Cognitive Computation
https://doi.org/10.1007/s12559-023-10188-7

Assessing the Potential of Data Augmentation in EEG Functional Connectivity for Early Detection of Alzheimer's Disease

Hao Jia[1,2] · Zihao Huang[2] · Cesar F. Caiafa[3] · Feng Duan[2] · Yu Zhang[4] · Zhe Sun[5] · Jordi Solé-Casals[1,6]

Received: 21 February 2023 / Accepted: 30 July 2023

Abstract

Electroencephalographic (EEG) signals are acquired non-invasively from electrodes placed on the scalp. Experts in the field can use EEG signals to distinguish between patients with Alzheimer's disease (AD) and normal control (NC) subjects using classification models. However, the training of deep learning or machine learning models requires a large number of trials. Datasets related to Alzheimer's disease are typically small in size due to the lack of AD patient samples. The lack of data samples required for the training process limits the use of deep learning techniques for further development in clinical settings. We propose to increase the number of trials in the training set by means of a decomposition–recombination system consisting of three steps. Firstly, the original signals from the training set are decomposed into multiple intrinsic mode functions via multivariate empirical mode decomposition. Next, these intrinsic mode functions are randomly recombined across trials. Finally, the recombined intrinsic mode functions are added together as artificial trials, which are used for training the models. We evaluated the decomposition–recombination system on a small dataset using each subject's functional connectivity matrices as inputs. Three different neural networks, including ResNet, BrainNet CNN, and EEGNet, were used. Overall, the system helped improve ResNet training in both the mild AD dataset, with an increase of 5.24%, and in the mild cognitive impairment dataset, with an increase of 4.50%. The evaluation of the proposed data augmentation system shows that the performance of neural networks can be improved by enhancing the training set with data augmentation. This work shows the need for data augmentation on the training of neural networks in the case of small-size AD datasets.

Introduction

Alzheimer's disease (AD) is a clinical syndrome characterized by the progressive deterioration of the memory and cognitive functions, particularly in elderly people. The disease usually appears silently, and the process is slow and irreversible. According to the 2019 Alzheimer's World Report [1], there are more than 50 million people with AD. The figure may rise to 152 million by 2050.

✉ Feng Duan
duanf@nankai.edu.cn

✉ Zhe Sun
z.sun.kc@juntendo.ac.jp

✉ Jordi Solé-Casals
jordi.sole@uvic.cat

Yu Zhang
yuzhang@lehigh.edu

[1] Data and Signal Processing Research Group, University of Vic-Central, University of Catalonia, Vic, Catalonia, Spain

[2] College of Artificial Intelligence, Nankai University, Tianjin, China

[3] CONICET CCT La Plata/CIC-PBA/UNLP, Instituto Argentino de Radioastronomía, Villa Elisa, Argentina

[4] Department of Bioengineering and the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem 18015, PA, USA

[5] Faculty of Health Data Science, Juntendo University, Urayasu, Chiba, Japan

[6] Department of Psychiatry, University of Cambridge, Cambridge, UK

Published online: 09 September 2023

In recent years, the attention paid to AD has been gradually increasing. So far, only five drugs have been approved by the Food and Drug Administration (FDA) for the treatment of AD [2], and all of them can only delay the development of AD and alleviate symptoms, but not cure or even treat AD. Consequently, early diagnosis is important to delay the symptoms through medication. Typically, AD is divided into four stages, and the best time to diagnose the disease is during the early stages of mild cognitive impairment (MCI) and mild AD [3–5].

Electroencephalography (EEG) is the non-invasive acquisition of signals corresponding to electrical activity in the brain using electrodes positioned directly on the scalp. Magnetoencephalography (MEG) is also a non-invasive technique which is used to acquire signals by recording the magnetic activity of the brain. Functional magnetic resonance imaging (fMRI) indirectly detects changes of the brain neuronal activity based on the linked alterations of cerebral blood flow as exhibited by the differentiated magnetic properties of the hemoglobin molecule between its oxygen saturated and desaturated states. The difference between AD patients and normal control subjects can be detected using these brain signals, each coming with different advantages and disadvantages. Machine learning methods related to the classification between AD patients and normal control subjects using EEG, MEG, and fMRI brain signals are listed in Table 1.

With the increasing use of deep learning techniques, many deep AD detection methods have recently emerged. Sarraf and Tofighi [14] used LeNet-5, a convolutional neural network (CNN) architecture, to classify fMRI data from AD subjects and normal controls, with an accuracy on the testing dataset of 96.85%. They used 5-fold cross-validation on a dataset containing 28 AD subjects and 15 normal controls. Kim and Kim [15] proposed a classifier based on deep neural networks using the relative power of EEG to fully exploit and recombine features through its own learning structure. Their dataset contained 10 MCI subjects and 10 normal controls, and leave-one-out cross-validation was used to evaluate the model's performance. The accuracy obtained on the testing dataset was 59.4%. Duan et al. [16] used EEG functional connectivity as the network input to train ResNet-18, achieving an accuracy of 93.42% and 98.5% on the MCI and mild AD datasets, respectively, where the former contained 22 MCI subjects and 38 normal controls, and the latter contained 17 mild AD subjects and 24 normal controls.

Among the aforementioned brain signals (EEG, MEG, and fMRI), EEG has the best temporal resolution. Nevertheless, since EEG signals are acquired from several locations on the scalp with electrodes, their spatial resolution is not as good as that of the measurements for the other two types of signals. Despite this, the spatial distribution of the signals can be optimized in the processing steps with the use of well-designed algorithms [17–21]. Given that EEG signals are easier to acquire and is less expensive than other techniques, EEG-based methods for AD detection are currently more popular.

In studies based on EEG signals, deep learning methods are trained on small datasets, as electrophysiological signals are more difficult to acquire in AD patients. The learning capability of deep learning models partially relies on their large number of hyper-parameters. A high amount of samples is required to fit these hyper-parameters and avoid the over-fitting problem [22, 23]. One way to deal with the issue is using data augmentation.

Data augmentation can be implemented by generating artificial data [24, 25]. The strategy of decomposing and recombining the original EEG signals is one possible way to create new artificial data for data augmentation [26–28]. EEG signals can be decomposed into different filter banks. In each filter bank, the frequency of the decomposed EEG signals is within a certain frequency band. All filter banks cover a wide range of frequencies. This strategy helps to achieve a better performance using deep-learning models in the enhancement of small-size datasets. Note that in studies where this particular data augmentation strategy has been implemented, the details about the models used are not entirely the same throughout, even though the same overall approach is being used. For instance, Zhao et al. [26] proposed a method of

Table 1 Summary of papers using EEG/MEG/fMRI signals to design a classification system for AD/MCI detection

Ref	Method	Signal	Disease type	Accuracy	Year
[6]	Correlation, phase synchrony, and Granger causality measures	EEG	MCI and mild AD	83% and 88%, respectively	2012
[7]	Hybrid feature selection	EEG	MCI and mild AD	95% and 100%, respectively	2015
[8]	Complex network theory and TSK fuzzy system	EEG	AD	97.3%	2019
[9]	Functional connectivity and effective connectivity analysis	MEG	AD	86%	2019
[10]	Phase locking value, imaginary part, and correlation of the envelope	MEG	MCI	75%	2019
[11]	High-order FC correlations	fMRI	MCI	88.14%	2016
[12]	Hierarchical high-order functional connectivity networks	fMRI	MCI	84.85%	2017
[13]	Strength and similarity guided GSR using LOFC and HOFC	fMRI	MCI	88.5%	2019

random recombination of EEG signals in different filter banks, which are decomposed by the discrete cosine transform. This approach enhances the classification performance of one-dimension convolutional neural networks in the epileptic seizure focus detection task. Zhang et al. [27] used the augmentation strategy to enhance the classification performance of motor imagery. Instead of decomposing signals with the discrete cosine transform, the empirical mode decomposition (EMD) technique was adopted [29]. In the decomposition–recombination strategy, EMD has the advantage that the signals can be recovered by simply adding up the decomposed intrinsic mode funtions (IMFs). Besides the decomposition–recombination strategy, generative adversarial networks (GANs) also offer a solution to generate artificial signals [30]. However, GANs require a large dataset to tune the parameters and fit the model. Since the goal of data augmentation in small Alzheimer's datasets is to solve the problem of insufficient samples, it is not possible to use GANs to generate artificial data.

