

Automated patient-specific classification of long-term Electroencephalography



Serkan Kiranyaz^{a,*}, Turker Ince^b, Morteza Zabihi^a, Dilek Ince^c

^a Department of Signal Processing, Tampere University of Technology, Tampere, Finland

^b Department of Electrical and Electronics Engineering, Izmir University of Economics, Turkey

^c Department of Pediatric Oncology, Dokuz Eylul University, Izmir, Turkey

ARTICLE INFO

Article history:

Received 29 June 2013

Accepted 3 February 2014

Available online 22 February 2014

Keywords:

EEG classification

Seizure event detection

Evolutionary classifiers

Morphological filtering

ABSTRACT

This paper presents a novel systematic approach for patient-specific classification of long-term Electroencephalography (EEG). The goal is to extract the seizure sections with a high accuracy to ease the Neurologist's burden of inspecting such long-term EEG data. We aim to achieve this using the minimum feedback from the Neurologist. To accomplish this, we use the majority of the state-of-the-art features proposed in this domain for evolving a collective network of binary classifiers (CNBC) using multi-dimensional particle swarm optimization (MD PSO). Multiple CNBCs are then used to form a CNBC ensemble (CNBC-E), which aggregates epileptic seizure frames from the classification map of each CNBC in order to maximize the sensitivity rate. Finally, a morphological filter forms the final epileptic segments while filtering out the outliers in the form of classification noise. The proposed system is fully generic, which does not require any *a priori* information about the patient such as the list of relevant EEG channels. The results of the classification experiments, which are performed over the benchmark CHB-MIT scalp long-term EEG database show that the proposed system can achieve all the aforementioned objectives and exhibits a significantly superior performance compared to several other state-of-the-art methods. Using a limited training dataset that is formed by less than 2 min of seizure and 24 min of non-seizure data on the average taken from the early 25% section of the EEG record of each patient, the proposed system establishes an average sensitivity rate above 89% along with an average specificity rate above 93% over the test set.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Epilepsy is a chronic disease of the central nervous system that predisposes patients to experiencing recurrent seizures which are abnormal brain activities and stem from many diseases. Seizures are not a disease themselves; rather a transient symptom of synchronous neuronal activity in the brain [1]. In contrast to methods such as Magnetic Resonance Imaging [2] and three-dimensional accelerometer [2], using Electroencephalography (EEG) is a more common method for long-term epileptic patient monitoring. EEG is a multi-channel recording of the electrical activity generated by neurons in the brain. The so-called scalp EEG, henceforth referred as EEG, is measured by non-invasive electrodes arrayed on an individual's head. One percent of population in the world is suffering from epilepsy and 80% of its burden is in the developing countries [2], and this shows the importance of seizure detection and epilepsy diagnosis.

In recent years, several feature-based epileptic seizure detection studies using EEG signals have been done. They can be divided

into two categories: patient specific and non-patient specific methods. EEG presents significant complications and serious drawbacks to the classification of the epileptic seizures due to the physical properties of the EEG signal, which is extremely sensitive to the activity of the neurons on the brain surface; however, the neurons in the deeper sections of the brain have only limited or no effect on the signal. The major problem in EEG that complicates the accurate classification of epileptic activities the most is the significant variations of the EEG signals for seizure and non-seizure states across the individuals [3,4]. The opposite is also true, i.e., the signal properties of one patient's seizure may closely resemble the characteristics of a normal EEG signal gathered from the same or a different patient [4,5]. This is obviously why the limited success achieved by those non-patient specific methods. Furthermore, EEG is susceptible to contamination of physiologic (e.g. involuntary body or organ movements such as eye blinks, heart beats and muscle contractions) and non-physiologic noise (e.g. the sway of electrodes, the coupling of AC harmonics of the machinery, etc.) [4]. As a result the features, too, are contaminated by a certain level of noise, which degrades their characterization power and the noise level can also vary among patients and/or the EEG devices used. Even

* Corresponding author.

for a patient-specific approach it is a well-known fact that the performance of EEG classification strongly depends on the characterization power of the features extracted from the EEG data and on the design of the classifier (classification model or network structure and its parameters) [4]. Due to the massive size and duration of the long-term EEG signal, the majority of the patient specific approaches were proposed on a single or few manually selected features in order to avoid the “Curse of Dimensionality” phenomenon, e.g., [4,12,15]. Unfortunately, the aforementioned variations in the EEG signal and in the level of noise contamination over the features make the characterization power of any feature significantly varying from patient to patient, meaning that no feature can guarantee a certain level of detection and/or classification accuracy for all patients. Fig. 1 demonstrates this fact for two features, coefficient of variation using wavelet Daubiches 2 and delta Short-Time Fourier Transform (STFT), commonly used for EEG classification where both of them have a clear discrimination of the seizure section with the highest peak on one recording (the plots on the top; left belonging to same patient and right to another) but fails to do so for the other (the plots at the bottom). The latter issue, the design of the classifier, is also critical since all patient-specific methods usually rely on a single and static (fixed configuration) classifier, which cannot scale well to such massive magnitude of data. Such *a priori* fixations on the features and the classifier configuration may further cause an unpredictable level of unreliability for any patient-specific method and this is perhaps why the most of those methods use the majority of the EEG dataset for training the classifier to achieve an acceptable level of sensitivity on seizure (onset) detection. This is obviously cumbersome or not even feasible at all for long-term EEG records and in fact, a true “patient-specific” approach should, therefore, search for the best possible feature set and the optimal classifier configuration for the patient’s EEG signal characteristics.

In order to address such deficiencies and drawbacks, in this paper we propose a generic and robust system for patient-specific classification of long-term EEG. The proposed system, first of all, is not only a seizure onset detector; rather attempts to classify the entire seizure section using an ensemble of the collective

network of binary classifiers (CNBC), each of which encapsulates binary classifiers (BCs) that are optimized for a particular feature set and EEG signal from an individual channel. The recently proposed multi-dimensional Particle Swarm Optimization (MD-PSO) [9] is used as the primary evolution technique and evolutionary feed-forward Artificial Neural Networks (ANNs) [10,11] are used as the BCs; however, any other classifier type can also be used within the CNBC topology. Our main objective is to maximize the detection accuracy especially for the seizure sections (i.e. *sensitivity*) with an acceptable level of false positives (i.e. *specificity*) whilst using the *minimum* feedback from the Neurologist. To accomplish this, we shall use the major state-of-the-art features proposed in this domain and for each patient the underlying network topology will learn to *weight* the features according to their ability to discriminate epileptic seizures. The proposed system is fully generic, which does not require any *a priori* information about the patient such as the list of relevant EEG channels. In addition to the learning the most discriminative features, the proposed system will also learn the most informative EEG channels during the training phase and favor their outputs accordingly during the classification phase. Moreover, in order to maximize the *sensitivity* rate, the output of each CNBC is then used to aggregate epileptic seizure frames and a morphological filter forms the final epileptic segments while filtering out the outliers within the non-seizure segments. At the end, the CNBC ensemble (CNBC-E), once evolved using the Neurologist’s labels over an earlier EEG record of a patient, can then be used to classify automatically the long-term EEG signal of that patient.

The rest of the paper is organized as follows. Section 2 presents the related work in this area. Section 3 outlines the EEG dataset used in this study and provides a detailed description of the feature extraction methodology for the proposed patient-specific EEG classification system. Section 4 briefly presents the recently proposed evolutionary search technique, The MD PSO and its application over the evolutionary feed-forward ANNs. In Section 5, the proposed classifier network topology, the CNBC, and the overall systematic approach for the EEG classification using CNBC ensembles (CNBC-E) are described in detail. The classification results and comparative performance evaluations using the benchmark

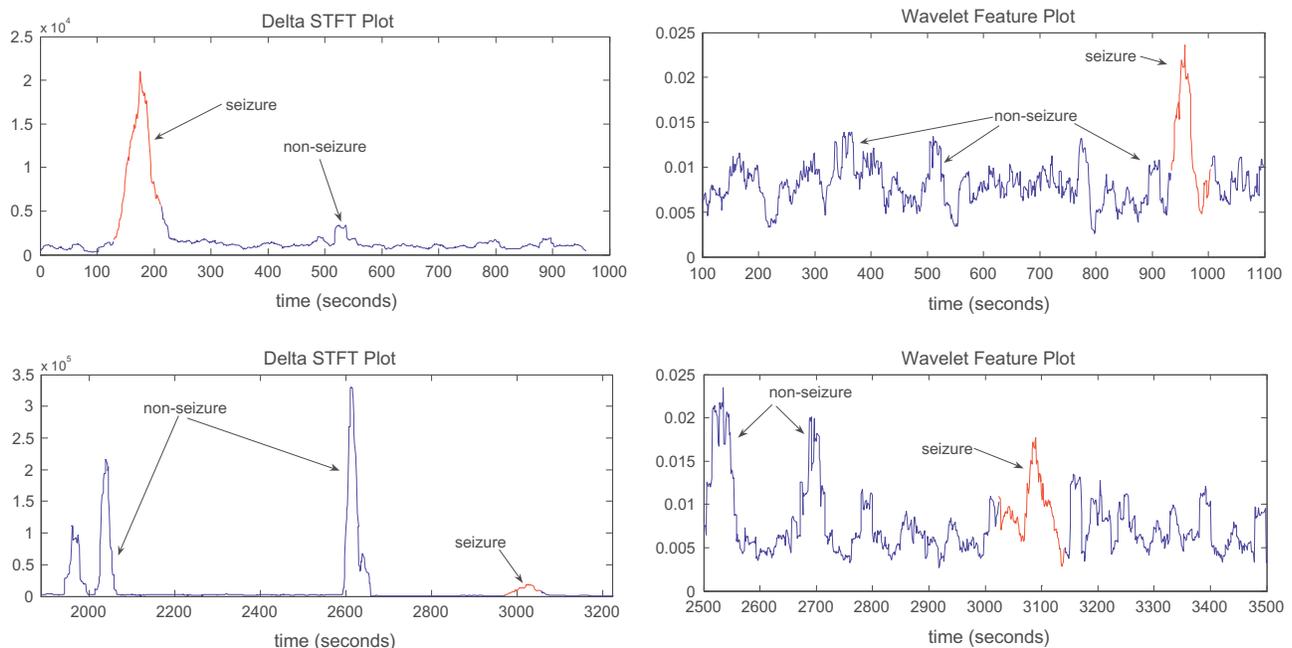


Fig. 1. Delta STFT feature plot for 2 different EEG recordings of patient 2 (left), Wavelet feature (coefficient of variation using db2 wavelet) plots for patient 13 (right-top) and patient 17 (right-bottom). All EEG recordings are from the benchmark CHB-MIT scalp EEG dataset [8].

CHB-MIT scalp long-term EEG dataset [8] are given in Section 4. Finally, Section 7 concludes the paper and suggests topics for future research.

2. Related work

Although there are several related studies in this area, most of them are either non-patient specific techniques with inferior performances or patient specific methods only applied over relatively short EEG records and/or few patients. There are also other studies in the literature that achieved high classification accuracies but they either focused on a specific seizure type, or selected subset of channels while some of them used training and test sets which have high temporal correlation. Therefore, in this section we shall focus on the major and the most recent works particularly on patient-specific long-term EEG classification. Shoeb and Guttag [12] designed a seizure detector to detect the seizure onset using CHB-MIT Scalp benchmark EEG dataset. They measured the energy of each 2 s epoch, which was passed through a filterbank that spans the frequency range, 0.5–25 Hz, as the feature. They also used Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel (kernel parameters: $\gamma = 0.1$ and $C = 1$) for the classification purpose. First, 20 s of seizure segments and all non-seizure segments were used for training. In order to evaluate the sensitivity and specificity performance of the designed detector they used the following approach. To compute the sensitivity (specificity) they used N_{Ns} non-seizure (N_s seizure) records and $N_s - 1$ seizure ($N_{Ns} - 1$ non-seizure) records of the patient EEG where N_{Ns} corresponds to the number of 1-h records without seizure events and N_s is the number of 1-h records with at least one seizure event. They repeated the training task N_s (N_{Ns}) times so that each seizure (non-seizure) record is tested. An average of 96% sensitivity with the mean latency of 4.6 s was reported.

In 2012, Ravish et al. [13] used CHB-MIT Scalp benchmark EEG dataset to detect period of the seizure events. They used 64 one-hour records, half with and half without seizure events. The power ratio of alpha (7.5–12.5 Hz) and delta (0.5–3.5 Hz) bands (Alpha power/Delta power Ratio, ADR) from each 2 s window with 1 s overlap were calculated. They only used four EEG channels: FP2–F8, T8–P8, FP1–F7 and T7–P7. In order to determine the seizure duration, the time duration between 10% of the peak ADR before and after the peak is measured. Their algorithm detected all 32 non-seizure segments and only 24 (out of 32) of seizure segments failing to detect 8 seizure segments.

