







Article

An EEG-Based Transfer Learning Method for Cross-Subject Fatigue Mental State Prediction

Hong Zeng^{1,2,3}, Xiufeng Li¹ , Gianluca Borghini³ , Yue Zhao¹, Pietro Aricò³ , Gianluca Di Flumeri³ , Nicolina Sciaraffa³, Wael Zakaria⁴, Wanzeng Kong^{1,2}  and Fabio Babiloni^{3,*} 

¹ School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China; jivon@hdu.edu.cn (H.Z.); lixiufengcn@hdu.edu.cn (X.L.); zhaoyue@hdu.edu.cn (Y.Z.); kongwanzeng@hdu.edu.cn (W.K.)

² Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province, Hangzhou 310018, China

³ Industrial NeuroScience Lab, University of Rome “La Sapienza”, 00161 Rome, Italy; gianluca.borghini@uniroma1.it (G.B.); pieter.arico@uniroma1.it (P.A.); gianluca.diflumeri@uniroma1.it (G.D.F.); nicolina.sciaraffa@uniroma1.it (N.S.)

⁴ Department of Mathematics-Computer Science, Faculty of Science, Ain Shams University, Abbassia, Cairo 11435, Egypt; wael.zakaria@sci.asu.edu.eg

* Correspondence: fabio.babiloni@uniroma1.it

Abstract: Fatigued driving is one of the main causes of traffic accidents. The electroencephalogram (EEG)-based mental state analysis method is an effective and objective way of detecting fatigue. However, as EEG shows significant differences across subjects, effectively “transferring” the EEG analysis model of the existing subjects to the EEG signals of other subjects is still a challenge. Domain-Adversarial Neural Network (DANN) has excellent performance in transfer learning, especially in the fields of document analysis and image recognition, but has not been applied directly in EEG-based cross-subject fatigue detection. In this paper, we present a DANN-based model, Generative-DANN (GDANN), which combines Generative Adversarial Networks (GAN) to enhance the ability by addressing the issue of different distribution of EEG across subjects. The comparative results show that in the analysis of cross-subject tasks, GDANN has a higher average accuracy of 91.63% in fatigue detection across subjects than those of traditional classification models, which is expected to have much broader application prospects in practical brain–computer interaction (BCI).

Keywords: cross-subject prediction; Domain-Adversarial Neural Network (DANN); electroencephalogram (EEG); Generative Adversarial Networks (GAN)



Citation: Zeng, H.; Li, X.; Borghini, G.; Zhao, Y.; Aricò, P.; Di Flumeri, G.; Sciaraffa, N.; Zakaria, W.; Kong, W.; Babiloni, F. An EEG-Based Transfer Learning Method for Cross-Subject Fatigue Mental State Prediction. *Sensors* **2021**, *21*, 2369. <https://doi.org/10.3390/s21072369>

Academic Editor: Wai Lok Woo

Received: 11 March 2021

Accepted: 24 March 2021

Published: 29 March 2021

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



Copyright: © 2021 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

Mental fatigue is a complex physiological and psychological state, which can lead to a decline in alertness, concentration and cognitive performance [1]. About 1.3 million people lose their lives in traffic accidents every year in the world [2], and fatigued driving is a leading factor in it [3]. Thus, performing mental state detection and prediction while driving is extremely important to reduce losses of lives and property caused by fatigued driving [4,5].

Numerous fatigued driving detection methods have been proposed. They can be divided into three categories: psychometric questionnaires (e.g., Karolinska Sleepiness Scale (KSS) [6] and Checklist Individual Strength questionnaire (CIS) [7]), behavioral methods (e.g., facial expressions [8], head movement [9], blink rate [10] and eye state [5]) and physiological measurements (e.g., EEG, electrocardiogram (ECG), electrooculogram (EOG) and surface electromyography [8,11,12]). Regarding these questionnaires, they are not only strongly subjective, but they are also unable to monitor the fatigue state in real time, while behavioral methods are extremely susceptible to interference from the road environment, resulting in a certain error of judgement. Among physiological measurements, EEG is regarded as the most effective method for mental state detection, since EEG records electric

activities of neural cells from the human cerebral cortex, which can directly reflect instant states of the brain and avoid the effects of human subjectivity [10,13].

Most existing EEG-based methods for fatigued driving detection focus on extracting suitable features and designing classifiers. Chai et al. used power spectral density (PSD) as the feature extraction method and Bayesian neural networks (BNN) as the classifier for classifying alert and fatigue [14]. Rahma et al. extracted the Common Spatial Pattern (CSP) feature to train extreme learning machine (ELM) for fatigue detection [15]. Huo et al. [16] combined EEG and forehead EOG to detect fatigue level of drivers by using a discriminative graph regularized extreme learning machine (GELM). San et al. [17] proposed a hybrid deep generic model (DGM)-based support vector machine for driver fatigue detection. Although effective, such methods mainly focus on EEG analysis under the situation of the same subject, which means the abilities for cross-subject detection are still insufficient.

Nowadays, the domain adaptive-method-based transfer learning models are extensively applied to many fields such as natural language processing and image classification and achieve very good performance. Domain-Adversarial Neural Network (DANN) [18] is a typical domain adaptive method, and the most important feature of DANN is to align the source domain and target domain without the labeled target samples. Although achieving better performance in these applications, it still has some drawbacks in EEG analysis across subjects. First, DANN requires the samples between source and target domains to be balanced, while for cross-subject EEG analysis, the samples in source domain are usually much larger than those in target domain, which shows severe imbalance. Second, since significant difference of EEG signals exists across subjects, some source domain samples may be extremely inconsistent with the distribution of the target domain, which will cause “negative transfer” [19,20] and make the performance of DANN decrease in cross-subject EEG analysis.

In view of the shortcomings encountered by DANN in cross-subject EEG analysis, we propose an improved model, Generative-DANN (GDANN), by combining Generative Adversarial Networks (GAN) [21] to improve the ability in EEG analysis across subjects through the following three aspects: (a) the hidden layer, optimizer and loss function of DANN are improved to meet the requirement of cross-subject EEG analysis; (b) random noise is selected through GAN to generate sufficient fake data, which are close to the data distribution in the target domain, to balance the data set in the source and target domains to assist model training; and (c) when faced with multiple source domain data, GDANN can select the samples from the subjects with the closest data distribution to ensure cross-subject analysis performance more effectively and avoid negative transferring to a certain extent.

The rest of this paper is organized as follows. Section 2 introduces the process of the experiment. Section 3 elaborates the proposed model in detail. In Sections 4 and 5, we present and discuss the experimental results. Finally, we conclude and give our prospects in the future in Section 6.

2. Experiments

2.1. Subjects

Informed consent was obtained from 13 healthy volunteers after the explanation of the study, which was approved by the local institutional ethical committee of University of Rome, La Sapienza (Rome, Italy). The group was selected with the aim to have a homogeneous sample in terms of age (26.8 ± 3.2 years old), and driving expertise (all the participants drove daily and regularly). All of them are asked to avoid alcohol the day prior to the measurements and caffeine 5 h before the experiment. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2008.

2.2. Experimental Protocol

The experiments were performed within the same hours of the day to avoid bias in the results due to circadian rhythms or meals. In particular, the participants took part in the driving simulations between 2 PM and 5 PM. The simulation consisted in driving an Alfa Romeo—Giulietta QV (1750 TB_i, 4 cylinders, 235 HP) on the Spa—Francorchamps (Belgium) track under different conditions. In order to modulate the difficulty of the driving simulation (i.e., primary task), the Attentional and Vigilance task (TAV) was employed as a secondary task [22,23]. When enabled, the TAV lasted for the entire corresponding experimental condition (i.e., 2 laps), and it worked as follows: the *vigilance* stimulus, a white X, was presented on the center of a monitor placed about 60 (cm) from the driver and below the main screen where the car cockpit was projected, as shown in Figure 1.

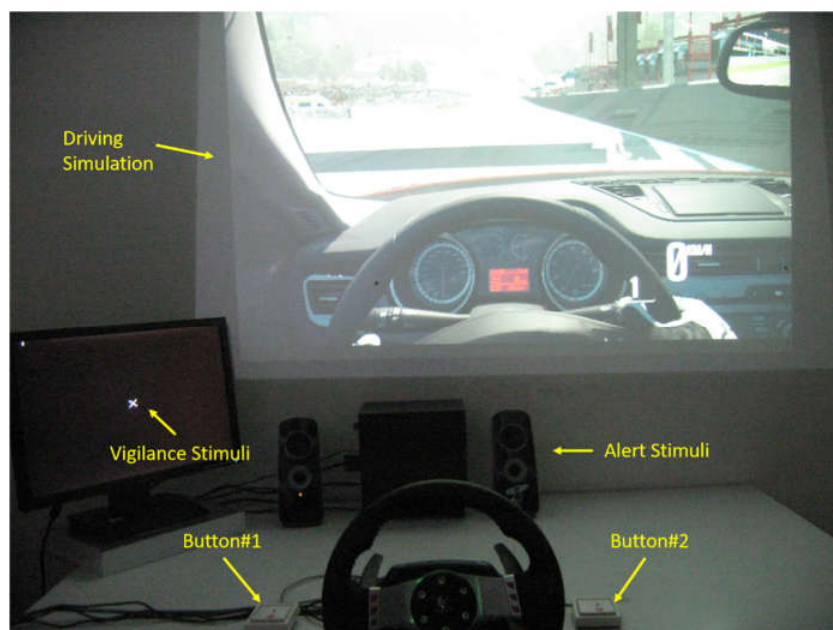


Figure 1. Experimental settings. (The experiment established five different Attentional and Vigilance tasks (TAV) to gradually increase the workload of the brain, allowing the subjects to drive a simulated car on the circuit under the same other conditions. When a white X appeared on the screen, subjects were asked to press Button#1, which represents “Vigilance Stimuli”. When the sound monitor emitted a stimulus, subjects were asked to press Button#2, which represents “Alert Stimuli”).

