HAL open science

Mental arithmetic task classification with convolutional neural network based on spectral-temporal features from EEG

Zaineb Ajra, Binbin Xu, Gérard Dray, Jacky Montmain, S. Perrey

To cite this version:

Zaineb Ajra, Binbin Xu, Gérard Dray, Jacky Montmain, S. Perrey. Mental arithmetic task classification with convolutional neural network based on spectral-temporal features from EEG. EMBS 2022 - 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Jul 2022, Glasgow, United Kingdom. pp.52-55, 10.1109/EMBC48229.2022.9870887 . hal-03711249

HAL Id: hal-03711249

https://imt-mines-ales.hal.science/hal-03711249v1

Submitted on 13 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9870887

Mental arithmetic task classification with convolutional neural network based on spectral-temporal features from EEG

Zaineb Ajra, Binbin Xu, Gérard Dray, Jacky Montmain, Stephane Perrey

***Abstract*— In recent years, neuroscientists have been interested to the development of brain-computer interface (BCI) devices. Patients with motor disorders may benefit from BCIs as a means of communication and for the restoration of motor functions. Electroencephalography (EEG) is one of most used for evaluating the neuronal activity. In many computer vision applications, deep neural networks (DNN) show significant advantages. Towards to ultimate usage of DNN, we present here a shallow neural network that uses mainly two convolutional neural network (CNN) layers, with relatively few parameters and fast to learn spectral-temporal features from EEG. We compared this models to three other neural network models with different depths applied to a mental arithmetic task using eye-closed state adapted for patients suffering from motor disorders and a decline in visual functions. Experimental results showed that the shallow CNN model outperformed all the other models and achieved the highest classification accuracy of 90.68%. It's also more robust to deal with cross-subject classification issues: only 3% standard deviation of accuracy instead of 15.6% from conventional method.**

I. INTRODUCTION

Nowadays, one of the medical challenges is to understand the diseased brain in the context of neurological diseases such as Alzheimer's, Parkinson's, consciousness disorders, epilepsy, etc., in order to develop suitable treatments. The brain and the spinal cord constitute the central nervous system, capable of integrating information, controlling motor skills / movements and ensuring cognitive functions [1]. Several technologies based on head movements, eye contact, etc have given people the chance to control a wheelchair or robot. However, for patients with brain injury who have lost muscle control voluntarily, it is not easy or even not possible to use these technologies.

The brain computer interface (BCI) technology offers those who are deprived of muscle control a way to communicate with their environment and become less dependent in their daily activities [2][3]. One of the most common BCI is through electroencephalography (EEG), a non-invasive neuroimaging technique that measures brain activity, which is more accessible and more accurate than functional magnetic resonance imaging (fMRI) [4] in time resolution. EEG signals reflect valuable information about the brain activity in measured scalp regions. This is crucial for the diagnosis and management of certain diseases or brain injuries.

In consequence, the first step is to better classify the acquired EEG signals, for example separating EEG signals containing cognitive activities from the resting state. However, classifying cognitive tasks using EEG signals necessitates solving difficult pattern recognition issues. A prior feature engineering step is required with conventional classification methods. Due to the highly complex property of EEG signal, these methods cannot always guarantee to produce the acceptable performance of classification. The neural network (NN) in computer science has reached human parity milestone in many applications. The most exciting feature of neural network is its self-learning ability which implies that manual feature extraction is not necessarily required. This approach is now more and more used in EEG classification to overcome the limits of conventional methods, especially when data volume is increasing in an exponential way [5][6].

One main issue when using NN is the choice of feature selection and architecture, as in other fields. In this work, we proposed to use joint spectral-temporal features instead of the ones extracted with common spatial pattern filtering and to benchmark different NN architectures for a mental task classification using EEG signals, from basic shallow NN to state-of-art deep NN models.

II. DATA AND MODELS

A. Data from cognitive tasks

1) Participants and data: In this study, we used the publicly available dataset collected by Shin et al. in 2018 [7]. In the original dataset, 22 EEG channels and 9 near-infrared spectroscopy (NIRS) channels were simultaneously recorded from the scalp in 12 participants performing mental arithmetic (MA) task and being relaxed (BL). Here, we only consider EEG recordings. Regarding the EEG signals acquired, 10 were placed on the frontal cortex and 12 other on parieto-occipital areas.

