Generative Adversarial Networks-Based Data Augmentation for Brain–Computer Interface

Fatemeh Fahimi, Strahinja Dosen, *Member, IEEE*, Kai Keng Ang, *Senior Member, IEEE*, Natalie Mrachacz-Kersting, *Member, IEEE*, and Cuntai Guan, *Fellow, IEEE*

Abstract—The performance of a classifier in a brain–computer interface (BCI) system is highly dependent on the quality and quantity of training data. Typically, the training data are collected in a laboratory where the users perform tasks in a controlled environment. However, users' attention may be diverted in real-life BCI applications and this may decrease the performance of the classifier. To improve the robustness of the classifier, additional data can be acquired in such conditions, but it is not practical to record electroencephalogram (EEG) data over several long calibration sessions. A potentially time- and cost-efficient solution is artificial data generation. Hence, in this study, we proposed a framework based on the deep convolutional generative adversarial networks (DCGANs) for generating artificial EEG to augment the training set in order to improve the performance of a BCI classifier. To make a comparative investigation, we designed a motor task experiment with diverted and focused attention conditions. We used an end-to-end deep convolutional neural network for classification between movement intention and rest using the data from 14 subjects. The results from the leave-one subject-out (LOO) classification yielded baseline accuracies of 73.04% for diverted attention and 80.09% for focused attention without data augmentation. Using the proposed DCGANs-based framework for augmentation, the results yielded a significant improvement of 7.32% for diverted attention ($p < 0.01$) and 5.45% for focused attention ($p < 0.01$). In addition, we implemented the method on the data set IVa from BCI competition III to distinguish different motor imagery tasks. The proposed method increased the accuracy by 3.57% ($p < 0.02$). This study shows that using GANs for EEG augmentation can significantly improve BCI performance, especially in real-life applications, whereby users' attention may be diverted.

Index Terms—Brain–computer interface (BCI), data augmentation, deep learning, diverted attention, electroencephalogram (EEG), generative adversarial networks (GANs), intersubject transfer learning.

I. INTRODUCTION

CLASSIFICATION algorithms are designed to deal with the decoding of information from a wide range of real-world data. However, there are situations where the classifiers fail to maintain their performance due to data-related problems such as corruption and/or insufficiency of training data. An active research area that is seriously affected by these problems is the brain–computer interfacing (BCI). The BCI aims at connecting the brain to an external device like a computer or a prosthesis with applications in motor rehabilitation, cognitive training, entertainment, and so on [1]–[3]. Since the brain signal is nonstationary in the context of STFT and varies from subject to subject and session to session [4], a large amount of data is required to effectively train a robust classifier. However, it is not feasible to run long calibration sessions especially when the BCI users are patients, children, or elderly. Moreover, the collected data are not always fully utilizable because data acquisition is usually prone to technical errors such as noise and/or human muscle artifacts such as blinking. In such situations, the need for generating artificial data, to resemble and augment the real data, arises. Generative methods with an emphasis on data generation rather than distribution estimation are the potential answer to this need. In computer vision, geometric transformation is commonly used to generate artificial images [5]. These transformations mainly include scaling to account for distance variations and rotation to account for angle variations. Although these kinds of transformation are irrelevant in the context of bio-signals such as EEG, they have inspired several researchers. Sakai *et al.* [6] shifted EEG trials in time and amplified the amplitude to generate artificial EEG for augmentation. The results showed that their augmentation method improved the classification when the size of training set was as low as 20 but it made no difference for larger sizes of training set. In another work, artificial EEG trials were generated by segmentation and recombination (S&R) in time and frequency domains [7] and the results were more convincing. Other studies have used more advanced techniques such as variational auto-encoders (VAE) [8] and generative adversarial networks (GANs) [9].

These existing works have had some signs of success but also several limitations that could be mitigated by leveraging data and techniques more effectively. In this article, therefore, we introduce a novel framework based on GANs for artificial EEG generation powered by a subject-specific conditioning vector and modified objective and loss functions. We also compare the performance of the novel method with the existing approaches in literature, namely, S&R and VAE methods. Importantly, the assessment was conducted in a realistic BCI scenario that included both focused and diverted attention conditions. The results showed that the augmentation with our proposed framework yielded improved performance, especially under the diverted condition. We have shown that the proposed method underwent stable training and did not face mode collapse [10] and, finally, we have analyzed the temporal, spatial, and frequency characteristics of the artificial samples to demonstrate that they were realistic and diverse. Overall, the comprehensive assessment of the novel method shows that it is a promising approach to generate subject-specific artificial EEG by learning from subject-independent data to reduce calibration time in EEG-based BCI applications.

II. RELATED WORK

In this section, we review the important GANs-based studies.

A. Generative Adversarial Networks (GANs)

GANs comprise two competing networks including the generator and the discriminator whose competition eventually leads to the generation of the artificial data of high quality. After its emergence, the method of GANs showed promising results in image generation as well as the potential for further improvement and applications in different domains.

Several researchers have contributed to solving the issues related to the first version of GANs. Mirza and Osindero introduced conditional GANs [11], where the idea was to feed conditioning data into both networks. They conditioned the generator and the discriminator on class labels and reported that conditional GANs trained on the MNIST data set generated images of superior quality than the regular GANs. Later, Salimon and colleagues [12] proposed techniques to enhance GANs training by investigating a range of training procedures and architectures. They put forward that the inception score (IS) is a proper metric for model comparison.

In other studies, a new version of GANs was presented [10], [13]–[15]. A very first modification to the original GANs was deep convolutional GANs (DCGANs) that used deep convolutional neural networks (CNN) for both the generator and the discriminator for a better training [13]. Mao *et al.* [14] introduced the method of least squares generative adversarial networks (LSGANs) which replaced the sigmoid loss function in the discriminator with a least square loss function. Based on their experiments, LSGANs performed better than regular GANs in terms of learning stability and image quality [14]. The method of Wasserstein GANs proposed by Arjovsky and colleagues is another version that is gaining more attention recently [10], [15]. The authors' concern was the problem

of vanishing gradients caused by minimizing the Jensen–Shannon divergence between the real and the generated data distributions in the original GANs. They showed that Wasserstein distance is a better choice than Jensen–Shannon divergence [10], [15]. Other attempts to improve the networks architecture, training stability, and image quality are presented in [16]–[19].

B. GANs for EEG

Although the GANs method was originally introduced for image generation, it can be extended to other types of data. For example, in [20], GANs have been implemented for synthesizing audio. In this article, we use GANs for electroencephalogram (EEG) signal which is the electrical activity of the brain recorded using electrodes placed on the surface of the scalp. Due to its noninvasiveness, EEG is a popular recording modality in BCI [21].

There are a few studies on the use of GANs for EEG [22]–[27]. In [22], GANs were conditioned on EEG features in order to improve image generation. The EEG data were first recorded, while the person was looking at target images. Then, EEG features were extracted using a recurrent neural network (RNN)-based encoder and fed as conditioning data to the generator and the discriminator networks [22]. Their main objective was image generation and EEG was used only as an auxiliary input. In another study, Wasserstein GANs were used to augment EEG differential entropy (DE) features in order to boost the classification of emotion [24]. The networks were conditioned on class labels and DE features were imported to the discriminator as real data. They applied the support vector machine (SVM) as the classifier, and the results showed that the classification benefited from the inclusion of the generated DE features in the training set [24]. Hartmann *et al.* [27] used GANs to generate single-channel EEG. They showed that GANs have potential applications in EEG generation and suggested that multichannel EEG generation should be explored as future work. Although only a handful of studies have applied GANs on EEG so far, their results suggest that GANs-based methods are a promising approach to deal with the issues associated with EEG.

