Computers in Biology and Medicine 166 (2023) 107517

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Multiband seizure type classification based on 3D convolution with attention mechanisms☆

Hui Huang *, Peiyu Chen, Jianfeng Wen, Xuzhe Lu, Nan Zhang

College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China

ARTICLE INFO

Keywords:
EEG
Seizure classification
Attention mechanisms
3D convolution

ABSTRACT

Electroencephalogram (EEG) signal contains important information about abnormal brain activity, which has become an important basis for epilepsy diagnosis. Recently, epilepsy EEG signal classification methods mainly extract features from the perspective of a single domain, which cannot effectively utilize the spatial domain information in EEG signals. The redundant information in EEG signals will affect the learning features with the increase of convolution layer and multi-domain features, resulting in inefficient learning and a lack of distinguishing features. To tackle these issues, we propose an end-to-end 3D convolutional multiband seizure-type classification model based on attention mechanisms. Specifically, to process preprocessed electroencephalogram (EEG) data, a multilevel wavelet decomposition is applied to obtain the joint distribution information in the two-dimensional time–frequency domain across multiple frequency bands. Subsequently, this information is transformed into three-dimensional spatial data based on the electrode configuration. Discriminative joint activity features in the time, frequency, and spatial domains are then extracted by a series of parallel 3D convolutional sub-networks, where 3D channels and spatial attention mechanisms improve the ability to learn critical global and local information. A multi-layer perceptron is finally implemented to integrate the extracted features and further map them to the classification results. Experimental results on the TUSZ dataset, the world's largest publicly available seizure corpus, show that 3D-CBAMNet significantly outperforms the state-of-the-art methods, indicating effectiveness in the seizure type classification task.

1. Introduction

Epilepsy is a chronic non-communicable disease of the brain that can be caused by a variety of etiologies, which is one of the most common neurological disorders worldwide [1–3]. The diagnosis of epilepsy and the accurate determination of seizure type are prerequisites for proper treatment, rational use of medication, and prognosis [4–6]. Electroencephalogram (EEG) is the main non-invasive clinical screening tool to detect abnormal electrophysiological activity in patients by directly reflecting changes in brain function through scalp electrical signals. The non-smooth and random nature of the EEG signal makes human diagnosis both labor-intensive and subjective to the diagnostician, resulting in a few intelligent diagnostic methods [7–10] for epileptic EEG.

With the improvement of computer technology, it is possible to achieve intelligent classification and detection of seizure types through algorithmic analysis of epileptic EEG signals [3,11]. The classification of different seizure types can greatly reduce the burden of clinical diagnosis and alleviate the subjective impact of the human diagnosis. The accurate classification of seizure types in the EEG signal may further facilitate targeted treatment and improve treatment outcomes, which helps researchers to understand the seizure mechanisms of different seizure types and formulate treatment plans.

Time–frequency analysis, a powerful tool for analyzing time-varying non-stationary signals, is often used to characterize the joint distribution of information in the time and frequency domains of epileptic EEG. It can describe the relationship between the frequency of non-stationary EEG signals over time. The wavelet transform reflects the frequency intensity corresponding to brain activity while retaining important transient information in the time domain and is widely used in EEG signal analysis. For example, Gandhi et al. [12] characterized the energy, variance, and information entropy of EEG signals within a frequency band and used wavelet functions to investigate the characteristics of EEG signals. Guo et al. [13] used the multiscale wavelet transform to decompose the EEG signal and used the approximate entropy of each frequency band as a feature of the EEG signal. Faust [14] used wavelet transform-based EEG signal denoising and feature extraction

☆ This work was supported by the National Natural Science Foundation of China (Grant No. 62072340 and 62006076).

* Corresponding author.
E-mail address: huanghui@wzu.edu.cn (H. Huang).

https://doi.org/10.1016/j.compbiomed.2023.107517
Received 24 May 2023; Received in revised form 28 August 2023; Accepted 19 September 2023
Available online 25 September 2023

techniques for an epilepsy diagnosis. The feature extraction method of obtaining joint distribution information in the time and frequency domains through time–frequency analysis is more comprehensive than the single-domain feature extraction method and is therefore widely used. However, EEG signal data is obtained from scalp electrical signals in different regions of the brain, and the acquisition electrodes have a spatial distribution pattern that reflects the joint activity relationship between electrode signals, which can be an important basis for differentiating seizure mechanisms between different seizure types. However, the above existing algorithms only used two-dimensional planar EEG features for classification and neglected the information on the spatial distribution, which limits the accuracy of the epilepsy EEG classification.

In the study of seizure type classification, EEG signals can be used as material for complex condition detection with the help of machine learning methods [15,16]. Saputro et al. [17] extracted the Mel inverse spectral coefficients and Hjorth as features and put them into an SVM classifier with cubic polynomial kernel function to get the classification accuracy of 91.4%. [18] is validated against a support vector machine (SVM)-based neonatal epilepsy detection algorithm. [19]introduces a method based on sample weighting to improve seizure prediction performance by assigning weights to different samples. Roy et al. [20] used a fast Fourier transform to extract frequency features. A combination of multiple classifiers KNN, SGD, XGBoost, Adaboost, etc. was used to explore the ability to perform the seizure classification task. Traditional machine learning models are less complex and more explanatory but rely on expert experience to extract valid classification features. Extracted features are limited to shallow layers, and deeper features with discriminative properties cannot be explored, resulting in limited model accuracy and time-consuming feature extraction of signal data, which affects model efficiency in practical applications. Deep learning methods have the ability to process raw data and extract deep features to eliminate the dependence on manual features. There are many methods [21–23] of intelligent diagnosis based on CNN. [24] utilizes an ensemble learning approach to improve the accuracy and robustness of seizure prediction by combining multiple deep-learning models. However, due to the high complexity of the model, a large amount of computing resources are required for training and reasoning. Simultaneously, these intelligent diagnostic methods often revolve around the classification and detection of seizure periods, with few studies addressing the classification of seizure types, i.e., partial/systemic seizures, etc. In this paper, we propose a 3D convolutional multi-band seizure-type classification model based on channel and spatial attention mechanisms with seizure-type classification as the theme. The main contributions of this paper are as follows:

