KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

| | |
|---|---|
| **Citation/Reference** | Milosevic M., Van de Vel A., Cuppens K., Bonroy B., Ceulemans B., Lagae L., Vanrumste B., Van Huffel S., ``Feature selection methods for accelerometry-based seizure detection in children'', *Medical and Biological Engineering and Computing*, vol. *, no. *, Apr. 2016, pp. 1-15. |
| **Archived version** | Final publisher's version / pdf |
| **Published version** | http://dx.doi.org/10.1007/s11517-016-1506-9 |
| **Journal homepage** | http://link.springer.com/journal/volumesAndIssues/11517 |
| **IR** | https://lirias.kuleuven.be/handle/123456789/537381 |

*(article begins on next page)*

Med Biol Eng Comput
DOI 10.1007/s11517-016-1506-9

ORIGINAL ARTICLE

# Feature selection methods for accelerometry-based seizure detection in children

**Milica Milošević[1] · Anouk Van de Vel[3] · Kris Cuppens[4] · Bert Bonroy[4] · Berten Ceulemans[3,5] · Lieven Lagae[5,6] · Bart Vanrumste[1,2] · Sabine Van Huffel[1]**

Received: 5 March 2014 / Accepted: 29 March 2016
© International Federation for Medical and Biological Engineering 2016

**Abstract** We investigate the application of feature selection methods and their influence on distinguishing nocturnal motor seizures in epileptic children from normal nocturnal movements using accelerometry signals. We studied two feature selection methods applied one after the other to reduce the complexity and computation costs of least-squares support vector machine (LS-SVM) models. Simultaneous feature selection analyses were performed for each seizure type individually and jointly. Starting from 140 features, a filter method based on mutual information was applied to remove irrelevant and redundant features. The obtained subset was further reduced through a wrapper feature selection strategy using an LS-SVM classifier with both forward search and backward elimination. The discriminative power of each feature subset was evaluated on the test data in terms of the area under the receiver operating characteristic curve, sensitivity, and false detection rate per hour. We showed that, by using only a filter method for feature selection, it was possible to obtain classification results of comparable or slightly reduced performance with respect to the complete feature set. The attained results could facilitate further development of accelerometry-based seizure detection and alarm systems.

**Keywords** Epilepsy · Children · Seizure detection · Accelerometers · Feature selection

## 1 Introduction

Epilepsy is the second most common neurological disorder, occurring in about 1 % of the world population [31]. It manifests itself in recurrent epileptic seizures, which are the result of abnormal excessive and/or synchronous neuronal activity in the brain. Seizures can affect at least one of the following: sensory, motor, and autonomic functions; consciousness; emotional state; memory; and cognition and behaviour [11]. Accordingly, epilepsy has a direct influence on the quality of life of epileptic patients. In addition, there are social and economic implications. A long-term home monitoring system would facilitate better supervision of the patient, improve insights into the effects of the prescribed medication or therapy, and enable an objective measure of seizure frequency.

In the past decade, efforts have been made to build a system for long-term home monitoring of epileptic patients that can offer an alternative to the standard video/electroencephalography (EEG), which can only be performed in specialised hospitals. The main focus has been on detection of convulsions related to epileptic seizures using different

**Electronic supplementary material** The online version of this article (doi:10.1007/s11517-016-1506-9) contains supplementary material, which is available to authorised users.

✉ Sabine Van Huffel
sabine.vanhuffel@esat.kuleuven.be

[1] Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics/iMinds Medical IT, KU Leuven, 3001 Leuven, Belgium

[2] ESAT, Advanced Integrated Sensing (AdvISe), KU Leuven, 2440 Geel, Belgium

[3] Department of Neurology-Paediatric Neurology, University Hospital Antwerp, 2650 Edegem, Belgium

[4] MOBILAB, Thomas More Kempen, 2440 Geel, Belgium

[5] Rehabilitation Center for Children and Youth Pulderbos, 2242 Pulderbos, Belgium

[6] Department of Child Neurology, University Hospitals KU Leuven, 3000 Leuven, Belgium

Published online: 22 April 2016

single modalities or their combinations: video, accelerometry (ACM), gyroscopes, magnetometers, and surface electromyography (sEMG). Many research groups have been trying to develop a monitoring and/or alarm system based on some of the aforementioned sensors [1, 4–8, 16, 17, 20, 21, 24, 26]. The process of development of these systems is rather slow for many reasons. The collection of data is a tedious process with slow progress owing to the small number of recorded seizures. Data annotation is a laborious and expensive task. Moreover, the detection system should be comfortable for the patients, easy to use for caregivers, and work with a perfect detection rate. However, these objectives are difficult to meet jointly, mostly due to the lack of data and huge inter- and intra-patient variability regarding the number, type, and clinical manifestation of the motor seizures.

Many systems have been proposed, but a perfect solution has yet to emerge. Research groups have employed different motion sensors and their combinations to detect specific seizure types. Most of the algorithms have been based on the use of classification algorithms, such as support vector machines. Different features were fed to a chosen classifier, and performance metrics were extracted [1, 6, 20, 24]. The number and type of features varied significantly from study to study. Various criteria were used to choose the features. Features were selected (1) by mimicking the expert's knowledge while screening and annotating the data, (2) from visualisation of the feature space and separation of the different classes, or (3) by borrowing features from similar studies. All these approaches have their disadvantages. Experts rely on video/EEG data to determine seizure occurrence and then video and motion sensor (in most cases ACM) data to determine the type of motor seizures. By examining only motion data, experts would not be able to correctly identify and characterise seizures. Visualisation of the feature space becomes demanding as the number of features grows. Moreover, the chosen features can be redundant and the used sets are not optimised for the employed classification method. So far, only a few studies have been performed to identify the most informative features for this application.

A recent study [18] investigated the performance of 65 features in terms of EEG-based seizure detection performance and computational complexity. The goal of this work was to perform a similar analysis on the largest ACM database with 1498 labelled nocturnal motor seizures of epileptic children. Instead of examining each feature individually, as in [18], features were combined in this paper. Additionally, to overcome the aforementioned feature selection disadvantages, feature selection approaches—namely filter and a combination of filter and wrapper selection methods—were applied. The main goal was to determine a significantly smaller subset of features that preserved the performance metrics at the same level as when all features were used. In total, 140 features were taken from publications that considered detection of motor seizures, EEG-based seizure detection, and ACM-based activity detection. The feature selection analysis was performed for each motor seizure type separately and then jointly. As part of a long-term research collaboration [7, 9], this study should provide new insights into ACM-based motor seizure detection, thereby facilitating the further design of methods for home monitoring of epileptic children.

## 2 Methods

### 2.1 Data collection and partition

Simultaneous EEG, video and audio, electrocardiography, sEMG, and ACM data were recorded. Only ACM signals were processed in this study. Four triaxial accelerometers were attached to the wrists and ankles with wristbands generating 12 ACM channels. The measured range of acceleration was $\pm 3G$ (where $G = 9.81$ m/s$^2$), and the sampling rate was 250 Hz. The whole system was installed in the Rehabilitation Center for Children and Youth in Pulderbos, Belgium, where all measurements were taken during the night, typically from 8 p.m. to 8 a.m. the following day.

In this study, 51 children and adolescents suffering from refractory epilepsy participated. Approval by the Medical Ethical Commission of the Antwerp University Hospital, Belgium, and signed informed consent forms from all parents were obtained prior to inclusion in the study, which was performed in accordance with the 1964 Declaration of Helsinki. Ultimately, data for 143 nights were collected. All epileptic seizures were annotated by a trained EEG specialist using the gold standard video/EEG monitoring and International League Against Epilepsy (ILAE) motor seizure classification terminology [2, 19]. In case of doubt or disagreement regarding the occurrence or type of seizure, two EEG experts and two neurologists agreed on the label in consensus. After excluding the patients without seizures and with frontal lobe seizures with hyperkinetic movements (analysed in [7]), 1498 labelled motor seizures of 25 patients were kept (Table 1). In addition, 18 of 83 recordings were excluded from the analysis because they contained nonfunctional (or broken) channels. A broken channel, in general, was the consequence of bad connections of the ACM sensor and cable and/or the cable and signal collection point. In this case, this channel only recorded the intrinsic noise of the sensor. To exclude the influence of broken channels, these recordings were removed from the analysis.

