


Article

Modeling Cognitive Load as a Self-Supervised Brain Rate with Electroencephalography and Deep Learning

Luca Longo ^{1,2,3} 

- ¹ Artificial Intelligence and Cognitive Load Research Lab, Technological University Dublin, Grangegorman Lower, D07 H6K8 Dublin, Ireland; luca.longo@tudublin.ie
- ² Applied Intelligence Research Center, Technological University Dublin, Grangegorman Lower, D07 H6K8 Dublin, Ireland
- ³ School of Computer Science, Technological University Dublin, Grangegorman Lower, D07 H6K8 Dublin, Ireland

Abstract: The principal reason for measuring mental workload is to quantify the cognitive cost of performing tasks to predict human performance. Unfortunately, a method for assessing mental workload that has general applicability does not exist yet. This is due to the abundance of intuitions and several operational definitions from various fields that disagree about the sources or workload, its attributes, the mechanisms to aggregate these into a general model and their impact on human performance. This research built upon these issues and presents a novel method for mental workload modelling from EEG data employing deep learning. This method is self-supervised, employing a continuous brain rate, an index of cognitive activation, and does not require human declarative knowledge. The aim is to induce models automatically from data, supporting replicability, generalisability and applicability across fields and contexts. This specific method is a convolutional recurrent neural network trainable with spatially preserving spectral topographic head-maps from EEG data, aimed at fitting a novel brain rate variable. Findings demonstrate the capacity of the convolutional layers to learn meaningful high-level representations from EEG data since within-subject models had, on average, a test Mean Absolute Percentage Error of around 11%. The addition of a Long-Short Term Memory layer for handling sequences of high-level representations was not significant, although it did improve their accuracy. These findings point to the existence of quasi-stable blocks of automatically learnt high-level representations of cognitive activation because they can be induced through convolution and seem not to be dependent on each other over time, intuitively matching the non-stationary nature of brain responses. Additionally, across-subject models, induced with data from an increasing number of participants, thus trained with data containing more variability, obtained a similar accuracy to the within-subject models. This highlights the potential generalisability of the induced high-level representations across people, suggesting the existence of subject-independent cognitive activation patterns. This research contributes to the body of knowledge by providing scholars with a novel computational method for mental workload modelling that aims to be generally applicable and does not rely on ad hoc human crafted models.

Keywords: cognitive load; deep learning; self-supervision; brain rate; convolutional neural network; recurrent neural network; mental workload; EEG bands; electroencephalography; spectral topology-preserving head-maps



Citation: Longo, L. Modeling Cognitive Load as a Self-Supervised Brain Rate with Electroencephalography and Deep Learning. *Brain Sci.* **2022**, *12*, 1416. <https://doi.org/10.3390/brainsci12101416>

Academic Editor: Gianluca Borghini

Received: 17 September 2022

Accepted: 14 October 2022

Published: 21 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The explosion of home-working and online interactions, the pervasive uses of technologies in daily activities and many working environments impose ever more mental workload upon operators and less physical load. The literature on the construct of mental workload (MWL) or, often interchangeably referred to as cognitive load (CL), has been vast and in constant evolution for the last half-century. Note that cognitive load and

mental workload might differ in little aspects, according to authors working in different fields, such as psychology, neuroscience or education. However, to my knowledge, no clear evidence has been found to address these differences formally. Thus, they are interchangeably used in the remainder of this article. Many definitions of MWL exist in the literature, as reported in [1]. However, a newly proposed operational and inclusive definition is: “Mental workload (MWL) represents the degree of activation of a finite pool of resources, limited in capacity, while cognitively processing a primary task over time, mediated by external stochastic environmental and situational factors, as well as affected by definite internal characteristics of a human operator, for coping with static task demands, by devoted effort and attention” [1]. The principal reason for measuring workload is to quantify the mental cost of performing tasks to predict human performance [2]. In turn, prediction of performance can be used for designing interfaces, interactive technologies [3], and information-processing activities [4] optimally aligned to the well-known human mental limited capacities [1]. Despite 50 years of effort, research on MWL has not been able to make major advances yet [5–7] failing at creating a clear, robust, transparent and explainable model [8–11]. Guess intuitions and several operational definitions from various fields have proliferated [12,13]. Still, these disagree about the MWL sources, their attributes, the mechanisms to aggregate these together and their impact on human performance [8]. Identifying these sources, attributes, and mechanisms and how they impinge on human performance are all open fundamental research problems. For instance, some researchers have considered task-specific attributes [14] while others chose a combination of task and user-specific attributes [5]. Primary researchers have employed self-reporting measurements [13] or a combination of psychophysiological techniques [15]. However, MWL is also influenced by the environment in which a human performs a task [16].

Currently, the literature on mental workload includes a plethora of hand-crafted knowledge-driven models grounded in different theories, employing different attributes and different strategies for aggregating these into indexes of workload, limiting their comparison [1,10,17,18]. This makes cognitive load a knowledge-dependent construct. This is also supported by the fact that cognitive load has been mainly investigated in the fields of ergonomics and psychology [8,13] with several applications in the aviation [5], automobile [15] and manufacturing industries [19]. In these fields, investigations are mainly conducted in labs and highly controlled settings, making cognitive load a field-dependent construct. Past research has had a tendency to focus on complex safety-critical systems [9] with many applications in the transportation [20,21], nuclear and manufacturing industries [5,15], making mental workload an application-driven construct. However, researchers have claimed the need for models of cognitive load in other ecological settings with real-world activities [6–8,22]. The vast majority of existing knowledge-dependent, field-dependent, and application-driven models aggregate attributes, believed to influence workload, in a linear fashion [5], or assume stationarity within a task, neglecting temporal dynamics [9], making cognitive load a static construct. Additionally, these models are largely built by fitting or correlating to some ad hoc indicator of human performance. This is either explicitly achieved by applying self-reporting techniques and correlating to subjective responses from experimental participants, or based on fitting human responses grouped by tasks of varying demands, often ad hoc and subjectively defined. This largely complicates research efforts attempted at modelling mental workload and increasing the generalisability of models because they are highly constrained on those subjective design choices from modellers that highly differ across experiments, disciplines and contexts. The aforementioned state of the art in cognitive load modelling has led to many definitions of workload [1,17,23,24] and the formation of ad hoc, knowledge-dependent, field-dependent, application-driven and static models with little chance of reconciliation [9]. Because of this, despite 50 years and more of research, the construct of workload is still ill-defined [1,7–9,17,18].

The goal of this research is to tackle the above issues and design a model of cognitive load that has wider applicability, facilitating comparison across studies, that is less

constrained to the context of application, that is not static, meaning it consider mental workload over time, that does not require any explicit ground truth, and that minimizes experimental design-choices of researchers. To achieve this goal, this research proposes to apply modern deep learning methods to avoid incrementally extending current knowledge-driven approaches and supporting automatic learning of salient features for cognitive load and their non-linear inner-relationship from data. In other words, this means that rather than producing ad hoc models of mental workload, by using the declarative knowledge and experience of a human modeller, it is possible to 'induce' such models from data, and 'learn' salient features and patterns from it automatically. Additionally, this research focuses on neurophysiological data collected in ecological settings and daily real-world activities not traditionally considered in cognitive load research. In detail, electroencephalography is employed for such a purpose. Experiments will be focused on simultaneously taking advantage of the temporal, spatial and spectral properties of physiological EEG data without making any assumption on the linearity of cognitive load, supporting the automated extraction of salient features and representations and their inner relationships with no explicit declarative knowledge from designers. Linearity here refers to the extent to which any effect on mental workload variation is exactly proportional to its cause, which is something it is assumed not to hold. This will allow moving beyond the knowledge-driven research approaches that have produced hand-crafted deductive knowledge and have dominated the research landscape on mental workload for the last 50 years. Additionally, without resorting to self-reporting subjective perceptions or task-performance measures but only employing physiological EEG data, it represents a more objective method for modelling cognitive load. Eventually, the proposed computational method does not require explicit ground truth for mental workload, usually achieved by subjectively settings two design conditions of increasing task demands, for example, 'low' and 'high', and then using this as a dependent variable to fit while developing a specific model. Instead, a self-supervised brain rate generated from data is proposed, supporting the automatic development of a method of cognitive load modelling that potentially has a higher degree of applicability and replicability. In other words, self-supervision is that property of learning methods that obtains supervisory signals from the data itself, often leveraging the underlying structure in it. Eventually, it is important to note that setting task conditions is still necessary while performing experimental studies to evaluate the variation of cognitive load. However, this is outside the scope of current research. As discussed in [1], the multiple resource theory could be used to design different task conditions by manipulating the task demands on each resource.

The remainder of this article is structured as follows. Section 3, introduces the design of a self-supervised mental workload model based on a brain rate, an index of cognitive activation, trained with deep learning techniques that are expected to identify recurrent patterns while fitting such a rate. Section 4 presents the results of the experiment, followed by a discussion in Section 5 and the identification of future research improvements.

2. State of the Art in Cognitive Load Modeling

The literature on cognitive load is vast, and recent work has attempted to collate the great amount of information surrounding this construct [1]. Mental workload management has always been one of the primary reasons for building interactive technologies, information-based procedures and user interfaces to predict and optimise their performance and that of humans interacting with them, improving their engagement and minimising errors. Sometimes, the acquisition of specific certifications or compliance with certain industrial standards of such interactive technologies is required [17]. All cognitive activities performed by humans, even the most simple and rudimentary, require some mental processing and, therefore, at least some degree of mental workload. Three main classes of measures of mental workload exist, including self-report measures, physiological (and neurophysiological) measures, and primary task performance measures. Widely used self-report measures, often multidimensional, are easy to administer across contexts and

domains, do not interfere with the primary task, and are computationally inexpensive. However, they are mainly administered post-task and provide an overall mental workload assessment with low granularity and reliability over time, especially for long tasks. Additionally, as scores are derived from subjective perceptions, it is difficult to use them for comparison across participants on an absolute scale [1,8]. Task performance measures can be gathered continuously, even for long tasks, representing reliable indicators of mental workload, thus, the most direct indicators of human performance. However, their diagnosticity is poor in spotting sources of workload, and if taken in isolation, they have low utility. Contrarily, if considered in conjunction with other measures, they can be useful in helping scholars establish cause and effect relationships [25]. Physiological measures are getting momentum thanks to advances in sensor-based technologies, with increasing applications in experimental settings. Their main utility is monitoring the body's physiological and neuro-physiological responses of the brain continuously over time, often not interfering with the primary task and with greater sensitivity. However, offline pre-processing pipelines on gathered signals are often required to reduce the presence of internal and external artefacts. These are computationally expensive, thus making them more difficult to administer and limiting mental workload assessment in real-time, ecological settings. Several evaluation criteria exist for measures of mental workload, with sensitivity, validity, reliability and diagnosticity as the primary ones [1]. Given the use of neuro-physiological measures in the current research study, this section is mainly devoted to reviewing related work on the application of Electroencephalography to the problem of cognitive load modelling. Readers can obtain further information on mental workload in [1,8,17,25].