The drivers were asked to press the Button#1 as soon as the X appeared on the screen regardless of the ongoing driving situation. The vigilance stimuli simulates traffic-related events like red-turning traffic lights, road crossing pedestrians, other cars or other uncontrollable traffic agents. The acoustic alert stimulus was presented by two speakers placed on the left and right sides of the driver. A sequence of frequent (with a 95% probability rate) and target (5% probability) tones at different acoustic frequencies were continuously delivered to the drivers. They were asked to press the Button#2 as soon as the target stimulus occurred. The frequent acoustic tones simulated the car’s radio or engine noise, while the target ones reproduced unexpected events like a phone call. Five levels of the TAV were designed by setting different stimulation rates in order to induce different levels of workload demands and an overall mental fatigue status in the driver [3,24,25]. Before starting the experiment, the participants took part in a training session of half an hour to get familiar with the simulator commands and interface. Afterward, they initially drove the car through the selected track without any requests in terms of speed but always maintaining the car within the path. Such a condition was named “warm up” (WUP) and was aimed at defining the driver’s baseline. Then, the drivers are asked to repeat the 2 laps by driving as fast as they could but always by ensuring high safety, that is avoiding driving off the path; this was the “performance” condition (PERFO). After the PERFO

condition, the five TAVs (TAV1 to TAV5) were presented in a random order across the participants, where TAV1 was the easiest condition (i.e., very slow stimulation), whilst TAV5 was the most difficult one (i.e., very fast stimulation). Finally, the last experimental condition consisted of monotonous driving, in which the participants had to drive without exceeding the speed of 70 km/h. The purpose of the consecutive TAVs was to induce mental fatigue, while the last one aimed to make the drivers hypo-vigilant or drowsy (DROW). Except for the last stage of the DROW, which took 10 min, the remaining 7 stages in the experiment only took 2.5 to 4 min, so the total time required for the experiment was approximately 32–45 min. At the end of each experiment, the participants were asked to fill in the NASA-TLX questionnaire to provide the subjective workload perception [26]. In addition, errors in terms of keeping the car within the path and missed or wrong TAV answers were gathered.

2.3. EEG Recording and Preprocessing

EEG was recorded by a digital monitoring system (Brain Products GmbH, Munich, Germany) with a sampling frequency of 200 Hz. All 61 EEG channels were referred to both earlobes, grounded to the FC_z channel, and their impedances were kept below 10 K Ω . After that, a band-pass filter was used for keeping EEG data between 1 Hz and 30 Hz, and Independent Component Analysis (ICA) [27] was adopted to remove the artifacts caused by eye blinking. After removing artifacts, we applied a 1 s sliding window with 10% overlap to segment EEG signals into 61 channels, and 400 segments of each condition were acquired for each subject.

Then, power spectral density (PSD), which is more sensitive in the range of frequency from 0.1 to 30 Hz [28,29], was chosen to transform three-dimensional time-series segments into two-dimensional sample data. The detailed schematic diagram of PSD is shown in Figure 2.

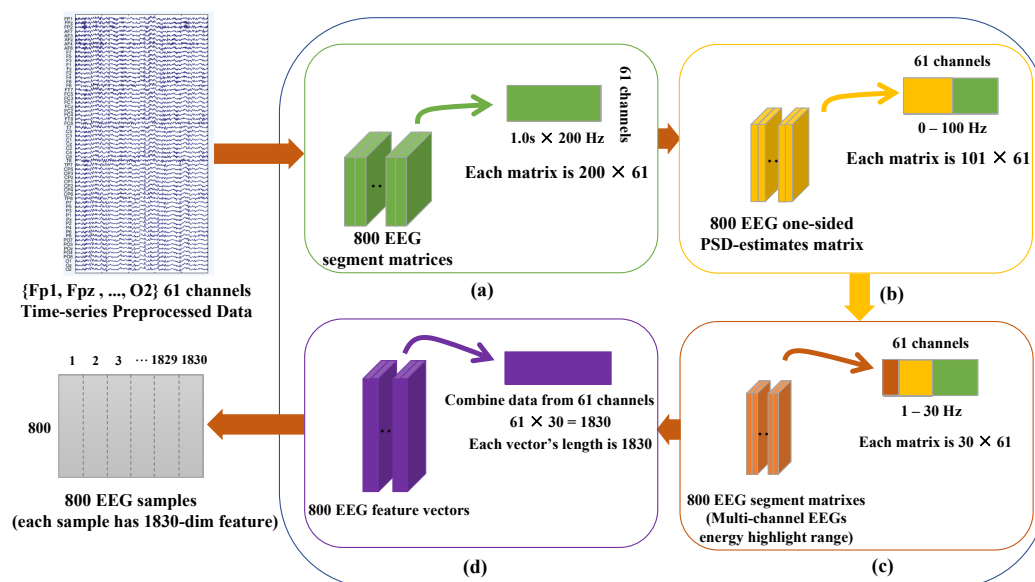


Figure 2. Schematic Diagram of electroencephalogram (EEG) power spectral density (PSD) extraction. ((a) shows that, we obtained 400 segment matrices for each condition regarding 61 EEG channels, and due to 200 Hz of sampling frequency and 1 s sliding window, the dimension of each matrix was 200×61 ; (b) reveals the one-side PSD estimate of the 100 Hz EEG signals, which means the logarithm of the signal power at an integer frequency between 0 and 100 Hz [29]. Regarding PSD-related frequency band, we chose the range from 0.1 to 30 Hz of EEG as the input signals (c). Each segment of the integer frequency between 0 and 30 Hz contained 30 EEG powers and 1830 powers for 61 channels. Correspondingly, the 400×1830 of dimensional feature vectors in each segment were obtained at the end, as shown in (d)).

Therefore, for 13 subjects, 400 samples could be obtained for each condition (TAV1-TAV5, PERFO and DROW), and each sample had 1830 dimensional features. Herein, the two conditions of TAV3 and DROW were selected as positive and negative examples for subsequent experiments, which is explained in Section 4.1.

2.4. Domain Adaptation Learning

Domain Adaptation is a representative method in transfer learning, which uses source domain samples that contain rich information to improve the performance for classifying other target domain samples. In domain adaptation, the samples in target domain for testing usually have no or only a few labels, while those in the source domain for training are rich in supervised information. The source domain and the target domain often belong to the same type of task, but the sample distribution between them is different.

For the domain adaptation task, if a common feature representation space can be extracted between the source domain and the target domain, then, in this feature space, the classifier model learned from the source domain features can also be used on the features of the target domain. In particular, the domain adaptation task is often converted into the task of finding the common feature representation space, which is the domain-invariant feature [30]. If the domain invariant features can be learned by the model, the classifier can be trained by the obtained invariant features to make it effective for both the source domain and the target domain.

DANN and GDANN work as described above, which is described in detail in Section 3 later. The data used for training and testing the model comes from the feature samples extracted in Section 2.3.

2.5. Cross-Subject Cross-Validation and Evaluation Index

For cross-subject cross-validation, unlike the previous intra-subject experiments, we retained the data of one subject from the data set, used them as test data (i.e., target domain data), and summarized the data of the remaining 12 subjects together as training data (i.e., source domain data). Obviously, there is no intersection of data between the two domains. In addition, it should be noted that in this experiment, our model required some unlabeled data from the target domain to assist training. After the model was trained, we evaluated the recognition performance on the retained subject's data and recorded the results. This process was repeated until each subject had been used at least once as a testing subject.

A confusion matrix was used to analyze obtained results, which used count values to summarize the number of correct and incorrect predictions and to subdivide them by category. The confusion matrix shows which part of the classification model will be confused when making predictions. It is this decomposition of the results that overcomes the limitations of using only classification accuracy. The binary classification analysis used in this experiment is shown in Table 1.

Table 1. Binary confusion matrix indicator.

	Predicted = 1	Predicted = 0
Label = 1	TP (True Positive)	FP (False Positive)
Label = 0	FN (False Negative)	TN (True Negative)

1. Recall = $TP / (TP + FN)$
2. Precision = $TP / (TP + FP)$
3. Accuracy = $(TP + TN) / (TP + TN + FN + FP)$
4. F1Score = $(2 \times Precision \times Recall) / (Precision + Recall)$

Herein, recall means, in all the samples that are actually 1 (positive samples), the probability of correct prediction. Precision means the probability of correct prediction in all samples with a prediction of 1 (positive sample). Accuracy is the proportion of the samples

that are predicted correctly in all samples. Furthermore, F1Score [31,32] is the harmonic mean of Precision and Recall.