The most frequent seizure types within the used database are as follows:

**Table 1** Database overview: in addition, 21 patients (31 nights) did not have seizures and five patients had only frontal lobe seizures (56) during 29 nights

| Study code | Complete recordings | | Incomplete recordings | |
|---|---|---|---|---|
| | Motor seizures[a] | # of nights | Motor seizures[a] | # of nights |
| 2 | 55S, 1S series | 2 | | |
| 3 | 26M, 9S, 1S series, 1T, 2C, 2F, 2U | 18 | 3M, 1M series, 3T | 2 |
| 7 | 5S, 1S series | 1 | | |
| 9 | 7M, 15S, 2T | 3 | 4M, 3S | 1 |
| 10 | 2M, 1S | 1 | | |
| 11 | 9M, 3TC, 3V | 1 | 40M, 106S, 51T, 6TC, 3V | 1 |
| 13 | | | 5M, 2T | 1 |
| 17 | | | 1M, 8S, 20T, 2U | 1 |
| 18 | 1U | 2 | | |
| 19 | 2M, 2F | 1 | | |
| 21 | 3S, 17T | 5 | | |
| 22 | 1S, 2S series, 1C | 2 | 1S series | 1 |
| 24 | 10M, 4S, 1S series | 3 | 4M, 2C, 1TC | 2 |
| 25 | 1M, 8U | 2 | | |
| 26 | 9M, 1S, 1C | 4 | | |
| 28 | 2M, 3S | 2 | 1S, 1U | 1 |
| 30 | 1S, 6C | 3 | 1M, 1M series | 2 |
| 32 | 1M, 5S, 3T | 4 | 7M, 4S | 2 |
| 35 | 11M, 10S, 2S series, 3T, 1TC, 1U | 2 | 11M, 3T | 1 |
| 39 | 1M, 1S, 24T | 2 | 1M, 1T | 1 |
| 40 | 2M, 10S, 5T, 13C | 2 | 2M, 1S, 1T, 2U | 2 |
| 46 | 23M | 1 | | |
| 47 | 1C | 1 | | |
| 48 | 9S, 1TC | 2 | | |
| 51 | 49C | 1 | | |
| Total: 25 patients | 105M, 133S, 8S series, 55T, 73C, 5TC, 3V, 4F, 12U | 65 | 78M, 2M series, 123S, 1S series, 81T, 2C, 7TC, 3V, 5U | 18 |

[a] *M* myoclonic, *S* epileptic spasm, *T* tonic, *C* clonic, *TC* tonic–clonic, *V* versive, *F* frontal lobe, *U* unclassified, *M/S* series myoclonic/spasm series (more than ten consecutive seizures in 1 min)

- Myoclonic seizures are sudden short (<0.5 s) twitching movements, mostly involving the shoulders and proximal limbs.
- Clonic seizures are a series of myoclonic contractions of agonist and antagonist muscles that regularly occur from 0.2 to 5 times a second. In general, a rhythmic pattern can be observed in the ACM data.
- Tonic seizures are tensions resulting in a change in posture. Usually, blockwise ACM signals are observed with low noise representing the tremor.
- Tonic–clonic seizures are a combination of tonic and clonic seizures.
- Epileptic spasms are seizures that mainly involve axial muscles, leading to flexion or extension of the neck (and legs) and abduction of both arms. The waveforms observed in the ACM signals can significantly differ depending on which limb(s) are involved and their position before the seizure starts.

- The characteristics of other seizure types are explained in the Supplementary Materials. As can be seen in Table 1, the number of seizures and types of seizures per patient vary significantly. Short seizures, such as myoclonic seizures and epileptic spasms, occur frequently, whereas longer, more severe seizures, such as tonic–clonic ones, appear only in a few patients and are usually rare events. However, these longer, more intense seizures are clinically important, and consequently, they are the main candidates for the alarm system.

Figure 1 illustrates the analysis applied to all seizure types individually (clonic, tonic, and myoclonic seizures spasms,) and then jointly (called motor seizures). The

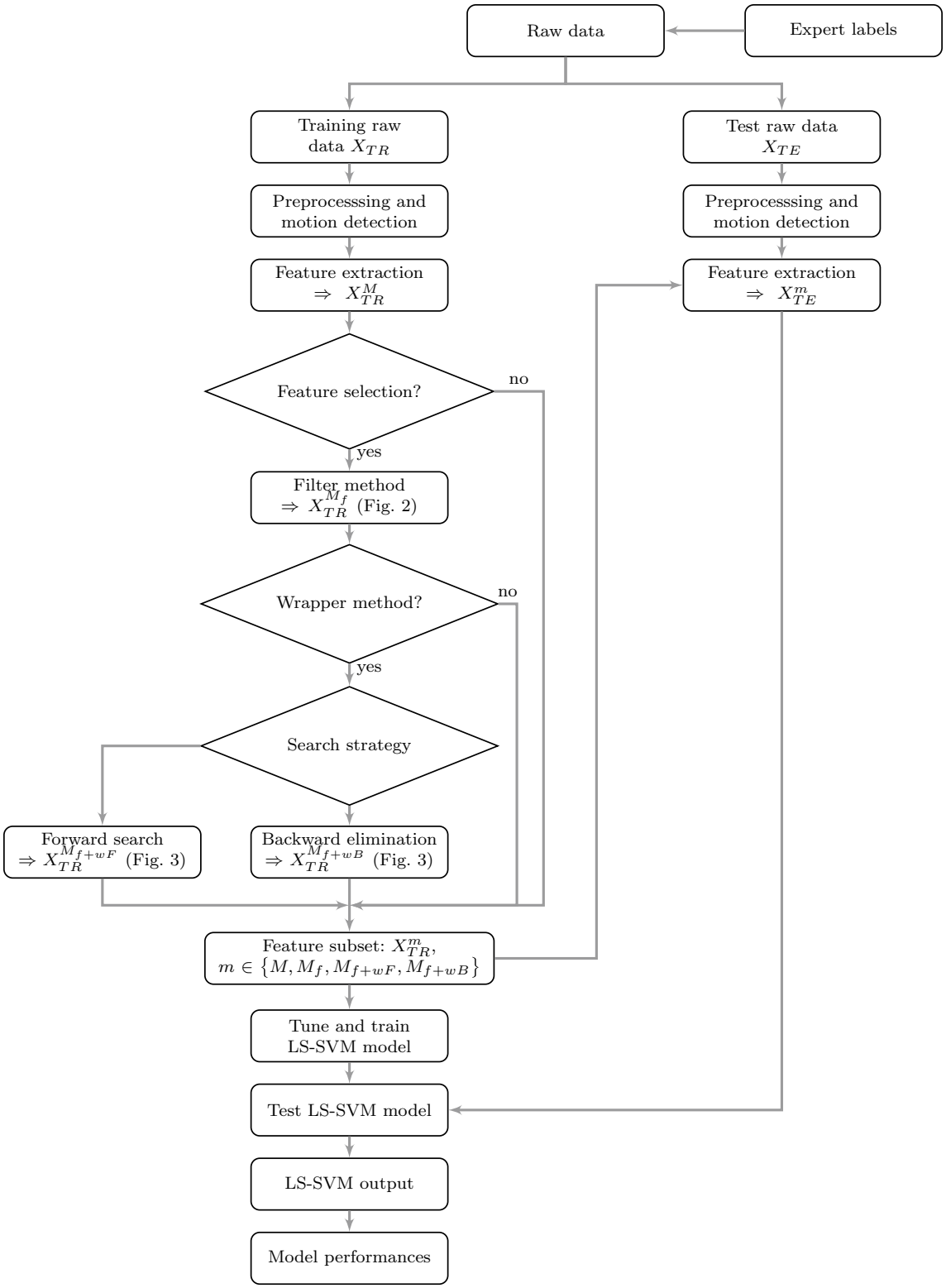



**Fig. 1** Methodology scheme from data pre-processing to feature selection to classification to evaluation. There are four possibilities for feature selection: no feature selection, only filter feature selection, and wrapper feature selection methods with both forward search and backward elimination strategies after applying the filter method. The corresponding feature matrices are $X^{M}$, ($M = 140$), $X^{M_{\mathrm{f}}}$, $X^{M_{\mathrm{f+wF}}}$, and $X^{M_{\mathrm{f+wB}}}$, respectively. $X_{\mathrm{TR}}$ data were used for feature selection, tuning, and training of the LS-SVM classification model, which was tested on test data $X_{\mathrm{TE}}$. Although all data in $X_{\mathrm{TE}}$ were used for testing, data for feature selection and training were randomly selected as much smaller data subsets of $X_{\mathrm{TR}}$. In corresponding sections, the number of data points used is indicated