Electroencephalography (EEG) is a technique for the direct assessment of brain electrical activity via electrodes placed on the scalp and, as a consequence, the inference of objective neuro-metrics of human mental activation and mental states [26]. The advantages behind the application of EEG data for cognitive load modelling are represented by its high portability, when compared to neuroimaging methods such as fMRI [27], its wider applicability in ecological settings [28,29], financial affordability [30], and its high temporal resolution [31]. Unfortunately, EEG-based cognitive load modelling methods must consider several technical issues. Firstly, variation in EEG signals exists mainly because of the slight differences in cortical mappings and brain functioning of subjects, leading to differences in spatial, spectral and temporal patterns or due to imperfect fitting of the EEG cap on heads of different shapes and sizes. Therefore, a key challenge in successfully recognizing mental states from EEG data is to create a model that is robust to deformation and translation of signal in space, frequency, and time due to inter and intra-subject differences and to the protocols or methods employed in signal acquisition. Fortunately, advances in machine learning [32] and particularly in deep learning methods [33] have proven useful for learning models from EEG data [34], and for minimising the above technical issues while learning from data. The advantage of these data-driven deep-learning methods is that they support the automatic extraction of meaningful high-level representations from complex, non-linear data [35], they can lead to the creation of learning architectures that have wider applicability, supporting replicability of experimental research, and are flexible enough to be adapted and extended, eventually supporting advances and research progresses. However, applications of deep learning methods with EEG data have barely attempted to jointly preserve the structure of EEG signals within space, frequency and time. Most studies have focused on spatio-temporal learning [36], time-frequency learning [37] or spatial-frequency learning [38]. Therefore, a challenge is to inductively learn a model capable of exploiting the spatio-temporal and frequency-based properties of EEG data.

The literature on cognitive load modelling with EEG and deep learning is recent, not vast and highly scattered [39–47]. Most of these models are supervised, which means they require a form of ground truth, usually in task-based categories or task-performance measures. Unfortunately, there is no agreement among researchers on how to form such categories systematically. This limits comparisons across studies because, on the one hand, some scholars might focus on building a model for classifying low or high levels of task load

for relatively simple tasks. On the other hand, others might focus, for example, on building models for assessing low, medium or high load of complex tasks. In other words, these models are context-dependent, and they learn high-level features from EEG data focused on fitting these application-specific target classes. Therefore they cannot be meaningfully used across studies, limiting their generalisability. Some recent work focused on applying unsupervised learning techniques such as auto-encoders to automatically learn relevant latent representations from EEG data in an unsupervised fashion or aimed at automatically reducing the presence of noise in the data itself [48,49]. However, these unique high-level representations are often used to learn a second model that, unfortunately, still often requires supervision, as the goal is to fit, as described earlier, categories of task load, these being the independent feature subjectively defined by researchers. State-of-the-art models manipulating EEG data often rely on frequency bands, such as the alpha or theta rhythms, deemed the alphabet for brain functions and mental state extraction. These have been individually used as cognitive load indicators [50,51], or aggregated together [21,52–54] because they have been shown to be sensitive to task difficulty manipulation, task engagement or memory load [55,56]. However, these approaches often discard some EEG bands in favour of other bands.

3. Design and Methods

A novel method is proposed to tackle the issues in modelling cognitive load, as discussed in the previous sections, followed by an empirical study to validate such a method. Contrary to all the existing methods of cognitive load modeling, the method proposed here is self-supervised [57,58]. Self-supervision is an approach that autonomously learns from the data itself, and that is in the middle between supervised and unsupervised learning methods within the discipline of artificial intelligence. It is not fully supervised because it does not require ground truth (an independent variable to fit), usually as a form of declarative knowledge. It is also not fully unsupervised because it is not used for discovering patterns in the EEG data that need to be subsequently labelled and categorised with human intervention. Rather, self-supervision refers to the fact that the ground truth is generated by some automatic methods applied to the available data itself. Subsequently, some supervised machine learning algorithm uses this ground truth as supervisory data to train a model. In other words, self-supervised machine learning can be seen as an autonomous form of supervised learning because it does not require explicit human declarative knowledge.

Starting from the definition of workload proposed in [1] whereby ‘Mental workload (MWL) represents the degree of activation of a finite pool of resources, limited in capacity, while cognitively processing a primary task over time, mediated by external stochastic environmental and situational factors, as well as affected by definite internal characteristics of a human operator, for coping with static task demands, by devoted effort and attention’, this study is built upon a simpler version. This simpler definition assumes that mental workload is the activation of the brain at a given point in time. In details, analogously to blood pressure and heart rate, seen as standard preliminary indicators of general bodily activation, a brain rate is proposed as an indicator of mental activation, and then used in this research as an indicator of cognitive load. This simpler definition indeed assumes that the brain is limited in its capacity, but it does not dissect it into multiple resources. Additionally, the influence of the internal characteristics of a human operator during task performance, and the influence of external environmental and situational factors are not modelled here, nor the effort and attention devoted to the primary task. Rather, the continuous cognitive activation, as measured by such brain rate, is assumed to be the means to explain the manifestation of certain mental states over time, such as effort or attention, and the influence of the internal and external factors of and to a human operator.

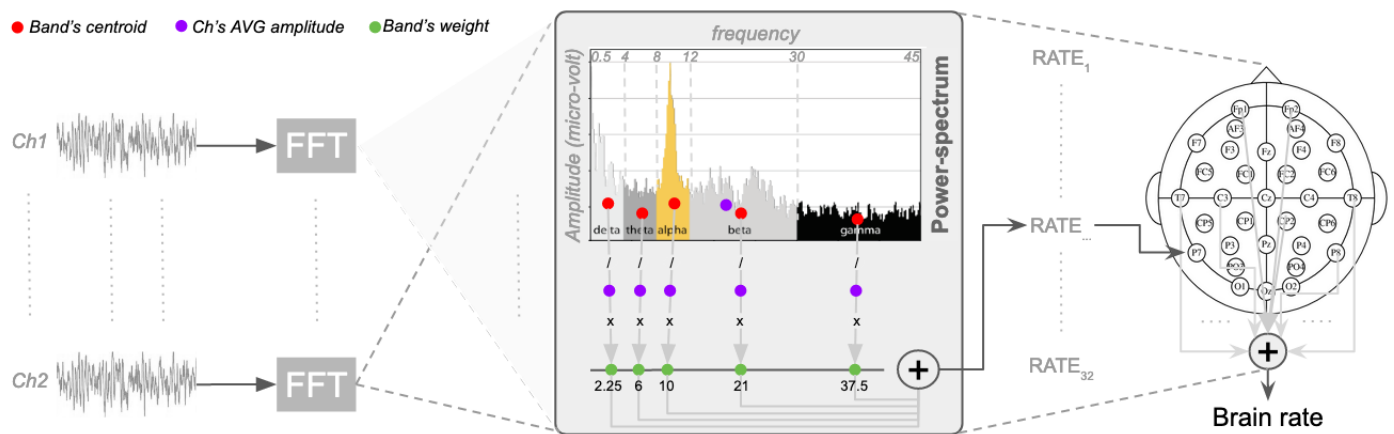


Figure 1. Diagrammatic illustration of the computation of the mean frequency of brain oscillations, via fast-Fourier Transformation (FFT), weighted over the EEG bands of potential (power) spectrum for each channel and their final aggregation towards a brain rate.

In contrast to the approaches that suppress or elevates some EEG band, as described in the previous section, the proposal is to fully use them, reasonably assuming that, whenever some band is modulated, the others are influenced too [59]. Analogously to the computations for the centre of gravity or the mean energy of a physical system [60], a spectrum-weighted frequency rate across the five canonical EEG bands (delta, theta, alpha, beta, gamma) is proposed [61], here on referred to as the brain rate (BR). This is the sum of the mean frequency of brain oscillations weighted over the EEG bands of the potential (power) spectrum for each channel, as illustrated in Figure 1). Formally:

$$BR = \sum_{ch=1}^n \sum_{b=1}^5 f_b \cdot P(b, ch)$$

where b is the index denoting the frequency band (for delta $b = 1$, theta $b = 2$, alpha $b = 3$, beta $b = 4$, gamma $b = 5$), ch is the index denoting a specific EEG channel, f_b is the weight associated with frequency band b , which is the mean frequency of each EEG band. Setting the boundaries for each band in hertz as $delta = [0.5, 4]$, $theta = (4, 8]$, $alpha = (8, 12]$, $beta = (12, 30]$ and $gamma = (30, 45]$, then $f_1 = 2.25$, $f_2 = 6$, $f_3 = 10$, $f_4 = 21$, $f_5 = 37.5$ (Figure 1). $P(b, ch)$ is the mean amplitude of the electrical potential for band b of a channel ch over the mean of all its amplitudes:

$$P(b, ch) = \frac{avg_b(FFT_{ch})}{avg(FFT_{ch})}$$

with FFT_{ch} is the vector containing the amplitudes of the fast-Fourier transformed channel ch , avg_b is the average (centroid) of only the amplitudes within the frequency band b . Note that f_b is in hertz, and $P(b, ch)$ is in microvolt, with the brain rate BR in hertz. P can be seen as the probability of having f_b , with the number of frequencies in the lower bands (example delta) smaller than the number of frequencies in the higher bands (example gamma). In other words, since the ranges of the frequency bands are different, then also the number of the frequencies within each of them is different, thus having different effect in the computation of the brain rate. Thus, P balances the importance of each band and then allows the fair identification of the dominant activated band in each EEG segment considered. By keeping the length of an EEG segment relatively short, in the order of seconds, then this rate can be used as a pseudo-real-time measure of cognitive load, since it is the mean activation of the brain response, as registered all over the scalp. Pseudo-real-time is because this rate is computed over a window of EEG data rather than a single point in time. This is also dictated by the fact that the Fourier transformation requires some data collected over time to produce meaningful translations in the frequency domain.

