




Predicting discrete-time bifurcations with deep learning

Received: 28 August 2023

Accepted: 27 September 2023

Published online: 10 October 2023

 Check for updates

Thomas M. Bury¹  , Daniel Dylewsky², Chris T. Bauch² , Madhur Anand³, Leon Glass¹, Alvin Shrier^{1,4} & Gil Bub^{1,4}

Many natural and man-made systems are prone to critical transitions—abrupt and potentially devastating changes in dynamics. Deep learning classifiers can provide an early warning signal for critical transitions by learning generic features of bifurcations from large simulated training data sets. So far, classifiers have only been trained to predict continuous-time bifurcations, ignoring rich dynamics unique to discrete-time bifurcations. Here, we train a deep learning classifier to provide an early warning signal for the five local discrete-time bifurcations of codimension-one. We test the classifier on simulation data from discrete-time models used in physiology, economics and ecology, as well as experimental data of spontaneously beating chick-heart aggregates that undergo a period-doubling bifurcation. The classifier shows higher sensitivity and specificity than commonly used early warning signals under a wide range of noise intensities and rates of approach to the bifurcation. It also predicts the correct bifurcation in most cases, with particularly high accuracy for the period-doubling, Neimark-Sacker and fold bifurcations. Deep learning as a tool for bifurcation prediction is still in its nascence and has the potential to transform the way we monitor systems for critical transitions.

Many systems in nature and society possess critical thresholds at which the system undergoes an abrupt and significant change in dynamics^{1,2}. In physiology, the heart can spontaneously transition from a healthy to a dangerous rhythm³; in economics, financial markets can form a ‘bubble’ and crash into a recession⁴; and in ecology, ecosystems can collapse as a result of their interplay with human behaviour^{5,6}. These events, characterised by a sudden switch to a different dynamical regime, are referred to as critical transitions.

Critical transitions can be better understood with bifurcation theory^{7,8}, a branch of mathematics that studies how dynamical systems can undergo sudden qualitative changes as a parameter crosses a threshold (a bifurcation). Many bifurcations are accompanied by critical slowing down—a diminishing of the local stability of the system—which results in systematic changes to properties of a noisy time series, such as its variance, autocorrelation and power spectrum^{9–11}. These properties can be approximated analytically in the presence of

different bifurcations^{10,12–14}, and a corresponding observation in data can be used as an early warning signal (EWS) for the bifurcation¹¹. Systematic changes in variance and lag-1 autocorrelation have been observed prior to transitions in climate^{15–17}, geological¹⁸, ecological^{19,20} and cardiac²¹ systems, suggesting the presence of a bifurcation. However, these EWS have limited ability to predict the type of bifurcation^{14,22} and can fail in systems with nonsmooth potentials²³ or noise-induced transitions²⁴.

More recently, deep learning techniques have been employed to provide EWS for bifurcations^{25–27}. This involves training a neural network to classify a time series based on the type of bifurcation it is approaching, as well as appropriate controls^{25–27}. Unlike many applications of deep learning, this approach does not require abundant data from the study system, which, in the context of critical transitions, is often unavailable. (Unfortunately we do not have data from thousands or more ecosystems or climate systems that went through bifurcation).

¹Department of Physiology, McGill University, 3655 Promenade Sir William Osler Montreal, Canada. ²Department of Applied Mathematics, University of Waterloo, Waterloo, Canada. ³School of Environmental Sciences, University of Guelph, Guelph, Canada. ⁴These authors jointly supervised this work: Alvin Shrier, Gil Bub. ✉ e-mail: thomas.bury@mcgill.ca

Instead, the approach generates a massive library of simulation data from generic models that possess each type of bifurcation. The neural network then learns generic features associated with each type of bifurcation, that can be recognised in an unseen time series of the study system. This is enabled by the existence of universal properties of bifurcations that are manifested in time series as a dynamical system gets close to a bifurcation^{7,9}. In our previous work, we trained a deep learning classifier to provide an EWS for continuous-time bifurcations, and found it was effective at predicting transitions for real thermo-acoustic, climate and geological transitions²⁵.

Bifurcations can be partitioned according to whether they occur in continuous or discrete-time dynamical systems^{7,8}. This distinction is important, since discrete-time dynamical systems (difference equations) can display very different behaviour to their continuous-time counterparts (differential equations). As an example, consider the logistic model for population growth. When set up in continuous time (appropriate for populations with overlapping generations e.g. humans), the population grows smoothly as the reproduction rate increases. Whereas, when set up in discrete time (appropriate for populations with non-overlapping generations, e.g. insects), the population displays a spectrum of dynamics across parameter values, including stable points, stable cycles, and chaos²⁸. It is therefore important to develop EWS suitable for both continuous and discrete-time bifurcations. While indicators like variance and lag-1 autocorrelation can provide EWS for discrete-time bifurcations, the ability of a deep learning classifier at this task has not been investigated.

As well as in ecology, discrete-time bifurcations arise naturally in physiology³, epidemiology²⁹, and economics³⁰, where events can take place on a discrete timeline. To illustrate our approach, we will use model simulations from ecology, physiology and economics, as well as experimental data from spontaneously beating chick heart aggregates^{21,31}. Following administration of a drug, in some aggregates the time interval between two heart beats begins to alternate i.e. there is a period doubling bifurcation (Fig. 1). Such transitions can also occur for the human heart in the form of T-wave alternans, which increases a patient's risk for sudden cardiac death³². The period-doubling

bifurcation is accompanied by critical slowing down, so systematic changes in variance and lag-1 autocorrelation are expected and have been shown to provide an EWS in this system²¹. The chick heart aggregates serve as a good study system to test the performance of EWS since we have multiple recordings, not all of which underwent a transition, allowing us to test for false positives.

Among discrete-time bifurcations, there are many types, each with an associated change in dynamics⁷. For this study, we focus on the five local bifurcations of codimension-one (Supplementary Note 1). In being 'local', these bifurcations are accompanied by critical slowing down, so systematic changes and variance and autocorrelation are expected. However, not all of these bifurcations result in a critical transition²². They can instead involve a smooth transition to an intersecting steady state (transcritical) or to oscillations with gradually increasing amplitude (supercritical Neimark–Sacker). Predicting the type of bifurcation provides information on the nature of the dynamics following the bifurcation, something variance and autocorrelation alone do not provide.

Here, we train a deep learning classifier to provide a specific EWS for bifurcations of discrete-time dynamical systems. We train the classifier using simulation data of normal form equations appended with higher-order terms and noise. We then test the classifier on simulation runs of five discrete-time models used in cardiology, ecology and economics, and assess its performance relative to variance and lag-1 autocorrelation. We vary the noise amplitude and rate of forcing in model simulations to assess robustness of the EWS. Finally, we test the classifier on experimental data of spontaneously beating chick-heart aggregates that go through a period-doubling bifurcation. A reproducible run of all analyses may be performed on Code Ocean (<https://codeocean.com/capsule/2209652/tree/v2>) where the code is accompanied by the necessary software environment.

