1. Brahim CHEROUATI[1], 2. Mohamed SENOUCI[2]

University of Oran1 Ahmed Ben Bella, Oran-Algeria (1), University of Oran1 Ahmed Ben Bella, Oran-Algeria (2)

doi:10.15199/48.2023.07.20

Patient specific channel optimization using entropy and CNN deep learning for epileptic seizure prediction

Abstract. *Predicting epileptic seizures in advance improves greatly the life of epileptic patients. In this paper we present a new approach based on patient specific channel optimization using four different features namely entropy, variance, kurtosis and skewness. After selecting three best channels for each method, we then use Convolutional Neural Network (CNN) to classify raw EEG signal in order to discriminate between interictal and preictal state. With entropy, our method achieves a good degree of prediction in terms of accuracy 97.09%, sensitivity 97.67% and specificity 96.51% for patient 01 using channels 4, 8 and 20.*

Streszczenie. *Przewidywanie napadów padaczkowych z wyprzedzeniem znacznie poprawia życie chorych na padaczkę. W tym artykule prezentujemy nowe podejście oparte na optymalizacji kanałów specyficznych dla pacjenta przy użyciu czterech różnych metod, a mianowicie entropii, wariancji, kurtozy i skośności. Po wybraniu trzech najlepszych kanałów dla każdej z metod, wykorzystujemy Neuronową Sieć Konwolucyjną (CNN) do klasyfikacji surowego sygnału EEG w celu rozróżnienia pomiędzy stanem międzynapadowym i przednapadowym. Dzięki entropii nasza metoda osiąga dobry stopień predykcji w zakresie dokładności 97,09%, czułości 97,67% i specyficzności 96,51% dla pacjenta 01 przy użyciu kanałów 4, 8 i 20.* (***Specyficzna dla pacjenta optymalizacja kanałów z wykorzystaniem entropii i głębokiego uczenia CNN do przewidywania napadów padaczkowych***)

Keywords: seizure prediction, EEG, entropy, CNN, classification.
Słowa kluczowe: predykcja napadów, EEG, entropia, CNN, klasyfikacja.

Introduction

Epilepsy is a chronic brain disease that affects people of all ages. Approximately 50 million people worldwide have epilepsy, making it one of the most common neurological diseases in the world [1]. Seizure prediction can help epilepsy patient have a better life by preventing accidents that occur during seizures [2]. Predicting epilepsy consists in detecting the preictal state, which is the state that directly precedes the onset of the seizure.

Electroencephalogram (EEG) is the most widely used tool for the analysis and diagnosis of epilepsy. The brain activity of an epileptic patient is classified into four states, the interictal state which is the period between two seizures, the preictal state which is the period before the seizure onset, the ictal state which defines the period during the seizure onset and the postictal state which is the period after the seizure onset [2]. In the study of epileptic seizure prediction, only the interictal state and the preictal are considered [4].

To make a classification between interictal and preictal states, EEG signals are analyzed and discriminated according to the following two methods: (a) Manual extraction of EEG signal features followed by classification using machine learning algorithms such as neural networks or SVMs, (b) Automatic extraction and classification of EEG signal features using deep learning algorithms such as CNN or LSTM [5].

Deep learning algorithms have shown their performance in analysis of EEG signal. In [6], the authors used the Convolutional Neural Network(CNN) as feature extraction algorithm and Long Short Term Memory(LSTM) as classification algorithm between preictal and interictal state . In [7], the authors used CNN, LSTM and recurrent neural networks (RNN) algorithms as classification in 30 reviewed articles.

The key problem with the previously described solutions is to find the most distinctive characteristics that best represent each category using all channels of the dataset. The time required to extract these features is dependent on the process complexity and is regarded as a further difficulty added to time used to classification and prediction. In response to these problems and the need for accurate seizure prediction, we developed a deep learning-based seizure prediction algorithm that combines channel reduction and selection, feature extraction and classification into one automated system.

