COMPARISON OF CNN-BASED EEG CLASSIFICATION IN SENSOR AND SOURCE SPACE

M. Maurer¹, D. Baumgarten^{1,2}, J. Vorwerk¹

¹ Institute of Electrical and Biomedical Engineering, UMIT TIROL, Hall in Tirol, Austria

2 Institute of Mechatronics, University of Innsbruck, Innsbruck, Austria

E-mail: magdalena.maurer@umit-tirol.at

ABSTRACT: Electroencephalography (EEG) is a popular tool in brain-computer interfacing (BCI), due to its unique time resolution and simplicity of application. For the design of BCIs, rapid and accurate classification algorithms are needed to classify the brain state correctly in real-time. Recent technological advancements facilitate the use of novel methods for signal processing and analysis such as real-time source estimation and classification via deep learning approaches. In this work a previously established convolutional neural network (CNN) architecture, the EEGNet, was applied to a publicly available motor imagery EEG dataset for classification of sensor measurements and source estimates that were computed with three different inverse approaches. Both for sensor signals and source estimates similar classification accuracies as in the literature could be achieved. However, no significant difference in performance between sensor and source space analysis was observed.

INTRODUCTION

In the field of brain-computer interfaces (BCI), the electroencephalogram (EEG) remains one of the most popular measurement tools for acquiring brain signals, due to its many positive characteristics for the BCI use case. BCIs offer the possibility to interact with an external machine such as a computer or a rehabilitation robot based primarily on modulations of brain activity. With EEG, electrical brain activity can be measured noninvasively on the surface of the scalp. EEG is comparatively cheap and easy to use, and it is portable. In addition, due to its high temporal resolution in the millisecond range, the EEG allows for measuring ongoing brain activity basically in real-time. EEG-based BCI systems are therefore used in a great variety of settings such as in rehabilitation, health and attention monitoring, entertainment and skill improvement [1, 2]. To allow for the control of an external device it is necessary to analyze data and derive a control command in real time. Various measurement paradigms for EEGbased BCIs exist to elicit specific EEG patterns from which control signals can be generated. Common paradigms include motor imagery (MI), visually evoked potentials or error-related potentials [2].

A control signal is extracted by detecting suitable features to distinguish the respective signal patterns and then classifying them. To train the classifier in the best possible way, a time-consuming recording of training data and an offline training phase are usually required. Since the goal is to provide BCI users with good, precise, and intuitive control of the external device, achieving high classification accuracies in real-time is essential [1, 3]. One of the greatest difficulties in the real-time derivation of control signals lies in the low signal-tonoise ratio (SNR) of single-trial evaluations. Further challenges are the large variability of signals between subjects, but also within subjects from session to session, as well as the large amount of training data required to train a classifier [3].

One frequently used BCI paradigm, especially in the field of rehabilitation, is based on Motor Imagery (MI), where the user has to imagine a movement without actually performing it. MI-based BCIs have been used for example to control a robotic rehabilitation device in motor training of the upper extremity after stroke [4], to trigger functional electrical stimulation (FES) in muscles when using an orthosis after spinal cord injury [5], or to elicit electrical stimulation with transcranial magnetic stimulation (TMS) during neurorehabilitation to enhance cortex excitability [6]. Typically, control commands are generated from these signal types using classification methods such as linear discriminant analysis (LDA) or support vector machines (SVM) [3, 7].

Recently, the use of convolutional neural networks (CNN) has been increasingly investigated for the classification of brain activity. It was stated that CNNs work reasonably well even in the presence of artifacts and noisy data. Their application is therefore interesting to tackle the challenge of poor SNR in real-time analyses. In addition, CNNs offer the possibility of being used in transfer learning, which can help to reduce timeconsuming data collection from individuals and lengthy training cycles. Here, the influence of possible differences in the measurement setup and electrode configurations between recordings must be considered, when data from multiple measurements and various individuals are combined [3, 8, 9].