Khan et al. [14] proposed a framework to detect the seizure onsets. For this purpose, they extracted 6 features from each 1 s epoch: kurtosis and skewness of the raw EEG, relative energy and variation of Daubechies 4 wavelet coefficients (A5 0–4 Hz, D5 4–8 Hz, and D3 16–32 Hz). In order to normalize the epoch features they formed a 25 s window as a background window and considered 15 s between each epoch and the background window to prevent any failure of the seizure onset detection. They applied the algorithm only to the first 10 patients of CHB-MIT Scalp benchmark EEG dataset. Using a linear classifier they for each patient they used 80% of the seizure segments for training and the other remaining 20% for testing, and repeated this process until every seizure segment was tested. They obtained mean latency of 3.2 s and reported 100% sensitivity.

Shoeb et al. [15] in a recent study proposed an algorithm to detect seizure terminations (offsets). Similar to their previous studies they used 5 s sliding window with 4 s overlap. In this study, they assumed that each seizure onset has been detected in advance. As for the features, they extracted average energy of the periodogram of each channel within 25 contiguous frequency bands (e.g., 0–1, 1–2, ..., 24–25 Hz) from each window. They performed both patient-specific and patient non-specific approaches for

classification. SVM with linear and RBF kernels were used for both modes, respectively. They used leave-one-record-out procedure for evaluation of the end detector's performance as in [12]. In patient non-specific mode, the classifier was trained on feature vectors of all records of all patients except patient i . In patient-specific detector, 132 out of 133 seizure ends were detected with an average, absolute end detection error of 10.3 ± 5.5 s. Also in patient non-specific detector, all seizure ends with an average absolute error of 8.9 ± 2.3 s were recognized.

All these prior works achieved an elegant sensitivity rate on seizure onset detection and thus are among the state-of-the-art techniques in this area; however, as discussed briefly in Section 1, there are serious flaws and feasibility drawbacks particularly about their practical use to ease the manual burden of the Neurologist. Except the work proposed in [13], they all used 80% or higher training rate to train classifiers. This obviously is quite hard, if feasible at all, in practice for long-term EEG recordings. Additionally, it is well known that for a larger set of data the training rate must be kept small in order to avoid overfitting and improve classifier's generalization performance. Note that for such onset detectors the sensitivity rate indicates the rate of correctly detected onsets of the seizure segments without any indication of the seizure duration. In addition to that for instance in [12], the 96% sensitivity rate seems quite successful; however, 7 seizure segments (with all frames within) were entirely missed, which might be critical for diagnosis.

3. EEG data processing

3.1. EEG dataset

The CHB-MIT scalp long-term EEG dataset [8] was recorded from pediatric patients (males with the ages varying between 3 and 22 and females with the ages varying between 3 and 19) with intractable seizures at Boston Children's Hospital. Table 1 presents the properties of seizure records (i.e., records with at least one seizure event) in this dataset. In this work, we selected 21 out of 24 patients excluding the patients 6, 12 and 16 because we failed to read some channel data from the EEG records of those patients. The total duration of the EEG data is 8756 min (~146 h) within which the total duration of the seizure segments is only around 165 min (<2.8 h). So even with the seizure record selection, the seizure segments cover only 1.88% of the EEG dataset, presenting a highly imbalanced data distribution, which makes the classification highly challenging. The sampling frequency of all recorded signals was 256 Hz with 16-bit resolution and each frame is annotated whether it is seizure or non-seizure. International 10–20 system for electrode positioning was used. The 18 processed channels are: FP1–F7, F7–T7, T7–P7, P7–O1, FP1–F3, F3–C3, C3–P3, P3–O1, FP2–F4, F4–C4, C4–P4, P4–O2, FP2–F8, F8–T8, T8–P8, P8–O2, FZ–CZ and CZ–PZ. In this study, EEG signals from all the channels are used to capture spatial distribution of EEG waveforms on the scalp within the feature vector. The time-resolution of each frame is 1 s.

3.2. Pre-processing and feature extraction

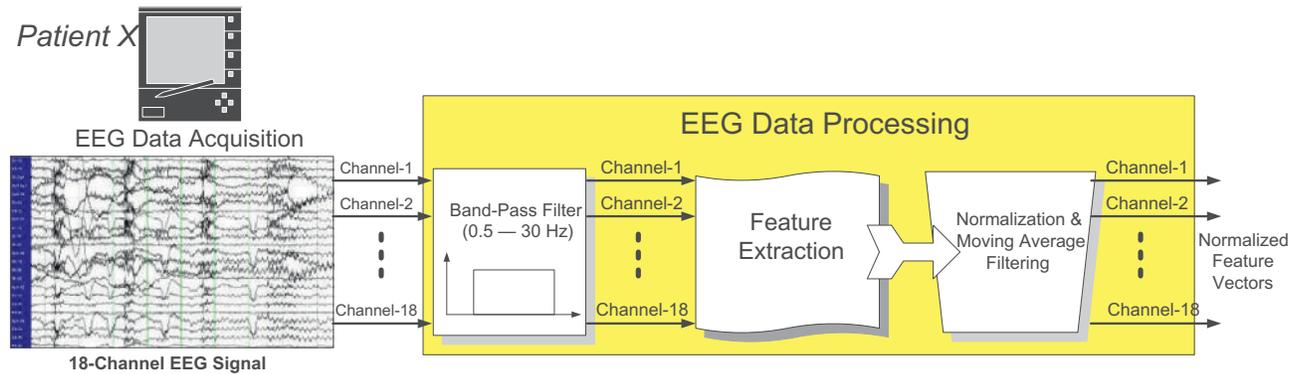
As the entire EEG data processing shown in Fig. 2, the EEG signal of each channel is first band-pass filtered between 0.5 and 30 Hz using a linear phase FIR filter with the Parks–McClellan algorithm [17]. This frequency band has been used in most seizure detection studies [16,18] to remove some of the well-known physiological and electrical noise and artifacts.

Four typical feature categories in EEG signal processing are time, frequency and time–frequency domain features and other

Table 1

The properties of the CHB-MIT scalp EEG benchmark dataset [8].

Pat. ID	Sex	Age	Total duration of seizure records (s)	Total duration of seizures (s)
1	F	11	23,925	442
2	M	11	8159	172
3	F	14	25,200	402
4	M	22	38,359	378
5	F	7	18,000	558
7	F	14.5	32,537	325
8	M	3.5	18,000	919
9	F	10	34,499	276
10	M	3	50,464	447
11	F	12	10,059	806
13	F	3	28,800	535
14	F	9	25,200	169
15	M	16	50,434	1992
17	F	12	10,824	293
18	F	18	20,274	317
19	F	19	10,546	236
20	F	6	20,036	294
21	F	11	13,790	199
22	F	13	10,800	204
23	F	9	32,254	424
24	F	6	43,200	511

**Fig. 2.** Pre-processing and feature extraction stages to obtain normalized feature vectors per EEG channel.

non-linear features. In this paper, in order to adopt a patient specific approach and to exploit the characteristics of any patient's EEG pattern, several features from all of these categories are extracted from each EEG channel. Once the EEG signal from a channel is filtered, it is then partitioned into non-overlapping frames with a 1 s duration from which the features are extracted. The first group of time domain features is morphological set used in [19], which contains 16 distinct features including, Latency time (LAT), Latency/amplitude ratio (LAR), Absolute amplitude (AAMP), Absolute latency/amplitude ratio (ALAR), Positive area (PAR), Negative area (NAR), Total area (TAR), Absolute total area (ATAR), Total absolute area (TAAR), Average absolute signal slope (AASS), Peak to peak (PP), Peak to peak time window (PPT), Peak to peak slope (PPS), Zero crossings (ZC), Zero-crossing density (ZCD) and Slope sign alteration (SSA).

Other time domain features extracted from each frame are: skewness, kurtosis, number of maxima and minima, root mean square, Shannon entropy [20,21], mean, variance, coefficient of variation [22,23], approximate entropy [21], energy and energy of auto-covariance.

The second group of features is in the frequency domain such as, maximum, minimum mean of power spectrum density, spectral entropy [16] and median frequency. In addition to such traditional features we also extract 6 Mel Frequency Cepstral Coefficients (MFCCs) [24], the first order and the second order derivatives of MFCCs from 5 frequency bands of 1–4 Hz, 4–8 Hz, 8–12 Hz,

12–20 Hz and 20–30 Hz. In [24], MFCCs were used to detect robust emotion in EEG signal. MFCCs are widely used in several speech and speaker recognition systems due to the fact that they provide a decorrelated, perceptually-oriented observation vector in the cepstral domain. First the incoming frames are Hamming windowed in order to enhance the harmonic nature. In addition, Hamming window can reduce the effects of discontinuities and edges that are introduced during the framing process. Especially in logarithmic domain, the windowing effects can be encountered significantly. In order to perform filtering in the time domain, the frame is zero-padded to get the size as a power of 2 and then FFT is applied to get into the spectral domain for plain multiplication with the filterbank. The mel (melody) scaled filterbank is a series of filterbank, which has the central frequencies uniformly distributed in mel-frequency ($\text{mel}(f)$) domain where,

$$\text{mel}(f) = m_f = 1127 \log \left(1 + \frac{f}{700} \right) \text{ and } f = 700 \left(e^{\frac{m_f}{1127}} - 1 \right). \quad (1)$$

Fig. 3 illustrates a sample mel-scaled filterbank in the frequency domain. The number of bands is reduced for the sake of clarity. The shape of the band filters in the filterbank can be Hamming window or plain triangular shaped. As clearly seen in Fig. 3 the resolution is high for low frequencies and low for higher frequencies. Once the filtering is applied, the energy is calculated per band and the Cepstral transform is applied on the band energy

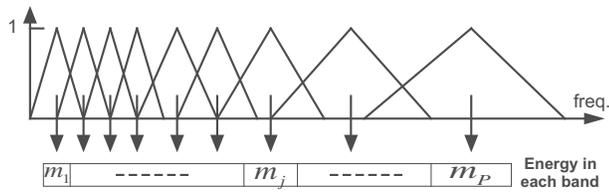


Fig. 3. The derivation of mel-scaled filterbank amplitudes.

values. Cepstral transform is a discrete cosine transform of log filterbank amplitudes:

$$c_i = (2/P)^{1/2} \sum_{j=1}^P \log m_j \cdot \cos\left(\frac{\pi \cdot i}{N} (j - 0.5)\right), \quad (2)$$

where $0 < i \leq P$ and P is the number of filter banks. A subset of c_i is then used as the feature vector for this frame.

The third group of features in time–frequency domain include, relative scale energy, Shannon entropy, and coefficient of variation for both approximation and detail coefficients in 5 levels with Daubechies wavelets (db1, db2, db3 and db4), frequency regularity index [26], maximum, minimum, variance, mean, number of extrema and energy [27] of both approximation and detail coefficients with db4. Also the energy of each Delta (1–4 Hz), Theta (4–7 Hz), Alpha (7–13 Hz) and Beta (13–30 Hz) band are calculated using Short-Time Fourier Transform (STFT). In addition, the energy of each Delta (1–4 Hz), Theta (4–7 Hz), Alpha (7–13 Hz) and Beta (13–30 Hz) band from the Wigner-Ville distribution of each frame are extracted. Lyapunov exponent is also extracted as a non-linear feature with the method used in [28]. All the extracted features according to their types are listed in Table 2.

After extracting the features from each of the 18 channels, all feature vectors were normalized between -1 and 1 as follows:

$$\begin{aligned} \min &= \overline{(x - 4\sigma(x))}, \quad \max = \overline{(x + 4\sigma(x))}, \quad \text{Normalized} \\ &= 2 \frac{x - \min}{\max - \min} - 1 \end{aligned} \quad (3)$$

where x is an individual feature value, $\sigma(x)$ is the standard deviation and the bar sign indicates the average value. Moreover, in order to smoothen the features and enhance the discrimination between the seizure and non-seizure segments, we applied a moving average filter over the feature vectors with a 20 s window with 1 s overlap. In Fig. 4 the energy of the Theta band (4–7 Hz) is shown before and after using the moving average filter. It is fairly obvious that after using the moving average filter the discrimination between the seizure and non-seizure segment is enhanced and the noise level in the non-seizure segment is significantly reduced. Note that in this example the seizure occurs between 2996 and 3036 s.

Table 2

List of extracted features.

