Nonconvulsive Epileptic Seizure Monitoring with Incremental Learning

Yissel Rodríguez Aldana[1,2], Enrique J. Marañón Reyes[1], Frank Sanabria Macias[3], Valia Rodríguez Rodríguez[4,5,6], Lilia Morales Chacón[7], Sabine Van Huffel[2], and Borbála Hunyadi[2,8]

[1]*Universidad de Oriente, Center of Neuroscience and Signals and Image Processing. Santiago de Cuba, Cuba. email:yraldana@gmail.com*

[2]*KU Leuven, Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics. Leuven, Belgium.*

[3]*Escuela Politécnica Superior, Universidad de Alcalá, Alcalá de Henares, Spain*

[4]*Aston University. Birmingham, United Kingdom.*

[5]*Cuban Neuroscience Center. Havana, Cuba.*

[6]*Clinical-Surgical Hospital "Hermanos Almeijeiras". Havana, Cuba*

[7]*International Center of Neurological Restoration. Havana, Cuba.*

[8]*Department of microelectronics, Delft University of Technology. Delft, Netherlands*

September 2, 2019

1 Abstract

Nonconvulsive epileptic seizures (NCSz) and nonconvulsive status epilepticus (NCSE) are two neurological entities associated with increment in morbidity and mortality in critically ill patients. In a previous work, we introduced a method which accurately detected NCSz in EEG data (referred here as 'Batch method'). However, this approach was less effective when the EEG features identified at the beginning of the recording changed over time. Such pattern drift is an issue that causes failures of automated seizure detection methods. This paper presents a support vector machine (SVM)-based incremental learning method for NCSz detection that for the first time addresses the seizure evolution in EEG records from patients with epileptic disorders and from ICU having NCSz. To implement the incremental learning SVM, three methodologies are tested. These approaches differ in the way they reduce the set of

This work has been supported by the Belgian foreign Affairs-Development Cooperation through VLIR-UOS (2013-2019) (Flemish Interuniversity Council-University Cooperation for Development) in the context of the Institutional University Cooperation program with Universidad de Oriente.

The research leading to these results has received funding from imec funds 2017 and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no.339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

potentially available support vectors that are used to build the decision function of the classifier. To evaluate the suitability of the three incremental learning approaches proposed here for NCSz detection, first, a comparative study between the three methods is performed. Secondly, the incremental learning approach with the best performance is compared with the Batch method and three other batch methods from the literature. From this comparison, the incremental learning method based on maximum relevance minimum redundancy (MRMR_IL) obtained the best results. MRMR_IL method proved to be an effective tool for NCSz detection in a real-time setting, achieving sensitivity and accuracy values above 99%.

Keywords Nonconvulsive epileptic seizures, Hilbert Huang Transform, Multiway Data Analysis, Incremental Learning.

2 Introduction

Nonconvulsive epileptic seizures (NCSz) and nonconvulsive status epilepticus (NCSE) are two related neurological entities that are frequently found in critically ill patients [1, 2]. Despite the nonconvulsive nature, they are associated with increment in morbidity and mortality in the intensive care unit (ICU). Since NCSz/NCSE present subtle or no overt clinical signs, it is not uncommon that in patients with altered mental status or coma they remain unnoticed and untreated for long periods of time. Studies carried out in this population have reported though that NCSE lasting more than 10 hrs are associated with permanent disabilities, while mortality is very high in NCSE lasting more than 20 hrs [3].

When suspected, NCSz/NCSE diagnosis is carried out using continuous EEG (cEEG) monitoring. However, several studies have reported that the likelihood to detect the first NCSz increases in patients at risk (i.e.: comatose patients and children) when EEG is recorded for more than 24 hours [4]. Hence, seizure detection in ICU could be an exhausting and time-consuming process. To assist on the visual identification of changes, quantitative trends summarizing EEG amplitude and frequency composition as well as annotating the presence of seizures have been recently introduced to the continuous EEG monitoring technique.

Previous algorithms developed for NCSz detection combined wavelet analysis [5, 6, 7, 8], entropy [9, 10], nonlinear parameters [6, 9, 11], statistical and spectral features of the EEG [9, 10, 11, 12, 13, 14] with various machine learning techniques [8, 9, 10, 11, 13, 15] or thresholds [6, 7, 12, 14] to detect the NCSzs. These algorithms obtained a reasonable sensitivity (over 90% in most cases) during the test process [16]. Among the methods with better results in the context of NCSz detection in patients with epileptic etiology are those proposed by Kollialil at al (2013) [9], Sharma et al (2014) [8] and Fatma et al (2016) [14].

Figures 1, 2 and 3 display the block diagrams of Kollialil's, Sharma's and Fatma's methods respectively. As can be appreciated, the cited methods iterate over the EEG channels at least until the feature extraction step. This means the features are extracted from a single channel without considering important characteristics of the seizure as the synchronization and spread out/in of the seizure activity over the EEG channels [17]. These methods intended to exploit the possible cross information of the channels by combining the features computed individually into one classifier, as in Kollialil's method, or imposing hard thresholds, as in Sharma's and Fatma's methods.

Kollialil et al. proposed a patient independent training for a linear SVM. The NCSz characteristics vary enormously across patients. This implies that the number of patterns used to train a patient independent classifier must

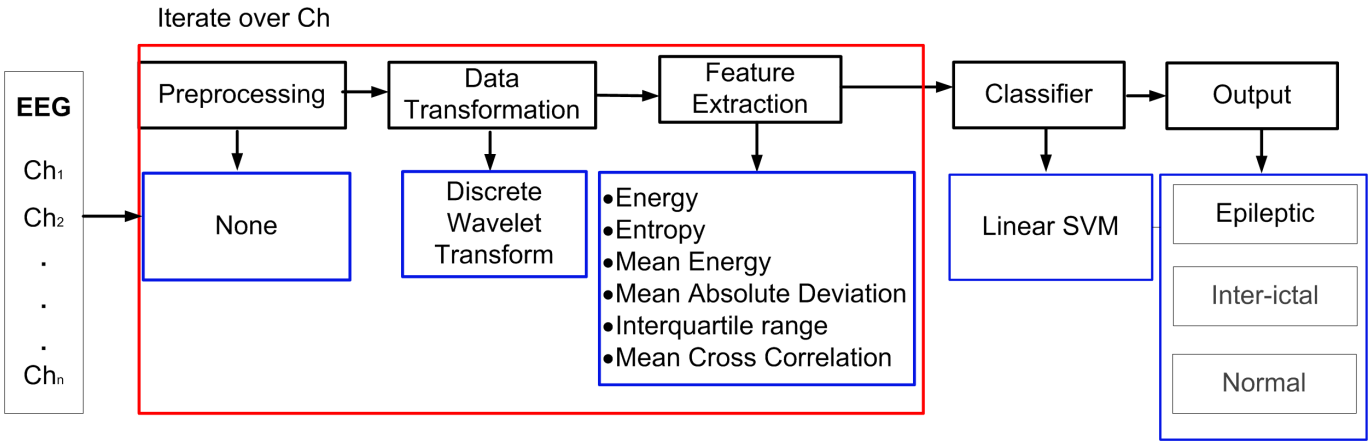


Figure 1: Block diagram of the method proposed by Kollialil et al. [9] absence seizure detection. The proposal uses a multiclass SVM and a single feature vector computed from the fifth detail wavelet coefficients. Features as energy, mean energy, entropy, mean cross-correlation, mean curve length, the coefficient of variation, interquartile range (IQR) and median absolute deviation (MAD) are compared to obtain an optimal single feature. The data for this experiment consisted of normal, epileptic and interictal EEG data from 100 subjects, from a reputed Neurology Clinic. The best performances were found for the energy, entropy, MAD, and IQR with an accuracy above 95%. The best feature was the IQR with an accuracy value of 99.66 %.

be quite large. Given the nonconvulsive nature of NCSz the existent databases are small. With such databases it is not likely to successfully train a classifier capable to generalize the acquired knowledge to other NCSz data, especially in case the data originate from ICU patients. Taking into account this characteristic of the NCSz, Sharma et al. and Fatma et al. proposed patient specific methods. However, their approaches ignore the fact that the EEG patterns present in a specific record could also differ. Having this in mind, the methods proposed by Sharma et al. and Fatma et al. cannot guarantee to maintain their performance in longer records where these changes are more likely to occur.

