



# Convolutional neural network for detection and classification of seizures in clinical data

Tomas Iešmantas<sup>1</sup> · Robertas Alzbutas<sup>1</sup>

Received: 15 July 2019 / Accepted: 31 May 2020 / Published online: 12 June 2020  
© International Federation for Medical and Biological Engineering 2020

## Abstract

Epileptic seizure detection and classification in clinical electroencephalogram data still is a challenge, and only low sensitivity with a high rate of false positives has been achieved with commercially available seizure detection tools, which usually are patient non-specific. Epilepsy patients suffer from severe detrimental effects like physical injury or depression due to unpredictable seizures. However, even in hospitals due to the high rate of false positives, the seizure alert systems are of poor help for patients as tools of seizure detection are mostly trained on unrealistically clean data, containing little noise and obtained under controlled laboratory conditions, where patient groups are homogeneous, e.g. in terms of age or type of seizures. In this study authors present the approach for detection and classification of a seizure using clinical data of electroencephalograms and a convolutional neural network trained on features of brain synchronisation and power spectrum. Various deep learning methods were applied, and the network was trained on a very heterogeneous clinical electroencephalogram dataset. In total, eight different types of seizures were considered, and the patients were of various ages, health conditions and they were observed under clinical conditions. Despite this, the classifier presented in this paper achieved sensitivity and specificity equal to 0.68 and 0.67, accordingly, which is a significant improvement as compared to the known results for clinical data.

**Keywords** Epilepsy · Seizure detection · Convolutional neural networks · Deep learning · EEG

## 1 Introduction

Almost two decades ago, it was observed that 63% of patients under medication in an epilepsy service were rendered as seizure-free patients [1]. Since then, little has changed; despite that new antiepileptic drugs became available in the market, it was found [2] that the percentage of patients getting rid of seizures have changed only slightly. Better drug delivery systems to the patient organism might help to increase the efficiency of epilepsy treatment. In addition, considering the possible toxicity of the epilepsy drugs [3], a smart system, able to predict seizure in advance and deliver a small dosage of drugs to interrupt the development of that seizure, would be a significant addition to the arsenal. At least one could expect a system, able to detect the onset of the seizure and inform

patient or medical personnel about it—this would help to reduce the stress or development of more severe physical or mental conditions like depression [4].

In fact, many such systems have already been patented. Patents describe various devices, able to detect and predict the developing seizures using nonlinear chaotic time series techniques [5], or by tracking the time evolution of chaotic entrainment [6]. Yet, another concept uses a wavelet transform maximum modulus algorithm applied to data from the entire body [7] or tracks impedance of the brain [8]. These examples barely scratch the surface of many patented tools (devices/methods/algorithms) able to detect (at least for the dataset they were validated) and predict seizures.

A large number of patents correlate with an even larger set of research reporting high accuracy of seizure detection and prediction. For example, F. Morman et al. [9] reported results of seizure detection for five epilepsy patients as follows: area under the curve (AUC), as a measure of the accuracy of the algorithm, equal to 0.83, with sensitivity and specificity measure correspondingly equal to 0.95 and 0.65. This algorithm was trained on intracranial EEG data first extracting so-called mean phase coherence, more commonly known as phase locking

---

✉ Tomas Iešmantas  
tomas.iesmantas@gmail.com

<sup>1</sup> Department of Mathematics and Natural Sciences, Kaunas University of Technology, 44249 Kaunas, Lithuania

value, which measures the degree of brain synchronisation. M. J. Cook et al. [10] trained a classification algorithm on intracranial EEG data of 15 patients and obtained sensitivity 0.78, although authors did not report what kind of EEG dataset features were used. M. D'Alessandro et al. [11] considered prediction of seizures for four patients based on intracranial EEG data—authors reported sensitivity equal to 0.625 and specificity equal to 0.91; numerous features were calculated every 1 min with 2.5-s displacement and extracted considering the time domain, frequency domain and nonlinear dynamics.

Often used Freiburg dataset consisting of 21 patients (data collected at Epilepsy Center of the University Hospital of Freiburg, Germany and no longer freely available, <http://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database>) was used in various studies with variable, although very high accuracies:

- [12] reports 0.975 sensitivity of algorithm, trained on spectral power features (20 s window of intracranial EEG).
- [13] obtained 0.71 sensitivity with perfect specificity. Authors used convolutional neural networks on various features (20 min window).
- [14] reports 0.9 sensitivity and 0.21 false positives per hour.

High accuracy was reported on scalp EEG as well by A. Shoeb [15], where the author considered seizure prediction for 22 pediatric patients and sensitivity equal to 0.9 was obtained. On the same dataset (from scalp EEG), S. Nasehi and H. Pourghassem [14] obtained sensitivity equal to 0.98 with 0.06 false positives per hour.

The research on seizure detection is very active, and during recent years, there have been many different approaches. Convolutional neural networks are being applied to this problem quite extensively nowadays with very promising results (see Table 1).

From this brief review of seizure detection results, one can make several observations: most of the results are reported for

intracranial EEG data, records are obtained under laboratory conditions, often during long-term monitoring for presurgical evaluation. High accuracies were achieved for seizure vs non-seizure classification algorithms for intracranial EEG. Features, representing a state of the brain, are often calculated using time windows significantly larger than 1 s, even though this is a time resolution often used in the fine-grained analysis of EEG [23].

EEGs are not the only option to use in seizure reporting. Other modalities may also be used for this purpose, like accelerometry [24], extracerebral movements [25], audio recordings [26], worn wrist devices [27], etc. (more broad review of the methods can be found in [28]). However, in this paper, the focus is on EEG signals.

However, given all the advances in seizure prediction and all the patents for seizure detection/prediction systems, then one may ask why commercially available tools perform only at around 0.3 sensitivity with high false positive rate [29]. The answer probably lies in the fact that most of the results of various classifiers were obtained for “nice” EEG data, i.e. EEG obtained under laboratory conditions when seizures observed are often of the same type, and they are within homogeneous age groups. A similar concern was recently voiced by D. R. Freestone et al. [30], where authors noted that “data that are recorded in the epilepsy monitoring unit are not representative of ‘normal’ seizures”. Also, as observed above, many algorithms were evaluated on intracranial EEG data, and it is known that intracranial electrodes can change the dynamics of seizures [31].

A recently released dataset and still actively under further development as Temple University Hospital (TUH) EEG Corpus [32], which is the world’s largest publicly available database of clinical EEG data, might fill the void of more realistic seizure data—it was obtained under clinical conditions, spans over different age groups, different types of seizure events, different health conditions of patients. The first reported benchmark based on this database produced the following results [23]: sensitivity (0.292) and specificity (0.667). Authors used a classification algorithm based on the hidden

**Table 1** Convolutional neural network-based seizure detection from EEG features

Reference (year)	Accuracy measures (accuracy—A; sensitivity—Se; specificity—Sp; false positives per hour—FP/h)	Time window length, seconds	EEG type
[16] (2018)	A = 88.67%, Se = 95%, Sp = 90%	23.6 s	Intracranial (Freiburg)
[17] (2019)	Se = 86.6%, 0.84 FP/h	5 s	Scalp (KK WCH)
[18] (2019)	Se = 89.4%, 0.5 FP/h and Se = 97%, 0.6 FP/h	30 s	Scalp (CHB-MIT) and Scalp (SNUH)
[19] (2018)	Se = 88.9%, Sp = 93.78%	10 s	Scalp (laboratory conditions)
[20] (2018)	Se = 87.8%, 0.142 FP/h	1 s	Scalp (MSSM) + Scalp (CHB-MIT)
[21] (2018)	Se = 81.4%, 0.06 FP/h and Se = 81.2%, 0.16 FP/h	30 s	Intracranial (Freiburg) and Scalp (CHB-MIT)
[22] (2020)	Se = 100%, in average Sp = 94.39%	5 s	Scalp (CHB-MIT)

Markov chain model. The low accuracy of these methods shows their inadequacy for this TUH EEG Corpus. However, it is the first baseline to overcome for other machine learning algorithms.

