


Article

Performance Evaluation of Epileptic Seizure Prediction Using Time, Frequency, and Time–Frequency Domain Measures

Debiao Ma [†], Junteng Zheng [†] and Lizhi Peng ^{*,†}

Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, China; debiaoma@mail.ujn.edu.cn (D.M.); finddreamzz@gmail.com (J.Z.)

* Correspondence: plz@ujn.edu.cn

† These authors contributed equally to this work.

Abstract: The prediction of epileptic seizures is crucial to aid patients in gaining early warning and taking effective intervention. Several features have been explored to predict the onset via electroencephalography signals, which are typically non-stationary, dynamic, and varying from person-to-person. In the former literature, features applied in the classification have shared similar contributions to all patients. Therefore, in this paper, we analyze the impact of the specific combination of feature and channel from time, frequency, and time–frequency domains on prediction performance of disparate patients. Based on the minimal-redundancy-maximal-relevance criterion, the proposed framework uses a sequential forward selection approach to individually find the optimal features and channels. Trained models could discriminate the pre-ictal and inter-ictal electroencephalography with a sensitivity of 90.2% and a false prediction rate of 0.096/h. We also present the comparison between the classification accuracy obtained by the optimal features, several features summarized from optimal features, and the complete set of features from three domains. The results indicate that various patient interpretations have a certain specificity in the selection of feature-channel. Furthermore, the detailed list of optimal features and summarized features are proffered for reference to those who research the corresponding database.

Keywords: EEG; frequency domain; seizure prediction; time domain; time–frequency domain



Citation: Ma, D.; Zheng, J.; Peng, L. Performance Evaluation of Epileptic Seizure Prediction Using Time, Frequency, and Time–Frequency Domain Measures. *Processes* **2021**, *9*, 682. <https://doi.org/10.3390/pr9040682>

Academic Editor: Alessandro Tonacci

Received: 17 March 2021

Accepted: 10 April 2021

Published: 13 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Epilepsy is the fourth most common neurological disorder with approximately 50 million patients worldwide, and it affects people of all ages [1]. Many of the pre-onset symptoms are not visible to observers; thus, family and friends may inadvertently witness them all the time [2]. It is an irrefutable fact that a method to predict the occurrence of epileptic seizure would significantly improve the possibility of treatment.

Electroencephalography (EEG) can be used to study changes in brain activity, which is commonly applied in epilepsy research [3]. Epilepsy studies distinguish two types of EEG signals, based on the way they are recorded, namely intracranial EEG (iEEG) and scalp EEG (sEEG). The iEEG is obtained by invasive electrodes, while sEEG is recorded through electrodes attached to specific locations on the scalp. Since sEEG is obtained in a non-invasive way, it has the advantages of easy access and greater safety.

There are four states for EEG recording of a patient with epilepsy, namely inter-ictal, pre-ictal, ictal, and post-ictal (as shown in Figure 1). The seizure prediction problem involves designing models to distinguish between pre-ictal state and inter-ictal state, while seizure detection focuses on discriminating the ictal state. Although the model in seizure detection can detect seizures, it cannot be used for monitoring and treatment.

Nowadays, machine learning is an advanced technology for epilepsy prediction. Among them, feature extraction is the key procedure. In this area, feature extraction is one of the key procedures for improving performance of classification. Previous work mainly used one or a few features, or focused on the improvement of a certain feature.

Even when multiple features are used, the same features are extracted from different patients. However, due to the nonlinear and non-stationary aspects of EEG signals and the different seizure types between individuals, few and fixed features may not be suitable for all patients. Hence, the aim of this study is to examine whether a patient-specific feature design principle will achieve relatively high improvement rates. To address these issues, we extract features of EEG signals from 18 channels using time, frequency, and time–frequency domains. After that, we use a sequential forward selection method to optimize specific feature-channels for each patient based on the output of minimal-redundancy-maximal-relevance (mRMR). Beyond that, several features statistically analyzed from optimal features and all features extracted from these three domains are also explored to discriminate between pre-ictal and inter-ictal states. In addition, the models trained with the optimal

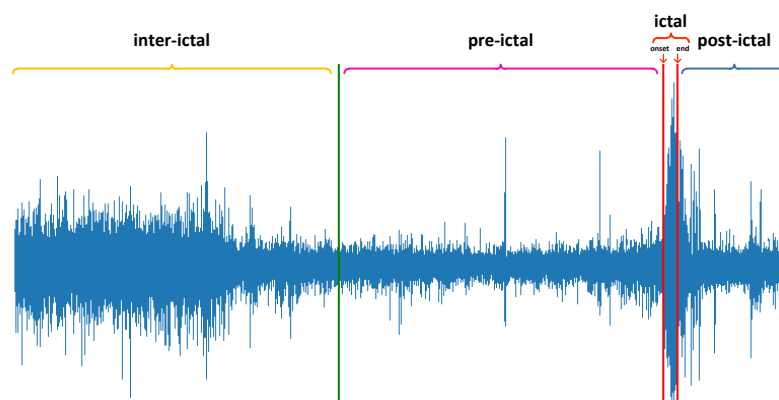


Figure 1. The state of the epileptic seizure. Inter-ictal, pre-ictal, ictal, and post-ictal state. This signal segment comes from channel FP1-F7 of chb01_03.edf file of case chb01 in the CHB-MIT Scalp EEG database. It contains a 15-min inter-ictal period, a 15-min pre-ictal period, a 40-s ictal period, and a 5-min post-ictal period.

features perform well on extra undefined pre-ictal window data of majority of patients. The contributions of this study lie in the following:

1. We verify the importance of feature-specific in seizure prediction.
2. We comprehensively summarize the features of time, frequency, and time–frequency domains and their interpretations in predicting seizures using EEG signals.
3. The optimal features can provide guidance for studying each patient separately, and the several features summarized in these have implications for the general design principle of an epileptic prediction system.

The remainder of the paper is organized as follows. Section 2 provides the details of our proposed method. Section 3 presents the results of this method, discussions of the results, and comparisons with related work. Finally, the paper is concluded in Section 4.

2. Materials and Methodology

The framework suggested in the present study consists of four stages: preprocessing, feature extraction, feature ranking, and classification. The overall implementation process of this methodology is depicted in Figure 2.

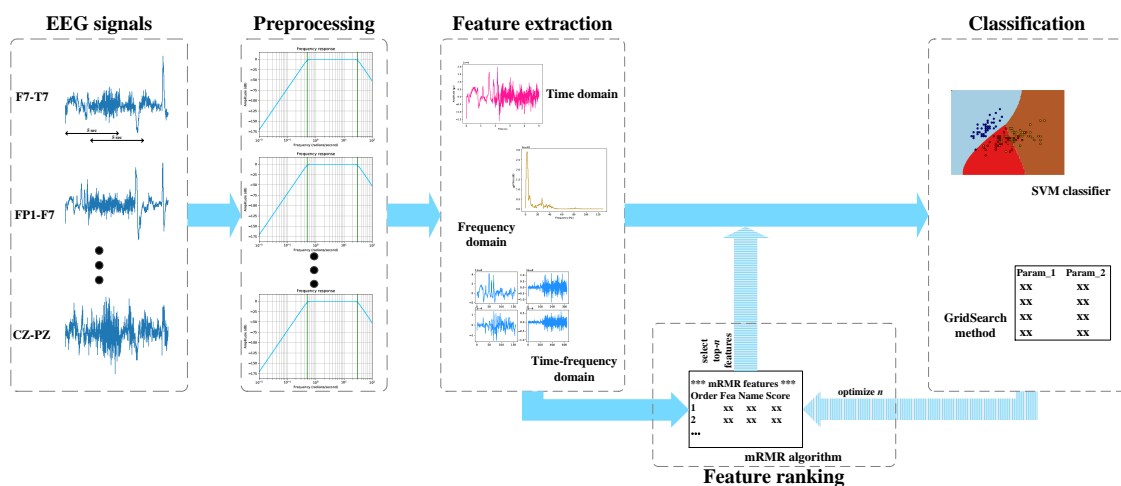


Figure 2. Schematic representation of the proposed framework for seizure prediction.

2.1. EEG Data

The CHB-MIT Scalp EEG database [4] used in this work includes data from 22 pediatric patients with intractable seizures at the Children’s Hospital Boston. EEG signals sampled at 256 samples per second usually contain 23 channels but in some cases contain 24 or 26 channels. The start and end time of seizures judged by clinical experts are stated in annotation files.

Each case contains between 9 and 42 recordings (edf files) from a single patient. All recordings were grouped into 23 cases. In particular, case chb21 was obtained from the same patient 1.5 years after obtaining chb01. For convenience, “patient” and “case” in this work have the same meaning. The detailed information on these 23 cases is listed in Table 1.

2.2. Preprocessing

We picked 18 channels common to all cases: 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 particular, since the different electrodes were used by each patient in multiple experiments, a few recordings in case chb12 do not contain the common channels. Therefore, we drop these recordings in chb12.

As shown in Figure 1, seizure prediction is used to distinguish pre-ictal and inter-ictal states. The pre-ictal interval needs to be defined in the study of seizure prediction; nevertheless, it is still controversial. Mormann [2] believes that, when univariate measures are used to classify the EEG signal, it changes 30 min before onset. Taking into account that activities of EEG signals vary from patient to patient, we use 15 min as the pre-ictal interval. Other principles are as follows:

1. When the second seizure occurs within 15 min after the end of the first seizure, the data of the last seizure are discarded.
2. Use the previous consecutive recording to supplement the data that do not satisfy 15 min.
3. Any previous recording with a gap larger than 5 s from this recording is marked as disconsecutive.
4. In the end, a duration of less than 15 min will also be used as a pre-ictal state of this seizure.

Table 1. Detailed information on the CHB-MIT Scalp EEG database and the data used in this study. Sex: Female (F), Male (M). Seizure type: simple partial seizure (SP), complex partial seizure (CP) and generalized tonic-clonic seizure (GTC). No. of channels: Number of channels included in this database. No. of recordings: Number of edf files. No. of seizures (all/used): number of seizures included in this database (all), number of seizures used in this study (used). No. of samples: Number of 5-s signal segments used in this study.

Case	Sex	Age	Seizure Type	No. of Seizures (All/Used)	No. of Channels	No. of Recordings	No. of Samples
chb01	F	11	SP, CP	7	23	42	4566
chb02	M	11	SP, CP, GTC	3	23	36	1538
chb03	F	14	SP, CP	7	23	38	4218
chb04	M	22	SP, CP, GTC	4	23/24	42	2872
chb05	F	7	CP, GTC	5	23	39	3202
chb06	F	1.5	CP, GTC	10	23	18	6404
chb07	F	14.5	SP, CP, GTC	3	23	19	2154
chb08	M	3.5	SP, CP, GTC	5	23	20	3590
chb09	F	10	CP, GTC	4	23	19	2872
chb10	M	3	SP, CP, GTC	7	23	25	5026
chb11	F	12	SP, CP, GTC	3	23	35	1672
chb12	F	2	SP, CP, GTC	40/11	23/25/24	24	7020
chb13	F	3	SP, CP, GTC	12/10	23/20/18	33	5968
chb14	F	9	CP, GTC	8	23	26	5744
chb15	M	16	SP, CP, GTC	20/17	26/32	40	10,524
chb16	F	7	SP, CP, GTC	10/9	23/18	19	5702
chb17	F	12	SP, CP, GTC	3	23/18	21	2154
chb18	F	18	SP, CP	6/5	18/23	36	3240
chb19	F	19	SP, CP, GTC	3	18/23	30	1672
chb20	F	6	SP, CP, GTC	8	23	29	4690
chb21	F	13	SP, CP	4	23	33	2872
chb22	F	9	-	3	23	31	2154
chb23	F	6	-	7	23	9	4566
Total	-	-	-	185/149	-	664	94,420

Inter-ictal state accounts for 99% of most epilepsy patients' lives; hence, we balance the data by selecting the inter-ictal signals as far away as possible from the ictal state. Each trimmed recording including pre-ictal and inter-ictal state is divided into 5-s segments with 2.5 s overlapping as samples for classification. The details on the data we used are listed in Table 1.