In this paper, we propose a decomposition and recombination model for data augmentation in a small Alzheimer's data set, which is used to distinguish AD patients from normal controls. The decomposition and recombination approach consists of three steps. First, empirical multivariate mode decomposition (MEMD) is used to decompose EEG signals into IMFs. These IMFs are then randomly recombined within each of the two groups. Finally, in each group, the IMFs are added up to generate a new artificial trial. These artificial trials are used to extend the AD training dataset.

This work is organized as follows. "Method" includes the description of the small Alzheimer's datasets used, the scheme of the proposed decomposition and recombination approach, and the neural networks used for classification. "Results" presents the experimental results, including the classification performance of the neural networks during the training process and the effects of data augmentation in the datasets. Then, these results are discussed in "Discussion", together with the limitations associated with the method. Finally, the conclusions are presented in "Conclusion".

Method

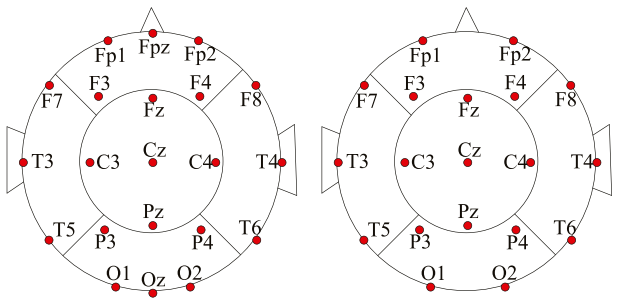
Alzheimer's Datasets

All experiments in this work use two datasets: the MCI dataset, containing 22 subjects with MCI and 38 normal controls, and the mild AD dataset, containing 17 subjects with mild AD and 24 normal controls. Other studies have been conducted based on these datasets [5, 7, 31].

The MCI Dataset

The MCI dataset is comprised of data from subjects who complained of memory impairment and of control subjects who did not have memory impairment or other diseases. The patient group included 53 subjects who underwent a comprehensive neuropsychological test; the results showed quantitative and objective evidence of memory impairment, but their overall cognitive, behavioral or functional status was not significantly lost. The classification of mild dementia impairment requires a score of at least 24 in the Mini-Mental State Examination (MMSE) [32], a score of 0.5 on the Clinical Dementia Rating (CDR) scale [33] and a standard deviation lower than the normal memory performance reference value. All subjects met these criteria. Then, these subjects underwent an initial assessment, and their progress was monitored in the clinic during the subsequent 12–18 months. According to the criteria defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA), 25 of these 53 mild AD patients might develop into AD. The average age of the 25 subjects in the MCI data set is 71.9 ± 10.2 years old, and the MMSE score is 28.5 ± 1.6. The control group had 56 age-matched healthy subjects with an average age of 71.7 ± 8.3 years old and an MMSE score of 26 ± 1.8.

Twenty-one electrodes from Biotop 6R12 (NEC-Sanei, Tokyo, Japan) were placed on the subject's scalp in a 10–20 international system with a sampling frequency of 200 Hz. In addition, Fpz and Oz electrodes were added to the system, as shown in Fig. 1a. After the data was collected, analog bandpass filtering was used



(a) Placement of the EEG electrodes in the MCI dataset (b) Placement of the EEG electrodes in the mild AD dataset

Fig. 1 Schematic display of the electrode positions from above

to retain data between 0.5 and 250 Hz, and then third-order Butterworth filters (forward and reverse filtering) were used to perform digital band-pass filtering between 0.5 and 30 Hz.

The Mild AD Dataset

The mild AD dataset is comprised of data from 17 mild AD patients (age: 69.4 ± 11.5 years) and 24 healthy subjects (age: 77.6 ± 10.0 years). The patient group underwent a full set of cognitive tests (MMSE, Rey auditory verbal learning, Benton visual retention, and memory recall tests) along with psychological tests. The results were graded and interpreted by psychologists and then discussed in meetings with multidisciplinary teams. The subjects in the control group were all healthy volunteers, and their EEG was judged to be normal by the clinical neurophysiology consultants.

Nineteen electrodes were placed on the subject's scalp using the Maudsley system, which is similar to the international 10–20 system. The sampling frequency was 128 Hz, as shown in Fig. 1b. After data acquisition was carried out, a third-order Butterworth filter (forward filter and reverse filter) was used for digital band-pass filtering between 0.5 and 30 Hz.

Recording Conditions in Both Datasets

During the collection process of the two aforementioned datasets, the subjects were awake and with their eyes closed. The whole process lasted for 5 min. After that, the EEG data was checked by EEG experts, and the data containing artifacts were discarded. Finally, only clean EEG data of 20 s of length was saved for each subject, discarding the subjects whose data did not meet this condition. Based on this procedure, the MCI dataset finally comprised of 22 subjects with MCI and 38 normal controls, while the mild AD dataset comprised of 17 subjects with mild AD and 24 normal controls.

A Decomposition and Recombination System

In small data sets, neural networks often face overfitting problems. Data augmentation is used to enlarge the size of the training set, as shown in Fig. 2.

In this work, we propose a decomposition and recombination system to generate artificial trials and thus enlarge the training set. For the decomposition part, the empirical mode decomposition (EMD) method is used. EMD can divide a signal into multiple intrinsic mode functions (IMFs). These IMFs cover different frequency bands, with low overlap. The original signal can then be

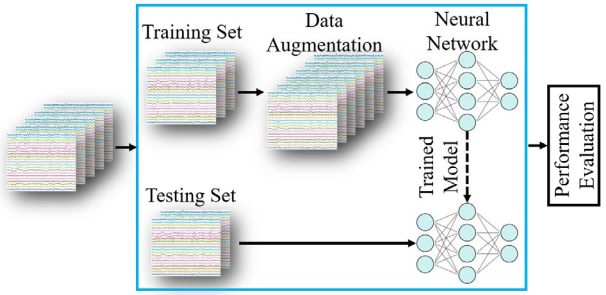


Fig. 2 The concept of data augmentation. In a small data set, the training set is small in size, since it is generated from only a portion of the (few) original data. When a neural network is used to fit the training set, there is a potential overfitting problem. Data augmentation is used to mitigate this issue by enlarging the size of the training set

recovered by adding up these IMFs [29]. The recombination part consists of adding IMFs from different trials, taking each of the IMFs from a different one.

The simplest EMD method is classical empirical mode decomposition (CEMD), which is the original version of EMD, as shown in the algorithm 1. A faster version of EMD is serial EMD (SEMD), which is used to deal with multi-channel signals. SEMD converts multi-channel signals into a single channel by concatenating them over time, ensuring the continuity of the signals by suitably adding a transient part between channels. CEMD is then used to decompose the single (long) channel. Multivariate EMD (MEMD) is also a method used for decomposing multi-channel signals, as shown in the algorithm 2. First, it places the multi-channel signals in a tangent space and then decomposes these signals into IMFs. The IMFs are finally reverted to normal space. Figure 3 shows the original multi-channel signals and the signals decomposed by MEMD. MEMD ensures that IMFs with the same index (shown in Fig. 3) cover the same frequency band.

