scientific reports

Check for updates

OPEN

Time–frequency time–space LSTM for robust classification of physiological signals

Tuan D. Pham

Automated analysis of physiological time series is utilized for many clinical applications in medicine and life sciences. Long short-term memory (LSTM) is a deep recurrent neural network architecture used for classification of time-series data. Here time–frequency and time–space properties of time series are introduced as a robust tool for LSTM processing of long sequential data in physiology. Based on classification results obtained from two databases of sensor-induced physiological signals, the proposed approach has the potential for (1) achieving very high classification accuracy, (2) saving tremendous time for data learning, and (3) being cost-effective and user-comfortable for clinical trials by reducing multiple wearable sensors for data recording.

Analysis and classification of clinical time-series data in physiology and disease processes are considered as a catalyst for biomedical research and education. Innovative computerized tools for physiological data classification are increasingly needed to facilitate investigations on new unsolved challenging problems in clinical and life sciences with respect to both basic and translational perspectives. Conventional methods for classification of physiological time series to detect abnormal conditions include fractals, chaos, nonlinear dynamics, signal coding, pattern matching, and machine learning. The current surge of modern artificial intelligence (AI) opens a new approach for sequential data classification with long short-term memory (LSTM) networks[1], which are an architecture of deep learning. LSTM networks are a type of recurrent neural networks that learn order dependence in sequential data.

There are many methods developed for classification of time series in different fields of applications. Time-series classification algorithms based on discriminatory features can be categorized into six main groups[2]: (1) whole series, (2) intervals, (3) shapelets, (4) dictionary, (5) combinations, and (6) model. For the whole-series approach, classification is performed by comparing the similarity between two time series using a distance measure. The methods of intervals choose one or multiple intervals of the series and use summary measures as features for classification. The methods of shapelets define a class with phase-independent patterns called shapelets, then a class is identified by the existence of one or more shapelets in the whole time series. The dictionary-based methods classify time series based on the frequency of its recurring subseries. The methods of combinations try to combine two or more methods of the whole series, intervals, shapelets, and dictionary for classification. The model-based methods fit a time series to mathematical models constructed for the classes and then assign the time series to the class that has the largest similarity score given by the class model. Most recently, deep-learning methods or deep neural networks have been reported to outperform many baseline time-series classification approaches and appear to be the most promising techniques for classifying temporal data[3].

Because LSTM networks can capture long-term temporal dependencies, they have been applied to provide solutions for many difficult problems in bioinformatics and computational biology[4]. As a state-of-the-art method for learning physiological models for disease prediction, many applications of LSTM and other deep-learning networks have recently been reported in literature, such as classifying electroencephalogram (EEG) signals in emotion, motor imagery, mental workload, seizure, sleep stage, and event related potentials[5], non-EEG signals in Parkinson's disease (PD)[6], learning and synthesis of respiration, electromyograms, and electrocardiograms (ECG) signals[7], decoding of gait phases using EEG[8], and early prediction of stress, health, and mood using wearable sensor data[9].

The present work presents a time–frequency time–space LSTM tool for robust and efficient classification of physiological time series, while solutions obtained from conventional LSTM networks would result in lower accuracy and higher data training time. Furthermore, for the case of clinical gait analysis with the use of measurement sensors to assess biomechanical patterns and therapeutic plan for rehabilitation in patients disabled from

Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 31952, Saudi Arabia. email: tpham@pmu.edu.sa

nature portfolio

1

conditions such as PD and post stroke, long walk trials are recommended to obtain at least 370 strides[10]. Such long-distance walks result in long records of physiological measurements, cause discomfort to the patients, and may be impractical to perform in many clinical settings[11].

Differentiating patients with PD from healthy controls using gait data was studied in[12], which trained fuzzy neural networks with wavelet features extracted from the gait data. Another study extracted gait features with the short-time Fourier transform and used the support vector machines (SVMs) for the classification task[13]. To capture the local changes in the dynamics of gait signals, the feature-extraction method of shifted 1-D local binary patterns and a multilayer perceptron, which is a class of feed-forward artificial neural networks, were used for the classification of PD and healthy controls[14]. The extraction of time-domain and frequency-domain features of gait data for training with random decision forests, which are an ensemble machine-learning method for classification, was reported in a more recent study for detecting patients with PD[15]. All these studies employed shallow neural networks or SVMs. However, deep neural networks are known to be the most advanced models of the neural-network approach and shown to be of performance superior to other types of statistical classifiers[16].

The novel idea for classification of physiological data with LSTM presented herein is the creation of complementary time–frequency and time–space features of time series. In signal processing, instead of viewing a time series as a one-dimensional signal, time–frequency analysis studies a signal in both time and frequency domains simultaneously by some function whose domain is the two-dimensional real plane to extract transient features from the signal by a time–frequency transform. Time–frequency signal processing for feature extraction was reviewed as a useful approach for pattern recognition[17] that provided successful applications, including EEG seizure detection and classification[17], classification of ultra-high-frequency signals[18], classification of vibration events[19], and classification of EEG signals and episodic memory[20].

In nonlinear dynamics, the time–space analysis attempts to transform one-dimensional signal into a two-dimensional space to enable the visualization of the recurrences of states of a dynamical system at certain times and enable the extraction of distinctive features representing behaviors of different dynamical mechanisms underlying nonlinear time series. The extraction of novel features from time series not only facilitate the power of signal compression for deep learning but also enhances the capability of LSTM networks for robust signal classification. In chaos theory, the method of recurrence plots (RPs) was developed for nonlinear time-series analysis[21]. RPs and extended methods were further addressed for the analysis of complex systems[22,23] and dynamical features of nonlinear time series[24]. While an RP is a binary visualization of recurrences of states of a dynamical system at certain pairs of time, a fuzzy recurrence plot (FRP)[25] displays the visualization as a grayscale image. Because of being much richer in texture than RPs, the technique of FRPs of time series is a preferred approach for texture analysis and has been successfully applied to extract texture features for pattern recognition, including classification of PD and control subjects using deep learning[6,26], tensor decomposition[27], and SVMs[28]; and other neuro-degenerative diseases[29].

In general, the time–frequency analysis is known a preferred approach for the representation and essential feature extraction of non-stationary signals because it is effective for estimating the underlying characteristics composing the signals[17], whereas the time–space analysis provides another kind of visual information about the signals by detecting hidden dynamical features being inherent in the data. The combination of complementary features generated by both time–frequency and time–space analysis methods is therefore promising for enhancing the classification power of the sequential deep learning.

