PSR-based research of feature extraction from one-second EEG signals: a neural network study

Aleksander Dawid¹

Received: 1 August 2019 / Accepted: 25 October 2019 / Published online: 4 November 2019
© The Author(s) 2019

Abstract

The speed and accuracy of signal classification are the most valuable parameters to create real-time systems for interaction between the brain and the computer system. In this work, we propose a schema of the extraction of features from one-second electroencephalographic (EEG) signals generated by facial muscle stress. We have tested here three sorts of EEG signals. The signals originate from different facial expressions. The phase-space reconstruction (PSR) method has been used to convert EEG signals from these three classes of facial muscle tension. For further processing, the data has been converted into a two-dimensional (2D) matrix and saved in the form of color images. The 2D convolutional neural network (CNN) served to determine the accuracy of the classifications of the previously unknown PSR generated images from the EEG signals. We have witnessed an improvement in the accuracy of the signal classification in the phase-space representation. We have found that the CNN network better classifies colored trajectories in the 2D phase-space graph. At the end of this work, we compared our results with the results obtained by a one-dimensional convolution neural network.

Keywords Phase space reconstruction · Signal processing · Facial expressions · EEG · CNN

1 Introduction

In 1924, Hans Berger developed the electroencephalography method of human brain study. [1]. The activity of the brain cortex, in this method, is visible in the change of electric potential on the scalp. Amplitude and frequency describe the oscillatory character of the signals obtained by EEG devices. Earlier research by Hans Berger has shown that there are frequency bands highly connected with the activity of the brain [2]. These frequency bands were named delta (< 4 Hz), theta (4–7 Hz), alpha (8–15 Hz), beta (16–31 Hz), and gamma (> 32 Hz). The original EEG signal is the compilation of all neuron's activities. In this method, we are not able to distinguish the activity of single neurons or even a small group of neurons. What we get is preferably the general state of the brain. The second kind of signal

is known as electromyographic (EMG) signal. The EMG is usually studied in the case of muscle disorders [3, 4], in controlling robotic limbs [5, 6] or face emotion recognition [7, 8]. As we see, the signal obtained by an EEG device represents the state of the brain and skull muscles. The signal must be processed if we want to extract information about the state of the brain or face muscles. Recently the phase space reconstruction method became the most popular in analyzing EEG data [9, 10]. The basic concept of this method is the multidimensional phase space [11]. The time-dependent EEG signal represents a single point in this space. The dimension of this space depends on the amount of data representing the signal. Nevertheless, analyzing data in such a multidimensional space could be quite challenging. The projection of this vector on a given surface creates a two-dimensional pattern that can

✉ Aleksander Dawid, adawid@wsb.edu.pl | ¹Department of Transport and Computer Science, WSB University, 1c Cieplaka St, 41-300 Dabrowa, Gornicza, Poland.

say more about the non-linear nature of the EEG signal and is easier to analyze [12, 13]. This method was adopted to process a short-timed EEG series [14]. These short-term signals are very significant in real-time brain-computer interface (BCI) development [15]. The PSR method of signal processing is widely used in transport control [16]. All those methods of signal processing are usually introducing before the classification process of the patterns. Nowadays, the neural network (NN) techniques getting more significant in expert systems. For example, in breast cancer detection [17] or medical diagnostic applications [18]. The accuracy of NN depends on its topology and the quality of input data. Currently, there are many results from multi-channel EEG mapping, where spatial and temporal data of EEG signals are studied, mainly for applications in neurology [19–21]. One of the good examples may be the work of Jiao et al. In which the pattern recognition methods were used to classify EEG signals coming from working memory during cognitive tasks [22]. The installation of multi-electrode EEG devices usually requires the help of another person, which hinders the use of these devices in everyday applications. EEG devices that are currently gaining increasing popularity are mobile devices with only one active electrode and one reference electrode fastened on the ear. These systems, due to the user's convenience, are potential candidates for use in remotely controlled computer devices.