**Table 2** Number of seizures in training and test sets per seizure type

| Seizure type | # of seizures in training set | # of seizures in test set |
|---|---|---|
| Myoclonic seizures | 86 | 19 |
| Epileptic spasms | 79 | 420 |
| Tonic seizures | 24 | 31 |
| Clonic seizures[a] | 19 | 54C + 5TC |
| Motor seizures | 316 | 551 |

[a] *C* clonic, *TC* tonic–clonic

number of tonic–clonic seizures is significantly lower than the number of other seizure types; therefore, the feature selection and classification was not performed separately for this type of seizure. Nevertheless, these seizures (and corresponding recordings) were added as test data to the clonic seizure detection algorithms, as it was expected that the algorithm would be able to detect the clonic part of a tonic–clonic seizure. In the case of clonic and tonic seizures, data recorded till the end of the fifth seizure of each patient were used for both feature selection and building the least-squares support vector machine (LS-SVM) classifier. The remaining data were used to test a seizure detection algorithm with a reduced number of features on the independent data. Owing to the higher number of seizures for epileptic spasms and myoclonic and motor seizures, the first 15 seizures were used for training and the rest for the testing. These divisions were chosen to select enough data for reliable feature selection and training and still retain some data for testing. As a consequence, the number of test seizures was always greater than the number of training seizures, except for myoclonic seizures (see Table 2). Once the data were divided into the training $X_{\mathrm{TR}}$ and test $X_{\mathrm{TE}}$ sets, the analysis was performed as indicated in Fig. 1. All data in $X_{\mathrm{TE}}$ were used for testing, but data for feature selection and training were randomly selected from $X_{\mathrm{TR}}$ as much smaller data subsets. Each block in Fig. 1 will be explained in the corresponding subsections.

Before explaining the feature selection procedure, we briefly introduce our notation. From each movement epoch of the considered dataset, a feature value is calculated from each ACM channel and stored in a single-feature row vector of size $1 \times 12$. For $N$ epochs, a matrix $x_j$ of size $N \times 12$ for each single feature $j$ is built, $j = 1, \ldots, M$. Hence, the set of all $M$ features can be written as a feature set $\{x_1, \ldots, x_M\}$. Subscripts *TR* and *TE* are added to denote, respectively, the training set and the test set. Superscript $M$ denotes the number of 12-dimensional features in $X$. Additional subscripts are added to denote specific subsets. In particular, $M_{\mathrm{f}}$, $M_{\mathrm{w}}$, and $M_{\mathrm{f+w}}$ denote, respectively, the subset of features kept after applying a filter method ($f$), a wrapper method ($w$), or a combination of both ($f + w$). Subscripts $wF$ and $wB$, respectively, specify that the wrapper is combined with a forward search or backward elimination.

### 2.2 Pre-processing

Pre-processing was performed to remove the artefacts and to reduce the data for further processing. First, a Chebyshev type II low-pass filter with a cut-off frequency of 47 Hz was applied to remove the power-line frequency of 50 Hz and its harmonics. Next, after downsampling the ACM signals by factor of 2, a Chebyshev type II high-pass filter with a 0.2-Hz cut-off frequency was applied to remove the gravitational component of acceleration. The pre-processed data were divided into two-s epochs with 75 % overlap. Finally, whenever the standard deviation in each channel was lower than 10 m$G$, the corresponding epoch was removed from further analysis. The value of 10 m$G$ was experimentally determined by measuring the smallest movements of the limbs [8]. This "movement detection" resulted in a $79.96 \pm 15.97$ % reduction in the initial number of epochs per recording.

### 2.3 Feature extraction

A systematic literature review was performed to gather possible features for detection of motor seizures. Papers related to ACM-based seizure detection algorithms were taken into consideration [1, 6, 7, 20, 21, 24]. In addition, features used in EEG-based seizure detection [18, 29] and ACM-based activity detection algorithms [14, 25] were added. All these features, provided they were fully described and were applicable to ACM measurements, were included in the initial feature subset. In total, $M = 140$ features from each ACM channel were included: 33 in the time domain, 33 in the frequency domain, 10 derived from the continuous wavelet transform, 62 from the packet wavelet transform (PWT), and two from a recurrence plot analysis. The full list and definitions of features are given in the Supplementary Materials.

### 2.4 Feature selection

Identification of the most characteristic features, i.e. feature selection, is an important step in reducing classification error and computational costs. The process of feature selection consists of the removal of irrelevant and redundant features and/or optimisation of the feature subset for the particular classification algorithm.

In general, there are three groups of feature selection methods: filter, wrapper, and embedded methods. Filter methods include methods performed in the pre-processing phase as an initial analysis of the feature's relevance. Features are ranked according to a predefined criterion, such as

mutual information, entropy, and cross-correlation. Since filter methods ignore the effects of the selected feature on the classification algorithm, they are quite fast. Wrapper methods regard the classification algorithm as a black box [15]. The classification performance is used as a criterion for the feature subset selection. Thus, the obtained subset is optimised for the particular classification algorithm. An exhaustive search can be performed if the number of features is low. However, with an increase in the number of features, a full search quickly becomes computationally unfeasible. Therefore, heuristic search strategies are employed using forward search or backward elimination. In forward selection, features are incorporated one by one into larger subsets, until the performance metrics start to saturate. In contrast, a backward elimination strategy starts with all features and progressively eliminates the least significant feature until the performance metrics start to degrade. The main disadvantage of this strategy is that the computational cost significantly increases with increased number of features. Embedded methods are performed in the training phase, and these methods are usually specific to the given classification algorithm and computationally demanding. They are beyond the scope of this paper.

In this paper, we focused on two feature selection methods, a filter method based on mutual information and a wrapper method based on an LS-SVM classifier. These methods were combined in the two-stage process. First, the filter method was applied to quickly remove a high number of irrelevant and redundant features, and then, the significantly smaller subset was even more reduced by applying the wrapper method with both forward selection and backward elimination. Both feature selection methods were extended to work with multivariate features, since all 12 ACM channel values were taken as one feature and their relevance was evaluated as a whole. Since the number of seizures and their types significantly vary among the different patients, a bootstrap technique was applied [12]; each feature selection algorithm was run 50 times on different randomly chosen subsets from $X_{\mathrm{TR}}$ (300 seizure epochs and 1500 nonseizure epochs) and the importance of each feature was evaluated using the number of times it appeared in these 50 optimal feature subsets. The features that appeared in more than 75 % of the number of runs (i.e. in more than 35 of the 50 runs) were kept for further analysis.

#### 2.4.1 *Maximal relevance minimal redundancy filter feature selection method*

Ding and Peng [10] demonstrated that the maximal relevance minimal redundancy (mRMR) feature selection method is especially useful for large-scale feature selection problems [10]. The mRMR method is presented in [27]. Moreover, prior application of mRMR can enhance the wrapper feature selection method, thereby achieving better classification performance [23].

