

Received July 26, 2020, accepted August 16, 2020, date of publication September 3, 2020, date of current version September 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3021580

A Data Augmentation Scheme for Geometric Deep Learning in Personalized Brain-Computer Interfaces

FOTIS P. KALAGANIS^{1,2}, (Graduate Student Member, IEEE), NIKOLAOS A. LASKARIS^{1,3},
ELISAVET CHATZILARI^{1,2}, SPIROS NIKOLOPOULOS^{1,2},
AND IOANNIS KOMPATSIARIS^{1,2}, (Senior Member, IEEE)

¹AIIA Laboratory, Department of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

²MKLab, Centre for Research and Technology Hellas, Information Technologies Institute, 57001 Thessaloniki, Greece

³Neuroinformatics Group, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Corresponding author: Fotis P. Kalaganis (kalaganis@csd.auth.gr)

This work was supported in part by the Project MAMEM funding from the European Union's Horizon 2020 Research and Innovation Program under Grant 644780, and in part by the Project Evison that has been Co-Financed by the European Regional Development Fund of the European Union and Greek National Funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call RESEARCH-CREATE-INNOVATE under Project T1EDK-02454.

ABSTRACT Electroencephalography signals inherently deviate from the notion of regular spatial sampling, as they reflect the coordinated action from multiple distributed overlapping cortical networks. Hence, the observed brain dynamics are influenced both by the topology of the sensor array and the underlying functional connectivity. Neural engineers are currently exploiting the advances in the domain of graph signal processing in an attempt to create robust and reliable brain decoding systems. In this direction, Geometric Deep Learning is a highly promising concept for combining the benefits of graph signal processing and deep learning towards revolutionising Brain-Computer Interfaces (BCIs). However, its exploitation has been hindered by its data-demanding character. As a remedy, we propose here a novel data augmentation approach that combines the multiplex network modelling of multichannel signal with a graph variant of the classical Empirical Mode Decomposition (EMD), and which proves to be a strong asset when combined with Graph Convolutional Neural Networks (GCNNs). As our graph-EMD algorithm makes no assumptions with respect to linearity and stationarity, it appears as an appealing solution towards analysing brain signals without artificially imposing regularities in either temporal or spatial domain. Our experimental results indicate that the proposed scheme for data augmentation leads to substantial improvement when it is combined with GCNNs. Using recordings from two distinct BCI applications and comparing against a state-of-the-art augmentation method, we illustrate the benefits from its use. By making it available to BCI community, we hope to further foster the application of geometric deep learning in the field.

INDEX TERMS Brain-computer interfaces, BCI, data augmentation, EEG, geometric deep learning, graph CNN, GCNN, graph signal processing, neural engineering.

I. INTRODUCTION

Research on Brain-Computer Interfaces (BCIs) has experienced an impressive growth in the recent past. The main objective in BCIs is to provide a direct communication pathway between the human brain and an external device. A typical BCI system consists of a signal processing module which can be further decomposed into three submodules

The associate editor coordinating the review of this manuscript and approving it for publication was Easter Selvan Suvishamuthu.

(i.e. pre-processing, feature extraction and feature selection) and a classification module which converts the resulting features into machine commands.

The most common neuroimaging modality that is employed in BCIs is the electroencephalography, a typically non-invasive neuroimaging technology that measures the brain's electrical activity using electrodes placed on the human scalp. The produced recording, called electroencephalogram (EEG), is not easy to interpret as it has a low signal to noise ratio and its statistical properties change

substantially with the course of time [1]. Moreover, EEG is known to vary significantly across individuals and even to depend on subject's state during the recording.

Researchers within the computational neuroscience and machine learning communities have put a lot of effort into developing signal processing techniques and computational intelligence algorithms that perform robust brain decoding from EEG signals despite the aforementioned challenges. The current state-of-the-art techniques include among others Riemannian geometry-based classifiers [2], filter banks [3], adaptive classifiers [4] and graph signal processing [5]. On top of them, and in complete harmony with the concurrent trends in empirical data analysis, deep learning has come into the picture. After radically changing the field of machine learning in many aspects of our digitalized modern world (like computer vision and speech recognition), by providing flexible and general-purpose models, deep learning nowadays facilitates brain decoding practices undertaking jointly the signal processing and classification modules [6]. The corresponding models however, require large amounts of data to directly learn patterns and capture the "true" information structure in the data in an efficient way that can be then transferred and/or adapted to similar tasks.

Since electroencephalographic activity is a "fuzzy" signal coming from a complex system and governed by the underlying structure of (and the functional connectivity within) the cortical networks, neuroscientists and BCI researchers have started to exploit the recent advances in the domain of graph signal processing [7] so as to incorporate the functional principles of the networked brain within signal analysis and build reliable brain decoding systems [8]–[10]. In this context, geometric deep learning, which collectively refers to adapting and deploying deep learning on data manifolds, graph patterns and signals registered over irregular grids, could significantly enhance the performance in existing BCI protocols and implementation pipelines. Indeed, in the last few years, graph deep learning architectures have been very successful in processing complex data such as social networks, meshes and sensor-array signals, leading to state-of-the-art performance on multiple public datasets [11]. Although geometric deep learning appears to be most suitable for classifying EEG signals, it is the lack of large datasets that constantly limits its use in BCI applications.

As a matter of fact, neuroimaging data collection is still expensive, time-consuming and the availability of large scales of such data is even more restricted due to personal data regulations. Consequently, the corpus of data in neuroimaging is rather small in terms of size compared to other domains such as computer vision or speech recognition. Many public EEG datasets typically contain only a small number of participants up to a few dozens. Although some specific topics such as sleep and epilepsy studies do have larger datasets publicly available [12], in the particular field of BCI-related applications the data are even more limited. In addition, poor signal to noise ratio limits the amount of available information contained in the recordings which are often inextricably

connected to the data collection protocol and hence do not facilitate dataset curation by aggregation across different laboratories. Finally, the models that have been developed for images and speech even though they appear as technically generic, they are not suitable for EEG recordings. This also holds for many well-established strategies for training deep learning models, which cannot be adopted per se in the BCI domain, such as image augmentation methods [13].

An undeniable fact in the field of deep learning is that more data can offer a substantial improvement in the classification accuracy of a model. For the typically small EEG datasets, it is difficult to use deep learning methods with satisfactory results. Therefore, creating artificial EEG signals for deep learning classification schemes emerges as a necessity. Common signal processing tools like the Discrete Fourier and the Wavelet transforms that are typically used in the domain of signal processing are not adequate for augmenting existing EEG datasets (e.g. by generating surrogate data), due to the non-linear and non-stationary character of EEG signals. Previous works, in the context of generating artificial EEG signals, mainly employ stationary spectrum approaches, such as adding Gaussian noise to the spectrum of the signal [14], that oversee the inherent temporal characteristics of the EEG signals. On the other hand, studies that augment the EEG datasets by operating on the temporal domain, such as concatenating different temporal EEG segments [15], maintain most of the temporal aspects of the EEG signals but fail to preserve the properties of the EEG spectrum. Recently, a more suitable strategy for EEG data augmentation, based on the Empirical Mode Decomposition (EMD), has been proposed [16] and tested successfully in conjunction with Convolutional and Wavelet Neural Networks [17]. However, all of the aforementioned methods share as shortcoming their inadequacy to fully preserve the underlying complex dynamics of the original EEG data, and this in turn constitutes their employment in geometric deep learning an inefficient strategy.

In this work we propose a data augmentation methodology with no assumptions regarding stationarity and linearity, capable of capturing and preserving the inherent structural and functional characteristics of the superficially observed cortical activity. The novelty of our work is on the exploitation of the spatiotemporal character of EEG signals which is taken into consideration by constructing a sparse binary graph that incorporates both the topological arrangement of the sensor array and the temporal continuity between consecutive signal samples (by means of multiplex graph modelling). Subsequently, we use the aforementioned sparse binary graphs in conjunction with the Graph EMD (GEMD) [18] method for data-augmentation in order to improve the classification accuracy in Graph Convolutional Neural Networks (GCNNs). Our approach is validated based on two distinct BCI-related datasets, where GCNNs are trained, at a personalised level, with only few dozens of trials initially available. The first dataset concerns the classification of the reaction time of a driver, in a simulation

TABLE 1. Notation table. Lowercase italics refer to scalars, functions or signals whereas lowercase bold letters refer to vectors and uppercase bold letters to matrices. Matrix elements are denoted as uppercase characters accompanied by the corresponding indices as subscripts. Calligraphic uppercase letters denote sets. The same symbol appearing under different fonts usually refers to the same entity but in a different form, e.g. a graph signal and its corresponding vector representation.