The required classification steps for BCIs are usually carried out directly with sensor space data, i.e., with measurements obtained directly from the scalp surface.

However, EEG data can be analyzed not only in sensor space but also in source space. In general, source modeling is widely used across many fields of EEG analysis. This involves transforming the superficially measured signals into the source space using inverse operations. The signals derived at the scalp are influenced and distorted on their way from the brain to the head surface by the different tissue layers and their diverse electrical properties. In order to counteract this so-called volume conduction effect, an attempt is made to reconstruct the underlying activity of electrical sources in the brain using the signals measured externally on the scalp. Popular inverse methods for the transformation into the source space are minimum norm estimation (MNE), weighted MNE, and beamformers [10–13]. While source estimation has already been used occasionally in the context of BCIs, the combination with CNN offers a new perspective.

The aim of this preliminary work is to combine CNNbased classification of EEG data with three different methods of source analysis to compare the classification performance between sensor space and source space. For this purpose, a preestablished CNN architecture, the EEGNet [9] is applied to a publicly available MI-based EEG dataset from the Graz BCI competition IV [14]. CNN-based classification is conducted both directly on sensor measurements and source estimates generated with three different inverse approaches: Minimum Norm Estimation (MNE), weighted MNE (wMNE), and Beamformer.

METHODS

During this work, all analyses were performed using the Graz BCI competition IV dataset 2A [14]. This dataset contains EEG measurements from nine healthy subjects, who performed four different MI tasks. The four classes include cue-based imagination of left-hand, right-hand, feet, and tongue movement. There were two sessions recorded on different days each containing 288 trials in total, yielding 72 trials per class. EEG measurements were derived from 22 electrodes arranged according to the international 10-20 system with a sampling frequency of 250Hz. A more detailed description of this dataset can be found in [14].

In the context of this work classification of only two classes was desired, therefore only data from left-hand and right-hand MI were used. For each trial, ten overlapping 2s MI-epochs were extracted from the data spanning the MI period from 0.5s to 3.5s after cue onset, as indicated in [14].

To perform classification in sensor space, EEG measurements from the BCI dataset were provided as input to the EEGNet-based CNN, and the performance was evaluated using a 10-fold cross-validation. Before classification in source space, three different representations of source activity were computed using MNE, wMNE and beamformer. Then, classification was performed on each source estimation using CNN again in combination with a 10-fold cross-validation.

The three different inverse approaches for source transformation were computed using MATLAB version R2023a (The MathWorks Inc., https://www.mathworks.com). As individual head geometries were not available, the required lead field matrix for inverse computation was taken from the New York Head [15], which is a standardized finite element human head model that provides lead fields for 231 electrode locations and approximately 75,000 source locations. The scalar lead field matrix for fixed dipole orientation orthogonal to the cortex surface was used. First, the 22 channels that were also used in the BCI dataset were selected and the entire head model was down-sampled using cortex2K provided by the New York Head to reduce the number of sources. Next, regions of interest (ROI) were selected including postand precentral gyrus, central sulcus and paracentral lobule according to the Destrieux atlas [16]. In the end, a total of 122 source locations remained for estimation. For ease of comparison, the same number of sources was used for all three inverse operations. Prior to classification and source estimation EEG data were preprocessed with common average referencing (CAR) and the lead field matrix was restricted to the 22 channels from the BCI dataset and rereferenced. No further preprocessing or artifact removal was performed.

