Algorithm for automatic classification of benign childhood epilepsy and structural focal epilepsy patient EEGs

Andrius Vytautas Misiukas Misiūnas¹, Tadas Meškauskas¹, Rūta Samaitienė²

¹ Vilnius University Faculty of Mathematics and Informatics 47 Didlauko Street, LT-08303 Vilnius
² The Childrens Hospital of Vilnius University Hospital Santariskiu Klinikos 7 Santariskiu Street, LT-08406, Vilnius

Abstract

A novel algorithm for automatic classification of: 1) benign childhood epilepsy (Group I) and 2) structural focal epilepsy (Group II) patient electroencephalograms (EEGs) is presented. The algorithm consists of three stages: EEG spike detection, determining EEG spike parameters and classifying EEGs by epilepsy type based on the spike parameters. The classification algorithm is based on artificial neural network, the algorithm has been trained and tested on large data sample provided by The Childrens Hospital of Vilnius University Hospital Santariskiu Klinikos. The algorithm has been tested only with EEGs from both patient groups that have indistinguishable (for neurologists) epileptiform discharges. The algorithm has been trained with data of one subset of patients and tested with another non overlapping set containing different patients. This methodology gives approximately 72% reliability of classification with EEGs containing 90 or more epileptiform discharges.

Keywords: EEG, Electroencephalogram, Epilepsy, Epileptiform discharge, Machine learning

1. Introduction

Few studies have attempted to develop automatic algorithms for detecting and classifying interictal epileptiform discharge (ED, or spike) according to

Preprint submitted to Journal of Biomedical Informatics       October 15, 2017
the epilepsy type. We have focused on 2 different types of epilepsy, which could have the similar EEG pattern. Manually distinguishing this type of EDs is time-consuming, and sometimes impossible (without clinical record of the illness), as EDs have similar morphology. Automatic detection algorithms introduced in the literature mostly focus on detection of spikes itself [1, 2], while other algorithms are used in order to classify EEGs of patients diagnosed with epilepsy or similar disease and healthy people, like [3]. Epileptiform sharps are detected using a variety of algorithms such as mathematical morphology used in this work [4, 5, 6], while others use wavelets and other methods summarised in [1, 2].

In this work we are going to deal with two different groups of patients.

Group I will be defined as benign childhood epilepsy with centrotemporal spikes (in literature reffered as BCECT), for example rolandic epilepsy exhibiting the EEG hallmark – the benign epileptiform discharge (BED) frequently appearing as rolandic discharge. These sharp waves are called benign because they are found predominantly in the benign partial epilepsies of childhood [7]. It is sharp wave or spike detected over the centrotemporal brain regions (T3, T4, C3, C4 regions), but other regions as central, centroparietal or centrofrontal, can also be involved [8]. A number of less well-defined syndromes of idiopatic focal epilepsies (IFE) have been proposed, including benign childhood epilepsy with parietal spikes, benign childhood seizures with frontal or midline spikes [9].

Group II will be defined as structural focal epilepsy patients with cerebral palsy, dysplastic brain lesion, gliosis etc. following ILAE criteria [10].

The problem is that benign epileptiform discharges seem not to be pathognomonic of for rolandic epilepsy. They may represent both a functional focus [11] and an expression of a focus secondary to an organic brain damage. Both rolandic epilepsy and symptomatic focal epilepsy with a lesion in rolandic area may present with BEDs [12, 7].

It is not clear if there are specific EEG parameters that help to distinguish structural and idiopathic spikes arising in rolandic area. The waveform of the BEDs is distinctive. The most prominent element is a negative sharp wave/spike duration of more than 80 ms. This sharp wave is preceded by a short duration prepositivity and followed by a prominent positive wave. The amplitude is nearly 50% of the preceding negative sharp wave, and almost always of lower amplitude than the preceding negative sharp wave. There is tendency of sharp waves to have a constant, stereotyped waveform and field of distribution for any given recording sample. Usually there is no diffuse or
localized slowing of the EEG [13, 14, 15].

