Cross-Subject Zero Calibration Driver's Drowsiness Detection: Exploring Spatiotemporal Image Encoding of EEG Signals for Convolutional Neural Network Classification

João Ruivo Paulo, Gabriel Pires, *Member, IEEE*, and Urbano J. Nunes, *Senior Member, IEEE*

***Abstract*— This paper explores two methodologies for drowsiness detection using EEG signals in a sustained-attention driving task considering pre-event time windows, and focusing on cross-subject zero calibration. Driving accidents are a major cause of injuries and deaths on the road. A considerable portion of those are due to fatigue and drowsiness. Advanced driver assistance systems that could detect mental states which are associated with hazardous situations, such as drowsiness, are of critical importance. EEG signals are used widely for brain-computer interfaces, as well as mental state recognition. However, these systems are still difficult to design due to very low signal-to-noise ratios and cross-subject disparities, requiring individual calibration cycles. To tackle this research domain, here, we explore drowsiness detection based on EEG signals' spatiotemporal image encoding representations in the form of either recurrence plots or gramian angular fields for deep convolutional neural network (CNN) classification. Results comparing both techniques using a public dataset of 27 subjects show a superior balanced accuracy of up to 75.87% for leave-one-out cross-validation, using both techniques, against works in the literature, demonstrating the possibility to pursue cross-subject zero calibration design.**

***Index Terms*— Driver's drowsiness detection, electroencephalography, recurrence plot, Gramian angular fields, convolutional neural network.**

I. Introduction

MENTAL fatigue and drowsiness are associated with a substantial number of road traffic accidents. It is estimated to reach as much as 20% of the total number globally [1]. Its detection can help prevent a significant number of these events and with them injuries and deaths.

Fatigue and drowsiness detection has been extensively studied to better understand and prevent their negative effects in performance activities where vigilance is required, like heavy-machine operation [2], surgery [3], and driving [4]. Different modalities have been used to assess and record data related with drowsiness. The subjective approach makes use of psychometric questionaires [5]–[8], however, a self-assessment questionnaire-based methodology is unpractical in the driving context and does not assess sudden variations in sleepiness level [9]. More practical solutions that can be used in the driving context resort to visual monitoring or detection of specific behaviors like pupil dilation and eye blinking frequency [10], however, these last approaches are affected by occlusions, light changes, and facial apparel. Another approach is the use of physiological measurements, such as electrocardiography (ECG) [11], electrooculography (EOG) [12], and electroencephalography (EEG) [9]. These approaches have the capability of early detection and allow anticipatory intervention. Considering the physiological approach, EEG is the most researched and promising modality, as it directly measures brain activity associated with the underlying processes of drowsiness, and features high temporal resolution. EEG has been used in the driving context much beyond just drowsiness detection. It has been used for applications like vehicle control (brain-computer interface) [13], motor anticipatory potentials to predict driving actions [14], correct automatic erroneous decisions (e.g., driving cues provided by intelligent car systems) [15] based on error-related potentials [16], and detection of stressful situations [17]–[19].

Looking at the EEG modality, a problem intrinsically present is the inter-subject variability. The EEG signals differ from subject to subject and even for the same subject, they change from trial to trial. The signals are non-stationary, non-stable, and non-linear [20]. This makes it difficult to create an efficient drowsiness detection system that can be used with everyone. Most works on drowsiness and fatigue detection in the driving context test their approaches using subject-dependent models, which require a calibration step for each subject [21]–[23]. However, systems should be

Manuscript received January 13, 2021; revised March 19, 2021 and April 20, 2021; accepted May 6, 2021. Date of publication May 12, 2021; date of current version May 18, 2021. This work was supported in part by the Portuguese Foundation for Science and Technology (FCT) through the funded projects under Grant B-RELIABLE-PTDC/EEIAUT/30935/2017 and in part by the Institute of Systems and Robotics, University of Coimbra (ISR-UC) FCT grant, Portugal, under Grant UIDB/00048/2020. *(Corresponding author: João Ruivo Paulo.)*

João Ruivo Paulo is with the Institute of Systems and Robotics, University of Coimbra, 3030-290 Coimbra, Portugal (e-mail: jpaulo@isr.uc.pt).

Gabriel Pires is with the Institute of Systems and Robotics, University of Coimbra, 3030-290 Coimbra, Portugal, and also with the Engineering Department, Polytechnic Institute of Tomar, 2300-313 Tomar, Portugal (e-mail: gpires@isr.uc.pt).

Urbano J. Nunes is with the Institute of Systems and Robotics, University of Coimbra, 3030-290 Coimbra, Portugal, and also with the Department of Electrical and Computer Engineering, University of Coimbra, 3030-290 Coimbra, Portugal (e-mail: urbano@isr.uc.pt).

Digital Object Identifier 10.1109/TNSRE.2021.3079505

calibration-free since it is unpractical to be required to acquire calibration data related to non-fatigue and fatigue conditions for each new user, which can take hours. Hence, drowsiness detection systems should focus on subject-independent development and validation.

In this work, we contribute to the state of the art by presenting an approach to drowsiness detection, while driving, focusing on the design of a system that can be used by any user without the need for user-specific calibration, hence cross-subject zero calibration. We present novel EEG data representations based on encoding for deep learning classification. We exploit recurrence plots and gramian angular fields to transform EEG signals' band powers into an image-like structure. This image is fed to a convolutional neural network (CNN) with a single convolutional layer architecture, which is less computationally expensive then the other approaches found in the EEG literature, but still efficient.

Results with a public dataset of 27 individuals performing a simulated sustained-attention driving task, reveal a cross-subject accuracy above 75%, which is a superior result comparatively to the works found in the literature.

This paper is organized as follows: Section II is dedicated to the related works of drowsiness detection, particularly focused on driving tasks. Section III describes the proposed methodology, followed by its validation in Section IV, where results are presented. Lastly, Section V presents the final remarks and future lines of research.

II. Related Work

Drowsiness detection has been extensively studied using neurocognitive information, particularly, through EEG, which has been used to evidence changes in global brain dynamics related to changes in alertness during driving [24], [25]. EEG was also studied as a robust measurement for driver's cognitive state estimation [26], and mental state changes through the exploration of working memory [27] and inhibitory control [28]. However, EEG acquisition setups still face many challenges, from the electrode types, to their cumbersome montages, as well as the specific EEG signals' properties that make them difficult to handle due to low signal-to-noise ratio and non-stationary nature, with the added trait of high variability between individuals.

In the literature, we can find studies addressing drowsiness and fatigue detection using EEG, like the work of Jap *et al*. [29] who demonstrate that the ratio of slow to fast EEG waves increased when the subject is influenced by fatigue. Gharagozlou *et al*. [30] suggested that different levels of fatigue can be estimated using band power features and EEG signal entropy features, reporting a significant increase in alpha power associated with driver fatigue. In [31], 4 different types of entropy combined with different classifiers were used for subject-dependent driver fatigue classification. The work in [32] used approximate entropy and Kolmogorov complexity to discriminate between fatigue states. In [33], entropy features were used combined with Gradient Boosting Decision Tree Model. The use of deep learning models for fatigue classification was also proposed, as in [22], through a Residual Convolutional Neural Network (EEG-Conv-R), using data collected from 10 healthy subjects over 16 channels. In [23] a combination of a deep neural network with support vector machine (SVM) classifier at the last layer was also proposed. In [34], [35], the authors used graph convolutional networks and multi-layer perceptron, respectively, to address fatigue detection in pre-event and during steering wheel operation on the same dataset of our work.

Most of the previous works are based on intra-subject approaches. Cross-subject approaches have been proposed, but mixing up EEG samples from all subjects, followed by splitting them into training and testing randomly, like [22]. This approach, however, due to its random nature, ultimately, ends up mixing some training subjects' samples with the testing ones, not being truly cross-subject. Other works perform cross-session instead of cross-subject [36], but this is only useful for the same subject in future trials. To the best of our knowledge, it could only be found 4 preliminary works which performed cross-subject validation, considering the driving context [34], [35], [37], [38]. In [37], the authors perform domain adaptation, a branch of transfer learning, to adapt the data distributions of source and target so that the classification could be more efficient in a cross-subject scenario. In [38], EEG features, statistics, higher order crossing, fractal dimension, hjorth, signal energy, and spectral power were extracted and combined with several classifiers, such as logistic regression, linear discriminant analysis, 1-nearest neighbor, linear SVM, and naïve Bayes. While in [34], [35], as described above, the authors pursue driver's drowsiness detection using the same dataset as in this paper, relying neural network-based methodologies.