Morphological	LAT, LAR, AAMP, ALAR, PAR, NAR, TAR, ATAR, TAAR, AASS, PP, PPT, PPS, ZC, ZCD, SSA
Time	Skewness, Kurtosis, No. maxima and minima, mean, variance, standard deviation, coefficient of variation, RMS, Shannon entropy, approximate entropy, energy; standard variation, variance and energy of auto-covariance
Frequency	Maximum, minimum, and mean of power spectrum, spectral entropy, median frequency
Time–frequency	Relative scale energy, Shannon entropy, coefficient of variation of 5 approximations and 5 details coefficients with db1, db2, db3, and db4, frequency regularity index, maximum, minimum, variance, mean, standard deviation, No. of extrema and energy of 5 approximations and 5 details using db4, energy of STFT in 4 frequency bands of delta (1–4 Hz), theta (4–7), alpha (7–13 Hz), and beta (13–30 Hz), energy in Winger-Ville distribution in 3 frequency bands of delta (1–4 Hz), theta (4–7), alpha (7–13 Hz), and beta (13–30 Hz)
Non-linear	Lyapunov exponent
MFCC	MFCC, first and second order derivative of MFCC coefficients in 5 frequency bands of 1–4 Hz, 4–8 Hz, 8–12 Hz, 12–20 Hz

4. Evolutionary neural networks

As mentioned earlier, evolutionary ANNs are used as binary classifiers (BC) within the CNBC topology that is primarily used for the classification of EEG data from each individual patient in the database. In this section, the MD PSO technique [9–11], which is used for evolving ANNs, will briefly be introduced and we shall present its application for evolving the feed-forward ANNs next.

4.1. Multi-dimensional particle swarm optimization

The Particle Swarm Optimization (PSO) was introduced by Kennedy and Eberhart [29] in 1995 as a population based stochastic search and optimization process. In a PSO process, a swarm of particles, each of which represents a potential solution to an optimization problem, navigate through the search space. The particles are initially distributed randomly over the search space and the goal is to converge to the global optimum of a function or a system. Each particle keeps track of its position in the search space and its best solution so far achieved. This is the personal best value (the so-called *pbest* in [29]) and the PSO process also keeps track of the global best solution so far achieved by the swarm with its particle index (the so called *gbest* in [29]). So during their journey with discrete time iterations, the velocity of each particle in the next iteration is computed by the best position of the swarm (position of the particle *gbest* as the *social* component), the best personal position of the particle (*pbest* as the *cognitive* component), and its current velocity (the *memory* term). Both *social* and *cognitive* components contribute randomly to the position of the particle in the next iteration.

As the evolutionary method, we shall use the multi-dimensional (MD) extension of the basic PSO (*bPSO*) method, the MD PSO, proposed in [11]. Instead of operating at a fixed dimension N , the MD PSO algorithm is designed to seek both positional and dimensional optima within a dimension range, $\{D_{\min}, D_{\max}\}$. In order to accomplish this, each particle has two sets of components, each of which has been subjected to two independent and consecutive processes. The first one is a regular positional PSO, i.e. the traditional velocity updates and due positional shifts in N dimensional search (solution) space. The second one is a dimensional PSO, which allows the particle to navigate through dimensions. Accordingly, each particle keeps track of its last position, velocity and personal best position (*pbest*) in a particular dimension so that when it re-visits the same dimension at a later time, it can perform its regular “positional” update using this information. The dimensional PSO process of each particle may then move the particle to another dimension where it will remember its positional status and will be updated within the positional PSO process at this dimension, and so on. The swarm, on the other hand, keeps track of the *gbest* particle in each dimension, indicating the best (global) position so far achieved. Similarly, the dimensional PSO process of each particle uses its personal best dimension in which the personal best fitness

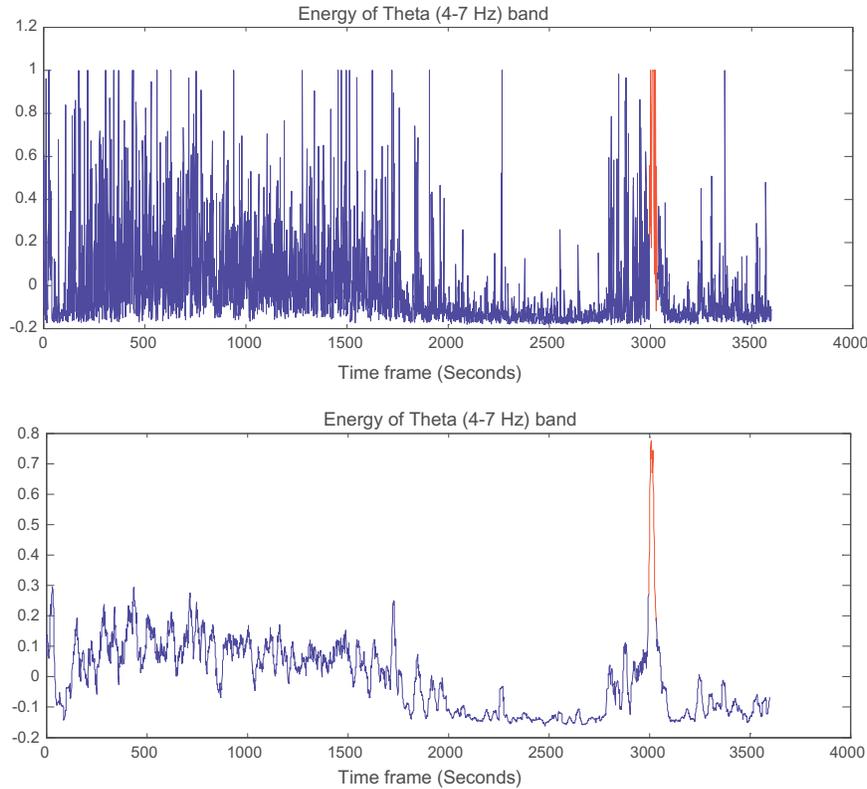


Fig. 4. Before (top) and after (bottom) using the moving average filter. The seizure segment is shown by red color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

score has so far been achieved. Finally, the swarm keeps track of the global best dimension, $dbest$, among all the personal best dimensions. The $gbest$ particle in the $dbest$ dimension represents the optimum solution and dimension, respectively.

In a MD PSO process at time (iteration) t , each particle a in the swarm with S particles, $\xi = \{x_1, \dots, x_a, \dots, x_S\}$, is represented by the following characteristics:

- $xx_{aj}^{xd_a(t)}(t)$: j th component (dimension) of the position of particle a , in dimension $xd_a(t)$.
- $vx_{aj}^{xd_a(t)}(t)$: j th component (dimension) of the velocity of particle a , in dimension $xd_a(t)$.
- $xy_{aj}^{xd_a(t)}(t)$: j th component (dimension) of the personal best position of particle a , in dimension $xd_a(t)$.
- $gbest(d)$: Global best particle index in dimension d .
- $x\tilde{y}_j^d(t)$: j th component (dimension) of the global best position of swarm, in dimension d .
- $xd_a(t)$: Dimension component of particle a .

- $vd_a(t)$: Velocity component of dimension of particle a .
- $xd_a(t)$: Personal best dimension component of particle a .

Let f denote the fitness function that is to be optimized within a certain dimension range, $\{D_{min}, D_{max}\}$. Without loss of generality assume that the objective is to find the minimum of f at the optimum dimension within a multi-dimensional search space. Assume that the particle a visits (back) the same dimension after T iterations (i.e. $xd_a(t) = xd_a(t+T)$), then the personal best position can be updated in iteration $t+T$ as follows,

$$xy_{aj}^{xd_a(t+T)}(t+T) = \begin{cases} xy_{aj}^{xd_a(t)}(t) & \text{if } f(xx_a^{xd_a(t+T)}(t+T)) > f(xy_a^{xd_a(t)}(t)) \\ xx_{aj}^{xd_a(t+T)}(t+T) & \text{else} \end{cases} \quad j = 1, 2, \dots, xd_a(t+T) \quad (4)$$

Furthermore, the personal best dimension of particle a can be updated in iteration $t+1$ as follows,

$$x\tilde{d}_a(t+1) = \begin{cases} x\tilde{d}_a(t) & \text{if } f(xx_a^{xd_a(t+1)}(t+1)) > f(xy_a^{x\tilde{d}_a(t)}(t)) \\ xd_a(t+1) & \text{else} \end{cases} \quad (5)$$

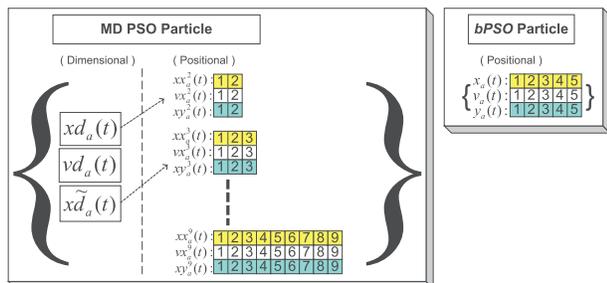


Fig. 5. Sample MD PSO (right) vs. $bPSO$ (left) particle structures. For MD PSO $\{D_{min} = 2, D_{max} = 9\}$ and at time t , $xd_a(t) = 2$ and $x\tilde{d}_a(t) = 3$.

Fig. 5 shows sample MD PSO and $bPSO$ particles denoted as a . Particle a in $bPSO$ particle is at a (fixed) dimension, $N = 5$, and contains only positional components; whereas in MD PSO particle a contains both positional and dimensional components, respectively. In the figure the dimension range for MD PSO is given by $\{D_{min}, D_{max}\} = \{2, 9\}$, therefore the particle contains 9 sets of positional components. In this example the particle a currently resides at dimension 2 ($xd_a(t) = 2$); whereas its personal best dimension is 3 ($x\tilde{d}_a(t) = 3$). Therefore, at time t a positional PSO update is first performed over the positional elements, $xx_a^{xd_a(t)}(t)$, and then the

particle may move to another dimension with respect to the dimensional PSO. Recall that each positional element, $xx_a^j(t)$, $j \in \{0, 1\}$, represents a potential solution in the data space of the problem. Further details and the pseudo-code of the MD PSO can be found in [9].

4.2. MD PSO for evolving MLPs

As a stochastic search process in multi-dimensional search space, MD PSO seeks (near-) optimal networks in an architecture space (AS), which can be defined over any type of ANNs with any properties. In this work, the focus is particularly drawn on automatic design of the multilayer perceptrons (MLPs). All network configurations in the AS are enumerated into a hash table with a proper hash function, which ranks the networks with respect to their complexity, i.e. associates higher hash indices to networks with higher complexity. MD PSO can then use each index as a unique dimension of the search space where particles can make inter-dimensional navigations to seek an optimum dimension ($dbest$) and the optimum solution on that dimension, xy^{dbest} . The former corresponds to the optimal architecture and the latter encapsulates the optimum network parameters (connections, weights and biases). Suppose for the sake of simplicity, a range is defined for the minimum and maximum number of layers, $\{L_{min}, L_{max}\}$ and number of neurons for the hidden layer l , $\{N_{min}^l, N_{max}^l\}$. The sizes of both input and output layers are determined by the problem and hence fixed. The AS can then be defined only by two range arrays, $R_{min} = \{N_i, N_{min}^1, \dots, N_{min}^{L_{max}-1}, N_o\}$ and $R_{max} = \{N_i, N_{max}^1, \dots, N_{max}^{L_{max}-1}, N_o\}$, one for minimum and the other for the maximum number of neurons allowed for each layer of a MLP. The size of both arrays is naturally $L_{max} + 1$ where the corresponding entries define the range of the l th hidden layer for all those MLPs, which can have an l th hidden layer. The size of the input and output layers, $\{N_i, N_o\}$, is fixed and is the same for all configurations in the AS. $L_{min} \geq 1$ and L_{max} can be set to any value meaningful for the problem encountered. The hash function then enumerates all potential MLP configurations into hash indices, starting from the simplest MLP with $L_{min} - 1$ hidden layers, each of which has minimum number of neurons given by R_{min} , to the most complex network with $L_{max} - 1$ hidden layers, each of which has a maximum number of neurons given by R_{max} .