(1) To address the drawbacks of single feature extraction methods and effectively utilize the spatial domain distribution information, this paper proposes a method to construct three-dimensional data in the time–frequency-spatial domain. The joint multi-band time–frequency distribution information is obtained through multi-stage wavelet decomposition and reconstructed into a more comprehensive time–frequency-spatial domain data representation according to the EEG signal spatial domain location, providing the possibility of subsequent analysis of joint signal spatial domain activity information.
(2) To achieve targeted processing of each frequency band, a series of parallel 3D convolutional sub-networks are used instead of a single-path two-dimensional convolutional process. The number of branching layers is reduced to alleviate the loss caused by continuous convolution while ensuring improved accuracy. Three-dimensional convolution structure can acquire the joint time-domain, frequency-domain, and spatial-domain active features simultaneously respectively, which is free from multi-domain feature fusion and reduces the potential loss to achieve the purpose of effectively coordinating multi-domain features.
(3) A three-dimensional channel and spatial attention mechanism are used to generate the corresponding global and local information embedding vectors for the channel and spatial features, which improves the ability to selectively learn critical global and local information. The module can effectively counter the problem of inefficient learning due to the influence of redundant information in the dataset for the features learned in the shallow framework. Moreover, the end-to-end deep convolutional structure is designed to mine deep features of epilepsy EEG signals, which can improve model accuracy, reduce manual selection, and enhance model efficiency in practical applications.

We organize the remainder of the paper as follows. Section 2 introduces the related works. Section 3 details the proposed end-to-end 3D convolutional multiband seizure-type classification model. In Section 4, experiment results show the feasibilities of the proposed model. Finally, some conclusions are given in Section 5.

2. Related work

Deep learning methods are the result of advances in machine learning research [25–28], with the ability to process raw data and extract deep features to eliminate the reliance on manual features, and are widely used in seizure-type classification tasks. Raghu et al. [29] converted EEG into spectrograms and used AlexNet, VGG16, VGG19, and basic convolutional neural network models to handle the 8-classification problem. Covert [30] used a variety of convolutional neural networks with different layers to form a temporal graph convolutional network, which effectively utilized temporal and structural information for automatic epilepsy detection. Cao et al. [31] used a squeeze and excitation network to extract short-term features, then extracted long-term features and constructed a classifier by a long short-term memory network, with an average classification accuracy of 87.5%. However, these networks also have some drawbacks: firstly, many modules need manually selected, and tuning is also complicated. Secondly, the continuous convolution process of traditional convolutional neural networks loses important information about the input or gradient as the number of layers increases. Features learned from shallow frameworks are affected by redundant information in the dataset, leading to inefficient learning. Li et al. [32] proposed an end-to-end EEG seizure detection framework that combines multi-level wavelet [33] decomposition and multi-scale time-domain analysis to extract time–frequency domain features of epileptic EEG signals and achieves recognition of epileptic EEG signals through a series of parallel sub-networks stitched together with two-dimensional convolution and squeeze and excitation modules. The multi-domain feature fusion enhances the classification performance of the model but may also be accompanied by the loss of auxiliary discriminative features. The limited success of current methods in EEG signal recognition may be due to the failure to simultaneously coordinate information from different domain features. Therefore, there is a demand to refine epilepsy EEG signal classification models that can simultaneously capture and analyze feature information from different domains and address the issue of data redundancy. We propose that spatial three-dimensional data based on multi-level wavelet decomposition and electrode arrangement can well solve the information coordination problem of different domain features, and can capture good feature information through the designed 3D convolutional network and attention mechanism.

3. Methodology

Three-dimensional seizure type classification based on channel and spatial attention is divided into three parts. Firstly, multi-band time–frequency domain joint distribution information is isolated from the input data using time–frequency analysis. Secondly, the EEG time–frequency-spatial 3D data of each frequency band is constructed based

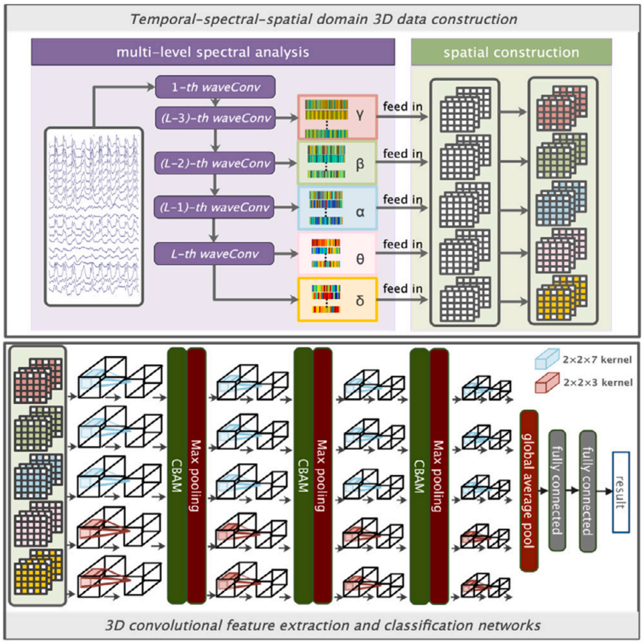

Fig. 1. 3D Convolutional Multiband Seizure Type Classification with Attention Mechanisms. The Temporal-Spectral-Spatial domain 3D data construction module applies multi-level wavelet decomposition to generate two-dimensional representations of 19-channel EEG data, preserving spatial position distribution information. These two-dimensional channel data are then combined for each sampling point, resulting in three-dimensional time–frequency-space domain data input. The 3D Convolutional feature extraction and classification networks extract spatial position features and time–frequency information using three-dimensional convolutions while capturing important features through attention mechanisms. The final predictions are obtained through fully connected layers for classification.

on the spatial-domain electrode arrangement. Finally, each frequency band data is input into a series of parallel sub-networks composed of 3D convolutional layers and channel spatial attention modules respectively to obtain classification results, and the flow is shown in Fig. 1.