Considering the methodologies used in this work (recurrence plots and gramian angular fields), that have been successfully applied in computer vision algorithms combined with deep learning [39]–[41], these are now starting to be used in recent works in the EEG research domain, but still are relatively unexplored. In [42]–[44] recurrence plots are used for epileptic seizures detection, Alzheimer's resting state analysis, and motor imagery, respectively. In [45], [46] gramian angular fields are used in combination with deep learning techniques for epileptic seizure detection.

The use of these methodologies for a cross-subject application is considered to be a good fit to our purpose since recurrence plots and gramian angular fields are time series analysis techniques, which are not limited by data stationarity and size constraints of dynamical nonlinear systems' features which are present in EEG signals due to subjects' variability. These methodologies may find application in any rhythmical system, whether they are mechanical, electrical, neural, hormonal, chemical, or even spacial [47].

EEG calibration and non-practical systems are, together, two of the most critical issues that limit the use of EEG in scenarios outside the laboratory. Additionally, the very low signal-to-noise ratio (SNR) of EEG signals is one of the main challenges faced by EEG-based systems. Here, the CNN acts as a statistical temporal and spatial filter in search of relevant features that maximize the discrimination between classes, indirectly improving the SNR, which, however, can not be

quantified as in other domains. The scenario in our work is even more challenging, as it deals with the variability between subjects that may result from different electrode locations (due to different sizes of the head), and different mental and cognitive states. The signals of interest to be found are, therefore, unique patterns common to all subjects.

III. Methodology

A. Dataset Description and Processing

1) Sustained-Attention Driving Dataset: To study drowsiness detection in the driving context we used the publicly available sustained-attention driving task dataset detailed in [48]. In a sustained-attention task, reaction-times to events are directly related with drowsiness [49]. This is also true for the driving context [36], [37], [50]. The authors of the dataset collected data from 27 participants, aged from 22 to 28 years, who enrolled in 90-minute sustained-attention driving sessions at different times on the same or different days. The participants drove a car in a VR-based driving simulator, where an event-related lane-departure paradigm was implemented. The subject's reaction time to perturbations during a continuous driving task was quantitatively measured. The experimental paradigm simulated night-time driving on a four-lane highway, and the subject was asked to keep the car cruising in the center of the lane. Lane-departure events were randomly introduced within a 5 to 10 second period following a previous trial. These events forced the car to drift from the cruising lane towards the sides (deviation onset). The participants were instructed to immediately correct the trajectory to the original cruising lane (response onset).

During the experiments, EEG signals were collected using the Scan SynAmps2 Express system (Compumedics Ltd., VIC, Australia). Recorded EEG signals were collected using a wired EEG cap with 32 Ag/AgCl electrodes, including 30 EEG electrodes and 2 reference electrodes (opposite lateral mastoids). The EEG electrodes were placed according to a modified international 10–20 system. The EEG recordings were sampled at 500 Hz.

The EEG dataset was band-pass filtered between 1 Hz and 50 Hz. Artefact rejection and apparent eye blink contamination removal was performed by the authors manually, followed by Automatic Artifact Removal (AAR) for automatic correction of ocular and muscular artefacts in the signals.

2) Data Labeling: With the goal of drowsiness detection through a classification strategy, we adopted an approach to label observations as either 'Alert' or 'Drowsy', as proposed in [36]. Considering the pre-event period before the lane-departure event, we extract a 3-second window of EEG data before the event, as shown in Fig. 1. These windows are the observations to be classified, defining if the subject is either 'Alert' or 'Drowsy'. Their labeling is based on the reaction time (RT) of the subject for the respective trial, which is the time between the lane-departure event and the response onset. The RT at each lane-departure event was named local RT, which represents the short-term level of drowsiness. On the other hand, the long-term level of drowsiness was defined by global RT, which was calculated by averaging the RTs across all trials within a 90-second window before the respective trial in study. For each driving session, the 'alert RT' was measured as the 5th percentile of local RTs across the entire session, representing the RT that the subject could perform during alertness. Considering these variables, the labeling of the observations is divided in 'Alert' when both local and global RT are shorter than 1.5 times the alert RT, whereas they are labeled 'Drowsy', when the windows are associated with trials with local and global RT both above 2.5 times alert RT [36]. This categorization excludes the transitioning trials that correspond to a moderate performance of the driver which do not require intervention. In this sense, this work focuses on the detection of low performance (drowsiness) "where intervention is crucial" by training the CNN with well-defined 'Alert' and 'Drowsy' observations. The total number of 'Alert' and 'Drowsy' windows for the 27 subjects were 8903 and 3976, respectively.

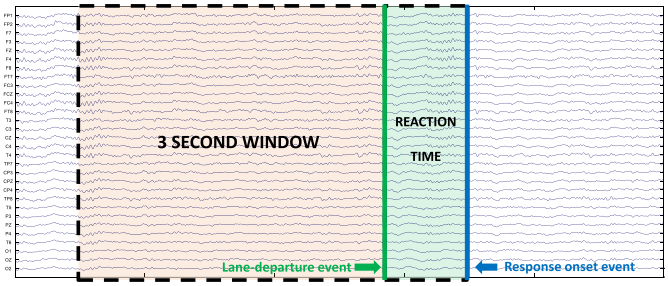

Fig. 1. EEG signals for all the channels with a 3-second window and visual events registration, the time frame between the lane departure event (green) and the response onset (blue) is the reaction time of the subject which is used to label the 3-second windows.

3) Neurophysiological Analysis: To understand the behavior of the EEG rhythms, we performed topographical analysis comparing both scenarios ('Alert' and 'Drowsy'). Figure 2 presents the topographical views for the band power difference of specific EEG frequency sub-bands to evaluate their relation to drowsiness and alertness states (δ, θ, α, β, and γ), showing the power differences between 'Drowsy' and 'Alert' observations from all subjects over 4 time intervals within a 3-second observation window. For this particular analysis, it was used 100 ms frames of cumulative power per channel averaged across all subjects using

$$P^c_{all} = \frac{\sum_{s=1}^{27}\sum_{n=n_1}^{n_f} X^c_{D,b}(n)^2 - \sum_{s=1}^{27}\sum_{n=n_1}^{n_f} X^c_{A,b}(n)^2}{27} \tag{1}$$

where $X^c_{D,b}$ and $X^c_{A,b}$ are the filtered signals for each respective frequency sub-band b per channel c for 'Drowsy' and 'Alert' observations, respectively, for all the 27 users. n_1 and n_f are the beginning and end of the particular 100 ms frame, respectively. This width of 100 ms allows us the analysis of the neurophysiological temporal dynamics of the brain during the sustained-attention driving task at different sections of the observation. It supported our selection of frequency sub-bands and groups of channels to use.

This aimed to understand the spatial dynamics between the two scenarios, hence identifying predominant discriminative

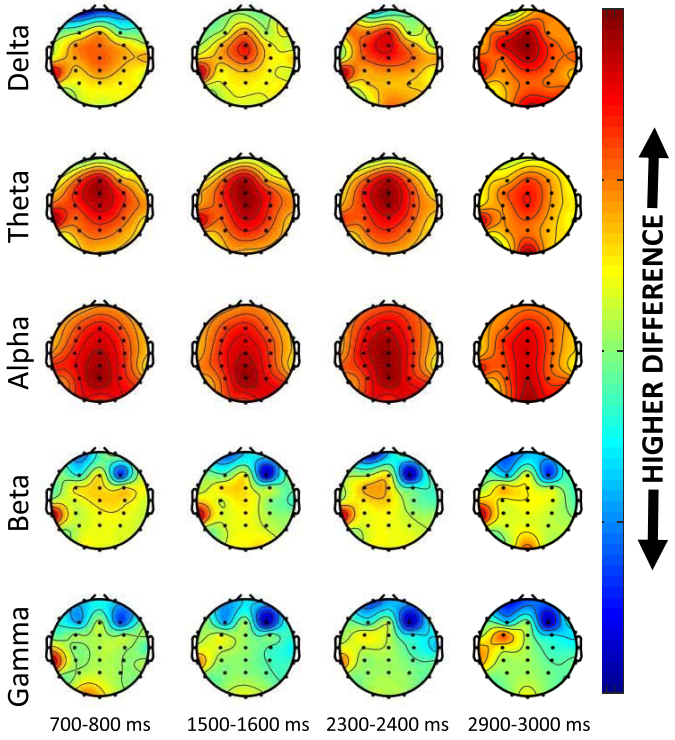

Fig. 2. Topographical view of the difference of the band powers of the EEG signals between 'Drowsy' and 'Alert' in different moments of a trial (3-second window) using data from the whole trials of all the subjects. The difference is obtained for each band after computing the mean power for each context from all subjects.