Let N_h^l be the number of hidden neurons in layer l of a MLP with input and output layer sizes N_i and N_o , respectively. The input neurons are merely fan-out units since no processing takes place. Let F be the activation function applied over the weighted inputs plus a bias, as follows:

$$y_k^{p,l} = F(s_k^{p,l}) \text{ where } s_k^{p,l} = \sum_j w_{jk}^{l-1} y_j^{p,l-1} + \theta_k^l \quad (6)$$

and where $y_k^{p,l}$ is the output of the k th neuron of the l th hidden/output layer when the pattern p is fed, w_{jk}^{l-1} is the weight from the j th neuron in layer $l - 1$ to the k th neuron in layer l , and θ_k^l is the bias value of the k th neuron of the l th hidden/output layer. The training Mean Squared Error, MSE , is formulated as follows:

$$MSE = \frac{1}{2PN_o} \sum_{p \in T} \sum_{k=1}^{N_o} (t_k^p - y_k^{p,o})^2, \quad (7)$$

where t_k^p is the target (desired) output and $y_k^{p,o}$ is the actual output from the k th neuron in the output layer, $l = o$, for pattern p in the training dataset T with size P , respectively. Suppose that particle a , has the positional component formed as, $xx_a^{xd_a(t)}(t) = \Psi^{xd_a(t)} \{ \{w_{jk}^0\}, \{w_{jk}^1\}, \{w_{jk}^2\}, \{w_{jk}^{o-1}\}, \{w_{jk}^o\}, \{0_k^0\}, \{0_k^1\}, \dots, \{0_k^{o-1}\}, \{0_k^o\} \}$ at a time t , where $\{w_{jk}^l\}$ and $\{0_k^l\}$ represent the sets of weights and biases of the layer l of the MLP configuration, $\Psi^{xd_a(t)}$. Note that the input layer ($l = 0$) contains only weights whereas the output layer ($l = o$) has only biases. By means of such a direct encoding scheme, particle a

represents all potential network parameters of the MLP architecture at the dimension (hash index) $xd_a(t)$. As mentioned earlier, the dimension range, $\{D_{min}, D_{max}\}$, where MD PSO particles can make inter-dimensional jumps, is determined by the AS defined. Apart from the regular limits such as (positional) velocity range, $\{V_{min}, V_{max}\}$, dimensional velocity range, $\{VD_{min}, VD_{max}\}$, the data space can also be limited with some practical range, i.e. $X_{min} < xx_a^{xd_a(t)}(t) < X_{max}$. Setting MSE in Eq. (7) as the fitness function enables MD PSO to perform evolutions of both network parameters and architectures within its native process. Further details and an extensive set of experiments demonstrating the optimality of the networks evolved with respect to several benchmark problems can be found in [10,11].

5. EEG classification system

This section describes in detail the proposed EEG classification system starting from the core classifier topology: the Collective Network of (Evolutionary) Binary Classifiers (CNBC), which uses Neurologist's labels as the Ground Truth Data (GTD) in the training dataset to configure its internal structure and to evolve its binary classifiers (BCs) individually. Before going into details of the topological characteristics and the evolutionary process of the CNBC, a general overview of the proposed classification system will first be introduced in the next sub-section.

5.1. The system overview

As shown in Fig. 6, the long-term EEG records are divided into individual channels and into 1 s time frames and then as described in detail in Section 3, EEG data processing in a sequential order of pre-processing, feature extraction and post-processing operations are applied to each frame. From each EEG channel, normalized and smoothed features in the encapsulating feature vectors (FVs) of each time frame are then fed to the CNBC ensemble (CNBC-E) that has already been created and evolved specially for the patient. Note that as a "divide-and-conquer" paradigm, the proposed system for long-term EEG classification divides the problem into many "manageable" parts and the division is performed over EEG channels, features, temporal frames and binary classifiers. Such a broad division is managed mainly within each CNBC, which will be covered in the next section. Accordingly, the CNBC-E strives for minimizing the misclassifications of the epileptic seizure frames. In other words, the purpose of the *ensemble* structure is to capture and aggregate the "seizure" decisions from each CNBC, i.e., any frame is *initially* classified as "seizure" if at least one CNBC classifies it as such and thus epileptic seizures are aggregated and then morphologically filtered as the final post-processing step. Each CNBC is individually (and independently) evolved and in order not to miss any seizure frame out, achieving *divergent* decision boundaries among them is crucial. Therefore, it is of utmost importance to perform evolutionary search process based on stochastic optimization technique, such as MD PSO; otherwise, it is obvious that static classifiers such as SVMs or Random Forest (RF) cannot provide such a divergence since identical decision boundaries will result due to their deterministic training methods.

Morphological filtering (MF) is a process to filter out noisy outliers in seizure and non-seizure segments. For instance one or few seizure frames in a non-seizure segment (or vice versa) are obviously the classification noise and hence should be removed. In order to accomplish this we develop a fuzzy rule-based morphological filtering technique, which operates according to the morphological structure of both segment types such as *continuity*, *neighborhood similarity* (except the boundaries) and *closure*

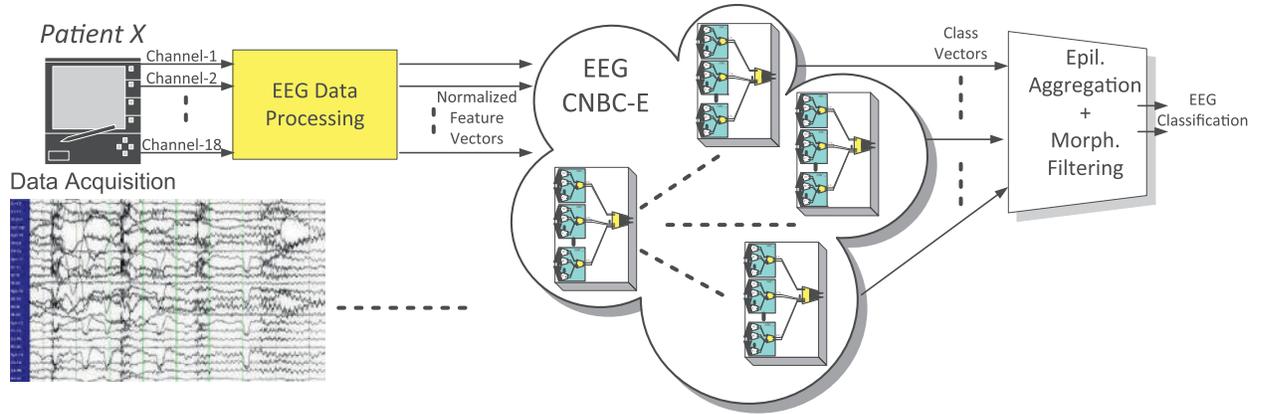


Fig. 6. The illustration of the proposed EEG classification system.

(seizure segments have a finite duration). So after the last block shown in Fig. 6, the aim is to produce continuous and noise free segments for both types. In order to accomplish this, a consecutive application of morphological filtering is performed after the epileptic seizure aggregation (ESA) over the EEG classification array of each CNBC output. In order to address global properties such as *continuity* and *neighborhood similarity*, we design MF in a top-down approach, which operates on the analysis windows rather than the frames. Each analysis window will then be classified as non-seizure (N), seizure, (S) or fuzzy (F) according to its seizure frame count. Before going into details of the ESA and MF, we shall first define the terminology used for the classification of the EEG record of the patient p with duration of T_p seconds by a CNBC-E with N_{CNBC} number of CNBCs, as follows:

- W_S : window duration in seconds,
- $W_N = [T_p W_S]$: number of analysis windows in the EEG record.
- $\psi_{CNBC}^c(i)$: the classification of the i th frame, $i \in [1, T_p]$, as either N or S by the c th CNBC, $c \in [1, N_{CNBC}]$.
- $\psi(i)$: the final classification of the i th frame as either N or S after ESA and MF operations.
- $W(j)$: the type of the j th analysis window, $\forall j \in [1, W_N]$, as the non-seizure (N), seizure, (S) or fuzzy (F),
- N_W^j : number of frames in ψ classified as N in the j th analysis window.
- S_W^j : number of frames ψ classified as S in the j th analysis window,
- ΔW : number of analysis windows for verifying the *continuity* and *neighborhood similarity*.

Once the c th CNBC is evolved and then used to classify the EEG record, the resultant classification array, $\psi_{CNBC}^c(i), \forall i \in [1, T_p]$, can then be used to update the final classification array, $\psi(i), \forall i \in [1, T_p]$.

Accordingly, the pseudo-code of the consecutive application of ESA and MF operations after the c th CNBC classification is given in Table 3. The only input variable is c where the consecutive operations of ESA + MF are performed for $\forall c \in [1, N_{CNBC}]$. The input parameters are the analysis window properties, W_S , W_N , and ΔW . The first loop between steps 1 and 2 aggregates the epileptic seizure segments within $\psi_{CNBC}^c(i)$ to $\psi(i)$. The rest of the steps 3–6 then perform the MF operation. The loop between steps 3 and 4 divides the EEG record into analysis windows, $W(j) \forall j \in [1, W_N]$, and assigns their class types as either N , S or F . The windows with no seizure frames ($S_W^j = 0$) are classified as the non-seizure (N) type whereas the others are assigned as either fuzzy (F) or seizure (S). Note that these are only the classification of the analysis windows, not the frames of the record, which will then be classified by

Table 3

Pseudo-code of epileptic seizure aggregation (ESA) and morphological filtering (MF).

```

ESA + MF ( $c, W_S, W_N, \Delta W$ ).
1. For  $\forall i \in [1, T_p]$  do:
  1.1. If  $(\psi_{CNBC}^c(i) = S)$  then  $\psi(i) \leftarrow S$ 
2. End For.
3. For  $\forall j \in [1, W_N]$  do:
  3.1. Compute  $N_W^j$  and  $S_W^j$  where  $N_W^j + S_W^j = W_S$ 
  3.2. Assign  $W(j)$  as:  $W(j) = \begin{cases} N & \text{if } S_W^j = 0 \\ S & \text{if } S_W^j > N_W^j \\ F & \text{if } S_W^j \leq N_W^j \text{ and } S_W^j \neq 0 \end{cases}$ 
4. End For.
5. For  $\forall j \in [\Delta W + 1, W_N - \Delta W]$  do:
  5.1. If  $(W(k) \neq S \forall k \in [j - \Delta W, j + \Delta W] - \{j\})$ 
    5.1.1. Then  $\psi(i) \leftarrow N \forall i \in [(j - 1)W_S + 1, jW_S]$ 
  5.2. Else if  $\exists k1, k2: W(k1) = S \& W(k2) = S \forall k1 \in [j - \Delta W, j - 1] \& \forall k2 \in [j + 1, j + \Delta W]$ 
    5.2.1. Then  $\psi(i) \leftarrow S \forall i \in [(j - 1)W_S + 1, jW_S]$ 
  End For.

```

verifying these analysis windows for *continuity* and *neighborhood similarity*. This is basically performed in the loop between the steps 5 and 6, over the classified analysis windows, $W(j)$. The step 5.1 verifies whether or not a window $W(j)$ -regardless of its initial classification- has an S type window within the ΔW neighborhood. If not, according to the *neighborhood similarity* rule, it cannot be an S or F type window even though if it was initially classified as such and thus all S frames within – if exist – should be filtered out realizing that they are the classification noise encountered within ψ_{CNBC}^c . In this case they are all assigned as type N . On the other hand, the step 5.2 verifies whether or not the window $W(j)$ -also regardless from its initial classification- has both S type windows within the ΔW neighborhood. If so, according to the *continuity* rule, the window $W(j)$ should be an S type window even though if it was not initially classified as such and thus all N frames within –if exist- should be filtered out realizing that they are, too, the classification noise. The fuzzy (F type) windows are not conclusive at this stage due to the presence of seizure frames that are in the minority within the window; however, with the integration of other CNBC classification results, seizure frames can be aggregated via ESA operations and hence this F window can be transformed to an S window. Note further that the MF operation basically strives for detecting and thus forming *continuous* seizure segments and therefore, any S or F type window is kept intact except whenever it only has N type neighbors. We empirically set $W_S = 9$ s so that the minimum seizure duration is required to be 5 s in order to assign an analysis window as S type. Moreover, we set $\Delta W = 2$ so that

a minimum two neighbor windows are required for the verification of analysis windows for *continuity* and *neighborhood similarity*. This allows a sufficient range of $2 \times 9 = 18$ s that can be aggregated by the later ESA operations to fill seizure segments within the range. Therefore, if there is a seizure section with one or more *S* type windows already detected, the undetected *N* or *F* type windows can thus be filled within this range.

A sequence of ESA and MF operations is shown in Fig. 7 over the classification arrays of the first two CNBCs of the CNBC-E for the Patient 5. The classification map on the top is the ground-truth labeling by the Neurologist and note that the map is shown in a 2D array as a B/W image where black and white pixels represent the *N* and *S* type frames (each with duration of 1 s), respectively. Each row of the image represents 4 min of the EEG record and therefore, the width of the image is $4 \times 60 = 240$ pixels and the total duration of the EEG recording is 4 h, therefore, the height of the images are all $4 \times 60/4 = 60$ pixels. As the first CNBC, the classification map of the CNBC-1 is subject to only MF operation, which removes some noisy (isolated) *N* type frames and merged some *S* type frames. The resultant classification map shown in the left-bottom side of the figure is then epileptic seizure aggregated (ESA) using the classification map of the CNBC-2, and as a result some of the missing seizure frames are now aggregated; however, some noisy ones are too emerged as clearly seen in the classification map on right-top. The following MF operation then successfully filters those noisy *S* frames as shown in the right-bottom and then the ESA operation is, once again, performed using the classification map of CNBC-2, and so on.