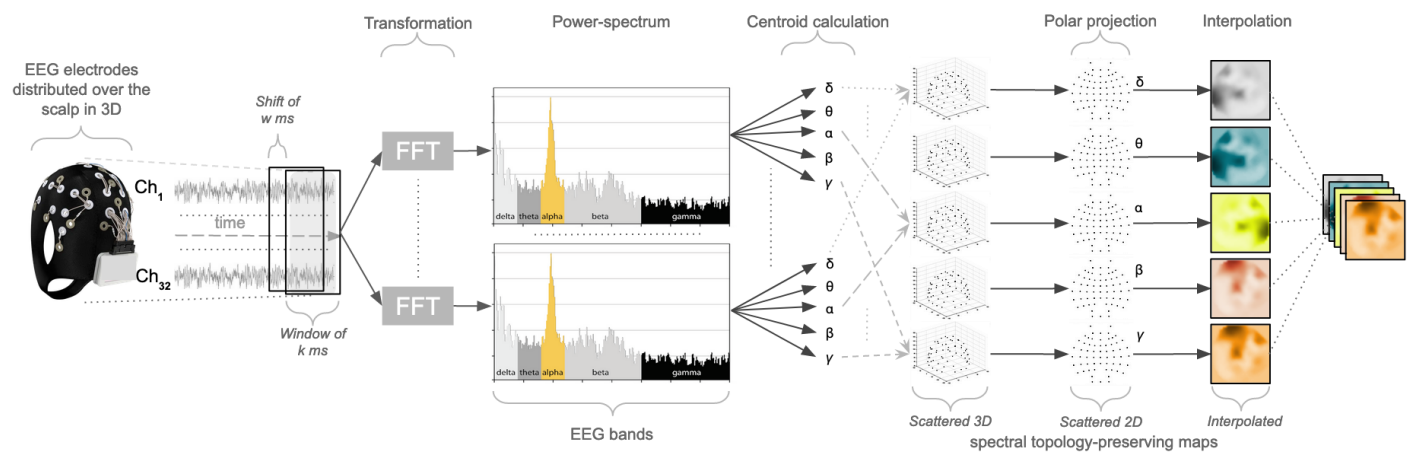


Figure 2. Processing pipeline for producing topology-preserving head maps from windows of EEG data. (I) The electrodes distributed over the scalp in a 3D space produce neural signals continuously over time; (II) these are segmented into windows; (III) for each signal in a window, fast-Fourier transformation (FFT) is applied to obtain information in the power spectrum; (IV) each power spectrum is divided into the five EEG bands (delta, theta, alpha, beta, gamma); (V) the centroid of the frequency amplitudes for each band is computed; (VI) all the centroids are positioned in a 3D space to produce a scattered head map, one for each EEG band; (VII) polar projection is applied to each scattered map to produce 2D head maps; (VIII) each 2D map is interpolated; (IX) the 5 2D maps, one for each EEG band are aggregated into a tensor.

One common problem within neuroscience, in general, and for the specific technical challenge of creating a robust model of cognitive load, in particular, is the limited availability of EEG data. This is often due to the difficulties in recruiting participants, or faulty recordings, or the presence of various artefacts in the EEG signal, leading researchers to discard significant portions of collected data. Unfortunately, when employing machine learning methods, in general and deep learning methods in particular, limited training data might often not benefit a robust model formation. For these reasons, this work proposes to use a sliding-window technique [62]. The available EEG data are segmented into windows of k seconds, shifted by w milliseconds. For each window, a pre-processing pipeline has been designed for producing 2D spatial-spectral preserving images, as summarised in Figure 2. Fast Fourier transformation is run for each EEG channel in each window, obtaining a power spectrum in the frequency domain. For each spectrum, the five EEG bands (delta, theta, alpha, beta, gamma) are defined by employing the same boundaries used to compute the brain rate. For each band, the centroid (geometric centre) is computed, which equates to the arithmetic mean of all the power values within that band. For a given band, all the computed centroids, one for each channel, are positioned in a 3-dimensional space, following the coordinates of each electrode position on the scalp, forming a scattered 3D spectral topology-preserving map. Azimuthal Equidistant Projection (polar) is subsequently used to transform this map into a scattered 2D map, preserving the relative distance between adjacent electrodes. Eventually, the Clough-Tocher method [63] is applied to fill the scattered 2D maps by estimating the values in-between the electrode over a new interpolated map, an image of 32×32 . The aggregation of the five 32×32 maps, one for each EEG band, creates a tensor of $32 \times 32 \times 5$. The sequence of these tensors can be seen as an ‘EEG movie’, a stream of data over time in the frequency domain that preserves information in space. This stream can then be processed with deep learning methods, inspired by state-of-the-art video classification methods for spatio-temporal feature learning [64,65].

The aforementioned justifications and design choices have led to the design of a novel self-supervised convolutional, recurrent deep neural network trained to fit the brain rate introduced above. The proposed architecture, as depicted in Figure 3, is built upon a first part, the Convolutional Network (CNN), due to its ability to learn robust compressed representations of EEG data, and upon a second part, the Recurrent Network (RNN) to account for

temporal variations. From a higher perspective, the overall architecture contains z parallel convolutional networks with shared weights, which are useful for representational learning. Their outputs, high-level representations referred to as feature maps, are concatenated into a sequence of length z , respecting their time order. This sequence is subsequently injected into a recurrent network composed of Long Short-term Memory units (LSTM) aimed at temporal feature learning. The feature maps, the output of each CNN parallel network, are injected into a final convolutional one-dimensional layer, and along with the output of the last LSTM unit, they are used to fit the brain rate extracted from the $z + 1$ EEG window (hence self-supervision).

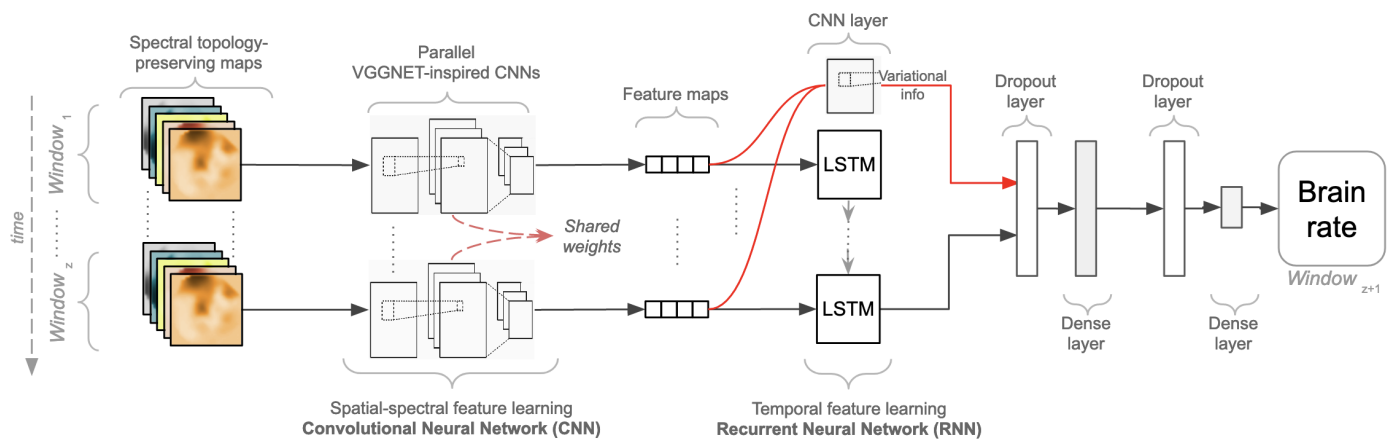


Figure 3. A self-supervised Convolutional-recurrent deep neural network for spatio-temporal learning with spectral topology-preserving head maps and a brain rate. Parallel Convolutional Neural Networks (CNN), inspired by the Visual Geometry Group Neural Network (VGGNET), are injected into a Recurrent Neural Network (RNN) with Long-Short Term Memory (LSTM) layers to learn a brain rate.

In more detail, the CNN architecture was inspired by the VGG-NET architecture designed and used in the Imagenet classification challenge [66,67]. In detail, this network, as depicted in Figure 4, is composed of 7 stacked convolutional layers with small receptive fields of size 3×3 and stride of 1×1 pixel, with Rectified Linear Unit (ReLU) selected as the activation function. To preserve the spatial resolution of each of the $32 \times 32 \times 5$ topology-preserving spectral maps of each convolutional block, each layer's inputs are padded with 1 pixel. Each stacked block of convolutional layers is followed by a max-pooling layer over a 2×2 window with a stride of 2×2 pixels. The number of kernels in each convolutional block doubles for every consecutive block, expecting to create effective receptive fields of higher dimensions while requiring fewer parameters [66]. In summary, this network contains 4 consecutive 2D CNN layers with 32 filters, each with a kernel size of 3×3 , a stride of 1×1 and no padding ('valid' padding), followed by a max pooling layer with a stride size of 2×2 and zero-padding ('same' padding, results in padding with zeros evenly to the left/right or up/down of the input). This block is followed by another one containing two 2D-CNN layers with 64 filters, with a kernel size of 3×3 , a stride of 1×1 and no padding (valid padding), followed by a max pooling layer with a stride size of 2×2 and zero-padding (same padding). Eventually, the last block contains a single 2D-CNN layer with 128 filters, with a kernel size of 3×3 , a stride of 1×1 and no padding (valid padding), followed by a max pooling layer with a stride size of 2×2 and zero-padding (same padding).

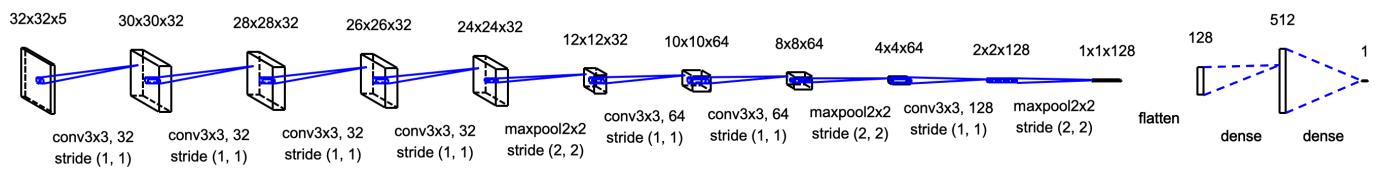


Figure 4. Single VGGNET-inspired Convolutional Neural Network (CNN) architecture for feature maps learning with spectral topology-preserving head-maps with brain rate as a target feature.