3. Method

3.1. The Existing DANN

The architecture of DANN consists of three kinds of networks, as shown in Figure 3: (a) Feature Extractor Network $N_f(\cdot; \mu_f)$, which adopts a shallow network model to extract the potential features with the parameter μ_f from both the source and the target domain; (b) Label Predictor Network $N_y(\cdot; \mu_y)$, which is essentially a binary classifier and in charge of training the predictor with the labeled source domain data with the parameter μ_y and loss value L_y ; and (c) Domain Classifier Network $N_d(\cdot; \mu_d)$, which extracts the invariant features through the parameter μ_d and loss value L_d and then judges whether the features come from the source domain or the target domain. During the training period of back-propagation, DANN implements unsupervised domain adaptation by adding a Gradient Reversal Layer (GRL) between N_f and N_d [18].

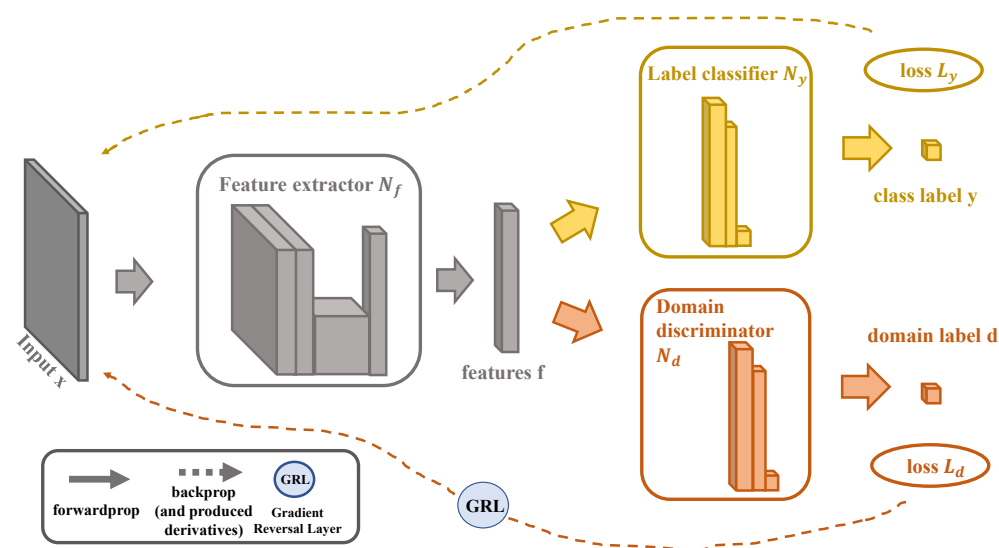


Figure 3. The Architecture of Domain-Adversarial Neural Network (DANN). (DANN combines three networks (N_f , N_y , and N_d) to align data from different distributions of source and target domains).

3.2. The Architecture of GDANN

On the basis of DANN, we added GAN networks, which consists of (a) the generator network $N_{tg}(\cdot; \mu_{tg})$, which is used to generate the fake target domain data with its parameter of μ_{tg} and loss value L_{tg} , and (b) the classifier network $N_{td}(\cdot; \mu_{td})$, which is used to identify the real target domain data and the fake target domain data with its parameter of μ_{td} and loss value L_{td} . The parameter μ should contain the weight W and the bias b of each layer network. Their network structure is shown in Figure 4, which includes the input layer, output layer, several hidden layers and the bias node.

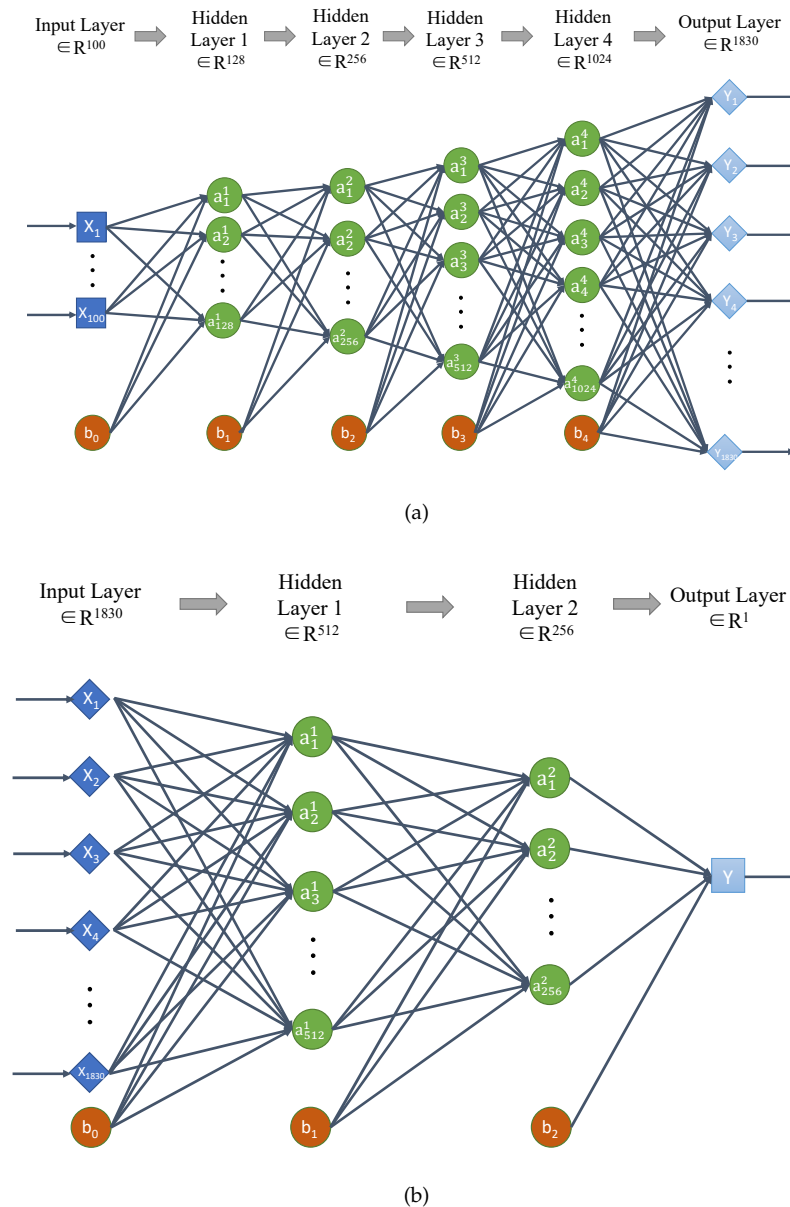


Figure 4. The network structure of Generative Adversarial Networks (GAN) ((a) N_{tg} and (b) N_{td}).

For the improvement of all layers of GDANN, the *LeakyRelu* function was employed as the activation function, as shown in Equation (1), where $leak \in (0, 1)$ is a constant. In this way, some negatives are retained, and the correspondingly effective information will not be completely lost.

$$LeakyRelu(\alpha) = \max(0, \alpha) + leak \times \min(0, \alpha) \quad (1)$$

The final output layer of the N_{tg} and N_f uses the Tanh function to output the result, which could effectively reduce the probability of saturation compared to the traditional Sigmoid function used in DANN [33]. In the newly added N_{tg} and N_{td} networks, the adaptive time estimation method (Adam function [34]) is used to perform the gradient descent, to make GDANN have better convergence. Meanwhile, N_f , N_y and N_d networks adopt *SGD* function [18] for parameter update optimization which is used by DANN.

As for the loss function of L_{tg} , L_{td} , L_d , and L_y , the CrossEntropy is used to predict the difference between the prediction data and true data. As shown in Equation (2), where

x represents a sample, $p(x)$ and $q(x)$ are the true sample distribution and the predicted sample distribution of x , respectively.

$$\text{CrossEntropy}(p, q) = - \sum_x p(x) \log q(x). \quad (2)$$

Thus, the network architecture of GDANN is combined with the three networks of DANN (N_f , N_y , and N_d), and GAN (N_{td} and N_{tg}).

3.3. The Training Process of GDANN

We define X as the input space, $Y = 0, 1, \dots, L - 1$ as a collection of L labels. Therefore, data D can be expressed as $D = \{(x_i, y_i)\}_{i=1}^n \in R^{x*y}$, where $x \in X, y \in Y$. In this case, we take binary classifications as an example (0 for wakefulness and 1 for fatigue), i.e., $L = 2$. Suppose that we have two different distributions on $X * Y$, which are called the source domain D_s (consisting of data from 12 subjects) and the target domain D_t (consisting of data from the remaining subjects). The data distribution of the source domain and target domain are P_s and P_t , respectively.

The goal is to train GDANN by using (1) the data on suitable N_s ($1 \leq N_s < 13$) subjects in D_s , which is close to the distribution of P_s ; (2) a certain proportion (denote λ as the scaling factor) of randomly selected unlabeled data X_{t1} in D_t , which is line with the distribution P_t ; and (3) the parameter μ , wherein we define n_s as the number of samples in source domain data set, n_t as the number of X_t (i.e., 800), n_{t1} as the number of the randomly selected unlabeled X_{t1} . Then we can get: $n_{t1} = \lambda n_t$ and $n_s = N_s \times 800$. In addition, random noise Z that conforms to Gaussian distribution P_z is denoted as Equation (3) and is used to generate equivalent number of fake data compared with those of source data, where r is a random number.

$$Z = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{r^2}{2}\right)} \in P_z \quad (3)$$

As shown in Figure 5, the training process of GDANN includes two steps: training GAN and training DANN. The design of this process embodies the idea of a two-two game and optimization, including N_{td} vs. N_{tg} , and N_d vs. N_f and N_y . First, N_{td} hopes to distinguish between real samples and fake samples as much as possible, while N_{tg} tries its best to generate fake data conforming to the P_t distribution to deceive N_{td} —this is a zero-sum game. Second, N_d wants to distinguish the samples from the source domain and the target domain to the greatest extent, while N_f extracts the domain features to fool N_d for not discerning which domain those features come from—this is another zero-sum game. The optimization and competition of GDANN interplay with each other and finally reach the performance of global optimum, which will be shown through N_y label classifier network.