Data
ECG data.
ECG signals capture the electrical activity of a human heart over a period of time. ECG signals are used by physicians for examining the condition of a patient's heartbeat to detect if the condition is normal or irregular. Atrial fibrillation (AF) is a type of irregular heartbeat that occurs when the upper chambers of the heart (atria) beat out of coordination with the lower chambers (ventricles). The ECG data[30] used in this study are publicly available from the *PhysioNet: The Research Resource for Complex Physiologic Signals*. The data consist of ECG signals sampled at 300 Hz and classified by a group of experts into normal sinus rhythm, AF, alternative rhythm, and noise. The purpose of the creation of this challenging database was to call for the development of new methods for classifying these types of cardiac arrhythmias. Information about the number of participants in the recordings of normal rhythm, AF rhythm, and other rhythms is not available from the data source[31].

Gait in Parkinson's disease data.
The Gait in Parkinson's Disease database[32] consists of time series of vertical ground reaction force in Newtons of gait dynamics from 93 patients with idiopathic PD and 73 healthy controls. This database is also publicly available from the *PhysioNet: The Research Resource for Complex Physiologic Signals*. The data consist of the vertical ground reaction force (in Newtons) signals of the subjects as they walked at their usual, self-selected pace for approximately 2 minutes on level ground. The force was measured as a function of time with 8 sensors placed underneath each foot. The force signals of each of the 16 sensors placed under the two feet of each subject were digitized and recorded at 100 samples per second.

Time–frequency and time–space analysis
Instantaneous frequency.
The instantaneous frequency (IF) of a non-stationary signal is a time-varying parameter that relates to the average of the frequencies f present in the signal as it evolves over time instants t[33,34]. The IF function estimates the IF of a signal at a sampling rate by computing the spectrogram power spectrum $P(t, f)$ and estimating the IF as

$$IF(t) = \frac{\int_{-\infty}^{\infty} fP(t,f)df}{\int_{-\infty}^{\infty} P(t,f)df}. \tag{1}$$

The power spectrum is a mathematical expression of the amount of the signal at a frequency f. For a periodic signal, peaks at the fundamental frequency and its harmonics are observed at the spectrum; for a quasiperiodic signal, peaks at linear combinations of related frequencies observed; and a chaotic signal yields broad band components to the spectrum. In practice, the exact solution for the power spectrum cannot be determined because a signal $x(t)$ is not infinitely long but measured over a finite interval $0 \leq t \leq T$. Therefore, the power spectrum needs to be numerically estimated. A method for estimating the power spectrum of a time series x_k, $k = 0, \ldots, N - 1$ is described as follows.

The spectral density of a time series of length N can be approximated as[35]

$$P_N(f) = \frac{\Delta t}{N} \left| \sum_{k=0}^{N-1} x_k e^{-i2\pi fk\Delta t} \right|^2, \tag{2}$$

where Δt is the sampling interval.

If the spectral value is calculated at $f = j\Delta f$, where $\Delta f = 1/(N\Delta t)$, and $\Delta t = 1$, then

$$P_j = \frac{1}{N} \left| \sum_{k=0}^{N-1} x_k e^{-i2\pi \frac{jk}{N}} \right|^2 = \frac{1}{N} |X_j|^2, \tag{3}$$

which indicates the discrete Fourier transform (DFT), X_j, as

$$X_j = \sum_{k=0}^{N-1} x_k e^{-i2\pi \frac{jk}{N}}, j = 0, \ldots, N - 1. \tag{4}$$

However, it was proved that the power spectrum estimate expressed in Eq. (3) is not properly scaled[35]. Therefore, the estimate is modified as

$$P_j = \frac{1}{WN} \left| \sum_{k=0}^{N-1} w_k x_k e^{-i2\pi \frac{jk}{N}} \right|^2, j = 0, \ldots, N - 1; \tag{5}$$

in which

$$W = \frac{1}{N} \sum_{j=0}^{N-1} w_j^2, \tag{6}$$

where w_j, $j = 0, \ldots, N - 1$, are the weights or coefficients of a window function (the Kaiser window[36] is applied in this study).

The estimate of P_j expressed in Eq. (5) using the fast Fourier transform (FFT) can be sequentially carried out as follows[35].

- Truncate the time series or pad with zeros so that $N = 2^n$, where n is a positive integer.
- Weight the time series with a window function.
- Calculate the DFT of the weighted time series $(w_k x_k)$ using the FFT.
- Calculate P_j using Eq. (5).

Spectral entropy.   The spectral entropy (SE) of a signal is a measure of its spectral power distribution[33,34]. The SE treats the normalized power distribution of the signal in the frequency domain as a probability distribution and calculates its Shannon entropy. The Shannon entropy in this context is known as the spectral entropy of the signal. Given a time–frequency power spectrogram $P(t,f)$, the probability distribution at time t, $0 \leq t \leq T$; and frequency point m, $m = 1, \ldots, N$; denoted as $p(t, m)$, is

$$p(t, m) = \frac{P(t, m)}{\sum_f P(t, f)}, \tag{7}$$

where $f \in [0, fs/2]$ is specified in this study, and fs is the sampling frequency.

