EEG Synchronization Analysis for Seizure Prediction: A Study on Data of Noninvasive Recordings

Paolo Detti [1],*,† , **Giampaolo Vatti [2],†** and **Garazi Zabalo Manrique de Lara [1],†**

1   Department of Information Engineering and Mathematics, University of Siena, 53100 Siena, Italy; garazizml@gmail.com
2   Department of Medicine, Surgery and Neuroscience, University of Siena, 53100 Siena, Italy; giampaolo.vatti@gmail.com
*   Correspondence: detti@dii.unisi.it
†   These authors contributed equally to this work.

Abstract: Objective: Epilepsy is a neurological disorder arising from anomalies of the electrical activity in the brain, affecting ~65 million individuals worldwide. Prediction methods, typically based on machine learning methods, require a large amount of data for training, in order to correctly classify seizures with small false alarm rates. Methods: In this work, we present a new database containing EEG scalp signals of 14 epileptic patients acquired at the Unit of Neurology and Neurophysiology of the University of Siena, Italy. Furthermore, a patient-specific seizure prediction method, based on the detection of synchronization patterns in the EEG, is proposed and tested on the data of the database. The use of noninvasive EEG data aims to explore the possibility of developing a noninvasive monitoring/control device for the prediction of seizures. The prediction method employs synchronization measures computed over all channel pairs and a computationally inexpensive threshold-based classification approach. Results and conclusions: The experimental analysis, performed by inspection and by the proposed threshold-based classifier on all the patients of the database, shows that the features extracted by the synchronization measures are able to detect preictal and ictal states and allow the prediction of the seizures few minutes before the seizure onsets.

Keywords: EEG data; epilepsy; synchronization measures; threshold-based classifier; data classification

1. Introduction

Epilepsy is a disease characterized by an enduring predisposition to generate epileptic seizures, due to abnormal excessive or synchronous neuronal activity in the brain [1]. Very importantly, the main feature of epilepsy, epileptic seizures, is associated with a number of negative consequences at both short- and long-term, including the risk of falls and injuries; eventual death; psychiatric disturbances; cognitive deficits; and difficulties in achieving academic, social, and employment goals. Pharmacological and surgical treatments are available that can stop or control seizures for most people with seizures and epilepsy. However, antiepileptic drugs have limitations and side effects [2], and often fail to control seizures in ~30% of the cases, while surgery cannot always applied. Therefore, seizure prediction methods could be an important option. In fact, an early prediction might allow patients or caregivers to take suitable actions, such as warning an alarm, applying short-acting drugs, and activating stimulating devices.

Seizure prediction is a challenging research topic that has been explored for over 25 years. Many approaches are based on the analysis of electroencephalogram (EEG) signals (e.g., see, in [3,4]), and on the development of machine learning methods [5–8], usually requiring a large amount of data for training. However, physiological data (e.g., EEG and EKG) including seizures are in general scarce,

often including artifacts and noise. Furthermore, preictal and ictal morphologies may vary both within the same patient (intra-patient specificity) and across different patients (inter-patient specificity). Finally, the length of the preictal states are not univocally defined in the literature [9], implying that a clear partition between inter-ictal and preictal states is difficult to perform.

Several works in the literature investigated the relationship between ictal and preictal states and synchronization patterns in EEG through the use of bivariate measures (i.e., involving two or more EEG channels). Such studies showed that synchronization phenomena in EEG may help to highlight the dynamics of epilepsy [6,10–14]. Mormann et al. [12] measured synchronization in EEG channels by the mean phase coherence, finding a decrease of the measure before seizure onset [13]. Mirowski et al. [6] combined nonlinear classifiers with phase-synchrony measures. Kuhlmann [11] applied the mean phase coherence as a measure of bivariate phase synchrony, too. In this case, the seizure prediction performance was evaluated for all possible channel pairs and for different prediction methods, in order to find the best performing channel pair and method for both preictal decreases and increases in synchrony. Myers [14] employed measures to determine synchronization both in phase and in amplitude between pairs of EEG signals.

In general, bivariate measures can be naturally embedded on a graph model, i.e., a network, in which nodes correspond to sources of the EEG signals and arcs between two nodes are weighted by value corresponding to the synchronization degree between couple of EEG signals. Note that the arc weights vary over time according to the changes in synchronization. Such a network model is also justified by new findings of recent research in epilepsy, showing that the epileptiform activity in one part may influence brain activity in other parts [15–17].

In this paper, we present a classification algorithm for seizure prediction that employs phase-synchronization measures computed for all pairs of EEG signals, namely, Phase Lag Index (*PLI*) and Weighted Phase Lag Index (*WPLI*) [18,19]. The phase-synchronization measures can be also embedded on a network model that allows studying how epileptic events propagate throughout the brain. An earlier version of the algorithm has been proposed and tested in [20], while in this paper we present a new modified version in which new feature selection rules and postprocessing procedures are employed, with the aim of reducing the number of false alarms. Furthermore, we also introduce a new freely accessible database, containing scalp EEG data of different epileptic patients, and present an experimental campaign, in which the seizure prediction method and the phase synchronization measures are tested on all the data of the database. The data has been collected by the Unit of Neurology and Neurophysiology at the University of Siena, Italy, during a regional research project, called PANACEE [21], aiming at the development of noninvasive patient-specific monitoring/control low-cost devices for the prediction of epileptic seizures.

In fact, the prediction method has been designed to be used on noninvasive (scalp) EEG data, generally noisier than those obtained with other invasive (intracranial) techniques. However, the choice of noninvasive techniques for data acquisition would allow a real-time use of the method also in portable and wearable devices with limited discomfort for the patients.

The article is organized as follows. Section 2 is organized into two sections: In Section 2.1 the database is presented and described. The synchronization measures and the related network model developed to study the synchronization patterns in EEG are introduced in Section 2.2. The seizure prediction approach employing the synchronization measures presented in Section 2.2 is described in Section 2.5. Section 3 provides experimental results on all the 14 epileptic patients acquired at the Unit of Neurology and Neurophysiology at University of Siena. In particular, the trends of the synchronization measures are shown (Section 3.1) and the performances of our seizure prediction method are presented (Section 3.2). In Section 4, the experimental results are analyzed and discussed. Finally, conclusions follow in Section 5.

2. Methods

2.1. A Scalp EEG Database

The Unit of Neurology and Neurophysiology at the University of Siena, Italy, acquired the data of 14 epileptic patients, collected during the national interdisciplinary research project PANACEE [21], aiming to describe, detect, and predict in a sufficiently short time epileptic seizures and wake–sleep transitions in order to take appropriate actions to ensure the safety of the subjects. The data contain recordings of cerebral (EEG) activity of pathological subjects acquired with a Video scalp EEG with a sampling rate of 512 Hz, with electrodes arranged according to the international 10–20 system. Most of the recordings contain EKG signals, too. The data were acquired employing EB Neuro and Natus Quantum LTM amplifiers, and reusable silver/gold cup electrodes. Patients were asked to stay in the bed as much as possible, either asleep or awake. Clinical and electrophysiological data of each patient have been carefully revised by expert clinicians, that performed the diagnosis of epilepsy and the classification of the seizures according to the criteria of the International League Against Epilepsy [22]. The Ethical Committee of the University of Siena approved the data in accordance with the Declaration of Helsinki. At the time of admission at the clinics, each patient signed a written informed consent in which agrees to the video registration and to the use of the data for a possible scientific divulgation.