In this paper, we planned to create entropy based channel optimization and deep Convolutional neural network (CNN) to classify spatial features between interictal and preictal states while reducing computational time. Because computing complexity is important for real-time application, we utilize a channel reduction and selection technique to choose three best representative channels from the 23 multi-channel EEG recording. CNN is used to learn the discriminative spatial features between interictal and preictal states. The testing method employed demonstrates the accuracy of the suggested algorithm above various seizures.

Methodology

For seizure prediction, we present a channel reduction and selection strategy as well as CNN deep learning models. The seizure prediction problem is given as a classification problem between interictal and preictal state. In many research studies, there is no standard duration for the preictal state. In our experiments, the preictal length was chosen to be 15 minutes before the seizure onset as in [8] and interictal length was chosen to be at least four hours before seizure start or after seizure end as in [2,3].

EEG data is used as the model input without any preparation or manual feature extraction. CNN deep learning is used to automatically learn and classify features in order to minimize costs and improve classification process. There is also an inequality between preictal and interictal samples due to each patient's limited number of seizures. The number of interictal examples is higher than the number of preictal samples, and classifiers tend to be more efficient at classifying with the largest number of training instances, therefore interictal and preictal samples are taken equally, and the data is balanced. Figure 1 shows an example of 15 seconds of interictal, preictal and ictal state of patient 01.

In our model, raw EEG signals are divided into 30 sec windows with an overlap of 8 sec, and continuous wavelet transform(CWT) is applied to each segment, which is a temporal frequency representation of signal. CWT is used to convert raw segmented signal to scalogram images which are then fed into 2D CNN models. Interictal and preictal discrimination was performed using the extracted features. Fig. 2 shows the flow chart of our study.

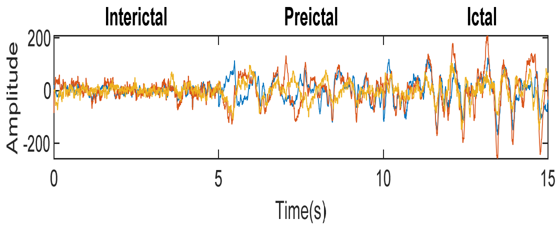

Fig.1 An example showing patient01 within his interictal, preictal and ictal state from channels P7-O1, P3-O1 and T7-FP9 for 15 sec, 5 sec for each state.

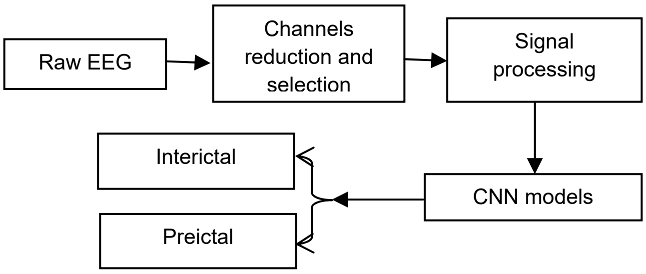

Fig.2. Bloc diagram of seizure prediction method

Dataset

We used the CHBMIT scalp EEG dataset, which may be found at physionet.org [9]. This dataset, generated by Children's Hospital Boston, contains 664 files, 129 of those files contain one or more seizures and the remainder files contain no seizure activity. EEG is recorded from pediatric patients suffering from intractable seizures. Subjects were observed for many days after discontinuing anti-seizure medication to define their seizures and determine their candidacy for surgical surgery. The scalp EEG recordings of 22 epilepsy patients are included in this collection. EEG signals were captured for one hour using the conventional electrode placement of 10–20 system; these signals were captured using 23 different channels. Each file containing data in which a seizure occurred has been annotated for the start and finish times of the seizure. EEG signals were captured at 256 Hz, with 23 electrodes utilized in most sessions [7].