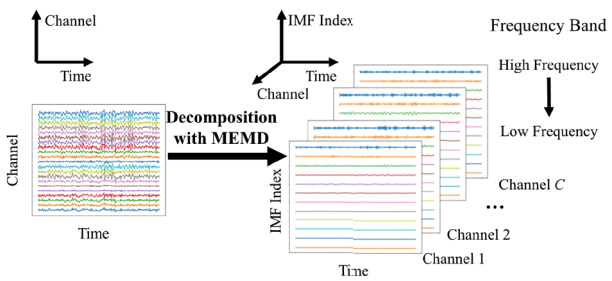


Fig. 3 Data decomposition with MEMD. MEMD can decompose multi-channel signals into IMFs. The IMFs are located in different frequency bands, but in all the decomposed channels, the *k*th IMF covers the same frequency band. In this figure, the IMFs are sorted in descending order in the frequency domain

Algorithm 1 CEMD

Require: $T \in \mathbb{Z}$
Ensure: $\mathbf{x} \in \mathbb{R}^{T \times 1}$
  while number of extreme values in $\mathbf{x}$ is greater than 3 **do**
    interpolate the maxima extremes (minima extremes) of $\mathbf{x}$
    obtain the envelop of extremes e_{max} (e_{min})
    $\mathbf{m} = (e_{max} + e_{min})/2$
    $\mathbf{h} = \mathbf{x} - \mathbf{m}$
    if $\mathbf{h}$ is an IMF **then**
      save $\mathbf{h}$ as an IMF
      $\mathbf{x} = \mathbf{x} - \mathbf{r}$
    else
      $\mathbf{x} = \mathbf{h}$
    end if
  end while
  save $\mathbf{x}$ as an IMF

Algorithm 2 MEMD

Require: $T \in \mathbb{Z}$
Ensure: $\mathbf{x} \in \mathbb{R}^{T \times C}$
  obtain N_{dir} uniformly-distributed sequences ($\boldsymbol{\theta} = \theta_1, \theta_2, ..., \theta_P$) from Hammersley sequence, $N_{dir} > P$;
  cast $\boldsymbol{\theta}$ into tangent space and obtain direction vector $\mathbf{v}$.
  while number of extreme values in $\mathbf{x}$ is greater than 3 **do**
    for $n = 1 : N_{dir}$ **do**
      cast X into tangent space with inner product $(X \cdot \mathbf{v})$
      locate the positions of extremes in $X \cdot \mathbf{v}$, which are assumed to be the positions of extremes in X
      interpolate the maxima extremes (minima extremes) of $\mathbf{X}$
      obtain the envelop of extremes e^n_{max} (e^n_{min}) of the n-th uniform-distributed sequence.
    end for
    $e_{max} = mean(e^n_{max})$ for $n = 1 : N_{dir}$
    $e_{min} = mean(e^n_{min})$ for $n = 1 : N_{dir}$
    $\mathbf{m} = (e_{max} + e_{min})/2$
    $\mathbf{h} = \mathbf{x} - \mathbf{m}$
    if $\mathbf{h}$ is an IMF **then**
      save $\mathbf{h}$ as an IMF
      $\mathbf{x} = \mathbf{x} - \mathbf{r}$
    else
      $\mathbf{x} = \mathbf{h}$
    end if
  end while
  save $\mathbf{x}$ as an IMF

In order not to decompose each trial separately, which would result in IMFs covering non-equal frequency bands in the same trial, and also to decrease the processing time, we combine the MEMD and SEMD methods as shown in Fig. 4. Multi-channel signals from several trials are first concatenated along the time axis as in SEMD, and then MEMD is used to decompose the concatenated signals, ensuring that each trial has the same number of IMFs. Figure 5 presents an example of generating an artificial trial with the original EEG signals.

Neural Network Classifiers

In the analysis of EEG signals, there are two traditional options used as inputs for the neural networks. In the first case, the original multi-channel signals are used as inputs. In the second, the multi-channel signals are converted into a functional connectivity (FC) matrix [34]; this is an EEG-based connectivity matrix between brain regions obtained by calculating the inter-channel EEG similarity, e.g., by means of the coherence measure. The degree of similarity between two brain regions can be reflected in the FC matrix. In this way, the generated matrix preserves the spatial information of the multi-channel signals. To distinguish between controls and AD patients, EEG is often analyzed in four frequency bands: *delta* (0.1–4 Hz), *theta* (4–8 Hz), *alpha* (8–13 Hz), and *beta* (13–30 Hz). The signal in each band contains different information about brain connectivity and synchronization [35]. In this work, however, we adopt slightly different frequency bands, namely 4–8 Hz, 8–10 Hz, 10–13 Hz, and 13–30 Hz. These bands are derived from a previous work [16] and are optimized for the datasets used [7].

The main goal of this work is to measure the effect of the data augmentation method on the performance of the classifiers when functional connectivity matrices are used as inputs to the models. Therefore, it is not in the scope of this work to determine the best possible model. To evaluate the effects of the data augmentation method on the small AD datasets, three neural networks are used: BrainNet CNN [36], ResNet [37], and EEGNet [38]. To simplify the explanation of the networks, some symbols are defined here. In the following, B is the batch size, C is the number of input EEG signals, and T is the number of sample points of the EEG signals.

Methods such as Pearson's correlation coefficient or coherence can be used to compute the correlation or relationship between channels. Here, we adopt the coherence to compute the FC matrices. EEG coherence measures the degree of phase synchronization of EEG spectral activity between two electrodes [39]. For two temporal signals $x(t)$ and $y(t)$, the coherence between them can be defined as follows:

$$C_{xy} = \frac{\left|G_{xy}(f)\right|^2}{G_{xx}(f)G_{yy}(f)}, \tag{1}$$

where G_{xy} is the cross-spectral density between x and y, and G_{xx} and G_{yy} are the power-spectral density of x and y, respectively. Considering an EEG sample that has 21 channels containing data of 20 s of length, we can obtain an FC matrix with a size of $C \times C$ by calculating the coherence between each pair of EEG signals. Here, we first divide the original signals into the four aforementioned frequency bands, namely 4–8 Hz, 8–10 Hz, 10–13 Hz, and 13–30 Hz. As a consequence, the input of the neural networks is of size $4 \times C \times C$ (where C is the channel number of EEG signals). The inputs for BrainNet CNN and ResNet are the FC matrices of the four frequency bands. The input for EEGNet is the original multi-channel time series.

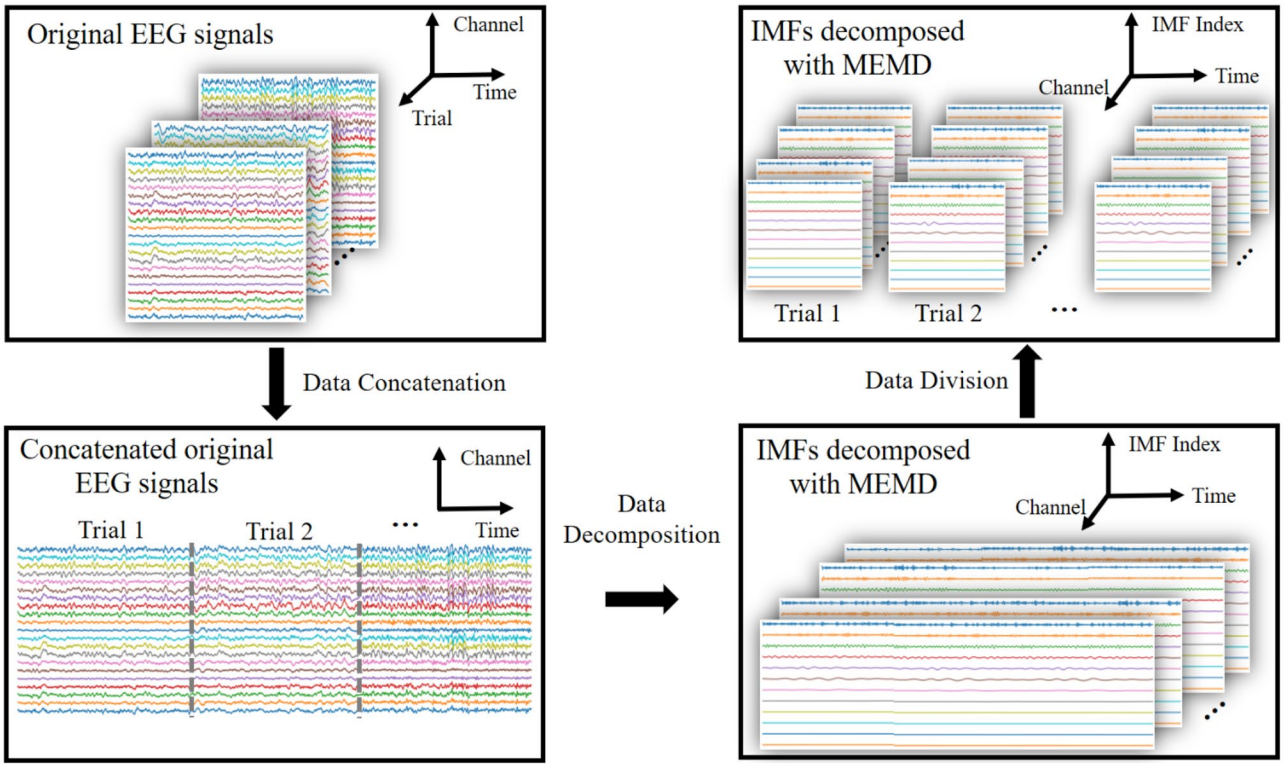


Fig. 4 The procedure of SEMD-MEMD decomposition for multiple trials of multi-channel EEG signals. Trials of EEG signals are concatenated along the time axis and then decomposed

