IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 5 September 2023, accepted 13 October 2023, date of publication 23 October 2023, date of current version 31 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3326720

RESEARCH ARTICLE

A Novel Small-Data Based Approach for Decoding Yes/No-Decisions of Locked-In Patients Using Generative Adversarial Networks

PASCAL PENAVA, (Member, IEEE), AND RICARDO BUETTNER, (Senior Member, IEEE)
Chair of Information Systems and Data Science (BWL XVIII), University of Bayreuth, 95447 Bayreuth, Germany

Corresponding author: Pascal Penava (Pascal.Penava@uni-bayreuth.de)

This work was supported in part by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant 491183248; and in part by the Open Access Publishing Fund of the University of Bayreuth.

ABSTRACT We demonstrate how to use generative adversarial networks to improve the small data problem when training brain-computer-interfaces. The new approach is based on finely graded frequency bands, which are extracted from motor imagery electroencephalography data by using power spectral density method to synthetically generate electroencephalography data using generative adversarial networks. We evaluate our approach using one of the currently largest publicly available electroencephalography datasets, by first checking the synthetic and real data for statistical and visual similarity, and secondly, by training a random forest classifier, once using only the real data and then using the real data augmented with the synthetic data. With similarity scores of 95.72 % in the subject-dependent case and 83.51 % in the subject-independent case, and a predictive gain of 17.53 % in the subject-dependent case, and 7.51 % in the subject-independent case, we were able to achieve promising results. The results show that our approach can make it possible to research rare diseases for which there is too little patient data. Also, synthetic data can be a way for many electroencephalography-based brain-computer interface applications to obtain the required data more cost- and time-efficiently.

INDEX TERMS Brain-computer-interface, decision prediction, generative adversarial networks, motor imagery tasks, electroencephalography, machine learning.

I. INTRODUCTION

Severe motor disabilities strongly restrict the affected persons in their communication [1]. Especially people suffering under total locked-in syndrome (LIS) completely lack verbal communication [2]. Total LIS is a neurological impairment that prevents patients from performing cognitive functions [3]. Patients can only move their eyes and partially move their eyelids to communicate [4]. It is precisely for this group of patients that it is necessary to enable a different form of communication, since there is no cure for the disease [5].

The progress of IT-enabled hardware was huge over the last years [6], [7], still its potential is being advanced by using modern Machine Learning (ML) build upon big data [8], [9]. Brain Computer Interfaces (BCIs), whose concept was introduced in the 1970, thus could be a promising approach for affected individuals [10], [11]. Since this millennium, EEG data have proven to be a suitable basis for BCIs [12]. In such a system, recorded brain activities are used to link the brain and a computer [13].

Motor imagery (MI) is the imagination of physical actions [14], sensorimotor rhythms (SMR) emerge by modeling MI in humans [15], [16]. Thus, SMR-based BCIs make it possible for a subject to control a device through his imagination, without external stimuli. For patients with motor disorders, SMR-based BCIs can therefore be very useful [17]. Thus, classification of SMRs in EEG enables non-invasive BCIs, which are used to provide connectivity between the brain and external devices [18].

The associate editor coordinating the review of this manuscript and approving it for publication was Md Kafiul Islam.

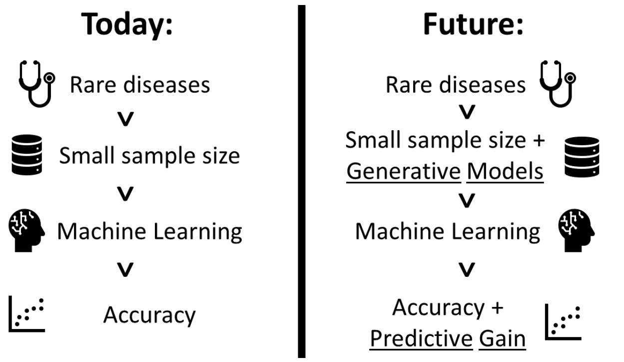

FIGURE 1. Problem.

EEG based BCIs have multiple applications, such as emotion recognition [19], motor imagery task recognition [20], [21], [22], decoding yes/no decisions [23], [24], cursor control [25], [26], computer control [25], robotic arms [27], wheelchairs [28] and many more [29]. Because of the many proven application areas, EEG data is a solid basis for BCI applications, and therefore a good alternative communication tool for individuals with severe disabilities [30].

Due to individual differences in EEG signals, most BCI systems are calibrated specifically for individual users, i.e., they are subject-dependent [21], resulting in long and strenuous training times [31] making data-acquisition a rather complex process.

Still, MI decoding via EEG data is a promising approach for practical BCI applications [32], why there is a strong need for an solution that reduces the training burdens and at the same time increases the subject-specific classification accuracy. In recent years, Generative Adversarial Networks (GANs) have emerged that can be used to generate new similar data from existing data. In this way, data acquisition can be made easier in certain areas of operation, as not so much data is required initially [33]. GANs operate through two models (networks) that compete with each other [34]. Here, one network serves to generate new data sets. The other network then attempts to distinguish the artificial data from the real data. And so, iteratively, better and better new data is created [35].

In recent research, GANs have been proven to be good for many application areas. Data generation has been proven to be a promising area [36], [37], [38], [39], also image generation [40], [41], text generation [42], [43] and many more [35].

Neurons in our brain produce voltage potential that can be measured by electroencephalograms over the time axis, producing EEG Data [44]. In research, GANs could also be successfully used to generate such time-series data [45], [46]. Further, in current research, GANs are already successfully used for EEG applications [47], [48], [49]. GANs must preserve temporal dynamics for the generation of time-series data in that the dependencies of the different variables are preserved over time. Thus, GANs bring great challenges in the area of time-series data generation, which have an impact on performance [50].

Spectral analysis can be used to take the time dependence from sequential data. The Fast Fourier Transform (FFT) can be used to perform time averaging and thus turn sequential data into tabular data [51]. Reference [52] introduced a novel approach using spectral analysis for a finer graded analysis of EEG data. Most current research classifies EEG data in the traditional frequency bands delta, theta, alpha, beta and gamma [53]. Promising results could already be achieved by using a fine graded approach [54], [55]. The finer graded spectral analysis of EEG data also enables a better research for corresponding diseases [52].

There are many rare diseases where less than 1 in 2,000 people is affected [56]. In fact there are over 7,000 rare diseases [57]. The small number of patients suffering from this kind of disease prevents adequate research into it [58]. If a GAN could generate EEG data for rare diseases, it would alleviate that problem.

Current research shows that the investigation of mental illnesses such as schizophrenia [52], [59], [60], or the investigation of certain personality traits based on EEG data [61], [62], [63], [64], [65], show similar challenges, but can also be supported by the help of modern ML approaches.

Figure 1 summarizes the problem described over the last few paragraphs and shows a potential solution for the future, which is supported by our work.

Most of the current ML approaches show long calibration and classification times [66]. Random forest (RF) classifiers could achieve high classification accuracy's with a short application time. Also, they could already show promising results by classification in non-invasive BCIs [67]. Still, RF classifiers got little attention in BCI research [68].

Therefore, we want to investigate whether GANs can be used for EEG data generation in the context of non-invasive EEG based BCIs. Thereby, we are going to use a fine graded analysis of the EEG data using spectral analysis and use a RF Classifier for classification. Also, we will use a FFT for time averaging to avoid the classical problems of GANs with sequential data. Therefore, in this paper we want to investigate the following research question: Is it possible to use GANs for EEG data generation in the context of non-invasive EEG based BCIs, more specific for binary-class (Yes/No Decision) SMR-based BCI task classification? Using a tabular data GAN in combination with fine-graded frequency bands and power spectral density (PSD), we provide a RF based classifier for EEG-based BCIs.

Our most important contributions are:

1) We have succeeded in generating synthetic data that is statistically similar to the real data. We achieved a similar ity score of 95.72 % in the subject-dependent case and 83.51 % in the subject-independent case.

2) We have managed to achieve a predictive gain of 17.53 % in the subject-dependent case, and 7.51 % in the subject independent case for an RF classifier when augmenting the real data with the synthetic data (The GANs were trained with data from the same session that we used for classifier evaluation).