activation regions in the brain (scalp), and observe the discrimination in the power of the EEG frequency sub-bands. From Fig. 2, we can observe a wide discriminative activation of different regions of the brain across the whole scalp, being most widely distributed on the theta and alpha frequency bands. In these bands the frontal, central, and parietal regions are very active, but as well prefrontal, temporal, and occipital regions show high activity. All channels show this activity (Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T3, C3, Cz, C4, T4, TP7, CP3, CPz, CP4, TP8, T5, P3, Pz, P4, T6, O1, Oz, and O2). In the delta band, it is observed a more significant activity in the central and frontal regions (Fz, FCz, and Cz). Looking at the beta and gamma frequency bands, there is a stronger discriminative presence in the frontal and prefrontal regions (Fp1, Fp2, F7, F3, Fz, F4, and F8). Increases in alpha and theta rhythms have been shown to be related to decreases in vigilance and performance degradation [51], which is very evident in the topographic map, showing the 'Drowsy' state with a higher power (positive difference) in theta and alpha sub-bands. The delta band also shows an increase of power, but with some time dependency associated. On the other hand, the beta and gamma bands show a power decrease, as the alertness level decreases, which is also consistent to the literature [52], [53].

We have also analyzed the two scenarios using a r^2 statistical measurement. For the entire 3-second observation,

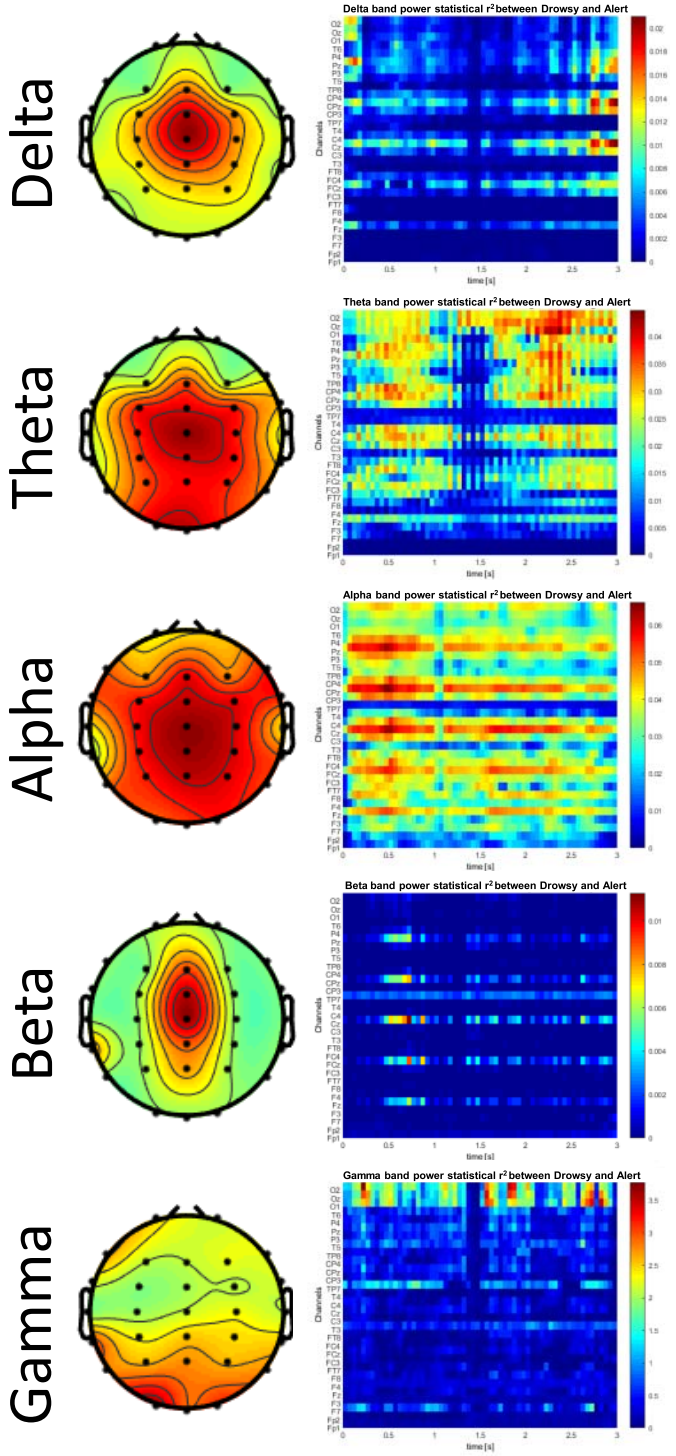

Fig. 3. On the left, the topographical view of the r^2 of the band powers of the EEG signals between 'Drowsy' and 'Alert' for the whole 3-second window using data from the whole trials of all the subjects. On the right the view of r^2 per channel along the 3-second window with a sample of 50 ms cumulative power.

we analyzed the discrimination at each channel per band power, as seen in Fig. 3. In this case, a 50 ms cumulative power step is used, similar to the proposed approach explained below in this paper. Moreover, it is provided a topographical view of the cumulative power of the whole 3-second window

using the r^2 statistics. The θ and α frequency sub-bands show discriminative features across all channels, consistent over the time window. δ shows a time-dependent discrimination higher at the end of the 3-second window. The β and γ frequency sub-bands are less discriminative having less consistency. When analyzing their temporal evolution along the window, β is very focused on central the region and γ manifests a light focus in the occipital region.

The two scenarios are distinct in their distribution over space and time, as demonstrated by this cumulative analysis of all the subjects. Looking at the literature [52], [53] and from our analysis, we have elected to use the frequency sub-bands θ, α, and β, which are consensually used in this context [36], across the 30 channels, with the goal of having the discriminative information to account for cross-subject zero calibration. The β band is less discriminative, as is the γ one, than the two other bands, but given its widespread use in the context of attention analysis in the literature [52], [53], it was decided to include it.

B. Methodology Overview

From the previous subsection, it is defined as the input observation a 3-second sliding window of 30 channels. This translates into an input X of dimension 30×1500, being the sampling rate of 500 Hz. An overview of the proposed approach is illustrated in Fig. 4. The goal is to classify if an observed window of data is 'Alert' or 'Drowsy'. The data is processed according to the following steps: (i) the data X is band-pass filtered into 3 EEG frequency sub-bands (θ, α, and β); (ii) For every 3-second window observation, in each frequency sub-band, the window is divided into frames of 50 ms, followed by cumulative band power computation for each channel's frame. A metric of attention is also computed, resulting from the combination of the 3 band powers; (iii) The resulting band powers (3 band power vectors and the attention metric power vector, per channel) are used to compute an image encoding representation, which result in a multiple-channel image; (iv) The CNN handles this image and outputs the class of the observation.

The following subsections provide the details of the aforementioned blocks.

C. Image Encoding Representations

This subsection details the computational steps of the pre-processing group of blocks of Fig. 4. The goal is to generate an image-like representation of the data, at each observation, to be used by a CNN to classify drowsiness.

As referred in section II, studies on drowsiness detection have shown that the change of cognitive state is strongly correlated with changes of specific EEG frequency bands such as theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz), [29], [54]. Therefore, the EEG signals of the 3-second window observation X are initially band-pass filtered for each frequency sub-band with a fourth-order Butterworth filter. The result are three 30-channel filtered EEG signal windows X_b, being b either θ, α, or β.

