2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

An Ensemble CNN for Subject-Independent Classification of Motor Imagery-based EEG

Irina Dolzhikova[1], Berdakh Abibullaev[2], Reza Sameni[3] and Amin Zollanvari[4]

Abstract— **Deep learning methods, and in particular Convolutional Neural Networks (CNNs), have shown breakthrough performance in a wide variety of classification applications, including electroencephalogram-based Brain Computer Interfaces (BCIs). Despite the advances in the field, BCIs are still far from the subject-independent decoding of brain activities, primarily due to substantial inter-subject variability. In this study, we examine the potential application of an ensemble CNN classifier to integrate the capabilities of CNN architectures and ensemble learning for decoding EEG signals collected in motor imagery experiments. The results prove the superiority of the proposed ensemble CNN in comparison with the average base CNN classifiers, with an improvement up to 9% in classification accuracy depending on the test subject. The results also show improvement with respect to the performance of a number of state-of-the-art methods that have been previously used for subject-independent classification in the same datasets used here (i.e., BCI Competition IV 2A and 2B datasets).**

I. INTRODUCTION

Brain-Computer Interface (BCI) aims to establish a robust channel between user and device, bypassing normal neuromuscular pathways by interpreting the user's brain activity patterns and decoding them into useful commands for external machine/computer control [1], [2], [3].

One of the most commonly used brain signal acquisition methods is by electroencephalogram (EEG) recordings [4]. As a non-invasive data collection method, the EEG is estimated to measure only around 5 % of the original electrical brain activities. At the same time, the remaining 95 % are attenuated and smoothed by multiple layers such as the skull and tissues, or obscured by background activities [4], [5]. The issues associated with such a low signal-to-noise ratio are further amplified by extreme variability in the signal patterns across various subjects. As a result, most of the available BCIs apply subject-specific (subject-dependent) training, which requires the user to spend time on the training/calibration, so that the subject's data could be used to adapt the system for future use [6], [7]. This procedure is time-consuming and especially inconvenient for people with disabilities. Therefore, recent research attempts to move towards calibration-free (subject-independent) systems [8].

[1]I. Dolzhikova is with the Electrical and Computer Engineering Department, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan, IEEE Student Member, ifedorova@nu.edu.kz

[2]B. Abibullaev is with the Robotics and Mechatronics Department, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan, IEEE Member, berdakh.abibullaev@nu.edu.kz

[3]R. Sameni is with the Department of Biomedical Informatics, School of Medicine, Emory University, USA, IEEE Senior Member, rsameni@dbmi.emory.edu

[4]A. Zollanvari is with the Electrical and Computer Engineering Department, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan, IEEE Senior Member, amin.zollanvari@nu.edu.kz

This work was supported by the Faculty Development Competitive Research Grants Program of Nazarbayev University under grant number 021220FD1151

In this paper, we investigate the capabilities of an ensemble CNN in the context of subject-independent binary classification of motor imagery (MI) tasks. MI, which is the process of imagining the movement of specific parts of the body, is an endogenous process (i.e. stimuli independent), which uses the sensory-motor rhythms (SMRs, i.e. changes in the oscillatory activity) to control the BCI device [9], [10]. Generally, measuring the brain's biometric information allows to reflect active and passive mental states of the subjects and recognize their psychological and physical status. In case of MI-based BCI, event-related synchronization (ERS) and event-related desynchronization (ERD) patterns elicited in SMRs are measured [11], [12]. These patterns are believed to be responsible for the changes in the energy levels in the contralateral/ipsilateral areas, due to certain MI tasks.

An overview of the existing predictive models in the area of BCI suggests that deep learning algorithms outperform traditional machine learning methods in EEG classification [13], [14]. In particular, Convolutional Neural Networks (CNNs) have demonstrated their high potential in decoding complex ERD/ERS patterns in various BCI applications with efficient training time [13], [14], [15], [16], [17]. In some research CNN has been used as a classifier that is trained on the pre-processed features fed from Common Spatial Pattern (CSP) [18] or Short-Time Fourier Transform (STFT) [19], while others take advantage of CNN efficiency in terms of capturing latent features from raw EEG [16], [17], [20]. CNNs are able to learn from raw data by first extracting the local low-level features from the input and then learn global high-level features in the deeper layers [20].