III. MATERIALS AND METHODS

In this study, we use conditional GANs for artificial generation of raw EEG in order to augment the classification. The task is the detection of subject's movement intention (MI). The traditional methods that were used for MI detection are readiness potentials [28], [29] and common spatial patterns [30]–[32]. These methods are suitable for subject-specific classification rather than subject-independent classification, which is in the focus of this article. To evaluate the proposed approach in a more realistic scenario, we designed an experiment in which subjects performed a motor task under two conditions: 1) focused attention condition and 2) diverted attention condition. The first condition is commonly used in BCI studies; however, the performance of BCI under the second condition is not studied that well. This is a more realistic scenario, as the subject is likely

to be exposed to many distractions during a real-life BCI, and this might affect the EEG patterns [33]. The general hypothesis is that with attention diversion, the classification performance would decrease [34]. We first test this hypothesis and then use the proposed method for augmentation to improve the classification. For comparative analysis, we implement a few existing methods for artificial EEG generation. Moreover, we implement the proposed method on a public data set to evaluate its generalizability.

A. Participants

Fourteen individuals with the age range between 21 and 29 years old (24.71 $\pm$ 2.49, six males and eight females) participated in the experiment. All participants were healthy, right-handed, and without any hearing or vision abnormalities. The experiment was approved by the local ethical committee and the participants signed an informed consent form. They all could fluently speak English and understand the instructions. The experiment was performed at the Brain–Computer Interface Laboratory of the Department of Health, Science, and Technology, Aalborg University, Denmark.

B. Protocol and Data

Many BCI experiments rely on a controlled condition in which the subject is instructed to fully focus on the main task. This is different from what normally happens in real-life situations where various internal and external factors can make it difficult to stay focused on the task. By bringing the BCI system from the laboratory to home (or clinics), these factors can impair the classifier performance and BCI system efficacy. In this experiment, we considered both scenarios by including conditions with focused and diverted attention.

The main task was right hand movement (opening and closing). In both conditions, the subjects were seated in a comfortable chair, placed 1 m away from the screen, with their right hand on the desk, as can be seen in Fig. 1(a). In each condition, subjects performed 40 trials; ten hand openings followed by ten closings, 15 s of rest, and then ten openings and ten closing trials [Fig. 1(b)]. Every trial consisted of five phases including focus, preparation, execution, hold, and rest. Fig. 1(c) shows the experiment flow. This was a cue-based, nonrandomized paradigm in which the subjects were told when and what type of movement to execute.

In the focus phase, the subjects were instructed to focus on the screen and avoid blinking or moving. In the preparation phase, the type of movement was indicated to the subjects. The cue prompted the subjects to execute the movement. After the movement execution, they maintained the movement (hand opened/closed) during the hold phase. In the rest phase, the subjects relaxed.

In the focused condition, there was no distraction, while during the diverted condition, a random sequence of beeps was played. The beeps were of different frequencies with a duration of 0.5 s and a random interstimulus interval of 1–2 s. In this way, the subjects' attention was diverted by the external noise. Importantly, we wanted to assure that the participants could not simply ignore the auditory stimulus while fully focusing

(a)

(b)
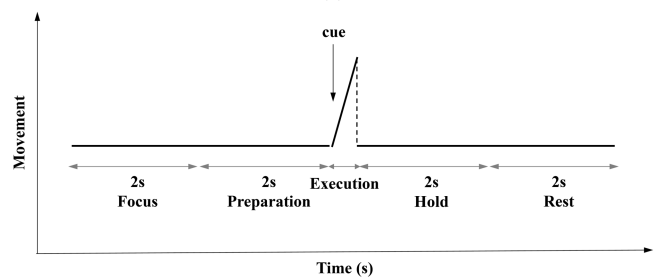
(c)
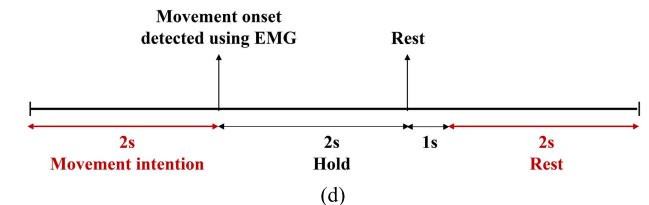
(d)

Fig. 1. (a) Demonstration of the experimental settings including EEG and EMG electrodes, (b) protocol, (c) experiment flow, and (d) segmentation diagram: movement intention and rest are the two classes for binary classification.

on the task (as in focused condition). Therefore, we asked them to count the number of times each tone (beep of a certain frequency) was repeated over ten consecutive trials. After every block of ten trials, the subjects were asked to

report how many tones of each frequency they have heard [see Fig. 1(b)]. The task difficulty was gradually increased over blocks by starting from only 2 tones (500 and 1000 Hz) played over block 1, 3 tones (500, 750, and 1000 Hz) over block 2, and 4 tones (250, 500, 750, and 1000 Hz) over blocks 3 and 4. Based on the feedback we received from the participants, their attention was indeed diverted with this oddball paradigm.

The EEG signals were recorded using a g.HIamp-research amplifier and 62 gel-based active electrodes placed in a g.Gamma cap. The recorded EEG channels were Fp1, Fp2, Fpz, AF3, AF4, AF7, AF8, F1-8, Fz, FC1-6, FCz, FT7, FT8, C1-6, Cz, T7, T8, CP1-6, CPz, TP7-10, P1-8, Pz, PO3, PO4, PO7, PO8, POz, O1, O2, and Oz, and they were referenced to the right earlobe. AFz channel was used as the ground. Moreover, 2 bi-polar electromyography (EMG) channels (four EMG electrodes) were used to detect the movement onset. They were placed on the hand flexor and extensor muscles that were located using palpation [see Fig. 1(a)]. Before placing the electrodes, the skin was cleaned using an alcohol swab. The data were recorded at 1200 Hz by g.Recorder (gtec bio-signal recording software).

C. Data Preparation

Raw EEG can be contaminated by several artifacts including eye movements and muscle activity [35]. Hence, the data were bandpass filtered at 0.01–100 Hz by g.Recorder during acquisition to remove artifacts and a notch filter was applied to remove the line frequency of 50 Hz. Moreover, independent component analysis (ICA) and artifact subspace reconstruction (ASR) were applied to remove electrooculogram (EOG) and EMG artifacts, and the data were high-pass filtered with the cutoff at 0.5 Hz. These methods were implemented in MATLAB R2013b using EEGLab [36].

After pre-processing, the data were segmented. First, the exact movement onset was determined from the EMG signals by thresholding based on the instantaneous power of EMG. The power was computed by high-pass filtering of the recorded signals (2nd order Butterworth, cutoff at 80 Hz), squaring and applying a moving average filter (25 samples). The movement executions were clearly visible as the bursts in the power signal. For each participant, the EMG threshold was set to the minimum power burst during movement execution. In this way, the intentional movements that have a power above this threshold were determined and the unintentional hand tremors were filtered out. The threshold was set based on the first half of trials (training set) and applied to the second half of trials (test set). Then, the EEG signals were segmented into MI and rest samples. The MI samples comprised 2 s of the EEG before the movement onset. The rest samples comprised 2 s of the EEG starting 1 s after the rest onset. As described in section III.B, the subjects were asked to hold their hand open (or close) during the hold phase and then release it when the rest cue was shown. Therefore, we assumed that the rest EEG segments started from 1 s after the rest onset to make sure that the subjects have indeed finished the movement. Fig. 1(d) shows the segmentation and marks the time intervals under analysis. Moreover, we checked the recorded EMG signals to

assure that there was no visible muscle activity in the time periods corresponding to both classes.

After the segmentation, the data were down-sampled to 250 Hz. In total, each subject performed 40 trials in each condition. Therefore, there were 40 samples per class per condition for each subject. The EEG data for each subject in each condition was thus a matrix of the size $80 \times 500 \times 62$, i.e., 80 samples (40 MI and 40 rest), 500 time points per sample, and 62 EEG channels.

D. Nonaugmented Classification

We implemented two methods for nonaugmented classification to compare the results with the proposed augmented classification method.

