Federated deep transfer learning for EEG decoding using multiple BCI tasks

Xiaoxi Wei
Dept. of Computing, Imperial College London
Brain & Behaviour Lab
London, UK
xiaoxi.wei18@imperial.ac.uk

A. Aldo Faisal
Brain & Behaviour Lab, Imperial College London
London, UK
Chair in Digital Health & Data Science, University of Bayreuth
Bayreuth, Germany
aldo.faisal@imperial.ac.uk

***Abstract*—Deep learning has been successful in BCI decoding. However, it is very data-hungry and requires pooling data from multiple sources. EEG data from various sources decrease the decoding performance due to negative transfer [1]. Recently, transfer learning for EEG decoding has been suggested as a remedy [2], [3] and become subject to recent BCI competitions (e.g. BEETL [4]), but there are two complications in combining data from many subjects. First, privacy is not protected as highly personal brain data needs to be shared (and copied across increasingly tight information governance boundaries). Moreover, BCI data are collected from different sources and are often based on different BCI tasks, which has been thought to limit their reusability. Here, we demonstrate a federated deep transfer learning technique, the Multi-dataset Federated Separate-Common-Separate Network (MF-SCSN) based on our previous work of SCSN [1], which integrates privacy-preserving properties into deep transfer learning to utilise data sets with different tasks. This framework trains a BCI decoder using different source data sets obtained from different imagery tasks (e.g. some data sets with hands and feet, vs others with single hands and tongue, etc). Therefore, by introducing privacy-preserving transfer learning techniques, we unlock the reusability and scalability of existing BCI data sets. We evaluated our federated transfer learning method on the NeurIPS 2021 BEETL competition BCI task. The proposed architecture outperformed the baseline decoder by 3%. Moreover, compared with the baseline and other transfer learning algorithms, our method protects the privacy of the brain data from different data centres.**

***Index Terms*—Machine Learning, Transfer Learning, Domain Adaptation, Brain-Computer-Interfaces (BCI), Electroencephalography, Ethic, Privacy-preserving, Federated Machine Learning**

I. Introduction

Deep Learning based EEG decoding has become a standard in BCI [5], [6] and unlocked state-of-the-art machine learning ideas to benefit neural engineering research. Deep learning approaches are usually data-hungry. Some previous studies deal with the lack of available EEG data by data-efficient approaches [7]–[9]. In recent years, the development of transfer learning in EEG decoding [2], [3] enables algorithms to learn more from combining different EEG data sets. However, most methods focus on only a single data set with a unified experiment setup or task. The scale of one EEG data set is usually limited to only dozens of subjects, unlike biomedical data sets with thousands, due to the difficulty and cost of EEG data collection. The international BEETL EEG competition [4] held at NeurIPS 2021 focused on cross-dataset EEG transfer learning and brought academic attention to utilising many EEG data sets across tasks for transfer learning with around 30 international competing teams. Several successful solution algorithms were proposed to tackle the BEETL challenge, which has provided some fundamental design principles for cross-dataset and cross-task EEG transfer learning. With examples showing that heterogeneous EEG data sets from different data centres and sources could be utilised for large-scale machine learning algorithms, EEG data sharing privacy becomes the next concern. Brainwaves contain rich privacy information that could be potentially decoded by algorithms, e.g. images, words and identities. Moreover, data sharing across data centres or countries is usually under strict restrictions, e.g. the EU GDPR data policy and the data dispute between China and the US.

There were some strategies for privacy-preserving in the machine learning literature. Federated learning [10] trains models on edge servers without exchanging the data. Data encryption methods encrypt raw data or parameters of the model [11]. This requires an encoding-decoding procedure which introduces extra computational cost. Users also need to decrypt the data following a certain protocol given by the algorithm provider, in which case the protocol could be potentially hacked. Similarly, transformation [12] adds cancelable noise to local gradients or parameters before uploading them to a central server. Methods of model splitting [13] are based on allocating different parameters to different data sets, thus protecting the model privacy of each individual. Multi-party computation [14] trains models locally first, then aggregation is done securely by a third party. However, the above methods are sometimes unsuitable or have not been tested for brainwave decoding, e.g. the robustness of adding noise to the low signal-to-noise ratio EEG has not yet been proved. And they introduce extra computational consumption,

Preprint Submitted to 2023 IEEE NER [©2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.] Address for correspondence: aldo.faisal@imperial.ac.uk . We acknowledge funding from UKRI Turing AI Fellowship to AAF.

which disadvantages the training for large-scale EEG data.