The advantages of CNN could be enhanced by integrating several CNN architectures into a single predictive model via ensemble method as it has potential to deal with variability of the data [23]. The summary of the studies exploring the performance of various ensemble models towards MI application is presented in Table I. Accordingly, most prior research in this area use the traditional machine learning methods, such as Support Vector Machine (SVM) [23], Linear Discriminant Analysis (LDA), K Nearest Neighbor (KNN), decision trees [21], [22], [24], [25], [26] and rarely take the advantage of deep learning architectures [27], [28]. Therefore, in this study in order to cope with substantial

TABLE I

SUMMARY OF PREVIOUS RESEARCH EXPLORING THE ENSEMBLES FOR THE MOTOR IMAGERY TASKS

Study	Base classifiers	Type of ensemble
[21]	SVM	Dynamic Weighted Ensemble classification
[22]	SVM, LDA, Decision Tree	Bagging, Stacking
[23]	LDA, Artificial Neural Network	Stacked Ensembles
[24]	Decision Stump, KNN	Boosting, Bagging
[25]	LDA, SVM, Radial Basis Function, KNN	Stacking
[26]	SVM, Adaptive boost, Logitboost	Boosting, Bagging
[27]	Artificial Neural Network	Simple Averaging, Stacked Generalization
[28]	Feed Forward Neural Network	Adaboost, LPBoost, RUSBoost, Bagging, Random Subspace

TABLE II

NUMBER OF OBSERVATIONS USED AS A TRAINING/VALIDATION/TESTING SET FOR EACH ENSEMBLE METHOD TO ESTABLISH BASE CLASSIFIERS

Dataset	Original number of trials per subject	Number of trials after segmentation	Number of trials in training set	Validation set/Test set
2A	288	864	6048	864
2B	680, 720, 740, 760	2040, 2160, 2220, 2280	14280, 15120, 15540,15960	2040, 2160, 2220, 2280

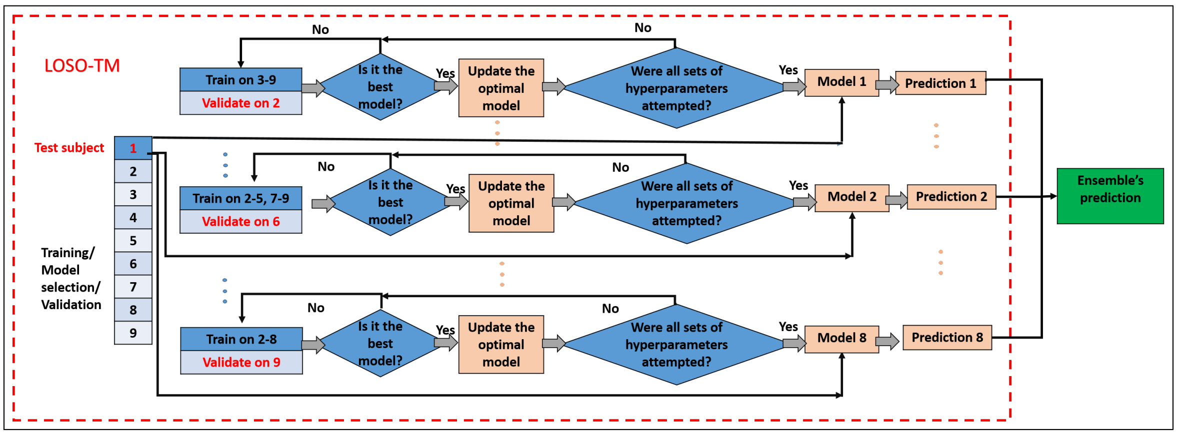

Fig. 1. LOSO-TM method for base classifiers training: leave out one subject with remaining subjects being held-out one at a time to create the validation set such that the training is performed on remaining 7 subjects

inter-subject variability in constructing accurate subject-independent classifiers, we develop an ensemble of CNN classifiers trained on data collected from multiple subjects excluding the test subject. The proposed *multi-subject ensemble CNN* is a promising tool as it achieves good performance in subject-independent classification, which is *per se* an indicator of robustness to the inter-subject variability.