Results

Performance of classifiers on withheld test data

We train two different types of classifiers and use their ensemble average to make predictions. Classifier 1 is trained to recognise bifurcation trajectories based on middle portions of the time series, whereas Classifier 2 is trained on end portions (see Methods). In this way, Classifier 1 provides an earlier signal of a bifurcation and Classifier 2 provides a more specific signal, as more information is revealed closer to the bifurcation. To quantify performance of the classifiers, we use the F1 score, which is a combined measure of sensitivity (how many of the true positives were predicted correctly) and specificity (how many of the positive predictions were actually true positives). On the withheld test data, Classifier 1 and 2 achieved an F1 score of 0.66 and 0.85, respectively. On the simpler, binary classification problem of predicting whether or not there will be any bifurcation, the classifiers achieved an F1 score of 0.79 and 0.97, respectively. Classifier 2 has a higher performance as it has the easier task of classifying data closer to the bifurcation where fluctuations are more pronounced. Performance on individual bifurcation classes is shown by confusion matrices (Supplementary Fig. 1). The period-doubling, Neimark–Sacker and fold bifurcations are correctly classified with high sensitivity and specificity. On the other hand, the transcritical and pitchfork bifurcations are often mistaken for one another, likely due to having very similar normal forms (identical linear terms). Despite this, Classifier 2 can distinguish them at better than random, suggesting it is capable of recognising the different higher order terms in the data. From here onward, we report results using the ensemble prediction of the two classifiers, referred to collectively as the deep learning classifier.

Performance of EWS on theoretical models

We monitor variance, lag-1 autocorrelation and the deep learning classifier as progressively more of the time series is revealed. Variance and lag-1 autocorrelation are considered to provide an EWS if they

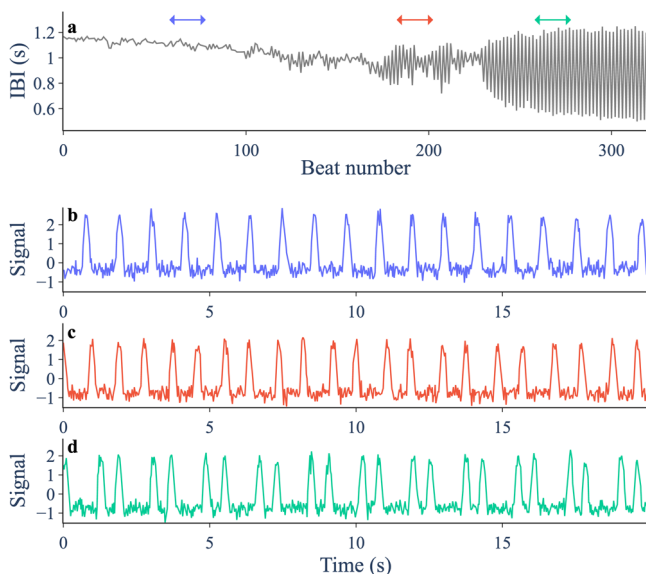


Fig. 1 | Period-doubling bifurcation in a spontaneously beating aggregate of embryonic chick heart cells following treatment with a potassium channel blocker (E-4031, 1.5 μmol). **a** Interbeat intervals (IBI) for consecutive beats. A period-doubling bifurcation occurs at approximately beat 230. Arrows mark intervals plotted in lower panels. **b–d** Normalised signal from optical imaging of the aggregate's motion. Traces are from a section well before (**b**), just before (**c**), and after (**d**) the period-doubling bifurcation.

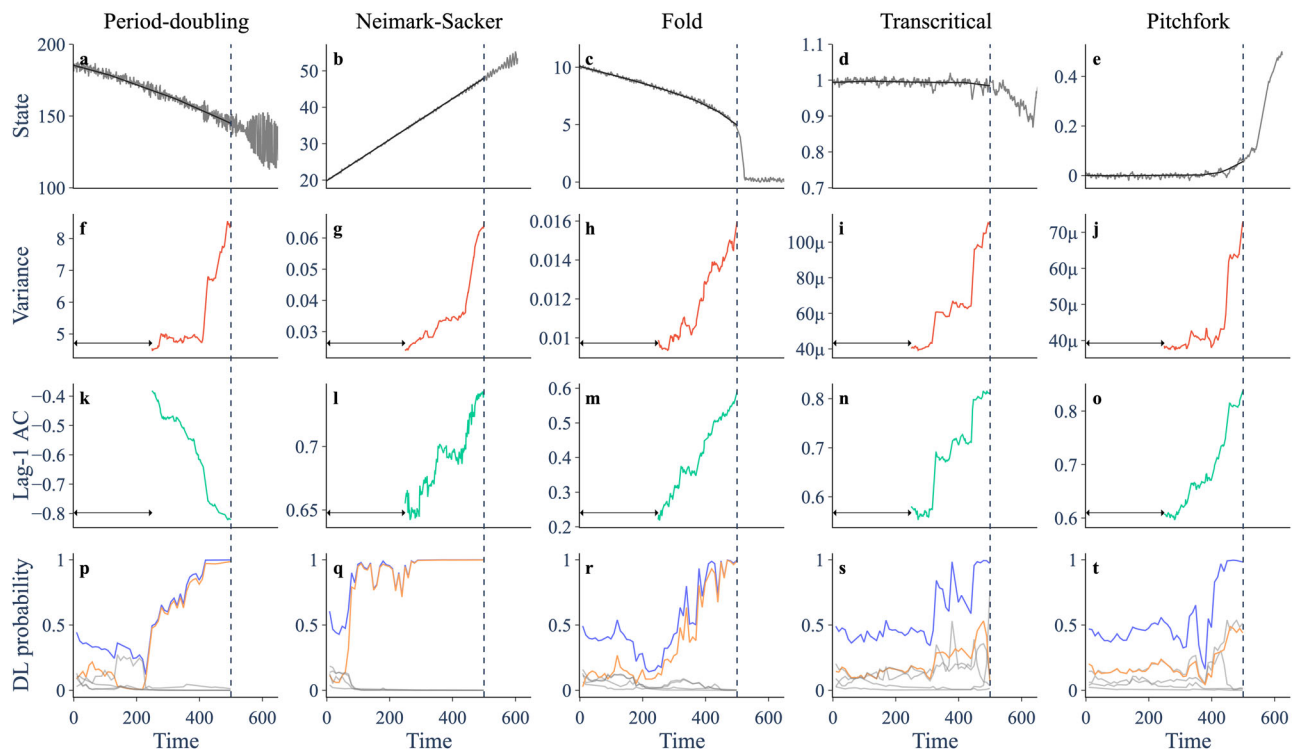


Fig. 2 | Trends in indicators prior to five different bifurcations in the theoretical models. a–e Trajectory (grey) and smoothing (black) of a simulation going through a period-doubling, Neimark–Sacker, fold, transcritical and pitchfork bifurcation, respectively. **f–j** Variance of residual dynamics after smoothing, computed over a rolling window (arrow) of size 0.5 times the length of the pre-transition data.

k–o Lag-1 autocorrelation. **p–t** Probabilities assigned by the deep learning (DL) classifier when given all preceding data. Orange shows probability assigned to the true bifurcation. Grey shows probabilities assigned to the other bifurcations. Blue shows the sum of the five bifurcation probabilities.

display a strong trend, which we quantify using the Kendall tau statistic. For each of the five theoretical models (Fig. 2(a–e)), we observe an increasing trend in variance (Fig. 2f–j), and an increasing or decreasing trend in lag-1 autocorrelation (Fig. 2k–o). The direction of the trend in lag-1 autocorrelation prior to a bifurcation depends on the frequency of oscillations (θ) at the bifurcation—equivalently the angle of the dominant eigenvalue in the complex plane (Supplementary Note 1). For $\theta \in [0, \pi/2)$ lag-1 autocorrelation increases, whereas for $\theta \in (\pi/2, \pi]$ it decreases—insights that can be obtained from analytical expressions of the autocorrelation function^{10,14}. The period-doubling bifurcation is characterised by $\theta = \pi$, and the Neimark–Sacker bifurcation shown here has $\theta \approx \pi/4$. The trends in variance and lag-1 autocorrelation therefore behave as expected and can be used as an EWS.