Since the nature of neural responses is dynamic over time, a suitable method for modelling the temporal evolution of brain activity is recurrent neural networks (RNNs). Technically, Long Short-Term Memory (LSTM) appears to be an appropriate modelling choice [68]. It is a specific type of RNN that uses memory cells with internal memory, and gated inputs/outputs which have led to the creation of models that are efficient in capturing long-term dependencies. The hidden layer function for LSTM is calculated by applying the following equations:

$$\begin{aligned}
 i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\
 c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
 o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \\
 h_t &= o_t \tanh(c_t)
 \end{aligned}$$

σ represents the logistic sigmoid function, i as the input gate of the LSTM model, f as its forget gate, o as the output gate and c as the cell activation vectors. As shown in [69] where various trials were performed with EEG data, a reasonable number of LSTM units seems to be only one, with 128 cells in it. This architecture was adopted to capture the temporal relationship of the feature maps obtained from each parallel CNN and shaped as a sequence of feature maps. However, only the output made by the LSTM after seeing the complete sequence of the feature maps produced by each parallel CNN was propagated to a fully connected layer. This fully connected layer also gets the output of a CNN layer that receives the concatenation of the features maps computed by each of the parallel CNNs. This is because of the reasonable assumption that variations between these may contain additional information about the underlying mental state experienced by a subject. This is a single 2D-CNN layer containing 64 filters with a stride of dimension 1×1 with valid padding and ReLU as the activation function. The output of this layer was concatenated to the output of the last LSTM, followed by a drop-out layer with a probability of 0.5, and its output was injected to a dense layer with 512 neurons and ReLU as an activation function. Another dropout layer with a probability of 0.5 followed, and a final dense layer with a linear activation function was devised for fitting the brain rate computed for the next window of EEG data following the sequence in time ($z + 1$). Concerning the hypothesis that this study seeks to test, this is:

H: IF a convolutional-recurrent deep neural network architecture is trained with spatio-temporal spectral topology-preserving head maps, derived from multi-channel EEG data, to fit a brain rate, an index of cognitive activation, in a self-supervised fashion.

THEN within-subject and across-subjects models can be induced with low error rates, highlighting recurrent patterns of cognitive activation, thus cognitive load.

In order to test such a research hypothesis, data from a well-known dataset of EEG recording is employed, namely, the DEAP dataset [70], as described in the following section.

3.1. Dataset and Pre-Processing

The DEAP dataset [70] has been widely used for various experimental purposes in the past few years. The electroencephalographic (EEG) data in this dataset was recorded from 32 participants while watching 40 one-minute-long excerpts of music videos [70]. Cortical activity was recorded at 512 Hz using a Biosemi ActiveTwo system using 32 active AgCl electrodes placed according to the international 10–20 system, with participants sitting 1 m away from a 17-inch screen. A 5-s fixation cross was run before each video to act as a baseline. Participants watched two blocks of 20 videos each, separated by a short break. EEG data was recorded from the following 32 channels: Fp1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, P03, O1, Oz, Pz, Fp2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, PO4, O2. Biosemi ActiveTwo system was used to record data, with Cz used as the reference channel. A pre-processing procedure using the EEGLab toolbox was applied to data, including (i) downsampling to 128 Hz (ii) EOG artefact removal using the ICA blind-source separation technique (iii) band-pass filtering between 0.5 Hz to 45 Hz (iv) common average referencing. Further information on how the dataset has been formed, details on the study and consent forms, and the decision taken to select videos, can be found in [70]. For the current research, the most important reason behind the selection of this dataset was that the data was recorded for a prolonged time, which means 1 min, and not in the order of seconds, as often the case for event-related potential studies. The reasonable assumption was that while cognitively processing excerpts of videos, participants would have also experienced different levels of cognitive processing [71]. In other words, the dataset was selected because of the presence of various participants, executing different tasks, and cognitively processing information in variables ways, thus activating the brain differently. It is assumed that with this variability in brain responses, also the resulting computation of the brain rates are variable, which is an important property to train cognitive load models with the method described in the previous section.

3.2. Training

After the pre-processing pipeline is applied to selected EEG data, a new procedure (as depicted in Figure 5) is designed and run to generate training instances for the specific convolutional/recurrent neural network described in the previous section. Here, each video that participants watched lasted for 63 s (60 for the actual video and 3 for baseline). A time window of 2 s is set for producing spectral topology-preserving maps by applying the processing pipeline described in Figure 2. This length is deemed short enough for producing a meaningful power spectrum that contains enough points well distributed across the five EEG bands. In detail, given a final sample rate of 128 Hz, each window contains 256 points (128×2) spread across the EEG bands for each channel. This means that each video contains 8064 points ($63 \times 128 = 8064$). A sliding-window technique is applied across these points, and a shift of 125 ms is used (8 points per second), which translates into a shift of 16 points (128×0.125). This generates 489 windows of 2 s ($63 \times 8 - 16 + 1$) for each video in the dataset. The neural network designed in Figure 3 is a specific convolutional-recurrent neural network accepting a sequence of windows. As mentioned before, this sequence is set to $z = 7$ windows, equating to 14 s of neural activity. This is believed to be short enough for the expectation of detecting some variations in cognitive load, and not too long for hampering the automatic learning of temporal dependencies across points. Each of these sequences represents a training input instance. Thus, 482 of these instances (sequences) were produced for each video ($489 - 7$). As previously mentioned, the designed architecture is a specific self-supervised many-to-one network. The target output is the brain rate computed for the subsequent window outside the sequence, next in time (the 8th). The goal is to learn this rate from past information, which in other words, is the estimation of a brain rate from the neural activity of the previous 14 s (7×2).

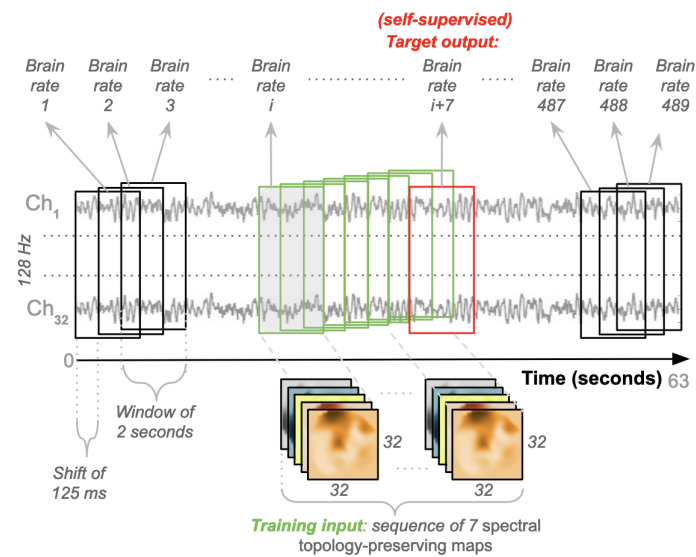


Figure 5. Pipeline for generating sequences for the convolutional-recurrent neural network.

Several models are trained within and across subjects to test the research hypothesis, as listed in Table 1. Since each participant watched 40 videos, then the number of total sequences associated with each participant equates to 19,280 (482×40). The canonical approach employed in machine learning to create generalisable models would be to shuffle these sequences and split them into training, validation and test sets. However, although technically valid, performing such a shuffle for training a within-subject model would generate a training set that will likely contain some sequence from each video. In other words, each video would have a certain amount of representative data in the training, validation and test sets. To further increase generalisability, it is decided that the training set contains entire data from random 70% of the possible videos, and the validation and test sets, respectively, 15% of the data associated with the remaining videos. Thus, the shuffle is done at the video level, and data associated with 28 random videos are selected as the training set ($482 \times 28 = 13,496$ training sequences), data from 6 different random videos for the validation set ($482 \times 6 = 2892$ training sequences), and the data from the remaining videos for the test set. In this way, the generalisability is exploited across unseen test videos, expected to lead to different cognitive load fluctuations than those used for training and validating models. The same rationale is applied to across-specific models. The only difference is that the training, validation and test sets contain data from a random number of participants, as listed in Table 1. In other words, for example, for a 3-persons model, 3 splits are performed for each participant individually. Then the resulting individual training, validation and test sets are concatenated to produce larger sets. 32 within-subject CNN models (Figure 4) are trained for participants twice with different batch sizes (32 and 100). This step aims to understand batch-size manipulation to validate and test errors.

Table 1. Details of within and across-subjects models with number of training, validation and test instances, as well as the number of Monte Carlo repetitions.

Models	Type	Instances (Training Sequences)				Repetitions
		Total	Training	Validation	Test	
1-person	within subject	19,280	13,496	2892	2892	2
3-persons	across-subjects	57,840	40,488	8676	8676	10
5-persons	across-subjects	96,400	67,480	14,460	14,460	10
7-persons	across-subjects	134,960	94,472	20,244	20,244	10
9-persons	across-subjects	177,570	125,514	26,028	26,028	10

The rationale is to analyse the trade-off between generalisability and computational resource consumption since it is known that larger batches lead to better convergence to the global optima of the objective function but at the cost of slower convergence since more memory is requested and more computations are performed. Instead, smaller batches allow the model to start learning earlier, before seeing all the data, with lower consumption of computational resources. Still, it is not guaranteed that the model converges to the global optima, thus with a negative impact on its generalisability. After assessing the ideal batch size, across-subject models are trained with incremental complexity, in terms of a higher volume of data coming from an increasing number of participants, to assess whether their generalisability still holds with a higher heterogeneity in the EEG signals. Additionally, to reinforce the analysis, repeated Monte Carlo sampling is performed for each across-subject model, with a random selection of participants at each repetition. Table 1 summarises the number of training, validation and test sequences used and the number of repetitions for each training configuration. The training dataset is not augmented in any way, for example, by employing image zooming or flipping techniques, because of the distinct interpretations of direction and location in the EEG topographic-maps that correspond to specific cortical regions. Training is conducted by optimising the Mean Squared Error (MSE) loss function:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

with n the number of sequences (of length 7), y_i the observed brain rate for that sequence (in the 8th position) and \hat{y}_i the predicted brain rate for that sequence. Validation and test MSEs is monitored during and after training and Mean Absolute Percentage Error (MAPE) is also computed:

$$\frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

where y_i is the observed brain rate and \hat{y}_i is the predicted one. Their difference is divided by the actual observed brain rate y_i . The absolute value in this ratio is summed for every predicted brain rate and divided by the number of sequences n . MAPE comes under percentage errors and it has been selected because these errors are scale independent, thus especially suitable for across-subject models and because it is easier to interpret and explain. As mentioned earlier, the parallel CNNs share weights, thus potentially producing different gradients in different internal layers. As a consequence, a smaller learning rate, set to 1×10^{-3} , was employed when applying the Stochastic Gradient Descent (SGD) to the CNNs. Similarly, the whole neural network was trained with a small learning rate of 1×10^{-3} , optimised with the Adam algorithm [72], shown to achieve reasonable fast convergence rates, with decay rates of first and second moments set to 0.9 and 0.999, respectively.