1) Slow Cortical Potentials-LDA: We implemented the method used in [28] where they employed slow cortical potential (SCP) to discriminate MI from rest. In this method, SCP is computed by filtering the EEG within the band of 0.1–1 Hz from 6 pre-defined channels, namely, C1, C2, Cz, CP1, CP2, and CPz. The EEG is divided into 500-ms segments; the MI segment contains $[-750 \ -250]$ ms of EEG before the movement onset and the rest segment contains 500 ms of EEG starting 1 s after the rest onset. By down-sampling to 8 Hz, 4 features per channel from 500-ms segments are extracted (8 Hz $\times$ 0.5 s). Thus, 24 features are extracted from each sample (6 channels $\times$ 4 features). In total, each subject has 80 samples in each condition. The first 40 samples forming a 40×24 feature matrix are used for training the classifier and the last 40 samples are used for testing. A linear discriminant analysis (LDA) is used for subject-specific classification as in [28].

2) End-to-End DCNN: We used the modified end-to-end DCNN as proposed in [37]. Two intersubject classification strategies were implemented: leave-one subject-out (LOO) and subject adaptation, hereafter called adaptive. In the adaptive method, the LOO models were updated based on the first half of the target subject's data and were tested on the second half.

Performing LOO with a total of 14 subjects formed a training set of 1040 samples (13×80) and a test set of 80 samples. In the adaptive method, the first half of the target subject's data was used to update the LOO model. Hence, size of the training and test sets changed to 1080 and 40, respectively.

The end-to-end DCNN takes the EEG segments as input and passes them through three convolution layers. The first layer is followed by a max-pooling of size 2. In these layers, convolution is done with a 1-D kernel over time [37]–[39] with 60, 40, and 20 filters, kernel sizes of 10, 8, and 6 with stride sizes of 2, 1, and 1, respectively. After the last convolution layer, the features are flattened and then sent to a fully connected layer of 100 nodes. The output is then sent to the last fully connected layer with a sigmoid activation function for the classification. In all other layers, ReLU is used as the activation function. Two dropout layers with the probability of 0.2 and 0.3 are inserted respectively before and after the first fully connected layer to avoid over-fitting. We added batch-normalization layers after the convolution

and max-pooling layers. We used Adam method [40] with a learning rate of 0.001 and beta1 of 0.9 for optimization and a batch size of 20 for training.

E. Baseline Methods for Artificial EEG Generation

In addition to the proposed method, we implemented other generative methods for augmentation to serve as the benchmark.

1) Variational Auto-Encoders: We used VAE inspired in [8] where VAE was employed for artificial steady-state visual evoked potential (SSVEP) generation. The encoder consisted of a 1-D-convolution layer, batch normalization, and a max-pooling layer, and the decoder consisted of three 1-D-transpose convolution layers. Leaky ReLU with the slope of 0.2 was used as the activation function except in the last layer of the decoder where the hyperbolic tangent was used. The method of Adam with a learning rate of 0.0001, beta 1 of 0.1, and beta 2 of 0.999 was used for optimization [8].

2) Segmentation and Recombination (S&R): We also implemented EEG generation using S&R methods presented in [7] to minimize the calibration time. The authors developed three methods: 1) S&R in the time domain; 2) S&R in the time-frequency domain; and 3) generation based on signals analogy. They reported that S&R methods significantly outperformed the analogy method. Thus, we implemented these two S&R methods for benchmarking.

In the time domain S&R, EEG trials were divided into several segments and the artificial samples were generated by concatenating random segments from randomized trials of the same class. We divided the 2-s EEG trials into eight segments, as suggested in [7]. In the time-frequency domain S&R, short-time Fourier transform (STFT) of each trial was first computed and then the STFT windows from randomized trials of the same class were concatenated to form the STFT of the artificial trial. Finally, the artificial EEG trial was computed by taking the inverse STFT. For STFT calculation, we used a 250-ms Hamming window with a 50% shift, as it was suggested in [7].

F. Proposed Method for Artificial EEG Generation

In this article, we propose a novel approach based on the conditional DCGANs to generate subject-specific artificial EEG by learning from subject-independent data.

1) Conditional DCGANs: GANs include two neural networks: a generator G and a discriminator D. In an analogy, these networks can be considered as counterfeiter and police, respectively, where the counterfeiter tries to deceive the police with fake money. In GANs, G generates the artificial samples and D identifies which samples are real and which are generated. The training target for G is to eventually generate the samples that are no longer distinguishable from the real samples by the discriminator D. At this point, the generated samples closely resemble the real samples. Two opposing networks are simultaneously being trained to maximize $\log(D(x))$ and minimize $\log(1\text{-}D(G(z)))$. This adversarial training proce-

dure is formulized as a minimax problem [9]

$$\min_G \max_D V(G, D) = E_{x \sim p_x}\big[\log(D(x))\big]$$
$$+ E_{z \sim p_z}\big[\log(1 - D(G(z)))\big] \quad (1)$$

where E denotes the expectation operator, $D(x)$ is the probability of x belonging to the real data, and $G(z)$ is the generated sample produced by G from a random noise input z

$$x_g = G(z). \quad (2)$$

The cross-entropy loss is used to calculate the discriminator loss and the generator loss as formulated in the following equations, respectively:

$$\text{Loss}_D = -\log D(x_r) - \log\big(1 - D\big(x_g\big)\big) \quad (3)$$
$$\text{Loss}_G = -\log D\big(x_g\big). \quad (4)$$

In this article, we use conditional GANs. In this type, the networks are conditioned on some information, for instance, class labels [11]. Since the objective is to generate samples for a target subject, we conditioned the networks on a feature vector extracted from a subset of the target subject's data. In this way, the generated samples will not only resemble the training set (other subjects' data) but also inherit the specific characteristics of the target subject's data. Given the conditioning vector y, (1)–(4) change to (5)–(8), respectively

$$\min_G \max_D V(G, D) = E_{x \sim p_x}\big[\log(D(x|y))\big]$$
$$+ E_{z \sim p_z}\big[\log(1 - D(G(z, y)|y))\big] \quad (5)$$
$$x_g = G(z, y) \quad (6)$$
$$\text{Loss}_D = -\log D(x_r|y) - \log\big(1 - D\big(x_g|y\big)\big) \quad (7)$$
$$\text{Loss}_G = -\log D\big(x_g|y\big). \quad (8)$$

To improve the performance of GANs, we used one-sided label smoothing which replaces the 0 and 1 labels (fake and real, respectively) for the discriminator with smoothed values [12]. Thus, the discriminator loss is modified as follows:

$$\text{Loss}_D = -\alpha \log D(x_r|y) - (1 - \alpha) \log(1 - D(x_r|y))$$
$$- \log\big(1 - D\big(x_g|y\big)\big). \quad (9)$$

We followed [12] to set the smoothed values to 0.9 and 0.1 (i.e., $\alpha = 0.9$ in the above formula).