5.2. Collective network of binary classifiers

5.2.1. The topology

The long term EEG recordings are inevitably the source of massive amount of data in high dimensions due to multiple channels, and several features extracted. Therefore, the main motivation behind the CNBC topology is to divide this massive problem into several “manageable” parts each of which can be “learned” by a simple classifier optimized for this purpose by the evolutionary technique in order to maximize the performance. This is indeed a

“Divide and Conquer” type approach which uses as many classifiers as necessary, so as to divide such a large-scale learning problem into many NBC units along with the binary classifiers (BCs) within, and thus prevent the need of using complex classifiers as the learning (both training and evolution) performance degrades significantly as the complexity rises due to the well-known *curse of dimensionality* phenomenon. Recall from the earlier discussion that the proposed EEG classification framework based on CNBC is fully generic, which neither requires any *a priori* information about the patient such as the list of relevant EEG channels, nor manually selects the most discriminative features among the large set. Therefore, it is a crucial design objective to select or more accurately to “learn” the relevancy of each EEG channel and each feature so as to “weight” them accordingly during the classification process.

As the CNBC topology shown in Fig. 8, each NBC corresponds to a *unique* EEG channel and strives to learn and classify the signal using the FVs only from that channel. The output of each NBCs will then be learned and weighted accordingly by a dedicated classifier, the “master fuser BC”, which will make the final classification decision of each EEG frame. Note that this is the *fusion* of the individual decisions by each NBC unit, which *divides* the problem per channel. A similar *division–fusion* paradigm also exists; this time for the features within the NBCs. Each NBC contains a set of evolutionary *binary classifiers* (BCs) in the input layer where each BC performs binary classification over an individual FV. In this “feature division” scheme, each FV of a particular feature extraction method is only fed to its corresponding BC in each NBC. Furthermore, such a flexible scheme allows the number of BCs varies, and whenever a new FV is extracted by a new method, its corresponding BC will be created, evolved and inserted into each NBC, yet keeping other BCs intact. In this way *scalability* with respect to varying number of features is achieved and the overall system can adapt to the change without requiring re-forming and re-evolving from scratch. On the fusion side, each NBC has a “fuser BC” in its output layer, which *fuses* the binary outputs of all BCs in the input layer and generates a single binary output, indicating the classification decision of each EEG frame according to the NBC’s corresponding EEG channel.

As shown in Fig. 8, 1-of-*n* encoding scheme is applied in all BCs, therefore, the output layer size of all binary classifiers is always 2

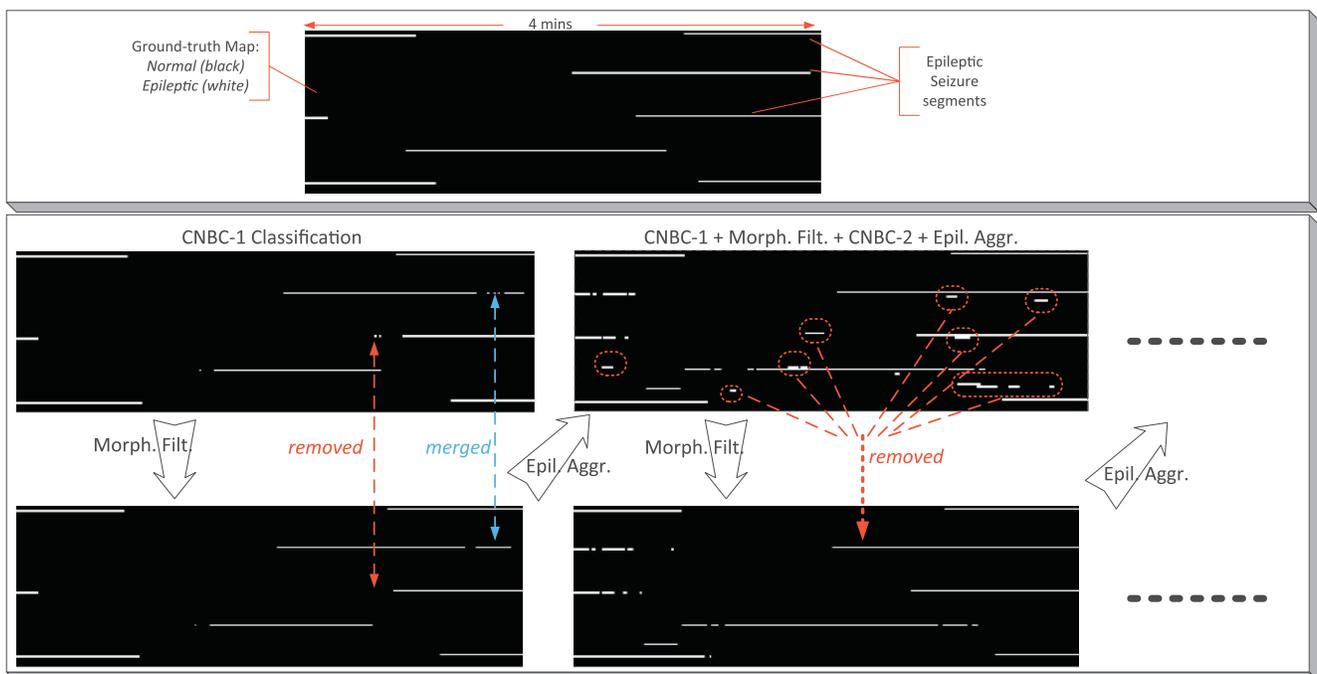


Fig. 7. ESA and MF operations over the two CNBCs’ classification arrays of the EEG record for the patient 5 in the benchmark CHB-MIT scalp EEG dataset.

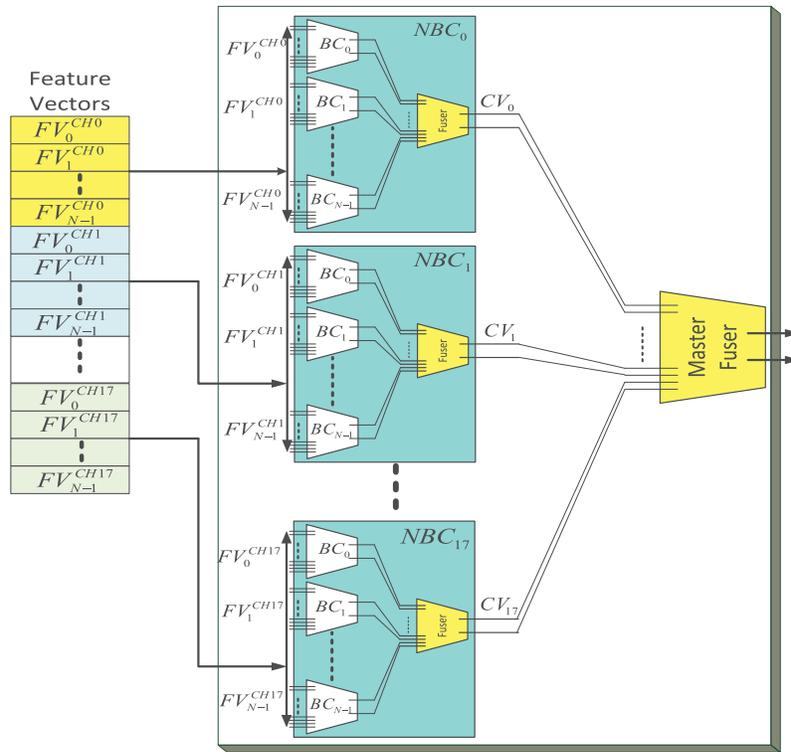


Fig. 8. Topology of the proposed CNBC framework for 18 EEG channels and N FVs.

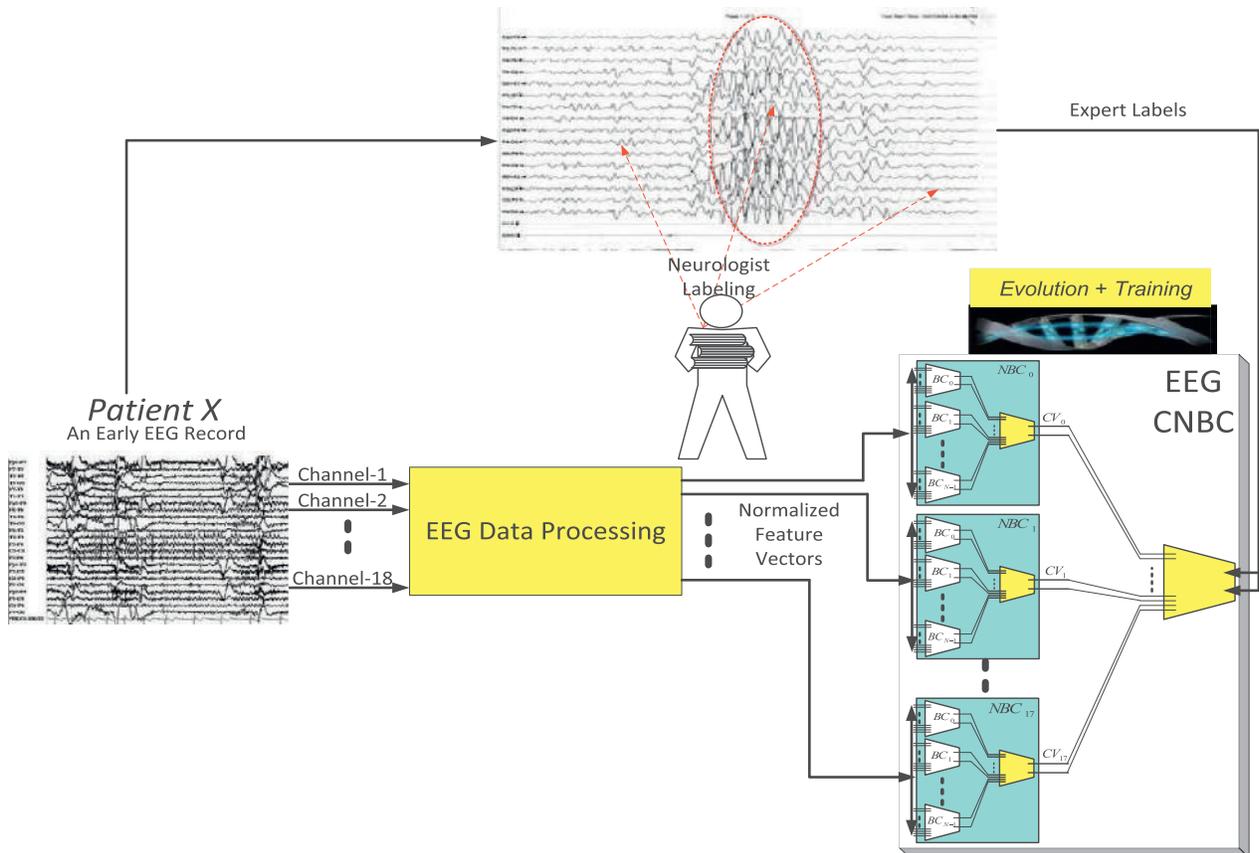


Fig. 9. The illustration of the evolution process of a CNBC.

