PAPER • **OPEN ACCESS**

Improving the generalization of patient non-specific model for epileptic seizure detection

To cite this article: Gustav Munk Sigsgaard and Ying Gu 2024 *Biomed. Phys. Eng. Express* **10** 015010

Biomedical Physics & Engineering Express

PAPER

Improving the generalization of patient non-specific model for epileptic seizure detection

Gustav Munk Sigsgaard and Ying Gu

Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

E-mail: yingu@dtu.dk

Keywords: epilepsy, feature extraction, machine learning, random forest, electroencephalography (EEG)

OPEN ACCESS

RECEIVED
25 June 2023

REVISED
25 October 2023

ACCEPTED FOR PUBLICATION
3 November 2023

PUBLISHED
6 December 2023

Abstract

Epilepsy is the second most common neurological disorder characterized by recurrent and unpredictable seizures. Accurate seizure detection is important for diagnosis and treatment of epilepsy. Many researches achieved good performance on patient-specific seizure detection. However, they were tailored to each specific individual which are less applicable clinically than the patient non-specific detection, which lacked good performance. Despite several decades of research on automatic seizure detection, seizure detection is currently still based on visual inspection of video-EEG (Electroencephalogram) in clinical setting. It is time consuming and prone to human error and subjectivity. This study aims to improve patient non-specific seizure detection to assist neurologist with efficient and objective evaluation of epileptic EEG. The clinical data used was from the open access Siena Scalp EEG Database which consists of 14 patients. First the data were pre-processed to remove artifacts and noises. Second the features from time domain, frequency domain and entropy were extracted from each channel and then concatenated into a feature vector. Finally, a machine learning approach based on random forest was employed for seizure detection with leave-one-patient-out cross-validation scheme. Automatic seizure detection was carried out with the trained model. The study achieved a specificity of 99.38%, sensitivity of 81.43% and 3.61 FP/h (False Positives per hour), which outperformed some other patient non-specific detectors found in literature. The findings from the study shows the possibility of clinical application of automatic seizure detection and indicate that further work should focus on dealing with reducing false positives.

1. Introduction

Around 1% of the world's population suffer from epilepsy which makes it the second most common neurological disorder only surpassed by stroke [1]. Around 30% of epilepsy patients do not respond to medication. In some cases, surgery to remove the area of the brain that is causing seizures may be an option for them. Epilepsy is characterized by recurrent and unpredictable seizures, which can significantly impact the patients' lives. Due to loss of control and consciousness, patients often experience serious injuries. Moreover, sudden unexpected deaths in epilepsy (SUDEP) which is the major cause of epilepsy related mortality is 2-3 times higher for epileptic patients [2]. Despite the underlying pathogenesis of SUDEP being poorly understood, a key contributing factor seems to be lack of seizure control. Therefore, an early and correct detection of seizures is crucial for management of epilepsy. Currently Video- EEG (Electroencephalogram) monitoring of epilepsy is the golden standard at hospital for pre-surgical evaluation [3]. EEG is a non-invasive recording of brain activities and has high temporal resolution, which makes it efficient to accurately detect the onset of seizure and its type. The research has shown that long-term video-EEG monitoring for hours to days can contribute positively to the diagnostic yield due to unpredictability of seizure occurrence [4]. However, it is a highly time-consuming and labour-intensive process for a neurologist to conduct a manual review on multi-channel scalp EEG. In addition, visual inspection is not only inefficient and ineffective but also prone to human errors because epileptic seizures may look similar to artifacts and noises [5]. For the above reasons, there is a clear need for an automation of epileptic seizure detection which

can assist neurologists' EEG evaluation and greatly reduce their time spent on diagnosing each patient.

For several decades, research has focused on development of automatic seizure detectors [6–11]. Seizure detectors are typically divided into two categories: patient specific or patient non-specific. Generally, the patient-specific is tailored to the seizure characteristics of one patient and applied for this patient. In contrast, non-specific detectors are trained with a group of patients and later applied directly to detect seizures for new patients. One of the earliest patient non-specific detectors was developed by Gotman *et al* in 1982 [6] and later improved in [12]. Gotman *et al* validated their methods in a separate dataset containing 54 patients' EEG data from three medical centers and reached a seizure detection rate of 69% with a false detection rate of 2.3 per hour [7]. As the research area gained more attention, people found that patient specific algorithms outperformed the non-specific ones due to a large patient-to-patient variability in EEG seizure patterns [8, 13]. In a study by Shoeb *et al* from 2004, they presented a patient-specific method to exactly utilize this consistency of seizure and non-seizure EEG within the same patient [13]. Their method rested on extracting features from a 2-second epoch using a multi-level wavelet decomposition filter bank to achieve four desired frequency sub-bands and extract their energy. They managed to detect 131 of 139 seizure events which corresponds to a sensitivity of 94% using support vector machine (SVM). Hunyadi *et al* [14] presented another patient-specific seizure detection algorithm in which they extracted both time- and frequency domain features. Among three different integration approaches, an integration approach where features were extracted and preserved in their original arrangement in the form of a matrix performed best with a median sensitivity of 100%. The research on patient non-specific detectors is limited. A study by Orosco *et al* [15] proposed a pediatric non-specific algorithm based on data collected from 22 subjects from the CHB-MIT scalp EEG database [16]. The EEG signal was divided into epochs of 2 seconds and decomposed into specific sub-bands by wavelet transform. For each band and channel, spectral and energy features were extracted prior to a feature selection process. Finally, the selected features were fed to two different classifiers, one based on linear discriminant analysis and another on a neural network. Their best performing method achieved an average sensitivity of 87.5%, a specificity of 99.9%, and a false positives per hour (FP/h) of 0.9. Part of study done by Greene *et al* [17] investigated a non-specific model using 7-12 EEG channels from 12 records of 10 term neonates. For every eight-seconds EEG epoch, a feature vector containing six features per channel was produced. With a linear discriminant classifier, the non-specific model reached a sensitivity of 80.41% with 3.42 FP/h. The work by Sridevi *et al* [18] used multiday video-scalp EEG monitoring on 18 drug-resistant patients with temporal lobe epilepsy showing 29 seizure events. They used a 4-seconds window and extracted signal energy, spectral energies in the low (1-25 Hz), mid (25-100 Hz), and highband (100-200 Hz), and the spectral entropy. For classifying their features, they investigated five different classifiers. The best performing classifier SVM achieved a sensitivity of 80% and a specificity of 86%.