3.1. Multi-band extraction based on multi-level wavelet decomposition

In this paper, wavelet packet decomposition is used to extract time–frequency features from EEG data. While having the advantages of the wavelet transform, wavelet packet decomposition can simultaneously decompose high and low-frequency information, to obtain both comprehensive and non-redundant results for better time–frequency localization analysis. The period-filled wavelet decomposition is achieved by a generic convolution operator, the wavelet convolution layer, which for a one-dimensional input x of length N can be defined as:

$$x_p = x\left(N-\frac{R}{2}+1\right), \ldots, x(N-1)\,|\,x(0), \ldots, x\left(\frac{R}{2}-2\right), \tag{1}$$

$$y_A(i) = \left(x_p \otimes g\right)(i) = \sum_{r=0}^{R} x_p(s\times i-r)\times g(r), \tag{2}$$

$$y_D(i) = \left(x_p \otimes h\right)(i) = \sum_{r=0}^{R} x_p(s\times i-r)\times h(r), \tag{3}$$

where | represents the concatenation, $x(i)$ denotes the ith element of input x, $\otimes$ denotes the convolution, g and h represent a set of scales and wavelet filters, R and s represent the convolution kernel size and step size in the convolution. Compared with the mode containing 0 padding, periodic padding can avoid data boundary distortion while ensuring that the length of the wavelet decomposition result is not too short. The wavelet packet decomposition is performed simultaneously on T

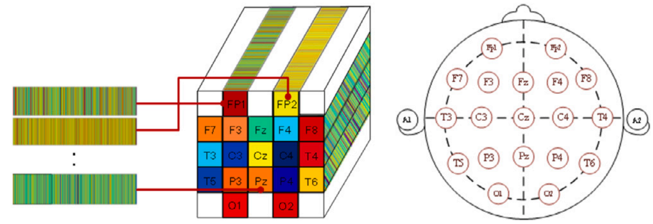

Fig. 2. Temporal-spectral-spatial domain 3D data construction.

electrode signals of the $T \times N$ matrix, and the wavelet convolution layer performs an iterative decomposition of the results generated by the upper layer, denoted as y_A, to isolate five frequency bands, 32–64 Hz (band), 16–32 Hz (band), 8–16 Hz (band), 4–8 Hz (band) and 0–4 Hz (band), which are often used for analysis in medical diagnosis and effectively express the brain activity state, the number of steps L depends on the sampling frequency of the input data. The wavelet transform was carried out using Daubechies order-4 (Db4).

3.2. Time–frequency-spatial 3D feature construction

The two-dimensional time–frequency feature data of the C electrode signals were reconstructed into 3D time–frequency-spatial data according to the electrode distribution. The correspondence between the electrode data and the position matrix is shown in Fig. 2, the $t \times t$ spatial position matrix is constructed according to the international 10/20 system, and the spatial position is corresponded to the signal time–frequency information one by one. The two-dimensional signal data of each sampling point is reconstructed into 3D time–frequency-spatial features. T electrode N' sampling points of time–frequency data, reconstructed into the $t \times t \times N'$ three-dimensional time–frequency-spatial information.

3.3. Multi-band 3D convolutional feature extraction and classification network

Two-dimensional convolution extract features from image dimensions only [34]. When applied to multi-dimensional data, information encoded in multiple consecutive frames needs to be captured. 3D convolution is achieved by convolving 3D kernels into a cube formed by superimposing multiple consecutive frames. By this construction, feature maps in the convolution layer are connected to the layer above the multiple consecutive frames in the convolution layer, thereby capturing the information. Formally, the value at position (x, y, z) on the jth feature map of the ith layer is defined as

$$v_{ij}^{xyz} = \tanh\left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right), \tag{4}$$

where R_i is the size of the 3D convolutional kernel along the time dimension and w_{ijm}^{pqr} is the (p, q, r)-th value of the kernel connected to the mth feature mapping in the previous layer. The 3D convolutional feature extraction network performs further feature extraction with the flow shown in Fig. 1. For the input 5-band time–frequency-spatial domain 3D EEG signal data, a 3D convolutional kernel is selected to extract three dimensions of features simultaneously for the 5-band data. The $k_h \times k_w \times k_d$ 3D convolutional kernel is used to implement 3D feature extraction with $k_h \times k_w$ convolution obtaining data information on the joint activity between electrode signals at $t \times t$ sampling points, while the convolutional kernel's third-dimensional sliding convolution obtains all pixels for which time–frequency feature mapping.

Most of the core ideas of deep learning combined with attention mechanisms are to highlight key information in the data through a new layer of weights, called masks, that are trained and learned, causing the neural network to focus on key regions in the data and form attention. In the visual system, this can be expressed as

$$Attention = f\left(g\left(x\right), x\right), \tag{5}$$

where $g(x)$ processes the input features and generates attention, and $f(g(x), x)$ represents the process of processing the input features in conjunction with attention. The channel and spatial attention module (CBAM block) [35] model the interdependencies between the convolutional feature channels, aggregates the spatial information of the feature mapping and improves the quality of the representation obtained from the network training. The CBAM performs a feature recalibration mechanism that gives the network the ability to learn global and local information and selectively emphasize important discriminative information while suppressing features that do not contribute significantly to classification. To interface with 3D convolution, the model extends CBAM into 3D spatial, and for input X mapped to a feature mapping, the corresponding channel attention module is constructed to achieve the highlighting of channel discriminative features, and the channel attention module is formulated as

