Benchmarking EEG-based Cross-dataset Driver Drowsiness Recognition with Deep Transfer Learning

Jian Cui[1]*, Liqiang Yuan[2], Ruilin Li[2], Zhaoxiang Wang[1], Dongping Yang[1], Tianzi Jiang[1,3]

***Abstract*—It usually takes a long time to collect data for calibration when using electroencephalography (EEG) for driver drowsiness monitoring. Cross-dataset recognition is desirable since it can significantly save the calibration time when an existing dataset is used. However, the recognition accuracy is affected by the distribution drift problem caused by different experimental environments when building different datasets. In order to solve the problem, we propose a deep transfer learning model named Entropy-Driven Joint Adaptation Network (EDJAN), which can learn useful information from source and target domains simultaneously. An entropy-driven loss function is used to promote clustering of target-domain representations and an individual-level domain adaptation technique is proposed to alleviate the distribution discrepancy problem of test subjects. We use two public driving datasets SEEG-VIG and SADT to test the model on the cross-dataset setting. The proposed model achieved an accuracy of 83.3% when SADT is used as source domain and SEED-VIG is used as target domain and 76.7% accuracy on the reverse setting, which is higher than the other SOTA methods. The results are further analyzed with both global and local interpretation methods. Our work illuminates a promising direction of using EEG for calibration-free driver drowsiness recognition.**

Index Terms— Deep transfer learning, cross-dataset, driver drowsiness recognition, EEG, CNN

I. Introduction

Driver drowsiness is one of the major contributing factors of road accidents. Development of driver drowsiness monitoring systems is of utmost importance for road safety and prevention of car accidents. Electroencephalogram (EEG), which is thought to be the most practical non-invasive method for collecting brain dynamics due to its high temporal resolution and low cost, has been extensively investigated for monitoring driver drowsiness. To accurately recognize the mental states with EEG, a long calibration time is usually spent to collect data from the same subject (within-subject setting) or from other subjects under the same experiment condition (cross-subject setting) [1] for training the classifiers. The prolonged time spent for calibration severely hinders wide adoption of EEG. Cross-dataset recognition is crucial for development of a calibration-free system since it can significantly save the calibration time when an existing dataset is used. However, the task is challenging since the recognition accuracy is affected by the problem of distribution drift which is caused by various factors, such as different driving environments, capturing devices, background noise, types of electrodes, etc.

In order to solve the problem, we propose a deep transfer learning model named Entropy-Driven Joint Adaptation Network (EDJAN) for EEG-based cross-dataset driver drowsiness recognition. The model is novel in the following aspects:

- The model uses state-of-the-art (SOTA) EEG-based driver drowsiness recognition model Interpretable Convolutional Neural Network (ICNN) [1] as backbone under a joint adaptation training framework.
- An entropy-driven loss function is plugged in the model to promote clustering of the target samples and thus encourage the model to learn domain-invariant features from the target unlabeled data.
- An individual-level domain adaptation technique named 'Individual Batch Normalization (IndBN)' is proposed to adapt the trained model to different test subjects alleviating the problem of individual-level distribution discrepancy.

II. Related work

Deep transfer learning techniques improve the performance of a deep learning model on the target domain containing unlabeled data by encouraging the model to learn domain-invariant representations across different domains. In the context of EEG, deep transfer learning techniques have been used for improving the cross-subject classification accuracy, where the source domain is formed by labeled data from several subjects and the target domain is the set of unlabeled data from the test subject. For example, Zhao *et al.* [2] proposed to use joint distribution discrepancy (JDD) as a regularizer in deep networks to alleviate domain shifts from different subjects for cross-subject motor imagery (MI) classification. Bao *et al.* [3] reduced maximum mean discrepancy (MMD) between source and target domain features to improve cross-subject emotion recognition. Cai *et al.* [4] projected EEG signals into a hierarchical subspace and explored the domain invariant features in each layer, with the objective to improve the performance of EEG-based epilepsy classifications. Peng *et al.* [5] attempted to learn domain-invariant features between different epileptic patients by performing an adversarial training procedure to match the aggregated posterior of the embedding space to a Riemannian

[1] Research Center for Augmented Intelligence, Research Institute of Artificial Intelligence, Zhejiang Lab, Hangzhou 311121, China

