BCINet: An Optimized Convolutional Neural Network for EEG-Based Brain-Computer Interface Applications

Avinash Kumar Singh
Australian Artificial Intelligence Institute
University of Technology Sydney
Ultimo, Australia
avinash.singh@uts.edu.au

Xian Tao
Research Center of Precision Sensing and Control
Institute of Automation
Chinese Academy of Sciences
Beijing, China
taoxian2013@ia.ac.cn

***Abstract*—EEG based brain-computer interface (BCI) allows people to communicate and control external devices using brain signals. The application of BCI ranges from assisting in disabilities to interaction in a virtual reality environment by detecting user intent from EEG signals. The major problem lies in correctly classifying the EEG signals to issue a command with minimal requirement of pre-processing and resources. To overcome these problems, we have proposed, BCINet, a novel optimized convolution neural network model. We have evaluated the BCINet over two EEG based BCI datasets collected in mobile brain/body imaging (MoBI) settings. BCINet significantly outperforms the classification for two datasets with up to 20% increase in accuracy while fewer than 75% trainable parameters. Such a model with improved performance while less requirement of computation resources opens the possibilities for the development of several real-world BCI applications with high performance.**

***Index Terms*—Convolutional neural network, deep learning, EEG, brain-computer interface, MOBI, cognitive conflict, BCINet.**

I. Introduction

Brain-Computer Interface (BCI) allows people to communicate and control external devices using brain signals. Originally BCI was envisioned to help individuals with disabilities such as amyotrophic lateral sclerosis, cerebral palsy, stroke, or spinal cord injury. BCI has proven to be significantly useful in rehabilitation and social cognition after strokes such as Attention deficit hyperactivity disorder (ADHD) and Autism [1]. In recent years, more application has been proposed not limited to the user with a disability but also health users, e.g., BCI based drone flying, in gaming like car racing, three color matching, interaction in virtual reality, etc. A BCI recognizes the user's intent through the electrophysiological signals, which are usually recorded over the scalp, underneath the scalp, or within the brain. Other types of physiological signals can be recorded by magnetic sensors, infrared sensors, or by other means. One such method of recording physiological signals is the electroencephalogram (EEG). EEG records the change in the postsynaptic potential of cortical neurons across the scalp using electrodes [2]. EEG is widely famous in BCI research, and its applications, because of its noninvasive property with high temporal resolution. EEG has been a preferred choice because of its portability and affordability, but on the other hand, it is highly susceptible to electrical noise and has low spatial resolution [3]. The use of EEG signals to develop BCI poses several challenges which usually overcome with the help of machine learning.

The past few years have seen an increased number of deep learning applications in understanding and classifying brain signals [4]. Deep learning is already shown a high number of successful applications in the field of natural language processing [5] and computer vision [6]. Deep learning has a property to learn valuable information from raw data without manual labour [7], which is very useful in EEG signal processing for BCI. Convolutional neural network (CNN) is one of the prevalent methods in the field of deep learning. It has proven to be effective model in several applications of BCI, such as epilepsy/seizures prediction [8], [9], for detection of visual-evoked responses [10], motor imagery classification [11], and speller [12]. Although deep learning can learn from raw EEG data, it requires pre-processing signals to reach the optimal performance level. The processing method is highly dependent on individual data sets and domain knowledge, which could vary task by task. In addition, it is also possible that the pre-processed data does not contain all the information and potentially could be excluded the relevant EEG features depend on the choice of pre-processing such as different filtering [13], channel referencing [14], etc. The use of methods that can learn from raw EEG data without the hand-crafted method for pre-processing is highly desirable, particularly for BCI applications.

However, deep learning methods have shown promising results to automate the process to learn from raw EEG data [10]–[12] with high performance but come with a considerable resources cost such as computation time and memory usage. Such a cost poses a hurdle to develop real-world BCI applications with limited resources. It is also crucial to

978-1-7281-2547-3/20/$31.00 ©2020 IEEE

develop a method that can show high performance by utilizing deep learning to get richer representation while minimizing resources requirements. There could be two significant ways to reduce the resource cost with deep learning methods. By using the fixed weight matrices with deep learning methods like CNN using approximations such as SVD [15]. Another possible way, which this paper is focused on, by reducing the complexity of deep learning network itself while preserving the high performance. There have been several works proposed particularly in the past [16]–[18] to reduce the complexity of model but no work found focused on the model developed for EEG-based BCI applications.