As explained in [3] many aspects, such as the interictal period, preictal period, number of channels, and recording continuity, differ between subjects in this dataset. As a result, we select eight subjects for this study such the calculated interictal and preictal durations are satisfied, and that the full channels' recordings are available. The details of the EEG recordings used in our investigations are summarized in Table 1.

Table 1. Description of the selected subjects

Subject	Gender	Age	# of seizure
chb01	F	11	7
chb02	M	11	3
chb03	F	14	7
chb04	M	22	4
chb05	F	7	5
chb06	F	1.5	10
chb07	F	14.5	3
chb08	M	3.5	5

We define the length of the Preictal signal to be 15 min, starting 20 min before the onset of the seizure thus leaving a period of 5 min immediately preceding the seizure as shown in Fig. 3. This period is necessary to give the patient sufficient time to be treated quickly.

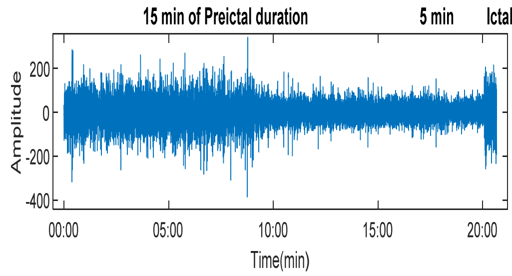

Fig.3. An example showing 15 min of preictal of patient01 followed by 5 min which immediately precede the 40 sec of the ictal state from channel p3-o1.

EEG channel reduction and selection

EEG channel reduction and selection are an important step in seizure prediction to reduce the computational complexity and thus reducing the time seizure prediction. The acquired EEG signals are typically from multi-channel recordings; however, the number of channels must be reduced because a large number of channels cost time and causes discomfort to the subject, on the other hand, channel selection refers to selecting the best channels that provide better information about the subject seizure.

In this study, four statistical methods were used for channels reduction and selection, including, entropy, variance, kurtosis and skewness, those methods were applied to the ictal segments of each patient. Only the ictal segment of the first file containing the signal in which a seizure occurred has been used. These methods are applied on all 23 channels and then only 3 channels that correspond to highest entropy, variance, kurtosis and skewness were selected. The algorithm 1 shows the steps to channel reduction and selection.

Algorithm 1. Algorithm for channel reduction and selection

```
For Patient = 1:8
    Take first file which contain seizure onset session
    Extract segment of seizure start and seizure end period
    For channel=1:23 of extracted segment
            Apply entropy
            Apply variance
            Apply kurtosis
            Apply skewness
    End
    Sort (entropy, descendant)
    Sort (variance, descendant)
    Sort (kurtosis, descendant)
    Sort (skewness, descendant)
    Take three channels of each sorted method
End
```

Entropy

The entropy H is a function that is used to better understand the uncertainty and the randomness from a probability distribution for a random variable X[10], it is defined by the equation eq. (1) [11].

$$H(c) = \sum_{i=1}^{n} P(x_c(i)) \log P(x_c(i)) \tag{1}$$

Where, $P(x_c(i))$ is the probability distribution of value i of channel c and n represents the size of c.

Variance

The variance V is a function that measures the variability of the data, i.e. the distance of each variable from the mean value of its data set. It is calculated by taking the average of the squared deviations from the mean, this variance was applied to each channel in the EEG ictal signal by the equation eq. (2) [11].

$$V(c) = \frac{1}{n}\sum_{i=1}^{n}(x_c(i) - \mu_c)^2 \tag{2}$$

Where, $x_c(i)$ represents the data value of index i of the channel c and μ_c the mean values of the same channel with size n.

Kurtosis

The kurtosis K is a function that is used to provide a measure of the spikiness of signals [12], It is represented by the equation eq.(3)

$$K(c) = \frac{\frac{1}{n}\sum_{i=1}^{n} x_c(i)^2}{\frac{1}{n}\sum_{i=1}^{n} x_c(i)^4} - 3 \tag{3}$$

Here, $x_c(i)$ represents the data value of index i of the channel c with size n.