Starting from a complete feature set $X^M = \{x_1, \ldots, x_M\}$, where $x_i, i = 1, \ldots, M$ are feature matrices of dimension $N \times 12$, and denoting the newly generated set of chosen features by $S$, the mRMR algorithm maximises the mutual information between the feature $x_j \in X^M \backslash S$ and the target class $c$ (extending a one-dimensional label column to a 12-column matrix) and ascertains that the mutual information between the new feature $x_j$ and the ones that are already chosen, $x_i \in S$, is minimal. In practice, an incremental search method is employed that optimises the following condition:

$$x_j = arg \max_{x \in X^M \backslash S} \left( I(x; c) - \frac{1}{j-1} \sum_{x_i \in S} I(x; x_i) \right), \tag{1}$$

where $I$ is the mutual information defined as

$$I(x; y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \mathrm{d}x\mathrm{d}y. \tag{2}$$

with $p(x)$, $p(y)$, and $p(x,y)$ as probability density functions. These are approximated by diffusion-based kernel density estimation, an adaptive kernel density estimator that has proven to outperform existing methods in terms of accuracy and reliability [3]. Figure 2 illustrates this selection process, which was repeated 50 times on 300 randomly selected seizure epochs and 1500 randomly selected nonseizure epochs chosen from rows from $X^M_{\mathrm{TR}}$, represented by $X^M_n$. Features appearing more than 35 times were kept in the final mRMR feature subset $X^{M_{\mathrm{f}}}_{\mathrm{TR}}$.

#### 2.4.2 *LS-SVM wrapper feature selection method*

A wrapper feature selection method with forward and backward search strategies described in Sect. 2.4 was implemented in MATLAB (MathWorks Inc.) using the LS-SVM classifier as a black box (see Sect. 2.5). The $F_1$-measure (Eq. 3) was used as an optimisation and feature subset evaluation metric. Figure 3 illustrates the feature selection process in more detail. Feature selection was repeated 50 times on 50 randomly chosen subsets of training data $X^{M_{\mathrm{f}}}$ from $X^{M_{\mathrm{f}}}_{\mathrm{TR}}$ and features appearing more than 35 times composed the final feature subset $X^{M_{\mathrm{f+w}}}_{\mathrm{TR}}$.

### 2.5 LS-SVMs

Support vector machines (SVMs) are modern machine learning methods that use kernels to construct a linear classification hyperplane in higher dimensional space for the separation of two classes of data points [30]. In this way, even really complex nonlinear classification tasks in the data point space become simple and linear in the higher

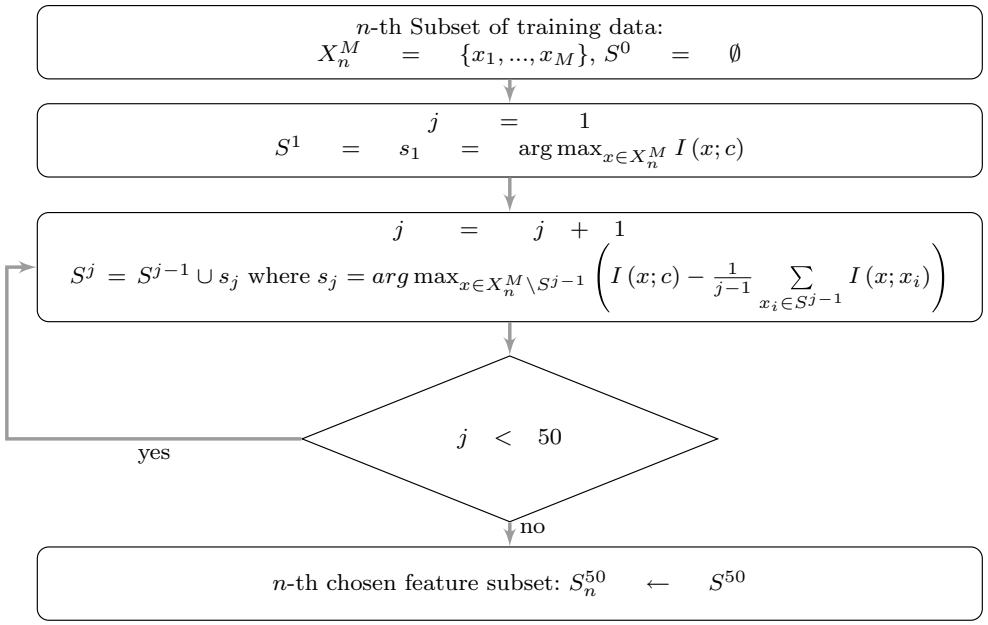



**Fig. 2** Filter feature selection algorithm based on mutual information starts from a complete feature set $X_n^M$ ($M = 140$) and an empty set of chosen features $S^0$. $X_n^M$ consists of the features as calculated epoch-wise from the $n$th subset of randomly chosen 300 seizure epochs and 1500 nonseizure epochs from $X_{\mathrm{TR}}^M$ data. In iteration $j < 50$, chosen feature $s_j$ is added to the feature subset from the previous iteration $S^{j-1}$, by maximising the difference between the mutual information $I$ between feature $s_j$ and class target $c$ and the mean of mutual information values between feature $s_j$ and already selected features $S^{j-1}$. The described process was repeated 50 times using different randomly chosen subsets $X_n^M$ from $X_{\mathrm{TR}}^M$, $n = 1, \ldots, 50$, and features that appeared more than 35 times within $S_n^{50}$, $n = 1, \ldots, 50$, were included in $X_{\mathrm{TR}}^{M_{\mathrm{f}}}$, which was used in further analysis

dimensional space. The mapping is performed with a suitable choice of a kernel function. Least-squares support vector machines are a modification of SVM methodology, entailing the introduction of a least-squares loss function and error variables [28]. Now, instead of solving a quadratic programming problem as required in classical SVM, LS-SVM reduces the problem to solving a set of linear equations. Therefore, this modification leads to a significant reduction in complexity and computation costs. The mathematical formulation of SVM and LS-SVM methods can be found in the Supplementary Materials, and for more details, the reader is referred to the original publications [28, 30].

LS-SVMlab [22] was used to perform all LS-SVM calculations in MATLAB. The regularisation parameter and the width of the radial basis function (RBF) kernel were obtained using fivefold cross-validation with a grid search technique. In the LS-SVMlab toolbox, by default, the optimisation is performed with minimisation of the misclassification error. However, in the case of seizure detection, the data are highly unbalanced; seizures are rare events, which should be distinguished from all other human motion and in-bed movements. Optimising the LS-SVM boundary by minimising the full classification error can result in a negligible classification error of nonseizures and high misclassification of seizures. To avoid these phenomena, the LS-SVM regularisation parameters were estimated by maximising the $F_1$-measure parameter [24]:

$$F_1\text{-measure} = 2\frac{\text{sen} \cdot \text{ppv}}{\text{sen} + \text{ppv}}, \tag{3}$$

which is the harmonic mean of the sensitivity (sen) and positive predictive value (ppv) of epoch detection. Once the smaller feature subsets were obtained, the LS-SVM model was retrained on the full training set $X_{\mathrm{TR}}$ with the obtained parameters (randomly chosen 300 seizure epochs and 1500 nonseizure epochs) and tested on the new data $X_{\mathrm{TE}}$ (see Fig. 1).

### 2.6 Performance metrics

The performance metrics were based on a confusion matrix, which reports the numbers of true positives, TP, true negatives, TN, false positives, FP, and false negatives, FN, between true and predicted labels. To evaluate the chosen features, the following four metrics were calculated: (1) the area under the receiver operating characteristic (ROC) curve for epoch detection, which gives better insight into the distribution of the decisions; (2) seizure detection sensitivity when at least one epoch within the seizure was detected; (3) false detection rate per hour; and (4) delay between the first positively detected epoch and the clinical

