Real-time detection of epileptic seizures in animal models using reservoir computing

Pieter Buteneers[a,*], David Verstraeten[a], Bregt Van Nieuwenhuyse[b], Dirk Stroobandt[a], Robrecht Raedt[b], Kristl Vonck[b], Paul Boon[b], Benjamin Schrauwen[a]

[a]*Electronics and Information Systems, Ghent University,*
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
[b]*Laboratory for Clinical and Experimental Neurophysiology, Ghent University,*
De Pintelaan 185, 9000 Ghent, Belgium

Abstract

In recent years, an increasing number of studies have investigated the effects of closed-loop anti-epileptic treatments. Most of the current research still is very labour intensive: real-time treatment is manually triggered and conclusions can only be drawn after multiple days of manual review and annotation of the electroencephalogram (EEG). In this paper we propose a technique based on reservoir computing (RC) to automatically and in real-time detect epileptic seizures in the intra-cranial EEG (iEEG) of epileptic rats in order to immediately trigger seizure treatment.

The performance of the system is evaluated in two different seizure types: absence seizures from genetic absence epilepsy rats from Strasbourg (GAERS) and limbic seizures from post status epilepticus (PSE) rats. The dataset consists of 452 hours iEEG from 23 GAERS and 2083 hours iEEG from 22 PSE rats.

In the default set-up the system detects 0.09 and 0.13 false positives per seizure and misses 0.07 and 0.005 events per seizure for GAERS and PSE rats respectively. It achieves an average detection delay below 1 second in GAERS and less than 10 seconds in the PSE data. This detection delay and the number of missed seizures can be further decreased when a higher false positive rate is allowed.

Our method outperforms state-of-the-art detection techniques and only a few parameters require optimization on a limited training set. It is therefore suited for automatic seizure detection based on iEEG and may serve as a useful tool for epilepsy researchers. The technique avoids the time-consuming manual review and annotation of EEG and can be incorporated in a closed-loop treatment strategy.

Keywords: automatic seizure detection, experimental animal models for

---

*Corresponding author. Tel.: +32 9 264 33 68; fax: +32 9 264 35 94.
Email address: `Pieter.Buteneers@UGent.Be` (Pieter Buteneers)

epilepsy, reservoir computing, neural networks, EEG classification

1. Introduction

Interictal and ictal EEG from patients with epilepsy is characterised by typical patterns that are usually visually detected for diagnostic purposes. In a research setting the EEG is often used to evaluate therapeutic efficacy. In this case, in humans and especially in animal research settings, many hours of EEG require manual review and analysis. Automated seizure detection decreases the workload and may also be more reliable compared to hours of visual analysis.

Many [1, 2, 3, 4, 5, 6] automated seizure detection programs, based on various techniques are available. Only a few of them allow real-time seizure detection [3, 6]. The advantage of accurate real-time seizure detection is the potential to incorporate this detection into a so called closed-loop system that allows immediate triggering of an intervention at the time of seizure occurrence such as: fast working anti-epileptic drugs, Deep Brain Stimulation (DBS) [7, 8], Vagus Nerve Stimulation (VNS) [9], ...

In recent years machine learning in general and more specifically artificial neural networks are more commonly used to build seizure detection algorithms. Reservoir Computing (RC) is a recurrent neural network (RNN) training technique which has been shown to achieve state-of-the-art performance for annotating EEG [1]. It keeps the performance of regular RNNs but reduces the training time dramatically [10]. For seizure detection RC has the advantage that it is a learned non-linear dynamical system which is opposed to the mostly linear classification techniques used in literature. This paper improves the annotation accuracy of [1] and extends it with the ability to detect the seizures with a low latency. To investigate the accuracy of RC to detect epileptiform discharges in real-time, we apply it to the EEG of genetic absence epilepsy rats from Strasbourg (GAERS) and post status epilepticus (PSE) rats.

GAERS and PSE rats are well known and widely used models for human absence and temporal-lobe epilepsy respectively [11]. The EEG is characterised by very stereotyped ictal epileptiform discharges shown in Figures 1 and 2. The models have a very good electroclinical correlation and the animals have a high seizure frequency. For these reasons they are a useful model for drug screening and the application of seizure detection algorithms.

2. Materials

This study compares several methods for epileptic seizure detection on two different seizure types: absence seizures from GAERS and limbic seizures from PSE rats. In both cases, the EEG was recorded with a custom-built amplifier. Afterwards it was subsampled to 200Hz before being evaluated by experienced encephalographers.

The complete dataset consists of 454 hours of data from GAERS and 2083 hours of data from PSE rats. For GAERS the training set, 5.75 hours in total,

Figure 1: An example of a spike and wave discharge (SWD) caused by an absence seizure in genetic absence epilepsy rats from Strasbourg. In (A) the EEG signal of one intra-cranial channel is shown. The seizure starts at time = 2 s and stops at time = 8.9 s. (B) shows the spectrogram of the EEG signal with a Hamming-window of 128 samples and an overlap of 120 samples.

consists of the first 15 minutes of EEG per rat that contained at least 90 seconds of ictal EEG. The training set for the PSE data, 44 hours of data, consists of the first 10 seizures of each rat in the dataset combined with about 5 minutes of pre-ictal and post-ictal EEG. The rest of the data following the training data was used for testing.

