Heliyon 9 (2023) e14858

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A novel method for diagnosing Alzheimer's disease using deep pyramid CNN based on EEG signals

Wei Xia, Ran Zhang *, Xiao Zhang **, Muhammad Usman

School of Medical Informatics and Engineering, Xuzhou Medical University, Xuzhou 221000, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Electroencephalography
Deep learning
Deep pyramid convolutional neural network
Mild cognitive impairment

ABSTRACT

Background: The diagnosis of Alzheimer's disease (AD) using electroencephalography (EEG) has garnered more attention recently.
New methods: In this paper, we present a novel approach for the diagnosis of AD, in terms of classifying the resting-state EEG of AD, mild cognitive impairment (MCI), and healthy control (HC). To overcome the hurdles of limited data available and the over-fitting problem of the deep learning models, we studied overlapping sliding windows to augment the one-dimensional EEG data of 100 subjects (including 49 AD subjects, 37 MCI subjects and 14 HC subjects). After constructing the appropriate dataset, the modified DPCNN was used to classify the augmented EEG. Furthermore, the model performance was evaluated by 5 times of 5-fold cross-validation and the confusion matrix has been obtained.
Results: The average accuracy rate of the model for classifying AD, MCI, and HC is 97.10%, and the F1 score of the three-class classification model is 97.11%, which further proves the model's excellent performance.
Conclusions: Therefore, the DPCNN proposed in this paper can accurately classify the one-dimensional EEG of AD and is worthy of reference for the diagnosis of the disease.

1. Introduction

Alzheimer's disease (AD) is a common neurodegenerative disease among the elderly, affecting about 51.6 million people worldwide currently [1]. The number is expected to reach 132 million by the mid-21st century [2]. From 2000 to 2018, the number of deaths caused by AD increased by 146.2% [3], and AD has become a major threat to the life and health of the people over 65 years old. At the same time, the total cost of AD worldwide is as high as 1 trillion US dollars, which is expected to rise to 2 trillion US dollars in 2030 [4]. This hugely burdens both the families of patients and the finance of countries.

In its early clinical stages, AD is known as mild cognitive impairment (MCI), and patients in this stage show symptoms consistent with normal ageing, except for certain deficits in the cognitive domain. Trying to accurately diagnose AD, especially at the MCI stage, requires histological analysis of the brain or other complex tests, which can be extremely difficult for elderly patients.

The main pathological factors of AD are the deposition of amyloid β in human brain and the neuronal degeneration caused by protein tau phosphorylation [5]. The former can be measured by positron emission tomography and the latter by cerebrospinal fluid examination. Imaging analysis and cerebrospinal fluid examination are the current gold standard for the screening and diagnosis of AD

* Corresponding author.
** Corresponding author.
E-mail addresses: zhangran@xzhmu.edu.cn (R. Zhang), zhangxiao@xzhmu.edu.cn (X. Zhang).

https://doi.org/10.1016/j.heliyon.2023.e14858
Received 6 September 2022; Received in revised form 7 March 2023; Accepted 20 March 2023
Available online 24 March 2023

[6]. However, the purchase price and maintenance cost of imaging equipment are usually millions, resulting in high cost of imaging examination, while cerebrospinal fluid examination is an invasive examination with certain risks. Both have limitations in the front line of AD diagnosis and screening. At present, neurologists assess the cognitive status based on mini-mental state examination score and the Montreal cognitive test, which are the most common clinical screening method [7]. Nevertheless, this method is greatly influenced by the subjective effect of clinicians and is prone to misdiagnosis and missed diagnosis.

Electroencephalography (EEG), a real-time recording technology of the electrical activities inside the brain, is widely used in the diagnosis of mental disorders. As early as 2005, Kannathal et al. used a computer to carry out feature extraction and data analysis on EEG to assist in the diagnosis of epilepsy [8]. Chen et al. proposed a recognition framework for attention deficit and hyperactivity disorder based on the combination of EEG brain network and convolutional neural network, and achieved a recognition accuracy rate of 94.67% [9]. Cai et al. established a multi-modal model based on EEG data under different audio stimuli to distinguish depression patients [10]. A large number of studies have shown that EEG has the advantages of simple collection, low cost, objective recording data and quantification, which make it a suitable tool for the analysis of mental diseases [11].

Evidence suggests that AD is closely related to the deficiency of synapses in the brain [12]. EEG, a widely used noninvasive method to reflect synaptic activities in human brain, can be used as a potential marker for the assessment of AD [13]. However, EEG is complicated, changeable and easy to be disturbed by noise. It is difficult to extract valuable information to diagnose AD from the original EEG data by ordinary mathematical statistics methods. With the development of artificial intelligence technology in the 21st century, more and more researchers are applying this technology to EEG analysis. Machine learning is a subset of artificial intelligence, among which many classical algorithms are valuable for diagnosing AD. Acharya et al. extracted approximate entropy and sample entropy from EEG as input data of support vector machine (SVM) classifier, successfully achieving an accuracy rate of 91.70% in the classification of AD patients [14]. Tavares et al. took the power spectral density (PSD) of different frequency bands of EEG as the input features of eight different machine learning algorithms and achieved an average classification accuracy rate of 95.60% [15]. By using machine learning algorithms to analyze EEG to diagnose AD, the screening time and human intervention are reduced. As a result, the diagnostic results are more efficient and objective.

Classical machine learning algorithms usually require complex feature engineering (FE) to achieve good performance, which means that researchers need to spend plenty of time to explore the original EEG and select appropriate features. As a subset of machine learning, deep learning utilizes neural network and hidden layer to realize the automation of FE, greatly saving researchers' time and energy [16]. In addition, artificially selected features have the risk of losing key information of EEG data, but the deep learning algorithm can make more efficient use of EEG data and improve the robustness of the algorithm. In recent years, using the deep learning algorithm to analyze EEG to diagnose AD is a promising field.

The EEG data used in this study were published on line by the Fiscon team [17]. Firstly, the EEG data were augmented. Secondly, based on the deep pyramid convolutional neural network (DPCNN) [18], a one-dimensional convolutional neural network model suitable for the three-class classification of EEG of AD, MCI and healthy control (HC) was proposed. Finally, confusion matrices were used to evaluate the performance of the model.

The remaining sections of this paper are as follows: Relevant researches in the same field are introduced in Section 2. The EEG data, data augmentation methods and the deep learning model architecture used in this study are described in Section 3. The research results are presented in Section 4, and the results are discussed in Section 5. In Section 6, conclusions and prospects to the future research are given.