BrainNet CNN

BrainNet CNN is a network architecture that analyzes the FCs of different frequency bands [36]. This network has three basic convolutional blocks: edge-to-edge (E2E), edge-to-node (E2N), and node-to-graph (N2G), which are specially designed for FC matrix processing. The three blocks are convolutional layers with different kernels. E2N is a convolutional layer with kernel size (1, C) which converts the edges in FC matrices to nodes. N2G is a convolutional layer with kernel size (C, 1) which suppresses the output nodes of the E2N layer. Finally, E2E is the added-up output of convolutions with kernel size (1, C) and (C, 1). An illustration of E2E is given in Fig. 6. The structure of the BrainNet CNN is given in Table 2.

ResNet

In the training process of deep learning methods, the back-propagation of multiple layers faces the problem of gradient vanishing [40]. The residual module of the deep residual network can reduce the influence of gradient vanishing by introducing a shortcut connection [37]. The deep residual network is a network that has already been validated on a large number of classification problems. Compared with that of deep neural networks without shortcut connections, the shortcut connection of the deep residual network allows raw input information to be sent directly to a later layer. Assuming that the input of the residual block is x, the expected output is $H(x)$. The learning target of the deep residual network is then $F(x) = H(x) - x$, which is called residual, and then the input and output of this block are added together through the shortcut (Fig. 7). This approach greatly increases the training speed of the model, improves the training effect, and effectively solves the vanishing problem when the number of layers is deepened without adding extra parameters and calculations to the network. In this study, we employed the ResNet-18 deep residual network.

EEGNet

EEGNet is a universal solution to the classification of multi-channel EEG signals, which has been validated in the classification of other brain activity signals such as motor imagery and movement-related cortical potential [38]. EEGNet takes the original multi-channel EEG signals as the input instead of the FC matrices. Even though EEGNet has not been validated in the classification of early AD, in this work, we use

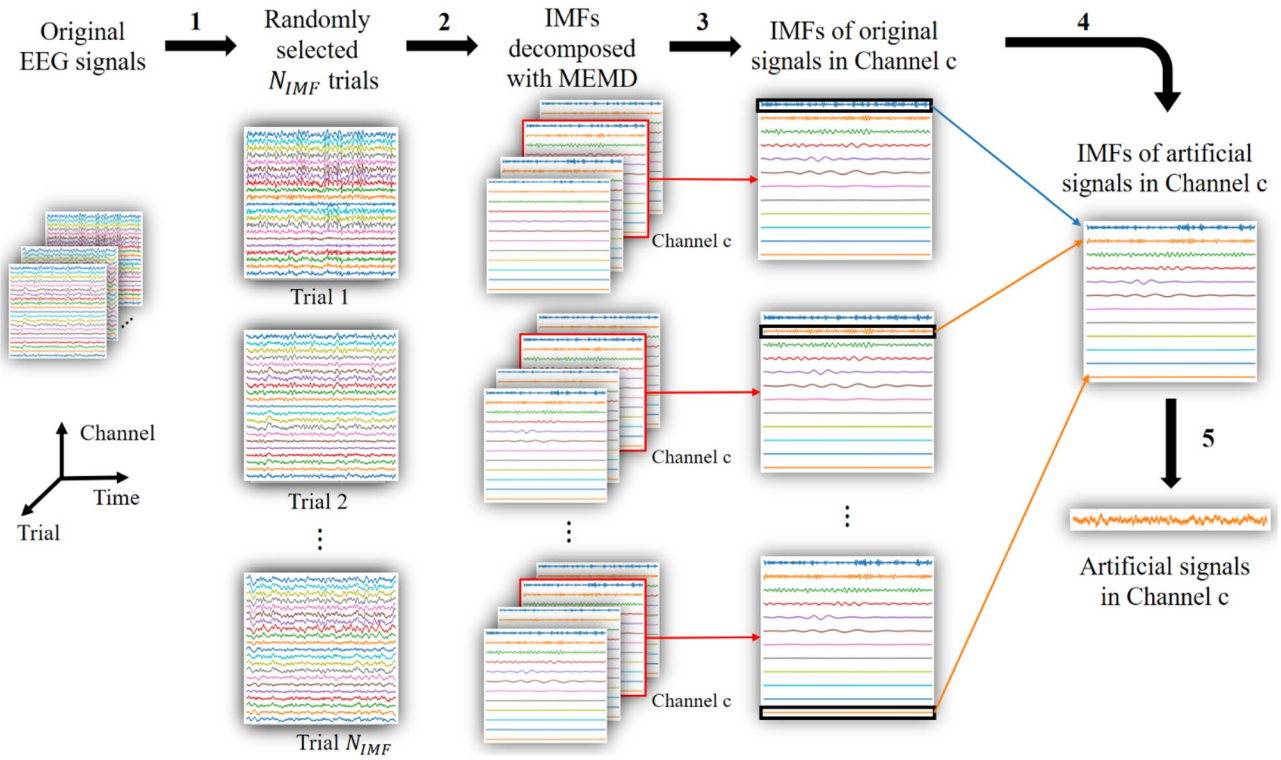


Fig. 5 Outline of the proposed decomposition and recombination system. As an example, for an artificial signal generated in channel c, the procedure consists of **i** randomly selecting N_{IMF} trials from the original EEG signals; **ii** obtaining the IMFs, which are decomposed using the method outlined in Fig. 4; **iii** collecting the decomposed IMFs of channel c from randomly selected N_{IMF} trials; **iv** recombining the IMFs in channel c. The n_{imf}-th IMF of the artificial signal is the n_{imf} -th IMF of the n_{imf}-th randomly selected trial; and **v** adding the IMFs and obtaining the artificial signal of channel c

it to test and explore the data augmentation performance. The structure of EEGNet is given in Table 3.

Parameter Setting

In the training of these neural networks, the adaptive moment estimation (Adam) optimizer was used, with $\beta_1 = 0.9$, $\beta_2 = 0.99$ and 0.0001 for the learning rate. ResNet and BrainNet CNN were trained using 100 epochs, and EEGNet was trained using 200 epochs. The mini-batch size was 50.

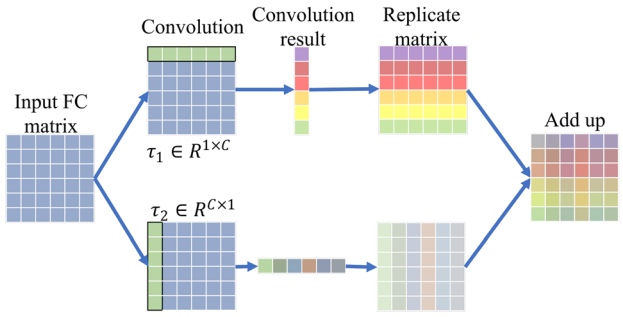


Fig. 6 A schematic depiction of the E2E block in BrainNet CNN. The output of the block is the sum of two convolution results

Results

The experiments aim to explore the effects of data augmentation on the small AD dataset with the decomposition and recombination strategy using FC matrices as inputs and with three different neural networks as classifiers. In Table 4, the number of trials in the training and testing sets is given. In the training set, 10 trials are randomly selected from the original EEG signals of AD patients and controls to avoid the imbalance of the training set. Five hundred artificial trials are generated from the 10 original trials for each class. The rest of the original trials are used in the testing set. The chance level is calculated with the stratified dummy classifier in Python's scikit-learn toolbox [41]. The training set consists of both original and artificial EEG signals. Artificial EEG signals in the training set are generated exclusively from the real EEG data of this set (Fig. 5). The original EEG signals in the training set

Table 2 The structure of BrainNet CNN

Layer	Output size	Parameter
Input layer	[*B*, 4, *C*, *C*]	
BatchNorm	[*B*, 4, *C*, *C*]	
ReLU	[*B*, 4, *C*, *C*]	
E2E	[*B*, 16, *C*, *C*]	(*C*, 1)
BatchNorm	[*B*, 16, *C*, *C*]	
ReLU	[*B*, 16, *C*, *C*]	
E2E	[*B*, 32, *C*, *C*]	
ReLU	[*B*, 32, *C*, *C*]	
E2N	[*B*, 64, *C*, 1]	(1, *C*)
N2G	[*B*, 512, 1, 1]	(*C*, 1)
Flatten	[*B*, 512]	
Linear and softmax	[*B*, 2]	

are randomly selected ten times, and the classification is repeated as a cross-validation procedure.