Skewness

The skewness S is a function that is used to calculate the asymmetric variability of signal from its mean value [13], it is represented by the equation eq(4)

$$S(c) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_c(i) - \mu_c\right)^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_c(i) - \mu_c)^2}\right)^3} \tag{4}$$

Where, $x_c(i)$ represents the data value of index i of the channel c and μ_c the mean values of the same channel with size n.

EEG signal processing

In this section, 15 min of preictal signals were extracted from each file containing seizure onset as shown in figure 2, this step is repeated for the 8 patients used in this study. The same calculation operations are applied on the interictal part contained in the no epileptic seizure files by respecting the same length as that used for the preictal state, i.e. 15 min.

A moving window of 30 sec with an 8 sec overlap was applied to the segments extracted from the interictal and preictal state, the resulting segments were then transformed into frequency-temporal domain using the continuous wavelet transform (cwt). Fig.4 displays 30 seconds of interictal and preictal data from patient01.

a) b)

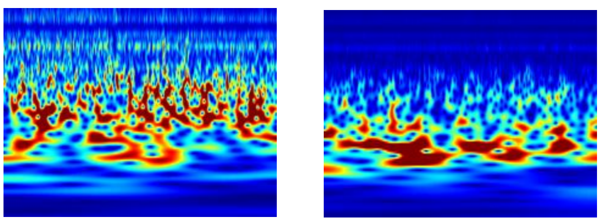

Fig.4. 30 sec of interictal cwt image (a), and 30 sec of preictal cwt image (b)

CNN for features extraction and classification

In recent years, deep learning has become a field that interests many researchers. Convolutional neural networks are deep learning techniques frequently used to solve difficult problems [8]. CNN is used for feature extraction and classification, CNN is made up of three layers: a convolutional layer, a pooling layer, and a fully connected layer. The two first layers are used to extract features, while the final layer is used for classification as shown in Fig.5

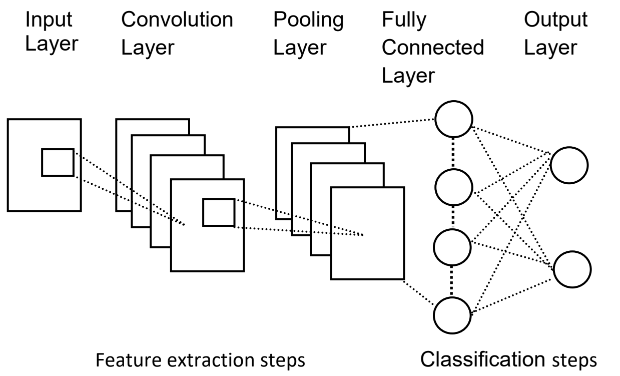

Fig.5. CNN architecture

In this work, three convolutional layers are used with 8 kernels for the first one, 16 for the second, and 32 for the third one, all of which had 3x3 kernel sizes with stride 1. Two maximum pooling layers are also used after each convolutional layer to reduce the feature map; the pool size for the maximum pooling layer is 2x2 with a stride 2. All convolutional layers use the ReLU (Rectified Linear Unit) activation function with batch normalization, table 2 summarizes the structure of our proposed CNN.

Table 2. Structure of proposed CNN

Layer	Feature map	Configuration	Total parameters
Convolution	224x224x8	3x3, stride 1, padding 1	224
Batch normalization	224x224x8	-	16
ReLU	224x224x8	-	0
Max pooling	112x112x8	2x2, stride	0
Convolution	112x112x16	3x3, stride, padding 1	1168
Batch normalization	112x112x16	-	32
ReLU	112x112x16	-	0
Max pooling	56x56x16	2x2, stride 2	0
Convolution	56x56x32	3x3, stride 1, padding 1	4640
Batch normalization	56x56x32	-	64
ReLU	56x56x32	-	0
Fully connected	1x1x2	-	200706
Softmax	1x1x2	-	0

Convolution layer

The convolution layer is the most important layer of the CNN architecture whose main purpose is feature extraction using different kernels and a combination of linear and non-linear operations namely convolution operations and activation function [2,14].