Hence, the purpose of this study was to probe this clinical reference database with algorithms, which have shown high accuracy of seizure detection on other specific datasets. More specifically, we selected convolutional neural network as a method, which showed most promise when applied to other datasets (see Table 1 above). We adapted the pattern construction from EEGs in order to be able to use very short 1-s epochs, as opposed to most of the other studies. The main contribution of this paper is the application and investigation of convolutional neural network architectures to the very large and highly heterogeneous EEG dataset. It must be stressed that this dataset was obtained not under laboratory conditions, but in various intensive care units. The proposed approach of classification enabled to achieve sensitivity and specificity equal to 0.68 and 0.67, respectively—a considerable improvement from the known result on this dataset.

The structure of the paper is as follows: in Sect 2, the considered TUH EEG Corpus database is described in detail. Also, features of brain synchronisation, entropy and power spectrum, as well as brain activity patterns and methods relevant for the considered convolutional neural network, are defined and described. In Sect. 3, an extensive analysis of TUH EEG Corpus by the convolutional neural network classifier is presented.

## 2 Material and methods

### 2.1 Data

As was pointed out in Sect. 1, most seizure detection and prediction algorithms were validated on the relatively “nice” and homogeneous EEG datasets. However, one clearly cannot expect to have always the same type of seizures and the same conditions under which EEGs are recorded—clinical setting produces much more variable EEG recordings (influenced by the environment, equipment, electrode location, noise, etc.) as opposed to the tightly controlled research setting. Therefore, there was a need for a dataset, which would become a testing ground for the machine learning algorithms. The Temple University Hospital EEG Corpus [32] could be such a reference dataset. The entire corpus contains 16,986 sessions from 10,874 unique patients, and each session additionally includes one physician report.

Part of the corpus is devoted specifically to the seizure detection problem (including classification by seizure types). It is also subdivided into two segments: training data and evaluation data. The training dataset is composed of 196 patients with total seizure time intervals being equal to 14 h and a

non-seizure time interval—244 h. The evaluation dataset is constructed from recordings of 50 unique patients with total seizure and non-seizure time intervals being equal to 15 h and 152 h correspondingly. Furthermore, there were several types of seizures observed in the training and evaluation datasets (see Table 2), which also adds up to the variability of the dataset.

The data set is highly imbalanced: only 6% of all the data points are annotated as belonging to one of the seizures described in Table 2. This affects sensitivity as well as specificity severely. Therefore, these measures will be reported together with overall AUC (area under curve) measure, which is not sensitive to the balance of data.

### 2.2 Seizures: synchronisation, entropy and power spectrum

In this subsection, several features of EEG dataset are defined. In previous studies, these features demonstrated as carrying a significant amount of information regarding the differentiation between seizure and non-seizure events, namely phase-locking value [33], average power and entropy for different EEG frequency bands. After the definition follows the initial exploration of these features regarding distributions in the seizure/non-seizure classes.

One of the most investigated aspects of seizures is the phenomena of synchronisation. Spontaneous seizures have been interpreted as abnormal excitability and synchronisation of large neuronal populations [34, 35]. However, recent advances have shown [36] that the picture is more complicated: desynchronisation at the onset of the seizure has been observed in neural activity both in vitro and in vivo [37, 38]. This initial desynchronisation is followed by the increase of the synchronisation until it reaches its peak during the final moments of the seizure, and at last, the seizure simultaneously terminates over large areas of the brain [39, 40]. Thus, there is a complex interplay of synchronisation and desynchronisation of the brain during the evolution of the seizure.

On the other hand, the seizure is not the only time when synchronisation occurs. In general, it has been hypothesised [41] that synchronisation phenomena are inherent in the brain’s large-scale integration. For example, in the visual binding problem [42], when the different attributes of an object brought together in a unified representation given that its various attributes—edges, colour, motion, texture, depth and so on—are treated separately in specific visual areas [42]. Also, the increase in synchronisation has been observed during the following activities: flicker stimulation [43], process related to pattern vision [44] and reading [45]. Given this brief review, it could be hypothesised that synchronisation features possibly are not very strong predictors as in the cases with more homogeneous seizure EEG datasets. Also, the relative success of such features, as described in the introduction,

**Table 2** Types of seizures considered in the training and evaluation of TUH EEG Corpus dataset

Type of seizure	Brief description
1. Focal non-specific seizure	Focal seizures which cannot be specified by its type
2. Generalised non-specific seizure	Generalised seizures which cannot be further classified into one of the groups below
3. Simple partial seizure	Partial seizures during consciousness; type specified by clinical signs only
4. Complex partial seizure	Partial seizures during unconsciousness; type specified by clinical signals only
5. Absence seizure	Absence discharges observed on EEG; patient loses consciousness for a few seconds (Petit Mal)
6. Tonic seizure	Stiffening of the body during a seizure (EEG effects disappears)
7. Clonic seizure	Jerking/shivering of the body during a seizure
8. Tonic-clonic seizure	At first stiffening and then jerking of the body (Grand Mal)

See webpage [https://www.isip.piconepress.com/projects/tuh\\_eeg/downloads/tuh\\_eeg\\_seizure/v1.1.1/](https://www.isip.piconepress.com/projects/tuh_eeg/downloads/tuh_eeg_seizure/v1.1.1/)

probably is simply due to the aforementioned EEG data cleanliness, which is not the case in the TUH EEG Corpus case.

One of the commonly used measures of dynamic system synchronisation is the *phase-locking value* (PLV). The locking of phases of two oscillators occurs when the phase entrainment condition is satisfied:

$$\phi = |\varphi_1 - \varphi_2| < \text{const.},$$

where  $\varphi$  is the instantaneous phase of the oscillator and  $\phi$  is their relative phase. A phase entrainment condition can be defined more generally, but this definition is sufficient for our purposes. Instantaneous phases can be estimated in various ways, but two methods are mostly used: a signal is transformed by Hilbert transform or using complex Gabor wavelets. However, it was shown [46] that these methods are equivalent. For our calculations, the Hilbert transform method was chosen, i.e. initially the original signal  $x(t)$  is filtered by a bandpass filter  $g$  and then this filtered signal  $x_g(t)$  is Hilbert-transformed to obtain a complex signal  $X_g(t) = x_g(t) + iH\{x_g(t)\}$ . The instantaneous phase is  $\varphi = \arg(X_g)$ .

Phase locking value is defined as follows

$$\text{PLV} = \left| \frac{1}{N} \sum_{n=1}^N e^{i\phi_n} \right| \quad (1)$$

where  $\phi_n$  is the relative phase at an  $n^{\text{th}}$  time point;  $N$  is the length of a time interval over which averaging is performed.

Another measure that quantifies the strength of synchronisation is based on *Shannon entropy* and is defined [47] as

$$\rho = \frac{S_{\max} - S}{S_{\max}}, \quad S = - \sum_{k=1}^K p_k \ln p_k \quad (2)$$

where  $S$  is the entropy of the distribution of the relative phase  $\phi$ ,  $S_{\max} = \ln K$ ,  $K$  is the number of bins.

Another kind of feature is spectral covariance or coherence. However, these features will not be used in the study. It is due to the fact that these features are suitable only for stationary signals [46], and one cannot expect this stationarity from EEG records.

In addition to the PLV and  $\rho$  features, a *power spectrum* of EEG is taken into consideration. The power spectrum was divided into seven frequency bands (i.e. delta, theta, alpha, low beta, high beta, low gamma and high gamma) and logarithms of averages over each of the band are calculated. The reasoning behind the power spectrum consideration as an informative feature is the changes in spectral power during various phases of the seizure evolution, e.g. the increase of spectral power of high frequencies during the start of the seizure has been observed [48].