In this work, we want to propose a different approach to short-timed signal processing. As a scientific background, we will use the phase space theory. We will show that processed EEG signal, as an input to the neural network, can give better validation accuracy than raw EEG signal. Three kinds of signals will be analyzed here. The first one is the

signal without any facial muscle tension. The second signal is coming from the jaw movement. The last signal registers the tightening of the mouth. The auto-mutual information theory will be used to show the separation of these three classes of signals. Finally, we will compare our results to that obtained from one-dimensional CNN.

2 Signals acquisition

In this work, a device from NeuroSky Mindwave Mobile [23] was used to collect EEG signals. This device is an inexpensive wireless EEG headset with a single electrode powered by one 1.5 V AAA battery. The sampling frequency of this device is 512 samples per second. This device can transmit both raw data and mental state data [24]. It is also able, at the hardware level, to filter the basic frequency bands of brain activity, such as bands; delta, theta, alpha, beta, and gamma. The conversion of the signal of analog activity of neurons to a digital signal takes place inside the EEG headset. The Bluetooth (BT) wireless system connects the device and a computer. Receiving data from the device is handled by the ThinkGear library procedure written in the C# programming language. The author of this work is responsible for the development of this software. Figure 1 is showing the scheme of the EEG signal acquisition system. The time-dependent raw signal in the form of two columns as a text file is stored. The first column represents the time, and the second column represents the values of the EEG signals. This data then creates a two-dimensional pattern. In chapter 4 will be more information about the construction of this pattern. Next, the neural network classifies the patterns. As the output of NN, we obtain a

Fig. 1 Overview of project workflow

probability vector, which characterizes the probability of similarity of the tested signal to three previously described classes. In chapter 5 will be more about the neural network used in this work.

3 The biological source of the signal

The subject of research in this work is a short-term EEG signal created during the change of muscle tone in the mouth and gums. Generally, these signals come from the facial muscles that are responsible for the facial expressions of the human face. Signals collected in this experiment creates three different classes corresponding to the activation of other parts of the facial muscles. The first class of signals corresponds to the lack of tension of any facial muscles. This class has been named NONE. The second class of signals corresponds to the condition of tightening and loosening the mouth muscles. The name of this class is LIP. The third class is from clamping and loosening the jaws. JAW is this class name. The biological object of research in this work was its author. After a single session for specific facial expression, we have collected 400 one-second signals. The total number of samples is equal to the number of samples multiplied by the number of classes. We have here in a total of 1200 samples. The author has conducted ten such sessions at different times of the day. Each item has a number. Data was not collected from the device when the signal condition was weak, which was usually caused by a low battery level. In this project, the placing of the measuring electrode is the frontal lobe of the brain (two inches above the brow). According to the

international 10–20 electrode placement standard [25], the plate was in the Fp1 position. The reason for such placement of the electrode in this test is its native location when installing the EEG device. Employing a clip on the ear of the examined subject places the reference electrode. The author conducted these non-invasive EEG tests on himself. All samples collected in this study are available on the GitHub website in the EDF format used in the medical series [26].

4 Patterns construction

Neural networks usually process data in the form of a tensor. An image is one such example of a tensor form. The raw image data structure consists of a three-dimensional tensor denoted as $\text{Image}[w][h][c]$, where the indices w , h , and c means the width, height, and color of an image, respectively. The dimension represented by the value of c corresponds to the number of channels associated with the selected color model. The most popular RGB model has three channels; red, green, and blue. The design of the pattern consists of projecting functional dependencies between the data constituting the components of the series representing the EEG signal on the image plane. The first pattern studied here comes from a raw time-dependent output EEG signal. We see here the dependence of EEG (t) in the form of a white curve on a black background. The original signal data has been rescaled here so that the curve is inside a 512×128 pixel frame. The value of the image width is related to the amount of data in a one-second sample. In the first picture of Fig. 2c,