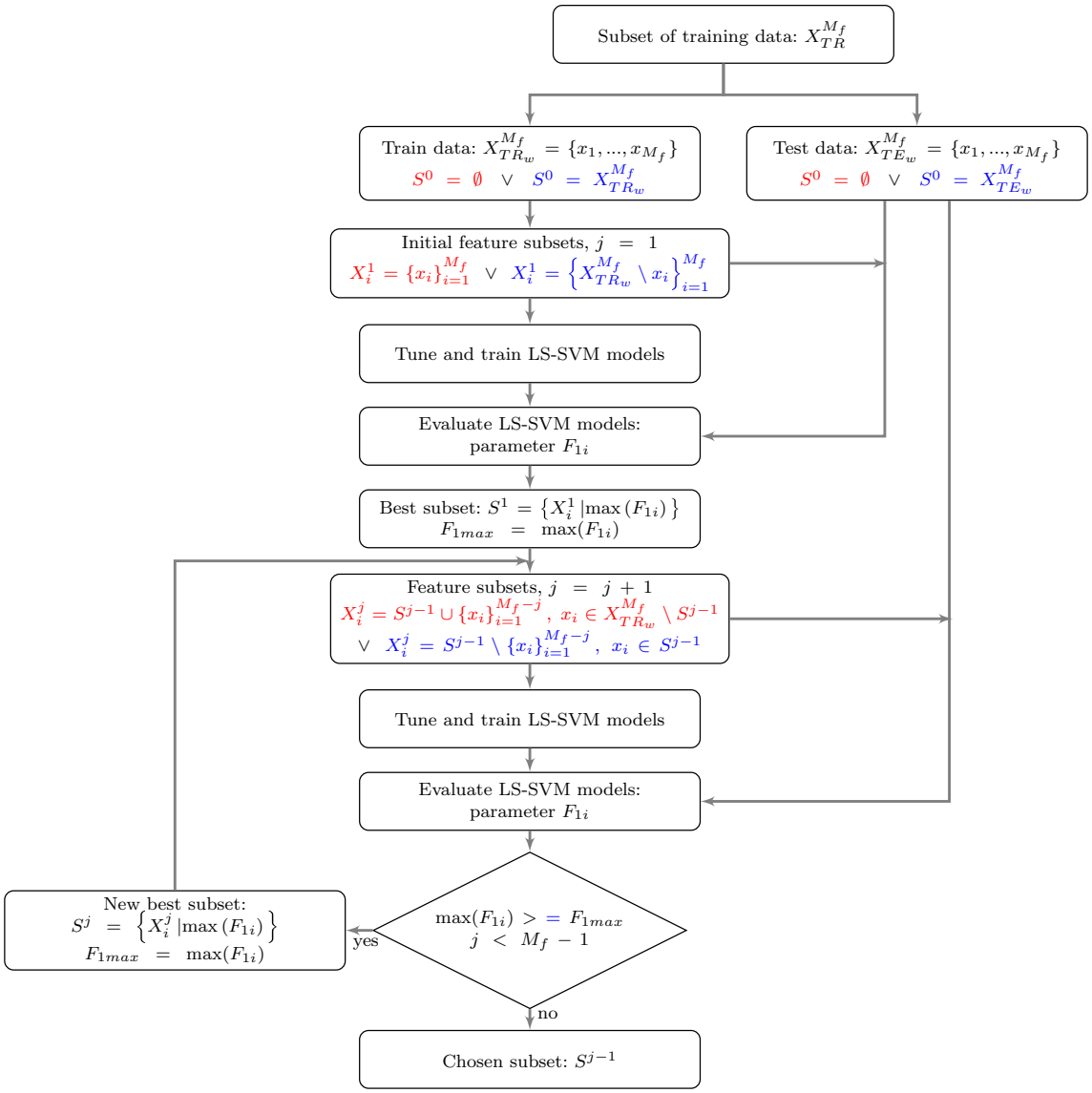



**Fig. 3** Wrapper feature selection algorithm with LS-SVM classifier starts from training data with $M_f$ features obtained by the mRMR method. $X^{M_f}$ represents the feature set extracted from one subset of randomly chosen 300 seizure epochs and 1500 nonseizure epochs from $X_{TR}^{M_f}$ data. $X^{M_f}$ data are divided into the LS-SVM training $X_{TR_w}^{M_f}$ and test $X_{TE_w}^{M_f}$ sets (in the ratio 2:1). $X_i^j$ are $i$ feature sets in iteration $j$ that are built by adding or removing feature matrices of dimension $N \times 12$ to or from previous feature set $S^{j-1}$ (with the *red colour* corresponding to forward search and the *blue* to backward elimination). LS-SVM models are built for each $X_i^j$ feature set, and $F_{1i}$ parameters are obtained (see Eq. 3). The set with the greatest $F_{1i}$ is selected, and the procedure is continued until $F_{1i}$ of the chosen set is greater than or equal to $F_{1max}$ from the previous iteration $j - 1$. The described process was repeated 50 times using different randomly chosen training subsets $X_{TR_{w,n}}^{M_f}$ from $X_{TR}^{M_f}$, $n = 1, \ldots, 50$ (with the rest of the data, $X_{TE_{w,n}}^{M_f}$, being used for testing), and features that appeared more than 35 times within $S_n^{j-1}$ were included in $X_{TR}^{M_{f+w}}$ and used in further analysis

onset of the seizure. The value of the false detection rate per hour is calculated as the number of successively grouped false-positive epochs. In addition, sensitivity (sen = TP/(TP + FN)), specificity (spec = TN/(TN + FP)), and positive predictive value (ppv = TP/(TP + FP)) for epoch detection were calculated for the $F_1$-measure-optimised LS-SVM hyperplane. While the epoch sensitivity, specificity, and positive predictive value were calculated for the specific decision threshold (or the bias term in the case of the LS-SVM classifier) obtained within the training process, the ROC curve is a function of sensitivity versus 1-specificity obtained by varying the decision threshold.

All metrics obtained with the complete feature set, the feature subset after the filter method, and the feature subsets after filter and wrapper methods with both forward search and backward elimination were statistically

compared using the Friedman paired test and Bonferroni post hoc test. The Friedman test is a nonparametric paired test working with ranks. The assumption of normality is not acceptable. When multiple significance tests are performed, the chance of finding a significant difference just by chance increases. The Bonferroni test is one of several methods of ensuring that the chance of finding a significant difference in any comparison (under a null model) is maintained at the alpha level of the test. In addition, the purpose of the Bonferroni test was to determine which groups in the sample differ. Although a Friedman test can tell whether groups in the sample differ, it cannot tell which groups differ.

## 3 Results

The features obtained with the filter method using mutual information are listed in Table 3. The number of chosen features for different seizure types ranged from 15 to 28. Some features, such as complexity, laminarity, and some PWT features, were relevant for multiple seizure types, whereas others were seizure-type-specific. Continuous wavelet transform (CWT) features used in many ACM-based seizures were completely excluded [1, 20], in contrast to PWT features, which contain similar information [6].

Next, a wrapper feature selection method incorporated with an LS-SVM classifier was performed on the feature subset obtained by the filter method. Both strategies, forward search and backward elimination, were applied. Table 4 lists the chosen features. In general, more features were selected by a backward elimination strategy, since it takes into consideration the relation between the different features. Most of the features were time-domain features, which in general have a low computational complexity (time needed to calculate the feature) [18]. Motor seizures that contain twitching contractions (myoclonic and clonic seizures) can be detected with simple statistical features, mobility, and complexity. For the detection of tonic seizures, two features based on zero-crossing rates were the most optimal features.

The area under the ROC curve for epoch detection (AUC), sensitivity when at least one epoch per seizure was detected, latency of the first detected epoch compared to the clinical onset of the seizure, and false detection rate per hour (FDR/h) are presented in Fig. 4. This figure displays global trends when different feature sets are used; numerical values (median, 25th, and 75th percentile) sorted in the tables can be found in the Supplementary Materials. In addition, other epoch-based metrics are presented in the Supplementary Materials. Owing to the small number of test data, reliable statistical comparisons were not possible for myoclonic and tonic seizure types.

The AUC on the epoch detection level gives the overall performance of the classifier. The AUC for (tonic–)clonic seizures did not significantly degrade with the reduction in the number of features; however, the median values dropped from 72.17 to 68.05 %. In contrast, when all motor seizures are combined, the AUC significantly decreased with feature reduction ($p < 0.001$), which is probably due to the increased diversity. The sensitivity on epoch detection level varied significantly from patient to patient. In case of (tonic-)clonic seizures, sensitivity values cover the range from zero to 100 %, since many patients had only one test seizure that was detected or missed. The median latency per patient was quite low for detected tonic, myoclonic, and spasm seizures, whereas for clonic and tonic–clonic seizures, the latency increased up to 25 s. The false detection rate per hour was quite high for most of the patients, and in the case of (tonic–)clonic and motor seizures, it increased with feature reduction ($p < 0.01$ and $p < 0.001$, respectively).