Different classifier methods have been proposed for epileptic seizure detection in EEG. In a recent paper from 2022, Sánchez-Hernández *et al* evaluated various classifiers in combination with different feature selection methods for two different feature sets [19]. Classifiers included a decision tree (DT), SVM, K-nearest neighbour (KNN), random forest (RF), and an artificial neural network (ANN). The feature selection methods comprised a DT, a SVM, local interpretable model-agnostic explanations (LIME), shapley additive explanations (SHARP), an embedded RF (ERF), and reciprocal ranking. Combinations were evaluated on two open access scalp EEG databases, namely the CHB-MIT and the Siena Scalp EEG [20]. For the CHB-MIT database, they reported an optimal F1-score of 0.9 for the KNN-SVM combination and for the Siena Scalp database a F1-score of 0.86 with the combination RF-ERF.

Many patient-specific seizure detectors have achieved good performance as previously described. However, they need to be tailored to each patient prior to their employment, which limits their applicability for clinical setting. The strength of patient non-specific detectors is their unspecific nature and attempt to generalize among patients, which offers high clinical practicability. However, little research has gone into non-specific detection. Its major shortcoming is lower performance compared to patient-specific. This can explain why automatic seizure detection hasn't been applied in clinical setting. Our study aims to improve the performance of a patient non-specific model for clinical application. We developed a new approach for improving the generalization of the model. The features extracted from time domain, frequency domain and entropy and from multiple EEG channels served as an input to a detector based on random forest. Those features reflecting different perspectives of the signal were used to represent the variability among patients and provide random forest with rich features. The randomly selected subset from those rich features for developing decision tree in random forest will improve the generalization of the model and reduce the over-fitting, therefore facilitate a precise seizure detection performance among patients. The main purpose of patient non-specific seizure detector is to assist the neurologist for an efficient and effective evaluation of epilepsy.

The remainder of the paper is organized as follows. Section 2 first presents the Siena Scalp EEG Database. Then the steps of automatic seizure detection involving signal pre-processing, feature extraction and

Table 1. Patient Information from Siena Scalp EEG Database (Pat.: patient; Locali: Localization; Laterali.: Lateralization; #chan.: nr. of channels; #Seiz.: nr. of seizures; Time: monitoring time in minutes; IAS: focal onset impaired awareness, WIAS: focal onset without impaired awareness, FBTC: focal to bilateral tonic-clonic; T: temporal lobe, F: frontal lobe; R: right, L: left, Bilateral: right and left simultaneously.).

Pat. Id	Age	Gender	Seizure	Locali.	Laterali.	# Chan.	# Seiz.	Time
PN00	55	Male	IAS	T	R	29	5	192
PN01	46	Male	IAS	T	L	29	2	809
PN03	54	Male	IAS	T	R	29	2	1453
PN05	51	Female	IAS	T	L	29	3	362
PN06	36	Male	IAS	T	L	29	5	723
PN07	20	Female	IAS	T	L	29	1	524
PN09	27	Female	IAS	T	L	29	3	410
PN10	25	Male	FBTC	F	Bilateral	20	10	1122
PN11	58	Female	IAS	T	R	29	1	145
PN12	71	Male	IAS	T	L	29	4	366
PN13	34	Female	IAS	T	L	29	3	520
PN14	49	Male	WIAS	T	L	29	4	1408
PN16	41	Female	IAS	T	L	29	2	293
PN17	42	Male	IAS	T	R	29	2	308

classification with random forest are described. The results are showed in section 3 and followed by discussion in section 4. Section 5 concludes the study and proposes future research perspectives.

2. Methodology

2.1. Siena scalp EEG database

The clinical data used in this project was from the open access Siena Scalp EEG Database [20]. The data was collected by the Unit of Neurology and Neurophysiology at the University of Siena during the national interdisciplinary research project called PANACEE [21]. The database consists of recordings from 14 patients with focal epilepsy including nine males (ages 36-71) and five females (ages 20-58). The patients were monitored with a non-invasive scalp Video-EEG with electrodes arranged according to the international 10-20 system using a sampling rate of 512Hz. The monopolar EEG was recorded. Patients were told to be in bed for the most of the recording time, either asleep or awake. Along with the recordings, the database holds a seizure list file for each patient specifying the start and stop recording times, start and stop times of seizure, and a list of all electrodes being used. The exact times of seizure events were carefully annotated by expert clinicians. In addition, the database comes with a patient info file containing relevant information about each patient such as age, gender, seizure type, number of seizures etc. Table 1 shows the basic information about all patients. About seizure, IAS refers to focal onset impaired awareness, WIAS is focal onset without impaired awareness, and lastly, FBTC is focal to bilateral tonic-clonic. Localization refers to which brain lobe the seizure originates, here T is temporal and F is frontal. The lateralization refers to the side of origin in the brain, here R is right, L is left, and bilateral is both sides simultaneously. Other metrics are self-explanatory. In total, the database contains 47 seizure instances with a mean seizure duration of 61 seconds on approximately 144 hours of monitoring time. Figure 1 shows the seizures' duration among patients.

One change was made to the available data. In the seizure file for patient PN00, it is stated that the recording starts at 18:15:44 and ends at 18:57:13. The seizure period, however, is 18:28:29-19:29:29 which was assumed to be a typo and instead the seizure ought to end at 18:29:29. Several arguments justify this assumption. Firstly, the seizure end time extents the recording period. Secondly, the mean seizure duration (computed based on the patient's four other seizures) for PN00 is about 66 seconds. A maximum seizure duration is 151 seconds across all patients. Therefore, it would be highly unlikely that PN00 had a seizure duration of 3600 seconds (one hour). Lastly, after a visual inspection of the data, it does look like the seizure ends at 18:29:29. Consequently, the altered time of seizure end was used in the further analysis.

2.2. Automatic seizure detection

Seizure detection aims to classify an epoch as either seizure or non-seizure. In this study, the seizure detection involved pre-processing, features extraction from time domain, frequency domain and entropy and classification carried out with random forest. A schematic overview of the seizure detection architecture can be seen in figure 2.