[2] Nanyang Technological University, Singapore

[3] Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

*Corresponding author: Jian Cui (email: cuijian@zhejianglab.com)

* This work was partially supported by the STI2030-Major Projects 2021ZD0200201, the National Natural Science Foundation of China (Grant No. 62201519), and Scientific Projects of Zhejiang Lab (No. 111012-AA2301, No. 2022KI0AC02, No. 2022ND0AN01).

manifold-based prior that contains cross-domain information. For cross-subject motor imaging classification, Tang *et al*. [6] first trained a dense convolutional neural network to obtain high-level discriminative features from MI data. A classifier and a conditional domain discriminator were trained in an adversarial manner to learn common intra-subject EEG features. Hong *et al*. [7] leveraged one global discriminator to align the marginal distribution between source and target domain, and one local discriminator to reduce the conditional distribution discrepancy between sub-domains.

In summary, existing work has shown promising results on using deep transfer learning to discover common EEG features across different subjects under the same experimental conditions, while the more challenging problem of cross-dataset driver drowsiness recognition has not yet been considered. Actually, in practical application limited by space and resources, the data collected from real driving environment are usually different from that collected for training the classifier, which could cause group-level drifts of the data due to different recording devices, background noise, types of electrodes, etc. Therefore, in this paper we consider the challenging topic of cross-dataset driver drowsiness recognition, which is a crucial step for delivering a calibration-free driver drowsiness detection system.

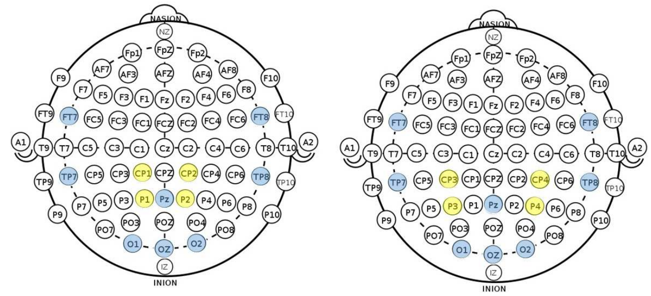
III. MATERIALS AND METHODS

A. Data preparation

SADT Dataset: The first dataset is a public dataset that was collected from subjects conducing a sustained-attention driving task (SADT) in a virtual reality simulator [8]. In the experiment, lane-departure events were randomly introduced to drift the car away from the central lane. Reaction time was used to measure the level of drowsiness of the subjects. EEG signals were recorded with a Scan SynAmps2 Express system providing 32 Ag/AgCl electrodes (30 EEG electrodes and 2 reference electrodes). The signals were processed with 1-50Hz band-pass filter and down-sampled to 128 Hz. EEG samples of 3-second length prior to the car deviation events for each trail were extracted. The samples were labeled according to the corresponding reaction time and balanced for each subject. Finally, we got a dataset containing 112 samples from 11 subjects, as it is shown in TABLE I. The processed dataset is publicly available from [9]. Details on the processing steps can be found in [1, 10, 11].

TABLE I. NUMBER OF SAMPLES FOR EACH SUBJECT

Subject ID	**SADT Dataset**		**SEED-VIG Dataset**	
	Alert	***Drowsy***	***Alert***	***Drowsy***
1	94	94	114	114
2	66	66	83	83
3	75	75	226	226
4	74	74	249	249
5	112	112	97	97
6	83	83	321	321
7	51	51	196	196
8	132	132	71	71
9	157	157	202	202
10	54	54	352	352
11	113	113	210	210
12	-	-	162	162
Total	1011	1011	2283	2283



(a) (b)

Figure 1. The selected channels for (a) SADT dataset and (b) SEED-VIG dataset. The channels in blue are common ones in both datasets. The channels in yellow colors are additionally selected for a less strict match.