3) For subject-dependent data, we have managed to achieve a predictive gain with an RF classifier of 2 % for unseen data (The GAN was trained with a different session than the one we used to evaluate the classifier).

These contributions allow us to add a few things to the current state of research. The use of PSD allows us to generate EEG data with a fine spectrum that can eventually be used to study rare diseases where there is not enough patient data. By synthetically generating the data and adding entropy, especially in the subject-dependent case, new methods of EEG data acquisition could be possible, where significantly less time and thus costs have to be spent to obtain the data.

The paper is organized as follows: Next, we address the research background and related work. After that, we describe our ML method as well as the applied dataset. Subsequently, we present the results of our implemented method and discuss it, including theoretical and practical implications. Finally, we draw a conclusion that contains the limitations of our work and propose possible future research directions.

II. RESEARCH BACKGROUND

BCIs are a way of transmitting information from the brain to the outside via a non-muscular channel. As a non-invasive method to realize BCIs, EEG data can be recorded over the scalp [69]. In EEG-based BCIs the EEG signals are then translated into machine-readable outputs [26]. Specifically, in MI-BCI, the EEG signals are used to convert the motor intent of the brain into a signal, since MI generates similar EEG patterns as the real movements [70]. Through MI, frequency bands in the EEG data are affected by increase or decrease of power [71]. These power fluctuations can then be assigned to different tasks, such as opening and closing the hands.

A. SPECTRAL ANALYSIS

For a long time, these frequency bands were divided into the classic delta, theta, alpha, beta, and gamma bands [53]. Welch's method for spectral analysis enables a finer graded analysis of the frequency bands. The method uses a sliding window over overlapping segments in the EEG data, for estimation of the respective periodograms. By averaging all estimates, you get the power of a signal at different frequencies. Thereby, the frequency range can be picked individually, and the classic frequency bands do not have to be used. Rather, the complete frequency range is divided into equally sized self-selected bands, which allows for finer analysis. Thereby, the time series domain of the EEG signals can be converted into frequency domain [51].

Promising results have already been obtained in the literature using this approach. For example, to detect daytime sleepiness [72], working memory assessment [54] or early detection of alcohol use disorders [73].

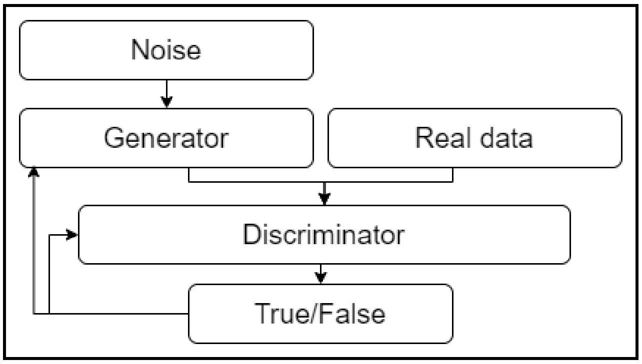

FIGURE 2. GAN structure.

B. EEG BASED BCI APPLICATIONS

As with the use cases just described, the focus of BCIs has long been in the medical field [74]. For example, in the field of prevention, as in the study of the influence of alcohol on brain waves [75], [76] or in the field of motion sickness [77]. Or also in the field of diagnosis, where for example brain tumor detection is performed using EEG data [78]. And last but not least, EEG-based BCIs are used for rehabilitation, for example in the field of neuroprosthetic devices [79], [80].

Since this century, however, BCIs have also become a topic in other research areas [81], and thus have applications in other domains as well. For example, as an interface for an augmented reality-based inspection system [82], or as a tool for human robot interaction [83]. Also, for example, to control robots via a BCI [84].

C. GENERATIVE ADVERSARIAL NETWORKS

GANs have already proven their added value in many areas where they support data generation where it was originally very complex and expensive [33].

The functionality of a classical GAN is based on the counter-play of two parties, a generator and a discriminator. As figure 2 illustrates, the generator has random variables (noise) at its disposal, from which it creates samples. The discriminator then has these generated and samples and the real data at its disposal. The task of the discriminator is then to correctly classify the sample and tell whether it is a real or synthetic sample. The goal of the generator is to make the synthetic data consistent with the real data. The performance of the two improves gradually and is based on each other [35].

GANs work in a kind of game, where on one side there is a neural network that is fed with training data to learn its distribution and generate realistic synthetic samples based on it in several iterations. And on the other side is a neural network that receives samples and has to decide whether they come from the training samples or from the generator (real or fake) [34].

D. GANS FOR EEG DATA

Also, EEG data have complex and long training times [31] due to the strong subject dependency of the EEG signals [21],

TABLE 1. Accuracy scores of TGAN & CTGAN at different machine learning classifiers with real and synthetic data [91].

	Real data	CTGAN data (99.63 % similarity)	TGAN data (98.76 % similarity)
Random Forest	64.75	49.8	50.57
XGBoost	61.13	49.17	48.75
LightGBM	62.8	49.79	49.3
Catboost	64.2	49.66	49.17

which results in high acquisition costs. Much of the research is concerned with augmenting real data with the GAN's synthetic data. This research investigates whether, for example, a classification algorithm works better with the addition of synthetic data than without. In the area of emotion recognition, an improvement of 1, 4 and 5 % was achieved in the classification of 3 emotions using a conditional Wasserstein GAN. Likewise, the performance of a Support Vector Machine (SVM) could be improved by 3, 9 and 20 % each for 3 datasets [85]. Another work also tries emotion recognition, among others, with a multiple generator conditional Wasserstein GAN. Again, using an SVM classifier with the augmented data, a performance increase of 1 % could be achieved in the best case. With the same data set, the previous approach achieved a 3 % improvement, but also used more data for augmentation [86]. Yet another paper attempts to extend the EEG data using a conditional Boundary Equilibrium GAN (cBGAN) on the same data set. cBGAN also uses the Wasserstein loss function for training. Here, using SVM classification, a performance increase of about 6 % could be achieved in the best case [87]. Another approach tries to improve the classification of autistic and non-autistic subjects by means of synthetic data. The approach is also based on a Wasserstein GAN, with which an improvement of between 8 and 10 % could be achieved in SVM classification [88]. Almost all the cited papers and many others in research use CTGAN for data augmentation in the area of EEG signals [85], [86], [88], [89], [90]. Also in all works, an evaluation of the augmentation is done by means of a classification comparison between a classification with only real data and one with the augmentation with synthetic data. A visual evaluation of the distribution of the real and synthetic data is also performed in the most cases.

In another work, the performance between conditional tabular GAN (CTGAN) and tabular GAN (TGAN) with synthetic data augmentation is compared. Also in this work, a visual comparison of the synthetic and real data was performed, comparing the distribution of the features. In the visual evaluation, CTGAN could generate the synthetic data better, also CTGAN could achieve a similarity score of 99.63 %, while TGAN achieved 98.76 %. In this work, an additional ML based evaluation was performed. The difference to the previous works is that a ML based classifier was trained once with the real data, and then with the synthetic data, and the results were compared against each other. For all classifiers used, the accuracy for the synthetic data is significantly worse than for the real data. The performance of the two classifiers is shown in Table 1. For both approaches, the complete available 12,811 samples were used to train the GANs [91].

E. GANS FOR TABULAR DATA

To the best of our knowledge, there is no scientific work that investigates the generation of synthetic EEG data using GANs with the preprocessing of the data using PSD. As described in the spectral analysis subsection, PSD can be used to convert the time series domain into a frequency domain [51]. This simplifies the data structure and eliminates the time dependency, which is why a tabular GAN can be chosen for the data. There are many different GAN approaches to generate synthetic tabular data.

MedGAN [92], ehrGAN [93] or tableGAN [94] are only 3 different approaches that generate tabular data. CTGAN is another model with which promising results have also been achieved [91], and that has out-performed all the named approaches in up to 8 different datasets regarding the discrete probability distributions of the synthetic data [95].

F. RANDOM FOREST CLASSIFIER

Classification accuracy is a very commonly used method to measure the quality of the EEG data generated by a GAN [96]. RF classifiers are one way to measure classification accuracy. RFs use several simple decision tree classifiers, each with a portion of the dataset, and then use averaging to bring the partial results together to achieve better results. The classifiers are trained with part of the data, and the remaining data is used to evaluate the model [68].