2) Experimental protocol: Experiments were conducted in 3 sessions. In each session, a pre-rest period of $15\,\mathrm{s}$ was followed by 20 trials of task phase (27–29 s each) and then a final $15\,\mathrm{s}$ post-rest period. A fixation cross was shown on the monitor through the pre-rest and post-rest periods. During each trial, participants were instructed to look at a visual instruction ($2\,\mathrm{s}$) indicating the type of task, followed by task period of $10\,\mathrm{s}$ and a rest period (15–17 s). In the instruction period, the type of task was randomly displayed on the screen (MA or BL). For MA task, an arbitrary one-digit number between 6 and 9 was subtracted from three digit number. For BL, participants were asked to relax and gaze at the fixation cross displayed in the middle of the monitor. A total of 60 trials (20 trials $\times$ 3 sessions) were performed by each subject.

[1]EuroMov Digital Health in Motion, Univ. Montpellier, IMT Mines Ales, Ales, France. zaineb.ajra@umontpellier.fr binbin.xu@mines-ales.fr

This work is licensed under a Creative Commons Attribution 3.0 License.
For more information, see http://creativecommons.org/licenses/by/3.0/

3) Data preprocessing: The raw EEG signals from 22 channels were down-sampled to 200 Hz and then band-pass filtered within the range of 0.5–50 Hz using a 3rd-order Butterworth filter. Ocular artifacts were rejected manually. The artifact-free data was segmented into epochs using time intervals from -2 s (2 s before the beginning of the task) to 10 s (the end of the task). The baseline in the range of -1 s to 0 s were removed from data. In total, 60 epochs of 12 s duration were obtained for each participant.

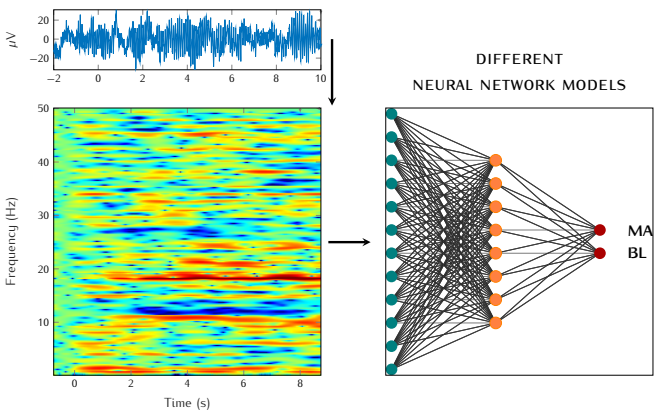

Fig. 1. The study work-flow: from signal-spectrum to different models

B. Spectral-temporal feature extraction

In the work [7], Shin et al. used features extracted from EEG signals with common spatial pattern filter (CSP). The time-frequency analysis allows to obtain richer information which are more appropriate to neural network. Thus, in this work, spectral-temporal features were used. These features are extracted with event-related spectral power (ERSP) approach from EEG data with EEGLab. ERSP is applied at epoch level. As most known DNN have fixed input dimensions such as (224×224), to take the advantage of these powerful models, the ERSP output (time $\times$ frequency domain) is set with the same dimensions. In total 15 480 spectrograms (12 subjects $\times$ 22 electrodes $\times$ 60 epochs) were extracted. The main workflow can be found in Fig. 1.

C. Models

Shin et al. used shrinkage linear discriminant analysis (sLDA) to classify the features obtained from CSP on EEG. While classifying MA task from BL, the performance by sLDA reached $77.3 \pm 15.9\%$ (mean $\pm$ std) in terms of accuracy. In our work, we opted for neural network models. Two types of NN models were evaluated: 1) proposed shallow neural network: basic recurrent neural network (RNN) and convolutional neural network; 2) common state-of-art deep neural networks: GoogLeNet and ResNet-50.

1) Long Short-Term Memory (LSTM) model: RNN is capable to learn temporal dependence in the sequential data. It's the common architecture to deal with time series data, and often serves as the baseline model in bench-marking. Here, we proposed to use a first model based on LSTM which is one of the most used RNN variants. This model is consisted of two LSTM layers having respectively 256 and 128 hidden neurons. Each of them is followed by a Dropout layer (drop rate 0.5). The output classification layer used SOFTMAX as activation. The spectral-temporal features are thus considered as for example 224 time series in the frequency domain. The model architecture can be found in Fig. 2.