The remainder of the paper is structured as follows. Section II reviews the description of the datasets used in this study. Section III describes ensemble learning, training base classifiers for an ensemble, and the systematic model selection strategy used in this study. Results and discussion are presented in Section IV. Section V concludes the paper.

II. DATASETS

This study uses two publicly available MI datasets: BCI Competition IV 2A [29] and BCI Competition IV 2B [30]. These datasets were collected under different experimental conditions that include several types of motor imagery tasks. For the current study, only those limb imagery data corresponding to left-hand and right-hand imagery were selected. The brief description of these datasets is presented below.

1) BCI Competition IV 2A (BCI IV 2A): Nine subjects participated in the cue-based BCI experiment performing four types of imagery tasks (imagination of left-hand, right-hand, both feet, and tongue movement). The data were obtained using 22 Ag/AgCl electrodes according to the 10-20 electrode placement system, with the reference electrode being set to the left mastoid, while the ground was set to the right mastoid. The sampling frequency was 250 Hz and band-pass filtering range is 0.5 Hz-100 Hz. More details could be found in [29].

2) BCI Competition IV 2B (BCI IV 2B): This dataset contains EEG from nine healthy, right-handed subjects. During the experiment, the subjects performed two-class motor imagery of the left-hand and right-hand movements. The data were collected at a sampling frequency of 250 Hz with three channels (C3, Cz, and C4), while electrode Fz was used as the EEG ground (reference lead). The frequency range 0.5 Hz-100 Hz was used for bandpass filtering (see [30] for more details).

For both of the datasets we segmented original trials of 4 seconds duration into a series of shorter segments of 2 second with an overlap of 50%. This was done with the aim of increasing the sample size, as with small number of training samples, deep neural networks, such as the CNN, are prone to over-fitting.

III. Multi-Subject Ensemble CNN

A. Ensemble Technique

Ensemble learning integrates several predictive and possibly unstable models to reduce their variance and to improve the overall predictive performance. An ensemble classifier $\psi^{\mathrm{E}}(\mathbf{x})$ is given by

$$\psi^{\mathrm{E}}(\mathbf{x}) = \mathcal{C}\left(\bigcup_{i=1}^{M}\{\psi_i(\mathbf{x})\}\right) \tag{1}$$

where $\psi_i(\cdot)$, $i = 1, \ldots, M$ denotes the set of M base classifiers, $\mathcal{C}(\cdot)$ is the combination rule, and $\mathbf{x}$ denotes the p-dimensional feature vector to be classified. In the present study, to design an ensemble we combine several CNN architectures that are trained on multiple subjects and are outcomes of a systematic model selection process. The voting mechanisms (combination rules, or combiners) used to form the ensemble include majority vote (MV) and probabilistic classifier ensemble weighting (PCEW) [31].

Using the MV rule, $\psi^{\mathrm{E}}(\mathbf{x})$ is determined based on the decision of the majority of the base classifiers [31], such that

$$\psi^{\mathrm{E}}(\mathbf{x}) = \underset{y\in\{0,1\}}{\arg\max} \sum_{i=1}^{M} \mathrm{I}_{\{\psi_i(\mathbf{x})=y\}}\,, \tag{2}$$

where $\mathrm{I}_{\{S\}}$ is 1 if statement S is true, and zero otherwise [31].