2. Literature review

Recognizing EEG of AD, MCI and HC is a matter of classification. This section details the latest techniques for diagnosing AD through the combination of EEG and deep learning. These techniques can be divided into two-class classification and three-class classification, and the latter has been the focus of research in recent years.

Ferri et al. [19] collected the resting-state EEG signals from 89 AD subjects and 10 HC subjects, and preprocessed them by using Fast Fourier Transform (FFT) and eLORETA source estimation. They extracted spectral power density calculated by FFT and neural current density by LORETA as the input features of the neural network, gaining a two-class classification accuracy rate of 89.00% on the stacked ANN with autocoder. Although this study innovatively used neural current density as a classification feature, the classification accuracy rate is relatively low.

Duan et al. [20] collected two groups of resting-state EEG signals, i.e., 22 MCI subjects and 38 HC subjects in Group 1 and 17 AD subjects and 24 HC subjects in Group 2 to achieve two two-class classifications of MCI vs. HC and AD vs. HC. Both groups of EEG data were preprocessed with the same band-pass filter. The functional connectivity of EEG was extracted as the input to CNN by calculating the coherence between two electrodes. In their paper, the recognition accuracy rate of CNN model ResNet-18 is 93.42% for MCI and 98.54% for AD. This study is the first to consider the difference of spatial attribute of EEG in AD classification and achieve excellent classification performance, but it lacks direct comparative analysis of AD vs MCI and the challenge of accurately diagnosing MCI remains unresolved.

Two-class classification is less complex than three-class classification, but the latter is more significant in research. MCI is an intermediate stage between HC and AD. At present, it is uncertain whether MCI will definitely develop into AD [21], but the diagnosis of MCI can help to prevent and treat AD. Therefore, it has more clinical application value.

Ieracitano et al. [22] collected EEG signals of 63 AD subjects, 63 MCI subjects and 63 HC subjects, and used two-dimensional grayscale PSD images of EEG as classification features to construct a CNN-based three-class classification model. Although this model is superior to traditional machine learning techniques such as SVM and MLP, its accuracy rate is only 83.33%. Therefore, the

study considered using the average PSD image of the entire EEG as a classification feature in the future to reduce the effect of epoch variability. It will also investigate mapping spectral feature representations that are more sensitive to AD and MCI EEG features to improve the accuracy of the classification model.

You et al. [23] proposed a cascade neural network which used the sequential characteristics of EEG and the human gait characteristics to classify AD, MCI and HC, with a classification accuracy rate of 91.70%. This attention-based spatiotemporal graph convolutional network can automatically extract the features of EEG and gait data, reducing human intervention and achieving good classification results. Also, the study used only gait data when differentiating between HC and patients, eliminating the need to use EEG data from HC and saving a lot of data collection time. Although this study provides a new way of thinking about which data to collect to diagnose AD, it is difficult to reproduce the EEG and gait fusion data used when subdividing patients into AD and MCI.

Rad et al. [24] extracted different bands of frequency characteristics, statistical properties, Lyapunov index of task-state EEG, and selected optimal features that had the greatest impact on AD, MCI and HC classifications with variance analysis. The multi-channel deep convolutional neural network constructed by the author reached a three-class classification accuracy rate of 97.50%. This study presents the first analysis method based on the four task states of closed-eye, open-eye, recall and stimulation EEG, opening up a new way of using task-state EEG to diagnose AD. However, the EEG was collected in three self-defined electrode positions (Pz, Cz and Fz), rather than 19 or 64 channels in the 10–20 international system. The three-channel recording of human brain electrical activities is not comprehensive, and the data used in this paper cannot represent the real situation of AD and MCI patients, so the experimental results are somewhat fortuitous.

A paper published by Amini et al. [25] used the time-dependent power spectrum descriptor to extract EEG features of each channel as CNN input. The accuracy rate is 82.30% in the data set consisting of the resting-state EEG of 64 AD subjects, 64 MCI subjects and 64 HC subjects. The TD-PSD method well extracts time-domain and frequency-domain features from the 180 s EEG signals at each epoch (one epoch every 300 s, from 60 to 240 s), minimizing EEG background artifacts. However, the paper did not improve or innovate in CNN. Instead, it only proved that CNN has obvious advantages compared with SVM and LDA in this data set. Moreover, the accuracy rate of the algorithm used in the paper is low, especially for HC (only 75%).

Huggins et al. [26] recently published a paper about the AlexNet-based classification of EEG in dementia patients. In this paper, continuous wavelet transform was used to transform EEG into time frequency graph which was input into the deep learning model. The three-class classification accuracy rate of the model is up to 98.90%. The method proposed in this paper is excellent, but the process of converting one-dimensional EEG into a three-dimensional RGB image is complicated and there is too much human intervention in the research process. Therefore, the research team in this paper is still exploring alternative methods of creating RGB images from EEG signals, and in the future may combine other signals to construct multimodal data to quantify features associated with AD.

To sum up, the current use of EEG combined with deep learning technology for AD diagnosis is mostly to convert one-dimensional EEG into high-dimensional data for discrimination, but the one-dimensional deep learning model is rarely used directly. This not only

Table 1
Summary of the use of EEG combined with the deep learning technology for AD diagnosis in recent years.

Author	Time	Size of data set	Feature extraction method	Deep learning model	Result
Ferri et al.	2020	89AD 10HC	Extract the features by FFT and eLORETA source analysis	ANN	AD vs. HC = 89.00%
Duan et al.	2020	22MCI 38HC 17AD 24HC	Calculate the EEG coherence between channels and extract the EEG functional connectivity as the feature	CNN	MCI vs. HC = 93.42% AD vs. HC = 98.54%
Ieracitano et al.	2019	63AD 63MCI 63data	Extract the two-dimensional gray PSD image of EEG	CNN	AD vs. MCI vs. HC = 83.33%
You et al.	2020	35AD 35MCI 17HC	Extract the fusion features of EEG and gait data	GCN	AD vs. MCI vs. HC = 91.70%
\Rad et al.	2021	10AD 11MCI 19HC	Extract the optimal classification features from EEG frequency features and statistical attributes through the analysis on variance	CNN	AD vs. MCI vs. HC = 97.50%
Amini et al.	2021	64AD 64MCI 64HC	Extract the EEG features of each channel by D-PSD	CNN	AD vs. MCI vs. HC = 82.30%
Huggins et al.	2021	52AD 37MCI 52HC	Extract the RGB image by CWT	CNN	AD vs. MCI vs. HC = 98.90%

AD = Alzheimer's Disease, MCI = Mild Cognitive Impairment, HC=Healthy Control, FFT=Fast Fourier Transform, EEG = Electroencephalogram, LORETA = Low-Resolution Brain Electromagnetic Tomography, ANN = Artificial Neural Network, CNN=Convolutional Neural Network, CWT=Continuous Wavelet Transform, TD-PSD=Time-Dependent Power Spectrum Descriptor, RGB = Red, Green, Blue.

makes data preprocessing more difficult, but also increases the human intervention in the research process, reducing the reproducibility of research results. The key information of each paper above is listed in Table 1.