As shown in Figure 5a, the first step is the confrontation between N_{tg} and N_{td} . Their optimization process is expressed in Equations (4) and (5), which are the evolution of Equation (2). Here, noise z and the target domain data X_{t1} (which are randomly selected from X_t and the number is n_{t1}) are used to assist training. Z is the random noise (please see Equation (3)). The $E(\cdot)$ function means the optimization process, here is the process of maximizing L_{td} and minimizing L_{tg} .

$$E(L_{tg}) = \min_{N_{tg}} L_{tg} \leftarrow \sum_{i=1}^{n_{t1}} \log \frac{1}{n_{t1} \sqrt{N_{tg}(Z_i; \mu_{tg})}} \quad (4)$$

$$E(L_{td}) = \max_{N_{td}} L_{td} \leftarrow \sum_{i=1}^{n_{t1}} \log \left[\frac{1}{2^{n_{t1}} \sqrt{(1 - N_{td}(N_{tg}(Z_i; \mu_{tg}); \mu_{td}))}} \times \frac{1}{2^{n_{t1}} \sqrt{N_{td}(X_{t1}^i; \mu_{td})}} \right] \quad (5)$$

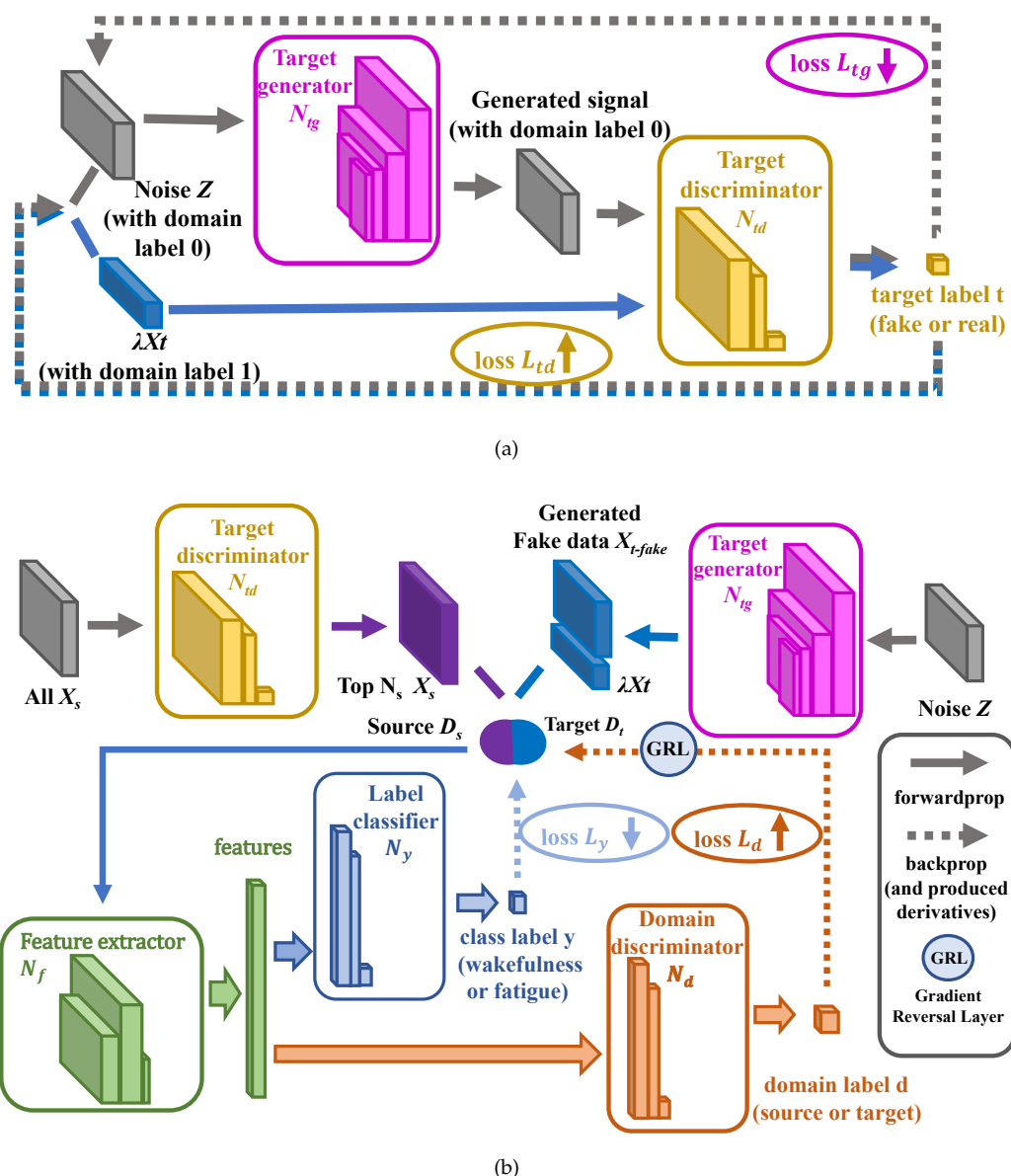


Figure 5. The Training Process of Generative-DANN (GDANN). ((a) is the process of training GAN with minimizing N_{tg} and maximizing N_{td} , and (b) is the process of training DANN with minimizing N_f , N_y and maximizing N_d , which contains the process of screening the top N_s source domain subjects' data that are most similar to the target domain and use random noise to train the domain alignment model together).

In an ideal state, the confrontation training between N_{td} and N_{tg} will reach the Nash equilibrium (i.e., $L_{td} = 0.5$). At this time, N_{tg} can generate fake data that conform to the P_t distribution, and N_{td} also has a certain ability to discriminate the true and fake data; that is, N_{td} can roughly judge the similarity between the data and the P_t distribution.

Then, the training process enters the second stage which is shown in Figure 5b. After the training of N_{tg} and N_{td} , each subject's data are passed through the N_{td} network to obtain their distribution probability. Thus, N_s subjects with the highest probability ranking are selected by the model from the X_s data, which are closest to the P_t distribution and good for transfer learning.

The training data on N_y are derived from the domain features, which come from the top N_s subjects' data, and are generated through N_f . It is just a standard classifier based on neural network technology. In the process of the global optimization, the loss should

be minimized, as shown in Equation (6) (the evolution of Equation (2)), where n_s is the total number of N_s subjects' data, and Y_s is the label of target domain's feature matrix, i.e., $Y_s \in \{0, 1\}$.

$$E(L_y) = \min_{N_f, N_y} L_y \leftarrow \sum_{i=1}^{n_s} \log \left[\left(1 - N_y \left(N_f \left(X_s^i; \mu_f \right); \mu_y \right) \right)^{\frac{Y_s^i - 1}{n_s}} \times \left(N_y \left(N_f \left(X_s^i; \mu_f \right); \mu_y \right) \right)^{\frac{-Y_s^i}{n_s}} \right] \quad (6)$$

The auxiliary fake data, X_{t-fake} , which also obey P_t distribution in the target domain X_t , is generated by a large number of noise Z through N_{tg} , whose number is $(n_s - n_t)$; thus, it makes the number of sample data in X_t ($n_t \ll n_s$, where n_s is the number of data in X_s) be increased, so that the number of data in both domains is balanced. Then, a total of approximate $2n_s$ in both domains can be used for training, as shown in Equation (7) (the evolution of Equation (2)).

$$E(L_d) = \max_{N_d} L_d \leftarrow \sum_{i=1}^{n_s} \log \left(1 - N_d \left(N_f \left(X_s^i; \mu_f \right); \mu_d \right) \right)^{\frac{-1}{n_s}} + \sum_{i=1}^{n_t} \log \left(N_d \left(N_f \left(X_t^i; \mu_f \right); \mu_d \right) \right)^{\frac{-1}{n_t}} \\ + \sum_{i=1}^{n_s - n_t} \log \left(N_d \left(N_f \left(N_{tg} \left(Z_i; \mu_{tg}^* \right); \mu_f \right); \mu_d \right) \right)^{\frac{-1}{n_t - n_s}} \quad (7)$$

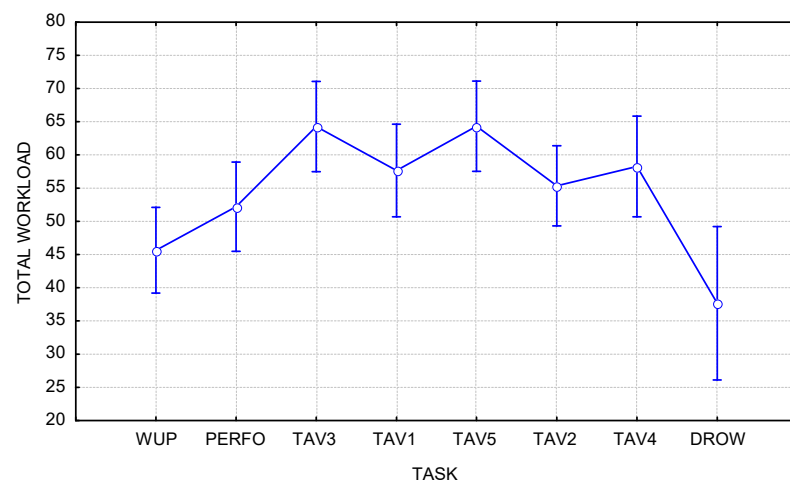
4. Results

This section presents the performance of the proposed approach on samples from the industry and neural science laboratory in University of Rome, La Sapienza', which include the EEG recordings of 13 subjects. The work was executed on a computer with 16 GB RAM, NVIDIA GeForce GTX 1660 graphic memory with 6 GB, and Intel Core i5-9400F @ 2.9 GHz processor. Python 3.6.8 tools were adopted to verify GDANN algorithm under Linux environment with Ubuntu 5.4 operating system.

