

Article

Applying Neural Networks with Time-Frequency Features for the Detection of Mental Fatigue

Ioannis Zorzos ¹, Ioannis Kakkos ^{1,2}, Stavros T. Miloulis ¹, Athanasios Anastasiou ¹, Errikos M. Ventouras ² and George K. Matsopoulos ^{1,*}

¹ Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, 9, Iroon Polytechniou Street, Zografos, 15780 Athens, Greece

² Department of Biomedical Engineering, University of West Attica, 17, Ag. Spyridonos Street, Egaleo, 12243 Athens, Greece

* Correspondence: gmatso@biomed.ntua.gr; Tel.: +30-210772288

Abstract: The detection of mental fatigue is an important issue in the nascent field of neuroergonomics. Although machine learning approaches and especially deep learning designs have constantly demonstrated their efficiency to automatically detect critical features from raw data, the computational resources for training and predictions are usually very demanding. In this work, we propose a shallow convolutional neural network, with three convolutional layers, for fatigue detection using electroencephalogram (EEG) data that can alleviate the computational burden and provide fast mental fatigue detection. As such, a deep learning model was created utilizing time-frequency domain features, extracted with Morlet wavelet analysis. These features, combined with the higher-level characteristics learnt by the model, resulted in a resilient solution, able to attain very high prediction accuracy (97%), while reducing training time and computing costs. Moreover, by incorporating a subsequent SHAP values analysis on the characteristics that contributed in the model creation, indications of low frequency (theta and alpha band) brain wave characteristics were indicated as prominent mental fatigue detectors.

Keywords: deep learning; CNN; fatigue; Morlet wavelet; EEG; time-frequency analysis



Citation: Zorzos, I.; Kakkos, I.; Miloulis, S.T.; Anastasiou, A.; Ventouras, E.M.; Matsopoulos, G.K. Applying Neural Networks with Time-Frequency Features for the Detection of Mental Fatigue. *Appl. Sci.* **2023**, *13*, 1512. <https://doi.org/10.3390/app13031512>

Academic Editor: Habib Hamam

Received: 14 December 2022

Revised: 5 January 2023

Accepted: 11 January 2023

Published: 23 January 2023



Copyright: © 2023 by the authors. 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

Sleep deprivation comprises a state provoked by either a complete lack of sleep or a suboptimal duration of rest. Evidence-based studies have demonstrated a significant impact of sleep deprivation on information processing and overall cognitive performance [1], with its effects varying based on the cognitive requirements and the brain areas involved. Under this state, mental fatigue emerges as a key outcome, resulting in concentration difficulties and low vigilance. Consequences are particularly evident in the context of working memory, which by definition entails rapid information manipulation and task execution [2]. As such, mental fatigue is associated with compromised performance from a working memory perspective, which often proves crucial in specialized settings, such as healthcare institutions. Specifically, hospital staff typically undergo prolonged periods of demanding cognitive activity during their shifts, presenting increased probability of burnout, which in turn poses a potential risk on patient safety [3]. In such circumstances, prompt identification of mental fatigue levels that may induce declined performance and related errors in hospital settings is essential.

In order to assess mental fatigue states, several electrophysiological recording modalities have been utilized, providing indications of the complex brain interactions and therefore insights of the cognitive mechanisms governing cognitive exhaustion [4,5]. In this view, electroencephalography (EEG) has been showcased as an invaluable apparatus due to its ability to detect brain activity (non-invasively) in the form of neuronal oscillations, enabling researchers to quantify the impact on cognitive functions under various levels

of mental load [6]. The main classification problem lies on the state of the individual being mentally rested (herein referred to as Rested class) or mentally fatigued (Fatigued class). Most frequently, signal processing techniques are employed in order to develop mental fatigue detection algorithms. These techniques involve the use of mathematical algorithms to analyze and extract relevant information from EEG signals, without the use of data-driven development. For instance, several studies include wavelet transformation and independent component analysis to analyze EEG data and identify changes in brain activity that are indicative of mental fatigue [7,8]. However, these approaches require manual intervention and refactoring that might deter a fully automated detection of mental fatigue. Another widely used approach is machine learning, which involves the use of statistical models to analyze and interpret (EEG) data in order to identify patterns and features (e.g., power spectrum) that are indicative of mental fatigue [9]. However, such techniques require manual feature extraction in order to perform optimally, while the extracted features are usually subject dependent [9] and thus not ideal for real-world and real-time paradigms.

On this premise, artificial intelligence (AI) has become a powerful tool, with deep learning (DL) designs delivering automatic feature detection and extraction, unleashing outstanding advances in classification, segmentation, and prediction tasks. As such, many in fatigue detection studies have utilized DL algorithms due to their high classification performance (compared to conventional machine learning methods), especially when utilizing solely EEG signals (disregarding other modalities, such as electrooculograms or facial video data) [10,11]. Furthermore, in pursuit of metrics with high discriminative power, even more complex architectures have been proposed in literature, with convolutional layers and recursive neural networks achieving promising results, however at the cost of increasing complexity [12]. In this regard, most of these implementations use multi-layered deep convolutional neural networks (CNNs) that require high-end hardware and a substantial amount of time for both training and prediction phases. Although, such configurations usually attain high classification accuracy, the overall computational cost deems them inapplicable for real-world application (e.g., brain-computer interfaces), since minimal computational resources and prediction time are essential in order to achieve real-time functionality.

Another important factor that presents a limitation in the current DL architectures is their explainability. DL entails the need for enormous, multi-layered networks that are hard to train and even harder to explain. The models are perceived as black boxes with little understanding of the actual system usability [13]. Explainability is crucial, since it increases the trustworthiness of the system. Understanding the reason behind a model's predictions can provide confidence in the model's results, while providing indications of the flow of reasoning behind the results produced.