The recent successful implementations of CNN in GANs [13], [22], [25], [41] and EEG applications [37]–[39], [42]–[46] motivated us to use CNN architecture for discriminator and generator. The generator network starts with two fully connected layers followed by a batch-normalization layer. Then, the output is first reshaped to (1, 100, 62) and then up-sampled with a size of 5. The output is then passed through two convolution layers with a kernel size of 5 and 62 filters. Please note that as suggested by the original DCGAN work [13], we used convolution layers for the generator, not deconvolution layers that were used by some studies [22], [41]. Eventually, the output of the generator (artificial EEG) has the same shape as the real EEG (500 time points, 62 channels). The discriminator consisted of a convolution layer with a kernel size of 5 and 62 filters followed by a max-pooling of size 2, another convolution
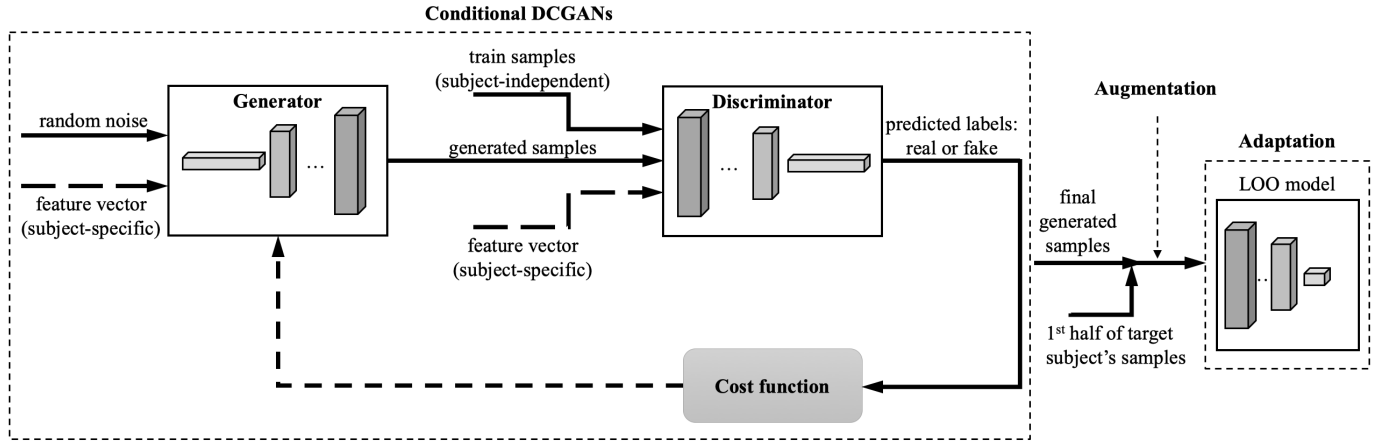
Conditional DCGANs

Fig. 2. Augmented classification with the conditional DCGANs. The generator takes a random noise and the subject-specific feature vector as inputs and generates some artificial samples. The discriminator takes these generated samples, train samples (real), and the feature vector as inputs and detects whether its input samples are real or generated. Through back-propagation, the generator learns to generate the samples that highly resemble the train samples so that the discriminator can no longer distinguish them from the real samples. After reaching training stability, the generated samples are appended to the first half of the target subject's samples and the pretrained LOO model is adapted based on this augmented set.

layer with a kernel size of 5 and 128 filters followed by a max-pooling of size 2, a flattening layer, a fully connected layer of size 400, and finally a fully connected layer of size 1. The hyperbolic tangent activation function is used in both networks except in the last layer of the discriminator where a sigmoid activation function was used for classification (real or fake). For optimization, we used Adam [40] with learning rate and beta1 parameters initialized at 0.0001 and 0.2, respectively. The hyper-parameters were selected by trial and error [47].

2) Artificial EEG Generation with Conditional DCGANs: Fig. 2 shows the conditional DCGANs framework. The GANs are trained on the other subjects' data (training samples) while conditioned on the subject-specific feature vector to generate the artificial EEG from a random noise input. Thus, the inputs to the generator are noise and the subject-specific feature vector, which are concatenated, and the inputs to the discriminator are the train samples, the subject-specific feature vector, and the generated samples. The output of the first fully connected layer in the discriminator (400 features) is concatenated with the subject-specific feature vector and the result is passed to the final fully connected layer to assign a label (real or fake) to its input samples. Noise was sampled from a normal distribution with mean 0 and standard deviation 1 following the approach in other studies [10], [15]. The training samples include data from all subjects excluding the target subject (subject-to-subject transfer learning). Thus, the training set for each class is of size (520, 500, 62) where 520 is the number of samples per class (13×40), 500 is the number of time points, and 62 is the number of EEG channels. After training, any number of the artificial EEG trials with the same shape as the real EEG can be generated by the trained generator.

3) Extracting the Subject-Specific Feature Vector for Conditioning: As mentioned above, a subject-specific feature vector is used as conditioning vector. To extract this vector, we use the pretrained LOO models as described in Section III-D2.

The first half of the target subject's samples are passed to the LOO model and the output of the first fully connected layer (of size 100) based on which the classification is done is extracted. Given 40 samples per class per subject, taking the first half of the samples produced a feature matrix of size 20×100 per class per subject. This matrix is then averaged across samples to produce the final feature vector of size 1×100 per class per subject. A separate DCGANs is trained for each condition using the features extracted from that condition (focused or diverted). Please note that the subset of samples used for feature extraction is excluded from the test set for classification.

G. Augmented Classification

The overall framework of the augmented classification with the conditional DCGANs is depicted in Fig. 2. A subject-independent DCGANs conditioned on the subject-specific feature vector are trained for the target subject. In this way, the training set is the other subjects' data, the conditioning vector is extracted from the first half of the target subject's data, and the second half which is not seen is used as the test set for classification. After reaching training stability, the generator is used to generate the artificial EEG samples with the same number of samples as in the training set ($13 \times 40 = 520$ samples per class). These generated samples are then appended to the first half of the target subject's samples and the LOO models are updated based on this augmented set. Therefore, we refer to this classification as augmented adaptive. The test set is the same as in adaptive classification. Since the augmented training set is larger, we increase the batch size to 40 for training.

Similarly, the augmented classification with VAE or S&R used the samples generated by VAE or S&R for augmentation. Fig. 3 plots the flow diagram of the different classification approaches: LOO, adaptive, and augmented adaptive.
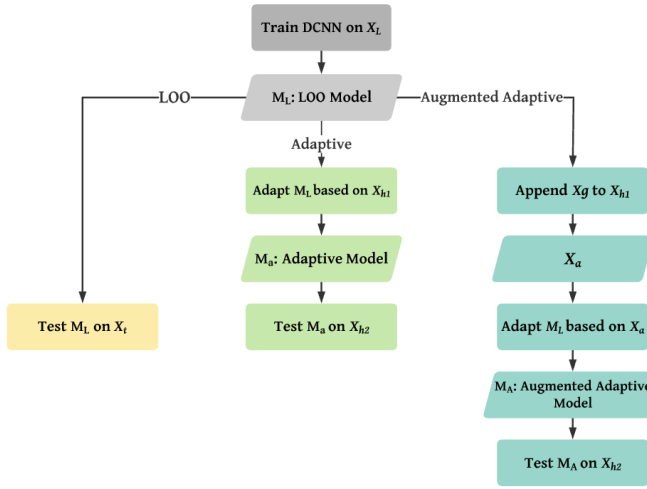
Fig. 3. Diagram of LOO, adaptive, and augmented adaptive classification approaches. In this diagram, X_L denotes EEG samples from all subjects excluding target subject (LOO), X_t denotes target subject's samples, X_{h1} and X_{h2}, respectively, denote the first and second half of X_t, and X_g denotes the generated EEG samples. The rectangles show the process and the parallelograms show the output of the process or input to the next process.

H. Evaluating Quality of the Generated Samples

It is important to ensure that the generated samples are of high quality; in other words, they are realistic and diverse. Lack of diversity among the generated samples is an indicator of mode collapse [10], [15], meaning that the generator has collapsed into generating only limited modes of the real data. In this study, we use several qualitative and quantitative measures to evaluate the quality of the samples generated by the conditional DCGANs in terms of diversity and similarity with the real samples.

1) GAN-Test: We train a classifier on the real samples (defined as X_L in Fig. 3) and test the trained model on the generated samples. The obtained classification accuracy is named GAN-test, as it is called in [48]. A high value of the GAN-test denotes that the test set, i.e., generated, is similar to the train set, i.e., real. The end-to-end DCNN is used for classification.

2) KL Divergence: We also calculate the Kullback–Leibler (KL) divergence to investigate the mode collapse. We follow the instruction given in [49] to obtain the KL divergence for EEG. There, band-passed EEG trials were represented by a Gaussian distribution with mean μ (~ 0) and covariance matrix Σ. The KL divergence between E_{μ_1, Σ_1} and E_{μ_2, Σ_2} was then formulated as

$$\text{KLD}\left(E_{\mu_1, \Sigma_1}, E_{\mu_2, \Sigma_2}\right)$$
$$= \frac{1}{2}\left[(\mu_2 - \mu_1)^T \Sigma_2^{-1}(\mu_2 - \mu_1) + tr\left(\Sigma_2^{-1}\Sigma_1\right) - \ln\left(\frac{\delta(\Sigma_1)}{\delta(\Sigma_2)}\right) - d\right]$$
(10)

where tr denotes the trace, δ denotes the determinant, and d denotes the dimensionality. We calculate the KL divergence using this equation. In successful GANs training, the KL divergence between the generated samples should be close to the KL divergence between the real samples.