The deep learning classifier assigns a probability to each of the six possible outcomes (null, period-doubling, fold, Neimark–Sacker, transcritical and pitchfork). It is considered to provide an EWS when there is a heightening in the sum of the bifurcation probabilities (blue line, Fig. 2p–t). The type of bifurcation predicted is then taken as the highest individual bifurcation probability. For each simulation, the classifier becomes more confident of an approaching bifurcation as time goes on, and its assigned bifurcation probability for the true bifurcation increases. The period-doubling, Neimark–Sacker and fold bifurcations are identified with high confidence well before the transition. The transcritical and pitchfork bifurcations are assigned roughly equal probability on their respective time series, suggesting they are difficult to tell apart—an observation consistent with the classifier’s performance on its within-sample test data.

To obtain a measure of performance for the EWS, we need to test their predictions on both ‘forced’ time series (where a bifurcation is approached) and ‘null’ time series (where no bifurcation is approached). For the theoretical models, we generate null time series by

keeping the bifurcation parameter fixed. Sample null time series and their EWS are shown in Supplementary Fig. 3. We also test the robustness of the EWS to the rate of forcing and the noise amplitude of the simulations—two factors that have been shown to influence the performance of variance and lag-1 autocorrelation as an EWS^{33,34}. To this end, we simulate 100 forced and null time series at five different values of noise intensity and five different values of rate of forcing, resulting in a total of 5000 time series for each theoretical model. Sample trajectories illustrating the different noise amplitudes and rates of forcing are shown in Supplementary Fig. 4. We compute the probabilities assigned by the classifier and the Kendall tau values for variance and lag-1 autocorrelation at 80% of the way through the pretransition time series, and use these values as discrimination thresholds to construct ROC curves (Fig. 3a–e).

Using the AUC score (area under the ROC curve) as a measure of performance, we find that the classifier outperforms variance and lag-1 autocorrelation for each theoretical model. When evaluated for each combination of noise amplitude and rate of forcing separately, the classifier has the highest AUC score in 100% of cases for the Neimark–Sacker, fold, and pitchfork models, 84% of cases for the period-doubling model, and 80% of cases for the transcritical model (Supplementary Fig. 5). Similar to variance and lag-1 autocorrelation, the performance of the classifier is lower at higher rates of forcing. Noise amplitude affects performance differently depending on the model. In terms of predicting the correct bifurcation, the classifier typically performs better at slower rates of forcing (Supplementary Fig. 6) and was able to classify the period-doubling and Neimark–Sacker bifurcations to high accuracy at all noise amplitudes and rates of forcing considered. Finally, we evaluate the EWS for a range of parameter values in the period-doubling model that yield period-doubling bifurcations of different locations and morphology (Supplementary Fig. 7). We find that the deep learning classifier

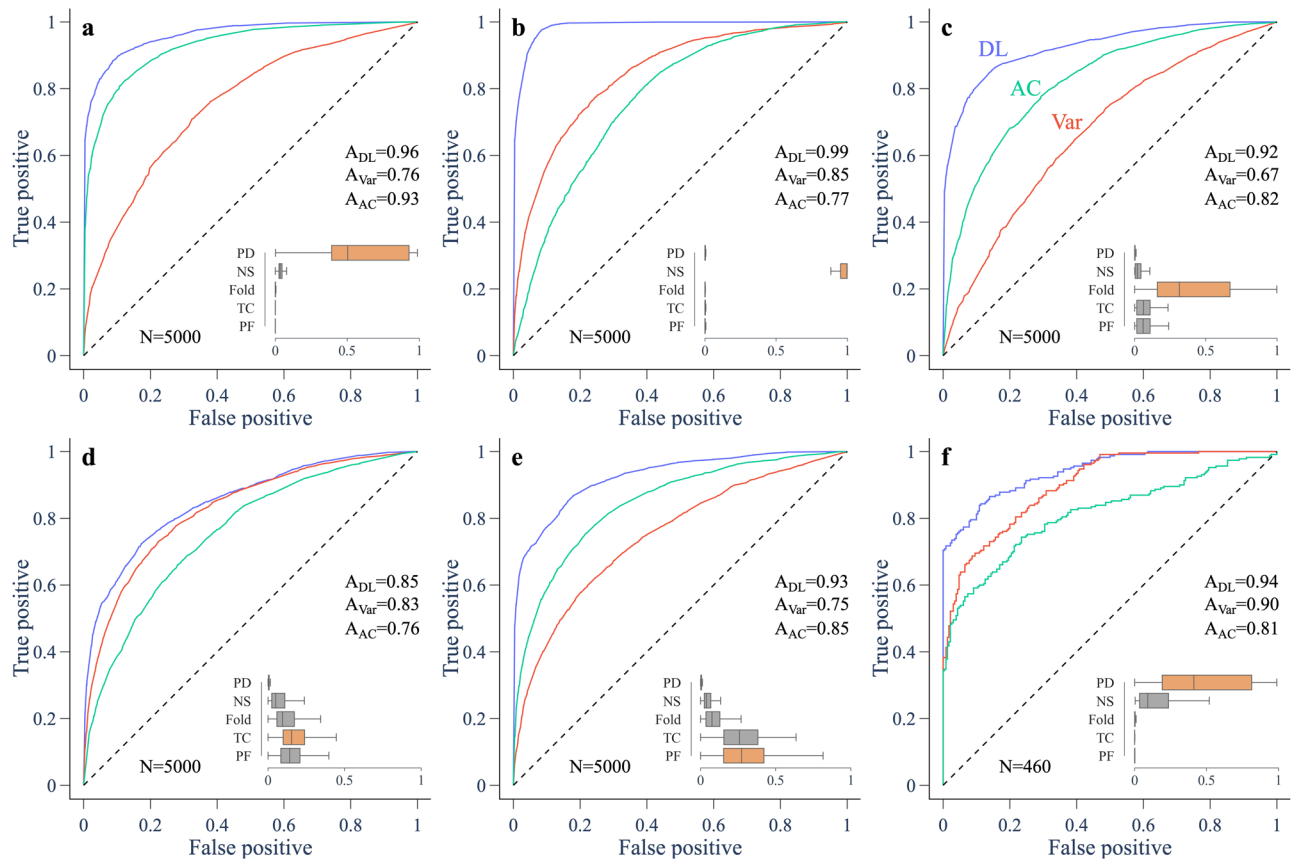


Fig. 3 | ROC curves for predictions of an upcoming transition in model and experimental data. ROC curves compare the performance of the deep learning classifier (DL, blue), variance (Var, red) and lag-1 autocorrelation (AC, green). For models (a–e), performance is assessed on 2500 forced and 2500 null simulations with different noise amplitudes and rates of forcing, with predictions made 80% of the way through the pretransition data. For experimental data (f), performance is assessed on 46 experimental runs, with 10 equally spaced predictions made between 60 and 100% of the way through the pre-transition data. The area under the curve (AUC), abbreviated to A, is a measure of performance. Insets show box plots for the probabilities assigned by the classifier to each type of bifurcation

(orange being the true bifurcation) among the trajectories approaching a transition. Box centre line is the median, box limits are the upper and lower quartiles, and whiskers capture the range up to 1.5 times the interquartile range. **a** Fox model going through a period-doubling bifurcation⁴⁹. **b** Westerhoff model going through a Neimark–Sacker bifurcation³⁰. **c** Ricker model going through a fold bifurcation⁵⁰. **d** Lotka–Volterra model going through a transcritical bifurcation⁵¹. **e** Lorenz model going through a pitchfork bifurcation⁵³. **f** Chick heart aggregates going through a period-doubling bifurcation³¹. PD: period-doubling. NS: Neimark–Sacker. TC: transcritical. PF: pitchfork.

outperforms variance and lag-1 autocorrelation in each case and correctly identifies the period-doubling bifurcation.