For eliminating low frequency activity and high frequency noise to obtain quasi-stationary signals, a 5th-order Butterworth filter band-passing between 0.5 and 30 Hz is employed.

2.3. Feature Extraction

Feeding the raw EEG signals directly into the classifier may adversely affect the output quality. Thus, we extract features from preprocessed EEG signals. Several features have been determined to discriminate the changes in EEG signals. These features can be categorized based on time domain, frequency domain, and time–frequency domain. A variety of the most common features from these three domains is extracted in this study. The extracted features are summarized in Table 2. Features are described and explained with mathematical expressions in Appendix A.

Table 2. The list of the features extracted from time, frequency, and time–frequency domains.

Time Domain	Basic statistics	Mean, skewness, kurtosis, peak-to-peak amplitude, coefficient of variation
	Energy related	Energy, average power, root mean square, nonlinear energy
	Line length	
	Entropy based	Approximate entropy, sample entropy, singular value decomposition entropy
	Randomly related	Hurst exponent, Higuchi fractal dimension
	Hjorth parameters	Hjorth activity, Hjorth mobility, Hjorth complexity
	Detrended fluctuation analysis	
Frequency Domain	Number of zero-crossings	
	Number of local extrema	
	Energy	
	Intensity weighted	Mean frequency, bandwidth
	Spectral edge frequency	
Time-frequency Domain *	Spectral entropy	
	Peak frequency	
	Basic statistics	Mean of absolute value, skewness, kurtosis, peak-to-peak amplitude
	Energy related	Energy, average power, root mean square
	Line length	
	Randomly related	Hurst exponent, Higuchi fractal dimension

* All features in this domain are extracted on four frequency sub-bands (δ , θ , α , β).

Time domain analysis uses the EEG signal directly, that is, the analysis of electrode voltage amplitude with reference to time series. Time domain analysis can show how the signal changes according to time which is commonly used in other fields [5]. It can reflect the physical meaning of the signal straightforwardly. In terms of time domain features, the basic statistical characteristics (e.g., mean value, skewness and kurtosis) are computed, as well as the energy related, entropy related, randomly related, Hjorth parameters, and the number of zero-crossings and local extrema. In addition, line length can reflect changes in signal amplitude and frequency. The fluctuations in time series can be studied by a detrended fluctuation analysis.

Frequency domain analysis is the analysis of the signal with reference to frequency, while the spectral information of EEG is used. It can reflect the distribution of frequency components or signal power in the whole frequency range. Research in other fields [6] often benefits from frequency domain analysis. Power Spectral Density (PSD) is the measure of a signal's power content versus frequency. The fast Fourier transform is applied using Welch's method [7] to obtain the PSD of signals. The energy and entropy of PSD, intensity weighted mean frequency and bandwidth, edge frequency, and peak frequency are extracted in the frequency domain.

Time-frequency domain analysis, which is available for other fields [8,9], can obtain the relationship between frequency and time. It is necessary to have time–frequency analysis,

as the EEG signals' frequency characteristics are changing constantly. Wavelet transform is known to capture both frequency and location in signals. Delta (0.5–30 Hz), Theta (4–7 Hz), Alpha (8–13 Hz), and Beta (14–30 Hz) are four basic patterns of EEG signals. Following the discrete wavelet transform method and EEG patterns, we use a 3-level filter bank (as shown in Figure 3) with Daubechies 1 (db1) mother wavelet function to decompose each raw EEG signal. Finally, four different frequency sub-bands, which are detail coefficients and approximation coefficients, are formed. Basic statistics, energy based, randomly based, and line length introduced in the time domain are also used in the time–frequency domain. The difference is that features are extracted from four sub-bands in time–frequency domains.

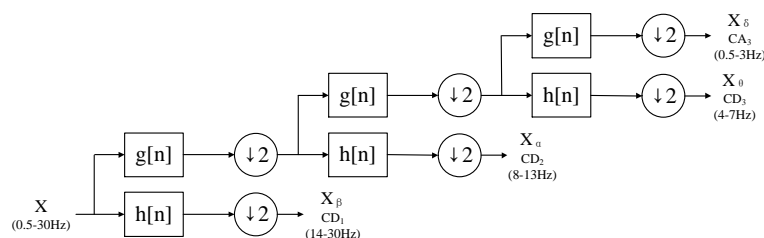


Figure 3. Wavelet decomposition. X : preprocessed EEG signal. h : high-pass filter. g : low-pass filter. Down arrow: subsampled by 2. CD_i : detail coefficients. CA_i : approximation coefficients.

The selection of EEG signal channels is still an open research topic. In order to capture the characteristics of signals on all channels, we extract these 67 features from every channel. Having calculated the measurements using the above parameters, the 18×67 feature matrix is obtained. After aggregating this feature matrix, we acquire a 1206-dimensional vector (called channel-based feature vector or feature-channel vector) for every sample.

2.4. Feature Ranking

The minimal-redundancy-maximal-relevance algorithm [10] selects features not only based on the dependency between features and sample categories, but also considers the relationship between features. In this research, we reorder the variables in the extracted channel-based feature vector based on the mRMR criterion. Suppose there is a dataset D with N samples and M (1206) features. The purpose of feature ranking is to generate a new set S from an M -dimensional set $F = \{f_i, i = 1, \dots, M\}$.

The Max-Dependency criterion uses mutual information as the dependency of features on target class c to select features. Since the Max-Dependency is hard to implement, the mean value of all mutual information values between individual feature f_i and class c , Max-Relevance, is used as an approximation of Max-Dependency:

$$\max D(S, c), D = \frac{1}{|S|} \sum_{f_i \in S} I(f_i; c), \quad (1)$$

where $I(\cdot)$ is the mutual information function, f_i is the i th feature in F , and c is the class label vector. The mutual information of two random variables x and y is defined by their probability density functions $p(x)$, $p(y)$ and the joint probability density function $p(x, y)$:

$$I(x; y) = \iint p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy. \quad (2)$$

Only considering dependency could have rich redundancy, so the relationship between features should be considered. Min-Redundancy condition, which is the mean value of all mutual information values between every two features in S , can be added to select mutually exclusive features. It has the following form:

$$\min R(S), R = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i, f_j). \quad (3)$$

There are two selection schemes, the mutual information quotient (MIQ) and mutual information difference (MID), to combine D and R for discrete data. We denote the operator $\Phi(D, R)$ scheme, and use MIQ in this study. The final optimization is:

$$\max \Phi(D, R), \Phi = D/R. \quad (4)$$

The incremental search method is used to obtain the near-optimal solution. Then, when we want to select m th promising feature from the set $\{F - S_{m-1}\}$, the formula of (4) can be written as:

$$\max_{f_j \in F - S_{m-1}} \left[I(f_j; c) / \frac{1}{m-1} \sum_{f_i \in S_{m-1}} I(f_j; f_i) \right]. \quad (5)$$

2.5. Classification

Support vector machines are classical classifiers most commonly applied in the seizure prediction problem. In this work, we combine a sequential forward selection approach and Support Vector Machine (SVM) to optimize feature subsets and train models.

SVM finds an optimal hyperplane by maximizing the distance between two categories [11]. Given training set $T = \{(x_i, y_i) \mid i = 1, \dots, N\}$ with N samples of n -dimensional vector, where $x_i \in \mathbb{R}^n$, $y_i \in \{+1, -1\}$. The problem of maximizing the margin can be written as a dual problem:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i, \\ \text{s.t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N, \end{aligned} \quad (6)$$

where C is the penalty term that acts as an inverse regularization parameter. $K(\cdot)$ is the kernel function. The radial basis function (RBF) kernel used in this study on two samples x and z is defined as:

$$K(x, z) = \exp(-\gamma \|x - z\|^2), \quad (7)$$

where γ is kernel coefficient. After solving this problem, the optimal b^* can be obtained using α^* . Then, the decision function of classifying sample x becomes:

$$f(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i^* y_i K(x, x_i) + b^*\right). \quad (8)$$

Since hyperparameters directly control the behavior of the SVM algorithm, choosing hyperparameters plays a crucial role in the success of classifiers. Two hyperparameters (C and γ) should be tuning in our classifier. We use a grid-search method to find the most optimal hyperparameter combination. Grid-search exhaustive searches over specified parameter values, which is crude but effective.

Various combinations of feature-channels (n) are tried. The one with the highest model accuracy will be regarded as the optimal subset of a patient. The whole algorithm for finding the optimal combination is as shown in Algorithm 1.

2.6. Performance Evaluation

Cross-validation (CV) is a model validation technique for assessing how the results of a model will generalize to an independent dataset. In order to obtain a reliable outcome, a leave-one-out CV method is used. Compared with k -fold cross-validation, leave-one-out is more practical and deterministic.

Algorithm 1 Finding optimal feature-channels using a sequential forward selection approach

Require:

The reordered feature-channel set $S = \{s_i, i = 1, \dots, N\}$;

Ensure:

The optimal feature-channel subset S' ;

Initial the model accuracy $Acc \leftarrow 0$;

for each $n \in [1, N]$ **do**

 Feed the feature subset $\{s_1, \dots, s_n\}$ to SVM classifier;

 Tune the parameters of SVM using grid-search;

 Use leave-one-out to get model accuracy acc ;

if $acc > Acc$ **then**

$Acc \leftarrow acc$

$S' \leftarrow \{s_1, \dots, s_n\}$

end if

end for

In leave-one-out CV, one observation contains some pre-ictal and inter-ictal samples, which correspond to a seizure. Suppose there are P observations for a certain case, where the value of P depends on the number of seizures (used) in Table 1. Leave-one-out uses one observation as the validation set and the remaining $P - 1$ observations as the training set. This is repeated on all observations to obtain P validation results. The average of these results is defined as the performance of this model.

In order to produce the best and most robust models, classifiers with different hyper-parameters are tried. It is a practical method that the candidate parameter sequence of SVM conforms to exponential growth. Therefore, we do a grid search on $\gamma \in [2^{-15}, 2^{-13}, \dots, 2^3]$ and $C \in [10^{-4}, 10^{-3}, \dots, 10^4]$.

Seven metrics are applied to measure the performance. Accuracy is defined as the percentage of correct classification in the total samples, which can mirror overall performance of the model when using balanced data. The false prediction rate (FPR) is defined as the proportions of the duration which is wrongly predicted as the pre-ictal in the inter-ictal period per hour. Sensitivity (SEN) is the rightness rate of pre-ictal prediction. Moreover, Area Under Curve (AUC), F1, and kappa are used to measure the classification ability. Cost is the average prediction time spent on a sample, including preprocessing, feature extraction, and classification. All experiments are implemented in Python 3.6 on a server of six 3.7 GHz Intel Core (TM) CPUs running Ubuntu 16.04.

