Biomedical Signal Processing and Control 68 (2021) 102720

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Classification and analysis of epileptic EEG recordings using convolutional neural network and class activation mapping

Abdulnasir Yildiz [a], Hasan Zan [b,*], Sherif Said [c]

[a] Dicle University, Department of Electric and Electronics, Turkey
[b] Mardin Artuklu University, Vocational School, Turkey
[c] College of Engineering and Technology, American University of the Middle East, Kuwait

ARTICLE INFO

Keywords:
Electroencephalogram
Epileptic EEG signal classification
Epilepsy
Seizure detection
Convolutional neural networks
Spectrogram
Scalogram
Hilbert-Huang transform
Class activation mapping
CAM

ABSTRACT

Electrical bio-signals have the potential to be used in different applications due to their hidden nature and their ability to facilitate liveness detection. This paper investigates the feasibility of using the Convolutional Neural Network (CNN) to classify and analyze electroencephalogram (EEG) data with their time-frequency representations and class activation mapping (CAM) to detect epilepsy disease. Several types of pre-trained CNNs are employed for a multi-class classification task (AlexNet, GoogLeNet, ResNet-18, and ResNet-50) and their results are compared. Also, a novel convolutional neural network architecture comprised of two horizontally concatenated GoogLeNets is proposed with two inputs scalograms and spectrogram of the eplictic EEG signal. Four segment lengths (4097, 2048, 1024, and 512 sampling points) with three time-frequency representations (short-time Fourier, Wavelet, and Hilbert-Huang transform) are statistically evaluated. The dataset used in this research is collected at the University of Bonn. The dataset is reorganized as normal, interictal, and ictal. The maximum achieved accuracies for 4097, 2048, 1024, and 512 sampling points are 100 %, 100 %, 100 %, and 99.5 % respectively. The CAM method is used to analyze discriminative regions of time-frequency representations of EEG segments and networks' decisions. This method showed CNN models used different time and frequency regions of input images for each class with correct and incorrect predictions.

1. Introduction

Epilepsy is a chronic non-infectious disease of the brain [1]. Any person can develop epilepsy, and this disease affects both genders regardless of their race, ethnic background, and age [2]. Around 50 million people living in low- and middle-income countries suffer from epilepsy, which is higher than the number of people affected by any other neurological diseases [1]. This disease causes repetitive seizures that lead to unintentional body parts or whole-body movement and sometimes, loss of consciousness and cognitive functions [3]. The extreme electrical discharge in a group of brain cells causes epileptic seizures [4]. Such discharges can be originated in a distinct region of the brain [1]. The severity of seizures can vary from a short loss of attention to long-lasting muscle contraction [5]. The frequency of seizures can also alter from one seizure in a year to many seizures in a day, making seizure prediction very difficult [5]. Because of these variations, seizures at certain times can cause dangerous situations such as falling, drowning, and car accidents for people with epilepsy.

Electroencephalogram (EEG) signal is an effective clinical apparatus for determining brain diseases such as epilepsy, Alzheimer's, Parkinson's disease, and sleep disorders [6]. Therefore, analysis and classification of EEG signals can help physicians to diagnose epilepsy. Generally, specialists classify EEG segments into three classes: ictal, interictal, and normal for analysis [6]. Ictal EEG is recorded during seizures and has characteristics of the hyper-synchronous electrical activity of the brain. EEG segment, which is recorded when the epileptic subject does not have a seizure is called interictal EEG. This EEG segment is of small spikes. Finally, normal EEG is recorded from a healthy subject. Fig. 1 depicts EEG segments of these classes.

The application of deep learning especially Convolutional Neural Network (CNN) has gained popularity among researchers due to its state-of-art performance and its extensive applications in many research areas such as computer vision, image processing, biomedical image and signal classification, and authentication system [7]. CNN is a type of neural network that automatically extracts spatial features from its inputs by convolution, activation, and pooling layers. It does not require

* Corresponding author.
E-mail address: hasanzan@artuklu.edu.tr (H. Zan).

https://doi.org/10.1016/j.bspc.2021.102720
Received 28 January 2021; Received in revised form 14 April 2021; Accepted 1 May 2021
Available online 19 May 2021

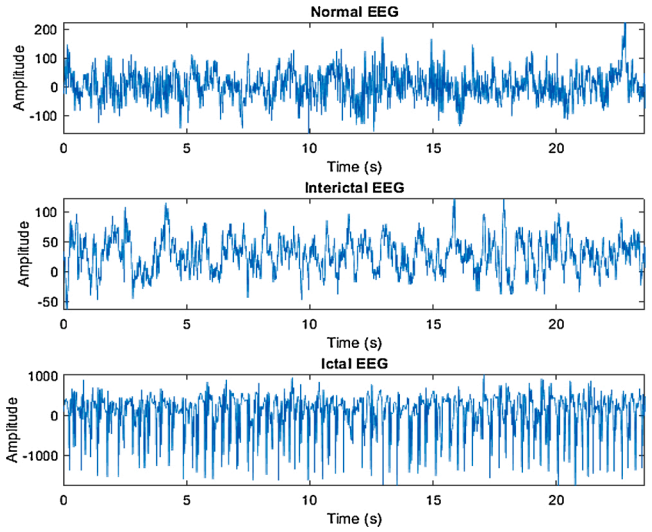

Fig. 1. Normal, interictal, and ictal EEG segments (amplitudes in μV).

hand-crafted features and its applications demand much less domain knowledge compared to other classification algorithms. Therefore, this paper materializes two main goals by employing CNN and class activation mapping (CAM):

- Classification of EEG segments into three classes defined above using a convolutional neural network, and improve the accuracy achieved by other studies using CNN.
- Analyzing important regions of time-frequency representations of EEG segments and interpret predictions using class activation mapping (CAM).

Recent developments in computer hardware systems and introducing new algorithms have enabled the use of deep learning [7]. Recently, deep learning algorithms, especially, CNN have been intensely exploited due to state-of-art performance in object recognition, object detection, computer vision, automatic speech recognition (ASR) applications [8,9].

Furthermore, CNN has been employed in biomedical applications [10]. Li et al. [11] proposed an end-to-end CNN classifier comprised of 9 layers for binary classification of normal and ictal EEG segments. In this study, they reported an accuracy of 100 % using 2600 EEG segments. In another study, Acharya et al. [12] introduced an end-to-end CNN classifier comprised of 13 layers for the multiclass classification of EEG signals. In this study, they used 300 signals consisting of normal, interictal, and ictal EEG classes taken from the University of Bonn EEG dataset [13]. Researchers reported an overall accuracy of 88.67 %. Yildirim et al. [14] proposed a deeper 1D CNN model with 23 layers for EEG signal classification. The data used in this study are a collection of 2993 EEG signals containing both normal and abnormal classes. These signals were recorded from 2383 subjects, including males and females. The researcher reported an overall accuracy of 79.34 %. Thomas et al. [15] proposed a system that involves three modules: pre-preparing, EEG-level grouping, and waveform-level classification. They utilized a CNN system for waveform-level grouping and a support vector machine (SVM) for EEG-level characterization. Researchers assessed the proposed framework on a dataset of 156 EEGs recorded at Massachusetts General Hospital (MGH), Boston. The framework accomplished a mean 4-fold classification precision of 83.86 % for grouping EEGs. Hu et al. [16] proposed a CNN model and SVM system instead of an end-to-end CNN model. They employed CNN for feature extraction from the amplitude spectrum of EEG signals and SVM to classify EEG signals using extracted features. They reported a classification accuracy of 86.25 %. San-Segundo et al. [5] proposed the CNN model, which consists of 12 layers for the classification of epileptic EEG recordings. They evaluated several EEG signal transforms (raw data, Fourier transform, wavelet transform, and empirical mode decomposition) for generating the inputs to the CNN. The analysis was performed using Bern-Barcelona EEG and Epileptic Seizure Recognition datasets, obtaining noticeable improvements in the accuracy.