The spectral entropy at time t, denoted as $H(t)$, is given as

$$H(t) = -\sum_{m=1}^{N} p(t, m) \log_2 p(t, m). \tag{8}$$

Fuzzy recurrence plot.   In the study of dynamical systems, a sequence of values in time can be transformed into an object in space. This transformation allows the sequence to be analyzed in space. Such space is called
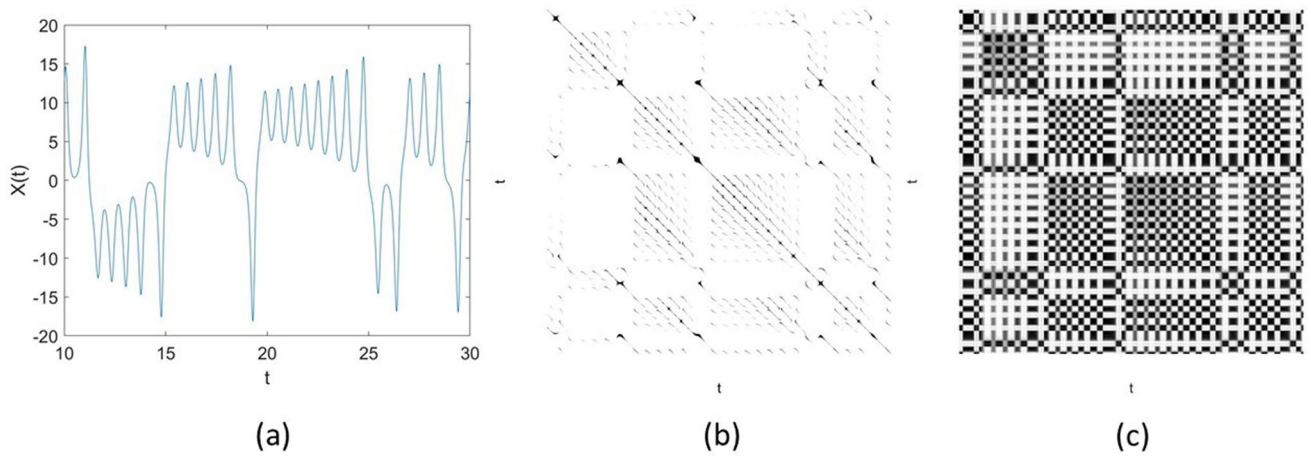
Figure 1. Recurrence of a dynamical system: (**a**) time series of the Lorenz X-component, (**b**) recurrence plot, and (**c**) fuzzy recurrence plot.

the phase space. The object in the phase space is called the phase space set. The transformation of a sequence of values in time into an object in the phase space can be done using the time-delay embedding[37]. The embedding dimension describes the space (such as a line, an area, or a volume) that contains the object[38]. Time delay, which is also called lag, expresses the amount of offset in a time series. Mathematically, the phase-space reconstruction using time-delay embedding for a time series $(z_1, z_2, \ldots, z_I)$ can be performed as $\mathbf{y}_i = (z_i, z_{i+\phi}, \ldots, z_{i+(d-1)\phi})$, $i = 1, \ldots, I - (d-1)\phi$, where ϕ and d are time delay and embedding dimension, respectively.

In fuzzy logic[39], a fuzzy set is defined as a collection of distinct objects whose membership grades in the set are expressed with real numbers. In mathematic terms, let U be a universe of discourse and F a subset of U. The fuzzy set F is characterized by a fuzzy membership function $\mu_F(x)$ that maps each element $x \in U$ to the interval $[0, 1]$: $\mu_F(x) : U \rightarrow [0, 1]$. The real value of $\mu_F(x)$ is called the fuzzy membership grade of x in F. The notion of a fuzzy set can be expressed in the following three cases: 1) $\mu_F(x) = 0$ if x is not totally in F, 2) $\mu_F(x) = 1$ if x is totally in F, and 3) $0 < \mu_F(x) < 1$ if x is partially in F. Thus, the greater value of the fuzzy membership grade is, the more certain x is a member of F.

In cluster analysis, data points can be assigned to different groups or clusters. Points that are most similar to each other belong to the same cluster. Based on the concept of fuzzy sets, fuzzy clustering assigns the data points to all clusters with different degrees of fuzzy membership. In other words, the fuzzy membership value of a data point for a certain cluster indicates how positive the data point belongs to that cluster.

Now let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{Nm}$ with $\mathbf{x}_i \in \mathbb{R}^m$ be a phase-space collection of a signal transformed by the time-delay embedding method, c a pre-defined number of clusters, $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_c\}$ a set of clusters, and $\mu(\mathbf{x}_i, \mathbf{v}_q)$, $i = 1, \ldots, N, q = 1, \ldots, c$, fuzzy membership grades expressing the degrees of phase-space vectors $\mathbf{x}_i$ belonging to cluster centers $\mathbf{v}_q \in \mathbf{V}$. These fuzzy membership grades can be determined using the fuzzy c-means algorithm[40]. An FRP, denoted by $\tilde{\mathbf{R}}$, is defined as[25]

$$\tilde{\mathbf{R}}(i, j) = \mu(\mathbf{x}_i, \mathbf{x}_j), \ i, j = 1, \ldots, N, \tag{9}$$

where $\mu(\mathbf{x}_i, \mathbf{x}_j) \in [0, 1]$ is the fuzzy membership of similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$.

The elements of an FRP, $\tilde{\mathbf{R}}(i, j)$, $i = 1, \ldots, N$, $j = 1, \ldots, N$, can be inferred using three properties of fuzzy relations as follows.

1. Reflexivity:

$$\mu(\mathbf{x}_i, \mathbf{x}_i) = 1, \ i = 1, \ldots, N. \tag{10}$$

2. Symmetry:

$$\mu(\mathbf{x}_i, \mathbf{v}_q) = \mu(\mathbf{v}_q, \mathbf{x}_i), \ i = 1, \ldots, N, q = 1, \ldots, c. \tag{11}$$

3. Transitivity:

$$\mu(\mathbf{x}_i, \mathbf{x}_j) = \max[\min\{\mu(\mathbf{x}_i, \mathbf{v}_q), \mu(\mathbf{x}_j, \mathbf{v}_q)\}], q = 1, \ldots, c. \tag{12}$$

As an example, to illustrate some difference in the visual display of an RP and an FRP, Fig. 1 shows a time series of 2000 points of the X-component of the Lorenz (chaotic) system[41], and its RP and FRP. The RP was constructed using the embedding $= 3$, time delay $= 1$, and a conventional value for the similarity threshold $= 5\%$ of the standard deviation of the signals. The FRP was constructed using the embedding $= 3$, time delay $= 1$, and number of clusters $= 3$. The grayscale image of the FRP is much richer in texture than the binary image of the RP.
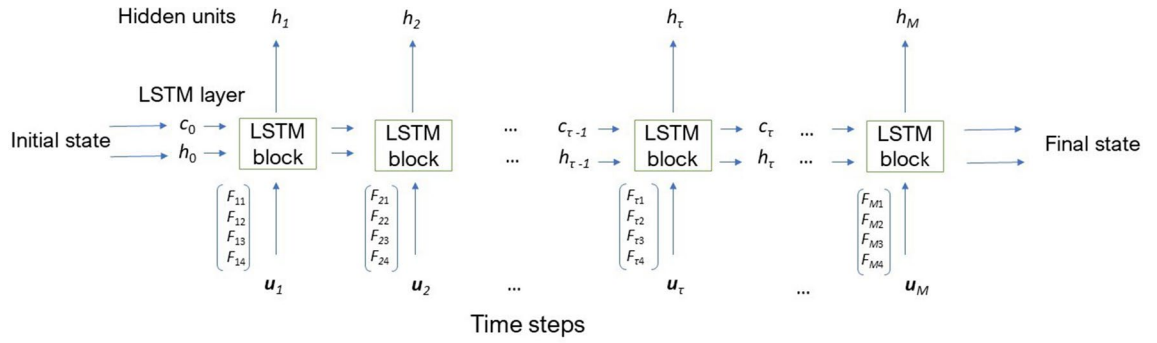
Figure 2. TF–TS LSTM layer architecture.