2.2.1. EEG Signal Pre-processing

The information about seizure periods and channels were extracted for each patient. Out of all available channels, 19 channels (Fp1, F3, C3, P3, O1, F7, T3, T5, Fz, Cz, Pz, F4, C4, P4, O2, F8, T4, T6 and Fp2) remained the same across all 14 patients and were used for further analysis. The filtering scheme is similar to the one presented in [19]. The signal was first filtered

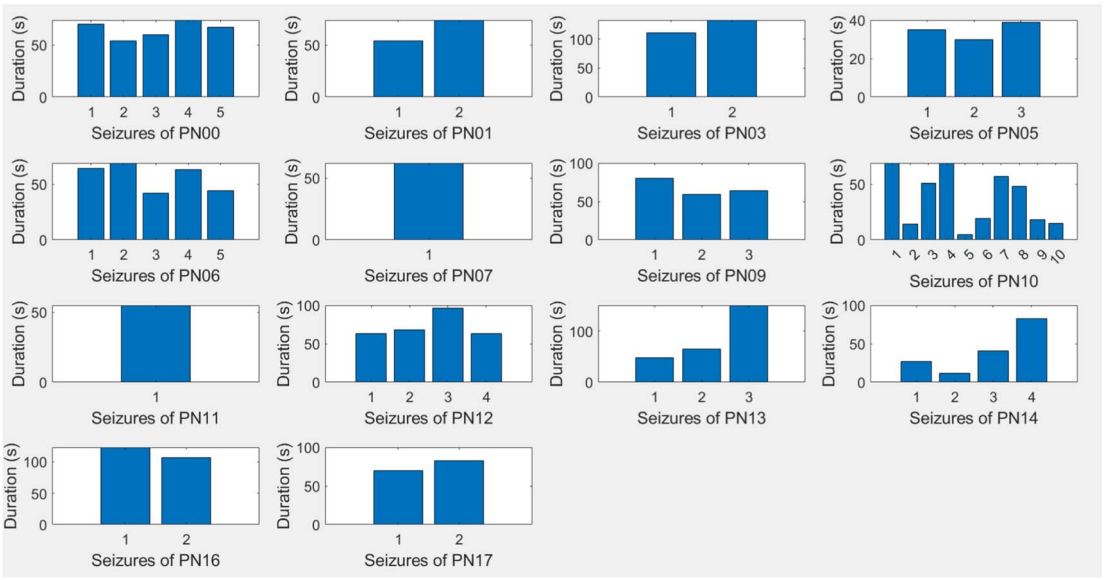

Figure 1. Seizure duration among patients.

with a fourth-order Butterworth high-pass filter at 0.5Hz to remove possible baseline wander and low frequency noises, and a 50 Hz notch filter to remove the power line interference. The signal was then divided into shorter epochs, from which features will be extracted. This permits a tracking of the evolution of the signal. For EEG, it is important to choose a small epoch size due to the non-stationary of EEG signals [19, 22]. From literature, epochs between 2-12 seconds are typically being used, sometimes with overlap [13, 21, 23, 24]. In this study, an epoch length of 6 seconds was chosen as a trade-off between dealing with the non-stationary and computational complexity. The 6-seconds epoch was then moved along the signal extracting $l = 1,\ldots,L$ epochs where L denotes the total number of 6-seconds epochs in the recording. Epochs outside the seizure period or partly overlapping with seizure period were labelled with a 0 (non-seizure) while full overlapping epochs were labelled with a 1 (seizure). Furthermore, as there were far more non-seizure data compared to seizure data, a 50% overlap was applied to full overlapping seizure epochs. A representation of the signal segmentation process can be seen in figure 3.

2.2.2. Feature extraction

Extracting meaningful features is important for seizure detection. In order to deal with variability of patients with non-specific model, time domain features, frequency-domain features, and entropy derived features which reflect signal characteristics from different perspectives were extracted from every 6-seconds epoch of every channel (N = 19 channels). Those features have been found in literature and are widely used for adult scalp EEG [14]. In total, $M = 22$ features extracted from every channel were concatenated into a feature vector X_l with length of $N \cdot M = 418$. Below, each feature is described along with their corresponding equations.

Time-domain features: The following features were extracted. Here, x represent EEG time series.

- **Number of zero crossings** refers to the number of times the signal goes from a positive voltage value to a negative one or vice versa.
- **Maximum** represents the largest voltage value.
- **Minimum** represents the smallest voltage value.
- **Skewness** is a measure of the lack of symmetry in the amplitude distribution. Larger values mean more skewness. It is computed as: $skew = \frac{E(x-u)^3}{\sigma^3}$, where u is the mean of x, σ is the standard deviation of x.
- **Kurtosis** gives information about the extent of the peak in the amplitude distribution. It is computed as: $kurt = \frac{E(x-u)^4}{\sigma^4}$, where u is the mean of x, σ is the standard deviation of x.
- **Root mean square** measures the magnitude of the varying quantity [25] and is computed as in [26]:
$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$
- **Mobility** is the second of Hjorth's three statistical time-domain parameters introduced by Bo Hjorth in 1970 [27]. The mobility parameter measures the standard deviation of the slope with reference to the amplitude's standard deviation [27]. It is calculated as: $Mobility = \sqrt{\frac{\sigma^2(\dot{x})}{\sigma^2(x)}}$, where $\dot{x}$ denotes the derivative of the signal x and σ the standard deviation.
- ***Complexity*** is the third Hjorth's parameter and gives an estimate of the bandwith of the signal indicating the similarity of the shape of the signal