Symbol	Meaning/Definition
$x(t)$	Arbitrary signal in time domain.
x	Arbitrary graph signal indexed by vertices.
\mathbf{x}	The vector representation of the graph signal x .
\mathcal{G}	An arbitrary connected, undirected and weighted graph.
\mathcal{V}	The set containing the vertices of \mathcal{G} .
\mathcal{E}	The set containing the edges of \mathcal{G} .
\mathbf{W}	The weighted adjacency matrix of \mathcal{E} .
\mathbb{R}	The set of real numbers.
\mathbf{I}_N	The $N \times N$ identity matrix.
\mathbf{L}	The graph Laplacian, either combinatorial or normalized, matrix that corresponds to \mathcal{G} .
$\mathbf{\Lambda}$	The diagonal matrix that holds the eigenvalues of \mathbf{L} .
$\tilde{\mathbf{\Lambda}}$	The diagonal matrix that holds the scaled, in range $[-1, 1]$, eigenvalues of \mathbf{L} .
$\tilde{\mathbf{L}}$	The graph Laplacian matrix resulting by reconstructing \mathbf{L} with the scaled eigenvalues $\tilde{\mathbf{\Lambda}}$.

environment, into fast and slow driving responses. The second dataset includes EEG recordings of event related responses and concerns the differentiation between attentive and passive condition during a driving pc-game. The selection of these two datasets was dictated by the need to examine and validate the introduced data-augmentation scheme using brain activity signals reflecting different cognitive processes and recorded via distinct BCI paradigms (with the first/second dataset concerning endogenous reactions/evoked responses and corresponding to asynchronous/synchronous BCI). Furthermore, their inclusion in this article opens the possibility for this work to pave the way for the adoption of geometric deep learning in the realm of brain-to-vehicle technology [19].

II. METHODOLOGY

In this section we initially describe briefly the basic notions in the field of graph signal analysis and then we present the employed GCNN architecture [20] that is used for classification. We note that the term GCNN refers to a convolutional neural network that operates on graphs. Next, we introduce the proposed data augmentation approach starting with the presentation of the GEMD [18] method which leads to a novel strategy for generating artificial EEG epochs with respect to the underlying graph structure that governs the EEG signals. The term EEG epoch is used to express a segment extracted from the continuous EEG traces. Finally we present the construction of a spatiotemporal graph that expresses the spatial and temporal relationships jointly. The code which implements the proposed methodology is available at https://github.com/fkalaganis/graph_emd.

A. GRAPH SIGNAL PROCESSING PREREQUISITES

Let us denote by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ a connected, undirected and weighted graph where \mathcal{V} denotes a finite set of $|\mathcal{V}| = N$

vertices, $\mathcal{E} = \{\mathcal{V} \times \mathcal{V}\}$ denotes the set of edges and $\mathbf{W} \in \mathbb{R}^{N \times N}$ the corresponding weight adjacency matrix. The entry $W_{i,j}$ of the adjacency matrix indicates the weight of the (i, j) edge whereas the absence of an edge is represented by a zero value. Then, $x : \mathcal{V} \rightarrow \mathbb{R}$ represents a signal indexed on the vertices of \mathcal{G} and is usually termed as graph signal. An alternative representation of the graph signal x is achieved by the vector $\mathbf{x} \in \mathbb{R}^N$ where \mathbf{x}_i is the value of x at vertex i .

With the above formulations, we can now define the graph Fourier Transform (GFT) that serves as the basis for the graph signal convolution and filtering. The spectral analysis on graphs is achieved by exploiting the graph Laplacian operator [21]. Let $\mathbf{D} \in \mathbb{R}^{N \times N}$ be the degree matrix which is diagonal and its elements are calculated as $D_{ii} = \sum_j W_{ij}$. Then, the combinatorial and the normalized Laplacian matrices are defined as $\mathbf{L}_C = \mathbf{D} - \mathbf{W}$ and $\mathbf{L}_N = \mathbf{I}_N - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$ respectively with $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ denoting the identity matrix. As \mathbf{L} , which denotes either the combinatorial or the normalized graph Laplacian matrix, is a symmetric and positive semidefinite matrix it admits an eigendecomposition, $\mathbf{L} = \mathbf{U}^T \mathbf{\Lambda} \mathbf{U}$, with a complete set of orthonormal eigenvectors \mathbf{u}_l appearing in columns of \mathbf{U} , known as the graph Fourier modes, and the corresponding nonnegative ordered eigenvalues, λ_l appearing as the elements of the diagonal matrix $\mathbf{\Lambda}$, known as the graph frequencies, with $l = 1, \dots, N$. Then, the GFT of a signal is defined as $\hat{\mathbf{x}} = \mathbf{x} \mathbf{U}^T$ and the inverse GFT as $\mathbf{x} = \mathbf{U} \hat{\mathbf{x}}$ [7]. We note that the GFT is valid for both the combinatorial and the normalized Laplacian matrices whose choice concerns the exploitation of different bases.

B. GRAPH CONVOLUTIONAL NEURAL NETWORKS

Since a meaningful shifting operation cannot be directly defined on the vertex domain, the convolution operator on a graph \mathcal{G} , denoted by $*_{\mathcal{G}}$, is defined on the spectral domain of the graph as $(x * y)_{\mathcal{G}} = \mathbf{U}((\mathbf{x} \mathbf{U}^T) \circ (\mathbf{y} \mathbf{U}^T))$ with (\circ) denoting the Hadamard product operator. Hence, a filter h_{θ} can be applied on x as

$$y = h_{\theta}(\mathbf{L})x = h_{\theta}(\mathbf{U} \mathbf{\Lambda} \mathbf{U}^T)x = \mathbf{U} h_{\theta}(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{x} \quad (1)$$

A non-parametric filter, where all parameters are free and unconstrained, would have the form of

$$h_{\theta}(\mathbf{\Lambda}) = \text{diag}(\boldsymbol{\theta}) \quad (2)$$

where $\boldsymbol{\theta} \in \mathbb{R}^N$ is a vector containing the graph filter coefficients. However, the non-parametric filter approach leads to filters that are not spatially (in terms of graph neighborhood) localized. On top of that, the learning complexity of a non-parametric filter is $O(N)$ with N expressing the dimensionality of data which may be forbidding for the scale of filters required by the GCNN. In order to tackle these problems, one can employ a polynomial filter

$$h_{\theta}(\mathbf{\Lambda}) = \sum_{k=0}^{K-1} \theta_k \mathbf{\Lambda}^k \quad (3)$$

where the parameter $\theta \in \mathbb{R}^K$ now is a vector of polynomial coefficients. Since (3) is a spectral filter defined by the K -th order polynomial of the Laplacian matrix it is exactly K -localized [22]. Moreover, the filtering complexity is now in $O(K)$, the size of the filter as in classical CNNs.

The cost to apply the polynomial filter on a signal is $O(N^2)$ as it involves the multiplication with the GFT basis. So as to further reduce the complexity of GCNNs the Chebyshev expansion is typically employed to approximate the polynomial filter of (3). As the Chebyshev polynomials form an orthogonal basis for the Hilbert space of square integrable functions over the interval $[-1, 1]$, (3) can be parameterized as

$$h_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\Lambda}^k) \quad (4)$$

where $T_k(\tilde{\Lambda})$ is the k -th order Chebyshev polynomial evaluated at $\tilde{\Lambda} = 2\Lambda/\max\{\lambda_l\} - I_N$. $\tilde{\Lambda}$ denotes the diagonal matrix that holds the scaled eigenvalues within the range $[-1, 1]$, with $l = 1, \dots, N$. We note that in (4), $\theta \in \mathbb{R}^K$ is a vector holding the Chebyshev coefficients. As the Chebyshev polynomials can be computed by the stable recurrence sequence $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ with $T_0 = 1$ and $T_1 = x$, the filtering operation can be computed via sparse matrix multiplications with cost $O(K|\mathcal{E}|)$ as

$$h_\theta(\Lambda)x = [\bar{x}_0, \bar{x}_1, \dots, \bar{x}_{K-1}]\theta \quad (5)$$

where $\bar{x}_k = 2\tilde{\Lambda}\bar{x}_{k-1} - \bar{x}_{k-2}$ with $\bar{x}_0 = x$ and $\bar{x}_1 = \tilde{\Lambda}x$. This rationale becomes more evident by realizing that $\bar{x}_k = T_k(\tilde{\Lambda})x$. Having the above formulations established, the backpropagation algorithm can be used efficiently in order to learn the graph filters [20], i.e. the Chebyshev coefficients.