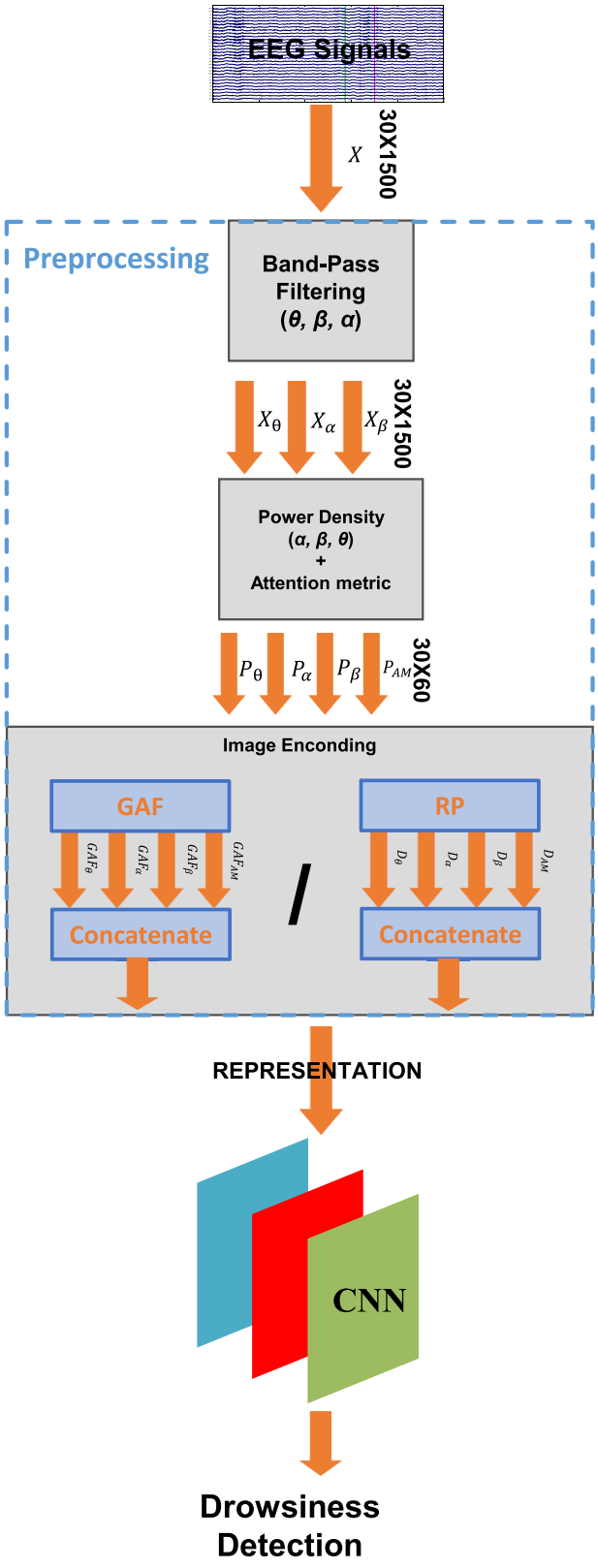

Fig. 4. Drowsiness detection methodology's flowchart using data representations as input for the CNN.

Each window X_b is divided into 50 ms intervals followed by a computation of the cumulative band power $P^c_{b,f}$ for each

of the 30 channels, per frame, according to

$$P^c_{b,f}(X^c_b) = \sum_{n=1}^{25} X^c_{b,f}(n)^2 \tag{2}$$

where c is one of the 30 channels, b is the frequency sub-band, f is the sequential frame of 50 ms, n is the sample within a frame, and $X^c_{b,f}$ is the 50 ms EEG signal of a given channel and frequency sub-band. Since the sampling rate is 500 Hz, 3 seconds are represented by 1500 samples, each 50 ms corresponds to 25 samples (n), resulting in $1500/25 = 60$ frames (f). The band power vector for each channel is computed using $P^c_b = [P^c_{b,1}, \ldots, P^c_{b,60}]$.

An attention metric (AM) is additionally computed which is correlated to fatigue, as described in the work of Cao *et al.* [55]. It consists of the power ratio $(\theta + \alpha)/\beta$, computed at each channel c per frame f, forming the vector $P^c_{AM} = [P^c_{AM,1}, \ldots, P^c_{AM,60}]$.

From the computation of P^c_b and P^c_{AM} results $4 \times 30 = 120$ vectors of 60 elements

$$P = \begin{bmatrix} P^c_\theta \\ P^c_\alpha \\ P^c_\beta \\ P^c_{AM} \end{bmatrix}, \quad c = 1, \ldots, 30 \tag{3}$$

P is a matrix resultant of the concatenation of these vectors that are encoded into an image as proposed below.

For the image encoding, two techniques are used in this work, the recurrence plot and the gramian angular fields. Each method explores different perspectives. Recurrence plots explore the recurrence over time of the phase space trajectory of a dynamical system, while gramian angular fields try to reveal temporal correlations of a time series.

1) Recurrence Plot: A recurrence plot is a visual tool to inspect the periodic nature of a trajectory through a phase space. It is a technique of nonlinear data analysis through a square matrix, in which the matrix elements correspond to those times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of times). It shows for each moment i in time, the times at which a phase space trajectory visits roughly the same area in the phase space as at time j [56]. The benefit of recurrence plots is that they can also be applied to rather short and even nonstationary data.

In our work, we explore three recurrence approaches, temporal recurrence for each separate channel, spatiotemporal recurrence between pairs of channels over time, and spatial recurrence.

For our temporal approach, a recurrence plot is a square matrix where each element is represented by

$$R^c_{b,i,j}(P^c_b) = \Theta\left(\varepsilon - \left\| P^c_{b,i} - P^c_{b,j} \right\|\right), \quad P^c_{b,i} \in \mathbb{R}, \quad i, j = 1, \ldots, N \tag{4}$$

where $N = 60$, which is the number of frames corresponding to samples within P^c_b, ε is a threshold distance and $\Theta(\cdot)$ is the Heaviside function. The vector P^c_b is a band power vector of 60 elements for one channel c, where b is either θ, α, β, or AM. From (4), the result is a binary image (black and white), being the binary difference determined by the user-defined threshold parameter ε_i.

In this work, we use a variation of the recurrence plot, also called global recurrence plot or unthreshold recurrence plot [57], which is computed as follows

$$D^c_{b,i,j}(P^c_b) = \left\| P^c_{b,i} - P^c_{b,j} \right\|, \quad i, j = 1, \ldots, 60 \tag{5}$$

With this variation, there is no need to choose a threshold, which in this case could disguise important information. As such, the whole spectrum of the difference between states is used. At the end of the recurrence plots computation, 120 square matrices of 60×60 elements are obtained, corresponding to each band power and attention metric for each channel. An example of such a recurrence plot is provided in Fig. 5. To produce a usable representation for a CNN, the final step is the combination of all 120 matrices in a $120 \times 60 \times 60$ image-like structure.

Looking at the spatiotemporal approach, we aim at exploring recurrences between channels over time, using a pairwise channel comparison. Recurrence plots are computed at each frame of 50 ms (60 frames in total), instead of being computed for each individual channel over time, using

$$D^f_{b,i,j}(P^c_b) = \left\| P^i_{b,f} - P^j_{b,f} \right\|, \quad i, j = 1, \ldots, C \tag{6}$$

where f is the frame of the band power vector P^c_b, i and J are channel indexes, and $C = 30$ which is the total number of channels. After this step, we have 240 square matrices of 30×30 elements, corresponding to each band power and attention metric for each of the 60 frames. As before, the final step is the combination of all 240 matrices in a $240 \times 30 \times 30$ image-like structure.

The spatial approach uses the spatiotemporal representation of 240 plots of dimension 30×30 to compute their density. A single density matrix of dimensions 30×30 for each band power and the attention metric is computed, using the 60 respective frames as follows

$$Den_{b,i,j} = \frac{\sum_{f=1}^{60} D^f_{b,i,j}}{60}, \quad i, j = 1, \ldots, C \tag{7}$$

resulting in a final representation of $4 \times 30 \times 30$. An illustrative example of the three representations is presented in Fig. 5.