## 2.3 Patterns of brain activity

In the introduction, the importance of short detection window (1–2 s, [23]) was discussed as most of the algorithms rely on the longer time windows: lengths vary from 20 s [12] to 5 min [13]. Broader detection horizons (or epochs) enable extraction of more features from time subintervals in one epoch. Thus, a single feature vector embodies not only a momentary state of the brain during the seizure but also the specific evolution of it (see the discussion above and in [36] for the different phases of synchronisation changes). On the other hand, a 1-s window (as a length of detection horizon is selected in this study) is too short to reflect significant changes of EEG before, during and after the seizure. Hence, authors were forced to use features that do not reflect the time evolution of the EEG signal, i.e. each time-dependent element in the vector for any single feature reflects the average state over this short detection horizon. This poses a great challenge—the shorter the period, the smaller the amount of information

is contained in it; also, short time periods are the reason for higher variance of feature vectors as obtained using different time windows. A notable example of detection of seizures for very short time windows is the Kaggle challenge [49], where 1-s non-overlapping time windows were also used for seizure detection, and very high accuracies were achieved even with such short time slices of specific EEGs.

Because the combinations of features (as pointed out in the introduction part) play a significant role in the efficiency and accuracy of classification algorithms, in this study, they are used additionally. A feature matrix, rather than feature vector, is extracted from each 1-s time window of EEG—such a matrix is called a *pattern*. The same name is also used by other authors [13] when combinations of features are formed for the application of their convolutional neural network (more details follow below). However, authors [13] had the advantage of having very long detection horizons (5 min): they divided the entire horizon into smaller windows, and each bivariate feature vector is placed into a separate pattern matrix column (i.e. columns represented the time evolution of features). In this paper, the construction of patterns is adapted for short time windows.

Consider 1 s (or any other length) epoch of EEG signal matrix  $X \in \mathbb{R}^{n \times t}$ , where  $n$  is the number of EEG channels and  $t$  is the number of discrete-time moments, which in our case is equal to 250 points—250 Hz was the frequency at which the EEG sampled, at least in the TUH EEG Corpus portion which was devoted to the seizure detection.

Three types of features were calculated for each 1-s epochs (no intersection). For the PLV and Shannon entropy features:

1. First, each EEG channel was filtered in 7 frequency intervals ([50]: delta (< 4 Hz), theta (4–7 Hz), alpha (7–13 Hz), low beta (13–15 Hz), high beta (14–30 Hz), low gamma (30–45 Hz) and high gamma (45–70 Hz));
2. Then for each filtration result, Hilbert transform  $H(\cdot)$  was calculated from which instantaneous phase  $\varphi$  was extracted. In other words, from 1 s of EEG of one channel, one gets 7 time series values of instantaneous phases.
3. The pairwise channel differences of instantaneous phases are calculated to get relative phases. This is done for each of the seven filtration intervals. And then PLVs and Shannon entropies are calculated for each result, collapsing a series of 250 points to a single number.

So, for example, suppose there are 10 channels. Then, for each 1-s time series, there will be  $7 \times 10 = 70$  frequency filtration results. The Hilbert transform is performed and instantaneous phases extracted. Then, take instantaneous phases 10 channels corresponding to, let us say, delta band filtration and obtain  $\frac{10 \times 9}{2} = 45$  pairwise differences—these differences will

be relative phases. Then, for each of 45 relative phases, calculate PLV and Shannon entropies. In other words, there will be  $45 \times 7 = 315$  PLV values and 315 Shannon entropy values.

In order to form a pattern, a picture of the brain state during the particular 1-s epoch, feature values are represented not as a vector but rather as a symmetric square matrix. For the above example, those 45 values of PLV and Shannon entropy values are put into symmetric  $10 \times 10$  format matrix (where entry with address  $(i, j)$  is the feature value obtained from the relative phase of channels  $i$  and  $j$ ).

These patterns are then stacked together on top of each other. Referring to the example, final output for 1-s epoch will be two PLV and Shannon entropy matrices of dimensions  $(7 \times 10) \times 10 = 70 \times 10$ .

To form power spectrum features, take the pairwise differences of each channel, obtain power spectrum for 1-s epoch and then obtain averages of spectrum values over the 7 band intervals. Again, relating to the above example, the final result would be a  $70 \times 10$  matrix. These matrices are calculated for each second of the EEG for each patient.

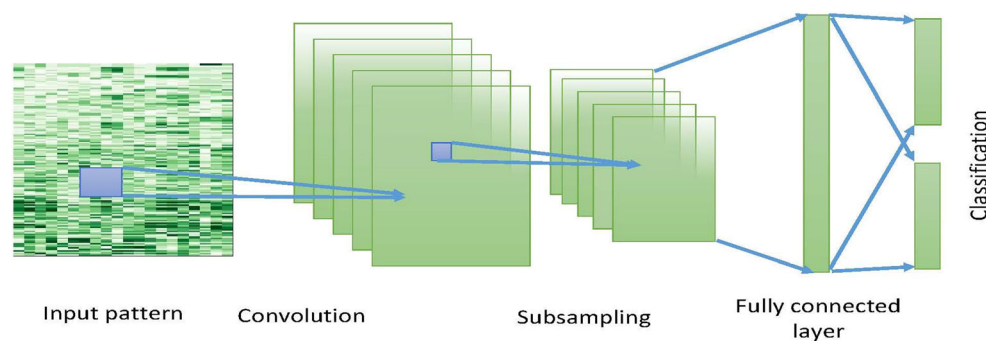
The above-constructed patterns were then used by the learning algorithm. Several feature combination settings were considered for training the algorithm: using only one feature, using pairs of features and using all features together. In the setting when two or three features are used, then their matrices simply are stacked by sides (e.g. if we combine PLV and Shannon entropy feature matrices of dimensions  $70 \times 10$ , then we will get after combining a matrix of size  $70 \times 20$ ). More details on the learning and convolutional neural network (CNN) are given in the next subsection. The reasons of more extensive CNN presentation are that other classification methods like support vector machines or k-nearest neighbour algorithms were already applied in many seizure detection cases, while CNN with specific features is a newer concept in this type of problems (although CNN has a long history in image recognition applications [51, 52]). In addition, this classification method for a long time window showed a significant superiority over other methods in seizure classification tasks (see the reference [13]). So, it is important to focus on it and describe it in greater detail keeping in mind the importance and issues for short time windows.

## 2.4 Convolutional neural network

As was pointed out in the previous subsection, convolutional neural networks are superior over logistic regression and support vector machines—both methods, often in the bucket of machine learning methods in this context. Thus, it is reasonable to introduce CNN in greater detail to facilitate more widespread applications of it within the EEG dataset classification community.

Historically, CNN was developed for image classification and was based on the theory (which is now outdated) on how





**Fig. 1** Visualisation of the simple architecture of the convolutional neural network: one convolutional, one subsampling and two fully connected layers, including the final classification or output layer. The convolutional layer is obtained by convolving the image with a number of small filters. Most often, the size of filters is  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , i.e. very small as

visual cortex works. The significant turning point for the theory of CNN was several seminal works by Yann LeCun [51, 52] in which the application of CNN was made for the recognition of handwritten digits. The original image was convolved with several filters impacting the parts of small size (usually it is  $2 \times 2$ ,  $3 \times 3$ ,  $5 \times 5$ , etc.) to form layers of feature maps. Usually, nonlinear functions like rectified linear unit (ReLU) and sigmoid are applied to the convolved images. Then, these feature maps are subsampled (often by a factor of 2) to form the next layer—at first, it may look like a reduction of useful information, but in fact, it facilitates the generalisation process. One may put as many convolution/subsampling layers as it is necessary. In addition, one can put fully connected layers at the end of the network. A general structure of the convolutional neural network is presented in Fig. 1 below.

Main ingredients of a convolutional neural network are

1. A filter. This is a small matrix (usually of dimensions  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ), with unknown (therefore, learnable) entries. Filter entries can be viewed as neurons in a classical neural network, the only thing is that they have shared parameters. Parameter sharing means that rather than learning a separate set of parameters for every location, only one set is learned. Most often, the numbers of filters in each convolutional layer are 8, 16, 32, etc.
2. The discrete convolution operation is used to convolve the input (initial pattern, if it is the first layer or output from the previous layer) with filters. An analogous operation in classical (or densely connected) artificial neural networks is the multiplication of input by a weight matrix.
3. The nonlinear activation function is applied after the convolution with filters. In a recent CNN application, the so-called rectified linear activation function is used. It became popular after it was shown to enable faster network training for a large-scale image recognition task [53].

compared to the whole image. Subsampling serves as dimensionality reduction operation with the extra advantage being that this operation makes the network position invariant. The last layer produces a two-class probability for non-seizure image and seizure image.

