFFW<sup>+</sup>19] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.A. Muller. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.
- [KHW89] M.S. Kurt Hornik and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Netw.*, 2(5):359–366, 1989.
- [KKK98] Y.A. Kuznetsov, I.A. Kuznetsov, and Y. Kuznetsov. *Elements of applied bifurcation theory*, volume 112. Springer, 1998.
- [KMA<sup>+</sup>21] K. Kashinath, M. Mustafa, A. Albert, J.L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194):20200093, 2021.
- [KPDV92] J.M. Kuo, J.C. Principe, and B. De Vries. Prediction of chaotic time series using recurrent neural networks. In *Proc. 1992 IEEE Workshop of Neural Networks in Signal Processing*, pages 436–443. Citeseer, 1992.
- [Lor63] E.N. Lorenz. Deterministic nonperiodic flow. *Journal of atmospheric sciences*, 20(2):130–141, 1963.
- [LY03] Y.C. Lai and N. Ye. Recent developments in chaotic time series analysis. *International Journal of Bifurcation and Chaos*, 13(06):1383–1422, 2003.
- [MOG97] S. Mukherjee, E. Osuna, and F. Girosi. Nonlinear prediction of chaotic time series using support vector machines. In *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pages 511–520. IEEE, 1997.
- [OCRT22] G. Ortali, A. Corbetta, G. Rozza, and F. Toschi. Numerical proof of shell model turbulence closure. *Physical Review Fluids*, 7(8):L082401, 2022.
- [OD08] D.L. Olson and D. Delen. *Advanced data mining techniques*. Springer Science & Business Media, 2008.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

- [PC12] T.S. Parker and L. Chua. *Practical numerical algorithms for chaotic systems*. Springer Science & Business Media, 2012.
- [PHG<sup>+</sup>18] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical review letters*, 120(2):024102, 2018.
- [PJSF04] H.O. Peitgen, H. Jürgens, D. Saupe, and M.J. Feigenbaum. *Chaos and fractals: new frontiers of science*, volume 106. Springer, 2004.
- [PLDJ21] A. Panday, W.S. Lee, S. Dutta, and S. Jalan. Machine learning assisted network classification from symbolic time-series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(3):031106, 2021.
- [PRK92] J.C. Principe, A. Rathie, and J.M. Kuo. Prediction of chaotic time series with neural networks and the issue of dynamic modeling. *International Journal of Bifurcation and Chaos*, 2(04):989–996, 1992.
- [RCDL93] M.T. Rosenstein, J.J. Collins, and C.J. De Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1-2):117–134, 1993.
- [RKJ<sup>+</sup>19] J. Rabault, M. Kuchta, A. Jensen, U. Réglade, and N. Cerardi. Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control. *Journal of fluid mechanics*, 865:281–302, 2019.
- [Rob76] M. Robert. Simple mathematical models with complicated dynamics. *Nature*, 261:459–467, 1976.
- [SCGH05] N. Sebe, I. Cohen, A. Garg, and T.S. Huang. *Machine learning in computer vision*, volume 29. Springer Science & Business Media, 2005.
- [SG64] A. Savitzky and M.J.E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.
- [SHM<sup>+</sup>16] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.
- [Str18] S.H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC press, 2018.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.  
 PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0160813

- [TKL<sup>+</sup>20] Y. Tang, J. Kurths, W. Lin, E. Ott, and L. Kocarev. Introduction to focus issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(6):063151, 2020.
- [VK86] J.A. Vastano and E.J. Kostelich. Comparison of algorithms for determining lyapunov exponents from experimental data. In *Dimensions and entropies in chaotic systems: quantification of complex behavior*, pages 100–107. Springer, 1986.
- [WAB10] J.W. Woolley, P.K. Agarwal, and J. Baker. Modeling and prediction of chaotic systems with artificial neural networks. *International journal for numerical methods in fluids*, 63(8):989–1004, 2010.
- [Wan19] Z. Wang. Github. [https://github.com/cauchyturing/UCR\\_Time\\_Series\\_Classification\\_Deep\\_Learning\\_Baseline](https://github.com/cauchyturing/UCR_Time_Series_Classification_Deep_Learning_Baseline), 2019.
- [WBGL15] J. Wieting, M. Bansal, K. Gimpel, and K. Livescu. Towards universal paraphrastic sentence embeddings. *arXiv preprint arXiv:1511.08198*, 2015.
- [WLTMK22] D. Wenkack Liedji, J.H. Talla Mbé, and G. Kenne. Classification of hyperchaotic, chaotic, and regular signals using single nonlinear node delay-based reservoir computers. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(12), 2022.
- [WSSV85] A. Wolf, J.B. Swift, H.L. Swinney, and J.A. Vastano. Determining lyapunov exponents from a time series. *Physica D: nonlinear phenomena*, 16(3):285–317, 1985.
- [WYO17] Z. Wang, W. Yan, and T. Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, pages 1578–1585. IEEE, 2017.
- [ZGS<sup>+</sup>23] Y. Zhou, S. Gao, M. Sun, Y. Zhou, Z. Chen, and J. Zhang. Recognizing chaos by deep learning and transfer learning on recurrence plots. *International Journal of Bifurcation and Chaos*, 33(10):2350116, 2023.