2) Gramian Angular Fields: This encoding is based on a transformation of time series into polar coordinates followed by the construction of a quasi-Gram matrix. This technique allows to preserve the temporal dependencies and correlations within the time series [58]. Given the band power vector of each channel $P^c_b = [P^c_{b,1}, P^c_{b,2}, \ldots, P^c_{b,60}]$ of 60 real-valued observations, first P^c_b is rescaled into the interval $[-1, 1]$ using a min-max scaler. This allows to represent the rescaled vector $\tilde{P}^c_b$ in polar coordinates by encoding the value as the angular cosine and the time line as the radius using

$$\begin{cases} \phi^c_{b,i} = \arccos\left(\tilde{P}^c_{b,i}\right), -1 \le \tilde{P}^c_{b,i} \le 1, & \tilde{P}^c_{b,i} \in \tilde{P}^c_b \\ r = \dfrac{i}{N}, & i \in \mathbb{N} \end{cases} \tag{8}$$

where $i = 1, \ldots, 60$ is the frame and $N = 60$ is the constant factor to regularize the span of the polar coordinate system.

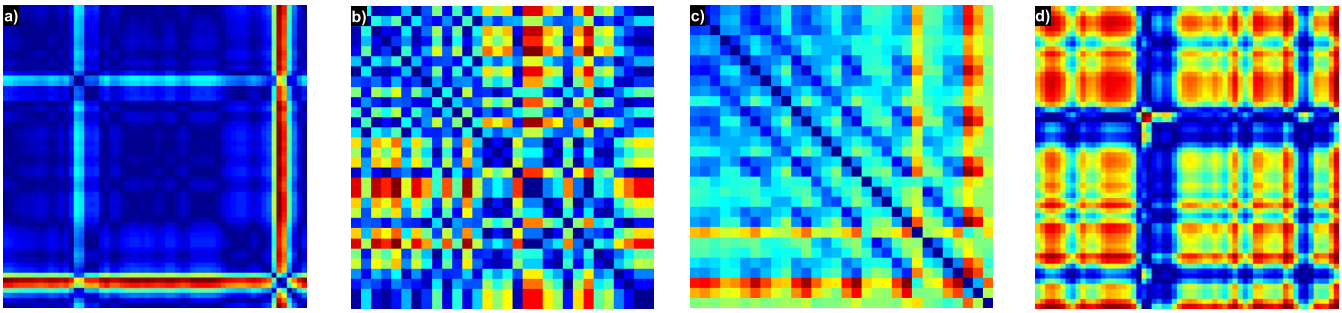

Fig. 5. Illustrative examples of image representations of a 3-second 'Drowsy' observation shown in a visible color spectrum scale. a) Temporal RP of 60 × 60 elements, as time progresses along the diagonal we observe that the time disparities in terms of distance are larger between the last time steps against the rest (lower right corner). b) Spatiotemporal RP of 30 × 30 elements, where it is observed the large disparities between specific channels that appear in a brighter color, at a given time. c) A result of the spatial RP, density is showing larger disparities for channels 20 and up. d) GAF showing the temporal correlations along the diagonal, with brighter regions presenting higher correlation.

As i increases, corresponding values warp among different angular points. The rescaled vector results in cosine values in the interval $[-1, 1]$ which fall into the angular bounds $[0, \pi]$. This encoding has two properties: (i) it is bijective as $\cos(\phi)$ is monotonic when $\phi \in [0, \pi]$. For a given band power vector, the proposed map produces one and only one result in the polar coordinate system with a unique inverse map; (ii) Polar coordinates preserve absolute temporal relations.

After rescaling the band power vectors into polar coordinates, we identify the temporal correlation within different frame intervals by exploiting the angular perspective by considering the trigonometric sum between each frame. Each element of the Gramian Angular Field (GAF) is defined as follows

$$\begin{aligned} GAF^{c}_{b,i,j} &= \left[\cos\left(\phi^{c}_{b,i} + \phi^{c}_{b,j}\right)\right] \\ &= \tilde{P'}^{c}_{b,i} \cdot \tilde{P}^{c}_{b,j} - \sqrt{I - \tilde{P}^{2^{c}}_{b,i}} \cdot \sqrt{I - \tilde{P}^{2^{c}}_{b,j}}, \\ & i, j = 1, \ldots, 60 \end{aligned} \tag{9}$$

where c is each EEG channel, b either θ, α, β, or AM, I is the unit row vector. For the above operation, we take a band power vector at each frame as a 1-D metric space, and apply a redefined inner product $< x, y >= x \cdot y - \sqrt{1 - x^2} \cdot \sqrt{1 - y^2}$. In this manner, gramian angular fields are actually quasi-Gram matrices, because the redefined function used does not satisfy the property of linearity in the proper inner-product space. The size of the Gram matrix is 60 × 60 since the length of the band power vector is 60. The result is, alike the above-mentioned recurrence plots combination, an image-like structure of dimension 120 × 60 × 60. Figure 5 shows a visualization of the GAF for each band power and attention metric for the first channel of a single 3-second window observation.

Considering the image encoding stage, it can be summarized, for each observation, by Algorithm 1.

Algorithm 1 Image Encoding

input : P
output: D, Den, GAF
$D = \emptyset$;
$Den = \emptyset$;
$GAF = \emptyset$;
if *Temporal RP* **then**
  for $c \leftarrow 1$ **to** 30 **do**
    $D^{c}_{b}(P^{c}_{b})$;
    D =concatenate(D, D^{c}_{b});
  end
if *Spatiotemporal RP* **then**
  for $f \leftarrow 1$ **to** 60 **do**
    $D^{f}_{b}(P_{b,f})$;
    D =concatenate(D, D^{f}_{b});
  end
if *Spatial RP* **then**
  for $f \leftarrow 1$ **to** 60 **do**
    $Den_{b}(D^{f}_{b})$;
  end
if *GAF* **then**
  for $c \leftarrow 1$ **to** 30 **do**
    $GAF^{c}_{b}(P^{c}_{b})$;
    D =concatenate(GAF, GAF^{c}_{b});
  end

D. Drowsiness Detection

The presented CNN architecture in Fig. 6 was designed to classify the data representation for drowsiness detection. The network is composed by one convolutional layer and three fully-connected layers. This architecture is used as an image classification architecture processing our representations as an image classification problem, following typical designs found in the literature [59]. The convolutional layer is followed by Batch Normalization and a Rectified Linear Unit (ReLU). The network is trained with a cross-entropy loss function and uses the Adam optimizer [60] with a learning rate of 0.0001. Between the convolutional layer and fully-connected layers a dropout of 25% is applied (except for the last fully-connected transition with a dropout of 50%). A kernel of 5 × 5 is used in the convolutional layer with 1 stride and 2 padding. Data are normalized after the convolutional layer and max pooling of 2 × 2 with 2 stride is applied. The input of the CNN is the

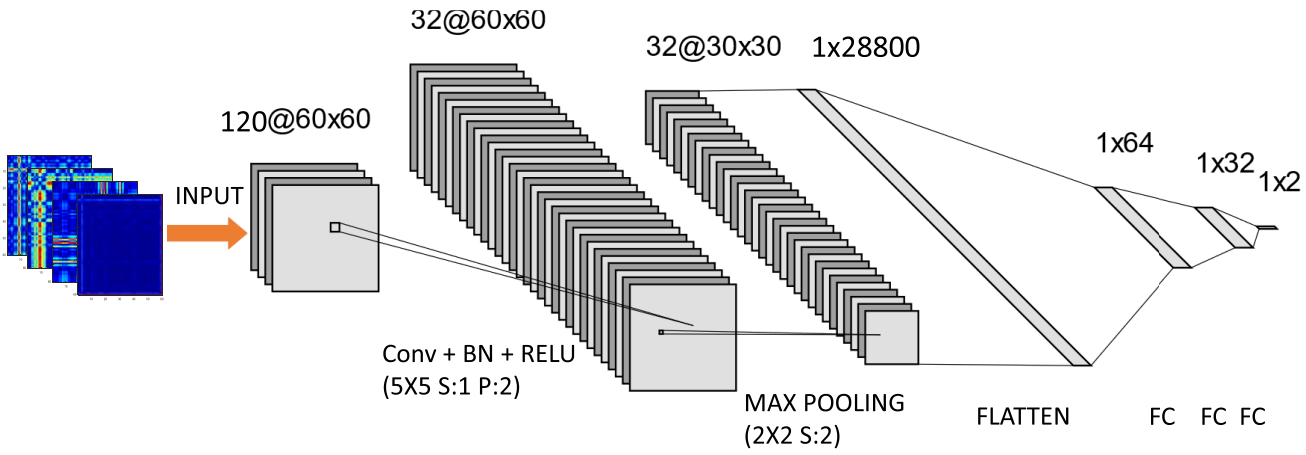

Fig. 6. Convolutional neural network architecture.

proposed representation and the output assumes one of two classes 'Alert' or 'Drowsy'.