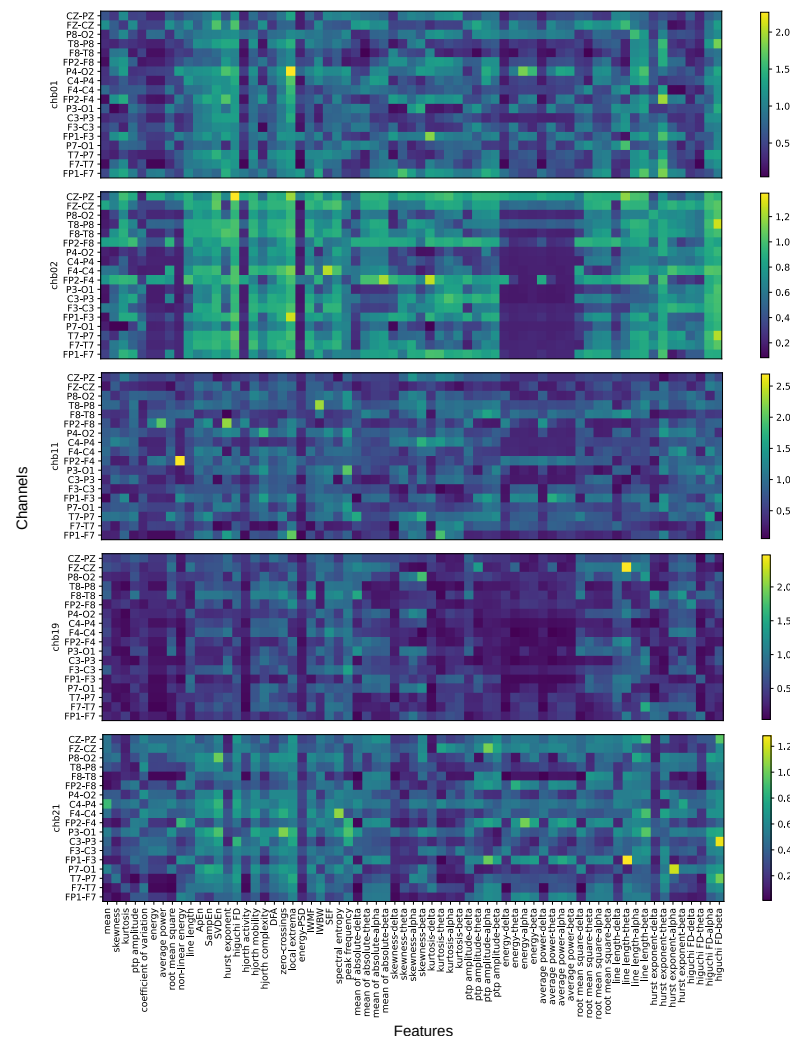
3. Results and Discussion

This section presents the results of the proposed feature design method applied to the scalp EEG data, and compares it with other design principles. Then, visualization of required features and the generalization capacity of models are studied.

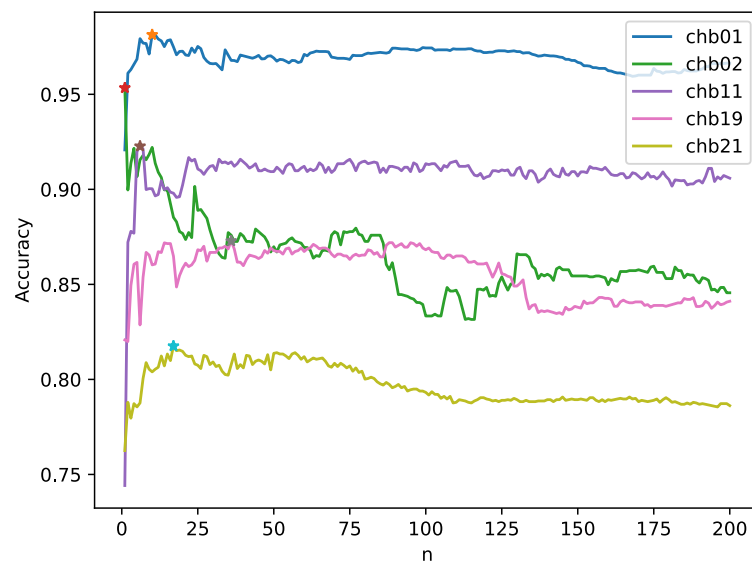
3.1. Performance of Different Numbers of Feature-Channels

As shown in Algorithm 1, combinations with different feature-channels are fed into the classifier to predict seizures in a patient-specific approach. In order to study the prediction performance of different n on different patients, five cases are randomly selected for detailed analysis.

Figure 4a visualizes the mRMR scores of these five cases, which are chb01, chb02, chb11, chb19, and chb21, from top to bottom. The vertical axis represents 18 channels, and the horizontal axis represents 67 features. A brighter color indicates that the feature-channel is more valuable to this patient. The accuracies of five models with different n from 1 to 200 are presented in Figure 4b. The selection order of first n -ranked feature-channels is based on mRMR score.



(a)



(b)

Figure 4. The mRMR scores and accuracy changes of cases chb01, chb02, chb11, chb19, and chb21. (a) visualization of mRMR scores. The higher the score, the more valuable the feature-channel is to the patient; (b) accuracy with various top n feature-channels subsets, where the selection of feature-channel is based on (a).

As can be seen from Figure 4, finding a specific feature-channel subset for each patient is crucial to seizure prediction. The effects of same feature and channel on different patients are diverse. In addition, the number of required feature-channels differ from patient to patient. According to the accuracy curve of each patient, as features are added, the model will gradually reach the optimum. However, with the further increase of features, model accuracy will decline and tend to flatten. Hence, in order to highlight the difference, only the accuracy with n from 1 to 200 are displayed. The selection of feature subset is based on the evaluation of the mRMR algorithm. Therefore, the feature-channel that is added to the feature subset firstly has the maximal relevance with the category. For case chb11, the accuracy of model is 74.4% when only nonlinear energy-FP2-F4 was used in the model. When the fourth feature-channel is added, the accuracy of the model will decrease from 87.8% to 87.7%. This kind of fluctuation is very common in the training and evaluation process of some machine learning algorithms (e.g., SVM), for which the model is not guaranteed to achieve the best performance. However, this fluctuation will not affect the overall trend. Compared to the first six feature-channels, the added feature-channels are redundant as the feature subset is optimal. In particular, case chb01 only needs local extrema-FP1-F3 to achieve the optimal performance. The new redundant feature-channels will reduce the accuracy of the model. Finally, when using first 10, 1, 6, 36, and 17 feature-channels combination, respectively, the optimal model can be obtained for cases chb01, chb02, chb11, chb19, and chb21.

The model trained with these optimal feature-channels is considered the best model for the patient. After 23 patients with 146 seizures in the CHB-MIT sEEG database were evaluated, the number of items in optimal feature-channel combination and classifier parameters of the best model are presented in Table 3. The complete list of optimal feature-channel combinations is described in Appendix B. The features and channels that contribute most to the seizure prediction in each patient are provided for future research. We can see that the number of optimal feature-channels and the combination varies greatly from patient to patient.

3.2. Comparison of Different Feature Design Principles

To examine whether the proposed prediction method is significantly different from the general feature design principle in previous works, comparison and discussion between optimal feature-channel subset, summarized feature subset, and complete feature-channel set are presented.

Summarized feature subset is the combination of the top 11 features (except channel) with the most occurrences in the optimal feature-channel combinations in Table A1. The statistics of a total of 62 features are presented in Figure 5, in which the features with a sky-blue bar are regarded as summarized features. These 11 features are extracted from 18 channels for all patients. Then, 198-D feature vectors are fed into the classifier in a patient-specific method.

The complete channel-based feature set is the combination of all feature-channels extracted from time, frequency, and time–frequency domains. Feature vectors with 1206-D are used to train the SVM classifier.

Figure 6 presents the comparison of classification accuracy obtained by these three feature sets. The optimal feature-channel subset gives an overall accuracy of 90.4%, while summarized feature subset and complete feature set achieve an accuracy of 84.8% and 85.0%, respectively.

Other metrics are also applied to measure the performance of these three feature sets (as shown in Table 4). Models trained using optimal feature subsets reach a sensitivity of 90.2% with FPR of 0.096/h.

Table 3. Optimal number of features and classifier parameters for all cases. n: the number of items in optimal channel-based feature combination. C, gamma: SVM classifier hyperparameters.

Case	n	C	Gamma
chb01	10	10^{-2}	2^{-5}
chb02	1	10^2	2^{-9}
chb03	19	10^4	2^{-11}
chb04	58	10^1	2^{-5}
chb05	4	10^{-4}	2^{-11}
chb06	28	10^{-2}	2^{-11}
chb07	57	10^1	2^{-3}
chb08	30	10^3	2^{-9}
chb09	1	10^{-1}	2^3
chb10	7	10^{-4}	2^{-3}
chb11	6	10^4	2^{-15}
chb12	31	10^1	2^{-5}
chb13	37	10^1	2^{-7}
chb14	55	10^1	2^{-5}
chb15	58	10^1	2^{-5}
chb16	56	10^4	2^{-11}
chb17	6	10^{-1}	2^{-9}
chb18	60	10^2	2^{-5}
chb19	36	10^{-4}	2^{-7}
chb20	26	10^1	2^{-5}
chb21	17	10^{-4}	2^{-9}
chb22	22	10^{-4}	2^{-9}
chb23	37	10^1	2^{-5}

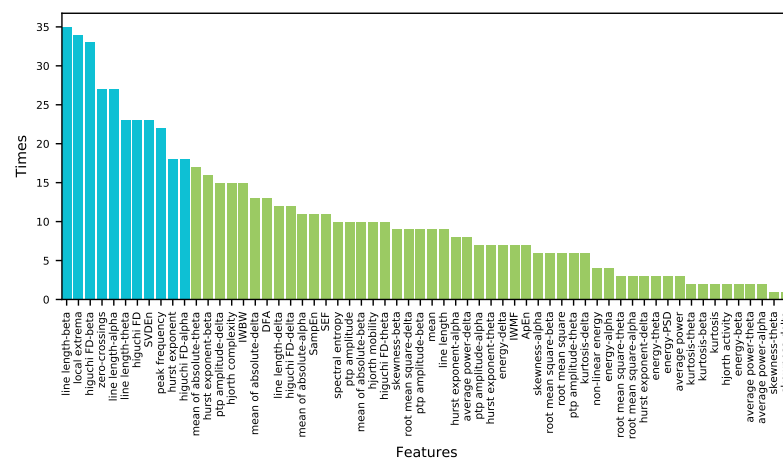


Figure 5. Times of feature appearance in optimal channel-based feature combinations. The top 11 of 62 features with sky-blue bar are regarded as summarized features.

In summary, the overall performance of each patient model obtained when using the optimal feature-channel subset is higher than the other two. Comparing summarized feature subset and complete feature set, even when a generous amount of features in the complete feature set is used, the upgrade in performance is negligible. Features in the complete set may have shared similar contributions, making them redundant in the training process. Especially in chb19, there is a significant difference between the performance of these two feature sets. It also indicates that only specific features can effectively discriminate the signal states in some patients.

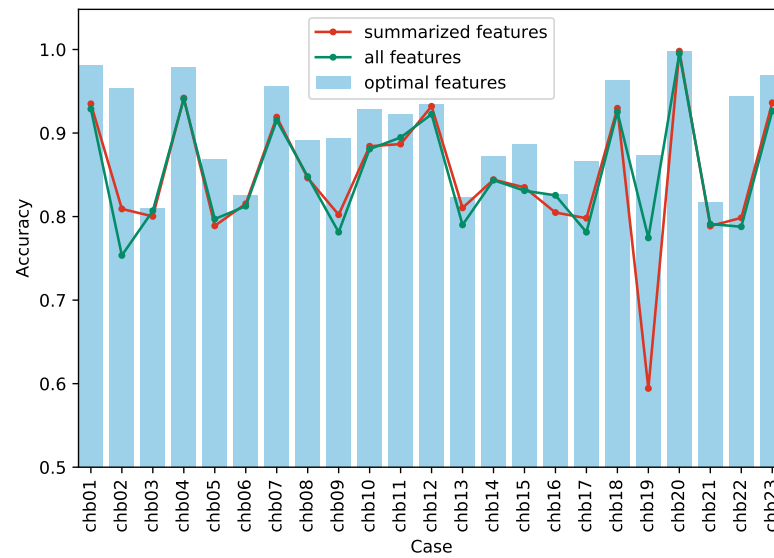


Figure 6. Comparison between the classification accuracy obtained using optimal, summarized, and complete feature sets.