SEED-VIG Dataset: The second dataset is the public SEED-VIG dataset [12], which was collected from subjects conducting a monotonous driving task in a virtual simulator. The EEG data from 18 channels were recorded using the Neuroscan system. The eye closure (PERCLOS) [13] information was obtained from Senso-Motoric-Instrument (SMI) eye-tracking glasses, which was used for labeling the data. We down-sampled the EEG signals to 128Hz and processed the data with a low-pass filter of 1Hz. EEG samples with a 3-second length were extracted prior to the PERCLOS evaluation event. We followed the procedure adopted in [14] by labeling the samples as 'alert' when PERCLOS is lower than 0.35 and samples as 'drowsy' when PERCLOS is higher than 0.7, while the samples in the middle range were discarded. We further discarded sessions with less than 50 samples of either class and balanced the class for each session by selecting the most alert or drowsiest ones. In this way, we have 4566 samples in total from 12 subjects for the test, as it is shown in TABLE I.

Channel selection: The two datasets have eight channels in common which are 'FT7', 'FT8', 'TP7', 'TP8', 'Pz', 'O1', 'Oz', and 'O2'. We additionally selected four channels from each dataset, which are 'CP1', 'CP2', 'P1', and 'P2' from SADT dataset corresponding to 'CP3', 'CP4', 'P3', and 'P4' from SEED-VIG dataset, respectively, since they are nearby topologically, as it is shown in Fig 1. The final selected channels for SADT dataset are 'FT7', 'FT8', 'TP7', 'CP1', 'CP2', 'TP8', 'P1', 'Pz', 'P2', 'O1', 'Oz', and 'O2'. The final selected channels for SEED-VIG dataset are 'FT7', 'FT8', 'TP7', 'CP3', 'CP4', 'TP8', 'P3', 'Pz', 'P4', 'O1', 'Oz', and 'O2'.

B. Model design

The structure of the proposed EDJAN model is shown in Fig. 2. We have adopted several novel techniques in the model design to boost performance on the cross-dataset recognition task. Details are illustrated as below.

Backbone model: ICNN [1] is used as backbone in the framework. The model uses pointwise convolutional layers, mimicking spatial filtering techniques [15], to extract a set of new signals related to the classification task from the raw highly correlated multi-channel recordings of EEG, which is followed by depthwise convolutional layers to extract temporal features. The model contains a ReLU activation layer

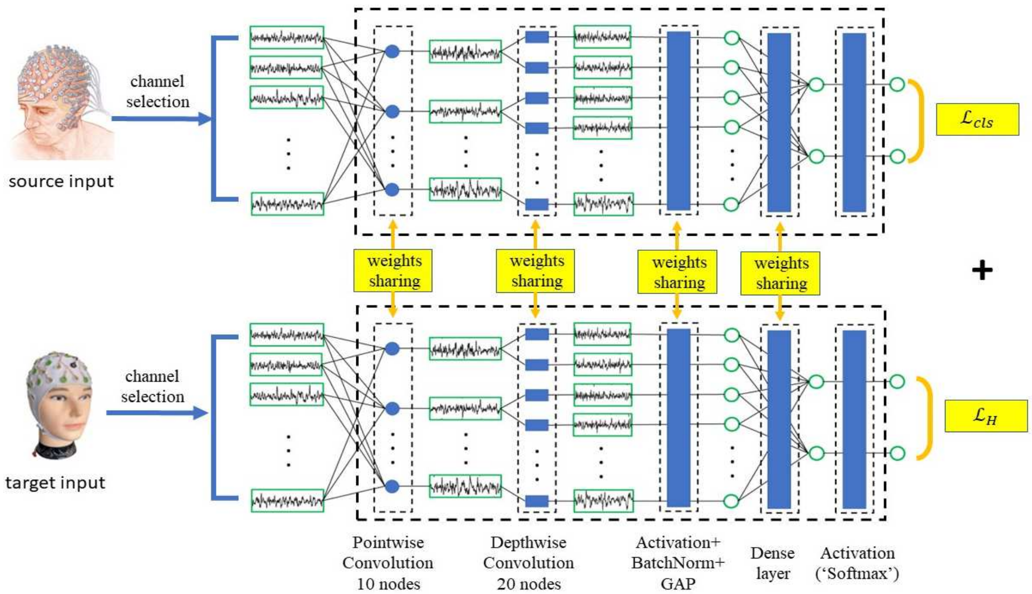


Figure 2. The structure of the proposed model EDJAN.

and a batch normalization layer, followed by a global average pooling (GAP) layer to summarize the features into a single vector. The model has shown superior performance on extracting shared features across different subjects for the driver drowsiness recognition task [1, 16]. In addition, the model comes with an interpretation technique which allows sample-wise analysis of the results.

