Data Augmentation

Data augmentation is the technique of increasing the size of data used for training a model. For reliable predictions, deep learning models often require a lot of training data, which is not always available. Therefore, the existing data is augmented in order to make a better-generalised model.

<https://iq.opengenus.org/data-augmentation/>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **When Should You Use Data Augmentation?**

1. To prevent models from overfitting.
2. The initial training set needs to be bigger.
3. To improve the model accuracy.
4. To Reduce the operational cost of labelling and cleaning the raw dataset.

<https://www.datacamp.com/tutorial/complete-guide-data-augmentation>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Techniques used in “Data augmentation for deep-learning-based electroencephalography“

<https://www.sciencedirect.com/science/article/abs/pii/S0165027020303083?via%3Dihub>

* Noise addition
* Generative adversarial networks
* Sliding windows
* Sampling
* Fourier transform
* Recombination of segmentation
* And others

Result

-Noise addition and sliding windows provided the highest accuracy boost; mental workload most benefited from DA. Sliding windows, noise addition, and sampling methods are most common for seizure detection, mental workload, and sleep stages, respectively.

### 

### Comparing with existing methods

Percent of decoding accuracy explained by DA beyond unaugmented accuracy varied between 8 % for recombination of segmentation and 36 % for noise addition and from 14 % for motor imagery to 56 % for mental workload—29 % on average.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# 

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

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10070085/>

Mentioned more DA technique: reorganisation

And the results show that compared with other data augmentation methods in various EEG tasks, sliding window boasts the optimal effect

(didn’t find any related information in this ref:<https://www.sciencedirect.com/science/article/pii/S1877042813037749?via%3Dihub> no:30)

The results show that while decreasing the step length the more data samples we will have to train on, furthermore the less step size gets the best average accuracy, and when using a bigger sliding window size with less step length the accuracy gets more accurate.

How the data sets are divided for training and testing

The accurately divided data were input into the deep learning model for 5-fold cross-validation and the model performance was evaluated by a confusion matrix.

The difference between traditional algorithms and time series domain algorithms :

not all techniques can be applied to every dataset.

<https://link.springer.com/article/10.1007/s00521-023-08459-3>

Paper Conclusion :

Due to the significant evolution that DA has undergone in recent years, more and more fields are emerging in which to apply and improve the results. This article is focused on giving a comprehensive overview of the main algorithms used for data augmentation (DA) in the field of time series. The review is organized in a taxonomy, consisting of basic and advanced approaches, where it summarises representative methods of each algorithm (traditional, VAEs, and GANs) comparing them empirically, disaggregating by application areas, and highlighting advantages/disadvantages for future research.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

“**Data Augmentation Effects on Highly Imbalanced EEG Datasets for Automatic Detection of Photoparoxysmal Responses”(2023)**

One-Hidden-Dense-Layer Neural Network were employed to evaluate the performance of this DA stage

<https://www.mdpi.com/1424-8220/23/4/2312>

-performance improvement of around 20% for the Accuracy and Specificity measurements without Sensitivity suffering any losses.

ad hoc method for DA that merges two actual PPR windows. The idea is to split the selected windows into **n** intervals, generating a new PPR window by collecting alternating intervals from each parent, a similar approach to the recombination method applied in [[**40**](https://www.mdpi.com/1424-8220/23/4/2312#B40-sensors-23-02312)].

(similar to sliding windows)(method for images not numerical)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DA: autoencoder (AE) and generative adversarial networks (GAN).

Mentioned in

<https://www.frontiersin.org/articles/10.3389/fnhum.2021.765525/full>

<https://link.springer.com/article/10.1007/s00521-023-08459-3>