While there are many different ML algorithms for classifying data, RF classifiers provide a solution that can indicate which variables are decisive for the results and achieve good classification accuracy while keeping model complexity to a minimum [52].

III. METHODOLOGY

Our approach was to use a GAN to generate subject dependent and subject independent artificial data and then evaluate the results. Since CTGAN is one of the most promising architectures for both tabular and especially tabular EEG data, we decided to use CTGAN for data generation in this work. The EEG Data used for our approach consist of both categorical and numerical features. As shown in the architecture of the GAN in Figure 3, CTGAN is designed to generate synthetic data depending on the categorical values. First, the model selects a vector of categorical data, in our case there is only one categorical vector. Then a category is taken out of this vector. From the training data, a row with the corresponding category is then used to compare it to the synthetically generated row. The generator tries to create the synthetic row with the condition of the categorical variable. From the comparison, a score is generated, with which the generator can improve. The generator uses a variational Gaussian Mixture Model

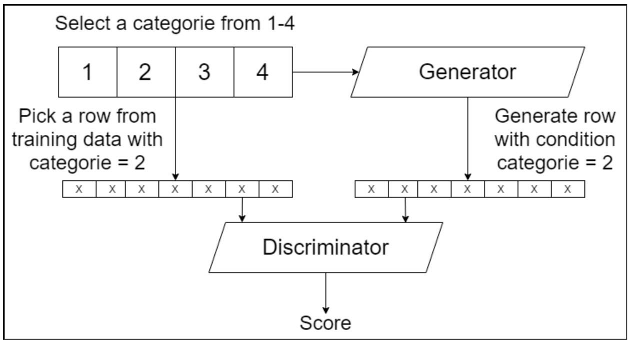

FIGURE 3. CTGAN architecture.

(VGMM) [97] for generating the data, and a Wasserstein GAN loss function [98] for optimizing the model, thereby fully-connected layers are being used. Next to the VGMM, one-hot encodings and the softmax activation function with added uniform noise are being used for categorical features [95].

Following the classic design science research approach [99], the first step of our methodology was to identify the lack of an artifact, that is able to generate EEG data for SMR-based BCI applications, where the temporal complexities can be neglected. Using the underlying dataset here, we developed the artifact in this paper based on this misconception. Once we created the artifact, we were able to use the underlying data to test and evaluate it. Figure 4 shows our methodical approach used in this paper, starting by reading in the EEG-data and preprocessing it with different filters and the application of Independent Component Analysis (ICA). From the pre-processed data, we then extracted the features using the PSD method and then selected the appropriate ones using feature importance analysis. We then used the introduced CTGAN for synthetic data generation. At one hand, we generated subject-dependent data, i.e. data from only one person. And on the other hand, we also generated subject-independent data, i.e. data from several persons as a basis. For Evaluation, we then made a table evaluation by comparing the synthetic and real table first visually then statistically. And finally, for both the subject-independent and subject-dependent approaches, we then trained an RF classifier and evaluated it with a 10-k cross-validation (CV).

A. DATA PREPROCESSING

The first step of data preprocessing was to remove noise and artifacts from the data. If a subject was, e.g., blinking his eyes or had muscular activities, that can be considered as biological artifacts. Computers, cable movements, etc. can create electrical noises which can be understood as non-biological artifacts [53]. These artifacts and noises cannot be prevented when recording EEG data via the scalp using multiple electrodes, so they must be filtered. There are different procedures to increase feature extraction efficiency by increasing signal-to-noise ratio [100], [101].

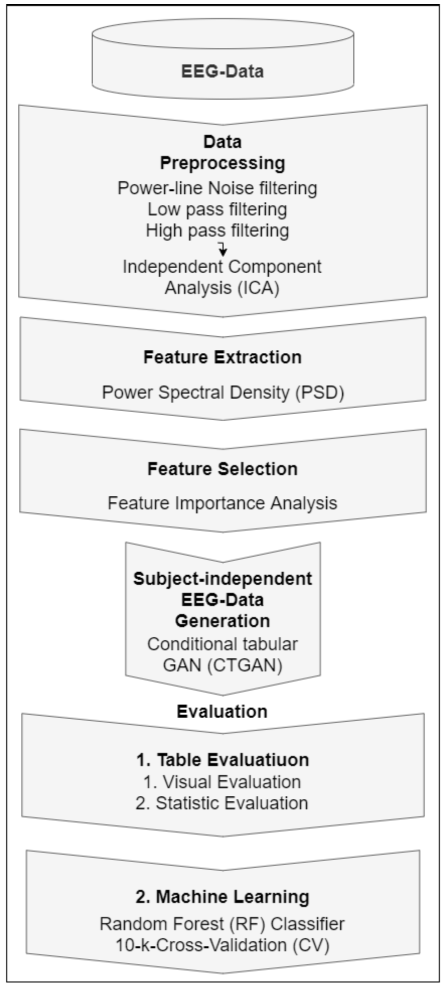

FIGURE 4. Methodical approach.

Our approach was to first remove the power-line noise by applying a notch filter [102], and then use a high-pass filter at 0.5Hz and a low-pass filter at 50 Hz, to reduce artifacts [53]. To then extract statistically independent components from the mixture of signals recorded by the electrodes, we used ICA [103]. To be more precisely, we used standardized automatic ICA [53], which is suitable for EEG data [104].

B. FEATURE EXTRACTION AND SELECTION

Feature Extraction is being used to reduce the dimensionality and thereby computation time of the SMR-based BCI system [105]. Spectral analysis can be used to convert the time series domain into frequency domain [52]. Thereby, the EEG data gets independent of the time domain.

Peter D. Welch introduced a method which uses a sliding window over overlapping segments, for estimation of the respective periodograms. By averaging all estimates, you get the power of a signal at different frequencies [51]. In our approach, we use a PSD based on a non-parametric approach of Welch's method. For the PSD we did not use the usual frequency bands of gamma, beta, alpha, theta and delta [53], but we used the whole frequency bandwidth up to 64.5 Hz

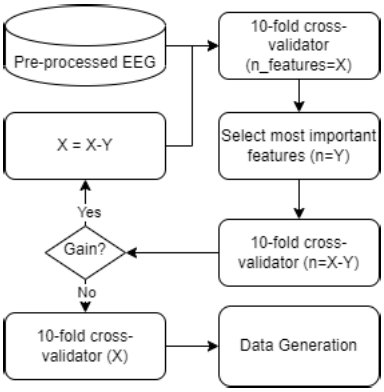

FIGURE 5. Feature selection approach.

and divided it into equally sized sub-bands of 0.5 Hz. Such a approach promises better classification results due to the higher information content, which has already been used for detecting schizophrenia [55] or detection of yes/no decisions based on EEG data [106]. Through PSD the EEG data is no longer strictly sequential but more frequency based, since sequential time periods are averaged in frequency sub-bands by the method.

In the initial dataset were four target classes (Left, Right, Up, Down). Left/Right represents the imagination of opening and closing from the left- (right-) hand. Up is represented by the imagination of opening both hands, and Down by the imagination of a rest state. For our approach, we merged Left and Right classes to one common class, and Up and Down classes to one common class, so we create a binary class problem. We achieved our goal by creating two new classes that accordingly consist of the EEG Signals of the initial classes. Thus, the first class corresponded to a horizontal movement and therefore the aversion (No Decision), and the second class corresponded to a vertical movement and therefore to agreement (Yes Decision) [107].

Finally, after creating a binary class problem, we were able to perform a feature importance analysis, for selecting the most important features and thereby reduce the complexity and dimensionality of our data. To identify the most important features for classification, we used the average reduction in impurity across all trees of a RF classifier model [108]. Therefore a RF from the scikit-learn library [109] was trained with the pre-processed data and used for the classification.

Figure 5 shows the chosen approach, in which the pre-processed data are run in several iterations through a 10-fold CV with a random forest and the feature importances are measured until there is no more predictive gain. This determines the number of features with which the best predictive performance can be achieved, and these features can be used in the next step.