Fuzzy recurrence image entropy. Entropy of a grayscale image is a statistical measure of randomness to characterize the texture of the image. As an FRP is a grayscale image, the entropy of an FRP image is defined as

$$E_{FRI} = -\sum_{k=1}^{K} p_k \log_2 p_k, \tag{13}$$

where $K = 256$, which is the number of gray levels of the FRP (obtained by converting real values of pixels in [0, 1] to integers in [0, 255]), and p_k is the probability associated with the intensity level k, $k = 1, \ldots, K$, obtained from the normalized histogram for the k-th bin.

Fuzzy recurrence entropy. Based on the definition of the non-probabilistic entropy of a fuzzy set[42], the entropy of an $N \times N$ FRP or fuzzy recurrence entropy that is a measure of the degree of uncertainty of recurrences of the reconstructed phase space of a signal is defined as[43]

$$E_{FR} = \sum_{i=}^{N} \sum_{j=1}^{N} -\mu(\mathbf{x}_i, \mathbf{x}_j) \log_2 \mu(\mathbf{x}_i, \mathbf{x}_j) - [1 - \mu(\mathbf{x}_i, \mathbf{x}_j)] \log_2[1 - \mu(\mathbf{x}_i, \mathbf{x}_j)], \tag{14}$$

where $\mu(\mathbf{x}_i, \mathbf{x}_j)$ corresponds to $\tilde{\mathbf{R}}(i, j)$ defined in Eq. (9).

Time–frequency time–space long short-term memory networks

Based on LSTM networks[1,4,44], in which the proposed input time–frequency (TF) and time–space (TS) features are included, the architecture for a TF–TS LSTM block is graphically described in Fig. 2. This figure illustrates the flow of an input time series $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_M) \in \mathbb{R}^{MQ}$ through an LSTM layer, where M is the number of segments split from the original time series of length L, and Q the number of features. In this study, $M = \lceil L/N \rceil$, where $N = 128$, $\lceil \rceil$ denotes the ceiling function, and $Q = 4$. The input at a time point is the concatenation of the four features extracted for the segment at the same time point, i.e., $\mathbf{u}_\tau = (F_{\tau 1}, F_{\tau 2}, F_{\tau 3}, F_{\tau 4})^T$, $\tau = 1, \ldots, M$, where $F_{\tau 1}, F_{\tau 2}, F_{\tau 3}$, and $F_{\tau 4}$ are the instantaneous frequency, spectral entropy, fuzzy recurrence image entropy, and fuzzy recurrence entropy extracted from segment $\mathbf{u}_\tau$, respectively.

The learnable weights of an LSTM layer are the input weights, denoted as $\mathbf{a}$, recurrent weights, denoted as $\mathbf{r}$, and bias, denoted as b. The matrices $\mathbf{A}$, $\mathbf{R}$, and vector $\mathbf{b}$ are the concatenations of the input weights, recurrent weights, and bias of each component, respectively. The concatenations are expressed as

$$\mathbf{A} = [\mathbf{a}_i, \mathbf{a}_f, \mathbf{a}_g, \mathbf{a}_o]^T, \tag{15}$$

$$\mathbf{R} = [\mathbf{r}_i, \mathbf{r}_f, \mathbf{r}_g, \mathbf{r}_o]^T, \tag{16}$$

$$\mathbf{b} = [b_i, b_f, b_g, b_o]^T, \tag{17}$$

where i, f, g, and o denote the input gate, forget gate, cell candidate, and output gate, respectively.

The cell state at time step τ is defined as

$$\mathbf{c}_\tau = f_\tau \circ \mathbf{c}_{\tau-1} + i_\tau \circ g_\tau, \tag{18}$$

where $\circ$ is the Hadamard product.

The hidden state at time step τ is given by

$$\mathbf{h}_\tau = o_\tau \circ \sigma_c(\mathbf{c}_\tau), \tag{19}$$

where σ_c is the state activation function that is usually computed as the hyperbolic tangent function (tanh).

At time step τ, the input gate (i_τ), forget gate (f_τ), cell candidate (g_τ), and output gate (o_τ) are defined as
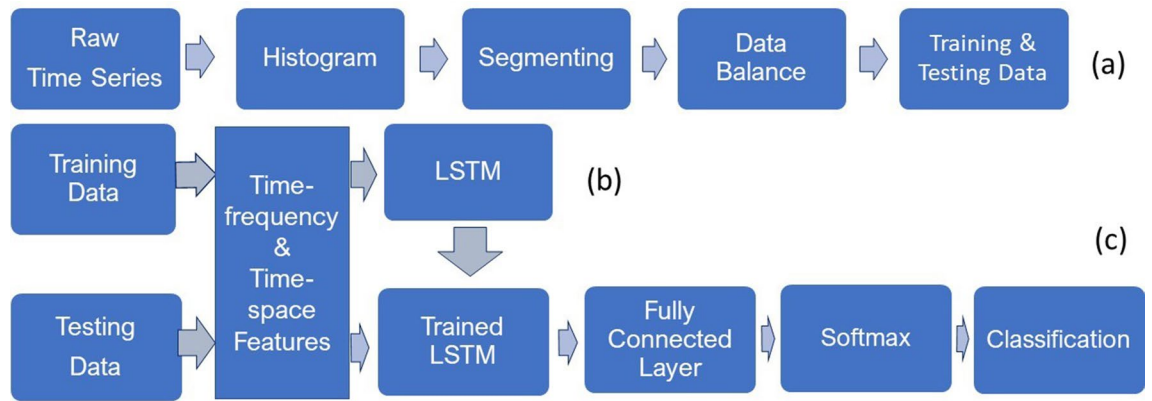
Figure 3. Procedure for classification of physiological time series with TF–TS LSTM: (**a**) from raw data to data balance in training and testing, (**b**) transformation of raw time series into time–frequency and time–space features, and (**c**) classification of testing data.