Joint Adaptation Training: In order to encourage the model to learn useful information from both the source and target datasets simultaneously, we design a joint adaptation training framework which uses dual backbone models with shared parameters. In each training iteration, one model takes mini-batch from source domain as input and the classification loss is given by

$$L_{cls} = -\sum y_S log\hat{y}_S, \tag{1}$$

where y_S is the one-hot ground-truth label vector and $\hat{y}_S$ is the predicted probability. The other model takes mini-batch from target domain as input and use the entropy loss L_E for domain adaptation. Updating the final loss function

$$L = L_{cls} + L_E \tag{2}$$

encourages the model to learn domain invariant features from the data.

Entropy-Driven Loss Function L_E : The entropy minimization technique is used to regulate training with the target domain data, and the loss function L_E is given by

$$L_E = -\sum \hat{y}_T log\hat{y}_T \quad , \tag{3}$$

where $\hat{y}_T$ is the predicted probability. Entropy minimization is a classical semi-supervised learning method that can encourage well-clustered feature representation of unlabeled data by penalizing high-density regions lying around classification boundaries in the latent space. The method is well-suited to the task of drowsiness recognition since the EEG signals exhibit distinctive patterns under the alert and drowsy states [1], which could form strong clusters in a proper feature space.

Individual Batch Normalization (IndBN): When the training is completed, the model still cannot be satisfactorily applied for every subject from the target domain, since the feature distribution of each subject varies significantly. In order to solve the problem, we propose an individual-level domain adaptation technique named 'Individual Batch Normalization (IndBN)' and the method is defined by

$$y = \frac{x_{TS} - E[x_{TS}]}{\sqrt{Var[x_{TS}] + \epsilon}} * \gamma + \beta, \tag{4}$$

where x_{TS} is the set of samples from the same subject of the target domain, and γ and β are trained parameters of the model. The data from each subject is aligned with the global distribution learned by the model by removing mean and divided by variation. The method is applied to the batch normalization layer and only used in the classification phase.

C. *Methods for comparison*

We compare the performance of the proposed model with baseline methods and several well-known deep transfer learning techniques, and they are illustrated as below.

BandPower+SVM: Band power features are extracted from four frequencies bands of Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz) and Beta (12–30 Hz) from each of 12 channels forming a vector of length 4×12=48. Support Vector Machine (SVM) is used as the classifier. This method without any transfer learning technique is used as baseline.

Vanilla ICNN: the backbone model is used as another baseline to evaluate the effect of the proposed entropy loss L_E on the proposed model.

AdaBN: As a variation of the standard batch normalization layer, this deep transfer learning technique was designed by Li *et al* [17] to solve the distribution drift problem between different domains. We have implemented this method in order to compare it with our proposed IndBN technique.

MMD: Maximum Mean Discrepancy (MMD) is a well-known distribution distance metric widely used in deep transfer learning frameworks [18, 19]. The MMD distance of features after the global average pooling layer of ICNN is measured and minimized during training.

Deep CORAL: Similar to MMD, the CORAL method [20] aligns the second-order statistics of the source and target distributions with a linear transformation [21]. The CORAL distance of features after the global average pooling layer of ICNN is measured and minimized during training.

Adversarial [22]: The method uses a gradient reversal layer and a domain classier connected to the feature extractor via a gradient reversal layer to encourage the model to learn domain invariant features.

Upper Bound: We obtain cross-subject recognition accuracy within the same dataset using Vanilla ICNN as the upper limit in this study. Since the training and testing data are from the same dataset (under the leave-one-subject-out training scheme), which is free of the problem of distribution shift, the obtained accuracy will be in principle higher than the other methods tested across different datasets.

D. Implementation details

For the backbone ICNN model [1], we use N_1=10 spatial pointwise filters in the first layer and $2N_1$=20 depthwise filters in the second layer. We use the Adam method to train the models. We set batch size as 50 and used default parameters (η = 0.001, β_1= 0.9, β_2= 0.999) of Adam for optimization. IndBN is an effective technique and used for all the methods except for AdaBN for comparison. For each model, we train with 11 epoches (which is the optimal epoch for ICNN [1]) for 10 times and the average accuracies are reported.