The current state-of-art method was proposed by Ullah et al. [17]. They used the University of Bonn EEG dataset. This dataset has five classes of EEG signals (A, B, C, D, E). A and B consist of signals recorded from healthy subjects with eyes open and closed, respectively. C, D, and E have EEG signals recorded from epileptic subjects. Signals in C and D were collected during seizure-free periods while E was created during seizures. They proposed a memory-efficient CNN model that has few layers and no pooling layer. Inputs to the models were generated by a fixed-size sliding window that operates over raw EEG signals. To increase the number of samples, overlapping EEG segments were employed during the training of the models. However, different window sizes were not investigated. Ensemble of the models and a majority vote was used for final classification. Three classification problems were investigated: AB vs CD (normal vs interictal), AB vs CDE (normal vs epileptic), and AB vs CD vs. E (normal vs interictal vs ictal). Best cross-validated accuracies for AB vs. CD, AB vs CDE, AB vs CD vs. E are obtained as 99.8 %, 99.95 %, and 99.1, respectively. Even though they achieved a high accuracy for multi-class classification, it still lower than accuracies of binary classifications. Therefore, we think the accuracy can be improved more by utilizing combination of time-frequency transformations. Furthermore, they did not discuss using different CNN architectures nor how CNN models make decisions.

A class activation map (CAM) for a certain category indicates the CNN's discriminative image regions to recognize that category [18]. Zhou et al. [18] introduced class activation mapping to visualize discriminative image regions. They concluded that CNN models with global average pooling and trained for classification could be used to localize important image subregions. Jia et al. [19] used CAM to improve a CNN model's classification performance for skin lesion classification. They localized important image regions using CAM and extracted these regions. Then, they retrained the CNN with the regions and improved the classification performance of the CNN. Cai et al. [20] applied an advanced convolutional neural network to achieve accurate muscular dystrophy image classification and then performed CAM to highlight the important magnetic resonance imaging (MRI) image textures. Liu et al. [21] proposed to adopt CNN to classify ultrasound images and predict tumor malignancy. After network training was performed in this study, a radiologist visually inspected CAM of the last convolutional layer of the trained network to assess the output. They reported that the proposed classifier could not only give a reasonable performance in predicting breast cancer but also present potential lesion regions.

This research achieves the goals by using several CNN structures and analysis of several time-frequency representations (TFR) to create inputs to CNN models. The main contributions are:

- Analysis of four well-known CNN architectures (AlexNet, GoogLeNet, ResNet-18, and ResNet-50).
- Proposal and analysis of a novel two-inputs CNN architecture consisting of the horizontal concatenation of two GoogLeNet models.
- Evaluation of several TFRs as inputs to CNN models: short-time Fourier transform (STFT), wavelet transform (WT), and Hilbert-Huang transform (HHT).
- Evaluation of four segment lengths: 4097, 2048, 1024, and 512 sampling points.
- Usage of CAM for identification of discriminative regions of TFRs and interpretation of prediction of the CNN. This is the first study analyzing TFRs using CAM.
- Segment length, TFRs, and CNN architectures were evaluated in the above-mentioned classification problem. EEG dataset of the University of Bonn has been used in this study [13], achieving

Table 1
Summary of the EEG dataset with respect to segment length (N: segment length).

Set	Subset	Number of Samples			
		$N = 4097$	$N = 2048$	$N = 1024$	$N = 512$
Training	Normal	100	200	400	800
	Interictal	100	200	400	800
	Ictal	50	100	200	400
Test	Normal	100	200	400	800
	Interictal	100	200	400	800
	Ictal	50	100	200	400
Total		500	1000	2000	4000

noteworthy improvement in differentiating normal, interictal, and ictal EEG segments. CAM revealed time-frequency regions that are most benefited from by CNN models when predicting each class.

This paper proposes employment of different and novel CNN architectures for the classification of epileptic EEG segments, evaluate several TFRs to create inputs to the networks, analyze different EEG segment lengths, and interpret essential regions of TFRs of EEG segments and classification results using CAM.

The paper is organized as follows. Section 2 details the dataset used in this work and data segmentation. Section 3 describes the methods used in this study, including the definition of TFRs and CNN. Section 4 outlines the study and its results and compares them with previous works. Lastly, section 5 reviews the conclusions of this study.

2. Dataset

To evaluate our proposed methods' performance, this study used a benchmark EEG dataset, which is the University of Bonn EEG dataset [13]. The complete dataset contains 500 single-channel EEG signals. All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference and 10–20 system of electrode placement. After 12 bit analog-to-digital conversion, the data were written continuously onto the disk of a data acquisition computer system. The EEG was filtered using a bandpass filter. Band-pass filter settings were 0.5340 Hz ~12 dB/octave. The dataset includes five subsets (A, B, C, D, and E) and, each subset contains 100 EEG segments of 23.6 s. These segments were extracted from continuous single-channel EEG signals after visual inspection for artifacts. A and B's EEG segments were recorded from five able-bodied volunteers with no health issues with eyes open and closed, respectively. Subsets C, D, and E were acquired from volunteers diagnosed with epilepsy. EEG segments of C and D contain activity during seizure-free intervals, and EEG segments of E contain seizure activity, selected from all recordings with ictal activity. All segments have a sampling frequency of 173.61 Hz and, thus, have 4097 sampling points. More details can be found in [13].

This study followed the same protocol used in [12,22–25] such that datasets were grouped into three classes: (i) Normal (A and B); (ii) Interictal (C and D); and (iii) Ictal (E). Each EEG signal was segmented to (i) increase the number of samples for training CNN models and (ii) investigate the effects of segment length on the performance of the proposed method. Segment lengths of 4097 (original length), 2048, 1024, and 512 were chosen for evaluation. Due to hardware limitation and long training time, training, validation, and a testing scheme was used instead of cross-validation. Each subset was equally and randomly divided into training and test sets. 10 % of the training set was used for validation during training. Table 1 lists a summary of the dataset with respect to segment length.

3. Methodology

This section introduces an overview of the methodology. It describes TFRs evaluated in this work, convolution neural network architectures used for epileptic EEG classification, and class activation mapping employed to localize discriminative regions of TFRs of EEG segments and interpretation of CNN prediction.

3.1. Overview

EEG segments used in this study were described in Section 3. EEG segments with different lengths were independently treated for analysis. Firstly, EEG segments are converted to images using three TFRs given in the introduction section. Secondly, these images are used in two ways:

- First, images of each TFR were separately fed to pre-trained CNN models with one input (AlexNet, GoogLeNet, ResNet-18, and ResNet-50). This strategy is depicted in Fig. 2.

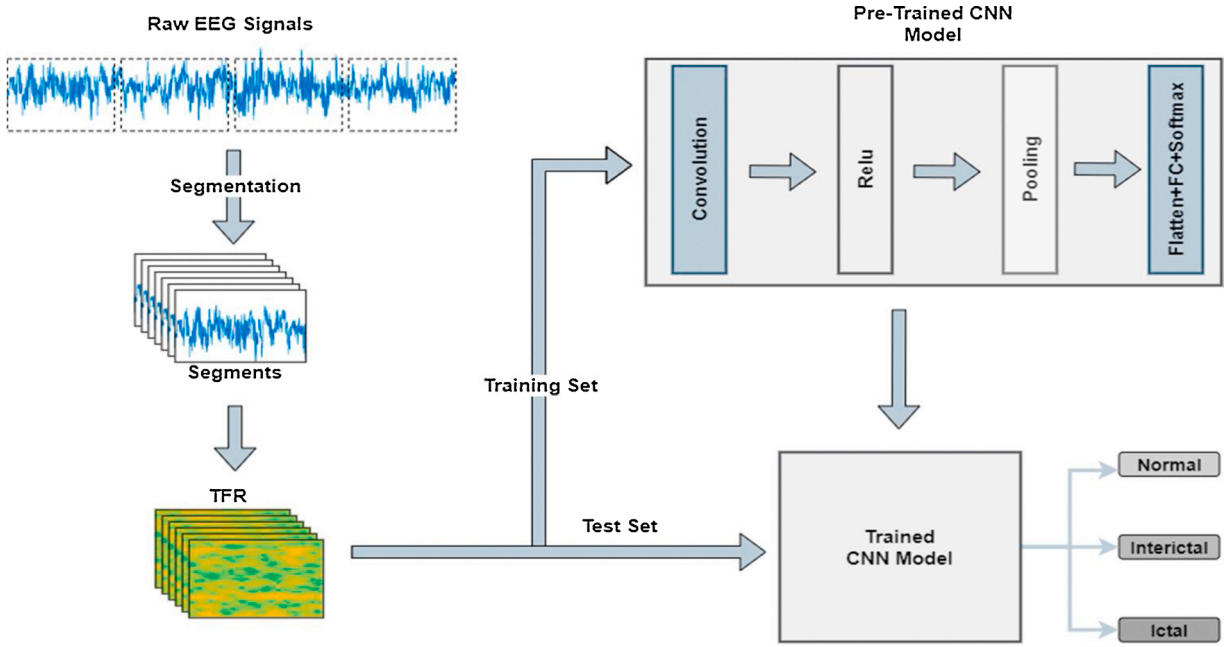

Fig. 2. Overview of strategy 1: CNN with one input.