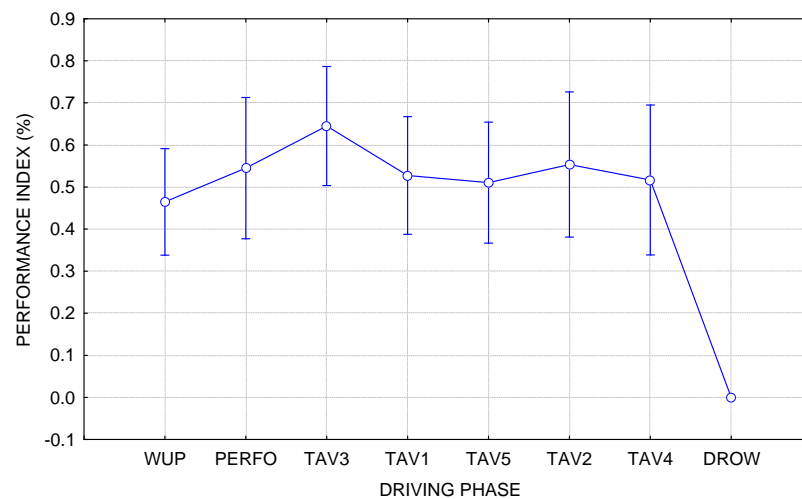
We used four models, DANN, GDANN, Support Vector Machine (SVM) [35] and Easy Transfer Learning (EasyTL) [36], to perform the cross-subject cross-validation process, which was mentioned in Section 2.5. In addition, the obtained EEG data were pre-processed by the method of Section 2.3. The above process was repeated 5 times to get the average result and analyzed.

4.1. Selection of Regression Labels

The subjective (i.e., NASA-TLX) and performance (i.e., Driving Performance) data were used to identify the experimental conditions by which our deep-learning model was trained. In particular, the results reported in Figure 6 show that the TAV3 and DROW required, respectively, the lowest and highest cognitive demand; therefore, they were used as calibration data set. Our previous work [37,38] also proved that TAV3 and DROW are the two most different states, which are suitable for regression labels for fatigue state classification. This means that each subject (13 subjects in total) will get a total of 800 samples, of which 400 are from TAV3 with a label of 0, and the rest are from DROW with a label of 1.



(a)



(b)

Figure 6. Reasons for the choice of TAV3 and DROW as the regressor labels. ((a) is the process of NASA-TLX (subjective) with $p < 1 \times 10^{-5}$, and (b) is the process of Behavioural (Performance) with $p < 1 \times 10^{-5}$).

4.2. Parameter Sensitivity

In order to obtain a more stable performance, we then studied the sensitivity of the parameters of λ (scaling factor, which was used to choose the samples of unlabeled λX_t from X_t to assist training), N_s (source subject number, which was used to select the best N_s subjects from 13 to assist training) in GDANN. We then adopted a kind of grid search method to observe their sensitivity. That is, when optimizing one parameter, the other parameter remained unchanged, thus we could observe the optimal training performance. These two parameter values are determined from the following two ranges, which are: $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ for λ , and $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ for N_s . The results are shown in Figure 7.

The “Acc” values in Figure 7 indicate that the average accuracy obtained by using the GDANN model for experiments under these parameter values (namely λ in Figure 7a, N_s in Figure 7b), which can exclude the differences in model performance among different subjects. Choosing different values of λ , and N_s has the following effects:

- (a) In Figure 7a, as λ increases, the average accuracy of GDANN tends to increase slowly. To better reflect the robustness of the model, $\lambda = 0.5$ was selected.
- (b) In Figure 7b, as the number of source subjects increases from 1 to 9, the accuracy increases sharply. When the number reaches 9, the accuracy remains stable, and the accuracy curve may fluctuate slightly. Thus, we set $N_s = 9$ in the following experiments.

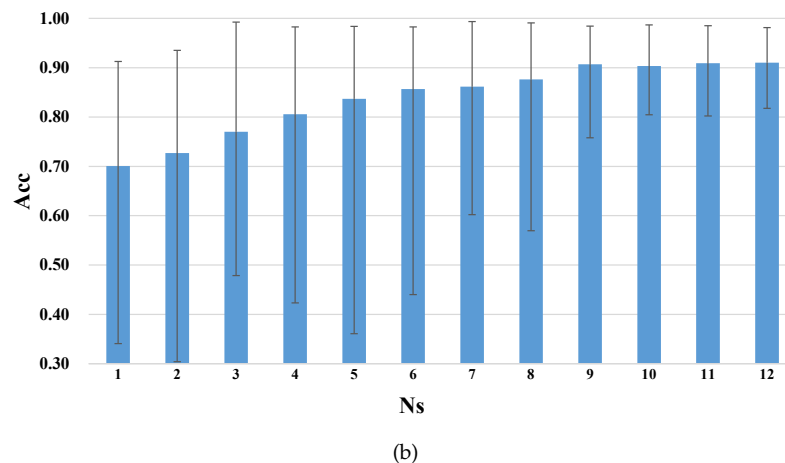
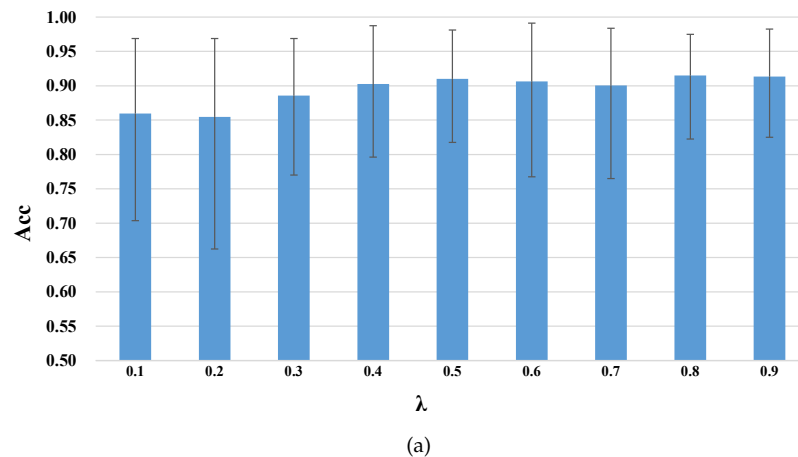


Figure 7. Parameter sensitivity of GDANN ((a) shows the parameter influence of λ with $N_s = 12$, and (b) shows the parameter influence of N_s with $\lambda = 0.5$).

In addition, the model also includes the following parameter settings: $Epoch = 50$, $Batch_Size = 64$, $Adam_Learning_Rate = 0.00001$, $Adam_b1 = 0.5$, $Adam_b2 = 0.999$, $SGD_Learning_Rate = 0.005$ and $SGD_Momentum = 0.9$.

4.3. High-Dimensional Feature Visualization

Since the training of the model relies on the auxiliary fake data from N_{tg} , we used the t-SNE method [39] to measure the quality of the data generated by GAN. For example, we took Subject#1 as the target domain data and used the default parameters of t-SNE; after 1000 iterations, the high-dimensional feature visualization results shown in Figure 8 were obtained. Here, tag 0 represents the unlabeled target domain data (Figure 8a), tag 1 represents the target domain data generated in GAN (Figure 8b), and tag 2 represents the source domain data (Figure 8c). It can be seen that the compact and dense generated fake data are roughly centered on the center of the unlabeled target domain data, while the distribution of the source domain data is relatively scattered and extensive.