$$\begin{aligned} M_C(F) &= \sigma(MLP(AvgPool(F) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \end{aligned}, \tag{6}$$

where σ is the sigmoid function, the feature F is compressed by maximum pooling and average pooling to obtain the maximum channel information embedding vector and the average channel information embedding vector, respectively, after MLP, the maximum information embedding vector and the average information embedding vector corresponding to the channel are summed and activated by the sigmoid function to generate the channel attention descriptors. The feature mapping on the dimension of the aggregation spatial $t' \times t' \times N' \times C$ generates each channel descriptor containing the corresponding global information embedding vector of the channel features, which is global in nature. The channel attention output feature map is obtained by multiplying it with the corresponding channel of the input feature map. The spatial attention mechanism is to compress the channels, and the formula of the spatial attention module is defined as:

$$\begin{aligned} M_C\left(F\right) &= \sigma\left(f^{7\times7}\left(\left[AvgPool\left(F\right); MaxPool\left(F\right)\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \end{aligned}. \tag{7}$$

The output feature map of the channel attention module is used as the input feature map of the spatial attention module, and the maximum pooling and average pooling compression of the channel dimension is done to obtain the maximum spatial information embedding vector and the average spatial information embedding vector, and after concatenation and convolutional dimensionality reduction, the spatial attention descriptors are generated after sigmoid activation and multiplied with the corresponding positions of the output feature map of the channel attention module. The learning of the importance of each channel and spatial location is achieved, and the discriminative features are enhanced to suppress the role of weak features.

The classification network consists of a multilayer perceptron (MLP), and the result with the highest output probability is the class of classification, achieved by 4 times 3D convolution and CBAM connection for split-band spatial–temporal feature extraction.

4. Experiments and analysis

To validate the performance of the proposed model in the seizure type classification task, seven classification experiments were conducted on the TUSZ dataset, the world's largest publicly available seizure corpus. A five-fold cross-validation method is used to analyze the overall classification effect and the model's seizure-type classification performance. Furthermore, the confusion matrix under 2D and 3D convolution was used to verify the effectiveness of 3D feature construction combined with 3D convolutional structure. By comparing the confusion matrices under different attention mechanisms, the effectiveness of the channel and spatial attention mechanism incorporation can be verified. Finally, the models are compared with state-of-the-art algorithms on the seizure type classification task. We train the model with a warm-up learning rate and optimize the parameters using a stochastic gradient descent optimization function. We compare the proposed method with state-of-the-art in a workstation with a core Inter i7-12700k CPU and GeForce RTX 3060 GPU.

Table 1
Details of TUSZ dataset.

Seizure type	Count
Absence seizure (AB)	99
Complex partial seizure (CP)	367
Focal non-specific seizure (FN)	1836
Generalized non-specific seizure (GN)	583
Myoclonic seizure (MY)	3
Simple partial seizure (SP)	52
Tonic clonic seizure (TC)	48
Tonic seizure (TN)	62

4.1. EEG dataset and preprocessing

The TUSZ dataset [36] used for the experiment includes clinical EEG data from 642 cases collected at the Temple University Hospital over the past 14 years and contains a rich variety of seizure patterns. Table 1 gives an overview of the data for the different seizure types in the TUSZ dataset. There are eight seizure categories in the TUSZ dataset, in which only three segments are available for myoclonic seizures, so data from seven seizure categories other than these were selected for the experiment. Due to the long span of clinical collection in the dataset, the criteria of the parameters collected changed. The seizure part was cut out from the original data to facilitate neural network training according to the seizure events marked in the instruction document. The preprocessing procedure of EEG data includes three steps. Firstly, considering the difference in sampling rate inter-patients, we resample EEG recordings to 128 Hz. Then the EEG signal data was selected according to the international EEG 10/20 system mainly used in clinical practice, as shown in Fig. 2, the red circle is the electrode channel used in the experiment. Finally, data were cut into 4s segments, and data smaller than 4s were discarded.

4.2. Evaluation indicators

Precision, recall, and F1-score were used to evaluate the classification performance of the proposed network, and the formula is expressed as follows:

$$Precision = \frac{TP}{(TP + FP)}, \tag{8}$$

$$Recall = \frac{TP}{(TP + FN)}, \tag{9}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{(Precision + Recall)}, \tag{10}$$

where TP is the number of samples correctly classified, FP is the number of samples misclassified as positive cases and FN is the number of samples misclassified as negative cases. A confusion matrix is a situation analysis table in machine learning that summarizes the predicted results of a classification model, in the form of a matrix that summarizes the records in the dataset according to two criteria: the true category and the category judgment predicted by the classification model.

Table 2
Result of model in epilepsy classification.

Seizure type	Precision	Recall	F1-score
ABSZ	0.9994	0.9718	0.9854
CPSZ	0.9699	0.9081	0.9380
FNSZ	0.7951	0.9224	0.8540
GNSZ	0.8772	0.9109	0.8937
SPSZ	0.9994	0.9679	0.9834
TCSZ	0.9863	0.9388	0.9620
TNSZ	0.9964	0.9838	0.9901
Average	**0.9462**	**0.9434**	**0.9438**

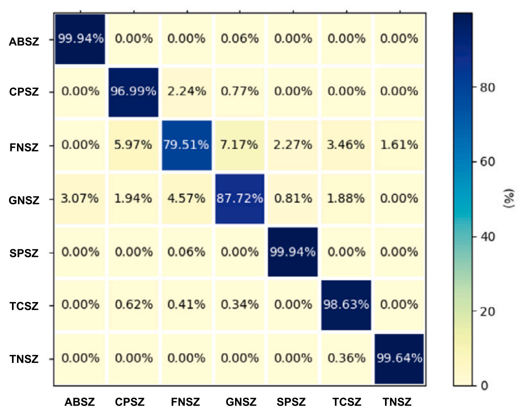

Fig. 3. Confusion matrix of the model in epilepsy classification.