Fig. 2 Phase space reconstruction patterns; **a** harmonic oscillator, **b** random oscillator, **c** EEG data, **d** classical mechanics trajectory

we see such an EEG image. The modern way of processing EEG signals refers to the theory of deterministic chaos [11]. The concept of phase space is the main idea of this theory. Consider the EEG signal as a data set, where each of them is a single time point in this set. These data are representing an n -dimensional vector in the phase space. Each data point represents one dimension. This type of vectors is complicated to present in the form of a pattern. In the analysis of dynamic systems, the phase-space graph is coming from a delayed dependence in the data set. This means that for each pair of points $(z_k, y_l) \in X$ elements y_l and z_k are linked by the following relationship $y_l = z_{k+T}$ and form two subsets of Z and Y within the set X , defined by indexes $k = 1, 2, \dots, n - T$ and $l = T, T + 1, \dots, n$, where T is the phase shift. This constant T is an integer. The theory of mutual information is usually used to approximate the

model with a dimension of 128×128 pixels. We are talking here about the image representing the phase space for a given time signal. If we have a regular orbit in the phase space, we know that we have harmonic oscillations in our dynamic system (Fig. 2a). If the system behaves chaotically, we are unable to see any regular orbit in the phase space plot (Fig. 2b). In this representation for the EEG signal, we see irregular orbits, and the graph is less chaotic than in the case of random data (Fig. 2c). On the black and white picture, we can not distinguish the beginning and the end of the line in the plot of created phase space. The truth is that time is not visible for this pattern. To increase the visibility of the time direction, we have introduced colors to the model in such a way that the colors of the trajectory lines change continuously from red to blue. We can convert the total time values into an RGB vector using these three simple functions

$$F_R(t) = \begin{cases} t & 0 \leq t < 256 \\ 255 & 256 \leq t < 512 \\ -t & 512 \leq t < 768 \\ 0 & 768 \leq t \end{cases}, \quad F_G(t) = \begin{cases} 0 & 0 \leq t < 256 \\ t & 256 \leq t < 512 \\ 255 & 512 \leq t < 1024 \\ -t & 1024 \leq t < 1280 \end{cases},$$

$$F_B(t) = \begin{cases} 0 & 0 \leq t < 768 \\ t & 768 \leq t < 1024 \\ 255 & 1024 \leq t < 1280 \\ -t & 1280 \leq t < 1536 \end{cases}$$

value of T [27]. This value is commonly between 3 and 7 for EEG signals [28, 29]. The shared information equation between the sets Z and Y will look as follows.

$$I(Z;Y) = \sum_{k,l} p(z_k, y_l) \log \left(\frac{p(z_k, y_l)}{p(z_k)p(y_l)} \right) \quad (1)$$

where indices $k \in Z$ and $l \in Y$, $p(z_k)$ is the probability of finding z_k in the set Z and $p(z_k, y_l)$ is the probability of finding a pair z_k, y_l in Z and Y sets. For the EEG signal, the sets Z and Y are subsets of the set X , and their constituent elements remain in the relation for $z_k = x_k$ and $y_l = x_{k+T}$ for $k = 1, 2, \dots, n - T$. If we now substitute this relationship to Eq. 1, we get

$$I(Z \subset X; Y \subset X) = \sum_{k=1}^{n-T} p(x_k, x_{k+T}) \log \left(\frac{p(x_k, x_{k+T})}{p(x_k)p(x_{k+T})} \right). \quad (2)$$

Minimization of this auto-mutual information (AMI) function concerning T estimates the optimal value for the delay with the given data stream [30]. Based on the delay T , we can create a new pattern on the 2D image. It is a set of pairs of points (x_k, x_{k+T}) scaled so that it fits in a

where $t \in \mathbb{N}$. In Fig. 2c, we see an example of a colored trajectory in the phase space (the first image from the right). This method is less expensive (time-consuming) than entering another dimension, as is the case in the 3D phase space classification [31]. Another interpretation of the phase space comes from classical mechanics in which a two-dimensional phase space is a function of momentum relative to location. The oscillation of the EEG signal is one-dimensional in the sense of a classic oscillator. If we assume that mass is equal to one, then momentum can be equivalent to speed. $v(t) = \lim_{h \rightarrow 0} (x(t+h) - x(t))/h$ of change of the oscillating signal. Using the graph $v(x)$, we can create a different pattern representing the EEG signal (Fig. 2d). All conversions from the raw signal to the image were made using the original program developed by the author of this work. [32].