The anonymous data of the 14 patients, 9 males (ages 36–71) and 5 females (ages 20–58), are freely available. The data are freely available at [23]. The database, that we call the *Siena scalp database*, consists of 14 folders containing EEG recordings in European Data Format (EDF) format from epileptic patients. Each folder refers to a specific subject including between 1 and 5 data files with a maximum size of 2.11 GB each, and a text file containing data sampling rate (in Hz), a list of electrodes positioned on the scalp from which the EEG signals are extracted, start and end time of the registration, and start and end time of each seizure.

Table 1 reports detailed overviews of the data. More precisely, Columns 2–6 report the clinical information of the patients, including seizure classification according to the criteria of the International League Against Epilepsy [22]. Columns 7–9 of the table show, for each patient, the number of EEG channels (# EEG Chan.) and seizures (# Seiz.), and the total registration time in minutes (Time), respectively.

Table 1. Data of the patients in the Siena Scalp Database.

Pat. Id	Age (Years)	Gender	Seizure	Localization	Lateralization	# EEG Chan.	# Seiz.	Time
PN00	55	Male	IAS	T	R	29	5	198
PN01	46	Male	IAS	T	L	29	2	809
PN03	54	Male	IAS	T	R	29	2	752
PN05	51	Female	IAS	T	L	29	3	359
PN06	36	Male	IAS	T	L	29	5	722
PN07	20	Female	IAS	T	L	29	1	523
PN09	27	Female	IAS	T	L	29	3	410
PN10	25	Male	FBTC	F	Bilateral	21	10	1002
PN11	58	Female	IAS	T	R	29	1	145
PN12	71	Male	IAS	T	L	29	4	246
PN13	34	Female	IAS	T	L	29	3	519
PN14	49	Male	WIAS	T	L	29	4	1408
PN16	41	Female	IAS	T	L	29	2	303
PN17	42	Male	IAS	T	R	29	2	308

IAS: focal onset impaired awareness; WIAS: focal onset without impaired awareness; FBTC: focal to bilateral tonic–clonic; T: temporal; R: right; L: left.

2.2. Synchronization Measures and Graph Model

Synchronization patterns in EEG have been studied in the literature [6,10–14,24] to highlight connections between preictal/ictal states and synchronization. Many works employed phase synchronization measures, such as mean phase coherence [12], and the recently introduced Phase Lag Index (*PLI*) and Weighted Phase Lag Index (*WPLI*) [18,19]. More precisely, *PLI* has been introduced

in [18] for the assessment of functional connectivity in the brain, and *WPLI* has been introduced in [19] as an extension of *PLI*. All the above measures can be applied to EEG signal pairs and return values in $[0, 1]$, where 0 indicates that no phase locking occurs and 1 is a perfect phase locking.

In this work, we employ the recently introduced PLI and WPLI to compute the synchronization levels of pairs of EEG signals. Given two time series $x_h(t)$ and $x_k(t)$ (e.g., series related to two EEG signals) and a time window Δ_t containing N time instants, PLI and WPLI can be defined as follows,

$$PLI_{h,k,\Delta_t} = \left| \frac{1}{N} \sum_{p=1}^{N} \text{sign}(\phi_h(p) - \phi_k(p)) \right|, \tag{1}$$

$$WPLI_{h,k,\Delta_t} = \frac{\left| \frac{1}{N} \sum_{p=1}^{N} \sin\left(\phi_h(p) - \phi_k(p)\right) \right|}{\frac{1}{N} \sum_{p=1}^{N} \left| \sin\left(\phi_h(p) - \phi_k(p)\right) \right|}, \tag{2}$$

where $\phi_h(p)$ and $\phi_k(p)$ are the phases at time instant p of signals $x_h(t)$ and $x_k(t)$, respectively, determined by the Hilbert transformation (see in [20]).

PLI is based on the idea of discarding the phase differences that center around 0 (mod π), and quantify the asymmetry of the distribution of phase differences around zero. In *WPLI*, each phase difference is weighted according to the magnitude of the lag. In this way, phase differences around zero only marginally contribute to the calculation of the wPLI. This procedure reduces the probability of detecting "false positive" connectivity in the case of volume conducted noise sources with near zero phase lag and increases the sensitivity in detecting phase synchronization.

2.3. Embedding Synchronization Measures into a Network Model

The synchronization measures PLI and WPLI are bivariate measures that can be computed for each pair of EEG signals extracted from the electrodes positioned on the scalp. In this context, a natural network model of the brain connectivity can be defined [19,25], where the nodes of the network represent the physical sites from which the EEG signals are extracted and the arcs between the nodes and their weights model the brain connectivity. More formally, the network can be defined by an undirected complete weighted graph $G = (V, A)$, where V is set of nodes and A is the set of arcs. One or more weights $w_{h,k}$ can be associated to each arc $(h, k) \in A$, that, in our case, can be set to the values provided by the synchronization measures (e.g., PLI, WPLI, or both). Note that, the synchronization measures and, therefore, the edge weights $w_{h,k}$ of the graph vary over time. Consequently, observing how the graph structure is modified over time (i.e., the weights of the arcs), we have an overall picture on how epileptiform events evolve throughout the brain. Two easily computable graph measures of brain connectivity are the *degree* of a node, i.e., the number of edges (with weights larger than a given threshold) incident to a node in G, and the *strength* of a node, i.e., the sum of the weights of the edges incident to a node in G. In the next sections, this last measure will be used to detect variations in synchronization with the aim of predicting seizure occurrences.

2.4. Employing Synchronization Measures PLI and WPLI for Seizure Prediction

We developed a seizure prediction approach, in which the synchronization indices *PLI* and *WPLI*, computed as in (1) and (2), respectively, and the graph embedding them are used to distinguish preictal and inter-ictal states.

However, in our approach, *PLI* and *WPLI* are computed on signals obtained by raw EEG after some preprocessing steps reported in Figure 1 and explained in the following.
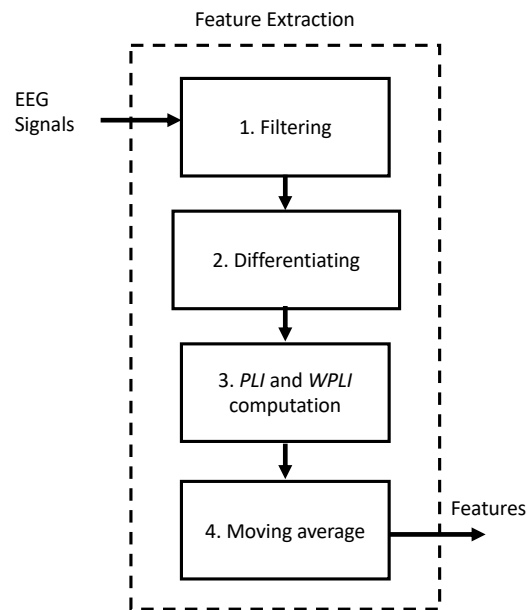
Figure 1. Block diagram of the feature extraction procedure.

