Nonconvulsive epileptic seizure monitoring with incremental learning

Yissel Rodríguez Aldana, Enrique J. Marañón Reyes, Frank Sanabria Macías, Valia Rodríguez Rodríguez, Lilia Morales Chacón, Sabine Van Huffel, Borbála Hunyadi

ARTICLE INFO

Keywords:
- Nonconvulsive epileptic seizures
- Hilbert Huang transform
- Multway data analysis
- Incremental learning

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 NCSE. To implement the incremental learning SVM, three methodologies are tested. These approaches differ in the way they reduce the set of 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%.

1. 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 that NCSE lasting more than 10 h are associated with permanent disabilities, while mortality is very high in NCSz lasting more than 20 h [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 h [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 [5–8], entropy [9,10], nonlinear parameters [6,9,11], statistical and spectral features of the EEG [9–14] with various machine learning techniques [8–11,13,15] or thresholds [6,7,12,14] to detect the NCSzs.

* Corresponding author. Universidad de Oriente, Center of Neuroscience and Signals and Image Processing, Santiago de Cuba, Cuba.
E-mail address: yaldana@uo.edu.cu (Y. Rodríguez Aldana).

https://doi.org/10.1016/j.compbiomed.2019.103434
Received 6 May 2019; Received in revised form 2 September 2019; Accepted 3 September 2019
Available online 6 September 2019
0010-4825/© 2019 Elsevier Ltd. All rights reserved.
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].

Figs. 1–3 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 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 cases that 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 [7,8,13,15]. 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 Ref. [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 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%. Fig. 4 shows the block diagram of the method proposed in Ref. [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 [22–24]. 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 [25–27]. 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 [8,9,14,18,30]. 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 Ref. [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 Ref. [28] and not requiring the user intervention in the updating process as in Ref. [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
Table 1

<table>
<thead>
<tr>
<th>Patient data</th>
<th>Recording Seizures</th>
<th>Protocol Channels</th>
<th>EKG</th>
<th>EOG</th>
<th>EMG</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>Temporal lobe epilepsy</td>
<td>vEEG</td>
<td>6</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>Temporal lobe epilepsy</td>
<td>vEEG</td>
<td>3</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>Temporal lobe epilepsy</td>
<td>vEEG</td>
<td>13</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>Temporal lobe epilepsy</td>
<td>vEEG</td>
<td>5</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>Lennox-Gastaut syndrome</td>
<td>vEEG</td>
<td>3</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>Temporal lobe epilepsy</td>
<td>vEEG</td>
<td>2</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>Temporal lobe epilepsy</td>
<td>vEEG</td>
<td>12</td>
<td>17</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>Juvenile myoclonic epilepsy</td>
<td>vEEG</td>
<td>6</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>Frontal lobe epilepsy</td>
<td>vEEG</td>
<td>34</td>
<td>17</td>
<td>x</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>Coma/Subarachnoid Haemorrhage</td>
<td>ICU</td>
<td>40</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>Myoclonic Seizures/Brain Tumor</td>
<td>ICU</td>
<td>2</td>
<td>13</td>
<td>x</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>Coma/Systemic Vasculitis</td>
<td>ICU</td>
<td>6</td>
<td>8</td>
<td>x</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>Seizures/Brain Tumor/Sepsis</td>
<td>ICU</td>
<td>5</td>
<td>19</td>
<td>x</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>Generalized Tonic-Clonic Seizures</td>
<td>ICU</td>
<td>3</td>
<td>14</td>
<td>x</td>
</tr>
</tbody>
</table>

performance than the Batch method in some sense (for the results of this comparison, see Ref. [18]).

2. Materials and methods

2.1. 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 14 min to more than 21 h (mean 280 min 21 s) 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.

2.2. Batch method

The Batch method in Ref. [18] analyses the EEG data in non-overlapping segments (epochs) of 3 s long. All epochs are expanded in the time-frequency domain using an Hilbert Huang Transform (HHT).

2.3. Incremental learning

2.3.1. 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 evolving 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 $t = 10 \text{min}$ 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 towards the beginning of the recording. In other words, if the first seizure is of length $L$ ‘epochs’, then $S = 2L/3$.

\[ y_{\text{new}} = \text{sign} \left[ w^T \varphi(x_{\text{new}}) + w_0 \right] \]
collected during this time is defined as a training round.

### 2.3.2. 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 $x_{tnew}$ and their corresponding labels $y_{tnew}$ 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 the 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 $y_{tnew}^{PLS}$ labels of the $x_{tnew}$ 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 Ref. [36].

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

### 2.3.3. Conservative updating

The selected samples $x_{tnew}$ 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 $M_{opt}$ that is tested on the validation set (An explanation on why and how to choose the optimal subset will follow below in section 3.3.3). Then, the model $M_{t-1}$ is also tested on the validation set. Finally, after comparing $M_{t-1}$ and $M_{opt}$, 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.