5 Model of neural network

Artificial neural networks are currently very successful in such a problem as the classification of features. The main reasons for this are the development of the convolutional neural network [33], deep learning method [34], and fast

Fig. 3 The diagram of the neural network used in this project

increase of computing power, especially the use of graphics cards for numerical calculations [35]. There is the use of CNN in recognition of handwriting [36], speech [37], and drug molecules [38]. The 2D CNN model used in this study consists of 6 layers. The CNN input tensor has the shape of $512 \times 128 \times 3$ or $128 \times 128 \times 3$, but the output tensor has the form of 1×3 . The output represents the probability of finding one of three types of EEG signals; LIP, JAW, and NONE. There are four convolutional layers and two fully connected layers in this model (Fig. 3). The first mentioned type of layer consists of three elements, such as a convolution filter, a subsample, and an activation function. The purpose of the convolution filter is to detect the presence of specific features or patterns present in the original data. The first layer in the CNN network was designed to detect (large) functions that are reasonably easy to recognize and interpret. Subsequent CNN layers are used to understand more advanced features that are more abstract. The max-pooling technique enables you to distinguish features of objects regardless of where and from what angle they appear in the image. In this network, the max-pooling operation involves downloading the maximum value from a 5×5 square matrix and creating a subsample of data. The goal is to sample the input representation, which in turn reduces its dimensionality. Max-pooling also reduces the number of free parameters in the model and calculation costs. The max-pooling output goes to the rectified linear unit (ReLU) [39], which in this case, is the activation function. Fully connected layers are responsible for the classification process. We used a

Table 1 The parameters of 1D and 2D convolutional neural networks

Name of layer	Filter size	#Filters	#Neurons	Stride	Padding
Conv1	5	32	$\times$	1	2
Conv2	5	32	$\times$	1	2
Conv3	5	64	$\times$	1	2
Conv4	4	64	$\times$	1	1
Fc5	1	$\times$	4096	1	0
Fc6	1	$\times$	256	1	0

similar model for CNN 1D. The only difference is that we have used a one-dimensional convolution filter. The input to this network is a vector of integers with the size of 512 data points. In Table 1 are the parameters of our system. We applied the Adam algorithm in the training process of our network [40]. This algorithm is a variant of the stochastic gradient decrease algorithm. The selected unit of time in the process of learning the system is an epoch. An epoch is defined as a single pass through the entire training set while training a machine learning model. The initial learning rate was equal to 2×10^{-5} . After every ten epochs, its value was divided by five. Initial weights and biases are random with 0.05 standard deviation. The random seed was equal to zero for all calculations. When the random seed is the same for all representations of EEG signals in the learning process, we can compare the results. The training process of our network has been carried out until the loss of validation decreases. Learning time varies

from 60 to 300 epochs. The TensorFlow library was used to build the CNN network [41]. Python programs using 1D and 2D CNN were placed in the GitHub repository to reproduce the results [42].

6 Results and discussion

The recognition of short-duration EEG signals (less than a second) is a fundamental part of building multiple systems for brain-computer interaction. Any improvements in classifying short-timed signals can help to create more useful applications. One of the advances in signal classification can be the method of phase space reconstruction. We want to show the PSR's ability to create more distinctive functions between different signal groups. The time delay T between the raw signal data points plays a fundamental role in the PSR. First, we want to calculate the mutual information (Eq. 2) between successive data points for each signal class. For this purpose, we collected 4000 samples for each group. Next, we calculate AMI concerning the delay T in the range from 1 to 200. Then we record the value of the delay T for the minimum AMI in the histogram of the probability $P_{AMI}(T)$. The distance T for the maximum of $AMI(T)$ function indicates the most independent data in the data set. Figure 4 presents the probability results for three facial muscle states (JAW, LIP, NONE). The results show that the most likely value of the T -delay for jaw and lip movements is 5. The maximum values of these two probabilities are