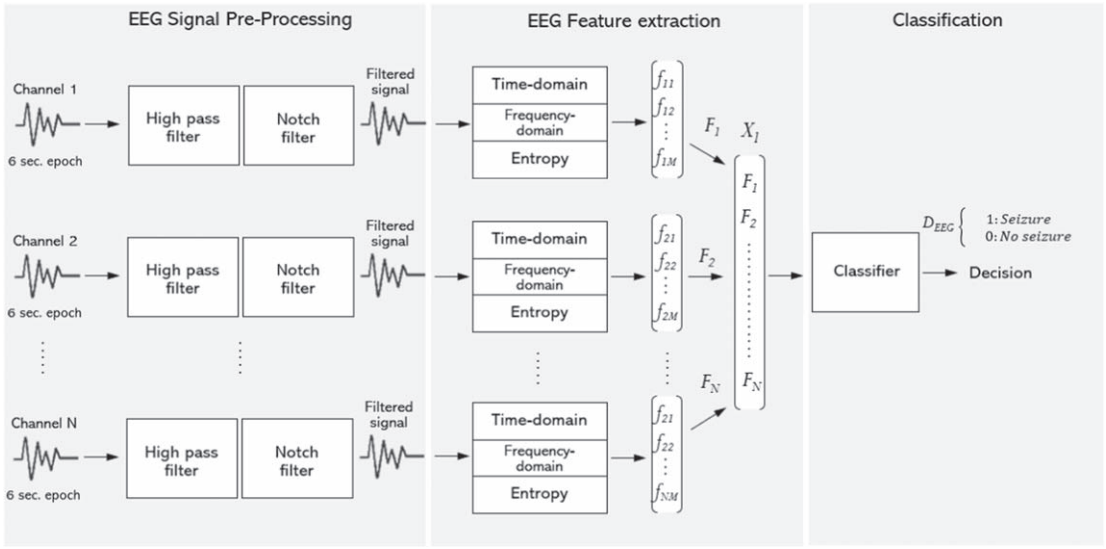

Figure 2. Architecture of the seizure detector. f_{NM} corresponds to the *Mth* feature of channel *N*. F_N is a vector with all features from channel *N*, and X_l is the concatenated feature vector where $l = 1,\ldots,L$ number of epochs in the EEG recording. D_{EEG} is the decision outputted by the detector.

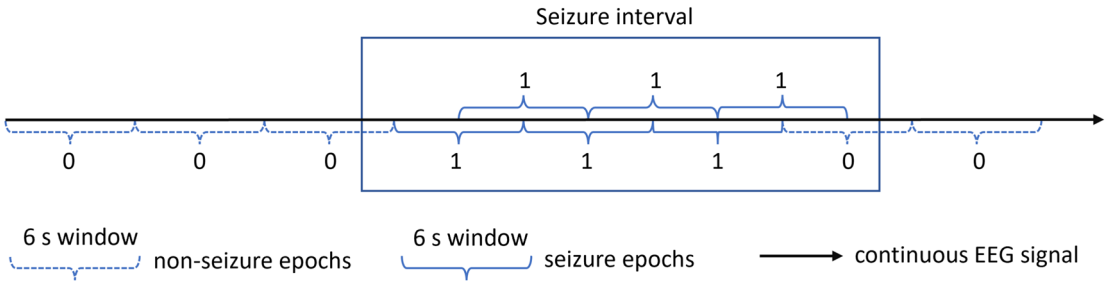

Figure 3. Segmentation of the EEG signal into seizure (1) and non-seizure (0) epochs.

compared to a pure sine wave [28]. It is calculated as: $Complexity = \frac{\sqrt{\frac{\sigma^2(\ddot{x})}{\sigma^2(\dot{x})}}}{\sqrt{\frac{\sigma^2(\dot{x})}{\sigma^2(x)}}}$, where $\ddot{x}$ denotes the double derivative of the signal x, $\dot{x}$ denotes the derivative of the signal x and σ the standard deviation.

Frequency-domain features First, power spectral density (PSD) was calculated using the fast fourier transform (FFT) with the Welch's method [29]. According to brain status and cognitive activities, the EEG can be classified into five different frequency sub-bands namely δ $(0.5-4Hz)$; θ $(4-8Hz)$; α $(8-13Hz)$; β $(13-30Hz)$; γ $(30-80Hz)$. In the frequency domain, the following features were extracted where f represents the frequency and $S(f_i)$ is the spectral power of that frequency:

- **Total power** $totalPow = \sum_{f_i=0.5}^{30} S(f_i)$.
- **Peak frequency** is the frequency with highest spectral power. $peakFreq = arg \max_{f\in[0.5,30]} S(f)$.
- **Mean power** in five frequency sub-bands δ, θ, α, β and high frequency (HF). Here n_δ, n_θ, n_α, n_β, n_{HF} represent the number of the samples in those sub-bands respectively.

$$meanPow_\delta = \frac{1}{n_\delta}\sum_{f_i=0.5}^{4} S(f_i),$$
$$meanPow_\theta = \frac{1}{n_\theta}\sum_{f_i=4}^{8} S(f_i),$$
$$meanPow_\alpha = \frac{1}{n_\alpha}\sum_{f_i=8}^{13} S(f_i),$$
$$meanPow_\beta = \frac{1}{n_\beta}\sum_{f_i=13}^{30} S(f_i),$$
$$meanPow_{HF} = \frac{1}{n_{HF}}\sum_{f_i=40}^{80} S(f_i)$$

- **Normalized power** in five frequency sub-bands δ, θ, α, β and high frequency

$$normPow_\delta = \frac{\sum_{f_i=0.5}^{4} S(f_i)}{\sum_{f_i=0.5}^{30} S(f_i)},$$
$$normPow_\theta = \frac{\sum_{f_i=4}^{8} S(f_i)}{\sum_{f_i=0.5}^{30} S(f_i)},$$
$$normPow_\alpha = \frac{\sum_{f_i=8}^{13} S(f_i)}{\sum_{f_i=0.5}^{30} S(f_i)},$$
$$normPow_\beta = \frac{\sum_{f_i=13}^{30} S(f_i)}{\sum_{f_i=0.5}^{30} S(f_i)},$$
$$normPow_{HF} = \frac{\sum_{f_i=40}^{80} S(f_i)}{\sum_{f_i=0.5}^{30} S(f_i)}$$

Entropy features In general, entropy is an information measure that represents the amount of uncertainty associated with events from a given distribution [30, 31]. Two entropy measures were extracted namely sample entropy and the spectral entropy.

- **Sample entropy** is proposed by Richman and Moorman in [32] and is used to assess the complexity of physiological time-series signals. It was computed as: $SampEn(x, m, r) = -log(\frac{C(m+1,r)}{C(m,r)})$, where x is a signal, m is the embedding dimension (m = 2 in the study), r is the radius of the neighbourhood ($r = 0.2std(x)$ in the study), $C(m+1, r)$ is the number of embedded vectors of length $m+1$ having a Chebyshec distance inferior to r and $C(m, r)$ is the number of embedded vectors of length m having a Chebyshev distance inferior to r.
- **Spectral entropy** is a measure of the degree of EEG irregularity because the entropy of the power spectrum represents the relative peakedness or flatness of the spectral distribution [31]. It is calculated as:
  $SpecEn = -\sum_{f=0}^{f_s/2} P(f)\ log_2[P(f)]$, where P is the normalized PSD and f_s is the sampling frequency.