Privacy-preserving is drawing increasing attention in EEG decoding with the development of more accurate human intention decoders. There are a few studies in the EEG literature on privacy-preserving based on the above methods [15]–[18], or based on better protocol or user level system design [19]. Our previous work and some other recent studies [1], [18] utilised distributed feature extractors to deal with individual EEG differences while maintaining private information without extra cost for encryption. However, both studies can not handle learning from different tasks and protect inference-level privacy, thus limiting the use of large-scale data. Therefore, it is still a challenge for cross-data-centre transfer learning with different tasks in a privacy-preserving way.

In this study, we propose an architecture to combine privacy-preserving machine learning with deep transfer learning. The Multi-dataset Federated Separate-Common-Separate Network (MF-SCSN) integrates privacy-preserving properties into deep transfer learning EEG decoding on multiple tasks.

II. Method Development

A benchmark architecture, the shallow ConvNet [6], is used as the baseline model. The network includes a temporal layer to extract time-scale information, followed by a spatial layer to extract cross-channel features. Square non-linearity, average pooling and log non-linearity are then performed.

In this study, a privacy-preserving cross-dataset deep transfer learning architecture, the MF-SCSN, is proposed based on our previous study on inter-subject deep transfer learning [1]. The variability of EEG comes from several aspects. There are superficial variabilities like sensor locations, sensor impedance and devices. These differences could be handled in shallow layers of a transfer learning network. For the intrinsic variability of individual brains, functionalities could be potentially learnt and handled more precisely in deeper layers.

In light of this, the MF-SCSN consists of three main components. As shown in 2, the MF-SCSN separates both shallow layers and deeper layers to handle the variabilities while performing a joint feature extractor to learn common transferable knowledge across data sets.

The first set of components is the local branches (left side of the figure) as both feature extractors and 'keys' for data encryption. The shallow ConvNet above is used here as a feature extractor. Raw data from different data centres is encrypted into EEG features through the local branches. Local servers of data centres conduct the computation of feature extraction, and parameters (keys) are stored locally. In this way, the proposed architecture preserves both data-level privacy and parameter-level privacy.

The second component of the MF-SCSN is a common transfer network located in a cloud server. This is where transfer learning across different data sets happens. Encrypted features are received from data centres in the cloud for common feature extraction. Previous local feature extractors handle variabilities in data set distributions. The design purpose of common layers is to find a common distribution to which different data sets and distributions could transfer to. Unlike other encryption methods for privacy-preserving, the MF-SCSN has no encryption costs since the federated feature extractors encrypt data automatically. Moreover, it does not have a 'decryption' procedure because the encrypted features are exactly the inputs required by the common transfer layer. Therefore, the cloud server does not need to know any information about the local feature extractors (the 'keys'). This further increases the parameter-level security of the model.

Finally, the transferred features are delivered to the third component of MF-SCSN, i.e. the deep separate layers and heterogeneous classifiers. Its design is motivated by [4] showing that combining different classifiers for cross-dataset transfer learning can help overcome label inconsistency. In light of this, the third set of components contains some separate layers to deal with further differences across data sets, followed by local classifiers specified for different tasks. Besides the benefit of handling label inconsistency, predictions and labels are preserved locally in this way.

III. Evaluation Method

The proposed architecture was tested on the motor imagery task of the BEETL competition. Details of the data can be found in [4]. There are three source data sets with different devices and data collection protocols. Cho2017 [20] has 52 subjects performing left and right-hand motor imagery. Subjects 32,46 and 49 were not used due to data problems. BCICIV2a [21] contains nine subjects performing left-hand, right-hand, feet and tongue motor imagery. Subject 1,3,7,8 and 9 with higher data quality were selected as sources. PhysioMI [22], [23] has 109 subjects, among which we selected subject 1,7,17,24,28,31,33,34,35,42,49,52, 54,55,56,60,62,63,68,71,72,73,85,91,93,94 and 103 as sources. Test sets consist of 5 subjects with 200 trials each (1000 trials in total). Labels are left-hand, right-hand and feet motor imagery and rest state (250 trials each). As in the BEETL, final accuracies are reported on the weighted accuracy of the three classes - left-hand (LH), right-hand (RH) and 'other'.

To align and make use of all data sets, we extracted 3 seconds windows for all trials, because Cho2017 had a maximum trial length of 3 seconds. Note that this may not utilise the full potential of the 4-second target test data. 17 common EEG channels of the four data sets were selected (Fz, FC1, FC2, C5, C3, C1, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2). All trials are down-sampled to 200Hz. A 5th-order bandpass filter was applied between 4Hz and 32Hz, where motor imagery usually occurs. Normalisation across channels is done, followed by temporal normalisation across time steps.