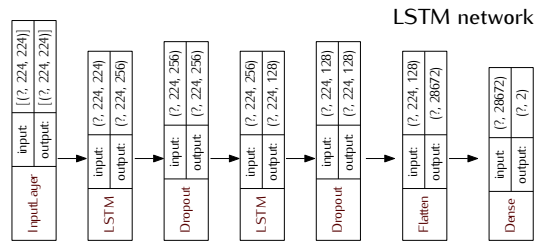

Fig. 2. Baseline LSTM model architecture

2) Convolutional Neural Networks: CNN is currently the most used neural network architecture. The main advantage of CNN is its self-learning ability of feature extraction – using a series of convolution filters, CNN layer can generate invariant features from two-dimensional / three-dimensional data. There's no necessarily further step of manual feature extraction. Another advantage of CNN models is that it also takes the spatial (or cross-dimensional in case of data other than images) constraints which may yield better features. The two-dimensional spectral-temporal features ERSP features of dimension 224×224 from EEG signals can be considered as one type of image. Application of CNN layers on its classification can then be considered.

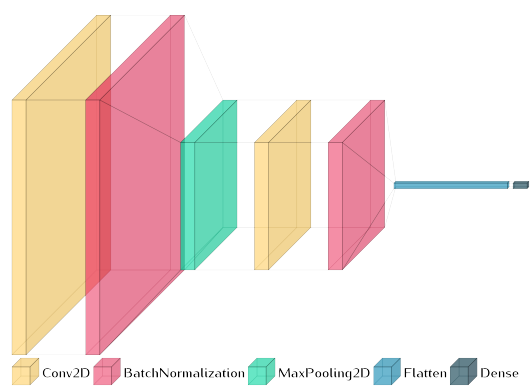

Fig. 3. CNN model architecture

The proposed CNN model (Fig. 3) followed the same principle – shallow neural network. This model has essentially two 2D convolutional layers (Conv2D) which are each attached to a batch normalization layer to address the issue of internal covariate shift. The Batch normalization layer can also be considered as regularizer, in many cases it may help to eliminate the need of a Dropout layer. Both Conv2D layers using a Rectified Linear Unit (ReLU) activation have 10 filters with kernel of size 3×3 and zero-padding is applied to match the same input dimension. These two blocks are

linked by a max pooling layer with 2×2 size and 2×2 stride.

3) Deep neural network: Considering the data size limit, two proposed models are both basic shallow network. The ultimate goal is to acquire enough data to re-train the state-of-the-art DNN models. At the first stage, it's worth to investigate how these DNN models perform in this application. Two models with the same input image size are considered: GoogLeNet and ResNet-50. As shown in Table I, they have considerably larger sizes in comparison to the proposed shallow networks which have only 0.74 / 0.25 million parameters.

TABLE I
MODELS PROPERTIES

Network	Depth	Parameters (Millions)	Input Size
GoogLeNet	22	7	224-by-224
ResNet-50	50	25.6	224-by-224
LSTM	2	0.74	224-by-224
CNN	2	0.25	224-by-224

D. Training options and performance evaluation

All models were trained in the same configuration. Stochastic gradient descent with momentum (SGDM) optimizer is used for all. The initial learning rate is fixed at 0.001. The maximum number of training epoch is 50, with batch size 64. The same early stopping rule is applied – validation patience of 20 and a validation frequency of per 8 iterations.

Keeping the original subject / task ratios from the raw data, the 15 840 samples are randomly split into TRAINING (70%), VALIDATION (15%) and TEST (15%). In order to minimize the risk of over-fitting, 20 random triple-sets were generated and used for all models. The median values of commonly used metrics (Accuracy, Specificity, Sensitivity and F1-score) on TEST sets are reported from the 20 splits.

III. RESULTS / DISCUSSION

For the four NN models with same training conditions – shallow LSTM / CNN, deep GoogLeNet and ResNet-50, their performance metrics are presented in Table II. The shallow LSTM model showed poor performance, with a median accuracy only at 67.76% which is quite lower than the baseline results of 80.1% from [7]. Deep NN models performed similarly. Both models showed neither satisfactory results with 73.34% / 72.77% accuracy. The shallow CNN model, despite its smallest number of parameters, achieved an accuracy of 90.68% and outperformed all the listed models. This model is also much more robust with a standard deviation accuracy only at 2.3%, while it was 15.9% in [7].