There are two main drawbacks in the methods proposed in the literature for NCSz detection. First, they employ a patient independent training of the classifiers and, second, thresholds for the detection are arbitrarily set. In general, the duration of the seizure and the number of channels displaying seizure activity are the most popular thresholding criteria [7, 15, 13, 8]. If the seizure is too short in time or affects just a few channels it is not detected. Furthermore, NCSz characteristics vary across patients. EEG patterns present in a specific record could also differ depending on the patient disease's etiology. Therefore, a threshold or classifier which works for one patient will not necessarily work for another. Additionally, a more meaningful description of the seizure's spatial localization should be considered, for instance its whole head topography instead of its distribution in a limited number of channels.

In [18] we proposed a patient-specific method that mitigates the need of thresholds to detect the NCSz. This method identifies the NCSz by exploiting the similarity between the first NCSz detected by the physician on the EEG and the rest of the NCSz in the recording [19].The method expands the EEG using the Hilbert Huang Transform (HHT) into a third-order tensor. This multiway representation of the data exploits the EEG high-dimensional structure by analyzing its spectral, temporal and spatial properties simultaneously. This is a fundamental difference compared to the methods of Sharma [8] and Fatma [14] : in our approach, multichannel information is integrated at the

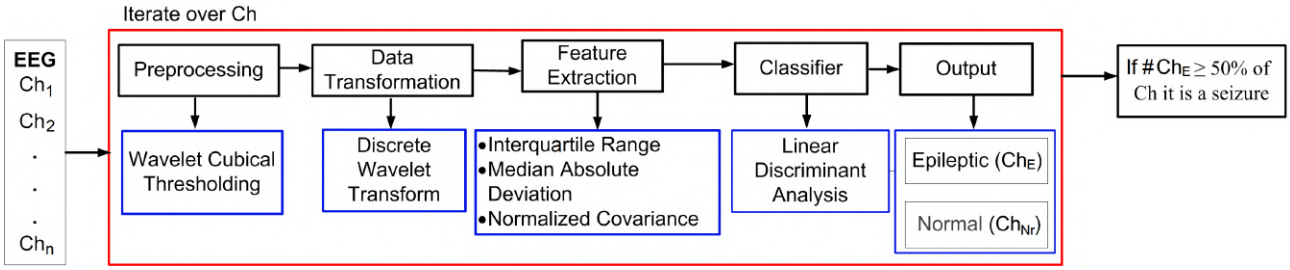


Figure 2: Block diagram of the patient specific method proposed by Sharma et al. [8] for nonconvulsive epileptic seizures detection (♯ denotes "number of "). The proposed algorithm analyzed the EEG in epochs of 1-second duration. Each epoch was denoised using wavelet analysis applying cubic thresholding. The extracted features are the IQR, the MAD, and the Normalized Covariance, normalized by the median of their background EEG features. The dataset used for testing consisted of 24 seizures recorded from the EEG of 9 subjects in the All India Institute of Medical Sciences. This method requires the seizure activity to be present in at least 50% of the channels. Otherwise, the seizures will be missed. The method reported 100% of sensitivity and 99.3% of specificity.

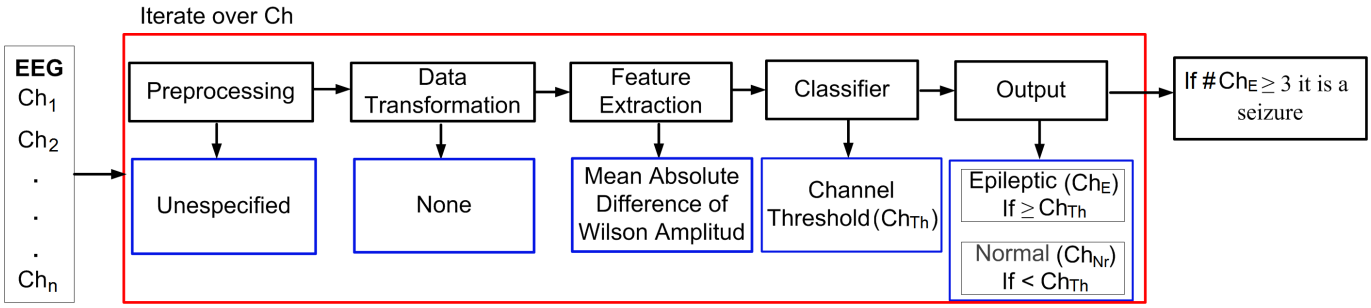


Figure 3: Block diagram of the patient specific method proposed by Fatma et al. [11] for nonconvulsive epileptic seizures detection (♯ denotes "number of "). The algorithm analyzes the EEG in epochs of 1-second duration from which the mean absolute difference is computed. To differentiate seizure from normal EEG, the method uses thresholds over the computed parameter. The method reported sensitivity and specificity of 100% and 99.21% respectively.

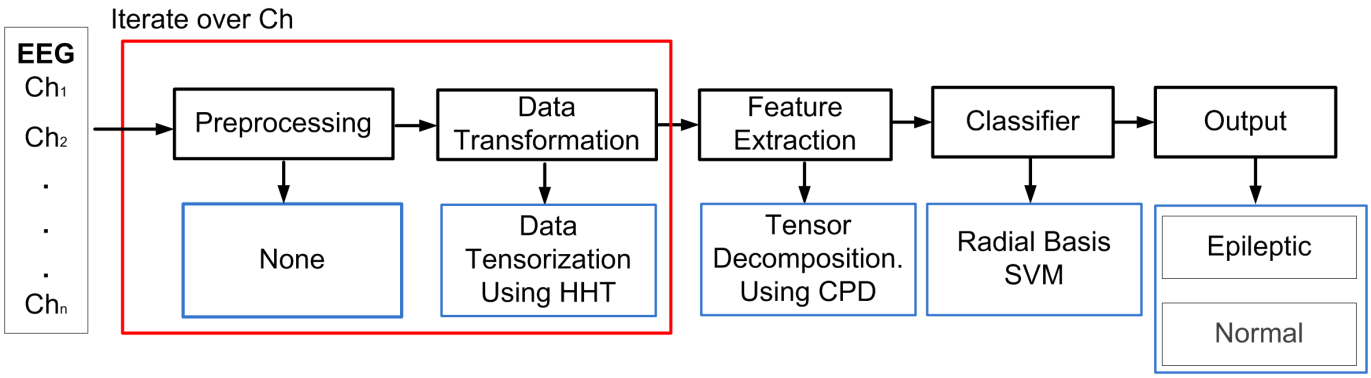


Figure 4: . General block diagram of the patient specific method proposed by Rodríguez et al. [18] for nonconvulsive epileptic seizures. This method identifies the NCSz by exploiting the similarity between the first NCSz detected by the physician on the EEG and the rest of the NCSz in the recording. The explored features are obtained by means of a multiway analysis of the EEG signal represented as a third-order tensor $X \in \mathbb{R}^{(F \times T \times Ch)}$ with modes $frequency \times time \times channels$. The tensors are computed by expanding EEG segments of 3 s duration using Hilbert Huang Transform (HHT). The tensor decomposition is performed with Canonical Polyadic Decomposition (CPD). The method uses the spatial component from the CPD as features of a SVM to discriminate between seizure and non-seizure segments.

level of feature extraction via a tensor decomposition, as opposed to performing separate feature extraction per channel. There is evidence that exploiting the (multi)-linear structure inherently present in multichannel EEG achieves superior performance compared to methods that ignore such structure [20, 21].The spatial component extracted from this multiway EEG representation with canonical polyadic decomposition (CPD) accurately characterized the seizure pattern. The algorithm achieved average sensitivity and specificity values over 98%. Figure 4 shows the block diagram of the method proposed in [18]. However, this approach also disregarded the seizure pattern changes within an EEG record. It showed to be less accurate for records where the EEG morphology evolved over time and the morphological characteristics varied with respect to the beginning of the record.

Nonconvulsive epileptic seizures and status in epileptic and critically ill patients present a different temporal and morphological evolution. It is common, for instance, that seizures in a critically ill patient develop from an ictal-interictal continuum composed of periodic discharges or rhythmic activity, waxing and waning over long periods of time, or fluctuations from low amplitude and/or low frequency seizures. These phenomena are the reason why seizure detection algorithms developed for epilepsy studies fail in the critical care context [22, 23, 24]. Trained and experienced medical specialists are able to identify epileptic seizure EEG patterns in complex scenarios, e.g. when the background EEG resembles ictal activity. Similarly, automatic seizure detection methods need appropriate training to solve a complex signal processing and pattern recognition task.