 Source estimation: In general, the relationship between the cortical sources and the resulting signals at sensors on the head surface can be described by the formulation in equation (1), where y denotes sensor signals, x denotes the underlying source activity, L denotes the lead field matrix, and n represents noise. The underlying sources \hat{x} can be reconstructed from measurements y following equation (2), where M describes the inverse operator that maps the sensor signals to the source space.

$$
y = Lx + n \tag{1}
$$

$$
\hat{x} = My \tag{2}
$$

The first and simplest approach for transforming the sensor measurements to the source space is MNE which uses Tikhonov regularization and minimizes the expression described in equation (3).

$$
\min_{x} \|Lx - y\|^2 + \lambda \|Wx\|^2 \tag{3}
$$

For the simple case of MNE, the weight matrix W corresponds to the identity matrix I. By solving this minimization problem, the inverse mapping operator M_{MNE} can be computed as described in equation (4). Here, λ denotes the regularization parameter, which was estimated from the SNR, and I corresponds to the identity matrix.

$$
M_{MNE} = L^T (LL^T + \lambda I)^{-1}
$$
 (4)

A more refined version of MNE is weighted MNE where additional weights W are introduced to compensate for the preference of weak superficial sources. The resulting

mapping operator M_{wMNE} is presented in equation (5). The weights W were computed based on the norm of the columns of the lead field matrix L.

$$
M_{wMNE} = W^{-1}L^{T}(LW^{-1}L^{T} + \lambda I)^{-1}
$$
 (5)

Finally, another approach for source estimation is provided by the linearly constrained minimum variance (LCMV) beamformer. In beamforming, a set of weights is computed for each predefined source location constructing a spatial filter that relates each sensor's contribution to each source estimate. These weights serve as the inverse operator M_{LCMV} and the computation of its components $m_{LCMV}(r)$ for the scalar case is described by equation (6). Here, R denotes the data covariance matrix and l(r) the column of the lead field matrix for a dipole of fixed orientation at location r.

$$
m_{LCMV}(r) = (l^T(r)R^{-1}l(r))^{-1}l^T(r)R^{-1}
$$
 (6)

The resulting three mapping operators, M_{MNE} , M_{wMNE} , and M_{LCMV} were applied to EEG measurements from the BCI dataset according to equation (2), to compute three different time course representations in source space.

 Classification with CNN: Subsequently, feature extraction and classification were performed with an EEGNet-based [9] CNN on four different signal modalities: sensor signals, source estimates generated with MNE, source estimates generated with wMNE, and source estimates from LCMV beamformer. The EEGNet is a compact CNN architecture consisting of two blocks, where at first a 2D temporal convolution is performed to learn frequency-specific features followed by depthwise convolution to learn spatial filters. In the second block 2D separable convolution is performed to summarize the individual features. In the end, classification is performed using a softmax algorithm, however, in the context of this work classification based on sigmoid function was used. As optimization algorithm the Adam optimizer was used. A more detailed explanation of the network can be found in [9]. Prior to training the CNN, a hyperparameter search was performed to find the most suitable model parameters. Based on that eight temporal and two spatial filters were implemented, the learning rate was set to 0.001, batch size was set to 32, and the number of epochs was set to 12 both for sensor and source space data. The EEGNet-based CNN was implemented in Python using TensorFlow (https://www.tensorflow.org/), Keras (https://keras.io/), and SciKit-Learn (https://scikitlearn.org/stable/) libraries.

In the end, classification accuracies of all four signal modalities were compared via the means of Kruskal-Wallis-Test, to determine whether there is a statistically significant difference between the methods.

RESULTS

For the classification in sensor space, an average classification accuracy of 79.25 \pm 13.90 % for all nine subjects was found. Individual classification results for

each subject for sensor space data are presented in Tab. 1, where the second column shows testing accuracies highlighted in bold, and column three shows training accuracies. The lowest accuracy in sensor space was found for subject 02 with 56.87 ± 2.92 %, and the highest accuracy was obtained for subject 08 with $96.04 \pm 1.35\%$.