4.3. Analysis of model seizure type classification

The results of the seizure-type classification task are shown in Table 2. In the 3D convolutional seizure type classification based on channel and spatial attention, the classification precision of all categories except FNSZ is higher than 0.85 (e.g., the classification precisions on ABSZ, CPSZ, SPSZ, TCSZ, and TNSZ were 0.9994, 0.9699, 0.9994, 0.9863, and 0.9964 respectively, all above 0.95), which demonstrates that the model could correctly predict positive samples. Moreover, the classification recall for all categories was higher than 0.90, demonstrating that the model had good recognition ability for all types of samples. The F1-score was higher than 0.85, with the model performing better on the seizure-type classification task. The overall classification results were superior, demonstrating that the model was able to classify seizure types more accurately. However, the model's precision for the FNSZ seizure class was observed to be significantly lower than the other classes, differing from the highest precision class by 0.23. The model's ability to correctly predict the FNSZ and GNSZ classes was significantly lower than the other classes, but the difference with the other classes was not too great, similar to GNSZ, indicating that the model's ability to correctly predict the FNSZ and GNSZ classes was significantly lower than the other classes.

The confusion matrix was plotted against the classification results to observe the confusion between different classes. As shown in Fig. 3, the five classes ABSZ, CPSZ, SPSZ, TCSZ, and TNSZ were identified with strong ability, and the probability of being correctly identified reached over 95%, while the probability of being misidentified as other classes was also low. This demonstrates that the channel and spatial attention-based 3D convolutional neural network is a powerful method for seizure-type classification. It can be seen that the probability of FNSZ and GNSZ being misidentified is 7.17% and 4.57%, respectively, which may be due to the high similarity and similar characteristics between the two types of seizure signals. In other words, FNSZ and GNSZ are easily confused, which is explored subsequently.

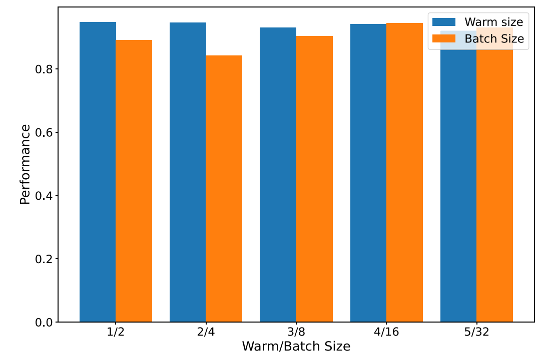

Fig. 4. Hyperparameter Analysis of Batch and Warm.

Table 3
Results under 2D and 3D convolution.

Seizure type	Precision		Recall		F1-score	
	2D	3D	2D	3D	2D	3D
ABSZ	**0.9988**	0.9967	0.9535	**0.997**	0.9756	**0.9968**
CPSZ	0.7606	**0.8802**	**0.9117**	0.8837	0.8293	**0.8819**
FNSZ	0.716	**0.8036**	**0.9099**	0.8171	0.8014	**0.8103**
GNSZ	0.8687	**0.8797**	0.744	**0.8466**	0.8015	**0.8628**
SPSZ	**0.9877**	0.9709	0.9401	**9827**	**0.9633**	1.9416
TCSZ	**0.9956**	0.9688	0.8864	**0.969**	0.9378	**0.9689**
TNSZ	**0.9986**	0.9915	0.888	**0.9965**	0.9401	**0.9940**
Average	0.9037	**0.9273**	0.8905	**0.9275**	0.8971	**0.9274**

4.4. Hyperparameter analysis

To investigate the issue of hyperparameter optimization for our model, we conducted an analysis on two hyperparameters: batch size (batch) and warm-up learning rate(warm). We explored the impact of varying batch sizes (2, 4, 8, 16, 32) and warm-up values (1, 2, 3, 4, 5) on the model's performance. As shown in Fig. 4, the overall model performance exhibited an increasing trend with the increase in batch size until it reached saturation at batch size 16, beyond which performance started to decline. Conversely, increasing the warm-up value resulted in a decrease in model performance. We attributed this effect to the potential incomplete learning of crucial features due to excessive warm-up rounds.

4.5. Three-dimensional convolutional validity analysis

To verify the effectiveness of 3D convolution for seizure classification, we ran 2D convolutional neural networks without the construction of time–frequency-spatial domain 3D information and only after multi-stage wavelet decomposition, and time–frequency-spatial domain 3D convolutional neural networks, respectively, and the results are given in Table 3.

By comparing the experimental results under the two settings, it was found that the precision of the CPSZ, FNSZ, and GNSZ categories under 3D convolution is higher than that of 2D convolution in terms of the accuracy index, which indicates that 3D convolution enhanced the ability of the model to correctly identify samples to a certain extent. The accuracy of ABSZ, SPSZ, TCSZ, and TNSZ in the case of 2D convolution was better than that of 3D convolution, but the difference did not exceed 0.03, and the average accuracy was 0.0236 lower than that of 3D convolution. As shown in Table 3, the F1-score values under 3D convolution were all better than those of 2D convolution, which fully demonstrated the effectiveness of 3D convolution for the epilepsy classification task.

The confusion matrix of the model seizure type classification is shown in Fig. 5, where the confusion matrix was plotted against the

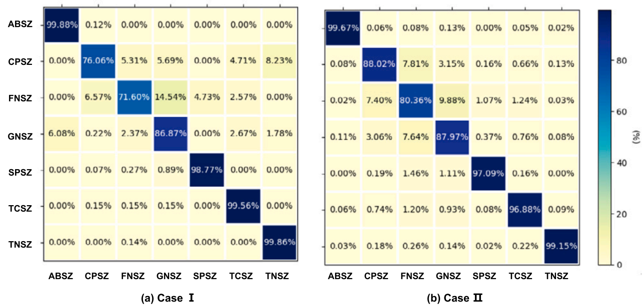

Fig. 5. Confusion matrix of the model in epilepsy classification.