The initial hypothesis for this research has been made that some statistical properties of BED series (of Group I and Group II) could be detected by machine learning type methods. This assumption has been confirmed by the results of this work. In the previous study we have defined some quantifiable parameters of BEDs [16]. In this paper the tests of artificial neural network (ANN) based classifier (see Fig. 3) and the statistical analysis (see Fig. 5 and Fig. 6) showed that the series of BED upslopes and downslopes (see Fig. 2) have the highest significance. Tests involving other parameters (defined in [16]) were also carried out, but such approach did not improve the accuracy of the classifier (for details see section 3.2), therefore these additional parameters were discarded from the proposed methodology.

Thus, the main aim of our study has been to explore machine learning type methods for automatic classification between two different epilepsy groups (both presenting with visually indistinguishable series of BEDs). By achieving this goal we have proposed the algorithm, based on three consecutive steps: 1) EEG spike detection, 2) EEG spike measurement, 3) ANN based classifier, as described in more detail in section 3. In Section 4 we show that reliability of the automatic ANN based classifier (trained and tested on non overlapping inputs, obtained from different patients’ EEGs) is about 72%.

The presented algorithm is implemented with Python 2.7.10 programming language, employing NumPy [17], SciPy [18], OpenCV [19], Mpi4Py [20] and EdfTools [21] libraries. The algorithm was tested and deployed in high performance parallel computing environments, including supercomputer at Digital Science and Computing Center, Faculty of Mathematics and Informatics, Vilnius University.

2. Data

A retrospective search has been done of the EEG database (The Childrens Hospital of Vilnius University Hospital Santariskiu Klinikos) over an 2010-2016 period to identify children having benign epileptiform discharges. It has been defined as any focal epileptiform discharge that involved the central central regions and mid-central or mid-parietal regions (T3, T4, C3, C4, Cz, Pz). Focal discharges that had maximal negativity at the mid-temporal regions (T5 or T6) or the biparietal regions (P4 or P3) were also included
if the discharges had constant, stereotyped waveform, spread to central regions, no diffuse or localized slowing. Inclusion criteria were: (a) hard to distinguish (visually identical) benign epileptiform discharges in idiopathic rolandic epilepsy and rolandic like discharges in structural focal epilepsy, (b) artifact-free EEG recording of $2 - 13$ min, (c) $\geq 50$ spikes in the raw EEG, none of which should be preceded or followed by artifacts.

All EEG recordings were performed with electrodes placed according to the international 10-20 system.

2.1. EEG selection

In this research we have processed 94 EEGs of 86 different patients, divided into two groups:

- **Group I:** benign childhood epilepsy with centrotemporal spikes, in our case rolandic epilepsy (62 EEGs or about 66% of all EEGs);

- **Group II:** structural focal epilepsy patients with cerebral palsy, dysplastic brain lesion, gliosis etc. (32 EEGs, or about 34% of all EEGs).

It should be noted, that we have omitted from further analysis trivial cases: Group II EEGs, whose BEDs exhibit visual dissimilarities (easily detectable by neurologists) from those in Group I. By including such data into presented methodology one would only decrease error of automatic classification. The reason for such decision was to demonstrate that discussed methodology is capable to classify even challenging data.

2.2. ANN training and testing strategy

In this study we applied strict policy of assigning the same patient either to training or testing pool. If multiple EEGs (recorded at different times) of the same patient were available, then all of them went into the same (training or testing) set. This has been done in order to prevent biasing of results by ANN over-fitting [22]. An experiment to determine significance of over-fitting effect has also been carried out, resulting in approximately 95-99% pseudo-reliability of classifier (compared to about 72% reliability of non over-fitted methodology, see section 4).

The training and testing (of ANN based classifier) pools have been defined in the following way:

- 62 EEGs from Group I – 21 for training (about 34%) and 41 for testing (about 66%);
32 EEGs from Group II – 11 for training (about 34%) and 21 for testing (about 66%).

Training dataset has been manually cleaned from artefacts (caused by patient movements, rapid eye movement, etc.) by neurologist to avoid ANN training errors.

Note that the same training and testing strategy has been applied for both ANN and supported vector machine (SVM, see section 3.3.2) based classifiers.

3. Algorithm for EEG processing and automatic classification

In this section we discuss the three steps of the EEG processing and classification algorithm in more detail. These steps are:

1. EEG spikes (BEDs) detection, for details see section 3.1 and references [4, 5, 6].
2. Determining EEG spikes parameters, for details see section 3.2 and our previous work [16].
3. ANN based classifier, for details see section 3.3.1.