3. Materials and methods

3.1. Research design

The overall process of this study is shown in Fig. 1. Firstly, with EEG being preprocessed by FFT, Fourier coefficients were extracted as EEG frequency domain features. Subsequently, the data were augmented by using overlapping sliding window segmentation of EEG frequency domain features. Afterwards, the data set was randomly divided into the training set and the test set at a ratio of 4:1. Next, the model was validated with 5-fold cross validation, and finally, evaluated by the confusion matrix.

3.2. EEG data set

The data used in this study are open-source data posted online by Fiscon et al. [17] (https://github.com/tsyoshihara/Alzheimer-s-Classification-EEG/tree/master/data). The data set is composed of the EEG from 100 subjects, including 49 AD subjects, 37 MCI subjects and 14 HC subjects. As human cognitive ability naturally degrades with age, older HC subjects and younger AD subjects are prone to misclassification [27]. For this reason, the study ensured that AD and MCI subjects were older than HC subjects to rule out the effect of age on classification results. Besides, the study took into account the influence of subjects' gender, but the two-tailed test proved that gender differences and clinical status are not statistically significant [17]. Detailed information of subjects is shown in Table 2.

According to the international 10–20 system, 19 electrode positions were selected on the scalp (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1 and O2), and 300 s of the resting-state EEG with eyes closing were collected by using the monopolar connections with earlobe electrode landmark, with the sampling frequency being 1024 Hz.

In order to reduce the artifacts, the EEG signal were preprocessed. The 19-channel EEG signals were first band-pass filtered from 0.5 to 48 Hz using the EEGLAB toolbox in Matlab, and then the average reference means commonly used in resting-state EEG is used to re-reference the signal to reduce the interference of electrical and power frequency noise. Next, the EEG signals were downsampled to 256 Hz and selected from 60s to 240s. Finally, the independent principal component analysis (ICA) method in the EEGLAB toolbox was used to remove oculoelectric and electromyographic artifacts from the EEG signals. The average EEG length after artifact rejection was 3 min.

Analyzing frequency domain features of EEG is one of the best methods to identify brain activities [28]. FFT was applied to each sample to convert one-dimensional EEG to the frequency domain. The formula is as follows (Equation (1)):

$$\begin{cases} X(k) = X_1(k) + W_N^k X_2(k) \\ X\left(k + \frac{N}{2}\right) = X_1(k) - W_N^k X_2(k) \end{cases} \tag{1}$$

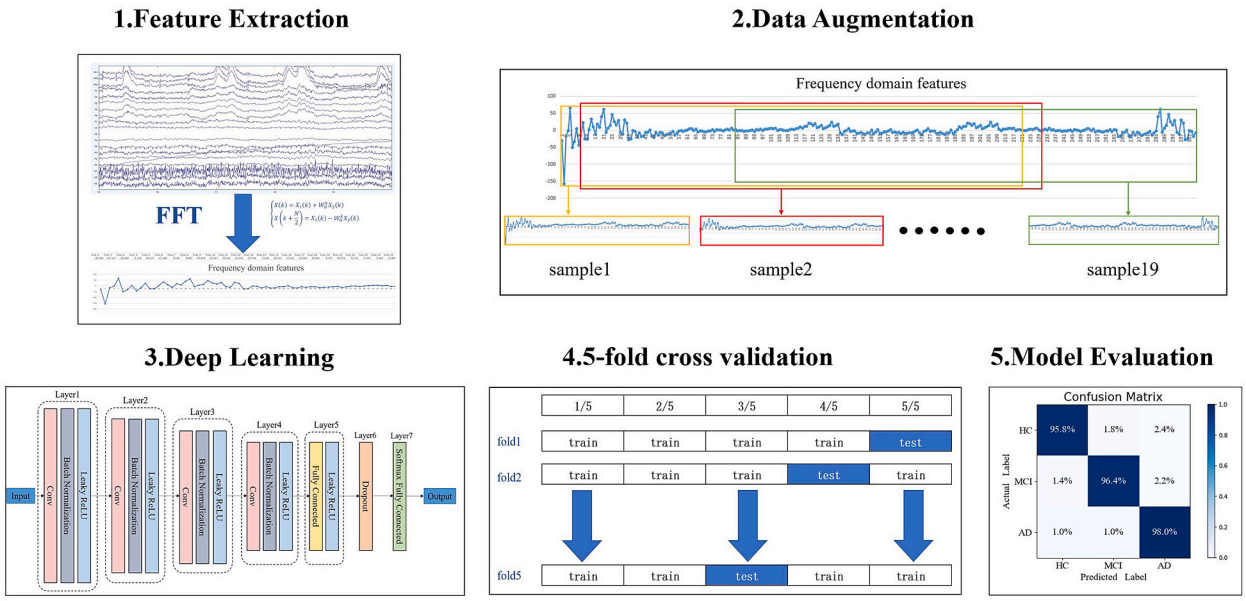

Fig. 1. Overall flow of the study.

Table 2
Subjects' information.

Type	Number	Male	Female	Age (Mean ± standard deviation)
AD	49	20	29	78.4 ± 6.4
MCI	37	17	20	74.1 ± 9.4
HC	14	9	5	64.1 ± 9.4

where N = 2 M, and M is a natural number; k = 0,1, …, $\frac{N}{2}$-1; $W_N = e^{-\frac{j2\pi}{N}}$; X is the frequency domain representation of time-series signals, as follows (Equation (2)):

$$\begin{cases} X_1(k) = \sum_{n=0}^{\frac{N}{2}-1} x_1(n) W_{\frac{N}{2}}^{kn} \\ X_2(k) = \sum_{n=0}^{\frac{N}{2}-1} x_2(n) W_{\frac{N}{2}}^{kn} \end{cases} \tag{2}$$

The signal x(n) is decomposed into $x_1(n)$ and $x_2(n)$ by the parity of n; By means of FFT, 16 Fourier coefficients were selected as frequency domain features of EEG in each channel, and the number of features for each subject was 19 × 16 = 304. The data set obtained is composed of 30,400 frequency domain features, including 14,896 AD features, 11,248 MCI features and 4256 HC features. One of the subject's frequency domain features of all channels is presented in Fig. 2.