TABLE II
THE PERFORMANCES OF DIFFERENT METHODS (MEDIAN %)

Methods	ACC	Sensibility	Specificity	F1 score
sLDA [7]	80.10	–	–	–
LSTM	67.76	67.74	68.37	68.38
GoogLeNet	73.34	77.56	72.45	72.22
ResNet-50	72.77	74.63	72.00	72.07
CNN	**90.68**	**94.34**	**90.85**	**90.21**
STANDARD DEVIATION OF CNN MODEL				
CNN *std*	**2.30**	**5.53**	**6.29**	**2.28**

TABLE III
PERFORMANCES BY PARTICIPANT WITH CNN MODEL (MEDIAN %)

	sLDA [7]	proposed CNN model			
Participant	ACC EEG	ACC	Sen	Spe	F1
1	96.5	90.3 ± 4.4	92.7	91.3	90.1
2	79.0	91.2 ± 3.0	94.4	93.4	91.3
3	58.2	92.0 ± 2.6	94.6	94.0	91.8
4	90.7	90.9 ± 2.9	93.7	91.2	90.9
5	95.7	90.3 ± 3.4	91.1	91.6	89.9
6	83.0	89.4 ± 3.6	91.6	89.1	89.6
7	50.7	92.1 ± 2.6	95.1	90.8	92.1
8	66.2	90.3 ± 2.8	92.2	90.6	89.9
9	81.2	88.7 ± 3.0	91.7	90.6	88.3
10	76.8	90.1 ± 2.7	95.1	91.3	89.8
11	96.0	92.4 ± 2.4	95.7	92.5	92.2
12	53.7	89.1 ± 2.6	90.5	90.1	88.6
std	15.9	1.20	1.80	1.36	1.30

Another common issue of classification study in neuroscience is that the performance varies strongly from one subject to another. So, we tested the trained models for each participant in test set as well. As shown in Table III, the CNN model demonstrated its robustness. The classification accuracy is quite similar for all the 12 participants. The sLDA approach with CSP extracted features failed to classify the MA task from BL, with accuracy close to 50% – the random selection in case of binary classification. The inter-participant standard deviation of accuracy is 15.6%, while the new CNN model showed significantly small as 1.2%. At participant level, the inter-participant std values are all close to 3%.

Further observation at EEG channel level revealed that the performance varied from one brain region to another (Fig. 4, topological plot with interpolation from EEGLab; detailed values in Tab. IV). The good accuracy in the frontal cortex region (especially left) is consistent with literature on mental arithmetic tasks [8]. In occipital region, similar accuracy was obtained. This suggested that this region might be a promising candidate in the classification of this type of tasks.

The poor performance from LSTM models is generally related to one of its drawbacks - lack of spatial (or cross-dimensional) learning. More specifically in this study, LSTM model took only the temporal information into account, the structural correlation in frequency domain being ignored. This would make it fail to classify correctly this type of data in cognitive task. Applying the same early stopping rules as shallow NN, the deep neural networks showed only moderate performance. However, if they are trained differently (with less strict early stopping rule and considerably more fine-

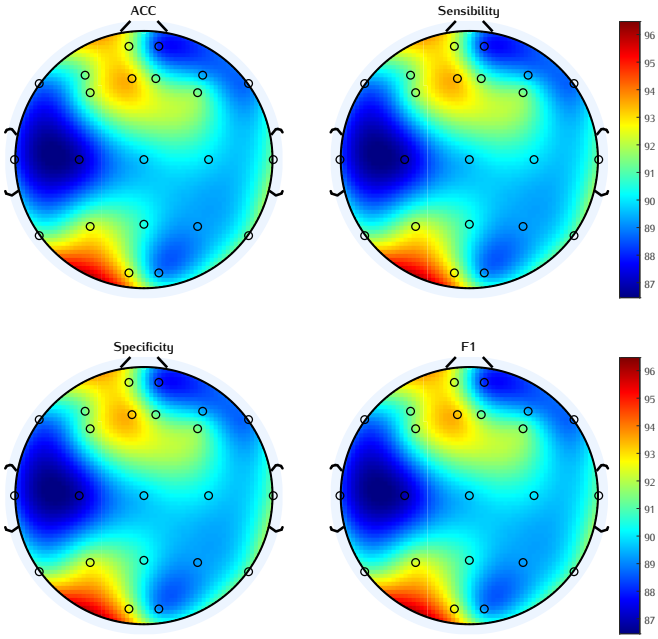

Fig. 4. Topological representation of performances by Channels

TABLE IV
PERFORMANCES BY CHANNELS (MEDIAN %)

Channel	ACC	Channel	ACC	Channel	ACC
F7	89.1	AFF1h	93.3	POO1	92.6
AFF5h	91.3	AFF2h	91.3	POO2	89.1
F3	91.7	Cz	89.7	P4	89.2
AFp1	92.2	Pz	90.5	P8	91.4
AFp2	88.1	T7	88.5	C4	90.0
AFF6h	89.3	C3	86.9	T8	91.4
F4	91.4	P7	91.5		
F8	88.8	P3	92.3		

tuning), these models started to show promising performance (some case better than baseline result with sLDA). They should perform better when the data size is sufficiently large.