Table 1: Classification accuracies in sensor space

Subject	Test accuracies	Training accuracies
#	in $\%$	in $%$
01	85.49 ± 2.71	88.26 ± 7.95
02	56.87 ± 2.92	79.86 ± 7.39
03	90.21 ± 3.10	98.19 ± 1.54
04	65.42 ± 5.65	85.76 ± 4.32
05	73.68 ± 5.06	88.26 ± 1.72
06	67.29 ± 4.09	87.92 ± 2.97
07	84.44 ± 2.41	93.06 ± 2.51
08	96.04 ± 1.35	99.03 ± 0.82
09	93.82 ± 0.89	93.06 ± 3.29
AVG	79.25 ± 13.90	90.38 ± 6.09

The classification accuracies for source estimates generated with MNE and wMNE can be seen in Tab. 2 and Tab. 3, respectively. Again, obtained accuracies on the test data are presented in column two and training accuracies are presented in column three.

Table 2: Classification accuracies in source space generated with MNE

Subject	Test accuracies	Training accuracies
#	in $\%$	in $%$
01	71.39 ± 8.87	73.47 ± 12.56
02	50.21 ± 2.62	78.06 ± 8.33
03	83.68 ± 13.13	91.46 ± 12.36
04	51.11 ± 2.08	56.67 ± 9.33
0 ₅	74.44 ± 6.24	83.61 ± 6.60
06	52.71 ± 2.33	62.36 ± 15.05
07	69.31 ± 10.25	86.94 ± 12.30
08	83.33 ± 11.95	94.72 ± 5.44
09	88.19 ± 3.91	82.85 ± 11.03
AVG	69.37 ± 14.84	78.90 ± 12.79

Table 3: Classification accuracies in source space generated with wMNE

The average classification accuracy over all subjects for the MNE-based source space was 69.37 ± 14.84 % and 69.94 \pm 17.63 % for wMNE. Here, the minimal accuracy

for MNE was found for subject 02 with 50.21 ± 2.62 %, and in wMNE for subject 04 with 50.00 ± 2.07 %. Similarly, the maximal accuracy in MNE source space was obtained for subject 09 with 88.19 ± 3.91 % and in wMNE source space for subject 03 with 93.40 ± 0.94 %. In Tab. 4 the classification accuracies for sources reconstructed using LCMV beamformer are provided. The average accuracy achieved with this inverse method was 78.14 ± 13.12 %. The lowest accuracy was achieved for subject 02 with 55.83 ± 2.60 % and maximum accuracy was achieved for subject 08 with 95.56 ± 1.10 %.

Table 4: Classification accuracies in source space generated with LCMV beamformer

Subject	Test accuracies	Training accuracies
#	in $\%$	in $%$
01	85.83 ± 3.63	95.35 ± 3.09
02	55.83 ± 2.60	89.17 ± 5.00
03	85.07 ± 6.87	98.47 ± 2.29
04	64.51 ± 6.28	87.71 ± 7.34
0 ₅	81.46 ± 4.31	91.11 ± 2.61
06	65.00 ± 3.50	90.69 ± 2.89
07	83.26 ± 5.86	94.58 ± 4.62
08	95.56 ± 1.10	99.51 ± 0.74
09	86.74 ± 3.72	97.99 ± 1.62
AVG	78.14 ± 13.12	93.84 ± 4.34

The distribution of classification accuracies across all four modalities for all nine subjects is depicted in Fig. 1. The comparison of all four methods with the Kruskal-

Wallis-Test showed that there is no significant difference (Chi square $= 2.93$, $p = 0.40$, df $= 3$) in classification performance between the different approaches.

DISCUSSION

In this study, the classification performance of a previously introduced CNN architecture, the EEGNet, was assessed for four different input datasets generated from a publicly available MI BCI dataset. The first input provided were unprocessed sensor measurements and the other three modalities were corresponding source estimates computed with three distinct inverse methods. Considering the classification accuracies for both sensor and source space data, the proposed EEGNet-based CNN was able to achieve a comparable performance to other standard classification methods from previous studies that were also evaluated on the BCI competition IV dataset [9, 17–19], with a relatively small number of epochs. The presented network yielded an average classification accuracy of about 79% on sensor space data and thereby shows slightly higher classification accuracies than e.g. Schirrmeister et al. who reported an average accuracy of about 74% with their ShallowConvNet [18] or Kar et al. who achieved an average accuracy of about 70% with their CNN-model [19]. It has to be considered, however, that the present work discriminated only two classes whereas most other studies performed classification of all four tasks contained in the dataset.