almost the same with a slightly higher value in the case of lip motion signals. The probability of the lack of tension in the facial muscles is much flatter than in the two previous functions. The function reaches its maximum at $T = 13$. It means that the distance between independent data [30] is much more prominent. The reason for this may be a lower frequency of signals. Looking again at Fig. 4, we see that the case with the lack of facial muscle tension is significantly different from those in which facial muscle tension occurs. Another significant question that we want to answer in this work concerns the relationship between the T -delay step and the accuracy of recognizing signals that do not participate in learning the neural network. To achieve this goal, we have prepared 400 images of the EEG signal in the phase space with a given T shift with a data set for each class (JAW, LIP, NONE). The network was set to take 20% of the images as a validation (test) set, what gives 80 images per class. The rest of the pictures (80%) take part in the machine learning process. The seed of random algorithm is constant for all calculations of the network weights. We have applied the same T -delay shift for all classes. The program is recording validation accuracy value for a given set of images at the point of minimum loss of validation. The presented here validation accuracy comes from averaging over ten different sets of EEG signal data. We have repeated this calculation for nine values of T -delay steps ranged from 1 to 9. The validation accuracy has been estimated both for black&white and colored images. In general, the increase of T -delay in phase space

Fig. 4 Minimization of auto-mutual information concerning time delay T for three different EEG signals

patterns makes a decrease in the validation accuracy of the neural network both for B&W and colored images (Fig. 5). We can observe the best value of validation accuracy for $T=1$. For colored trajectories, validation accuracy reached a maximum value (90.54%) (Table 2). It means that painted phase space trajectories give better results at low T values. Other valuable information for future application of given patterns is the learning time. As we know, the learning rate can be adjusted to provide a longer or shorter learning time. We have here applied the same rule of the learning rate for all types of patterns to compare the learning time results. The learning process in our experiment finishes when validation loss reaches its minimum value. Although the best accuracy is at $T=1$, the learning time at this point is the longest one both for B&W and colored trajectories (Fig. 6). For $T=1$, the neural network learning time is 2160 epochs. Next, the learning time falls exponentially to 950 for $T=9$. If we look at the

inset of Fig. 6, we can see that the best ratio of accuracy to learning time is for $T=4$ and $T=6$, which is close to maxima of class JAW and LIP in $P_{AMI}(T)$ probability. In Fig. 7 we see the diversity of data samples. The learning time of PSR pattern for $T=1$ varies with the sample number, randomly. Generally, the process of learning a colored model takes longer than the process of learning a black and white pattern. The reason for this may be the definition of an input tensor. For black and white images, we have binary data, but for color images, the input tensor accepts many different floating-point numbers from 0 to 1. Another definition of phase space gives classical mechanics, where the x-axis represents the position and y-axis velocity of an oscillating point. Speed is a derivative of location. We calculated this value as the difference between successive EEG signal values divided by a fixed period of $1/512$ s. Figure 2d is an example of this phase space trajectory. The validation accuracy of these

Fig. 5 The dependency of validation accuracy of time delay for monochromatic and color phase space patterns

Table 2 Classification accuracy of different patterns averaged over all samples

Pattern name	Image size	Validation accuracy (%)	Error (%)
Raw signal 2D	512 × 128	85.67	8.91
Classical mechanics phase space 2D	128 × 128	86.79	5.39
Classical mechanics phase space (color) 2D	128 × 128	87.08	5.01
Shifted phase space [t + 1] 2D	128 × 128	86.12	5.01
Shifted phase space [t + 1] (color) 2D	128 × 128	90.54	3.98
1D CNN	–	87.37	3.77

Fig. 6 Graph of learning time dependent on delay for black and white and colored phase space trajectories. The insert shows the accuracy divided by the learning time relative to the step of delay

Fig. 7 The learning time dependence for nine different samples

images we can see in Fig. 8. The difference between B&W and colored images is less pronounced in this case. If we look at the sample series, the B&W and color patterns follow the same path. In contrast, the PSR constructed as T-delay in data series shows some striking differences in sample 1 and 7 (Fig. 8). The validation accuracy is, on average, 4.5% higher for colored than for B&W images.