C. DATA GENERATION USING CTGAN

There are many different approaches to create synthetic data using GAN [85], [86], [90]. But since we were able to convert the time series domain of the EEG data into frequency domain using PSD, we can use a more effective GAN architecture that is able to generate tabular data. Also, for synthetic tabular data generation there are many different GAN architectures. CTGAN and TGAN outperformed the other models for synthetic tabular data generation [110]. Between those two, ML performance after data generation was equal, but CTGAN was able to outperform TGAN by table evaluation [91], which is why we decided to use that model for data generation.

As briefly described in the introduction of this section, CTGAN uses a generator and a discriminator for synthetic data generation. CTGAN uses a VGMM for the numerical values [95], for estimation of the number of modes and fit a Gaussian mixture [111]. In our case, all columns except the target are numeric values. A Wasserstein GAN [98] loss function is used for gradient penalty.

For training the CTGAN we made two different approaches. The first approach was subject-independent, thus we took the second session of every subject for training the CTGAN. The subject-independent part here refers to the fact that the data from several subjects can subsequently be used independently in the ML evaluation, and thus in other applications. The second approach was subject-dependent, thus we took two randomly picked sessions (3 and 5) of a randomly picked subject (29) for training the CTGAN. Before starting to train the model we used the list of the most important features, whose origin was described in the last chapter, to reduce the dimensionality of the data, by only taking the most important features. In both approaches, we trained one GAN for the left/right targets and one GAN for the up/down targets. Next, we defined the target column of the training data as categorical. The CTGAN function requires the information about the categorical features for generating the data correctly. For the subject-independent approach we trained a GAN with 100 epochs, and for the subject-dependent approach we used 200 epochs.

The trained GANs can then be used to generate data samples. In both cases we created half of the samples with the GAN for left/right targets and the other half with the GAN for up/down targets. Subsequently we merged these data to have our final synthetic samples.

D. EVALUATION

For this paper we had two primary objectives, first, we wanted to show whether EEG data converted from a time domain to a frequency domain using PSD can be used so CTGAN can successfully learn to generate realistic synthetic data. Second, we wanted to investigate whether the synthetically generated data could be used for classification of Yes/No decisions. For this purpose we evaluated our approach with the EEG based MI dataset by Stiegern et al. [112]. Thereby, on the one hand we performed a visual and general evaluation of the tables and on the other hand we used a RF classifier to compare the classifications' performance between the real and synthetic data. Before evaluation, we used our trained CTGAN models to generate a table of synthetic data with the same size as the training table and then merged the outcomes. Last but not least, we will evaluate the diversity of the data in the subject-dependent case by calculating the Euclidean distance (ED) and the Wasserstein distance (WD).

1) TABLE EVALUATION

The visual and general evaluation of the tables is both done with the TableEvaluator library, which was built to evaluate how similar a synthesized dataset is to a real dataset. As a first general step, we use TableEvaluator's basic-statistical-evaluation method with both, the real data table and the synthetic data table. The method gives us a similarity score, which is the aggregation of mean, median, standard deviation and deviance values between the datasets. For the visual evaluation, in both approaches, we used the visual-evaluation method. The method compares the cumulative sums per feature, as well as the absolute log mean and standards of the numerical data [113]. For the cumulative sums per feature, we randomly selected 12 features, since a visual evaluation of over 1,000 features is not appropriate for this work.

2) MACHINE LEARNING

The general and visual evaluation using TableEvaluator library provides a general comparison of the datasets, but gives little information about the correlations between features. Also, the second objective of this paper was to investigate whether the artificial data is suitable for an ML approach for the subject-independent/dependent detection of Yes/No decisions. As described in the Feature extraction and selection subsection, we first created a two class problem in both approaches (Yes/No). Then in both cases, we used a RF classifier for evaluation [68]. For the subject-independent as well as the subject-dependent approach, we first trained and evaluated a RF classifier only with real data. Second, we also trained and evaluated a RF classifier with the real data combined with the synthetic data. Therefore, we simply concatenated the datasets. To evaluate how well the predictions made by the model match the observed data, cross validation (cv) is the most practical and best-known approach [21], [114]. By applying a 10-fold cv [115], the data is randomly divided into 10 equally sized parts. One of these parts is withheld while the other parts are used to train the classifier. This process is repeated 10 times until each part has been used for testing once [116]. The resulting CV matrix then indicates how robust a model is [117]. The training and testing data is running through the pre-processing steps by their own. Therefore methods such as ICA are also used on testing data independently. As shown by Song et al. [118], next to the most used leave-on-out approach for achieving subject-independence, there is another way to diminish subject-dependency by training the classifier with data across all subjects and sessions. In comparison to the approach of the authors, we use a 10-fold CV instead of strictly pre-training with subject-independent data and then fine-adjusting with data from the target subject, to achieve subject-independence [118]. The 10-fold CV was therefore used across all subjects, and it computed the classification accuracy trial by trial. Subsequently it averaged the final accuracy across subjects.

To finally test whether we could achieve a predictive gain with the synthetic data, we then took a new session to test the RF classifier in the subject-dependent case, and trained the model once with only the real data and for comparison with the concatenated real and synthetic data. In the subject-independent case, we also took data with which we had not trained the GAN and performed the same procedure.

3) DIVERSITY EVALUATION

To further evaluate the diversity of the generated data and thus our model, we will apply part of the methodology shown by Hartmann et al. [119] and calculate the ED and WD.

The WD describes how much effort is needed to convert two data distributions into each other. Thus, a small WD means that two distributions are similar. The ED also describes how similar two distributions are to each other. In both cases, we examine whether the GAN simply copies the data or actually adds variance.

Like Hartmann et al. [119] we will scale the data in the range -1 to 1 for comparability by calculating and subtracting the mean and then dividing by the absolute maximum value. We then calculate the WD between real and synthetic data for all trials together and add them up. To get the average WD, we divide this value by the number of channels.

For the ED, we use the same procedure, except that we also use the ED for real data only, and compare this value with the ED between real and synthetic data, since here a similar value speaks for similarity and variance [119].

Manhattan distance (MD) and cosine similarity (CS) are two other metrics that we can use to measure the similarity of synthetic and real data [120], [121]. Both metrics are calculated with the same procedure as the ED and WD. Again, for both MD and CS, we compare the achieved results between synthetic and real data, with the results between real data only.

The diversity evaluation only takes place in the subject-dependent case, as it is a comparison of the similarity of the data, which provides more information for the same subjects. The results can be transferred to the subject-independent case.

E. DATASET

We used the open-access EEG based MI dataset by Stiegern et al. [112] to validate our method. We decided to use this particular dataset, because we wanted to create an approach for both subject-independent and subject-dependent methods with high generalizability, and therefore needed data from many different subjects with a binary class problem. The proposed dataset contains EEG data from 64 electrodes that were digitized at 1,000 Hz and filtered between 0.1-200 Hz with an additional notch filter

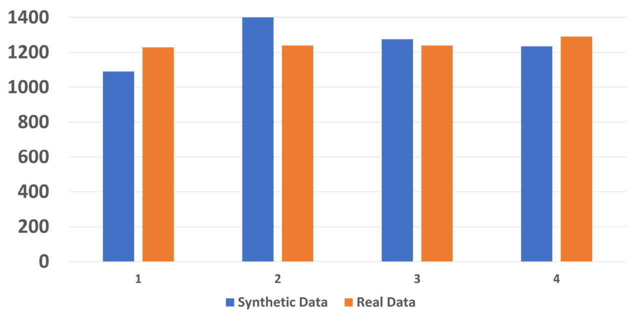

FIGURE 6. Subject-independent feature distribution of synthetic data vs real data.

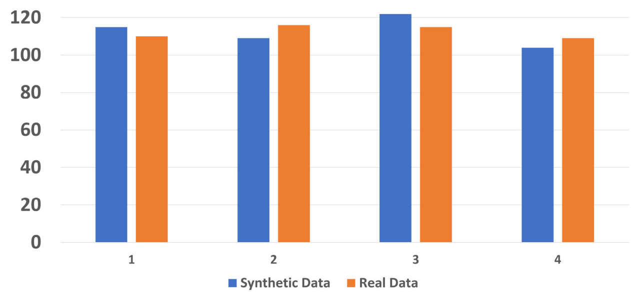

FIGURE 7. Subject-dependent feature distribution of synthetic data vs real data.

at 60 Hz. The internationally standardized 10-10 system was used to place the electrodes accordingly. 7–11 Sessions, in which a subject had to control a cursor to a target on a 2D screen, were performed by each of the 62 adult subjects. The subjects were instructed to imagine opening and closing their left (right) hand to move the cursor left (right), to imagine the opening and closing of both hands to move the cursor up, and to imagine a resting state or clear their mind to move the cursor down, which created the four classes of the dataset. Thus, in total, 600 hours of EEG recordings, consisting of 598 recording sessions with 269,099 trials for continuous 2D control, were recorded, which has made the dataset one of the largest and most complex datasets in the field [112]. The full dataset is available at: https://www.nature.com/articles/s41597-021-00883-1#SecVI.