Feature Distribution

First, the feature distributions of the artificial data generated by data augmentation are assessed. To clearly illustrate this, the FC matrices of mild AD patients vs controls are depicted in Fig. 8 using the uniform manifold approximation and projection method (UMAP) [42, 43]. There are four FC matrices for each trial: 4–8 Hz, 8–10 Hz, 10–13 Hz, and 13–30 Hz. Since FC matrices are symmetric, their upper triangle is taken as the feature of said matrix. For each trial, we have $4 \times C \times (C-1)/2$ features. The UMAP model is first trained with features from 10 mild AD trials and 10 control trials. Then, 100 artificial mild AD trials and 100 artificial control trials generated with SEMD-MEMD, SEMD, or CEMD are transformed with the trained UMAP model. In the UMAP setting, the size of the local neighborhood used for manifold approximation is set to 10, and the effective minimum distance between embedded points is set to 1; the training epoch number for embedding optimization is 1000. The dimension of the features is reduced and projected onto

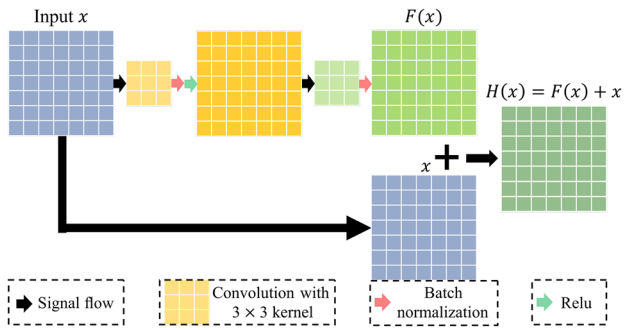


Fig. 7 A residual block with a shortcut in ResNet

a two-dimensional map with UMAP. Figure 8 shows that artificial data of the two classes generated with MEMD are more easily separable than those generated with SEMD or CEMD.

Performance Analysis

The evolution of the classification accuracy of the classifiers during the training process is depicted in Fig. 9. The training set is augmented with SEMD-MEMD. For the mild AD dataset, EEGNet has the worst classification performance with an average accuracy of around 53%. The data augmentation deteriorates the performance of EEGNet compared to the case of not using artificial data. On the other hand, the classification accuracy for BrainNet CNN improves with data augmentation when the number of artificial trials is greater than 20, as the accuracy converges faster than without data augmentation. The ResNet performance also improves with data augmentation.

In Fig. 10, the trend of the accuracy of the classification is given. The accuracies of ResNet and BrainNet CNN in this figure are obtained after 100-epoch training, while the number of training epochs of EEGNet is 200. We note that data augmentation does not always help to improve the training of neural networks.

Finally, Fig. 11 shows the confusion matrices, with only real data (before) or with 10 artificial trials per class (after), respectively. The number of artificial trials generates an increase of 100% of samples in the training dataset (factor of 2). These confusion matrices are calculated using MATLAB's "confusionmat" function [44]. The results were

Table 3 The structure of EEGNet

Layer	Output size	Parameter
Input layer	[*B*, 1, *C*, *T*]	
ZeroPad2d	[*B*, 1, *C*, *T*+63]	(31, 32, 0, 0)
Conv2d	[*B*, 8, *C*, *T*]	(1, 64)
BatchNorm2d	[*B*, 8, *C*, *T*]	
Conv2d	[*B*, 16, 1, *T*]	(*C*, 1), *grouped*
BatchNorm2d	[*B*, 16, 1, *T*]	
ELU	[*B*, 16, 1, *T*]	
AvgPool2d	[*B*, 16, 1, *T*//4]	(1, 4)
Dropout	[*B*, 16, 1, *T*//4]	0.25
ZeroPad2d	[*B*, 16, 1, *T*//4+15]	(7, 8, 0, 0)
Conv2d	[*B*, 16, 1, *T*//4]	(1, 15), *grouped*
Conv2d	[*B*, 16, 1, *T*//4]	(1, 1)
BatchNorm2d	[*B*, 16, 1, *T*//4]	
ELU	[*B*, 16, 1, *T*//4]	
AvgPool2d	[*B*, 16, 1, *T*//32]	(1, 8)
Dropout	[*B*, 16, 1, *T*//32]	0.25
Flatten	[*B*, 16**T*//32]	
Linear	[*B*, *K*]	*bias* = *False*

Table 4 Distribution of the number of trials

	Training set		Testing set	Chance level
Data type	Artificial	Original	Original	
Data type	Artificial	Original	Original	
Mild AD	0–500	10	7	0.3333
Control	0–500	10	14	
MCI	0–500	10	12	0.3000
Control	0–500	10	28	

obtained by averaging over ten folds, and the final values were normalized by dividing by the sum of each row. The experiment was carried out using only a small number of artificial trials, as the results depicted in Fig. 10 pointed out that this was a good value in almost all the models. Table 5 contains the accuracy, sensitivity, and precision calculated using the "confusionmat" function.

Discussion

In this work, we proposed a decomposition and recombination system to enlarge the size of two AD datasets and explored the data augmentation performance on three different neural networks. This work is based on the following two assumptions:

1. The AD dataset is a small dataset.
2. Neural networks need a considerable amount of data to tune the parameters.

Most patients affected by AD are elderly people. In contrast to the EEG signal acquisition of healthy people, AD patients are easily exhausted, weak, or less willing in the process of acquiring EEG signals. Sometimes, the acquisition can even be interrupted for unexpected reasons such as the non-collaboration of the patients. Therefore, AD datasets are very valuable and are usually small in size. To protect the health of the patients and to facilitate data acquisition in experiments, a data augmentation method is needed to process small AD datasets.

When it comes to the second assumption, note that deep neural networks can accurately find the unknown relationship between the raw data and the corresponding labels because of their intrinsic nature and huge number of parameters. At the same time, these parameters can only be learned from the available data, but the higher the number of parameters, the higher the number of signals needed to train the model. Therefore, data augmentation on small AD datasets is again of great interest.

In addition to the decomposition and recombination strategy in data augmentation, generative adversarial networks (GANs) are also a universal solution for time series data augmentation. However, in these, both the generator and discriminator parameters require a certain amount of data to be tuned. For an AD dataset of limited size, this requirement on the amount of data is not met, and hence, GANs are not suitable in this case.

In the classification of mild AD, data augmentation has a positive effect on the training of ResNet. When the number of artificial trials increases, the average accuracy of ResNet increases from 72.38 to 77.62%, with a consistent performance. In the BrainNet CNN case, a positive outcome is also obtained in the classification performance when using data augmentation in the mild AD dataset. However, this effect is only positive for a small number of artificial trials in the MCI dataset; if the number of artificial trials increases above 30, the mean accuracy decreases. Finally, the EEGNet network is the one with the poorest results for the mild AD dataset, and artificial trials only have a moderate positive effect for the MCI dataset again when the number of artificial trials is small.