Taking the above into consideration, a DL architecture was employed in this study, which utilized EEG data from a working memory task triggering visual memory processes. The overall procedures included two recordings (before and after on-call shifts) of medical and nursing staff eliciting high mental fatigue due to the influence of sleep deprivation. In this regard, a fairly shallow neural network architecture (with three layers of cascading CNNs and a fully connected layer) was applied, based on non-invasive EEG recordings and an appropriate data (pre-)analysis. Our results indicate the efficiency of the proposed methodology towards both classification performance and computational cost. In detail, our workflow achieved an accuracy of 97% (97% F-score) within a particularly cost-effective setup, bearing notable implementation prospects for real-time applications. Moreover, the application of an explainability design was able to provide indication of the input brain activity features utilized by the DL algorithm to assess mental fatigue. The main contributions of this study are: (a) to effectively classify fatigue states with high overall accuracy while employing an explainable and trustworthy AI architecture; (b) to emphasize the importance of hand-engineered features in the performance of deep learning models.

2. Materials and Methods

2.1. Participants

This study included 22 healthy participants (9 female, mean age 27.3 ± 4.1) recruited from the 401 General Military Hospital of Athens. All participants were doctors and staff members, reporting normal or corrected-to-normal vision. Before the experiment, all participants were prescreened to ensure no sleep disorder, no history of any mental disease or ADHD and no long-term medication intake. The experiment was approved by the Institution's Review Board in accordance with the Declaration of Helsinki, while written informed consent was obtained from all subjects.

2.2. Experimental Design

The participants underwent a working memory task twice, before on-call shifts (up to 28 h with little or no sleep—Rested) and after work shift ended (Fatigued). As such, to assess the influence of mental fatigue on the working memory capacity, a visual n-back task was employed under an $N = 2$ design [14]. During this task, the participants were asked to remember and compare the image displayed in 2 trials prior to the current visual stimulus. In each trial, an image would appear in one of the four corners of the screen, with the participants indicating among 1 of the 4 possible conditions by pressing the corresponding button. Each condition required comparisons with regards to the image content and location as follows: (a) same image and same location; (b) same image (different location); (c) same location (different image); and (d) no similarities (different image, different location). The experiment consisted of 72 tests (4 conditions balanced) and lasted approximately 5 min. The visual stimuli were displayed for 3.5 s, interposing a fixed cross for 1 s. Practice trials were conducted before the EEG recordings to ensure individuals' comprehension regarding the execution of the task.

2.3. Data Acquisition and Preprocessing

EEG data recordings were performed using a 64-channel electrode cap (Biosemi, ActiTwo System, Amsterdam, Netherlands), according to the standard 10–20 positioning system at a 512 Hz sampling rate. Bipolar electrooculogram signals were recorded from electrodes placed at the outer canthi, as well as above and below the eyes. The raw EEG data were down sampled to 256 Hz, band-pass filtered from 1 to 40 Hz, and re-referenced to the average of the all electrodes. Additional artifact correction was performed by utilizing Independent Component Analysis (ICA) and rejecting the components highly correlated with the electrooculogram signals [15]. Due to significant artifact contamination, data from 2 participants were excluded, resulting in 20 subjects for further analysis. Signals were then segmented into trials and adjusted relative to a 100-ms pre-stimulus baseline. An additional removal of the first 0.5 sec stimulus onset data was employed to alleviate influences due to stimuli effects [16], resulting in 18 (per condition) 3-sec trial-based epochs. Electrode Iz was removed from analysis for symmetry reasons, leaving 63 channels for subsequent analysis. In order to negate the contribution of cognitive processing irrelevant to working memory, only correct answers were included in the following analysis, resulting in 2401 samples: 1234 samples of Fatigued data and 1167 of Rested data. Preprocessing was implemented in Matlab 2022b (Mathworks Inc., Natick, MA, USA) using the EEGLAB tool [17].

2.4. Feature Extraction

Despite the trigger-related nature of the data, identification of the mental fatigue states, while omitting the visual stimuli information, was decided (thus the 0.5 sec post-trigger time frame was removed, as mentioned in the previous section). However, such a configuration creates a non-phase-locked dataset. This could be a significant hindrance in time-locked analysis, since the important temporal information of the signal would be ignored and lost. On the contrary, time-frequency (TFR) analysis can harness the temporal and spectral resolution for the non-phase-locked trials [18]. Therefore, TFR analysis was

implemented as the base analysis prior to artificial intelligence learning to address potential data inconsistencies that could prove detrimental for the predictive model performance.

Specifically, the preprocessed data were analyzed by employing complex Morlet wavelet [19] creating the final feature dataset. The utilization of Morlet wavelets was done due to their prominence in time-frequency analysis, as they provide equal variance in time and in frequency [20]. With this approach, the raw power values for each frequency bin were computed with a varied time window (that rely on frequency), using a Gaussian taper with a specified width. The width of the Morlet taper was selected as a hyperparameter with a value of 4, resulting in a temporal resolution of 13 time bins. This analysis outputs a power spectral-temporal activation map to be used as input for the subsequent DL model. As such, a 3D TFR matrix was created (per subject and trial) incorporating the Morlet power features, with dimensions $[C \times F \times T]$, where C is the number of channels, F is the frequency bins, and T is the time bins. The number of channels in the proposed configuration is 63, with 40 frequency bins (from 1–40 Hz, with 1 Hz frequency windows) and 13 time bins (from 0.5 s to 3.5 s with a time window of 0.25 s). Morlet wavelet calculations were performed using the FieldTrip Toolbox [21].