IV. Results and Discussion

Considering the objective of cross-subject drowsiness detection, a leave-one-subject-out cross-validation strategy was performed to validate the proposed methodologies. Two subjects had to be discarded from the group, due to the absence of 'Drowsy' observations, resultant from the data labeling described in section III-A.2. The first test focused on understanding the contribution of the use of the attention metric resultant from the combination of the band powers. We performed a classification test using the band power vectors as inputs to the CNN without being encoded by any type of image representation. The contribution of the attention metric from the leave-one-subject-out cross-validation using just band power vectors is undetectable. Accuracy values for data with and without attention metric are 69.64% and 69.99%, respectively. Taking into account this result, we performed the following tests without the use of the attention metric.

A validation was carried out at the subject level, to understand each subject's contribution to the classification performance, so that by selecting groups of subjects with the highest individual contributions to train the network with, a better cross-subject generalizable model could be created. For this purpose, a single training subject against the others cross-validation strategy was performed. We obtained for each subject a classification accuracy from the test with the remaining 24 subjects, and then sorted a list of all subjects by increasing accuracy. For the purpose of cross-subject model creation, groups of subjects from the above list were created for the training stage [61]. 3 groups were created: all subjects (control group); subjects with individual contribution above 50% accuracy; subjects with individual contribution above 60% accuracy. The results of a leave-one-subject-out cross-validation in each scenario, compared to each other, did not offer statistical differences between them (balanced accuracy values of 69.99%, 69.48%, and 70.08%, respectively). From these results, the remainder of the experiments were performed considering all subjects. Considering the data-driven nature of CNN training, it is expected that for a performance improvement that could be noticeable it would be required more data from a higher number of training subjects.

TABLE I
Results With Leave-One-Subject-Out Cross Validation for the Modalities and Montages Addressed

Method	Accuracy	Precision	Recall
	All channels		
Power	69.99 ± 11.13	79.67 ± 17.75	81.52 ± 20.38
GAF	74.53 ± 9.10	85.40 ± 20.89	89.36 ± 9.95
RP temporal	75.87 ± 9.31	79.04 ± 21.38	76.81 ± 19.69
RP spatiotemporal	75.24 ± 11.27	83.17 ± 19.76	88.82 ± 12.54
RP spatial	68.36 ± 13.22	74.64 ± 19.96	88.72 ± 14.58
	Sleep staging setup		
Power	60.07 ± 7.249	77.12 ± 22.96	54.03 ± 30.37
GAF	67.74 ± 9.27	74.58 ± 21.37	81.51 ± 7.74
RP	70.56 ± 10.57	82.10 ± 18.04	76.95 ± 19.20
	Non-hair-bearing setup		
Power	59.20 ± 8.25	72.74 ± 20.84	79.90 ± 15.41
GAF	64.82 ± 9.13	76.45 ± 20.71	80.29 ± 11.45
RP	64.11 ± 8.94	77.60 ± 21.00	60.77 ± 22.89

To understand the specific contribution of the proposed image representations, we performed a test using the non-encoded band power vectors against the representations. The comparison results are shown in Fig. 7 presenting the balanced accuracy for each subject, and in Table I showing the comparison of the average of those balanced accuracies, precision, and recall. The balanced accuracy is used, due to the imbalanced nature of the dataset towards the 'Alert' observations. The obtained balanced accuracy without any of the representations is 69.99%, while for the gramian angular fields and recurrence plots representations (temporal, spatiotemporal and spatial) we achieved 74.53%, 75.87%, 75.24%, and 68.36%, respectively. These results demonstrate an added contribution for the use of the proposed representations, except when using the spatial approach. This result shows that discriminative information is found mainly in the time domain, corroborating our previous analysis and that of the literature [62], showing an increase in theta and alpha rhythms during fatigue over the entire cortex.

As a comparison, we also present results from works using the same dataset with a cross-subject approach [34], [35], [37],

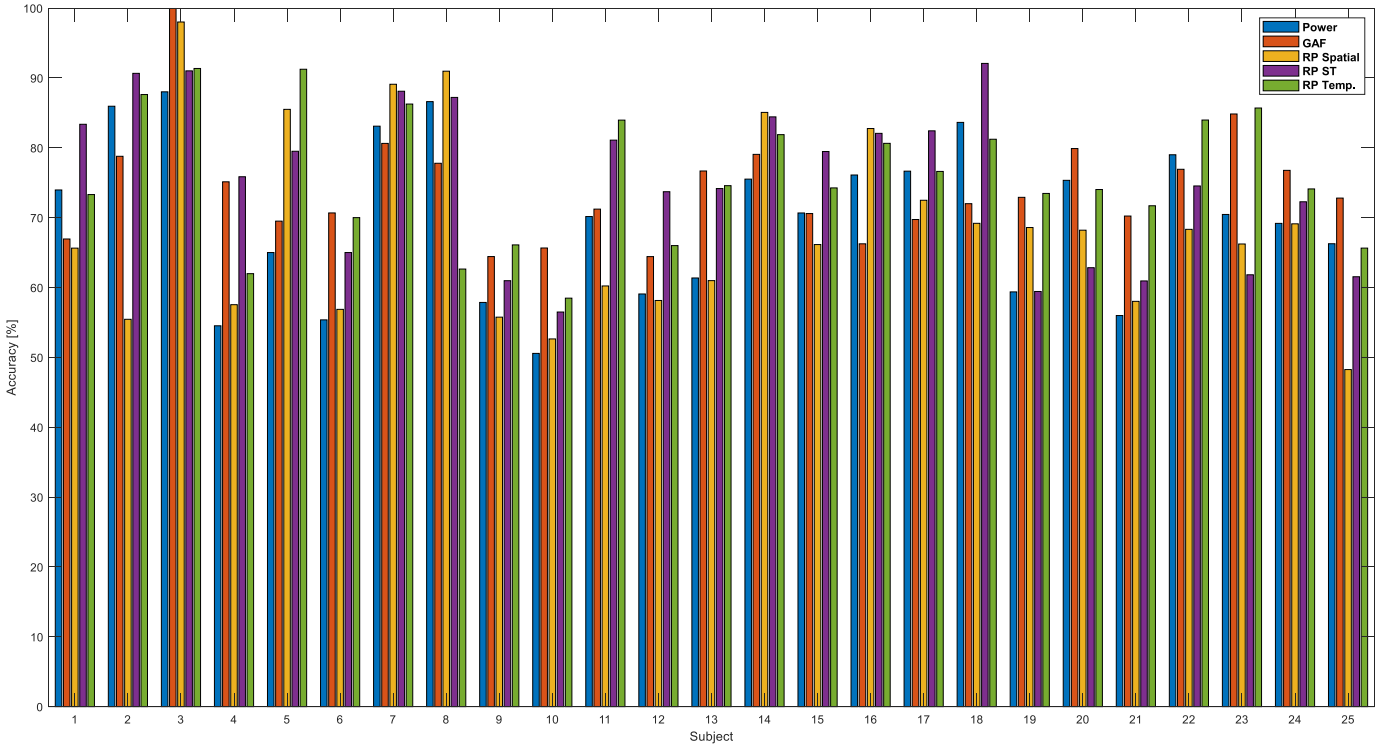

Fig. 7. Comparison results for leave-one-subject-out cross-validation for each subject for GAF, RP, and band power approaches. Each subject has one accuracy value per approach.

TABLE II
COMPARISON OF THE ACCURACY BETWEEN METHODS

Method	Mean (%)	Std. Dev. (%)	# Test Subjects
RP temporal	**75.87**	**9.31**	**25**
GAF	**74.53**	**9.10**	**25**
Zou *et al.* [35]	73.99	-	14
Liu *et al.* [37]	72.70	9.42	11
Li *et al.* [34]	70.60	-	11
EEGNet-8,2 [63]	63.98	11.10	11

whose authors used a reduced group of subjects from the dataset, as shown in Table II. When put alongside our results, shown in Table II, a superior performance was achieved against those approaches, particularly when considering the use of a more robust metric, balanced accuracy, and the use of more test subjects, 25.