3.1. EEG spike detection

In this section we provide details of the first step of the algorithm. In this step mathematical morphological filter based algorithm is used in order to find occurrences of spikes in EEG. The EEG spike detection algorithm is based on mathematical morphological filters, first described in [4], later in [5] with further improvements [6, 16]. The basic idea behind this algorithm has been to filter out known normal brain activity, for example brain rhythms, measurement noise generated by EEG measurement machine and others.

The EEG is pre-processed using moving window (for example, of 4 seconds duration), normal brain activity is re-evaluated in each window. This allows the algorithm to adapt to changing patient’s brain activity and recognise it correctly. In each window the dominating brain rhythm is determined as well as other its characteristics, for example exact period and strength. Then according to mentioned characteristics, the structuring element (for morphological filter) is generated. Afterwards the signal is processed by applying morphological filter and generated structuring element (see Fig. 1).

At this stage the EEG channel with highest number of spikes detected is chosen and passed to the second step of the algorithm. The channel with
Figure 1: Demonstration of principles behind the morphological filter (for details see [6]). The EEG has two spikes at 6.2 and 7.0 seconds. A) Original EEG at T3 channel, B) the same EEG signal with morphological filter applied.

maximum number of spikes is coherent to the focus of epilepsy. It is worth to point out that the spikes detected in this channel have mostly distinguishable derivative parameters, discussed in section 3.2.

The spike (BED) detection procedure is described in greater detail in [4, 5, 6].

3.2. Measuring of EEG spike derivative parameters

We briefly describe step 2 (for details see our previous work [16]) of the algorithm.

In order to distinguish between Group I and Group II patients some quantifiable parameters (obtained by measuring spike geometrical characteristics) will be employed. Experiments with measurements of many features (upslope, downslope, width at half maximum, baseline level) of EEG spike were done, but upslope and downslope of EEG spike have prevailed as the most significant parameters in ANN based classifier (see section 3.3.1).

Upslope and downslope are measured cutting of $2\sigma$ from both lower and upper parts of the sharp wave of EEG spike and line is fitted to the parts remaining (as shown in Fig. 2). Here $\sigma$ is a standard deviation of signal fluctuations at baseline neighbourhood. Note that the signs of both upslope and downslope values are dependent purely on EEG measurement apparatus.
configuration and should have different signs [23] (otherwise the spike is considered falsely detected and rejected from further analysis). Therefore only absolute values of upslope and downslope are dealt with further.

More detailed description of spike parameters measurement methods is available at our previous paper [16].

3.2.1. Post-processing of EEG spikes detections

The initial spike detection algorithm still had some problems – for example it might detect patient’s rapid eye movement as EEG spike. This is remedied by post-processing as described in [16]. The basic idea behind the post-processing is to eliminate false detections of spikes based on their estimated characteristics (see section 3.2). These characteristics are validated against their neurological properties:

- Upslope and downslope values should have different signs;
- Length of sharp wave of spike should be between 20 ms and 80 ms [23];
- Total length of spike should not exceed 200 ms [23].

Only the spikes meeting all three above conditions are considered detected successfully and the values of upslopes as well as downslopes (computed from
these spikes) are included in further analysis. Such post-processing procedure improves specificity of EEG spike detection algorithm.

3.3. Machine learning type methods for automatic classification of EEGs

In this subsection we describe third and final step of the algorithm.

3.3.1. ANN based classifier

A feed-forward perceptron based neural network without backpropagation [22] has been applied in order to classify EEGs between Group I and Group II using parameters of their BEDs (spikes). An artificial neural network, defined by an input layer, a single hidden layer, and an output layer has been employed for the classification. We have set the number of input neurons to be equal to the length of precomputed series of upslopes and downslopes (see section 3.2). Also, the classifier under consideration has been defined by 20 hidden neurons, 1 output neuron, scaled conjugate gradient as a training algorithm (for minimization of some cost function), and cross entropy

\[
C = -\frac{1}{n} \sum_x \left[ y \ln a + (1 - y) \ln (1 - a) \right],
\]

as the cost function [22], here \(n\) is length of the list, containing training data, the sum is over all training inputs \(x\), \(y\) is the corresponding desired output and \(a\) is the activation function. We have applied the so-called softmax function [22]

\[
a^L_j = \frac{e^{z^L_j}}{\sum_k e^{z^L_k}},
\]

as the activation function for \(j\)-th neuron in \(L\)-th layer (\(L = 1\) in the input layer, \(L = 2\) in the hidden layer and \(L = 3\) in the output layer). Here

\[
z^L_j = \sum_k w^L_{jk} a^{L-1}_k + b^L_j,
\]