$$i_\tau = \sigma_g(\mathbf{a}_i \mathbf{u}_\tau + \mathbf{r}_i \mathbf{h}_{\tau-1} + b_i), \tag{20}$$

$$f_\tau = \sigma_g(\mathbf{a}_f \mathbf{u}_\tau + \mathbf{r}_f \mathbf{h}_{\tau-1} + b_f), \tag{21}$$

$$g_\tau = \sigma_c(\mathbf{a}_g \mathbf{u}_\tau + \mathbf{r}_g \mathbf{h}_{\tau-1} + b_g), \tag{22}$$

$$o_\tau = \sigma_g(\mathbf{a}_o \mathbf{u}_\tau + \mathbf{r}_o \mathbf{h}_{\tau-1} + b_o), \tag{23}$$

where σ_g denotes the gate activation function that usually adopts the sigmoid function.

A bidirectional LSTM (bi-LSTM)[45] is an extension of traditional LSTM that can improve performance on sequence classification problems. Instead of being trained with one LSTM on the input time series, a bi-LSTM architecture is trained with both time directions simultaneously with hidden forward and backward layers. The first on the input time series as it is and the second on a reversed copy of the time series. This architecture learns bidirectional long-term dependencies between time steps of time series and therefore can provide additional context to the network and result in fuller learning on the data.

The procedures for obtaining data balance for training and testing sets, and the transformation of raw time series into TF and TS features for LSTM learning and classification are outlined in Fig. 3.

To obtain signals of the same length contained in both training and testing datasets, the histogram of the distribution of the lengths of the signals is observed to detect the majority length. Signals of lengths that are less than the majority are discarded, and those that are longer than the majority are split into segments of the majority length and the remaining samples of the signal are ignored if there are any. Creating signals of equal length is particularly useful for the training of the networks that breaks the data into mini-batches. In the same mini-batch, the training pads or truncates the signals to have the same length. However, it is known that the process of padding or truncating can reduce the performance of the networks because of the added or missed information caused by the padding or truncating, respectively. To obtain the data balance in each class for both training and testing, copies of the signals of the minority class are repeated to achieve the same size of the signals of the majority class. This step is described in Fig. 3a. The next step is to extract the TF features of the signals using the instantaneous frequency and spectral entropy and the TS features of the signals using the fuzzy recurrence image entropy and fuzzy recurrence entropy for training the networks (Fig. 3b). The same TF and TS features are extracted from the testing signals as the input for the trained TF–TS LSTM networks to carry out the classification task (Fig. 3c).

Performance measures

Let condition positive *P* be the total number of disease signals, condition negative *N* the total number of healthy control signals, true positive *TP* the number of disease signals correctly identified as disease, false positive *FP* the number of healthy control signals incorrectly identified as disease, true negative *TN* the number of healthy control signals correctly identified as healthy control, and false negative *FN* the number of the disease signals incorrectly identified as healthy control.

Accuracy (*ACC*) is defined as

$$ACC = \frac{TP + TN}{P + N}. \tag{24}$$

Sensitivity (*SEN*) is defined in this study as the portion of the disease signals that are correctly identified as having the condition:

Method	Sensors	% ACC	% SEN	% SPE	% PRE	F_1	Time (mins)
ECG of atrial fibrillation and normal sinus rhythm							
LSTM	1	58.37 ± 0.01	82.86 ± 0.03	33.88 ± 0.09	55.62 ± 0.06	0.67 ± 0.02	3506 ± 19
TF–TS LSTM	1	$\mathbf{93.77 \pm 0.07}$	91.43 ± 0.05	96.12 ± 0.01	95.93 ± 0.01	0.94 ± 0.02	$\mathbf{1} \pm 0.11$

Table 1. Ten-fold cross-validation metrics for ECG of classification of atrial fibrillation and normal sinus rhythm. LSTM = direct use of time series by bi-LSTM, TF–TS LSTM = use of time–frequency and time–space features of time series by bi-LSTM.

Method	Sensors	% ACC	% SEN	% SPE	% PRE	F_1	Time (mins)
Lee and Lim[12]	8	77.33	81.10	65.48	n/m	n/m	n/m
Daliri[13]	16	91.20	91.71	89.92	n/m	n/m	n/m
Ertugrul et al.[14]	16	88.89	93.93	77.66	n/m	n/m	n/m
Acici et al.[15]	16	98.04	99.10	95.70	n/m	n/m	n/m
Zeng et al.[47]	4	96.99	96.77	97.26	n/m	n/m	n/m
LSTM	1	78.95 ± 0.03	100.00 ± 0.00	60.00 ± 0.10	69.23 ± 0.05	0.82 ± 0.04	111 ± 9.95
TF–TS LSTM	1	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1 ± 0.00	< 1

Table 2. Ten-fold cross-validation metrics for classification of gait of patients with Parkinson's disease and healthy controls. LSTM = direct use of time series by bi-LSTM, TF–TS LSTM = use of time–frequency and time–space features of time series by bi-LSTM, and n/m = not mentioned.

$$SEN = \frac{TP}{P}. \tag{25}$$

Specificity (SPE) is the portion of the healthy control signals that are correctly identified as not having the disease:

$$SPE = \frac{TN}{N}. \tag{26}$$

Precision (PRE) is calculated as

$$PRE = \frac{TP}{TP + FP}. \tag{27}$$

F_1 score is the harmonic mean of precision and sensitivity and calculated as

$$F_1 = \frac{2TP}{2TP + FP + FN}. \tag{28}$$

Results

Tables 1 and 2 list the tenfold cross-validation results of two physiological databases: ECG, and Gait in Parkinson's Disease, respectively. For the ECG database, this experiment used normal sinus rhythm (5050 signals) and AF (738 signals) for binary classification. For the Gait in Parkinson's Disease data, this study used the time series recorded from only one sensor under the left foot labeled as L5 on the database. The purpose of selecting the sensor data recorded at the L5 location was to compare with the work reported in[47], which used four sensors at L5, L7, R7, and R8 for the classification of gait patterns. The LSTM used in this study was the bi-LSTM (LSTM will be used as bi-LSTM subsequently). To extract the TF features, sampling frequency was set as 300 Hz. To extract the TS features, the embedding dimension = 1, time delay = 1, and number of clusters = 3 for computing the FRPs. The specifications of the FRP parameters were based on previous studies[25,43], which provided satisfactorily results and were not as sensitive for constructing FRPs as for RPs[25].