4. Pooling operation. In order to decrease the dimensions of a convolutional layer, one uses so-called pooling operations. This operation replaces the output of a layer at a certain location with a summary statistic of the neighbouring outputs. Most often used are two such summary statistics: maximum of a neighbourhood or an average.

There are no restrictions on how many filters should be within one layer and how many convolutional layers should one put. However, this freedom of choice is not a very good thing, because there are also no guidelines on what architecture will best fit the problem at hand.

Most often, several architectures significantly differing from each other (e.g. obtained by doubling the number of convolutional layers) are evaluated against each other by 10-fold cross-validation procedure: divide the data set into 10 equal parts, use 9 parts for training and 1 part for testing and repeat 10 times—the final architecture is selected by the largest average accuracy.

The training could be done by the variant of the backpropagation algorithm or with gradient-based deep learning environments like TensorFlow [54]. In this study, TensorFlow is used together with Adam optimiser [55] and learning rate equal to  $1e-3$ . To obtain nonlinearity at each convolutional layer, the ReLU (rectified linear unit) function  $f(x) = \max(0, x)$  is used. Later, for the classification layer, the SoftMax function with two classes is used. Then, the output from this last layer is provided as input to the cross-entropy loss function. If the output of SoftMax function is a vector reflecting probabilities of  $K$  classes  $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_K)^T$ , then the loss due to one pattern is equal to

$$L(y, \tilde{y}) = - \sum_{j=1}^K y_j \log \tilde{y}_j,$$

where  $y = (y_1, y_2, \dots, y_K)^T$  is the true label of a pattern.

One may consider using another loss function. For example, it is known that in the case of imbalanced classes, the squared error loss function might work well [56], although in our case, it did not work well.

Finally, in the study, the convolutional neural network was compared with the support vector machine (SVM) algorithm [57]. SVM algorithm is a powerful general purpose algorithm, based on a comparison of observed pattern to support vectors.

## 2.5 Evaluation of detection algorithm

To be able to judge the performance of the detection or classification algorithm, often sensitivity and specificity measures are used [58]. In the study, algorithm sensitivity (also called the true positive rate, the recall or probability of detection), is defined as the probability of correctly anticipated seizure within a time horizon or epoch, and it is vital for judging the performance of seizure prediction. On the other hand, a classical definition of specificity (also called the true negative rate) in the seizure prediction context is the probability of correctly indicating a non-seizure state at any given time [59]. However, one can obtain very high sensitivity by simply tampering with a cut-off threshold, i.e. the point in the interval [0;1] when the judgment about the epoch changes from non-seizure to seizure is performed. Alternatively, another measure, for which the cut-off value is irrelevant, is a receiver operating characteristic (ROC) curve [60] and a derived measure as an area under the ROC curve or area under curve (AUC). ROC is not affected by the imbalance of classes (a very severe issue in all seizure detection cases), and it can be used to obtain an optimal cut-off value (e.g. optimality might be defined as a maximum of a sum of sensitivity and specificity).

In this study, ROC and AUC measures together with optimal sensitivity and specificity measures were used exclusively. AUC measure provided the information on how better the detection algorithm is as compared to random guessing of the classes or labels of epochs.

## 3 Results and discussion

In this section, the results of the study and the main task of seizure detection and classification are discussed. First, the data sample is explored in terms of synchronisation (PLV), Shannon entropy and power spectrum features—discussion on the classification of classes is given, based on distributions of the features mentioned above. Next, the results of the trained classifier are presented. The main task was to classify each 1-s duration epoch to classes “seizure” and “non-seizure”. The more specific results of the classification of all types of seizures defined in Table 1 are also presented and discussed.

Two additional subtasks of smaller scale are considered as well: (1) seizure detection using first 10 s of each seizure, i.e. annotations are modified in such a way, that first 10 s of the seizure epochs could be labeled as “seizure” and the rest as only “non-seizure” epochs; (2) seizure detection using 50% of the testing set when the training set is extended by the first 50% of the test set, i.e. 50% of EEG records for each patient in the test set was added to the training set. The classifier trained in the second subtask is patient-specific and is opposed to the patient-non-specific classifiers trained during the first subtask and other initial cases.

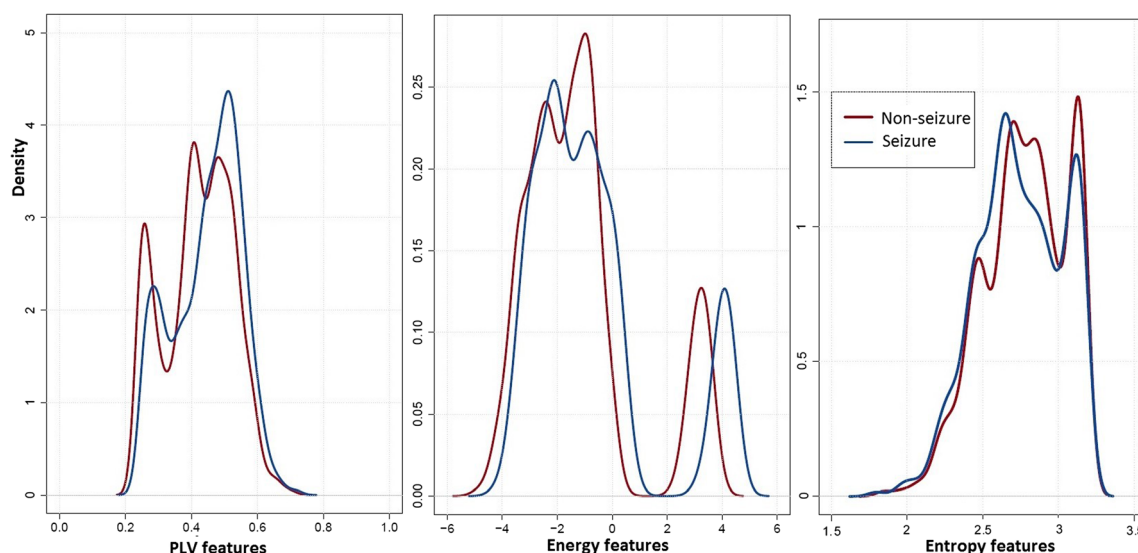
These 10 s and 50% values are arbitrary and should not be taken as some kind of a rule. The additional options were chosen only for the sake of demonstration and evaluation of the classifier from various angles. The first subtask serves as an indication of how well the classifier picks up initial phases of seizures, while the second subtask has the purpose of evaluating our approach and CNN architecture to obtain patient-specific classifier, as opposed to the patient-non-specific classifier. Such kind of early detection or more accurate patient-specific detection, for example, is critical for successful intervention with a responsive neuro-stimulation device.

In each case, the analysis of sensitivity to the input (not just AUC or sensitivity in relation to prediction) is presented as well. Sensitivity to the network inputs is defined as a squared gradient of a loss function, averaged over all testing samples. This sensitivity to the network inputs must not be confused with sensitivity as a measure of the classifier—which one used should be clear from the context where it is used.

### 3.1 Data analysis

In what follows, an overview of seizure detection task using the TUH EEG Corpus dataset, in terms of previously defined features of brain state, is given. This reflects the reasoning of classification to “seizure” vs “non-seizure” patterns.

Even though distributions cannot be a rule of thumb for detection, it is important to look at the distributions of various brain state features during the seizure and non-seizure periods. The distribution of each feature (Fig. 2) raises concerns, in which features are not capable of efficient differentiation between seizure and non-seizure classes. The phase-locking feature (PLV), as a feature related to synchronisation and initially assumed highly useful for prediction, actually, poorly reflects different classes—its distribution only has a slight shift and increase in values for seizure class. The same or even worse is for the entropy, which is also a feature of synchronisation. Just the energy (or power spectrum) feature carries more information to distinguish the two classes, although it is still not very well expressed.