2.2.3. Classification based on random forest

Random forest is an ensemble method and was firstly introduced by Breiman in 2001 [33]. An ensemble method is constructed from several models (in our case weaker classifiers) and decisions are made by taking a majority vote from all weaker classifiers. There are different ensemble learning methods, the most common being bagging and boosting. Random forest is based on bagged decision trees, where each tree is trained on a different subset with replacement from original data. Randomly selected subset of features are used at each split point of the decision tree to avoid high correlation among trees in ensemble. The following paragraph describes the decision tree.

Decision Tree: The algorithm behind decision trees was developed by Earl B. Hunt and others in 1963 [34]. The algorithm adopts a greedy, top-down approach starting at the root of the tree with the entire training dataset. From this point, the tree branches out meaning that the training set is recursively divided into smaller partitions, commonly referenced to as child nodes. There are two types of nodes, decision nodes and leaf nodes. In a decision node, the training data is split into child nodes according to some condition decided by the model. Depending on the attributes of the data, these splits may be binary, discrete, or continuous. Naturally, there are many different splits to choose from which begs the question of how the model finds the optimal one. It does so by trying all combinations of splits and for each one, it computes the impurity of the child nodes. The impurity is a measure of how the different classes are divided in the child nodes where optimally every child node only contains one class. Such a split would give an impurity of 0. To calculate the impurity, one of three different impurity functions $I(v)$ are often used. Those are entropy, Gini, and classification error [30]:

$$Entropy(v) = -\sum_{c=1}^{C} p(c|v) log_2 p(c|v), \quad (1)$$

$$Gini(v) = 1 - \sum_{c=1}^{C} p(c|v)^2, \quad (2)$$

$$ClassError(v) = 1 - \max_c\ p(c|v), \quad (3)$$

where C is the number of classes and $p(c|v)$ denotes the probability of a class in that branch v with $v_1,\ldots,v_K$. The model will then select the split that maximizes the *purity gain* Δ which is calculated by subtracting the impurity at the branches $I(v_1)$, $I(v_2)$, $\ldots I(v_K)$ from the impurity at the root $I(r)$ given by the formula [30]:

$$\Delta = I(r) - \sum_{k=1}^{K} \frac{N(v_k)}{N(r)} I(v_k), \quad (4)$$

where the term $\frac{N(v_k)}{N(r)} I(v_k)$ is the weight given by the fraction of total observations at the branch. Note that once the model has found its best split and moved forward, there is no backtracking. This is what greediness of algorithm entails. Furthermore, the decision tree will keep branching out until all nodes end in leaf nodes that is nodes containing only one (unless otherwise is specified) particular class. The decision tree can also be terminated earlier by setting different stop criteria such as a max depth, if the purity gain Δ of a split drops below some level, or if a the number of observations in a branch becomes less than or equal to some specified number. In case of the random forest, the number of trees corresponds to the number of decision trees being trained.

Random forest was chosen due to its widespread applicability in automatic seizure detection. In addition, the research [23] has shown that random forest outperformed other classifiers (ANN, KNN, SVM, single Decision Tree). In this study, the Ginientropy (see equation (2)) was used as a measure of impurity in the different branches and 150 decision trees were used in the random forest. The number of randomly selected features for each split of the decision tree was 21, which was calculated as rounding $\sqrt{418}$ to the nearest integer larger than $\sqrt{418}$. Here 418 is the number of the extracted features.

2.2.4. Model training and evaluation

Figure 4 shows the steps of model training and evaluation. To train and test the classifier, the data set was divided into training and testing subsets. In our study, due to the small number of patients, the leave-one-patient-out cross-validation scheme was used. One patient left-out (testing subset) was used to validate the classifier trained on the remaining patients (training set). This process was repeated until each patient had been validated once. The classification

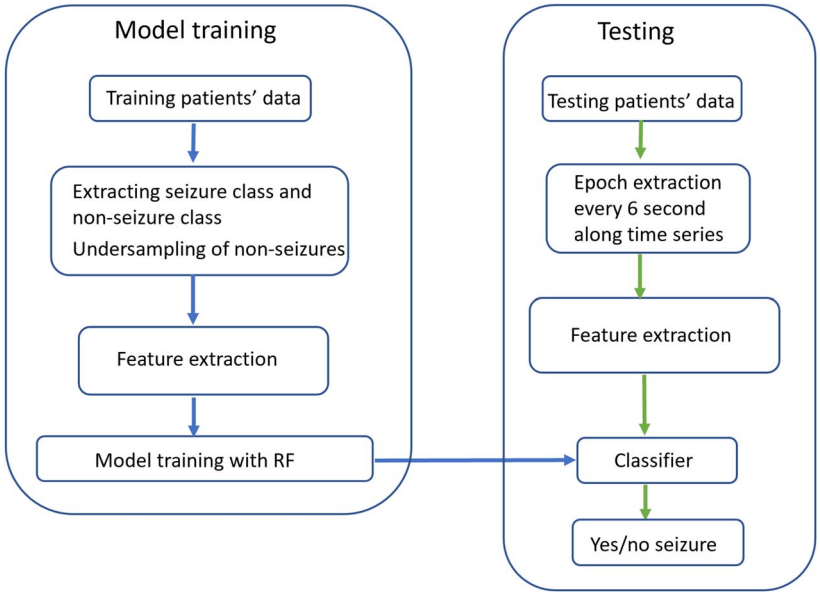

Figure 4. Block diagram of model training and evaluation.