2.1. Genetic absence epilepsy rats from Strasbourg

GAERS are a strain of Wistar rats that all exhibit spontaneous absence seizures characterized by paroxysmal unresponsiveness to environmental stimuli and cessation of ongoing activity. These absence seizures, which are displayed as synchronous spike and wave discharges (SWDs) on the EEG, occur mostly when the animal is in a state of quiet wakefulness. However, they are rare during periods of active arousal and sleep. The number of seizures and their duration increase with age, until it reaches a maximum at about 6 months. The EEG of SWDs shows a fundamental frequency in the range of 7 to 12 Hz and several harmonics (see Figure 1), an amplitude varying from 300 to 1000 µV and a duration from 0.5 to 120 s.

Dataset A was made during a study to evaluate the effect of acute and non-acute high (130 Hz) and middle high (60 Hz) frequency DBS on the occurrence of SWDs [7]. The rats from dataset B were part of a study to evaluate the effect of long-term VNS.

All EEG fragments were visually reviewed, the data contaminated with stimulation artefacts was removed, one EEG channel was selected and all present SWDs with a minimum seizure length of 0.5 s were marked by an experienced encephalographer. These annotations were used as the 'gold standard' in this study. From study A, 64.5 hours of single-channel depth EEG-data recorded in the anterodorsal thalamus from 12 different rats was used. 23% of the total time contained the 3468 seizures which lasted on average 15 seconds. Study B yielded 390 hours of single-channel scalp EEG-data recorded over the frontoparietal cortex from 11 rats. A total number of 6183 seizures made up 4.5% of the data and lasted 10 seconds on average. Each of the seizures lasted between 0.5 and 110 s.

2.2. Post status epilepticus rats

Kainic acid is a potent central nervous stimulant, isolated from the seaweed digenea simplex. This excitotoxic product is an agonist of a subclass of ionotropic glutamate receptors and a systemic injection in healthy rats triggers

Figure 2: An example of a limbic seizure in a post status epilepticus rat. In (A) the EEG signal is shown, the seizure starts at time = 7 s and stops at time = 31.2 s. (B) shows the spectrogram of the EEG signal with a Hamming-window of 128 samples and an overlap of 120 samples.

a cascade of molecular and cellular events eventually leading to status epilepticus, followed by a period of gradual increase in seizure frequency, which eventually stabilizes. Finally, rats display spontaneous, secondary generalized limbic seizures which resemble those seen in temporal-lobe epilepsy patients [12].

During annotation, spontaneous EEG seizures were recognized against background by their large amplitude (more than 3 times baseline amplitude), high-frequency EEG activity ($\geq$ 5 Hz), with characteristic high temporal correlation and progression of spike frequency. Figure 2 shows an example of a limbic seizure.

Dataset C was made during a study to evaluate the effect of long-term high frequency (130 Hz) and Poisson distributed high frequency (on average 130 Hz) DBS on the occurrence of limbic seizures [8]. An experienced encephalographer evaluated all EEG fragments visually and marked all present seizures in dataset C. This resulted in 913 hours of four channel EEG from 11 different rats. Approximately 2.5% of this data consisted of 1541 seizures which have a duration of 9 to 240 seconds with an average of 54 seconds. In five animals DBS was applied. Episodes from this subset (C*) contained some episodes of EEG contaminated with stimulation artefacts (in the rest of this work referred to as C^*_{stim}). An example of a stimulation artefact is shown in Figure 3.

Figure 3: An example of stimulation artefacts caused by stimulation at 130 Hz. Because of the low sample rate (200 Hz) this is only visible as spikes in the EEG for example at time = 3.05 s and 9.1 s (up arrows). These spikes are somewhat similar to the many epileptic spikes as for example at time = 7.6 s (down arrow).

Study D compared the therapeutic effect of DBS in the hippocampus and midline thalamic nuclei. Both experimental therapies were evaluated for their effect on the frequency of spontaneous seizures in the PSE model. The EEG fragments were visually evaluated and annotated. This resulted in 1105 hours of EEG from 7 different rats without simulation artefacts. 1374 seizures were recorded in total, which lasted on average 42 seconds or between 12 and 220 seconds and represent 1.4% of the data.

In study E the effect of introducing stem cells from foetal mice brains in the epileptogenic areas was studied on the occurrence of limbic seizures. All EEG fragments were visually evaluated and all seizures were marked. From this study 69 hours of 4 channel EEG from 4 different rats was used. Dataset E contained 113 seizures that were located in about 2.5% of the data and lasted

23 to 360 seconds with an average of 51 seconds.

Datasets C, D and E consist of 4 channel hippocampal EEG with a referential montage for each rat. From these 4 channels one was chosen with visually the most significant difference between ictal and inter-ictal EEG based on the first 4 seizures.

3. Methods

3.1. Seizure detection using reservoir computing

Figure 4: A schematic representation of the detection method presented in this work.

The proposed detection method, RC with Bayesian relevance regression (RC-BRR), is constructed out of three parts: a preprocessing stage where features get extracted from the EEG, a classification stage which is based on RC and uses BRR for classification and a post-processing stage where two thresholds are applied. A schematic representation is shown in Figure 4. It is an extension of the RC with ridge regression (RC-RR) method presented in [1]. The preprocessing stage is adapted such that it can be used in an online set-up, and the training of RC is done using BRR as opposed to RR. BRR has the advantage that it is able to train a better common model for all animals by automatically scaling the influence of each example according to its relevance. In this section a brief explanation will be given, for more details we refer to Appendix A.