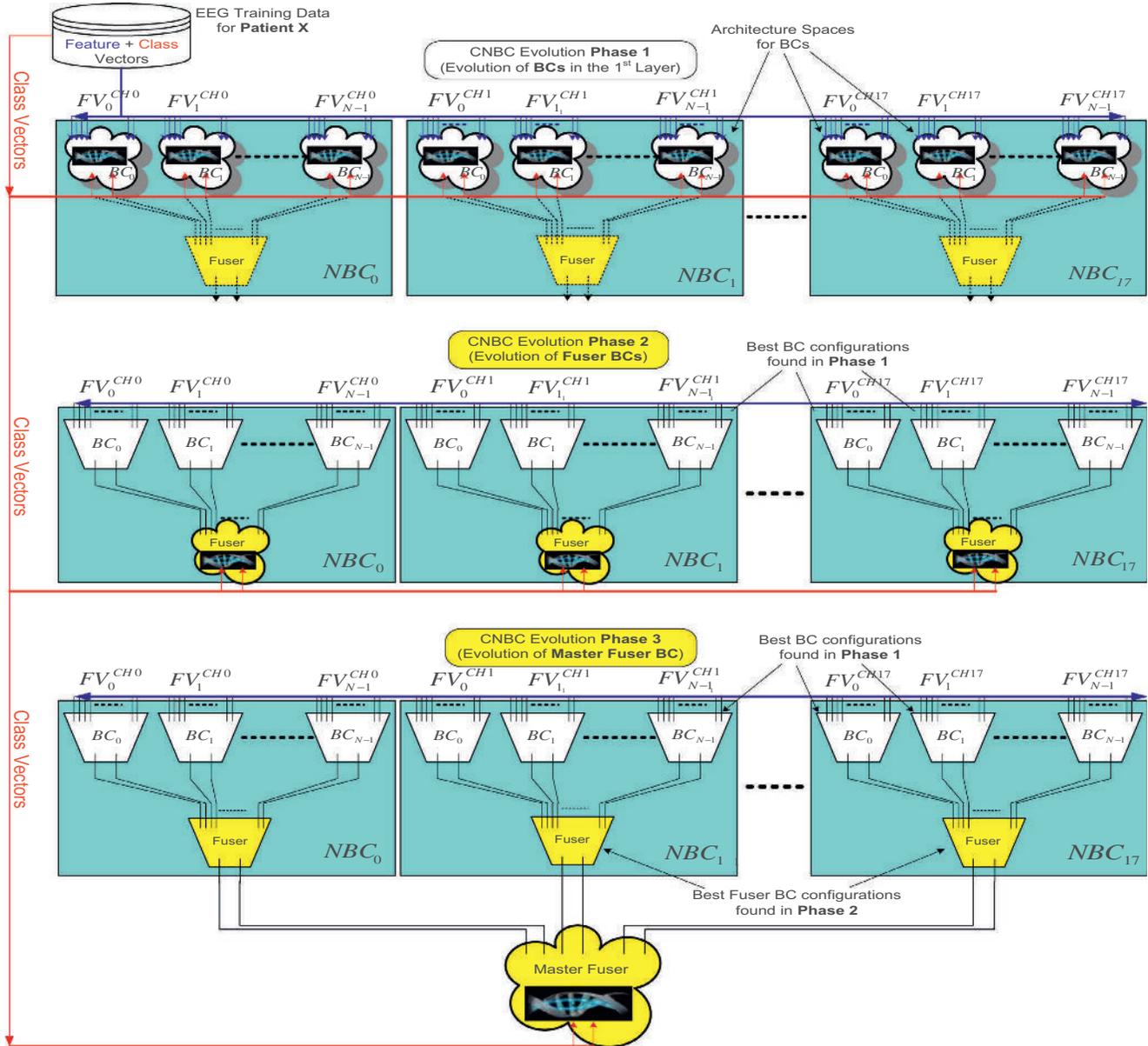


Fig. 10. Illustration of the three-phase evolution process for CNBC.

(i.e. $n = 2$). Let $CV_{c,1}$ and $CV_{c,2}$ be the first and second output of the c th (EEG channel's) fuser BC's class vector (CV). For each frame, its FVs of each channel are fed into the corresponding NBC in the CNBC. Each FV drives through (via forward propagation) its corresponding BC in the input layer of the NBC. The outputs of these binary classifiers are then fed to the fuser binary classifier of each NBC to produce the combined set of CVs in 36-D, i.e., $\{CV_{c,i}\}, \forall c \in [1, 18], i \in \{1, 2\}$, that are finally fed into the master fuser BC as the input. Its output is the CV indicating the classification of the input EEG frame. For the class selection over the CVs, 1-of- n encoding scheme is performed by comparing the individual outputs, e.g. say a seizure (S type) classification if $CV_{c,2} > CV_{c,1}$, and vice versa for the N type.

As a result such a division-fusion scheme implemented within the CNBC topology voids the need of selecting manually or having the *a priori* knowledge of the relevant channels and features in order to avoid the curse of dimensionality phenomenon. Besides this, another major benefit of this approach is that the configurations in the AS can be kept as *compact* as possible avoiding unfeasibly large storage and computation time requirements. Therefore, evolving

the CNBC may be further reduces since all the binary classifiers can be individually evolved by a separate process. This basically leads to the possibility of parallel processing that can be implemented on a Grid or Cloud computation environment [30]. We shall cover the details of the CNBC evolution next.

5.3. Evolution of the CNBC

The evolution of the entire CNBC-E is performed for each CNBC and all binary classifiers (BCs) in a CNBC are evolved using the same set of Neurologist's labeling of an early EEG record of that patient, which constitutes the *training dataset*. As illustrated in Fig. 9, using the Feature Vectors (FVs) extracted from each EEG channel along with the target Class Vectors (CVs) of the training dataset, the evolution process of each BC in a NBC is performed within the AS in order to find the best BC configuration with respect to a given criterion (e.g. training/validation MSE or CE). In 1-of- n encoding scheme, the target CVs for N and S type frames are set as, $CV_{c,1} = \{1, -1\}$ and $CV_{c,1} = \{-1, 1\}$, respectively that are set according to the labels of the Neurologist of each frame.

The evolution of the entire CNBC-E is performed for each CNBC individually in three consecutive phases, as illustrated in Fig. 10. In Phase 1, see top of Fig. 10, BCs in the first level of each NBC are evolved using the FVs extracted from the frames of its associated channel and a target CVs of the training dataset according to the frame type (N or S). Once the evolution process in Phase 1 is completed for all BCs in the input layer, the best BC configurations evolved by the MD PSO process are used to forward propagate all FVs of all the EEG frames in the training dataset to produce the combined set of CVs that is then fed into the fuser BC as an intermediate FV. In Phase 2, as shown in Fig. 10 in the middle, the fuser BCs of each NBC are evolved using these intermediate FVs and the same target CVs of the training dataset. As in Phase 1, the best configurations evolved will then be used as the fuser BCs. Note that in both Phase 1 and 2, the master fuser BC is excluded and hence not illustrated in the block diagrams (top and middle) of Fig. 10. When both phases are completed, all NBCs are then ready (with the best BCs and fuser BCs) and the FVs can now be forward propagated through them to produce the combined set of CVs in 36-D that can be fed to the master fuser BC. In Phase 3 the master fuser BC is finally evolved in a similar manner, using the combined CVs as

the intermediate FVs and the same target CVs of the training dataset.

In both Phases 2 and 3, apart from the difference in the generation of the FVs, the evolutionary method of the fuser BCs is same as any other BC has in the input layer. As discussed earlier, in phase 2, the fuser binary classifier learns the *significance* of each individual binary classifier (and its feature) over that particular channel. This can be viewed as a crucial way of applying a weighted *feature selection* scheme as some FVs may be quite discriminative for some patients and channels whereas others may not, and the fuser, if properly evolved, can “weight” each binary classifier (with its FV), accordingly. In this way the usage of each feature (and its BC) shall optimally be “fused” according to their discrimination power of each class. Similarly, the master fuser BC learns the significance (or discrimination capability) of each EEG channel and weights them accordingly while making the classification decision over these channels. For instance NBC of an associated channel, which generates irrelevant (or indiscriminant) outputs (CVs) is most likely discarded or weighted less among others by the master fuser BC, and vice versa for a NBC which can discriminate well between N and S type frames. Finally, note that each binary classifier in the first layer shall in time learn the significance of individual feature components of the corresponding FV for the discrimination of its class. In short the CNBC, if properly evolved, shall learn the significance of each EEG channel and the individual features with the best classifiers dedicated for the task.

As it is evident from Table 1, the amount of seizure and non-seizure frames is highly unbalanced and such an *unbalanced* numbers of positive (seizure) and negative (non-seizure) samples may cause a bias problem during the evolution process, i.e. for every positive sample; there will be a large number of negative samples which may bias the classifier (class imbalance problem). In order to prevent this, a negative sample selection is performed

Table 4
The architecture space (AS) used for MLPs.

Dim.	Conf.	Dim.	Conf.	Dim.	Conf.
0	$N_i \times 2$	6	$N_i \times 13 \times 2$	12	$N_i \times 19 \times 2$
1	$N_i \times 8 \times 2$	7	$N_i \times 14 \times 2$	13	$N_i \times 20 \times 2$
2	$N_i \times 9 \times 2$	8	$N_i \times 15 \times 2$	14	$N_i \times 21 \times 2$
3	$N_i \times 10 \times 2$	9	$N_i \times 16 \times 2$	15	$N_i \times 22 \times 2$
4	$N_i \times 11 \times 2$	10	$N_i \times 17 \times 2$	16	$N_i \times 23 \times 2$
5	$N_i \times 12 \times 2$	11	$N_i \times 18 \times 2$	17	$N_i \times 24 \times 2$

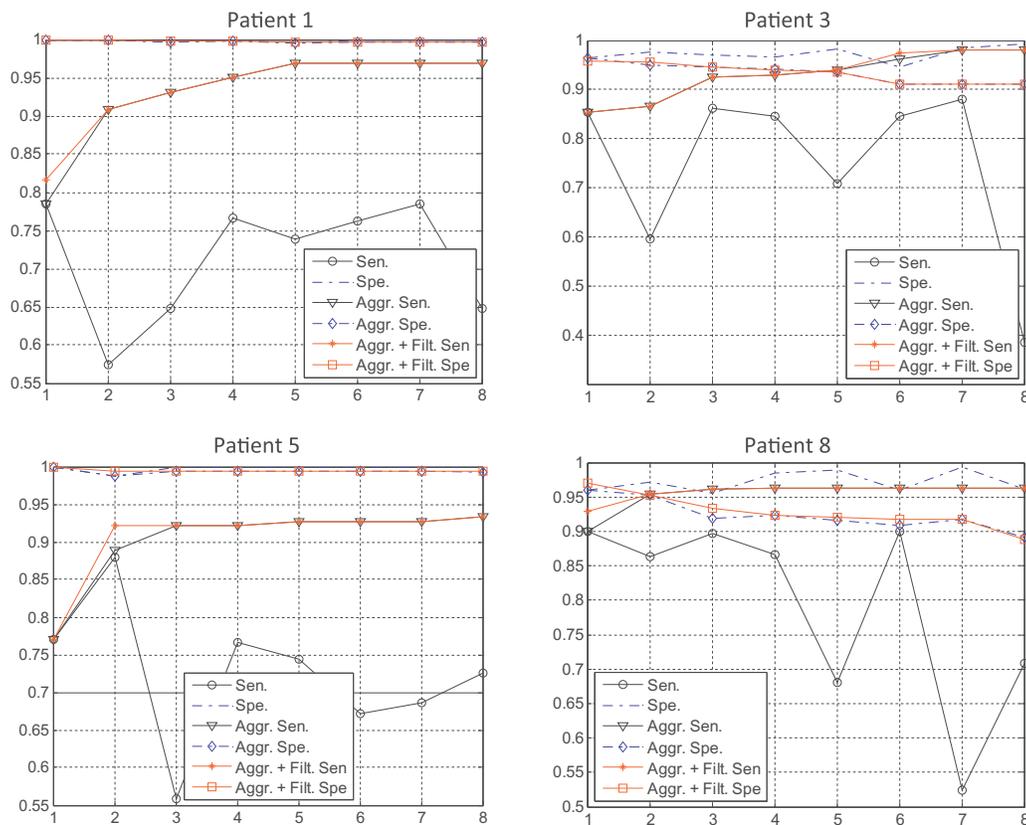


Fig. 11. The plots of sensitivity (Sen.) and specificity (Spe.) of the original, ESA and ESA + MF applied versus CNBC number obtained by the CNBC-E with 8 CNBCs for the 4 patients.

in such a way that for each positive sample, number of the negative samples (per positive sample) will be limited according to a pre-determined Positive-to-Negative Ratio (PNR). Selection of the negative samples is performed with respect to the closest proximity to the positive sample so that the classifier can be evolved by discriminating those negative samples (from the positive sample) that have the highest potential for the false-positive. Therefore, if properly evolved, the classifier can draw the “best possible” boundary between the positive and (PNR number of) negative samples, which shall in turn improve the classification accuracy. Besides the accuracy improvement, since the number of negative samples (non-seizure frames) is significantly reduced, this will also speed up the evolution process and reduce the computational cost.

6. Experimental results

In this section we shall present the overall results obtained from the EEG classification experiments over the CHB-MIT Scalp benchmark long-term EEG dataset. First we randomly select four

patients and use a total of 21 h of EEG records to demonstrate the sensitivity and specificity plots obtained from the proposed CNBC-E. Then we shall perform comparative evaluations against the state-of-the-art classification methods in this field, all of which use the same features and training dataset. Finally, computational complexity analysis will be performed for the evolution of the entire CNBC-E.