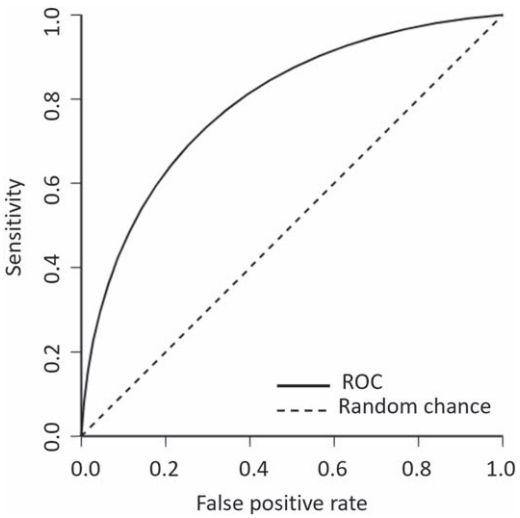

Figure 5. Illustration of a ROC-curve. [own creation].

performance was evaluated in terms of sensitivity, specificity, false positives per hour (FP/h), detection latency (DL) in seconds and the receiver operating characteristics curve (ROC). They can be expressed as:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{6}$$

$$\text{FalsePositiveRate} = \frac{FP}{TN + FP}, \tag{7}$$

$$\text{FP/h} = \frac{FP}{Recording\ time\ (in\ hours)}, \tag{8}$$

$$\text{DL} = T_{estimated} - T_{annotated}, \tag{9}$$

where TP is true positive, TN is true negative, FP is false positive, FN is false negative, $T_{estimated}$ is time of seizure detected with our model, and $T_{annotated}$ is time of seizure onset annotated by neurologists (ground truth). The ROC-curve is normally plotted by sensitivity against the false positive rate (FPR) at various thresholds. In figure 5, the concept of a ROC curve can be seen. As opposed to classic ROC-curve, the metric FP/h was used instead of the false positive rate in our study. This is a more tangible measure of a detector's proneness to produce false positives.

During the training process, as the seizure segments only accounted for approximately 1% of all extracted segments, an under-sampling of non-seizure segments was performed. This involved randomly removing non-seizure segments until a ratio of 9:1 between both classes was met. Testing on a patient was performed in a real-time manner with a sliding window approach. The size of the window was 6 seconds equal to the size used in the pre-processing step for extraction of training samples. The window was moved along the time series. For each epoch, the same filtering and feature extraction processes were applied followed by the classification. Afterwards, all detections occurring less than 30 seconds after each other were grouped together as one detection which is similar to what was done in [14].

3. Results

Figure 1 shows that seizures' duration varied greatly within PN10, PN13 and PN14. Table 2 shows the performance of seizure detection. It achieved an overall sensitivity of 81.43%. For 9 of the 14 patients, it managed to detect all seizures (100% sensitivity). However, sensitivities were significantly lower for, in particular, patients PN09, PN10, and PN14 which caused the overall reduction. The detector managed to

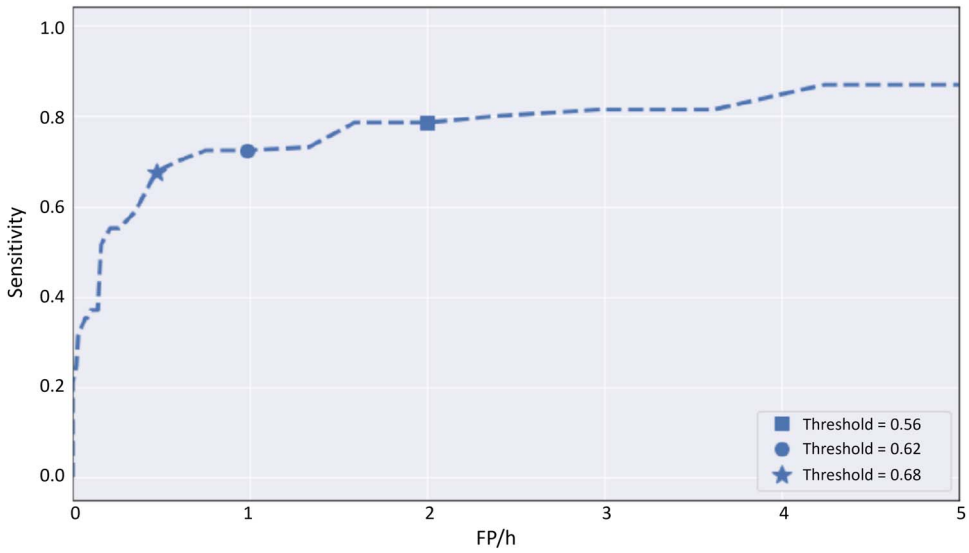

Figure 6. Sensitivity in relation to FP/h.

Table 2. Seizure detection performance.

Pat Id.	Sensitivity(%)	Specificity (%)	FP/h	DL (sec)
PN00	100	99.95	0.31	32.8
PN01	100	100	0	10.5
PN03	100	99.69	1.86	22
PN05	66.67	98.75	7.46	14.5
PN06	60	99.89	0.66	22
PN07	100	97.32	16.03	9
PN09	33	100	0	41
PN10	30	99.88	0.7	12.3
PN11	100	98.82	7.03	6
PN12	100	99.5	2.95	12.5
PN13	100	99.81	1.15	20
PN14	50	98.69	6.86	18
PN16	100	99.1	5.32	38
PN17	100	99.97	0.19	21.5
Avg.	81.43	99.38	3.61	20

detect the seizures with an average latency of 20 seconds. For many of the patients, the DL fell in the range of approximately (9-22) seconds. Lowest DL was 6 seconds from for PN11 and the highest was 41 seconds in PN09. In terms of specificity, it was close to 100% for all patients. The average was 99.38%. On average, the detector made 3.61 FP/h. For half of the patients the FP/h were actually below or close to 1. For PN07, however, 16.03 FP/h had a large negative effect on the average FP/h.

The performance of detection is illustrated in terms of its sensitivity in relation to FP/h in figure 6. Only thresholds yielding a maximum of 5 FP/h are showed. Comparing three different thresholds in the figure, the performance has lowest sensitivity but best FP/h with threshold of 0.68, while the performance has highest sensitivity but worse FP/h with threshold of 0.56.

4. Discussion

Machine learning methods have become increasingly popular in addressing complex healthcare challenges [35–39]. The complex data can be efficiently inspected and subtle patterns and correlations can be detected, which is impossible by humans. This study aims to improve the generalization of automatic seizure detection algorithm based on advanced machine learning technique to assist the neurologist for a better, faster and more objective evaluation of epilepsy.