In Fig. 11, the confusion matrices before and after data augmentation are given. Both ResNet and BrainNet CNN obtain a consistent accuracy, sensitivity, and precision

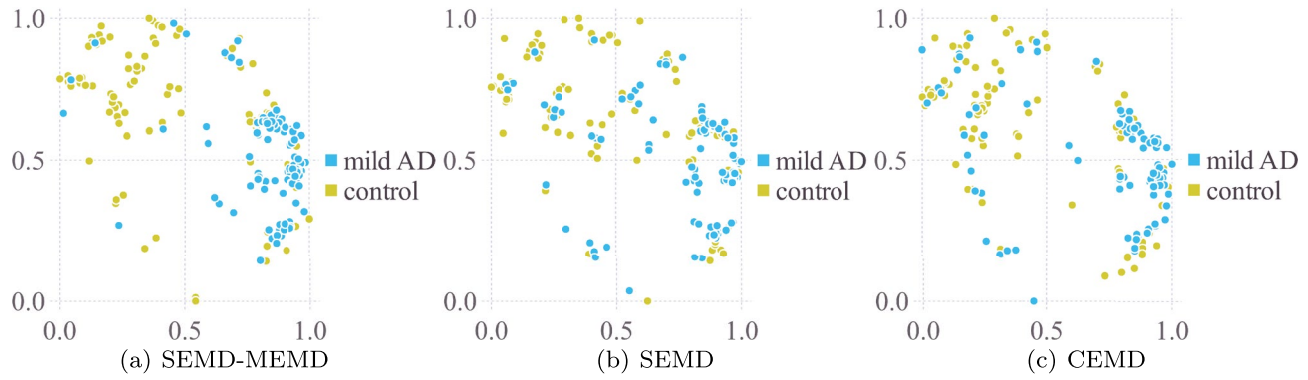


Fig. 8 Feature map of artificial mild AD patients vs controls, plotted with UMAP. For each class, 100 artificial samples are generated using MEMD **a**, SEMD **b**, and CEMD **c**. The obtained embedding is normalized with min-max normalization before visualization

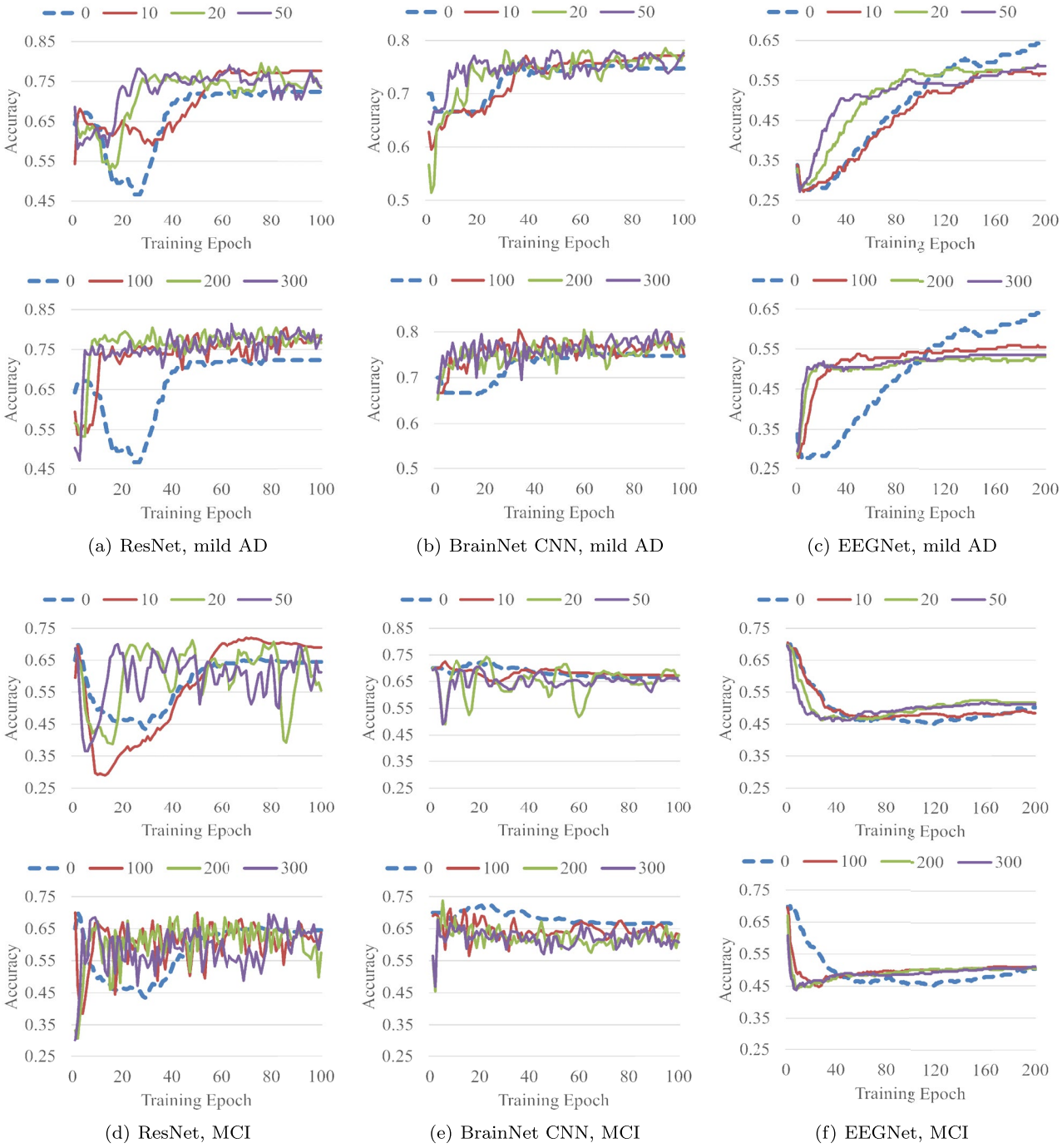


Fig. 9 The testing accuracy averaged across ten folds of the two datasets during the training process. A different number of artificial trials are generated, each one of them shown in a different color in the subplots. For each case, the upper panel contains the experiments with 0 to 50 artificial trials, and the lower panel contains the experiments with 0 to 500 artificial trials. The dashed line represents the 0 case, where no artificial trials are used

increase when 10 artificial trials per class are used. As expected, the improvement is more noticeable in the mild AD database, as the two classes (controls and patients) are more distant from each other when compared to the MCI case, in which the patients are closer to the control subjects.

Summarizing the above experiments, the proposed decomposition and recombination system helps the training of neural networks in small AD datasets, and it seems that just a factor of 2 is enough for that. Having more artificial data does not always provide a better result, as we have seen