3.3. Data augmentation

Deep learning model requires a large amount of data in the training process, but the EEG data of AD subjects are difficult to collect and the data set is small in size [29]. This results in poor generalization ability and obvious overfitting of the trained deep learning model. To solve this problem, a data augmentation scheme was introduced for the frequency domain features of each EEG signal.

Lashgari et al. used a variety of data augmentation methods for EEG, such as sliding window, noise addition, segmentation and reorganization. The results suggest that compared with other data augmentation methods in various EEG tasks, sliding window boasts the optimal effect [30]. In this study, overlapping sliding windows were used to segment the frequency domain features of each EEG signal. Assuming that the window size is m and the sliding step length is n, the frequency domain features with the size of 304 are divided into (304-m)/n sub-features with equal length. We took each sub-feature as an independent data sample to form a new data set to train the deep learning model.

Based on the window size and step length, a data augmentation scheme suitable for the EEG frequency domain features of AD was presented. In the research process, we set the window size to 25%, 50% and 75% of the original feature size, and the sliding step length to 22, 23, 24 and 25 respectively for research. Taking the window size of 228 (sub-sample accounts for 75% of the original sample) and the sliding step length of 22 as an example, the sample instance of the new data set was 1900% of the original data set. The detailed data augmentation effect is listed in Table 3.

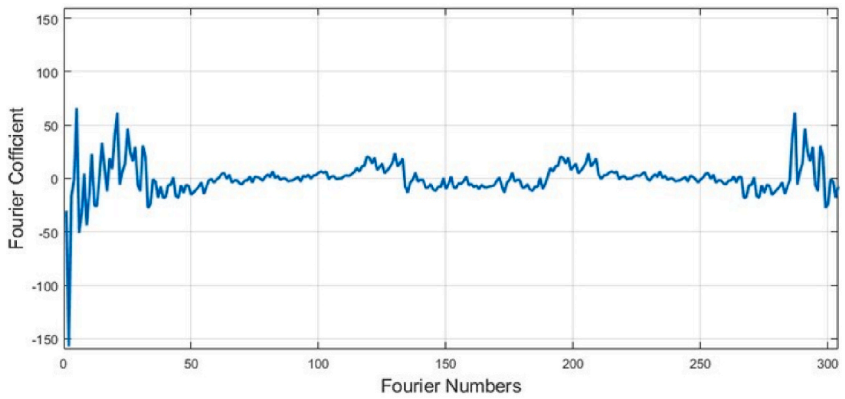

Fig. 2. One subject's frequency domain features of all channels.

3.4. Deep learning model

As the convolutional layer of CNN can efficiently extract features from input data, it has been widely and successfully applied in some fields including image recognition, natural language processing, etc. [31].

The original EEG, which belongs to one-dimensional time-series data, are *trans*-formed into numerical data reflecting the frequency domain features of EEG after FFT and feature extraction. Traditional CNN has fewer bottom-layer convolution kernels but more high-layer convolution kernels [32]. Such structure is prone to gradient explosion and over-fitting when dealing with discrete one-dimensional numerical data. Based on the DPCNN proposed by Johnson et al. [18], a deep learning model suitable for one-dimensional EEG feature classification of AD is proposed. It consists of 4 convolutional layers, 2 fully connected layers and 1 dropout layer. The model structure is given in Fig. 3. In order to reduce the learnable parameters and lower the risk of overfitting of the model, the number of convolutional kernels in each convolutional layer decreases with the increase of the number of layers (24 in the first layer, 16 in the second layer, 8 in the third layer and 4 in the fourth layer). In the training process of the model, various parameters of the model were constantly adjusted, and the specific situation is as follows:

- The optimizer. It determines the updating mode of model weight during training and plays a decisive role in the performance of the deep learning model. In the research process, we did a comparative study of SGD, Adagrad, PMSProp and Adam optimizers. Although the training of the model using SGD is fast, its convergence is slow due to the large variation of parameter iteration direction [33]. Adagrad is the earliest adaptive optimizer, with which the learning rate of the model will decrease progressively to 0 in a monotonous way, and the model training process tends to end in advance [34]. RMSProp calculates the second-order momentum by the exponential weighted-average method, which solves the monotonically decreasing problem of learning rate in Adagrad. Adam is a further improvement of RMSprop [35]. Hence, among the four optimizers, Adam was finally chosen.
- Activation function. It is the core of neural networks, which performs nonlinear transformation on input to enable it to execute complex tasks [36]. Three activation functions, TANh, ReLU and Leaky ReLU, were tested, among which Leaky ReLU was the best. The Leaky ReLU activation function, which corresponds to high computational efficiency, is able to help the deep learning model to converge quickly. When the input value is negative, this function still has a small gradient, which solves the problem of gradient disappearance facing the use of the ReLU activation function. The Leaky ReLU activation function was finally chosen in the model.
- Learning rate. As one of the most influential parameters, learning rate determines the learning efficiency of the model weight [26]. Multiple tests between 1×10^{-5} and 1×10^{-3} were performed, and finally, the optimal performance of the model was with the learning rate set to 2×10^{-5}.
- Batch size. It determines the sample size of each iteration during the training process. Too small batch size leads to reduced model training efficiency and non-convergence within a certain number of epochs. Too large batch size leads to poor model generalization ability and poor performance of trained models on the test set. The performances of models with batch sizes of 8, 16, 32 and 64 were compared, among which 8 and 16 did well. Considering the operation efficiency of the model, finally the batch size of 16 was chosen.

The model we used ran each fold for about 11 min with an RTX 3060 GPU.

4. Result

In this study, transformed each subject's the 180-s EEG signals of 19 channels of all the subjects were transformed to the frequency domain with FFT, from which 304 frequency domain features were extracted as the input of the deep learning model. After data augmentation for frequency domain features of all subjects, the data were randomly divided into 5 pieces of equal length, 4 of which were used as training set and 1 as test set in turn. The accurately divided data were input into the deep learning model for 5-fold cross validation and the model performance was evaluated by confusion matrix. The specific quantitative indexes of the model's classification performance are as follows [37]:

Table 3
Parameter setting and augmentation effect of the sliding window.