The previous tests considered the use of 30 channels, which is a whole scalp montage which, in practice, is cumbersome. Recently, some works have been proposing the use of more practical montages with a reduced number of electrodes. Specifically, we tested a non-hair-bearing (NHB) montage which considers electrodes on scalp areas without hair focusing on practical use of electrode montages, that are quicker and simpler to set up, and they also avoid the interference of electrode-skin contact caused by the hair [36], [64]–[66]. Since the electrodes in the dataset are referenced to the two mastoids, the remaining NHB electrodes are Fp1, Fp2, F7, and F8. Beyond this montage, we also applied our approach to a montage used in sleep staging, using the electrodes F3, C3, O1, F4, C4, and O2 [51]. We choose this montage to try and evaluate the approximation of drowsiness to sleep states, trying to find discriminative potential with the same electrodes. We performed a leave-one-subject-out cross-validation, like the previous test, where results are also presented in Table I. For this test, the recurrence plot approach used was the temporal since it previously yielded the best results.

Considering the results obtained, starting with the attention metric test, we attribute the absence of a performance improvement to the linear nature of the attention metric itself. Since we use directly the 3 sub-frequency bands, their linear combination does not practically add value. The obtained results using the proposed representations when compared to purely power vectors demonstrate that spatial and temporal correlations and dynamics are of important discriminative nature, in the power domain. Both, RP and GAF, take into account temporal recurrences and correlations into a single two-dimensional encoding, which is well suited for a CNN. The marginal difference between spatial and temporal RP approaches can be partly justified by the convolutional nature of the CNN, which explores the interaction between image channels also aiding in a spatial or temporal analysis, complementing each respective approach. The spatial approach being a summation of frames ends in poorer performance justified by the loss of information in the global sum. We also note that when looking at each subject's performance, there is a considerable variability. This is a result of the challenging nature of EEG signals that vary significantly even for the same subject in different sessions. Even though we achieved higher results than the literature, for a zero calibration system, performance still

needs to improve. As an alternative approach, we estimate that since the time-domain provides valuable information, sequential analysis using recurrent neural networks could be an appropriate approach, due to their handling of temporal features successfully.

Considering our neurophysiological analysis of Section III-A.3, the results here achieved with a complete electrode montage and the two electrode groups, show that the difference in the power domain between 'Alert' and 'Drowsy' states are not localized, but on the contrary are widely spread all over the regions of the brain. Even though the beta band frequency had a more noticeable difference in the frontal regions of the brain, with a NHB montage the classifier did not achieve a powerful enough performance as with the use of all channels.

The chosen CNN architecture used follows a common design found in the image classification domain with application on EEG [59], being able to achieve superior results compared to works in the same context. We consider this to be a validation of the contribution of the proposed data representations which efficiently encode the features associated with the mental state.

V. Conclusion

This paper proposes the application of image encoding representations that exploit time recurrences and correlations of EEG signals, to be used with deep learning strategies, with the objective of drowsiness detection in driving tasks. We validated our methodologies with a challenging dataset achieving superior performance against the works found in the literature. Overall, results still demonstrate a need to further research this domain to improve drowsiness detection so that we can have a reliable safety system for advanced driver assistance systems. A limitation of this work can be identified pertaining to the practical use of the proposed approach. For a real-world scenario the use of 32 channels is not a practical solution, this is a physical constraint that makes these systems hard to implement as a commercial product. As well, the lack of online validation is also of notice, being a necessary future step.

In the future, we plan to focus on the design of the CNN's architecture to further exploit the representation, extracting deeper features beyond the spatiotemporal domain. Considering the demonstrated potential of time recurrences and correlations found in our work, we plan to explore sequential analysis using techniques found in the domain of natural language processing such as recurrent neural networks with attention mechanisms, as well as the recently proposed transformer networks [67], using an approach that could replace the use of the representations.

References

[1] G. Zhang, K. K. W. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Anal. Prevention*, vol. 87, pp. 34–42, Feb. 2016.

[2] J. Li, H. Li, H. Wang, W. Umer, H. Fu, and X. Xing, "Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology," *Autom. Construct.*, vol. 105, Sep. 2019, Art. no. 102835.

[3] C. Sugden, T. Athanasiou, and A. Darzi, "What are the effects of sleep deprivation and fatigue in surgical practice?" *Seminars Thoracic Cardiovascular Surg.*, vol. 24, no. 3, pp. 166–175, 2012.

[4] M. Gastaldi, R. Rossi, and G. Gecchele, "Effects of driver task-related fatigue on driving performance," *Procedia-Social Behav. Sci.*, vol. 111, pp. 955–964, Feb. 2014.

[5] M. Basner and J. Rubinstein, "Fitness for duty: A 3 minute version of the Psychomotor Vigilance test predicts fatigue related declines in luggage screening performance," *J. occupational Environ. Med./Amer. College Occupational Environ. Med.*, vol. 53, no. 10, p. 1146, 2011.

[6] A. J. Beurskens, U. Bültmann, I. Kant, J. H. Vercoulen, G. Bleijenberg, and G. M. Swaen, "Fatigue among working people: Validity of a questionnaire measure," *Occupational Environ. Med.*, vol. 57, no. 5, pp. 353–357, 2000.

[7] K. A. Lee, G. Hicks, and G. Nino-Murcia, "Validity and reliability of a scale to assess fatigue," *Psychiatry Res.*, vol. 36, no. 3, pp. 291–298, Mar. 1991.

[8] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci.*, vol. 52, nos. 1–2, pp. 29–37, Jan. 1990.

[9] R. P. Balandong, R. F. Ahmad, M. N. M. Saad, and A. S. Malik, "A review on EEG-based automatic sleepiness detection systems for driver," *IEEE Access*, vol. 6, pp. 22908–22919, 2018.

[10] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

[11] Y. Tran, N. Wijesuriya, M. Tarvainen, P. Karjalainen, and A. Craig, "The relationship between spectral changes in heart rate variability and fatigue," *J. Psychophysiol.*, vol. 23, no. 3, pp. 143–151, Jan. 2009.

[12] S. Hu and G. Zheng, "Driver drowsiness detection with eyelid related parameters by support vector machine," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7651–7658, May 2009.

[13] Q. Zhao, L. Zhang, and A. Cichocki, "EEG-based asynchronous BCI control of a car in 3D virtual reality environments," *Chin. Sci. Bull.*, vol. 54, no. 1, pp. 78–87, Jan. 2009.

[14] Z. Khaliliardali, R. Chavarriaga, L. A. Gheorghe, and J. R. D. Millan, "Detection of anticipatory brain potentials during car driving," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 3829–3832.

[15] H. Zhang, R. Chavarriaga, Z. Khaliliardali, L. Gheorghe, I. Iturrate, and J. D. R. Millán, "EEG-based decoding of error-related brain activity in a real-world driving task," *J. Neural Eng.*, vol. 12, no. 6, Dec. 2015, Art. no. 066028.

[16] A. Cruz, G. Pires, and U. J. Nunes, "Double ErrP detection for automatic error correction in an ERP-based BCI speller," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 26–36, Jan. 2018.

[17] A. Cruz, G. Pires, A. C. Lopes, and U. J. Nunes, "Detection of stressful situations using GSR while driving a BCI-controlled wheelchair," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 1651–1656.

[18] Z. Halim and M. Rehan, "On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning," *Inf. Fusion*, vol. 53, pp. 66–79, Jan. 2020.

[19] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[20] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Hoboken, NJ, USA: Wiley, 2013.

[21] R. Chai *et al.*, "Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 715–724, May 2017.

[22] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "EEG classification of driver mental states by deep learning," *Cognit. Neurodynamics*, vol. 12, no. 6, pp. 597–606, Dec. 2018.

[23] P. P. San, S. H. Ling, R. Chai, Y. Tran, A. Craig, and H. Nguyen, "EEG-based driver fatigue detection using hybrid deep generic model," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 800–803.