\(w^L_{jk}\) is the weight from the \(k\)-th neuron in the previous \((L - 1)\)-th layer to the \(j\)-th neuron in the current \((L)\)-th layer, and \(b^L_j\) is the bias of the \(j\)-th neuron in the \((L)\)-th layer, and \(a^L_j\) is the activation function of the \(j\)-th neuron in the \((L)\)-th layer, and \(a^{L-1}_k\) is the activation function of the \(k\)-th neuron in the previous \((L - 1)\)-th layer.
It should be noted that we have experimented with some other activation functions (for example sigmoid), however the choice (2) has resulted in training within least amount of epochs, without any substantial impact on accuracy of classification.

This simple neural network is well suited for binary classification problems (note that we apply binary classification as well), as some related studies suggest, see for example [24]. Increasing ANN complexity would result in higher computational time, not necessarily achieving better accuracy, compared to 72% (see section 4.1) obtained by the classifier, dealt with in this work.

3.3.2. SVM based classifier

Despite the fact that for binary classification problems supported vector machine (SVM) based methods [25] are commonly used, we managed to achieve the reliability (of such classification techniques) to be of order 50-52% only, that is within statistical margin of error of random guess. The comparison of ANN and SVM classifiers for different (however, also binary) problem is provided in [24].

For readers with deeper interest in machine learning methods we discuss possible reasons for failure of SVM based classifier in section 4.2.

4. Results and discussion

We have proposed a novel three stage algorithm (defined in section 3) for EEG classification related to EEG epileptoform discharges (EDs) and their parameters between EEGs of patients diagnosed with benign childhood epilepsy with centrotemporal spikes, for example rolandic epilepsy (Group I) and structural focal epilepsy in patients with cerebral palsy, dysplastic brain lesion, gliosis etc. (Group II).

Our results suggest that the algorithm, based on: 1) spike detection, 2) derivative characteristics (upslopes, downslopes of spike) estimation, 3) machine learning (ANN), can be employed for classifying EEGs according the epilepsy type (we discuss the reliability of this methodology further). The mathematical morphology based EEG spike detection algorithm has already been explored in previous studies [4, 5, 6], while employment of EEG spike parameters is a novel approach for distinguishing between patients diagnosed with different epilepsy types (in this study Group I and Group II).
4.1. Accuracy of ANN based automatic classification

In order to evaluate the accuracy of the ANN based classifier (see section 3.3.1) between Group I and Group II patients, we have experimented with the following strategies (defining lists, containing parameters of spikes, see section 3.2, employed in training and testing):

- **Strategy A**: lists have been defined by upslope and downslope pairs of values;
- **Strategy B**: lists have been defined by the upslopes values only;
- **Strategy C**: lists have been defined by the downslopes values only.

In this analysis all the lists have been defined by partitions (of available spike parameters series) of equal length $N_{\text{spikes}}$.

We are now in position to discuss the main result of this paper, represented in Fig. 3. It allows us to conclude that the proposed methodology can be used for classification of benign childhood epilepsy and structural focal epilepsy patient EEGs, reaching approximately 72% accuracy. Other conclusions coming from Fig. 3 are discussed in section 5.
As we see, Strategy A (as compared to Strategy B or Strategy C) yields better or equal accuracy rate, especially with lower values of $N_{\text{spikes}}$. As additional mathematical regularisation – to eliminate statistical fluctuations of computed accuracy (achieved by Strategy A) we have derived analytical equation by fitting function:

$$A(N_{\text{spikes}}) = \frac{A}{B - C + D},$$

(4)

with estimated values $A = 1.1$, $B = 0.02$, $C = 1$ and $D = 0.01515$. The fitted function has an asymptote at 72.6% which means that at about 100 EDs the limit of the current classifier with current training data pool is approached. It should be noted that this limit could change if more training and testing data would be available.