Figure 5: An example of a SWD caused by an absence seizure in GAERS. In (A) the EEG signal of one intra-cranial channel is shown. The seizure starts at Time = 2 s and stops at Time = 8.9 s. The preprocessing is shown in (B), the wavelet filtered signal, and (C): the rescaled absolute value of the filtered signal. (D) shows the activation values of 5 of the 200 reservoir neurons and (E) shows the generated output that was trained using BRR together with the two thresholds. Every sample above the high threshold is a detection, illustrated with the bold part of the high threshold line (the upper horizontal line). Samples neighboring these detected samples that are above the second threshold are used for marking and illustrated by the bold part of the low threshold line.

The GAERS data is preprocessed by filtering with a FIR based approximation of a wavelet filter that was proposed for seizure detection in [13]. The PSE EEG is frequency filtered with a band-pass filter for the beta-band. These filters were selected using the feature selection method from [1]. The wavelet filtered signal of an SWD is shown in Figure 5.B. Next the signal is subsampled using an interval based median filter. The absolute value of this signal is called the foreground signal. Since there is a high variability in signal amplitudes between different rats, a background signal is estimated to serve as a reference level for

rescaling. This signal is estimated as the median of the foreground signal values of the past hour. The rescaled foreground signal is used as input for RC (Figure 5.C).

To classify the rescaled foreground signal a Recurrent Neural Network (RNN) is used. Most training techniques train all the weights between the neurons in the RNN. Reservoir Computing (RC) [14] on the other hand, makes use of a randomly created neural network, called a reservoir, from which only a single linear output is trained. In this way the long training time and stability issues of regular RNN training are avoided without losing the desired generalization abilities.

In the RC set-up each non-zero input sample will excite this dynamical system and push the reservoir to a new state. In Figure 5.D the activation values of 5 of the 200 neurons are shown. To generate the output a linear combination of these activations is made using BRR. This iterative process finds the optimal weights of the linear combination to create a general model for all rats. It scales the influence of each seizure example according to how statistically relevant it is for the common model, such that it does not overspecialize in detecting uncommon seizure examples.

The output of the linear combination trained using BRR results into the signal shown in Figure 5.E. To classify this continuous output value two thresholds are applied as post-processing: a high and a low threshold. The high threshold is used for seizure detection. Every sample above the high threshold is considered a seizure sample and thus part of a seizure (see Figure 5.E). Lowering the high threshold allows for a shorter detection delay at the cost of more false positives. To gain annotation precision without generating more false positives, a low threshold is used. Every sample neighbouring a seizure sample, that is above this low threshold, is also considered as part of a seizure.

3.2. Other detection methods

For comparison two state-of-the-art methods from literature were applied on the same dataset: the real-time Adapted Osorio-Frei Algorithm (AOFA) presented by Haas et al. in [6] and the annotation method RC-RR presented by Buteneers et al. in [1]. Although the AOFA is a detection method for humans, the original, simpler Osorio-Frei algorithm presented in [13] has been previously applied to animal data [1, 2] and can be considered as current state of the art for real-time seizure detection in animal models [1]. For more details on the techniques we refer to literature [1, 6]. To allow a fair comparison between the different methods the interval sizes and/or other free parameters were optimized on the same training set.

A third method that was used for comparison is a linear seizure detection method based on the presented technique but where the reservoir is left out and the same feature selection algorithm and preprocessing stage is used. In the next step a linear classifier is trained on these features using BRR. The output of this classifier is then post processed using the two-threshold method of the presented technique. For more details we refer to Appendix A and [1].

3.3. Evaluation measures

The gold standard used for comparing the different detection methods used in this study was the scoring by experienced encephalographers. The following measures were used to compare the detection delays of the real-time seizure detection methods:

- Δ_{delay} is the average detection delay in seconds. It is only determined for correct seizure detections. As a lower bound, the first inter-ictal sample after the previous seizure is used and as an upper bound, the last marked sample of the to be detected seizure.

- σ_{delay} is the standard deviation on the detection delay.

As a measure for the number of seizures that were falsely or not detected, the following two measures were used:

- FPPS or False Positives Per Seizure is the number of falsely detected seizures divided by the total number of true seizures.

- FNPS or False Negatives Per Seizure is the number of missed detections divided by the total number of true seizures.

To compare the sample based annotation precision, the following 3 measures were used:

- Sensitivity is the percentage of correctly classified seizure samples.

- Specificity is the percentage of correctly classified inter-ictal samples.

- BER or Balanced Error Rate is the average of the error on the sensitivity and the specificity. This error measure is used during training to optimize the parameters.

Each of these measures is calculated for each rat individually. Then the mean and standard deviation is calculated over all the rats and used for comparison. Each rat thus has the same influence on the results, independent of the amount of data that was recorded for this rat.

4. Results

4.1. Default performance

In order to evaluate the overall performance each method was trained on the training set and tested on the test set. During training, meta parameters were selected to achieve minimal BER. In Table 1 the results are given for the different methods. Because of the random initialisation of RC, 10 different systems were trained on each training set and the system that performed best on the training set was used.

Table 1: The results for the default settings. The averages and standard deviations (between brackets) over the different rats of the following measures is given: detection delay, standard deviation on the detection delay, false positives per seizure, false negatives per seizure, the sensitivity, specificity and balanced error rate.