All TF and TS features were standardized to improve the network training and testing[46]. For the LSTM specifications, the network layer with an output size = 100, fully connected layer = 2 (two classes), followed by a softmax layer and a classification layer. Training options of the bi-LSTM were set as optimizer = 'Adam' (adaptive moment estimation), including L_2 regularization factor, maximum number of epochs = 80, minimum batch size = 150, initial learning rate = 0.01, and gradient threshold = 1.

For the ECG data, the TF–TS LSTM significantly outperformed conventional LSTM in terms of classification accuracy (58% and 94% for conventional LSTM and TF–TS LSTM, respectively), other statistical measures (sensitivity, specificity, precision, and F_1 score), and training time (3506 minutes and 1 minute for LSTM and TF–TS LSTM, respectively, where the time for computing the four features was excluded in the TF–TS LSTM training). The specificity (34%) is much lower than the sensitivity (83%) obtained from the conventional LSTM, while these two measures are much more balanced using the TF–TS LSTM (sensitivity = 91% and specificity = 96%).
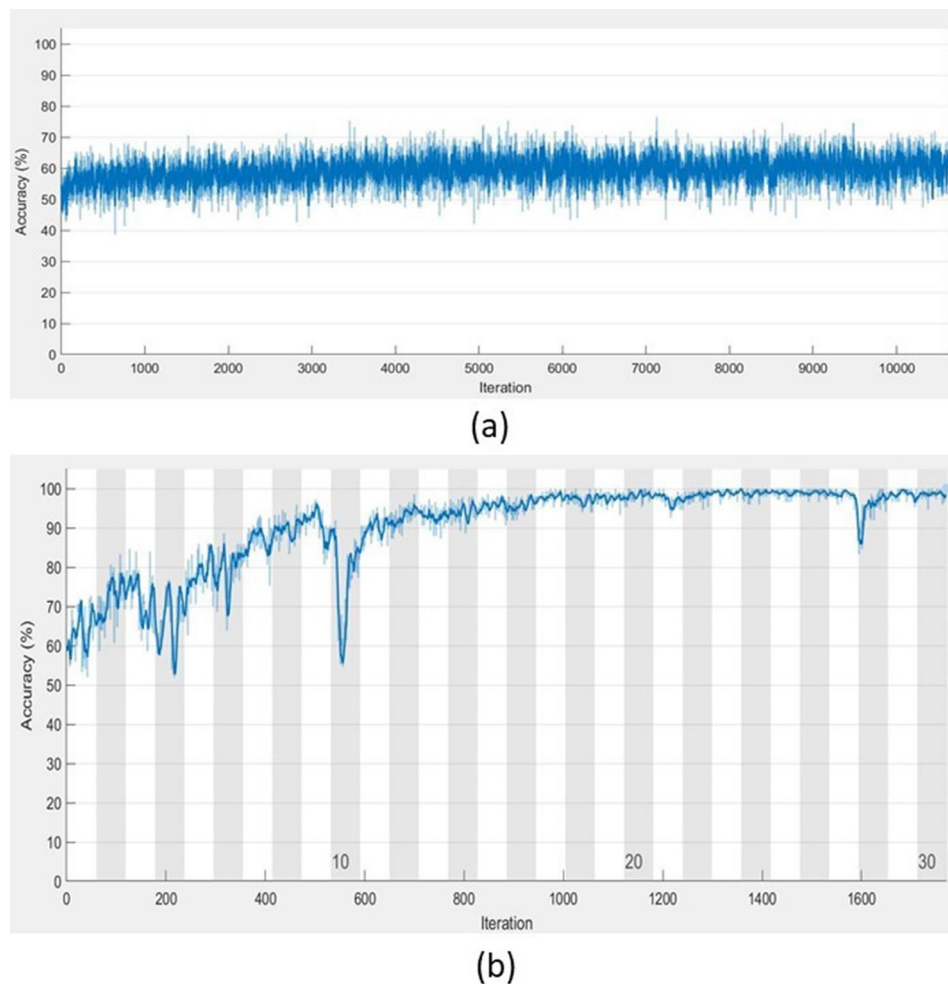
Figure 4. Training processes of (**a**) bi-LSTM and (**b**) TF–TS bi-LSTM, using tenfold cross-validation of the ECG data.

For the gait data, using the signals recorded from only one sensor, TF–TS LSTM provided perfect classification metrics (accuracy = 100%, sensitivity = 100%, specificity = 100%, precision = 100%, and F_1 score = 1) with the training time of < 1 minute (the time for computing the four features was excluded). The use of conventional LSTM yielded the accuracy = 79% with 111 minutes for data training. Other five previous methods[12–15,47] that studied the same database used the number of sensors between 4 and 16 obtained accuracy rates between 77%[12] and 98%[15] (standard deviations of classification results obtained from these five methods were not given in literature[47]).

Discussion

Computer experiments have shown that TF–FS LSTM achieved very high performance in the classification task and saved tremendous training time in comparison with the conventional implementation of the conventional LSTM. As an example, Fig. 4 shows the contrast of the training processes of conventional LSTM and TF–TS LSTM with respect to the convergence of accuracy and the number of iterations. Not only the TF–TS LSTM out-performed conventional LSTM, classification results of gait in Parkinson's disease in terms of accuracy, sensitivity, specificity, precision, and F_1 score obtained from the TF–TS LSTM are higher than those previously reported in literature. In particular, the TF–TS LSTM used the data recorded from only one sensor. The significant reduction in biomedical sensors to measure human physiological parameters in real time for disease detection has an implication for promising the user's comfort and contributing to the low cost, simplicity, and portability in wearable sensor technology.