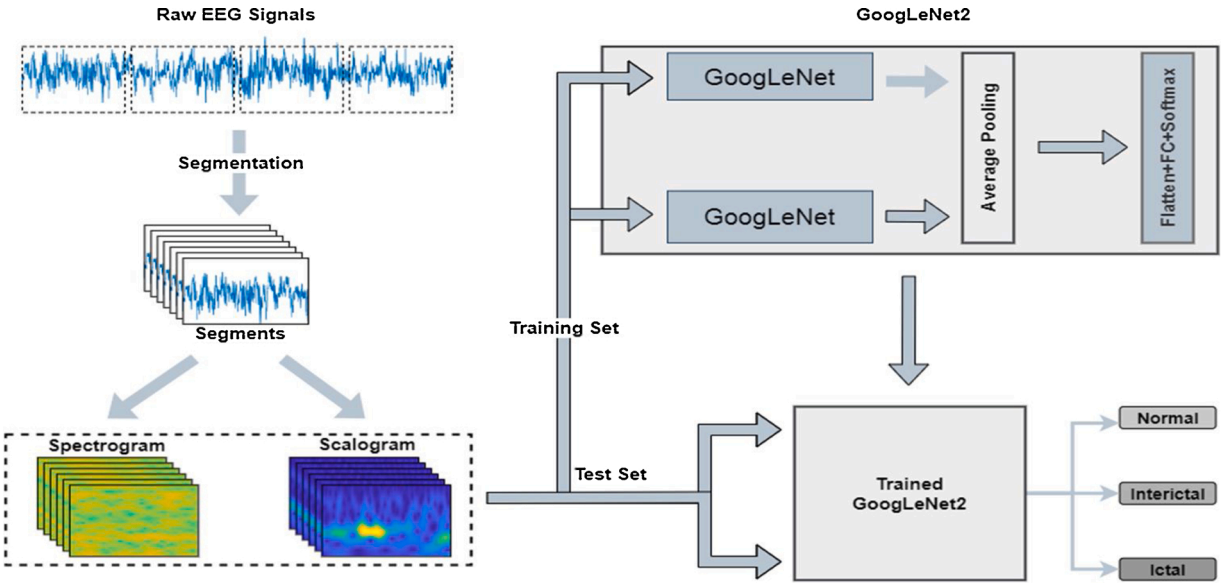

Fig. 3. Overview of strategy 2: CNN with two inputs.

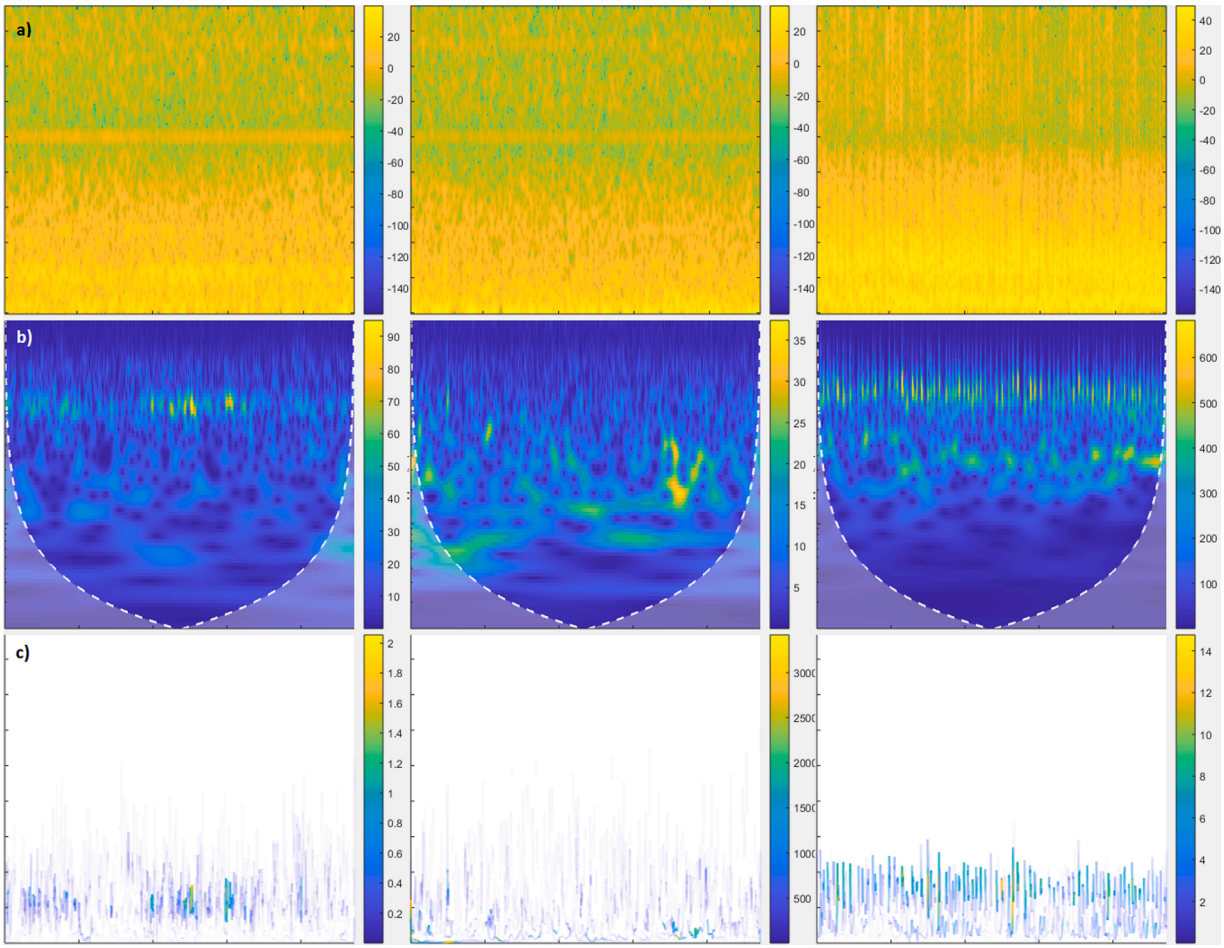

Fig. 4. TFRs of EEG segments in original resolution without labels and titles (x-axis is time in seconds, the y-axis is the frequency in Hz) a) spectrogram, b) scalogram, c) HHS. The first column is a normal segment, the second is interictal, and the third is ictal.

- Second, STFT and WT images were fed to a pre-trained CNN model with two inputs (a model consisting of two horizontally concatenated GoogLeNets) together. This strategy is depicted in Fig. 3.

Finally, data were categorized into normal, interictal, and ictal classes using CNN models.

3.2. Time-frequency representations

The frequency domain of a signal is mostly calculated by the well-known Fourier transform. However, its firm presumption on stationarity is usually breached by biological signals. To deal with such signals, Fourier transform is applied to a small portion of the signals that hold stationarity. This procedure is known as STFT. Although STFT weakens stationary requirement, it is hard to determine optimal window length. To overcome window length or temporal and spectral resolution trade-off issues, wavelet transform (WT) is developed. Thanks to varying window lengths, WT provides a better trade-off in temporal and spectral resolution. Despite both methods ease stationarity conditions, they are linear techniques unlike Hilbert-Huang transform which is suitable for both non-linear and non-stationary signals. Therefore, three mentioned transformations are employed to generate the inputs to CNN models. Fig. 4 shows TFRs of EEG segments of normal, interictal, and ictal classes.

The first approach is to use STFT transform to generate the inputs as a spectrogram. STFT of a discrete-time signal is achieved by Fourier transform of windowed signal and calculated as

$$X(m,k) = \sum_{n=-\infty}^{\infty} x(n)w(n-m)e^{-j2\pi kn} \tag{1}$$

where n and k are time and frequency indices, $x(n)$ is the input signal, w is the window function, and m is the window interval centered around zero. Visualization of log-spectra of STFT as heat-map yields spectrogram. In this case, the inputs to CNN models are spectrogram saved as an image.