PCEW first calculates the probabilities of class $y \in \{0, 1\}$ given feature vector $\mathbf{x}$ and base classifier $i \in \{1, ..., M\}$. These probabilities, denoted as $\hat{p}_{iy}$, are then weighted by the estimated accuracies of individual classifiers ($\hat{\mathrm{acc}}_i$) raised to the power of α, which shows the degree of trust to the estimated accuracies. Finally, $\mathbf{x}$ is classified to the class with the highest weighted sum of the probabilities $\hat{p}_{iy}$. Formally, this combination rule is presented as

$$\psi^{\mathrm{E}}(\mathbf{x}) = \underset{y\in\{0,1\}}{\arg\max} \sum_{i=1}^{M} \hat{\mathrm{acc}}_i^{\alpha}\,\hat{p}_{iy}\,. \tag{3}$$

B. Training the Base Classifiers

Let S_i and $S = \cup_{i=1}^{K} S_i$ denote the EEG data collected from subject i, and from all K subjects, respectively ($K = 9$ for both BCI IV 2A and 2B datasets). To ensure the subject-independent evaluation of the proposed design, we apply the leave-one-subject-out cross-validation (LOSO-CV) procedure, where the data for each subject S_i is successively held out and not seen during the training/model selection process, and is used only to assess the performance of the constructed ensemble classifier. In addition to this *external* LOSO-CV, there is an *internal* LOSO-CV that is used for model selection and training each ensemble classifier, which is utilized to classify observations for the held-out subject S_i. The effect of the inner LOSO-CV is to simulate the subject-independent context within the training data $S - S_i$, i.e., $K - 1$ subjects at a time, and estimate the best set of hyperparameters to be used in training the CNN base classifiers within the ensemble. That is to say, in order to construct the ensemble classifier (III-A), $K - 1$ base classifiers are constructed where each classifier is trained using data for $K - 2$ subjects ($S - S_i - S_j, j \neq i$) with the hyperparameters being tuned using data for subject S_j, which is held out due to the internal LOSO-CV process.

Table II shows the number of observations available in two datasets prior to and after segmentation, the number of trials used during training, validation, and testing as described before. Multiple numbers specified in some entries associated with BCI IV 2B dataset indicate different number of trials across subjects who participated in the BCI experiment. Fig. 1 shows a schematic diagram of the entire process of the internal LOSO for training and model selection (LOSO-TM) as well as the subject-independent assessment via the external LOSO.

C. Model Selection and Hyperparameter Tuning

Within the LOSO-TM process, we used a brute-force (exhaustive) search to perform model selection. In this regard, similar to [16], [17], we first defined a limited space of hyperparameters within which the exhaustive search was conducted to tune the hyperparameters of the constructed CNN base classifiers used in the ensemble learning. The following assumptions were made to define the possible search space of hyperparameters: (i) assuming $L \in \{2, 3, \ldots, 6\}$ denotes the total number of layers used in CNN architectures, we considered both increasing and decreasing patterns for the number of output channels across layers. In particular, for $j = 1, \ldots, L$, it was assumed that the number of output channels in each layer is 2^{2+j} and $2^{(L+3-j)}$ for an increasing and decreasing pattern, respectively; (ii) kernel shapes of size 3×8, 3×24, 3×40; (iii) the mini-batch gradient decent with Adam optimizer [32] with a learning rate of 10^{-4} and decay of 10^{-5}; (iv) a cross-entropy loss function [33]; (v) a maximum number of epochs of 150, interrupted by early stopping with a patience parameter of 30 [34]; (vi) RELU activation function in all layers; and (vii) the last convolutional layer was followed by two fully-connected layers with 256 and 2 output features, respectively. The aforementioned structural and algorithmic hyperparameters set the cardinality of hyperparameter space to 5 (number of layers) $\times$ 2 (increasing or decreasing patterns of output channels) $\times$ 3 (kernel size) $\times$ 150 (maximum number of epochs) $\times$ 1 (learning rate) $\times$ 1 (decay factor) $\times$ 1 (batch size) = 4500. Recall that due to the external LOSO used for subject-independent assessment of the proposed ensemble classification rule, K ensemble classifiers were constructed, each composed of $K - 1$ base CNN classifiers. As part of the model selection conducted within the aforementioned space of hyperparameters, each base classifier is the one (among at most 4500 classifiers) that led to the highest estimated accuracy on the corresponding validation set (i.e., the data for the held-out subject as part of the internal LOSO described in Section III-B).