Table 4. Performance of models when using optimal, summarized, and complete feature set.

Case	Optimal Feature Subset					Summarized Feature Subset					Complete Feature Set				
	FPR (h)	SEN	AUC	F1	kappa	FPR (h)	SEN	AUC	F1	kappa	FPR (h)	SEN	AUC	F1	kappa
chb01	0.024	0.986	0.999	0.982	0.963	0.006	0.875	0.996	0.926	0.870	0.103	0.961	0.998	0.939	0.858
chb02	0.055	0.962	0.987	0.954	0.907	0.362	0.980	0.803	0.866	0.618	0.447	0.954	0.846	0.814	0.507
chb03	0.084	0.698	0.876	0.750	0.619	0.155	0.755	0.899	0.772	0.601	0.116	0.735	0.897	0.771	0.614
chb04	0.024	0.981	0.997	0.978	0.956	0.052	0.936	0.974	0.942	0.884	0.054	0.936	0.954	0.941	0.882
chb05	0.152	0.888	0.947	0.869	0.736	0.210	0.788	0.917	0.772	0.578	0.169	0.763	0.949	0.769	0.594
chb06	0.156	0.807	0.827	0.775	0.651	0.210	0.840	0.903	0.787	0.631	0.187	0.812	0.869	0.775	0.625
chb07	0.060	0.973	0.988	0.957	0.913	0.082	0.920	0.957	0.919	0.838	0.085	0.916	0.947	0.915	0.830
chb08	0.116	0.899	0.937	0.894	0.783	0.172	0.865	0.898	0.853	0.692	0.173	0.869	0.902	0.858	0.696
chb09	0.112	0.899	0.903	0.889	0.787	0.124	0.728	0.842	0.741	0.604	0.085	0.648	0.818	0.662	0.563
chb10	0.064	0.920	0.989	0.926	0.856	0.093	0.861	0.971	0.861	0.768	0.092	0.854	0.982	0.857	0.762
chb11	0.075	0.921	0.983	0.922	0.845	0.127	0.900	0.966	0.888	0.774	0.083	0.872	0.963	0.891	0.789
chb12	0.077	0.946	0.975	0.937	0.869	0.071	0.935	0.984	0.932	0.864	0.061	0.906	0.983	0.917	0.845
chb13	0.208	0.855	0.893	0.840	0.647	0.162	0.783	0.828	0.779	0.621	0.172	0.752	0.815	0.758	0.580
chb14	0.126	0.871	0.942	0.872	0.745	0.156	0.844	0.896	0.844	0.688	0.156	0.843	0.898	0.844	0.687
chb15	0.102	0.875	0.954	0.885	0.773	0.163	0.833	0.906	0.835	0.670	0.167	0.829	0.896	0.830	0.662
chb16	0.160	0.813	0.905	0.810	0.653	0.176	0.785	0.866	0.790	0.610	0.161	0.812	0.890	0.811	0.651
chb17	0.108	0.839	0.965	0.862	0.732	0.347	0.943	0.808	0.851	0.596	0.333	0.896	0.826	0.832	0.563
chb18	0.045	0.971	0.993	0.963	0.926	0.074	0.933	0.976	0.930	0.859	0.080	0.931	0.966	0.926	0.851
chb19	0.187	0.933	0.920	0.886	0.746	0.346	0.535	0.563	0.533	0.189	0.168	0.717	0.865	0.765	0.549
chb20	0.002	0.998	1.000	0.998	0.996	0.001	0.997	1.000	0.998	0.996	0.004	0.994	0.999	0.995	0.990
chb21	0.151	0.786	0.886	0.810	0.635	0.289	0.866	0.826	0.816	0.577	0.265	0.847	0.850	0.809	0.582
chb22	0.069	0.955	0.964	0.945	0.887	0.064	0.661	0.937	0.710	0.597	0.063	0.639	0.938	0.686	0.576
chb23	0.041	0.979	0.994	0.971	0.939	0.114	0.987	0.990	0.950	0.872	0.126	0.979	0.987	0.942	0.852
Total	0.096	0.902	0.949	0.899	0.807	0.155	0.850	0.900	0.839	0.696	0.146	0.846	0.915	0.839	0.700

The method of iteratively selecting the optimal features for each patient still suffers from the problem of great time complexity. Therefore, in the case of ensuring acceptable performance, the 11 summarized features (*line length-beta* [12], *local extrema* [13], *Higuchi FD-beta*, *zero-crossings*, *line length-alpha*, *line length-theta*, *Higuchi FD* [14], *SVDEn* [15], *peak frequency* [16], *Hurst exponent* [15,17], and *Higuchi FD-alpha*) that are commonly applied in seizure prediction can still be used as a guide for general features of all patients.

3.3. Visualization of Optimal Feature-Channels

To investigate how the optimal feature-channels contributed to the classification, we project the optimal feature-channel vector of cases chb01, chb20, chb23 and chb14 onto a 2D plane using t-SNE [18] algorithm. The selection principle is based on various *n*, which are 10, 26, 37, and 55, respectively.

Figure 7 shows the dimension reduction results of optimal features of these four cases. The blue dots represent pre-ictal samples, and the orange ones represent inter-ictal samples.

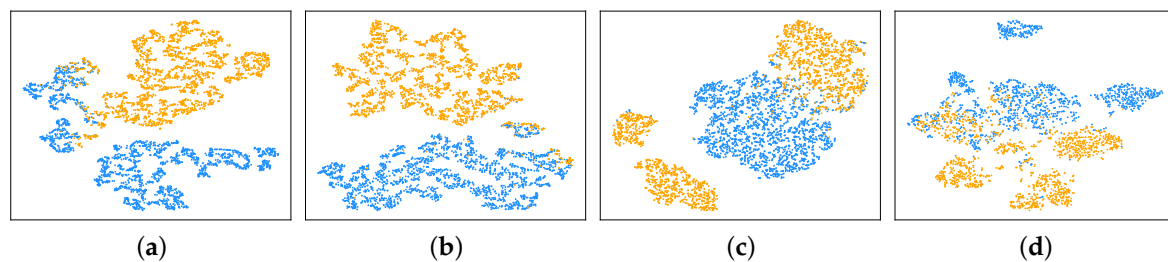


Figure 7. Visualization of optimal feature-channel vector of various patients. (a–d) illustrate the 10-feature, 26-feature, 37-feature and 55-feature subsets from chb01, chb20, chb23 and chb14 embedded onto a 2D space, respectively. The points with blue correspond to pre-ictal samples, and the others represent the inter-ictal samples.

It can be seen that the samples represented by the optimal feature-channels can be well divided into two categories. This indicates that these features are also friendly to other classifiers, not only to SVM.

3.4. Evaluation of Generalization Ability

In order to verify the generalization ability of models, the optimal models are evaluated on the extended pre-ictal data. Many studies have found that characteristic changes in EEG signals occur within 30 min before onset [2,19]. Since the data 15 min before onset are used to train and evaluate the models, we use the same criteria in Section 2.2 to collect the EEG signals from 15 to 30 min before onset as the extended pre-ictal data. These data are divided into three datasets at 5 min intervals, which are –30 to –25 min dataset, –25 to –20 min dataset, and –20 to –15 min dataset.

The evaluation of models trained with optimal feature-channels on these three datasets is presented in Table 5.

Table 5. Evaluation Results of extended pre-ictal data using optimal models. The extended data 15 to 30 min before the onset is divided into three datasets. These three datasets with same number of seizures are –30 to –25 min dataset, –25 to –20 min dataset and –20 to –15 min dataset. # of samples: Number of 5-s segments used for prediction. SEN: The percentage of the samples that are correctly predicted as pre-ictal state. Cost: Average time to predict a sample. AVG SEN: Average sensitivity of these three dataset on each case.

Case	# of Seizures	–30 to –25 min Dataset			–25 to –20 min Dataset			–20 to –15 min Dataset			AVG SEN
		# of Samples	SEN	Cost (sec)	# of Samples	SEN	Cost (sec)	# of samples	SEN	Cost (sec)	
chb01	4	416	0.827	8.48×10^{-2}	476	0.887	9.06×10^{-2}	476	0.964	8.02×10^{-2}	0.893
chb02	2	238	0.996	7.64×10^{-4}	238	0.954	7.50×10^{-4}	238	0.975	7.87×10^{-4}	0.975
chb03	5	565	0.979	7.50×10^{-2}	595	0.887	8.94×10^{-2}	595	0.911	6.91×10^{-2}	0.926
chb04	4	427	0.970	2.69×10^{-1}	476	0.950	2.71×10^{-1}	476	0.968	2.70×10^{-1}	0.963
chb05	3	357	0.706	1.84×10^{-2}	357	0.812	2.19×10^{-2}	357	0.868	1.86×10^{-2}	0.796
chb06	8	921	0.722	6.09×10^{-2}	952	0.808	6.40×10^{-2}	952	0.834	5.81×10^{-2}	0.788
chb07	3	357	0.922	2.74×10^{-1}	357	0.835	2.72×10^{-1}	357	0.846	2.76×10^{-1}	0.867
chb08	5	595	0.839	2.71×10^{-1}	595	0.830	2.72×10^{-1}	595	0.820	2.70×10^{-1}	0.830
chb09	4	476	0.855	6.53×10^{-4}	476	0.815	6.53×10^{-4}	476	0.813	6.66×10^{-4}	0.828
chb10	6	714	0.542	2.92×10^{-2}	714	0.573	2.99×10^{-2}	714	0.731	2.95×10^{-2}	0.615
chb11	1	119	0.916	3.64×10^{-2}	119	0.891	3.71×10^{-2}	119	0.773	3.53×10^{-2}	0.860
chb12	5	395	0.820	6.29×10^{-2}	595	0.805	6.28×10^{-2}	595	0.825	6.19×10^{-2}	0.817
chb13	4	476	0.908	2.73×10^{-1}	476	0.975	2.73×10^{-1}	476	0.977	2.73×10^{-1}	0.953
chb14	5	595	0.677	2.70×10^{-1}	595	0.655	2.70×10^{-1}	595	0.608	2.69×10^{-1}	0.647
chb15	4	372	0.704	2.71×10^{-1}	476	0.681	2.69×10^{-1}	476	0.708	2.69×10^{-1}	0.698
chb16	2	238	0.777	2.73×10^{-1}	238	0.811	2.72×10^{-1}	238	0.777	2.73×10^{-1}	0.789
chb17	3	357	0.675	1.28×10^{-3}	357	0.734	1.22×10^{-3}	357	0.790	1.29×10^{-3}	0.733
chb18	4	476	0.790	2.71×10^{-1}	476	0.651	2.67×10^{-1}	476	0.750	2.70×10^{-1}	0.730
chb19	2	238	0.634	3.20×10^{-2}	238	0.655	3.10×10^{-2}	238	0.840	3.16×10^{-2}	0.710
chb20	2	238	1.000	2.69×10^{-1}	238	1.000	2.69×10^{-1}	238	1.000	2.70×10^{-1}	1.000
chb21	3	357	0.549	2.67×10^{-1}	357	0.437	2.69×10^{-1}	357	0.423	2.70×10^{-1}	0.470
chb22	2	238	0.832	2.71×10^{-1}	238	0.714	2.71×10^{-1}	238	0.748	2.74×10^{-1}	0.765
chb23	5	498	0.994	6.53×10^{-2}	595	0.971	6.11×10^{-2}	595	0.983	5.82×10^{-2}	0.983
Total	86	9663	0.810	1.50×10^{-1}	10234	0.797	1.51×10^{-1}	10234	0.823	1.49×10^{-1}	0.810

Since the pre-ictal state exists within 30 min before onset, most models can also accurately predict the data that is not involved in training. The prediction time of our models are much shorter than 5 s, so it can be easily applied in the application of real-time prediction.