Shallow ConvNet was used as the feature extractor. For both the baseline model(Shallow ConvNet) and the MF-SCSN, the feature size was aligned to 50 after the feature extractor. The common cloud network consists of three fully-connected layers with a feature size of 50 each. Before the classifiers, the separate layers consist of three fully-connected layers with a feature size of 50. All four data sets use individual classifiers according to their own tasks. During training, the

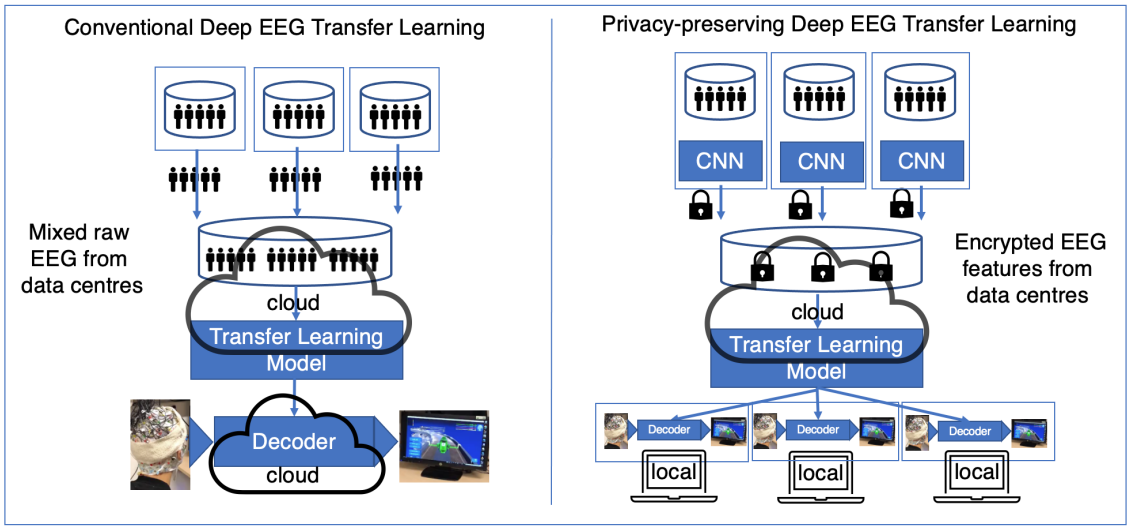

Fig. 1. The figure illustrates the differences between conventional deep transfer learning and privacy-preserving deep transfer learning with federated models.

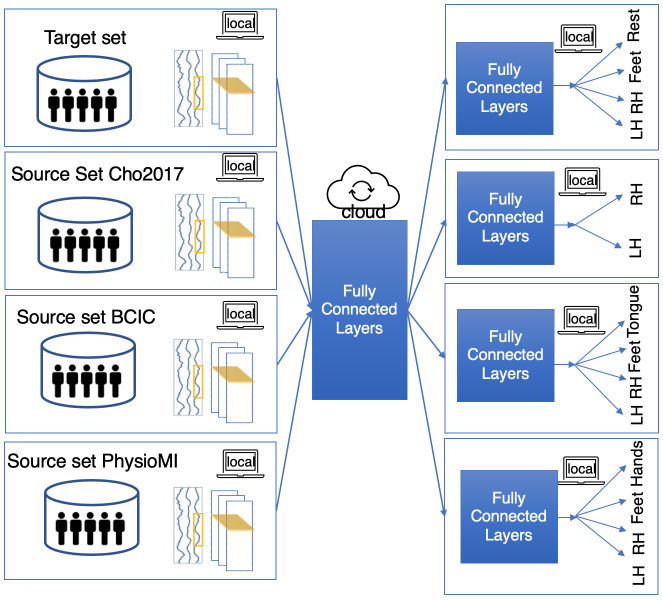

Fig. 2. Multi-dataset Federated Separate-Common-Separate Network (MF-SCSN). Raw data, feature extractors (the encryption keys), labels and predictions are all preserved locally.

target classier kept the original four classes. After prediction, feet and rest labels were combined as 'other' to report the final accuracy. For all four data sets, trial numbers are balanced to 2880 trials (the size of BCIC sources).To balance training sizes, a random sampling was done for Cho2017 (originally 9880 trials). Similarly, we augment the PhysioMI (Originally 2399 trials) and target data set to 2880 trials. A batch size of 10, a learning rate of 0.001 and a weight decay factor of 0.0005 were used for both shallow CovNet baseline and MF-SCSN. During the training of MF-SCSN, four batches of size 10 from branches (40 in total) were delivered to the cloud layers simultaneously and distributed back to each separate local branch after common feature extraction. In the target training set, subject 1-3 has 100 trials each. Subject 4-5 has 120 trials each. 20 trials each were used as the validation set for model selection. All randomization and initialization were conducted with a typical random seed for machine learning of 42 for reproducibility in the same setup and environment.