In the line of this work, we have introduced BCINet for BCI applications with the property of significantly reduced complexity of model with better performance. The proposed model has been evaluated with two noisy EEG datasets collected in mobile brain/body imaging (MoBI) [19] settings of different sizes. Our results from the BCINet show a significant reduction in trainable parameters such that an optimized model and significantly improved performance to classify EEG signals compared to some of the popular deep learning model [20]–[22]. The significant contributions of presented work are as follows:

1) A novel optimized deep learning model, BCINet is proposed for EEG based BCI applications.
2) The significantly less trainable parameters compared to existing models for EEG based BCI applications.
3) The significantly improved performance compare to existing models for EEG based BCI applications.

II. MATERIALS AND METHOD

A. Data description

The proposed method has been tested on two datasets collected in MoBI settings in cognitive conflict task. The description of datasets is below:

1) CC-dataset: Cognitive Conflict in the 3D object selection task: The 62-channel EEG dataset (sampling rate 1000 Hz) is collected with 16 participants while performing 3D object selection tasks with their dominant hand tracked by the Leap Motion controller in virtual reality (VR). Fig. 1 (top) displays the scenario where the user is performing the task wearing an EEG cap together with a VR headset, while Fig. 1 (bottom) displays the whole task in a trial. The trial starts with a cube appearing on the table, and the participants were instructed to reach out and select (touch) the cube in VR. The cube would turn red when it was touched by the participant and identified as a normal condition in EEG signals. In other cases, the cube would turn red prematurely and identified as a cognitive conflict condition in EEG signals. For more detail about the experiment and data, please see [23], [24].

For the proposed CNN model, the trial has been extracted from EEG signals 200 ms prior to the onset of cube touch to 1000ms after it, and about 500 trials have been extracted from each participant data. Overall, the dimension of the total data was 62 x 1200 x 6841, where 5,075 trials belonged to non-conflict class and 1,766 for conflict class.

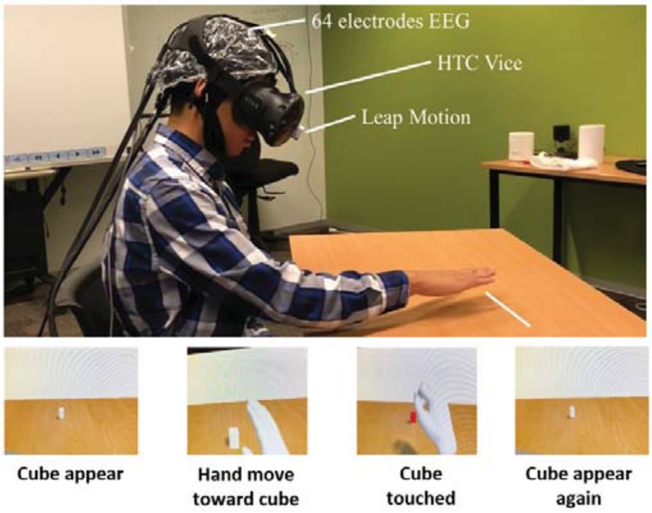

Fig. 1. An experiment scenario in CC-dataset where the participant is trying to touch a cube in VR environment. (see more detail in [23], [24])

2) pHRC-dataset: Cognitive Conflict in physical human-robot collaboration (pHRC): The 32-channel EEG dataset (sampling rate 1000Hz) is collected with 10 participants while performing a swing game task using pHRC. Fig. 2 (top) displays the participant is performing the task holding a robotic arm in the hand and Fig. 2 (bottom) displays the structure of the task. In the task, the participant needs to move the robotic arm from the center to left or right randomly as instructed on the project screen represented by a green target. Once the user reaches the target, the target would turn red as a sign to finish. The blue circle represents the point the nozzle of the robot is aiming at. In order to keep the user more engaged additional input from the user was asked to perform like holding the back handle known as the robot end-effector. All the trials where participants successfully move and reach the target identified as normal in EEG signals while in other cases, the task was suddenly stopped by a virtual object and identified as a conflict in EEG signals. For more detail about the whole experiment, please check [25].

For the proposed CNN model, the 250 trials have been extracted from 200ms prior to the onset of cube touch to 1000ms after it. The overall dimension of the data for a participant was 32 x 1200 x 5600, where 3,354 trials belonged to non-conflict class and 2,246 for conflict class.