Several authors addressed this phenomenon in methods proposed for convulsive epileptic seizure detection allowing the user to control the classification process using thresholds or tuning some parameters [25, 26, 27]. However, these algorithms are not able to learn new seizure patterns that may appear on the EEG. Other authors [28, 29] used online learning techniques in their proposals to incorporate novel patterns to be added to the already known ones by the classifier. None of the methods proposed for NCSz addressed this issue.

The most commonly used training strategy for a machine learning algorithm is the batch method [18, 8, 9, 14, 30]. In the batch method, the algorithm has a fixed number of samples that are used to train a classifier. The trained classifier is then applied to new samples without further updating. Most classifiers are trained with batches of data coming from several patients. However, patient-specific solutions are expected to perform much better. In a better solution, the patients' EEG should be recorded for some time to train the classifier. As new information is continuously becoming available, and the patients' seizures may evolve in morphology, the classifier must be updated online using the latest EEG data. However, simply including all new incoming data is not practical as it would continuously increase memory and computational requirements.

This paper presents an SVM-based incremental learning method for NCSz detection that for the first time addresses the seizure evolution in EEG records from patients with epileptic disorders and from ICU having NCSz. The method proposed here, provides improved performance compared to the method introduced in [18] (referred here as 'Batch method'), while maintaining similar memory requirements. To introduce the incremental learning step in the Batch method, three methodologies are tested. These approaches differ in the way they reduce the set of potentially available support vectors (i.e. training samples) that are used to build the decision function of the classifier: 1) Discard a fixed number of support vectors after the classifier retraining, based on a predefined threshold. From now referred as 'Hard_IL', 2) Select an optimal support vector subset using cross-validation. From now referred as 'Cross_IL', and 3) Select an optimal support vector subset using an approximate technique for incremental SVM proposed by Yang [31] based on the maximum relevance minimum redundancy (MRMR) feature selection method [32]. From now referred as 'MRMR_IL'. To our knowledge this is the first time Yang's algorithm is used for a practical implementation. The MRMR_IL approach provides an algorithm to select appropriately the patterns to update the classifier, avoiding to incorporate random patterns as done in [28] and not requiring the user intervention in the updating process as in [29]. To evaluate the suitability of the three incremental learning approaches proposed here for NCSz detection, first, a comparative study between the three methods is performed. Secondly, the incremental learning approach with the best performance is compared with the Batch method and three other batch methods from the literature [8, 9, 14] that reported better performance than the Batch method in some sense (for the results of this comparison, see [18]).

3 Materials and methods

3.1 EEG data

The EEG data were collected at the Epilepsy Unit of the Cuban International Neurological Restoration Center (CIREN) and the ICU of the Clinical Surgical Hospital "Hermanos Ameijeiras ", both in Havana City. A video-telemetry EEG (vEEG) study was performed on all patients. For clarity of exposition, we will denote the patients recorded at the Epilepsy Unit as vEEG, and the ones recorded at the ICU, as ICU. The dataset comprised EEG clips of about $14min$ to more than $21h$ (mean $280min\ 21s$) of 14 patients with ages between 18 and 57 years and different brain disorders leading to NCSz. The visual inspection and seizure labeling were performed by two pairs of neurophysiologists (including the authors VRR and LMC). All recordings were re-analyzed for the purpose of this study. Each pair of neurophysiologists independently labeled the ICU or vEEG seizures. In case of disagreement it was resolved by discussion. A total of 117 NCSz were identified (55/117 were associated to coma or other acute brain dysfunction).

Table 1 presents a more detailed description of the dataset. The data were anonymized before their use in this study. All procedures were reviewed and approved by the Ethical Committees of the CIREN and "Hermanos Ameijeiras" Hospital respectively.

Table 1: Description of the EEG Database. In the table 'x' and 'o' indicate whether the signal was recorded or not together with the EEG for each patient. vEEG stands for video-telemetry EEG. ICU stands for continuous EEG recorded in the intensive care unit.

Patient data				*Recording*	*Protocol*				
Patient	*Gender*	*Diagnosis*	*Type*	*Seizures*	*Channels*	*EKG*	*EOG*	*EMG*	*Video*
1	*M*	*Temporal lobe epilepsy*	*vEEG*	*6*	*19*	*o*	*o*	*o*	*x*
2	*M*	*Temporal lobe epilepsy*	*vEEG*	*3*	*19*	*o*	*o*	*o*	*x*
3	*M*	*Temporal lobe epilepsy*	*vEEG*	*13*	*19*	*o*	*o*	*o*	*x*
4	*F*	*Temporal lobe epilepsy*	*vEEG*	*5*	*19*	*o*	*o*	*o*	*x*
5	*F*	*Lennox-Gastaut syndrome*	*vEEG*	*2*	*19*	*o*	*o*	*o*	*x*
6	*F*	*Temporal lobe epilepsy*	*vEEG*	*2*	*19*	*o*	*o*	*o*	*x*
7	*F*	*Temporal lobe epilepsy*	*vEEG*	*12*	*17*	*x*	*x*	*x*	*x*
8	*F*	*Juvenil myoclonic epilepsy*	*vEEG*	*6*	*17*	*x*	*x*	*x*	*x*
9	*M*	*Frontal lobe epilepsy*	*vEEG*	*34*	*17*	*x*	*x*	*x*	*x*
10	*F*	*Coma /Subaracnoid Haemorrhage*	*ICU*	*40*	*19*	*x*	*x*	*x*	*x*
11	*M*	*Myoclonic Seizures /Brain Tumor*	*ICU*	*2*	*13*	*x*	*o*	*x*	*x*
12	*F*	*Coma/ Systemic Vasculitis*	*ICU*	*6*	*8*	*x*	*x*	*x*	*x*
13	*F*	*Seizures/Brain Tumor /Sepsis*	*ICU*	*5*	*19*	*x*	*x*	*x*	*x*
14	*F*	*Generalized Tonic-Clonic Seizures*	*ICU*	*3*	*14*	*x*	*x*	*o*	*x*

3.2 Batch method

The Batch method in [18] analyses the EEG data in non-overlapping segments (epochs) of 3 seconds long. All epochs are expanded in the time-frequency domain using an Hilbert Huang Transform (HHT). A 3^{rd} order tensor is built from every epoch with modes $frequency \times time \times channels$.

The tensors built in this way are decomposed using a canonical polyadic decomposition (CPD) [33] with rank one. In other words, we model the EEG data as the outer product of three vectors, that describe the signature of the EEG in time, frequency and across channels. The values of these signature vectors can be used as features for classification, as we will describe below. From now on, we will refer to the channel mode vector as 'spatial signature'. In [18] we have shown that the spatial signature is a powerful feature to discriminate between NCSz and NCSz-free (n-NCSz) epochs. Therefore, we will use only the spatial signature in this study.

To discriminate between the two classes NCSz and n-NCSz (seizure and seizure free) a support vector machine (SVM) classifier is used. Given a set of S training data $\{\mathbf{x}_s\}_{s=1}^{S}$, $\mathbf{x}_s \in \mathbb{R}^{Ch\times 1}$ (Ch is the number of channels) with labels $\{y_s\}_{s=1}^{S} \in \{\pm 1\}$ (seizure or seizure-free), the SVM attempts to infer a model M_0 that correctly estimates the

labels $\hat{y}_{new}$ of a new test vector $\mathbf{x}_{new}$ based on a function of the form

$$\hat{y}_{new} = \text{sign}[\mathbf{w}^T \phi(\mathbf{x}_{new}) + w_0] \tag{1}$$

where $\mathbf{w}$, a set of weights and ϕ is a nonlinear transformation that maps the input data to a higher dimensional feature space. The objective of the SVM formulation is to construct a separating hyperplane in the feature space with maximal margin. This can be translated to a convex optimization problem. In the dual space, the classifier takes the form

$$\hat{y}_{new} = \text{sign} \sum_{s=1}^{S} \alpha_s y_s k(\mathbf{x}_s, \mathbf{x}_{new}) + b \tag{2}$$

where $k(\mathbf{x}_s, \mathbf{x}_{new}) = \phi(\mathbf{x}_{new})\phi(\mathbf{x}_s)$ is a symmetric and positive definite kernel function that defines the inner product of $\mathbf{x}_{new}$ and $\mathbf{x}_s$ in the higher dimensional space. Here, we use a Gaussian kernel. The $\mathbf{x}_s$ input vectors corresponding to non-zero α_k values are called support vectors.

