Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis

Berardino Barile, Aldo Marzullo, Claudio Stamile, Françoise Durand-Dubief, Dominique Sappey-Marinier

▶ **To cite this version:**

Berardino Barile, Aldo Marzullo, Claudio Stamile, Françoise Durand-Dubief, Dominique Sappey-Marinier. Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis. Computer Methods and Programs in Biomedicine, 2021, 206, pp.106113. 10.1016/j.cmpb.2021.106113 . hal-03241649

HAL Id: hal-03241649
https://hal.science/hal-03241649

Submitted on 28 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Data Augmentation using Generative Adversarial Neural Networks on Brain Structural Connectivity in Multiple Sclerosis

Berardino Barile[a], Aldo Marzullo[b], Claudio Stamile[c], Françoise Durand-Dubief[a,d] and Dominique Sappey-Marinier[a,e]

[a]CREATIS (UMR 5220 CNRS & U1206 INSERM), Université Claude Bernard Lyon 1, Université de Lyon, Villeurbanne, France
[b]Department of Mathematics and Computer Science, University of Calabria, Rende, Italy
[c]R&D Department CGnal, Milan, Italy
[d]Hôpital Neurologique, Service de Neurologie A, Hôpital Civils de Lyon, Bron, France
[e]CERMEP - Imagerie du Vivant, Université de Lyon, Bron, France

ARTICLE INFO

Keywords:
Brain Connectivity
Multiple Sclerosis
Data Augmentation
Generative Adversarial Networks

ABSTRACT

Background and objective: Machine learning frameworks have demonstrated their potentials in dealing with complex data structures, achieving remarkable results in many areas, including brain imaging. However, a large collection of data is needed to train these models. This is particularly challenging in the biomedical domain since, due to acquisition accessibility, costs and pathology related variability, available datasets are limited and usually imbalanced. To overcome this challenge, generative models can be used to generate new data.
Methods: In this study, a framework based on generative adversarial network is proposed to create synthetic structural brain networks in Multiple Sclerosis (MS). The dataset consists of 29 relapsing-remitting and 19 secondary-progressive MS patients. T1 and diffusion tensor imaging (DTI) acquisitions were used to obtain the structural brain network for each subject. Evaluation of the quality of newly generated brain networks is performed by *(i)* analysing their structural properties and *(ii)* studying their impact on classification performance.
Results: We demonstrate that advanced generative models could be directly applied to the structural brain networks. We quantitatively and qualitatively show that newly generated data do not present significant differences compared to the real ones. In addition, augmenting the existing dataset with generated samples leads to an improvement of the classification performance ($F1_{score}$ 81%) with respect to the baseline approach ($F1_{score}$ 66%).
Conclusions: Our approach defines a new tool for biomedical application when connectome-based data augmentation is needed, providing a valid alternative to usual image-based data augmentation techniques.

1. Introduction

Artificial intelligence has revolutionized many areas of research, from economics and law to health-care. However, a large collection of data is essential for statistical evaluation and machine learning applications, particularly in the field of deep learning (DL). Indeed, DL frameworks have achieved remarkable results in many fields, such as pattern recognition, natural language processing, image processing, among others [1, 2]. The main advantage of using DL applications lies in their great ability to recognize hidden patterns in the data, thanks to the multiple nonlinear transformations produced by the sequential stacks of multiple layers [3, 4] However, huge amount of data are required for training this kind of models while in the context of biomedical domain, and particularly in medical imaging, extensive datasets are challenging to obtain due to systems availability, costs constraints, acquisition methodology, and pathology related variability [5, 6], resulting in small and imbalanced dataset. Notwithstanding, when dealing with image data, different solutions have been proposed to overcome these limitations [7]. A general and widely accepted solution is to impose meaningless perturbations to the original data [8] or to apply more advanced techniques, like rotation, reflection, scaling among others. These approaches offer straightforward alternatives for augmenting the training set, allowing DL models to reach better performance and/or more stable training [9]. Recently, with the rise of DL, interesting alternatives have appeared and new generative DL-based models were proposed to obtain synthetic data with characteristics spanning the original data manifold [10]. Therefore, in this study we refer to *generative models* as a subclass of DL frameworks able to generate complex data data structure, including the recent modeling approach used to characterize brain networks by means of graph theory [11, 12, 13]. Given the great capability of graphs to represent complex relations among different areas of the brain, such relational data structure started to be widely employed in many contexts, including social behavioral studies. Additionally, advances in brain image acquisition and computer assisted methods have begun to provide meaningful results in support of clinicians, leading to a steadily growing use in the neuroscience community, particularly in brain imaging [14]. Using magnetic resonance imaging (MRI), functional or structural brain connectivity can be obtained by analyz-

berardino.barile@creatis.insa-lyon.fr (B. Barile); marzullo@mat.unical.it (A. Marzullo); cstamile@cgnal.com (C. Stamile); francoise.durand-dubief@chu-lyon.fr (F. Durand-Dubief); sappey-marinier@univ-lyon1.fr (D. Sappey-Marinier)
ORCID(s):

ing temporal correlations of gray matter (GM) activity with resting-state functional MRI (fMRI) or reconstructing white matter (WM) fiber-bundles with diffusion tensor imaging (DTI), respectively. Such network-like structure of the human connectome consists of nodes, defined by parcellisation of the brain grey matter (GM), and edges, corresponding to functional or structural links between the network nodes. These new approaches paved the way for a better characterization of brain networks, particularly in brain diseases such as Multiple Sclerosis (MS).

MS is a demyelinating, inflammatory, chronic disease of the central nervous system [15]. While its etiology remains unknown, MS is the most frequent disabling neurological disease in young adults. Disease onset is characterized in about 85% of cases [15], by a first acute episode called clinically isolated syndrome (CIS) or a relapsing-remitting course (RRMS) followed by a secondary-progressive course (SPMS), while the remaining 15% of MS patients evolve directly into a primary-progressive course (PPMS). The course of the disease and the risk for developing permanent disability are very different from one patient to another. Thus, the neurologist's challenge is to predict the disease evolution based on early clinical, biological and imaging markers available from disease onset. However, the complexity brought by conectome data is more cumbersome with respect to the grid-like pixel-by-pixel representation found in images. In fact, due to the multiple interconnections between different nodes, connectome data represent a challenge for synthetic data generation for which simple operations, like edge swapping, would end up changing the entire structure of the graph network, jeopardizing the information they convey. [16].

In this study, a generative adversarial network framework is proposed, namely Generative Adversarial Neural Network AutoEncoder (AAE). The framework is able to automatically generate synthetic structural brain connectivity data of MS patients. To achieve this, a prior is imposed to the latent space of the autoencoder network by means of an adversarial model. Moreover, a consistency loss is also introduced in order to increase the stability of the training process. New samples of brain connectivity data are generated by drawing from the parametrized latent space. An overfitting analysis over generated graphs, by exploiting graph properties, is proposed for model evaluation. The synthetic generated data can be used to augment the MS brain networks dataset to improve classification performances of classical machine learning methods like the Random Forest Classifier.

The paper is structured as follows. In Section 2, we illustrate the related literature, and in section 3, we provide a detailed description of our methodological approach. In Section 4, we describe our experimental results and finally, in Section 5, we draw our conclusions.