In this study, only the gait data recorded by one sensor located at L5 were used to compare with the other work[47] that included the data recorded by four sensors located at L5, L7, R7, and R8. The gait classification with the use of a single sensor located at L5 obtained from the proposed TF–TS LSTM outperformed the use of the four sensors for the classification obtained from the methods of phase space reconstruction, empirical mode decomposition, and neural networks[47]. Tests of the TF–TS LSTM for the gait classification using data recorded from other single sensors were not carried out. However, the current comparison has shown the better per-formance of the TF–TS LSTM. As the five methods[12–15,47], which were compared with the TF–TS LSTM using

the gait data, were proposed and implemented by other authors, it would be difficult to fairly implement these methods for the classification of the ECG data without the provision of the source codes. However, it is shown that the test results obtained from the TF–TS LSTM are significantly higher than the LSTM using the two data-sets, and the classification accuracy obtained from the LSTM using the gait data from only one sensor (79%) is higher than the result reported in[12] using the gait data from 8 sensors (77%).

Here the signal lengths were made to be the same length of the majority. In case, if the majority does not exist or the histogram has a uniform distribution, the signal lengths can be made to be equal to the length of the shortest signal. In general, signals of lengths that are shorter than the majority can be included for the classification. However, it has been mentioned earlier, creating signals of equal length can be more effective for the network training and testing. In practice, the recording of physiological signals that meet some standard length for testing is feasible because it is based on the majority.

As shown in Fig. 4, the high accuracy of the TF–TS LSTM training could be reached while the training of the LSTM with raw time series could not improve much in accuracy. Furthermore, the TF–TS LSTM requires much shorter time for training in comparison with the training of raw long time series. This is because it is trained with sequential features of the time series instead of the time series, where the length of the features is much shorter than that of the original data and the effectiveness of the standardized features is an important factor for improving the network performance during training.

Feature extraction can be related to dimensionality reduction by which multivariate data can be reduced to lower-dimensional space for more manageable data processing. The physiological time series used in this study are one-dimensional time series. On the contrary, these time series were split into equal segments from which the four features were extracted for learning and classification by the TF–TS LSTM. In other words, the one-dimensional time series were transformed into much shorter sequences of 4 feature dimensions as shown in Fig. 2. The extracted features provide essential information of the data in time–frequency and time–space domains, which are intended to be complementary, informative, and non-redundant responses. Thus, the transformed data can facilitate the subsequent learning and leverage discriminative power of the sequential deep learning, leading to better class predictions. The results obtained in this study have shown the TF–TS LSTM outperformed other statistical classifiers, including SVMs and multilayer perceptron.

In summary, the finding is that training the LSTM network with raw time series produce poor classification results but training the network with TF and TS features extracted from the signals can both significantly enhance the classification performance and reduce the training time.

The Matlab-based TF–TS LSTM software for classification of physiological signals is designed to be easily utilized by biomedical and life science users who do not have technical knowledge in AI, signal processing, and general physics by following provided step-by-step instructions (Supplementary Note). In biomedical data, the problem of data imbalance is common, which can significantly prevent classifiers from achieving good results. The software suggests how to design a balance of class samples for training and testing datasets when minority classes exist.

Conclusions

An AI-based approach for improving the performance in detecting diseases using physiological signals have been presented and discussed. The proposed method takes advantages of information extracted from both frequency and space out of the temporal data for effective deep learning to increase the classification task and lower computational complexity. Although the method was developed for classifying time series in physiology, it can be readily applied to the classification of other biological and clinical signals, such as time series in gene expression[48], neurology[49], and epidemiology[50].

The AI-based method presented in this work was tested using the records obtained from a single-sensor measurement of gait in PD. The results suggest the method has potential to be able to reduce the need of using multiple sensors for recording physiological data, thus resulting in both cost-saving and comfort to the participants. Further tests of the method with other multiple-sensor data would be necessary to confirm the finding. Wearable sensors are useful devices for evaluating patient outcomes in clinical trials. However, the devices need to provide physical ease to participants so that they are prepared to wear them. Otherwise, the deployment of such tools will not be practically feasible, particularly when applying to the older adult (> 50 years) population[51].

Software availability

MATLAB software, ECG data for AF and normal sinus rhythm, and Supplementary Note for running the ECG data used in this paper are publicly available at the author's personal homepage: https://sites.google.com/view/tuan-d-pham/codes under the title "TF–TS LSTM".

References

1. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
2. Bagnall, A., Lines, J., Bostrom, A., Large, J. & Keogh, E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Disc.* **31**, 606–660 (2017).
3. Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L. & Muller, P. A. Deep learning for time series classification: a review. *Data Min. Knowl. Disc.* **33**, 917–963 (2019).
4. Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R. & Schmidhuber, J. LSTM: a search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **28**, 2222–2232 (2017).