The convolution operation consists in moving the kernel on the input data to have in output a features map which will be the input for the next layer by using the following equation (5) as used in [2], Where I represent the two dimensional image used as input to the CNN and K

represent the two-dimensional kernel, Y is the out of the convolution operations.

$$Y(i,j)=\sum_m\sum_n I(m,n)K(i-m,j-n) \quad (5)$$

After each convolutional operation, a batch normalization function is applied to speed up the training process of the CNN using a mini batch normalization of each dimension of features map, it is used before the non-linear operation such as ReLU activation function, the following equation (6) define the use of batch normalization transform [3].

$$BN_{\gamma,\beta}(x_i)=\gamma\frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}}+\beta \quad (6)$$

Where $B=\{x_1,x_2,x_3,\ldots,x_m\}$ is a mini batch of size m that contains a set of input values of vector x_i to be normalized, μ_B is a mini batch mean and σ_B^2 is mini batch variance of the vector x_i. ε is a constant added to the mini-batch variance for numerical stability, while γ and β are parameters to be learned.

The ReLU activation function is used in this study, which is the most used activation function among the nonlinear functions used in CNN [8], it is represented by the equation (7).

$$y=\max(x,0) \quad (7)$$

Where x is the input data and y is output data of the ReLU activation function.

Pooling layer

The max pooling operation is applied on the features map after each convolutional layer. This operation allows us to reduce the size of this features map and subsequently reduce the computational time [2]. This reduction is done by calculating the maximum value for each region using a square window of dimension 2x2 with a stride of 2 as used in our study.

Fully connected layer

The convolutional layer and the pooling layer are used for features extraction while the fully connected layer and output layer are used for classification between interictal and preictal state as shown in figure 4, Softmax activation function which is represented by equation (8) have been used in output layer in this study.

$$soft\,max(x_i)=\frac{e^{x_i}}{\sum_{j=1}^{n}e^{x_j}} \quad (8)$$

Where x_i is input data of i^{th} matched class which can either be interictal or preictal and n represent the number of classes which is equal to 2 in our case

Results and discussions

In this section, three statistical measures were used to evaluate our model, namely accuracy, sensitivity, and specificity. The accuracy, sensitivity, and specificity are calculated using the equations (9), (10) and (11) respectively.

(9) *Accuracy=((TP+TN)/(TP+FN+FP+FN))*100*
(10) *Sensitivity=(TP/(TP+FN))*100*
(11) *Specificity=(TN/(TN+FP))*100*

Where *TP*, *TN*, *FP* and *FN* stands for true positive, true negative, false positive and false negative respectively.

The aim of this experiment was to classify EEG signals into two classes, preictal and interictal, using four optimization methods. Our approach consists in taking only three most important channels instead of 23, which allows reducing considerably the computation time, and the obtained results are satisfactory.

The purpose of this experiment was to discriminate between preictal and interictal EEG signals using the four statistical functions used in channel reduction and selection on the one hand and using CNN deep learning for features extraction and classification on the other. Our approach was to make a patient specific channels selection for the eight patients treated in this study.

Thus, for each patient three most important channels were identified and used.

With this approach, the evaluation of our model starts only after selecting the three most important channels. This leads to a considerable computational time saving.

In Table4, the accuracy, sensitivity and specificity based channel selection for patient 1 was presented. The variance and specificity functions show maximum accuracy of 97.48% and specificity of 98.06% using channels 5, 14 and 21. The kurtosis function gives minimum accuracy which was 92.44% using channels 1, 16 and 22. Our proposed method showed highest sensitivity given by entropy which was 97.67% using channels 4, 8, 20.

Table 04
Accuracy, sensitivity and specificity using entropy, variance, kurtosis and skewness for patient 01.