2. Related Work

Due to their ability to generate new data, generative models have gained a lot of interest in the computer vision and medical imaging research communities. The Generative Adversarial Network framework (GAN) has been previously used for generating realistic training images that synthetically augment datasets. Radford *et al.* [17] introduced a class of generative model called deep convolutional generative adversarial networks (DCGAN) to generate 2D brain MR images followed by an AE neural network for image denoising. Makhzani *et al.* [18] proposed a new method for regularizing AutoEncoders (AE) by imposing an arbitrary prior on the latent representation. Calimeri *et al.* [19] proposed a GAN for the automatic generation of artificial MR images of the human brain. They demonstrated that the power of adversarial training could be exploited for the generation of brain networks data, which are more complex than usual images.

GAN frameworks have also shown to improve accuracy of image classification via generation of new synthetic training images. Frid-Adar *et al.* [20], for instance, used synthetic medical image augmentation with GAN for the classification of liver lesions. Similarly, Salehinejad *et al.* [21] used this framework to simulate pathology across five classes of chest X-rays in order to augment the original imbalanced dataset and improve the performance of a convolutional model in chest pathology classification. In the context of MS, Shui-Hua W. *et al.* [22] proposed a new transfer-learning-based approach to identify MS patients with higher accuracy, comparing three different types of neural networks (DenseNet-121, DenseNet-169, and DenseNet-201), which make use of composite learning factors to different layers. Yu-Dong Z. *et al.* [23] exploited the AlexNet model to classify MS patients and studied the best transfer-learning settings (i.e. number of layers transferred and replaced) obtaining high level of performance.

Interestingly, other applications of adversarial variational training frameworks have been reported. For example, Zhang *et al.* [24] proposed a semi-supervised learning Adversarial Variational Embedding for leveraging both the power of GAN as a high quality generative model and Variational AutoEncoder (VAE) as a posterior distribution learner. They demonstrated that the combination of VAE and GAN provided significant improvements of semisupervised classification. Imran *et al.* [25] used a network architecture that incorporates an ensemble of discriminators in a VAE-GAN network using datasets from the computer vision and medical imaging domains in order to generate new realistic images of medical data. They showed that the combination of this two generative models can lead to superior performances against state-of-the-art semi-supervised models both in image generation and classification tasks. However, the generation process become more cumbersome in the case of highly structured graph data. In order to address this challenge, many approaches have been reported. Chawla *et al.* [26] proposed a GraphVAE method for generating small graphs using a Variational approach. This model is composed by a simple linearized decoder output, which produces a probabilistic fully-connected graph. Pan *et al.* [27] proposed a new architecture for which an adversarial train-

ing is combined with a graph autoencoder structure (ARAE). The framework encodes the topological structure and node content in a graph to a compact representation, on which a decoder is trained to reconstruct the graph structure.
Freund *et al.* [28] proposed an approach based on adversarial regularization of the latent space for generating graph structured data. They could demonstrate the ability of the model to embed graph-based data coherently, and at the same time, generate meaningful samples. Thus, Graph AE and VAE constitute today the best approach for embedding nodes and learn a low dimensional vector representation with applications to link prediction, node clustering and matrix completion. However, much less attention has been spent on generating the entire structure of graphs. Khoshgoftaar *et al.* [29] proposed a simple graph AE structure which does not use the graph convolutional network. They demonstrated that a straightforward linear models with adjacency matrices as inputs performed equally well in benchmark datasets like Cora, Citeseer and Pubmed citation networks.

3. Materials and Methods

3.1. Dataset Description and Preprocessing

Forty-eight MS patients distributed across the two most frequent clinical courses, namely the RRMS course, which is followed, between 10 to 20 years later, by the SPMS course [1]. If these two clinical forms are distinguished by the status of the patient, mainly expressed by its Expanded Disability Status Scale (EDSS), they can also be differentiated by their biological and imaging markers reflecting two underlying pathological processes, such as inflammation, and neurodegeneration. Each patient underwent multiple brain MRI examinations over different periods, ranging from 2.5 to 6 years. The minimum number of scans per patient is 3 while the maximum is 10. The gap between two consecutive scans is either 6 or 12 months. The total number of MRI scans in the dataset is 270. This study was approved by the local ethics committee (CPP Sud-Est IV) and the French national agency for medicine and health products safety (ANSM). Written informed consent was obtained from all patients and the control subject prior to study initiation.
For each subject, the brain structural connectivity graphs are generated by combining brain GM parcellation extracted from T1-weighted MRI and the white matter (WM) fiber-tracking obtained from DTI acquisition. An undirected graph $G = (V, E)$ representing the WM fiber-tracks of the brain is created, where V defines the set of nodes (GM regions) and E represents the set of connections (WM fiber-tracks) between these nodes. Each graph G is represented by an adjacency matrix. More in detail, the data processing includes an atlas parcellation of cortical and sub-cortical GM regions, performed after segmentation of the T1-weighted MRI in four classes [WM, cortical GM, sub-cortical GM, cerebro-spinal fluid (CSF)], as described in Kocevar *et al.* [30]. Meanwhile, a pre-processing of the diffusion images is performed by applying the Eddy-current distortions correction [31] and a probabilistic streamline tractography algorithm is applied to generate WM fiber-tracks that combined with the T1-image parcellation leads to a symmetrical connectivity matrix $A \in \mathbb{N}_{+}^{q \times q}$ where $q = 84$ for each subject. Finally, due to the symmetry of the matrix, only the upper triangular part was considered in order to reduce the dimensionality of the problem. This implies that a single matrix can be represented as a vector $x \in \mathbb{R}^{(1,d)}$ where $d = 3486$ excluding the diagonal which is imposed to zero values.

Table 1
Population description by clinical profiles

	RRMS	SPMS
Patients (M\F)	29 (20\80)	19 (61\39)
Age at first scan (years)	35.1 (7.4)	42.3 (4.4)
Disease duration (years)	6.75 (4.81)	13.12 (5.84)
EDSS median (range)	2.0 (0-4.5)	5.0 (3-7)
Total number of scans	182	88

3.2. MRI Data Acquisition

MS patients underwent a MR examination on a 1.5T Siemens Sonata system (Siemens Medical Solution, Erlangen, Germany) using an 8-channel head-coil. The MR protocol consisted in the acquisition of a sagittal 3D-T1 sequence ($1 \times 1 \times 1\ mm^3$, $TE/TR = 4/2000\ ms$) and an axial 2D-spin-echo DTI sequence ($TE/TR = 86/6900\ ms$; 2×24 directions of gradient diffusion; $b = 1000\ s.mm^{-2}$, spatial resolution of $2.5 \times 2.5 \times 2.5\ mm^3$) oriented in the AC-PC plane.

3.3. Generative Adversarial Neural Network

GAN is a generative model approach based on differentiable neural networks where two actors are involved: a *Generator* $[G_{q_\theta(z|x)}(x)]$ and a *Discriminator* $[D(v)]$ [8]. The former is a neural network mapping the input x to the output z by training a network with structure q and parameters θ. In most applications, brand new data are generated by defining a prior on input noise variables. The latter is a network which takes as input v and outputs the probability that the input is coming from the true data distribution instead of being synthetically generated [19]. Formally, the adversarial game can be defined as a *min-max* problem following Eq. (1).

$$\min_G \max_D = \mathbb{E}_{v \sim p(v)}[log D(v)] + \mathbb{E}_{z \sim p(z)}[1 - log D(G_{q_\theta(z|x)}(x))] \quad (1)$$

Here, the first term represents the discriminator network's probability that true instances v from distribution $p(v)$ are rightly classified. The second term in the summation, identifies the generator network's ability to fool the discriminator by producing data with probability distribution $p(z)$ indistinguishable from that of the true data.