IV. Results

We tested the baseline shallow ConvNet and the MF-SCSN on the BEETL motor imagery task. Five subjects from two data sets were tested. The first three subjects (S1 S2 S3) are from the CybathlonIC data set [4] collected in an online closed-loop format. The latter two subjects are from the Weibo2014 data set [24]. Weibo2014 uses instructions on the screen to inform subjects to perform offline motor imagery without feedback and real-time control. The CybathlonIC used the Cybathlon 2020 BCI game for data collection. The intention of controlling the virtual car in a real-world setup with real-time feedback and interference made the brain signal more complex and noisier to decode.

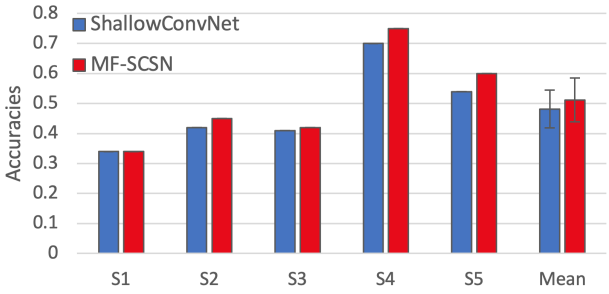

Fig. 3. Label-weighted decoding accuracy of subjects, their mean and standard error of the mean (SEM shown as error bar).

Considering the differences between the two data sets, we trained two data sets separately. Prediction results were reported based on their own models. In our previous work, a significant accuracy drop could be observed by simply adding more subjects to train the ShallowConvNet [1]. Therefore, here we used only the target subjects to train the ShallowConvNet as the baseline. For the MF-SCSN, S1-S3 are regarded as the target set and trained together with the source branches. Similarly, we trained another model combining S4 and S5 to classify each. As shown in figure 3, the decoding accuracies of S1-S3 are significantly lower than S4 and S5 in both the

baseline method and MF-SCSN. This reflects the challenge of real-world BCI decoding and the variance of the two data sets.

The main observation of this study is that, as in figure 3, the MF-SCSN outperformed the baseline methods. Among the 1000 trials (200 trials per subject), the MF-SCSN correctly classified 555 samples compared with 518 samples by the shallow ConvNet. In the 1000 testing samples, there are 250 left-hand/right-hand trials and 500 trials labelled as 'others' (the feet and rest). Therefore, by giving half weight to 'other' trials, weighted average decoding accuracies are computed for both methods. As shown in the last column of figure 3, the shallow convNet yielded a decoding accuracy of 48.2%, and the MF-SCSN outperformed the baseline with an accuracy of 51.2%.

V. Discussion

As in the result, MF-SCSN outperformed the baseline method. This indicates that the MF-SCSN has the potential to utilise cross-dataset federated features from different tasks to increase the performance of EEG decoding. Below we also highlight five privacy-preserving properties of the MF-SCSN.

First, subjects and dataset-specific information are stored locally with data owners, which preserves data-level privacy. To preserve parameter-level privacy, feature extractors are stored locally. Another property is that local feature extractors encrypt raw data naturally, so there is no extra encryption cost. Additionally, there is no need for a protocol of decryption, because the cloud network only uses the encrypted features for transfer learning. Finally, labels and classifiers are stored and predicted locally. This also preserves inference-level privacy.

One limitation of this study is that some parts of the source and target data sets were discarded to align the input shape, e.g. the window length and channels. Further experiments should be conducted on transfer learning with different input shapes. A potential solution based on MF-SCSN could be exploring if the local feature extractors with different kernels could handle inputs of different shapes. This is possible once the output shapes of the encryption are unified across branches. Another future direction could be exploring the flexibility of MF-SCSN as a meta-architecture, by changing the feature extractors to other models.