2.5. Artificial Intelligence Modeling

2.5.1. Deep Learning Model

The AI architecture consisted of three 2-dimensional convolutional layers in a sequential order (Figure 1). In the first layer, a 3×3 kernel channel-wise (63 channels) convolution is conducted, with 32 output filters (each filter is producing one output matrix; thus, resulting in 32 first layer matrices). No padding was performed and a standard stride of 1 was used. The kernel size was specifically chosen to incorporate a 3 Hz frequency band and 0.75 sec time band features. The 3 Hz frequency window was selected as it is a wide bandwidth that can correlate and therefore be explained (as it is a good divisor) with the most commonly used EEG bands (delta, theta, alpha, beta, and gamma). As far as the time window duration is concerned, a time window of 750 ms was selected based on similar experimental paradigms (for 2-back working memory tasks) that present the mean reaction time of approximately 700 ms [22]. The second layer is comprised of 16 output filters with a kernel size of 3×3 , also computing channel-wise convolution, capturing higher-level features of the 32 first-layer input matrices. The final convolution layer consists of 4 output filters with a kernel size of 4×4 . Usually in DL architectures, a pooling layer is used after the convolutional layers to down-sample the feature space [23]. However, the reduction of the dimension of the outputs was not required in the proposed approach, and thus no pooling layer was utilized. Since the scope of this study includes the application of AI methods with minimal computational resources, the hyperparameters of the created shallow network were selected for fast execution time per time step during training. Thereafter, the selected number of layers was 3, taking into account the medium size of the dataset. Furthermore, the filter sizes follow a logarithmic pattern, reducing the number of features after each layer to half, thus decreasing the training and the inference time. The optimizer, the learning rate, the batch size, and the activation function were selected based on common DL practices and testing with a variety of values. The optimizer that was selected was Adam [24], the initial learning rate was selected to be 0.001, the batch size was 32 samples, and the sigmoid function was chosen as the activation function.

Finally, a fully connected (FC) layer with 792 input nodes and 2 output nodes for each of the classes was utilized. In order to introduce non-linearity to our model, a bounded logistic function (sigmoid function) was employed as an activation function for the output probabilities. Following each convolutional layer, a batch normalization layer was utilized to increase the stability and robustness of the AI network during the training processes. Furthermore, a dropout configuration with a probability of a zero-element set to 0.25 was used to reduce overfitting and handle uncertainty. Dropout is a regularization technique that can reduce uncertainty in deep learning models by randomly dropping out

neurons during training. All algorithms were implemented with custom code in the Python programming language and the PyTorch, version 1.13 [25] deep learning framework.

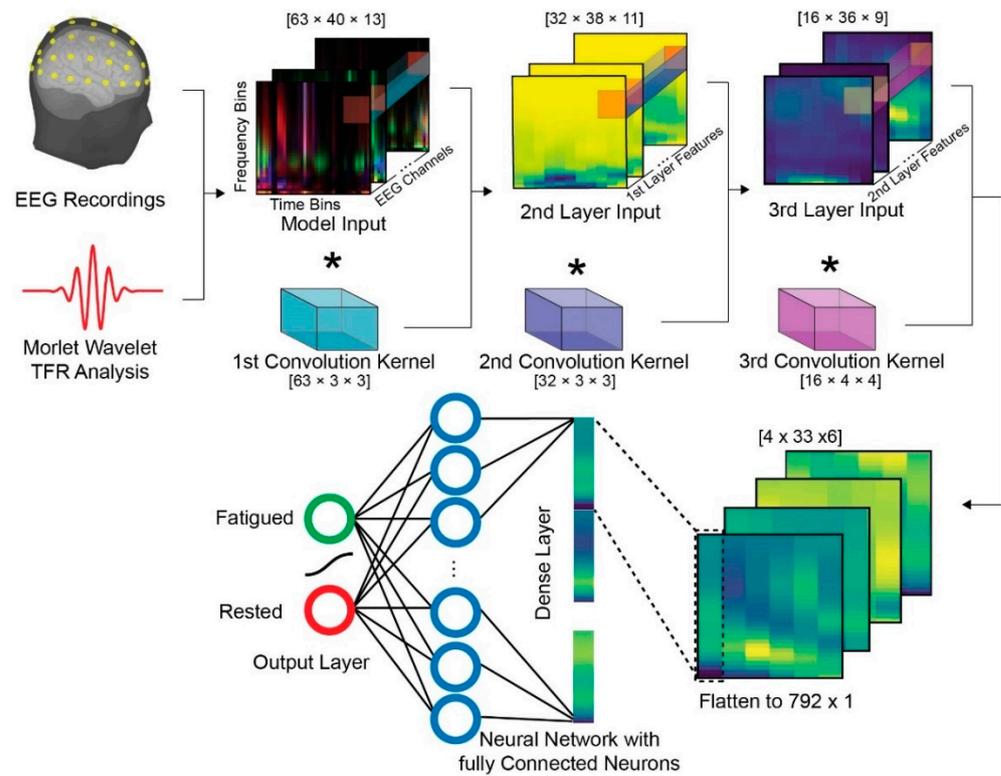


Figure 1. A schematic of the proposed framework. The * symbolizes the convolution function.

2.5.2. Deep Learning Training

The computing resources employed for training the aforementioned model consist of an Intel (R) Core (TM) i7-9700 CPU with 3.00 GHz and 16 GB of 2667 MHz RAM on a Windows 11 Pro OS Version 22H2 (Microsoft, Redmond, Washington, U.S.). No CUDA compatible GPU was utilized, so the AI training was conducted completely on the CPU. The created feature dataset consists of 2401 trials: 1234 samples for the Fatigued class and 1167 for the Rested class. Even though all subjects completed the same number of tasks for each state, only correct responses were considered for the creation of the dataset, introducing a minor class imbalance. However, the differences were negligible and, henceforth, it cannot be suggested that they contributed to the corresponding bias. To train the designed deep learning model, a combined-subject training strategy was employed in order to exploit features that provide generalized (subject independent) information on mental fatigue. As such, all the observations of Fatigued and Rested data points of all the subjects were combined and randomly partitioned into the three data subsets; i.e., the training, the validation and the test subsets. In detail, the full dataset was split into a training subset (containing 75% of the full dataset; 925 samples of Fatigued class and 875 of Rested class), a validation set (containing 15% of the full dataset; 185 of Fatigued class and 175 of Rested) and a test set (containing 10% of the full dataset; 124 of Fatigued class and 117 of Rested). Assignment was performed randomly to avoid any selection bias among participants. The neural network model was trained repeatedly using the training dataset for a specific number of epochs. Each epoch (which is a hyperparameter of the training process) is a complete pass of the entire training dataset from the learning algorithm. This is required for a model until the learning process converges. Usually, the number of epochs is in hundreds or thousands. However, our model converged in almost 150 epochs.