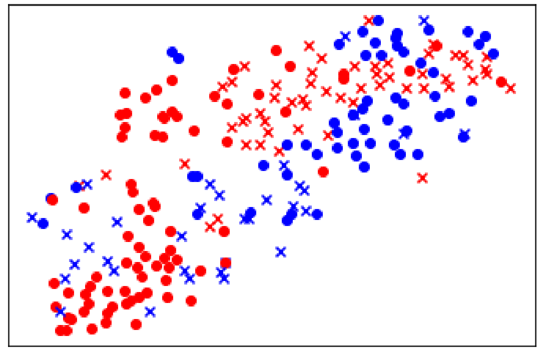
IV. Evaluation on the Proposed Method

A. Results

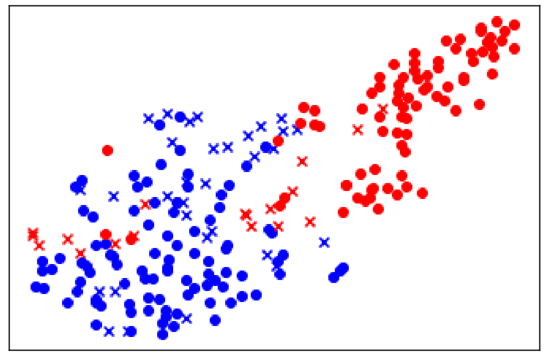
As it can be seen from the results shown in TABLE II, the proposed model achieved an accuracy of 83.3% when SADT is used as source domain and SEED-VIG is used as target domain, which is 5.9%-10.9% higher than the other deep transfer learning methods. On the reverse setting, the proposed method achieved an accuracy of 76.7%, which is 1.3%-7.7% higher than the other deep transfer learning methods. The backbone ICNN model has shown advantage in extracting shared features across different subjects over the conventional BandPower+SVM method. The Vanilla ICNN model also has a higher accuracy than the AdaBN method, showing the advantage of our proposed IndBN technique. It can also be noticed that not all the deep transfer learning techniques are better than baseline Vanilla ICNN as negative transfer happens, e.g., Deep CORAL on the SEED-VIG to SADT setting. In comparison, the proposed method beats the baseline on both settings, validating the effectiveness of the proposed entropy-driven loss function over the other deep transfer learning techniques. The proposed method is approaching the upper bound 78.0% on the SEED-VIG to SADT setting, while there is still an 8.2% room for improvement on the reverse setting.

TABLE II. classification accuracies of different methods

Method	SADT to SEED-VIG	SEED-VIG to SADT
BandPower+SVM	52.1%	67.0%
Vanilla ICNN	76.8%	72.4%
AdaBN	72.4%	69.6%
MMD	77.3%	75.4%
Deep CORAL	77.1%	71.5%
Adversarial	77.4%	72.3%
Propose method	**83.3%**	**76.7%**
Upper Bound	*91.5%*	*78.0%*



(a)



(b)

Figure 3. Visualization of feature distribution of Subject 5 for (a) Vanilla ICNN and (b) the proposed model from SEED-VIG to SADT using TSNE dimension reduction. The points with red and blue colors are predicted with ‘drowsy’ and ‘alert’ labels by the models, respectively, and the cross markers indicate that the predictions are wrong.

B. Interpretation

In this section, we analyze the results with both global and local interpretation techniques. We take the data from Subject 5 for example on the SEED-VIG to SADT setting and investigate how the proposed method improves the recognition accuracy over the baseline Vanilla ICNN model. We visualize the learned feature distribution obtained after the GAP layer of the model using TSNE dimension reduction. As it can be seen in Fig. 3 (a), the Vanilla ICNN model fails to separate the two classes with a clear decision boundary, resulting a large number of wrongly classified samples. In comparison, the proposed method could push the features away from the boundary and thus two dense clusters are formed with a clear boundary between them, as it is shown in Fig. 3 (b). There are less wrongly classified samples in comparison to the Vanilla ICNN model.