Performance of EWS on chick heart data

In the chick heart data, we mostly observe an increasing trend in variance and a decreasing trend in lag-1 autocorrelation prior to the period-doubling bifurcation, as previously reported²¹. This is consistent with analytical approximations for variance and lag-1 autocorrelation prior to a period-doubling bifurcation in a noisy dynamical system^{10,14}. The 46 records and their EWS are shown in Supplementary Figs. 8–11, and a sample of five period-doubling records are shown in Fig. 4. The classifier correctly predicts a period-doubling bifurcation in 16 of the 23 period-doubling records. In other cases, it incorrectly predicts a Neimark–Sacker bifurcation (e.g. Fig. 4e). This seems to be linked to an early increase in lag-1 autocorrelation, perhaps caused by a non-monotonic approach to the period-doubling bifurcation. For predictions made at 60–100% of the way through the chick heart data, the classifier obtains the highest AUC score (Fig. 3f), a slight improvement on variance, with the advantage of also providing the bifurcation type. We find this result is robust to smoothing method (Gaussian or Lowess), a range of different smoothing parameters, different rolling window sizes for variance and lag-1 autocorrelation, and sample error in the

experimental data (Supplementary Figs. 12–15). However, smoothing with a bandwidth that is too small diminishes the ability of the classifier to identify a period-doubling bifurcation, presumably since fluctuations that enable identification of the bifurcation type are being removed.

Discussion

Many systems that evolve on a discrete timeline can undergo a sudden change in dynamics via a discrete-time bifurcation. We have found that a deep learning classifier is an effective tool for predicting discrete-time bifurcations in systems with a range of noise levels and rates of approach to the bifurcation. The classifier provides higher sensitivity and specificity than variance and lag-1 autocorrelation—two commonly used EWS for bifurcations. Moreover, the classifier provides early indication of the type of bifurcation—an important piece of information given the qualitatively different dynamics associated with each bifurcation. A reliable early warning signal that specifies the type of bifurcation will help us prevent harmful bifurcations (e.g. dangerous heart rhythms³) and promote favourable transitions (e.g. ecosystem recovery³⁵).

It may be possible to design a deep learning classifier that achieves a higher performance on our test data. First, there are many neural network architectures that could be investigated. For example,

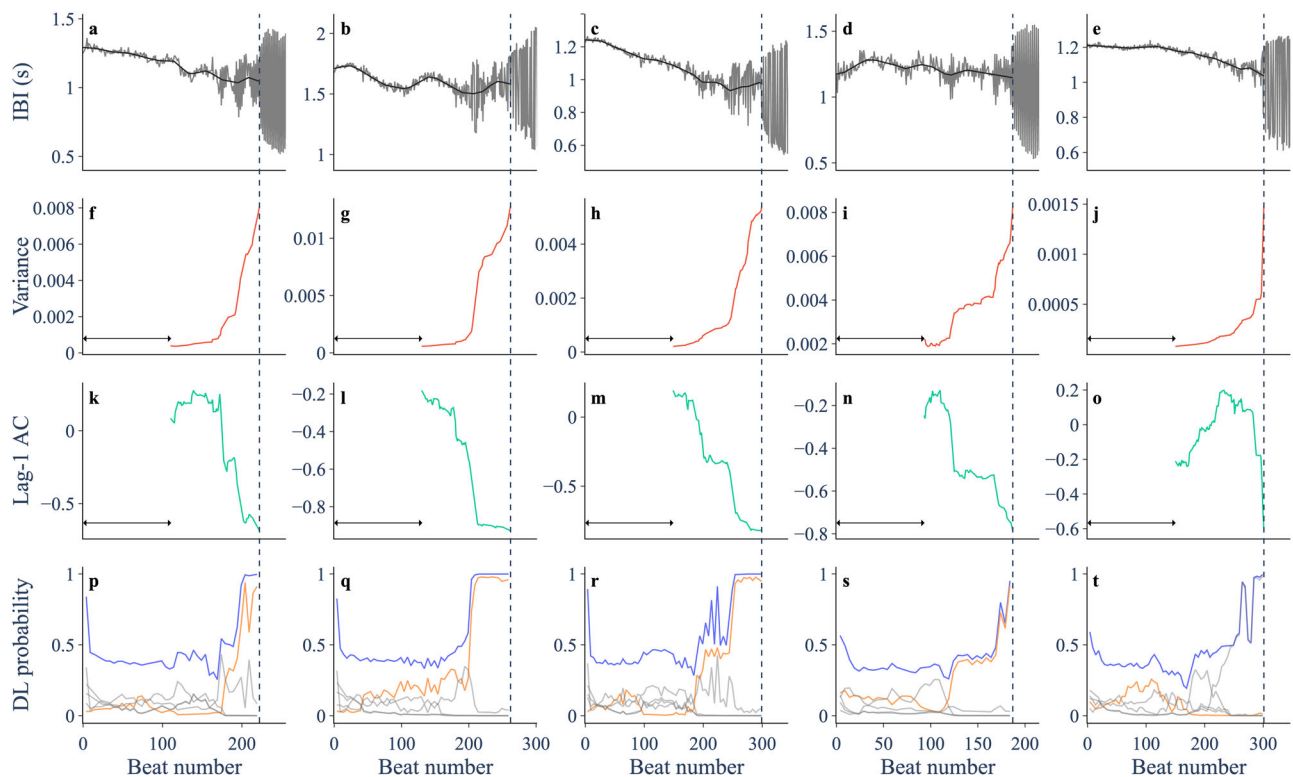


Fig. 4 | Trends in indicators prior to a period-doubling bifurcation in five chick heart aggregates treated with a potassium channel blocker. a–e Inter-beat interval (IBI) trajectory (gray) and smoothing (black). **f–j** Variance of residual dynamics after smoothing, computed over a rolling window (arrow) of size 0.5

times the length of the pre-transition data. **k–o** Lag-1 autocorrelation. **p–t** Probabilities assigned by the deep learning (DL) classifier to the period-doubling bifurcation (orange) and the other bifurcations (gray). Blue shows the sum of the five bifurcation probabilities.

transformers, which are the current state-of-the-art for language models like GPT³⁶, may also be useful for time series classification³⁷. Second, the hyperparameters of the classifier could be systematically tuned to optimise performance. Third, there may be benefit to reframing bifurcation prediction as a hierarchical classification problem³⁸. One classifier could address the binary problem of flagging an approaching bifurcation, and a second classifier could address the multi-class problem of classifying the type of bifurcation given that a bifurcation is approaching. (The same way one might want to distinguish images of dogs and cats before attempting to classify dog breeds). Finally, performance could be improved by training a larger ensemble of classifiers³⁹.