GAERS	RC-BRR	AOFA	Linear	RC-RR
Δ_{delay} (s)	**0.97 (0.33)**	2.45 (0.6)	1.01 (1.05)	n/a
σ_{delay} (s)	**0.93 (0.44)**	1.78 (1.25)	4.77 (10.91)	n/a
FPPS	**0.091 (0.163)**	1.40 (2.96)	2.23 (3.01)	0.96 (1.05)
FNPS	**0.065 (0.055)**	0.177 (0.200)	0.020 (0.037)	0.067 (0.064)
Sens (%)	**96.2 (2.4)**	67.6 (20.4)	95.9 (5.1)	96.4 (2.0)
Spec (%)	**98.2 (1.6)**	97.8 (2.0)	88.2 (5.8)	96.2 (2.3)
BER (%)	**2.8 (1.4)**	17.3 (10.1)	7.9 (3.3)	3.7 (1.8)

PSE	RC-BRR	AOFA	Linear	RC-RR
Δ_{delay} (s)	**9.4 (2.1)**	20.3 (3.5)	15.1 (4.2)	n/a
σ_{delay} (s)	**5.7 (6.5)**	7.4 (8.7)	12.0 (16.3)	n/a
FPPS	**0.128 (0.120)**	3.43 (3.77)	1.11 (1.14)	0.256 (0.211)
FNPS	0.005 (0.016)	0.032 (0.067)	0.026 (0.056)	**0.003 (0.064)**
Sens (%)	**95.8 (2.3)**	75.4 (13.8)	81.7 (13.2)	87.2 (9.7)
Spec (%)	**98.1 (1.1)**	96.4 (3.0)	96.9 (2.9)	97.9 (2.4)
BER (%)	**3.1 (1.3)**	14.1 (7.6)	10.7 (6.9)	7.4 (4.8)

4.2. Depth versus epidural EEG

Datasets A, C, D and E were recorded using depth electrodes, dataset B on the other hand was recorded using epidural electrodes. To compare whether the system performs better when the EEG is recorded using depth or epidural electrodes we present the results for the RC-BRR method on datasets A and B separately in Table 2. It shows that when depth electrodes are used a lower detection delay is achieved with less FPPS and FNPS.

4.3. EEG with or without stimulation artefacts

The EEG of several animals from study C contains stimulation artefacts. In Table 3 the results are shown on the data without stimulation artefacts C^*_{clean} and the data with simulation artefacts C^*_{stim}. It shows that artefacts have no significant influence on the detection delay but they increase the number of FNPS and the sample based detection error while only slightly reducing the number of FPPS.

4.4. Performance versus detection delay

The presented method, the AOFA and the linear method can be altered to achieve a lower detection delay. Each method has one specific parameter that has the most influence. For the presented method the threshold changes

Table 2: A comparison in performance of the RC-BRR method on depth and epidural EEG from datasets A and B respectively. The averages and standard deviations (between brackets) over the different rats of the following measures is given: detection delay, standard deviation on the detection delay, false positives per seizure, false negatives per seizure, the sensitivity, specificity and balanced error rate.

GAERS	A (depth)	B (epidural)
Δ_{delay} (s)	0.82 (0.22)	1.14 (0.36)
σ_{delay} (s)	0.85 (0.45)	1.02 (0.44)
FPPS	0.039 (0.026)	0.147 (0.226)
FNPS	0.024 (0.021)	0.109 (0.046)
Sens (%)	97.2 (2.1)	95.2 (2.6)
Spec (%)	97.2 (1.6)	99.3 (0.2)
BER (%)	2.8 (1.7)	2.8 (1.2)

Table 3: A comparison in performance of the RC-BRR method on EEG with or without stimulation artefacts. It was tested on the animals from study C that were stimulated: C^*_{clean} represents the EEG data without stimulation artefacts and C^*_{stim} the EEG data with stimulation artefacts. The averages and standard deviations (between brackets) over the different rats of the following measures is given: detection delay, standard deviation on the detection delay, false positives per seizure, false negatives per seizure, the sensitivity, specificity and balanced error rate.

PSE	C^*_{clean}	C^*_{stim}
Δ_{delay} (s)	7.1 (2.5)	6.7 (3.3)
σ_{delay} (s)	3.9 (2.3)	4.1 (5.9)
FPPS	0.084 (0.080)	0.048 (0.069)
FNPS	0.013 (0.027)	0.043 (0.043)
Sens (%)	96.6 (2.0)	89.8 (13.3)
Spec (%)	98.6 (0.3)	96.6 (1.3)
BER (%)	2.4 (0.9)	7.3 (6.4)

the response time. Lowering the threshold allows for an earlier detection but increases the number of falsely detected seizures. Increasing the threshold delays the detection and increases the number of missed seizures but decreases the number of false detections. To allow the AOFA method to detect seizures with a lower delay, the minimal duration of a seizure needs to be lowered. The threshold plays a less significant role and is thus optimized on the training set for each minimal duration setting. The most important parameter for the linear method is the interval length L (see Section Appendix A.1). The shorter this interval the shorter the detection delay, but also the more falsely detected seizures.

In the left graphs of Figures 6 and 7 the FPPS is plotted as a function of the detection delay for the GAERS and PSE dataset respectively. The right graphs of Figures 6 and 7 show the FNPS as a function of the delay. Settings

that resulted in more than 4 FPPS were not included.

Figure 6: On the left the number of False Positives Per Seizure and on the right the number of False Negatives Per Seizure in function of the detection delay in seconds for the GAERS dataset

Figure 7: On the left the number of False Positives Per Seizure and on the right the number of False Negatives Per Seizure in function of the detection delay in seconds for the PSE dataset.

4.5. Comparison of the 'golden standards'

Figure 8: An example of a scratch artefact which starts at time = 3s and ends at time = 6s.