C. GRAPH EMPIRICAL MODE DECOMPOSITION

We commence by reviewing the classical EMD algorithm. Given a signal, EMD decomposes it into a finite set of Intrinsic Mode Functions (IMFs) which are time-varying roughly mono-component (i.e single frequency) functions [23]. An IMF holds the following two properties: i) the number of its extrema must be equal or differ mostly by one compared to its number of zero crossings and ii) its upper and lower envelopes, defined by the local maxima and minima respectively, are symmetric with respect to zero.

Let $x(t)$ be a signal in the time domain. The EMD algorithm starts by separating a local low frequency component, $m_1(t)$, referred to as “the trend” from an IMF, denoted as $d_1(t)$, which corresponds to a local high frequency. By applying this step recursively to the remaining trend, the $x(t)$ can be rewritten as:

$$x(t) = m_K(t) + \sum_{k=1}^K d_k(t) \quad (6)$$

The iterative process terminates when every IMF of $x(t)$ has been extracted. The separation of the slow oscillation

trend from the fast oscillating IMF is performed within the EMD algorithm with the so-called *sifting* process. The most conservative sift-stopping criterion for the EMD algorithm is that the extracted fast oscillation is indeed an IMF (i.e. holds all the aforementioned IMF properties). As this is a very strong constraint, more relaxed sifting criteria are typically employed that yield approximate IMFs [24].

In order to extend the EMD algorithm to graph signals we will first provide the definition of extrema and an interpolation method for graph signals as we will need to calculate the graph signal envelopes from local minima and maxima. For a graph signal x defined on $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ the signal at node i is a local maximum (or minimum) if its value is higher (or lower respectively) than every value of its neighbouring vertices.

Having the local extrema identified, the graph signal should be interpolated (i.e generate the signal values on the unknown, non-extrema, vertices) in order to obtain the upper and lower graph envelopes. In order to maintain the assumption-free characteristics of the classical EMD method, the interpolation is treated as a discrete partial differential equation on the graph [25]. As the envelopes are slow varying components, the interpolated signal s needs to minimize the total graph variation, $s^\top \mathbf{L}s$, with \mathbf{L} being the graph Laplacian matrix under the constraint that the graph signal values of the known vertices remain unchanged. Let us denote by \mathcal{K} the set of vertices where the graph signal is known and by \mathcal{U} the set of unknown vertices. Then, in order to calculate the new, interpolated, graph signal s we need to solve

$$\text{minimize } s^\top \mathbf{L}s \text{ subject to } s(\mathcal{K}) = x(\mathcal{K}) \quad (7)$$

By a simple rearrangement of vertices s can be rewritten, in its equivalent vector expression, as $s^\top = [s_{\mathcal{K}}^\top \ s_{\mathcal{U}}^\top]$ with $s_{\mathcal{K}}$ and $s_{\mathcal{U}}$ being the vector representations of $s(\mathcal{K})$ and $s(\mathcal{U})$ respectively alongside with the rearranged Laplacian matrix

$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{\mathcal{K}} & \mathbf{R} \\ \mathbf{R}^\top & \mathbf{L}_{\mathcal{U}} \end{bmatrix}$. Ultimately, the graph interpolation is a Dirichlet problem on the graph and whose solution relies on the following system of linear equations

$$\mathbf{L}_{\mathcal{U}}s_{\mathcal{U}} = -\mathbf{R}s_{\mathcal{K}} \quad (8)$$

Having the graph extrema and the graph interpolation process defined, the classical EMD algorithm can be easily extended to its graph signal counterpart. More concretely, the GEMD method [18] is defined via the following algorithm:

- 1) Initialize $m = x$.
- 2) **If** m meets the stopping criteria:
 - a) Return stored IMFs.
 - b) Stop the decomposition and terminate.
- 3) Set $m_{prev} = m$
- 4) Calculate the upper and lower envelopes, e_+ and e_- respectively, of m_{prev}
- 5) Subtract the average envelope value $\mu = \frac{e_+ + e_-}{2}$ from m_{prev} as $m_{curr} = m_{prev} - \mu$

- 6) If m_{curr} does not meet the soft sifting criteria:
 - a) Set $m_{prev} = m_{curr}$
 - b) Repeat from step 4.
- Else:**
 - a) Store m_{curr} as an IMF
 - b) Set $m = m - m_{curr}$
 - c) Repeat from step 2.

In order to avoid conservative sifting criteria that yield exact graph-IMFs, inspired by [26], we employ the relaxed graph-counterparts that yield approximate graph-IMFs. More specifically the relaxed sifting criteria concern the Sift Relative Tolerance, which is a Cauchy-type criterion, and the Energy ratio criterion which is based on the ratio of the energy of the signal at the beginning of sifting and the average envelope energy. Fig. 1A illustrates the decomposition of a sample graph signal into its graph IMFs.

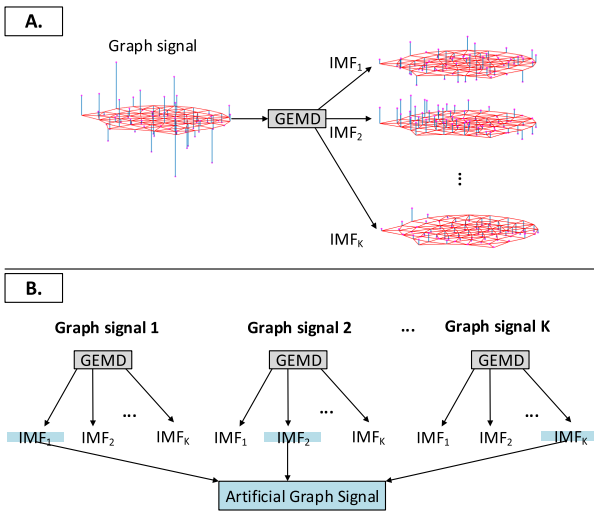


FIGURE 1. A) Decomposition of a graph signal into graph IMFs using the GEMD algorithm. The red lines indicate the graph structure and the vertical blue lines the graph signal value in the corresponding node. B) Schematically depicting the proposed artificial graph signal generation strategy.

D. EEG DATA AUGMENTATION

Given an arbitrary number of EEG epochs, the GEMD method can be utilized in order to generate artificial EEG epochs. Each EEG epoch is initially decomposed into a finite set of graph IMFs with respect to a graph structure \mathcal{G} . Then, an artificial EEG epoch can be generated by combining the graph IMFs from different epochs. Since the graph IMFs are mono-component graph functions, the artificial EEG epochs are expected to exhibit similar characteristics with the originating signals that contribute with their IMFs. In order to create EEG epochs under the scope of improving a classifier, the corresponding class information of each EEG epoch and consequently its corresponding IMFs is taken into account. Therefore, each artificial EEG epoch is generated by graph IMFs stemming from a single class, hence, it is assigned with the corresponding label.

More specifically, the proposed data augmentation is as follows:

- 1) Randomly select the class-specific EEG epochs that will contribute with their IMFs. As the number of IMFs extracted from each signal is finite, the maximum number of IMFs that a signal segment holds indicates the number of randomly selected contributing EEG epochs.
- 2) In order to generate an artificial EEG epoch, select the first IMF from the first contributing EEG epoch, the second IMF from the second contributing EEG epoch and so on. If a contributing EEG epoch holds less IMFs than required we consider its additional graph IMFs to be the zero graph signals.