We further analyze individual interpretation results with the technique proposed in [1]. The interpretation technique allows the local areas of the signals that contribute most to the classification to be highlighted. Two samples are shown in Fig. 4. The first sample (Fig. 4 (a) (b)) has the ground-truth label of 'alert'. There are a high portion of Beta waves in the beginning part of the sample, which could be caused by electromyography (EMG) activities [10]. The proposed model sensitively captures these features as evidence of alertness, as it is shown in Fig. 4(b). Despite the Vanilla ICNN model identified similar features for this sample, it reaches the opposite conclusion due to the problem of distribution drift of the two domains. The second samples shown in Fig. 4 (c) (d) has the ground-truth label of 'drowsy'. The proposed model has correctly located local features from CP4, Pz and O2 areas which are featured with Alpha spindles [23] as evidence of the 'drowsy' label [1], as it is shown in Fig. 4 (d). In comparison, the features captured by the Vanilla ICNN model are slightly different from the proposed model – it concentrated more on the TP8 channel while ignored the features contained in the occipital channels, which could explain the reason for the wrong decision.

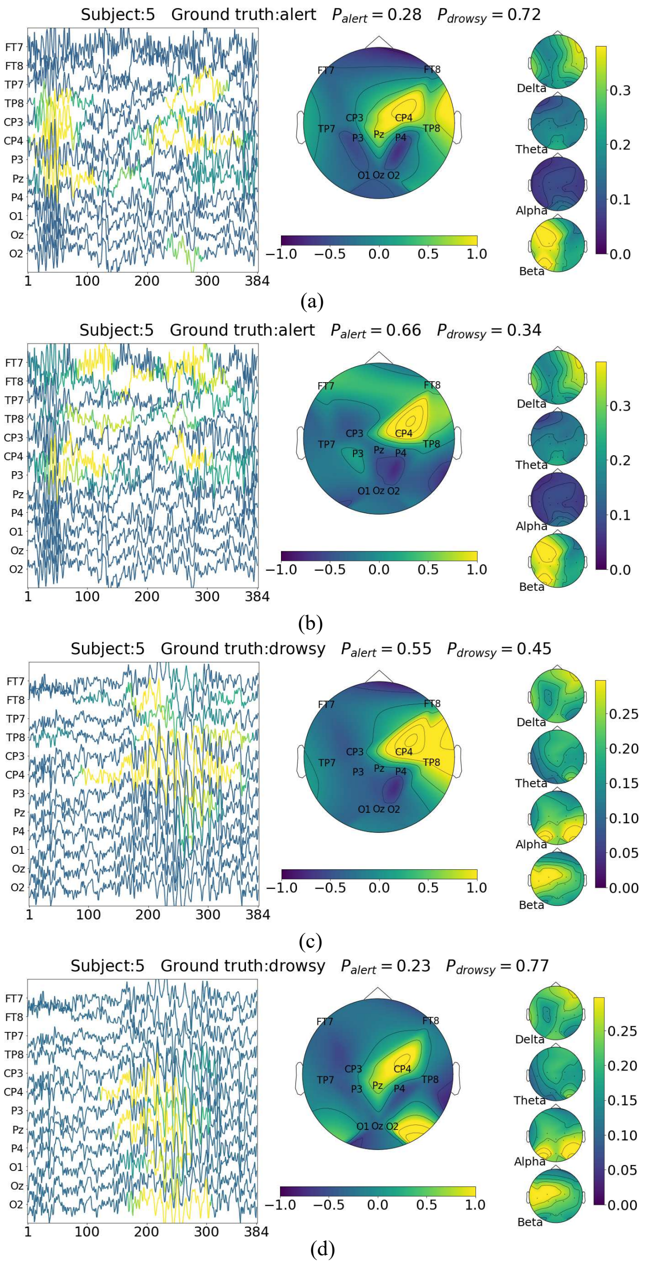


Figure 4. Individual interpretation results of two samples from Subject 5 on the SEED-VIG to SADT setting. The first sample is shown in (a) and (b), while the second sample is shown in (c) and (d). (a) and (c) are the results produced by Vanilla ICNN. (b) and (d) are the results produced by the proposed method. In each sub-figure, the interpretation result for the original sample is shown in the first column. The topological map obtained by averaging the values for each channel is shown in the second column. The band power features calculated for each channel are shown in the third column.