5. Craik, A., He, Y. & Contreras-Vidal, J. L. Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.* **16**, 031001 (2019).
6. Pham, T. D., Wardell, K., Eklund, A. & Salerud, G. Classification of short time series in early Parkinson's disease with deep learning of fuzzy recurrence plots. *IEEE/CAA J. Autom. Sin.* **6**, 1306–1317 (2019).
7. Belo, D., Rodrigues, J., Vaz, J. R., Pezarat-Correia, P. & Gamboa, H. Biosignals learning and synthesis using deep neural networks. *BioMed. Eng. OnLine* **16**, 115 (2017).
8. Tortora, S., Ghidoni, S. S., Chisari, C., Micera, S. & Artoni, F. Deep learning-based BCI for gait decoding from EEG with LSTM recurrent neural network. *J. Neural Eng.* **17**, 046011 (2020).
9. Umematsu, T., Sano, A. & Picard, R. W. Daytime data and LSTM can forecast tomorrow's stress, health, and happiness. In *Proc. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2186–2190 (2019).
10. Hollman, J. H. *et al.* Number of strides required for reliable measurements of pace, rhythm and variability parameters of gait during normal and dual task walking in older individuals. *Gait Posture* **32**, 23–28 (2010).
11. Kribus-Shmiel, L., Zeilig, G., Sokolovski, B. & Plotnik, M. How many strides are required for a reliable estimation of temporal gait parameters? Implementation of a new algorithm on the phase coordination index. *PLoS ONE* **13**, e0192049 (2018).
12. Lee, S. H. & Lim, J. S. Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. *Expert Syst. Appl.* **39**, 7338–7344 (2012).
13. Daliri, M. R. Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease. *Biomed. Signal Process. Control* **8**, 66–70 (2013).
14. Ertugrul, O. F., Kaya, Y., Tekin, R. & Almali, M. N. Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait. *Expert Syst. Appl.* **56**, 156–163 (2016).
15. Acici, K., Erdas, C. B., Asuroglu, T., Toprak, M. K., Erdem, H. & Ogul, H. A random forest method to detect Parkinson's disease via gait analysis. In *Proc. Int. Conf. Engineering Applications of Neural Networks* 609–619 (2017).
16. Dargan, S. *et al.* A survey of deep learning and its applications: a new paradigm to machine learning. *Arch. Comput. Methods Eng.* **27**, 1071–1092 (2020).
17. Boashash, B., Khan, N. A. & Ben-Jabeur, T. Time–frequency features for pattern recognition using high-resolution TFDs: a tutorial review. *Digit. Signal Proc.* **40**, 1–30 (2015).
18. Wang, K., Li, J., Zhang, S., Qiu, Y. & Liao, R. Time-frequency features extraction and classification of partial discharge UHF signals. In *Proc. 2014 International Conference on Information Science, Electronics and Electrical Engineering* 1231–1235 (2014).
19. Xu, C., Guan, J., Bao, M., Lu, J. & Ye, W. Pattern recognition based on time-frequency analysis and convolutional neural networks for vibrational events in ϕ-OTDR. *Opt. Eng.* **57**, 016103 (2018).
20. Anderson, R. & Sandsten, M. Time-frequency feature extraction for classification of episodic memory. *EURASIP J. Adv. Signal Process.* **2020**, 19 (2020).
21. Eckmann, J. P., Kamphorst, S. O. & Ruelle, D. Recurrence plots of dynamical systems. *Europhys. Lett.* **5**, 973–977 (1987).
22. Marwan, N. *et al.* Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329 (2007).
23. Zou, Y., Donner, R. V., Marwan, N., Donges, J. F. & Kurths, J. Complex network approaches to nonlinear time series analysis. *Phys. Rep.* **787**, 1–97 (2019).
24. Goswami, B. A brief introduction to nonlinear time series analysis and recurrence plots. *Vibration* **2**, 332–368 (2019).
25. Pham, T. D. Fuzzy recurrence plots. *EPL* **116**, 50008 (2016).
26. Canturk, I. Fuzzy recurrence plot-based analysis of dynamic and static spiral tests of Parkinson's disease patients. *Neural Comput. Appl.* **33**, 349–360 (2021).
27. Pham, T. D. & Yan, H. Tensor decomposition of gait dynamics in Parkinson's disease. *IEEE Trans. Biomed. Eng.* **65**, 1820–827 (2018).
28. Pham, T. D. Pattern analysis of computer keystroke time series in healthy control and early-stage Parkinson's disease subjects using fuzzy recurrence and scalable network features. *J. Neurosci. Methods* **307**, 194–202 (2018).
29. Pham, T. D. Texture classification and visualization of time series of gait dynamics in patients with neuro-degenerative diseases. *IEEE Trans. Neural Syst. Rehabilit. Eng.* **26**, 188–196 (2018).
30. AF classification from a short single lead ECG recording–The PhysioNet Computing in Cardiology Challenge 2017. *PhysioNet*. https://physionet.org/content/challenge-2017/1.0.0/.
31. Clifford, G. D. *et al.* AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017. *Comput. Cardiol.* **44**, 11. https://doi.org/10.22489/CinC.2017.065-469 (2017).
32. Gait in Parkinson's disease. *PhysioNet*. https://physionet.org/content/gaitpdb/1.0.0/.
33. Boashash, B. Estimating and interpreting the instantaneous frequency of a signal-Part 1: fundamentals. *Proc. IEEE* **80**, 520–538 (1992).
34. Boashash, B. Estimating and interpreting the instantaneous frequency of a signal-Part 2: algorithms and applications. *Proc. IEEE* **80**, 540–568 (1992).
35. Buttkus, B. *Spectral Analysis and Filter Theory in Applied Geophysics* (Springer, 2000).
36. Kaiser, J. F. & Schafer, R. W. On the use of the I_0-sinh window for spectrum analysis. *IEEE Trans. Acoust. Speech Signal Process.* **28**, 105–107 (1980).
37. Takens, F. Detecting strange attractors in turbulence. *Lect. Notes Math.* **898**, 366–381 (1981).
38. Liebovitch, L. S. *Fractals and Chaos Simplified for the Life Sciences* (Oxford University Press, 1998).
39. Zadeh, L. A. Fuzzy sets. *Inf. Control* **8**, 338–353 (1965).
40. Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, 1981).
41. Lorenz, E. N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
42. de Luca, A. & Termini, S. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Inf. Control* **20**, 301–312 (1972).
43. Pham, T. D. Fuzzy recurrence entropy. *EPL* **130**, 40004 (2020).
44. Yu, Y., Si, X., Hu, C. & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **31**, 1235–1270 (2019).
45. Schuster, M. & Paliwal, K. K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **45**, 2673–2681 (1997).
46. Brownlee, J. How to scale data for long short-term memory networks in Python. *Machine Learning Mastery*, 07 July 2017. https://machinelearningmastery.com/how-to-scale-data-for-long-short-term-memory-networks-in-python/.
47. Zeng, W., Yuan, C., Wanga, Q., Liu, F. & Wang, Y. Classification of gait patterns between patients with Parkinson's disease and healthy controls using phase space reconstruction (PSR), empirical mode decomposition (EMD) and neural networks. *Neural Netw.* **111**, 64–76 (2019).
48. Qian, L., Zheng, H., Zhou, H., Qin, R. & Li, J. Classification of time series gene expression in clinical studies via integration of biological network. *PLoS ONE* **8**, e58383 (2013).
49. Costa, I. G., Schonhuth, A., Hafemeister, C. & Schliep, A. Constrained mixture estimation for analysis and robust classification of clinical time series. *Bioinformatics* **25**, i6–i14 (2009).
50. Perkins, T. A. *et al.* Heterogeneous local dynamics revealed by classification analysis of spatially disaggregated time series data. *Epidemics* **29**, 100357 (2019).

51. Keogh, A., Dorn, J. F., Walsh, L., Calvo, F. & Caulfield, B. Comparing the usability and acceptability of wearable sensors among older Irish adults in a real-world context: observational study. *JMIR mHealth uHealth* **8**, e15704 (2020).

Author contributions

T.D.P. conceptualized, designed the study, implemented the methods, and carried out the computer experiments. T.D.P. wrote the manuscript.

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-86432-7.

Correspondence and requests for materials should be addressed to T.D.P.

Reprints and permissions information is available at www.nature.com/reprints.

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