TABLE III

TEST ACCURACY RESULTS FOR DIFFERENT ENSEMBLING SCHEMES USING DATA FROM BCI IV 2A DATASET. THE COLUMNS LABELED "MINIMUM", "MAXIMUM", AND "AVERAGE" SHOW THE MINIMUM, MAXIMUM, AND AVERAGE ACCURACY OF 8 CONSTRUCTED BASE CLASSIFIERS ON THE DATA FOR THE TEST SUBJECT (I.E., THE SUBJECT WHO IS UNSEEN DURING THE TRAINING)

Test subject	**Minimum % (Base Classifiers)**	**Maximum % (Base Classifiers)**	**Average % (Base Classifiers)**	**MV %**	**PCEW %** ($\alpha = 1$)	**PCEW %** ($\alpha = 5$)
1	63.31	70.72	66.35	70.60	71.87	72.22
2	48.26	53.70	51.46	53.00	52.78	53.12
3	64.93	77.31	71.08	78.82	77.89	76.16
4	54.40	64.58	58.75	59.61	60.76	60.65
5	52.08	56.71	54.01	56.02	56.48	54.86
6	55.21	62.50	58.93	59.26	60.30	60.07
7	53.36	61.80	58.09	61.57	61.34	60.42
8	66.78	74.07	69.69	76.85	76.16	77.43
9	58.56	66.43	62.50	61.69	61.46	62.6
avg±std	**57.43±6.00**	**65.32±7.30**	**61.20±6.37**	**64.16±8.60**	**64.34±8.31**	**64.17±8.42**

TABLE IV

TEST ACCURACY RESULTS FOR DIFFERENT ENSEMBLING SCHEMES USING DATA FROM BCI IV 2B DATASET. THE COLUMNS LABELED "MINIMUM", "MAXIMUM", AND "AVERAGE" SHOW THE MINIMUM, MAXIMUM, AND AVERAGE ACCURACY OF 8 CONSTRUCTED BASE CLASSIFIERS ON THE DATA FOR THE TEST SUBJECT (I.E., THE SUBJECT WHO IS UNSEEN DURING THE TRAINING)

Test subject	**Minimum % (Base Classifiers)**	**Maximum % (Base Classifiers)**	**Average % (Base Classifiers)**	**MV %**	**PCEW %** ($\alpha = 1$)	**PCEW %** ($\alpha = 5$)
1	59.63	65.00	62.39	64.17	65.00	64.58
2	52.11	55.34	53.49	53.87	54.26	54.61
3	52.31	54.91	53.69	54.12	53.84	53.70
4	77.25	81.89	80.41	83.29	82.61	82.79
5	57.88	64.46	62.53	63.96	65.81	64.37
6	63.56	70.55	67.77	71.06	71.02	70.14
7	61.25	66.85	64.37	65.74	65.83	66.02
8	66.80	72.41	70.45	70.53	71.84	71.23
9	65.51	70.32	68.09	69.77	70.79	70.51
avg±std	**61.81±7.34**	**66.86±7.94**	**64.80±7.87**	**66.28±8.52**	**66.78±8.43**	**66.44±8.37**

IV. RESULTS AND DISCUSSION

A. Results of the Proposed Method

This section presents the results of subject-independent evaluation achieved by applying ensemble classifiers developed using two combination rules (MV and PCEW [$\alpha = 1$ and $\alpha = 5$]). The minimum, the maximum, and the average test accuracies of the base classifiers, as well as the test accuracies of the ensemble classifiers are presented in Tables III and IV for BCI IV 2A and BCI IV 2B datasets, respectively. Comparing the results achieved for the MV and the PCEW shows that there is virtually no difference between these methods. Therefore, hereafter, we will use the results of the MV combination rule that required no additional hyperparameter such as α used in PCEW. Comparing the accuracies of the ensemble classifiers with the average accuracies of the constructed base classifiers on test data shows that the ensemble scheme significantly outperforms the base classifiers. This observations is verified statistically by a one-sided (paired) Wilcoxon signed rank test with the alternative hypothesis that the classification accuracy achieved by an ensemble classifier is greater than the average classification accuracy obtained by its base classifiers. Table V shows the *P*-values of the tests for all pairwise comparison between the performance of the ensemble classifiers and the average performance of base classifiers. It is observed that all P-values are smaller than 0.01, which is far below the common significance level of 0.05, indicating the significant improvement achieved by the ensemble classifiers.