Figure 1: Distribution of classification accuracies across all four modalities (sensor measurements in blue, MNE estimates in orange, wMNE estimates in yellow, beamformer estimates in purple) for all nine subjects. For better visualization the y-axis was cut to contain only accuracy values between 45 and 100 %.

10.3217/978-3-99161-014-4-051

To keep conditions across models as similar as possible the same hyperparameters, including number of epochs, were used for all four modalities. Overall, the training accuracies in all four cases were quite high while still matching the general trend, meaning that subjects with lower testing accuracies also had lower training accuracies. In some cases, like for example with subject 02, the training accuracies were considerably higher than the accuracies achieved in testing, however, this is likely due to a higher intra-subject variability between the first and second recording session rather than due to overfitting. When comparing the classification results from sensor space to those achieved with source estimates in Fig. 1, it is evident that the accuracy values do not differ noticeably between the different modalities, even though it seems that source estimation with MNE and wMNE yielded slightly lower accuracies for most of the subjects. The Kruskal-Wallis-Test confirms that there is no significant difference between sensor space and any of the source space data. These findings are in contrast to other studies that found an improvement in BCI performance when applying source analysis prior to classification, although these studies did not use CNNs for classification [11, 12]. The results of the present work support the idea that CNNs are good at finding optimal separation criteria despite noisy data, as the compensation for the volume conduction effect via source estimation did not seem to significantly affect the classification performance. One reason for the comparatively weaker classification performance of source estimates could be the rather small number of measurement channels that are provided with the BCI dataset. In this work, 122 source estimates were generated using a standardized head model and only 22 channel EEG measurements with poor head coverage. No individual head geometries and exact electrode positions were available. It was previously shown, however, that an increased electrode density and good head coverage allow for improved source reconstruction [20, 21]. Other studies that found an improvement in BCI performance in combination with source estimation used a larger number of sensors distributed over the entire head [11, 12]. Results of source estimation could further be improved by additional preprocessing such as data whitening with noise covariances, regional clustering for optimally reducing the source space, or introducing spatial normalization strategies to beamformers [12, 13, 22, 23]. So, with measurements providing better electrode configurations spanning the entire head surface, computation of more accurate source estimates and thereby an improvement of classification performance might be feasible.

Nonetheless, as comparable classification accuracies could be achieved with source estimates, future implementations using CNN and source reconstruction for transfer learning in BCIs are promising. The transformation of sensor measurements to the source space offers the possibility to jointly process data of several individuals and different recording sessions in a common signal space, as the effect of changes in electrode configuration can be circumvented.

CONCLUSION

Accurate real-time classification of EEG data is essential to provide users with precise and intuitive BCI control. In this study, it was shown that CNNs can achieve comparable classification performances based on sensor and source space data respectively. Future investigations will focus on the real-time applicability of CNNs and transfer learning for BCI systems will be explored. The reliable classification of source estimates as shown by this work is promising for the application in transfer learning. Furthermore, CNN-based classification of source estimates will also be performed on own recorded data during MI and arithmetic learning tasks, to explore whether alternative tasks can be discriminated from EEG measurements and could be used for providing neurofeedback.