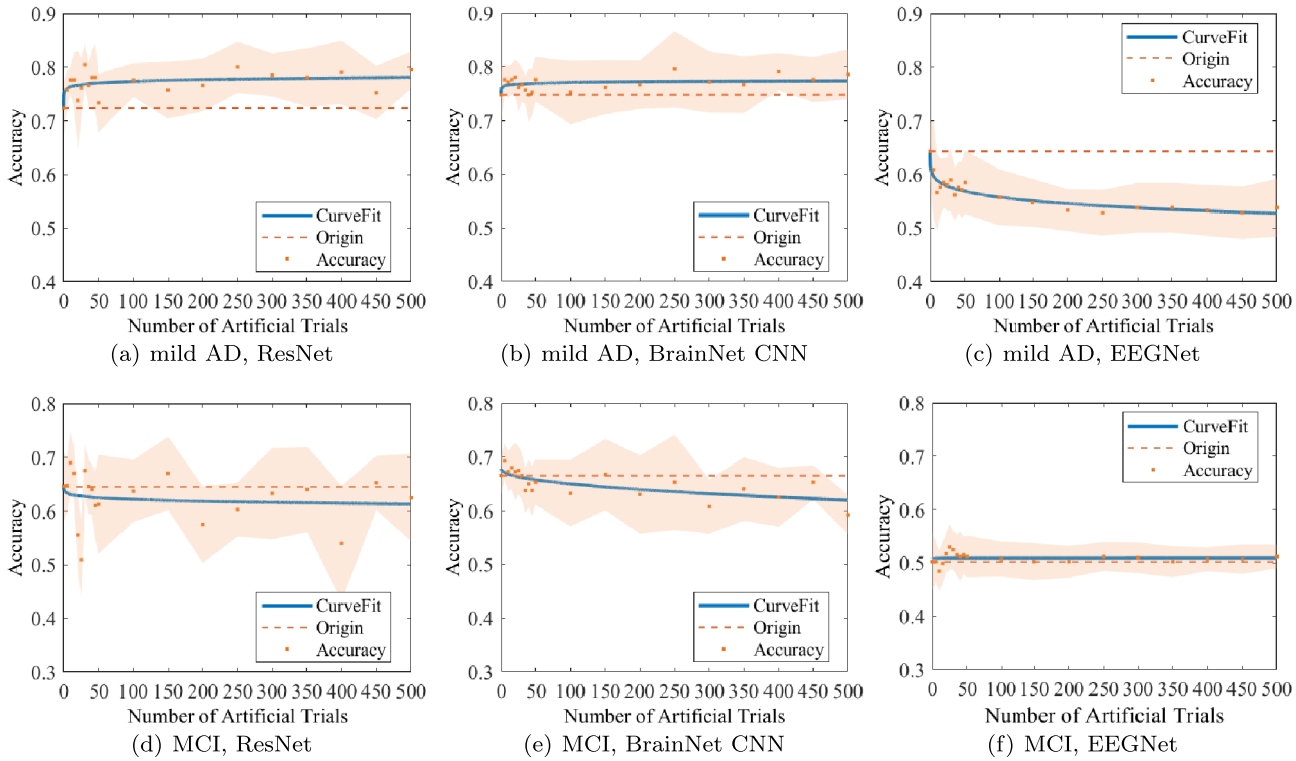


Fig. 10 Accuracy evolution when the number of artificial trials increases from 0 to 500. The trend of the accuracy is fitted with the power function $f(x) = ax^b + c$. The dotted line represents the accuracy without data augmentation

in our experiments. The effects of the data augmentation depend on two factors: (i) the type of neural networks and (ii) the data set. Determining the number of artificial trials is influenced by these two factors, and ascertaining how to obtain an optimal value requires further experiments.

One possible reason for why the proposed data augmentation method does not always improve the accuracy results is due to the different characteristics of the two datasets. In Fig. 9a, the accuracy of ResNet in the mild AD dataset converges as the number of training epochs increases, and the result is stable in the training, with a small variance around the mean accuracy. However, in Fig. 9d, the accuracy in the MCI dataset still fluctuates in a larger range, especially compared with the mild AD

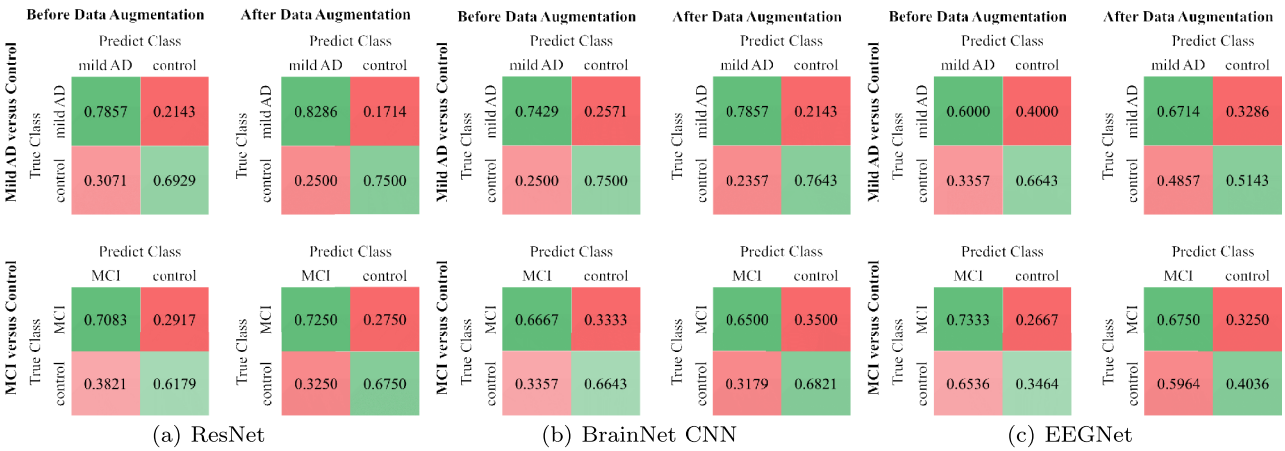


Fig. 11 Comparison of the confusion matrices before and after data augmentation for the two datasets. The confusion matrices are averaged across ten folds and normalized by dividing by the sum of each row

Table 5 Performance measurement before and after data augmentation, calculated using the "confusionmat" function in Fig. 11. The best result in each case is highlighted in bold

ResNet		Accuracy	Sensitivity	Precision
Mild AD	Before	0.7238	0.7393	0.7243
	After	**0.7762**	**0.7893**	**0.7628**
MCI	Before	0.6450	0.6631	0.6450
	After	**0.6900**	**0.7000**	**0.6775**
BrainNet CNN		Accuracy	Sensitivity	Precision
Mild AD	Before	0.7476	0.7464	0.7338
	After	**0.7714**	**0.7750**	**0.7850**
MCI	Before	0.6650	0.6655	0.6469
	After	**0.6725**	**0.6661**	**0.6517**
EEGNet		Accuracy	Sensitivity	Precision
Mild AD	Before	**0.6429**	**0.6321**	**0.6336**
	After	0.5667	0.5929	0.5887
MCI	Before	0.4625	**0.5399**	**0.5399**
	After	**0.4850**	0.5393	0.5337

dataset. This means that the network is more difficult to fit for the MCI dataset or that perhaps the quality of the data is also worse in that case. Although data augmentation improves the accuracy in the MCI dataset very slightly when the number of artificial trials is small, it still helps to train the ResNet: when the accuracy converges, the number of training epochs needed after data augmentation is smaller than without data augmentation, as shown in Fig. 9d. Similar fluctuations can be observed for the BrainNet CNN network in both datasets (Fig. 9b, e). This could explain why data augmentation is not helping in this case.

The proposed decomposition and recombination system has its own limitations. No pre-processing was used to remove artifacts or noise in the databases used in the experiments. Since the proposed method recombines all existing information in the data to enlarge the size of the training data, it is possible that artifacts or noise may also be replicated, which would negatively affect the results. Another aspect that can play a role is the decomposition method used. Here, we combine SEMD and MEMD, but other EMD-based methods have been proposed in the literature. Each method has different properties which impact the frequency mixing effect (overlapping of IMFs) and hence may influence the quality of the artificial frames. Moreover, the number of required artificial trials is unknown, as has been shown, and should be further investigated. More experiments are also needed to determine the number of epochs in the training phase, as our results indicate that the use of artificial trials may help to reduce the number of epochs in training and thus control possible overfitting. All of these aspects are now under consideration, and we expect to propose more reliable methods in future works.

Conclusion

In this paper, we proposed a decomposition and recombination system for data augmentation of the small AD data set as a way to solve the problem of insufficient data in neural network training.

This system consists of signal decomposition with SEMD-MEMD and a random recombination of the decomposed IMFs. The performance of this system is evaluated using three classifiers on two datasets. The main results show that the proposed system improves the accuracy of ResNet on the mild AD dataset with an increase of 5.24% and on the MCI dataset with an increase of 4.50%. Furthermore, BrainNet CNN results improve on the mild AD dataset with an increase of 2.38% and an increase of 0.75% on the MCI dataset. This work is expected to help the training process of detection methods for early diagnosis of Alzheimer's disease.

Acknowledgements This work was carried out as part of the doctoral program in Experimental Science and Technology at the University of Vic - Central University of Catalonia. The authors would like to thank Pau Solé-Vilaró for the English proofreading of this manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was supported by the National Natural Science Foundation of China (Key Program) (No. 11932013) and the Tianjin Science and Technology Plan Project (No. 22PTZWHZ00040). J.S-C. work is also based upon work from COST Action CA18106, supported by COST (European Cooperation in Science and Technology). C.F.C work is partially supported by grants PICT 2020-SERIEA-00457 and PIP 112202101 00284CO (Argentina).

Data Availability Not applicable. No new data were collected for this study.

Code Availability The code used for the data augmentation is freely available here: https://github.com/Sole-Casals/DR-EMD.

Declarations

Informed Consent Not applicable. No new data were collected for this study.

Conflict of Interest The authors declare no competing interests.