3) Visualization: Furthermore, we visually inspect the quality of the artificial samples by mapping the generated and real samples into two dimensions using t-distributed stochastic neighbor embedding (t-SNE), spectrogram plots, and temporal distribution.

I. Evaluating Generalizability of the Proposed Method

We implement the proposed method on the data set IVa from Berlin BCI competition III [50] to evaluate the generalizability of the method. The data set includes EEG recorded by 118 channels from five subjects during two classes of motor imagery: right hand and foot. There were 140 trials per class per subject. Each trial had a length of 3 s with a sampling frequency of 100 Hz that was down-sampled from 1000 Hz.

IV. RESULTS

The input data preparation including the artifact removal and segmentation, SCP-LDA, and EEG generation with S&R methods was implemented in MATLAB R2013b. The end-to-end DCNN, VAE, and the conditional DCGANs were implemented in Python 3.6 with Keras 2.1.2 and Tensorflow 1.2.1. The reported *p*-values were obtained using Wilcoxon signed-rank tests. In this text, the terms "artificial samples" and "generated samples" are used interchangeably. The main results are presented in Tables I and II.

A. Nonaugmented Classification

Table I shows the results of nonaugmented classification. When performing subject-specific classification using SCP-LDA [28], as a traditional method, the accuracy in the focused condition was 75.71% which dropped significantly to 66.07% in the diverted condition (-9.64%, *p*-value = 0.01). Four subjects out of 14 had less than 70% accuracy in the focused condition. This number increased to 6 in the diverted condition.

Performing LOO with DCNN, the accuracy in the focused condition was 80.09% which dropped significantly to 73.04% in the diverted condition (-7.05%, *p*-value < 0.01). Only 1 subject out of 14 had less than 70% accuracy in the focused condition. This number increased to 4 in the diverted condition.

Similarly, with adaptive classification, the accuracy decreased from 82.32% in the focused condition to 76.43% in the diverted condition, and this decrease was statistically significant (-5.89%, *p*-value < 0.01). None of the subjects had less than 70% accuracy in the focused condition, while for 3 of 14, the accuracy was below 70% in the diverted condition.

The adaptive classification significantly outperformed LOO in the focused condition ($+2.23\%$, $p < 0.05$), while the observed difference was not statistically significant in the diverted condition (*p*-value > 0.05). These observations denote that classification in the diverted attention condition was more challenging so that the adaptation did not help.

To provide a fair comparison with the adaptive classification, we evaluated the performance of LOO on the second half of the test subject's data. The accuracy in the focused condition was 80.89 (SD: 9.07) and in the diverted condition

TABLE I

BASELINE CLASSIFICATION ACCURACIES (%)

	Condition 1: Focused		
	SCP-LDA	**DCNN-LOO**	**DCNN-Adaptive**
Average (SD)	75.71 (11.74)	80.09 (6.13)	82.32 (3.60)
Range (min-max)	35 (57.50-92.50)	25.00 (62.50-87.50)	12.50 (75.00-87.50)
No of subjects with accuracy<70%	4 of 14	1 of 14	0 of 14
	Condition 2: Diverted		
	SCP-LDA	**DCNN-LOO**	**DCNN-Adaptive**
Average (SD)	66.07 (11.59)	73.04 (7.38)	76.43 (7.83)
Range (min-max)	32.50 (45.00-77.50)	25.00 (60.00-85.00)	22.50 (65.00-87.50)
No of subjects with accuracy < 70%	6 of 14	4 of 14	3 of 14

SD – Standard Deviation

TABLE II

AUGMENTED ADAPTIVE CLASSIFICATION ACCURACIES (%)

	Condition 1: Focused			
	VAE	**S&R in time domain**	**S&R in time-frequency domain**	**Conditional DCGANs**
Average (SD)	84.10 (8.47)	83.75 (7.05)	84.64 (6.71)	**85.54 (5.02)**
Range (min-max)	32.50 (62.50-95.00)	27.50 (70.00-97.50)	22.50 (72.50-95.00)	**15.00 (77.5-92.50)**
No of subjects with accuracy <70%	1 of 14	0 of 14	0 of 14	**0 of 14**
	Condition 2: Diverted			
	VAE	**S&R in time domain**	**S&R in time-frequency domain**	**Conditional DCGANs**
Average (SD)	76.43 (8.53)	74.82 (9.12)	74.64 (8.54)	**80.36 (7.46)**
Range (min-max)	32.50 (57.50-90.00)	32.50 (55.00-87.50)	27.50 (57.50-85.00)	**20.00 (70.00-90.00)**
No of subjects with accuracy <70%	2 of 14	3 of 14	4 of 14	**0 of 14**

SD – Standard Deviation
The test set for all these methods is the same.

was 71.43 (SD: 8.47). There was no statistically significant difference between these results and the results of LOO tested on the full test data (p-value > 0.05). Hence, the results hold independently of the test set.

B. Augmented Classification

The difference between adaptive and augmented adaptive classification is in the data that were used for adapting the LOO model—in the adaptive approach, a subset of target subject's data was used for adaptation, while in the augmented adaptive approach, this subset was augmented with the generated samples. Note that the same test set was used in both methods. Thus, the baseline here is the adaptive method which is denoted as DCNN-adaptive in Table I. The results of the augmented adaptive classification using different generative methods are presented in Table II.

In the focused condition, augmented classification with the conditional DCGANs significantly outperformed the baseline with 3.22% improvement (p-value $= 0.01$), while augmen-

tation with other generative methods did not make a significant improvement over the baseline. Nonetheless, performing ANOVA showed that none of the methods including conditional DCGANs were significantly better than the other one. In fact, the results were comparable.

In the diverted condition, performing ANOVA showed that there is a statistically significant difference between the methods. Pairwise comparisons followed by Benjamini–Hochberg FDR procedure for multiple tests correction revealed that the conditional DCGANs outperformed the other methods (corrected p-value < 0.05). There was no statistically significant difference between the other methods.

The results show that both conditions benefited from the augmentation with the samples generated by the proposed method. Applying the novel method resulted in several improvements in the focused condition. The classification accuracy increased by 5.45% compared with LOO (p-value < 0.01) and 3.22% compared with adaptive (p-value $= 0.01$). In addition, the accuracy in all subjects increased to above 70%. Table III summarizes the confusion matrix. The

TABLE III

CONFUSION MATRIX

Baseline (non-augmented)				
	Condition1: Focused		Condition2: Diverted	
	MI	Rest	MI	Rest
MI	85.36%	14.64%	71.79%	28.21%
Rest	20.71%	79.29%	18.93%	81.07%

Augmentation with VAE				
	Condition1: Focused		Condition2: Diverted	
	MI	Rest	MI	Rest
MI	86.79%	13.21%	85.71%	14.29%
Rest	18.57%	81.43%	32.86%	67.14%

Augmentation with S&R in time				
	Condition1: Focused		Condition2: Diverted	
	MI	Rest	MI	Rest
MI	81.43%	18.57%	70.71%	29.29%
Rest	13.93%	86.07%	21.07%	78.93%

Augmentation with cond. DCGANs				
	Condition1: Focused		Condition2: Diverted	
	MI	Rest	MI	Rest
MI	83.21%	16.79%	76.43%	23.57%
Rest	12.14%	87.86%	15.71%	84.29%

MI – Movement Intention

Rows and columns are respectively associated with actual and predicted classes. Note that since the results of S&R methods were similar, only confusion the matrix of S&R in time domain is reported in this table.