3. Results

3.1. Performance

The overall AI model performance results are presented below (Table 1, Figure 2), as measured with four standard machine learning metrics: accuracy, precision, recall and F-score. As such, the attained classification accuracy was 97%, with an additional very high F-score (reaching 97%), thus indicating that mental fatigue was identified effectively (the confusion matrix of the AI model results is presented in Figure 3a). For the 150 epochs of our model, training was completed in 44 min. Figure 3b presents the validation history in the form of validation accuracy over the course of epochs. Moreover, the estimated processing time for the approximated prediction time of a single instance was approximately 4 ms (following the suggested methodology and the computing resources setup described above).

Table 1. The performance of the proposed model.

	Precision	Recall	F-Score	Support (No. of Samples)
Fatigued	98%	97%	97%	124
Rested	97%	97%	97%	117
Overall Accuracy	-	-	97%	241

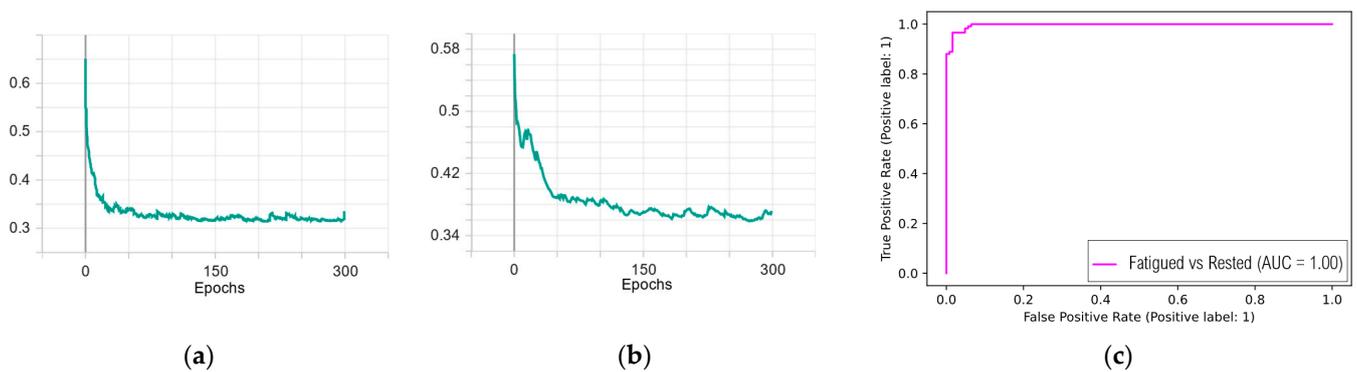


Figure 2. The proposed model performance regarding: (a) train loss over the epochs; (b) validation loss over the epochs; (c) receiver operating characteristic (ROC) curve.

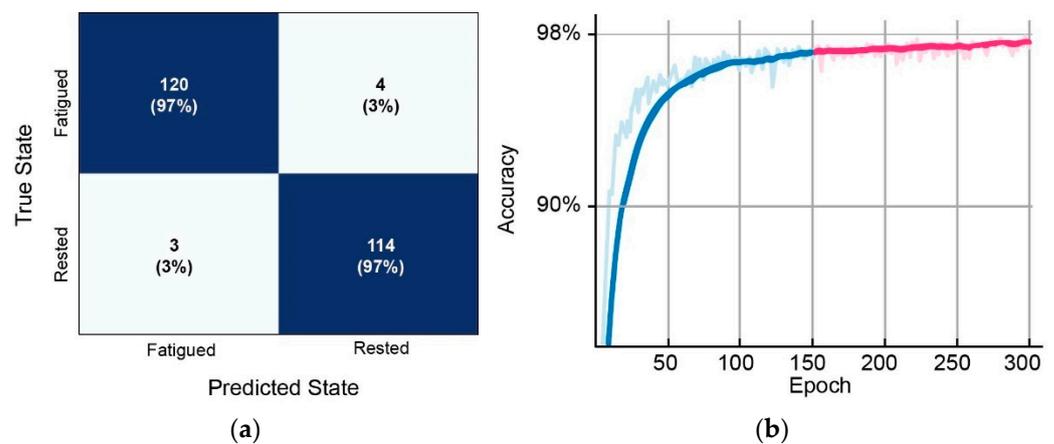


Figure 3. (a) Proposed model confusion matrix; (b) validation curve for 150 epochs (blue) and 300 epochs (magenta). Darker colors represent a smoother curve.

To evaluate whether the resulting performance could be further improved, an additional 300 epoch model was evaluated. However, despite the fact that in a separate training session the overall accuracy increased to 97.6 ($\approx 0.1\%$ increase) the test dataset required double the amount of training time in the same computational setup. As such, the extra

time could be considered unnecessary since the validation accuracy reaches a plateau and the model converges (Figure 3b).

In AI-applications, the computational cost is an important factor in order to balance the performance and the time needed for the model solution. In this regard, the requirements of the proposed model can be considered minimal, both in terms of computational power and in the time required for training and evaluating the AI model. Specifically, as described above, the model was trained in 44 min with no CUDA cores, while each consequent prediction was performed in 4 ms. On this premise, to provide an indication of the computational and the temporal efficiency of the proposed configuration, a thorough comparison in the same computation settings was conducted. As such, various state-of-the-art AI models that examine mental fatigue classification were recreated and trained with the same computational setup, with the algorithmic cost-efficiency evaluation including the time required for training and testing (Table 2). Specifically, the recreated AI models were fed with the same input data as the proposed approach and the duration of training and prediction time were measured. The models included: (a) a dual convolutional neural network (CNN) [11]; (b) a 1-dimensional u-net combined with a long short term memory (LSTM) network [12]; and (c) a modified principal component analysis network (PCANet) with a support vector machine (SVM) classifier [26].

Table 2. Algorithmic Cost-Efficiency Comparison.

AI Model	Authors	Training Time Required	Prediction Time
UNET(CNN) + LSTM	[12]	281 min	25 ms
Dual CNN	[11]	112 min	12 ms
PCANet + SVM	[26]	221 min	45 ms
This study (TF + CNN)	-	44 min	4 ms

The input of each of these models was a modification of the dataset described in Section 2.4 in order to comply with the parameters and designs provided by the studies' authors. In detail the dual CNN and UNET(CNN) + LSTM models require single channel time-series and, thus, a single channel time-series was provided. Similarly, the input for the PCANet was a $C \times r$ matrix, where C is the number of channels and r is the number of the PCA components of the reduced time-series data. The aforementioned AI designs were recreated using PyTorch to minimize any implementation differences. Interestingly, even though dual CNN and UNET(CNN) + LSTM require only a single channel as input, the amount of time required for training is several times the amount required for the proposed AI model.