ACKNOWLEDGEMENTS

This research was funded in whole or in part by the Austrian Science Fund (FWF), Grant DOI 10.55776/P35949, and the Tyrolian Science Fund (TWF), project F.45065. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

REFERENCES

- [1] K. Värbu, N. Muhammad, and Y. Muhammad, "Past, present, and future of EEG-based BCI applications," *Sensors*, vol. 22, no. 9, p. 3331, 2022.
- [2] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain‐computer interface paradigms," *Journal of Neural Engineering*, vol. 16, no. 1, p. 11001, 2019.
- [3] F. Lotte *et al.,* "A review of classification algorithms for EEG-based brain‐computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 31005, 2018.
- [4] K. K. Ang *et al.,* "A clinical study of motor imagery-based brain-computer interface for upper limb robotic rehabilitation," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 5981– 5984.
- [5] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "'Thought'‐control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neuroscience letters*, vol. 351, no. 1, pp. 33–36, 2003.
- [6] A. Gharabaghi *et al.,* "Coupling brain-machine interfaces with cortical stimulation for brain-state dependent stimulation: enhancing motor cortex excitability for neurorehabilitation," *Frontiers in Human Neuroscience*, vol. 8, p. 122, 2014.
- [7] P. Wierzgała, D. Zapała, G. M. Wojcik, and J. Masiak, "Most popular signal processing methods in motor-imagery BCI: a review and metaanalysis," *Frontiers in Neuroinformatics*, vol. 12, p. 78, 2018.
- [8] 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. 31001, 2019.
- [9] 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 EEGbased brain‐computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 56013, 2018.
- [10] R. Grech *et al.*, "Review on solving the inverse problem in EEG source analysis," *Journal of Neuroengineering and Rehabilitation*, vol. 5, no. 1, pp. 1–33, 2008.
- [11] B. J. Edelman, B. Baxter, and B. He, "EEG source imaging enhances the decoding of complex righthand motor imagery tasks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 4–14, 2015.
- [12] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss, "Beamforming in noninvasive brain‐ computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1209– 1219, 2009.
- [13] B. U. Westner *et al.,* "A unified view on beamformers for M/EEG source reconstruction," *NeuroImage*, vol. 246, p. 118789, 2022, doi: 10.1016/j.neuroimage.2021.118789.
- [14] C. Brunner, R. Leeb, and G. Muller-Putz IV, "BCI Competition IV Dataset 2a: 4-Class Motor Imagery," *Graz University of Technology: Graz, Austria*, 2008.
- [15] Y. Huang, L. C. Parra, and S. Haufe, "The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting," *NeuroImage*, vol. 140, pp. 150–162, 2016.
- [16] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *NeuroImage*, vol. 53, no. 1, pp. 1–15, 2010.
- [17] I. Dolzhikova, B. Abibullaev, R. Sameni, and A. Zollanvari, "An Ensemble CNN for Subject-Independent Classification of Motor Imagerybased EEG," in *IEEE 2021 – 2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 319– 324.
- [18] R. T. Schirrmeister *et al.,* "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017, doi: 10.1002/hbm.23730.
- [19] A. Kar, S. Bera, S. K. P. Karri, S. Ghosh, M. Mahadevappa, and D. Sheet, "A deep convolutional neural network based classification of multi-class

motor imagery with improved generalization," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 5085–5088.

- [20] J. Song *et al.*, "EEG source localization: Sensor density and head surface coverage," *Journal of Neuroscience Methods*, vol. 256, pp. 9–21, 2015, doi: 10.1016/j.jneumeth.2015.08.015.
- [21] G. Lantz, R. G. de Peralta, L. Spinelli, M. Seeck, and C. M. Michel, "Epileptic source localization with high density EEG: how many electrodes are needed?," *Clinical Neurophysiology*, vol. 114, no. 1, pp. 63–69, 2003.
- [22] F.-H. Lin, T. Witzel, S. P. Ahlfors, S. M. Stufflebeam, J. W. Belliveau, and M. S. Hämäläinen, "Assessing and improving the spatial accuracy in MEG source localization by depthweighted minimum-norm estimates," *NeuroImage*, vol. 31, no. 1, pp. 160–171, 2006, doi: 10.1016/j.neuroimage.2005.11.054.
- [23] C. Dinh *et al.*, "Real-Time MEG Source Localization Using Regional Clustering," *Brain Topography*, vol. 28, no. 6, pp. 771–784, 2015, doi: 10.1007/s10548-015-0431-9.