B. Comparison with the State-of-the-Art Methods

Zhang *et al.* [39] and Roy *et al.* [40] have recently reported the subject-independent classification performance of their developed classifiers on BCI IV 2A and 2B datasets, respectively. In particular, Zhang *et al.* [39] proposed the Convolutional Recurrent Attention Model (CRAM), which encodes the high-level representation of EEG data via CNN and learns the temporal patterns using a recurrent mechanism. They showed that CRAM outperforms a number of state-of-the-art methods including: EEGNet [15], Cropped-Training CNN (CTCNN) [35], EEG-Image [36], AE-XGboost [37], Filter Bank Common Spatial Pattern (FBCSP) [38], a three-layered CNN proposed in [39], and a two-layer Long Short-Term Memory (LSTM) proposed in [39]. Roy *et al.* [40], on the other hand, proposed a CNN architecture and evaluated its performance on the BCI IV 2B dataset. Table VI presents the average (over all subjects) subject-independent classifi-

TABLE V

P-VALUES CALCULATED USING ONE-SIDED WILCOXON SIGNED RANK TEST FOR PAIRWISE COMPARISON BETWEEN THE CLASSIFICATION ACCURACY ACHIEVED BY AN ENSEMBLE VS. AVERAGE ACCURACY OBTAINED USING THE BASE CLASSIFIERS FOR TWO DATASETS

Combiner vs. Average	BCI IV 2A	BCI IV 2B
MV vs. Average	0.00586	0.00195
PCEW ($\alpha = 1$) vs. Average	0.00391	0.00195
PCEW ($\alpha = 5$) vs. Average	0.00195	0.00195

TABLE VI

PERFORMANCE COMPARISON OF THE SUBJECT-INDEPENDENT (SI) METHODS USING BCI IV 2A AND 2B DATASETS

Method	Dataset	Mean±STD %	Range (MIN–MAX) %
EEGnet [15]	BCI IV 2A	51.32±5.18	21.84 (39.54–61.38)
CTCNN [35]	BCI IV 2A	47.67±6.18	44.62 (26.04–70.66)
EEG Image [36]	BCI IV 2A	32.70±7.18	12.67 (26.40–39.06)
AE Xgboost [37]	BCI IV 2A	33.18±8.18	8.16 (30.9–39.06)
FBCSP [38]	BCI IV 2A	35.69±9.18	23.09 (24.83–47.92)
CNN [39]	BCI IV 2A	47.20±10.18	16.66 (41.15–57.81)
RNN [39]	BCI IV 2A	35.48±11.18	6.49 (32.63–39.12)
CRAM [39]	BCI IV 2A	59.22±12.18	30.76 (42.35–73.11)
Proposed	**BCI IV 2A**	**64.16±8.60**	**25.81 (53.00–78.82)**
Roy [40]	BCI IV 2B	67.78±14.18	27.22 (54.2–81.42)
Proposed	**BCI IV 2B**	**66.28±8.52**	**29.42 (53.87–83.29)**

TABLE VII

P-VALUES CALCULATED USING ONE-SIDED WILCOXON SIGNED RANK TEST FOR PAIRWISE COMPARISON BETWEEN THE CLASSIFICATION ACCURACY ACHIEVED BY THE PROPOSED METHOD AND OTHER METHODS FOR BCI 2A AND 2B DATASETS

State-of-the-art methods	Dataset	*P*-values
EEGnet [15]	BCI IV 2A	0.00195
CTCNN [35]	BCI IV 2A	0.00195
EEG Image [36]	BCI IV 2A	0.00195
AE Xgboost [37]	BCI IV 2A	0.00195
FBCSP [38]	BCI IV 2A	0.00195
CNN [39]	BCI IV 2A	0.00195
RNN [39]	BCI IV 2A	0.00195
CRAM [39]	BCI IV 2A	0.02734
Roy [40]	BCI IV 2B	0.95120

cation accuracies for achieved different methods reported in [39] and [40] over BCI IV 2A and 2B datasets, respectively.