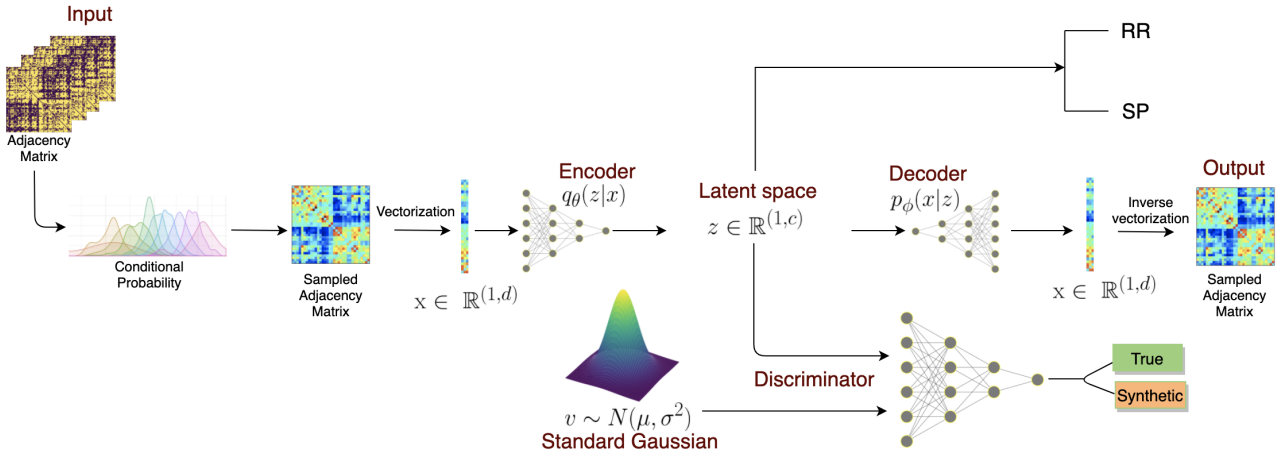

Figure 1: Schematic representation of the proposed AAE model. Starting from the brain connectome data representation (adjacency matrix), conditional probability distributions were calculated, from which new batches of connectome data were sampled. From the vectorized representation of the sampled adjacency matrix, the encoder network compresses the input into a latent lower dimensional representation, while the decoder reconstructs the input from its compressed latent representation. The combination of the two networks defines the autoencoder generator of the adversarial framework. Conversely, the discriminator network takes as input the latent representation and a random noise vector and tries to discriminate between the two, effectively imposing a constraint on the latent distribution of the autoencoder. Finally, from the latent space, an additional classifier discriminates between RRMS and SPMS patients.

3.4. Generative Adversarial Neural Network Autoencoder

In this study, the adversarial training is used to train the proposed AAE model at generating synthetic structural brain networks. Fig. 1, illustrates the adversarial process. The structure of the AAE model is defined by two adversarial neural networks: the generator and the discriminator. The former is an autoencoder composed by 13 layers for which fully connected and batch normalization layers alternate between one another except for the output layer. The input layer of the encoder is the number of upper triangular nodes in the graph ($d = 3486$). Subsequent fully connected layers have a number of neurons of 512, 256, 128, 100. This last encoder layer ($q_\theta(z|x)$) maps the input vector $x \in \mathbb{R}^{(1,d)}$ to a lower dimensional space $z \in \mathbb{R}^{(1,c)}$ with $c = 100$. The decoder $p_\phi(x|z)$ is defined as a mirror representation of the encoder with the aim of reconstructing the original input. Furthermore, the encoder is provided with an additional branch, a fully connected layer with a single neuron, used as regularization with respect to the clinical form.

On the other hand, a second neural network is introduced, which takes two inputs. The first is a random standard gaussian vector $v \in \mathbb{R}^{(1,c)}$ with $c = 100$ where:

$$v \sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{2}$$

with $\mu = 0$ and $\sigma = 1$. The second is $z \in \mathbb{R}^{(1,c)}$ obtained as the output of the encoder $q_\theta(z|x)$. The second model (discriminator) produces a probability score, which defines the likelihood that the two input vectors are coming from the same underlying data distribution. Its architecture is composed by 6 layers in which fully connected and dropout layers alternates between one another. The LeakyReLU activation function with an alpha parameter of 0.2 is used for all of the middle layers in both the generator and the discriminator while for the output layers a sigmoid activation function is employed. Only for the generator, batch normalisation with momentum 0.8 is added after each feedforward layer except for the output layer. For the discriminator, dropout with parameter 0.2 is used between each layer excluding the output layer. Finally, the encoder model $q_\theta(z|x)$ is connected to a second mirrored model $p_\phi(x|z)$, the decoder, whose objective is to learn the inverse mapping function of the encoder. In order to maximize the reconstruction quality of the input data, an additional penalty is imposed to final loss function (3).

$$MSE = \frac{1}{kn} \sum_{i=1}^{n} \sum_{j=1}^{k} \left(\frac{x_{ij} - p_\phi(x_{ij}|z_{ij})}{\sigma_{ij}} \right)^2 \tag{3}$$

where k defines the total number of possible connections in the upper triangular of the connectivity matrix while n defines the size of the batch used for training the network. In other words, we penalized the loss function each time the reconstructed matrix is far from the original data matrix in terms of mean squared error (MSE). This constraint ensures that while the hidden space z is forced to follow a standard normal distribution, the output of the model will produce results that span the entire input space. Moreover, in order to improve the classification performance we imposed an additional form of regularization which bind the latent space to be "coherent" [32] between the encoder and the decoder

network. Mathematically, this translates in Eq. (4):

$$MSE_{Coherence} = \frac{1}{cn}\sum_{i=1}^{n}\sum_{j=1}^{c}\left(\frac{q_\theta(z|x) - \Psi_{\theta_{ij}}}{IQR}\right)^2 \quad (4)$$

with $IQR = Q_3 - Q_1$ where Q_1 and Q_3 are respectively the first and the third quartile of the distribution given by $q_\theta(z|x) - \Psi_\theta$ where $\Psi_{\theta_{ij}} = q_\theta(z_{ij}|p_\phi(x_{ij}|q_\theta(z|x_{ij}))$.

As long as the adversarial loss is concerned, let $D(v)$ be the discriminator network, where $v \sim \mathcal{N}(\mu = 0, \sigma^2 = 1) = p(v)$, is a standard normal distribution. The related loss function will thus be defined as $\mathbb{E}_{v\sim p(v)}[log D(v)]$ for positive cases and $\mathbb{E}_{z\sim p(z)}[1-log D(G_{q_\theta(z|x)}(x))]$ for negative cases. In this last case, the generation of the latent space z is defined as $G_{q_\theta(z|x)}(x)$ with $z \sim q_\theta(z|x) = p(z)$. We would like that $p(z) \approx p(v)$, which implies that the latent space is distributed as a standard gaussian. On the contrary, we define with $G(x)$ the final generator (composed of an encoder and a decoder) and its respective loss function as $\mathbb{E}_{z\sim q_\theta(z|x)}[log\ p_\phi(x|z)]$. Henceforth, the adversarial loss will be defined as shown in Eq. (5).

$$L(\theta, D(v), G_{q_\theta(z|x)}(x)) = \mathbb{E}_{z\sim q_\theta(z|x)}[log\ p_\phi(x|z)] + \mathbb{E}_{v\sim p(v)}[log D(v)] + \mathbb{E}_{z\sim p(z)}[1 - log D(G_{q_\theta(z|x)}(x))] \quad (5)$$

Roughly speaking, the *Kullback-Leibler*, usually employed in a VAE framework [33], is now substituted with the adversarial loss. This model allows us to provide probabilistic descriptions of observations in latent space, which translates in the ability of the model to store latent attributes as probability distributions. In order to take into account the clinical form for each graph, an additional constraint has been imposed and defined as follows:

$$L_{CrossEntropy} = -\sum_{j=1}^{C} y_{i,j} \log\left(\frac{exp(a_i)}{\sum_j exp(a_j)}\right) \quad (6)$$

where C is the number of the clinical forms.