**Fig. 2** Comparison of distributions of all features averaged overall datasets and separated between two classes. This figure compares three features in terms of probability distributions differentiated into two classes: the non-seizure and seizure. The features appear to be not very informative with regard to two classes (non-seizure and seizure). This is contrary to the observations in literature—many papers have shown that

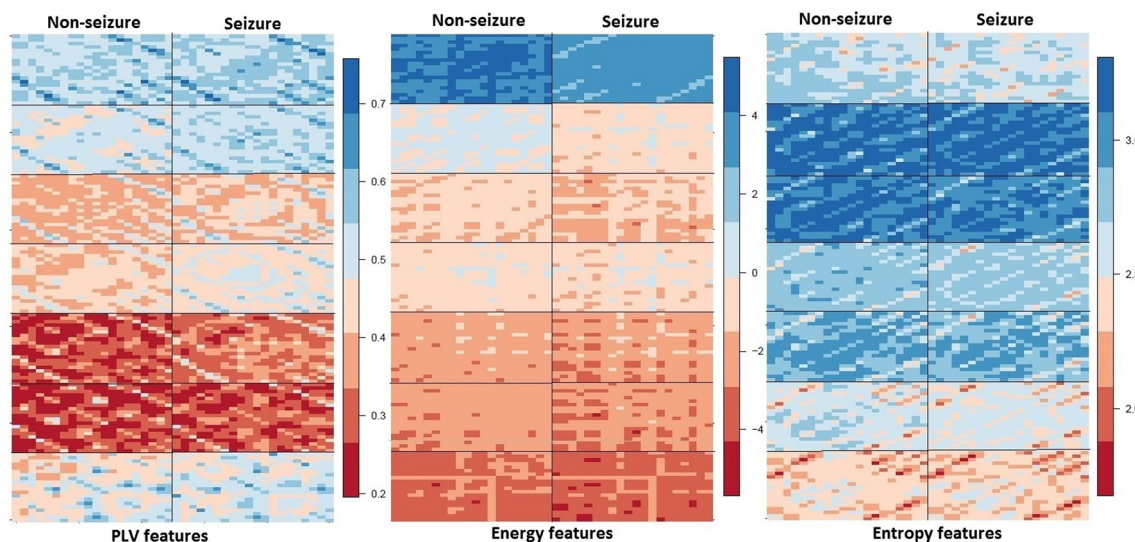
these features can differentiate well between seizure and non-seizure. The most likely reason that in this particular case, these features do not show strong differentiation that the EEG data used to construct them is very noisy, obtained from the highly variable population (age, health condition, gender variability), while other papers used data obtained under strictly controlled laboratory conditions

Separation into different frequency bands (Fig. 3) again shows minor differences in feature values. Thus, it is difficult to hope that the trained classifier will provide high accuracy, given that two classes are not very well distinguished using visual inspection of brain activity patterns. On the other hand, deep learning methods are known to have the property of learning features (deriving new features), which have high predictive power even though the interpretation of those new

features cannot be interpreted like initially defined PLV, energy or entropy features.

### 3.2 Classification and detection results

As the main task is the classification of seizure vs non-seizure epochs, initially, the feature patterns, formed and described previously, were fed to the convolutional neural network



**Fig. 3** Comparison of all feature patterns, formed as described in Sect. 2.3, averaged over an entire training dataset and separated between two classes. Images from left to right: PLV, energy and entropy averaged features. From top to bottom, there are seven zones corresponding to

seven bandpass filtrations of EEG: 1–4 Hz, 4–7 Hz, 7–13 Hz, 13–15 Hz, 15–30 Hz, 30–45 Hz, 45–70 Hz. Each image has two sides—non-seizure and seizure. One can observe that different frequency bands differ for non-seizure and for seizure zones



trained by the Adam Optimizer [55] with learning rate equal to  $1\text{E-}3$ . However, the high imbalance of the classes (6% of the time is in seizure state) caused great difficulty for the algorithm. This is so because distributions of seizure and non-seizure features are very similar (Fig. 2) and a high imbalance of the classes prevented proper learning. For example, there are many episodes with high synchronisation of the brain not only during the time interval with a seizure but also during a seizure-free time interval, which is much longer. Hence, the algorithm learns that a higher phase-locking value is not an indicator of the seizure epoch. This effect is hardly reduced by the usual technique based on the oversampling of the seizure epoch. Practically, it is simply not enough seizure epochs as compared to the non-seizure epochs. However, random under-sampling of the non-seizure epochs and such balancing of classes resulted in better detection accuracy.

Besides, 10-fold cross-validation is used to choose the architecture of the convolutional neural network. Cross-validation is done patient-wise, i.e. patients were divided into ten random groups. In each fold, a check is performed whether the accuracy is affected by the above-mentioned random under-sampling of non-seizure labels or not. It was observed that there is no impact of the random selection of non-seizure labels on the results of classification. In addition, random under-sampling was done to have a 50%/50% distribution of class labels.

Several architectures were considered:

1. CNN1 architecture: one convolutional layer (10 filters of size  $5 \times 5$ ), followed by max-pooling by a factor of 2 followed by a densely connected layer with 1000 neurons;
2. CNN2 architecture: two convolutional layers (10 filters of size  $5 \times 5$ ), followed by max-pooling by a factor of 2 followed by a densely connected layer with 1000 neurons;
3. CNN3 architecture: two convolutional layers (10 filters of size  $5 \times 5$ ) each followed by max-pooling by a factor of 2 (e.g. the first convolutional layer + max-pooling + the second convolutional layer + max-pooling) followed by a densely connected layer with 1000 neurons;
4. CNN4 architecture: two convolutional layers (10 filters of size  $5 \times 5$ ) each followed by max-pooling by a factor of 2 (e.g. the first two convolutional layers + max-pooling + the third and the fourth convolutional layers + max-

pooling) followed by a densely connected layer with 1000 neurons;

The average AUC values of 10-fold cross-validation procedure are reported in Table 3. In addition, results for the support vector machine are presented as well. The cross-validation was performed on the training dataset, by dividing it randomly into 10 equal parts; then, the average of 10 AUC values is calculated.

Further follows a more detailed analysis of testing results with the best CNN architecture (CNN2). Although CNN2 and CNN3 architectures show almost identical results, the simpler architecture (CNN2) was selected for further analysis.

For comparison, at first, each feature can be considered separately. This allows the assessment of the explanatory power of each feature. Convolutional neural network applied to the phase-locking value (PLV) feature patterns produced AUC = 0.67 value with optimal sensitivity and specificity being 0.62 and 0.64, respectively. Average energy features produced accuracy close to the PLV features' case (AUC = 0.66 with optimal sensitivity and specificity being 0.62 and 0.61). The similar results are for entropy features: AUC = 0.65, with sensitivity and specificity being 0.59 and 0.62.

In addition to seizure detection, the authors checked how CNN detects eight different types of seizures. The results for each feature and each type of seizures are presented in Table 3. Combinations of features improve overall results only slightly: for PLV + energy AUC = 0.71, for PLV + entropy AUC = 0.67, for energy + entropy AUC = 0.7, for all features, i.e. PLV + energy + entropy AUC = 0.74. The sensitivity and specificity of the classification algorithm, trained on all features, are 0.68 and 0.67, respectively.

It is particularly interesting to compare classifiers for different types of seizures. For each type of seizures, the evaluation was done by considering patterns of other types of seizures and non-seizure patterns as belonging to one class (Table 4).

Detection accuracy for simple partial seizures (the most frequent type of seizures) is homogeneous between different features and combinations of them. A more significant improvement can be achieved by combining all features although the improvement is counteracted by significantly increased dimensionality and the computational burden.