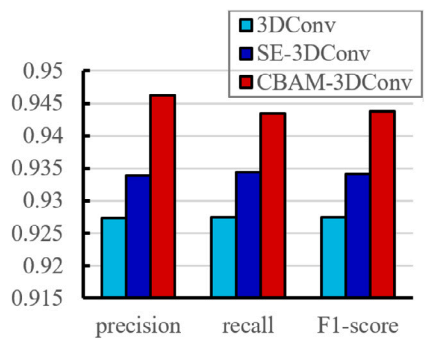

Fig. 6. Results under different attention.

classification results to observe the confusion under the two experimental settings. The probability of confusing GNSZ with ABSZ decreased by 5.97%, the probability of confusing CPSZ with TCSZ decreased by 4.05% and the probability of confusing TNSZ decreased by 8.1%. However, this was accompanied by a slight increase in the probability of confusion for some categories, but its comprehensive performance was better than the classification results under 2D convolution. This shows that 3D convolution effectively coordinates multi-domain information and improves the performance of the seizure classification model.

4.6. Analysis of the effectiveness of channel and spatial attention mechanisms

To verify the effectiveness of 3D convolution for seizure classification, we ran (1) 3D convolutional neural networks incorporating the channel attention mechanism (SE) and (2) incorporating the channel and spatial attention mechanism (CBAM) and without the attention mechanism, respectively, and the results are shown in Fig. 6.

By comparing the experimental results under the three settings, the average precision, recall, F1-score, and accuracy of 7 classifications were significantly improved after adding the attention mechanism. This shows that the attention mechanism enhances the model's ability to correctly predict samples and effectively highlights the discriminant features, which verifies the effectiveness of the attention mechanism. By comparing the confusion matrix of adding only the channel attention mechanism and adding channel and spatial attention mechanism, it was found that the classification accuracy and F1-score improved by 1.06% and 0.0097 respectively when both channel and spatial dimensions were considered compared to when only the channel dimension was considered, which proves that the channel and spatial attention modules can aggregate feature mapping by modeling the interdependencies between them. In this way, the convolutional feature channels and the spatial information allow the network to learn global and local information and selectively emphasize important discriminative information, effectively improving the quality of the representation obtained from the network training.

As shown in Fig. 7, although CPSZ is still easily confused with FNSZ and GNSZ in the inter-class comparison after adding the channel and spatial attention mechanisms, there is a very significant improvement compared to the non-inclusion of the channel and spatial attention mechanisms, resulting in a 5.57% decrease in the probability of CPSZ being falsely identified as FNSZ, the probability of misidentification of CPSZ as GNSZ decreased by 2.38%, the probability of misidentification of FNSZ as CPSZ decreased by 1.03%, the probability of misidentification of FNSZ as GNSZ decreased by 2.71%, the probability of misidentification of GNSZ as CPSZ decreased by 1.12%, the probability of misidentification of GNSZ as FNSZ decreased by 3.07%. Although there was a slight increase in the probability of confusion in some categories, the classification results were generally better than when the channel and spatial attention mechanisms were not included. This demonstrates that the channel and spatial attention mechanisms have an effect on the performance of the seizure classification model.

4.7. Comparison with advanced methods

In addition to the methods mentioned above, Wang [37] proposed an EEG band selection method using natural language processing features to compare three types of input data: bands extracted from clinical reports using NLP, EEG bands, and the complete spectrum for the role of seizure type classification, with classifiers selected from the random forest and support vector machine. Li et al. [32] also conducted the following experiments: (1) time domain and spectral analysis based on empirical pattern decomposition, using support vector machine classifier pairs for classification; (2) continuous wavelets transform combined with a greyscale co-generation matrix of the feature extraction method using SVM classifier for classification; (3) Retraining the first and last layers of ResNet18 to identify epileptic EEG signals. We compared our method with these approaches, as shown in Table 4. Our method achieved an accuracy of 94.47%, outperforming the best baseline by 2.47%. Additionally, it achieved the highest F1-score of 0.9438 among all other methods. To provide a clearer demonstration of the superiority of our algorithm, we conducted a critical difference analysis, as illustrated in Fig. 8. The critical difference analysis clearly shows that our method outperforms the other approaches significantly. In conclusion, our method achieved remarkable results, with a substantial improvement in accuracy (2.47%) compared to the best baseline. Moreover, it demonstrated the highest F1-score, highlighting its superiority over the alternative methods.

5. Conclusion

This paper builds a three-dimensional convolutional model based on channel and spatial attention, which is suitable for the seizure-type classification task. It takes advantage of multi-stage wavelet decomposition to obtain multi-band time–frequency information and constructs time–frequency-spatial domain 3D features according to the electrode distribution. Furthermore, the three-dimensional convolutional neural network containing channel and spatial attention is used to extract joint activity information of spatial domain signals and convolutional filtering of time–frequency domain features, where the addition of the channel and spatial attention mechanism generates global and local information embedding vectors giving the network the ability to emphasize important discriminative information selectively. Experimental results on the TUSZ dataset demonstrate that the proposed model achieves state-of-the-art results in the epileptic seizure type classification task. In the future, we plan to investigate the use of better feature extraction methods that can detect more fine-grained differences between EEG seizures.

Table 4
Comparison of algorithms with advanced algorithms.