Fig. 4. Augmented adaptive with the conditional DCGANs versus adaptive. The abbreviation "acc" stands for accuracy. The augmented adaptive increases the accuracy for most of the subjects. Note that the number of circles in the left plot is 12 and in the right plot is 13 instead of 14. This is because some subjects had the same pair of accuracies in 2 methods and thus their circles fully overlap (subject pairs of (1, 5) and (6, 13) in condition 1, and (8, 14) in condition 2).

Fig. 5. Generator and discriminator losses over training. Both losses converge after approximately 300 iterations.

false-positive rate (FPR) decreased to as low as 12.14%, and the true-positive rate (TPR) increased to as high as 83.21%. Note that in the focused condition, the results of other generative methods were comparable with the results of DCGANs. VAE achieved the highest TPR (86.79%) and conditional DCGANs achieved the highest specificity (87.86%); however, these differences were not statistically significant.

The proposed method also enhanced the classification performance in the diverted condition. There was a statistically significant improvement of 7.32% in classification accuracy compared with LOO (p-value < 0.01) and 3.93% compared with adaptive (p-value $= 0.02$). The method achieved the classification accuracy higher or equal to 70% for all subjects. The TPR was as high as 76.43% and the FPR, 15.71%, was the lowest among all the discussed methods. Note that although VAE achieved the highest TPR (85.71%) in the diverted condition, this method had the lowest specificity (67.14%).

Overall, the results showed that augmentation with conditional DCGANs improved the classification. It is interesting to see how augmentation has changed the accuracy of each subject compared with the baseline. Fig. 4 shows the scatter plot for this comparison. Each circle represents one subject. The vertical axis is the classification accuracy of the augmented adaptive with the conditional DCGANs and the horizontal axis is the accuracy of the baseline (DCNN-adaptive). In the focused condition, the accuracy for eight subjects increased and for four subjects did not change. In the diverted condition, the accuracy for 12 subjects out of 14 increased. Overall,
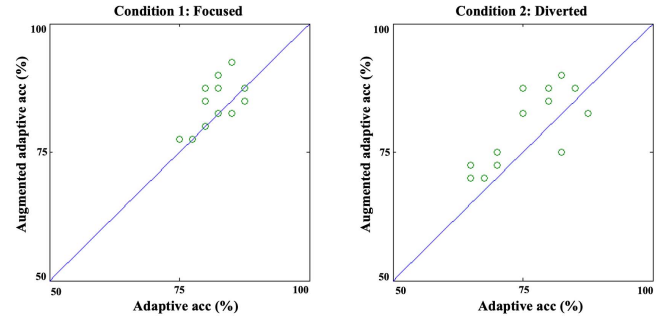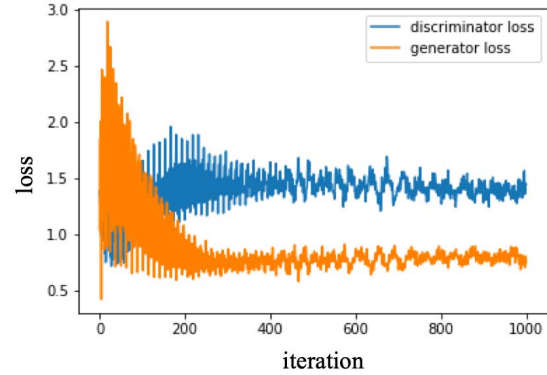
in both conditions, augmented adaptive achieved a better performance than adaptive for the majority of the subjects.

C. Adversarial Networks' Training

In training GANs, achieving the training stability is important. The networks' loss over the iterations is a good indicator of how the training proceeds. Fig. 5 plots the generator and the discriminator loss for one randomly selected subject over 1000 iterations. In successful training, it is expected to see a gradual drop in the generator loss and convergence to constant values for both networks. These criteria can be seen in Fig. 5; the generator loss gradually decreases and both losses converge to constant values after approximately 300 iterations. The same trend exists in other subjects' results. A careful choice for the type and order of the layers, input preparation, single-side label smoothing for the discriminator loss, and so on yielded a stable training.

D. Quality of the Generated EEG

Here, the results of the quantitative and qualitative evaluation measures defined in Section III-H to assess the validity of the artificially generated samples are presented.

TABLE IV
QUANTITATIVE MEASURES FOR QUALITY EVALUATION

		Condition 1: Focused	Condition 2: Diverted
GAN-test		99.16%	97.87%
KL divergence	*real*	2.01	2.25
	GAN	2.53	2.12

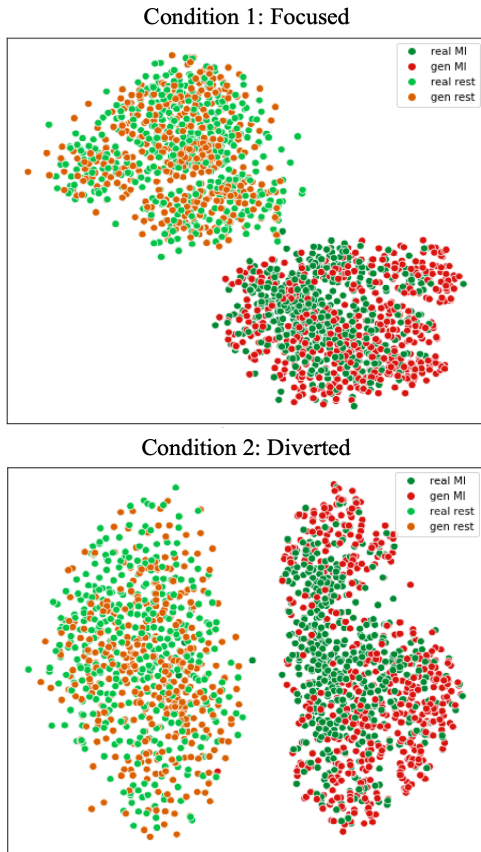GAN in the last row refers to the samples generated by conditional DCGANs.

Fig. 6. T-SNE embedding of real and generated samples. Abbreviation "gen" in the legend stands for "generated."

1) GAN-Test: The results are reported in Table IV. The GAN-test was 99.16% in the focused condition and 97.87% in the diverted condition, and this indicates that the generated samples are similar to the real samples.

2) KL divergence: The results are presented in Table IV. In both conditions, the KL divergence between the generated samples is fairly close to the KL divergence between the real samples, indicating thereby that the generated samples are as diverse as the real samples.

3) Visualization: T-SNE is applied to map the high dimensional real (training) and generated EEG samples into 2-D space. Fig. 6 shows the results. It can be seen that t-SNE embedding of real MI and generated MI is similar. In addition, real and generated rest samples have similar distributions.

Besides comparing the generated samples with the training samples, it is interesting to compare them with the test
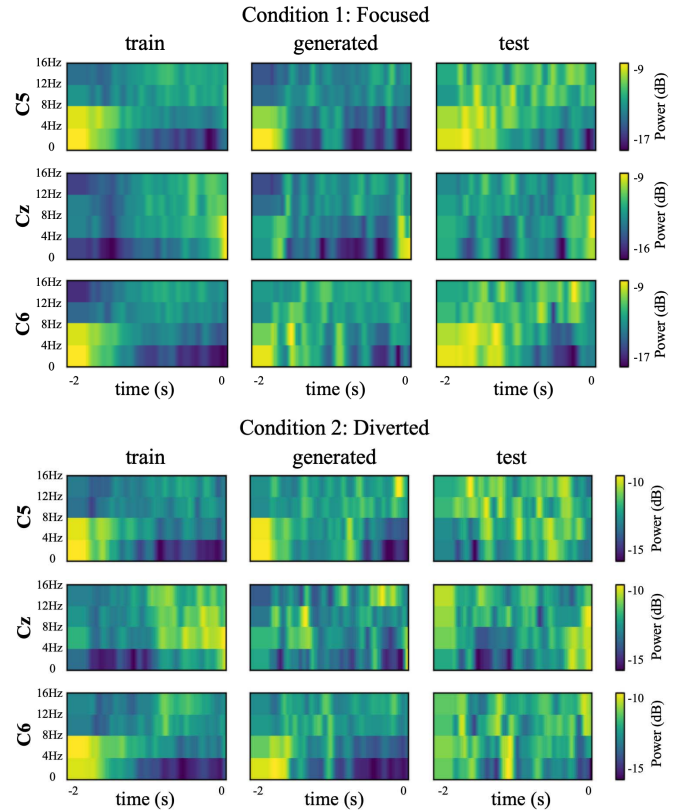

Fig. 7. Average spectrogram of the train, test, and generated samples in MI class. Time 0 shows the movement onset.

samples. In fact, the similarity between the generated and the test samples will explain the classification improvement made by the augmentation. For this purpose, the average spectrogram of the training, generated, and test samples in channels C5, Cz, and C6 for a random subject is depicted in Fig. 7. The training set includes the samples from all subjects excluding the target subject, the generated set includes the generated samples for the target subject, and the test set includes the second half of the target subject's samples that were not seen during training. Thus, no overlap between the training, test, and generated sets exists.