[24] C.-S. Huang, N. R. Pal, C.-H. Chuang, and C.-T. Lin, "Identifying changes in EEG information transfer during drowsy driving by transfer entropy," *Frontiers Hum. Neurosci.*, vol. 9, p. 570, Oct. 2015.

[25] S.-W. Chuang, L.-W. Ko, Y.-P. Lin, R.-S. Huang, T.-P. Jung, and C.-T. Lin, "Co-modulatory spectral changes in independent brain processes are correlated with task performance," *NeuroImage*, vol. 62, no. 3, pp. 1469–1477, Sep. 2012.

[26] Y.-T. Liu, Y.-Y. Lin, S.-L. Wu, C.-H. Chuang, and C.-T. Lin, "Brain dynamics in predicting driving fatigue using a recurrent self-evolving fuzzy neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 347–360, Feb. 2016.

[27] L.-W. Ko, R. K. Chikara, Y.-C. Lee, and W.-C. Lin, "Exploration of user's mental state changes during performing brain–computer interface," *Sensors*, vol. 20, no. 11, p. 3169, Jun. 2020.

[28] R. K. Chikara and L.-W. Ko, "Neural activities classification of human inhibitory control using hierarchical model," *Sensors*, vol. 19, no. 17, p. 3791, Sep. 2019.

[29] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, Mar. 2009.

[30] F. Gharagozlou *et al.*, "Detecting driver mental fatigue based on EEG alpha power changes during simulated driving," *Iranian J. Public Health*, vol. 44, no. 12, p. 1693, 2015.

[31] J. Hu, "Comparison of different features and classifiers for driver fatigue detection based on a single EEG channel," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–9, Jul. 2017.

[32] J. Liu, C. Zhang, and C. Zheng, "EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters," *Biomed. Signal Process. Control*, vol. 5, no. 2, pp. 124–130, Apr. 2010.

[33] J. Hu and J. Min, "Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model," *Cognit. Neurodynamics*, vol. 12, no. 4, pp. 431–440, Aug. 2018.

[34] R. Li, Z. Lan, J. Cui, O. Sourina, and L. Wang, "EEG-based recognition of driver state related to situation awareness using graph convolutional networks," in *Proc. Int. Conf. Cyberworlds (CW)*, Sep. 2020, pp. 180–187.

[35] B. Zou, M. Shen, X. Li, Y. Zheng, and L. Zhang, "EEG-based driving fatigue detection during operating the steering wheel data section," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 248–251.

[36] C.-S. Wei, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "Toward drowsiness detection using Non-hair-Bearing EEG-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 400–406, Feb. 2018.

[37] Y. Liu, Z. Lan, J. Cui, O. Sourina, and W. Müller-Wittig, "Inter-subject transfer learning for EEG-based mental fatigue recognition," *Adv. Eng. Informat.*, vol. 46, Oct. 2020, Art. no. 101157.

[38] Y. Liu, Z. Lan, H. H. G. Khoo, K. H. H. Li, O. Sourina, and W. Mueller-Wittig, "EEG-based evaluation of mental fatigue using machine learning algorithms," in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2018, pp. 276–279.

[39] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Proc. Workshops 29th AAAI Conf. Artif. Intell.*, vol. 1, 2015, pp. 1–8.

[40] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," *Proc. SPIE*, vol. 10696, Apr. 2018, Art. no. 106960Y.

[41] Z. Zhao, Y. Zhang, Z. Comert, and Y. Deng, "Computer-aided diagnosis system of fetal hypoxia incorporating recurrence plot with convolutional neural network," *Frontiers Physiol.*, vol. 10, p. 255, Mar. 2019.

[42] X. Gao, X. Yan, P. Gao, X. Gao, and S. Zhang, "Automatic detection of epileptic seizure based on approximate entropy, recurrence quantification analysis and convolutional neural networks," *Artif. Intell. Med.*, vol. 102, Jan. 2020, Art. no. 101711.

[43] P. Nunez *et al.*, "Characterization of EEG resting-state activity in Alzheimer's disease by means of recurrence plot analyses," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 5786–5789.

[44] P. G. Rodrigues, C. A. S. Filho, R. Attux, G. Castellano, and D. C. Soriano, "Space-time recurrences for functional connectivity evaluation and feature extraction in motor imagery brain-computer interfaces," *Med. Biol. Eng. Comput.*, vol. 57, no. 8, pp. 1709–1725, Aug. 2019.

[45] K. P. Thanaraj, B. Parvathavarthini, U. J. Tanik, V. Rajinikanth, S. Kadry, and K. Kamalanand, "Implementation of deep neural networks to classify EEG signals using Gramian angular summation field for epilepsy diagnosis," 2020, *arXiv:2003.04534*. [Online]. Available: http://arxiv.org/abs/2003.04534

[46] A. Shankar, H. K. Khaing, S. Dandapat, and S. Barma, "Epileptic seizure classification based on Gramian angular field transformation and deep learning," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Oct. 2020, pp. 147–151.

[47] C. L. Webber and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *J. Appl. Physiol.*, vol. 76, no. 2, pp. 965–973, Feb. 1994.

[48] 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.

[49] R. Langner, M. B. Steinborn, A. Chatterjee, W. Sturm, and K. Willmes, "Mental fatigue and temporal preparation in simple reaction-time performance," *Acta Psychologica*, vol. 133, no. 1, pp. 64–72, Jan. 2010.

[50] E. J. Cheng, K.-Y. Young, and C.-T. Lin, "Temporal EEG imaging for drowsy driving prediction," *Appl. Sci.*, vol. 9, no. 23, p. 5078, Nov. 2019.

[51] T. Sousa, A. Cruz, S. Khalighi, G. Pires, and U. Nunes, "A two-step automatic sleep stage classification method with dubious range detection," *Comput. Biol. Med.*, vol. 59, pp. 42–53, Apr. 2015.

[52] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehavioral Rev.*, vol. 44, pp. 58–75, Jul. 2014.

[53] R. Chavarriaga *et al.*, "Decoding neural correlates of cognitive states to enhance driving experience," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 4, pp. 288–297, Aug. 2018.

[54] P. Bashivan, G. M. Bidelman, and M. Yeasin, "Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity," *Eur. J. Neurosci.*, vol. 40, no. 12, pp. 3774–3784, Dec. 2014.

[55] T. Cao, F. Wan, C. M. Wong, J. N. D. Cruz, and Y. Hu, "Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces," *Biomed. Eng. Online*, vol. 13, no. 28, pp. 1–13, 2014.

[56] J. Eckmann *et al.*, "Recurrence plots of dynamical systems," *World Sci. Ser. Nonlinear Sci. Ser. A*, vol. 16, pp. 441–446, Sep. 1995.

[57] J. S. Iwanski and E. Bradley, "Recurrence plots of experimental data: To embed or not to embed?" *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 8, no. 4, pp. 861–871, Dec. 1998.

[58] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," 2015, *arXiv:1506.00327*. [Online]. Available: http://arxiv.org/abs/1506.00327

[59] 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.

[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[61] T. Duan, M. Chauhan, M. A. Shaikh, J. Chu, and S. Srihari, "Ultra efficient transfer learning with meta update for cross subject EEG classification," 2020, *arXiv:2003.06113*. [Online]. Available: http://arxiv.org/abs/2003.06113

[62] A. Craig, Y. Tran, N. Wijesuriya, and H. Nguyen, "Regional brain wave activity changes associated with fatigue," *Psychophysiology*, vol. 49, no. 4, pp. 574–582, Apr. 2012.

[63] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[64] Y.-T. Wang, M. Nakanishi, Y. Wang, C.-S. Wei, C.-K. Cheng, and T.-P. Jung, "An online brain-computer interface based on SSVEPs measured from non-hair-bearing areas," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 1, pp. 14–21, Jan. 2017.

[65] J. J. S. Norton *et al.*, "Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 13, pp. 3920–3925, Mar. 2015.

[66] H.-T. Hsu *et al.*, "Evaluate the feasibility of using frontal SSVEP to implement an SSVEP-based BCI in young, elderly and ALS groups," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 5, pp. 603–615, May 2016.

[67] A. Vaswani *et al.*, "Attention is all you need," 2017, *arXiv:1706.03762*. [Online]. Available: http://arxiv.org/abs/1706.03762