This procedure can be used to create a large number of artificial EEG epochs (up to the number of EEG epochs to the power of graph IMFs) and therefore augment the dataset. Fig. 1B depicts the proposed artificial graph signal generation process.

Finally, we present a note on the computational complexity of the proposed data augmentation approach which is based on the graph EMD. Since the most computationally demanding operation in the graph EMD method is the matrix inversion operation performed up to a constant number times, the computational complexity of the proposed method is $O(n|\mathcal{V}|^3)$ with $|\mathcal{V}|$ expressing the number of vertices in \mathcal{G} and n the number of EEG epochs. This constitutes the introduced methodology more computationally demanding than the ones presented in [16], [17] which have a computational complexity of $O(nET^2)$, with E and T denoting respectively the number of sensors and samples of the EEG signal. We note that according to the graph construction we employ in our study (refer to section II-E), the relationship that connects $|\mathcal{G}|$, T and E is $|\mathcal{V}| = ET$.

E. SPATIOTEMPORAL GRAPH CONSTRUCTION

Although EEG signals are governed by a presumably static underlying spatial structure, at least with respect to the topological structure of the recording sensor array, they also are time-varying signals that significantly change in the course of time. Hence, when constructing a graph to express the overall characteristics of an EEG signal not only the spatial but also the temporal dependencies should be considered. As the edges express the relationship between vertices, the most straightforward way to embed temporal and spatial relationships jointly in a graph is by employing binary edges only [27].

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ be a graph that expresses the topology of the EEG recording sensor array with $\mathbf{W} \in \{0, 1\}^{E \times E}$ and E being the number of sensors. The spatiotemporal graph that expresses jointly the relationship of an EEG signal with T time samples is defined through its adjacency matrix as

$$\mathbf{W}_{ST} = \mathbf{W} \otimes \mathbf{I}_E + \mathbf{S} + \mathbf{S}^T \tag{9}$$

where (\otimes) denotes the Kronecker product operator, \mathbf{I}_E the $E \times E$ identity matrix and \mathbf{S} an $ET \times ET$ matrix whose

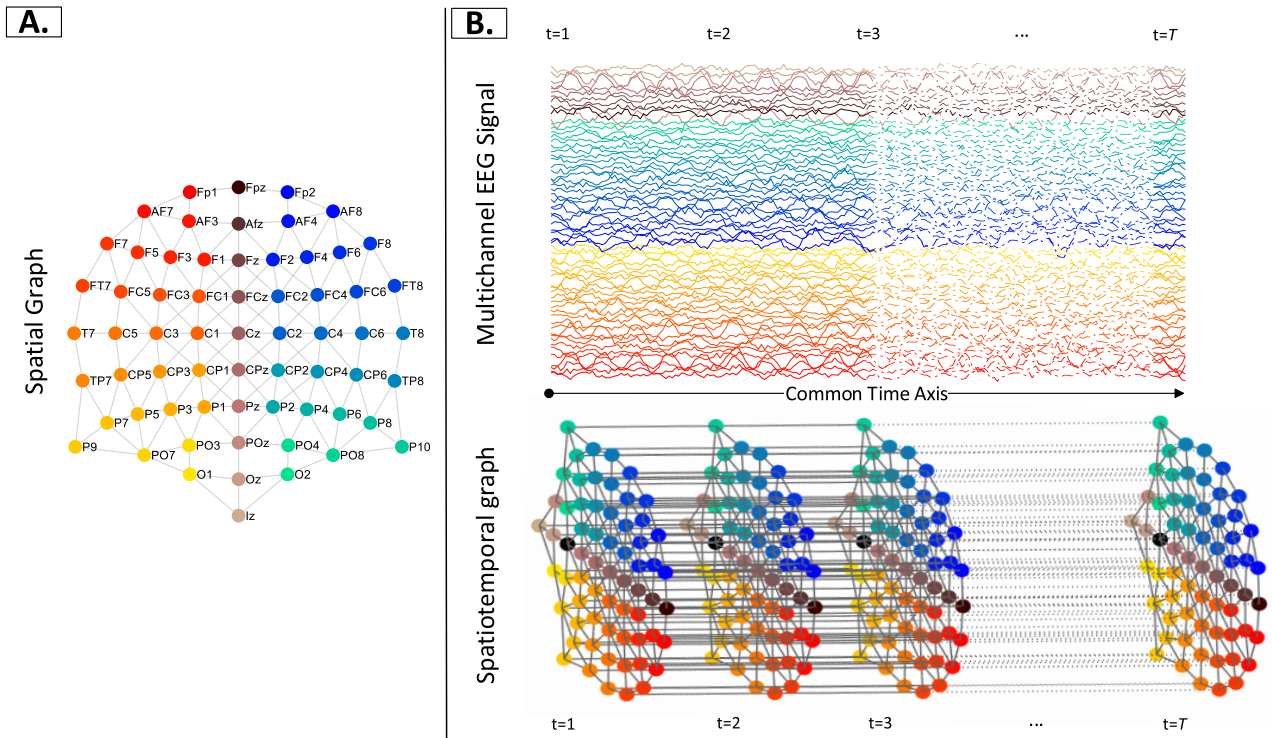


FIGURE 2. A) Spatial graph that represents the spatial connections of the recording sensor array. B) A multichannel EEG signal (top) can be defined on a spatiotemporal graph (bottom) where each time-slice corresponds to the spatial connectivity and expresses one sample in time of the multichannel EEG signal. All connections, in both spatial and spatiotemporal graphs, are binary. The colorcode indicates the electrode correspondence. We note that the white vertical line at $t=3$ appears since the visualization of the EEG signal switches, from solid to dashed lines, simultaneously for all sensors.

elements are calculated as $S_{i,j} = \delta_{i,i+E}$ with $\delta_{i,j}$ being the Kronecker delta. Actually, the Kronecker product appearing on the right-hand side of (9) creates a multilayer graph. Then, the additive terms, $\mathbf{S} + \mathbf{S}^T$, transforms the multilayer graph into a multiplex one [28], which contains bidirectional binary connections along the spatial dimension according to \mathbf{W} and temporal bidirectional binary connections only among consecutive time samples recorded at the same spatial location (i.e. sensor). Figure 2 demonstrates the spatiotemporal graph modeling of a given EEG epoch, which actually constitutes a principal contribution of this article.

III. DATASET DESCRIPTION

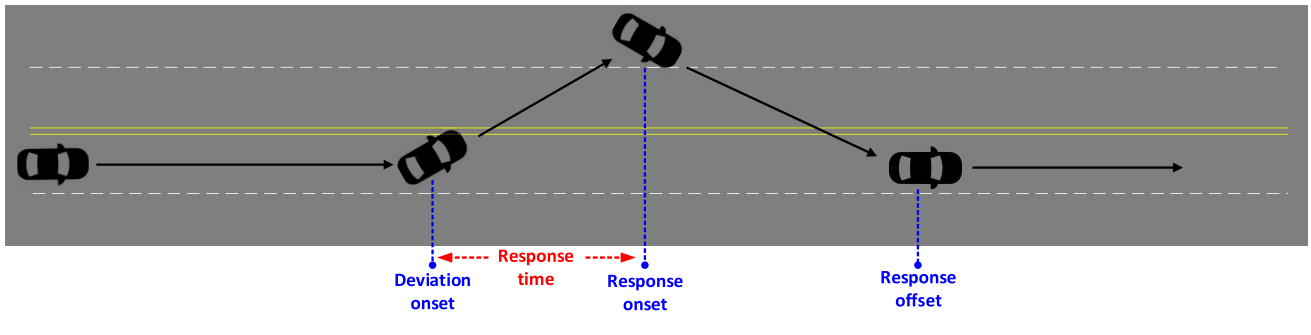
A. PREDICTING DRIVERS' RESPONSES: FAST VS SLOW

Twenty seven subjects participated in a sustained-attention driving task which took place multiple times on the same or different days with a total duration of 90 minutes [30]. An extremely realistic Virtual Reality environment was employed in order to simulate a realistic driving experience (Fig. 4). The experimental paradigm was based on a visually monotonous driving experience during nighttime on a trafficless highway with four lanes. The participants were instructed to maintain the car's course in the middle of the lane. At random time instants a lane-departure event was taking place causing the car to drift from the

central lane to one of its adjacent lanes (deviation onset). The drivers were instructed to immediately perform the corresponding driving maneuver by steering the wheel accordingly (response onset) in order to bring the car back into the central cruising lane (response offset). The elapsed time between the deviation onset and the response onset, referred to as response time, indicated the readiness and alertness of the driver. An illustration of the experimental paradigm events is depicted in Fig. 3A.