For a classifier to be effective, it must be trained on sufficiently diverse training data. As such, the method by which training data is obtained needs careful consideration. Our previous work on continuous-time bifurcations obtained training data from randomly generated dynamical systems with polynomial terms²⁵ and labelled the data using the bifurcation continuation software AUTO. This approach is appealing as it imposes relatively few restrictions on the models that are generated, and may include features associated with higher-order terms. Here, we opted for a more restricted approach that uses normal form models to generate the training data. This method has the advantage of being faster computationally, since the location and type of bifurcation in the model is known a priori. It also alleviates the need to detrend the training data, which can make the classifier reliant on receiving data that has been detrended using a specific method⁴⁰. We have found that even with this more restricted training data, a classifier can generalise to detecting bifurcations in more complex model and empirical systems.

An important consideration in building a training library for a bifurcation predictor is how to define a ‘null’ trajectory. We opted for a

simple approach that uses model simulations with a fixed bifurcation parameter, where the bifurcation parameter is sampled randomly from values that yield $|\lambda| < 0.8$, where λ is the eigenvalue of the Jacobian matrix. Larger values of λ result in a significant portion of simulations going through noise-induced transitions, and therefore not deemed appropriate. Upon investigating how the classifier performs on null trajectories specifically, we find that it is more confident in its prediction for null trajectories that are longer, and further away from the bifurcation (Supplementary Fig. 16), as seems logical. A useful extension to the training data could be a richer set of null trajectories, where the bifurcation parameter is allowed to move around, perhaps stochastically, as one would expect in real systems.

We trained a classifier to provide EWS for a subset of bifurcations, namely local, codimension-one, discrete-time bifurcations. While these bifurcations are present in many systems of interest, the real world presents many other classes of bifurcation in both continuous and discrete-time, including global bifurcations (e.g. homoclinic and heteroclinic), codimension-two bifurcations (e.g. cusp and Bogdanov-Takens), and bifurcations of attractors. For systems on attractors that explore a large portion of their phase space, empirical dynamical modelling⁴¹, reservoir computing^{42,43} and deep neural networks⁴⁴ can be used to make forecasts that may help predict critical transitions. In cases where spatial information is available, concepts from statistical physics may be useful⁴⁵, particularly in combination with deep learning²⁷.

A limitation of the present classifier is that it is only trained to predict discrete-time bifurcations. Therefore, one needs to know ahead of time whether continuous or discrete time is a better description for the system. In the case of the chick heart cells, we had prior knowledge that they are well described by a discrete-time dynamical system⁴⁶, and therefore appropriate for the classifier. An

interesting avenue for future research is to build a classifier that works for both continuous and discrete-time bifurcations. This may be achieved by generating a training library from models with a range of discretised timesteps, from very large steps that generate discrete-time bifurcations, down to the limit of a discrete timestep of zero, where continuous-time bifurcations occur. With a large enough training set, one would not need to assume ahead of time whether continuous or discrete time is a better description for the system.

Our results demonstrate that combining dynamical system and deep learning methodologies can provide EWS for critical transitions that are both more reliable and more descriptive than non-hybrid approaches. This study has set a baseline for prediction performance across a variety of popular discrete-time models and an experimental data set. In providing a code capsule that reproduces this study, we hope to facilitate the development, testing and comparison of related methods. In particular, the development of interpretable (as opposed to ‘black box’) models that achieve a similar performance would be highly desirable, especially in safety critical domains⁴⁷, although this will likely require new analytical and algorithmic insights. Techniques such as deconvolutional networks⁴⁸ make it possible to map the learned space of a deep learning algorithm back onto the original temporal dataset. This allows one to visualise the features that the algorithm is using to make its decision, which could serve as a starting point for an interpretable model. Building a universal predictor for critical transitions is not a job for a single research team⁴⁴, and will benefit from a variety of approaches and open source code. Depending on context, critical transitions can be devastating or highly desirable. Improved EWS would allow us to better prevent or promote such transitions.

Methods

Generation of training data for the deep learning classifier

Training data consists of simulation data from a library of 50,000 models. The models are generated at random from five different model frameworks, each possessing one of the bifurcations studied (period-doubling, Neimark–Sacker, fold, transcritical, pitchfork). The models are composed of the normal form of the bifurcation⁷, higher-order polynomial terms up to degree 10 with coefficients drawn from a normal distribution, and additive Gaussian white noise (ϵ_t) with amplitude (σ) drawn from a uniform distribution. In each case, the bifurcation occurs at $\mu = 0$.

The model for the period-doubling bifurcation is

$$x_{t+1} = -(1 + \mu)x_t \pm x_t^3 + \sum_{i=4}^{10} \alpha_i x_t^i + \sigma \epsilon_t, \tag{1}$$

where $\alpha_i \sim \mathcal{N}(0, 1)$. The positive (negative) cubic term yields a supercritical (subcritical) bifurcation, and is chosen at random. The model for the Neimark–Sacker bifurcation is

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = (1 + \mu)R(\theta) \begin{pmatrix} x_t \\ y_t \end{pmatrix} \pm (x_t^2 + y_t^2)R(\theta) \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \sum_{i=4}^{10} \sum_{j=0}^i \begin{pmatrix} \alpha_{ij} \\ \beta_{ij} \end{pmatrix} x_t^{i-j} y_t^j + \sigma \begin{pmatrix} \epsilon_t^{(1)} \\ \epsilon_t^{(2)} \end{pmatrix} \tag{2}$$

where $\alpha_{ij}, \beta_{ij} \sim \mathcal{N}(0, 1)$, $R(\theta)$ is the rotation matrix

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \tag{3}$$

and $\theta \sim \mathcal{U}[0, \pi]$ is the angular frequency of oscillations at the bifurcation. The positive (negative) cubic term yields a subcritical (supercritical) bifurcation, and is chosen at random. The model for the fold

bifurcation is

$$x_{t+1} = -\mu + x_t - x_t^2 + \sum_{i=3}^{10} \alpha_i (x_t - \sqrt{-\mu})^i + \sigma \epsilon_t, \tag{4}$$

where $\alpha_i \sim \mathcal{N}(0, 1)$. The model for the transcritical bifurcation is

$$x_{t+1} = (1 + \mu)x_t - x_t^2 + \sum_{i=3}^{10} \alpha_i x_t^i + \sigma \epsilon_t, \tag{5}$$

where $\alpha_i \sim \mathcal{N}(0, 1)$. Finally, the model for the pitchfork bifurcation is

$$x_{t+1} = (1 + \mu)x_t \pm x_t^3 + \sum_{i=4}^{10} \alpha_i x_t^i + \sigma \epsilon_t, \tag{6}$$

where $\alpha_i \sim \mathcal{N}(0, 1)$. The positive (negative) cubic term yields a subcritical (supercritical) bifurcation, and is chosen at random.