It is a well known fact that humans often disagree when marking epileptic seizures. We took a small dataset used to train students in marking epileptic seizures of 3 PSE rats. These 24 hours for each rat contain in total 72 seizures and 183 artefacts, mostly scratch artefacts as shown in Figure 8. As golden standard we asked 4 experienced encephalographers to mark the data and come to an agreement. In table 4 we show the performance of our method, trained on the training sets of dataset C, D and E, and the results of 2 experienced encephalographers, who had never seen the data before, against this golden standard.

5. Discussion

The main findings of this study are shown in Table 1. We can conclude that, when the techniques are optimized for optimal BER, the presented RC-BRR method significantly outperforms state-of-the-art techniques for annotation as well as real-time detection purposes. For the absence seizures in GAERS, which last on average 10 to 15 seconds, an average detection delay of less than a second was achieved. This resulted in 0.09 FPPS and 0.065 FNPS which means that only 1 out of 12 detections is a false detection, there are 1.9 false detections per hour and only 1 out of 16 seizures is missed. RC-BRR achieves an average BER of 2.8% with a sensitivity of 96% and a specificity of 98%. A detection delay of 9.2 seconds was achieved for the limbic seizures in the PSE dataset which last on average 50 seconds. This resulted in 0.13 FPPS and 0.005 FNPS, so that 1 out of 9 detections is a false detection, there are 0.19 false positives per hour and only 1 out of 200 seizures is missed. On a sample basis a BER of 3.1% was achieved with a sensitivity of 96% and a specificity of 98%.

RC-BRR not only outperforms the other techniques in average performance it also has a significantly lower standard deviation. This shows that the performance of the presented method is less influenced by differences between each animal in the dataset. It is thus better able to create a general model for seizure

10

Table 4: A comparison in performance of the presented method and two experienced encephalographers (E1 and E2) on a small student dataset. As golden standard 4 different encephalographers were asked to come to an agreement. The averages and standard deviations (between brackets) over the different rats of the following measures is given: false positives per seizure, false negatives per seizure, the sensitivity, specificity and balanced error rate. The data contains 183 artefacts, on average 4.7 (4.8) artefacts per seizure, most of which are scratch artefacts.

$\mathbf{PSE}_{student}$	RC-BRR	E1	E2
FPPS	0.15 (0.17)	0.010 (0.016)	0.13 (0.18)
FNPS	0.096 (0.140)	0.029 (0.029)	0.010 (0.017)
Sens (%)	85.0 (4.5)	91.4 (3.2)	90.8 (5.4)
Spec (%)	99.5 (0.3)	100 (0.0003)	99.9 (0.0002)
BER (%)	7.9 (2.3)	4.3 (1.6)	4.6 (2.7)

detection. The comparison with the RC-RR method shows that the common model created by BRR significantly outperforms the method without BRR. Because the presented technique is a non-linear method and most methods from literature are linear, a second seizure detection method was presented here. In this linear method RC is replaced with a linear classifier. Because RC-BRR outperforms this linear method it shows the advantage of using RC as a classifier. In practice the presented linear method is a more advanced version of the AOFA where multiple features can be used instead of one. It was not only able to outperform AOFA because the feature selection algorithm selected more than 1 feature but also because it was trained using BRR.

From Table 2 one can infer that there is a difference in performance when depth electrodes are used as opposed to epidural electrodes. The system achieves a slightly lower detection delay with a significantly lower number of FPPS and FNPS. In our experience this is due to the fact that there is less noise in the EEG which makes that the difference between ictal and inter-ictal EEG is more profound in terms of signal strength and signal shape. For optimal performance it is thus advised to use the technique in combination with depth electrodes.

Counter intuitively stimulation artefacts do not increase the number of FPPS as shown in Table 3. This is possibly due to the fact that a stimulation of 130 Hz was applied which does not fall within the beta band that was used for pre-processing. The artefacts however do increase the background signal level which results in a higher number of FNPS and a lower sensitivity and number of FPPS. If the exact time and duration of the stimulation is known, one could ignore these samples while estimating the background signal. However, the increased background scaling has no influence on the detection delay.

If the algorithms are altered to result in a lower detection delay, this comes at the cost of more false detections as shown in Figures 6 and 7. This is because the first few samples of ictal EEG better resemble normal EEG. The number of missed seizures on the other hand decreases because less outspoken seizures will also be detected. Figures 6 and 7 show that, for a given number of FPPS or

FNPS, RC is able to detect the seizures with less delay than the other methods. If 4 FPPS are allowed, i.e. 1 in 5 detections is a true positive, RC achieves an average detection delay of 0.2s and 2.1s, the linear method needs 0.8s and 13s, and the AOFA-method will take more than 2s and 15s to detect a seizure on the GAERS and PSE data respectively. When minimal FPPS and FNPS are required we see that RC-BRR is the only method that was able to achieve less than 10% FPPS and FNPS at the same time.

In [3] a technique to detect epileptic seizures in PSE rats was presented that was able to predict seizures by about 10 to 15 seconds. However, this resulted in more than 100 FPPS[1]. Even though our technique achieves similar performance if such a high number of FPPS is allowed, we do not believe it has any practical value. If for instance DBS or VNS is applied during a seizure, which lasts about 1 minute in PSE rats, you apply continues stimulation once you reach more than 15 seizures per 24 hours. PSE rats however have on average 35 seizures per 24 hours.

Adapting the methods to achieve a lower detection delay deteriorates the sample based annotation accuracy of the methods. Because annotation is mostly done retrospectively, the results are not shown here. One could use a set-up for real-time detection and another set-up for annotation. Since for the presented method no new training is required and only the threshold is altered to achieve a lower detection delay, the same set-up can be used for both tasks. One only needs to use two different high thresholds: one for real-time detection and one for annotation.