Table VII shows the P-values obtained using a one-sided (paired) Wilcoxon signed rank test to compare the subject-independent performance of the proposed ensemble scheme (i.e., results presented in Table III) with those reported in [39]. As it can be seen in this table, in all cases the P-values are less than the significance level of 0.05, which indicates a significant improvement in classification performance achieved by the proposed multi-subject ensemble CNN over all methods assessed on BCI IV 2A dataset in [39].

On the other hand, the result presented in Table VI for BCI IV 2B dataset shows no improvement in average classification performance achieved by our method with respect to that of Roy et al. [40] (corroborated with a P-value close to 1 in Table VII). Nevertheless, a two-sided (paired) Wilcoxon signed rank test for the hypothesis of distinction between our results with those of Roy et al. [40] reported on BCI IV 2B dataset has a P-value of 0.12890. In other words, on BCI IV 2B dataset, we observe no significant difference between the performance of our multi-subject ensemble CNN with the performance CNN-based Mega Blocks proposed in [40]. Altogether, these observations confirm the efficacy of the proposed CNN ensemble scheme with respect to a number of state-of-the-art methods for subject-independent classification of motor imagery based EEG.

V. CONCLUSION

In this work, we examined the feasibility of using a multi-subject ensemble CNN architecture for classification of motor imagery based EEG collected from individuals that have been entirely unseen during the training phase (i.e., subject-independent classification). The results show performance improvement achieved by the proposed ensemble CNN scheme with respect to the base CNN classifiers as well as a number of state-of-the-art methods proposed previously in the literature. These results, in turn, imply the efficacy of the proposed scheme in tackling substantial inter-subject variability that is the major impediment to achieving high accuracy in subject-independent classification of EEG records in BCI applications. An important research issue for the future is to study the effect of the number of subjects that are used in training (including model selection) on the performance of the proposed multi-subject ensemble CNN classification rule.

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, Jun. 2002.

[2] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran, "A brain–computer interface using electrocorticographic signals in humans," *Journal of neural engineering*, vol. 1, no. 2, p. 63, Jun. 2004.

[3] A. Vallabhaneni, T. Wang, and B. He, "Brain—computer interface," in *Neural Engineering*. Springer, 2005, pp. 85–121.

[4] X. Zhang, L. Yao, X. Wang, J. J. Monaghan, D. Mcalpine, and Y. Zhang, "A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers," *Journal of Neural Engineering*, 2020.

[5] T. Ball, M. Kern, I. Mutschler, A. Aertsen, and A. Schulze-Bonhage, "Signal quality of simultaneously recorded invasive and non-invasive eeg," *Neuroimage*, vol. 46, no. 3, pp. 708–716, 2009.

[6] 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.

[7] M. Rashid, N. Sulaiman, A. PP Abdul Majeed, R. M. Musa, A. F. Ab Nasir, B. S. Bari, and S. Khatun, "Current status, challenges, and possible solutions of EEG-based brain-computer interface: A comprehensive review," *Frontiers in Neurorobotics*, 2020.

[8] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural networks*, vol. 22, no. 9, pp. 1305–1312, 2009.

[9] G. Pfurtscheller, "Spatiotemporal erd/ers patterns during voluntary movement and motor imagery," *Supplements to Clinical neurophysiology*, vol. 53, pp. 196–198, 2000.

[10] S. Aggarwal and N. Chugh, "Signal processing techniques for motor imagery brain computer interface: A review," *Array*, vol. 1, p. 100003, 2019.

[11] G. Pfurtscheller and C. Neuper, "Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments," *Progress in brain research*, vol. 159, pp. 433–437, 2006.

[12] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[13] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051001, 2019.

[14] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. Mcalpine, and Y. Zhang, "A survey on deep learning based brain computer interface: Recent advances and new frontiers," *arXiv preprint arXiv:1905.04149*, 2019.

[15] V. Lawhern, A. Solon, N. Waytowich, S. Gordon, C. Hung, and B. Lance, "EEGNet: a compact convolutional network for EEG-based brain-computer interfaces. arXiv," *arXiv preprint arXiv:1611.08024*, 2016.