EEG signals are first filtered in the $[8, 13]$ Hz band (such a band has been identified by preliminary experimental campaign in which different bands have been considered): The filtering process allows the selection of the band of frequencies of interest, possibly removing undesired artifacts. Then, signals are differentiated (i.e., the absolute value of the time-derivative of the signal is computed [26]): the differentiation of the signal makes the basic noise nearly flat and sharpens the regions where the signal exhibits its peaks (see [26,27]), which, as shown below, are most likely to be the regions where seizures occur. As an example, Figure 2 shows the behaviors of *PLI* and *WPLI* without and with differentiation on a time span of 1000 s preceding the first seizure of patient PN01 of the database, computed on the channel pair $(T5, Pz)$, and filtered in the $[8, 13]$ Hz band. In the figure, the seizure starts (ends) at the first (second) vertical dotted line.

Note that no significant changes occur in *PLI* and *WPLI* without differentiation, whereas, with differentiation, *PLI* and *WPLI* increase as the seizure is approaching and peak just at the start of the seizure (denoted by the first vertical dotted line).

Figure 3 reports the behavior over time of the strength of the node/channel $F10$, computed by *PLI*, on a time period of ~1500 s containing the first seizure of patient $PN00$, starting at the second 1143. The two vertical dotted lines delimit the ictal period. Note that the strength of the node increases as the seizure's onset is approaching and then decreases at the end of the ictal period.

On the other hand, such evident trends for *PLI* and *WPLI* on the filtered and differentiated signals are not easy to detect in the original EEG data, as shown in Figure 4. In the upper parts of Figure 4a,b, EEG data segments of given pairs of EEG channels of patients PN00 and PN05, respectively, are reported. The segments refer to a time span of 300 s preceding the first seizures of patients PN00 and PN05, starting and ending at the first and at the second vertical dotted line, respectively. The lower parts of Figure 4a,b shows *PLI* and *WPLI* on the same time span computed on the channels reported in the upper parts of the figures. In all the figures, *PLI* and *WPLI* are computed using a time window Δ_t of 6 s (with an overlap of 1 s). Note that *PLI* and *WPLI* increase even before the seizure onset.
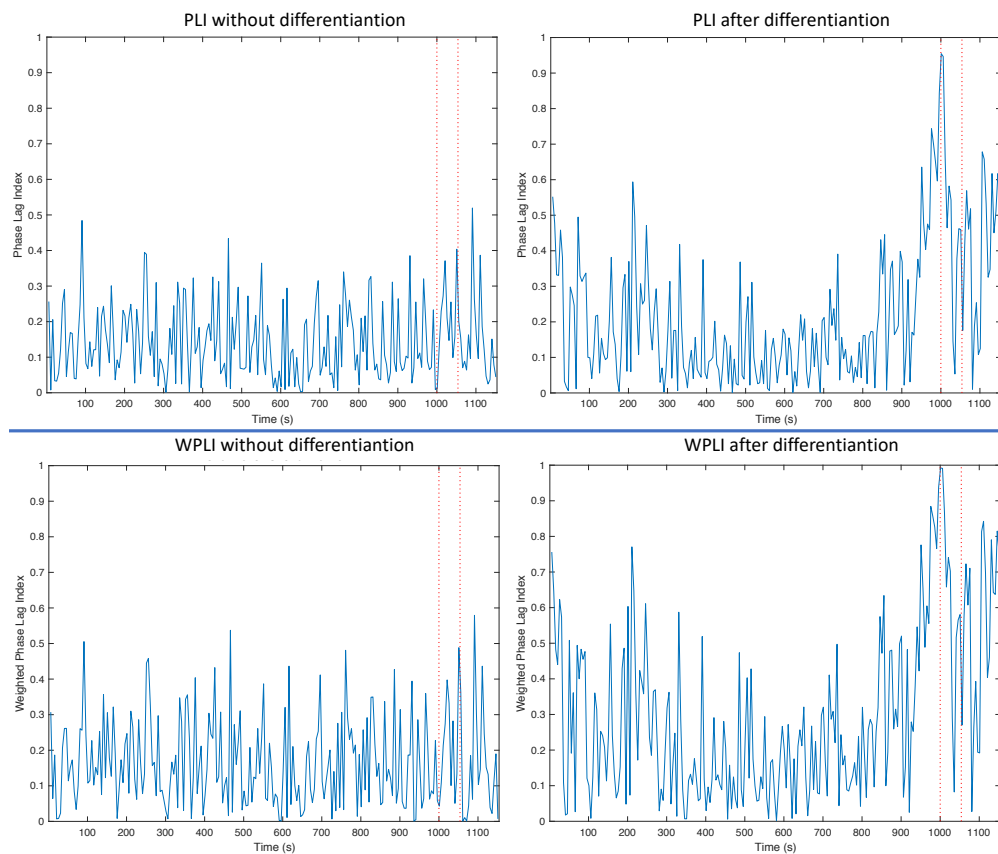
Figure 2. $PLI_{h,k}$ and $WPLI_{h,k}$ with $h = \{T5\}$ and $k = \{Pz\}$ on the 1st epileptic seizure of patient $PN01$ before and after differentiation.

Furthermore, in our seizure prediction approach, a moving average procedure (described in [20] in detail) is applied to all the measures introduced above. The moving average procedure aims to *smooth* (e.g., by removing isolated peaks) the *PLI* and *WPLI* measures. Figure 5 shows the strength of Figure 3 after the application of the moving average procedure.

In Figure 1, a block diagram summarizing all the operations performed on the EEG signals are reported. At the end of the process, we get time-series, hereafter called *features*, that will be given as an input to our for seizure prediction method.
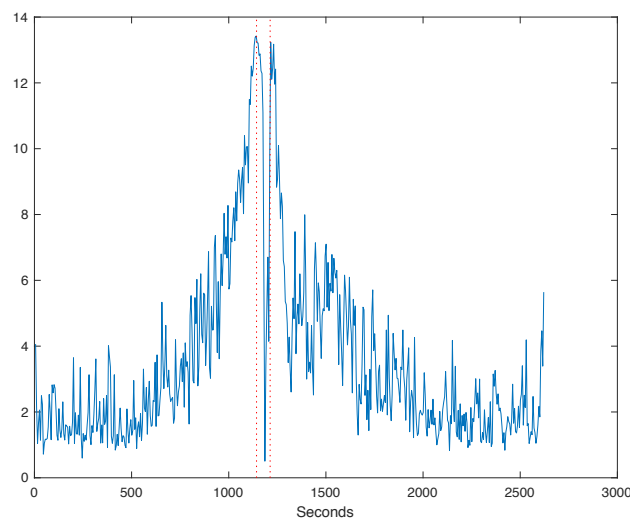