However, some patient models (e.g., chb21) have negative results on these data. The possible reason is that the signal 15 min before onset of this patient is very different from the signal 15 to 30 min before onset. This suggests that the pre-ictal window varies greatly from patient to patient.

Generally, we believe that the closer the distance to onset, the more obvious characteristics. However, many models (e.g., chb03, chb07 et al.) have better prediction results on the dataset of -30 to -25 min than on the one from -25 to -20 min. Therefore, a detailed analysis of time on epilepsy may be worthy of examination in future studies.

3.5. Comparison to Prior Works

A comparison of the proposed framework and other methodology from the works in pre-ictal/inter-ictal classification is presented in Table 6. The focus of the comparison is studies such as ours that have been evaluated within the CHB-MIT database. From the methods perspective, a similar workflow is used, with each work using different sets of features and classifiers. Models trained with optimal features show a promising performance with high sensitivity and low FPR.

Table 6. Comparison with prior works.

Study	# of Used Cases	Pre-Ictal Window (Minutes)	Features	Classifier	SEN	FPR (/h)
Zandi et al., 2013 [20]	3	40	Positive zero-crossing intervals	Bayesian Gaussian mixture model	88.34	0.155
Chu et al., 2017 [21]	13	86	Spectral measure	Warning threshold	86.67	0.367
Alotaiby et al., 2017 [22]	24	120	CSP	LDA	89	0.39
Truong et al., 2018 [23]	13	5	STFT	CNN	81.2	0.16
A. Agboola et al., 2019 [24]	17	60	Normalized Logarithmic Wavelet Packet Coefficient	SVM	87.26	0.08
The proposed framework	23	15	Energy Ratios Time, frequency, time–frequency domain features	SVM	90.2	0.096

4. Conclusions

In this study, mRMR is employed to determine the quality of each channel-based feature from time, frequency, and time–frequency domains. Based on this, we use the prediction accuracy of SVM combined with a sequential forward selection method to determine the optimal subset of features for each case. These models trained with optimal features achieve an overall sensitivity of 90.2% and an FPR of 0.096/h in the classification of pre-ictal and inter-ictal on the CHB-MIT database. In addition, a comparison of the optimal features with the summarized feature subset and the complete set of features from three domains show that finding an optimal feature-channel for each patient represents an important step in seizure prediction. This suggests a valuable future research trajectory of applying a method that combines feature quality measurement with classification to save training time.

Author Contributions: D.M. developed the theory and performed the computational experiments. D.M. and J.Z. wrote the manuscript with support from L.P. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the National Natural Science Foundation of China under Grant No. 61972176, No. 61472164, No. 61672262, No. 61572230, and No. 61573166, the Shandong Provincial Key R&D Program under Grants No. 2018CXGC0706, and No. 2017CXZC1206.

Conflicts of Interest: The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Appendix A. Detailed Description of Features

In this section, the features expressed in italics are briefly described and explained with mathematical expressions.

Appendix A.1. Time Domain Features

Time domain features are directly estimated on raw EEG signals that are a function of time and amplitude. Amplitude refers to the instantaneous energy, which emphasizes the changes. Therefore, the time domain features are the numerical embodiment of the changes of signals. $X = \{x_1, x_2, \dots, x_n\}$ refers to the signal with time n .

- **Basic statistics:**

Basic statistics features have been frequently used to distinguish pre-ictal pattern from an inter-ictal pattern. *Mean*, *skewness*, *kurtosis*, *peak-to-peak amplitude*, and *coefficient of variation* are used in this work. *Skewness* counting the asymmetry of the signal distribution is calculated as:

$$Skew = \frac{1}{n} \sum_{i=1}^n \left[\left(\frac{x_i - \mu}{\sigma} \right)^3 \right],$$

where μ is the mean value, and σ is the standard deviation of signal X . *Kurtosis* can be used to measure the steepness of a signal as follows:

$$Kurtosis = \frac{1}{n} \sum_{i=1}^n \left[\left(\frac{x_i - \mu}{\sigma} \right)^4 \right].$$

Peak-to-peak amplitude quantifies the range of signal. *Coefficient of variation* is a normalized measures of signal dispersion, which is defined as the ratio of standard deviation to average value.

- **Energy related:**

The *energy* is the sum of squares of the signal. The *average power* is the mean square of the signal, which is the average energy. The *root mean square* is the root mean square of the signal, which is the square root of the average power. The last energy related feature is *nonlinear energy*, which was originally presented in [25]. This feature can track the instantaneous frequency of the signal effectively, and its output is given by:

$$NE = \sum_{i=1}^{n-2} (x_i^2 - x_{i-1} \cdot x_{i+1}).$$

- **Line length:**

Line length derived from Katz's fractal dimension was first proposed by [26]. It increases as the amplitude or frequency of the signal increase. The normalized *line length* can be represented as:

$$LL = \frac{1}{n-1} \sum_{i=2}^n |x_i - x_{i-1}|.$$

- **Entropy based:**

Entropy, as the measure of uncertainty and disorder in the data, has been verified in a lot signal processing research. *Approximate entropy (ApEn)*, *sample entropy (Sam-pEn)*, and *singular value decomposition entropy (SVDEn)* are the features used in this work. *ApEn* [27] can be used to quantify the regularity of the signal. The signal X

is cut into subsequence $u_m(i)$ for $\{i \mid 1 \leq i \leq n - m + 1\}$ using m as the sliding window. Subsequence i is presented by $u_i = \{x_i, x_{i+1}, \dots, x_{i+m-1}\}$. Then, the $ApEn$ is defined as:

$$ApEn(m, r, n) = \frac{1}{n - m + 1} \sum_{i=1}^{n-m+1} \ln C_i^m(r) - \frac{1}{n - m} \sum_{i=1}^{n-m} \ln C_i^{m+1}(r),$$

where $C_i^m(r)$ is the proportion of the distance between subsequence u_i and all subsequence less than tolerance value $r = 0.2 * std(X)$. It can be calculated by:

$$C_i^m(r) = \frac{1}{n - m + 1} \sum_{j=1}^{n-m+1} H(r - d(u_i, u_j)),$$

where $H(t)$ is the Heaviside step function, and $d(t)$ is the distance function. $SampEn$ [28] is based upon concepts similar to $ApEn$. It is only for calculated other subsequences when calculation distance, so the $B_i^m(r)$ corresponding to $C_i^m(r)$ is:

$$B_i^m(r) = \frac{1}{n - m} \sum_{j=1}^{n-m+1} H(r - d(u_i, u_j)), j \neq i.$$

In addition, it performs logarithmic operations in the final entropy calculation. Therefore, the $SampEn$ is defined by:

$$SampEn(m, r, n) = \ln \frac{1}{n - m} \sum_{i=1}^{n-m} B_i^m(r) - \ln \frac{1}{n - m} \sum_{i=1}^{n-m} B_i^{m+1}(r).$$

Smaller values indicate more self-similar and regular signals, while larger values characterize higher complexity. $SVDEn$ [29] measures the dimensionality of the signals. It uses the embedding matrix Y , which can be written as:

$$y_i = [x_i, x_{i+j}, \dots, x_{i+(m-1)j}], \\ Y = [y_1, y_2, \dots, y_{n-(m-1)j}]^T,$$

where m is the embedding dimension, and j is the time delay. Then, this entropy is described as follows:

$$SVDEn = - \sum_{i=1}^M \bar{\sigma}_i \log_2 \bar{\sigma}_i,$$

where M is the number of the singular values of the embedding matrix Y and $\bar{\sigma}_i$ is the normalized singular value. Since $SVDEn$ is calculated on all channels, the value can indicate a pattern of pre-ictal signal in both space and time.

- *Hurst exponent:*

Hurst exponent can evaluate the predictability of a time series, and it has been proven that the epileptic brain is long term anticorrelated [30]. Rescaled range analysis is a commonly used calculation method in time series. The cumulative deviate series can be defined as follows:

$$Z_i = \sum_{u=1}^i (x_u - \mu), i = 1, 2, \dots, n,$$

then compute the range:

$$R_n = \max_{1 \leq t \leq n} \{Z_t\} - \min_{1 \leq t \leq n} \{Z_t\}.$$

In addition, the standard deviation is:

$$S_n = \sigma.$$

A *Hurst exponent* complies with the following rules:

$$\left(\frac{R}{S}\right)_n = c * n^{HE}.$$

- *Higuchi Fractal Dimension*:
Fractal dimension can be used to measure the signal complexity. *Higuchi fractal dimension* [31] calculates the fractal dimension of time series in the time domain. Use discrete time interval k to construct a set of new time series $U = \{u_i \mid 1 \leq i \leq k\}$, where $u_i = \{x_i, x_{i+k}, \dots, x_{i+\lfloor \frac{n-i}{k} \rfloor k}\}$, $i = 1, 2, \dots, k$. The average length of new time series u_i is computed as:

$$L_i = \sum_{j=1}^{\lfloor (n-i)/k \rfloor} |x_{i+jk} - x_{i+(j-1)k}| \frac{n-1}{\lfloor \frac{n-i}{k} \rfloor k}.$$

Thus, the total average length of all new series can be written as:

$$L = \sum_{i=1}^k L_i.$$

Finally, *Higuchi fractal dimension* is the slope of the linear regression between $\ln L$ and $\ln \frac{1}{k}$.

- *Hjorth parameters*:
Three Hjorth parameters proposed by [32] can together characterize the EEG signal in terms of amplitude, time scale, and complexity. Hjorth parameters were calculated on raw signal series X , the first derivative of the series X' , and the second derivative X'' . The derivatives were obtained as differences, namely:

$$\begin{aligned} X' &= x_{j+1} - x_j, j = 1, 2, \dots, n-1, \\ X'' &= x'_{j+1} - x'_j, j = 1, 2, \dots, n-2. \end{aligned}$$

The first parameter, *Hjorth activity*, is the variance σ_X^2 of the signal X . The second parameter, *Hjorth mobility*, can be expressed as:

$$mobility = \frac{\sigma_{X'}}{\sigma_X}.$$

The third parameters called *Hjorth complexity* is defined as:

$$complexity = \frac{\sigma_{X''}/\sigma_{X'}}{\sigma_{X'}/\sigma_X}.$$

- *Detrended fluctuation analysis (DFA)*:
DFA [33] is another long-range correlations analysis method, which is similar to the rescaled range analysis. Construct a new series using the mean of X , $u_i = \sum_{j=1}^i (x_j - \mu)$, then segmented into k subintervals using length list $v = \{v_1, v_2, \dots, v_k\}$. In each

subinterval, the data are fitted by polynomial regression to obtain the function \hat{u}_i , and the mean-squared residual is found:

$$F_j = \sqrt{\frac{1}{v_j} \sum_{l=1}^{v_j} (u_l - \hat{u}_l)^2}, j = 1, 2, \dots, k.$$

Finally, *DFA* is the slope of the 1D least-square regression between $\ln v$ and $\ln F$.