**Table 3** Ten-fold cross-validation accuracies expressed as AUC values for considered CNN classifiers and SVM

Classifier	PLV	Energy	Entropy	PLV + energy	PLV + entropy	Energy + entropy	All
CNN1	0.65	0.64	0.64	0.65	0.62	0.67	0.70
CNN2	0.67	0.66	0.65	0.71	0.67	0.70	0.74
CNN3	0.66	0.68	0.63	0.69	0.70	0.71	0.73
CNN4	0.68	0.65	0.68	0.70	0.68	0.69	0.72
SVM	0.62	0.51	0.59	0.63	0.63	0.60	0.64

**Table 4** Detection results for each seizure type annotated in the evaluation dataset. Three measures (top to bottom)—AUC, sensitivity and specificity values were reported

Features	Focal non-specific seizure (0.6%)	General non-specific seizure (16%)	Simple partial seizure (45.75%)	Complex partial seizure (26.18%)	Absence seizure (0.68%)	Tonic seizure (6.82%)	Clonic seizure (1.5%)	Tonic-clonic seizure (3.93%)
PLV (AUC = 0.67)	0.78	0.74	0.62	0.67	0.68	0.75	0.64	0.39
	0.72	0.68	0.6	0.62	0.67	0.71	0.57	0.35
	0.74	0.69	0.58	0.62	0.62	0.66	0.67	0.50
Energy (AUC = 0.66)	0.88	0.71	0.62	0.6	0.64	0.68	0.54	0.84
	0.61	0.65	0.58	0.61	0.78	0.69	0.57	0.76
	0.80	0.69	0.61	0.5	0.54	0.58	0.51	0.77
Entropy (AUC = 0.65)	0.86	0.72	0.65	0.63	0.63	0.6	0.46	0.46
	0.74	0.68	0.6	0.58	0.59	0.64	0.39	0.58
	0.85	0.65	0.61	0.59	0.66	0.5	0.56	0.38
PLV + energy (AUC = 0.71)	0.98	0.63	0.65	0.8	0.65	0.96	0.6	0.70
	0.95	0.65	0.64	0.74	0.76	0.89	0.52	0.56
	0.94	0.50	0.61	0.56	0.56	0.91	0.67	0.77
PLV + entropy (AUC = 0.67)	0.87	0.78	0.63	0.67	0.72	0.80	0.40	0.54
	0.76	0.70	0.57	0.63	0.70	0.73	0.30	0.65
	0.82	0.71	0.62	0.60	0.62	0.70	0.60	0.42
Energy + entropy (AUC = 0.70)	0.86	0.72	0.65	0.63	0.63	0.6	0.46	0.46
	0.74	0.68	0.6	0.58	0.59	0.64	0.39	0.58
	0.85	0.65	0.61	0.59	0.66	0.5	0.56	0.38
PLV + energy + entropy (AUC = 0.74)	0.97	0.63	0.73	0.78	0.62	0.95	0.55	0.70
	0.90	0.62	0.69	0.72	0.80	0.89	0.50	0.65
	0.92	0.58	0.66	0.72	0.50	0.90	0.60	0.60

Another frequent type is a complex partial seizure, and here, no improvement is made with learning from all features. In this case, the feature of synchronisation (i.e. PLV) and energy over different bands of frequencies is sufficient to achieve  $AUC = 0.8$ .

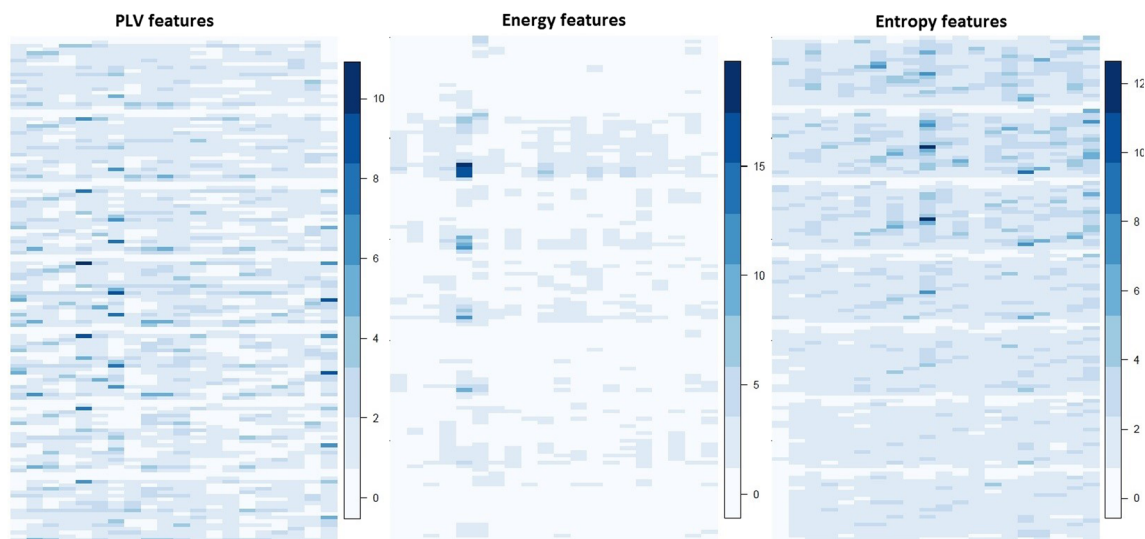
Also, the high accuracy of results for non-specific focal seizures probably is due to the ability of the classifier to distinguish its patterns accurately. Significantly poorer results for other rare types (absence and clonic seizures) allow rejecting the hypothesis that high results for non-specific focal seizures are solely due to the small sample effect; surely, small sample plays its role, but probably not definitive—the deeper investigation is out of the scope of this paper. Further analysis of the table reveals that energy is a highly informative feature for this type of seizures, while entropy provides almost no additional information.

The sensitivity of loss function (here defined as the square of the loss function gradient taken w.r.t. input features) to various features provides an additional level of information. Results show (Fig. 4) that phase-locking value carries most information about seizure/non-seizure classes in the middle and higher frequencies (7–70 Hz), while entropy (also a feature of phase synchronisation) is more sensitive to the lower frequencies in the range of 1–30 Hz. On the other hand, energy considering very low frequencies (1–4 Hz) is almost irrelevant with most of the information contained in the middle frequency ranges (4–30 Hz). Also, each feature has a small group of channels (specific zones) mostly involved in the discrimination between seizure and non-seizure classes—groups of channels are different for different features. This

provides implications to the dimensionality reduction—one could use a smaller number of frequency ranges as well as channels. This would be of great importance for the seizure detection and prediction tool, as such reduction could potentially reduce the calculation time and thus decrease the prediction timing.

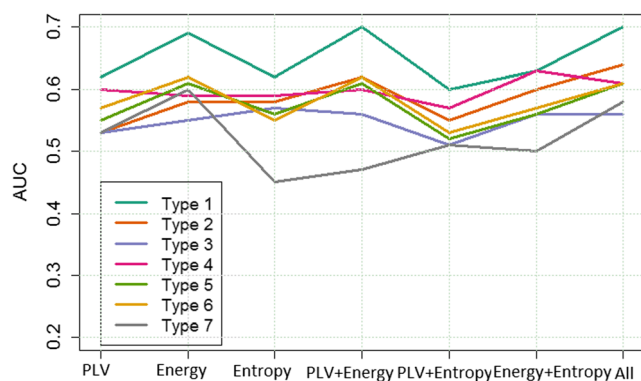
Further, the authors tested how difficult it is to detect the start of the seizure, i.e. the seizure during the first 10 s. Detection results were even more pessimistic (Fig. 5). The highest AUCs were achieved combining energy and entropy features, besides PLV features gave no further boost in the detection accuracy. Synchronisation-related PLV feature, alone and in combination with entropy, poorly detected the start of the seizure. This might be due to the desynchronisation during the initial phase of the seizure—this kind of feature is of value for the detection of later phases of seizure. However, one of the most important issues in epilepsy is the prediction of seizure onset, while the later phases are less relevant. Hence, the application of energy features should be a priority in seizure detection. Detection of seizure during the first 10 s just before the seizure onset (as judged by a trained clinician) failed completely: AUC values were even lower and concentrated around 0.5.