Figure 3. Strength for channel $F10$ on the first seizure of patient PN00.
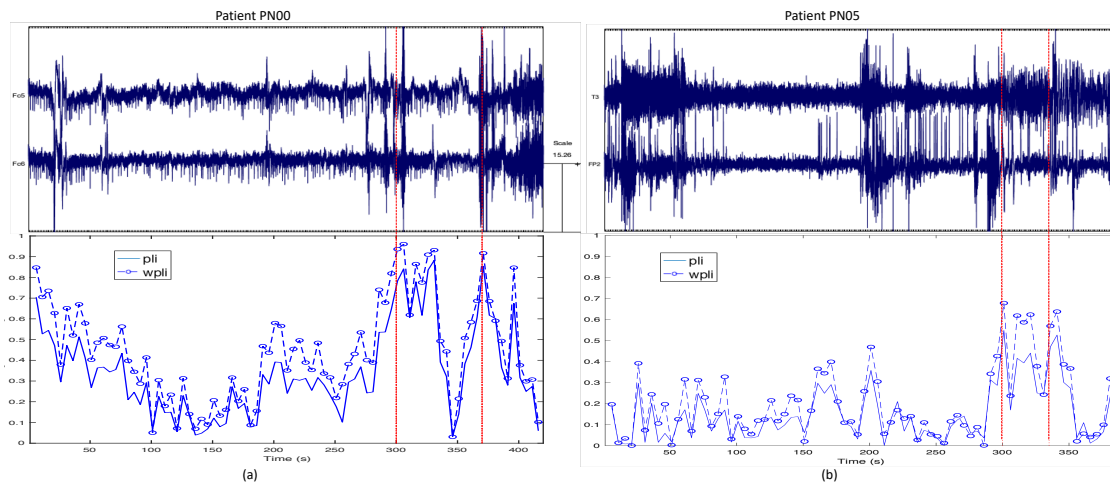
Figure 4. (**a**) Electroencephalogram (EEG) on channels $h = \{Fc5\}$ and $k = \{Fc6\}$; $PLI_{h,k}$ and $WPLI_{h,k}$ on the 1st seizure of patient $PN00$. (**b**) EEG on channels $h = \{T3\}$ and $k = \{FP2\}$; $PLI_{h,k}$ and $WPLI_{h,k}$ on the last seizure of patient $PN05$.
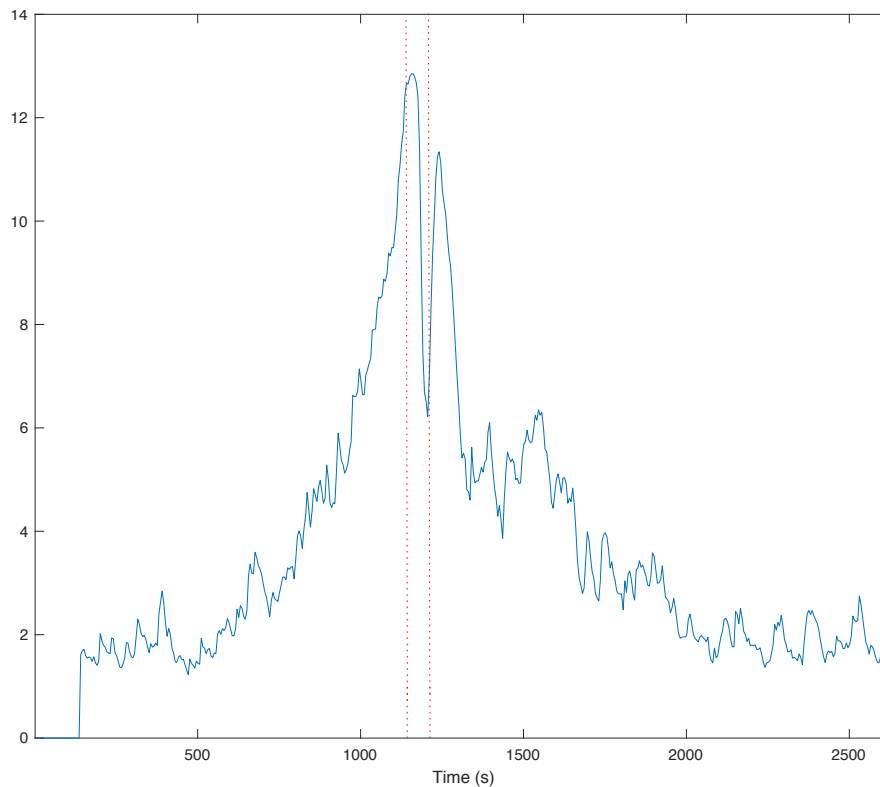

Figure 5. The strength reported in Figure 3 after the application of the moving average procedure.

2.5. A Classification Approach for Epileptic Seizures Prediction

The problem of predicting epileptic seizures can be formulated as a classification problem, which can be solved by a classifier (i.e., an algorithm) operating in two phases, called *Training* and *Test phases*. In the Training phase, the classifier is learned to recognize positive and negative records on a given set of data (*Training data*). In the Test phase, the classifier is tested on a different set of data (*Test data*) in order to evaluate its ability of correctly classifying (as positive or negative) the unlabeled records of the Test data.

The classifier we propose is a modified version of the threshold-based algorithm, called *ThAlgo*, introduced in [20]. In the modified version, denoted as *ModThAlgo* in the following, new feature

selection rules and a postprocessing procedure are considered, with the aim of reducing the number of false alarms

In Training and Test phases, *ModThAlgo* works on the features obtained by applying all the operations of Figure 1. Namely, as already stated in Section 2.2, the EEG signals are first filtered (by means of a pass-band FIR filter) on prespecified bands (Step 1) and differentiated (Step 2); in Step 3, the resulting time series are then used to extract the synchronization measures *PLI* and *WPLI* for each pair of signals (i.e., EEG channels) and each time-window through Formulas (1) and (2). Finally, a moving average procedure is applied to each measure (Step 4).

Classification and Postprocessing

As the length of the preictal states are not univocally defined in the literature, *ModThAlgo* employs the concept of *prediction interval*, i.e., a time window immediately preceding a given seizure. The length of the time window is a parameter given as an input to the algorithm and it is experimentally detected. All records inside a prediction interval belong to the positive class, and all the records outside the prediction interval belong to the negative class. During the Training phase, *ModThAlgo learns* from the positive and negative records of the Training data and employs such information to compute a set of thresholds to separate as accurate as possible the positive and negative classes. During the Test phase, *ModThAlgo* is applied on unlabeled Test data and employs the thresholds computed in the Training phase to discriminate positive and negative classes. A seizure is correctly predicted if at least a record inside the prediction interval is classified as positive. In this case, the *prediction time* is the time span between the first record correctly classified inside the prediction interval of a seizure and its onset. On the other hand, a *false positive* is generated when the algorithm classifies as positive a record outside the prediction interval. Note that, actually, at least a positive record of a prediction interval must be correctly classified (i.e., separated) to predict the corresponding seizure.

During the Training phase, a feature selection procedure is carried out in order to identify subsets of features that are *relevant* and *reliable* for correctly classifying the Training data. Such selected features are then used in the Test phase. In *ModThAlgo*, the feature selection phase has two steps: In the first step, a set of thresholds $\theta_{h,k}$ is computed, one for each feature derived by the synchronization measures $PLI_{h,k}$ and $WPLI_{h,k}$, by taking the minimum of the maximum values of the features over all the prediction intervals (or a suitable subset of them) of the training data. In the second step, the features are ranked according to given rules. More precisely, while in *ThAlgo* features are selected by simple ranking rules based on the number of times each feature is above (below) its threshold inside (outside) the prediction intervals, in *ModThAlgo* a function is used to rank the features, depending on the difference between their thresholds and the maximum of the values of the features outside the prediction intervals. In fact, the bigger this difference is, the easier the computation of a line separating negative and positive records is.