- *Number of zero-crossings:*
Zero-crossing, a favorite feature, is a point where the sign of the signal amplitude changes. Therefore, *number of zero-crossings* is these points' quantity in a signal series. It indirectly reflects the change of signal frequency. When this number is large, it infers that there are relatively high frequency components in this signal.
- *Number of local extrema:*
Local extrema consist of local maxima and minima. The local maxima, the so-called peaks, is obtained by a simple comparison of neighboring values. In the same way, local minima is found. *Number of local extrema* is also an indirect measurement of signal frequency similar to the *number of zero-crossings*.

Appendix A.2. Frequency Domain Features

Power Spectral Density describes how the power of a signal is distributed over frequency. We use $P = \{p_1, p_2, \dots, p_n\}$ to denote the signal's PSD with a frequency range from 1 to n and use \hat{P} to denote the normalized the PSD to the total power.

- *Energy-PSD:*
The same as the energy in time domain, the energy of the PSD named *energy-PSD*, similar to the energy in time domain, is the sum of the squares of the PSD.
- *Intensity weighted mean frequency (IWMF):*
IWMF, also known as the mean frequency, is the weighted mean of the frequencies present in the normalized PSD estimation for signal series. It is defined as:

$$IWMF = \sum_{i=1}^n i \hat{p}_i.$$

- *Intensity weighted bandwidth (IWBW):*
IWBW, also called standard deviation frequency, is a measure of the normalized PSD width expressed in standard deviation, and is defined to be:

$$IWBW = \sqrt{\sum_{i=1}^n i (\hat{p}_i - IWMF)^2}.$$

- *Spectral edge frequency (SEF):*
We use SEF_{50} [34], the median frequency, defined as the minimum frequency that can reach 50% of the total spectral power of reference frequency f_{ref} :

$$SEF_{50} = \min\{f^* \mid \sum_{i=1}^{f^*} p_i > \sum_{j=1}^{f_{ref}} p_j \cdot 0.5\}.$$

- *Spectral entropy:*
The entropy of power spectrum [35] can determine EEG's irregularity. *Spectral entropy* is calculated on the normalized PSD using the classical entropy [36]. Thus, it is defined to be:

$$SE = - \sum_{i=1}^n \hat{p}_i \log_2 \hat{p}_i.$$

- **Peak frequency:**
Peak frequency [37], also called dominant frequency, is the frequency at the peak which has the largest average power in its full-width-half-maximum (FWHM) band. The FWHM band is defined by two frequencies, which are within the rising slope and falling slope, respectively, and their amplitudes are equal to half of the peak's amplitude. This feature can find the most prominent rhythmic component of the signal.

Appendix A.3. Time–Frequency Domain Features

Features in this domain are applied to four frequency sub-bands (X_δ , X_θ , X_α , X_β).

Table A1. Optimal feature-channel combinations for every patient. In time and frequency domains, *feature_name-channel_name* denotes the channel-based feature. In time–frequency domain, *feature_name-pattern_name-channel_name* to represent it.