Window proportion	Window size	Step size	Augmentation effect
25%	76	4	5700%
		8	2900%
		16	1500%
		32	800%
50%	152	4	3800%
		8	1900%
		16	1000%
		32	500%
75%	228	4	1900%
		8	1000%
		16	500%
		32	300%

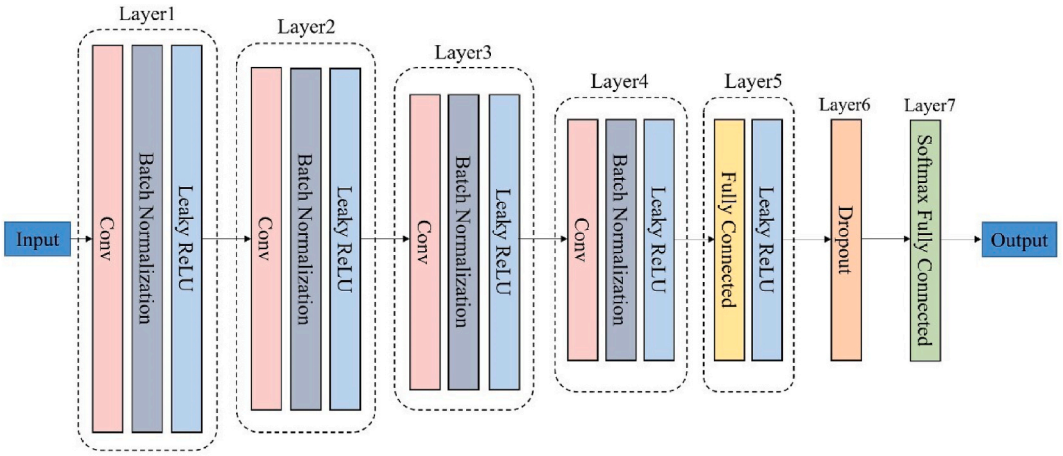

Fig. 3. Structure diagram of the deep learning model suitable for one-dimensional frequency domain feature classification of AD.

The ratio of actual true prediction to the total amount of data determines accuracy (Equation (3)).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

The ratio of actual positive prediction to the total positive forecast determines precision (Equation (4)).

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

The ratio of actual positive prediction to the positively labelled data determines recall (Equation (5)).

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F1-score (Equation (6)) indicates the harmonic mean of the two performance parameters precision and recall.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{6}$$

When AD is positive and MCI and HC are negative, TP = true positive, that is, when the sample is predicted to be positive, the sample is actually positive. For example, when AD is predicted to be AD, TN = true negative, that is, when the sample is predicted to be negative, the sample is actually negative. For example, when MCI is predicted to be MCI, HC is predicted to be HC, FP = false positive, that is, when the sample is predicted to be positive, the sample is actually negative. For example, when MCI and HC are predicted to be AD, FN = false negative, that is, when the sample is predicted to be negative, the sample is actually positive. For example, AD is

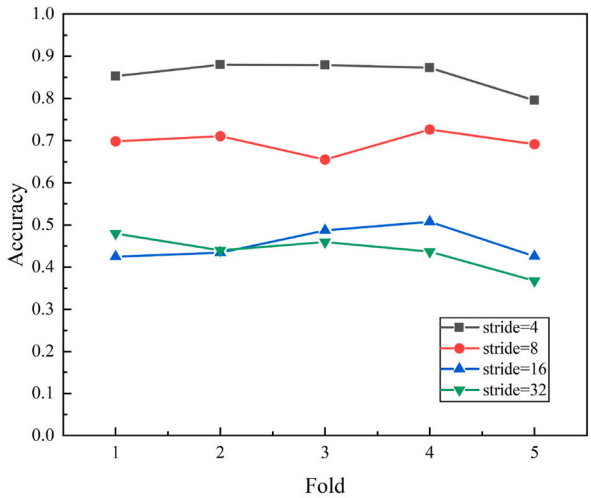

Fig. 4. Performance of the model (with different step lengths but a fixed window size of 76) using the data augmentation scheme.

predicted to be MCI and HC.

4.1. Results of the different data augmentation schemes

When the research data are augmented with a sliding window of 76 in size, the average accuracy rates of 5-fold cross validation for the scheme with a step length of 4, 8, 16 and 32 are 85.61%, 69.63%, 45.60% and 43.69%, respectively. The performance of the scheme with different step lengths but a fixed window size of 76 is displayed in Fig. 4.

When the research data are augmented with a sliding window of 152 in size, the average accuracy rates of 5-fold cross validation for the scheme with a step length of 4, 8, 16 and 32 are 95.87%, 93.29%, 74.13% and 57.14%, respectively. The performance of the scheme with different step lengths but a fixed window size of 152 is displayed in Fig. 5.

When the research data are augmented with a sliding window of 228 in size, the average accuracy rates of 5-fold cross validation for the scheme with a step length of 4, 8, 16 and 32 are 97.17%, 87.16%, 59.82% and 49.54%, respectively. The performance of the scheme with different step lengths but a fixed window size of 228 is illustrated in Fig. 6.

Compared with the above 12 data augmentation schemes, the proposed deep learning model achieves the best performance when the window size is 228 and the step length is 4. The average loss and average accuracy rate of 5-fold cross validation of the model are 0.097 and 97.16%. The specific results of each fold of the 12 augmentation schemes are exhibited in Table 4.

4.2. Results of the confusion matrix evaluation

After the application of the optimal data augmentation scheme, in order to reduce the impacts of fortuity and experimental error on the research results, this study carries out five 5-fold cross validations, and a confusion matrix is generated each time. The final result is composed of the average of each part of the five confusion matrixes [38]. Results indicate that 95.80% of HC, 96.40% of MCI, and 98.00% of AD subjects can be correctly classified. The model confusion matrix is illustrated in Fig. 7.

When the classification takes AD as positive samples while MCI and HC as negative samples, the accuracy rate, recall rate and F1-score of the model are 97.66%, 98.00% and 97.83%, respectively. When MCI is taken as positive samples while AD and HC are as negative samples, the three rates are 97.96%, 96.40% and 97.18%, respectively. When HC is taken as positive samples while AD and MCI are as negative samples, the three rates are 93.01%, 95.80% and 94.38%, respectively. The accuracy rate, recall rate and F1-score of the model in the case that AD, MCI and HC are taken as positive samples respectively are presented in Table 5. Due to the uneven proportions of AD, MCI and HC subjects in the samples, the accuracy rate, recall rate and F1-score of the three-class classification model are calculated by using the weighted-average method [39], with 97.12% accuracy rate, 97.10% recall rate and 97.11% F1-score.

The evaluation indicators for the five 5-fold cross-validation of the model are plotted as box plots in Fig. 8. The mean values of accuracy, sensitivity, specificity and AUC for the five experiments of the model were 97.10%, 97.08%, 98.14% and 0.9754 respectively. There was one outlier of 95.90% for accuracy, two outliers of 94.30% and 99.10% for sensitivity and one outlier of 0.992 for AUC. This box plot shows that although there are a certain number of outliers for each indicator, the results of multiple runs of the model have excellent performance and high authenticity of the model running results.