For comparison, the same figure shows the results for a one-dimensional convolution network. We can see that the 1D CNN also gives worse results when recognizing signals than colored PSR. Another way to measure the quality of the introduced algorithm is to determine the confusion matrix [43]. In Fig. 9, we see the confusion matrices calculated from the model based on the fourth

Fig. 8 The 1D CNN, classical mechanics and data shifted phase space plot against sample number

Fig. 9 The 1D CNN, classical mechanics and data shifted phase space confusion matrices of sample four

	LIP	JAW	NONE
LIP	72	0	1
JAW	31	57	1
NONE	3	1	74

(a) 1D CNN

	LIP	JAW	NONE
LIP	64	4	0
JAW	8	76	0
NONE	0	1	87

(b) PSR Dynamics B&W

	LIP	JAW	NONE
LIP	59	9	0
JAW	8	76	0
NONE	0	1	87

(c) PSR Dynamics Color

	LIP	JAW	NONE
LIP	62	6	0
JAW	9	75	0
NONE	0	1	87

(d) PSR T=1 B&W

	LIP	JAW	NONE
LIP	65	3	0
JAW	1	83	0
NONE	0	1	87

(e) PSR T=1 Color

sample. In this test sample, we have 240 random patterns of three different kinds. The smaller the value on the extra-diagonal of the confusion matrix, the higher the quality of the signal classification. The best quality of the model we can observe in the case of PSR with $T=1$ and introduction of coloring algorithm (Fig. 9e).

The total learning time, counted in epoch, for all types of images we can observe in Fig. 10. Looking at that figure, we can conclude that the phase space reconstruction based on classical mechanics is less demanding in the case of the learning process. The learning process lasted the longest for 1D CNN. On the other hand, the time for constructing 1D vectors is considerably shorter than the time

Fig. 10 Total learning time of different patterns

for creating phase space images. Vector construction time is very significant in systems for recognizing short-time signals.

7 Conclusions

In summary, the phase space reconstruction method is useful in the classification of short-timed EEG signals. We found that for the delay $T = 1$ in the time series, the precision of the validation of the EEG test signals has the highest value. We have also proved that incorporating color into PSR images increases the overall accuracy of validation of the test set of EEG data for the tested network by 4.87% compared to raw signal images and 3.17% compared to 1D CNN. These studies show that even with such a simple EEG device, facial expression signals are fairly well recognizable. In future research, we want to combine methods related to the reconstruction of the phase space and neural networks to create a new network layer aimed at the recognition of EEG signals.

Acknowledgement The author would like to thank Professor Walica's fund for supporting this work.

Compliance with ethical standard

Conflict of interest The author declares that there is no conflict of interests.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided 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.

References

1. Haas L (2003) Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. *J Neurol Neurosurg Psychiatry* 74:9. <https://doi.org/10.1136/jnnp.74.1.9>
2. Berger H (1929) Über das Elektrenkephalogramm des Menschen. *Arch Für Psychiatr Nervenkrankh* 87:527–570. <https://doi.org/10.1007/BF01797193>
3. Visser B, van Dieën JH (2006) Pathophysiology of upper extremity muscle disorders. *J Electromyogr Kinesiol* 16:1–16. <https://doi.org/10.1016/j.jelekin.2005.06.005>
4. Stålberg E, Dioszeghy P (1991) Scanning EMG in normal muscle and in neuromuscular disorders. *Electroencephalogr Clin Neurophysiol Potentials Sect* 81:403–416. [https://doi.org/10.1016/0168-5597\(91\)90048-3](https://doi.org/10.1016/0168-5597(91)90048-3)
5. Au SK, Bonato P, Herr H (2005) An EMG-position controlled system for an active ankle-foot prosthesis: an initial experimental study. In: 9th International conference on rehabilitation robotics, 2005. ICORR 2005, pp 375–379. <https://doi.org/10.1109/icorr.2005.1501123>
6. Rafiee J, Rafiee MA, Yavari F, Schoen MP (2011) Feature extraction of forearm EMG signals for prosthetics. *Expert Syst Appl* 38:4058–4067. <https://doi.org/10.1016/j.eswa.2010.09.068>
7. Daimi SN, Saha G (2014) Classification of emotions induced by music videos and correlation with participants' rating. *Expert Syst Appl* 41:6057–6065. <https://doi.org/10.1016/j.eswa.2014.03.050>