The final loss function to optimize is thus obtained by summing up all the losses as defined in Eq. (7).

$$L_{final} = L(\theta, D(v), G_{q_\theta(z|x)}(x)) + MSE_{Model} + + MSE_{Coherence} + L_{CrossEntropy} \quad (7)$$

As long as the parameters used for training the AAE are concerned, 500 iterations with a batch size of 64 are used for training the generator and the discriminator in an alternating fashion. The process terminates when the capability of the discriminator to distinguish synthetic samples from true samples remains stable approximately around 50%. The Adam optimizer is used for both models while the learning rate (lr), imposed for the discriminator, was chosen to be 10 times smaller than the generator ($lr = 0.001$). These settings, have been empirically observed to lead to a more stable training of the adversarial network, providing better results.

3.5. AAE Adversarial Training Pipeline

The whole pipeline for training the GAN framework and generating synthetic structural brain networks is summarized in Fig. 2. Generally speaking we can divide the entire workflow in three main phases: *i)* Training the AAE model *ii)* Using the AAE model for generating synthetic MS structural brain networks *iii)* Data augmentation for MS clinical form classification.

3.5.1. Training the AAE model

In order to properly train the proposed AAE model, a naïve data augmentation procedure is needed. The original dataset is split in training and test set by a leave-one-subject-out cross validation strategy (step 1). The training set was then exploited to calculate the conditional probability distribution defined in Eq. (8):

$$P(X = v|Y = y, Q = q) = \binom{n}{v} p_{iyq}^{v}(1 - p_{iyq})^{n-v} \quad (8)$$

Here, p_{iyq} defines the probability of an edge i to be present in the vectorized representation of the upper triangular matrix $x \in \{0,1\}^{(1,d)}$ with dimensionality d and $i \in [0, d]$ given a class label y and a degree quantile q. The letter v defines the number of times the edge i is present and n the number of trials (number of subjects drawn).

It is worth noting that outlier probabilities can be present. In fact, given that our training set is only the realization of a stochastic process, the fact that two regions are always ($p_{iyq} = 1$) or never ($p_{iyq} = 0$) connected might not be true in general. For example, it can be due to lack of data or biases in the collected dataset. To overcome this issue Eq. (9) is applied:

$$P(X|Y, Q) = \begin{cases} 0.95, & if\ p_{iyq} > 0.95 \\ 0.05, & if\ p_{iyq} < 0.05 \end{cases} \forall\ i, y, q \quad (9)$$

It is important to notice that the described anomalies represents a tiny fraction of the total number of connections ($\ll 1\%$) and do not mine the final result of the work. Yet, the operation is useful for a better generalisation capability and avoid overfitting.

The calculated probability density function (pdf) is then used as a "stamp" from which sampling new batches of data at each iteration (step 2 and 3). More in detail, given a class label y and a degree quantile q, a new vectorized representation of an upper triangular matrix can be obtained $x \in \{0,1\}^{(1,d)}$, where x_i is assigned 1 with probability p_{iyq} or 0 with probability $1 - p_{iyq}$. Here, x_i denotes the presence or the absence of a connection in the i-th edge of the y-th MS class in the q-th degree quantile. Finally, a *continuous transformation* of the connectivity matrices is applied as following (Eq. (10)):

$$x_i = \begin{cases} x_i + U(0.01, 0.05), & \text{if } x_i = 0 \\ x_i * U(0.95, 0.99), & \text{if } x_i = 1 \end{cases} \quad (10)$$

where U is the continuous uniform distribution. In other words, a noise component is introduced to the new generated connectivity matrix. This strategy is usually implemented in the context of an adversarial model and it has been shown to provide more stable training [34].
This naïve procedure for extracting new instances from the underlining likelihood distribution provides a double advantage: first, several samples can be generated, thus addressing the problem of limited training size. Second, the obtained instances resemble, at each iteration, the percentile distribution of the original dataset. Henceforth, this second advantage is important to overcome the problem of mode collapse, which constitute a usual challenge when it comes to train an adversarial network [12]. It is worth noting that this data augmentation technique is only used to train the AAE model but cannot be implemented for actual MS connectivity data augmentation. Indeed, this naïve method is not able to produce enough qualitative results, due to the hypothesis of conditional independence imposed between pairs of nodes inside the graph. However, the resampling strategy guarantees that the dataset effectively used to train the adversarial framework is perfectly balanced with respect to a-priori information of the clinical profile and graph density that we are interested in generating.

3.5.2. Using the AAE model for generating synthetic MS structural brain networks

Once the model is trained, we are ready for the second phase in which the synthetic graphs are generated by sampling new instances from a standard gaussian distribution (Figure 2 step 4), with shape $\mathbb{R}^{(1,c)}$, and then used to feed the already trained decoder, obtaining new realistic connectome data.

3.5.3. MS clinical form classification

Finally, in the third phase, the synthetic dataset was used to augment the original training set (step 5) and fed to a classifier (step 6) in order to enhance the classification performance of MS clinical profiles (step 7). The predicted labels obtained from the classifier were compared with the left-out samples for performance evaluation. It is important to notice that neither the training set nor the test set were ever used by the adversarial model to generate synthetic data. The proposed pre-processing approach thus reduces the overfitting tendency of the generative model.

3.6. Experimental Protocol

In this section, the results obtained by evaluating both the structural property of the connectivity matrices and their respective graph-derived metrics are reported. In order to perform the evaluation, a sample of data from a random normal distribution $v \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ were drawn, where $v \in \mathbb{R}^{(1,c)}$ and $c = 100$. N defines the number of samples to be generated and passed to the decoder $p_\phi(x|z)$ to obtain a new sample of synthetic data. Finally, we demonstrate the usefulness of our approach by improving the classification performance of MS clinical forms, even in presence of strong imbalance between classes. Classical approaches, like ROS and the more efficient SMOTE [26] are used for comparisons as well as the more recent adversarial model ARAE [27]. The performance metrics used for the evaluation are $F1_{score}$, Precision and Accuracy which are defined in Eq. (11, 12, 13) respectively.

Table 2
Performances of MS Clinical Forms Classification using different data augmentation methods

Strategy	$F1_{score}$	Precision	Recall
True Data	65.65 ± 12.34	77.49 ± 12.8	61.68 ± 12.63
ROS	65.84 ± 11.97	77.49 ± 12.76	61.76 ± 12.08
SMOTE	72.32 ± 11.18	83.04 ± 11.28	68.45 ± 11.39
ARAE	70.0 ± 11.58	81.1 ± 12.14	64.8 ± 11.83
AAE (ours)	$\mathbf{81.0 \pm 10.37}$	$\mathbf{86.25 \pm 10.36}$	$\mathbf{79.65 \pm 10.51}$

Average classification performance (with standard errors) based on a leave-one-subject-out cross validation strategy on the original dataset (True Data) and after data augmentation of the training set using the Random Over Sampling (ROS) technique, Synthetic Minority Oversampling Technique (SMOTE), Adversarially Regularized Graph Autoencoder (ARAE) and our approach (AAE)

$$F1_{score} = \frac{2TP}{2(TP + FP + FN)} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Precision = \frac{TP}{TP + FN} \tag{13}$$

where the abbreviations TP, TN, FP, FN represent the True Positive, True Negative, False Positive and False Negative of instances respectively.