Other datasets in this field had either way less data and samples [122] or fewer classes than the proposed dataset [123]. Even though BCI competition datasets are among the best known publicly available MI datasets, they are also mostly small and not complex with only nine subjects and no online feedback [20], [21], [124], [125].

IV. RESULTS

A. GENERAL EVALUATION

As described in the Table Evaluation subsection, we use the similarity score for the basic statistical analysis. Using the subject-independent approach, we achieved a score of 83.51 %. For the subject-dependent approach, we could even achieve 95.72 %. The analysis of the artificially generated data showed that there are slight differences in the distribution of the initial 4 targets as shown in figure 6 and figure 7, but this does not affect the final distribution of the two targets. In the figures, the y-axis represents the number of trials, while the x-axis represents the 4 different target classes.

B. VISUAL EVALUATION

Both figure 8 and figure 9 show that for both approaches the absolute log values follow the diagonal line and thus comparable means and standard deviations are present for the datasets [113]. The y-axis (Cumsum) in both figures represents the cumulative sum, i.e. the cumulative percentile frequency of the corresponding value on the x-axis and all values below it. So similar curves means similar data distributions between real and fake data. Nevertheless, some deviations can be seen, especially in the subject-independent approach. Especially the real data have some outliers, which were not taken into account by the GAN, but still differ from the synthetic data. The reason for the differences might be that the GAN focuses on the broadness of the data in order to best mimic it. In both figures it can also be seen that by the feature by feature evaluation between the real and fake data, most of the data points of the synthetic data in the features match the real data.

C. MACHINE LEARNING EVALUATION

Tables 2 and 3 show the results of the RF classifier, one subject-independent and one subject-dependent. It can be seen that all performance indicators in both the subject-independent and -dependent cases improve significantly when we concatenate the synthetic and real data. In the subject-dependent case, we can obtain a predictive gain of 17.53 %. In the subject-independent case, 7.51 %. In the test of the subject-independent case, with data with which the GAN model was not trained, we could not achieve a predictive gain. However, in the subject-dependent case with a new session, we were able to achieve a predictive gain of 2 %.

D. DIVERSITY EVALUATION

For measuring the diversity of the data, we calculated both the ED and the WD. For the ED in the subject-independent case, we obtained a value of 2.63 between the real distributions and the synthetic distributions. The ED between only real distributions was 2.18, giving us a difference of 0.45 between the two measurements.

For the WD we could achieve a value of 0.047, between the real and synthetic distributions. The MD scores between real and synthetic, as well as between real data only, was 0.0046. And last but not least the CS was 0.00059 between real and synthetic distributions and 0.00064 between only real distributions.

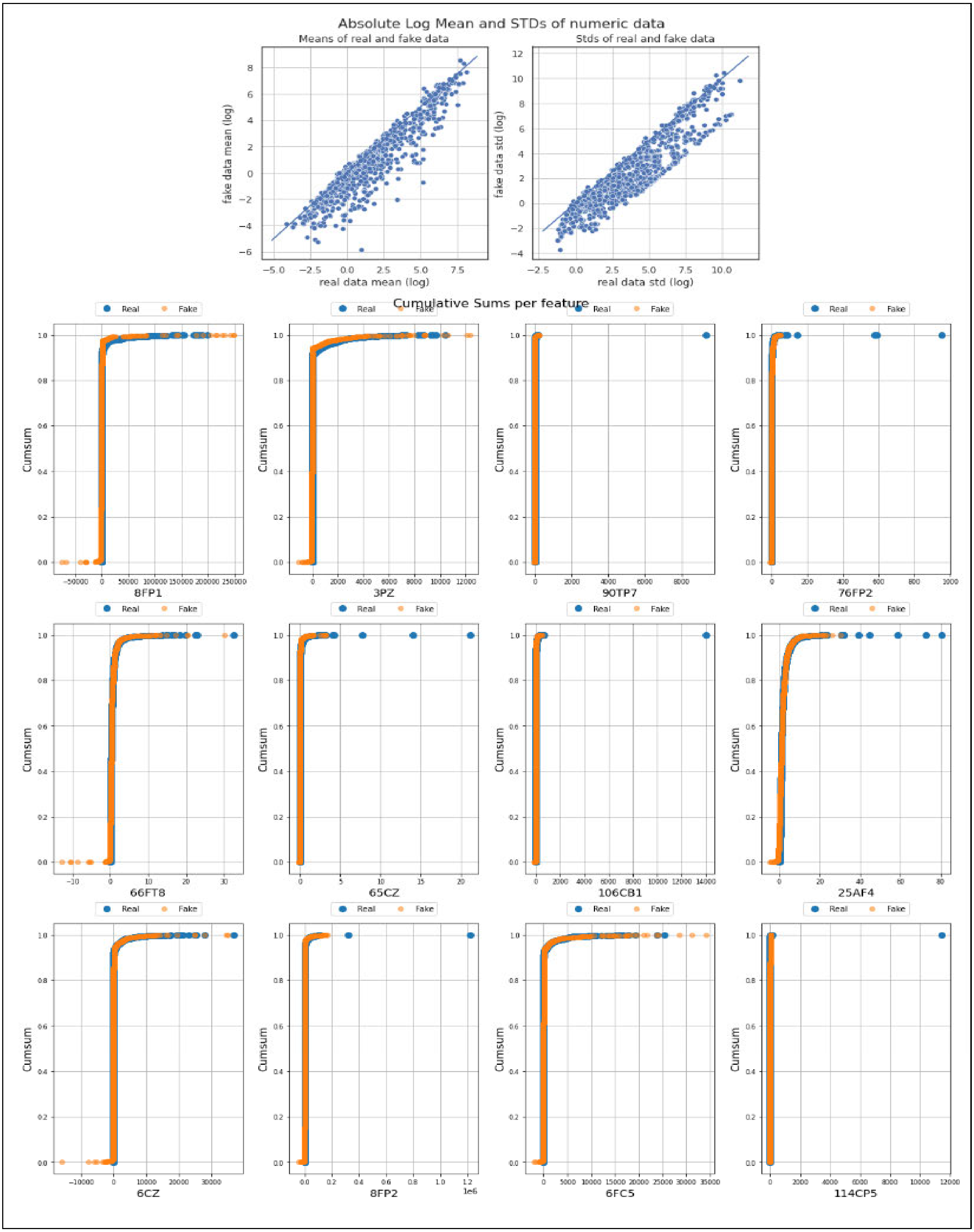

FIGURE 8. Visual results of subject-independent approach.