The overall final neural network devised contains a large number of parameters (1.62 million) and considering that a different number of models are trained with an increasing amount of training instances per model, with each instance being a tensor of $32 \times 32 \times 5 \times 7$ (where 32×32 is the size of the spatial-preserving topographic maps, 5 is the number of EEG bands, 7 is the number of EEG windows, that means the length of the trainable sequence), a significant demand on computational resources, in terms of memory and processing power, is required. Additionally, many parameters can make each trained model susceptible to overfitting. Therefore, several measures are taken into account. As mentioned earlier, all the CNN networks share parameters across the 7 frames. Thus, a good amount of parameters in the overall architecture were removed. Dropout layers were added after each fully connected layer, with a probability of 0.5 to minimise overfitting [73,74]. Similarly, an early stopping training mechanism is employed to avoid training models when it is no longer necessary, thus saving a significant amount of time. This is an optimization procedure that is also used to minimise overfitting without compromising on model accuracy. In detail, this is a regularization technique that stops training when the updates of the model's parameters no longer yield improvement on a validation set after consecutive

E epochs. The value E is called patience, and in this study it was set to 6, after some trials. This means that the training phase early stops automatically when the error associated with the validation set does not reach a lower value for 6 consecutive epochs, and the E th-last model is retained as the final model.

Data up to 9 people are considered to train a single across-subject model since this is the maximum amount of data that the selected machine has been estimated to process with its resources. In particular, this machine is an Alienware Aurora R8 (model: 02XRCM), Intel Core i7-8700 (6-core, 12 threads), 64 bits, 12Mb L2-cache, 32GB DDR-SRAM, 2 additional graphics cards (GeForce RTX 2070), with the Linux Mint 19.2 operating system, and an internal local total storage of 4 TeraBytes, comprising a primary 1TB SSD (Solid State Drive) hard-disk (model: SK Hynix, PC601 NVMe), a 3.5-inch 2TB hard-drive (model: Seagate BarraCuda ST2000DM008-2FR102) and an additional 1TB SSD hard-disk (model: 2-Power SSD2044A). For allowing training of across-specific models (up to 9 persons), a Swap RAM of 0.5TB was created.

4. Results

Figure 6 depicts the density plots of the validation and test mean squared errors (MSEs) for the 32 within-subject models trained only by employing the CNN architecture (Figure 4), respectively, with batch size of 32 and 100. Similarly, Figure 7 depict the density plots of the number of epochs necessary to train the within-subject CNN architectures, respectively, with a batch size of 32 and 100, with a minimum of 7 epochs to a maximum of 60. No significant difference exists in the validation and test errors, with the batch size of 32 leading to slightly better (lower) MSEs. However, although not significantly different, on average, the number of epochs necessary to train CNN models with batch size 32 is lower than that associated with batch size 100. Every epoch for the within-subject model, with the current machine, required on average 300 s (5 min), thus, the finalisation of training, according to the minimum and a maximum number of epochs (7 and 60), required between 2100 to 18,000 s (35 and 300 min). Therefore, 32 was the batch size selected for training the subsequent within-subject and across-subject models with the full architecture (Figure 3) since it leads to a lower number of training sequences in one forward/backwards pass, thus lower consumption of memory, as well as a lower number of training epochs, saving a great amount of time.

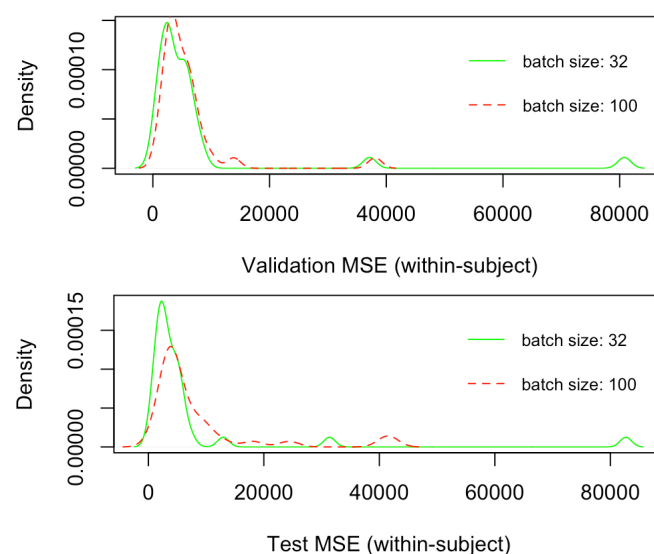


Figure 6. Comparison of validation and test Mean Squared Error for within-subjects CNN models trained, respectively, with batch size of dimension 32 and 100.

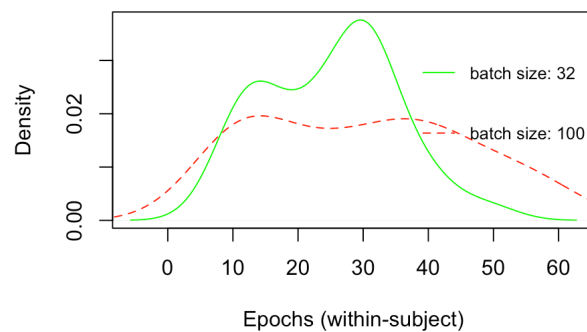


Figure 7. Comparison of the number of epochs to train the within-subjects CNN models, respectively, with batch size of dimension 32 and 100.

Figures 8 and 9 depict the Mean Absolute Percentage Errors (MAPE) for the test data of the within-subject models for the 32 participants, trained first with the single CNN architecture of Figure 4 for learning the weights (in full red), and with the convolutional/recurrent neural network with the parallel CNNs, sharing such weights, and the LSTM component for temporal learning (Figure 3) (in dashed blue). As it is possible to notice, the test MAPE has mean 0.111 (Std: 0.073) for the single CNN models and mean 10.75 (Std: 0.070) for the CNN+LSTM models. These results demonstrate that the brain rate prediction for each participant's unseen test data is good because the forecast is only off by roughly 10%. However, at first glance, it seems that the impact of the addition of the recurrent component (the Long Short Term Memory), as in the architecture depicted in Figure 3, does not add much value in minimising the MAPE. This seems to point to the individual capability of the single CNN architecture (Figure 4) to learn the relevant patterns, intricacies and relationships in the data in the shape of topographic head maps containing information in the 5 EEG frequency bands for the specific window length used (2 s). However, the LSTM layer takes a sequence of 7 outputs from the single CNNs (in addition to a vector containing their variational information) and tries to fit the brain rate associated with the next window (the 8th after the sequence). The fact that the MAPE of the CNN+LSTM does not significantly change (decrease) does not mean that the LSTM did not learn any temporal relationship and dependency in the input sequences. This can be demonstrated by inspecting Figure 10, whereby the brain rate index, the predictions of the single CNN model and those of the CNN+LSTM for some within-subject models associated with random participants and a random video in their respective test sets, are compared. In detail, these figures show that the brain rates (green), computed for each of the 482 instances, as explained in Section 3.2 (and depicted in Figure 5), associated with a specific video that a participant has watched, not used for training the within-subject model of that participant, are reasonably approximated by the single-CNN within-subject model (red). However, the brain rate indexes seem better approximated by the CNN+LSTM within-subject model (blue).

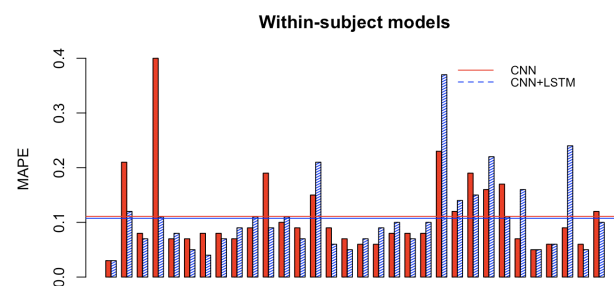


Figure 8. Paired histogram of the Mean Absolute Percentage Errors (MAPE) of the test data of the 32 within-subject models, respectively, trained only with the single Convolutional Neural Network (CNN), and the Convolutional/Recurrent Neural network (CNN+LSTM).

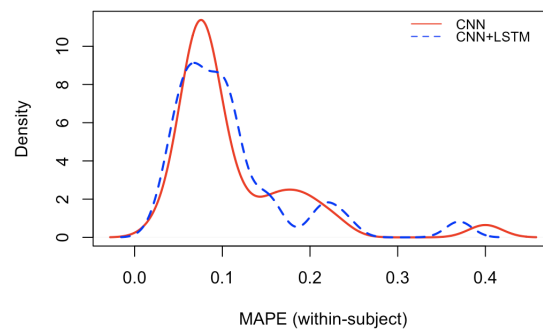


Figure 9. Density plot of the Mean Absolute Percentage Errors (MAPE) of the test data of the 32 within-subject models, respectively, trained only with the single Convolutional Neural Network (CNN), and the Convolutional/Recurrent Neural network (CNN+LSTM).

The comparisons of Figure 10 highlight a number of things. Firstly, the main bursts in the brain rates are also grasped by the CNN and the CNN+LSTM models. However, those associated with the CNN (red) are shifted a bit to the right (x time axis) when compared to those associated with the CNN+LSTM (blue), which seem to be more aligned to the brain rates (green) over time. This is confirmed by the Person correlation coefficient, which on average for participants and testing videos, is 0.5 for the CNN models and 0.7 for the CNN+LSTM models. This means that the LSTM layer in the CNN+LSTM architecture did learn some temporal relationships and long/short-term dependencies. The CNN+LSTM predictions are smoother than those produced by the single CNN, and this might be justified by the fact that they are based on the information taken from the precedent 7 consecutive EEG windows over time. For the same reasons, this might be the reason why the scale (y-axis) of the predictions of the CNN+LSTM (blue) is a bit lower than the others (blue and green).