From now on, we refer to the connectivity graphs generated through the adversarial network as *synthetic data*, while the original dataset will be labelled as *true data*.

4. Results

4.1. Comparison of Data Augmentation Methods for MS Classification

The classification task was performed with a Random Forest Classifier (RF) with 100 trees, due to its robustness to overfitting and unbalanced dataset. Table 2 reports the average classification performances (with standard errors) between our method and the three oversampling techniques previously introduced. Fig. 3 shows the corresponding confusion matrices. Compared to the true unbalanced date used as reference, our method obtained higher performance, reaching an $F1_{score}$ of 81% instead of 65.7%. The ROS and SMOTE methods provided a score of 65.8% and 72.3% respectively, while the ARAE model reports a value of 70% showing a marginal improvement over the unbalanced baseline.

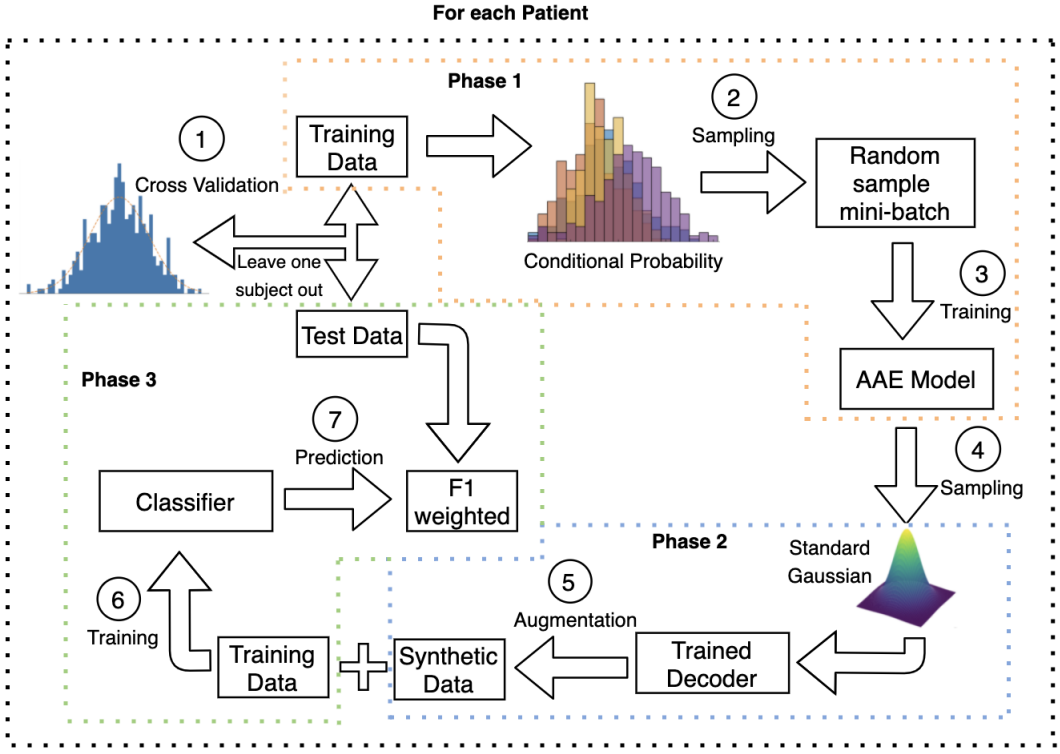

Figure 2: Schematic Representation of the entire workflow. (1) The original dataset is split in training and test set by means of a leave-one-subject-out cross validation strategy. Considering only the training set, conditional probabilities were calculated and mini-batch random samples were drawn (2), at each cycle, for training the AAE model (3). Once the training process was completed, a batch of random instances were sampled from a standard gaussian distribution (4) and fed to the trained decoder to produce synthetic graphs, used to augment the original training dataset (5). The resulting augmented dataset was used to train a classifier (6) to predict MS clinical profiles (7). The entire process was repeated for each patient and the predicted MS class was compared with the actual class from the left-out subjects by means of F1-score.

Figure 3: Confusion matrices for the classification of MS clinical profiles

4.2. Evaluation of Synthetic Data Based on Graphs Matrices

We want to evaluate both the coherence and the difference between true and synthetic data. Ideally, we aim at producing synthetic graphs that span the entire range distribution of the true data sample. In order to evaluate the properties of synthetic graphs, an equal number of data were generated with respect to the true data samples obtaining a perfectly balanced dataset. For both true and synthetic data, the global assortativity degree metric was calculated along with a percentile distribution of 1% width. In other words, the distributions of true and synthetic data were computed with the highest degree of precision following the idea that larger bandwidth will produce less precise comparisons by smoothing the distributions and providing a too optimistic result. In order to measure the distance between the two distributions, the mean absolute point-wise deviation was calculated (77.82 ± 46.07). The overlap proportion (OP) between the two distributions (true vs synthetic) is calculated by employing Eq. (14).

$$OP = \frac{\sum_{j=1}^{n} min(I_j, M_j)}{\sum_{j=1}^{n} M_j} \tag{14}$$

Here, n represents the number of comparison while I and M represent the synthetic and true data distribution respectively. A value of 96.26% is obtained. A graphical analysis of the true and synthetic data was also performed by

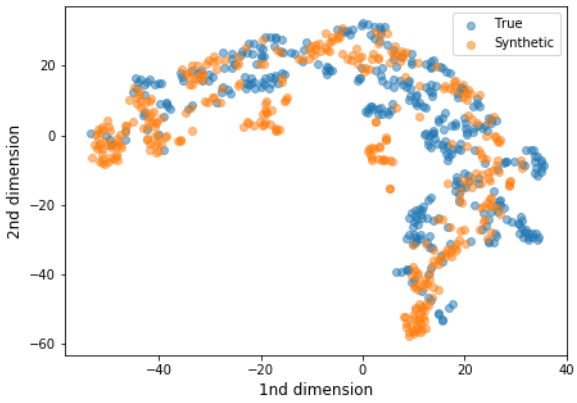

Figure 4: Embedded t-SNE representation of structural graphs: True vs Synthetic data

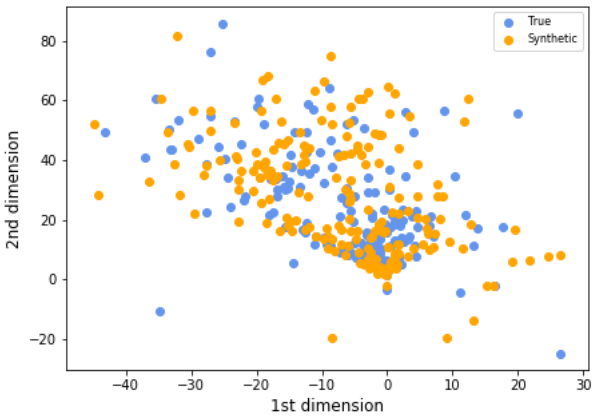

Figure 5: Graph2Vec embedding comparison of structural graphs: True vs Synthetic data

means of the t-distributed Stochastic Neighbour Embedding (t-SNE) model [35, 36]. This algorithm applies a non-linear transformation of the original multidimensional data. It performs an embedding of data, mapping them in a lower dimensional space. Specifically, the algorithm ensures that, each high-dimensional element is mapped to a lower dimensional space in such a way that similar objects are modelled by nearby points and dissimilar objects are modelled by distant points with high probability [37, 38].