## 4 Discussion

In this work, two feature selection algorithms were applied in cascade: a filter method and a wrapper method. Whereas the mRMR filter method gave the most relevant nonredundant feature subset, the LS-SVM wrapper gave a suboptimal subset of features optimised for the LS-SVM classifier. With the first method, the number of features was significantly reduced, from 140 to 15–28 features, which enabled the application of a wrapper method.

In many publications, the three axes of an ACM sensor are combined into an acceleration magnitude, which was further used for feature extraction [1, 6–8, 24]. In that case, one can apply feature selection methods to reduce both the number of features and the number of sensors. This approach was tested, and it resulted in inconsistency over the features and sensor locations. We opted to keep the original information available within ACM channels for distinguishing motor seizures from other movements. In addition, considering 12 values calculated from all ACM channels as one feature could be the input for tensor learning algorithms that can take into account the spatial relations between the channels [13]. A separate study will be performed to assess whether the incorporation of the structural information from multichannel ACM recordings improves the detection of epileptic convulsions.

Similar studies using ACM (in combination with other modalities) have never used 140 features together. In this study, this number was significantly reduced after applying feature selection. In previous publications, mostly wavelet-decomposition-derived features, among which PWT yielded the most relevant ones, were used alone or

**Table 3** mRMR filter method: selected features per seizure type

| Domain | Feature names[a] | Myoclonic seizures | Epileptic spasms | Tonic seizures | Clonic seizures | Motor seizures |
|---|---|---|---|---|---|---|
| Time | 25th percentile value | | | × | × | |
| | 50th percentile value | | | | × | |
| | 75th percentile value | | | × | × | |
| | Range of $x$ | × | | | × | |
| | Range of $x_{fst}$ | × | | | | |
| | Entropy | | × | | | |
| | Variance of $x''$ | | | | × | |
| | Skewness | × | × | | | |
| | Kurtosis | | | × | × | × |
| | Zero-crossing rate of $x$ | | × | × | | |
| | Zero-crossing rate of $x''$ | | | × | | × |
| | Mobility | × | × | | × | |
| | Complexity | × | × | × | × | × |
| Frequency | Dominant frequency | | | | × | × |
| | First peak in spectrum | | × | | | |
| | Third peak in spectrum | × | | | | |
| | Fourth peak in spectrum | × | × | | × | × |
| | Spectral centroid | | | | | × |
| | Spectral edge frequency, 10 % | | | × | × | |
| | Spectral edge frequency, 85 % | | × | | | × |
| | Power in 12- to 14-Hz band | × | | | | |
| PWT1[b] | AAD3 | | | × | × | |
| | DAAA4 | | × | × | | × |
| | DDDA4 | | | × | | |
| | AAAD4 | | | | × | |
| | DAAD4 | | | × | × | × |
| | DDAD4 | | × | × | × | × |
| | DDDD4 | | | | × | × |
| PWT2[b] | AD2 | × | | × | | × |
| | DAA3 | × | × | × | | × |
| | ADA3 | | | × | × | |
| | DDA3 | | | × | | |
| | AAD3 | × | × | × | × | |
| | DAD3 | × | × | × | × | |
| | DDD3 | × | × | × | × | |
| | DAAA4 | × | | × | × | × |
| | ADAA4 | × | | × | | |
| | DDAA4 | | | × | | |
| | AADA4 | | | × | × | |
| | DADA4 | × | | × | × | |
| | DDDA4 | | | × | × | |
| | AAAD4 | | | × | × | |
| | DAAD4 | | | × | × | |
| | DDAD4 | | | | × | |
| CRP | Entropy | × | × | | | |
| | Laminancy | × | × | × | × | × |
| | Number of features ($M_f$) | 18 | 16 | 28 | 28 | 15 |

[a] See the definition of individual features in the Supplementary Materials.

[b] PWT1 features are based on the sum of the absolute PWT coefficients, whereas PWT2 features are based on the energy of these coefficients; PWT features are extracted from the packet wavelet transform decomposition of the input signal until the fourth level

**Table 4** mRMR filter and LS-SVM wrapper method: selected features per seizure type and search strategy

| Domain | Feature names[a]/strategy[b] | Myoclonic seizures | | Epileptic spasms | | Tonic seizures | | Clonic seizures | | Motor seizures | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | B | F | B | F | B | F | B | F | B |
| Time | Range of $x$ | | | | | | | × | × | | |
| | Entropy | | × | × | × | | | | | | |
| | Skewness | | × | | | | | | | | |
| | Zero-crossing rate of $x$ | | | | × | × | × | | | | |
| | Zero-crossing rate of $x''$ | | | | | × | × | | | | × |
| | Mobility | × | × | × | × | | | × | × | | |
| | Complexity | | | | | | | | × | | × |
| Frequency | Spectral edge frequency, 10 % | | | | | | | | × | | |
| | Spectral edge frequency, 85 % | | | | | | | | | × | × |
| PWT1 | DAAA4 | | | | | | | | | | × |
| CRP | entropy | | | | × | | | | | | |
| | Number of features ($M_{f+w}$) | 1 | 3 | 2 | 4 | 2 | 2 | 2 | 4 | 1 | 4 |

[a] See the definition of individual features in the Supplementary Materials.

[b] Wrapper search strategy: F for forward search and B for backward elimination

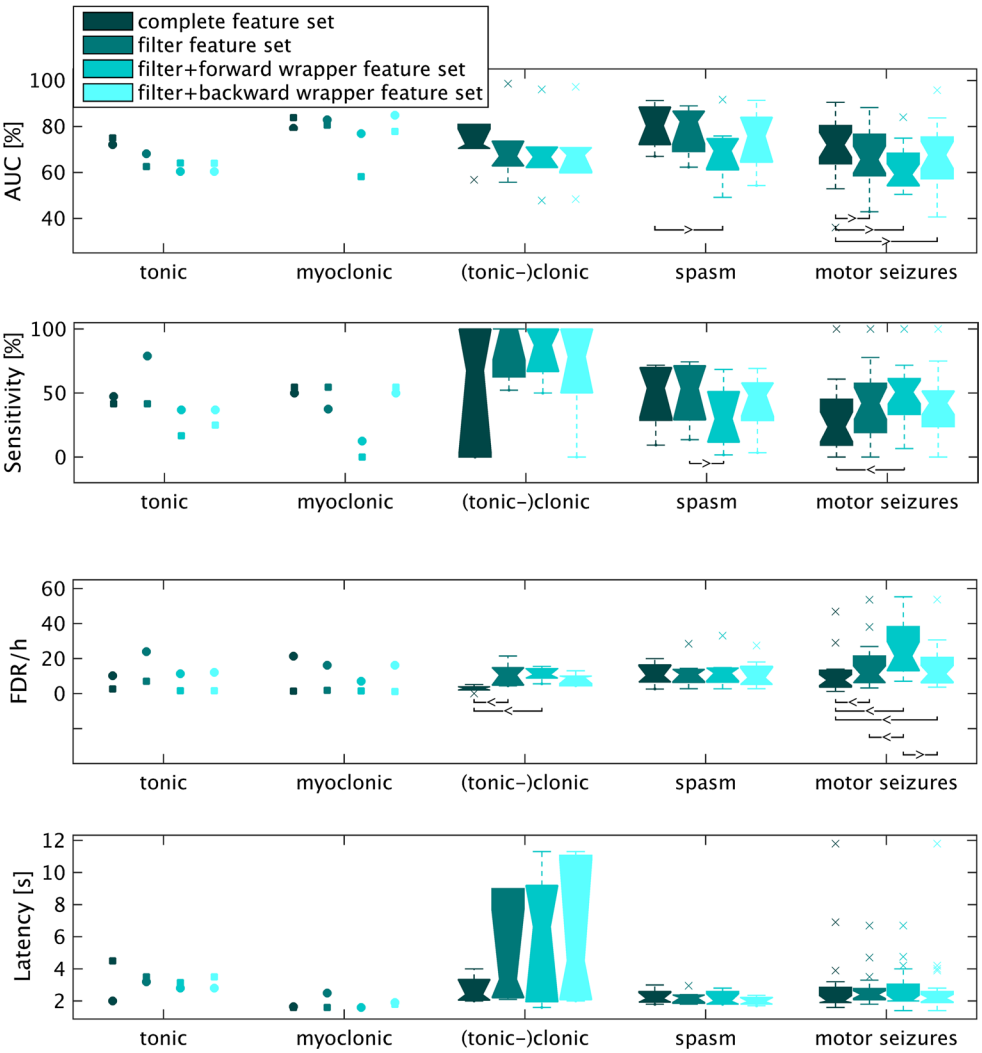



**Fig. 4** Seizure detection performance: area under the ROC curve (AUC), sensitivity (at least one epoch detected), false-positive rate per hour (FDR/h), and median latency for all test datasets when a complete feature set is used, after the filter method, and after the filter and wrapper methods (both for forward search and backward elimination). The significant differences between the groups are annotated with > or < depending on the relation between the groups. In case of myoclonic and tonic seizures, two test patients are denoted by a *square* and a *circle*