[16] B. Abibullaev, I. Dolzhikova, and A. Zollanvari, "A brute-force CNN model selection for accurate classification of sensorimotor rhythms in BCIs," *IEEE Access*, 2020.

[17] B. Abibullaev and A. Zollanvari, "A systematic deep learning model selection for p300-based brain-computer interfaces," *IEEE Transactions on Systems, Man, and Cybernetics: Systems, DOI: 10.1109/TSMC.2021.3051136*, 2021.

[18] X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, and P. Xu, "Separated channel convolutional neural network to realize the training free motor imagery bci systems," *Biomedical Signal Processing and Control*, vol. 49, pp. 396–403, 2019.

[19] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 23, p. e4413, 2018.

[20] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5619–5629, 2018.

[21] S. R. Liyanage, C. Guan, H. Zhang, K. K. Ang, J. Xu, and T. H. Lee, "Dynamically weighted ensemble classification for non-stationary EEG processing," *Journal of neural engineering*, vol. 10, no. 3, p. 036007, 2013.

[22] M. Rahimi, A. Zarei, E. Nazerfard, and M. H. Moradi, "Ensemble methods combination for motor imagery tasks in brain computer interface," in *2016 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 2016, pp. 336–340.

[23] B. Liu and H. Hao, "Application of ensemble classifier in eeg-based motor imagery tasks," in *MIPPR 2007: Medical Imaging, Parallel Processing of Images, and Optimization Techniques*, vol. 6789. International Society for Optics and Photonics, 2007, p. 678913.

[24] M. Mohammadpour, M. Ghorbanian, and S. Mozaffari, "Comparison of EEG signal features and ensemble learning methods for motor imagery classification," in *2016 Eighth International Conference on Information and Knowledge Technology (IKT)*. IEEE, 2016, pp. 288–292.

[25] A. C. Ramos, R. G. Hernández, M. Vellasco, and P. Vellasco, "Ensemble of classifiers applied to motor imagery task classification for BCI applications," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 2995–3002.

[26] A. Datta and R. Chatterjee, "Comparative study of different ensemble compositions in EEG signal classification problem," in *Emerging Technologies in Data Mining and Information Security*. Springer, 2019, pp. 145–154.

[27] R. Ebrahimpour, K. Babakhani, and M. Mohammad-Noori, "EEG-based motor imagery classification using wavelet coefficients and ensemble classifiers," in *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012)*. IEEE, 2012, pp. 458–463.

[28] S. Bhattacharyya, A. Konar, D. Tibarewala, A. Khasnobish, and R. Janarthanan, "Performance analysis of ensemble methods for multi-class classification of motor imagery EEG signal," in *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*. IEEE, 2014, pp. 712–716.

[29] 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," *Frontiers in neuroscience*, vol. 6, p. 55, 2012.

[30] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 4, pp. 473–482, 2007.

[31] J. Large, J. Lines, and A. Bagnall, "A probabilistic classifier ensemble weighting scheme based on cross-validated accuracy estimates," *Data mining and knowledge discovery*, vol. 33, no. 6, pp. 1674–1709, 2019.

[32] C. M. Bishop *et al.*, *Neural networks for pattern recognition*. Oxford university press, 1995.

[33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[34] R. Caruana, S. Lawrence, and L. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," *Advances in neural information processing systems*, pp. 402–408, 2001.

[35] 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.

[36] 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.

[37] X. Zhang, L. Yao, D. Zhang, X. Wang, Q. Z. Sheng, and T. Gu, "Multi-person brain activity recognition via comprehensive EEG signal analysis," in *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2017, pp. 28–37.

[38] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 2390–2397.

[39] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent EEG signal analysis," *IEEE Signal Processing Letters*, vol. 26, no. 5, pp. 715–719, 2019.

[40] S. Roy, A. Chowdhury, K. McCreadie, and G. Prasad, "Deep learning based inter-subject continuous decoding of motor imagery for practical brain-computer interfaces," *Frontiers in Neuroscience*, vol. 14, 2020.