Fig. 4 illustrates the embedded representation of the synthetic and true data. From the image, it can be observed that the two groups are fairly similar. It is worth to note that in the t-SNE procedures, the *perplexity* parameter dictates the shape of the mapping function. For this reason, multiple evaluations of this parameter has been performed using value from 10 to 60 at 10 units increment. In Fig. 4 the t-SNE results are illustrated for a perplexity parameter of 30. To be noticed that the author pointed out that as far as the perplexity parameter remains in the usual range $(5, 50)$ the model is quite robust [36]. In addition to the t-SNE representation, the embedding of the true and synthetic graphs was also performed using the Graph2Vec algorithm [39], which is optimised for working with graphs (Fig. 5).

It is a transductive neural embedding framework used to learn from data-driven distribution representations of arbitrary sized graphs. This framework ensures that structurally similar graphs are represented close to one another, while dissimilar graph are depicted far apart. In other words, the model is able to preserve the first and second order proximity. The former is the local pairwise similarity between nodes linked by edges, while the latter indicates the similarity of the nodes neighbourhood structures. In order to numerically compare the true and synthetic data, the $F1_{score}$ metric was computed by linearizing the upper triangular part of the binary adjacency matrices ($A_t, A_s \in \mathbb{R}^{(1,d)}$ respectively with $d = 3486$). Each true adjacency matrix is compared with every synthetic vector. A minimum $F1_{score}$ value of 63% and a maximum value of 82% was obtained with an average score of 76%.

4.3. Evaluation of Synthetic Data Based on Graphs Features

Analyzing the metrics of synthetic graphs is important to capture meaningful features characterizing the structural brain connectome. For this purpose, six graph-based global metrics are considered: Transitivity, Global Efficiency, Modularity, Density, Betweenness Centrality and Assortativity. Such metrics are indeed widely used to characterize brain connectivity [40], and could provide a reliable measure of the quality of generated data.

In Fig. 6, the boxplot distribution between true and synthetic data is presented for each metric. Comparable values of median and interquartile range are observed for the two distributions.

As in the previous section, the t-SNE analysis has been repeated varying the *perplexity* parameter in the range between 10 and 60 at steps of 10. In Fig. 7, the t-SNE embedded representation shows similar distributions between true and synthetic data without obvious discrepancies.

In order to assess the similarity between the two groups, an additional evaluation based on the Kernel Density Estimation (KDE) function was performed. The distributions of true and synthetic datasets were estimated by means of the KDE function and their likelihoods were compared. This approach was originally introduced by Breuleux *et al.* [41] and applied in the context of generative adversarial networks in two reports [19][42]. The method estimates the probability of the synthetic data, by fitting a Gaussian Parzen window to the generated samples and reports the likelihood under this distribution. The bandwidth of the Gaussian window is obtained by cross-validating the training data. Afterword, the similarity between the two datasets has been computed by estimating the pdf of the KDE estimation, so that similar datasets could be represented by similar distributions. Value of 2773.66 and 2657.13 were obtained respectively by comparing the log-likelihoods for the true and

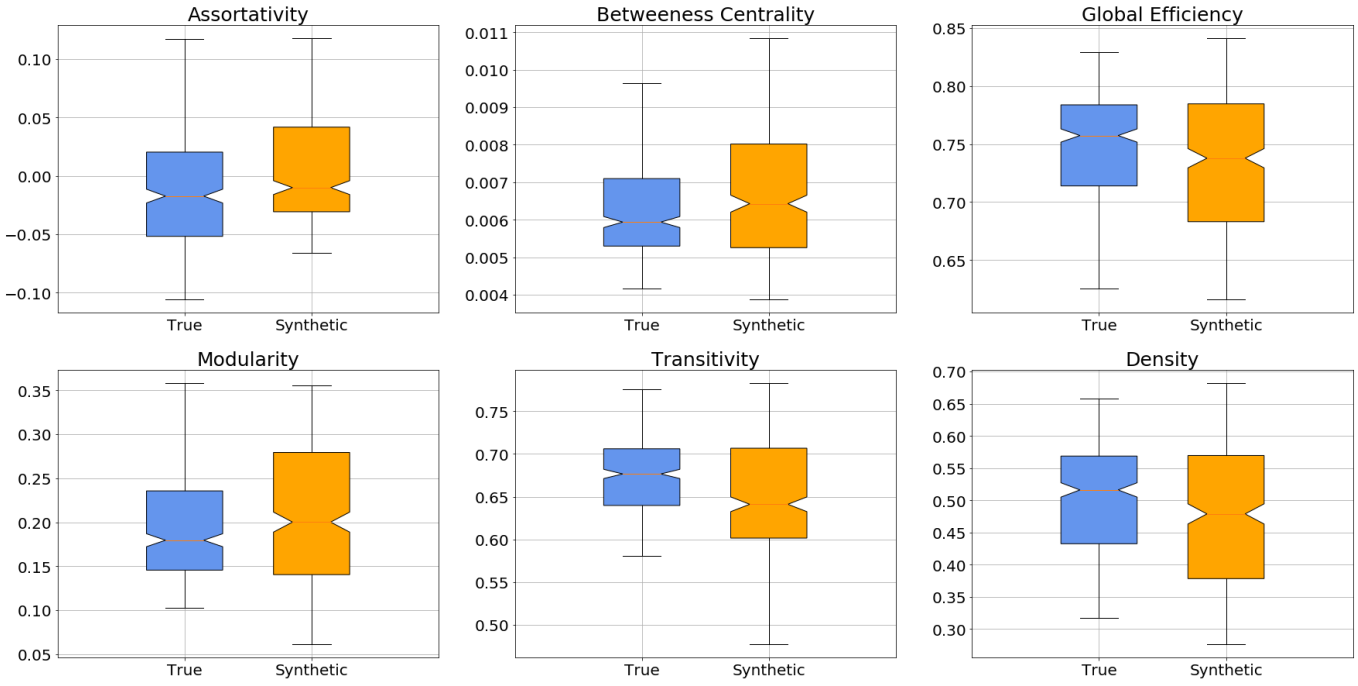

Figure 6: Boxplot distributions of graphs metrics: True vs Synthetic data

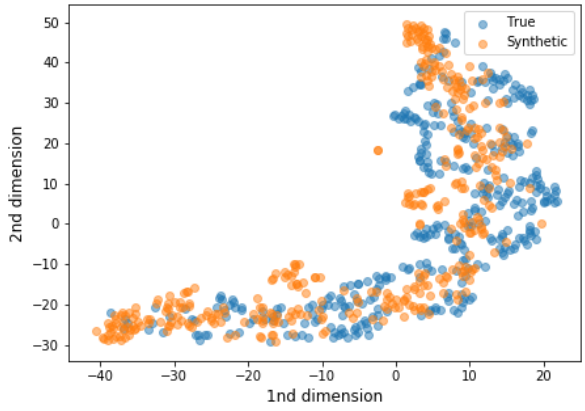

Figure 7: Embedded t-SNE representation of graph metrics: True and Synthetic data

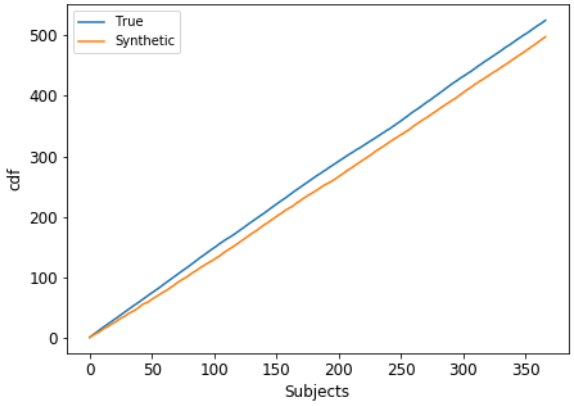

Figure 8: KDE cumulative density function estimation of structural graphs: True and Synthetic data

synthetic data. This result suggests, once more, that the two groups of data are similar. Indeed, the two cumulative functions (true in blue and synthetic in orange) are rather close to one another (Fig. 8). Furthermore, they both follow approximately a straight monotonic-increasing path, which means that the probability mass is evenly distributed across all data samples in both groups.