The library is composed of 10,000 models from each framework. For each model, we run a ‘forced’ simulation where the bifurcation parameter μ is increased linearly across the interval $[\mu_0, 0]$, and a ‘null’ simulation where μ is fixed at μ_0 . The initial value for the bifurcation parameter μ_0 is drawn from a uniform distribution across all values that correspond to $|\lambda| < 0.8$, where λ is the eigenvalue of the Jacobian matrix in the model. This ensures that the training data contains simulations that start close to and far from a bifurcation. For the period-doubling, Neimark–Sacker, transcritical and pitchfork models, this means drawing μ_0 from $\mathcal{U}[-1.8, -0.2]$. For the fold bifurcation, this means drawing μ_0 from $\mathcal{U}[-0.9, -0.1]$. After a burn-in period of 100 iterations, we simulate each model for 600 iterations and keep the last 500 data points, or the 500 data points immediately preceding a transition if one occurs. We define a transition as a time when the deviation from equilibrium exceeds ten times the noise amplitude σ . We simulate one forced and one null simulation from each model, resulting in 50,000 forced and 50,000 null trajectories. To balance the number of entries for each class, we take 10,000 null simulations at random, resulting in a total of 60,000 entries in the training data set. Example trajectories for each class are shown in Supplementary Fig. 2.

Architecture and training of the deep learning classifier

We use a neural network with a CNN-LSTM architecture and hyperparameters as in ref. 25. This consists of a single convolutional layer with max pooling followed by two LSTM layers with dropout followed by a dense layer that maps to a vector of probabilities over the six possible classes. For training, we use Adam optimisation with a learning rate of 0.0005, a batch size of 1024, and sparse categorical cross entropy as the loss function. We use a training/validation/test split of 0.95/0.025/0.025. We found 200 epochs was sufficient to obtain optimal accuracy on the validation set.

To expose the classifier to time series of different lengths, we censor each time series in the training data. We train two classifiers independently using different censoring techniques. Classifier 1 is trained on time series censored at the beginning and the end, forcing it to learn from data in the middle of the time series. Classifier 2 is trained on time series only censored at the beginning, allowing it to learn from data right up to the bifurcation. The length for each censored time series L is drawn from $\mathcal{U}[50, 500]$. Then, for Classifier 1, the start time of the censored time series $t_0 \sim \mathcal{U}[0, 500 - L]$ and for Classifier 2, $t_0 = 500 - L$. The censored time series are then normalised by their mean absolute value and prepended with zeros to make them 500 points in length. We report results using the average prediction of the two classifiers.

Theoretical models

To test the deep learning classifier on out-of-sample data, we simulate a variety of nonlinear, discrete-time models, each containing one of the studied bifurcations. To account for stochasticity, we include additive Gaussian white noise. We run forced simulations, where the bifurcation parameter is increased linearly up to the bifurcation point, and null simulations, where the bifurcation parameter remains fixed. To create a diverse set of test data, we vary the noise amplitude (σ) and rate of forcing (rate of change of the bifurcation parameter). We run 100 forced and null simulations at five different noise amplitudes and five different rates of forcing, resulting in 5000 simulations of each model. Values for the noise amplitude are on a logarithmic scale and values for the rate of forcing result in time series of length 100, 200, 300, 400, and 500. Sample simulations for each model at different noise amplitude and rate of forcing are shown in Supplementary Fig. 4. The transition time for each forced simulation is taken as the moment when the bifurcation parameter crosses the bifurcation, or the moment when the state variable crosses a threshold, if specified.

Fox model. To test the detection of a period-doubling bifurcation, we use a model of cardiac alternans⁴⁹ with additive Gaussian white noise. This is given by

$$D_{n+1} = (1 - \alpha M_{n+1}) \left(A + \frac{B}{1 + e^{-(I_n - C)/D}} \right) + \sigma \epsilon_n, \tag{7}$$

$$M_{n+1} = e^{-I_n/\tau} [1 + (M_n - 1)e^{-D_n/\tau}], \tag{8}$$

$$I_n = T - D_n, \tag{9}$$

where D_n is the action potential duration of the n th beat, M_n is a memory variable, I_n is the rest duration following the action potential, T is the stimulation period, τ is the time constant of accumulation and dissipation of memory, α is the influence of memory on the action potential duration, and A, B, C and D are parameters governing the shape of the restitution curve. Following⁴⁹, we take $A = 88, B = 122, C = 40, D = 28, \tau = 180, \alpha = 0.2$, which give dynamics in good agreement with a complex ionic model. This yields a period-doubling bifurcation at approximately $T = 200$. Forced simulations are run with T decreasing linearly on the interval [300,150] and null simulations are run with $T = 300$. Values for noise amplitude are $0.1 \times \{2^0, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}\}$.

To test the robustness of EWS to different model parameter values, we simulate trajectories with different values of α and a multiplicative scaling factor of A, B, C and D (Supplementary Fig. 7). In each case, we simulate 100 forced and null trajectories for 300 time steps and a noise amplitude of 0.1. Forced simulations are run with T decreasing linearly from 300 to the bifurcation point and null simulations are run with $T = 300$.

Westerhoff model. To test detection of a Neimark–Sacker bifurcation, we use a simple model of business cycles based on consumer sentiment³⁰ with additive Gaussian noise. This is given by

$$Y_t = a + (b - d)Y_{t-1} + dY_{t-2} + \frac{cY_{t-1}}{1 + \text{Exp}[-(Y_{t-1} - Y_{t-2})]} + \sigma \epsilon_t, \tag{10}$$

where Y_t is the national income at time step t , a is the level of autonomous expenditures of agents, b and c govern a curve that determines the fraction of income consumed by the agents, and d is the policymaker’s control parameter to offset income trends. We take $b = 0.45, c = 0.1$, and $d = 0.2$, which yields a Neimark–Sacker bifurcation at $a = 24$ corresponding to the onset of business cycles. Forced simulations are run with a increasing linearly on the interval [10,27] and null

simulations are run with $a = 10$. Values for noise amplitude are $0.1 \times \{2^0, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}\}$.

Ricker model. To test detection of a fold bifurcation, we use the Ricker model⁵⁰ with a sigmoidal harvesting term and additive Gaussian noise. This is given by

$$x_{t+1} = x_t e^{r(1-x_t/k)} - F \frac{x_t^2}{x_t^2 + h^2} + \sigma \epsilon_t, \tag{11}$$

where x_t is the population size at time step t , r is the intrinsic growth rate, k is the carrying capacity, F is the harvesting rate, and h governs the steepness of the sigmoidal harvesting term. We take $r = 0.75, k = 10, h = 0.75$, which yields a fold bifurcation at $F = 2.36$. Forced simulations are run with F increasing linearly on the interval [0, 3.54] and null simulations are run with $F = 0$. We define a transitions as the time when x_t drops below 0.45. Values for noise amplitude are $0.2 \times \{2^0, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}\}$.

Lotka–Volterra model. To test detection of a transcritical bifurcation, we use the discrete-time analogue of the Lotka–Volterra model, first studied by Maynard Smith⁵¹. This system is especially relevant to arthropod predator-prey and host-parasitoid interactions. The rescaled equations⁵² are

$$x_{t+1} = (r+1)x_t - rx_t^2 - cx_t y_t + \sigma \epsilon_t^{(1)}, \tag{12}$$

$$y_{t+1} = cx_t y_t + \sigma \epsilon_t^{(2)}, \tag{13}$$

r relates to the growth rate of the prey (x_t), and c relates to the foraging efficiency of the predator (y_t). We take $r = 0.5$, which yields a transcritical bifurcation at $c = 1$, the critical foraging efficiency beyond which the predator population can sustain themselves. Forced simulations are run with c increasing linearly on the interval [0.5,1.25] and null simulations are run with $c = 0.5$. We look for early warning signals in the prey population. Values for noise amplitude are $0.01 \times \{2^0, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}\}$.