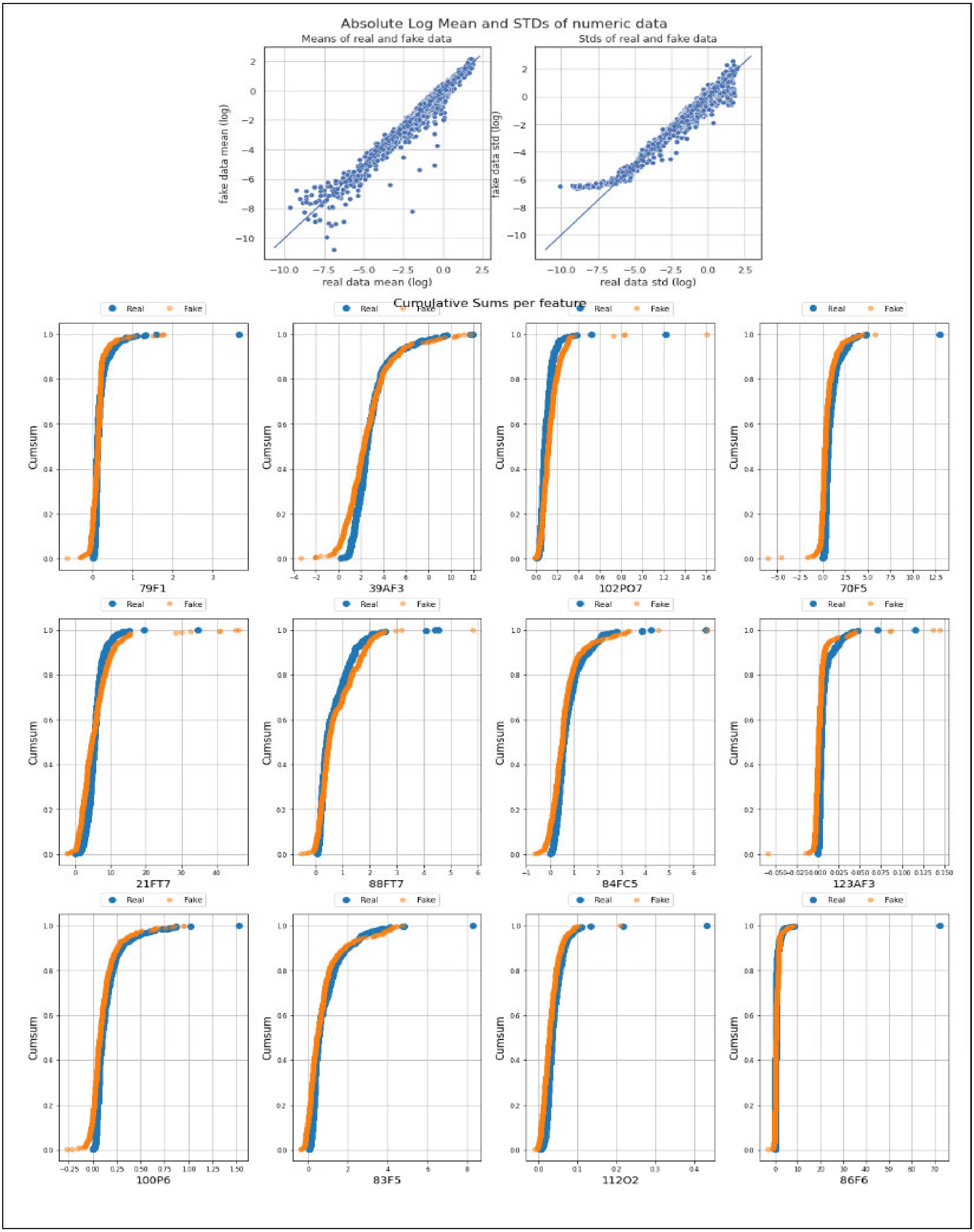

FIGURE 9. Visual results of subject-dependent approach.

TABLE 2. Subject-independent performance indicators.

Performance Indicator	Real data	Real + Synthetic data
Balanced Accuracy	65.6 %	83.13 %
Sensitivity (true positive rate)	65.64 %	83.12 %
Specificity (true negative rate)	65.64 %	83.12 %
Kappa	31.3 %	66.25 %

TABLE 3. Subject-dependent performance indicators.

Performance Indicator	Real data	Real + Synthetic data
Balanced Accuracy	85.14 %	92.65 %
Sensitivity (true positive rate)	85.14 %	92.65 %
Specificity (true negative rate)	85.14 %	92.65 %
Kappa	70.23 %	85.33 %

For the subject-independent case, as described in the methodology, we did not calculate the distances, because the values for ED and WD could not be compared here, due to the strong difference between EEG data of different subjects.

V. DISCUSSION

To the best of our knowledge, we are the first to use a GAN to generate artificial data for a subject-independent and subject-dependent algorithm to predict decisions based on binary-class MI EEG data for a BCI application employing a fine-grained EEG spectrum. Therefore, in the subject-independent case, with a predictive gain of 7.51 %, and in the subject-dependent case, with a predictive gain of 17.53 %, compared to the approach with only real data, we set a new benchmark.

We were able to identify two research studies where a GAN was used to generate MI EEG data and augment the real data with it. Both studies also used an RF classifier to test whether the augmentation of the real data with the synthetic data produced a predictive gain. As table 6 shows, Abdelfattah et al. were able to achieve a predictive gain of 13.1 %, while Debie et al. were able to achieve a predictive gain of 3.43 %. Both studies follow a subject-dependent approach. In this case, we were able to achieve a predictive gain of 17.53 % using the data augmentation. Uniquely, we used a fine-graded EEG spectrum for both generating and classifying the data, which increased accuracy and quality. Also unique is the removal of the dimensionality of the MI EEG data by the PSD, which differentiates the methodology from other work, and also has a positive impact on performance.

In addition, according to our knowledge, we are the first to achieve a predictive gain in decision prediction based on binary-class MI EEG data. Therefore, with a predictive gain of 2 % we set a new benchmark.

For both approaches, we used a 10-fold CV. In the subject-independent case, there is also the possibility to train the model first with purely subject-independent data and then fine-tune it with data from the subject. However, in both cases there are no major differences in achieving subject independence [118].

TABLE 4. Confusion matrix with mean values over ten folds. Subject-dependent approach with real data/real + synthetic data.

		Predicted	
		Yes	No
Actual	Yes	19.4 / 41.8	3.2 / 3.3
	No	3.5 / 3.3	18.9 / 41.6

TABLE 5. Confusion matrix with mean values over ten folds. Subject-independent approach with real data/real + synthetic data.

		Predicted	
		Yes	No
Actual	Yes	155.2 / 406.9	91.0 / 89.3
	No	80.6 / 79.4	173.2 / 424.4

TABLE 6. Comparison of subject-dependent data augmentation with previous work.

	Predictive Gain with RF Classifier
This work	17.53 %
Abdelfattah et al. [126]	13.1 %
Debie et al. [127]	3.43 %

In both cases, we used less than 10 % of the available data in the dataset to train the GANs. This means that we used little data compared to the comparative literature. Nevertheless, we achieved promising results with similarity scores of 83.51 % in the subject-independent case and 95.72 % in the subject-dependent case. Even if CTGAN takes the distribution of the features into account, the true distribution of the synthetically generated features shows that they were generated unequally. The used approach to generate the data in two parts leads to the fact that in the final result with only 2 features these are equally distributed, however, it can be that the similarity of the real and the synthetic data suffers from this inequality, which needs to be evaluated in future research. Rather, however, the GANs themselves may have been optimized, since they have been evaluated so far with only 100 epochs in the subject-independent case and 200 epochs in the subject-dependent case (CPU: Intel(R) Core(TM) i9-10885H CPU @ 2.40GHz).

The visual analysis of the feature-by-feature comparison between the real and synthetic data shows that while the synthetic data can represent most of the data points well in both cases, the real data contains more outliers, which could possibly be improved by further optimization of the GANs. Since outliers have a large influence on mean values [128], this could be an explanation for the differences in visual evaluation, but this needs further investigation. This could also be a reason why the standard deviations of the synthetic data are on average lower than those of the real data.

A more detailed feature-by -feature analysis would offer better conclusions. In the subject-dependent case, it can be seen that especially with the low standard deviations, the synthetic data sometimes have higher standard deviations, which could be due to the fact that the synthetic data can partially better represent the outliers.

The ML evaluation showed that the synthetic data can also be used for classification. Especially in the subject-independent case, a predictive gain could be achieved even with completely unseen data. The results need to be confirmed in further research and it can be investigated whether the inclusion of additional data has a positive effect on the results. The values of the confusion matrices do not allow any conclusions to be drawn about the particular distribution of the classification, but this must also be confirmed by a larger amount of data.

The diversity evaluation once again confirmed the quality of the generated data. Even if the ED of 0.45 would be the worst in comparison with the models of Hartmann et al. [119], the value is still relatively low and shows the similarity of the generated data. The comparatively worse result could be due to the inclusion of all channels, since only one specific channel was used for the ED in the comparison paper. For the WD we could achieve a result of 0.047, which is the best value compared to Hartmann et al's [119] results, and thus again for the quality of the synthetic data and the GAN, since the similarity and variance are best when the WD is low.