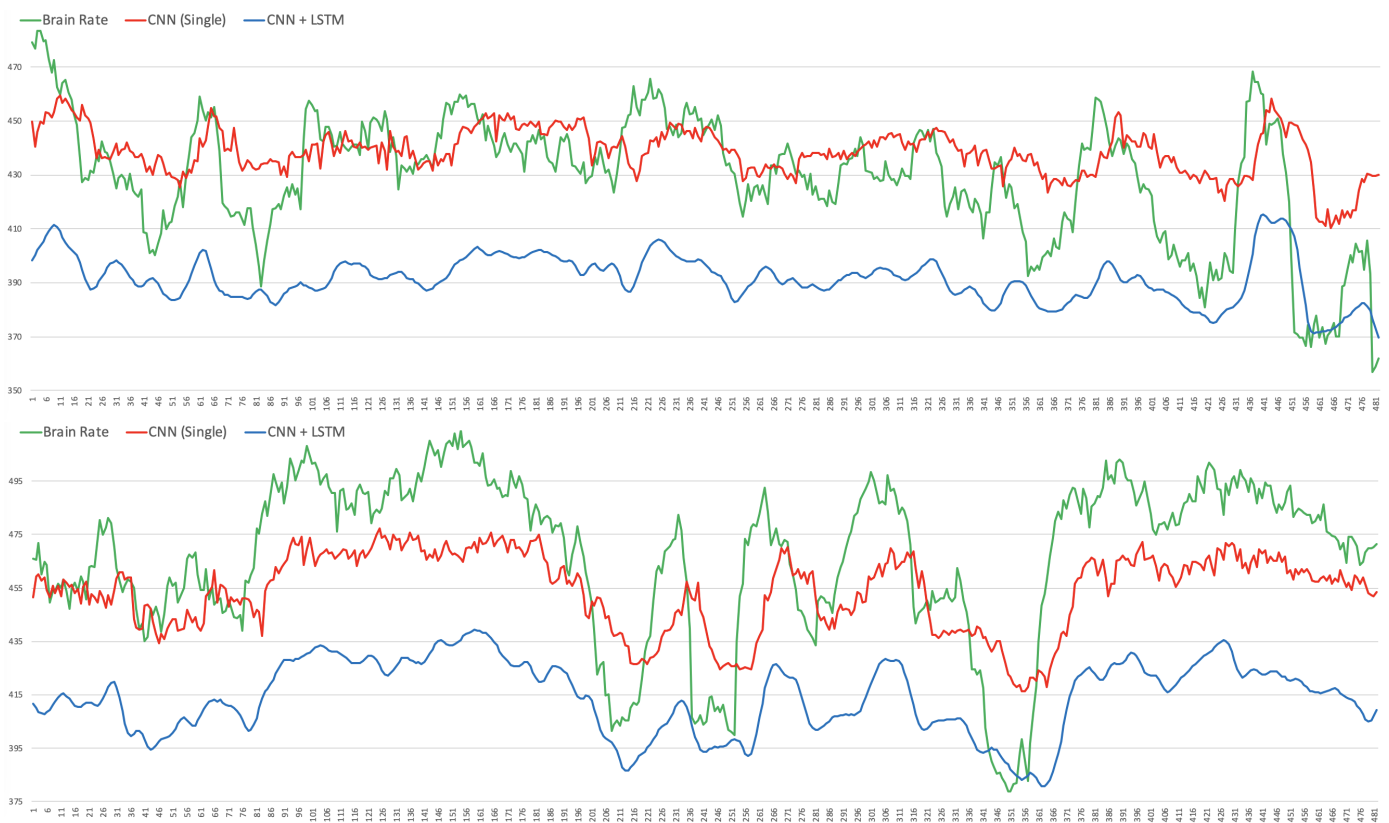


Figure 10. Illustrative comparisons of the brain rate index, the single Convolutional Neural Network (CNN) predictions and the Convolutional/Recurrent Neural Network (CNN+LSTM) predictions for two random participants and a random video used in the test set.

Regarding the across-subjects models, as planned in Table 1, Figure 11 depicts the density plots of their Mean Absolute Percentage Errors (MAPEs) on the test sets. In detail, each density curve contains the MAPEs associated with the test sets of 10 models, each trained with the respective number of random people. As it is possible to see, the test MAPEs are lower on average for those models trained with material taken from 10 people (black), followed by those trained with 7 (brown), 5 (grey) and 3 people (yellow). Additionally, the standard deviations (width of each curve) are smaller (thinner) for those trained with data from more people and larger for those trained with data from fewer people. This means that smaller standard deviations are associated with more steady models because these are capable of predicting brain rates on the test data more consistently. These results might seem intuitive because it can be argued that the more training material, the higher capacity a model has to learn. However, training material comes from different numbers of people, selected randomly at each run, and their cerebral responses are different while watching videos, exhibiting different power activations and temporal dynamics. This introduces a higher variability within data, thus making a model prone to confusion while learning. Despite this, across-subject models can mitigate the influence of such an increasingly higher variability and can learn consistent higher-level representations that are more generalisable across people.

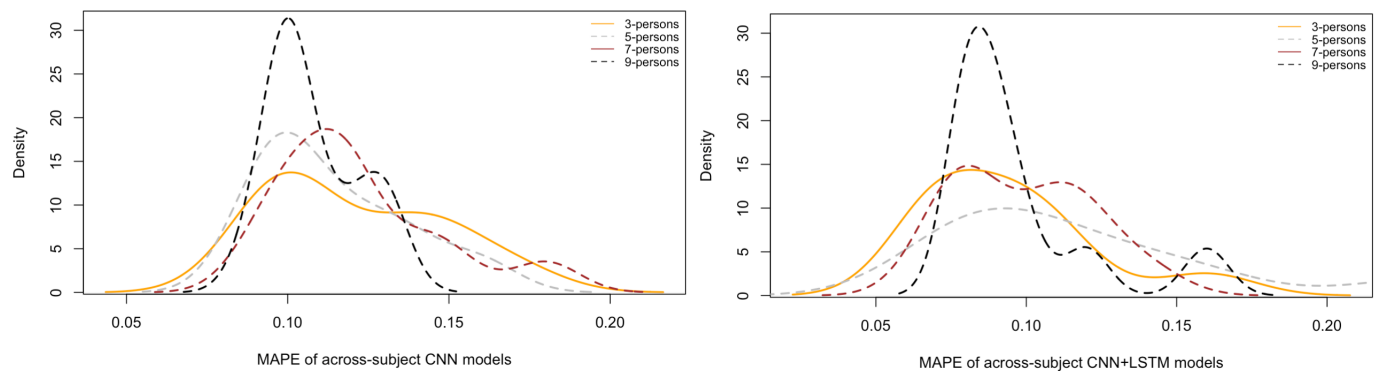


Figure 11. Comparisons of the test Mean Absolute Percentage Error (MAPE) of the across-subject models grouped by the type of architecture which is the single convolutional neural network (CNN) and the convolutional/recurrent neural network (CNN+LSTM).

Figure 12 plots the pair-wise comparison of the across-subject models trained with the single CNN and the CNN+LSTM architectures, grouped by the number of people, and the density curve associated with the MAPEs of the within-subject models, used here as baseline. Noticeably, the density plots associated with those models trained with the CNN+LSTM architecture (dashed lines) contain lower MAPEs on the test sets than those associated with the models trained with the CNN only (continuous lines). This means that the addition of the Long-Short Term Memory (LSTM) layer for temporal learning had an impact on building more accurate models, although, in this study, not statistically significant. Additionally, these results suggest that the convolution of the topology-preserving topographic maps over space (down-sampling) could learn some repetitive high-level patterns within an EEG window (as set to 2 s). In other words, as expected in the research hypothesis set in Section 3, within-subject and across-subjects models can be induced from spatio-temporal spectral topology-preserving head maps derived from multi-channel EEG data to fit a brain rate, an index of cognitive activation, with low error-rates, demonstrating the existence of recurrent patterns of cognitive load over time. A more detailed interpretation of such results, along with a discussion of the strengths and limitation of the designed method for cognitive load modeling, is done in the following section.

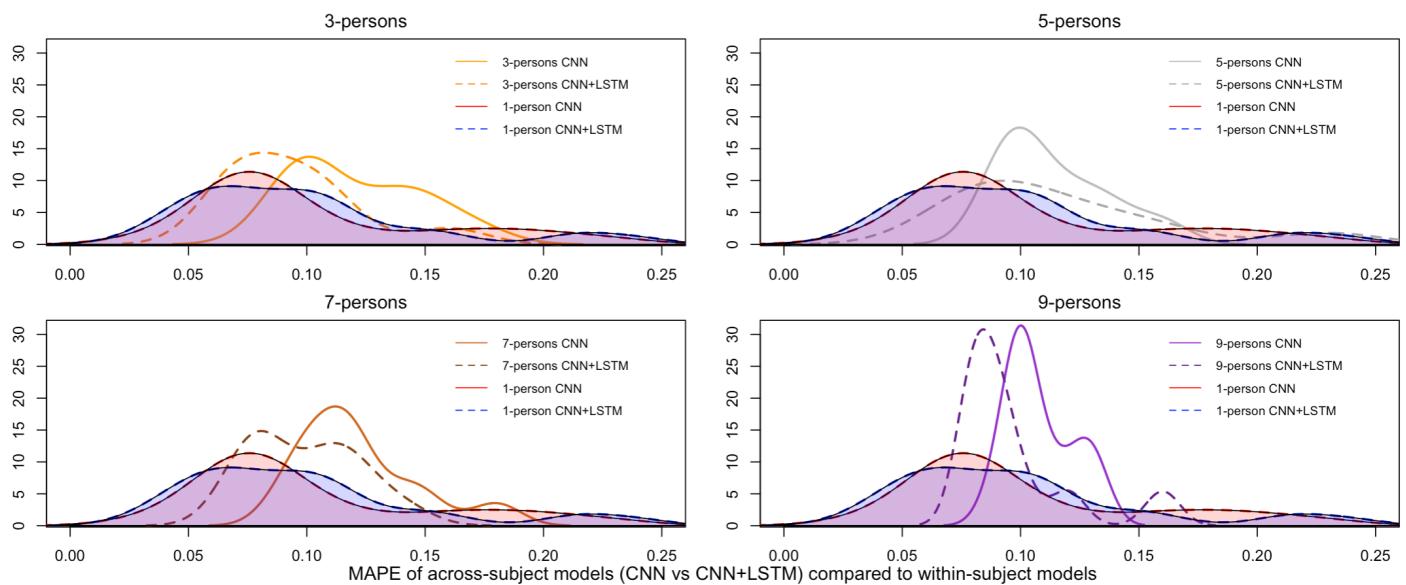


Figure 12. Pairwise comparisons of the test Mean Absolute Percentage Error (MAPE) of the across-subject models trained, respectively, with the single convolutional neural network (CNN) and the convolutional/recurrent neural network (CNN+LSTM) compared to the within-subject models.

5. Discussion

The computational method described and tested in the previous sections is fully automated and allows the induction of a model of cognitive load from EEG data based on deep learning without requiring human intervention. In summary, this novel method:

- is based on data-driven deep-learning techniques for automatic inductive learning [33];
- is built upon electroencephalography (EEG), a non-invasive method for gathering brain responses with high-temporal resolution [34];
- is sensitive to brain responses variation over time thanks to its recurrent neural network component [68];
- is robust to deformation and translation of signal in space, and frequency, thanks to the ability of its convolutional neural network component to learn meaningful representations [75];
- is built upon 2D spectral topology-preserving head maps that are rich in information and also more explainable than vectorial data, as discussed in [76–78];
- is self-supervised and does not require human intervention and explicit declarative knowledge [58];
- is constructed upon a brain rate, a measure of cognitive activation, and treated as an index of cognitive load that considers cortical brain oscillations weighted over the potentials of all the canonical EEG bands;
- is flexible with short EEG segments, thanks to its time-slicing procedure over cortical recordings;
- is adjustable and customisable because it can be trained on EEG data collected from a variable number of electrodes, it can be employed with different ranges for the five EEG bands (delta, theta, alpha, beta, gamma), and with EEG windows of varying size;
- is replicable and open to falsifiability [79], supporting the formation of models of cognitive load with higher generalisability.