As described in [18], we propose to start the NCSz monitoring after the clinicians identify the first epileptic seizure. The duration of the first seizure determines the number of NCSz epochs used for the training. The same number of n-NCSz epochs are selected as non-seizure training points, starting from the beginning of the first seizure and going back towards the beginning of the recording. In other words, if the first seizure is of length L 'epochs', then $S = 2\lfloor L/3 \rfloor$.

3.3 Incremental learning

3.3.1 Training rounds

The batch method uses a fixed model M_0 throughout the whole duration of the monitoring, that is trained based on the data up to the first seizure. As opposed to this, the purpose of incremental learning is to regularly update the model in order to ensure adaptability to the evoling EEG morphology within the same patient. We propose to update the model after regular time intervals of duration t using the EEG data collected during this time period. The value of t is selected arbitrarily as $t = 10min$ for short EEG recordings ($< 2h$) and $t = 2h$ for longer recordings ($> 2h$). These values were chosen based on evaluating the real chance of observing morphology changes in the EEG. The clinicians establish the recording time for a patient by considering how long it would take to register an epileptic event given the patient etiology or clinical state. It is assumed in this approach that the duration of the EEG will depend on when the clinician is expecting to see the EEG changes. That is, in short recordings we expect that EEG changes will develop sooner than in longer recordings. In practice, the parameter t can be specified by the specialist at the start of the EEG monitoring. The data collected during this time is defined as a training round.

3.3.2 Double labeling of training data

After t time has elapsed since the last seizure, it is time to obtain the new model M_t including the newly collected datapoints $x_{n_{new}}$ and their corresponding labels $y_{n_{new}}$ in the existing training dataset. However, in a practical setting, no expert is available to provide labels to the new datapoints. To overcome this problem and obtain training labels for computing the new model M_t, we will use the labels that are predicted by M_{t-1} (see Fig. 5 for the definition of

the notation used). However, the labels estimated by M_{t-1} could be erroneous. If this is the case, the error could be propagated into the classifier during the updating process, leading to a so-called concept drift [34]. Concept drifts are problematic since they lead to conflicts in the classification. The classifier performance will decrease until the model can be updated appropriately. To reduce the chances of generating a concept drift with the labels provided by the model M_{t-1}, a double set of labels are predicted, one provided by M_{t-1}, and additionally, another set using the partial least squares method (PLS) [35].

The PLS prediction model is trained with the same x_s samples used to estimate the model M_{t-1} and their corresponding y_s labels provided by the neurophysiologist. Then, the trained PLS model is used to predict the $y_{n_{new}}^{[PLS]}$ labels of the $x_{n_{new}}$ samples. In order to verify the suitability of the PLS method within this context, we first tested its performance using a batch approach, prior to applying it for the double labeling in the incremental learning setting. The PLS model was tested on a set of 14 EEG recordings and showed a positive predictive value (PPV) of 98.9% and a sensitivity of 97.6%. We describe the details of this study in [36].

Due to the double labeling, two sets of labels are available for $x_{n_{new}}$: $y_{n_{new}}$ provided by the model M_{t-1} and $y_{n_{new}}^{[PLS]}$ provided by PLS. The samples $x_{n_{sel}}$ for which both methods estimate the same label ($y_{n_{new}} = y_{n_{new}}^{[PLS]}$) are selected for estimating M_t.

3.3.3 Conservative updating

The selected samples $x_{n_{sel}}$ are split into training and validation sets. The updating is then performed in two steps. First, a temporary model M_{temp} is estimated only using the selected training samples. Secondly, the support vectors (SV) from M_{t-1} and M_{temp} are combined together. An optimal subset of the combined SVs are used to build the model M_{op}, that is tested on the validation set (An explanation on why and how to choose the optimal subset will follow below in section 3.3.4). Then, the model M_{t-1} is also tested on the validation set. Finally, after comparing M_{t-1} and M_{op}, the one with the better performance is assigned to be the new M_t model. The same updating procedure is performed after every training round.

3.3.4 Controlling model growth

Following the procedure above without the selection of an optimal subset of SVs, the size of the model, defined by the number of support vectors (SVs), would grow with the number of new training samples after each update.

Therefore, it is necessary to take some actions to limit the number of SVs in the solution. A regularization factor $\phi = N_{sv_{M_t}}/N_{sv}$ is introduced to limit the growth of the number of SVs, $\phi \in \mathbb{R}\colon 0 < \phi \leq 1$. $N_{sv_{M_t}}$ denotes the number of SVs from the model M_t, while N_{sv} denotes the sum of the SV from M_{temp} and M_{t-1}. Finally, the size of the SV subset is computed as $SV_{sub} = \phi N_{sv}$.

The three methodologies, defined in the introduction as Hard_IL, Cross_IL, and MRMR_IL, are tested, to select the optimal subset of SV. For all methodologies the new samples are divided into training and test set. A new model M_{temp} is trained with the training set. The SV from the models M_{t-1} and M_{temp} will become candidate SVs for the new model M_{op}.

To control the model growth, **Hard_IL** accumulates the SV of a number of training rounds, tr (user-specified parameter). Then, after tr training rounds, at every new training round, the oldest $N_{sv} - \phi N_{sv}$ SV are discarded to

obtain the SV subset for the model M_{op}. For this implementation, $tr = 2$ was selected. For $tr > 2$ the execution time increases without significant improvements in the algorithm accuracy. Fig. 5 (B) describes the Hard_IL algorithm.

For **Cross_IL** given the SV from the models M_{t-1} and M_{temp} a 5-fold (of length ϕN_{sv}) cross-validation is applied to find the optimal subset of SVs. This means that the set of all available SV is shuffled into five different subsets of size ϕN_{sv}. Hence, the performance of the five subsets is tested, and the subset with the best performance will become M_{op}. Fig. 5 (C) illustrates the diagram of the Cross_IL algorithm.

The **MRMR_IL** approach implements the incremental SVM algorithm proposed in [31] to estimate the optimal subset of SVs. In [31] the selection of the subset of SVs is formulated as equivalent to a feature selection problem. The equivalence can be observed if the SVM decision function in the dual space is written as

$$f(x) = \boldsymbol{\omega}^T \mathbf{K} + \boldsymbol{\omega_0}^T \tag{3}$$

where $\boldsymbol{\omega} = [\alpha_1 y_1, \alpha_2 y_2 \ldots, \alpha_{N_{sv}} y_{N_{sv}}]$ is the weight vector, $\mathbf{K} \in \mathbb{R}^{N_{sv} \times N_{sv}}$ is the kernel matrix, with the element in the i^{th} row and j^{th} column $K_{i,j} = k(x_i, x_j)$. N_{sv} is the number of SVs and $\boldsymbol{\omega}_0 \in \mathbb{R}^{1 \times N_{SV}}$, is a bias vector containing the bias term ω_0 in each element. The decisions are made according to sign$(f(x))$. This equation is very similar to the decision function of a simple linear classifier

$$g(x) = \boldsymbol{\beta}^T \mathbf{X} + \boldsymbol{\beta}_0^T \tag{4}$$

where $\mathbf{X} \in \mathbb{R}^{L \times M}$ is a data matrix with L number of features describing M data samples, $\boldsymbol{\beta} \in \mathbb{R}^{L \times 1}$ the corresponding weight vector and $\boldsymbol{\beta}_0 \in \mathbb{R}^{M \times 1}$ is the vector containing the bias term β_0 in each element.

In the context of feature selection, the aim is to drop some of the features (i.e. the rows of the data matrix $\mathbf{X}$) and the corresponding weights, but at the same time achieving correct decisions using $g(x)$.

By exploiting the similarities between (3) and (4) the kernel matrix $\mathbf{K} \in \mathbb{R}^{N_{sv} \times N_{sv}}$ can be interpreted as a data matrix in the feature selection context where each column of $\mathbf{K}$ corresponds to a data sample and each row of $\mathbf{K}$ (i.e. support vector) corresponds to a feature. Hence, the number of SVs can be reduced by dropping rows in the kernel matrix $\mathbf{K}$, while keeping the number of columns unchanged [31]. To select the best subset of SVs, Yang proposed the feature selection technique MRMR introduced in [32].