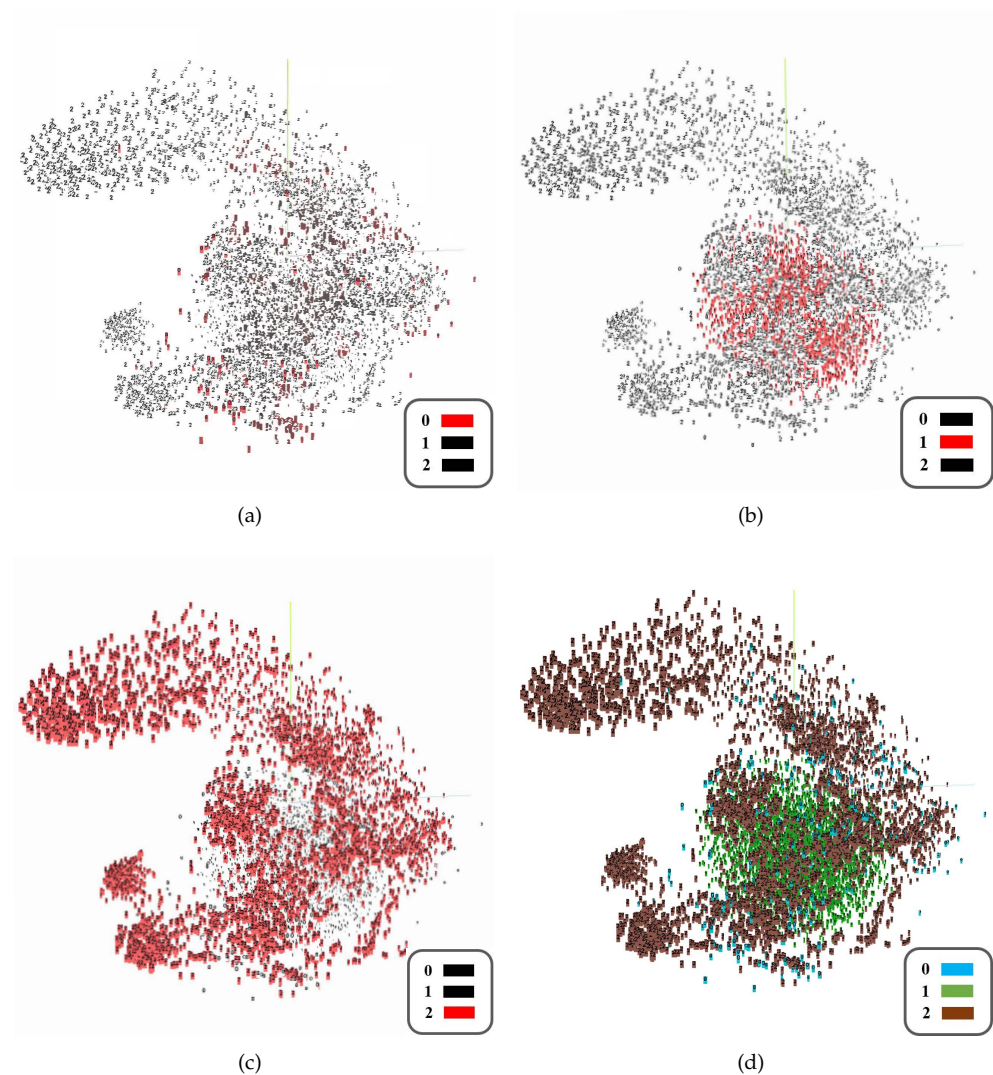


Figure 8. High-dimensional feature visualization results (t-SNE) on EEG DataSet Others_Subject#1. ((a) shows unlabeled target domain data with tag 0, (b) shows generated target domain data with tag 1, (c) shows source domain data with tag 2, and (d) is the collective display of data).

4.4. Performance Comparison between GDANN and DANN

In order to reflect the improvement of the proposed model, we first compared the performance between GDANN and DANN, then using the optimal parameters obtained above, we analyzed the performance of the related indexes of Accuracy, Precision, F1Score and Recall for it. In our experiment, PSD (Power Spectral Density) was extracted as the input to train GDANN and DANN. The same Training Set and Testing Set were used for the purposes of training and testing, respectively. Table 2 shows the average results obtained after five times of repeated experiments.

Table 2. Comparison results of the confusion matrix indicators between DANN and GDANN.

Others_Target Subject ID	Accuracy		Precision		F1Score		Recall	
	DANN	GDANN	DANN	GDANN	DANN	GDANN	DANN	GDANN
Others_Subject #1	0.6238	0.8294	0.6650	0.8800	0.6387	0.8376	0.6143	0.7991
Others_Subject #2	0.8838	0.9531	0.8900	0.9413	0.8845	0.9525	0.8790	0.9643
Others_Subject #3	0.6638	0.8831	0.6750	0.8425	0.6675	0.8782	0.6601	0.9171
Others_Subject #4	0.9363	0.9988	0.9550	0.9988	0.9374	0.9988	0.9205	0.9988
Others_Subject #5	0.8113	0.9019	0.7875	0.9150	0.8067	0.9032	0.8268	0.8916
Others_Subject #6	0.9313	0.9625	0.9325	0.9688	0.9313	0.9627	0.9302	0.9568
Others_Subject #7	0.6825	0.8831	0.7225	0.9038	0.6947	0.8840	0.6690	0.8695
Others_Subject #8	0.6663	0.8288	0.6600	0.8100	0.6642	0.8255	0.6684	0.8416
Others_Subject #9	0.9438	0.9719	0.9550	0.9725	0.9444	0.9719	0.9340	0.9713
Others_Subject #10	0.9375	0.9863	0.9150	0.9763	0.9361	0.9861	0.9581	0.9962
Others_Subject #11	0.7963	0.8556	0.7900	0.8200	0.7950	0.8503	0.8000	0.8831
Others_Subject #12	0.8488	0.9225	0.8225	0.9300	0.8447	0.9231	0.8681	0.9163
Others_Subject #13	0.9113	0.9356	0.9250	0.9500	0.9125	0.9365	0.9002	0.9235
Average	0.8182	0.9163	0.8227	0.9161	0.8198	0.9162	0.8176	0.9176

4.4.1. Statistical Analysis

For statistical validation of the results, a two-tailed Wilcoxon Signed Ranks Test [40] analysis was adopted to compare the performance of significant difference of Recall, Precision, Accuracy, and F1Score between DANN and GDANN using the data displayed in Table 2. The results of the two-tailed Wilcoxon Signed-Ranks Test are shown in Figure 9. The p -values of the four indicators are all less than 0.01, which means significant difference, validating the claim of better performance of GDANN.

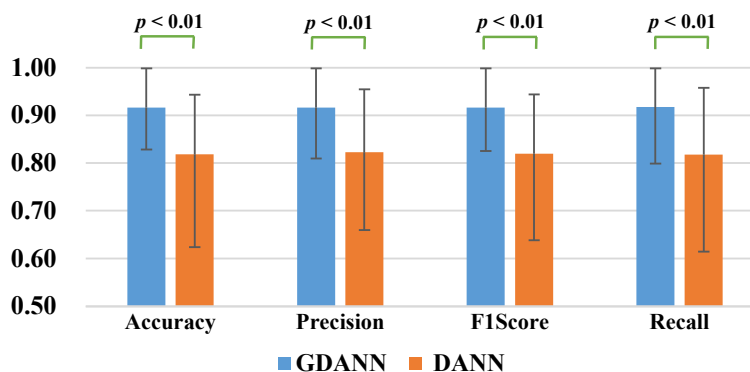


Figure 9. Paired Difference Analysis between four indicators of DANN and GDANN by Wilcoxon Signed Ranks.

4.4.2. Convergence Analysis

Additionally, the change in loss rate and convergence were analyzed between DANN and GDANN, as shown in Figure 10. When regarding the loss rate, we also take the Subject#1 as the target domain data and then observed the loss rate with the increase of the training epoch of the DANN and GDANN. Here, the X-axis represents the training epoch, and the Y-axis represents the total loss, which is the sum of the loss L_d of the domain discriminator N_d and the loss L_y of the label classifier N_y . The *Baseline* here is 1, because this is the process of maximizing L_d and minimizing L_y , and the ideal state of optimization is close to 1. Figure 10 shows that the total loss of GDANN and DANN changed significantly at the beginning, and then GDANN started to reach the baseline before DANN in the 15th epoch and fluctuated around the baseline; DANN reached balance in the 50th epoch, and the total loss stayed near 1.05; it was hard to continue training downwards.

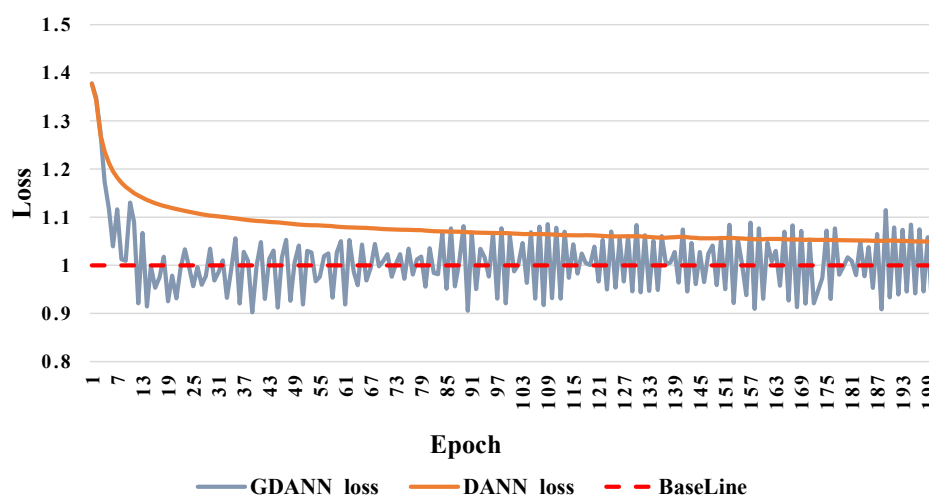


Figure 10. Loss of DANN and GDANN on EEG DataSet Others_Subject#1. (The Baseline refers to the saddle point where the training converged and reached the global optimal value, i.e., 1).

4.5. Performance Comparison between GDANN and Other Existing Models

In order to verify the efficacy of the proposed method, we also compared its performance with SVM and EasyTL. These methods were used as recommended in the referenced papers, and these were chosen for comparison for the following reasons. (a) SVM is the most common and efficient supervised machine learning method, and it can be used to highlight the powerful performance of transfer learning. (b) EasyTL is another useful transfer learning method which can be used to compare performance with GDANN.