Secondly, the wavelet transform (WT) is considered to create the inputs as a scalogram. Wavelet transform of a discrete-time signal is given as

$$c(j,k) = \sum_{n=-\infty}^{\infty} x(n)\Psi_{j,k}(n)^* \tag{2}$$

$$\Psi_{j,k}(n) = 2^{j/2}\Psi\left(2^j n - k\right) \tag{3}$$

where j and k are scale and translation indices, $x(n)$ is the input signal, Ψ is wavelet function. Visualization of the log-spectra of wavelet transforms as heat-map yields a scalogram. In this case, the inputs to CNN models are scalogram saved as an image. This study uses a scalogram generated by WT with an analytic Morlet wavelet because it has better for transient localization.

The final approach is to employ Hilbert-Huang transform to create the inputs to CNN models as Hilbert-Huang spectrum (HHS). Hilbert-Huang transform, which is suitable for non-stationary and non-linear signals like EEG, has two main parts: empirical mode decomposition (EMD) and Hilbert transform (HT).

The fundamental part of HHT is EMD, which decomposes a signal into an intrinsic mode function (IMF). IMF is defined as a function that satisfies: (i) difference between the number of extrema and the number of zero-crossing must be equal to or smaller than one in the entire dataset. (ii) mean value of envelopes defined by local maxima and minima must be zero at any point. EMD decomposes signal in terms of IMF as follow [26]:

1. Take the signal $x(t)$ and find extrema points
2. Define lower and upper envelopes by connecting local minima points with cubic spline and local maxima points, respectively.
3. Calculate the mean value of the lower and upper envelopes.
4. Calculate $h_1(t) = x(t) - m_1(t)$.
5. If h_1 does not satisfy IMF condition, use h_1 as new signals and repeat 1–4.
6. If h_1 satisfies IMF condition, rename it as c_1 and calculate $r_n(t) = x(t) - c_n(t)$.
7. Use r_n as new signal and repeat 1–6 until r_n becomes a monotonic function.

After all, IMFs are extracted the signal can be represented as

$$x(t) = \sum_{j=1}^{n} c_j + r_n \tag{4}$$

where c_j is IMFs and r_n is residue. The second part of HHT is the Hilbert transform, which is used to find instantaneous frequency and amplitude. Signal, $x(t)$ can be transformed into an analytic signal by adding a complex part, $y(t)$, given as

$$y(t) = \frac{1}{\pi}\boldsymbol{P}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{5}$$

Where $\boldsymbol{P}$ is the principal value of the singular integral. Then the analytic signal, its instantaneous amplitude, $a(t)$ and phase, $\theta(t)$ are given as

$$z(t) = x(t) + jy(t) = a(t)e^{j\theta(t)} \tag{6}$$

$$a(t) = \sqrt{x^2(t) + y^2(t)} \tag{7}$$

$$\theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right) \tag{8}$$

Instantaneous frequency can be obtained by differentiating $\theta(t)$ as

$$\omega(t) = -\frac{d\theta(t)}{dt} \tag{9}$$

Finally, the Hilbert-Huang spectrum of $x(t)$ can be found by applying HT to each IMF of $x(t)$ and it can be visualized as a heat-map. In this case, the inputs to CNN models are HHS saved as an image.

3.3. Convolutional neural networks

The convolutional neural network is a deep neural network class principally applied to images for classification, object detection, segmentation, and image processing [27,28]. A CNN can consist of several types of layers: convolution, rectified linear unit (ReLu), pooling, dropout, fully connected (FC).

- Convolution Layer: This layer is the building block of CNN. In this layer, image or feature maps from the previous layer are convolved with sliding kernels to extract new features.
- ReLu Layer: This layer removes negative values from feature maps by applying activation function $f(x) = \max(0,\ x)$ to introduce nonlinearity in feature maps.
- Pooling Layers: This layer reduces the dimensionality of feature maps by sliding windows, calculating the mean, max, or sum of values inside the window to make the network invariant to small transformations.
- Dropout Layer: This layer sets input elements to zero with a given probability to reduce overfitting.
- Fully Connected Layer: This layer is a traditional multi-layer perceptron that uses a softmax activation function in the output layer. It classifies input images using features extracted by previous layers.

In this study, two groups of CNN models were used for classification EEG segments represented by images generated by time-frequency transformations. First group involves pre-trained AlexNet [28], GoogLeNet [29], ResNet-18 [30] and ResNet-50 [30]. Models in this group are winners of the ImageNet Large Scale Visual Recognition Challenge [31] in different years and trained on more than one million images from the ImageNet database. Because these models were originally designed to classify 1000 object categories, the final 1000-unit layer was replaced

Table 2
Summary of the CNN models in the first group.

Model	Depth	Size	Parameters	Input Size
AlexNet	8	227 MB	61 M	227-by-227
GoogLeNet	22	27 MB	7 M	224-by-224
ResNet-18	18	44 MB	11.7 M	224-by-224
ResNet-50	50	96 MB	25.6 M	224-by-224

with a 3-unit layer and retrained on training datasets given in Table 1. The models in this group are summarized in Table 2.

The second group has a model consisting of two GoogLeNet (model named GoogLeNet2). GoogLeNet2 was formed by combining two GoogleNet using their last global average pooling layer, as shown in Fig. 3. Although total depth did not increase, it enlarged as much as the original model. Because it has two inputs, it can exploit two TFRs at the same time. The model was trained on training datasets given in Table 1 using spectrogram and scalogram images as inputs simultaneously. GoogLeNet was selected because it has slightly better performance than the other.

For a fair comparison, all models were trained on the same data using Adam optimizer with a learning rate of 0.001. The training was halted if validation accuracy had stopped improving over five epochs. 10 % of the training data is used for validation during training. While models were being trained, image augmentation that is random transformations such as translation, reflection, and scaling was applied to the training images to reduce overfitting. Such transformations are depicted in Fig. 5.

3.4. Class activation mapping

A class activation map (CAM) for a specific category indicates the discriminative image regions used by CNN to identify that category [18]. As illustrated in Fig. 6, the output of global average pooling is computed by averaging feature maps of each unit of the last convolution layer. The final input is created by the weighted sum of these values. Correspondingly, CAM is obtained by a weighted sum of feature maps at the previous convolution layer. Formally, for a given image I, CAM of class c is obtained as

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \tag{10}$$

where $f_k(x,y)$ is a kth channel of the feature map, (x,y) represents spatial location, w_k is the kth weight corresponding to class c.

In this study, CAM has been employed to highlight discriminative regions of TFRs and shed light on the prediction of CNN models used to classify EEG segments.

4. Results and discussion

4.1. Results and discussion for classification

The classification task described in this paper performed using two strategies mentioned in Section 4. The first is to separately feed the spectrogram, scalogram, and HHS generated from segments with four distinct lengths into AlexNet, GoogLeNet, ResNet-18, and ResNet-50, generating 48 classification results. The second is to input the same