CC and pHRC-dataset have been divided into 60%, 20%, 20% for training, validation, and testing respectively for binary conditions using stratified sampling method [26] to avoid the imbalanced class problem. See Table I for data description summary.

TABLE I
DESCRIPTION OF CC AND PHRC-DATASET USED

Datasets	Channels	Sampling Rate	Classes	Dimension
CC	64	1000	2	62 x 1200 x 6841
pHRC	32	1000	2	32 x 1200 x 5600

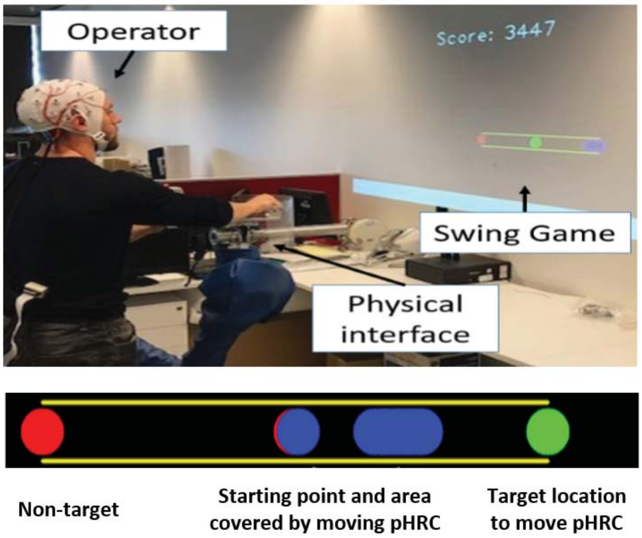

Fig. 2. An experiment scenario in pHRC-dataset where participant is using a robot to perform a task. (see more detail in [25])

B. Proposed model - BCINet

As shown in Fig. 3, the BCINet contains two parallel layers of the 2D convolution layer for the spatial conversion of data with a filter of size 64 and 32. The output of both 2D convolution layers has been concatenated. The output of 2D convolution layers also uses to create a squeeze and excitation layer with the help of global average pooling, followed by a full-connected layer. The concatenated layer and excited layer have been multiplied to get a richer representation of features. The resultant input is fed into one sequential and two parallel 2D depthwise separable convolution layers with different filter = 32, 32, and 64 with kernel = 3, 64, 32. The output of the 2D depthwise separable convolution has been flattened to receive reshaped tensor for classification. The resultant output use 'softmax' for classification. The activation layer utilizes the exponential linear unit (elu) with a linear dropout layer at each layer of 0.5 for classification.

The BCINet is fitted using Adam optimizer, using default parameters defined in [27]. We ran 300 training epochs and performed validation on each epoch, saving the model weights, which lowest validation set loss. All models were trained on NVIDIA Quadro P5000 GPU, with CUDA 9 and cuDNN v7, in TensorFlow [28], using the Keras API [29].

As mentioned before, BCINet uses two parallel convolution and two depthwise separable convolution layers. Our main reason behind these parallel layers in BCINet are as follows:

1) Parallel convolution layers: We have used two parallel convolution layers which convolves only on the feature maps obtained from kernel filters in its filter group, therefore reducing the number of computations to get output feature maps. The parallel convolution layers also increase overall feature space from the input data, which able to provide improved feature maps.

2) Parallel depthwise separable convolution layers: Following the similar intuition of two parallel convolution layers, we have further used two parallel depthwise separable convolution layers. Each layer was first performing a depthwise spatial convolution on each input channel, followed by pointwise convolution. The parallel layers again reduced the number of computations significantly while having increase features spaces.

C. Baseline model description

EEGNet. The EEGNet [20] is CNN based model that contains 2D convolution layer, 2D depthwise convolution layer, and 2D separable convolution layer followed by 'softmax' for classification with 'adam' as an optimizer. As per the model's recommendation, we have used a batch size of 16, kernel length of 500, and a 0.5 dropout rate. Other parameters used in EEGNet included F1 = number of EEG channels, D = 2, and F2 = F1*D. With F1 being the number of temporal filters, D the number of spatial and F2 the number of pointwise filters.

DeepConvNet. It consists of four convolutional blocks and a classification block. The first convolutional block is to handle EEG inputs, followed by three standard convolution layers. The classification is performed using a softmax with Adam optimizer. We have used batch size = 16, dropout rate = 0.50 with 300 epochs. The number of filters used in four convolution layers was 25, 50, 100, and 200, respectively, with each layer consist of five kernels [21].