This method allowed the fully-automated formation of within-subject and across-subject models of cognitive load from EEG signals. These models fit a brain rate, an index of cognitive activation, with good accuracy, measured by the Mean Absolute Percentage Error (MAPE) on the test sets, demonstrating a good degree of generalisability to unseen data. In detail, each within-subject model, trained with EEG material from a single person, could predict the brain rate of unseen EEG data—as encoded with spatially preserving

topographic head-maps built upon 32 channels—with a MAPE of 0.11 and 10.75 (std 0.073, 0.070), only using a convolutional neural network architecture for spatial learning, and its extension with a long-short term memory layer for temporal learning, respectively. The across-subject specific models, induced from an increasingly higher amount of EEG data from different people, confirmed these results and maintained the same testing accuracy as measured with MAPE, despite the increasing variability within training data. This perseverance in achieving similar testing accuracy, despite a higher variability in training data, can be seen as positive because it highlights the existence of some patterns within EEG data that are repetitive and stable. This observation might be linked to microstate theory which assumes that distributions of activity across the scalp persist for milliseconds before changing into a different pattern [80]. EEG microstates can be seen as transient, quasi-stable patterns of an electroencephalogram [81,82]. An analogy can be applied to the findings obtained in this current work, and the trained models might have learned quasi-stable patterns of mental activation fluctuations, as modelled with a brain rate. The convolution applied to the spatially preserving topographic head-maps, built over five EEG frequency bands, has already led to the development of within and across-subject models with good accuracy. This means quasi-stable high-level representations might be induced from the convolutional operations that can be successfully mapped to a brain rate. Furthermore, this view might be enforced by the minimal decrement of the test MAPEs obtained by those models trained with the LSTM layer in the neural network for temporal learning. The fact that it was minimal suggests that the sequence of convoluted representations over time is not as important as the actual representations alone, taken individually, which seem to be already rich in information and able to learn certain repetitive patterns of cognitive activation.

6. Conclusions

Cognitive Load, often referred to as Mental workload [11], is one of the most invoked concepts in the disciplines of human factors, with important utility within human–computer interaction, neuroscience and education [1]. Unfortunately, a reliable, generally applicable computational method for cognitive load modelling does not exist yet, complicating applied research. This research, the first of its kind, was aimed at developing a method for cognitive load modelling with generalisability in mind, supporting its application across disciplines, replicability, comparisons across studies and thus enabling falsifiability. All these advantages are aimed at supporting research on cognitive load modelling at a larger level, avoiding the creation of another ad hoc, field-dependent, knowledge-dependent and application-driven method of mental workload that has little chance of being generally applicable across empirical works. This novel method employs deep learning techniques of Artificial Intelligence, for the automatic formation of models of cognitive load, in a fully unsupervised way, drastically limiting human intervention and declarative knowledge. These models work on continuous EEG data, thus having a great temporal resolution. They are built upon a newly designed notion of brain rate, a particular index of cognitive load derived from the five EEG frequency bands (delta, theta, alpha, beta, gamma). This method works on spatially-preserving topographic head-maps of cognitive activation, offering spatial resolution and supporting diagnosticity. In this study, these maps are based on spectral information derived from the five EEG bands, which are known to be rich in information for deriving mental states and facilitating the analysis and interpretation of human behaviours.

Findings suggest that within-subject and across-subjects models of cognitive load, developed with the newly devised computational method, are accurate enough, exhibiting a low prediction error on unseen data, thus showing a good degree of generalisability. They suggest that certain high-level representations from EEG data in the frequency bands can be extracted automatically, frequently appearing over time. This can be related to the construct of cognitive load and these representations can be seen as patterns of cognitive activation that have a recurrent appearance. However, these existing repetitive blocks of

mental activation do not seem to be repetitive over time, in line with the non-stationary nature of brain activation. In other words, frequent, quasi-stable high-level representations of cognitive activation exist, but these are not repetitive over time. Additionally, these representations seem to be repetitive across-subjects, with important implications for the research field of mental workload. Their existence might suggest that general patterns of cognitive load exist, and these are subject-independent, therefore having a great generalisability. However, to confirm this claim, further studies are needed.

Future work will include replicating the method developed in this research study with varying time window sizes and investigating how these influence the accuracy of resulting cognitive load models. A layer of interpretability for the automatically extracted high-level representations will be deployed, considering the principles and notions of explainability from Explainable Artificial Intelligence (XAI) [77]. Similarly, by using argumentation theory and defeasible logic, as in [83], explainable rules can be created, these being more digestible and aligned to the way humans reasons under uncertainty and with contradicting pieces of information. This will help understand the shape of these high-level representations, and the recurrent activated brain regions, giving analysts a richer level of interpretability. It will also serve as a layer of explainability, providing analysts with tools for explaining spatial and temporal dynamic of cognitive activation. The inferences of these models of cognitive load can be compared against other indexes such as the theta-to-alpha or alpha-to-theta band ratios [54], increasing their meaningfulness and validity. Eventually, studies can be devoted to the development of additional recurrent neural networks for understanding the temporal aspects of the high-level representations of cognitive activation, and establishing if there exist sequences, and their lengths, that are repetitive and recurrent over time. These future avenues will expand the science of mental workload and support the formation of models of cognitive activation with an increasing accuracy and generalisability, in turn facilitating the analysis of human behaviour.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Thanks to IBM for supporting this research with its Power processor-based technology.

Conflicts of Interest: The author declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CL	Cognitive Load
MWL	Mental Workload
EEG	Electroencephalography
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory
VGGNET	Visual Geometry Group Network
MAPE	Mean Absolute Percentage Error
XAI	Explainable Artificial Intelligence

References

1. Longo, L.; Wickens, C.D.; Hancock, G.; Hancock, P.A. Human Mental Workload: A Survey and a Novel Inclusive Definition. *Front. Psychol.* **2022**, *13*, 883321. [[CrossRef](#)] [[PubMed](#)]
2. Wickens, C.D. Mental workload: Assessment, prediction and consequences. In *Proceedings of the International Symposium on Human Mental Workload: Models and Applications*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 18–29.

3. Longo, L. Designing medical interactive systems via assessment of human mental workload. In Proceedings of the 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, Sao Carlos, Brazil, 22–25 June 2015; pp. 364–365.
4. Orru, G.; Longo, L. Direct Instruction and Its Extension with a Community of Inquiry: A Comparison of Mental Workload, Performance and Efficiency. In Proceedings of the 11th International Conference on Computer Supported Education, CSEDU 2019, Heraklion, Greece, 2–4 May 2019; Volume 1, pp. 436–444. [\[CrossRef\]](#)
5. Hart, S.G. NASA-task load index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **2006**, *50*, 904–908. [\[CrossRef\]](#)
6. Young, M.S.; Stanton, N.A. Mental workload: Theory, measurement, and application. *Int. Encycl. Ergon. Hum. Factors* **2001**, *1*, 507–509.
7. Paas, F.; Tuovinen, J.E.; Tabbers, H.; Van Gerven, P.W. Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* **2003**, *38*, 63–71. [\[CrossRef\]](#)
8. Young, M.S.; Brookhuis, K.A.; Wickens, C.D.; Hancock, P.A. State of science: Mental workload in ergonomics. *Ergonomics* **2015**, *58*, 1–17. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Charles, R.L.; Nixon, J. Measuring mental workload using physiological measures: A systematic review. *Appl. Ergon.* **2019**, *74*, 221–232. [\[CrossRef\]](#)
10. Van Acker, B.B.; Parmentier, D.D.; Vlerick, P.; Saldien, J. Understanding mental workload: From a clarifying concept analysis toward an implementable framework. *Cognit. Technol. Work* **2018**, *20*, 351–365. [\[CrossRef\]](#)
11. Hancock, G.; Longo, L.; Hancock, P.; Young, M. Mental workload. In *Handbook of Human Factors & Ergonomics*, 5th ed.; Salvendy, G., Karwalski, W., Eds.; Taylor & Francis: Abingdon, UK, 2021; Chapter 7.
12. Longo, L. On the Reliability, Validity and Sensitivity of Three Mental Workload Assessment Techniques for the Evaluation of Instructional Designs: A Case Study in a Third-level Course. In Proceedings of the 10th International Conference on Computer Supported Education (CSEDU 2018), Funchal, Portugal, 15–17 March 2018; pp. 166–178.
13. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* **1988**, *52*, 139–183. [\[CrossRef\]](#)
14. Wickens, C.D. *Processing Resources and Attention*; CRC Press: Boca Raton, FL, USA, 2020.
15. Brookhuis, K.; De Waard, D. Assessment of drivers' workload: Performance, subjective and physiological indices. In *Stress, Workload and Fatigue*; Hancock, P., Desmond, P., Eds.; Lawrence Erlbaum Associates: Boca Raton, FL, USA, 2001; pp. 321–333.
16. Vidulich, M.A.; Tsang, P.S. Mental Workload and Situation Awareness. In *Handbook of Human Factors and Ergonomics*; John Wiley & Sons, Ltd: Hoboken, NJ, USA, 2012; Chapter 8, pp. 243–273. [\[CrossRef\]](#)
17. Cain, B. *A Review of the Mental Workload Literature*; Technical Report; Defence Research and Development Canada: Toronto, ON, Canada, 2007.
18. Longo, L. A defeasible reasoning framework for human mental workload representation and assessment. *Behav. Inform. Technol.* **2015**, *34*, 758–786. [\[CrossRef\]](#)
19. Bommer, S.C.; Fendley, M. A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *Int. J. Indust. Ergon.* **2018**, *63*, 7–17. [\[CrossRef\]](#)
20. Aricò, P.; Borghini, G.; Di Flumeri, G.; Colosimo, A.; Bonelli, S.; Golfetti, A.; Pozzi, S.; Imbert, J.P.; Granger, G.; Benhacene, R.; et al. Adaptive automation triggered by EEG-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Front. Human Neurosci.* **2016**, *10*, 539. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* **2014**, *44*, 58–75. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Burns, C.M. Understanding, supporting, and redesigning cognitive work. In Proceedings of the International Symposium on Human Mental Workload: Models and Applications, Amsterdam, The Netherlands, 20–21 September 2018; pp. 3–12.
23. Xie, B.; Salvendy, G. Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work & Stress* **2000**, *14*, 74–99.
24. Johannsen, G. Workload and workload measurement. In *Mental Workload*; Springer: Berlin/Heidelberg, Germany, 1979; pp. 3–11.
25. Hancock, P.A.; Matthews, G. Workload and performance: Associations, insensitivities, and dissociations. *Hum. Factors* **2019**, *61*, 374–392. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Richer, R.; Zhao, N.; Amores, J.; Eskofier, B.M.; Paradiso, J.A. Real-time mental state recognition using a wearable EEG. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 5495–5498.
27. Lemieux, L.; Allen, P.J.; Franconi, F.; Symms, M.R.; Fish, D.K. Recording of EEG during fMRI experiments: Patient safety. *Magn. Reson. Med.* **1997**, *38*, 943–952. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Xu, J.; Zhong, B. Review on portable EEG technology in educational research. *Comput. Hum. Behav.* **2018**, *81*, 340–349. [\[CrossRef\]](#)
29. Casson, A.J.; Yates, D.C.; Smith, S.J.; Duncan, J.S.; Rodriguez-Villegas, E. Wearable electroencephalography. *IEEE Eng. Med. Biol. Mag.* **2010**, *29*, 44–56. [\[CrossRef\]](#)
30. Mullen, T.R.; Kothe, C.A.; Chi, Y.M.; Ojeda, A.; Kerth, T.; Makeig, S.; Jung, T.P.; Cauwenberghs, G. Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 2553–2567. [\[CrossRef\]](#) [\[PubMed\]](#)