In order to avoid the implications of unrelated, with the driver's alertness, driving factors during the task participants had only to react to the lane-departure event by controlling the steering wheel solely. The accelerator and brake pedals in the experiment were deactivated and therefore had no affect in the car's operational behavior. Each lane-departure event defined a single-trial and included prestimulus activity, deviation onset, response onset and response offset. Each driver's brain activity was monitored by means of a 30-channel recording EEG at sampling rate of 500 Hz. For each participant the prestimulus brain activity corresponding to the 25% of fastest and slowest of response times was isolated and served as the data for the classification task were employed in order to validate the proposed approach. The basic idea revolves around predicting the fast and slow response times in a personalised manner by utilising the prestimulus brain electrical activity.

A. Fast vs Slow Drivers' Responses



B. Attentive vs Passive BCI-game

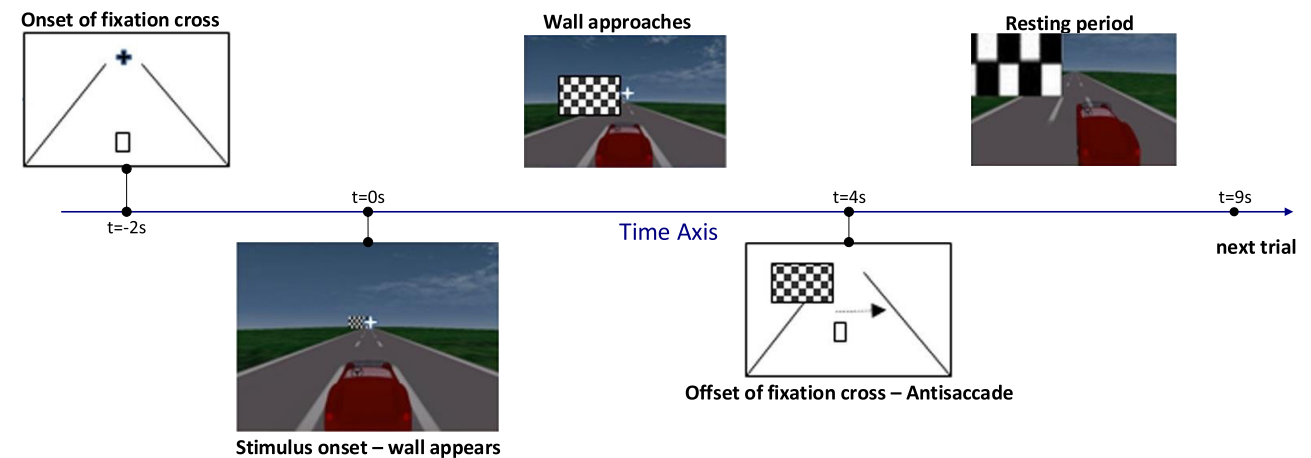


FIGURE 3. Illustrating the experimental procedures in the employed datasets. A) While the car moves in the central lane a drift towards a neighbouring lane occurs at random intervals. The driver has to restore the car's cruise into the central lane by steering the wheel accordingly. B) Illustrating the timeline of a single trial from the employed game-like BCI. The pc-game was developed by DIANA research group for the BRAINS project (<http://www.diana.uma.es/brains/> [29]).



FIGURE 4. Depicting the experimental setup that was used during the acquisition of the drivers' responses during the simulated driving experience. On top the driver's point of view is presented while the bottom images depict the employed simulation hardware. Image source: [30].

B. GAME-LIKE BCI: PASSIVE VS ATTENTIVE TASK

Six subjects (2 males, 4 females) participated in a pc-game-like BCI experiment [31]. During this experiment,

the participants were asked to drive a racing-car by using their eye-movements. As the car was moving, a wall was appearing suddenly either on the left or right side of the road and the participants had to avoid it by moving their gaze towards the opposite direction. At the beginning of each trial, a fixation cross was appearing in the center of the screen. Then, two seconds later, a wall with a checkerboard pattern appeared on either side of the road. After four seconds, the fixation cross disappeared, and the subject had to perform an anti-saccade (i.e. eye movement towards the opposite side of the checkerboard). A resting period of five seconds was taking place between consecutive trials. Fig. 3B illustrates the timeline of a single trial recording with actual images from the employed BCI-game. The origin of the time axis (i.e. 0-time instant) corresponds to the time instant of checkerboard-pattern onset. A 64-channel EEG recording device was employed in order to record the participants' brain activity with a sampling rate of 1024Hz. Moreover, an extra recording session took place where participants had been instructed to passively perceive the visual stimuli and to refrain from performing the anti-saccade. We will refer to the first condition as "attentive" condition, whereas to the second as "passive".

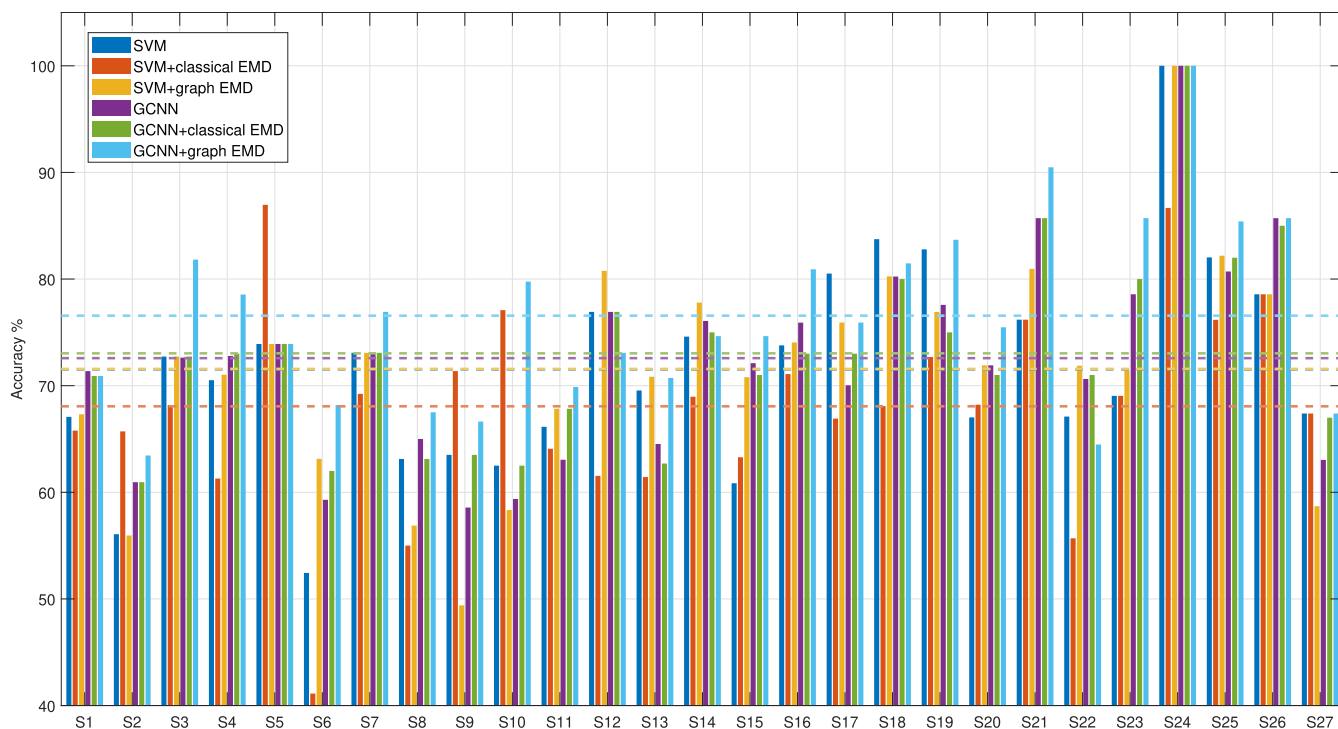


FIGURE 5. This bar diagram demonstrates the obtained classification accuracy (vertical axis) for the fast vs slow classification task for each of the participants (horizontal axis). The horizontal dashed lines indicate the mean classification values for the employed classification schemes. We note that the average classification accuracy of SVM and SVM+graph EMD is almost identical and therefore their corresponding dashed lines are overlapped in the figure.