8. Hess U, Blairy S (2001) Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *Int J Psychophysiol* 40:129–141. [https://doi.org/10.1016/S0167-8760\(00\)00161-6](https://doi.org/10.1016/S0167-8760(00)00161-6)
9. Lee S-H, Lim JS, Kim J-K, Yang J, Lee Y (2014) Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance. *Comput Methods Programs Biomed* 116:10–25. <https://doi.org/10.1016/j.cmpb.2014.04.012>
10. Klonowski W (2002) Chaotic dynamics applied to signal complexity in phase space and in time domain. *Chaos Solitons Fractals* 14:1379–1387. [https://doi.org/10.1016/S0960-0779\(02\)00056-5](https://doi.org/10.1016/S0960-0779(02)00056-5)
11. Mayer-Kress G (ed) (1986) Dimensions and entropies in chaotic systems: quantification of complex Behavior proceeding of an international workshop at the Pecos River ranch, New Mexico, September 11–16, 1985, Springer, Berlin. www.springer.com/br/book/9783642710032 Accessed 7 Sept 2018
12. Natarajan K, Acharya R, Alias F, Tiboleng T, Puthusserypady SK (2004) Nonlinear analysis of EEG signals at different mental states. *Biomed Eng Online* 3:7. <https://doi.org/10.1186/1475-925x-3-7>
13. Fell J, Rösche J, Beckmann P (1993) Deterministic chaos and the first positive Lyapunov exponent: a nonlinear analysis of the human electroencephalogram during sleep. *Biol Cybern* 69:139–146. <https://doi.org/10.1007/BF00226197>
14. Lutzenberger W, Preissl H, Pulvermüller F (1995) Fractal dimension of electroencephalographic time series and underlying brain processes. *Biol Cybern* 73:477–482. <https://doi.org/10.1007/BF00201482>
15. Atkinson J, Campos D (2016) Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst Appl* 47:35–41. <https://doi.org/10.1016/j.eswa.2015.10.049>
16. Shi T, Wang H, Zhang C (2015) Brain Computer Interface system based on indoor semi-autonomous navigation and motor imagery for Unmanned Aerial Vehicle control. *Expert Syst Appl* 42:4196–4206. <https://doi.org/10.1016/j.eswa.2015.01.031>
17. Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. *Expert Syst Appl* 36:3465–3469. <https://doi.org/10.1016/j.eswa.2008.02.064>
18. Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassis GD (2008) Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. *Neural Netw* 21:427–436. <https://doi.org/10.1016/j.neunet.2007.12.031>
19. Gupta V, Priya T, Yadav AK, Pachori RB, Rajendra Acharya U (2017) Automated detection of focal EEG signals using features extracted from flexible analytic wavelet transform. *Pattern Recognit Lett* 94:180–188. <https://doi.org/10.1016/j.patrec.2017.03.017>
20. Satapathy SK, Dehuri S, Jagadev AK (2017) EEG signal classification using PSO trained RBF neural network for epilepsy identification. *Inform Med Unlocked* 6:1–11. <https://doi.org/10.1016/j.imu.2016.12.001>
21. Mert A, Akan A (2018) Emotion recognition from EEG signals by using multivariate empirical mode decomposition. *Pattern Anal Appl* 21:81–89. <https://doi.org/10.1007/s10044-016-0567-6>
22. Jiao Z, Gao X, Wang Y, Li J, Xu H (2018) Deep convolutional neural networks for mental load classification based on EEG data. *Pattern Recognit* 76:582–595. <https://doi.org/10.1016/j.patcog.2017.12.002>
23. Neurosky EEG Sensors—EEG Headsets | NeuroSky, (n.d.). <http://neurosky.com/biosensors/eeeg-sensor/biosensors/>. Accessed 7 Sept 2018