5. Discussion

In this study, a data augmentation scheme for one-dimensional frequency domain features of EEG and a DPCNN-based deep learning model for classifying AD, MCI and HC are proposed. The model achieves a classification accuracy rate of up to 97.10% on the

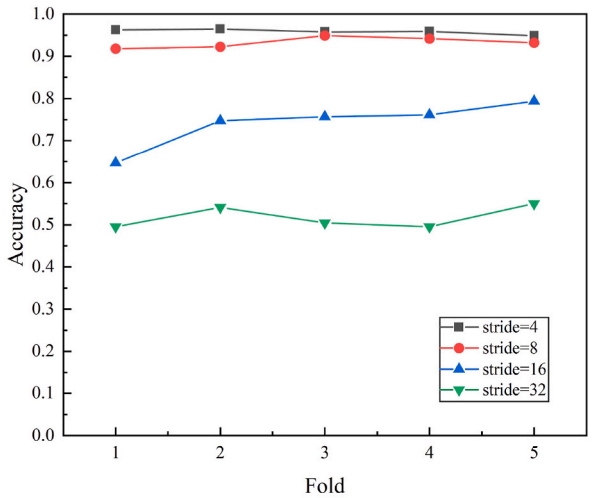

Fig. 5. Performance of the model (with different step lengths but a fixed window size of 152) using the data augmentation scheme.

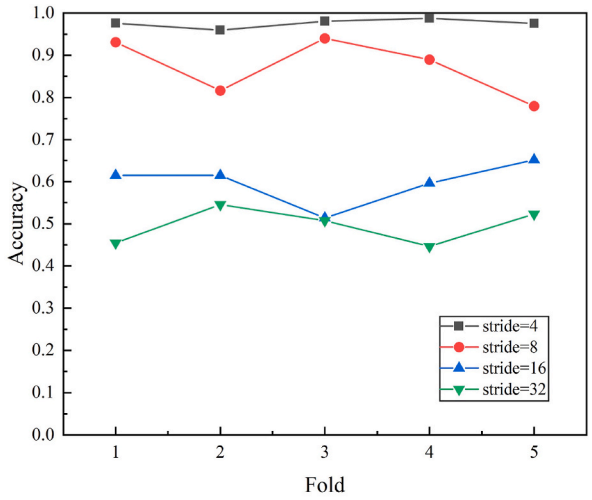

Fig. 6. Performance of the model (with different step lengths but a fixed window size of 228) using the data augmentation scheme.

Table 4
Performance of the deep learning model on 5-fold cross validation using different data augmentation schemes.

Window	Step size	1 fold		2 fold		3 fold		4 fold		5 fold	
		Loss	Acc	Loss	Acc	Loss	Acc	Loss	Acc	Loss	Acc
76	4	0.3558	85.28%	0.3239	88.01%	0.3020	87.93%	0.3371	87.28%	0.4939	79.55%
	8	0.9028	69.83%	0.7862	71.04%	0.8775	65.51%	0.7513	72.63%	0.8306	69.15%
	16	1.7883	42.51%	1.6897	43.43%	1.5569	48.73%	1.5440	50.76%	1.7150	42.58%
	32	2.6230	48.00%	2.5090	44.00%	2.1194	45.98%	2.2533	43.68%	2.3290	36.78%
152	4	0.0983	96.26%	0.1065	96.50%	0.1072	95.77%	0.1114	95.80%	0.1268	94.92%
	8	0.2950	91.81%	0.2966	92.27%	0.1791	94.93%	0.2568	94.20%	0.3498	93.24%
	16	1.3159	64.68%	1.3835	74.77%	1.0315	75.69%	1.1039	76.15%	0.8712	79.36%
	32	2.4952	49.54%	2.1432	54.13%	3.2865	50.46%	2.3459	49.54%	2.3647	55.05%
228	4	0.0802	97.59%	0.1485	95.97%	0.0792	98.07%	0.0286	98.79%	0.1493	97.58%
	8	0.2599	93.12%	1.1813	81.65%	0.1314	94.04%	0.3829	88.99%	1.3691	77.98%
	16	2.3029	61.47%	1.9817	61.47%	2.2511	51.38%	2.0074	59.63%	1.3725	65.14%
	32	2.7000	45.45%	2.0109	54.55%	2.6122	50.77%	2.9225	44.62%	2.7134	52.31%

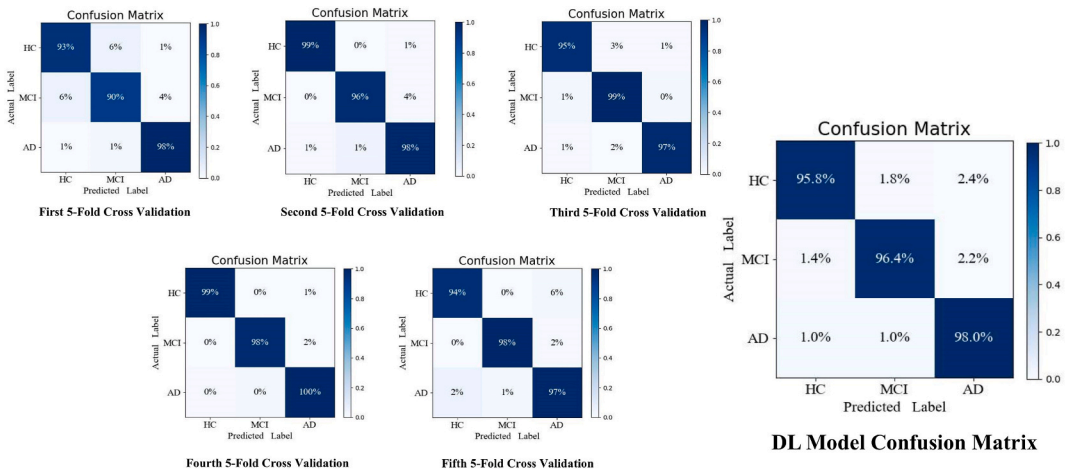

Fig. 7. Confusion matrix results under the optimal data augmentation scheme of the deep learning model.

data set under the optimal augmentation scheme. The F1-scores of the three-class classification model are 97.83%, 97.18 and 94.38% when AD, MCI and HC are taken as positive classifications respectively, and that calculated by the weighted-average method is 97.11%. It suggests that our research scheme is feasible, and the model can effectively learn and accurately classify AD, MCI and HC by

Table 5
Accuracy rate, recall rate, and F1-score of the model when different subjects are taken as positive classification.