ShallowNet. It is a shallow version of DeepConvNet, inspired by filter bank common spatial patterns. It consists of two convolution layers, followed by a fully connected layer. The classification is performed using a softmax with Adam optimizer. We have used batch size = 16, dropout rate = 0.50 with 300 epochs. The number of filters used in two convolution layers was 40, respectively, with each layer consist of kernel size =13 [22].

D. Evaluation metrics

The parameters for all the classifiers compared in this paper have been set up before training and testing for all participants. In this work, the classes are imbalanced; therefore, stratified random sampling [26] has been used on the data with the aforesaid machine learning algorithms. To compare the results of different classifiers, we have evaluated overall accuracy, precision, recall, and F1-score for a targeted class.

III. RESULTS AND DISCUSSION

A. Comparison between parameters

We have compared the parameters required by the BCINet model with EEGNet, ShallowNet, and DeepConvNet. As shown in Table II. It can be seen that BCINet needed about 72% less trainable parameters compare to EEGNet, 75% less trainable compare to DeepConvNet, and 55% less trainable parameters compare to ShallowNet for CC-dataset. Similarly, BCINet required about 70% less trainable parameters compared to DeepConvNet, 17% less trainable than ShallowNet,

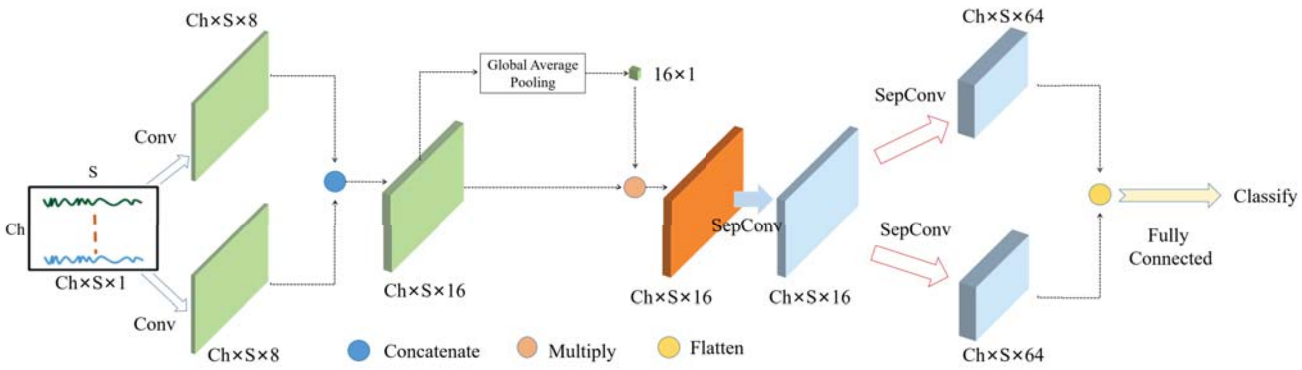

Fig. 3. Architecture of BCINet

TABLE II
COMPARISON OF PARAMETERS REQUIRED BY EEGNET, SHALLOWNET, DEEPCONVNET, AND BCINET FOR CC AND PHRC-DATASET

	EEGNet		ShallowNet		DeepConvNet		BCINet	
Params	**CC**	**pHRC**	**CC**	**pHRC**	**CC**	**pHRC**	**CC**	**pHRC**
Trainable	180,794	58,754	112,642	64,642	199,527	180,777	50,674	53,704
Non-Trainable	63,736	32,896	600	600	875	875	8,088	8,238
Total	244,530	91,650	113,242	65,242	200,402	181,652	58,762	61,942

and 8% less trainable parameters compare to EEGNet for pHRC-dataset.

In summary, BCINet required up to 75% fewer parameters for CC-dataset and about 70% fewer parameters for pHRC-dataset than baseline models. Less requirement of parameter also implies less necessity of computational resources such as computation time and memory usages. It is also known that fewer parameter requirements also suggest less need for data to learn new features in deep learning.

On the other hand, the BCINet has 69% less prediction time compare to EEGNet while no improvement compared to ShallowNet and DeepConvNet for CC-dataset. Similarly, BCINet has 50% less prediction time compared to EEGNet, while again, there is no improvement compared to ShallowNet and DeepConvNet for pHRC-dataset. The major reason behind no improvement in prediction time in BCINet compare to ShallowNet and DeepConvNet is the use of a depthwise separable convolution layer that generally requires significantly higher computation time compared to normal convolution layer [30]. The better prediction time could speed up the whole system's performance, which is very improvement for BCI applications with higher information transfer rate (ITR) [31].(Also see Fig. 4).