Researchers often disagree when marking EEG, which is reaffirmed in table 4. Even though our method performed worse than both experienced encephalographers, it is fair to say that the performance of our method is comparable with atleast one of the encephalographers. The artefacts that caused false positives were all scratch artefacts and often coincided with the errors made by the encephalographers. Other artefacts or sleep spindels did not generate any false positives.

In summary we can thus conclude that the presented method is well suited as an aid for research in animal models for epilepsy. It achieves state of the art performance for annotation purposes and can be used in a closed loop system where a low detection delay and a low number false and missed detections is desired.

Acknowledgement

---

[1]This was not specifically stated in [3], but it can be deduced from the results and was confirmed by one of the authors.

Appendix A. Seizure detection using Reservoir Computing

In this appendix the detection method is explained in more detail. It is constructed out of four main parts: a preprocessing stage where EEG features get extracted, a non-linear mapping stage which is based on RC, a classification stage that uses BRR for training and a thresholding stage.

Appendix A.1. Preprocessing

For preprocessing of the GAERS data a FIR-filter based on the level 3 DAUB4 wavelet filter, first presented for seizure detection in [13], is used to extract the relevant information from the EEG. The PSE data is frequency filtered between 12 and 30 Hz, the beta-band. These features were selected using the feature selection algorithm presented in [1]. This signal, sampled at 200 Hz, is then subdivided in non-overlapping intervals with length $L = 0.02$ s and $L = 0.2$ s, for the GAERS and PSE data respectively, from which the foreground input signal FG is calculated as follows:

$$FG_k = \mathrm{median}\{|s_k|, |s_{k-1}|, ..., |s_{k-200L+1}|\},$$

where s_k represents the filtered EEG signal. Since there is a high variability in signal amplitudes between different animals, a background signal is estimated for each of the EEG-features to serve as a reference level for rescaling. The background signal is estimated from the foreground signal as follows:

$$BG_k = \mathrm{median}\{FG_k, ..., FG_{k-H}\},$$

with H the number of FG-samples in one hour. This is a quantile based estimation technique as proposed in [15] and is based on the assumption that epileptic seizures occur less than half of the time. To save calculation time BG_k is only updated every minute and thus left unchanged for the next 59 s of FG-samples.

Appendix A.2. Reservoir computing

The operation of the RC [16] can be described as follows. We use $\mathbf{x}[k]$ to represent the current activation values of the neurons in the reservoir at time k, $\mathbf{u}[k]$ as the input vector, $\mathbf{y}[k]$ for the desired output and $\hat{\mathbf{y}}[k]$ for the output generated by the RC system. The inputs of the neurons in the reservoir are connected with the bias as well as with the input and the output of all the neurons in the reservoir. The weights of these connections are represented respectively by the weight matrices $\mathbf{W}_{bias}$, $\mathbf{W}_{inp}$ and $\mathbf{W}_{res}$. If n is the number of neurons these matrices respectively have the following shapes: n-by-1, n-by-N and n-by-n with N equal to the number of inputs. The elements of the bias weight matrix $\mathbf{W}_{bias}$ and the internal weight matrix $\mathbf{W}_{res}$ are initially uniformly distributed between -1 and 1. All elements in the input weight matrix $\mathbf{W}_{inp}$ are randomly set to -1 or +1. The sparseness of these matrices as well as the initialization process are however, not critical for the performance. If basic sigmoid

neurons, a weighted sum followed by the hyperbolic tangent function, are used, the state update equation is given by:

$$\mathbf{x}[k+1] = \tanh(\mathbf{W}_{res}\mathbf{x}[k] + \mathbf{W}_{inp}\mathbf{u}[k] + \mathbf{W}_{bias}).$$

In this work, we used leaky integrator neurons, i.e. basic sigmoid neurons followed by a first-order low-pass filter. The state equation now becomes:

$$\mathbf{x}[k+1] = (1-\gamma)\cdot\mathbf{x}[k] \; + \gamma\tanh(\mathbf{W}_{res}\mathbf{x}[k] + \mathbf{W}_{inp}\mathbf{u}[k] + \mathbf{W}_{bias}).$$

In this equation γ represents the leak rate which sets the cutoff frequency of the low-pass filter in the neurons. This extra parameter of leaky integrator neurons is used to tune the reservoir memory and timescales [17].

The processing power of reservoirs is the greatest when they operate at the edge of stability [18]. Therefore, the spectral radius, the largest norm of the eigenvalues of $\mathbf{W}_{res}$, the weights of $\mathbf{W}_{inp}$ and $\mathbf{W}_{bias}$ and the leak-rate are scaled to achieve optimal reservoir dynamics [19]. The meta parameters described in [1] were used in this work.

Appendix A.3. Bayesian relevance regression

To generate the output the following equation is used:

$$\hat{\mathbf{Y}}[k] = \mathbf{W}_{out}\begin{bmatrix}\mathbf{X}[k] \\ 1\end{bmatrix},$$

in which '1' represents the connection from the bias to the output. The weights of the output weight matrix $\mathbf{W}_{out}$ are trained using BRR. This form of regression was developed to provide a way to automatically scale the influence of each seizure example according to its statistical relevance. This allows to train one common model that is suited for every animal of the same seizure type without over-fitting to uncommon seizure examples.