Recall that our main objective is to maximize sensitivity with an acceptable level of specificity using the *minimum* Neurologist labels from the early EEG records. To mimic this objective, in all experiments in this section, we set the training rate as 25%, which means that the proposed system (and all the competing techniques) is trained with the early sections of the long-term EEG records using only the 25% of the seizure and non-seizure frames and tested with the remaining (75%) of the patient’s record. Recall that the total duration of the seizure sections is only around 165 min, which makes on the average of ~7.9 min per patient. Therefore, we, on the average, used less than 2 min of seizure data per patient, and with a PNR setting as 12, we practically used

Table 5

The sensitivity rates (%) of the proposed system and 8 competing methods using the best SVM classifier over the original features and features selected by 7 different techniques.

Patients	Proposed	Original	CMIM	FCBF	MIFS	MIM	MRMR	JMI	DISR
Patient-1	94.66	88.13	87.54	70.62	63.5	80.71	76.56	91.4	89.61
Patient-2	98.48	89.39	90.91	84.09	85.61	98.48	88.64	97.73	96.21
Patient-3	97.07	87.3	91.53	80.78	82.08	90.88	81.43	93.16	90.23
Patient-4	71.78	34.84	40.42	28.92	48.43	94.77	37.28	75.96	49.83
Patient-5	96.93	76.83	80.14	78.72	75.18	80.38	80.61	80.85	81.09
Patient-7	90.24	68.7	64.63	49.19	54.07	2.80	59.35	58.53	60.57
Patient-8	95.67	77.49	85.43	63.92	64.21	67.67	72.58	83.55	84.56
Patient-9	100	93.33	92.86	69.05	89.05	99.52	93.33	100	93.33
Patient-10	92.38	94.13	92.96	93.84	91.5	91.5	93.55	93.84	93.84
Patient-11	95.88	62.11	49.26	41.85	56.67	57.83	40.03	63.76	51.89
Patient-13	86.13	50.85	43.55	56.45	36.74	53.04	43.55	52.8	53.28
Patient-14	81.95	75.94	75.19	60.15	66.17	76.69	74.44	75.19	74.44
Patient-15	86.81	80.12	76.14	66.87	75.29	81.97	71.57	80.78	80.58
Patient-17	63.51	2.64	2.31	4.94	45.3	2.14	11.04	3.13	21.75
Patient-18	81.07	3.05	63.37	24.69	20.98	21.4	27.16	76.13	85.6
Patient-19	98.89	68.33	64.44	58.33	68.89	76.11	65.56	75.55	58.33
Patient-20	93.01	63.75	72.93	69.43	61.57	75.98	69	75.11	72.49
Patient-21	95.42	75.16	76.47	65.36	33.33	64.05	83.01	60.78	43.14
Patient-22	100	87.18	94.23	87.82	65.38	93.59	91.67	95.51	93.59
Patient-23	67.9	2.69	1.60	13.89	18.52	36.73	15.43	45.37	19.44
Patient-24	83.08	99.7	99.79	99.97	99.95	99.24	99.92	99.47	99.5
Average	89.01	68.49	69.58	61.36	64.43	70.55	65.21	76.14	72.77

Table 6

The specificity rates (%) of the proposed system and 8 competing methods using the best SVM classifier over the original features and features selected by 7 different techniques.

Patients	Proposed	Original	CMIM	FCBF	MIFS	MIM	MRMR	JMI	DISR
Patient-1	99.72	99.92	99.92	99.90	99.87	99.77	99.89	99.90	99.80
Patient-2	80.17	98.21	98.18	98.03	98.10	96.41	97.95	97.09	96.41
Patient-3	92.74	96.98	95.13	97.55	97.31	94.89	97.06	93.71	94.24
Patient-4	52.46	99.72	99.97	99.99	99.82	94.20	99.98	99.30	99.14
Patient-5	99.00	99.99	99.95	99.92	99.87	99.94	99.98	99.95	99.86
Patient-7	99.60	99.89	99.90	99.98	100.00	98.87	100.00	99.47	99.48
Patient-8	94.51	99.54	99.14	98.39	98.62	99.57	98.50	99.36	98.88
Patient-9	75.58	92.07	97.97	96.82	36.37	94.15	47.72	94.70	71.43
Patient-10	96.55	99.91	99.93	99.94	99.93	99.93	99.93	99.91	99.88
Patient-11	95.73	99.73	99.57	99.42	99.32	99.83	99.68	99.65	99.87
Patient-13	96.25	96.40	99.32	98.94	99.40	98.72	99.21	98.32	98.70
Patient-14	99.75	99.96	99.97	99.96	99.95	99.95	99.97	99.96	99.93
Patient-15	98.80	99.82	99.77	99.81	99.94	99.76	99.88	99.79	99.75
Patient-17	98.23	59.53	44.35	56.06	58.32	58.58	52.91	47.12	55.82
Patient-18	93.99	97.27	99.35	96.60	99.59	99.73	99.72	99.72	98.40
Patient-19	97.13	99.50	99.72	99.87	99.37	99.20	99.95	99.31	100.00
Patient-20	97.79	99.48	99.27	99.84	99.74	99.35	99.73	99.31	99.26
Patient-21	98.34	99.93	99.92	99.73	99.79	99.80	99.85	99.97	99.74
Patient-22	93.29	99.60	99.61	99.82	96.84	99.58	99.77	99.74	99.75
Patient-23	99.05	99.70	99.98	99.97	99.79	99.65	99.98	99.60	99.91
Patient-24	99.59	73.99	74.75	53.54	68.94	79.29	67.93	78.54	78.54
Average	94.71	96.63	96.49	95.65	94.88	96.75	95.07	96.53	96.10

24 min of non-seizure data as the negative samples to train our system, which is then used to classify a total duration of around 146 h of EEG data of the 21 patients. To our best knowledge, this is the first patient-specific system ever proposed in the literature, which uses such a low training rate and uses *only* the early recordings. The parameters of MD-PSO are set as follows: we use the termination criteria as the combination of the maximum number of iterations allowed ($iterNo = 100$) and the cut-off error ($\epsilon_C = 10^{-4}$). Other parameters were empirically set as: the swarm size, $S = 60$, $V_{max} = x_{max}/5 = 0.2$ and $VD_{max} = 5$, respectively. We use the typical activation function for MLPs: *hyperbolic tangent* ($\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$). For the AS, we used simple configurations with the following range arrays: $R_{min} = \{N_i, 8, 2\}$ and $R_{max} = \{N_i, 24, 2\}$, which indicate that besides the Single Layer Perceptron (SLP), all MLPs have only a single hidden layer, i.e. $L_{max} = 2$, with no more than 24 hidden neurons. Besides the SLP, the hash function enumerates all MLP configurations in the AS, as shown in Table 4.

6.1. CNBC-E performance plots

We shall start our performance analysis by demonstrating the continuous improvement of the sensitivity with the consecutive ESA and MF operations applied over the classification map of each CNBC within the CNBC-E. The plots in Fig. 11 belongs to the 21 h of EEG data from 4 patients (Patient 1, 3, 5 and 8) and presents both individual and ESA (alone and along with MF operations) sensitivity and specificity values. Since each CNBC is evolved with only 100 MD PSO iterations without a deep convergence, a significant variation can be observed within the individual sensitivity and specificity values indicating a highly divergent seizure frame classification, which in turn leads to significant sensitivity improvements via ESA. Note, for instance in plots for Patients 3 and 5, the individual sensitivity levels by each CNBC is less than 0.9 (90%) whereas the ESA (with or without the MF operation) can surpass 95% or even higher in couple of iterations. This clearly

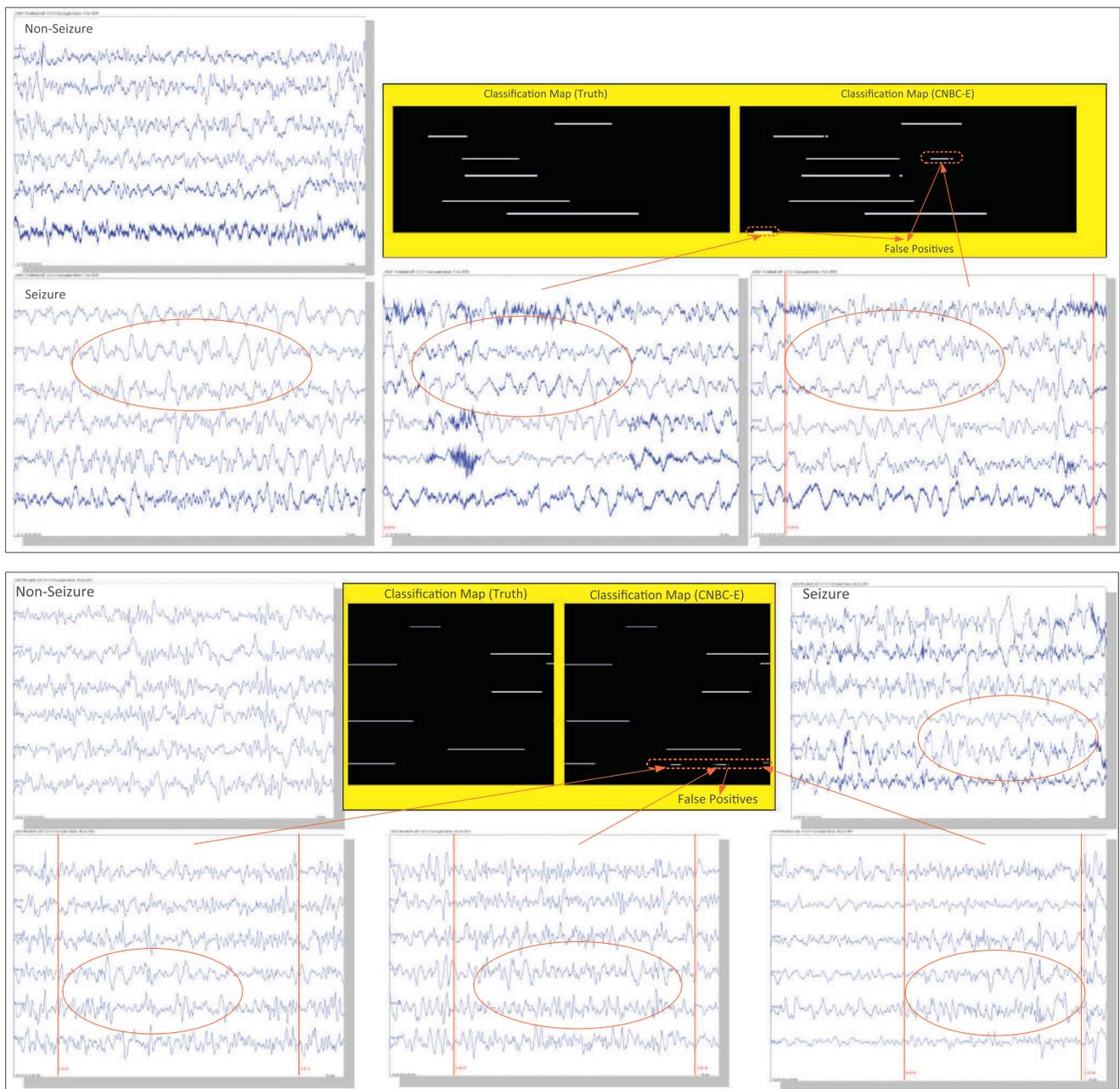


Fig. 12. EEG of seizure, non-seizure and the false-positive segments illustrated in classification map of the CNBC-E for Patients 1 (top) and 10 (bottom).

indicates that most of the seizure frames missed by the early CNBCs can still be detected and hence aggregated by the latter ones thanks to such a high divergence level reached among them. The MF operations, on the other hand, prevent the specificity level to significantly degrade.

6.2. Overall classification performance and comparative evaluations

The EEG classification performance is measured using the three standard metrics: classification accuracy (*Acc*), sensitivity (*Sen*), and specificity (*Spe*). While accuracy measures the overall system performance over both classes of the frames, the other metrics are specific to each individual class, seizure and non-seizure. The respective definitions of these three common metrics using true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) are as follows: *Accuracy* is the ratio of the number of correctly classified frames to the total number of frames classified, i.e., $Acc = (TP + TN)/(TP + TN + FP + FN)$; *Sensitivity* is the rate of correctly classified seizure frames among all seizure frames, $Sen = TP/(TP + FN)$; and *Specificity* is the rate of correctly classified non-seizures among all non-seizures, $Spe = TN/(TN + FP)$. Due to the highly unbalanced numbers of seizure and non-seizure frames, the accuracy is not a relevant performance metric and as discussed earlier, the accurate classification of the seizure frames is of utmost importance and hence makes the sensitivity rate as the primary performance metric. On the other hand, a reasonable level of specificity is acceptable since the Neurologist can easily discard or correct the false positives whenever necessary.