Case	Optimal n	Optimal Feature-Channel Subset
chb01	10	local extrema-P4-O2, hurst exponent-theta-FP2-F4, kurtosis-delta-FP1-F3, hurst exponent-theta-FP1-F7, energy-alpha-P4-O2, Higuchi FD-beta-T8-P8, line length-beta-FP1-F7, local extrema-C4-P4, hurst exponent-FP2-F4, SVDEn-FZ-CZ
chb02	1	local extrema-FP1-F3
chb03	19	hurst exponent-P3-O1, zero-crossings-F8-T8, Higuchi FD-beta-C3-P3, line length-alpha-P3-O1, peak frequency-P8-O2, local extrema-CZ-PZ, energy-delta-F7-T7, skewness-beta-T8-P8, ApEn-CZ-PZ, mean-P7-O1, mean of absolute-delta-P7-O1, Higuchi FD-beta-F3-C3, root mean square-theta-FZ-CZ, Hjorth complexity-P7-O1, root mean square-delta-P7-O1, hurst exponent-P4-O2, hurst exponent-delta-C4-P4, peak frequency-T8-P8, ptp amplitude-alpha-F8-T8
chb04	58	mean of absolute-delta-FP1-F3, mean-FP2-F4, line length-beta-FP1-F3, Higuchi FD-P4-O2, root mean square-delta-FP2-F4, Higuchi FD-alpha-FP2-F8, spectral entropy-FZ-CZ, peak frequency-C3-P3, mean-FP1-F7, peak frequency-P3-O1, ptp amplitude-delta-FP1-F3, Higuchi FD-beta-C4-P4, local extrema-FP2-F8, root mean square-delta-FP1-F3, local extrema-P4-O2, mean of absolute-theta-P3-O1, ptp amplitude-delta-FP2-F4, line length-alpha-FP1-F3, Higuchi FD-C4-P4, mean of absolute-delta-FP2-F4, line length-P3-O1, ptp amplitude-delta-FP1-F7, mean-FP1-F3, mean of absolute-delta-FP1-F7, peak frequency-C4-P4, mean of absolute-beta-P3-O1, zero-crossings-P3-O1, SVDEn-FP2-F4, line length-delta-FP2-F4, line length-beta-FP1-F7, Higuchi FD-beta-P4-O2, local extrema-FP1-F7, SVDEn-FZ-CZ, root mean square-delta-FP1-F7, ptp amplitude-FP2-F4, hurst exponent-P4-O2, line length-FP1-F3, local extrema-FP2-F4, root mean square-FP2-F4, peak frequency-P4-O2, zero-crossings-FZ-CZ, average power-FP1-F7, line length-alpha-FP1-F7, root mean square-F4-C4, Higuchi FD-beta-F8-T8, mean of absolute-beta-FP1-F3, peak frequency-FP1-F3, Higuchi FD-FP2-F8, local extrema-FZ-CZ, SampEn-C3-P3, SVDEn-C4-P4, ptp amplitude-delta-F4-C4, line length-theta-P3-O1, average power-delta-FP1-F3, energy-delta-FP1-F3, energy-delta-FP1-F3
chb05	4	energy-PSD-P7-O1, spectral entropy-FP1-F3, spectral entropy-P7-O1, local extrema-T7-P7
chb06	28	IWBW-FP2-F8, hurst exponent-beta-P3-O1, local extrema-F8-T8, local extrema-T7-P7, hurst exponent-T7-P7, Higuchi FD-delta-P4-O2, hurst exponent-T8-P8, ptp amplitude-beta-T8-P8, line length-beta-F7-T7, Higuchi FD-FP2-F8, DFA-P3-O1, hurst exponent-F8-T8, Higuchi FD-theta-T8-P8, Higuchi FD-beta-F7-T7, ptp amplitude-beta-F8-T8, local extrema-P7-O1, Higuchi FD-alpha-F8-T8, line length-beta-FP2-F8, Hjorth complexity-FP1-F7, line length-alpha-T7-P7, local extrema-T8-P8, Higuchi FD-beta-CZ-PZ, Higuchi FD-beta-T8-P8, Higuchi FD-alpha-T7-P7, skewness-alpha-F8-T8, spectral entropy-P4-O2, hurst exponent-FP2-F8, local extrema-F7-T7
chb07	57	ptp amplitude-delta-FP1-F3, hurst exponent-FP1-F7, Higuchi FD-beta-P3-O1, Hjorth complexity-C3-P3, ptp amplitude-beta-C4-P4, zero-crossings-T8-P8, line length-beta-F3-C3, ptp amplitude-alpha-F8-T8, peak frequency-T7-P7, hurst exponent-theta-P8-O2, ptp amplitude-theta-C4-P4, DFA-P3-O1, line length-beta-P3-O1, hurst exponent-F8-T8, local extrema-P4-O2, spectral entropy-FP1-F3, ptp amplitude-FP2-F4, line length-beta-FP1-F3, hurst exponent-FP2-F8, peak frequency-P3-O1, nonlinear energy-C3-P3, ptp amplitude-delta-FP1-F7, energy-alpha-FP1-F7, energy-alpha-F3-C3, root mean square-beta-FP1-F3, ptp amplitude-beta-F8-T8, ptp amplitude-alpha-C4-P4, Hjorth complexity-C4-P4, energy-theta-C3-P3, nonlinear energy-C4-P4, peak frequency-C3-P3, SampEn-P7-O1, hurst exponent-F7-T7, average power-alpha-C3-P3, SEF-P7-O1, Higuchi FD-theta-P4-O2, spectral entropy-P8-O2, ptp amplitude-delta-FP2-F4, line length-delta-CZ-PZ, Higuchi FD-T8-P8, IWBW-FP1-F7, peak frequency-FP2-F8, line length-theta-C4-P4, line length-alpha-P3-O1, hurst exponent-alpha-FP1-F7, Higuchi FD-beta-C3-P3, ptp amplitude-delta-FP2-F8, energy-beta-FP1-F7, energy-beta-C4-P4, peak frequency-P7-O1, root mean square-alpha-FP1-F3, mean of absolute-beta-P4-O2, IWMF-P3-O1, SVDEn-FP2-F4, Higuchi FD-beta-C4-P4, average power-delta-C3-P3
chb08	30	mean of absolute-delta-F3-C3, Higuchi FD-alpha-CZ-PZ, line length-alpha-C4-P4, SampEn-P3-O1, IWMF-F3-C3, Higuchi FD-delta-F8-T8, line length-delta-T8-P8, line length-delta-C3-P3, ptp amplitude-beta-T8-P8, zero-crossings-T8-P8, line length-theta-C4-P4, hurst exponent-FP2-F4, skewness-beta-P8-O2, hurst exponent-alpha-T8-P8, Higuchi FD-alpha-C4-P4, Hjorth mobility-CZ-PZ, zero-crossings-C4-P4, Higuchi FD-theta-C4-P4, Hjorth complexity-F4-C4, Higuchi FD-theta-P8-O2, hurst exponent-beta-F4-C4, ptp amplitude-delta-FZ-CZ, root mean square-P3-O1, ApEn-T8-P8, SampEn-C4-P4, mean of absolute-delta-P3-O1, mean of absolute-theta-T8-P8, line length-theta-P8-O2, Higuchi FD-alpha-F4-C4, mean of absolute-theta-C3-P3
chb09	1	peak frequency-P8-O2
chb10	7	mean of absolute-theta-T8-P8, line length-C4-P4, hurst exponent-T7-P7, ptp amplitude-T7-P7, line length-theta-FP1-F7, ptp amplitude-delta-FP1-F7, Higuchi FD-delta-F8-T8
chb11	6	nonlinear energy-FP2-F4, local extrema-FP1-F7, peak frequency-P3-O1, zero-crossings-T7-P7, IWBW-T8-P8, Hjorth complexity-P4-O2
chb12	31	root mean square-beta-F8-T8, ptp amplitude-FZ-CZ, local extrema-F4-C4, hurst exponent-beta-P3-O1, mean of absolute-theta-P3-O1, local extrema-T8-P8, Higuchi FD-theta-FP1-F7, line length-beta-FP2-F4, ptp amplitude-FP2-F8, ptp amplitude-beta-T8-P8, Higuchi FD-P3-O1, hurst exponent-F8-T8, local extrema-F8-T8, mean of absolute-theta-P4-O2, hurst exponent-delta-FP1-F7, ptp amplitude-theta-FP2-F8, Higuchi FD-F4-C4, IWBW-P8-O2, IWBW-F8-T8, ptp amplitude-theta-FZ-CZ, ptp amplitude-beta-T7-P7, zero-crossings-CZ-PZ, line length-beta-T8-P8, hurst exponent-alpha-P8-O2, IWBW-P4-O2, Higuchi FD-beta-P7-O1, root mean square-delta-F4-C4, Higuchi FD-beta-T8-P8, mean of absolute-alpha-P3-O1, ptp amplitude-delta-FP2-F8
chb13	37	SEF-C4-P4, skewness-delta-FP1-F7, mean of absolute-beta-C3-P3, Hjorth complexity-C3-P3, local extrema-P4-O2, average power-delta-C4-P4, ApEn-C3-P3, Higuchi FD-beta-C3-P3, kurtosis-delta-F7-T7, mean of absolute-delta-FP1-F7, Hjorth complexity-P3-O1, energy-delta-F8-T8, SVDEn-C4-P4, hurst exponent-C3-P3, local extrema-P8-O2, hurst exponent-beta-C3-P3, kurtosis-delta-FP1-F7, mean-C3-P3, zero-crossings-C4-P4, local extrema-P3-O1, hurst exponent-beta-F8-T8, SVDEn-P4-O2, SampEn-C3-P3, energy-delta-C4-P4, average power-delta-P4-O2, average power-delta-F8-T8, zero-crossings-C4-O2, kurtosis-delta-FP1-F3, hurst exponent-alpha-C4-P4, SVDEn-C3-P3, Higuchi FD-C3-P3, line length-theta-C3-P3, Hjorth mobility-C4-P4, local extrema-P7-O1, peak frequency-C3-P3, SEF-F8-T8, mean of absolute-delta-C4-P4
chb14	55	Higuchi FD-alpha-P8-O2, line length-alpha-FP1-F7, Hjorth activity-FP1-F7, Higuchi FD-delta-P7-O1, energy-PSD-FP1-F3, Higuchi FD-alpha-CZ-PZ, root mean square-beta-FP1-F7, hurst exponent-beta-FP1-F7, line length-beta-T8-P8, line length-beta-FP2-F8, ptp amplitude-delta-FP1-F7, local extrema-T8-P8, SampEn-T8-P8, line length-beta-FP1-F3, line length-alpha-F7-T7, mean-FP1-F7, SVDEn-F7-T7, peak frequency-P3-O1, mean of absolute-theta-FP1-F7, Higuchi FD-delta-P4-O2, line length-alpha-F8-T8, Hjorth activity-FP2-F8, IWMF-P7-O1, line length-beta-FP1-F7, line length-alpha-FP2-F4, SampEn-P8-O2, IWBW-FP1-F7, Hjorth complexity-FP1-F7, Higuchi FD-delta-F4-C4, line length-alpha-FP2-F8, root mean square-FP1-F7, IWMF-P3-O1, line length-alpha-T8-P8, Higuchi FD-CZ-PZ, ptp amplitude-theta-FP1-F7, zero-crossings-T8-P8, IWBW-FP2-F8, mean of absolute-alpha-FP1-F7, line length-beta-F8-T8, average power-FP1-F7, Higuchi FD-delta-C4-P4, zero-crossings-P7-O1, ptp amplitude-beta-F7-T7, DFA-FP1-F3, ptp amplitude-FP1-F7, ptp amplitude-F7-T7, mean of absolute-beta-FP1-F7, hurst exponent-beta-F3-C3, SVDEn-T8-P8, energy-theta-FP2-F8, zero-crossings-P8-O2, ptp amplitude-delta-FP1-F3, line length-theta-FP2-F8, ApEn-FP1-F7
chb15	58	mean of absolute-alpha-P4-O2, kurtosis-FZ-CZ, average power-delta-FP2-F4, line length-beta-P4-O2, Hjorth complexity-P3-O1, Higuchi FD-beta-P4-O2, line length-theta-FZ-CZ, IWBW-P8-O2, line length-alpha-P4-O2, mean-FP1-F3, SVDEn-T7-P7, kurtosis-delta-F8-T8, mean of absolute-theta-C4-P4, ptp amplitude-delta-FP2-F4, ApEn-C4-P4, skewness-beta-FP1-F3, mean of absolute-beta-P4-O2, average power-delta-FP1-F3, Higuchi FD-theta-FP1-F7, zero-crossings-P3-O1, local extrema-P4-O2, mean of absolute-alpha-C4-P4, root mean square-delta-P8-O2, local extrema-F4-C4, Higuchi FD-beta-C4-P4, Hjorth complexity-F3-C3, DFA-T7-P7, line length-alpha-C4-P4, line length-P4-O2, mean of absolute-beta-P8-O2, ptp amplitude-theta-FP1-F3, hurst exponent-theta-F8-T8, line length-theta-C4-P4, energy-delta-FP2-F4, mean-FP1-F7, kurtosis-FP2-F4, energy-PSD-FP2-F8, ptp amplitude-FP1-F7, DFA-P3-O1, hurst exponent-theta-F7-T7, line length-beta-P8-O2, ptp amplitude-P4-O2, Higuchi FD-C4-P4, average power-FP2-F4, root mean square-theta-P8-O2, spectral entropy-T7-P7, line length-C4-P4, Higuchi FD-alpha-FP1-F7, energy-delta-FP1-F3, local extrema-F3-C3, line length-P8-O2, line length-theta-F4-C4, mean of absolute-beta-C4-P4, Higuchi FD-theta-FP2-F4, root mean square-P8-O2, ApEn-P4-O2, SEF-P3-O1, ptp amplitude-delta-FP1-F3
chb16	56	line length-alpha-CZ-PZ, hurst exponent-P4-O2, Hjorth mobility-F7-T7, Higuchi FD-FZ-CZ, mean of absolute-delta-C3-P3, line length-alpha-P4-O2, Hjorth mobility-C3-P3, hurst exponent-alpha-P7-O1, SVDEn-P7-O1, peak frequency-P8-O2, Higuchi FD-CZ-PZ, line length-delta-C3-P3, Higuchi FD-beta-P7-O1, SEF-F8-O2, SVDEn-F4-C4, Higuchi FD-beta-C4-P4, hurst exponent-P3-O1, Higuchi FD-beta-FZ-CZ, ptp amplitude-beta-CZ-PZ, DFA-FP2-F8, line length-beta-P4-O2, DFA-P7-O1, SVDEn-F3-C3, hurst exponent-theta-P3-O1, SVDEn-F7-T7, Higuchi FD-beta-P4-O2, root mean square-beta-C3-P3, SVDEn-P8-O2, Higuchi FD-F8-T8, Higuchi FD-delta-T8-P8, SVDEn-P3-O1, root mean square-beta-FP1-F7, Higuchi FD-beta-CZ-PZ, Hjorth mobility-P7-O1, average power-theta-P8-O2, IWMF-C3-P3, hurst exponent-beta-P7-O1, DFA-F3-C3, Higuchi FD-alpha-FZ-CZ, Higuchi FD-beta-C3-P3, line length-beta-C4-P4, root mean square-theta-C3-P3, IWBW-T8-P8, Hjorth mobility-C4-P4, zero-crossings-F7-T7, SEF-P7-O1, DFA-P8-O2, ptp amplitude-alpha-P8-O2, DFA-F4-C4, Higuchi FD-P4-O2, Higuchi FD-alpha-CZ-PZ, root mean square-delta-C3-P3, Hjorth mobility-F3-C3, hurst exponent-beta-T7-P7, Higuchi FD-P3-O1, SVDEn-C3-P3
chb17	6	line length-alpha-FZ-CZ, energy-alpha-P8-O2, zero-crossings-FP1-F7, Higuchi FD-CZ-PZ, mean of absolute-theta-F7-T7, local extrema-FP1-F7
chb18	60	kurtosis-theta-FP1-F3, line length-alpha-FP2-F4, line length-delta-CZ-PZ, zero-crossings-FZ-CZ, Higuchi FD-theta-P3-O1, skewness-beta-FP2-F4, line length-beta-FP2-F4, mean of absolute-delta-P7-O1, skewness-beta-T7-P7, SVDEn-FP2-F4, line length-delta-P7-O1, Higuchi FD-alpha-P3-O1, line length-alpha-F4-C4, kurtosis-beta-FP2-F4, Higuchi FD-theta-P7-O1, local extrema-P3-O1, line length-beta-FP1-F3, line length-alpha-FP7-O1, root mean square-P7-O1, zero-crossings-FP2-F4, line length-theta-FP2-F4, ptp amplitude-alpha-P7-O1, skewness-alpha-FP1-F7, root mean square-beta-F4-C4, skewness-alpha-F7-T7, mean of absolute-theta-CZ-PZ, Higuchi FD-theta-FP2-F4, skewness-alpha-FP2-F4, line length-FP2-F4, mean of absolute-theta-P7-O1, line length-beta-F4-C4, line length-alpha-P3-O1, root mean square-delta-P7-O1, Higuchi FD-alpha-FP2-F4, kurtosis-theta-FP2-F4, line length-alpha-FP1-F3, root mean square-alpha-P7-O1, energy-theta-P7-O1, peak frequency-P3-O1, ApEn-FP2-F4, line length-F4-C4, mean of absolute-alpha-P3-O1, Higuchi FD-alpha-P7-O1, mean of absolute-beta-FP2-F4, zero-crossings-FP2-F8, mean of absolute-alpha-CZ-PZ, IWBW-P7-O1, kurtosis-beta-FP2-F8, mean of absolute-alpha-P7-O1, DFA-FP2-F4, Higuchi FD-alpha-F4-C4, average power-alpha-P7-O1, skewness-theta-FP1-F3, mean of absolute-alpha-FP2-F4, root mean square-alpha-F4-C4, SVDEn-P3-O1, line length-beta-FP2-F8, average power-theta-P7-O1, mean of absolute-theta-P3-O1, mean of absolute-beta-F4-C4

Table A1. Cont.