Lorenz model. To test detection of a pitchfork bifurcation, we use the reduced discrete Lorenz system, which was first introduced as a demonstration of computational chaos⁵³. This is given by

$$x_{t+1} = (1 + ah)x_t - hx_t y_t + \sigma \epsilon_t^{(1)} \tag{14}$$

$$y_{t+1} = (1 - h)y_t + hx_t^2 + \sigma \epsilon_t^{(2)} \tag{15}$$

where state variables and parameters are derived from the full Lorenz equations⁵³. We take $h = 0.5$, which yields a pitchfork bifurcation at $a = 0$. Forced simulations are run with a increasing linearly over the interval [−1, 0.25] and null simulations are run with $a = -1$. We look for early warning signals in x_t . Values for noise amplitude are $0.01 \times \{2^0, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}\}$.

Experiments with embryonic chick heart cell aggregates

Experiments were carried out in accordance with the ethical and Health and Safety regulations at McGill University. Aggregates were prepared using the method proposed by DeHaan⁵⁴. Ventricles of 7-day-old White Leghorn chicken embryo hearts were dissected and dissociated into single cells by trypsinization. The cells were then added to Erlenmeyer flasks containing a culture medium (818A) gassed with 5% CO₂, 10% O₂, 85% N₂ (pH = 7.4), and placed on a gyratory shaker for 24–48 hours at 37 °C. This generated aggregates with a diameter of approximately 100–200 μm that displayed a beating pattern of period

approximately 1–2 s. Experiments were conducted 2–6 h after the aggregates were plated, in dishes maintained at 37 °C.

The aggregates were treated with 0.5–2.5 μmol of E4031, a drug that blocks the human Ether-à-go-go-Related Gene (hERG) potassium channel⁵⁵. The beating of the aggregates was recorded using phase-contrast imaging sampled at 40 Hz using a CCD camera (NeuroCD-SM; RedShirtImaging, LLC) at an 80 × 80 pixel spatial resolution, focussing on light-intensity variation at the edge of the aggregate³¹. There are periodic stops in the recording (for 2–3 min) for data storage purposes. The light signal for each aggregate is processed through a band-pass filter (cutoff frequencies: 0.1–6.5 Hz). The timing of each beat is then determined as the moment when the signal passes a threshold (the mean of the record plus 0.7 times the standard deviation) with positive slope. The interbeat intervals are computed as the time between consecutive beats, and used in the analysis of this study.

We define the onset of the period-doubling bifurcation as the first time when the slope of a linear regression of the return map composed of a sliding window of interbeat intervals is below -0.95 for the next 10 beats. According to this definition, 43 of the 119 aggregates underwent period-doubling bifurcations. The remaining aggregates either went through no qualitative change in dynamics (18), or underwent a transition to more complex dynamics including irregular rhythms and bursting oscillations (58). Of the period-doubling aggregates, we captured the onset of the period-doubling bifurcation for 23 of them (Sup. Figs. 8, 9). The other period-doubling bifurcations were missed due to pauses in the recording. From the 18 aggregates that undergo no qualitative change, we extract 23 segments at random with a random length between 100 and 500 to serve as null time series (Supplementary Figs. 10, 11). Predictions are made at 10 equally spaced time points between 60–100% of the way through the 23 period-doubling (pre-bifurcation) and 23 null time series.

Computing and assessing the performance of EWS

EWS are computed using the Python package `ewstools`⁵⁶. This involves first detrending the (pretransition) time series. For the model simulations, we use a Lowess filter with a span of 0.25 the length of the data. For the heart cell data, we use a Gaussian filter with a bandwidth of 20 beats. Variance and lag-1 autocorrelation are then computed over a rolling window of 0.5, which had higher performance than a rolling window of 0.25. The deep learning predictions at a given point in the time series are obtained by taking the preceding data from the time series, normalising it, prepending it with zeroes to make it 500 points in length, and feeding it into the classifier.

To compare performance of variance, lag-1 autocorrelation and the deep learning classifier, we use the AUC (area under curve) score of the ROC curve. The ROC curve plots the true positive rate vs. the false positive rate as a discrimination threshold is varied. We use the Kendall τ value as the discrimination threshold for variance and lag-1 autocorrelation, and the sum of the bifurcation probabilities for the discrimination threshold of the deep learning classifier.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The chick heart data and the simulated data used to train and test the deep learning classifier have been deposited on Code Ocean <https://codeocean.com/capsule/2209652/tree/v2>⁵⁷.

Code availability

Code and instructions to reproduce the analysis are available at the GitHub repository https://github.com/ThomasMBury/dl_discrete_bifurcation. A reproducible run can be performed on Code Ocean at

<https://codeocean.com/capsule/2209652/tree/v2> where the code is accompanied by a compatible software environment⁵⁷.

References

- Scheffer, M. *Critical Transitions In Nature And Society* (Princeton University Press, 2020).
- Levin, S. A. Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* **1**, 431–436 (1998).
- Glass, L. & Mackey, M. C. *From Clocks to Chaos* (Princeton University Press, 2020).
- Sornette, D. *Why Stock Markets Crash* (Princeton University Press, 2017).
- Barlow, Lee-Ann, Cecile, J., Bauch, C. T. & Anand, M. Modelling interactions between forest pest invasions and human decisions regarding firewood transport restrictions. *PLoS ONE* **9**, e90511 (2014).
- Henderson, K. A., Bauch, C. T. & Anand, M. Alternative stable states and the sustainability of forests, grasslands, and agriculture. *Proc. Natl Acad. Sci.* **113**, 14552–14559 (2016).
- Kuznetsov, Y. *Elements of Applied Bifurcation Theory* (Springer, 1998).
- Strogatz, S. H. *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, 2018).
- Wissel, C. A universal law of the characteristic return time near thresholds. *Oecologia* **65**, 101–107 (1984).
- Wiesenfeld, K. Noisy precursors of nonlinear instabilities. *J. Statistical Phys.* **38**, 1071–1097 (1985).
- Scheffer, M. et al. Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
- Kuehn, C. A mathematical framework for critical transitions: normal forms, variance and applications. *J. Nonlinear Sci.* **23**, 457–510 (2013).
- O'Regan, S. M. & Burton, D. L. How stochasticity influences leading indicators of critical transitions. *Bull. Math. Biol.* **80**, 1630–1654 (2018).
- Bury, T. M., Bauch, C. T. & Anand, M. Detecting and distinguishing tipping points using spectral early warning signals. *J. Roy. Soc. Interface* **17**, 20200482 (2020).
- Dakos, V. et al. Slowing down as an early warning signal for abrupt climate change. *Proc. Natl Acad. Sci.* **105**, 14308–14312 (2008).
- Boers, N. Early-warning signals for Dansgaard-Oeschger events in a high-resolution ice core record. *Nat. Commun.* **9**, 2556 (2018).
- Boers, N. Observation-based early-warning signals for a collapse of the Atlantic meridional overturning circulation. *Nat. Clim. Change* **11**, 680–688 (2021).
- Hennekam, R. et al. Early-warning signals for marine anoxic events. *Geophys. Res. Lett.* **47**, e2020GL089183 (2020).
- Pace, M. L. et al. Reversal of a cyanobacterial bloom in response to early warnings. *Proc. Natl Acad. Sci.* **114**, 352–357 (2016).
- Wang, R. et al. Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature* **492**, 419–422 (2012).
- Quail, T., Shrier, A. & Glass, L. Predicting the onset of period-doubling bifurcations in noisy cardiac systems. *Proc. Natl Acad. Sci.* **112**, 9358–9363 (2015).
- Kéfi, S., Dakos, V., Scheffer, M., Van Nes, E. H. & Rietkerk, M. Early warning signals also precede non-catastrophic transitions. *Oikos* **122**, 641–648 (2013).
- Hastings, A. & Wysham, D. B. Regime shifts in ecological systems can occur with no warning. *Ecol. Lett.* **13**, 464–472 (2010).
- Ditlevsen, P. D. & Johnsen, S. J. Tipping points: early warning and wishful thinking. *Geophys. Res. Lett.* **37** (2010).
- Bury, T. M. et al. Deep learning for early warning signals of tipping points. *Proc. Natl Acad. Sci.* **118**, e2106140118 (2021).
- Deb, S., Sidheekh, S., Clements, C. F., Krishnan, N. C. & Dutta, P. S. Machine learning methods trained on simple models can predict