In the subsequent Test phase, the features are ordered according to the ranking obtained in the Training phase. Then, the first features in the ranking and their corresponding thresholds are linearly combined to obtain a single feature and a corresponding single threshold Θ. (The coefficients of the linear combination are parameters of the algorithm.) The classification in positive and negative records of the Testing data is performed by evaluating whenever the combined feature (computed on the Testing data) is above (below) the threshold Θ inside (outside) the prediction intervals.

At the end of the Testing phase, a postprocessing procedure is carried out on the classified records obtained so far. The procedure tries to remove *isolated* positive records, with the aim of reducing false positives (i.e., negative records classified as positive in the Testing phase). More precisely, the postprocessing procedure iteratively scans k consecutive classified records: If at least one of the k records is negative, then all the k records are set to negative. The procedure is applied several times, each time with a different value of the parameter k. (In the experimental campaign presented in Section 3, we execute the postprocessing procedure with different k values, i.e., $k = 0, 1, \ldots, 6$.)

3. Results

In this section, we apply the synchronization measures presented in the Section 2.2 on the data of the Siena Scalp Database introduced in Section 2.1. The aim is to evaluate whether changes in synchronization may help in predicting seizure occurrences through a patient-specific approach.

Note that classification algorithms require a large amount of data to be reliable and effective. Therefore, a sufficient number of seizures is required for each patient to properly training a patient-specific algorithm. As reported in Table 1, six patient datasets of the Siena Scalp database contain either one or two seizures, not enough to apply a training and a test phase of a patient-specific classification procedure. On these six datasets, we perform an analysis by inspection, by detecting the channel pairs on which the features extracted by the synchronization measures show useful trends for seizure prediction. On the data of the eight remaining patients (with at least three seizure occurrences), we apply the classification algorithm described in the previous section, and evaluate it through a cross-validation scheme.

In Section 3.1, the trends of the features (i.e., the synchronization measures) obtained as described in Figure 1 on selected channels for the six patients of the Siena Scalp database with one or two seizures are shown. In Section 3.2, results of our classification algorithm *ModThAlgo* are presented on the remaining eight patients.

3.1. Trends of the Features for Patient Datasets with One or Two Seizures

In this section, we show the trends of the features obtained as described in Figure 1 on selected channels (the selection has been performed by inspection) for the six patients of the Siena Scalp database with one or two seizures. According to Table 1, datasets PN01, PN03, PN16, and PN17 contain two seizures and PN07 and PN11 contain one seizure. On these datasets, Figures 6–11 show the trends of the most meaningful features (extracted either from PLI or WPLI) in a time span preceding the onset of each seizure of about one hour. In each figure, the first and the second dotted line correspond to the start and the end of the ictal period. As already stated, the meaningful channel pairs have been selected by inspection. We Note that, in all the cases, the selected features (that, by Definitions (1) and (2), cannot be bigger than 1 on a channel pair) sharply increase as the seizure is approaching, with a peak occurring approximately at the onset. The rising trends start before the seizure onsets, implying that the proposed features could be used for seizure prediction.
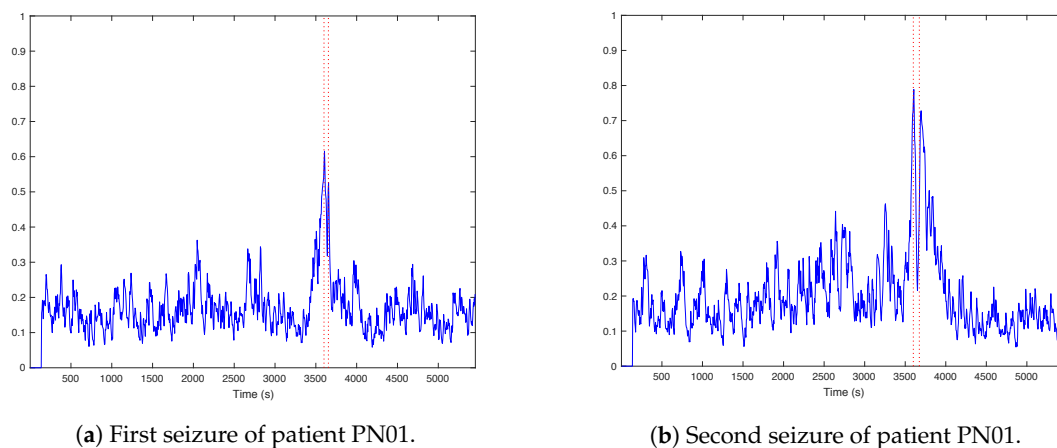

(**a**) First seizure of patient PN01.

(**b**) Second seizure of patient PN01.

Figure 6. Features computed on PLI for channel pair T5–C4 of patient PN01.

(**a**) First seizure of patient PN03.

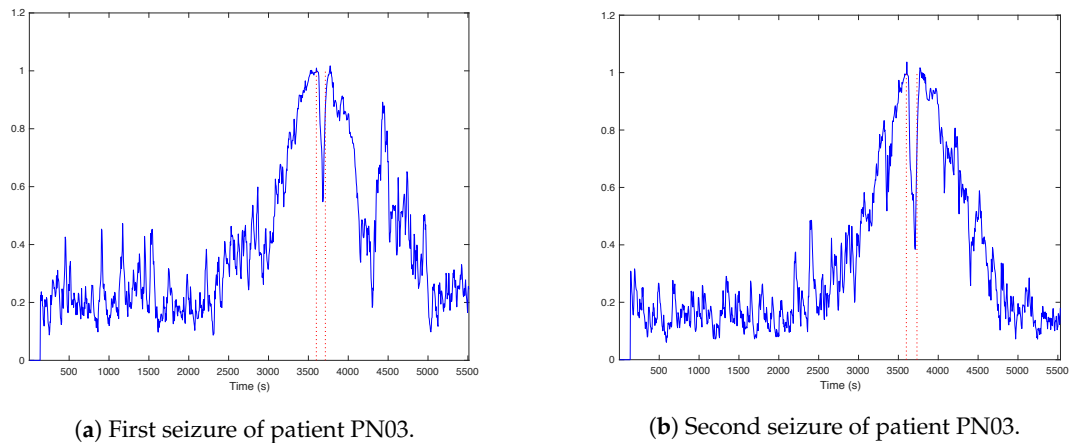(**b**) Second seizure of patient PN03.