Of note is that the additional information regarding the algorithmic performance was excluded from the cost-efficiency evaluation, since the focus of this paper is to assess the performance of fairly shallow neural network architectures based on appropriate data handling, and it is not a comparison to existing algorithmic designs. However, the performance of the additional AI models is included in Section 4, denoting their high discriminative ability (as presented in their respective studies).

3.2. Explainability

One of the major issues concerning DL AI designs is their innate "black box" nature, making it difficult to estimate the relationships between the input features, which provides reasoning behind the results that are computed. In this view, the colossal number of layers that are contained in the majority of DL architectures hinders the importance of each feature that is internally computed (as each layer transforms the input data). As a result, the interpretability and, therefore, the trustworthiness of such models can be compromised. Shallow artificial networks alleviate these concerns, since the relatively small number of layers allows the decisions incorporated in each step of the model to be estimated and thus explained. In this regard, explainability tools, such as Shapley additive explanations (SHAP), values provide a visualization of the internal model processes [27],

highlighting the importance of each region or feature. In this study, SHAP values estimation was performed by utilizing an enhanced method of DeepLift (deep learning important features) [28] with the DeepExplainer Tool. DeepExplainer uses a sampling technique to approximate the Shapley values for a DL model. Specifically, a large number of random samples is generated from the model's input space and using the resulting samples to estimate the Shapley values. In this study, 100 samples were utilized to generate the SHAP values. For each individual input/observation, the SHAP values are computed for each of the output classes, with the positive values indicating the contribution in model selection of the correspondent class, while the negative values denote the features inhibitory nature. In this study, SHAP values have been employed to identify what the model relies on (increasing the probability that the input feature map is the predicted class). An example regarding the positive and negative SHAP values observation is presented in Figure 4. Interestingly, in the majority of SHAP values computation, the range between 5–15 Hz is contributing positively in predicting the Fatigued class. In a similar way, activation in the range of 3–7 Hz on time-frequency map suggests an inhibitory factor for the Fatigued class and positive values for the Rested class. The importance of visually interpreting SHAP values comes along with visually inspecting TFR power maps. Visually interpreting TFR power maps is an invaluable tool for neuroscientists, facilitating the reasoning behind the AI model predictions and shedding light to the understanding of brain activity, EEG states, conditions, and mental illness [29].

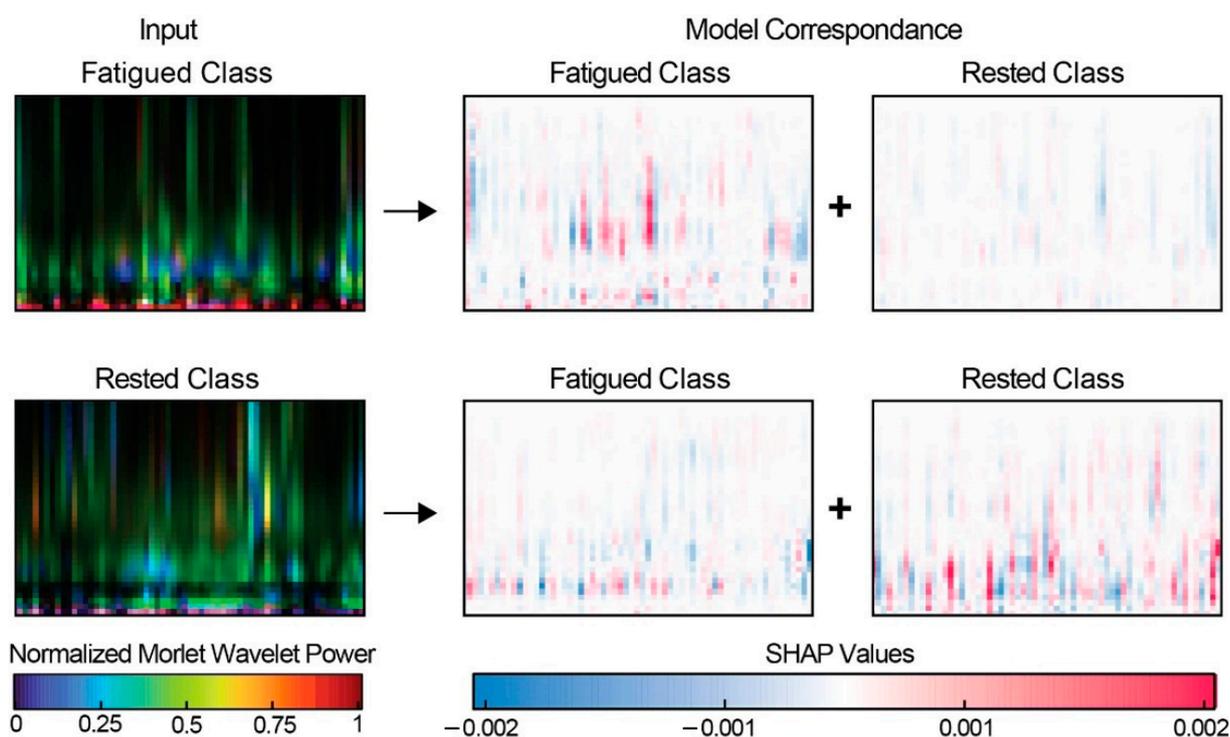


Figure 4. A paradigm of SHAP values representation. The inputs denote the TFR matrices, while the model correspondence reflects the effects of the colored regions in each specific class output. The upper row corresponds to the Fatigued class, while the lower row to the Rested class.

4. Discussion

In this study, a fairly shallow neural network to discriminate between the fatigued and rested state was utilized and tested, in an EEG working memory experiment. The proposed architecture was able to achieve high classification accuracy (97%), within a computationally inexpensive framework, emphasizing the utility of the data analysis required for real-time/real-world applications.