combined with others [1, 6, 20, 21]. However, it was unfeasible to test the importance of all of them from [6]. We could only calculate the ones within a two-s epoch until the fourth level of decomposition (instead of the sixth level as done in [6]) using Daubechies five wavelets.

Using cascade feature selection, mostly time-domain features have been obtained, facilitating faster algorithms. This methodology confirmed the importance of the zero-crossing rate (ZCR) as a feature for distinguishing nocturnal convulsions and normal in-bed movements, since the ZCR has already been used for sEMG-based detection of tonic–clonic seizures [5]. Apart from the ZCR, Hjorth's parameters (mobility and complexity) have been recognised as relevant features. To our knowledge, these parameters have never been used for ACM-based seizure detection. Other chosen features have already been employed in previous publications but as part of much larger feature sets [1, 20].

The classification performances significantly depended on the chosen wrapper search strategy. This difference was in direct relation to the number of chosen features. In the case of backward elimination, the number of chosen features was always greater than or equal to those obtained by forward search. Using forward search, one feature was permanently removed in each iteration. Thus, the effects of combined features were not always taken into consideration. However, the smallest degradation in all performances was observed using only the mRMR method, except in the case of the AUC for motor seizures, the FDR for (tonic–)clonic and motor seizures, and latency for (tonic–)clonic seizures. When all features were used, the sensitivity for some patients was zero while these seizures were detected when the mRMR feature selection method was applied, although with a higher latency. One possible explanation for the aforementioned degradation of seizure detection performance is that the criteria for feature selection were too strict. It is probable that a small set of features was obtained from the original set, which was very diverse; hence, information might had been lost as some of the features had been omitted.

The obtained results suggest that seizure-type-specific models have better performance. Moreover, some seizure classes are more easily detected with accelerometry. For instance, accelerometry is not suitable for the detection of tonic seizures since these seizures manifest themselves in a slow change in limb posture and usually there is not much acceleration activity during these seizures. In contrast, it is easier to detect seizures such as (tonic–)clonic ones using accelerometry owing to the specific rhythmic movements of the limbs.

In general, the number of seizures per patient used for the feature selection, LS-SVM training, and testing was limited. Moreover, patients with a small number of seizures did not participate in the testing. We chose to build one algorithm on mixed data (smaller number of patient-specific seizures and larger number of other patient seizures) as a compromise between algorithm performance and computational load (computational time). Nevertheless, the bias induced with this strategy sometimes resulted in large differences in classification performances between test patients. However, this is not the only reason for the inconsistency among patients. Seizures from the same patient were sometimes clinically diverse, and the trained models were unable to perform well for all possible clinical manifestations (see the Supplementary Materials for a few examples). One possible solution is to make subgroups of ACM similar seizures within one seizure type and build the models for each subgroup.

Other groups tried to detect motor seizures. Nijsen et al. [21] investigated the detection of myoclonic seizures by means of accelerometers. Various methods were tested, resulting in a maximal sensitivity of 80 %, but all with a high FDR. Whereas in [21] the algorithm was tested on segments containing seizures, normal movements, and no movements, we used all data of two test patients, reaching only 50 % sensitivity. More test data are needed to assess the real value of the model. Conradsen et al. [6] used a multimodal approach to detect simulated tonic–clonic seizures. A sensitivity of 100 % with an FDR between 0 and 18 $h^{-1}$ were obtained using ACM, magnetometer, gyroscope, and sEMG data. The developed models still have to be tested on real data. Poh et al. [24] developed a wrist-worn electrodermal activity and ACM biosensor to detect generalised tonic–clonic seizures. A total of 15 out of 16 GTC seizures were detected with 130 false alarms (0.74 alarm/h). We joined clonic and tonic–clonic seizures, and obtained a median sensitivity of 100 % with a median false detection rate of 10.5/h, using only features obtained with the mRMR method, although for some patients only half of the seizures were correctly identified. Within our database, we observed large inter- and intra-patient seizure morphology. This diversity is not present in simulated and generalised tonic–clonic seizures, which generally have stereotypical patterns. In addition, in the latter two publications, a multimodal approach was used that takes into consideration more aspects of epileptic seizures.

Generally, the obtained FDR was significantly higher than in the aforementioned publications, which is the consequence of the way this metric was estimated. Each segment of false-positive successive epochs was considered as one false alarm even when there are two segments only a few seconds apart. Thus, the output of the classifier was treated the same for both sensitivity and FDR estimation, which was not the case in other publications [6].

As mentioned in Sect. 2.1, 22 recordings (18 patients) with 631 seizures had been removed from the feature

selection analysis. This substantial amount of data had been removed to exclude the influence of nonfunctional channels on the feature selection. A separate study has to be performed to evaluate the influence of missing values and missing value imputation techniques on the feature selection and classification.

However, the hardware set-up used in this study was only a prototype system, designed to gather data for algorithm development. It was not intended to be as reliable as would be expected from clinical devices. Recently, different commercial systems have become available that can replace this hardware. The techniques explained in this paper are independent of these differences in hardware set-up.

Apart from the removal of part of the data from nonfunctional channels, one of the main limitations of the present study is that the obtained features are only optimal for the used ACM set-up and chosen population. Other ACM sensor locations and their total number will definitely affect the overall feature selection outcome. An initial feature selection analysis performed on limited data using different ACM sensor locations and numbers showed that some features were universal (i.e. not dependent on the location and sensor number), whereas others appeared only in ACM location- and sensor-number-dependent feature subsets. Therefore, in case of other measurement set-ups, detection goals, and population, the analysis should be redone and the paper guides us on how to perform it.

Although the performances reached are not clinically acceptable, this study showed that using solely an mRMR feature selection algorithm, which is not computationally demanding, led to the smallest degradation in classification performance among the used feature reduction methods. The best results are obtained for (tonic)–clonic seizures: despite a significant increase in the FDR, no significant reduction in the AUC has been observed. These seizures are the most relevant ones for a future alarm system, since they can potentially lead to physical injuries.

## 5 Conclusion

The aim of this study was to investigate the use of feature selection methods within the machine learning framework for detection of motor seizures using accelerometry. We proposed a two-stage feature selection strategy, including mRMR filter and LS-SVM wrapper selection methods applied one after the other. By applying only a filter feature selection method, the obtained classification results had comparable or slightly reduced performance with respect to the complete feature set, depending on the seizure type detected. Moreover, there was no need for the application of a wrapper method, which is computationally demanding. Leading performances were obtained for (tonic–)clonic seizures, with median sensitivity of 100 %, but a high FDR of 10.5/h. The main limitation of the present study is that the obtained features are only optimal for the used ACM measurement set-up and chosen population. Nevertheless, in this paper, we have explained in full detail the methodology that should be repeated for other ACM configurations. In addition, the proposed methodology and obtained results should be further validated by adding new data for the early detection of the most violent seizures, i.e. tonic–clonic seizures.