### 2.3.4. 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 $\varphi = N_{new}/N_SV$ is introduced to limit the growth of the number of SVs, $\varphi \in R: 0 < \varphi \leq 1. N_{new}$ 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

![Flow chart of the NCSz detection method with incremental learning](Image)

**Fig. 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, $\{x_1,y_1\}, \{x_2,y_2\} \ldots \{x_n,y_n\}$ refers to the samples resulting from the double labeling. SV stands for support vector.
\( SV_{ab} = \varphi N_{sv} \)

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 \( M_{\text{temp}} \) is trained with the training set. The SV from the models \( M_{-1} \) and \( M_{\text{temp}} \) will become candidate SVs for the new model \( M_{op} \).

To control the model growth, the 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 \( N_{sv} - \varphi N_{sv} \) 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 \( M_{-1} \) and \( M_{\text{temp}} \) an 5-fold (of length \( \varphi N_{sv} \)) 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 \( \varphi N_{sv} \). 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 Ref. [31] to estimate the optimal subset of SVs. In Ref. [31] the selection of the subsets 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

\[
 f(x) = \alpha^T K + \alpha^T_0
\]

where \( \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_{\varphi N_{sv}}, \varphi N_{sv}] \) is the weight vector, \( K \in \mathbb{R}^{N_{sv} \times N_{sv}} \) is the kernel matrix, with the element in the \( i \)-th row and \( j \)-th column \( K_{ij} = k(x_i, x_j) \). \( N_{sv} \) is the number of SVs and \( \omega_0 \in \mathbb{R}^{1 \times N_{sv}} \) is a bias vector containing the bias term \( \beta_0 \) 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

\[
 g(x) = \beta^T X + \beta_0^T
\]

where \( X \in \mathbb{R}^{L \times M} \) is a data matrix with \( L \) number of features describing \( M \) data samples, \( \beta \in \mathbb{R}^{L \times 1} \) the corresponding weight vector and \( \beta_0 \in \mathbb{R}^{1 \times 1} \) is the vector containing the bias term \( \beta_0 \) 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 \( X \)) 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 \( K \in \mathbb{R}^{N_{sv} \times N_{sv}} \) can be interpreted as a data matrix in the feature selection context where each column of \( K \) corresponds to a data sample and each row of \( K \) (i.e. support vector) corresponds to a feature. Hence, the number of SVs can be reduced by dropping rows in the kernel matrix \( K \), while keeping the number of columns unchanged [31]. To select the best subset of SVs, Yang proposed the feature selection technique MRMR introduced in Ref. [32].

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

\[
 F_i = \frac{\sum_{n=1}^{N_{sv}} n_i (K_{i} - \bar{K}_i)^2}{\sigma^2}
\]

where \( K_i \) denotes the \( i \)-th row of the feature matrix \( K \), \( n_i \) is the number of samples from the \( c \)-th class. \( \bar{K}_i \) is the mean value of the row \( K_i \) and \( \bar{K}_i \) is the mean value of \( K_i \) within the \( c \)-th class. \( \sigma^2 \) is the pooled variance defined as [31],

\[
 \sigma^2 = \frac{\sum(n_i - 1)\sigma^2}{N_{sv} - 2}
\]

where \( \sigma^2 \) 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],

\[
 \text{max} R_p \quad \text{with} \quad R_p = \frac{1}{N_{sv}} \sum_{i,c} F(i,c) \tag{7}
\]

and

\[
 \text{min} R_{off} \quad \text{with} \quad R_{off} = \frac{1}{N_{sv}} \sum_{i,j \in \mathcal{N}} C(i,j) \tag{8}
\]

respectively.

\( \mathcal{N}_{sv} \) denotes the size of the desired feature subset (i.e. \( \varphi N_{sv} \)). \( c \) 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,

\[
 \text{max} R \quad \text{with} \quad R = \frac{R_p}{R_{off}} \tag{9}
\]

The MRMR algorithm receives as input the kernel matrix \( K \in \mathbb{R}^{N_{sv} \times N_{sv}} \), a vector of length \( N_{sv} \) with the SV positions, and \( \varphi \) 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 \( M_{-1} \) and \( M_{\text{temp}} \) on the validation set. The model with better performance is selected to become the updated classifier \( M_{op} \). Fig. 5 (D) illustrates the algorithmic flow.

2.4. Alternative batch approaches

The three methods [8,9,14], 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 1s duration (length defined by the authors). In the case of Kolliarlis’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 veEGG 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.

2.5. Performance metrics

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

\[
 Sen = \frac{TP}{TP + FN} \times 100\% \tag{10}
\]

\[
 Spec = \frac{TN}{TN + FP} \times 100\% \tag{11}
\]

\[
 PPV = \frac{TP}{TP + FP} \times 100\% \tag{12}
\]

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.