Herein, it should be noted that in the process of comparison, EasyTL had the same training set and testing set as GDANN, while SVM did not need auxiliary training by unlabeled data but the training set from X_s and testing set from X_t . We can get the corresponding box plots of the confusion matrix, as well as their paired t-test's results, as shown in Figure 11. By comparing these four indicators, it can be seen that our model GDANN has a better performance than other three models. For example, in terms of accuracy, the improved model GDANN has an average accuracy of 91.63% and a performance improvement of 11.9% compared to the original model DANN, while EasyTL has an average accuracy of 72.95%, and SVM is 67.77%, reflecting the good performance of GDANN.

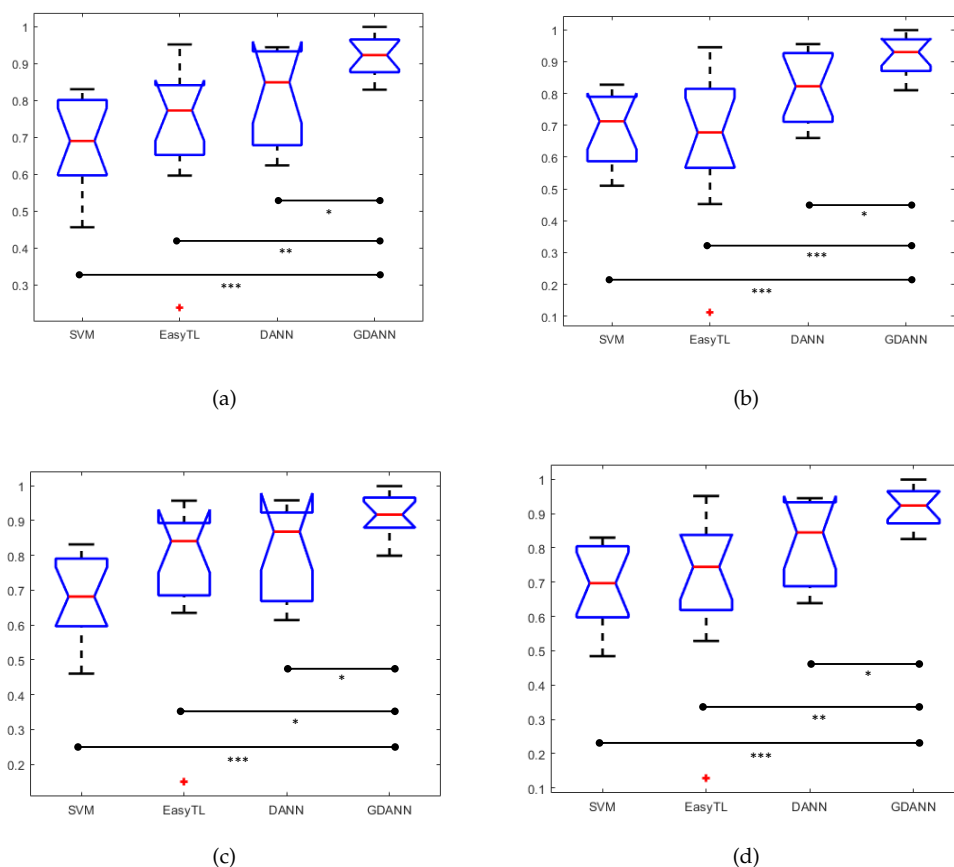


Figure 11. Box plots of confusion matrix ((a) Accuracy, (b) Precision, (c) Recall, and (d) F1Score) between SVM, EasyTL, DANN and GDANN. (* indicates $p < 0.05$; ** indicates $p < 0.01$; *** indicates $p < 0.001$).

Furthermore, we compared the testing time of the four models in Figure 12. The figure compares the average testing time required by the corresponding model, with positive and negative deviation error lines. It can be seen that GDANN (122.80 s) takes more time than SVM (102.3 s), EasyTL (27.44 s) and DANN (85.24 s).

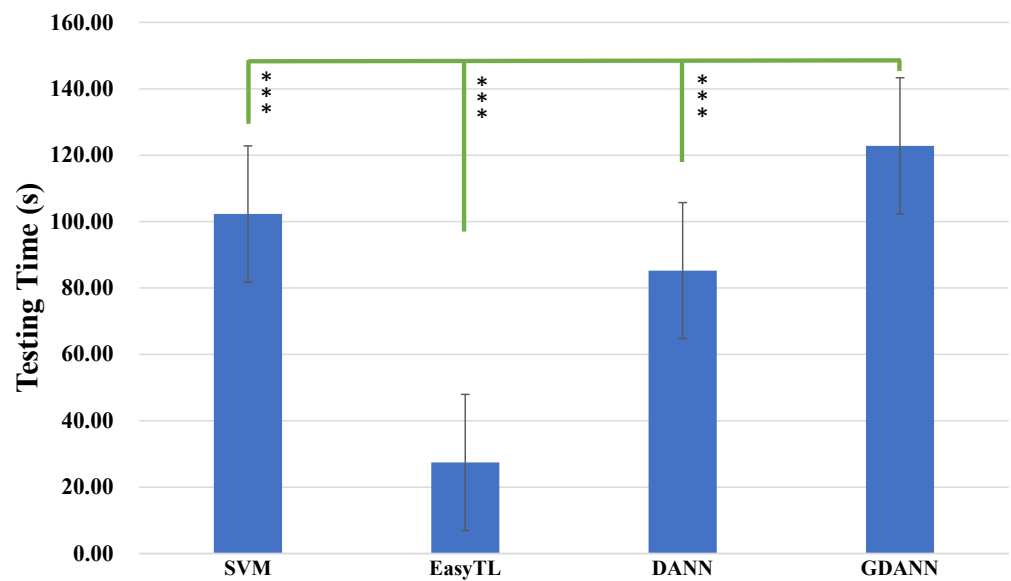


Figure 12. Testing time comparison between SVM, EasyTL, DANN and GDANN. (The testing time is the total time for one target subject to train the model and test results with the trained classifier, *** indicates $p < 0.001$).

5. Discussion

In the previous section, the proposed transfer learning model was been applied to data from the Industrial and NeuroScience Laboratory, University of Rome La Sapienza. Its performance was observed and compared with other standard classification techniques and transfer learning methods. The main observations of the results reported in Section 4 are as follows:

- (a) Analysis of the number of subjects in the source domain: As analyzed in Section 4.2, in multi-source transfer learning, the source number N_s is an important factor. More sources mean that we will integrate more data to predict fatigued driving. However, in view of the weak correlation between certain subjects, blindly increasing the number of sources may not improve accuracy and result in negative transfer and a calculation burden.
- (b) Analysis of the generated target domain data: As analyzed in Section 4.3, the generated fake data roughly conform to the distribution of the target domain data, effectively making up for the shortcomings of insufficient training data.
- (c) Comparison of the classification performance with the original model DANN: Due to the differences in the subjects, the classification performance of GDANN is also different. For those subjects who have good classification results in DANN, GDANN can give a slight improvement, while for those who do not perform well using the DANN method, GDANN will give a significant improvement. Since these accuracy values are not accidental (statistically verified), it can be said that a method for effectively performing EEG classification across subjects with multi-source training has been successfully proposed.
- (d) Swiftly approaching convergence of baseline: In the convergence comparison, GDANN quickly reaches the baseline of loss training (i.e., 1) before DANN, and fluctuates on this line. Furthermore, DANN can only converge to 1.05 and cannot go down, which is also a manifestation of its insufficient performance.
- (e) Comparison of classification accuracy with the state-of-the-art approaches: When comparing with some excellent related methods, such as supervised machine learning method SVM and transfer learning method EasyTL, as analyzed by Section 4.5, GDANN is seen to have better performance in terms of cross-subject EEG data prediction.

- (f) Comparison of the testing time: In terms of testing time, the proposed work takes more time than other methods. It should be noted that compared to DANN, it has an additional process to adapt to generate auxiliary fake data. In practical use, the training epoch can be appropriately reduced to reduce the time. Most of the time is spent on training the model, and when the model training is completed, its high-efficiency performance can always be used.

6. Conclusions

In this paper, we propose an improved DANN-based transfer learning model, GDANN, and apply it for EEG-based cross-subject fatigue mental state prediction. Our GDANN model combines GAN with transfer model to make appropriate improvements and optimizations by balancing the difference in the number of samples between the source domain and the target domain, selecting the best Top N source domain subjects to experiment and extracting the invariant features of the target domain as much as possible. The transfer learning model can be conducted across different domain and data tasks. The experimental results show that the performance of GDANN is better than that of DANN, SVM and EasyTL. GDANN improves the EEG classification accuracy by about 11.9% with the original model DANN, which proposes a potentially powerful solution for fatigue state detection during driving.

In the future, we intend to enhance the simulation capabilities (by trying to cover all the edge data of the target domain) of GAN (N_{tg} and N_{td}) to improve the performance of our model. Moreover, the size of the experimental group was not very big; therefore, one of the next steps could be to enlarge it to further validate the proposed method.