Our results show that ANN reliability is most sensitive to number of total parameters (only upslopes or only downslopes can be used, but twice more EDs are needed in order to classify EEG with same degree of accuracy, see Fig. 3). The order of EDs seems to carry less information, experiments with randomising order of EDs did not show larger than standard deviation drop in detection reliability. However noise (falsely positively detected spikes) reduces the positive detection rate of patient group. If an EEG has less than 90 spikes, algorithm still can be used with lower degree of accuracy (see Fig. 3), for example 20 spike the accuracy is 65%.

4.2. The reason why ANN works but not SVM

We are going to discuss the importance of some statistical properties of precomputed upslopes and downslope series to reliability of ANN based classifier.

We denote that $k_U$ and $k_D$ are coefficients near the $x$ in linear function $f(x) = kx + b$ when the function fitted to respective parts (upslope or downslope) of ED (see Fig. 2). $N$ is percentage of spikes in each $k_U$ or $k_D$ bin.

ANN classifier and its results show some important points. BED parameters seem to be scattered in similar intervals but their standard deviations seems to be different (see Fig. 5, Fig. 6 and Fig. 7). This is probably the reason why ANN based classifier worked better than SVM based classifier in this case. Furthermore this agrees with initial hypothesis that spikes of
Figure 4: Scatter plot of upslopes and downslopes from EDs of patients from Group I and Group II. The scattering area of both Group I and Group II slopes is similar, but their standard deviations are different. That is why ANN classifier is needed to distinguish between Group II and Group I patient EEGs.

patients of Group II are more varied in their shape that ones in Group I. This also proves that these results are achieved not due to data having different average but due to their different spread. That this probably happened for two reasons: 1) the position parameters of EEG spikes of both patient groups still overlap (see Fig. 4) in high dimensional spaces, or 2) we do not have an optimal kernel for SVM classifier. This means it is not possible to calculate parameters of multidimensional surface that separates Group I and Group II any better than random guess. This and the fact that [24] got better results from ANN based classifier lead to choice of artificial neural network based (ANN) based classifier.

5. Conclusions

We demonstrated that reliability of algorithm increases with amount of spike parameters (only upslopes or downslopes can be used, but twice more EDs are needed to achieve same reliability). We show on our available analysed data set the maximum accuracy is 72% see fig 3 with saturation reached with about 50 upslope-downslope pairs taken into account. This shows that
Figure 5: Histograms of upslopes of patient groups (Group I and Group II) investigated. A) part demonstrates upslopes of Group I patients and B) results of Group II patients. Upslopes of Group II patients seem to have higher standard deviation.

Figure 6: Histograms of downslopes of patient groups (Group I and Group II) investigated. A) part demonstrates downslopes of Group I patients and B) results of Group II patients. Downslopes of Group II patients again seem to have higher standard deviation but the difference is not as obvious as with upslopes.
Figure 7: Kernel density plot [26] of upslopes and downslopes: A) for Group I, B) Group II.

It is possible to use machine learning tools in order to distinguish between EEGs of patients suffering from different forms of epilepsy (in our case Group I and Group II).

It is also worth to point out that the final accuracy (which in our case reaches 72% of presented algorithm depends on: 1) reliability of spike detection, 2) reliability of upslope and downslope estimation and 3) selection and size of training data set. Reliability of first and second stage has been discussed in [6, 16]. The algorithm presented is a three stage algorithm consisting of morphological filter-based spike detection algorithm [6] and spike measuring algorithm [16] as well as ANN based EEG classifier. This is a novel approach to this problem since other studies did not consider EEG spike parameters used here. The algorithm is dependant on quality of data since many false positives or just noisy data would obfuscate the natural variance of EEG spike parameters.

This study shows that about the 72% maximum is reached with size of chunk is at least 90 EEG spikes. If spike count is lower algorithm still can be used but with lower reliability, for example 60% with 10 EEG spikes. The analytical function fit of the accuracy suggests that the maximum accuracy of current algorithm is 72.6% however this could change if more training and testing data is available.

We theorise that analogical methodology could be used for classification of EEGs of other central nervous system conditions which are characterised by EDs. This means that this methodology possibly has much more applications.
yet to be discovered.

References