Authors	Methods	Seizure types	ACC (%)	F1-score
Saputro et al. [17]	MFCC+Hjorth+SVM	3	91.40	/
Roy et al. [20]	FFT+KNN FFT+SGD FFT+XGBoost FFT+Adaboost FFT+ResNet50	7	/ / / / /	0.9070 0.7780 0.8440 0.7070 0.7230
Cao et al. [31]	SENet+LSTM	4 (Normal+3)	Average: 87.50	/
Wang et al. [37]	NLP+SVM NLP+RF	6	Average: 88.67 Average: 90.67	/ /
Raghu et al. [29]	AlexNet VGG16 VGG19 basic CNN	8	84.06 79.71 76.81 82.14	/ / / /
Li et al. [32]	EMD+SVM CWT+SVM ResNet18 Deep ConvNet CE-stSENet	7	85.69 90.89 79.58 88.03 92.00	0.8685 0.9190 0.7536 0.8940 0.9369
Jia et al. [38]	VWCNNs	7	91.71	0.9400
Zhang et al. [39]	VMD+IMFs+NLTWSVM	7	92.29	0.9230
This work	**3D-CBAMNet**	7	**94.47**	**0.9438**

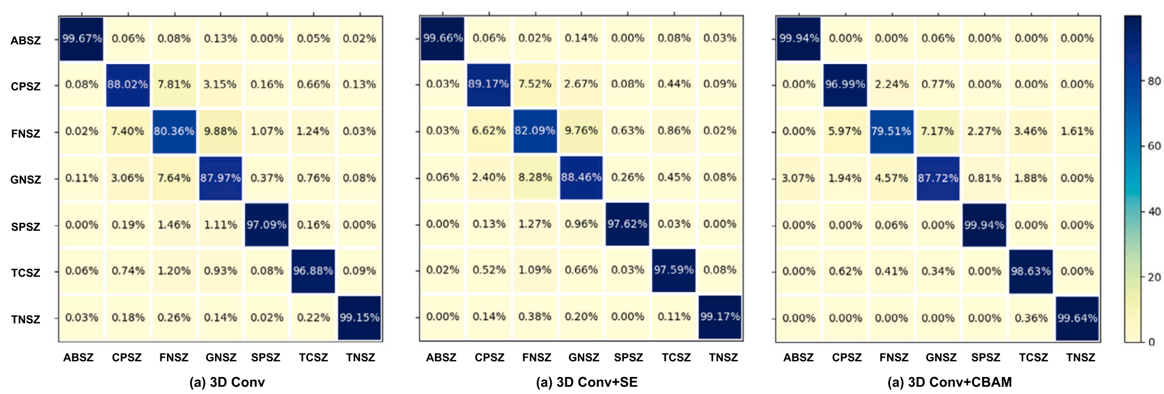

Fig. 7. Confusion matrix under different attention: (a) 3D Conv, (b) 3D Conv+SE, and (c) 3D Conv+CBAM.

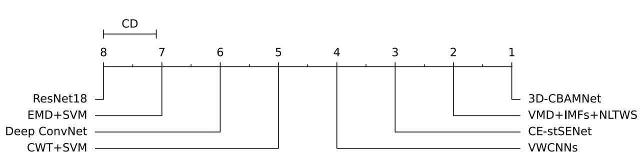

Fig. 8. Critical difference among 7 epileptic seizure type classification algorithms.

Declaration of competing interest

The authors declare that there is no conflict of interests regarding the publication of article.

References

[1] E. Perucca, The pharmacological treatment of epilepsy: recent advances and future perspectives, Acta Epileptol. 3 (2021) 22.
[2] D. Wu, J. Li, F. Dong, J. Liu, L. Jiang, J. Cao, X. Wu, X. Zhang, Classification of seizure types based on multi-class specific bands common spatial pattern and penalized ensemble model, Biomed. Signal Process. Control 79 (2023) 104118.
[3] R.K. Joshi, V. Kumar, M. Agrawal, A. Rao, L. Mohan, M. Jayachandra, H.J. Pandya, Spatiotemporal analysis of interictal EEG for automated seizure detection and classification, Biomed. Signal Process. Control 79 (2023) 104086.
[4] S.M. Usman, S. Khalid, S. Bashir, A deep learning based ensemble learning method for epileptic seizure prediction, Comput. Biol. Med. 136 (2021) 104710.
[5] H. Peng, C. Lei, S. Zheng, C. Zhao, C. Wu, J. Sun, B. Hu, Automatic epileptic seizure detection via stein kernel-based sparse representation, Comput. Biol. Med. 132 (2021) 104338.
[6] I. Megiddo, A. Colson, D. Chisholm, T. Dua, A. Nandi, R. Laxminarayan, Health and economic benefits of public financing of epilepsy treatment in India: An agent-based simulation model, Epilepsia 57 (3) (2016) 464–474.
[7] A. Hamlin, E. Kobylarz, J.H. Lever, S. Taylor, L. Ray, Assessing the feasibility of detecting epileptic seizures using non-cerebral sensor data, Comput. Biol. Med. 130 (2021) 104232.
[8] Z. Wu, S. Shen, X. Lian, X. Su, E. Chen, A dummy-based user privacy protection approach for text information retrieval, Knowl.-Based Syst. 195 (2020) 105679.
[9] Z. Wu, S. Xuan, J. Xie, C. Lin, C. Lu, How to ensure the confidentiality of electronic medical records on the cloud: A technical perspective, Comput. Biol. Med. 147 (2022) 105726.
[10] Z. Wu, G. Li, S. Shen, X. Lian, G. Xu, Constructing dummy query sequences to protect location privacy and query privacy in location-based services, World Wide Web 24 (1) (2021) 105726.
[11] H. Albaqami, G.M. Hassan, A. Datta, MP-SeizNet: A multi-path CNN bi-LSTM network for seizure-type classification using EEG, Biomed. Signal Process. Control 84 (2023) 104780.
[12] T.K. Gandhi, P. Chakraborty, G.G. Roy, B.K. Panigrahi, Discrete harmony search based expert model for epileptic seizure detection in electroencephalography, Expert Syst. Appl. 39 (4) (2012) 4055–4062.
[13] L. Guo, D. Rivero, A. Pazos, Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks, J. Neurosci. Methods 193 (1) (2010) 156–163.
[14] O. Faust, R.U. Acharya, A.R. Allen, C.M. Lin, Analysis of EEG signals during epileptic and alcoholic states using ar modeling techniques, Irbm 29 (1) (2008) 44–52.