Following the MRMR scheme to find the optimal subset of support vectors the first step is to select the row of $\mathbf{K}$ with the highest F-statistic F_i defined as [31],

$$F_i = \frac{[\sum_{c=1}^{2} n_c (\bar{K}_{i.}^c - \bar{K}_{i.})]}{\sigma^2} \tag{5}$$

where $K_{i.}$ denotes the i^{th} row of the feature matrix $\mathbf{K}$, n_c is the number of samples from the c^{th} class. $\bar{K}_{i.}$ is the mean value of the row $K_{i.}$ and $\bar{K}_{i.}^c$ is the mean value of $K_{i.}$ within the c^{th} class. σ^2 is the pooled variance defined as [31],

$$\sigma^2 = \frac{[\sum_c (n_c - 1)\sigma_c^2]}{\hat{N}_{sv} - 2} \tag{6}$$

where σ_c^2 is the variance of $K_{i.}$ within the class c.

The row with the highest F_i is the first element of the final subset. At each iteration, the rest of the unselected support vectors are evaluated with (5), and the one with the largest F_i is added to the subset. The subset is evaluated according to the relevance and the redundancy defined as [31],

$$\max_{c} R_F \quad \text{with} \quad R_F = \frac{1}{|\hat{N}_{sv}|^2} \sum_{i} F(i,c) \tag{7}$$

and

$$\min_{K \in \hat{N}_{sv}} R_{coff} \quad \text{with} \quad R_{coff} = \frac{1}{|\hat{N}_{sv}|^2} \sum_{i,j \in \hat{N}_{sv}} C(i,j) \tag{8}$$

respectively.

$\hat{N}_{sv}$ denotes the size of the desired feature subset (i.e.ϕN_{sv}). c is the target class and $F(i,c)$ is the F-statistic between the feature i and the class c. $C(i,j)$ is the correlation between the i^{th} and j^{th} rows of K.

Then, the algorithm selects the subset that maximizes the relation,

$$\max_{K \in \hat{N}_{sv}} R \quad \text{with} \quad R = \frac{R_F}{R_{coff}} \tag{9}$$

The MRMR algorithm receives as input the kernel matrix $\mathbf{K} \in \mathbb{R}^{N_{sv} \times N_{sv}}$, a vector of length N_{sv} with the SV positions, and ϕ to compute the length of the desired SV subset. The subset of SV selected by the MRMR algorithm are then the SV of the model M_{op}.

Finally, similarly as Hard_IL and Cross_IL, the MRMR_IL method compares the performance of the models M_{t-1} and M_{op} on the validation set. The model with better performance is selected to become the updated classifier M_t. Fig.5 (D) illustrates the algorithmic flow.

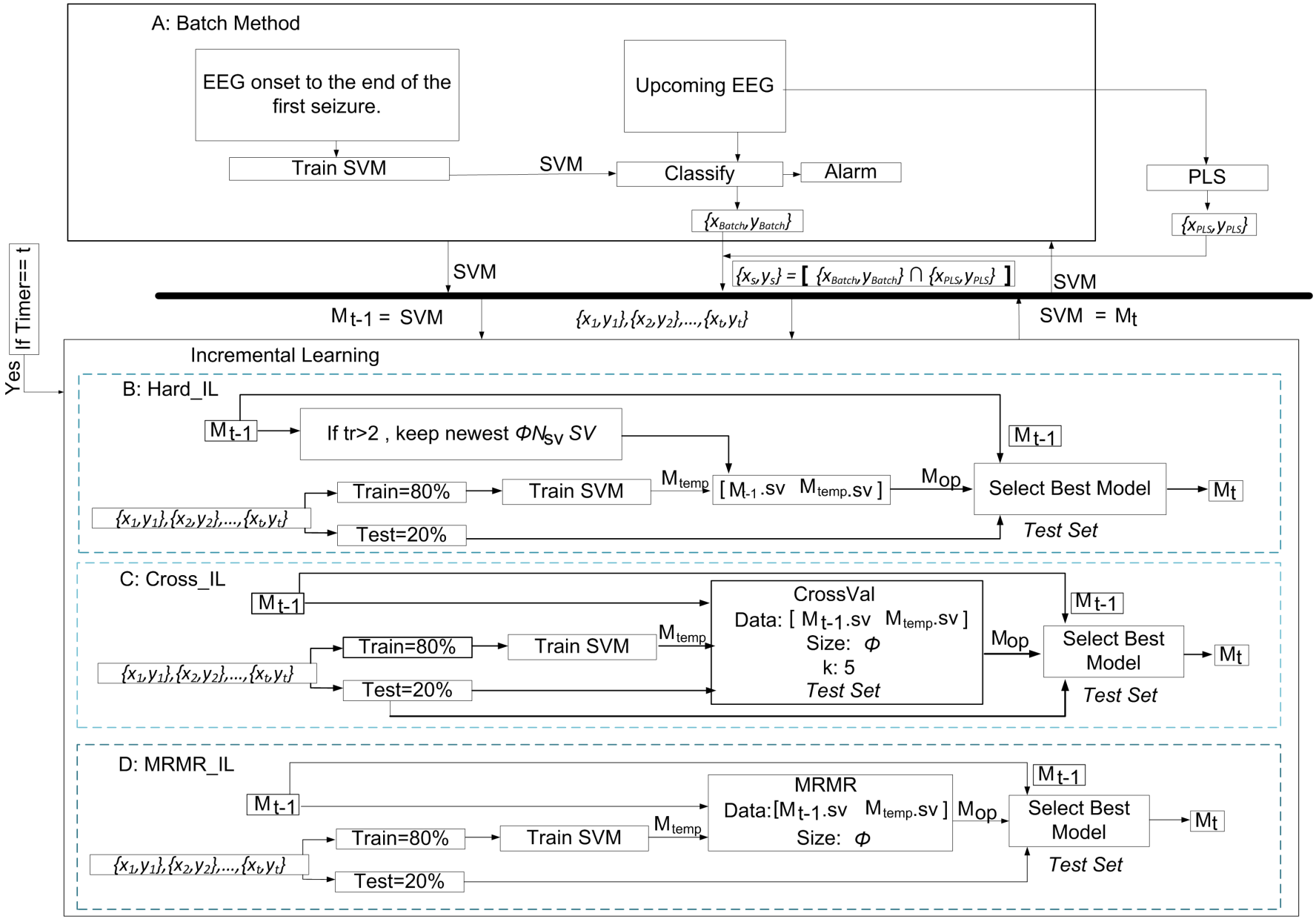


Figure 5: Flow chart of the NCSz detection method with incremental learning. Section (A) describes the Batch method. The three implementations of incremental learning approaches tested: HardIL, Cross_IL, and MRMR_IL are described in sections (B),(C), and (D) of the diagram. For the graph simplicity, $\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_t, y_t\}$ refers to the samples resulting from the double labeling. SV stands for support vector.

3.4 Alternative batch approaches

The three methods, [9],[8] and [14], were implemented and tested on our data. For each implementation, we tune the parameters to maximize the classification performance. We included some minimal variations in the training to tune the methods to the available dataset. First, the training set was composed of the first NCSz and the same number of epochs of n-NCSz EEG prior to it. For the methods of Sharma and Fatma, the EEG was segmented using non-overlapping epoch of $1\ s$ duration (length defined by the authors). In the case of Kollialil's method, two ways of training were implemented. First, the classifier was trained with all data from 7 of the 14 available cases, and the unseen cases were used to test the classifier. The cases were selected in such a way that each set included vEEG as well as ICU cases. Second, given that this method is not patient-specific, the training set assembled all training sets of all cases, i.e. the first NCSz and the same number of epochs of n-NCSz EEG prior to it.

3.5 Performance metrics

The performance for all methods was assessed by means of the sensitivity, specificity and positive predictive value (PPV) defined as,

$$Sen = \frac{TP}{TP + FN} * 100\% \tag{10}$$

$$Spec = \frac{TN}{TN + FP} * 100\% \tag{11}$$

$$PPV = \frac{TP}{TP + FP} * 100\% \tag{12}$$

respectively.

TP (true positives) is the number of samples identified as seizure by the algorithm and the human expert. FN (false negative) is the number of samples identified as nonseizure by the algorithm marked as seizure by the human expert. FP (false positive) is the number of samples identified as seizure by the algorithm marked as nonseizure by the human expert, and TN (true negative) is the number of samples classified as negative by the algorithm which are confirmed by the human expert.