Figure 7. Features computed on Phase Lag Index (PLI) for channel pair F4–F8 of patient PN03.
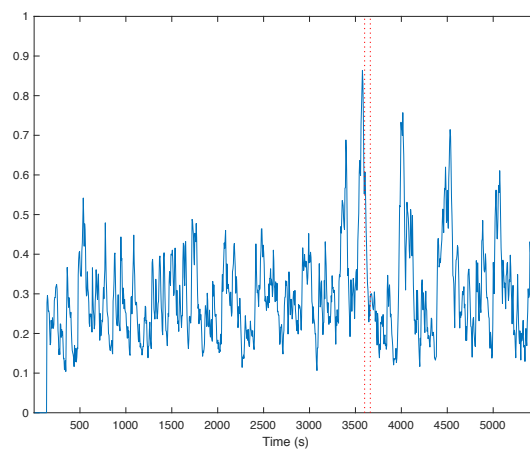

Figure 8. Features computed on Weighted Phase Lag Index (WPLI) for channel pair F9–Fc2 of patient PN07.
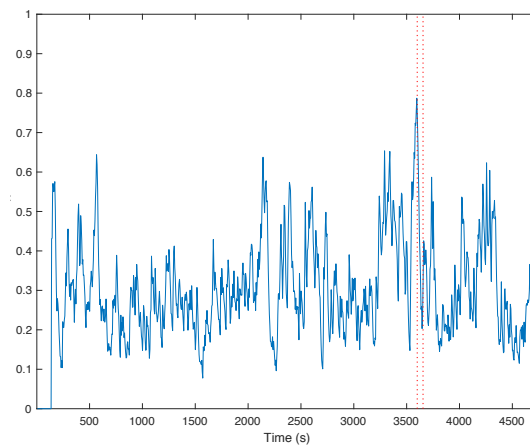

Figure 9. Features computed on WPLI for channel pair Fc1–FP2 of patient PN11.
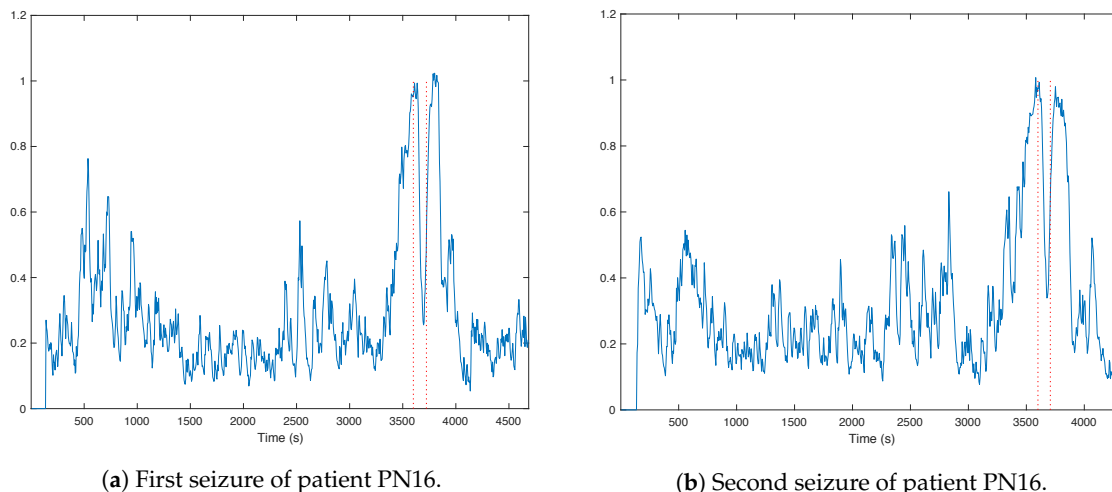
(**a**) First seizure of patient PN16.

(**b**) Second seizure of patient PN16.

Figure 10. Features computed on PLI for channel pair F9–Cp2 of patient PN16.

(**a**) First seizure of patient PN17.
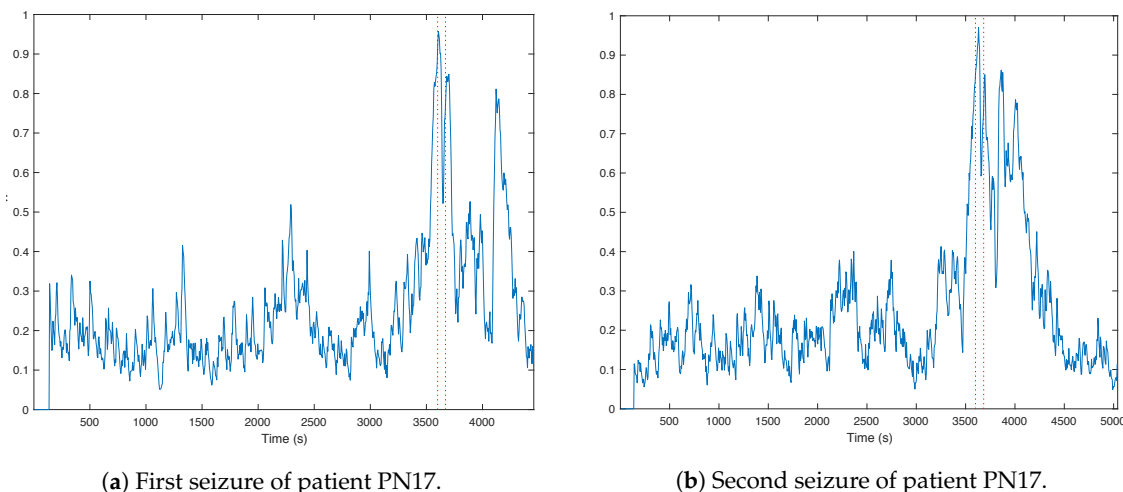

(**b**) Second seizure of patient PN17.

Figure 11. Features computed on WPLI for channel pair Fz–F10 of patient PN17.

3.2. Results of the Classification Algorithm

In this section, experimental results of the classification algorithm *ModThAlgo* introduced in Section 2.5 on the patients of the Siena Scalp database with at least three seizures (see Table 1) are presented.

To prepare the data for our classification algorithm, all data have been subsampled at 256 Hz, and the dataset of each patient has been segmented into a number of subsets, called *instances*, generally lasting about 1 h each, in such a way that each instance may contain at most one seizure occurrence. In what follows, we call *seizure instances* the instances containing one seizure, and *non-seizure instances* the instances that do not contain seizures. In the seizure instances, we discarded all the ictal periods and a postictal period of 1000 s after the end of each seizure. Therefore, a seizure is discarded if it lasts less than 1000 s from another seizure. Table 2 reports an overview of the patient datasets after the segmentation process described above. More precisely, for each patient, Columns 2 and 3 report the number of seizure and non-seizure instances obtained after the segmentation process, while Columns 3 and 4 report the number of seizure instances used in the Training and Testing phase, respectively, of algorithm *ModThAlgo*.

To assess the quality of the proposed seizure prediction algorithm, we use a cross-validation scheme on each patient, consisting in executing the algorithm many times, varying the data for training and testing at each execution. In such a scheme, the algorithm is executed different times on the data set of a given patient, by changing each time the Training and the Test set, in such a way that each seizure instance is included in the Test set at least once. At each execution, 60% of the instances of a

given patient dataset is used for training and the remaining 40% is used in the Test phase. Furthermore, at each execution, at least half of the seizures instances are included in the Training data (and the remaining in the Test data). For each patient, q rounds of cross-validation have been performed, with the additional condition that each seizure instance is selected for testing in at least one of the q rounds. Parameter q has been set equal to 5 for patients with more than three seizure instances, and equal to 3 for the other patients. Therefore, on each dataset we execute our classification algorithm q times, each time with a different training and test set randomly selected. Then, we evaluate the algorithm by averaging the performance indices over the q runs. Note that, on patient datasets with three seizures the algorithm is executed three times, each time using a Test set including a different seizure instance (while the training set contains the remaining two seizure instances). Columns 5 and 6 of Table 2, for each patient, report the average number of EEG hours over all the rounds of the cross-validation process used in the training phase and in the test phase, respectively.