Positive	Negative	Accuracy rate	Recall rate	F1-score
AD	MCI, HC	97.66%	98.00%	97.83%
MCI	AD, HC	97.96%	96.40%	97.18%
HC	AD, MCI	93.01%	95.80%	94.38%

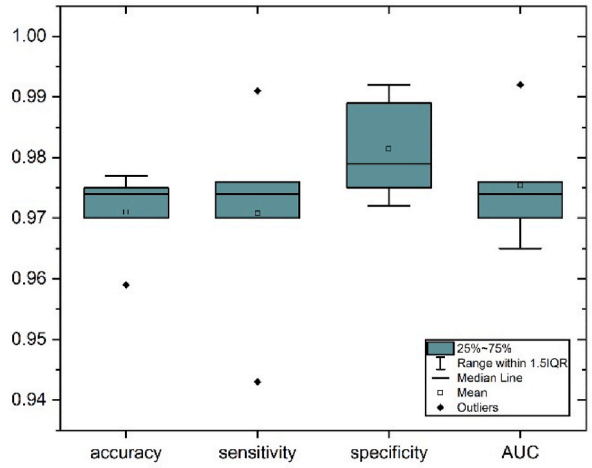

Fig. 8. Box plot of 5 times 5-fold cross validation of each evaluation indicator.

utilizing their EEG characteristics. The results of 5-fold cross validation and confusion matrix indicate that the three-class classification model is robust and the research results are reproducible.

Fiscon et al. [17] used three machine learning methods, i.e., SVM, decision tree and DMB classifier, to classify AD based on the same original EEG data set. Among them, decision tree performs the best in classification. The accuracy rate of the three two-class classifications of the J48 decision tree are AD vs. HC = 73.00%, AD vs. MCI = 80.00%, and MCI vs. HC = 90.00%. This study is obviously advantageous with respect to recognition accuracy, and the effective data augmentation method used in it can greatly enhance the robustness and generalization ability of the model.

In relevant studies in Section 2, Ferri et al. [19] and Duan et al. [20] conducted two-class classification studies on AD. Different from those previous studies, this study is about the three-class classification of AD, MCI and HC. As the two-class classification study does not accurately classify MCI which is more valuable for disease prediction [40], the three-class classification of AD is more significant than it [41].

In the other three-class classification studies, Ieracitano et al. [22] and Huggins et al. [26] converted one-dimensional EEG into high-dimensional images as input of the deep learning model, requiring a complicated process of data preprocessing and con-version. The accuracy rates of three-class classification conducted by Ieracitano et al. [22], You et al. [23] and Amini et al. [25] were low. Though Rad et al. [24] used one-dimensional EEG features, with less human intervention in EEG preprocessing and a three-class classification accuracy rate of 97.50%, the EEG data were collected in 3 channels of self-defined electrode positions. In this study, resting-state EEG of 19 channels collected in the international 10–20 system is used, the study samples being more representative and the deep learning model generalization ability being stronger. Our method balances the complexity of feature extraction and the representativeness of EEG samples while diagnosing AD with high accuracy, which is superior and efficient compared to the methods mentioned in the literature review.

Our main contributions are as follows:

- One-dimensional frequency domain EEG data were augmented based on the sliding window. Resultantly, the sample size was increased and the generalization ability of the deep learning model was improved.
- A deep learning model based on the pyramid convolutional neural network was proposed to automatically extract frequency domain features of EEG, which greatly reduces learnable parameters and data redundancy, and improves algorithm efficiency.
- The deep learning model proposed to classify AD, MCI and HC achieved an accuracy rate of 97.10%, specifically, AD 98.00%, MCI 96.40% and HC 95.80%.

Despite the good results, this research still has its limits. The selection of the optimal data augmentation scheme and the adjustment of parameters of the deep learning model are all realized in a manual way, which greatly increases the workload of the research and the risk of omitting the optimal scheme and parameters. Although the weighted-average method has been adopted for F1-Score of the

three-class classification to reduce the influence caused by uneven proportion of sample categories, F1-Score is still inevitably inclined to the value yielded when AD was positive classification (because AD samples are the largest in number). The indicators for evaluating the results of the three-class classification can be improved.

6. Conclusion

In conclusion, this study designed a deep learning model based on DPCNN to classify one-dimensional EEG features of AD, MCI and HC subjects. The accuracy rate of the model is up to 97.10% in three-class classification of the augmented data set. The innovative use of the data augmentation method for one-dimensional EEG frequency domain features of AD not only solves the problem of the small size of data set of EEG research on cognitive impairment, but also improves the robustness of deep learning model. The prospects for future work are as follows:

- The EEG data of AD, MCI and HC subjects are automatically collected under the condition that the proportions of AD, MCI and HC subjects are balanced.
- The generalization ability of the model is enhanced by using a variety of EEG devices to collect resting-state EEG signals with different channel numbers and building multiple data sets to train the model.
- Automatic parameter optimization methods, such as grid search and Bayesian optimization, are used when adjusting the parameters of the data augmentation method and the deep learning model.

The result of this study can significantly improve the efficiency of AD diagnosis. After being improved in the future, it is expected to become a new means of clinical diagnosis of AD, which will relieve the pressure of front-line neurologists and have a great prospect of clinical application.

Author contribution statement

Wei Xia: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. Ran Zhang: Conceived and designed the experiments; Wrote the paper. Xiao Zhang: Conceived and designed the experiments. Muhammad Usman: Contributed reagents, materials, analysis tools or data.

Funding statement

Xiao Zhang was supported by National Key Research and Development Program of China [No. 2020YFB1711500]. PhD Ran Zhang was supported by Natural Science Foundation of the Jiangsu Higher Education Institutions of China [No. 20KJB510025], Research Start-up Fund for Excellent Talents of Xuzhou Medical University [No. D2019032].

Data availability statement

Data will be made available on request.

Declaration of interest's statement

The authors declare no conflict of interests.