FURTHER TRAINING ON LSTM AND RESNET-50 MODELS

Methods	ACC	Sensibility	Specificity	F1 score
LSTM	78.87	82.69	79.01	75.86
ResNet-50	85.90	88.92	83.48	84.86

But one needs to note that the training time is significantly longer than shallow networks. And many deep NN models run only on graphics processing unit (GPU) in order to be trained in an acceptable time range. When the data size is limited, this study highlights that use of shallow CNN may already help to obtain satisfactory accuracy as high as 91%. There exists still slight room to improve the performance of this model by fine-tuning the architecture and training parameters.

IV. CONCLUSION

In this study, we examined the performance of four neural network models in cognitive tasks classification of EEG recordings at trial level. Our findings revealed that the proposed shallow CNN improves significantly classification performance reaching 90.68% cross-validation accuracy, which is 10.6% higher than with sLDA method in their original study [7]. The spectral and temporal features from EEG would be more representative than features extracted by spatial pattern filtering. In comparison with CNN model, the basic LSTM model in time domain only showed poor performance and failed to accomplish the classification task. This suggests that the structure across frequencies can indeed bring additional information.

Even though the DNN did not give satisfactory results when using the same training options, when trained with different configurations, their performances started to increase and can surpass the performance with sLDA. However, this will require much more training time and parameter fine-tuning. They are still very promising. Increasing the data volume will help to improve the performance and most importantly to propose some more generalized and robust models when dealing with unlearned new data. A project of this objective is currently on-going.

The high performance of the shallow CNN model showed that for cognitive task classification, simple shallow network might already be sufficient which requires much less training time and computational load. The personalised model is also conceivable when trial-level data number is enough. This can also help to reduce the inter-subjects issue.

ACKNOWLEDGMENT

Zaineb AJRA received a doctoral fellowship from AXIAUM Univ. Montpellier-ISDM (ANR-20-THIA-0005-01) and ED I2S in France.

REFERENCES

[1] M. Ito, "Controller-regulator model of the central nervous system," *Journal of integrative neuroscience*, vol. 1, no. 02, pp. 129–143, 2002.
[2] T. Ma, W. Chen, X. Li, Y. Xia, X. Zhu, and S. He, "fnirs signal classification based on deep learning in rock-paper-scissors imagery task," *Applied Sciences*, vol. 11, no. 11, p. 4922, 2021.
[3] M. J. Khan and K.-S. Hong, "Hybrid eeg–fnirs-based eight-command decoding for bci: application to quadcopter control," *Frontiers in neurorobotics*, vol. 11, p. 6, 2017.
[4] U. Asgher, K. Khalil, Y. Ayaz, R. Ahmad, and M. J. Khan, "Classification of mental workload (mwl) using support vector machines (svm) and convolutional neural networks (cnn)," in *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE, 2020, pp. 1–6.
[5] H. Ghonchi, M. Fateh, V. Abolghasemi, S. Ferdowsi, and M. Rezvani, "Deep recurrent–convolutional neural network for classification of simultaneous eeg–fnirs signals," *IET Signal Processing*, vol. 14, no. 3, pp. 142–153, 2020.
[6] M. Saadati, J. Nelson, and H. Ayaz, "Mental workload classification from spatial representation of fnirs recordings using convolutional neural networks," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing*. IEEE, 2019, pp. 1–6.
[7] J. Shin, K.-R. Müller, and H.-J. Hwang, "Eyes-closed hybrid brain-computer interface employing frontal brain activation," *PloS one*, vol. 13, no. 5, p. e0196359, 2018.
[8] G. Sammer, C. Blecker, H. Gebhardt, M. Bischoff, R. Stark, K. Morgen, and D. Vaitl, "Relationship between regional hemodynamic activity and simultaneously recorded eeg-theta associated with mental arithmetic-induced workload," *Human brain mapping*, vol. 28, no. 8, pp. 793–803, 2007.