Table 2. Patients used in the cross-validation scheme.

Pat. Id	Seiz. inst.	Nonseiz. inst.	Train Seiz.	Test Seiz.	Train Hours	Test Hours
PN00	5	0	3	2	0.86	0.54
PN05	3	2	2	1	2.93	1.96
PN06	5	7	3	2	5.45	4.05
PN09	3	3	2	1	2.99	2.98
PN10	8	4	4	4	6.69	4.44
PN12	4	3	2	2	3.15	2.17
PN13	3	4	2	1	4.08	3.09
PN14	4	12	2	2	8.84	6.92

A preliminary experimental campaign has been performed to determine the best setting of the parameters of the feature extraction phase (see Figure 1). According to the campaign, features are extracted by EEG previously filtered by a FIR filter with band 8–13 Hz, and PLI (1) and WPLI (2) are computed for each channel pair by using a time window of length $\Delta_t = 6$ s with an overlap of 1 s. No additional artifact suppression method is used.

In the algorithm, the length of the prediction interval is set to T seconds, where T is given as an input to the prediction algorithm: In the training phase, all the records contained in the time span of length T ending at the actual onset of each seizure are set to true. All the other records are set to false (interictal). The same prediction interval T has been used in the test phase to asses whether a seizure is correctly predicted or not. In our experiments, two different lengths of T have been considered, i.e., 300 and 900 s.

Tables 3 and 4 report the classification results for *ModThAlgo* on the data of Table 2 as well as a comparison with the previous version of the algorithm, i.e., *ThAlgo* presented in [20]. In the tables, *pred*% is the percentage of seizures of the test set correctly predicted and *spec*% is the specificity, computed over all the q runs of the cross-validation scheme, with $q \in \{3; 5\}$. Recall that a seizure is correctly predicted if at least one record inside the correspondent prediction interval is classified as true. The specificity is computed as $TN/(FP + TN) \times 100$, where TN and FP are the number of true negative and false positive records, respectively, detected by the algorithm. Finally, for each patient dataset, $\bar{t}_p$, t_p^m, and t_p^M report on the prediction time in seconds, computed over all the seizures and over the q runs of the algorithm. In particular, $\bar{t}_p$ is the average prediction time, and t_p^m and t_p^M are the minimum and the maximum prediction time in seconds, respectively.

In Table 3, we evaluate the benefit of introducing the postprocessing procedure in the old version of the algorithm *ThAlgo* [20]. To this aim, $T = 300$ is set. In the table, in Columns 2–6 and 7–8, the results of *ThAlgo* with and without the postprocessing procedure (described in Section 2.5) are, respectively, reported. Note that the postprocessing procedure allows to increase the percentage of correctly classified seizures from approximately 96.3% to 98.8%. In terms of total number of false positives, the postprocessing procedure is able to increase the specificity of ~0.7% on average, corresponding to a

false positive reduction of ~35% on average. In summary, the postprocessing procedure reduces the false positive rate and at the same time increases the number of correctly predicted seizures.

Table 3. Classification results of *ThAlgo* with and without postprocessing procedure with $T = 300$.

	ThAlgo [20]					*ThAlgo* [20] with Post Processing				
	pred%	spec%	$\bar{t}_p$	t_p^m	t_p^M	pred%	spec%	$\bar{t}_p$	t_p^m	t_p^M
PN00	100	100	142.80	29	213	100	100	142.8	29	213
PN05	100	98.20	29.67	15	37	100	98.94	24.67	10	32
PN06	100	95.46	152.50	0	294	100	97.09	157.60	0	293
PN09	100	99.98	7.67	0	19	100	100	12.67	0	29
PN10	100	96.28	94.90	0	290	100	96.73	131.05	0	291
PN12	90	96.63	221.67	0	292	100	98.78	157.40	41	289
PN13	100	100	67.00	1	157	100	100	67.00	1	157
PN14	80	98.06	131.50	0	289	80	98.86	134.38	0	269
Av.	96.25	98.08	105.96	5.63	198.88	97.50	98.80	103.44	10.13	196.63

Table 4. Classification results of *ModThAlgo* with $T = 300$ and $T = 900$.

	ModThAlgo with $T = 300$					*ModThAlgo* with $T = 900$				
	pred%	spec%	$\bar{t}_p$	t_p^m	t_p^M	pred%	spec%	$\bar{t}_p$	t_p^m	t_p^M
PN00	100	100	166.3	64	248	100	100	166.3	64	248
PN05	100	99.11	18	2	32	100	99.62	598	420	767
PN06	100	98.4	86.6	0	202	100	99.66	357.3	107	649
PN09	100	100	25.6667	4	64	100	100	25.67	4	64
PN10	100	94.11	140.25	0	294	100	98.69	578.15	216	887
PN12	100	100	58.5	0	293	100	100	58.5	0	293
PN13	100	100	58.6667	1	127	100	100	58.67	1	127
PN14	100	99.39	212.2	33	273	100	99.38	341.2	33	697
Av.	100	98.88	95.77	13	191.63	100	99.67	272.97	105.63	466.5

Table 4 shows the classification results of *ModThAlgo* with a prediction interval *T* of 300 and 900 s. Note that *ModThAlgo* always outperforms *ThAlgo* (Table 3) both in correctly predicted seizures and in specificity, as also discussed in Section 4.

4. Discussion

The experimental analysis performed both by inspection and by a simple classification algorithm performed on the patients of Siena Scalp Database shows that the EEG synchronization measures based on indices PLI (1) and WPLI (2) are effective for seizure prediction. In fact, in Figures 6–11, a rising trend of the indices is usually observed with a peak at the onset. Such trends are clearly evident in all the figures with the exception of Figure 9 (related to patient PN11), where the heights of the peaks immediately preceding the seizure's onset are comparable with the heights of other peaks located before and far from the seizure onset. Note that, during a real-time use, peaks located after the end of a seizure occurrence can be included in a post-ictal period and can be ignored if the seizure is correctly predicted. Furthermore, the classification results of Tables 3 and 4 show that the proposed prediction algorithm *ModThAlgo* always outperforms the older algorithm version (i.e., *ThAlgo*) and sensibly improves the performance both in correctly predicted seizures and in specificity: *ModThAlgo* is able to predict all the seizures with a specificity of 99.67% on average when $T = 900$, corresponding to about two false positives per hour. We also note that, when $T = 900$, *ModThAlgo* attains no false alarm (i.e., $spec\% = 100$) and a seizure prediction rate of 100% on four out eight patients. Concerning prediction times, *ModThAlgo* is able to predict seizures, on average, about 90 s in advance when $T = 300$ and more than 4 minutes in advance when $T = 900$. Obviously, even if larger *T* can be used, we point out that the use of too long prediction intervals might increase the prediction times too much,

introducing, during a real-time use, an uncertainty on the time in which suitable actions must be taken (e.g., to rest in a safe position).