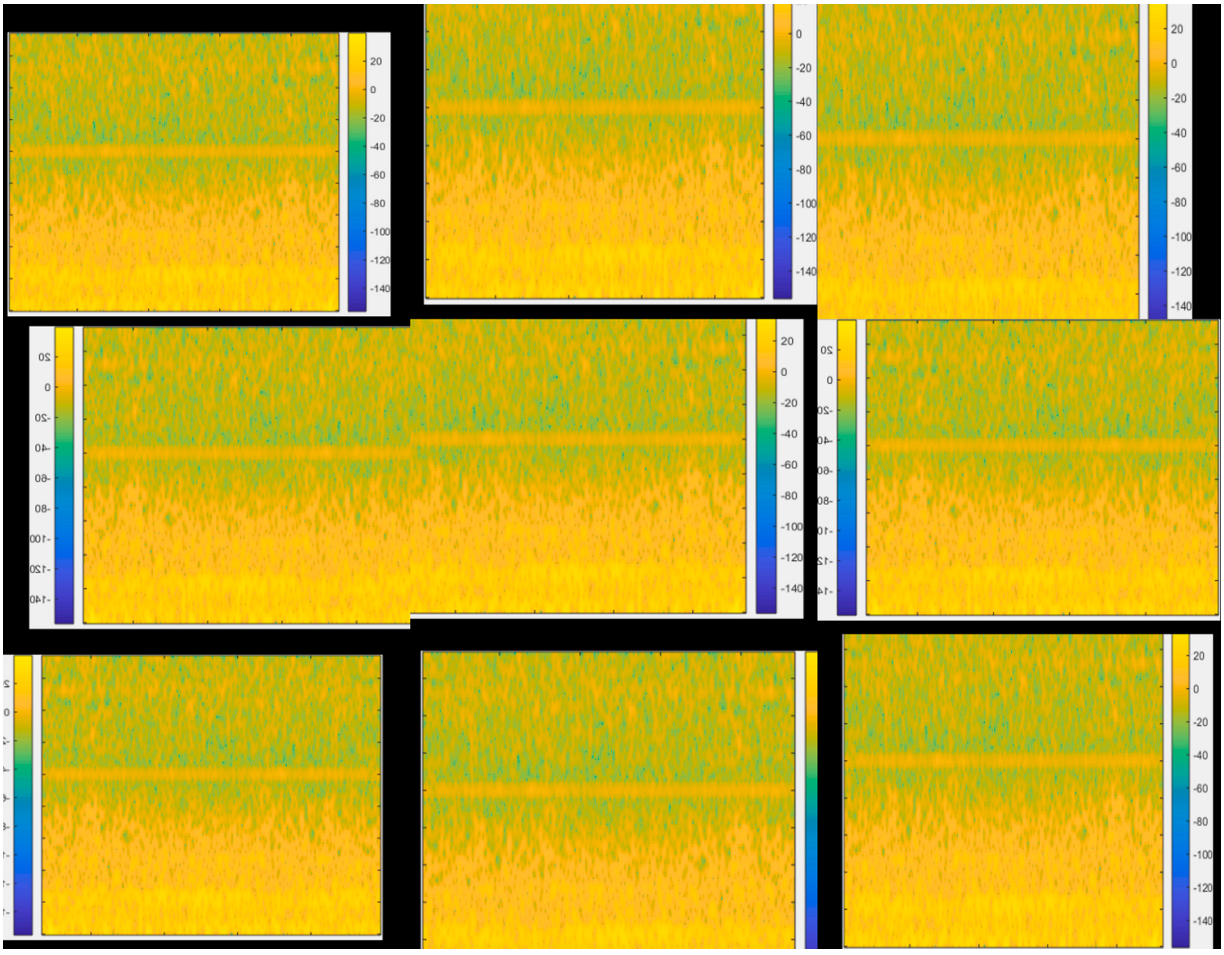

Fig. 5. Augmentation of training images.

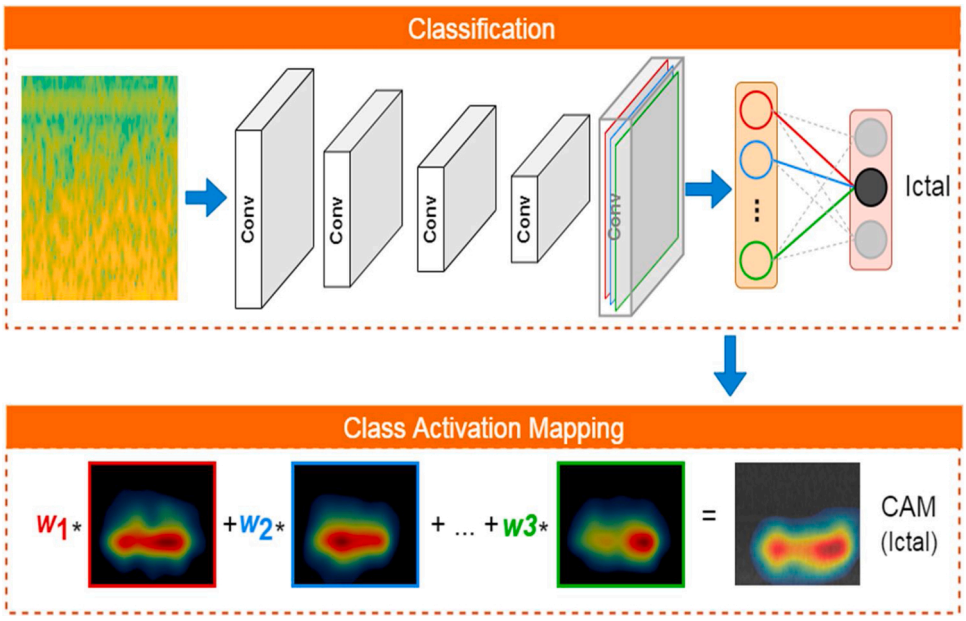

Fig. 6. Class activation mapping.

Table 3
Classification accuracies in percentage for strategy 1.

Model	N = 4097			N = 2048			N = 1024			N = 512		
	SPE	SCA	HHS	SPE	SCA	HHS	SPE	SCA	HHS	SPE	SCA	HHS
AlexNet	96.8	97.6	94.4	98.6	98.2	98.0	98.2	98.6	96.4	97.3	98.1	95.8
ResNet-18	97.6	94.0	93.8	98.0	98.8	97.4	98.0	98.1	95.5	96.9	96.6	94.5
GoogLeNet	97.2	96.0	94.0	98.2	98.6	97.4	98.4	**98.9**	95.7	98.1	97.7	95.9
ResNet-50	96.0	94.4	94.4	98.6	96.6	96.4	97.5	97.5	95.9	97.1	96.3	95.8

*SPE = spectrogram, SCA = scalogram, HHS = Hilbert-Huang spectrum.

Table 4
Classification accuracies in percentage for strategy 2.

Model	N = 4097	N = 2048	N = 1024	N = 512
GoogLeNet2	100	100	100	99.5

segment's spectrogram and scalogram into GoogLeNet2 simultaneously, creating four classification results. A total of 52 classification results is tabulated in Tables 3 and 4. Table 3 shows accuracies for the classification of normal, interictal, and ictal EEG segments depending on TFR, segment length, and CNN model using strategy 1. In strategy 1, The obtained accuracy ranges between a minimum of 93.8 % and a maximum of 98.9 %. Table 4 demonstrates the classification accuracies of GoogLeNet2 whose inputs are spectrogram and scalogram concerning segment length. In this case, the performance was considerably higher, obtaining accuracies between 99.5 % and 100 %.

By analyzing CNN models employed in strategy 1, the best performance was obtained by GoogLeNet, and the worst performance was obtained by ResNet-18. However, there is no significant difference between the mean accuracies of CNN models. Nevertheless, AlexNet and GoogLeNet were very similar and slightly higher than ResNet-18 and ResNet-50, whose performance was identical. Furthermore, strategy 2 proved itself to be effective by acquiring accuracy of 99.5 % for N = 512 and %100 for the rest because of the combination of best performer architecture and FTRs, which are GoogleNet, spectrogram, and scalogram, respectively.

Fig. 7 shows the performance of CNN models depending on segment length and TFR. Segment length of 1024 was the best in four scenarios for strategy 1, obtaining similar results to the case of segment length of 2048 (even if HHS was excluded, both would have the same mean accuracy) and segment length of 4097 was the worst because using too long segments causes a smaller number of training data and using too short segments causes an insufficient time-frequency representation of EEG signals. Even though using shorter segments or a greater number of segments was better than using longer segments or a smaller number of segments because CNN models are required a large amount of training data to avoid overfitting even input data augmentation described in Section 4.3 was used to increase an effective number of training data. Furthermore, segment length had almost no effect on the performance of GoogLeNet2, obtaining slightly lower in the case of segment length of 512. Therefore, segment lengths of 1024 and 2048 are the optimum length for all TFRs; using shorter EEG segments causes TFRs that carry inadequate information for EEG signals classification.

Scalogram with a combination of GoogLeNet and a segment length of 1024 was the best TFR in three scenarios for strategy 1. Although the mean accuracy of the scalogram was the highest, in some cases, its accuracy was lower than spectrograms, and the mean accuracies of both were almost the same. Therefore, there was no significant difference between them. Mean accuracy and performance in all cases of HHS were the worst among the three TFRs. Moreover, it was the most affected by segment length. Even though HHT was reported to be suitable for nonlinear and non-stationary data [32], it may not be applicable for this kind of task. Poor performance of HHT might be explained by mode mixing and end effect problems [33] of the method. Since GoogLeNet2 could benefit from both spectrogram and scalogram, it had superior results that are almost independent of segment length.