In many cases, such as computer vision, hidden features extracted by DL architectures were more efficient in certain tasks than the features extracted manually, which explains why DL approaches outperform the ones with machine learning. However, this does not hold in all tasks. Although, in some cases the utilization of raw EEG data with CNNs attains very high accuracy [11], supplying the network with analyzed data for EEG signal classification proves to be the optimal pathway. On the other hand, hand-crafted features (such as those proposed in this study) are usually employed with classical machine learning techniques, since the automatic feature extraction (the main advantage of deep learning architectures) is not required. This being the case, computation of higher level-features presents an inherent advantage, since the proposed model outperforms classical ML techniques in regards to the F-score [30].

Another important factor in machine learning approaches is the training strategies applied for the AI model creation. Typically, a subject-specific approach has been utilized by the majority of similar studies [12,26]. The subject-specific strategy includes the training and testing sets to be split for each subject separately, creating N results (where N is the number of subjects). The final model accuracy is calculated by the averaged result of all the subjects. However, although this subject-specific strategy usually yields high performance (since AI models are capable of capturing personalized features), this approach is not ideal for identifying generalized features that are applicable, albeit subject variations. In this study, a high universal (subject-independent) performance and a combined-subject training strategy was applied. As such, by including samples from all the subjects into our training dataset, the potential sampling bias in the data is minimized, making the AI model training more robust and less error prone [31].

To illustrate the prominence of the hand-crafted feature DL design, an overall comparison between the state-of-the-art AI methods used in the literature and the proposed methodology is presented in Table 3. Of note is that the comparison includes studies in which different datasets and analysis are applied in each method. As such, each of the models presented tackled different problems, domains, and feature spaces. It should be emphasized that each model's performance not only based on methodological architecture but also on data availability, quality, and variability. However, the overall results imply that the proposed framework provides high results in mental fatigue discrimination (under a generalized training procedure), comparable to those presented in literature.

Table 3. Mental fatigue classification methods comparison.

AI Model	Authors	Dataset (Training Strategy)	Classification Accuracy
UNET(CNN) + LSTM	[12]	9 subjects with 1 channel, 4 s segments (Subject-specific)	83%
SVM	[32]	8 subjects with 19 channels, 10 s segments (Combined-subject)	89%
Dual CNN	[11]	22 subjects with 1 channel, 1 s segments (Combined-subject)	93%
PCANet + SVM	[26]	6 subjects with 32 channels, 4 s segments (Subject-specific)	96%
This study (TF + CNN)	-	22 subjects with 64 channels, 3 s segments (Combined-subject)	97%

Equally important is the computational cost of the DL model employed. In this work, no CUDA compatible GPU was utilized, with the AI model being developed on the CPU. This is an important factor, since for a real-time apparatus, the presence of CUDA compatible GPUs is scarce. On this note, it is safe to infer that the AI model would perform equally well on an advanced microprocessor without the need of GPU cores. In detail, the TFR analysis for each subject was completed in under 2 ms, while the prediction time of a single instance was approximately 4 ms, making the system suitable for real-time classifications and appropriate for EDGE-AI systems (where training and inference is performed with minimal computational power) [33]. On the contrary, applications of different DL architectures (as presented in Table 2) or input calculation processes leads to increased (compared to the proposed framework) computational costs. For instance, the computation of PCA for the input requires more than 10× the time compared to TFR, exceeding 20 ms for each of the sample [26]. Based on the above, the presented approach

could be considered ideal for real-time fatigue detection with possible extensions to brain computer interfaces (BCI).

In addition to the performance of the proposed framework, the features that contributed most to the formation of AI model were inspected. In this regard, the calculated SHAP positive values highlight the TFR power in the 5–15 Hz bins as a rigorous indicator of mental fatigue development. In fact, previous similar studies have suggested that increased task demands in working memory paradigms can be reflected as theta (4–7 Hz) and alpha (8–12 Hz) neural oscillations [34,35]. Specifically, alterations in theta power have been consistently related to different states of fatigue level, and especially in memory manipulation [14,36]. In a similar way, visual attention and working memory aspects systematically report alpha activity alterations [37,38]. Although the fatigue state was elicited as a result of sleep deprivation rather than the experimental procedure, the information maintenance and item retrieval requirements of the n-back task involves working memory capacity and high attentional demands [35]. This is also reported in beta band (13–30 Hz) activity [39,40], although the very small overlap of our findings with the beta frequency range cannot provide a conclusive assumption. Regarding the SHAP negative values, the inhibitory elements indicated in the 3–7 Hz time-frequency map suggest theta wave memory load interactions, while implying task-related reactive control (as reactive control is less cognitive demanding than proactive control) [41,42].

Despite the fact that the proposed framework displayed an overall high performance, interpretation of the methods applied should be treated with caution. The main concern is that the dataset under study incorporated a small number of individual observations (20 subjects), which is deemed a moderate sample for DL designs. Nevertheless, it is similar (or larger) than most relevant similar studies [43]. Another possible limitation of this work is the definition of the Fatigued state and the interpretation of it as the ground truth. However, since the data comes from on-call doctors and nursing staff, it is safe to infer that mental fatigue is present after the shift. Hence, it can be assumed that there is no subjectivity bias in the current research.

Considering the results of this study, in the future we intent to expand the overall methodologies, both in terms of the experimental and methodological design. As such, the development of deep learning approaches for mental fatigue detection using EEG signals could be enhanced by including larger datasets (containing a variety of population and settings, subsequently increasing diversity/reducing population biases), cognitive load measures and behavioral observations, and optimization in the DL architecture. These, paired with explainability methods, are likely to yield valuable insights and improvements in the ability to detect and comprehensively investigate the underlying neural substrates that govern mental fatigue.

5. Conclusions

The use of deep learning approaches for mental fatigue detection is an active and promising area of research with the potential to have significant impacts on the detection and management of mental fatigue in various settings. In this paper, a shallow neural network architecture with EEG-derived hand-crafted features to discriminate between fatigued (sleep deprived) and rested states in a working memory task was created. Our model provided high overall results (97% classification accuracy), using computationally inexpensive and fast-to-train network properties. Furthermore, the data-driven SHAP analysis illustrated the distinct aspects of TFR Morlet power values between the two classes in 5–15 Hz and 3–7 Hz frequency ranges. Our results indicate that appropriate data (pre-)analysis with a carefully executed AI model design can achieve high performance, providing a progressive step forward towards the real-time mental fatigue detection.