To conclude, we have designed a cross-dataset federated deep transfer learning technique which combines privacy-preserving properties and deep transfer learning. Results show that the proposed method, with the advantage of both transfer learning and privacy-preserving, outperformed the baseline CNN. Our proposed method shows the potential to utilise larger heterogeneous data sets with different tasks for transfer learning while possessing better properties of privacy-preserving across data sets and data centres.

References

[1] X. Wei, P. Ortega, and A. A. Faisal, "Inter-subject deep transfer learning for motor imagery eeg decoding," in *2021 10th Int. IEEE/EMBS Conf. Neural Eng. (NER)*. IEEE, 2021, pp. 21–24.

[2] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2016.

[3] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.

[4] X. Wei *et al.*, "2021 beetl competition: Advancing transfer learning for subject independence & heterogenous eeg data sets," in *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*, ser. Proceedings of Machine Learning Research, vol. 176. PMLR, 06–14 Dec 2022, pp. 205–219.

[5] I. Walker, M. Deisenroth, and A. Faisal, "Deep convolutional neural networks for brain computer interface using motor imagery," *Imperial College, Tech Report*, p. 68, 2015.

[6] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017.

[7] A. Ferrante, C. Gavriel, and A. Faisal, "Data-efficient hand motor imagery decoding in eeg-bci by using morlet wavelets & common spatial pattern algorithms," in *2015 7thInt. IEEE/EMBS Conf. Neural Eng. (NER)*. IEEE, 2015, pp. 948–951.

[8] P. Ortega, C. Colas, and A. A. Faisal, "Compact convolutional neural networks for multi-class, personalised, closed-loop eeg-bci," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 2018, pp. 136–141.

[9] E. G. Ponferrada, A. Sylaidi, and A. A. Faisal, "Data-efficient motor imagery decoding in real-time for the cybathlon brain-computer interface race," 2018.

[10] L. Li *et al.*, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, p. 106854, 2020.

[11] M. Hao, H. Li, G. Xu, S. Liu, and H. Yang, "Towards efficient and privacy-preserving federated deep learning," in *2019 IEEE international conference on communications (ICC)*. IEEE, 2019, pp. 1–6.

[12] L. Lyu, X. He, Y. W. Law, and M. Palaniswami, "Privacy-preserving collaborative deep learning with application to human activity recognition," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1219–1228.

[13] H. Dong, C. Wu, Z. Wei, and Y. Guo, "Dropping activation outputs with localized first-layer deep network for enhancing user privacy and data security," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 662–670, 2017.

[14] W. Du and M. J. Atallah, "Secure multi-party computation problems and their applications: a review and open problems," in *Proceedings of the 2001 workshop on New security paradigms*, 2001, pp. 13–22.

[15] A. B. Popescu *et al.*, "Privacy preserving classification of eeg data using machine learning and homomorphic encryption," *Applied Sciences*, vol. 11, no. 16, p. 7360, 2021.

[16] K. Xia, W. Duch, Y. Sun, K. Xu, W. Fang, H. Luo, Y. Zhang, D. Sang, X. Xu, F.-Y. Wang *et al.*, "Privacy-preserving brain–computer interfaces: A systematic review," *IEEE Trans. Comput. Soc. Syst.*, 2022.

[17] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu, and C. Guan, "Federated transfer learning for eeg signal classification," in *2020 42nd Conf Proc IEEE Eng Med Biol Soc (EMBC)*. IEEE, 2020, pp. 3040–3045.

[18] D. Bethge, P. Hallgarten, T. Grosse-Puppendahl, M. Kari, R. Mikut, A. Schmidt, and O. Özdenizci, "Domain-invariant representation learning from eeg with private encoders," in *ICASSP 2022*. IEEE, 2022, pp. 1236–1240.

[19] M. Kapitonova, P. Kellmeyer, S. Vogt, and T. Ball, "A framework for preserving privacy and cybersecurity in brain-computer interfacing applications," *arXiv preprint arXiv:2209.09653*, 2022.

[20] H. Cho *et al.*, "EEG datasets for motor imagery brain–computer interface," *GigaScience*, vol. 6, no. 7, p. gix034, 2017.

[21] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz *et al.*, "Review of the bci competition iv," *Front. Neurosci.*, vol. 6, p. 55, 2012.

[22] G. Schalk *et al.*, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE. Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, 2004.

[23] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[24] W. Yi, S. Qiu, K. Wang, H. Qi, L. Zhang, P. Zhou, F. He, and D. Ming, "Evaluation of eeg oscillatory patterns and cognitive process during simple and compound limb motor imagery," *PloS one*, vol. 9, no. 12, p. e114853, 2014.