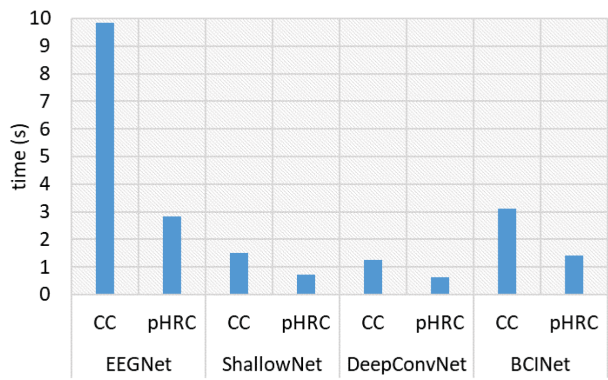

Fig. 4. Time required for prediction in EEGNet, ShallowNet, DeepConvNet, and BCINet on CC and pHRC-dataset

B. Classification performance

We have also evaluated the classification efficacy of the BCINet with EEGNet, ShallowNet, and DeepConvNet. It can be seen from Fig. 5 and Table III that, BCINet significantly outperformed compare to baseline models. BCINet shows an accuracy improvement by 10% for EEGNet, 5% for DeepConvnet, and a slight increase in ShallowConvNet for CC-dataset. Similarly, BCINet shows an accuracy improvement by 20% for EEGNet, 31% for ShallowNet, and 9% for DeepConvnet pHRC-dataset. This is a significant improvement in classification compare to the baseline model while using significantly fewer parameters.

As mentioned earlier, CC and pHRC-datasets has an imbalanced class problem. Therefore, we also look at other performance indicators such as precision, recall, and F1-score for BCINet and all baseline models. Again, as shown in Fig. 5, BCINet outperformed compare to all baseline models in these indicators. However, interestingly, BCINet only slightly improved compared to ShallowNet for CC-dataset similar to accuracy. One potential reason could be fewer convolution layers, which generally better if classes are imbalanced.

In addition to performance indicators, we also looked at the performance of BCINet while training epoch-wise. As shown in Fig. 6, it is also clear that during the training validation-loss of BCINet converge more compare to EEGNet, ShallowNet,

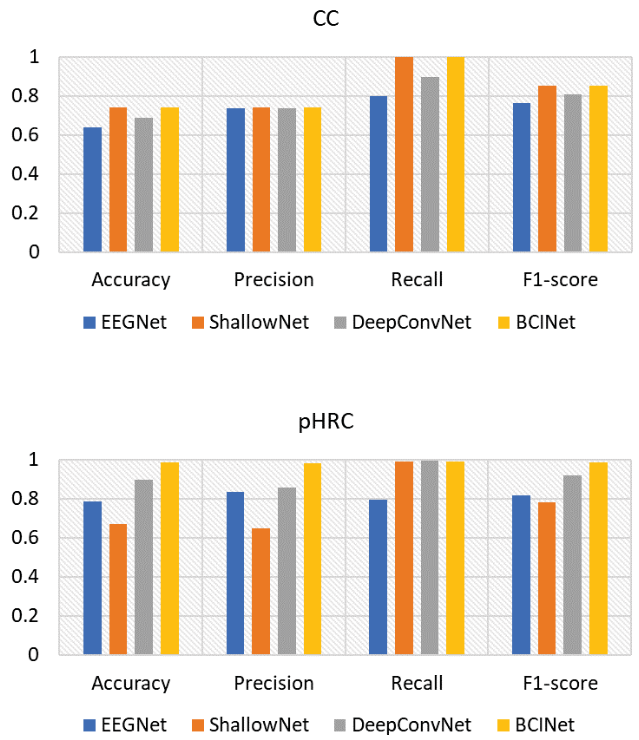

Fig. 5. Accuracy, precision, recall, and F1-score for EEGNet, ShallowNet, DeepConvNet, and BCINet on CC and pHRC-dataset