Method	#Channel	Accuracy	Sensitivity	Specificity
Entropy	4, 8, 20	97.09	97.67	96.51
Variance	5, 14, 21	97.48	96.90	98.06
Kurtosis	1, 16, 22	92.44	92.25	92.64
Skewness	5, 16, 22	93.22	93.41	93.02

For the eight patients used in this study, tables 5, 6, 7 and 8 showed that the entropy gives an average accuracy of 93.95% which represents a higher average than those given by the variance, kurtosis and skewness which are 87.53%, 89.71% and 90.78% respectively. Our proposed method gives higher sensitivity which was 96.20% using entropy.

Table5
Accuracy, sensitivity and specificity using entropy

#Patient	#Channel	Accuracy	Sensitivity	Specificity
1	4, 8, 20	97.09	97.67	96.51
2	14, 22, 10	97.75	100	95.50
3	7, 8, 18	90.88	89.86	91.89
4	4, 8, 20	86.49	97.30	75.68
5	3, 20, 22	100	100	100
6	9, 13, 21	81.79	86.41	77.17
7	16, 20, 22	98.65	100	97.30
8	3, 19, 20	98.91	98.37	99.46
	Average	93.95	96.20	91.69

Table6
Accuracy, sensitivity and specificity using variance.

#Patient	#Channel	Accuracy	Sensitivity	Specificity
1	5,14,21	97.48	96.90	98.06
2	7,8,11	95.50	95.50	95.50
3	5 ,2, 1	85.81	83.11	88.51
4	10,14,21	88.29	82.88	85.59
5	2,6,17	91.58	85.87	97.28
6	4,6,17	81.79	78.80	84.78
7	10,11,17	100	100	100
8	9,12,21	59.78	69.02	50.54
	Average	87.53	86.51	87.53

Table7
Accuracy, sensitivity and specificity for the eight patients using kurtosis.

#Patient	#Channel	Accuracy	Sensitivity	Specificity
1	1,16,22	92.44	92.25	92.64
2	8,13,14	85.59	72.07	99.10
3	10,11,13	94.26	94.59	93.92
4	3,19,20	93.24	88.29	98.20
5	1,13,14	82.88	92.39	73.37
6	5,8,9	88.04	88.59	87.50
7	3,4,22	100	100	100
8	1,11,13	81.25	90.22	72.28
Average		89.71	89.80	89.63

Table8
Accuracy, sensitivity and specificity using skewness.

#Patient	#Channel	Accuracy	Sensitivity	Specificity
1	5,16,22	93.22	93.41	93.02
2	14,16,19	94.14	92.79	95.50
3	1,14,22	92.23	87.16	97.30
4	1,8,19	96.40	94.59	98.20
5	6,14,17	86.14	92.39	79.89
6	8,10,12	84.24	90.22	78.26
7	10,11,13	100	100	100
8	1,9,13	79.89	64.13	95.65
Average		90.78	89.34	92.23

Table 9 presents the different channels found for each method and for each patient. We can easily deduce the channels that are used in common for all patients for each method. For entropy, we can see that channel 8 is used by patient 1, patient 3 and patient 4 as well as channel 20 which is used by the three patients 1, 4, 5 and 7 respectively. We can therefore use these channels for all patients by considerably reducing the channels to only 3 channels instead of 23.

Table 9
The patient channels selected based on the used methods.