Table 5 presents the sensitivity rates obtained per patient in the database using the proposed system and the 8 competing methods. In the literature, SVMs are commonly used for EEG classification and thus we used them as the competing classifiers. Among many kernel alternatives, we used the *linear* kernel since it gives the *best* results on the test set as compared to other kernel alternatives such as RBF, polynomial or Gaussian. Besides using the original feature set, we also used the state-of-the-art feature selection techniques to remove irrelevant features in order to avoid “Curse of Dimensionality” phenomenon for the classifier. These techniques are Conditional Mutual Information Maximization (CMIM) [31], Fast Correlation Based Filter (FCBF) [32], Mutual Information Feature Selection (MIFS) [33], Mutual Information Maximization (MIM) [34], Max-Relevance Min-Redundancy (MRMR) [35], Joint Mutual information (JMI) [36], and Double Input Symmetrical Relevance (DISR) [37]. In this study we performed CMIM to select 50 most relevant features among 342 features for each individual channel. The number 50 is empirically chosen based on trial and error. Recall that each competing method is also patient-specific where feature selection and classifier training are both performed over the same training dataset using the same features as for the proposed system. When the average sensitivity rates are compared, a significant performance gap between the proposed system and all the competing methods can be observed, i.e., the CNBC based EEG classification system achieved 13% or higher sensitivity rate on the average. Particularly for some patients, the competing methods entirely failed the classification, yielding <20% sensitivity rates. Such a failure may happen for any patient and thus this makes them unreliable for any practical usage. Note that the proposed system alone ensures a minimum of 63.51% sensitivity rate.

The specificity rates are presented in Table 6. All methods achieved >94% specificity rates that are within close vicinity of each other (~2% or less). The proposed system achieved an average specificity rate of 94.72%, which presents a considerable amount of false positives. A closer inspection, however, indicates that most of those false positive segments, although not labeled as the seizure in the benchmark database, show a high structural resemblance to the seizure sections and therefore, may still be interesting and

informative for some abnormal brain activity. For example, EEG waveforms of the typical seizure, non-seizure and the false-positive segments as illustrated in classification map of the CNBC-E for Patients 1 (top) and 10 (bottom) are shown in Fig. 12. It is obvious that the EEG waveforms of the relevant channels highlighted in the figure exhibit major similarities between the false positives and the labeled seizure segments, e.g., for Patient 1, frantic theta band rhythm with increased dynamic range can be observed on the central and parietal part of the brain (P7-O1 and C3-P3) as highlighted in red. Furthermore, such false positives may also be an indicator of an incoming or end of a seizure attack because it is a known fact that between seizures, the EEG of a patient with epilepsy may exhibit abnormal rhythmic activities or frequent discharges.

6.3. Computational complexity analysis

The CNBC-E system architecture has been developed in harmony with a distributed computing scheme, the Techila Grid [30]. Computational problems can be divided into two main categories; parallel problems and embarrassingly parallel problems. The entire frame-based EEG data processing including pre-processing and feature extraction over all channels and with different methods is an example of the latter and can be performed in parallel. Note that the evolution process of the CNBC-E consists of individual evolutions of the CNBCs that can be performed in parallel. Moreover, each CNBC evolution process is also an example of the embarrassingly parallel problems and therefore, this makes it an ideal work that can be distributed within the grid system, which contains a massive number of computers (i.e. the so-called “Workers” within the grid). In the current scheme each NBC evolution is assigned to an individual worker but in theory, each BC evolution can even be parallelized with a proper semaphore implementation so as to secure the order of evolutionary phases, 1, 2 and then 3 (e.g. see Fig. 10). In this case, the computational complexity of the entire CNBC evolution will only be proportional with the evolution of a single BC, regardless of the number of feature extraction methods and channels. The computational complexity analysis of the evolutionary ANNs were presented in [11] and as stated earlier, using a limited AS with a few compact MLP configurations and performing a shallow evolutionary process with only 100 iterations to improve divergence, the computational complexity is therefore, significantly reduced. In our MATLAB implementation, the average feature extraction time over a one-second frame is around 280 ms, which is also insignificant and can further be reduced with a dedicated and optimized implementation. As a result, the CNBC-E evolution is performed in parallel within a reasonably short time that is in the order of evolving a single BC. Note that besides the occasional incremental evolutions, if ever needed at all, the CNBC-E will be evolved only once for a patient, stored and then can conveniently be used any time to classify his/her EEG records, in offline mode or even in real-time since for the classification of 1 s EEG data, the time for data processing and propagation of the frame features through the CNBC-E is only a mere fraction of a second (typically less than 20 ms).

7. Conclusions

In this paper, we proposed a novel patient-specific classifier for long-term EEG signals. The classifier topology of this system, CNBC-E, is not only a seizure onset detector; rather attempts to accurately classify all seizure frames in the seizure sections in a generic way. Another major difference from the majority of the methods proposed in this domain is that it aims to maximize the sensitivity rate using the *minimum* feedback from the Neurologist.

To mimic this, we only used about 2 min of seizure and 24 min non-seizure labeled data per patient on the average to evolve the system and this training dataset is formed from the Neurologist's labels of the early 25% section of the patient's EEG record. To maximize the learning ability from such a limited training dataset and to learn as much patient's EEG pattern characteristics as possible, the CNBC-E uses the majority of the state-of-the-art features proposed in this domain. This is indeed the opposite path followed by many methods in the literature based on a single and static classifier so as to avoid the well-known "Curse of Dimensionality" phenomenon. CNBC-E, on the other hand, can truly take benefit of such a large feature collection due to its "Divide and Conquer" philosophy that is the basis of its topology. Furthermore, rather than relying on sub-optimum static classifiers fixed in advance; each binary classifier within all CNBCs is individually evolved to find out the optimal ANN configuration for the problem (i.e., for the patient, the EEG channel and the feature) in hand. In this way, each CNBC can "learn" and weight accordingly which features and channels are the most relevant and discriminative for the patient and the CNBC-E can aggregate the new seizure frames detected in each CNBC whilst filtering the classification noise as much as possible.

The results of the classification experiments, which are performed over the benchmark CHB-MIT scalp long-term EEG database show that the proposed system can achieve all the aforementioned objectives. Using such a limited training dataset, the proposed system established an average sensitivity rate of 89.01% along with an average specificity rate of 94.71% over the test set. The comparative evaluations against the competing methods indicate the superiority of the proposed system which achieves 13% higher average sensitivity rate over the best competitor. Another observation worth mentioning is that for more than half of the patients, the proposed system achieved the sensitivity rates above 90% and for only 3 patients (Patients 4, 17 and 23) out of 21 it is below 80%. A detailed investigation reveals that for these patients the seizure has a highly non-stationary nature. Therefore, the training set formed from the labels of the early 25% EEG recording do not have the entire seizure and non-seizure characteristics that can be learned. It is obvious that for those patients, further feedback from the Neurologist is needed for the misclassified seizure sections so as to incrementally evolve the CNBC-E system and thus to improve the overall classification performance. This will be the topic for our future research.

References

- [1] Fisher RS, Boas WVE, Blume W, Elger C, Genton P, Lee P, et al. Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE). *Epilepsia* 2005;46:470–2.
- [2] Kuzneicky RI. MRI in cerebral developmental malformations and epilepsy. *Magn Reson Imaging* 1995;13:1137–45.
- [3] Gotman J. Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr Clin Neurophysiol* 1982;530–40.
- [4] Shoeb A. Application of machine learning to epileptic seizure onset detection and treatment. Ph.D. thesis. Massachusetts Institute of Technology; 2009.
- [5] Hao Qu, Gotman Jean. A patient-specific algorithm for detection of seizure onset in long-term EEG monitoring: possible use as a warning device. *IEEE Trans Biomed Eng* 1997;44(2):115–22.
- [6] CHB-MIT Scalp EEG Database. 2010. <<http://www.physionet.org/pn6/chbmit/>>.
- [7] Kiranyaz S, Ince T, Yildirim A, Gabbouj M. Fractional particle swarm optimization in multi-dimensional search space. *IEEE Trans Syst Man Cybern – Part B* 2010;40(2):298–319.
- [8] Ince T, Kiranyaz S, Gabbouj M. A generic and robust system for automated patient-specific classification of electrocardiogram signals. *IEEE Trans Biomed Eng* 2009;56(5):1415–26.
- [9] Kiranyaz S, Ince T, Yildirim A, Gabbouj M. Evolutionary artificial neural networks by multi-dimensional particle swarm optimization. *Neural Networks* 2009;22:1448–62.
- [10] Shoeb A, Gutttag J. Application of machine learning to epileptic seizure onset detection. In: 27th international conference on machine learning (ICML), Haifa, 2010.
- [11] Ravish DK, Devi S. Automated seizure detection and spectral analysis of EEG seizure time series. *Eur J Sci Res* 2012;68:72–82.
- [12] Khan YU, Foroog O, Sharma P. Automatic detection of seizure onset in pediatric EEG. *Int J Embedded Syst Appl (IJESA)* 2012;2:81–9.
- [13] Shoeb A, Kharbouch A, Soegaard J, Schachter S, Gutttag J. An algorithm for detecting seizure termination in scalp EEG. In: 33rd annual international conference of the IEEE EMBS, Boston, 2011. p. 1443–46.
- [14] Gotman J. Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr Clin Neurophysiol* 1982;54:530–40.
- [15] McClellan JH, Parks TW. A personal history of the Parks–McClellan algorithm. *Signal Process Mag IEEE* 2005;22(2):82,86. doi: 10.1109/MSP.2005.1406492.
- [16] Shoeb A, Edwards H, Connolly J, Bourgeois B, Treves ST, Gutttag J. Patient-specific seizure onset detection. *Epilepsy Behav* 2004;5:483–98.
- [17] Kalatzis I, Piliouras N, Ventouras E, Papageorgiou CC, Rabavilas AD, Cavouras D. Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals. *Comput Methods Programs Biomed* 2004;75(1):11–22.
- [18] Temko A, Thomas E, Boylan G, Marnane W, Lightbody G. An SVM-based system and its performance for detection of seizures in neonates. In: 31st annual international conference of the IEEE EMBS, Minneapolis, Minnesota, USA, 2009. p. 2643–46.
- [19] Ferenets R, Lipping T, Anier A, Jantti V, Melto S, Hovilehto S. Comparison of entropy and complexity measures for the assessment of depth of sedation. *IEEE Trans Biomed Eng* 2006;53:1067–77.
- [20] Seng CH, Demirli R, Khuon L, Bolger D. Seizure detection in EEG signals using support vector machines. In: 38th annual northeast bioengineering conference (NEBEC), Philadelphia, PA, 2012. p. 231–32.
- [21] Kuhlmann L, Cook MJ, Fuller K, Grayden DB, Burkitt AN, Mareels IMY. Correlation analysis of seizure detection features. *Intelligent sensors, Sensor networks and information processing (ISSNIP 2008)*, Sydney, NSW, 2008. p. 309–14.
- [22] Beigi H. Mel frequency cepstral coefficients (MFCC). *Fundamentals of speaker recognition*. NY, USA: Springer; 2009. p. 173–176.
- [23] Pan Y, Guan C, Ang KK, Phua KS, Yang H, Huang D, Lim S. Seizure detection based on spatiotemporal correlation and frequency regularity of scalp EEG. In: World Congress on Computational Intelligence (WCCI 2012), Brisbane, Australia, 2012. p. 1–7.
- [24] Husain SJ, Rao KS. Epileptic seizure classification from EEG signals using neural networks. In: International conference on information and network technology (ICINT 2012), Singapore, 2012. p. 269.
- [25] Swiderski B, Osowski S, Rysz A. Lyapunov exponent of EEG signal for epileptic seizure characterization. In: European conference on circuit theory and design, Cork, Ireland, 2005. p. 153–56.
- [26] Kennedy J, Eberhart R. Particle swarm optimization. In: Proc. of IEEE Int. conference on neural networks. vol. 4, Perth, Australia, 1995. p. 1942–48.
- [27] Techila Grid. <<http://www.techila.fi/technology/>>.
- [28] Fleuret F. Fast binary feature selection with conditional mutual information. *J Mach Learn Res* 2004;5:1531–55.
- [29] Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy. *J Mach Learn Res* 2004;5:1205–24.
- [30] Battiti R. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans Neural Networks* 1994;5:537–50.
- [31] Lewis DD. Feature selection and feature extraction for text categorization. In: Proceedings of the workshop on speech and natural language. Association for Computational Linguistics Morristown, NJ, USA, 1992. p. 212–17.
- [32] Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of maxdependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1226–38.
- [33] Yang H, Moody J. Data visualization and feature selection: new algorithms for non-gaussian data. *Adv Neural Inform Process Syst* 1999;12.
- [34] Meyer P, Bontempi G. On the use of variable complementarity for feature selection in cancer classification. *Evol Comput Mach Learn Bioinform* 2006;91–102.