From a neurophysiological perspective, 100 milliseconds after the appearance of the checkerboard pattern (i.e. stimulus onset) a well-defined temporal pattern, known as P100 response, arises in the sensor-space that enormously contrasts between the “left” and “right” responses with respect to the topographical laterality that builds over occipital and parietal brain regions. In contrast, the differentiation between attentive and passive responses, with the wall appearing on the same side of the road, is a more challenging classification task [32], at least at the level of a single-trial analysis and is of great significance as it can be exploited in endogenous BCIs. It is exactly this demanding classification task (i.e. attentive vs. passive brain condition), that we attempt to manage with the proposed methodological framework.

IV. RESULTS

In this section we present the results for the two classification tasks that will serve as the basis for validating the proposed method. The introduced data augmentation method is used under two different classification schemes with emphasis in the GCNNs, so as to investigate whether and under which conditions the use of geometric deep learning can bring tangible benefits with respect to baseline machine learning schemes like Support Vector Machines. Moreover, the proposed augmentation strategy is compared against the classical EMD-based approach of [16] which has shown great potential when combined with deep learning [17].

A. PREDICTING DRIVER’S RESPONSES: FAST VS SLOW

As already stated previously, the classification task concerns the differentiation between fast and slow driver’s reaction times. Actually, the aforementioned classification task exhibits similar characteristics with detecting drowsiness and alertness during driving. Many studies have demonstrated that the analysis of the EEG spectrum is capable of accurately indicating a driver’s alertness level. Typical spectrum-based methodologies revolve around investigating the established EEG frequency bands with the most prominent for this task being the alpha (8-12Hz) and theta (4-8Hz) bands as they have shown strong correlation with one’s cognitive performance. The preliminary analysis we conducted indicated that the mental states of interest (alertness vs drowsiness) mainly contrast at 6-10Hz frequency range which overlaps with both alpha and theta bands.

Prior to obtaining the EEG epochs that will be used for the classification task, the signals were cleaned from artifacts using initially the Artifact Subspace Reconstruction (ASR) [33] in order to remove large magnitude artifacts followed by Wavelet-ICA denoising [34] for fine-grained artifact rejection. Then, the signals were bandpass filtered within the range of 6-10Hz and segmented into epochs. Each epoch contained four seconds of prestimulus activity that was converted into time varying energy signals using a one-second-long sliding window and a sliding step equal to 0.2 seconds. The extracted time varying energies constitute the, new,

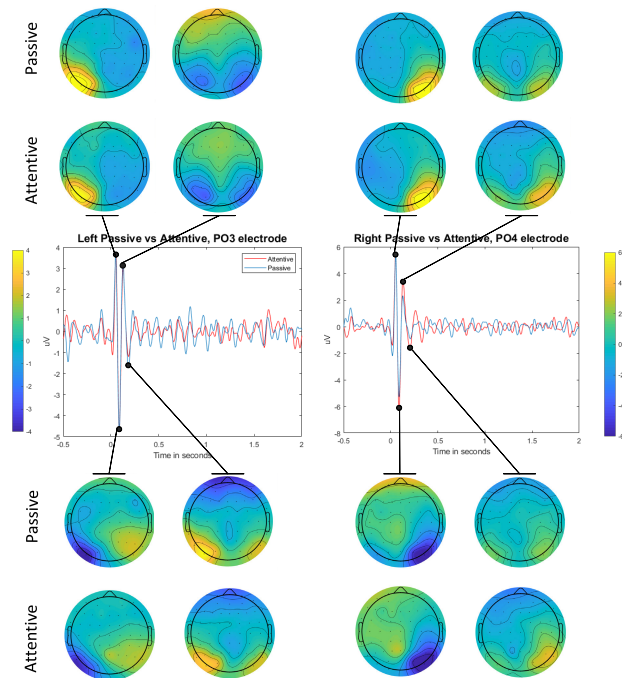


FIGURE 6. Averaged brain activation patterns, for the “Left” and “Right” responses for a single participant in both passive and attentive conditions, at electrodes that exhibit the maximum SNR (namely the PO3 and PO4 respectively). The corresponding topographic scalp potentials of the average EEG traces at selected time instants (marked with a dot), corresponding to peaks of the activation pattern, are also presented. All the topographies included in the left(right) column of the figure, share a common color-code, that extends according to the averaged signal as shown in the middle panel.

multichannel signal that will be used for inference. We must note here that the combination of ASR and Wavelet-ICA denoising leads to artifact reduced signals, by introducing a time delay that might be considerable in online scenarios.

The next essential step concerns the creation of a graph that will express initially the spatial information of the recording sensor array and then it will be extended so as to capture the time dependencies according to the spatiotemporal graph construction of section II-E. Based on the coordinates of the recording electrodes, the spatial graph construction is performed by means of a k -nearest neighbor graph. In our case, the k is equal to six as, for the employed sensor topology, it is the lowest value that leads to a connected spatial graph.

Having all the prerequisites for the classification task addressed, the EEG epochs for each participant were randomly split into three subsets, namely the training, the validation and the test set following a ratio of 80-10-10% respectively. The data of the training set were used in order to train the machine learning models whereas the validation set to investigate and uncover the most suitable hyperparameters of the classification schemes as well as the configuration for the GCNN. The test set was only put in use after all the models had been trained and established in order to produce the reported results.

One of the most important aspect that should be uncovered concerns the data augmentation ratio. As expected, the augmentation ratio (i.e. the number of artificial EEG epochs that should be generated) resulted by performing the classification task over the validation set under several different augmentation ratios. Our experiments have shown that the most suitable ratio for this task is equal to increasing the size of the training set by a factor of five. We noticed that up to the aforementioned augmentation ratio the classification accuracy in the validation set was increasing until it reached a plateau.

In order to validate the effectiveness of our method, we present the classification results under two different classification schemes. The first classification scheme concerns linear Support Vector Machines (SVMs) where the spatiotemporal structure of EEG is not taken into account and the extracted energy features from the multichannel EEG signal are treated as a large multidimensional vector. The second classification scheme corresponds to GCNN with two graph convolutional layers followed by two fully connected layers and the output layer (obtained through hyperparameter grid search). These two schemes are tested under three different augmentation approaches. The first approach concerns the original training set with no augmentation at all, the second concerns the classical EMD-based augmentation strategy of [16], where the signals’ spatial information is not considered, and the third concerns the proposed, GEMD-based, method. Figure 5 demonstrates that the baseline GCNNs manage to achieve slightly better classification results than all of the SVM-based approaches showing the importance of taking into account the spatiotemporal information of the signals. The best classification accuracy is actually achieved by employing the proposed, GEMD-based, augmentation methodology reaching on average 76.56% and surpassing the classical EMD-based by 4.02% and the baseline GCNN model (no augmentation) by 5.98%. As expected, the SVMs, which typically require less data to achieve their top-limit performance, do not exhibit the same improvement in classification performance by employing a data augmentation strategy.

By taking a closer look at the results contained in Fig 5, one can see that the classification performance varies significantly across the participants with accuracy values ranging from 65% to 100%. In an effort to explain this large variance we may consider the following. As we have already stated, the two class problem (slow vs fast reaction) was the result of keeping the 25% of slowest and the 25% of fastest responses for each participant. Although the problem degenerates into a binary classification task, there is no indication whether the fastest and slowest response times differ (e.g. a participant could have similar response times in each trial). By performing a post-classification analysis of the results we were able to associate the achieved classification accuracy with the difference in the response times for each class (e.g. subtracting the fastest response among the ones labeled as slow from the slowest response among the ones labeled as fast).