The small number of individual (20 subjects) could be a factor towards population bias, while the lack of behavioral observations could lead to a small number of misclassifications, however we intend to expand the experimental and methodological aspects described in this paper to address these limitations, with the intention of further increasing mental

fatigue detection accuracy (with minimal computational costs) and shedding light on the mental fatigue cognitive mechanisms.

Author Contributions: Conceptualization, I.Z. and I.K.; methodology, I.Z. and A.A.; software, I.Z.; validation, S.T.M., A.A. and I.K.; formal analysis, E.M.V.; investigation, G.K.M. and E.M.V.; resources, A.A., I.K. and S.T.M.; data curation, I.Z., I.K. and S.T.M.; writing—original draft preparation, I.Z.; writing—review and editing, I.K. and S.T.M.; visualization, I.K.; supervision, G.K.M.; project administration, G.K.M. and E.M.V.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of 401 General Military Hospital of Athens (protocol code 10598:19/06/17).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank for Nikolaos Fakas, and Dimitrios Karathanasis, (401 General Military Hospital of Athens) for their assistance in the data collection and experimental design.

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

References

1. Slimani, M.; Znazen, H.; Bragazzi, N.L.; Zguira, M.S.; Tod, D. The Effect of Mental Fatigue on Cognitive and Aerobic Performance in Adolescent Active Endurance Athletes: Insights from a Randomized Counterbalanced, Cross-Over Trial. *J. Clin. Med.* **2018**, *7*, 510. [\[CrossRef\]](#)
2. Pergher, V.; Vanbilsen, N.; Van Hulle, M. The Effect of Mental Fatigue and Gender on Working Memory Performance during Repeated Practice by Young and Older Adults. *Neural Plast.* **2021**, *2021*, e6612805. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Welp, A.; Meier, L.L.; Manser, T. Emotional Exhaustion and Workload Predict Clinician-Rated and Objective Patient Safety. *Front. Psychol.* **2015**, *5*, 1573. [\[CrossRef\]](#)
4. Boksem, M.A.S.; Meijman, T.F.; Lorist, M.M. Effects of Mental Fatigue on Attention: An ERP Study. *Cogn. Brain Res.* **2005**, *25*, 107–116. [\[CrossRef\]](#)
5. DeLuca, J.; Genova, H.M.; Hillary, F.G.; Wylie, G. Neural Correlates of Cognitive Fatigue in Multiple Sclerosis Using Functional MRI. *J. Neurol. Sci.* **2008**, *270*, 28–39. [\[CrossRef\]](#)
6. Li, G.; Huang, S.; Xu, W.; Jiao, W.; Jiang, Y.; Gao, Z.; Zhang, J. The Impact of Mental Fatigue on Brain Activity: A Comparative Study Both in Resting State and Task State Using EEG. *BMC Neurosci.* **2020**, *21*, 20. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Shou, G.; Ding, L.; Dasari, D. Probing Neural Activations from Continuous EEG in a Real-World Task: Time-Frequency Independent Component Analysis. *J. Neurosci. Methods* **2012**, *209*, 22–34. [\[CrossRef\]](#)
8. Dasari, D.; Shou, G.; Ding, L. ICA-Derived EEG Correlates to Mental Fatigue, Effort, and Workload in a Realistically Simulated Air Traffic Control Task. *Front. Neurosci.* **2017**, *11*, 297. [\[CrossRef\]](#)
9. Liu, Y.; Lan, Z.; Khoo, H.H.G.; Li, K.H.H.; Sourina, O.; Mueller-Wittig, W. EEG-Based Evaluation of Mental Fatigue Using Machine Learning Algorithms. In Proceedings of the 2018 International Conference on Cyberworlds (CW), Singapore, 3–5 October 2018; pp. 276–279.
10. Gao, Z.; Wang, X.; Yang, Y.; Mu, C.; Cai, Q.; Dang, W.; Zuo, S. EEG-Based Spatio-Temporal Convolutional Neural Network for Driver Fatigue Evaluation. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2755–2763. [\[CrossRef\]](#)
11. Balam, V.P.; Sameer, V.U.; Chinara, S. Automated Classification System for Drowsiness Detection Using Convolutional Neural Network and Electroencephalogram. *IET Intell. Transp. Syst.* **2021**, *15*, 514–524. [\[CrossRef\]](#)
12. Khessiba, S.; Blaiech, A.G.; Ben Khalifa, K.; Ben Abdallah, A.; Bedoui, M.H. Innovative Deep Learning Models for EEG-Based Vigilance Detection. *Neural Comput. Appl.* **2021**, *33*, 6921–6937. [\[CrossRef\]](#)
13. Zhang, Q.; Zhu, S. Visual Interpretability for Deep Learning: A Survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27–39. [\[CrossRef\]](#)
14. Kakkos, I.; Dimitrakopoulos, G.N.; Sun, Y.; Yuan, J.; Matsopoulos, G.K.; Bezerianos, A.; Sun, Y. EEG Fingerprints of Task-Independent Mental Workload Discrimination. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3824–3833. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Dimitrakopoulos, G.N.; Kakkos, I.; Thakor, N.V.; Bezerianos, A.; Sun, Y. A Mental Fatigue Index Based on Regression Using Multiband EEG Features with Application in Simulated Driving. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2017**, *2017*, 3220–3223. [\[CrossRef\]](#)