From the above results, one could make a general conclusion: seizures are not easily predictable with satisfactory accuracy by current state-of-art classification methods, as used on clinical EEG recordings. That is probably an answer to the observation, that seizure detection alarm tools used in hospitals have a very high false-positive rate, as opposed to the high accuracy of classifiers, reported in



**Fig. 4** Loss function sensitivity to different features. The structure of the images is the same as in Fig. 2, but here, average gradients of loss function are taken w.r.t. each entry of the brain activity pattern. The higher the value, the higher the sensitivity. High sensitivity indicates that the classifier strongly reacts to the change in that particular feature entry and therefore those entries are responsible for the detection of

seizures. Low sensitivity indicates that no matter the change of that particular feature entry, the classifier will not react strongly to that change. This indicates that those feature entries carry only a small amount of information. Given such information, it might be useful to only consider entries with the highest sensitivity and in this way to reduce the dimensions boosting the speed of calculations



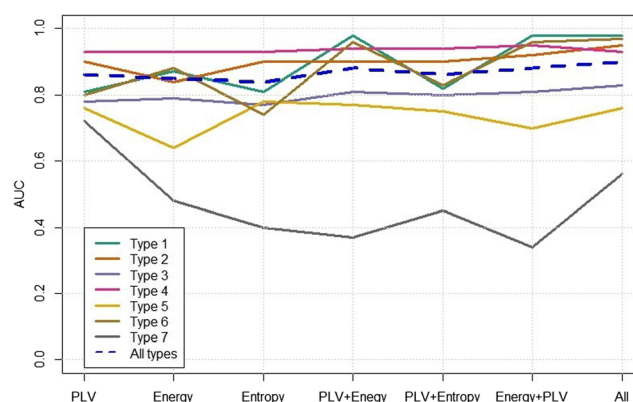
**Fig. 5** AUC results for each type of seizure (see Table 1 for seizure type names). From this figure, one can observe that accuracies of detection are very similar for all except type 1 and type 7 seizures. We discarded the 8th type of seizures (tonic-clonic) from this figure because for every combination of features, the detection accuracy (in terms of AUC) was less than 0.5 indicating that none of the features is able to differentiate this particular type of seizures

many research papers, but trained and tested on EEG had been obtained under non-clinical conditions. Then, a valid question would be: is it at all possible to have such an algorithm, which detects various types of seizures for patients of different ages and suffering different illnesses? It is challenging and at least hardly possible with the current widespread approach when many different features are constructed from EEG simply dividing recordings into short time windows and without the use of knowledge about the dynamics of seizures and relations between different phases and features.

On the other hand, the patient-specific detection system could be based on a classifier related to the one investigated above. In this case, 50% of each patient's EEG records (i.e. the first 50% of timing interval for each patient) from the test dataset is added to the training set. Then, the convolutional neural network is again trained to give much more optimistic results (Fig. 6). In this case, AUC is above 0.85 for various features and combinations (even achieving AUC = 0.9 using all types and all features). When each type of seizures is considered separately, a large deviation is observed, especially for clonic seizures (i.e. type 7), while accuracy for detection of other types of seizures had similarly high AUC values.

## 4 Summary and conclusions

In this paper, the clinical data from Temple University Hospital EEG Corpus database is studied. The main task of seizure detection and classification is solved by the application of convolutional neural network—state-of-the-art method, which has demonstrated to be of superior accuracy among other previously applied methods. Images of brain activity



**Fig. 6** Results of patient-specific classifier: the first 50% of EEG records in the test dataset for each patient were used to train the classifier (in addition to the previously used training dataset). One can observe that even though for the first six types of seizures, detection improved, for type 7, the effect was the reverse. We discarded the 8th type of seizures (tonic-clonic) from this figure because for every combination of features, the detection accuracy (in terms of AUC) was less than 0.5 indicating that none of the features is able to differentiate this particular type of seizures

were constructed using different features, i.e. phase-locking value, entropy, energy. These images then enabled the application of various deep learning methods and specifically focused on CNN.

The results are rather pessimistic, and we conclude that for clinical data, one cannot construct a universal seizure detector in a way how patient-specific data had been used for the past decades (i.e. extracting various features, e.g. energy and synchronisation, and then simply glueing them together). One could even go further and conclude that an accurate tool for clinical/hospital usage will not be constructed unless the shift of paradigm occurs and more attention will be paid to this challenge.

In our case for such a universal detector, the sensitivity and specificity reached 0.68 and 0.67, accordingly, and area under the curve was around 0.74 (with slight variations depending on the type of seizures and features used). When different types of seizures are considered separately, a very large variation of results was identified: from very poor AUC = 0.39 for tonic-clonic seizures (with PLV features) up to very high AUC = 0.98 for non-specific focal seizures (with PLV and energy features).

In future investigations, dimensionality reduction based on cost function sensitivity to inputs could increase the AUC values. However, the increase might be only marginal. The above-reported results for the universal seizure detector imply a certain lack of detectability of seizures if it is based on the energy and synchronisation features alone.

On the other hand, the patient-specific classifier (i.e. a classifier trained on each patient data) could be useful in personal or even mobile applications. For personal data treatment cases, AUC is around 0.9, which could be even slightly increased by dimensionality reduction.



**Funding information** Tomas Iesmantas was supported by the postdoctoral fellowship grant (2016–2018) from Kaunas University of Technology and Department of Mathematics and Natural Sciences.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants performed by any of the authors.