3. 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 Refs. [8,9,14] (from now referred in the text as ‘Kollialil’, ‘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).

3.1. 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.

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 one evaluated, and the implementation cost was found too high).

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 \(\tau = 10\text{min}\) and from 0.94s to 1.79min for \(\tau = 2\text{h}\). 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 3s epochs of a seizure (3s × 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 [37–39]. However, we discarded them after some analysis. Specifically, the method proposed in Ref. [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 \(\tau\) 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 Ref. [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 Ref. [39] was also evaluated, and the implementation cost was found too high.

3.2. MRMR_IL and batch methods comparison

Table 3 lists the classification results of the Batch, MRMR_IL, Kolliail, Sharma and Fatma methods for Group I and II. The MRMR_IL clearly outperformed the Batch method in all metrics for both groups. It

Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Group I</th>
<th>Group II</th>
<th>Group I + Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spec</td>
<td>Sen</td>
<td>PPV</td>
</tr>
<tr>
<td>Batch Method</td>
<td>98.7</td>
<td>98.2</td>
<td>84.6</td>
</tr>
<tr>
<td>Hard IL</td>
<td>99.1</td>
<td>96.0</td>
<td>88.2</td>
</tr>
<tr>
<td>Cross IL</td>
<td>99.7</td>
<td>94.8</td>
<td>93.3</td>
</tr>
<tr>
<td>MRMR_IL</td>
<td>99.4</td>
<td>99.2</td>
<td>92.3</td>
</tr>
</tbody>
</table>
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 Kolliilal, Sharma, and Fatma methods. From the two training processes performed for Kolliilal, the first one (using 7 cases for training and the rest for testing) (Kolliilal_1), achieved specificity and sensitivity values of 43.2% and 63.1% respectively for Group I. For Group II, Kolliilal_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 Kolliilal method (assembling for all cases the first NCSz and the same number of epochs of n-NCSz EEG prior to it) (Kolliilal_2), increased the performance significantly compared to Kolliilal_1. For Group I, specificity value of 77.6%, a sensitivity of 72.2% and a PPV of 23.3% were obtained.

For Group II, the specificity, sensitivity and PPV, using Kolliilal_2 was respectively 94.8%, 58.1% and 26%. The overall performance of the Kolliilal 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.

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

Table 3
Performance of MRMR_IL, Kolliilal(1 and 2), Sharma, and Fatma for Groups I and II. Group I + Group II is the performance when both groups are considered together.

<table>
<thead>
<tr>
<th>Method</th>
<th>Group I Spec</th>
<th>Group I Sen</th>
<th>Group I PPV</th>
<th>Group II Spec</th>
<th>Group II Sen</th>
<th>Group II PPV</th>
<th>Group I + Group II Spec</th>
<th>Group I + Group II Sen</th>
<th>Group I + Group II PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch Method</td>
<td>98.7</td>
<td>98.2</td>
<td>84.6</td>
<td>99.4</td>
<td>100</td>
<td>77.7</td>
<td>99.0</td>
<td>98.8</td>
<td>82.1</td>
</tr>
<tr>
<td>MRMR_IL</td>
<td>99.4</td>
<td>99.2</td>
<td>92.3</td>
<td>99.7</td>
<td>100</td>
<td>86.1</td>
<td>99.5</td>
<td>99.5</td>
<td>90.1</td>
</tr>
<tr>
<td>Kolliilal_1</td>
<td>43.2</td>
<td>63.1</td>
<td>7.7</td>
<td>65.9</td>
<td>74.0</td>
<td>15.8</td>
<td>51.3</td>
<td>67.0</td>
<td>10.6</td>
</tr>
<tr>
<td>Kolliilal_2</td>
<td>77.6</td>
<td>72.2</td>
<td>23.3</td>
<td>94.8</td>
<td>58.1</td>
<td>26.8</td>
<td>83.8</td>
<td>67.2</td>
<td>24.5</td>
</tr>
<tr>
<td>Sharma</td>
<td>55.4</td>
<td>100.0</td>
<td>28.7</td>
<td>93.6</td>
<td>100</td>
<td>47.9</td>
<td>69.0</td>
<td>100.0</td>
<td>35.6</td>
</tr>
<tr>
<td>Fatma</td>
<td>91.9</td>
<td>73.2</td>
<td>37.8</td>
<td>96.2</td>
<td>43.3</td>
<td>47.5</td>
<td>93.4</td>
<td>62.6</td>
<td>41.3</td>
</tr>
</tbody>
</table>

Fig. 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.
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 use a different length 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.

4. Conclusion

This paper proposed a method that uses incremental learning to improve the nonconvolutive epileptic seizures (NCSs) 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 retraining 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 [8,9,14] 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 NCSs detection in a real-time setting. The proposed method detected the NCSs 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 h).

Conflict of Interest Statement

None declared.

Funding

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.

References


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