As an additional test, the bandwidth value from 0.1 to 1 at steps of 0.1 has been increased in order to evaluate the robustness of our results. An increase in performances for every value greater than 0.1 (best cross-validation) has been observed. Finally, in order to offer a sample visualization of the true and synthetic connectivity matrices, Fig. 9 provides two visual examples from the RR and SP clinical forms. It is possible to notice that the true and synthetic data are very similar.

4.4. Evaluation of MSE Coherence

In order to evaluate the stability of the training process, Figure (10) depicts the generator (orange) and discriminator (blue) training loss. Panel (A) and panel (B) report the results obtained when the coherence loss in Eq. (4) was added or excluded from the final objective function, respectively. Less spikes and noise are noticeable in the former case compared to the latter, implying an higher degree of stability of the adversarial training.

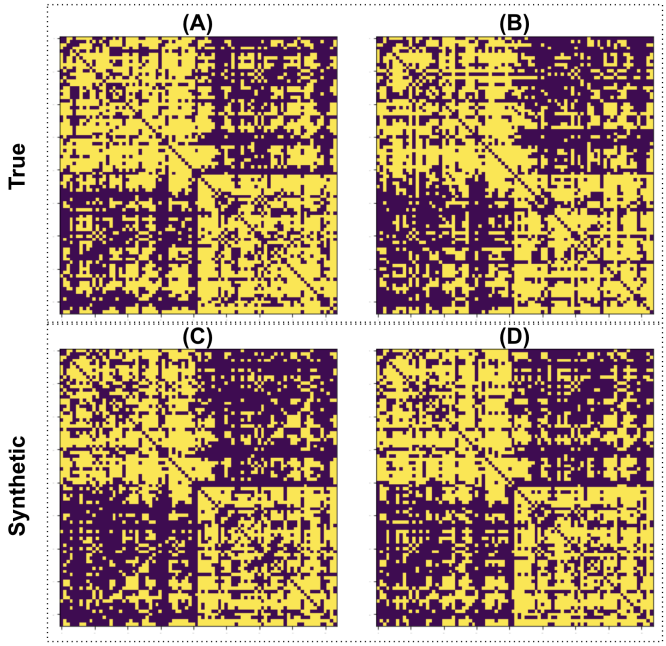

Figure 9: Structural Graph Comparison: True and Synthetic data

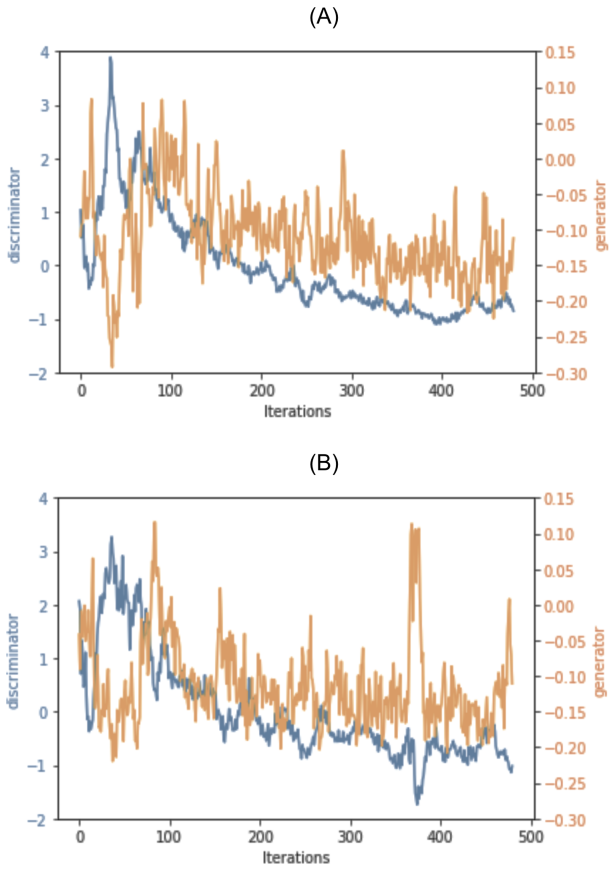

Figure 10: Training loss comparison *with* (A) and *without* (B) coherence loss

5. Discussion

In this work, an approach for generating new structural connectivity matrices of MS patients was presented. In a context of imbalanced data, the proposed framework was able to up-sample the minority class producing a much higher $F1_{score}$ (81%) with respect to the baseline unbalanced classification (66%). Furthermore, comparing to other classical oversampling techniques (ROS and SMOTE) and a graph-based adversarial network (ARAE), our method increases the classification performance by approximately +10%. The improvement can be related to the capability of our method to generate more biologically plausible connectomes that can better represent the different clinical forms. One of the possible explanations can be related to the additional classifier branch used as regularized factors. Indeed, it can help to preserve meaningful structural information which characterizes the different clinical forms.

Our method was evaluated by comparing true and synthetic data by means of visual and analytical techniques. In fact, the generated data should meet two requirements in order to be valid: first, they should preserve similar structural characteristics as the ones which can be observed in true MS brain networks and second, the new generated brain networks should not be simply copies of the original dataset (overfitting). In other words, while new synthetic graphs with enough diversity in terms of structure and properties need to be generated, they still have to be plausible and lie inside the manifold of the true data. In this work, we showed that both distributions (true and synthetic) were observed to be different but very similar to one another. Indeed, the average $F1_{score}$ obtained by comparing the binary adjacency matrix representation, is 76% (Range 63% to 82%). This means that a significant portion of the generated graphs is completely different with respect to the distribution of the true data samples. Notwithstanding, the boxplot comparison shown in Fig. 6 confirmed our visual observation of Fig. 7. The mean distributions of true and synthetic graph metrics are very close. However, one can notice a substantial variability in the synthetic group. From all this evidence the likelihood of overfitting the training set, by simply generating duplicates, is negligible. Instead, experiments highlighted the diversity of the synthetic samples, which lies inside the manifold of the true data, demonstrating that completely new instances have been generated. Moreover, the actual training set was never seen by the adversarial framework which was trained sampling from the conditional distribution proposed in Section 3.5, thus reducing the chance of overfitting even more.

Once completed, the model will allow to increment the size of the available dataset and thus perform a much robust training of machine learning algorithms even with limited amount and unbalanced data, which constitutes a common scenario in the medical field. In addition, by learning the complete underline data distribution, it is possible to perform meaningful bayesian statistical testing of hypothesis as well as generating graphs with desired characteristics. Finally, the combination with other embedding methods, which learn a meaningful representation of single nodes, is

of great interest. This study exploits the advantage of the AAE framework in the context of brain graphs data generation and can be easily expanded for the analysis of other brain diseases or other related domains.
Some limitations should be also mentioned. First, our dataset represents structural connectivity matrices in which connections are binarized, discarding the valuable information conveyed by weighted graphs. Second, due to its simple architecture, our method will not be efficient for very large graphs. Moreover, in order to sample random batches from Eq. (8), in this work the clinical class and the degree percentile were used for conditioning. In fact, the limited amount of data did not allow to add other covariates like age and gender which are worth exploring if one has a large enough sample size. Finally, the computational time needed to perform the leave-one-subject-out cross-validation is not negligible since for each subject the AAE model has to be trained. However, in the context of brain network analysis, we rarely deal with much larger networks as the actual MRI data are limited to a maximum of a few hundred nodes, justifying the simplification proposed in this study. Additionally, our method performs well even with a limited in number and strongly imbalanced data, in agreement with a previous report [29].