## References

- Kwan P, Brodie MJ (2000) Early identification of refractory epilepsy. *N Engl J Med* 342:314–319
- Brodie MJ (2017) Outcomes in newly diagnosed epilepsy in adolescents and adults: insights across a generation in Scotland. *Seizure*. 44:206–210
- Patsalos PN, Perucca E (2003) Clinically important drug interactions in epilepsy: general features and interactions between antiepileptic drugs. *Lancet Neurol* 2:347–356
- Kanner AM (2006) Depression and epilepsy: a new perspective on two closely related disorders. *Epilepsy Curr* 6:141–146
- Lee M, Hively N, Clap PE, et al Apparatus and method for epileptic seizure detection using non-linear techniques, US patent 5743860 A, April 28, 1996
- Iasemidis LD, Sackellaes JC Seizure warning and prediction, US patent 6304775 B1, October 16, 2001
- Osorio I, Lyubushin A, Sornette D Seizure detection methods, apparatus, and systems using a wavelet transform maximum modulus algorithm, US patent 9549677 B2, July 20, 2012
- Denison TJ, Santa WA Seizure prediction, US patent 8594779 B2, April 30, 2007
- Mormann F, Kreuz T, Rieke C, Andrzejak RG, Kraskov A, David P, Elger CE, Lehnertz K (2005) On the predictability of epileptic seizures. *Clin Neurophysiol* 116:569–597
- Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G et al (2013) Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *Lancet Neurol* 12: 563–571
- D'Alessandro M, Esteller R, Vachtsevanos G, Hinson A, Echauz A, Litt B (2003) Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *IEEE Trans Biomed Eng* 50(5):603–615
- Park Y, Luo L, Parhi KK, Netoff T (2011) Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia* 52:1761–1770
- Mirowski P, Madhavan D, Lecun Y, Kuzniecky R (2009 Nov) Classification of patterns of EEG synchronization for seizure prediction. *Clin Neurophysiol* 120(11):1927–1940
- Nasehi S, Pourghassem H (2013) A novel fast epileptic seizure onset detection algorithm using general tensor discriminant analysis. *J Clin Neurophysiol* 30(4):362–370
- Shoeb A, Kharbouch J, Soegaard J et al (2011) A machine learning algorithm for detecting seizure termination in scalp EEG. *Epilepsy Behav* 22(Suppl. 1):S36–S43
- Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H (2018) Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput Biol Med* 100:270–278
- Avcu MT, Zhang Z, Chan DWS Seizure detection using least EEG channels by deep convolutional neural network. *ICASSP* 2019–2019 IEEE International Conference on Acoustics, Speech and signal processing (ICASSP), Brighton, 2019, pp. 1120–1124
- Choi G et al. A novel multi-scale 3D CNN with deep neural network for epileptic seizure detection, 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, 2019, pp. 1–2
- Wei X, Zhou L, Chen Z, Zhang L, Zhou Y (2018) Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. *BMC Med Inform Decision Making* 18:111
- Khan H, Marcuse L, Fields M, Swann K, Yener B (2018) Focal onset seizure prediction using convolutional networks. *IEEE Trans Biomed Eng* 65(9):2109–2118
- Truong ND, Nguyen AD, Kuhlmann L, Bonyadi MR, Yang J, Ippolito S, Kavehei O (2018) Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Netw* 105:104–111
- Jana R, Bhattacharyya S, Das S (2020) Patient-specific seizure prediction using the convolutional neural networks. In: Bhattacharyya S, Mitra S, Dutta P (eds) *Intelligence enabled research. Advances in intelligent systems and computing*, vol 1109. Springer, Singapore
- Obeid I, Picone J (2017) Machine learning approaches to automatic interpretation of EEGs. *Biomedical Signal Processing in Big Data*. CRC Press
- Schulc E, Unterberger I, Saboor S, Hilbe J, Ertl M, Ammenwerth E, Trinka E, Them C (2011) Measurement and quantification of generalized tonic-clonic seizures in epilepsy patients by means of accelerometry—an explorative study. *Epilepsy Res* 95(1–2):173–183
- Van de Vel A, Verhaert K, Ceulemans B (2014) Critical evaluation of four different seizure detection systems tested on one patient with focal and generalized tonic and clonic seizures. *Epilepsy Behav* 37:91–94
- De Bruijne GR, Sommen PCW, Aarts RM (2009) Detection of epileptic seizures through audio classification. In: 4th European Conference of the International Federation for Medical and Biological Engineering, pp 1450–1454
- Regalia G, Onorati F, Lai M, Caborni C, Picard RW (2019) Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands. *Epilepsy Res* 153:79–82
- Bidwell J, Khuwatsamrit T, Askew B, Ehrenberg JA, Helmers S (2015) Seizure reporting technologies for epilepsy treatment: a review of clinical information needs and supporting technologies. *Seizure* 32:109–117
- Scheuer ML, Bagic A, Wilson SB (2017) Spike detection: inter-reader agreement and a statistical Turing test on a large data set. *Clin Neurophysiol* 128(1):243–250
- Freestone DR, Karoly PJ, Cook MJ (2017) A forward-looking review of seizure prediction. *Curr Opin Neurol*, vol. In Press
- Boex C, Seeck M, Vulliemoz S et al (2011) Chronic deep brain stimulation in mesial temporal lobe epilepsy. *Seizure* 20:485–490
- Choi SI, Obeid I, Jacobson M, Picone J The Temple University Hospital EEG Corpus. The Neural Engineering Data Consortium, College of Eng., Temple Univ., 2013. [Online]. Available: [http://www.isip.piconepress.com/projects/tuh\\_eeg](http://www.isip.piconepress.com/projects/tuh_eeg). [Accessed: 21-Feb-2018]
- Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. *Hum Brain Mapp* 8: 194–208
- Mormann F, Lehnertz K, David P, Elger CE (2000) Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D* 144:358–369
- Truccolo W, Ahmed OJ, Harrison MT, Eskandar EN, Cosgrove GR, Madsen JR, Blum AS, Potter NS, Hochberg LR, Cash SS (2014) Neuronal ensemble synchrony during human focal seizures. *J Neurosci* 34:9927–9944

36. Jiruska P, de Curtis M, Jefferys JG, Schevon CA, Schiff SJ, Schindler K (2013) Synchronization and desynchronization in epilepsy: controversies and hypotheses. *J Physiol* 591(Pt 4):787–797
37. Netoff TI, Schiff SJ (2002) Decreased neuronal synchronization during experimental seizures. *J Neurosci* 22:7297–7307
38. Cymerblit-Sabba A, Schiller Y (2012) Development of hypersynchrony in the cortical network during chemoconvulsant-induced epileptic seizures in vivo. *J Neurophysiol* 107:1718–1730
39. Topolnik L, Steriade M, Timofeev I (2003) Partial cortical deafferentation promotes development of paroxysmal activity. *Cereb Cortex* 13:883–893
40. Timofeev I, Steriade M (2004) Neocortical seizures: initiation, development and cessation. *Neuroscience* 123:299–336
41. Varela F, Lachaux J-P, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci* 2:229–239
42. Roskies A (1999) The binding problem: special issue. *Neuron* 24:7–125
43. Brechet R, Lecasble R (1965) Reactivity of mu-rhythm to tickler. *Electroencephalogr Clin Neurophysiol* 18:721–722
44. Koshino Y, Niedermeyer E (1975) Enhancement of rolandic mu-rhythm by pattern vision. *Electroencephalogr Clin Neurophysiol* 38:535–538
45. Pfurtscheller G, Klimesch W (1992) Functional topography during a visuoverbal judgement task studied with event-related desynchronization mapping. *J Clin Neurophysiol* 9:120–131
46. Le van Quyen M, Foucher J, Lachaux JP, Rodriguez E, Lutz A, Martinerie J, Varela F (2001) Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *J Neurosci Methods* 111:83–98
47. Tass P, Rosenblum MG, Weule J, Kurths J, Pikovsky A, Volkmann J, Schnitzler A, Freund HJ (1998) Detection of n:m phase locking from noisy data: application to magnetoencephalography. *Phys Rev Lett* 81:3291–3294
48. Fisher RS, Webber WR, Lesser RP, Arroyo S, Uematsu S (1992) Highfrequency EEG activity at the start of seizures. *J Clin Neurophysiol* 9:441–448
49. [Online] <https://www.kaggle.com/c/seizure-prediction> [Accessed on 2018-03-02]
50. Tatum WO (2014) Ellen R. grass lecture: extraordinary EEG. *Neurodiagnostic J* 54(1):3–21
51. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1(4):541–551
52. LeCun Y, Bottou L, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
53. Glorot X, Bordes A and Bengio Y Deep sparse rectifier neural networks (PDF). AISTATS, 2011
54. Abadi, M., Agarwal, A., Barham, P., et al. TensorFlow: large-scale machine learning on heterogeneous systems, 2015. Software is available from [tensorflow.org](https://www.tensorflow.org). Accessed 2019-08-01
55. Kingma D, Jimmy BA A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014
56. Mirowski P, Madhavan D, LeCun Y, Kuzniecky R (Nov 2009) Classification of patterns of EEG synchronization for seizure prediction. *Electroencephalogr Clin Neurophysiol* 120(11):1927–1940
57. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
58. Japkowicz N, Shah M (2011) Evaluating learning algorithms: a classification perspective. Cambridge University Press, Cambridge
59. Snyder DE, Echauz J, Grimes DB, Litt B (2008) The statistics of a practical seizure warning system. *J Neural Eng* 5:392–401
60. Hajian-Tilaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 4:627–635

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

**Tomas lešmantas** completed his Ph. D. in technological sciences (Bayesian assessment of reliability and vulnerability of complex systems, 2016, Lithuanian energy institute). Currently, he is working at the Kaunas University of Technology (Department of Mathematics and Natural Sciences). Recent interests are in deep learning and applications in medicine. More specifically, for biosignals as well as medical images—epilepsy, cardio signal classification, MRI, X-ray image classification, and segmentation, etc. He is currently participating in several research projects and collaborating with other academic institutions, industries, and governmental agencies.

**Robertas Alzbutas** completed his Ph. D. in technological sciences (Risk minimisation and reliability control under uncertainty, 2004, Lithuanian energy institute). Currently, he is working at the Kaunas University of Technology (Department of Mathematics and Natural Sciences). Recent interests are in connected health and precision medicine areas. His research includes the application of mathematical models in wearables for health state monitoring. He is leading the research group “Data Analytics and Modeling” that activities covers: big data analytics; numerical methods applications, analysis for risk and uncertainty, economics and business, medicine and sports, earth and environment.