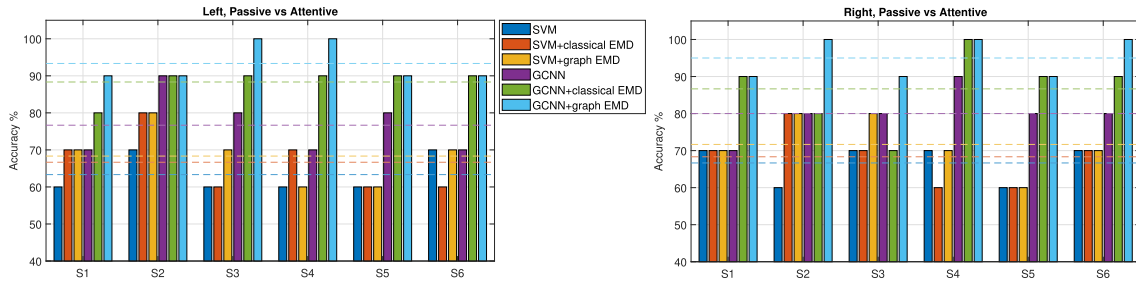


FIGURE 7. This bar diagram presents the obtained classification accuracy (vertical axis) for the attentive vs passive classification task in a personalized manner (subject ID - horizontal axis) for both left and right trials. The horizontal dashed lines indicate the mean classification value for each of the employed classification scheme.

Therefore, for participants where the achieved classification accuracy is slow, the gap between the slow and fast responses is very small (a few milliseconds) whereas in the cases of high classification accuracy, this gap extends to the order of several seconds. The aforementioned association is quantified by a Pearson's correlation coefficient $r = 0.74$.

B. GAME-LIKE BCI: PASSIVE VS ATTENTIVE TASK

We commence by presenting the neurophysiological findings in the “passive vs attentive” task. Fig. 6 presents the brain activation patterns in the most prominent electrodes (i.e. those where the brain activation pattern demonstrates the highest signal-to-noise ratio) that occur after the stimulus onset accompanied by the corresponding topographic scalp potentials (averaged across trials for a single participant). It becomes evident that although the “left vs right” trials are easily separable due to the laterality of brain activation patterns, the “passive vs attentive” task is more demanding as the brain activation patterns are similar both in sensor-space and the time-domain. Therefore, the classification task concerns the discrimination of the passive vs the attentive conditions when the checkerboard appeared on the same side of the screen.

For the single trial classification the EEG epochs contained 300 milliseconds-long multichannel EEG signals starting from the stimulus onset. The signals were bandpass filtered within the alpha (8-12Hz) frequency range so as to isolate the brainwaves that are mostly associated with the conditions we aim to separate [35], [36]. Then the EEG epochs were split into training, validation and testing sets in a 80-10-10% ratio respectively for each participant independently. As in section IV-B, the training set was used in order to train the classification models and the validation set so as to investigate and uncover the hyperparameters and specify the GCNN's specific configuration. The test set was used solely for the purpose of reporting the classification results.

Concerning the spatiotemporal graph construction, the spatial graph was created as a k nearest neighbour ($k = 6$) graph with respect to the recording sensor topographical distribution and then it was extended to capture time dependencies also, according to the methodology of section II-E. We note that in the “attentive vs passive” task, the graph signals used

for classification are the filtered raw EEG signals. Previous studies have shown that these two conditions can be separated by considering the brain's functional connectivity [37] and it is within our expectation that the GCNN will be able to handle this task more effectively. The particular GCNN architecture, as obtained through hyperparameter grid search, for this task concerns two graph convolutional layer, followed by a fully connected layer with rectified linear units and an output layer.

Again, one important aspect of the classification procedure concerns the augmentation ratio. Our experimental results on the validation set uncovered that the best classification performance is achieved when the training set is augmented by a factor of five. The preliminary results showed that, increasing the training set further offered no improvement in terms of classification in the validation set.

In Fig. 7 we present the classification accuracy obtained using both SVMs and GCNNs. Each classification approach is combined with classical EMD as well as GEMD based augmentation strategies. In order to examine the improvement that each of the aforementioned augmentation strategies offers, the classification accuracy, when no augmentation strategy is applied, is also reported. The experimental results on this dataset reveal not only that the GCNNs once again meet a significant improvement in terms of classification accuracy by the proposed augmentation strategy, but also that the SVMs are inadequate for this task. It can be seen in Fig. 7 that the highest classification performance is achieved by a combination of GCNNs with the proposed data augmentation strategy in both “left” and “right” trials achieving 93% and 95% accuracy respectively. These results are significantly improved, by 5% and 9% for the “left” and “right” trials respectively, compared to the data augmentation approach of [16] when GCNNs are used for classification. Although these trends also hold for the SVMs' case, their top performance which is achieved by combining SVMs with GEMD-based augmentation strategy does not exceed the 72% in either “left” or “right” responses.

V. DISCUSSION AND CONCLUSION

In this article we introduce a novel data augmentation methodology suitable for graph signals and consequently their corresponding, suitable, classification schemes.

By exploiting the graph variant of empirical mode decomposition we generate artificial EEG signals in an effort to improve the classification accuracy of personalized BCIs where the training samples are truly limited. Our experiments indicate that the introduced augmentation strategy improves significantly the classification accuracy of the GCNN models. Unsurprisingly, this trend is not maintained when SVM models are employed. Although the performance of deep learning models is inextricably connected with the size of the training set this fact does not hold for the SVMs as they are less dependant to the size of the training data corpus. Moreover, the introduced augmentation strategy is tailored so as to preserve the underlying structure of the EEG signals that is inherently learnt and considered by the GCNN models whereas being neglected in the case of SVMs.

It was among the scopes of this work to perform a feasibility study concerning the exploitation of geometric deep learning in the field of BCIs. The presented experiments on the first dataset (i.e. fast vs slow driver's responses) demonstrate that the baseline machine learning schemes (SVMs in our case) are on par, in terms of classification performance, with the employed geometric deep learning models when no augmentation approach is applied. However, data augmentation strategies seem to significantly benefit the geometric deep learning models, in contrast to the classical machine learning schemes, allowing them to be effectively exploited by unlocking their great decoding potential. For the second dataset (e.g. attentive vs passive responses), which is known that the contrasting conditions become more separable when working with connectivity features, the baseline performance is in favour of the GCNNs. Although both classification schemes benefit from the proposed augmentation strategy the GCNNs demonstrate the most significant improvement.

It is within our expectations that the presented, encouraging, results will promote the employment of geometric deep learning not only in the field of BCI but also in computational neuroscience broadly. However, it should be mentioned that the theoretical validation of the introduced augmentation strategy remains an open issue. Moreover, in order to further foster the combination of geometric deep learning and computational neuroscience we leave as a future work the development of weighted, instead of binary, spatiotemporal brain modeling. Recent works that facilitate the analysis of graph structural data that also evolve in time [38] appear as promising alternatives capable to bring forth additional benefit.