There are many similarities between the spectrograms of the training and generated samples in both conditions. This is in fact in line with the result of t-SNE in Fig. 6. For example, in C5 and C6, there is a higher power across low frequencies around 2 s before the movement (time -2) in the spectrograms of both training and generated data. There are also some similarities between the spectrograms of the test and generated samples. Nevertheless, these are the results of one subject and the spectrograms may vary across subjects. Note that since the training set for different subjects varies only in 40 samples out of 520, this will not substantially affect the average spectrogram and thus the spectrogram of training samples for other subjects are similar to that shown in Fig. 7.

An interesting point is that unlike the focused condition where the spectrograms of test and training data are similar, in the diverted condition, they are different. This suggests
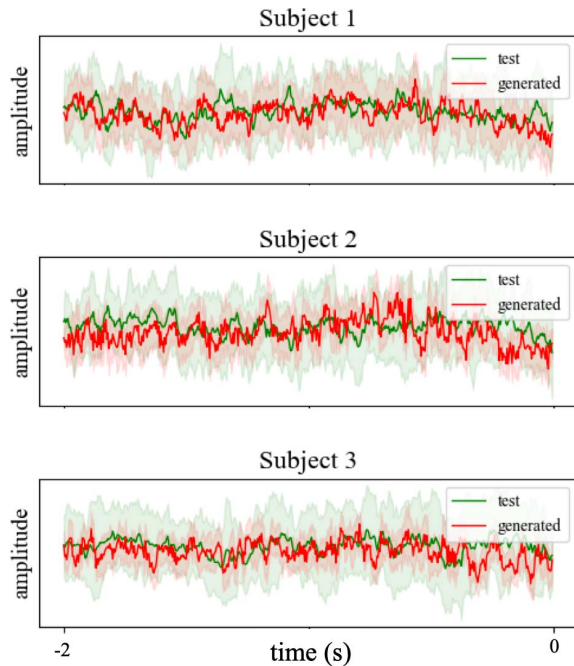
Fig. 8. Temporal distribution of the test samples (real) and samples generated by the conditional DCGANs. The plots are for the diverted condition, channel Cz, and class MI. Solid lines show the mean and faded colors show the standard deviation from the mean.

that the intersubject nonstationarity in the diverted condition is higher than in the focused condition. The generated spectrogram in the diverted condition inherits some features from the test. For example, in C5, it shows relatively high power above 8 Hz, which is more similar to the test than the training. To further inspect the generated against the test in this condition, the temporal distribution of the test and generated samples in channel Cz is plotted in Fig. 8.

E. Generalizability of the Proposed Method

Augmentation of data set IVa from Berlin BCI competition III with the proposed method significantly improved the classification between right hand and foot from 67.57% at baseline (adaptive) to 71.14% (p-value = 0.01). Note that since the classifier was not optimized specifically for this data set, the obtained accuracy may not be high. Nevertheless, the augmentation with the proposed method improves the classification. This suggests that the proposed method for EEG generation is not data-dependent.

V. DISCUSSION

In this article, we have developed a framework based on DCGANs to generate artificial EEG data from the recorded examples. The generated data can be used to supplement the training set for a BCI classifier and, indeed, we have demonstrated that the enhanced training significantly improved the performance of detection in challenging conditions. To the best of our knowledge, this study is among pioneers presenting a method to generate artificial EEG data for BCI applications. Previously, Lotte [7] explored several generative

methods based on S&R in time and time-frequency domain and employed them to generate artificial EEG to reduce the time for the calibration of a BCI system. They reported that the S&R methods performed better than the generative method based on analogy. Here, we showed that the conditional DCGANs outperformed the S&R methods, especially in the diverted condition. Furthermore, we implemented VAE, which is known as a powerful generative method. The results showed that the proposed conditional DCGANs significantly outperformed VAE in the diverted condition.

The artificial samples generated by conditional DCGANs improved the detection performance because they were generated specifically to contribute new information about unseen samples into the train set. To this aim, the GANs were conditioned on a feature vector learned through DCNN from a subset of samples. We used several quantitative and qualitative measures including GAN-test, KL divergence, 2D visualization using t-SNE, spectrogram, and temporal distribution to evaluate the quality of the generated samples. The results verified that the generated samples were indeed realistic and diverse.

A common problem associated with GANs is the training instability [17], [18]. However, the proposed GANs did not suffer from this problem and both networks gradually converged. Each training iteration on a CPU [Intel(R) Xeon(R) CPU E5-1650 v2 at 3.5 GHz] took approximately 0.58 s, and the generation of 520 samples took about 1.24 s. These numbers slightly varied in each run. Once the networks are trained, DCGANs can be used to easily generate hundreds of samples within a few seconds. Importantly, collecting the same amount of data experimentally would take over an hour. In addition, the experimental data collection would require additional work and cognitive burden on the subject. Another challenge that GANs usually face is the low quality of the generated samples [18]. In this work, by conditioning GANs on subject-specific features, the quality of the generated samples improved.

This article should be considered in the context of some limitations. The proposed framework can be further improved by applying automatic hyper-parameter selection methods, such as the one presented in [51]. Moreover, simultaneous optimization of DCNN and conditional DCGANs could improve the integrity of the framework. It is also worth investigating to see how the inclusion of the generated samples changes the features that the classifier learns and whether these changes are meaningful. Moreover, a deeper analysis could provide a better understanding of why augmentation with GANs improved the classification in the diverted condition.

VI. CONCLUSION

This article demonstrated that the proposed GANs-based approach was able to generate realistic EEG samples for a specific subject by learning from a pool of other subjects (intersubject transfer learning). Augmentation of the training set with these artificially generated samples significantly benefitted the classification tasks, which are known to be challenging in BCI systems (e.g., classification with distractions).

The proposed framework can be further extended to other applications such as restoration of the corrupted EEG and to online BCI designs.

ACKNOWLEDGMENT

The authors would like to thank Aalborg University, Denmark and the Institute for Infocomm Research, A-STAR, Singapore for supporting the project. They would also like to thank the cooperation of all participants and the members of the BCI Lab at Aalborg University.

REFERENCES

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

[2] J. J. Daly and J. R. Wolpaw, "Brain-computer interfaces in neurological rehabilitation," *Lancet Neurol.*, vol. 7, no. 11, pp. 1032–1043, 2008.

[3] J. R. Wolpaw *et al.*, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.

[4] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.*, vol. 3, no. 1, pp. R13–R23, Mar. 2006.

[5] E. Park, J. Yang, E. Yumer, D. Ceylan, and A. C. Berg, "Transformation-grounded image generation network for novel 3D view synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3500–3509.

[6] A. Sakai, Y. Minoda, and K. Morikawa, "Data augmentation methods for machine-learning-based classification of bio-signals," in *Proc. 10th Biomed. Eng. Int. Conf. (BMEiCON)*, Aug. 2017, pp. 1–4.