Many researches have investigated real-time implementation of

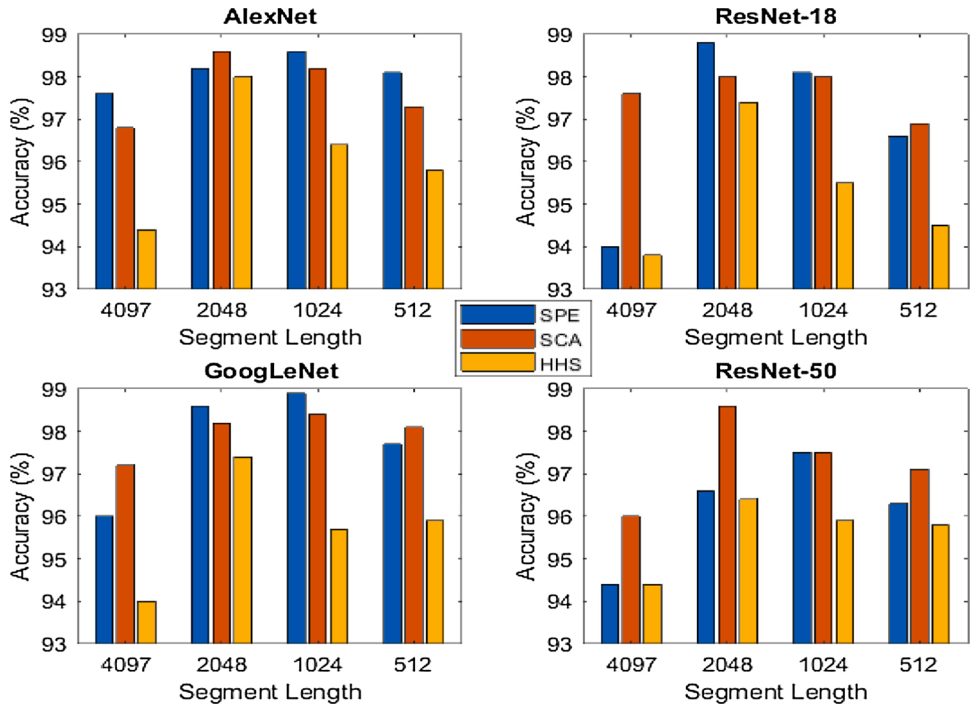

Fig. 7. Performance of CNN models with respect to segment length and TFR.

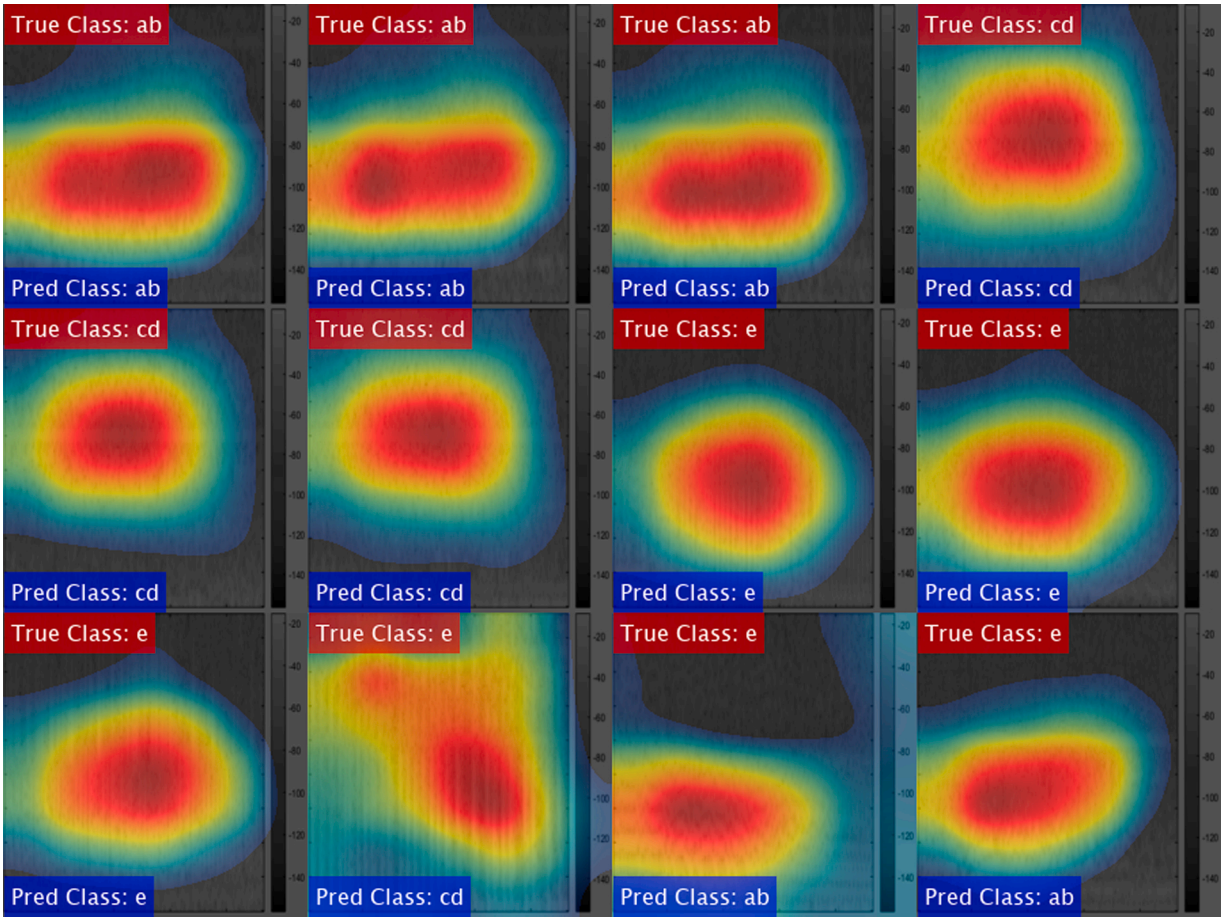

Fig. 8. CAM for spectrograms classified by GoogLeNet.

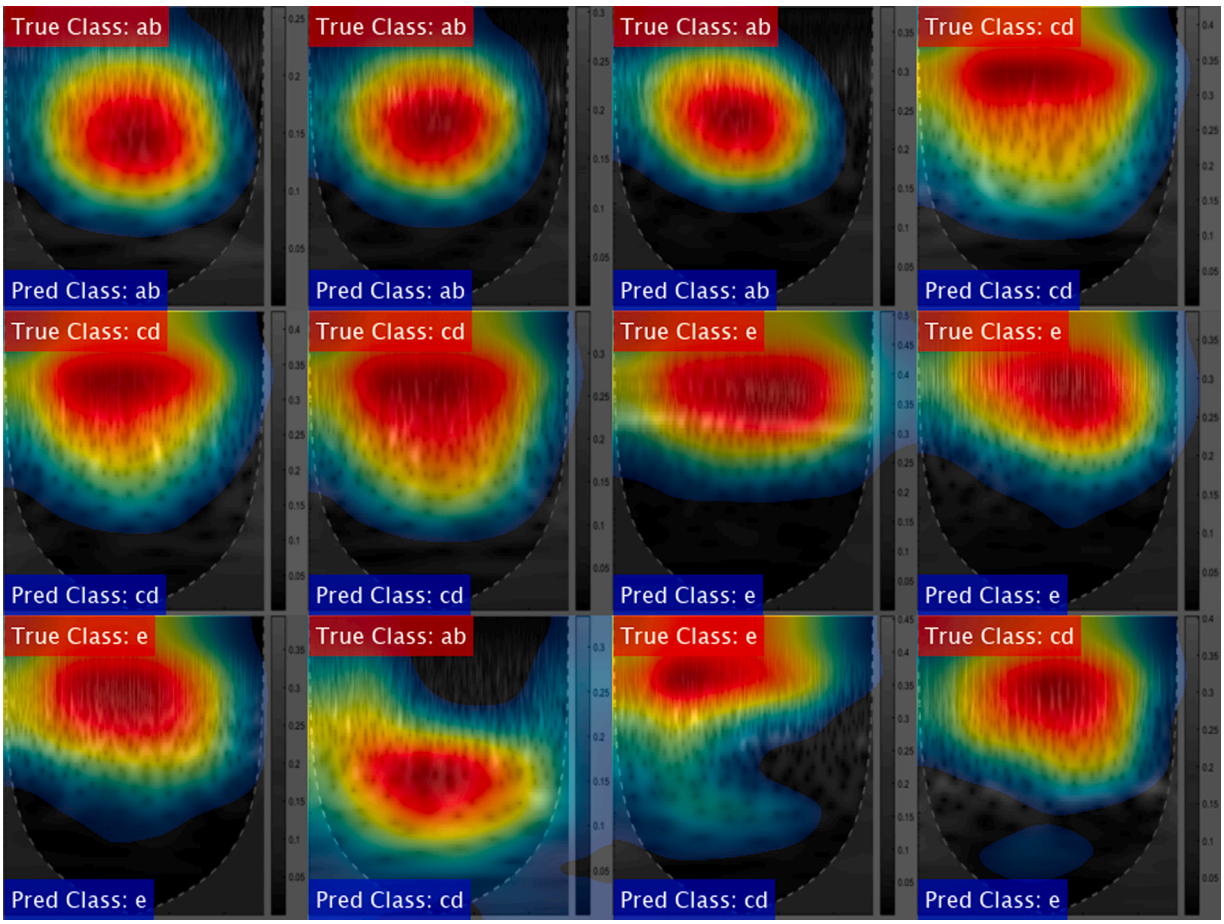

Fig. 9. CAM for scalograms classified by GoogLeNet.