6. Conclusion

In this study, a new data augmentation approach for connectome dataset was presented. Given the capability of graphs to represent complex brain networks, our approach provides a new tool for biomedical application, a domain in which the data availability is scarse and poorly behaving. Therefore, our connectome-based data augmentation approach represents a promising alternative to usual image-based techniques. Furthermore, the proposed data augmentation approach was capable to improve the MS classification performance even in cases of unbalanced data scenario. As future work, we aim to improve our approach by generating all clinical MS profiles and exploit weighted connectivity matrices in place of binary structural graphs. In addition, we plan to explore conditional adversarial neural network methods to perform meaningful bayesian statistical testing of hypothesis as well as generating graphs with desired characteristics. Finally the combination with other embedding methods, which learn a meaningful representation of single nodes, should be also explored.

Acknowledgement

This study is funded by the following projects: European Research Council within the grant 813120-2018 of Marie Skłodowska-Curie Innovative Training Networks (ITN) of the Horizon 2020 through the INSPiRE-MED project. French National Research Agency (ANR) within the national program "Investissements d'Avenir" through the OFSEP project (ANR-10-COHO-002).

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

References

[1] Young T., Hazarika D., Poria S., Cambria E. Recent trends in deep learning based natural language processing. *arXiv*, 2017.
[2] Milletari F., Navab N., Ahmadi S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *Fourth International Conference on 3D Vision (3DV)*, pages 565–571, 2016.
[3] Yun K., Huyen A., Lu T. Deep neural networks for pattern recognition. *arXiv*, 2018.
[4] Schwenker F., Abbas H.M., El Gayar N., Trentin E. *Artificial Neural Networks in Pattern Recognition*. 2016.
[5] Adibuzzaman M. , DeLaurentis P., Hill J., Benneyworth B.D. Big data in healthcare. *AMIA Annu Symp Proc.*, pages 384–392, 2018.
[6] Floca R. Challenges of open data in medical research. *Springer International Publishing*, pages 297–307, 2014.
[7] Perez L., Wang J. The effectiveness of data augmentation in image classification using deep learning. *arXiv*, 2017.
[8] Goodfellow I., Bengio Y., Courville A. *Deep learning Book*. MIT Press, 2016.
[9] Shorten C. and Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.
[10] Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 2019.
[11] Guo Y., Nejati H., Cheung N.M. Deep neural networks on graph signals for brain imaging analysis. *IEEE International Conference on Image Processing (ICIP)*, pages 3295–3299, 2017.
[12] Baldassarre A., Ramsey, L. E.; Siegel, J. S., Shulman, G. L., Corbetta, M. Brain connectivity and neurological disorders after stroke. *Current Opinion in Neurology*, 29(6):706–713, 2016.
[13] Rezazadeh M., I.; Frohlich, J., Loo S. K.; Jeste S. Brain connectivity in autism spectrum disorder. *Current Opinion in Neurology*, 29(2):137–147, 2016.
[14] Rafael G., Jaime R.S., Dante B., Jaime A. How artificial intelligence is supporting neuroscience research: A discussion about foundations, methods and applications. *arXiv*, pages 63–77, 2017.
[15] Ghasemi N., Razavi S. and Nikzad E. Multiple sclerosis: Pathogenesis, symptoms, diagnoses and cell-based therapy. *Cell Journal*, 19(1):1–10, 2016.
[16] Verma V., Qu M., Lamb A., Bengio Y., Kannala J., and Tang J. Graphmix: Regularized training of graph neural networks for semi-supervised learning. *ICRL 2020 Conference*, 2020.
[17] Radford A. , Metz L., Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *Preceeding at ICLR*, 1, 2016.
[18] Makhzani A., Shlens J., Jaitly N., Goodfellow I., Frey B. Adversarial autoencoders. *International Conference on Learning Representations ICLR*, 2016.
[19] Calimeri F. , Marzullo A. , Stamile C., Terracina G. . Biomedical data augmentation using generative adversarial neural networks. *Artificial Neural Networks and Machine Learning ICANN, Springer*, 10614:626–634, 2017.
[20] Frid-Adar M., Diamant I. , Klang E. , Amitai M. , Goldberge J. , Greenspan H. . Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321, 2018.
[21] Salehinejad H., Valaee S., Dowdell T., Colak E., Barfett J. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 990–994, 2018.
[22] Shui-Hua W., Yu-Dong Z. Densenet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *Association for Computing Machinery*, 16:1551–6857, 2020.

[23] Yu-Dong Z.. Vishnuvarthanan G., Chaosheng T. High performance multiple sclerosis classification by data augmentation and alexnet transfer learning model. *Journal of Medical Imaging and Health Informatics*, 9:1–10, 2019.

[24] Zhang X., Yao L., Yuan F. Adversarial variational embedding for robust semi-supervised learning. *Research Track in KDD*, 2019.

[25] Imran A., Terzopoulos D. Multi-adversarial variational autoencoder networks. *CoRR*, 2019.

[26] Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 6:321–357, 2002.

[27] Pan S., Hu R., Long G., Jiang J., Yao L. and Zhang C. Adversarially regularized graph autoencoder for graph embedding. *IJCAI*, 2018.

[28] Feng F., He X., Tang J., Chua T.S. Graph adversarial training: Dynamically regularizing based on graph structure. *arXiv*, 2019.

[29] Khoshgoftaar T.M., Golawala M., Hulse J.V. An empirical study of learning from imbalanced data using random forest. *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2:310–317, 2007.

[30] Kocevar G., Stamile C., Hannoun S., Cotton F., Vukusic S., Durand-Dubief F., Sappey-Marinier D. Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Frontiers in Neuroscience*, 13, 2019.

[31] Smith S.M., Jenkinson M., Woolrich M.W., Beckmann C.F. et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 62, 2019.

[32] Zhu, Jun-Yan and Park, Taesung and Isola, Phillip and Efros, Alexei A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Computer Vision (ICCV)*, 2017.

[33] Doersch C. Tutorial on variational autoencoders. *arXiv*, 2016.

[34] Sønderby C. K., Caballero J., Theis L., Shi W., Huszár F. Amortised MAP inference for image super-resolution. *arXiv*, 2016.

[35] Van der Maaten L.J.P. Learning a parametric embedding by preserving local structure. *Twelfth International Conference on Artificial Intelligence & Statistics (AI-STATS), JMLR W&CP*, 5:384–391, 2009.

[36] Van der Maaten L.J.P., Hinton G.E. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[37] Van der Maaten L.J.P. Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research*, 15:3221–3245, 2014.

[38] Van der Maaten L.J.P., Hinton G.E. Visualizing non-metric similarities in multiple maps. *Machine Learning*, 87:33–55, 2012.

[39] Narayanan A., Chandramohan M., Venkatesan R., Chen L., Liu Y., Jaiswal S. graph2vec: Learning distributed representations of graphs. *arXiv*, 2017.

[40] Beauchene C, Roy S, Moran R, Leonessa A, Abaid N. Comparing brain connectivity metrics: a didactic tutorial with a toy model and experimental data. *Neuroimage*, 15(5), 2018.

[41] Breuleux O., Bengio Y., Vincent P. Quickly generating representative samples from an RBM-derived process. *Neural Computation*, 23(8):2053–2073, 2011.

[42] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B. et al. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 23(8):2053–2073, 2011.