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

[8] N. K. N. Aznan, A. Atapour-Abarghouei, S. Bonner, J. D. Connolly, N. Al Moubayed, and T. P. Breckon, "Simulating brain signals: Creating synthetic EEG data via neural-based generative models for improved SSVEP classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[9] I. Goodfellow *et al.*, "Generative adversarial nets," presented at the Adv. Neural Inf. Process. Syst., 2014.

[10] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein GANs," presented at the Proc. 31st Int. Conf. Neural Inf. Process. Syst., Long Beach, CA, USA, 2017.

[11] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784

[12] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," presented at the Proc. 30th Int. Conf. Neural Inf. Process. Syst., Barcelona, Spain, 2016.

[13] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[14] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821.

[15] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875

[16] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," presented at the Adv. Neural Inf. Process. Syst., 2017.

[17] N. Kodali, J. Abernethy, J. Hays, and Z. Kira, "On convergence and stability of GANs," 2017, *arXiv:1705.07215*. [Online]. Available: http://arxiv.org/abs/1705.07215

[18] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: http://arxiv.org/abs/1710.10196

[19] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," 2017, *arXiv:1711.04340*. [Online]. Available: http://arxiv.org/abs/1711.04340

[20] C. Donahue, J. McAuley, and M. Puckette, "Synthesizing audio with GANs," presented at the 6th Int. Conf. Learn. Represent., Vancouver, BC, Canada, 2018.

[21] S. Machado *et al.*, "EEG-based brain-computer interfaces: An overview of basic concepts and clinical applications in neurorehabilitation," *Rev. Neurosci.*, vol. 21, no. 6, pp. 451–468, Jan. 2010.

[22] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, and M. Shah, "Generative adversarial networks conditioned by brain signals," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3410–3418.

[23] I. A. Corley and Y. Huang, "Deep EEG super-resolution: Upsampling EEG spatial resolution with generative adversarial networks," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 100–103.

[24] Y. Luo and B.-L. Lu, "EEG data augmentation for emotion recognition using a conditional wasserstein GAN," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2535–2538.

[25] F. Fahimi, Z. Zhang, W. B. Goh, K. K. Ang, and C. Guan, "Towards EEG generation using GANs for BCI applications," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, May 2019, pp. 1–4.

[26] Y. Lee and Y. Huang, "Generating target/non-target images of an RSVP experiment from brain signals in by conditional generative adversarial network," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 182–185.

[27] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals," 2018, *arXiv:1806.01875*. [Online]. Available: http://arxiv.org/abs/1806.01875

[28] E. Lew, R. Chavarriaga, S. Silvoni, and J. D. R. Millán, "Detection of self-paced reaching movement intention from EEG signals," *Frontiers Neuroeng.*, vol. 5, p. 13, Jul. 2012.

[29] M. Jahanshahi and M. Hallett, *The Bereitschaftspotential: Movement-Related Cortical Potentials*. Boston, MA, USA: Springer, 2003.

[30] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.

[31] W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown, and S. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 639–653, Mar. 2015.

[32] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Dec. 2008.

[33] S. Aliakbaryhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting, "Online Adaptive Synchronous BCI System with Attention Variations," in *Brain-Computer Interface Research: A State-of-the-Art Summary*, C. Guger, N. Mrachacz-Kersting, B. Z. Allison, Eds. Cham, Switzerland: Springer, 2019, pp. 31–41.

[34] S. Brandl, L. Frølich, J. Höhne, K.-R. Müller, and W. Samek, "Brain–computer interfacing under distraction: An evaluation study," *J. Neural Eng.*, vol. 13, no. 5, Oct. 2016, Art. no. 056012.

[35] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Technol.*, vol. 2, no. 2, pp. 1–11, 2002.

[36] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[37] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026007.

[38] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.

[39] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

[40] P. D. Kingma and J. Ba, "A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[41] I. Kavasidis, S. Palazzo, C. Spampinato, D. Giordano, and M. Shah, "Brain2Image: Converting brain signals into images," in *Proc. ACM Multimedia Conf. MM*, 2017, pp. 1809–1817.

[42] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[43] F. Fahimi, Z. Zhang, T.-S. Lee, and C. Guan, "Deep convolutional neural network for the detection of attentive mental state in elderly," presented at the 7th Int. BCI Meeting, Sacramento, CA, USA, 2018.

[44] S. L. Oh *et al.*, "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 10927–10933, Aug. 2018.

[45] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.

[46] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, "Deep learning based on batch normalization for P300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, Jan. 2018.

[47] J. Bergstra *et al.*, "Algorithms for hyper-parameter optimization," presented at the Proc. 24th Int. Conf. Neural Inf. Process. Syst., Granada, Spain, 2011.

[48] K. Shmelkov, C. Schmid, and K. Alahari, "How good in my GAN," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 213–229.

[49] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "EEG data space adaptation to reduce intersession nonstationarity in brain-computer interface," *Neural Comput*, vol. 25, no. 8, pp. 2146–2171, Aug. 2013.

[50] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[51] F. Ye, "Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data," *PLoS ONE*, vol. 12, no. 12, Dec. 2017, Art. no. e0188746.

Fatemeh Fahimi received the Ph.D. degree in computer science from the School of Computer Science and Engineering, Nanyang Technological University, Singapore, in February 2020.

She is currently a Scientist with the Institute for Infocomm Research, Agency for Science, Technology, and Research (A-STAR), Singapore. Her research interests include machine learning, brain–computer interfaces, artificial intelligence for healthcare, and digital health.

Dr. Fahimi received the BCI Student Award from the BCI Society, Research Presentation Award from Nanyang Technological University, and the Student Paper Award from Pattern Recognition and Machine Intelligence Association, Singapore.

Strahinja Dosen (Member, IEEE) received the Ph.D. degree in biomedical engineering from Aalborg University, Aalborg, Denmark, in 2009.

He has held research scientist position with the Medical Center Göttingen, Georg-August University, Göttingen, Germany. He is currently an Associate Professor with the Department of Health Science and Technology, Aalborg University. He has authored or coauthored more than 60 articles in peer-reviewed journals in the field of biomedical engineering. His research interests include closed-loop human–machine interfacing, movement restoration and control, rehabilitation robotics, and artificial sensory feedback.

Kai Keng Ang (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in 2008.

He is currently the Leader of the Signal Processing Group, SRO Department, and a Senior Scientist with the Institute for Infocomm Research, A-STAR, Singapore. He is also an Adjunct Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include brain–computer interfaces, signal processing, computational intelligence, machine learning and pattern recognition.

Dr. Ang received the BCI Annual Research Award, the IES Prestigious Engineering Achievement Award, and the BCI competitions Award. He also served as a Co-Chair and Session Chairs for international conferences, and a Reviewer for various journals and conferences.

Natalie Mrachacz-Kersting (Member, IEEE) received the Ph.D. degree in biomedical engineering from Aalborg University, Aalborg, Denmark, in 2005.

She has held positions with the Department of Health Science and Technology, Aalborg University, and also the University of Auckland, Auckland, New Zealand. She is currently a Professor with Neuroscience and Medical Technologies, Dortmund University of Applied Sciences and Arts, Dortmund. She has authored over 70 articles in peer-reviewed journals, over 120 conference articles/abstracts, and ten book chapters. Her research interests include brain–computer interfaces, neurorehabilitation for ALS and stroke patients, neurophysiology, and medical technology.

Dr. Mrachacz-Kersting received several awards including the international BCI award in 2017.

Cuntai Guan (Fellow, IEEE) received the Ph.D. degree in electrical and electronic engineering from Southeast University, Nanjing, China, in 1993.

He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include brain–computer interfaces, neural signal processing, neural image processing, machine learning, and artificial intelligence.

Dr. Guan received the Annual BCI Research Award, the IES Prestigious Engineering Achievement Award, the Achiever of the Year (Research) Award, and the Finalist of President Technology Award. He is an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, *Pattern Recognition*, *Neurocomputing*, *Frontiers in Human Neuroscience*, and *Brain-Computer Interfaces*.