epileptic seiuzure detection systems [34–38]. There are two approaches for such systems: software-based and hardware-based systems. Software-based systems often use computationally intensive algorithms while hardware-based systems that are wearable devices focus on low power consumption. Segmentation of continuous EEG signal, whose length defines minimum detection latency is a must regardless of the approach be used. Lower latency, namely shorter segment length in general causes lower accuracy. Therefore, choosing optimal segment length is important for both approaches. The discussion above has already investigated the effect of segment length on accuracy. Using GoogLeNet2 and a segment length of 512 samples ($\cong$ 2.95 s), epileptic seizures can be detected in real-time with low latency and high accuracy. This algorithm can be implemented as a software-based system with Matlab Graphical User Interface (GUI) or similar to [34,35,38] as a hardware-based system using a field-programmable gate array (FPGA).

4.2. Results and discussion for CAM

Class activation mapping is a technique that visualizes essential regions and the important degree of these regions of images classified by CNN. Figs. 8–10 depict important regions of randomly selected TFRs of normal, interictal, and ictal EEG segments whose lengths are 4097 were classified by GoogLeNet. True classes of each TFRs are shown on the left top corner of the images and predicted classes are on the left bottom. The first three images are correctly classified as normal EEG segments, the second three are correctly classified as interictal EEG segments, the third three are correctly classified as ictal EEG segments, and finally, the last three are incorrectly classified EEG segments. The horizontal axis is time in second (from 0 to 23 s), and the vertical axis is the frequency in Hertz (from 0 to 86 Hz) for all images. The red color indicates the most discriminative regions and blue indicates the least.

Fig. 8 shows CAM for spectrograms, and CAM indicates that regions with lower frequencies are the most important regions for normal and ictal EEG segments. These regions are larger for normal segments, revealing that more features are required for the identification of this class. Regions with higher frequencies are the most discriminative ones for interictal EEG segments. The last three images whose true classes are ictal were incorrectly classified as interictal, normal, and interictal, respectively, because either GoogLeNet could not properly focus on a specific region or focused on the wrong regions.

A similar analysis can be done for scalograms and HHS. However, since scalograms' frequency axis is on a logarithmic scale, it is more challenging to distinguish classes visually. Finally, HHS is the easiest to analyze, and they are very similar to the spectrograms.

4.3. Comparison with previous studies

Table 5 compares our results with previous studies that used CNN for the problem of classifying normal, interictal, and ictal signals in the University of Bonn EEG dataset. Because using CNN for the classification of EEG signals is a relatively new area of study, there are only a few studies. The first study using the same database was conducted by Acharya et al. [12]. They fed normalized raw EEG signals into a CNN model consisting of 13 layers. No signal transformation was employed in this study, the algorithm obtained a relatively low accuracy of 88.7 % for the same classification task. Ullah et al. [17] used an ensemble of pyramidal 1D-CNN with raw signals as input to the CNN. They employed a novel data augmentation scheme to increase the effective number of training data by using a sliding window to the segment EEG signal. They reported an accuracy of 99.1 %. San-Segundo et al. [5] proposed a CNN

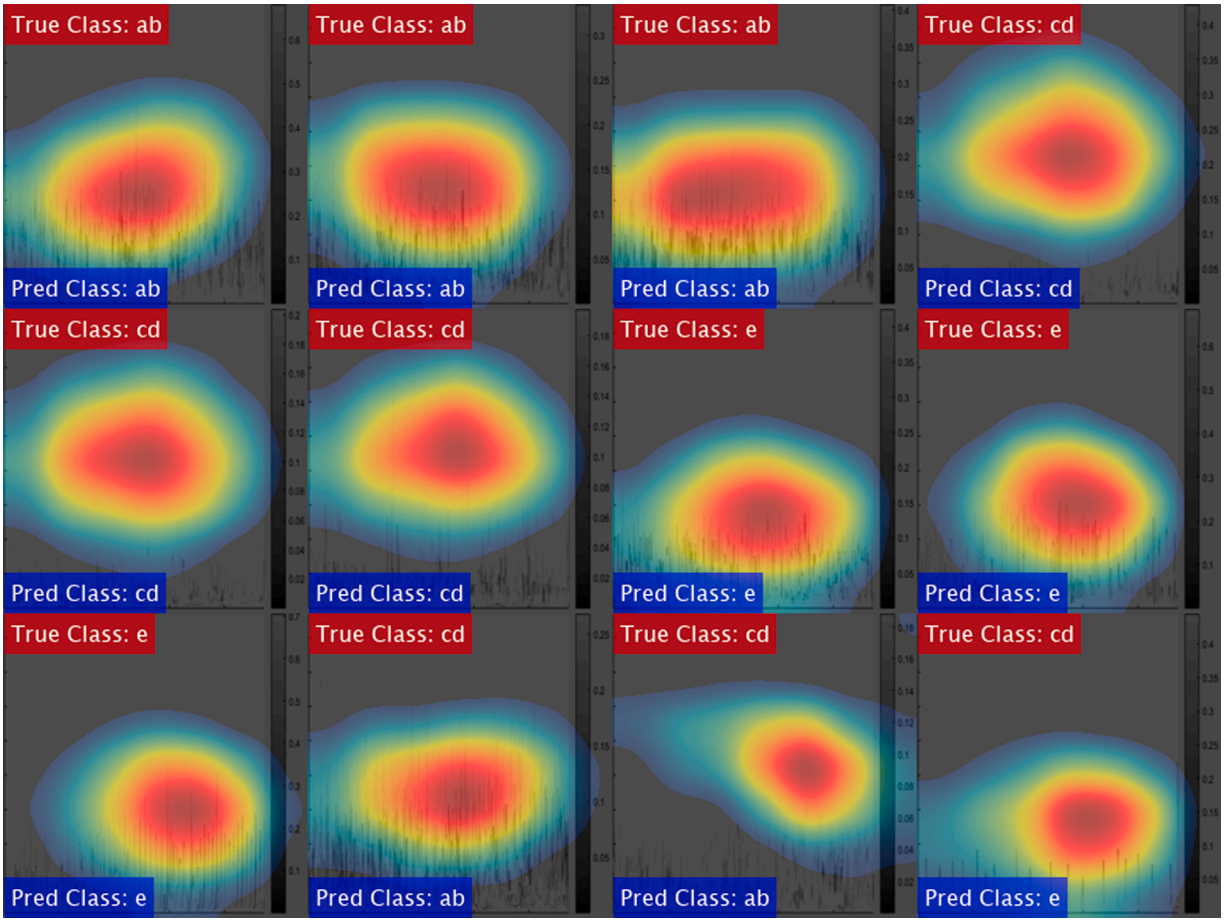

Fig. 10. CAM for HHSs classified by GoogLeNet.

Table 5
Accuracy comparison with previous studies using same database and CNN.

System (Year)		Accuracy (%)
Raw signal + 13-layers CNN (2018) [12]		88.7
Raw signal + Pyramidal 1D-CNN (2018) [17]		99.1
Raw signal + Wavelet + Fourier + EMD + 12-layers CNN (2019) [5]		96.5
This study	Spectrogram + AlexNet	98.6
	Spectrogram + ResNet-18	98.8
	Spectrogram + GoogLeNet	98.6
	Spectrogram + ResNet-50	98.6
	Scalogram + AlexNet	98.6
	Scalogram + ResNet-18	98.8
	Scalogram + GoogLeNet	98.9
	Scalogram + ResNet-50	97.5
	HHS + AlexNet	98.0
	HHS + ResNet-18	97.4
	HHS + GoogLeNet	97.4
	HHS + ResNet-50	96.4
	Spectrogram + Scalogram + GoogLeNet2	**100**

comprised of 12 layers and several signal transform (raw signal, wavelet transform, EMD, and Fourier transform) to generate input to the CNN. They used 2D signals instead of images as inputs. They achieved an accuracy of 96.5 % when all transformations were combined as one input. However, they did not investigate different segment lengths nor different CNN structures and unlike GoogleNet2, their model had one branch.