[15] M. Shen, P. Wen, B. Song, Y. Li, An EEG based real-time epilepsy seizure detection approach using discrete wavelet transform and machine learning methods, Biomed. Signal Process. Control 11 (2022) 103820.
[16] B. Maimaiti, H. Meng, Y. Lv, J. Qiu, Z. Zhu, Y. Xie, Y. Li, C. Yu, W. Zhao, J. Liu, M. Li, An overview of eeg-based machine learning methods in seizure prediction and opportunities for neurologists in this field, Neuroscience 481 (2022) 197–218.
[17] I.R.D. Saputro, N.D. Maryati, S.R. Solihati, I. Wijayanto, S. Hadiyoso, R. Patmasari, Seizure type classification on EEG signal using support vector machine, J. Phys. Conf. Ser. (2019) 012065.
[18] K.T. Tapani, P. Nevalainen, S. Vanhatalo, N.J. Stevenson, Validating an SVM-based neonatal seizure detection algorithm for generalizability, non-inferiority and clinical efficacy, Comput. Biol. Med. 145 (2022) 105399.
[19] Y. Gao, A. Liu, X. Cui, R. Qian, X. Chen, A general sample-weighted framework for epileptic seizure prediction, Comput. Biol. Med. 150 (2022) 106169.
[20] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T.H. Falk, J. Faubert, Deep learning-based electroencephalography analysis: a systematic review, J. Neural Eng. 16 (5) (2019) 051001.
[21] C.-f. Chen, Z.-j. Du, L. He, Y.-j. Shi, J.-q. Wang, W. Dong, A novel gait pattern recognition method based on LSTM-CNN for lower limb exoskeleton, J. Bionic Eng. 18 (2021) 1059–1072.
[22] J. Wang, D. Wu, Y. Gao, X. Wang, X. Li, G. Xu, W. Dong, Integral real-time locomotion mode recognition based on GA-CNN for lower limb exoskeleton, J. Bionic Eng. 19 (5) (2022) 1359–1373.
[23] M. Zhu, X. Guan, Z. Li, L. He, Z. Wang, K. Cai, sEMG-based lower limb motion prediction using CNN-LSTM with improved PCA optimization algorithm, J. Bionic Eng. 20 (2) (2023) 612–627.
[24] A. Shoeibi, P. Moridian, M. Khodatars, N. Ghassemi, M. Jafari, R. Alizadehsani, Y. Kong, J.M. Gorriz, J. Ramírez, A. Khosravi, et al., An overview of deep learning techniques for epileptic seizures detection and prediction based on neuroimaging modalities: Methods, challenges, and future works, Comput. Biol. Med. (2022) 106053.
[25] M.B. Qureshi, M. Afzaal, M.S. Qureshi, M. Fayaz, Machine learning-based EEG signals classification model for epileptic seizure detection, Multimedia Tools Appl. 80 (2021) 17849–17877.
[26] E. Tuncer, E.D. Bolat, Channel based epilepsy seizure type detection from electroencephalography (EEG) signals with machine learning techniques, Biocybern. Biomed. Eng. 42 (2) (2022) 575–595.
[27] Z. Wu, S. Shen, H. Zhou, H. Li, C. Lu, D. Zou, An effective approach for the protection of user commodity viewing privacy in e-commerce website, Knowl.-Based Syst. 220 (2021) 106952.
[28] Z. Wu, J. Xie, S. Shen, C. Lin, G. Xu, E. Chen, A confusion method for the protection of user topic privacy in Chinese keyword-based book retrieval, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22 (5) (2023) 1–19.
[29] N. Sriraam, Y. Temel, S.V. Rao, P.L. Kubben, A convolutional neural network based framework for classification of seizure types, in: Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society, 2019, pp. 2547–2550.
[30] I.C. Covert, B. Krishnan, I. Najm, J. Zhan, M. Shore, J. Hixson, M.J. Po, Temporal graph convolutional networks for automatic seizure detection, in: Proceedings of Machine Learning for Healthcare Conference, 2019, pp. 160–180.
[31] J. Cao, J. Zhu, W. Hu, A. Kummert, Epileptic signal classification with deep EEG features by stacked CNNs, IEEE Trans. Cogn. Dev. Syst. 12 (4) (2020) 709–722.
[32] Y. Li, Y. Liu, W.G. Cui, Y.Z. Guo, H. Huang, Z.Y. Hu, Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (4) (2020) 782–794.
[33] Q. Xin, S. Hu, S. Liu, L. Zhao, Y.D. Zhang, An attention-based wavelet convolution neural network for epilepsy EEG classification, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 957–966.
[34] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[35] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 3–19.
[36] V. Shah, E. Von Weltin, S. Lopez, J.R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, J. Picone, The temple university hospital seizure detection corpus, Front. Neuroinform. 12 (2018) 83.
[37] Z. Wang, P. Mengoni, Frequency bands selection for seizure classification and forecasting using NLP, random forest and SVM models, in: Proceddings of International Conference on Brain Informatics, 2021, pp. 310–320.
[38] G. Jia, H.-K. Lam, K. Althoefer, Variable weight algorithm for convolutional neural networks and its applications to classification of seizure phases and types, Pattern Recognit. 121 (2022) 108226.
[39] S. Zhang, G. Liu, R. Xiao, W. Cui, J. Cai, X. Hu, Y. Sun, J. Qiu, Y. Qi, A combination of statistical parameters for epileptic seizure detection and classification using VMD and NLTWSVM, Biocybern. Biomed. Eng. 42 (1) (2022) 258–272.