V. Conclusion

In this paper, we proposed a deep transfer learning model named Entropy-Driven Joint Adaptation Network (EDJAN) for cross-dataset driver drowsiness recognition. An entropy-driven loss function was used to promote clustering of target-domain representations and an individual-level domain adaptation technique was proposed to alleviate the distribution discrepancy problem of test subjects. The model showed better performance than the other deep transfer learning techniques on the cross-dataset recognition tasks. Our work illuminates a promising direction of using EEG for calibration-free driver drowsiness recognition.

References

[1] J. Cui, et al., "EEG-Based Cross-Subject Driver Drowsiness Recognition With an Interpretable Convolutional Neural Network," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[2] X. Zhao, et al., "Deep neural network with joint distribution matching for cross-subject motor imagery brain-computer interfaces," *BioMed research international*, 2020.

[3] G. Bao, et al., "Two-level domain adaptation neural network for EEG-based emotion recognition," *Frontiers in Human Neuroscience*, 2021. 14: p. 605246.

[4] W. Cai, et al., "Hierarchical Domain Adaptation Projective Dictionary Pair Learning Model for EEG Classification in IoMT Systems," *IEEE Transactions on Computational Social Systems*, 2022.

[5] P. Peng, et al., "Domain adaptation for epileptic EEG classification using adversarial learning and Riemannian manifold," *Biomedical Signal Processing and Control*, 2022. 75: p. 103555.

[6] X. Tang, and X. Zhang, "Conditional adversarial domain adaptation neural network for motor imagery EEG decoding," *Entropy*, 2020. 22(1): p. 96.

[7] X. Hong, et al., "Dynamic joint domain adaptation network for motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2021. 29: p. 556-565.

[8] Z. Cao, et al., "Multi-channel EEG recordings during a sustained-attention driving task," *Scientific data*, 2019. 6(1): p. 1-8.

[9] J. Cui, "EEG driver drowsiness dataset," 2021. Available from: https://figshare.com/articles/dataset/EEG_driver_drowsiness_dataset/14273687

[10] Cui, J., et al., "A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG, " *Methods*, 2021.

[11] J. Cui, et al. "Subject-Independent Drowsiness Recognition from Single-Channel EEG with an Interpretable CNN-LSTM model," in *2021 International Conference on Cyberworlds (CW)*. 2021. IEEE.

[12] BCMI, "SEED Dataset- SEED-VIG," 2019, Available from: https://bcmi.sjtu.edu.cn/home/seed/seed-vig.html.

[13] U. Trutschel, et al. "*PERCLOS: An alertness measure of the past*, " in *Driving Assesment Conference*. 2011. University of Iowa.

[14] W.-L. Zheng, and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *Journal of neural engineering*, 2017. **14**(2): p. 026017.

[15] M.Z. Baig, N. Aslam, and H.P. Shum, "Filtering techniques for channel selection in motor imagery EEG applications: a survey," *Artificial intelligence review*, 2020. **53**(2): p. 1207-1232.

[16] R. Li, et al., "Sample-Based Data Augmentation Based on Electroencephalogram Intrinsic Characteristics," *IEEE Journal of Biomedical and Health Informatics*, 2022. **26**(10): p. 4996-5003.

[17] Y. Li, et al., "Revisiting batch normalization for practical domain adaptation, " *arXiv preprint arXiv:1603.04779*, 2016.

[18] E. Tzeng, et al., "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.

[19] M. Long, et al. "Learning transferable features with deep adaptation networks. in International conference on machine learning," *PMLR*, 2015.

[20] B. Sun, J. Feng, and K. Saenko. "Return of frustratingly easy domain adaptation," *in Proceedings of the AAAI Conference on Artificial Intelligence*, 2016.

[21] B. Sun and K. Saenko. "Deep coral: Correlation alignment for deep domain adaptation," in *European conference on computer vision*. 2016.

[22] Y. Ganin, et al., "Domain-adversarial training of neural networks, " *The journal of machine learning research*, 2016. **17**(1): p. 2096-2030.

[23] M. Simon, et al., "EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions," *Clinical Neurophysiology*, 2011. **122**(6): p. 1168-1178.