To assess the significance of the improvement introduced by incremental learning, the different methods were statistically compared using a paired t-test.

4 Results and Discussion

The performance of Hard_IL, Cross_IL, and MRMR_IL was first mutually compared, to establish the best method to select the subset of support vectors to update the classifier. The method with the best performance was then compared with the Batch method and the batch methods proposed in [9],[8] and [14] (from now referred in the text as 'Koliallil', 'Sharma' and 'Fatma' respectively). Since the Batch method is the basis for all incremental learning methods proposed here, it is logical to compare their performances to assess the achieved improvement, if there is any.

To present the results, the recordings are subdivided in two groups taking into account the origin of the NCSz. The first group comprised all patients with an underlying epilepsy (Group I, recordings from 1 to 9 in Table 1). The second group comprised all patients that develop NCSz as a consequence of an acute brain dysfunction (Group II, recordings from 10 to 14 Table 1).

4.1 Comparison of Incremental Learning Approaches

All incremental learning approaches, as can be appreciated in Table 2, outperformed the Batch method in at least one of the metrics assessed.

The average performance of the Hard_IL approach showed an increase in specificity, PPV, and a lower number of false detections per hour compared to the Batch method. However, the method experiences a drop in sensitivity and specificity for Group I and II respectively. The arbitrary removal of support vectors seems to cause the forgetting of relevant information for the NCSz classification. It is possible that for more extended recordings, the catastrophic forgetting phenomenon [37] could appear if the EEG morphology changes too fast.

Table 2: Performance of the Batch Method, Hard_IL, Cross_IL and MRMR_IL in Group I and II. Group I + Group II is the performance when both groups are considered together. Spec: specificity, Sen: Sensitivity, PPV:positive predictive value, FP/h: false positive detected per hour.

	Group I				Group II				Group I + Group II			
Method	Spec	Sen	PPV	FP/h	Spec	Sen	PPV	FP/h	Spec	Sen	PPV	FP/h
Batch Method	98.7	98.2	84.6	5.9	99.4	99.98	77.7	4.9	99.0	98.8	82.1	5.5
Hard_IL	99.1	96.0	88.2	4.1	99.1	100	80.9	3.4	99.1	97.4	85.6	3.8
Cross_IL	99.7	94.8	93.3	1.4	99.6	100	85.6	1.8	99.7	96.7	90.5	1.6
MRMR_IL	99.4	99.2	92.3	3.3	99.7	100	86.1	0.8	99.5	99.5	90.1	2.4

The sensitivity values obtained by Cross_IL for Group I are lower than the ones obtained by the Batch method. Yet, Cross_IL displayed an increase in specificity and PPV for this group. Cross_IL outperformed the Batch methods in all the metrics assessed for Group II. This method achieved the lowest average number of false detections among the two groups with only 1.6 false positives per hour. The results obtained with Cross_IL demonstrate to be unstable since they depend on the support vector subset resulting after the cross-validation process (not always the same, and not always better). The performed tests show that the winning subset does not always show the best performance for new samples.

The MRMR_IL displayed values of specificity, sensitivity above 99% for the two groups assessed, and a PPV of 92.3 % and 86.1% for Group I and II respectively. The low average PPV value obtained for both groups, despite the high sensitivity and specificity values, may be caused by the unbalanced test sets in some cases. The MRMR_IL generates an average of 2.4 false detections per hour for both groups, 3.1 less than the Batch method. The MRMR_IL yields the best results among the compared methods, showing the highest average performances for all metrics.

Regarding the errors made by the MRMR_IL algorithm during the classification, it was found that the epochs

misclassified as false positives in Group I corresponded to preictal activity occurring just before a seizure. However, the preictal activity that appeared several seconds before the seizure onset (say $4-6s$), was not marked by the neurophysiologist as seizure activity. On the other hand, the end of the seizure is marked by the doctors immediately after the postictal activity. As can be seen in Fig. 6, the patterns of the preictal and postictal activity are very similar. Postictal activity was always labeled as part of the seizure by the doctors. Hence, if the patterns of the postictal activity (similar to preictal) are added to the training set, the algorithm recognizes the preictal activity as a seizure. In Group II the false positives are found in cases 10 and 11. The false positives detected in these cases are given by similarity between the background EEG and the seizure activity, making it difficult to detect the beginning of the seizure. False negative detections were found for recordings 1 and 9 (Group I). The false negatives identified occurred before the first updating of the classifier where the occurrence of errors will affect the performance of the incremental learning approach.

The execution time for the MRMR_IL approach is expected to be the same as the Batch method since the model updating is performed in the background while monitoring and seizure detection are running. Therefore, the execution time of this algorithm is assumed to be in the same time range as the Batch method ($0.37s$ to $3s$ for the classification task). The classifier retraining execution time ranged from $0.03s$ to $14.36s$ for $t = 10min$ and from $0.94s$ to $1.79min$ for $t = 2h$. This means that the highest delay in the classifier updating was approximately $14.36s$ and $1.79min$ for a time window of $10min$ and $2h$ respectively. The maximum detection delay found was $12.42s$ and occurred after the algorithm failed to detect the first three 3-seconds epochs of a seizure ($3s \times 4$) and took $0.42s$ to detect the fourth one. All tests were performed on a computer with an Intel Core-i3 processor at $1.70GHz$ with $8GB$ of RAM.

We have also considered other state-of-the-art incremental learning approaches to improve the Batch method [37, 38, 39]. However, we discarded them after some analysis. Specifically, the method proposed in [37] performs well when the classifier is trained with balanced training sets which happens in the initial training of most cases. However, in the retraining step this cannot be guaranteed; the upcoming samples during the retraining window t could be all from one class, negative samples in most of cases. After the retraining step this incremental learning method does not recognize properly the positive samples. The method introduced in [38] is proposed for a linear SVM. The linear SVM was excluded as possible classifier in our methodology in a previous work [30]. Finally, the method proposed in [39] was also evaluated, and the implementation cost was found too high.

4.2 MRMR_IL and Batch methods Comparison

Table 3 lists the classification results of the Batch, MRMR_IL, Kollialil, Sharma and Fatma methods for Group I and II. The MRMR_IL clearly outperformed the Batch method in all metrics for both groups. It should be noted that the performance of the Batch method for this database was already high before including incremental learning, which does not leave much room for improvement. In general, the MRMR_IL shows the same performance as the Batch method in the 5 cases for which the classification outcome was perfect, and improved the classification results for the other nine. However, the short duration of some EEG clips hindered the assessment of the real capabilities of the MRMR_IL approach for NCSz detection in comparison with the Batch method. MRMR_IL should be tested in longer EEG to assess the statistical significance of the improvement achieved with the addition of the incremental learning to the Batch method.

According to the paired t-test performed, the MRMR_IL improvement was statistically significant compared to Kollialil, Sharma, and Fatma methods. From the two training processes performed for Kollialil, the first one (using 7 cases for training and the rest for testing) (Kollialil_1), achieved specificity and sensitivity values of 43.2% and 63.1% respectively for Group I. For Group II, Kollialil_1 obtained specificity and sensitivity values of 65.9% and 74% respectively. This method obtained the lowest PPV among all the methods evaluated 10.6% (7.7% for Group I and 15.8% for Group II). The second way of training using the Kollialil method (assembling for all cases the first NCSz and the same number of epochs of n-NCSz EEG prior to it) (Kollialil_2), increased the performance significantly compared to Kollialil_1. For Group I, specificity value of 77.6%, a sensitivity of 72.2% and a PPV of 23.3% were obtained.