31. Burle, B.; Spieser, L.; Roger, C.; Casini, L.; Hasbroucq, T.; Vidal, F. Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *Int. J. Psychophysiol.* **2015**, *97*, 210–220. [[CrossRef](#)]
32. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)]
33. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
34. Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **2019**, *16*, 031001. [[CrossRef](#)]
35. Gómez-Tapia, C.; Bozic, B.; Longo, L. On the Minimal Amount of EEG Data Required for Learning Distinctive Human Features for Task-Dependent Biometric Applications. *Front. Neuroinform.* **2022**, *16*, 844667. [[CrossRef](#)]
36. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning spatiotemporal features with 3D convolutional networks. In Proceedings of the The IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4489–4497.
37. Boashash, B.; Ouelha, S. Automatic signal abnormality detection using time-frequency features and machine learning: A newborn EEG seizure case study. *Knowl.-Based Syst.* **2016**, *106*, 38–50. [[CrossRef](#)]
38. Ang, K.K.; Chin, Z.Y.; Wang, C.; Guan, C.; Zhang, H. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* **2012**, *6*, 39. [[CrossRef](#)]
39. Saha, A.; Minz, V.; Bonela, S.; Sreeja, S.; Chowdhury, R.; Samanta, D. Classification of EEG signals for cognitive load estimation using deep learning architectures. In *Proceedings of the International Conference on Intelligent Human Computer Interaction*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 59–68.
40. Jiménez-Guarneros, M.; Gómez-Gil, P. Custom Domain Adaptation: A new method for cross-subject, EEG-based cognitive load recognition. *IEEE Signal Proc. Lett.* **2020**, *27*, 750–754. [[CrossRef](#)]
41. Qayyum, A.; Khan, M.A.; Mazher, M.; Suresh, M. Classification of eeg learning and resting states using 1d-convolutional neural network for cognitive load assesment. In Proceedings of the 2018 IEEE Student Conference on Research and Development (SCoREd), Selangor, Malaysia, 26–28 November 2018; pp. 1–5.
42. Bashivan, P.; Yeasin, M.; Bidelman, G.M. Single trial prediction of normal and excessive cognitive load through EEG feature fusion. In Proceedings of the 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 12 December 2015; pp. 1–5.
43. Liu, Y.; Liu, Q. Convolutional neural networks with large-margin softmax loss function for cognitive load recognition. In Proceedings of the 2017 36th Chinese control conference (CCC), Dalian, China, 26–28 July 2017; pp. 4045–4049.
44. Jiao, Z.; Gao, X.; Wang, Y.; Li, J.; Xu, H. Deep convolutional neural networks for mental load classification based on EEG data. *Pattern Recogn.* **2018**, *76*, 582–595. [[CrossRef](#)]
45. Xiong, R.; Kong, F.; Yang, X.; Liu, G.; Wen, W. Pattern recognition of cognitive load using eeg and ecg signals. *Sensors* **2020**, *20*, 5122. [[CrossRef](#)] [[PubMed](#)]
46. Cabañero, L.; Hervás, R.; González, I.; Fontecha, J.; Mondéjar, T.; Bravo, J. Analysis of cognitive load using EEG when interacting with mobile devices. *Multidiscip. Digit. Publ. Inst. Proc.* **2019**, *31*, 70.
47. Yin, Z.; Zhang, J. Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. *Biomed. Signal Proc. Control* **2017**, *33*, 30–47. [[CrossRef](#)]
48. Yang, S.; Yin, Z.; Wang, Y.; Zhang, W.; Wang, Y.; Zhang, J. Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders. *Comput. Biol. Med.* **2019**, *109*, 159–170. [[CrossRef](#)] [[PubMed](#)]
49. Yin, Z.; Zhao, M.; Zhang, W.; Wang, Y.; Wang, Y.; Zhang, J. Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework. *Neurocomputing* **2019**, *347*, 212–229. [[CrossRef](#)]
50. Stipacek, A.; Grabner, R.; Neuper, C.; Fink, A.; Neubauer, A. Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load. *Neurosci. Lett.* **2003**, *353*, 193–196. [[CrossRef](#)] [[PubMed](#)]
51. Castro-Meneses, L.J.; Kruger, J.L.; Doherty, S. Validating theta power as an objective measure of cognitive load in educational video. *Educ. Technol. Res. Develop.* **2020**, *68*, 181–202. [[CrossRef](#)]
52. Chang, H.C.; Hung, I.C.; Chew, S.W.; Chen, N.S. Yet another objective approach for measuring cognitive load using EEG-based workload. In Proceedings of the 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), Austin, TX, USA, 25–28 July 2016; pp. 501–502.
53. Holm, A.; Lukander, K.; Korpela, J.; Sallinen, M.; Müller, K.M. Estimating brain load from the EEG. *Sci. World J.* **2009**, *9*, 639–651. [[CrossRef](#)]
54. Raufi, B.; Longo, L. An Evaluation of the EEG Alpha-to-Theta and Theta-to-Alpha Band Ratios as Indexes of Mental Workload. *Front. Neuroinform.* **2022**, *16*, 861967. [[CrossRef](#)]
55. Gevins, A.; Smith, M.E. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoret. Issues Ergon. Sci.* **2003**, *4*, 113–131. [[CrossRef](#)]
56. Antonenko, P.; Paas, F.; Grabner, R.; Van Gog, T. Using electroencephalography to measure cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 425–438. [[CrossRef](#)]
57. Jing, L.; Tian, Y. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 4037–4058. [[CrossRef](#)]
58. Banville, H.; Chehab, O.; Hyvärinen, A.; Engemann, D.A.; Gramfort, A. Uncovering the structure of clinical EEG signals with self-supervised learning. *J. Neural Eng.* **2021**, *18*, 046020. [[CrossRef](#)]

59. Ferri, R.; Rundo, F.; Bruni, O.; Terzano, M.G.; Stam, C.J. The functional connectivity of different EEG bands moves towards small-world network organization during sleep. *Clin. Neurophysiol.* **2008**, *119*, 2026–2036. [[CrossRef](#)] [[PubMed](#)]
60. Landi, G. Properties of the center of gravity as an algorithm for position measurements. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* **2002**, *485*, 698–719. [[CrossRef](#)]
61. Pop-Jordanova, N.; Pop-Jordanov, J. Spectrum-weighted EEG frequency (“brain-rate”) as a quantitative indicator of mental arousal. *Prilozi* **2005**, *26*, 35–42. [[PubMed](#)]
62. Ryang, H.; Yun, U. High utility pattern mining over data streams with sliding window technique. *Exp. Syst. Appl.* **2016**, *57*, 214–231. [[CrossRef](#)]
63. Mann, S. Cubic precision clough-tocher interpolation. *Comput. Aided Geom. Design* **1999**, *16*, 85–88. [[CrossRef](#)]
64. Yue-Hei Ng, J.; Hausknecht, M.; Vijayanarasimhan, S.; Vinyals, O.; Monga, R.; Toderici, G. Beyond short snippets: Deep networks for video classification. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4694–4702.
65. Wang, L.; Li, W.; Li, W.; Van Gool, L. Appearance-and-relation networks for video classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1430–1439.
66. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
67. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
68. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
69. Bashivan, P.; Rish, I.; Heisig, S. Mental State Recognition via Wearable EEG. In Proceedings of the 5th NIPS Workshop on Machine Learning and Interpretation in Neuroimaging (MLINI15), Montreal, QC, Canada, 11–12 December 2015; pp. 1–5.
70. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [[CrossRef](#)]
71. Plass, J.L.; Kalyuga, S. Four ways of considering emotion in cognitive load theory. *Educ. Psychol. Rev.* **2019**, *31*, 339–359. [[CrossRef](#)]
72. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
73. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv* **2012**, arXiv:1207.0580.
74. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Proc. Syst.* **2012**, *60*, 84–90. [[CrossRef](#)]
75. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
76. Longo, L.; Goebel, R.; Lécué, F.; Kieseberg, P.; Holzinger, A. Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions. In Proceedings of the Machine Learning and Knowledge Extraction—4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, 25–28 August 2020; pp. 1–16. [[CrossRef](#)]
77. Vilone, G.; Longo, L. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Inform. Fusion* **2021**, *76*, 89–106. [[CrossRef](#)]
78. Vilone, G.; Longo, L. Classification of Explainable Artificial Intelligence Methods through Their Output Formats. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 615–661. [[CrossRef](#)]
79. Popper, K. *The Logic of Scientific Discovery*; Routledge: London, UK, 2005.
80. Michel, C.M.; Koenig, T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *NeuroImage* **2018**, *180*, 577–593. [[CrossRef](#)] [[PubMed](#)]
81. Wackermann, J.; Lehmann, D.; Michel, C.; Strik, W. Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. *Int. J. Psychophysiol.* **1993**, *14*, 269–283. [[CrossRef](#)]
82. Khanna, A.; Pascual-Leone, A.; Michel, C.M.; Farzan, F. Microstates in resting-state EEG: Current status and future directions. *Neurosci. Biobehav. Rev.* **2015**, *49*, 105–113. [[CrossRef](#)]
83. Longo, L.; Rizzo, L.; Dondio, P. Examining the modelling capabilities of defeasible argumentation and non-monotonic fuzzy reasoning. *Knowl.-Based Syst.* **2021**, *211*, 106514. [[CrossRef](#)]