For the MD, small values mean greater similarity between the distributions, so the results we obtained show that the quality of the data generated is very good. In the case of CS, values in the zero range indicate that the data distributions are orthogonal to each other, which is the same in the case of the real data in relation to each other, as well as between the real and synthetic data. These results once more demonstrate the quality of the GAN.

A. APPLICATION IN REAL-WORLD BRAIN-COMPUTER INTERFACE SCENARIOS

The different evaluations show that our approach is a promising method to increase the classification accuracy of MI EEG data and to simplify the data acquisition process, but not how it can be used in real-world BCI applications.

As mentioned, the methodology can be used in BCIs to decode, for example, Yes/No decisions of total LIS patients. In this application, the patient would use a 64-electrode EEG cap as described in the dataset used, and imagine opening and closing their left (right) hand for a No decision. Or imagine the opening and closing of both hands or imagine a resting state for a Yes decision.

The raw EEG data is then divided into intervals as described in our methodology and converted into the frequency domain using PSD. By applying various methods such as ICA pre-processing is being done, and as we have shown in one of our previous works, the process of classification from data acquisition will be finished in 0.256 milliseconds [106].

The approach we have shown serves on the one hand to reduce the amount of data that needs to be collected for the BCI to achieve sufficient classification accuracy, and on the other hand to make data acquisition easier. For the subject-dependent calibration of a BCI system as just described, less data from the individual patient is needed, since most of the data can be generated through a GAN. Or, due to the synthetic data generated in advance, the system no longer needs to be generated in the future, since, as shown in our work, the decision can also be classified subject-independently without calibration.

VI. CONCLUSION

In this work, we have demonstrated a novel approach to generate four-class MI-BCI EEG data using GANs to be used for subject-independent/-dependent detection of Yes/No decisions. We have chosen a new methodology that uses a fine-graded EEG spectrum compared to previous work. Using this methodology, we were able to obtain similarity scores of 83.51 % in the subject-independent case and 95.72 % in the subject-dependent case. Further for the subject-dependent case, we were able to achieve a predictive gain of 2 % with unseen data, by augmenting the real wit the synthetic data. This shows the potential of our chosen approach to generate EEG data using a fine-graded EEG spectrum and GANs. Nonetheless, the practical real-world application is still constrained due to limited performance, but feasible. Therefore, our approach contributes significantly to the field of IT-enabled healthcare and gives a possibility to save time and costs in the acquisition of EEG data.

A. LIMITATIONS

Our proposed method does have certain limitations that should be acknowledged. Despite obtaining high internal validity through the use of 10-fold CV, external validation of the model has not been performed. To address this, further testing of the algorithm on datasets containing binary-class sensorimotor EEG data is necessary. In addition, subject-independence was achieved using a relatively new approach based on Song et al. [118]. This aspect could be confirmed by additional tests and, if necessary, by a leave-one-subject-out approach.

B. FUTURE WORK

In the future, we will further train the GANs with the remaining data of the dataset to improve the quality of the synthetically generated data and thus also to achieve better classification results. Furthermore, we will also optimize the GAN itself by optimizing the data generation approach, the number of epochs and the feature dependency within the data. The future visual analysis of the feature-by-feature comparison can and perhaps will also provide conclusions on which features the data can be better generated synthetically and which cannot, which can help to optimize the model accordingly.

Especially the promising results of the subject-dependent approach prove that our proposed methodology can help to generate data synthetically instead of cost-intensively acquiring them. We will evaluate these results in future research by conducting evaluations with more subjects and more data. In addition, we will try to obtain a predictive gain for completely unseen data by adding more trials in the subject-independent case.

In the future, we will test our model on other datasets for generalizability. Also, the performance of other state-of-the-art approaches on our given dataset should be compared to our results. We, therefore, encourage other researchers to use the method in their studies to advance the development of BCIs and GANs for EEG BCI data generation.

REFERENCES

[1] P. C. Scriba, M. Bauer, D. Emmert, A. Fateh-Moghadam, G. G. Hofmann, K. Horn, and C. R. Pickardt, "Effects of obesity, total fasting and realimentation on L-thyroxine (T4), 3, 5, 3'-l-triiodothyronine (T3),-3, 3', 5'-L-triiodorhyronine (rT3),-thyroxine binding globulin (TBG), cortisol, thyrotophin, cortisol-binding globulin (CBG), transferrin, α2-haptoglobin and complement C'3 in serum," *Eur. J. Endocrinol.*, vol. 91, no. 4, pp. 629–643, Aug. 1979.

[2] D. Lulé, C. Zickler, S. Häcker, M.-A. Bruno, A. Demertzi, F. Pellas, S. Laureys, and A. Kübler, "Life can be worth living in locked-in syndrome," *Prog. Brain Res.*, vol. 177, no. 9, pp. 339–351, Jan. 2009.

[3] A. Demertzi, S. Laureys, and M. Boly, "Coma, persistent vegetative states, and diminished consciousness," in *Encyclopedia of Consciousness*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 147–157. [Online]. Available: http://www.coma.ulg.ac.be/papers/vs/demertzi_VS2009.pdf

[4] American Congress of Rehabilitation Medicine, "Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness," *Arch. Phys. Med. Rehabil.*, vol. 76, no. 2, pp. 205–209, Feb. 1995.

[5] P. Masrori and P. Van Damme, "Amyotrophic lateral sclerosis: A clinical review," *Eur. J. Neurol.*, vol. 27, no. 10, pp. 1918–1929, Oct. 2020.

[6] K. Tsoi, P. Hung, and S. Poon, "Introduction to the minitrack on big data on healthcare application," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2021, p. 3389.

[7] G. Muhammad, F. Alshehri, F. Karray, A. E. Saddik, M. Alsulaiman, and T. H. Falk, "A comprehensive survey on multimodal medical signals fusion for smart healthcare systems," *Inf. Fusion*, vol. 76, pp. 355–375, Dec. 2021.

[8] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, "Uncovering the structure of clinical EEG signals with self-supervised learning," *J. Neural Eng.*, vol. 18, no. 4, pp. 1–32, 2021.

[9] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019.

[10] J. R. Wolpaw, "Brain–computer interfaces as new brain output pathways," *J. Physiol.*, vol. 579, no. 3, pp. 613–619, Mar. 2007.

[11] L. M. Nirenberg, J. Hanley, and E. B. Stear, "A new approach to prosthetic control: EEG motor signal tracking with an adaptively designed phase-locked loop," *IEEE Trans. Biomed. Eng.*, vol. BME-18, no. 6, pp. 389–398, Nov. 1971.

[12] S. Vaid, P. Singh, and C. Kaur, "EEG signal analysis for BCI interface: A review," in *Proc. 5th Int. Conf. Adv. Comput. Commun. Technol.*, Haryana, India, Feb. 2015, pp. 143–147.

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

[14] J. Decety, "The neurophysiological basis of motor imagery," *Behav. Brain Res.*, vol. 77, nos. 1–2, pp. 45–52, May 1996.

[15] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neurosci. Lett.*, vol. 239, nos. 2–3, pp. 65–68, Dec. 1997.

[16] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophys.*, vol. 113, no. 6, pp. 767–791, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245702000573

[17] J. N. Mak and J. R. Wolpaw, "Clinical applications of brain–computer interfaces: Current state and future prospects," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 187–199, 2009.

[18] A. Cruz, G. Pires, A. Lopes, C. Carona, and U. J. Nunes, "A self-paced BCI with a collaborative controller for highly reliable wheelchair driving: Experimental tests with physically disabled individuals," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 2, pp. 109–119, Apr. 2021.

[19] E. P. Torres, E. A. Torres, M. Hern?ndez-?lvarez, and S. G. Yoo, "EEG-based BCI emotion recognition: A survey," *Sensors*, vol. 20, no. 18, pp. 5083–5116, Sep. 2020.

[20] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Ḿuller, V. Kunzmann, F. Losch, and G. Curio, "The Berlin brain–computer interface: EEG-based communication without subject training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 147–152, Jun. 2006.

[21] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, vol. 35, no. 20, pp. 14681–14722, Aug. 2021.

[22] G. Townsend, B. Graimann, and G. Pfurtscheller, "Continuous EEG classification during motor imagery-simulation of an asynchronous BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no. 2, pp. 258–265, Jun. 2004.