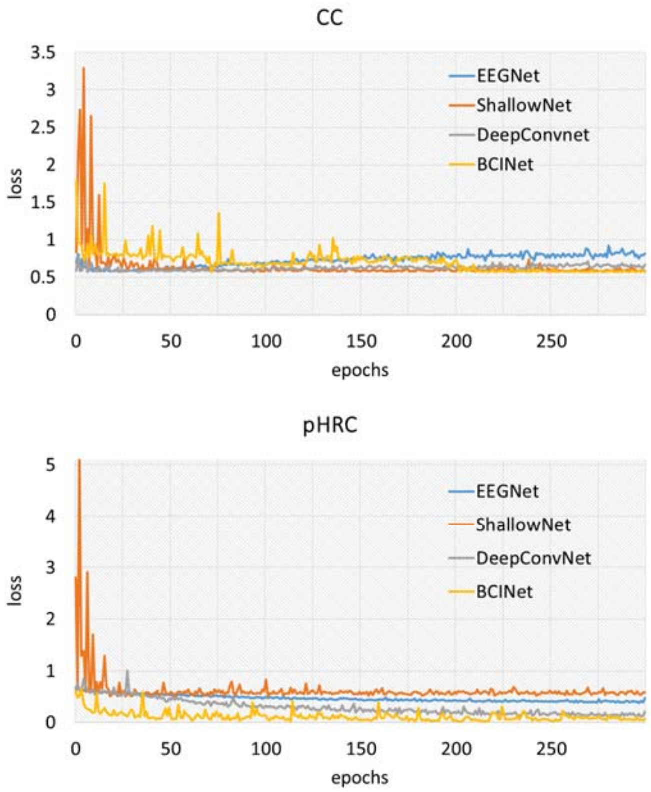

Fig. 6. Validation loss from EEGNet, ShallowNet, DeepConvNet, and BCINet for CC and pHRC dataset over 300 epochs

and DeepConvNet for CC and pHRC dataset. These results are again evident that BCINet is successfully able to learn features from CC and pHRC datasets.

It is also noted that EEG signals are known to have a poor signal to noise ratio, and the dataset used in evaluation has been collected in MoBI settings. The MoBI settings create additional challenges in terms of artifacts and even further degradation in the signal-to-noise ratio. Although, the BCINet can perform significantly better compared to the model like EEGNet, DeepConvNet, and ShallowNet in terms of performance and resource requirements. Such a model poses a high potential for BCI applications. These BCI applications do not necessarily need to be limited to a laboratory environment but could be from a real-world environment (MoBI settings).

TABLE III
ACCURACY FOR CC AND PHRC DATASET WITH RESPECT TO EEGNET, SHALLOWNET, DEEPCONVNET, AND BCINET

Datasets	**EEGNet**	**ShallowNet**	**DeepConvNet**	**BCINet**
CC	0.6377	0.7407	0.6881	0.7421
pHRC	0.7857	0.6705	0.8973	0.9857

IV. Limitations and Future Work

The BCINet shows several promising properties with high performance compare to the established models. However, to generalize and evaluate the efficacy with all type of BCI applications, several steps are required to take in the future as follows:

1) The current datasets were collected in two distinct MoBI settings but it lacks varieties of other BCI applications; in the future, we will evaluate the efficacy of the model with multiple kinds of BCI datasets such as motor imagery (MI), SSVEP, cognitive workload, etc. with extensive measures [1].
2) In this paper, we have only shown the performance of the BCINet as a whole in terms of performance form the machine learning point of view. In the future work, the output of different layers will be evaluated and compared with traditional EEG features such as topography, event-related spectral perturbation (ERSP) [32], etc. as well as evaluation will be performed for each layer such as the impact of hyper-parameters.
3) The current proposed model shows a significantly improved performance with EEG signals. In the future, we also wanted to evaluate this model with similar time-series data such as electrocardiography (ECG), electromyography (EMG), etc. to evaluate efficacy and robustness.

V. Conclusion

In this paper, we have proposed a novel BCINet model that outperforms the established model like EEGNet, ShallowNet, and DeepConvNet to classify data collected in MoBI setting. We have also shown that BCINet can outperform baseline models with up to 75% reduction in trainable parameters and up to 20% improvement in accuracy. Such a model has vast potential in the development of real-world (MoBI) BCI applications with higher performance but with comparatively limited resource requirements.

ACKNOWLEDGMENTS

This work was supported in part by the Australian Research Council (ARC) under discovery grant DP180100656, and the National Natural Science Foundation of China under Grant 61703399.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[2] M. Tudor, L. Tudor, and K. I. Tudor, "Hans berger (1873-1941)– the history of electroencephalography," *Acta medica Croatica: casopis Hravatske akademije medicinskih znanosti*, vol. 59, no. 4, pp. 307–313, 2005.