Using a novel GoogleNet2 model with spectrogram and scalogram, we achieved an accuracy of 100 %, outperforming San-Segundo et al. [5] by 3.5 % and the current state-of-art study Ullah et al. [17] by 0.9 %.

5. Conclusion

EEG signal is a powerful tool for identifying epilepsy. In this study, EEG signals were classified into three groups: normal, interictal, and ictal. This study's main contributions are analysis and evaluation of several CNN models for EEG signals classification, evaluation of various TFRs to generate inputs to CNN models, evaluation of several distinct segment lengths, and application of CAM for identifying a selective region of TFRs and assessing model prediction. Four pre-trained CNN architectures, which are AlexNet, ResNet-18, GoogLeNet, and ResNet-50, and a novel CNN architecture which is consisting of horizontally concatenating of two GoogLeNet models were used. Regarding the TFRs, a short-time Fourier transform was evaluated to generate spectrograms; wavelet transforms to generate scalograms and HHT to HHS. This analysis was carried out using segmented signals from the University of Bonn EEG dataset, obtaining significant improvements compared to previous studies. The best accuracy achieved is 100 % using GoogLeNet2, whose inputs are spectrogram and scalogram. GoogLeNet outperformed other models in group 1 in terms of maximum accuracy, and scalograms outperformed others. Results showed that the segment length of 1024 and 2048 are the optimum length for all TFRs. Finally, CAM shed light on the correct and incorrect prediction of CNN models by visualization of the discriminative region of the input images for each class.

CRediT authorship contribution statement

Abdulnasir Yildiz: Methodology, Conceptualization, Supervision, Resources. **Hasan Zan:** Software, Formal analysis, Writing - original draft, Visualization, Validation. **Sherif Said:** Writing - original draft,

Writing - review & editing, Methodology, Data curation.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] World Health Organization, Epilepsy, 2019 https://www.who.int/en/news-room/fact-sheets/detail/epilepsy [Accessed online: February 11th, 2020 18:00].
[2] Mayo Clinic, Epilepsy, 2019 https://www.mayoclinic.org/diseasesconditions/epilepsy/symptoms-causes/syc-20350093 [Accessed online: February 10th, 2020 17:00].
[3] K.S. Anusha, M.T. Mathews, S.D. Puthankattil, Classification of normal and epileptic EEG signal using time & frequency domain features through artificial neural network, in: 2012 International Conference on Advances in Computing and Communications, IEEE, 2012, August, pp. 98–101.
[4] B.E. Heldberg, T. Kautz, H. Leutheuser, R. Hopfengärtner, B.S. Kasper, B. M. Eskofier, Using wearable sensors for semiology-independent seizure detection-towards ambulatory monitoring of epilepsy, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, August, pp. 5593–5596.
[5] R. San-Segundo, M. Gil-Martín, L.F. D'Haro-Enríquez, J.M. Pardo, Classification of epileptic EEG recordings using signal transforms and convolutional neural networks, Comput. Biol. Med. 109 (2019) 148–158.
[6] R.B. Pachori, S. Patidar, Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions, Comput. Methods Programs Biomed. 113 (2) (2014) 494–502.
[7] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[8] O. Yildirim, R. San Tan, U.R. Acharya, An efficient compression of ECG signals using deep convolutional autoencoders, Cog. Syst. Res. 52 (2018) 198–211.
[9] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, S.W. Baik, Efficient deep CNN-based fire detection and localization in video surveillance applications, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2018) 1–16.
[10] Loris Nanni, Stefano Ghidoni, Sheryl Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification, Pattern Recognit. 71 (2017) 158–172.
[11] D. Li, G. Wang, T. Song, Q. Jin, Improving convolutional neural network using accelerated proximal gradient method for epilepsy diagnosis, in: 2016 UKACC 11th International Conference on Control (CONTROL), IEEE, 2016, August, pp. 1–6.
[12] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (2018) 270–278.
[13] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of non-linear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, Phys. Rev. E 64 (6) (2001), 061907.
[14] Ö. Yıldırım, U.B. Baloglu, U.R. Acharya, A deep convolutional neural network model for automated identification of abnormal EEG signals, Neural Comput. Appl. (2018).
[15] J. Thomas, L. Comoretto, J. Jin, J. Dauwels, S.S. Cash, M.B. Westover, EEG classification via convolutional neural network-based interictal epileptiform event detection, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, July, pp. 3148–3151.
[16] W. Hu, J. Cao, X. Lai, J. Liu, Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks, J. Ambient Intell. Humaniz. Comput. (2019) 1–11.
[17] I. Ullah, M. Hussain, H. Aboalsamh, An automated system for epilepsy detection using EEG brain signals based on deep learning approach, Expert Syst. Appl. 107 (2018) 61–71.
[18] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016) 2921–2929.
[19] X. Jia, L. Shen, Skin Lesion Classification Using Class Activation Map, 2017. ArXiv, abs/1703.01053.
[20] J. Cai, F. Xing, A. Batra, F. Liu, G.A. Walter, K. Vandenborne, L. Yang, Texture analysis for muscular dystrophy classification in MRI with improved class activation mapping, Pattern Recognit. 86 (2019) 368–375.
[21] J. Liu, W. Li, N. Zhao, K. Cao, Y. Yin, Q. Song, X. Gong, Integrate domain knowledge in training CNN for ultrasonography breast cancer diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2018, September, pp. 868–875.
[22] E.D. Übeyli, Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines, Comput. Biol. Med. 38 (1) (2008) 14–22.
[23] A. Subasi, M.I. Gursoy, EEG signal classification using PCA, ICA, LDA and support vector machines, Expert Syst. Appl. 37 (12) (2010) 8659–8666.
[24] N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines, Expert Syst. Appl. 39 (1) (2012) 202–209.
[25] N. Yalcin, G. Tezel, C. Karakuzu, Epilepsy diagnosis using artificial neural network learned by PSO, Turk. J. Electr. Eng. Comput. Sci. 23 (2) (2015) 421–432.
[26] N.E. Huang, Z. Wu, S.R. Long, Hilbert-Huang transform, Scholarpedia 3 (7) (2008) 2544.
[27] N.V. Mahajan, A.S. Deshpande, S.S. Satpute, Prediction of fault in gas chromatograph using convolutional neural network, in: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, 2019, April, pp. 930–933.
[28] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems (2012) 1097–1105.
[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015) 1–9.
[30] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, CoRR, 2015.
[31] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, A.C. Berg, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[32] N.E. Huang, M.L. Wu, W. Qu, S.R. Long, S.S. Shen, Applications of Hilbert–Huang transform to non-stationary financial time series analysis, Appl. Stoch. Models Bus. Ind. 19 (3) (2003) 245–268.
[33] W.-C. Shen, Y.-H. Chen, A.-Y. (Andy) Wu, Low-complexity sinusoidal-assisted EMD (SAEMD) algorithms for solving mode-mixing problems in HHT, Digit. Signal Process. 24 (2014) 170–186.
[34] M. Saleheen, H. Alemzadeh, A.M. Cheriyan, Z. Kalbarczyk, R.K. Iyer, An efficient embedded hardware for high accuracy detection of epileptic seizures, 2010 3rd International Conference on Biomedical Engineering and Informatics (2010).
[35] S.M. Kueh, T.J. Kazmierski, Low-power and low-cost dedicated bit-serial hardware neural network for epileptic seizure prediction system, IEEE J. Transl. Eng. Health Med. (2018), 1–1.
[36] L.S. Vidyaratne, K.M. Iftekharuddin, Real-time epileptic seizure detection using EEG, Ieee Trans. Neural Syst. Rehabil. Eng. 25 (11) (2017) 2146–2156.
[37] B. Karakaya, T. Kaya, A. Gulten, FPGA-based ANN design for detecting epilepticseizure in EEG signal, Balk. J. Electr. Comput. Eng. 6 (2) (2018) 15–19.
[38] R. Sarić, D. Jokić, N. Beganović, L.G. Pokvić, A. Badnjević, FPGA-based real-time epileptic seizure classification using Artificial Neural Network, Biomed. Signal Process. Control 62 (2020), 102106.