REFERENCES

- [1] W. Wu, S. Nagarajan, and Z. Chen, "Bayesian machine learning: EEG\MEG signal processing measurements," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 14–36, Jan. 2016.
- [2] F. P. Kalaganis, N. A. Laskaris, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "A Riemannian geometry approach to reduced and discriminative covariance estimation in brain computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 245–255, Jan. 2020.
- [3] S. Kumar, A. Sharma, and T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information," *BMC Bioinf.*, vol. 18, no. S16, p. 545, Dec. 2017.
- [4] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Adaptive multi-degree of freedom brain computer interface using online feedback: Towards novel methods and metrics of mutual adaptation between humans and machines for BCI," *PLoS ONE*, vol. 14, no. 3, Mar. 2019, Art. no. e0212620.
- [5] L. Rui, H. Nejati, and N.-M. Cheung, "Dimensionality reduction of brain imaging data using graph signal processing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1329–1333.
- [6] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.
- [7] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [8] K. Georgiadis, N. Laskaris, S. Nikolopoulos, and I. Kompatsiaris, "Connectivity steered graph Fourier transform for motor imagery BCI decoding," *J. Neural Eng.*, vol. 16, no. 5, Aug. 2019, Art. no. 056021.
- [9] V. P. Buch, A. G. Richardson, C. Brandon, J. Stiso, M. N. Khattak, D. S. Bassett, and T. H. Lucas, "Network brain-computer interface (nBCI): An alternative approach for cognitive prosthetics," *Frontiers Neurosci.*, vol. 12, p. 790, Nov. 2018.
- [10] F. De Vico Fallani and D. S. Bassett, "Network neuroscience for optimizing brain-computer interfaces," *Phys. Life Rev.*, vol. 31, pp. 304–309, Dec. 2019.
- [11] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.
- [12] A. Harati, S. Lopez, I. Obeid, J. Picone, M. P. Jacobson, and S. Tobochnik, "The TUH EEG CORPUS: A big data resource for automated EEG interpretation," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2014, pp. 1–5.
- [13] K. Gregor Hartmann, R. Tibor Schirrmester, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals," 2018, *arXiv:1806.01875*. [Online]. Available: <http://arxiv.org/abs/1806.01875>
- [14] A. Paris, G. K. Atia, A. Vosoughi, and S. A. Berman, "A new statistical model of electroencephalogram noise spectra for real-time brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1688–1700, Aug. 2017.
- [15] F. Lotte, "Generating artificial EEG signals to reduce BCI calibration time," in *Proc. 5th Int. Brain-Comput. Interface Workshop*, Graz, Austria, Sep. 2011, pp. 176–179.
- [16] J. Dinarès-Ferran, R. Ortner, C. Guger, and J. Solé-Casals, "A new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI," *Frontiers Neurosci.*, vol. 12, p. 308, May 2018.
- [17] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.
- [18] N. Tremblay, P. Borgnat, and P. Flandrin, "Graph empirical mode decomposition," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2014, pp. 2350–2354.
- [19] T. He, L. Bi, J. Lian, and H. Sun, "A brain signals-based interface between drivers and in-vehicle devices," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 1333–1337.
- [20] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
- [21] F. R. Chung and F. C. Graham, *Spectral Graph Theory*, vol. 92. Providence, RI, USA: American Mathematical Society, 1997.
- [22] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.
- [23] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
- [24] G. Rilling, P. Flandrin, and P. Goncalves, "On empirical mode decomposition and its algorithms," in *Proc. IEEE Workshop Nonlinear Signal Image Process. (EURASIP)*, vol. 3, Jun. 2003, pp. 8–11.
- [25] L. J. Grady and E. L. Schwartz, *Anisotropic Interpolation on Graphs: The Combinatorial Dirichlet Problem*. Boston, MA, USA: Boston Univ., Center for Adaptive Systems and Department of Cognitive and Neural Systems, 2003.

- [26] G. Wang, X.-Y. Chen, F.-L. Qiao, Z. Wu, and N. E. Huang, "On intrinsic mode function," *Adv. Adapt. Data Anal.*, vol. 2, no. 3, pp. 277–293, Jul. 2010.
- [27] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Phys. Rep.*, vol. 544, no. 1, pp. 1–122, Nov. 2014.
- [28] M. De Domenico, S. Sasaï, and A. Arenas, "Mapping multiplex hubs in human functional brain networks," *Frontiers Neurosci.*, vol. 10, p. 326, Jul. 2016.
- [29] R. Ron-Angevin and A. Díaz-Estrella, "Brain–computer interface: Changes in performance using virtual reality techniques," *Neurosci. Lett.*, vol. 449, no. 2, pp. 123–127, Jan. 2009.
- [30] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, no. 1, pp. 1–8, Dec. 2019.
- [31] D. Liparas, S. I. Dimitriadis, N. A. Laskaris, A. Tzelepi, K. Charalambous, and L. Angelis, "Exploiting the temporal patterning of transient VEP signals: A statistical single-trial methodology with implications to brain–computer interfaces (BCIs)," *J. Neurosci. Methods*, vol. 232, pp. 189–198, Jul. 2014.
- [32] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors," *Sensors*, vol. 13, no. 8, pp. 10273–10286, Aug. 2013.
- [33] C.-Y. Chang, S.-H. Hsu, L. Pion-Tonachini, and T.-P. Jung, "Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 1114–1121, Apr. 2020.
- [34] G. Inuso, F. La Foresta, N. Mammone, and F. C. Morabito, "Wavelet-ICA methodology for efficient artifact removal from electroencephalographic recordings," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2007, pp. 1524–1529.
- [35] J. J. Foxe and A. C. Snyder, "The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention," *Frontiers Psychol.*, vol. 2, p. 154, Jul. 2011.
- [36] W. Klimesch, "Alpha-band oscillations, attention, and controlled access to stored information," *Trends Cognit. Sci.*, vol. 16, no. 12, pp. 606–617, Dec. 2012.
- [37] F. P. Kalaganis, N. A. Laskaris, E. Chatzilari, D. A. Adamos, S. Nikolopoulos, and I. Kompatsiaris, "A complex-valued functional brain connectivity descriptor amenable to Riemannian geometry," *J. Neural Eng.*, vol. 17, no. 2, Apr. 2020, Art. no. 024001.
- [38] F. Grassi, A. Loukas, N. Perraudin, and B. Ricaud, "A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs," *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 817–829, Feb. 2018.



NIKOLAOS A. LASKARIS is currently an Associate Professor with the Department of Informatics, Aristotle University, Greece. He is a member of AIIA Laboratory and leads the Neuroinformatics.GROUP. He is a coauthor of more than 100 scientific publications. His current research interests include neuroinformatics, brain connectomics and the applications of machine learning, data mining and nonlinear dynamics in biomedicine and neuroscience.



ELISAVET CHATZILARI received the Diploma degree in electronics and computer engineering from the Aristotle University of Thessaloniki, in 2008, and the Ph.D. degree on social-based scalable concept detection from the University of Surrey, in 2014. She is currently a Postdoctoral Research Fellow with the Centre for Research and Technology Hellas (CERTH), Information Technologies Institute (ITI). Her research interests include machine learning, image processing, social media analysis, and bio-signal processing. Her scientific work has been published in peer-reviewed journals, international conferences, and book chapters.



SPIROS NIKOLOPOULOS received the Diploma degree in computer engineering and informatics and the M.Sc. degree in computer science and technology from the University of Patras, Greece, in 2002 and 2004, respectively, and the Ph.D. degree in semantic multimedia analysis using knowledge and context, from the Queen Mary University of London, in 2012. He is currently a Senior Researcher with the Centre for Research and Technology Hellas (CERTH), Information Technologies Institute (ITI). His research interests include computer vision, big data and visual analytics, visual and augmented reality, and brain–computer interfaces. His scientific work has been published in more than 24 peer-reviewed journals, 66 international conferences, and 15 book chapters.



IOANNIS (YIANNIS) KOMPATSIARIS is currently a Research Director with CERTH-ITI and the Head of Multimedia Knowledge and Social Media Analytics Laboratory. His research interests include multimedia, big data and social media analytics, semantics, human–computer interfaces (AR and BCI), eHealth, security and culture applications. He is the coauthor of 171 articles in refereed journals, 63 book chapters, eight patents, and more than 500 articles in international conferences. Since 2001, he has participated in 89 National and European research programs in 18 of which he has been the Project Coordinator and in 14 direct collaborations with industry. He has been the co-organizer of various international conferences and workshops and has served as a regular Reviewer, an Associate Member, and a Guest Editor for a number of journals and conferences currently being an Associate Editor of the *IEEE TRANSACTIONS ON IMAGE PROCESSING* and *BIG DATA JOURNAL*. He is a member of the Scientific Advisory Board of the CHIST-ERA funding program, an elected member of the *IEEE Image, Video, and Multidimensional Signal Processing-Technical Committee (IVMSP-TC)* and a member of ACM.



FOTIS P. KALAGANIS (Graduate Student Member, IEEE) graduated from the Department of Informatics of the Aristotle University of Thessaloniki, in 2013. He received the M.Sc. degree in digital media and computational intelligence from the Aristotle University of Thessaloniki, in 2016. He is currently pursuing the Ph.D. degree with the Department of Informatics, Aristotle University of Thessaloniki. During the master's program, he was awarded for excellence in academic studies.

He is currently working as a Research Associate with the Centre for Research and Technology Hellas (CERTH), Information Technologies Institute (ITI). His research interests include deep learning, complex-valued kernel methods, and non-Euclidean geometry with applications in computational neuroscience and brain–computer interfaces. His scientific work has been published in prestigious peer-reviewed journals and international conferences.

...