- critical transitions in complex natural systems. *Roy. Soc. Open Sci.* **9**, 211475 (2022).
27. Dylewsky, D. et al. Universal early warning signals of phase transitions in climate systems. *J. R. Soc. Interface.* **20**, 20220562 (2023)
 28. May, R. M. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. *Science* **186**, 645–647 (1974).
 29. Allen, Linda J. S. Some discrete-time SI, SIR, and SIS epidemic models. *Math. Biosci.* **124**, 83–105 (1994).
 30. Westerhoff, F. H. Consumer sentiment and business cycles: a Neimark–Sacker bifurcation scenario. *Appl. Econom. Lett.* **15**, 1201–1205 (2008).
 31. Kim, Min-Young et al. Stochastic and spatial influences on drug-induced bifurcations in cardiac tissue culture. *Phys. Rev. Lett.* **103**, 058101 (2009).
 32. Verrier, R. L. et al. Microvolt t-wave alternans: physiological basis, methods of measurement, and clinical utility—consensus guideline by international society for holter and noninvasive electrocardiology. *J. Am. College Cardiol.* **58**, 1309–1324 (2011).
 33. Clements, C. F. & Ozgul, A. Rate of forcing and the forecastability of critical transitions. *Ecol. Evol.* **6**, 7787–7793 (2016).
 34. Pavithran, I. & Sujith, R. I. Effect of rate of change of parameter on early warning signals for critical transitions. *Chaos: Interdisciplinary J. Nonlinear Sci.* **31**, 013116 (2021).
 35. Clements, C. F., McCarthy, M. A. & Blanchard, J. L. Early warning signals of recovery in complex systems. *Nat. Commun.* **10**, 1681 (2019).
 36. Vaswani, A. et al. Attention is all you need. *Advances in Neural Information Processing Systems* **30**, (2017).
 37. Wen, Q. et al. Transformers in time series: a survey. Preprint at <https://arxiv.org/abs/2202.07125> (2022).
 38. Silla, C. N. & Freitas, A. A. A survey of hierarchical classification across different application domains. *Data Mining Knowledge Discov.* **22**, 31–72 (2011).
 39. Polikar, R. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* **6**, 21–45 (2006).
 40. Dablander, F. & Bury, T. M. Deep learning for tipping points: pre-processing matters. *Proc. Natl Acad. Sci.* **119**, e2207720119 (2022).
 41. Ye, H. et al. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proc. Natl Acad. Sci.* **112**, E1569–E1576 (2015).
 42. Patel, D. & Ott, E. Using machine learning to anticipate tipping points and extrapolate to post-tipping dynamics of non-stationary dynamical systems. *Chaos: Interdisciplinary J. Nonlinear Sci.* **33**, 023143 (2023).
 43. Kong, Ling-Wei, Fan, Hua-Wei, Grebogi, C. & Lai, Ying-Cheng Machine learning prediction of critical transition and system collapse. *Phys. Rev. Res.* **3**, 013090 (2021).
 44. Lapeyrolerie, M. & Boettiger, C. Teaching machines to anticipate catastrophes. *Proc. Natl Acad. Sci.* **118**, e2115605118 (2021).
 45. Hagstrom, G. I. & Levin, S. A. Phase Transitions and the Theory of Early Warning Indicators for Critical Transitions. *How Worlds Collapse: What History, Systems, and Complexity Can Teach Us About Our Modern World and Fragile Future*, **358**, (2023)
 46. Quail, T. et al. Chaotic dynamics in cardiac aggregates induced by potassium channel block. *Chaos: Interdisciplinary J. Nonlinear Sci.* **22**, 033140 (2012).
 47. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
 48. Zeiler, M. D. & Fergus, R. Visualizing and Understanding Convolutional Networks. In *Computer Vision – ECCV 2014: 13th European Conference Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*. (eds Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) 818–833 (Springer, 2014).
 49. Fox, J. J., Bodenschatz, E. & Gilmour Jr, R. F. Period-doubling instability and memory in cardiac tissue. *Phys. Rev. Lett.* **89**, 138101 (2002).
 50. Ricker, William Edwin Stock and recruitment. *J. Fisheries Board Canada* **11**, 559–623 (1954).
 51. Smith, J. M. *Mathematical Ideas in Biology* (CUP Archive, 1968).
 52. Neubert, M. G. & Kot, M. The subcritical collapse of predator populations in discrete-time predator-prey models. *Math. Biosci.* **110**, 45–66 (1992).
 53. Lorenz, E. N. Computational chaos—a prelude to computational instability. *Phys. D: Nonlinear Phenom.* **35**, 299–317 (1989).
 54. DeHaan, R. L. Regulation of spontaneous activity and growth of embryonic chick heart cells in tissue culture. *Dev. Biol.* **16**, 216–249 (1967).
 55. Clay, J. R., Kristof, A. S., Shenasa, J., Brochu, R. M. & Shrier, A. A review of the effects of three cardioactive agents on the electrical activity from embryonic chick heart cell aggregates: TTX, ACh, and E-4031. *Prog. Biophys. Mol. Biol.* **62**, 185–202 (1994).
 56. Bury, T. M. ewstools: a Python package for early warning signals of bifurcations in time series data. *J. Open Source Softw.* **8**, 5038 (2023).
 57. Bury, T. M. et al. Predicting discrete-time bifurcations with deep learning [Source Code]. *Code Ocean* <https://doi.org/10.24433/CO.3359094.v2> (2003).

Acknowledgements

T.M.B. is supported by a Fonds de Recherche du Québec—Nature et technologies (FRQNT) postdoctoral fellowship. G.B. acknowledges support from the Heart and Stroke Foundation of Canada (RGPIN2018-05346) and the Natural Sciences and Engineering Research Council (NSERC) (G-18-0022123). A.S. acknowledges support from the Canadian Institutes of Health Research (CIHR) (#PJT-169008). The research was enabled in part by computing services from the Digital Research Alliance of Canada (www.alliancecan.ca). We thank Min-Young Kim, Alex Hodge and Thomas Quail for carrying out the experiments and providing the data.

Author contributions

T.M.B. conceived the study. T.M.B., D.D. and C.T.B. developed the methodology. T.M.B. performed the analysis. M.A., L.G., A.S. and G.B. provided resources. A.S. and G.B. provided project supervision. T.M.B. wrote the first draft. All authors revised and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-42020-z>.

Correspondence and requests for materials should be addressed to Thomas M. Bury.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023