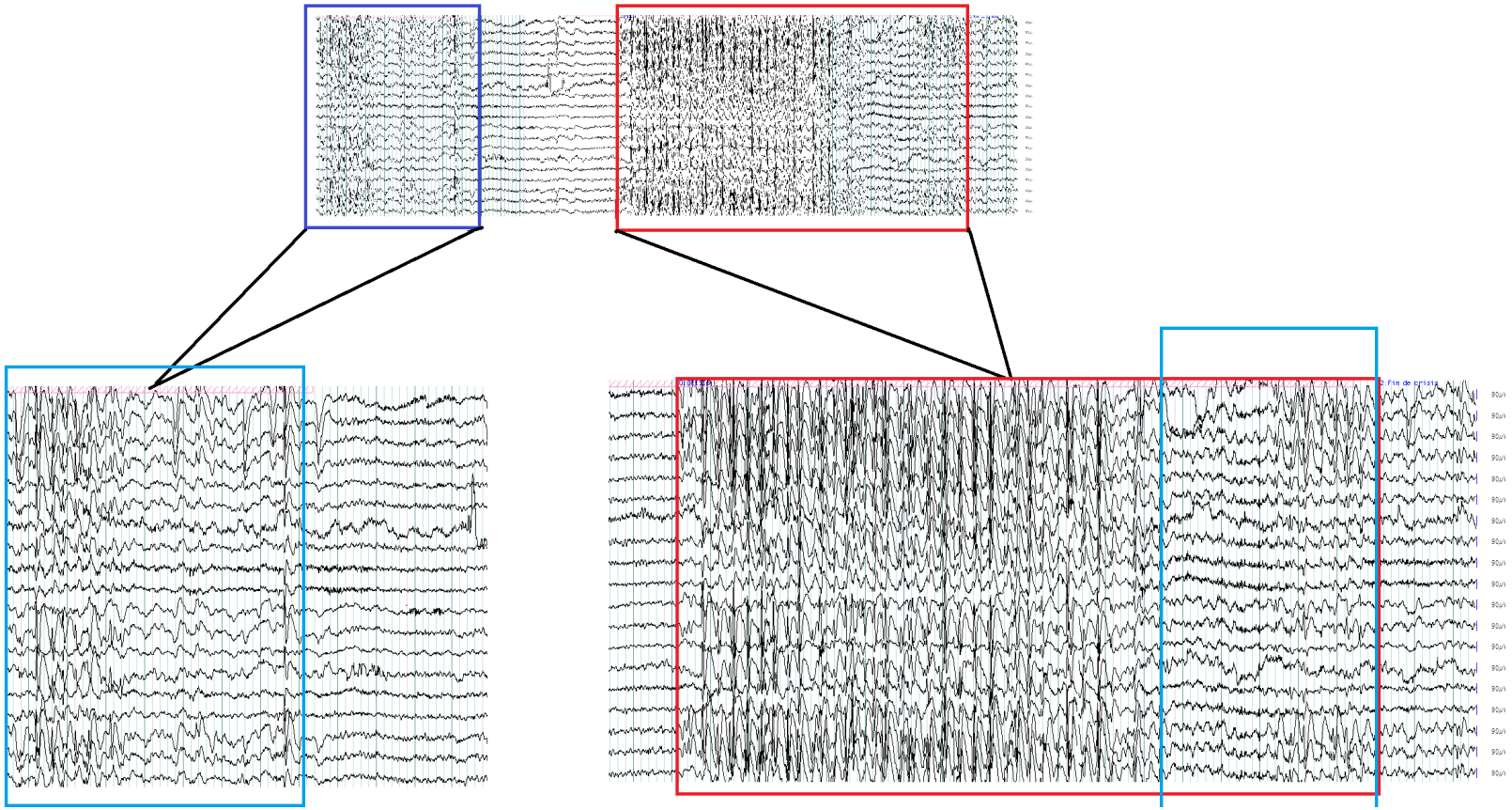


Figure 6: Epoch from Case 6 misclassified by the Incremental Learning approach. The red box indicates seizure activity. The blue boxes highlight the preictal and postictal EEG patterns. As can be observed the patterns from both preictal and postictal activities are very similar. Since the postictal was included in the classifier as seizure, its morphological similarity with the preictal activity induced this error.

For Group II, the specificity, sensitivity and PPV, using Kollialil_2 was respectively 94.8%, 58.1% and 26%. The overall performance of the Kollialil method was poor (average specificity and sensitivity values under 85%). These results, could be due to two possible causes. First, the training set was not big enough for classifier learning thereby deteriorating the classification performance. Second, the classifier was not able to recognize patterns that were not used during training, when using the Inter-quartile Range (IQR) as feature.

Table 3: Performance of MRMR_IL, Kollialil(1 and 2), Sharma, and Fatma for Groups I and II. Group I + Group II is the performance when both groups are considered together. Spec: specificity, Sen: Sensitivity, PPV:positive predictive value.

	Group I			Group II			Group I + Group II		
Method	Spec	Sen	PPV	Spec	Sen	PPV	Spec	Sen	PPV
Batch Method	98.7	98.2	84.6	99.4	100	77.7	99.0	98.8	82.1
MRMR_IL	99.4	99.2	92.3	99.7	100	86.1	99.5	99.5	90.1
Kollialil_1	43.2	63.1	7.7	65.9	74.0	15.8	51.3	67.0	10.6
Kollialil_2	77.6	72.2	23.3	94.8	58.1	26.8	83.8	67.2	24.5
Sharma	55.4	100.0	28.7	93.6	100	47.9	69.0	100.0	35.6
Fatma	91.9	73.2	37.8	96.2	43.3	47.5	93.4	62.6	41.3

Sharma method showed its best performance for Group II with specificity, sensitivity and PPV of 93.6%, 100% and 47.9% respectively. For Group I, a decrease in specificity and PPV was observed, 55.4% and 28.7% respectively. The performance of the Sharma method decayed for patients in which the training set was extremely unbalanced (case 3) or too small (cases 7, 8 and 9). The method obtained the lowest specificity values and PPV for these cases. The MRMR_IL outperformed the Sharma method in all metrics assessed except for the sensitivity, where Sharma displayed a 100% average value.

Fatma method achieved similar results in specificity (values over 90%) for the two groups analyzed. Regarding the sensitivity, a decrease was observed for Group II. Since this method does not use classifiers, the only plausible reason for the low sensitivity outcome is that the threshold defined for the seizure detection needs to be individually adjusted for each group. The method of Fatma was inferior to MRMR_IL for all the metrics evaluated, achieving an average specificity of 93.4% with a PPV of 41.3%. This algorithm displayed the lowest average sensitivity over all tested methods (around 62%).

Concerning the execution time, the MRMR_IL was only compared with Kollialil method in this regard since Sharma's and Fatma's methods uses a different lenght of analysis epochs. The algorithm from Kollialil executed the classification process of the $3\ s$ epochs in $0.28s$ to $0.19s$. An outlier of $25.25s$ was observed for one of the epochs of Case 10 which was disregarded for the time performance analysis. The superiority of Kollialil could be due to the fact that the tensorization and the tensor decomposition processes are more time consuming than the IQR range computation. Nevertheless, based on our results, we believe that both, MRMR_IL and the Batch method will perform properly in a real-time monitoring setting.

5 Conclusion

This paper proposed a method that uses incremental learning to improve the nonconvulsive epileptic seizures (NCSz) detection during continuous and long-term EEG monitoring. The proposed algorithm, namely MRMR_IL, is based on our previously proposed tensor-based batch solution [18]. The MRMR_IL retrains the original classifier periodically to improve the seizure recognition in case of changes in the EEG morphology over time. The obtained results show that the MRMR_IL outperforms the original method.

Three detection methods proposed in the literature [9],[8] and [14] were evaluated on the available database and compared to MRMR_IL. MRMR_IL was shown to outperform the three methods in all measured metrics.

In summary the MRMR_IL method proved to be an effective tool for NCSz detection in a real-time setting. The proposed method detected the NCSz caused by an epileptic disorder and those that appear as a consequence of an acute brain dysfunction with specificity and sensitivity values over 99%. For further application, it is necessary to test the method using EEG of longer duration (more than 12 hours).

References

[1] Ikuko Laccheo, Hasan Sonmezturk, Amar B Bhatt, Luke Tomycz, Yaping Shi, Marianna Ringel, Gina DiCarlo, DeAngelo Harris, John Barwise, Bassel Abou-Khalil, et al. Non-convulsive status epilepticus and non-convulsive seizures in neurological icu patients. *Neurocritical care*, 22(2):202–211, 2015.

[2] J Jirsch and LJ Hirsch. Nonconvulsive seizures: developing a rational approach to the diagnosis and management in the critically ill population. *Clinical Neurophysiology*, 118(8):1660–1670, 2007.

[3] Bryan G Young, Kenneth G Jordan, and Gordon S Doig. An assessment of nonconvulsive seizures in the intensive care unit using continuous eeg monitoring: an investigation of variables associated with mortality. *Neurology*, 47(1):83–89, 1996.