Analogously to the formulation of the automatic outlier detection technique presented in [20] we can deduct BRR as follows. If we consider a data set of input-target pairs (x_i, y_i), we can follow the probabilistic formulation of bayesian regression [21] and assume that the targets are samples from a model with additive noise:

$$y_i = \mathbf{w}^T x_i + \epsilon_i,$$

where ϵ_i is a sample from a noise process which we assume to be a zero mean Gaussian with a variance $\beta_i^{-1/2}$. As opposed to (sparse) Bayesian regression [21, 22] we consider each data point a sample of a different noise process. We get the following for the probability of $\mathbf{Y}$ given the data and the model and the probability of the model:

$$
\begin{aligned}
p(\mathbf{Y}|\mathbf{X}\mathbf{w},\boldsymbol{\beta}^{-\frac{1}{2}}) &= N(\mathbf{Y}|\mathbf{X}\mathbf{w},\boldsymbol{\beta}^{-\frac{1}{2}}) && \text{(A.1)}\\
&= \left(\frac{1}{2\pi}\right)^{\frac{M}{2}}|\boldsymbol{\beta}|^{\frac{1}{2}}exp\{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\mathbf{w})^T\boldsymbol{\beta}(\mathbf{Y}-\mathbf{X}\mathbf{w})\} && \text{(A.2)}\\
p(\mathbf{w}|\alpha) &= N(\mathbf{w}|\boldsymbol{\mu},\alpha) && \text{(A.3)}\\
&= \left(\frac{1}{2\pi}\right)^{\frac{N}{2}}|\alpha|^{\frac{N}{2}}exp\{-\frac{\alpha}{2}(\mathbf{w}-\boldsymbol{\mu})^T(\mathbf{w}-\boldsymbol{\mu})\}, && \text{(A.4)}
\end{aligned}
$$

with N the number inputs, M the number of data points and $\boldsymbol{\beta}$ a diagonal matrix with on the i'th diagonal element the variance β_i of the model on the i'th data point.

If we consider the posterior covariance matrix $\mathbf{S}_n^{-1} = \alpha\mathbf{I} + \mathbf{X}^T\boldsymbol{\beta}\mathbf{X}$, we can compute the weights and rewrite the evidence function, the probability of the data given the hyper parameters, as follows:

$$\mathbf{w} \;=\; \mathbf{S}_n\mathbf{X}^T\boldsymbol{\beta}\mathbf{Y} \tag{A.5}$$

$$p(\mathbf{Y}|\alpha,\boldsymbol{\beta}) \;=\; \left(\frac{1}{2\pi}\right)^{\frac{N+M}{2}} |\alpha|^{\frac{N}{2}}|\boldsymbol{\beta}|^{\frac{1}{2}}|\mathbf{S}_n|^{\frac{1}{2}}\,exp\{-E(\mathbf{w})\}, \tag{A.6}$$

with $E(\mathbf{w}) = \frac{1}{2}(\mathbf{Y}-\mathbf{X}\mathbf{w})^T\boldsymbol{\beta}(\mathbf{Y}-\mathbf{X}\mathbf{w}) + \frac{\alpha}{2}(\mathbf{w}-\boldsymbol{\mu})^T(\mathbf{w}-\boldsymbol{\mu})$. Maximising this evidence function is equivalent to maximising the logarithm of this evidence function. Taking the derivative of $\ln(p(\mathbf{Y}|\alpha,\boldsymbol{\beta}))$ to α and β_i and setting to zero gives the following:

$$\frac{\delta}{\delta\alpha}\ln(p(\mathbf{Y}|\alpha,\boldsymbol{\beta})) \;=\; \frac{N}{2\alpha} - \frac{1}{2}\mathrm{tr}(\mathbf{S}_n) - \frac{1}{2}||\mathbf{w}-\boldsymbol{\mu}||^2 = 0 \tag{A.7}$$

$$\frac{\delta}{\delta\beta_i}\ln(p(\mathbf{Y}|\alpha,\boldsymbol{\beta})) \;=\; \frac{1}{2\beta_i} - \frac{1}{2}\mathrm{tr}(\mathbf{S}_n(x_i x_i^T)) - \frac{1}{2}(y_i - \mathbf{w}^T x_i)^2 = 0. \tag{A.8}$$

These equations yield:

$$\alpha_i \;=\; \frac{N}{\mathrm{tr}(S_n) + (\mathbf{w}-\boldsymbol{\mu})^2} \tag{A.9}$$

$$\beta_i \;=\; \frac{1}{\mathrm{tr}(S_n(x_i x_i^T)) + (y_i - \mathbf{w}^T x_i)^2}. \tag{A.10}$$

$$\tag{A.11}$$

$\mathbf{S}_n$ in these equations still contains α and the β_is we try to find, but iteratively calculating $\mathbf{S}_n$ and $\mathbf{w}$ followed by α and each β_i converges to the optimal values [20, 21, 22].

To use this system for automatic relevance determination we cluster the data points such that we get one cluster for each seizure example i out of K examples. If N is the number of inputs, the update algorithm to find the optimal weights is given in Algorithm 1.

Appendix A.4. Classification based on two-thresholds

The output of the reservoir $\hat{\mathbf{Y}}[k]$ is a continuous output that approximates the desired output $\mathbf{Y}[k]$. To transform this continuous-valued output to a discrete classification label, a two-threshold setup was used. From the moment the output $\hat{\mathbf{Y}}[k]$ is higher than the first, high threshold a detection is made. To improve annotation precision, all direct neighbouring samples higher than the second, low threshold are then also annotated as a seizure. All other samples are considered as regular EEG.