[3] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, "Spatial and temporal resolutions of eeg: Is it really black and white? a scalp current density view," *International Journal of Psychophysiology*, vol. 97, no. 3, pp. 210–220, 2015.

[4] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: a review," *Journal of neural engineering*, vol. 16, no. 3, p. 031001, 2019.

[5] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Transactions on audio, speech, and language processing*, vol. 22, no. 10, pp. 1533–1545, 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[7] R. Chai, S. H. Ling, P. P. San, G. R. Naik, T. N. Nguyen, Y. Tran, A. Craig, and H. T. Nguyen, "Improving eeg-based driver fatigue classification using sparse-deep belief networks," *Frontiers in neuroscience*, vol. 11, p. 103, 2017.

[8] D. Lu and J. Triesch, "Residual deep convolutional neural network for eeg signal classification in epilepsy," *arXiv preprint arXiv:1903.08100*, 2019.

[9] A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, and H. Takahashi, "Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images," *NeuroImage: Clinical*, vol. 22, p. 101684, 2019.

[10] H. Cecotti and A. Graser, "Convolutional neural networks for p300 detection with application to brain-computer interfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 433–445, 2010.

[11] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating deep neural networks for online decoding of motor imagery movements from eeg signals," *Sensors*, vol. 19, no. 1, p. 210, 2019.

[12] H. Shan, Y. Liu, and T. P. Stefanov, "A simple convolutional neural network for accurate p300 detection and character spelling in brain computer interface." in *IJCAI*, 2018, pp. 1604–1610.

[13] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[14] S. Hu, Y. Lai, P. A. Valdes-Sosa, M. L. Bringas-Vega, and D. Yao, "How do reference montage and electrodes setup affect the measured scalp eeg potentials?" *Journal of neural engineering*, vol. 15, no. 2, p. 026013, 2018.

[15] J. Xue, J. Li, and Y. Gong, "Restructuring of deep neural network acoustic models with singular value decomposition." in *Interspeech*, 2013, pp. 2365–2369.

[16] C. W. Wu, "Prodsumnet: reducing model parameters in deep neural networks via product-of-sums matrix decompositions," *arXiv preprint arXiv:1809.02209*, 2018.

[17] H. Pratt, B. Williams, F. Coenen, and Y. Zheng, "Fcnn: Fourier convolutional neural networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2017, pp. 786–798.

[18] C. Ding, S. Liao, Y. Wang, Z. Li, N. Liu, Y. Zhuo, C. Wang, X. Qian, Y. Bai, G. Yuan *et al.*, "Circnn: accelerating and compressing deep neural networks using block-circulant weight matrices," in *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture*, 2017, pp. 395–408.

[19] E. Jungnickel and K. Gramann, "Mobile brain/body imaging (mobi) of physical interaction with dynamically moving objects," *Frontiers in human neuroscience*, vol. 10, p. 306, 2016.

[20] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[21] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[22] 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," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[23] A. K. Singh, "Cognitive conflict in virtual reality based object selection task," Ph.D. dissertation, 2019.

[24] A. K. Singh, K. Gramann, H.-T. Chen, and C.-T. Lin, "The impact of hand movement velocity on cognitive conflict processing in a 3d object selection task," *BioRxiv*, 2020.

[25] S. Aldini, A. Akella, A. K. Singh, Y.-K. Wang, M. Carmichael, D. Liu, and C.-T. Lin, "Effect of mechanical resistance on cognitive conflict in physical human-robot collaboration," in *2019 international conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 6137–6143.

[26] J. E. Trost, "Statistically nonrepresentative stratified sampling: A sampling technique for qualitative studies," *Qualitative sociology*, vol. 9, no. 1, pp. 54–57, 1986.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.

[29] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[30] Z. Qin, Z. Zhang, D. Li, Y. Zhang, and Y. Peng, "Diagonalwise refactorization: An efficient training method for depthwise convolutions," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[31] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "Brain–computer interface (bci) operation: optimizing information transfer rates," *Biological psychology*, vol. 63, no. 3, pp. 237–251, 2003.

[32] S. Makeig and T.-P. Jung, "Tonic, phasic, and transient eeg correlates of auditory awareness in drowsiness," *cognitive brain research*, vol. 4, no. 1, pp. 15–25, 1996.