Case	Optimal n	Optimal Feature-Channel Subset
chb19	36	line length-theta-FZ-CZ, skewness-beta-P8-O2, line length-beta-F7-T7, mean of absolute-delta-F8-T8, peak frequency-F3-C3, IWMF-F8-T8, ptp amplitude-theta-FZ-CZ, line length-beta-P8-O2, hurst exponent-beta-F8-T8, hurst exponent-delta-F8-T8, SEF-F4-C4, zero-crossings-F8-T8, mean of absolute-delta-P3-O1, line length-theta-FP2-F4, local extrema-FP2-F8, spectral entropy-F8-T8, Higuchi FD-delta-FP2-F8, skewness-beta-P7-O1, line length-alpha-FZ-CZ, skewness-alpha-P4-O2, mean of absolute-delta-FZ-CZ, SEF-F8-T8, line length-beta-C3-P3, Hjorth mobility-F8-T8, line length-delta-FP2-F4, Hjorth complexity-F8-T8, line length-theta-P4-O2, skewness-alpha-P8-O2, hurst exponent-beta-F4-C4, root mean square-delta-F8-T8, line length-theta-FP1-F3, peak frequency-P4-O2, DFA-F8-T8, line length-theta-FP1-F7, line length-beta-FZ-CZ, mean of absolute-theta-FZ-CZ
chb20	26	line length-beta-P4-O2, line length-theta-P4-O2, hurst exponent-beta-F4-C4, DFA-P3-O1, Hjorth complexity-T8-P8, line length-theta-F7-T7, line length-delta-FP2-F8, line length-beta-P3-O1, Higuchi FD-alpha-F4-C4, Higuchi FD-alpha-P4-O2, mean of absolute-alpha-FP1-F7, zero-crossings-CZ-PZ, Higuchi FD-delta-P8-O2, SEF-F3-C3, hurst exponent-alpha-F3-C3, mean of absolute-alpha-F8-T8, zero-crossings-P4-O2, Hjorth complexity-F4-C4, hurst exponent-beta-C3-P3, line length-beta-C3-P3, Higuchi FD-alpha-FZ-CZ, line length-delta-FP1-F3, spectral entropy-C4-P4, line length-alpha-P3-O1, line length-theta-T8-P8, SEF-C3-P3
chb21	17	line length-theta-FP1-F3, Higuchi FD-beta-C3-P3, SEF-P3-O1, IWMF-F4-C4, energy-alpha-FP2-F4, line length-beta-P3-O1, Higuchi FD-beta-T7-P7, ptp amplitude-alpha-FZ-CZ, zero-crossings-P3-O1, ptp amplitude-alpha-FP1-F3, SVDen-P8-O2, Higuchi FD-beta-CZ-PZ, line length-beta-F4-C4, line length-beta-CZ-PZ, SampEn-P7-O1, SVDen-P3-O1, nonlinear energy-FP2-F4
chb22	22	line length-theta-P8-O2, line length-alpha-P4-O2, SVDen-FZ-CZ, mean of absolute-theta-T7-P7, zero-crossings-CZ-PZ, hurst exponent-beta-T8-P8, Hjorth complexity-P8-O2, IWBW-P4-O2, SampEn-F7-T7, SampEn-P4-O2, skewness-beta-P7-O1, Higuchi FD-delta-F3-C3, Higuchi FD-delta-T7-P7, hurst exponent-delta-F7-T7, line length-beta-P4-O2, hurst exponent-alpha-T7-P7, zero-crossings-FZ-CZ, line length-theta-P7-O1, line length-delta-T8-P8, hurst exponent-beta-T7-P7, spectral entropy-FZ-CZ, hurst exponent-alpha-P8-O2
chb23	37	line length-alpha-P4-O2, mean of absolute-theta-F3-C3, skewness-beta-FP1-F3, Higuchi FD-beta-F4-C4, Higuchi FD-beta-P7-O1, mean of absolute-theta-FP1-F3, Higuchi FD-CZ-PZ, ptp amplitude-P4-O2, peak frequency-C3-P3, Higuchi FD-beta-F3-C3, hurst exponent-theta-F7-T7, Higuchi FD-beta-P3-O1, Hjorth mobility-P4-O2, Higuchi FD-beta-F8-T8, average power-delta-T8-P8, zero-crossings-P3-O1, Higuchi FD-F3-C3, Higuchi FD-P4-O2, SampEn-P3-O1, Higuchi FD-P3-O1, Higuchi FD-P7-O1, line length-theta-FP1-F7, Higuchi FD-beta-CZ-PZ, line length-delta-P4-O2, local extrema-F3-C3, Higuchi FD-T8-P8, line length-beta-P4-O2, Hjorth mobility-P3-O1, energy-delta-F8-T8, mean-T8-P8, zero-crossings-P4-O2, IWBW-CZ-PZ, local extrema-F4-C4, Higuchi FD-F4-C4, Higuchi FD-beta-P4-O2, local extrema-P3-O1, IWBW-FP1-F3

- **Basic statistics:**
Four statistical features are employed. *Mean of absolute value*, *skewness*, *kurtosis*, and *peak-to-peak amplitude* are mean of coefficients' absolute values, skewness, kurtosis, and peak-to-peak amplitude of the coefficients in every sub-band, respectively.
- **Energy related:**
Similar to time domain, *energy*, *average power*, and *root mean square* are used to observe every sub-band's coefficients amplitude.
- **Line length:**
Line length can efficiently measure the fractal dimension for each EEG pattern.
- **Randomly related:**
Hurst exponent and *Higuchi fractal dimension* can represent the randomness of each decomposed sub-band.

Appendix B. Optimal Feature-Channel Combinations

Below, we present the feature-channel subsets (as shown in Table A1) used to obtain the optimal models. In time and frequency domains, *feature_name-channel_name* denotes extracted items in channel-based feature vectors, e.g., *mean-FP1-F7*. Meanwhile, in the time–frequency domain, we use *feature_name-pattern_name-channel_name* to represent it, e.g., *skewness-delta-FP1-F7*.

References

1. Iasemidis, L.; Shiau, D.-S.; Chaovalitwongse, W.; Sackellares, J.; Pardalos, P.; Principe, J.; Carney, P.; Prasad, A.; Veeramani, B.; Tsakalis, K. Adaptive epileptic seizure prediction system. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 616–627. [CrossRef]
2. Mormann, F.; Kreuz, T.; Rieke, C.; Andrzejak, R.G.; Kraskov, A.; David, P.; Elger, C.E.; Lehnertz, K. On the predictability of epileptic seizures. *Clin. Neurophysiol.* **2005**, *116*, 569–587. [CrossRef] [PubMed]
3. Freestone, D.R.; Karoly, P.J.; Cook, M.J. A forward-looking review of seizure prediction. *Curr. Opin. Neurol.* **2017**, *30*, 167–173. [CrossRef] [PubMed]
4. Shoeb, A.; Guttag, J. Application of machine learning to epileptic seizure detection. In Proceedings of the ICML 2010—Proceedings, 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 975–982.
5. Aggarwal, Y.; Das, J.; Mazumder, P.M.; Kumar, R.; Sinha, R.K. Heart rate variability time domain features in automated prediction of diabetes in rat. *Phys. Eng. Sci. Med.* **2021**, *44*, 45–52. [CrossRef] [PubMed]
6. Benhassine, N.E.; Boukaache, A.; Boudjehem, D. Classification of mammogram images using the energy probability in frequency domain and most discriminative power coefficients. *Int. J. Imaging Syst. Technol.* **2020**, *30*, 45–56. [CrossRef]
7. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [CrossRef]
8. Cai, J.; Zhou, H.; Huang, W.; Wen, B. Ship Detection and Direction Finding Based on Time–Frequency Analysis for Compact HF Radar. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 72–76. [CrossRef]
9. Hassani Saadi, H.; Sameni, R.; Zollanvari, A. Interpretive time–frequency analysis of genomic sequences. *BMC Bioinform.* **2017**, *18*, 154. [CrossRef] [PubMed]
10. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef]
11. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]

12. Al-Bakri, A.F.; Villamar, M.F.; Haddix, C.; Bensalem-Owen, M.; Sunderam, S. Noninvasive seizure prediction using autonomic measurements in patients with refractory epilepsy. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–22 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2422–2425. [[CrossRef](#)]
13. Niknazar, H.; Maghooli, K.; Motie Nasrabadi, A. Epileptic Seizure Prediction using Statistical Behavior of Local Extrema and Fuzzy Logic System. *Int. J. Comput. Appl.* **2015**, *113*, 24–30. [[CrossRef](#)]
14. Khoa, T.Q.D.; Ha, V.Q.; Toi, V.V. Higuchi Fractal Properties of Onset Epilepsy Electroencephalogram. *Comput. Math. Methods Med.* **2012**, *2012*, 1–6. [[CrossRef](#)] [[PubMed](#)]
15. Zhang, Y.; Yang, S.; Liu, Y.; Zhang, Y.; Han, B.; Zhou, F. Integration of 24 Feature Types to Accurately Detect and Predict Seizures Using Scalp EEG Signals. *Sensors* **2018**, *18*, 1372. [[CrossRef](#)] [[PubMed](#)]
16. Minasyan, G.R.; Chatten, J.B.; Chatten, M.J.; Harner, R.N. Patient-Specific Early Seizure Detection From Scalp Electroencephalogram. *J. Clin. Neurophysiol.* **2010**, *27*, 163–178. [[CrossRef](#)] [[PubMed](#)]
17. Namazi, H.; Kulish, V.V.; Hussaini, J.; Hussaini, J.; Delaviz, A.; Delaviz, F.; Habibi, S.; Ramezanpoor, S. A signal processing based analysis and prediction of seizure onset in patients with epilepsy. *Oncotarget* **2016**, *7*, 342–350. [[CrossRef](#)] [[PubMed](#)]
18. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
19. Maiwald, T.; Winterhalder, M.; Aschenbrenner-Scheibe, R.; Voss, H.U.; Schulze-Bonhage, A.; Timmer, J. Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Phys. D Nonlinear Phenom.* **2004**, *194*, 357–368. [[CrossRef](#)]
20. Shahidi Zandi, A.; Tafreshi, R.; Javidan, M.; Dumont, G.A. Predicting Epileptic Seizures in Scalp EEG Based on a Variational Bayesian Gaussian Mixture Model of Zero-Crossing Intervals. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 1401–1413. [[CrossRef](#)]
21. Chu, H.; Chung, C.K.; Jeong, W.; Cho, K.H. Predicting epileptic seizures from scalp EEG based on attractor state analysis. *Comput. Methods Programs Biomed.* **2017**, *143*, 75–87. [[CrossRef](#)]
22. Alotaiby, T.N.; Alshebeili, S.A.; Alotaibi, F.M.; Alrshoud, S.R. Epileptic Seizure Prediction Using CSP and LDA for Scalp EEG Signals. *Comput. Intell. Neurosci.* **2017**, *2017*, 1–11. [[CrossRef](#)]
23. Truong, N.D.; Nguyen, A.D.; Kuhlmann, L.; Bonyadi, M.R.; Yang, J.; Ippolito, S.; Kavehei, O. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Netw.* **2018**, *105*, 104–111. [[CrossRef](#)] [[PubMed](#)]
24. Agboola, H.A.; Solebo, C.; Aribike, D.S.; Lesi, A.E.; Susu, A.A. Seizure Prediction with Adaptive Feature Representation Learning. *J. Neurol. Neurosci.* **2019**, *10*, 1–12. [[CrossRef](#)]
25. Kaiser, J. On a simple algorithm to calculate the ‘energy’ of a signal. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, NM, USA, 3–6 April 1990; IEEE: Piscataway, NJ, USA, 1990; pp. 381–384. [[CrossRef](#)]
26. Esteller, R.; Echauz, J.; Tchong, T.; Litt, B.; Pless, B. Line length: An efficient feature for seizure onset detection. In Proceedings of the 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Istanbul, Turkey, 25–28 October 2001; IEEE: Piscataway, NJ, USA, 2001; Volume 2, pp. 1707–1710. [[CrossRef](#)]
27. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)] [[PubMed](#)]
28. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)]
29. Roberts, S.J.; Penny, W.; Rezek, I. Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing. *Med. Biol. Eng. Comput.* **1999**, *37*, 93–98. [[CrossRef](#)] [[PubMed](#)]
30. Devarajan, K.; Jyostna, E.; Jayasri, K.; Balasampath, V. EEG-Based Epilepsy Detection and Prediction. *Int. J. Eng. Technol.* **2014**, *6*, 212–216. [[CrossRef](#)]
31. Esteller, R.; Vachtsevanos, G.; Echauz, J.; Litt, B. A comparison of waveform fractal dimension algorithms. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2001**, *48*, 177–183. [[CrossRef](#)]
32. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* **1970**, *29*, 306–310. [[CrossRef](#)]
33. Bryce, R.M.; Sprague, K.B. Revisiting detrended fluctuation analysis. *Sci. Rep.* **2012**, *2*, 315. [[CrossRef](#)]
34. Mormann, F.; Andrzejak, R.G.; Elger, C.E.; Lehnertz, K. Seizure prediction: The long and winding road. *Brain* **2007**, *130*, 314–333. [[CrossRef](#)] [[PubMed](#)]
35. Inouye, T.; Shinosaki, K.; Sakamoto, H.; Toi, S.; Ukai, S.; Iyama, A.; Katsuda, Y.; Hirano, M. Quantification of EEG irregularity by use of the entropy of the power spectrum. *Electroencephalogr. Clin. Neurophysiol.* **1991**, *79*, 204–210. [[CrossRef](#)]
36. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
37. Hao, Q.; Gotman, J. A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: Possible use as a warning device. *IEEE Trans. Biomed. Eng.* **1997**, *44*, 115–122. [[CrossRef](#)] [[PubMed](#)]