Research Involving Human Participants and/or Animals Not applicable. No new data were collected for this study.

References

1. International AD. World Alzheimer report 2019: attitudes to dementia. Alzheimer's Disease International London, UK; 2019.
2. Du X, Wang X, Geng M. Alzheimer's disease hypothesis and related therapies. Translational neurodegeneration. 2018;7(1):1–7.
3. The Alzheimer Association. Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2023Apr;19(4):1598–695. Available from: https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13016.
4. Hansson O, Edelmayer RM, Boxer AL, Carrillo MC, Mielke MM, Rabinovici GD, et al. The Alzheimer's Association appropriate use recommendations for blood biomarkers in Alzheimer's disease. Alzheimer's & Dementia. 2022 Dec;18(12):2669-86. Available from: https://onlinelibrary.wiley.com/doi/10.1002/alz.12756.
5. Musha T, Asada T, Yamashita F, Kinoshita T, Chen Z, Matsuda H, et al. A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease. Clinical Neurophysiology. 2002Jul;113(7):1052-8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1388245702001281.
6. Gallego-Jutglà E, Elgendi M, Vialatte F, Solé-Casals J, Cichocki A, Latchoumane C. Diagnosis of Alzheimer's disease from EEG by means of synchrony measures in optimized frequency bands. In, et al. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. 2012;2012:4266–70.
7. Gallego-Jutglà E, Solé-Casals J, Vialatte FB, Elgendi M, Cichocki A, Dauwels J. A hybrid feature selection approach for the early diagnosis of Alzheimer's disease. Journal of Neural Engineering. 2015;12(1).
8. Yu H, Lei X, Song Z, Liu C, Wang J. Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification. IEEE Transactions on Fuzzy Systems. 2019.
9. Mandal PK, Banerjee A, Tripathi M, Sharma A. A comprehensive review of magnetoencephalography (MEG) studies for brain functionality in healthy aging and Alzheimer's disease (AD). Frontiers in Computational Neuroscience. 2018;12:60.
10. Yang S, Bornot JMS, Wong-Lin K, Prasad G. M/EEG-based biomarkers to predict the MCI and Alzheimer's disease: a review from the ML perspective. IEEE Transactions on Biomedical Engineering. 2019;66(10):2924–35.
11. Chen X, Zhang H, Gao Y, Wee CY, Li G, Shen D, et al. High-order resting-state functional connectivity network for MCI classification. Human Brain Mapping. 2016;37(9):3282–96.
12. Chen X, Zhang H, Lee SW, Shen D, Initiative ADN, et al. Hierarchical high-order functional connectivity networks and selective feature fusion for MCI classification. Neuroinformatics. 2017;15(3):271–84.
13. Zhang Y, Zhang H, Chen X, Liu M, Zhu X, Lee SW, et al. Strength and similarity guided group-level brain functional network construction for MCI diagnosis. Pattern Recognition. 2019;88:421–30.
14. Sarraf S, Tofighi G. Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. 2016. arXiv preprint http://arxiv.org/abs/1603.08631arXiv:1603.08631.
15. Kim D, Kim K. Detection of early stage Alzheimer's disease using EEG relative power with deep neural network. In,. 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. 2018;2018:352–5.
16. Duan F, Huang Z, Sun Z, Zhang Y, Zhao Q, Cichocki A, et al. Topological network analysis of early Alzheimer's disease based on resting-state EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020.
17. Vialatte FB, Solé-Casals J, Maurice M, Latchoumane C, Hudson N, Wimalaratna S, et al. Improving the quality of EEG data in patients with Alzheimer's disease using ICA. In: International Conference on Neural Information Processing. Springer. 2008;979–86.
18. Vialatte FB, Solé-Casals J, Cichocki A. EEG windowed statistical wavelet scoring for evaluation and discrimination of muscular artifacts. Physiol Measure. 2008;29(12):1435.
19. Sanchez-Poblador V, Monte-Moreno E, Solé-Casals J. ICA as a preprocessing technique for classification. In: International Conference on Independent Component Analysis and Signal Separation. Springer. 2004;11657–2.
20. Solé-Casals J, Vialatte FB. Towards semi-automatic artifact rejection for the improvement of Alzheimer's disease screening from EEG signals. Sensors. 2015;15(8):17963–76.
21. Solé-Casals J, Caiafa CF, Zhao Q, Cichocki A. Brain-computer interface with corrupted EEG data: a tensor completion approach. Cognitive Computation. 2018;10(6):1062–74.
22. Caiafa CF, Solé-Casals J, Marti-Puig P, Zhe S, Tanaka T. Decomposition methods for machine learning with small. Incomplete or Noisy Datasets Applied Sciences. 2020;10(8481):1–20.
23. Alzubaidi L, Bai J, Al-Sabaawi A, Santamaría J, Albahri AS, Al-dabbagh BSN, et al. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. J Big Data. 2023;10(1):46. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00727-2.
24. Hazra D, Byun YC. SynSigGAN: generative adversarial networks for synthetic biomedical signal generation. Biology. 2020;9(12):441.
25. Bhattacharyya A, Singh L, Pachori RB. Fourier-Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals. Digit Signal Process. 2018;78:185–96.
26. Zhao X, Solé-Casals J, Li B, Huang Z, Zhao Q. Classification of epileptic IEEG signals by CNN and data augmentation. In: ICASSP 2020. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020.
27. Zhang Z, Duan F, Solé-Casals J, Dinares-Ferran J, Cichocki A, Yang Z, et al. A novel deep learning approach with data augmentation to classify motor imagery signals. IEEE Access. 2019;1.
28. Li B, Zhang Z, Duan F, Yang Z, Zhao Q, Sun Z, et al. Component-mixing strategy: a decomposition-based data augmentation algorithm for motor imagery signals. Neurocomputing. 2021;465:325–35. Available from: https://www.sciencedirect.com/science/article/pii/S0925231221013308.
29. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings Mathematical Physical & Engineering Sciences. 1971;1998(454):903–95.
30. Haradal S, Hayashi H, Uchida S. Biosignal data augmentation based on generative adversarial networks. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Honolulu, HI: IEEE; 2018;368–71. Available from: https://ieeexplore.ieee.org/document/8512396/.
31. Gallego-Jutgla E, Solé-Casals J, Vialatte FB, Dauwels J, Cichocki A. A theta-band EEG based index for early diagnosis of Alzheimer's disease. Journal of Alzheimer's Disease. 2015;43(4):1175–84.

32. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. Journal of the American Geriatrics Society. 1992;40(9):922-35. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.1992.tb01992.x.
33. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. International Psychogeriatrics. 1997;9(S1):173–6.
34. Venkatesh M, Jaja J, Pessoa L. Comparing functional connectivity matrices: a geometry-aware approach applied to participant identification. NeuroImage. 2020;207:116398. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1053811919309899.
35. Rodriguez G, Arnaldi D, Picco A. Brain functional network in Alzheimer's disease: diagnostic markers for diagnosis and monitoring. Int J Alzheimer's Dis. 2011.
36. Kawahara J, Brown CJ, Miller SP, Booth BG, Chau V, Grunau RE, et al. BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage. 2017;146(1038):1049.
37. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770-8.
38. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional network for EEG-based brain-computer interfaces. J Neural Eng. 2018;15(5):056013. Available from: http://arxiv.org/abs/1611.08024.
39. Ho MC, Chen TC, Huang CF, Yu CH, Chen JM, Huang RY, et al. Detect AD patients by using EEG coherence analysis. J Med Eng. 2014.
40. Hochreiter S. Recurrent neural net learning and vanishing gradient. International Journal Of Uncertainity, Fuzziness and Knowledge-Based Systems. 1998;6(2):107–16.
41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011;12:2825–30.
42. Mcinnes L, Healy J. UMAP: uniform manifold approximation and projection for dimension reduction. The Journal of Open Source Software. 2018;3(29):861.
43. McInnes L, Healy J, Saul N, Grossberger L. UMAP: uniform manifold approximation and projection. The Journal of Open Source Software. 2018;3(29):861.
44. Inc TM. MATLAB version: 9.9.0 (R2020b). Natick, Massachusetts, United States: The MathWorks Inc. 2020. Available from: https://www.mathworks.com.