#Patient	Entropy	Variance	Kurtosis	Skewness
1	4, 8, 20	5, 14 ,21	1, 16, 22	5, 16, 22
2	14, 22,10	7, 8, 11	8, 13, 14	14, 16, 19
3	7, 8,18	5, 2, 1	10, 11, 13	1, 14, 22
4	4, 8, 20	10, 14, 21	3, 19, 20	1, 8, 19
5	3, 20, 22	2, 6, 17	1, 13, 14	6, 14, 17
6	9, 13, 21	4, 6, 17	5, 8, 9	8, 10, 12
7	16, 20, 22	10, 11, 17	3, 4, 22	10, 11, 13
8	3, 19, 20	9, 12, 21	1, 11, 13	1, 9, 13

Conclusion

In this paper, a seizure prediction method was proposed, in which, four linear and non-linear parameters were used for channel optimization. 3 channels out of 23 channels were used. The EEG signals were divided into two classes namely interictal and preictal. After segmentation of the signals, they were transformed into a scalogram image using continuous wavelet transform (CWT). The obtained images were then run through the CNN classifier. The proposed method was evaluated on the CHB-MIT EEG database, and entropy gave better results for accuracy, sensitivity and specificity which are respectively 93.95%, 96.20% and 92.69%. With our method, the computational time was significantly reduced as well as the number of channels used, which facilitates the implementation of this approach in real time applications for the seizure prediction problem.

Authors:
Brahim Cherouati, Department of Computer Sciences, University of Oran1 Ahmed Ben Bella, E-mail: bcherouati@gmail.com, Mohamed Senouci, Department of Computer Sciences, University of Oran1 Ahmed Ben Bella, E-mail: msenouci@yahoo.fr

REFERENCES

[1] World Health Organisation, Accessed, Nov 19, 2022 Available, https://www.who.int/news-room/fact-sheets/detail/epilepsy
[2] R. Jana and I. Mukhetjee, Deep learning based efficient epileptic seizure prediction with EEG channel optimization, Biomed. Signal Process. Control, 68(2021), 102767.
[3] H.Daoud and M. A. Bayoumi, Efficient Epileptic Seizure Prediction Based on Deep Learning, IEEE Trans. Biomed. Circuits Sys., 13(2019) ,No. 5, 804-813.
[4] J. Liang, R. Lu, C. Zhang, and F. Wang, Predicting Seizures from Electroencephalography Recordings: A Knowledge Transfer Strategy, IEEE International Conference on Healthcare Informatics (ICHI), Chicago, USA(2016), 184-191.
[5] S. Toraman, Preictal and Interictal Recognition for Epileptic Seizure Prediction Using Pre-trained 2D-CNN Models, Trait. Signal, 37(2020), No. 6, 1045–1054.
[6] S. M. Usman, S. Khalid, and Z. Bashir, Epileptic seizure prediction using scalp electroencephalogram signals, Biocybern. Biomed. Eng., 41(2021), No. 1, 211–220.
[7] S. Paszkiel, Review of solutions for the application of example of machine learning methods for Motor Imagery in correlation with Brain Computer Interfaces, PRZEGLĄD ELEKTROTECHNICZNY, 1(2021), No. 11, 113–118.
[8] S. Ryu and I. Joe, A Hybrid DenseNet-LSTM Model for Epileptic Seizure Prediction, Appl. Sci., 11(2021), No 16, 7661.
[9] CHB-MIT Scalp EEG Database, Accessed, Jun. 09, 2022. Available, https://physionet.org/content/chbmit/1.0.0/
[10] Claude E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal, 27(1948), 379–423.
[11] J. Duun-Henriksen, T. W. Kjaer, R. E. Madsen, L. S. Remvig, C. E. Thomsen, and H. B. D. Sorensen, Channel selection for automatic seizure detection, Clin. Neurophysiol., 123(2012), No. 1, 84–92.
[12] A. Shoeb and J. Guttag, Application of Machine Learning To Epileptic Seizure Detection, ICML'10, Proceedings of the 27th International Conference on International Conference on Machine Learning, 2010, 975–982.
[13] G. Prendergast, G. G. R. Green, and M. Hymers, A robust implementation of a kurtosis beamformer for the accurate identification of epileptogenic foci, Clin. Neurophysiol., 124(2013), No. 4, 658–666.
[14] R. Yamashita, M. Nishio, R. K. G. Do, & K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights Imaging, 9(2018), No. 4, 611-629.