The seizure detection was based on clinical data from open access Siena scalp EEG database. Patient PN10 was different from the others in terms of seizure type, seizure localization and lateralization. The purpose of the study was to develop an algorithm generalized well among patients. So we included all the patients in our study. In order to deal with patients'

variability, the features from both time-domain, frequency-domain and reflecting entropy were extracted. A combination of those features with random forest can help to reduce over-fitting, ensure less correlation among decision trees and thus improve accuracy. The data were unbalanced between non-seizure and seizure because seizure was a rare event. Non-seizure epochs and seizure epochs for training were extracted differently. In order to increase the seizure epochs, the seizure epochs extracted overlapped with each other during seizure period. The non-seizure epochs were extracted without overlap. In this way, we still have unbalanced epochs between seizure and non-seizure. For training a robust model, we further down-sampled the non-seizure epochs to have a better balance. Then the model was trained with those epochs. Afterwards, every 6 seconds segment of data from unseen patients was evaluated with the trained model to make a detection/classification. Final detections were based on post-processing procedure. Then we compared final detections with annotations to evaluate the performance of the model. The trained model could be used to new patients directly in clinic when the performance is good enough, and the model would mark the detections of seizures. Then neurologists would review those detections for diagnosis of epilepsy without spending time on reading the rest of EEG recording.

The detector achieved a specificity of 99.38%, sensitivity of 81.43% and 3.61 FP/h. In 5 out of all 14 patients, the detector failed to detect all seizures. Low sensitivity was obtained for PN10, PN09, and PN14. The poor sensitivity is not overly surprising when examining data of PN10 with 10 seizures. The mean seizure duration for this patient is 36.5 seconds, which is shorter than the mean seizure duration (61 seconds) across all patients. Moreover, the seizure duration was below 20 seconds in 5 out of 10, with 1 as short as 5 seconds (see figure 1). In addition, an average detection latency of 20 seconds was observed, which could explain that some of the shorter seizures weren't caught. In this study, some of the DLs were quite high with range (6-32.8 seconds). The purpose of our detection is not meant to be used for timely control of electrical stimulation and seizure alert systems or for admitting a patient for an immediate scanning. As the objective of this detector was to reduce neurologist's time spent on the manual revision of patient recordings, a small DL is of course convenient but less crucial. The most important is detection of seizure events with high sensitivity. So doctors are capable of locating seizure onset and ending in detected segments. There is a trade-off between sensitivity and FP/h. Figure 6 showed that the trade-off could be adjusted by changing threshold of probability for the given task.

The performance of machine learning based patient non-specific seizure detection has been reported in several studies in the literature. The research by Greene *et al* [17] showed that their non-specific EEG-based detection achieved a sensitivity of 80.41% and a FP/h of 3.42. It is quite similar to our performance. In another study by Sridevi *et al* [18], they extracted features from both the time- and frequency-domain and investigated five different classifiers. With their best-performing SVM classifier, they achieved the highest sensitivity of 80% and specificity of 86%. Our performance is better in terms of sensitivity and specificity. Lastly, the work by Orosco *et al* [15] presented a pediatric patient non-specific method using linear discriminant analysis(LDA). They achieved a specificity of 99.9%, sensitivity of 87.5% and a FP/h of 0.9. They used the CHB-MIT scalp EEG database which is an open access database and composed of EEG recordings from 22 pediatric patients aged 3-22 years old [16]. Our study was based on 14 patients aged 20-71 years old, which represents a wider population. In addition, a stepwise analysis based on the statistical parameter Lambda of Wilks was implemented to select features that best detect epileptic seizures in their study. Our approach without feature selection was less complicated and straightforward.

The paper contributes to improve performance of patient non-specific seizure detection. In our approach, combination of rich features with random forest improved the generalization of the model. The result on 9 patients with sensitivity of 100% and specificity of close-to 100% showed strong evidence for clinical application of our proposed method. In addition, the extracted features could be easily extended by other methods and are fed into random forest based classifier. Our approach does not need feature selection step, which makes our algorithm less complicated and requires less resources.

Our study has two limitations. First the sample size is relatively small, which limits its generalization for all patients with epilepsy. Second, the quality of the data needs to be further improved by dealing with artifacts from eye movement and muscle movement with advanced method. In the future, our method will be examined on a bigger database. The model might generalize well with more training data and lead to better performance for PN10 and PN14. In order to reduce false positives, post-processing after detection will be further investigated. In addition, noise detection and removal will be developed to improve the quality of EEG for features extraction.

5. Conclusions

This study proposed a novel approach for improving performance of patient non-specific seizure detection using scalp EEG. With rich features extracted from time domain, frequency domain and entropy and using random forest, the study obtained a specificity of 99.38%, sensitivity of 81.43% and 3.61 FP/h, which outperformed some previous studies. For clinical application, further investigations are needed to limit the number of false positives while still maintaining high sensitivity.

Acknowledgments

The study received support from Lundbeck Foundation.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://physionet.org/content/siena-scalp-eeg/1.0.0/, https://doi.org/10.13026/5d4a-j060.