24. Crowley K, Sliney A, Pitt I, Murphy D (2010) Evaluating a brain-computer interface to categorise human emotional response. In: 2010 10th IEEE international conference on advanced learning technologies, pp 276–278. <https://doi.org/10.1109/icalt.2010.81>
25. Oostenveld R, Praamstra P (2001) The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol* 112:713–719. [https://doi.org/10.1016/S1388-2457\(00\)00527-7](https://doi.org/10.1016/S1388-2457(00)00527-7)
26. Dawid A (2019) This is one channel EEG dataset measured by Mindwave mobile device: alex386/EEG_Dataset_Dawid. https://github.com/alex386/EEG_Dataset_Dawid. Accessed 11 Jan 2019
27. Fraser AM (1989) Information and entropy in strange attractors. *IEEE Trans Inf Theory* 35:245–262. <https://doi.org/10.1109/18.32121>
28. Watt RC, Hameroff SR (1988) Phase space electroencephalography (EEG): a new mode of intraoperative EEG analysis. *Int J Clin Monit Comput* 5:3–13
29. Sharma R, Pachori RB (2015) Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Syst Appl* 42:1106–1117. <https://doi.org/10.1016/j.eswa.2014.08.030>
30. Pires CAL, Perdigão RAP, Pires CAL, Perdigão RAP (2012) Minimum mutual information and non-gaussianity through the maximum entropy method: theory and properties. *Entropy* 14:1103–1126. <https://doi.org/10.3390/e14061103>
31. Noponen K, Kortelainen J, Seppänen T (2009) Invariant trajectory classification of dynamical systems with a case study on ECG. *Pattern Recognit* 42:1832–1844. <https://doi.org/10.1016/j.patcog.2008.12.008>
32. Dawid A (2018) EEGPatternizer: An EEG signals converter to image pattern written in C# language: alex386/EEGPatternizer. <https://github.com/alex386/EEGPatternizer>. Accessed 13 Oct 2018
33. Zhang W, Itoh K, Tanida J, Ichioka Y (1990) Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl Opt* 29:4790–4797. <https://doi.org/10.1364/AO.29.004790>
34. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86:2278–2324. <https://doi.org/10.1109/5.726791>
35. Jang H, Park A, Jung K (2008) Neural network implementation using CUDA and OpenMP. In: 2008 Digital image computing techniques and applications, pp 155–161. <https://doi.org/10.1109/dicta.2008.82>
36. Niu X-X, Suen CY (2012) A novel hybrid CNN-SVM classifier for recognizing handwritten digits. *Pattern Recognit* 45:1318–1325. <https://doi.org/10.1016/j.patcog.2011.09.021>
37. Abdel-Hamid O, Mohamed A, Jiang H, Penn G (2012) Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In: IEEE International conference on acoustics, speech and signal processing (ICASSP), pp 4277–4280. <https://doi.org/10.1109/icassp.2012.6288864>
38. Gawehn E, Hiss JA, Schneider G (2016) Deep learning in drug discovery. *Mol Inform* 35:3–14. <https://doi.org/10.1002/minf.201501008>
39. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
40. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization, arXiv1412.6980 Cs. <http://arxiv.org/abs/1412.6980>
41. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser R, Kudlur M, Levenberg J, Mane D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viegas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2016) TensorFlow: large-scale machine learning on heterogeneous distributed systems, arXiv1603.04467 Cs. <http://arxiv.org/abs/1603.04467>

42. Dawid A (2018) EEGPatternRecognition: Tensorflow CNN for image recognition: alex386/EEGPatternRecognition. <https://github.com/alex386/EEGPatternRecognition>. Accessed 13 Oct 2018
43. Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>

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