**Acknowledgments** The research was supported by Bijzonder Onderzoeksfonds KU Leuven (BOF); Center of Excellence (CoE): PFV/10/002 (OPTEC); Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO): G.0108.11 (Compressed Sensing), G.0869.12N (Tumor imaging), and G.0A5513N (Deep brain stimulation); iMinds Medical Information Technologies: Dotatie-Strategisch basis onderzoek (SBO-2015); Belgian Federal Science Policy Office: IUAP P7/19/(DYSCO, 'Dynamical systems, control and optimization', 2012–2017); EU: European Union's Seventh Framework Programme (FP7/2007–2013): EU MC ITN TRANSACT 2012, 316679, ERASMUS EQR: Community service engineer, 539642-LLP-1-2013; and European Research Council: ERC Advanced Grant, 339804 BIOTENSORS. This paper reflects only the authors' views, and the Union is not liable for any use that may be made of the contained information.

## References

1. Becq G, Bonnet S, Minotti L, Antonakios M, Guillemaud R, Kahane P (2011) Classification of epileptic motor manifestations using inertial and magnetic sensors. Comput Biol Med 41(1):46–55
2. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, Van Emde Boas W, Engel J, French J, Glauser TA, Mathern GW et al (2010) Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. Epilepsia 51(4):676–685
3. Botev Z, Grotowski J, Kroese D et al (2010) Kernel density estimation via diffusion. Ann Stat 38(5):2916–2957
4. Carlson C, Arnedo V, Cahill M, Devinsky O (2009) Detecting nocturnal convulsions: efficacy of the MP5 monitor. Seizure 18(3):225–227
5. Conradsen I, Beniczky S, Hoppe K, Wolf P, Sorensen HB (2012) Automated algorithm for generalized tonic–clonic epileptic seizure onset detection based on sEMG zero-crossing rate. IEEE Trans Biomed Eng 59(2):579–585
6. Conradsen I, Beniczky S, Wolf P, Kjaer TW, Sams T, Sorensen HB (2012) Automatic multi-modal intelligent seizure acquisition (MISA): system for detection of motor seizures from electromyographic data and motion data. Comput Methods Programs Biomed 107(2):97–110
7. Cuppens K, Karsmakers P, Van de Vel A, Bonroy B, Milosevic M, Luca S, Croonenborghs T, Ceulemans B, Lagae L, Van Huffel S, Vanrumste B (2014) Accelerometry-based home monitoring for detection of nocturnal hypermotor seizures based on novelty detection. IEEE J Biomed Health Inform 18(3):1026–1033
8. Cuppens K, Lagae L, Ceulemans B, Van Huffel S, Vanrumste B (2009) Detection of nocturnal frontal lobe seizures in pediatric patients by means of accelerometers: a first study. In: Proceedings of the annual international conference of the IEEE

engineering in medicine and biology society (EMBC). IEEE, pp 6608–6611
9. Cuppens K, Lagae L, Ceulemans B, Van Huffel S, Vanrumste B (2010) Automatic video detection of body movement during sleep based on optical flow in pediatric patients with epilepsy. Med Biol Eng Comput 48(9):923–931
10. Ding C, Peng H (2005) Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol 3(02):185–205
11. Fisher RS, Boas WvE, Blume W, Elger C, Genton P, Lee P, Engel J (2005) Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia 46:470–472
12. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
13. Hunyadi B, Signoretto M, Van Paesschen W, Suykens JA, Van Huffel S, De Vos M (2012) Incorporating structural information from the multichannel EEG improves patient-specific seizure detection. Clin Neurophysiol 123(12):2352–2361
14. Karantonis D, Narayanan M, Mathie M, Lovell N, Celler B (2006) Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Trans Inf Technol Biomed 10(1):156–167
15. Kohavi R, John G (1997) Wrappers for feature subset selection. Artif Intell 97(1):273–324
16. Kramer U, Kipervasser S, Shlitner A, Kuzniecky R (2011) A novel portable seizure detection alarm system: preliminary results. J Clin Neurophysiol 28(1):36
17. Lockman J, Fisher R, Olson D (2011) Detection of seizure-like movements using a wrist accelerometer. Epilepsy Behav 20(4):638–641
18. Logesparan L, Casson AJ, Rodriguez-Villegas E (2012) Optimal features for online seizure detection. Med Biol Eng Comput 50(7):659–669
19. Lüders H, Acharya J, Baumgartner C, Benbadis S, Bleasel A, Burgess R, Dinner D, Ebner A, Foldvary N, Geller E et al (1998) Semiological seizure classification*. Epilepsia 39(9):1006–1013
20. Nijsen T, Aarts R, Arends J, Cluitmans P (2009) Automated detection of tonic seizures using 3-D accelerometry. In: Proceedings of the 4th European conference of the international federation for medical and biological engineering. Springer, pp 188–191
21. Nijsen T, Aarts R, Cluitmans P, Griep P (2010) Time–frequency analysis of accelerometry data for detection of myoclonic seizures. IEEE Trans Inf Technol Biomed 14(5):1197–1203
22. Pelckmans K, Suykens JA, Van Gestel T, De Brabanter J, Lukas L, Hamers B, De Moor B, Vandewalle J (2002) LS-SVMlab: a matlab/c toolbox for least squares support vector machines. Tutorial. KULeuven-ESAT, Leuven, Belgium
23. Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27(8):1226–1238
24. Poh MZ, Loddenkemper T, Reinsberger C, Swenson NC, Goyal S, Sabtala MC, Madsen JR, Picard RW (2012) Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. Epilepsia 53(5):e93–e97
25. Preece S, Goulermas J, Kenney L, Howard D (2009) A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans Biomed Eng 56(3):871–879
26. 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):173–183
27. Sheikholeslami N, Stashuk D (1997) Supervised mutual-information based feature selection for motor unit action potential classification. Med Biol Eng Comput 35(6):661–670
28. Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J (2002) Least squares support vector machines. World Scientific Pub. Co., Singapore
29. Temko A, Thomas E, Marnane W, Lightbody G, Boylan G (2011) EEG-based neonatal seizure detection with support vector machines. Clin Neurophysiol 122(3):464–473
30. Vapnik V (1999) The nature of statistical learning theory. Springer, Berlin
31. Van de Vel A, Cuppens K, Bonroy B, Milosevic M, Jansen K, Van Huffel S, Vanrumste B, Lagae L, Ceulemans B (2013) Non-EEG seizure-detection systems and potential SUDEP prevention: state of the art. Seizure 22:345–355



**Milica Milošević** born in Belgrade, Serbia, obtained her Ph.D. in Electrical Engineering in May 2015 at the KU Leuven. Her research interests are Biomedical Signal Processing and Machine Learning.



**Anouk Van de Vel** has a Master in Biomedical Sciences and Neurosciences. She is EEG analyst at the University Hospital Antwerp and researcher in the area of non-EEG-based epileptic seizure detection.



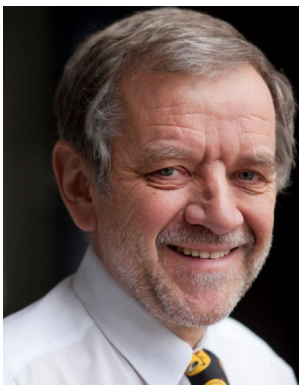
**Kris Cuppens** obtained his Ph.D. in Engineering Sciences in 2012 at the KU Leuven in collaboration with Thomas More University College. His research interests are Biomedical Signal and Image Processing.



**Bert Bonroy** obtained his Ph.D. in Engineering Technology in 2015 at the KU Leuven in collaboration with Thomas More Kempen University College. His research interest is Biomedical Engineering.



**Bart Vanrumste** is an Associate Professor at the Department of Electrical Engineering at KU Leuven, Leuven, Belgium, and Head of the Advanced Integrated Sensing Lab (ADVISE) at KU Leuven Technology Campus, Geel.



**Berten Ceulemans** is the Coordinator of the Department of Pediatric Neurology at the University Hospital Antwerp and Associate Professor at the University of Antwerp, Belgium.



**Sabine Van Huffel** is a Full Professor at the Department of Electrical Engineering, KU Leuven, Leuven, Belgium, and Head of the Biomedical Data Processing Research Group.



**Lieven Lagae** is a Full Professor at the Department of Development and Regeneration of the Medical School, KULeuven, and Head of the Paediatric Neurology section.