ORCID iDs

Ying Gu https://orcid.org/0000-0003-4220-8902

References

[1] Zhou M, Tian C, Cao R, Wang B, Niu Y, Hu T, Guo H and Xiang J 2018 Epileptic seizure detection based on EEG signals and CNN *Frontiers in neuroinformatics* **12** 95
[2] Kerr M P 2012 The impact of epilepsy on patients' lives *Acta Neurol. Scand.* **126** 1–9
[3] Bandopadhyay R *et al* 2021 Recent developments in diagnosis of epilepsy: scope of microrna and technological advancements *Biology* **10** 1097
[4] Stafstrom C E and Carmant L 2015 Seizures and epilepsy: an overview for neuroscientists *Cold Spring Harbor Perspectives in Medicine* **5** 6
[5] Sriraam N, Raghu S, Tamanna K, Narayan L, Khanum M, Hegde A S and Kumar A B 2018 Automated epileptic seizures detection using multi-features and multilayer perceptron neural network *Brain Informatics* **5** 10
[6] Gotman J 1982 Automatic recognition of epileptic seizures in the EEG *Electroencephalogr. Clin. Neurophysiol.* **54** 530–40
[7] Gotman J, Flanagan D, Rosenblatt B, Bye A and Mizrahi E 1997 Evaluation of an automatic seizure detection method for the newborn EEG *Electroencephalogr. Clin. Neurophysiol.* **103** 363–9
[8] Nasehi S and Pourghassem H 2012 Seizure detection algorithms based on analysis of EEG and ECG signals: a survey *Neurophysiology* **44** 174–86
[9] Wu D, Wang Z, Jiang L, Dong F, Wu X, Wang S and Ding Y 2019 Automatic epileptic seizures joint detection algorithm based on improved multi-domain feature of cEEG and spike feature of aEEG *IEEE Access* **7** 41551–64
[10] Zhang Y, Yao S, Yang R, Liu X, Qiu W, Han L, Zhou W and Shang W 2022 Epileptic seizure detection based on bidirectional gated recurrent unit network *IEEE Trans. Neural Syst. Rehabil. Eng.* **30** 135–45
[11] Naganur V, Sivathamboo S, Chen Z, Kusmakar S, Antonic-Baker A, O'Brien T J and Kwan P 2022 Automated seizure detection with noninvasive wearable devices: A systematic review and meta-analysis *Epilepsia* **63** 1930–41
[12] Gotman J 1990 Automatic seizure detection: improvements and evaluation *Electroencephalogr. Clin. Neurophysiol.* **76** 317–24
[13] Shoeb A, Edwards H, Connolly J, Bourgeois B, Treves S T and Guttag J 2004 Patient-specific seizure onset detection *Epilepsy Behav.* **5** 483–98
[14] Hunyadi B, Signoretto M, Van Paesschen W, Suykens J A k, Van Huffel S and De Vos M 2012 Incorporating structural information from the multichannel EEG improves patient-specific seizure detection *Clinical Neurophysiology* **123** 2352–61
[15] Orosco L, Correa A G, Diez P and Laciar E 2016 Patient non-specific algorithm for seizures detection in scalp EEG *Comput. Biol. Med.* **71** 128–34
[16] Goldberger A L, Amaral L A, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals *Circulation* **101** e215–20
[17] Greene B R, Boylan G B, Reilly R B, de Chazal P and Connolly S 2007 Combination of EEG and ECG for improved automatic neonatal seizure detection *Clinical Neurophysiology* **118** 1348–59
[18] Sridevi V, Reddy M R, Srinivasan K, Radhakrishnan K, Rathore C and Nayak D S 2019 Improved patient-independent system for detection of electrical onset of seizures *Journal of Clinical Neurophysiology* **36** 14
[19] Sánchez-Hernández S E, Salido-Ruiz R A, Torres-Ramos S and Román-Godínez I 2022 Evaluation of feature selection methods for classification of epileptic seizure EEG signals *Sensors* **22** 3066
[20] Detti P 2020 Siena scalp EEG database (version 1.0.0) *PhysioNet.*
[21] Detti P, Vatti G and Zabalo Manrique de Lara G 2020 EEG synchronization analysis for seizure prediction: A study on data of noninvasive recordings *Processes* **8** 846
[22] Shoeb A H and Guttag J V 2010 Application of machine learning to epileptic seizure detection *Proceedings of the XXVII international conference on machine learning (ICML-10)* pp 975–82
[23] Thangavel P *et al* 2022 Improving automated diagnosis of epilepsy from EEGs beyond IEDs *J. Neural Eng.* **19** 066017
[24] Mesbah M, Balakrishnan M, Colditz P B and Boashash B 2012 Automatic seizure detection based on the combination of newborn multi-channel eeg and hrv information *EURASIP Journal on Advances in Signal Processing* **2012** 1–14
[25] Fergus P, Hussain A, Hignett D, Al-Jumeily D, Abdel-Aziz K and Hamdan H 2016 A machine learning system for automated whole-brain seizure detection *Applied Computing and Informatics* **12** 70–89
[26] Kokoska S and Zwillinger D 2000 *CRC Standard Probability and Statistics Tables and Formulae* (CRC Press)
[27] Hjorth B 1970 EEG analysis based on time domain properties *Electroencephalogr. Clin. Neurophysiol.* **29** 306–10
[28] Cocconcelli M, Strozzi M, Molano J C C and Rubini R 2022 Detectivity: a combination of hjorth's parameters for condition monitoring of ball bearings *Mech. Syst. Sig. Process.* **164** 108247
[29] Welch P 1967 The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms *IEEE Trans. Audio Electroacoust.* **15** 70–3
[30] Herlau T, Schmidt M N and Mørup M 2016 Introduction to machine learning and data mining *Lecture Notes of the Course of The Same Name Given at DTU* (Technical University of Denmark)
[31] Inouye T, Shinosaki K, Sakamoto H, Toi S, Ukai S, Iyama A, Katsuda Y and Hirano M 1991 Quantification of EEG irregularity by use of the entropy of the power spectrum *Electroencephalogr. Clin. Neurophysiol.* **79** 204–10
[32] Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy *American Journal of Physiology-Heart and Circulatory Physiology* **278** H2039–49
[33] Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
[34] Attneave F 1963 *Earl B. Hunt. Concept Learning: An Information Processing Problem* vol 1962 (Wiley)
[35] Lipton Z C, Kale D C, Elkan C and Wetzel R 2015 Learning to diagnose with lstm recurrent neural networks arXiv:1511.03677
[36] Wahnoun R, Saigal R, Gu Y, Paquet N, DePauw S, Chen A C N, Ahmed-Khalid S S and Nielsen K D 2002 A real-time brain-computer interface based on steady-state visual evoked potentials *VII Annual Conference of the International Functional Electrical Stimulation Society, IFESS 2002 (Ljubljana, Slovenia, 25-29 June 2002)* pp 161–3

[37] Gu Y, Rasmussen S M, Mølgaard J, Haahr-Raunkjær C, Meyhoff C S, Aasvang E K and Sørensen H B 2021 Prediction of severe adverse event from vital signs for post-operative patients *2021 XLIII Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) pp 971–4

[38] Kristinsson Æ Ö, Gu Y, Rasmussen S M, Mølgaard J, Haahr-Raunkjær C, Meyhoff C S, Aasvang E K and Sørensen H B 2022 Prediction of serious outcomes based on continuous vital sign monitoring of high-risk patients *Comput. Biol. Med.* **147** 105559

[39] Prasad S, Tan Z H, Prasad R, Cabrera A F, Gu Y and Dremstrup K 2011 Feature selection strategy for classification of single-trial EEG elicited by motor imagery *2011 The XIV International Symposium on Wireless Personal Multimedia Communications (WPMC)* (IEEE) pp 1–4