[4] J Claassen, SA Mayer, RG Kowalski, RG Emerson, and LJ Hirsch. Detection of electrographic seizures with continuous eeg monitoring in critically ill patients. *Neurology*, 62(10):1743–1748, 2004.

[5] Hojjat Adeli, Ziqin Zhou, and Nahid Dadmehr. Analysis of eeg records in an epileptic patient using wavelet transform. *Journal of neuroscience methods*, 123(1):69–87, 2003.

[6] Arnaud Jacquin, Elvir Causevic, and E Roy John. Automatic identification of spike-wave events and non-convulsive seizures with a reduced set of electrodes. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 1928–1931. IEEE, 2007.

[7] Petros Xanthopoulos, Steffen Rebennack, Chang-Chia Liu, Jicong Zhang, Gregory L Holmes, Basim M Uthman, and Panos M Pardalos. A novel wavelet based algorithm for spike and wave detection in absence epilepsy. In *BioInformatics and BioEngineering (BIBE), 2010 IEEE International Conference on*, pages 14–19. IEEE, 2010.

[8] Priyanka Sharma, Yusuf Uzzaman Khan, Omar Farooq, Manjari Tripathi, and Hojjat Adeli. A wavelet-statistical features approach for nonconvulsive seizure detection. *Clinical EEG and neuroscience*, 45(4):274–284, 2014.

[9] Eldho S Kollialil, Gopika K Gopan, A Harsha, and Liza Annie Joseph. Single feature-based non-convulsive epileptic seizure detection using multi-class svm. In *Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), 2013 International Conference on*, pages 1–6. IEEE, 2013.

[10] Sheng-Fu Liang, Wan-Lin Chang, and Herming Chiueh. Eeg-based absence seizure detection methods. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–4. IEEE, 2010.

[11] Georgiy R Minasyan, John B Chatten, and Richard N Harner. Detection of epileptiform activity in unresponsive patients using ann. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 2117–2124. IEEE, 2009.

[12] J Chris Sackellares, Deng-Shan Shiau, Jonathon J Halford, Suzette M LaRoche, and Kevin M Kelly. Quantitative eeg analysis for automated detection of nonconvulsive seizures in intensive care units. *Epilepsy & Behavior*, 22:S69–S73, 2011.

[13] Swati Varshney, Yusuf U Khan, Omar Farooq, Priyanka Sharma, and Manjari Tripathi. Latency study of non-convulsive seizures. In *Multimedia, Signal Processing and Communication Technologies (IMPACT), 2013 International Conference on*, pages 108–111. IEEE, 2013.

[14] T. Fatma, O. Farooq, M. Khan, Y. U. Tripathi, and P. Sharma. Automatic detection of non-convulsive seizures: A reduced complexity approach. *Journal of King Saud University-Computer and Information Sciences*, 28(4):407–415, 2016.

[15] Yusuf U Khan, Omar Farooq, Manjari Tripathi, Priyanka Sharma, and Parvez Alam. Automatic detection of non-convulsive seizures using ar modeling. In *Power, Control and Embedded Systems (ICPCES), 2012 2nd International Conference on*, pages 1–4. IEEE, 2012.

[16] Abdul Quaiyum Ansari and Priyanka Sharma. A review on automated detection of non-convulsive seizures using eeg. In *Computational Intelligence & Communication Technology (CICT), 2016 Second International Conference on*, pages 283–286. IEEE, 2016.

[17] Eugen Trinka and Markus Leitinger. Which eeg patterns in coma are nonconvulsive status epilepticus? *Epilepsy & Behavior*, 49:203–222, 2015.

[18] Y. Rodríguez Aldana, B. Hunyadi, E.J. Marañón Reyes, V. Rodríguez Rodríguez, and S. Van Huffel. Nonconvulsive epileptic seizure detection in scalp eeg using multiway data analysis. *IEEE Journal of Biomedical and Health Informatics*, 23(2):660–671, 2019.

[19] Scott B Wilson. Algorithm architectures for patient dependent seizure detection. *Clinical Neurophysiology*, 117(6):1204–1216, 2006.

[20] Borbála Hunyadi, Marco Signoretto, Wim Van Paesschen, Johan AK Suykens, Sabine Van Huffel, and Maarten De Vos. Incorporating structural information from the multichannel eeg improves patient-specific seizure detection. *Clinical Neurophysiology*, 123(12):2352–2361, 2012.

[21] E Acar, C Schenker, Y Levin-Schwartz, VD Calhoun, and T Adali. Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data. *Clinical Neurophysiology*, 13(416), 2019.

[22] Mark L Scheuer and Scott B Wilson. Data analysis for continuous eeg monitoring in the icu: seeing the forest and the trees. *Journal of clinical neurophysiology*, 21(5):353–378, 2004.

[23] LD Iasemidis, D-S Shiau, Panos M Pardalos, Wanpracha Chaovalitwongse, K Narayanan, Awadhesh Prasad, Konstantinos Tsakalis, Paul R Carney, and J Chris Sackellares. Long-term prospective on-line real-time seizure prediction. *Clinical Neurophysiology*, 116(3):532–544, 2005.

[24] Stephen Faul, Geraldine Boylan, Sean Connolly, Liam Marnane, and Gordon Lightbody. An evaluation of automated neonatal seizure detection methods. *Clinical Neurophysiology*, 116(7):1533–1541, 2005.

[25] Gotman J. Automatic seizure detection: improvements and evaluation. *Electroencephalogr Clin Neurophysiol*, 76:317–324, 1990.

[26] Gabor AJ. Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies. *Electroencephalogr Clin Neurophysiol*, 107:27–32, 1998.

[27] Wilson SB, Scheuer ML, Emerson RG, and Gabor AJ. Seizure detection: evaluation of the reveal algorithm. *Clin Neurophysiol*, 115(10):2280–2291, 2004.

[28] Scott B Wilson. A neural network method for automatic and incremental learning applied to patient-dependent seizure detection. *Clinical Neurophysiology*, 116(8):1785–1795, 2005.

[29] Thomas De Cooman, Anouk Van de Vel, Berten Ceulemans, Lieven Lagae, Bart Vanrumste, and Sabine Van Huffel. Online detection of tonic-clonic seizures in pediatric patients using ecg and low-complexity incremental novelty detection. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5597–5600. IEEE, 2015.

[30] Y. Rodríguez Aldana, B. Hunyadi, E.J. Marañón Reyes, V. Rodríguez Rodríguez, and S. Van Huffel. Nonconvulsive epileptic seizures detection using multiway data analysis. In *Proc. of the 25th European Signal Processing Conference, Kos, Grece, Aug 2017*, pages 2344–2348. IEEE, 2017.

[31] Xiaoxing Yang, Ke Tang, and Xin Yao. The minimum redundancy–maximum relevance approach to building sparse support vector machines. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 184–190. Springer, 2009.

[32] Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005.

[33] Mikael Sørensen and Lieven De Lathauwer. Tensor decompositions with block-toeplitz structure and applications in signal processing. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pages 454–458. IEEE, 2011.

[34] Alexander Gepperth and Barbara Hammer. Incremental learning algorithms and applications. In *European Symposium on Artificial Neural Networks (ESANN)*, pages 1–13. ESANN, 2016, Bruges, Belgium, 2016. URL `https://hal.archives-ouvertes.fr/hal-01418129`.

[35] Richard G Brereton and Gavin R Lloyd. Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics*, 28(4):213–225, 2014.

[36] Yissel Rodríguez Aldana, Frank Sanabria Macías, Enrique J Marañón Reyes, Valia Rodríguez Rodríguez, Sabine Van Huffel, and Borbála Hunyadi. Using partial least squares for nonconvulsive epileptic seizure detection. *Revista Cubana de Ciencias Informáticas*, 13(1):1–13, 2019.

[37] Gert Cauwenberghs and Tomaso Poggio. Incremental and decremental support vector machine learning. In *Advances in neural information processing systems*, pages 409–415, 2001.

[38] Cheng-Hao Tsai, Chieh-Yen Lin, and Chih-Jen Lin. Incremental and decremental training for linear classification. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 343–352. ACM, 2014.

[39] Bin Gu, Victor S Sheng, Keng Yeow Tay, Walter Romano, and Shuo Li. Incremental support vector learning for ordinal regression. *IEEE Transactions on Neural networks and learning systems*, 26(7):1403–1416, 2015.