The thresholds were optimized for a minimal BER (see Section 3.3) during training. To lower the detection delay the high threshold can be decreased. This results in less missed detections but also results in more false positives. For annotation the output is up-sampled to 200 Hz, the original sample rate of the input, after applying the thresholds.

[1] P. Buteneers, D. Verstraeten, P. van Mierlo, T. Wyckhuys, S. Staelens, D. Stroobandt, and B. Schrauwen. Automatic Detection of Epileptic Seizures on Intra-cranial EEG from Rats using Reservoir Computing. *AI in Medicine*, 53:215–223, 2010.

Algorithm 1 Bayesian Relevance Regression
---
for i in [1,K] **do**
    $\mathbf{A}_i = \mathbf{X}_i^T \mathbf{X}_i$
    $\mathbf{B}_i = \mathbf{X}_i^T \mathbf{Y}_i$
    $M_i = \text{length}(\mathbf{X}_i)$
    $\beta_i = 1$
end for
$\alpha = 0$
while not converged **do**
    $\mathbf{S}_n^{-1} = \alpha \mathbf{I} + \sum_i^K \beta_i \mathbf{A}_i$
    $\mathbf{w} = \mathbf{S}_n \sum_i^K \beta_i \mathbf{B}_i$
    $\alpha = \frac{N}{\text{tr}(\mathbf{S}_n) + \mathbf{w}^T \mathbf{w}}$
    for i in [1,K] **do**
        $\beta_i = \frac{M_i}{\text{tr}(\mathbf{S}_n \mathbf{A}_i) + ||\mathbf{X}_i \mathbf{w} - \mathbf{Y}_i||^2}$
    end for
end while
---

[2] P. Van Hese, J-P. Martens, L. Waterschoot, P. Boon, and I. Lemahieu. Automatic Detection of Spike and Wave Discharges in the EEG of Genetic Absence Epilepsy Rats from Strasbourg. *IEEE Transactions on Biomedical Engineering*, 56(3):706–717, 2009.

[3] M. Nandan, S. Talathi, S. Myers, W. Ditto, P. Khargonekar, and P. Carney. Support vector machines for seizure detection in an animal model of chronic epilepsy. *Journal of Neural Engineering*, 7(3):036001, 2010.

[4] F. Westerhuis, W. Van Schaijk, and G. Van Luijtelaar. Automatic detection of spike-wave discharges in the cortical EEG of rats. In *Measuring Behavior '96, International Workshop on Methods and Techniques in Behavioral Research*, Utrecht, The Netherlands, 1996. Utrecht University.

[5] A. White, P. Willians, D. Ferraro, S. Clark, S. Kadam, and F. Dudek et al. Efficient unsupervised algorithms for the detection of seizures in continuous EEG recordings from rats after brain injury. *Journal of Neuroscience Methods*, 152:255–266, 2006.

[6] S. Haas, M. Frei, and Oso. Strategies for adapting automated seizure detection algorithms. *Medical Engineering & Physics*, 29:895–909, 2007.

[7] L. Waterschoot, S. Dedeurwaerdere, and T. Wyckhuys. Stimulation of the anterodorsal thalamus in genetic absence epilepsy rats from Strasbourg (GAERS). In Epilepsia, editor, *7th European Congress on Epileptology*, volume 47, page 70, Helsinki, Finland, 2006. Blackwell Publishing.

[8] T. Wyckhuys, P. Boon, R. Raedt, B. Van Niewenhuyse, K. Vonck, and W. Wadman. Suppression of hippocampal epileptic seizures in the kainate rat by poisson distributed stimulation. *Epilepsia*, 51:2297–2304, 2010.

[9] P. Boon, K. Vonck, J. De Reuckb, and J. Cammaert. Vagus nerve stimulation for refractory epilepsy. *Seizure*, 10:448–455, 2001.

[10] M. Lukoševičius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[11] L. Danober, C. Deransart, A. Depaulis, M. Vergnes, and Marescaux C. Pathophysiological mechanisms of genetic absence epilepsy in the rat. *Progress in Neurobiology*, 55:27–57, 1998.

[12] S. Baraban, editor. *Animal Models of Epilepsy: Methods and Innovations*, volume 40 of *Neuromethods*. Humana Press, New York, 2009.

[13] I. Osorio, M. Frei, and S. Wilkinson. Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia*, 39:615–627, 1998.

[14] B. Schrauwen, D. Verstraeten, and J. Van Campenhout. An overview of Reservoir Computing: theory, applications and implementations. In *Proceedings of the European Symposium on Artifical Neural Networks (ESANN)*, 2007.

[15] V. Stahl, A. Fisher, and R. Bippus. Quantile based noise estimation for spectral subtraction and Wiener filtering. In *International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, 2000. IEEE.

[16] M. Lukosevicius and H. Jaeger. Reservoir Computing Approaches to Recurrent Neural Network Training. *Computer Science Review*, 3(3):127–149, 2009.

[17] H. Jaeger, M. Lukosevicius, and D. Popovici. Optimization and applications of echo state networks with leaky integrator neurons. *Neural Networks*, 20:335–352, 2007.

[18] R. A. Legenstein and W. Maass. Edge of chaos and prediction of computational performance for neural microcircuit models. *Neural Networks*, pages 323–333, 2007.

[19] H. Jaeger. Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and the "echo state network" approach. Technical Report GMD Report 159, German National Research Center for Information Technology, 2002.

[20] J. Ting, A. D'Souza, and S. Schaal. Automatic outlier detection: A bayesian approach. In *Conference on Robotics and Automation, 2007 IEEE International*, 2007.

[21] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, August 2006.

[22] M. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.