16. Miloulis, S.T.; Kakkos, I.; Karampasi, A.; Zorzos, I.; Ventouras, E.-C.; Matsopoulos, G.K.; Asvestas, P.; Kalatzis, I. Stimulus Effects on Subject-Specific BCI Classification Training Using Motor Imagery. In Proceedings of the 2021 International Conference on e-Health and Bioengineering (EHB), Iasi, Romania, 18–19 November 2021; pp. 1–4.
17. Delorme, A.; Makeig, S. EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef]
18. Morales, S.; Bowers, M.E. Time-Frequency Analysis Methods and Their Application in Developmental EEG Data. *Dev. Cogn. Neurosci.* **2022**, *54*, 101067. [CrossRef]
19. Kiebel, S.; Kilner, J.; Friston, K. CHAPTER 16—Hierarchical Models for EEG and MEG. In *Statistical Parametric Mapping*; Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W., Eds.; Academic Press: London, UK, 2007; pp. 211–220. ISBN 978-0-12-372560-8.
20. D’Avanzo, C.; Tarantino, V.; Bisiacchi, P.; Sparacino, G. A Wavelet Methodology for EEG Time-Frequency Analysis in a Time Discrimination Task. *Int. J. Bioelectromagn.* **2009**, *11*, 185–188.
21. Oostenveld, R.; Fries, P.; Maris, E.; Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2010**, *2011*, e156869. [CrossRef]
22. Tsoneva, T.; Baldo, D.; Lema, V.; Garcia-Molina, G. EEG-Rhythm Dynamics during a 2-Back Working Memory Task and Performance. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 3828–3831.
23. Sellat, Q.; Bisoy, S.K.; Priyadarshini, R. Chapter 10—Semantic Segmentation for Self-Driving Cars Using Deep Learning: A Survey. In *Cognitive Big Data Intelligence with a Metaheuristic Approach*; Mishra, S., Tripathy, H.K., Mallick, P.K., Sangaiah, A.K., Chae, G.-S., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 211–238. ISBN 978-0-323-85117-6.
24. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization 2014. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 18 December 2022).
25. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035.
26. Ma, Y.; Chen, B.; Li, R.; Wang, C.; Wang, J.; She, Q.; Luo, Z.; Zhang, Y. Driving Fatigue Detection from EEG Using a Modified PCANet Method. *Comput. Intell. Neurosci.* **2019**, *2019*, e4721863. [CrossRef]
27. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Nice, France, 2017; Volume 30.
28. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning Important Features Through Propagating Activation Differences. In Proceedings of the 34th International Conference on Machine Learning, PMLR, Sydney, Australia, 17 July 2017; pp. 3145–3153.
29. Roach, B.J.; Mathalon, D.H. Event-Related EEG Time-Frequency Analysis: An Overview of Measures and An Analysis of Early Gamma Band Phase Locking in Schizophrenia. *Schizophr. Bull.* **2008**, *34*, 907–926. [CrossRef]
30. Venkata Phanikrishna, B.; Jaya Prakash, A.; Suchismitha, C. Deep Review of Machine Learning Techniques on Detection of Drowsiness Using EEG Signal. *IETE J. Res.* **2021**, 1–16. [CrossRef]
31. Scheurer, S.; Tedesco, S.; O’Flynn, B.; Brown, K.N. Comparing Person-Specific and Independent Models on Subject-Dependent and Independent Human Activity Recognition Performance. *Sensors* **2020**, *20*, 3647. [CrossRef] [PubMed]
32. Abidi, A.; Ben Khalifa, K.; Ben Cheikh, R.; Valderrama Sakuyama, C.A.; Bedoui, M.H. Automatic Detection of Drowsiness in EEG Records Based on Machine Learning Approaches. *Neural Process Lett.* **2022**, *54*, 5225–5249. [CrossRef]
33. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge Computing: Vision and Challenges. *IEEE Internet Things J.* **2016**, *3*, 637–646. [CrossRef]
34. Dimitrakopoulos, G.N.; Kakkos, I.; Dai, Z.; Wang, H.; Sgarbas, K.; Thakor, N.; Bezerianos, A.; Sun, Y. Functional Connectivity Analysis of Mental Fatigue Reveals Different Network Topological Alterations Between Driving and Vigilance Tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 740–749. [CrossRef] [PubMed]
35. Brouwer, A.-M.; Hogervorst, M.A.; van Erp, J.B.F.; Heffelaar, T.; Zimmerman, P.H.; Oostenveld, R. Estimating Workload Using EEG Spectral Power and ERPs in the N-Back Task. *J. Neural Eng.* **2012**, *9*, 045008. [CrossRef]
36. 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]
37. Crespo-Garcia, M.; Pinal, D.; Cantero, J.L.; Díaz, F.; Zurrón, M.; Atienza, M. Working Memory Processes Are Mediated by Local and Long-Range Synchronization of Alpha Oscillations. *J. Cogn. Neurosci.* **2013**, *25*, 1343–1357. [CrossRef]
38. Fukuda, K.; Mance, I.; Vogel, E.K. α Power Modulation and Event-Related Slow Wave Provide Dissociable Correlates of Visual Working Memory. *J. Neurosci.* **2015**, *35*, 14009–14016. [CrossRef]
39. Krause, C.M.; Pesonen, M.; Hämäläinen, H. Brain Oscillatory 4–30 Hz Electroencephalogram Responses in Adolescents during a Visual Memory Task. *Neuroreport* **2010**, *21*, 767–771. [CrossRef]
40. Dimitrakopoulos, G.N.; Kakkos, I.; Dai, Z.; Lim, J.; deSouza, J.J.; Bezerianos, A.; Sun, Y. Task-Independent Mental Workload Classification Based Upon Common Multiband EEG Cortical Connectivity. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1940–1949. [CrossRef] [PubMed]
41. Chiew, K.S.; Braver, T.S. Context Processing and Cognitive Control: From Gating Models to Dual Mechanisms. In *The Wiley Handbook of Cognitive Control*; Wiley Blackwell: Hoboken, NJ, USA, 2017; pp. 143–166. ISBN 978-1-118-92054-1.

42. Ardestani, A.; Shen, W.; Darvas, F.; Toga, A.W.; Fuster, J.M. Modulation of Frontoparietal Neurovascular Dynamics in Working Memory. *J. Cogn. Neurosci.* **2015**, *28*, 379–401. [[CrossRef](#)] [[PubMed](#)]
43. Stancin, I.; Cifrek, M.; Jovic, A. A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems. *Sensors* **2021**, *21*, 3786. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.