References

[1] S.F. Javaid, C. Giebel, M.A.B. Khan, et al., Epidemiology of Alzheimer's disease and other dementias: rising global burden and forecasted trends, F1000Research 10 (2021) 425–438.
[2] T. Dua, K.M. Dua, S. Sivananthan, et al., World health organization's global action plan on the public health response to dementia 2017–2025, Alzheimer's Dementia 13 (2017) 1450–1451.
[3] L. Zhao, Alzheimer's disease facts and figures, Alzheimer's Dementia 16 (3) (2020) 391–460.
[4] Alzheimers Association, Alzheimer's disease facts and figures, Alzheimer's Dementia 11 (3) (2015) 332–384.
[5] C.R. Jack, D.S. Knopman, W.J. Jagust, et al., Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade, Lancet Neurol. 9 (1) (2010) 4–5.
[6] C. Laske, H.R. Sohrabi, S.M. Frost, et al., Innovative diagnostic tools for early detection of Alzheimer's disease, Alzheimer's Dementia 11 (5) (2015) 561–578.
[7] K.D. Tzimourta, V. Christou, A.T. Tzallas, et al., Machine learning algorithms and statistical approaches for Alzheimer's disease analysis based on resting-state EEG recordings: a systematic review, Int. Neural. Syst. 31 (2020) 1–33.
[8] N. Kannathal, M.L. Choo, U.R. Acharya, et al., Entropies for detection of epilepsy in EEG, Comput, Meth. Prog. Bio. 80 (3) (2005) 187–194.
[9] H. Chen, Y. Song, X. Li, A deep learning framework for identifying children with ADHD using an EEG-based brain network, Neurocomputing 356 (3) (2019) 83–96.
[10] H. Cai, Z. Qu, Z. Li, et al., Feature-level fusion approaches based on multimodal EEG data for depression recognition, Inf. Fusion 59 (2020) 127–138.
[11] Q. Ge, Z.C. Lin, Y.X. Gao, et al., A robust discriminant framework based on functional biomarkers of EEG and its potential for diagnosis of Alzheimer's disease, Healthcare (4) (2020) 476–489.
[12] S.W. Scheff, D.A. Price, F.A. Schmitt, et al., Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment, Neurology 68 (18) (2007) 1501–1508.
[13] U. Smailovic, V. Jelic, Neurophysiological markers of Alzheimer's disease: quantitative EEG approach, Neurol. Ther. 8 (2) (2019) 37–55.
[14] U.R. Acharya, S.V. Sree, S. Chattopadhyay, et al., Automated diagnosis of normal and alcoholic EEG signals, Int. J. Neural Syst. 22 (3) (2012), 1250011.

[15] G. Tavares, R. San-Martin, J.N. Ianof, et al., Improvement in the automatic classification of Alzheimer's disease using EEG after feature selection, in: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019.
[16] M.J. Rivera, M.A. Teruel, A. Maté, Trujillo, et al., Diagnosis and prognosis of mental disorders by means of EEG and deep learning: a systematic mapping study, Artif. Intell. Rev. 55 (12) (2021) 1–43.
[17] G. Fiscon, E. Weitschek, G. Felici, et al., In Alzheimer's disease patients classification through EEG signals processing, Computational Intelligence & Data Mining (2014) 105–112.
[18] R. Johnson, T. Zhang, Deep pyramid convolutional neural networks for text categorization, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, pp. 562–570.
[19] F. Duan, Z. Huang, Z. Sun, et al., Topological network analysis of early Alzheimer's disease based on resting-state EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (10) (2020) 2164–2172.
[20] M.A. Kramer, F.L. Chang, M.E. Cohen, et al., Synchronization measures of the scalp electroencephalogram can discriminate healthy from Alzheimer's subjects, Int. J. Neural Syst. 17 (2) (2007) 61–69.
[21] C. Ieracitano, N. Mammone, A. Bramanti, et al., A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings, Neurocomputing 323 (2019) 96–107.
[22] Z. You, R. Zeng, X. Lan, et al., Alzheimer's disease classification with a cascade neural network, Front. Public Health (2020).
[23] E. Mazrooei Rad, M. Azarnoosh, M. Ghoshuni, et al., Diagnosis of mild Alzheimer's disease by EEG and ERP signals using linear and nonlinear classifiers, Biomed. Signal. Proces. (2021).
[24] M. Amini, M. Pedram, A. Moradi, et al., Diagnosis of Alzheimer's disease by time-dependent power spectrum descriptors and convolutional neural network using EEG signal, Comput. Math. Methods Med. (2021).
[25] C.J. Huggins, J. Escudero, M.A. Parra, et al., Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing, J. Neural. Eng. 18 (4) (2021) 1–14.
[26] J. Dukart, M.L. Schroeter, K. Mueller, Age correction in dementia–matching to a healthy brain, PLoS One 6 (7) (2011).
[27] A. Akrami, S. Solhjoo, A. Motie-Nasrabadi, et al., In EEG-based mental task classification: linear and nonlinear classification of movement imagery, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 4626–4629.
[28] N. Kasabov, E. Capecci, Spiking neural network methodology for modelling, classification and understanding of EEG spatio-temporal data measuring cognitive processes, Inf. Sci. 294 (2015) 565–575.
[29] E. Lashgari, D. Liang, U. Maoz, et al., Data augmentation for deep-learning-based electroencephalography, J. Neurosci. Methods 346 (2020) 1–36.
[30] W. Wang, J. Gang, In application of convolutional neural network in natural language processing, in: 2018 International Conference on Information Systems and Computer Aided Education, ICISCAE, 2019, pp. 64–70.
[31] Ihsan Ullah, Muhammad Hussain, Emad-ul-Haq Qazi, Hatim Aboalsamh, An automated system for epilepsy detection using EEG brain signals based on deep learning approach, Expert Syst. Appl. 107 (2018) 61–71.
[32] S. Ruder, An overview of gradient descent optimization algorithms, 2016 arXiv preprint arXiv: 1609.04747.
[33] G. Perin, S. Picek, On the influence of optimizers in deep learning-based side-channel analysis, in: 27th International Conference on Selected Areas in Cryptography, 2020, pp. 615–636.
[34] Y. Sun, W. Zhang, H. Gu, et al., Convolutional neural network based models for improving super-resolution imaging, IEEE Access 7 (2019) 43042–43051.
[35] P. Ramachandran, B. Zoph, Q. V. Le, Searching for Activation Functions, 2017 arXiv preprint arXiv: 1710.05941.
[36] D. M. W. Powers, Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation, 2010 arXiv preprint arXiv:2010.16061.
[37] K.H. Brodersen, C.S. Ong, K.E. Stephan, et al., The balanced accuracy and its posterior distribution, in: 2010 20th International Conference on Pattern Recognition, 2010, pp. 3121–3124.
[38] M. Grandini, E. Bagli, G. Visani, et al., Metrics for Multi-Class Classification: an Overview, 2008 arXiv preprint arXiv: 2008.05756.
[39] S.M. Rasquin, J. Lodder, P.J. Visser, et al., Predictive accuracy of MCI subtypes for Alzheimer's disease and vascular dementia in subjects with mild cognitive impairment: a 2-year follow-up study, Dement. Geriatr. Cogn. 19 (2–3) (2005) 113–119.
[40] D. Zhang, Y. Wang, L. Zhou, et al., Multimodal classification of Alzheimer's disease and mild cognitive impairment, Neuroimage 55 (3) (2011) 856–867.