Finally, although *ModThAlgo* has been designed to take into account patient's specificities, *ModThAlgo* could be also executed on Training and Test data belonging to different patients. Such a use could allow highlighting of the similarities in seizure and preictal morphologies of different patients and assessment of how the inter-patient specificity is important in seizure prediction. We leave this topic to a future study.

5. Conclusions

In this work, a database containing EEG scalp signals of epileptic patients acquired at the Unit of Neurology and Neurophysiology at University of Siena is presented. A patient-specific threshold-based algorithm for the prediction of epileptic seizures is also proposed, based on finding synchronization patterns in the EEG by computing synchronization measures on all channel pairs.

The experimental analysis, performed by inspection and by the proposed prediction algorithm, shows that the features extracted by the synchronization measures allow to detect preictal and ictal states and that the computationally viable threshold-based classification algorithm is able to detect preictal states few minutes before the onset. In order to reduce the false positive rates, and for enhancing the reliability of the predictions, directions for future research include the application of new preprocessing procedures able to reduce artifacts and noise in EEG signals, the use of state-of-the-art classification algorithms in place of the simple threshold-based classifier, and the use of new features derived from other physiological signals (e.g., EKG signals).

Author Contributions: P.D. and G.Z.M.d.L. developed the theory and performed the computational experiments. P.D. wrote the manuscript with support from G.V. and G.Z.M.d.L., G.V. worked to the selection of patients and performed the diagnosis of epilepsy. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

References

1. Epilepsy Foundations. Available online: https://www.epilepsy.com/ (accessed on 25 January 2019).
2. Deckers, C.L.P.; Genton, P.; Sills, G.J.; Schmidt, D. Current limitations of antiepileptic drug therapy: A conference review. *Epilepsy Res.* **2003**, *53*, 1–17. [CrossRef]
3. Assi, E.B.; Nguyen, D.K.; Rihana, S.; Sawan, M. Towards accurate prediction of epileptic seizures: A review. *Biomed. Signal Process. Control* **2017**, *34*, 144–157. [CrossRef]
4. Gadhoumi, K.; Lina, J.M.; Mormann, F.; Gotman, J. Seizure prediction for therapeutic devices: A review. *J. Neurosci. Methods* **2016**, *260*, 270–282. [CrossRef] [PubMed]
5. Chisci, L.; Mavino, A.; Perferi, G.; Sciandrone, M.; Anile, C.; Colicchio, G.; Fuggetta, F. Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1124–1132. [CrossRef]
6. Mirowski, P.; Madhavan, D.; LeCun, Y.; Kuzniecky, R. Classification of patterns of EEG synchronization for seizure prediction. *Clin. Neurophysiol.* **2009**, *120*, 1927–1940. [CrossRef]
7. Shoeb, A.H. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
8. Van Esbroeck, A.; Smith, L.; Syed, Z.; Singh, S.; Karam, Z. Multi-task seizure detection: Addressing intra-patient variation in seizure morphologies. *Mach. Learn.* **2016**, *102*, 309–321. [CrossRef]

9. Bandarabadi, M.; Rasekhi, J.; Teixeira, C.A.; Karami, M.R.; Dourado, A. On the proper selection of preictal period for seizure prediction. *Epilepsy Behav.* **2015**, *46*, 158–166. [CrossRef]

10. Fan, M.; Chou, C. Detecting Abnormal Pattern of Epileptic Seizures via Temporal Synchronization of EEG Signals. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 601–608. [CrossRef]

11. Kuhlmann, L.; Freestone, D.; Lai, A.; Burkitt, A.N.; Fuller, K.; Grayden, D.B.; Seiderer, L.; Vogrin, S.; Mareels, I.M.Y.; Cook, M.J. Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons. *Epilepsy Res.* **2010**, *91*, 214–231. [CrossRef]

12. Mormann, F.; Lehnertz, K.; David, P.; Elger, C.E. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Phys. D Nonlinear Phenom.* **2000**, *144*, 358–369. [CrossRef]

13. Mormann, F.; Kreuz, T.; Andrzejak, R.G.; David, P.; Lehnertz, K.; Elger, C.E. Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Res.* **2003**, *53*, 173–185. [CrossRef]

14. Myers, M.H.; Padmanabha, A.; Hossain, G.; de Jongh Curry, A.L.; Blaha, C.D. Seizure prediction and detection via phase and amplitude lock values. *Front. Hum. Neurosci.* **2016**, *10*, 80. [CrossRef]

15. Kramer, M.A.; Cash, S.S. Epilepsy as a disorder of cortical network organization. *Neuroscientist* **2012**, *18*, 360–372. [CrossRef] [PubMed]

16. Lemieux, L.; Daunizeau, J.; Walker, M. Concepts of connectivity and human epileptic activity. *Front. Syst. Neurosci.* **2011**, *5*, 12. [CrossRef] [PubMed]

17. Wendling, F.; Chauvel, P.; Biraben, A.; Bartolomei, F. From intracerebral EEG signals to brain connectivity: Identification of epileptogenic networks in partial epilepsy. *Front. Syst. Neurosci.* **2010**, *4*, 154. [CrossRef]

18. Stam, C.J.; Nolte, G.; Daffertshofer, A. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum. Brain Mapp.* **2007**, *28*, 1178–1193. [CrossRef]

19. Vinck, M.; Oostenveld, R.; Van Wingerden, M.; Battaglia, F.; Pennartz, C.M. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage* **2011**, *55*, 1548–1565. [CrossRef]

20. Detti, P.; de Lara, G.Z.M.; Bruni, R.; Pranzo, M.; Sarnari, F.; Vatti, G. A Patient-specific Approach for Short-term Epileptic Seizures Prediction through the Analysis of EEG synchronization. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 1494–1504. [CrossRef]

21. PANACEE Project. Available online: https://panacee.diism.unisi.it/ (accessed on 25 June 2020).

22. Fisher, R.S.; Cross, J.H.; French, J.A.; Higurashi, N.; Hirsch, E.; Jansen, F.E.; Lagae, L.; Moshé, S.L.; Peltola, J.; Perez, E.R.; et al. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia* **2017**, *58*, 522–530. [CrossRef]

23. Siena Scalp Database. Available online: https://drive.google.com/drive/folders/1Xk0de2pMJhilrCul72QX USOQ8DWuxT53?usp=sharing (accessed on 25 June 2020).

24. Zheng, Y.; Wang, G.; Li, K.; Bao, G.; Wang, J. Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition. *Clin. Neurophysiol.* **2014**, *125*, 1104–1111. [CrossRef]

25. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **2010**, *52*, 1059–1069. [CrossRef]

26. Majumdar, K.K.; Vardhan, P. Automatic seizure detection in ECoG by differential operator and windowed variance. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2011**, *19*, 356–365. [CrossRef]

27. Parvez, M.Z.; Manoranjan, P. Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 158–168. [CrossRef]