Author Contributions: Conceptualization, H.Z.; Data curation, G.B. and F.B.; Formal analysis, H.Z.; Funding acquisition, H.Z. and W.K.; Investigation, H.Z.; Methodology, H.Z., X.L. and G.B.; Resources, H.Z.; Software, X.L. and Y.Z.; Validation, X.L., G.B. and Y.Z.; Visualization, X.L.; Writing—original draft, H.Z.; Writing—review and editing, P.A., G.D.F., N.S., W.Z. and F.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China with grant No. 2017YFE0118200, also supported by NSFC with grant No. 62076083. The authors also thank the National International Joint Research Center for Brain-Machine Collaborative Intelligence with grant No. 2017B01020, Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province with grant No. 2020E10010, Fundamental Research Funds for the Provincial Universities of Zhejiang with grant No. GK209907299001-008, the Graduate Scientific Research Foundation of Hangzhou Dianzi University with grant No.CXJJ2020086, and Industrial Neuroscience Laboratory of University of Rome La Sapienza.

Institutional Review Board Statement: In this section, please add The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Ethics Committee of the Department of Physiology and Pharmacology of Sapienza University of Rome (Roma, 21/4/2016).

Informed Consent Statement: Any research article describing a study involving humans should contain this statement. Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The EEG data used to support the findings of this study are from the Industrial and NeuroScience Laboratory, University of Rome La Sapienza, and are restricted by the Ethics Committee of University of Rome La Sapienza, in order to protect SUBJECT PRIVACY. Data are available from the co-author: Gianluca Borghini (gianluca.borghini@uniroma1.it) for researchers who meet the criteria for access to confidential data.

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

Abbreviations

The following abbreviations are used in this manuscript:

EEG	Electroencephalogram
DANN	Domain-Adversarial Neural Network
GDANN	Generative-DANN
GAN	Generative Adversarial Networks
SVM	Support Vector Machine
EasyTL	Easy Transfer Learning
BCI	Brain-computer Interaction
PSD	Power Spectral Density
TAV	the Attentional and Vigilance task
WUP	warm up
DROW	drowsy
ICA	Independent Component Analysis
NASA-TLX	National Aeronautics and Space Administration–Task Load Index
t-SNE	t-distributed Stochastic Neighbor Embedding

References

- Lal, S.K.; Craig, A. A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* **2001**, *55*, 173–194. [[CrossRef](#)]
- World Health Organization. *Global Status Report on Road Safety: Time for Action*; World Health Organization: Geneva, Switzerland, 2009.
- 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](#)]
- Chaudhuri, A.; Routray, A. Driver Fatigue Detection Through Chaotic Entropy Analysis of Cortical Sources Obtained From Scalp EEG Signals. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 185–198. [[CrossRef](#)]
- Zhang, F.; Su, J.; Geng, L.; Xiao, Z. Driver fatigue detection based on eye state recognition. In Proceedings of the IEEE 2017 International Conference on Machine Vision and Information Technology (CMVIT), Singapore, 17–19 February 2017; pp. 105–110.
- Åkerstedt, T.; Gillberg, M. Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.* **1990**, *52*, 29–37. [[CrossRef](#)]
- Beurskens, A.J.; Bültmann, U.; Kant, I.; Vercoulen, J.H.; Bleijenberg, G.; Swaen, G.M. Fatigue among working people: Validity of a questionnaire measure. *Occup. Environ. Med.* **2000**, *57*, 353–357. [[CrossRef](#)]
- Liu, W.; Qian, J.; Yao, Z.; Jiao, X.; Pan, J. Convolutional Two-Stream Network Using Multi-Facial Feature Fusion for Driver Fatigue Detection. *Future Internet* **2019**, *11*, 115. [[CrossRef](#)]
- Smith, M.R.; Coutts, A.J.; Merlini, M.; Deprez, D.; Lenoir, M.; Marcora, S.M. Mental fatigue impairs soccer-specific physical and technical performance. *Med. Sci. Sports Exerc.* **2016**, *48*, 267–276. [[CrossRef](#)]
- Zeng, H.; Yang, C.; Dai, G.; Qin, F.; Zhang, J.; Kong, W. EEG classification of driver mental states by deep learning. *Cogn. Neurodyn.* **2018**, *12*, 597–606. [[CrossRef](#)]
- Nguyen, T.; Ahn, S.; Jang, H.; Jun, S.C.; Kim, J.G. Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Sci. Rep.* **2017**, *7*, 43933. [[CrossRef](#)] [[PubMed](#)]
- Di Flumeri, G.; Borghini, G.; Aricò, P.; Sciaraffa, N.; Lanzi, P.; Pozzi, S.; Vignali, V.; Lantieri, C.; Bichicchi, A.; Simone, A.; et al. EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Front. Hum. Neurosci.* **2018**, *12*, 509. [[CrossRef](#)]
- Zeng, H.; Yang, C.; Zhang, H.; Wu, Z.; Zhang, J.; Dai, G.; Babiloni, F.; Kong, W. A lightGBM-based EEG analysis method for driver mental states classification. *Comput. Intel. Neurosci.* **2019**, *2019*, 3761203. [[CrossRef](#)] [[PubMed](#)]
- Chai, R.; Tran, Y.; Naik, G.R.; Nguyen, T.N.; Ling, S.H.; Craig, A.; Nguyen, H.T. Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network. In Proceedings of the IEEE 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 4654–4657.
- Rahma, O.N.; Rahmatillah, A. Drowsiness analysis using common spatial pattern and extreme learning machine based on electroencephalogram signal. *J. Med. Signals Sens.* **2019**, *9*, 130. [[CrossRef](#)] [[PubMed](#)]
- Huo, X.Q.; Zheng, W.L.; Lu, B.L. Driving fatigue detection with fusion of EEG and forehead EOG. In Proceedings of the IEEE 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 897–904.
- San, P.P.; Ling, S.H.; Chai, R.; Tran, Y.; Craig, A.; Nguyen, H. EEG-based driver fatigue detection using hybrid deep generic model. In Proceedings of the IEEE 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 800–803.
- Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
- Yao, Y.; Doretto, G. Boosting for transfer learning with multiple sources. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1855–1862.
- Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]

21. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661.
22. Fairclough, S.H.; Venables, L.; Tattersall, A. The influence of task demand and learning on the psychophysiological response. *Int. J. Psychophysiol.* **2005**, *56*, 171–184. [[CrossRef](#)]
23. Vecchiato, G.; Borghini, G.; Aricò, P.; Graziani, I.; Maglione, A.G.; Cherubino, P.; Babiloni, F. Investigation of the effect of EEG-BCI on the simultaneous execution of flight simulation and attentional tasks. *Med. Biol. Eng. Comput.* **2016**, *54*, 1503–1513. [[CrossRef](#)]
24. Lal, S.K.; Craig, A. Driver fatigue: Electroencephalography and psychological assessment. *Psychophysiology* **2002**, *39*, 313–321. [[CrossRef](#)]
25. Maglione, A.; Borghini, G.; Aricò, P.; Borgia, F.; Graziani, I.; Colosimo, A.; Kong, W.; Vecchiato, G.; Babiloni, F. Evaluation of the workload and drowsiness during car driving by using high resolution EEG activity and neurophysiologic indices. In Proceedings of the IEEE 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 6238–6241.
26. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.
27. Jung, T.P.; Makeig, S.; Humphries, C.; Lee, T.W.; Mckeown, M.J.; Iragui, V.; Sejnowski, T.J. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **2000**, *37*, 163–178. [[CrossRef](#)]
28. Pfurtscheller, G.; Da Silva, F.L. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin. Neurophysiol.* **1999**, *110*, 1842–1857. [[CrossRef](#)]
29. Pal, M.; Bandyopadhyay, S.; Bhattacharyya, S. A Many Objective Optimization Approach for Transfer Learning in EEG Classification. *arXiv* **2019**, arXiv:1904.04156.
30. Tzeng, E.; Hoffman, J.; Saenko, K.; Darrell, T. Adversarial discriminative domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7167–7176.
31. Sasaki, Y. The Truth of the f-Measure. 2007. Available online: <https://www.cs.odu.edu/~mukka/cs795sum10dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf> (accessed on 30 July 2018).
32. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]
33. Zhang, H.; Weng, T.W.; Chen, P.Y.; Hsieh, C.J.; Daniel, L. Efficient neural network robustness certification with general activation functions. In Proceedings of the Thirty-Second Conference on Neural Information Processing Systems, Montreal, QC, Canada, 2–8 December 2018.
34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Burges, C.J. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]
36. Wang, J.; Chen, Y.; Yu, H.; Huang, M.; Yang, Q. Easy transfer learning by exploiting intra-domain structures. In Proceedings of the IEEE 2019 International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019; pp. 1210–1215.
37. Kong, W.; Zhou, Z.; Jiang, B.; Babiloni, F.; Borghini, G. Assessment of driving fatigue based on intra/inter-region phase synchronization. *Neurocomputing* **2017**, *219*, 474–482. [[CrossRef](#)]
38. Zeng, H.; Zhang, J.; Zakaria, W.; Babiloni, F.; Gianluca, B.; Li, X.; Kong, W. InstanceEasyTL: An Improved Transfer-Learning Method for EEG-Based Cross-Subject Fatigue Detection. *Sensors* **2020**, *20*. [[CrossRef](#)]
39. Wattenberg, M.; Viégas, F.; Johnson, I. How to Use t-SNE Effectively. *Distill* **2016**. [[CrossRef](#)]
40. Wilcoxon, F. Individual comparisons by ranking methods. In *Breakthroughs in Statistics*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 196–202.