[23] N. Naseer, M. J. Hong, and K.-S. Hong, "Online binary decision decoding using functional near-infrared spectroscopy for the development of brain–computer interface," *Exp. Brain Res.*, vol. 232, no. 2, pp. 555–564, Feb. 2014.

[24] D. A. Peterson, J. N. Knight, M. J. Kirby, C. W. Anderson, and M. H. Thaut, "Feature selection and blind source separation in an EEG-based brain–computer interface," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 19, pp. 1–13, Dec. 2005.

[25] Y. Li, J. Long, T. Yu, Z. Yu, C. Wang, H. Zhang, and C. Guan, "An EEG-based BCI system for 2-D cursor control by combining mu/beta rhythm and P300 potential," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2495–2505, Oct. 2010.

[26] D. J. McFarland and J. R. Wolpaw, "Brain–computer interfaces for communication and control," *Commun. ACM*, vol. 54, no. 5, pp. 60–66, Jun. 2011.

[27] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 51, pp. 17849–17854, Dec. 2004.

[28] D. Huang, K. Qian, D.-Y. Fei, W. Jia, X. Chen, and O. Bai, "Electroencephalography (EEG)-based brain–computer interface (BCI): A 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 379–388, May 2012.

[29] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based braincomputer interface paradigms," *J. Neural Eng.*, vol. 16, no. 1, pp. 1–21, Feb. 2019.

[30] S. Moghimi, A. Kushki, A. Marie Guerguerian, and T. Chau, "A review of EEG-based brain–computer interfaces as access pathways for individuals with severe disabilities," *Assistive Technol.*, vol. 25, no. 2, pp. 99–110, Apr. 2013.

[31] S. Aggarwal and N. Chugh, "Review of machine learning techniques for EEG based brain computer interface," *Arch. Comput. Methods Eng.*, vol. 29, no. 5, pp. 3001–3020, Jan. 2022.

[32] B. He, S. Gao, H. Yuan, and J. R. Wolpaw, "Brain–computer interfaces," in *Neural Engineering*, B. He, Ed. Boston, MA, USA: Springer, 2013, pp. 87–151.

[33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[34] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[35] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative adversarial networks: Introduction and outlook," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 4, pp. 588–598, Sep. 2017.

[36] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3774–3782.

[37] B. Chang, Q. Zhang, S. Pan, and L. Meng, "Generating handwritten Chinese characters using CycleGAN," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 199–207.

[38] L. Sixt, B. Wild, and T. Landgraf, "RenderGAN: Generating realistic labeled data," *Frontiers Robot. AI*, vol. 5, pp. 1–9, Jun. 2018.

[39] M. O. Turkoglu, W. Thong, L. Spreeuwers, and B. Kicanaoglu, "A layer-based sequential framework for scene generation with GANs," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 8901–8908.

[40] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[41] A. El-Nouby, S. Sharma, H. Schulz, R. D. Hjelm, L. E. Asri, S. E. Kahou, Y. Bengio, and G. Taylor, "Tell, draw, and repeat: Generating and modifying images based on continual linguistic instruction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10303–10311.

[42] M. Haidar and M. Rezagholizadeh, "TextKD-GAN: Text generation using knowledge distillation and generative adversarial networks," in *Proc. Can. Conf. Artif. Intell.* Montreal, QC, Canada: Springer, 2019, pp. 107–118.

[43] W. Nie, N. Narodytska, and A. Patel, "RelGAN: Relational generative adversarial networks for text generation," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–20.

[44] W. A. Chaovalitwongse, O. A. Prokopyev, and P. M. Pardalos, "Electroencephalogram (EEG) time series classification: Applications in epilepsy," *Ann. Oper. Res.*, vol. 148, no. 1, pp. 227–250, Nov. 2006.

[45] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks," in *Proc. Int. Conf. Artif. Neural Netw.* Springer, 2019, pp. 703–716.

[46] H. Ni, L. Szpruch, M. Sabate-Vidales, B. Xiao, M. Wiese, and S. Liao, "Sig-Wasserstein GANs for time series generation," in *Proc. 2nd ACM Int. Conf. AI Finance*, Nov. 2021, pp. 1–8.

[47] 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, pp. 1–12, 2019.

[48] F. Fahimi, Z. Zhang, T.-S. Lee, and C. Guan, "Deep convolutional neural network for the detection of attentive mental state in elderly," in *Proc. 7th Int. BCI Meeting*, Sacramento, CA, USA, 2018, pp. 1–2.

[49] 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.

[50] J. Yoon, D. Jarrett, and M. Van der Schaar, "Time-series generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Systems.*, vol. 32, 2019, pp. 5509–5519.

[51] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[52] R. Buettner, J. Frick, and T. Rieg, "High-performance detection of epilepsy in seizure-free EEG recordings: A novel machine learning approach using very specific epileptic EEG sub-bands," in *Proc. ICIS*, 2019, pp. 1–16, Art. no. 16.

[53] G. R. Müller-Putz, R. Riedl, and S. C. Wriessnegger, "Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications," *Commun. Assoc. Inf. Syst.*, vol. 37, no. 1, pp. 911–948, 2015.

[54] J. Breitenbach, A. Uzun, T. Rieg, H. Baumgartl, and R. Buettner, "A novel machine learning approach to working memory evaluation using resting-state EEG data," in *Proc. 25th Pacific Asia Conf. Inf. Syst. (PACIS)*, 2021, pp. 1–12, Art. no. 42.

[55] R. Buettner, D. Beil, S. Scholtz, and A. Djemai, "Development of a machine learning based algorithm to accurately detect schizophrenia based on one-minute EEG recordings," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2020, pp. 3216–3225.

[56] P. Sernadela et al., "Linked registries: Connecting rare diseases patient registries through a semantic web layer," *BioMed Res. Int.*, vol. 2017, pp. 1–13, Oct. 2017.

[57] S. Ekins, "Industrializing rare disease therapy discovery and development," *Nature Biotechnol.*, vol. 35, no. 2, pp. 117–118, Feb. 2017.

[58] S. Auvin, J. Irwin, P. Abi-Aad, and A. Battersby, "The problem of rarity: Estimation of prevalence in rare disease," *Value Health*, vol. 21, no. 5, pp. 501–507, May 2018.

[59] H. Baumgartl, S. Scholtz, D. Sauter, and R. Buettner, "Detection of schizophrenia using machine learning on the five most predictive EEG-channels," in *Proc. PACIS*, 2021, pp. 1–15, Art. no. 38.

[60] F. Janek, T. Rieg, and R. Buettner, "Detection of schizophrenia: A machine learning algorithm for potential early detection and prevention based on event-related potentials," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2021, pp. 3794–3804.

[61] J. W. Jeong, T. W. Wendimagegn, E. Chang, Y. Chun, J. H. Park, H. J. Kim, and H. T. Kim, "Classifying schizotypy using an audiovisual emotion perception test and scalp electroencephalography," *Frontiers Hum. Neurosci.*, vol. 11, Sep. 2017, Art. no. 450. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2017.00450

[62] A. Zandbagleh, S. Mirzakuchaki, M. R. Daliri, P. Premkumar, and S. Sanei, "Classification of low and high schizotypy levels via evaluation of brain connectivity," *Int. J. Neural Syst.*, vol. 32, no. 4, Apr. 2022, Art. no. 2250013.

[63] J. Trajkovic, F. Di Gregorio, F. Ferri, C. Marzi, S. Diciotti, and V. Romei, "Resting state alpha oscillatory activity is a valid and reliable marker of schizotypy," *Sci. Rep.*, vol. 11, no. 1, May 2021, Art. no. 42.

[64] H. Baumgartl, F. Dikici, D. Sauter, and R. Buettner, "Detecting antisocial personality disorder using a novel machine learning algorithm based on electroencephalographic data," in *Proc. 24th Pacific Asia Conf. Inf. Syst. (PACIS)*, vol. 48, 2020, pp. 1–14.

[65] H. Baumgartl, S. Bayerlein, and R. Buettner, "Measuring extraversion using EEG data," in *Information Systems and Neuroscience*. Aalen, Germany: Springer, 2020, pp. 259–265.

[66] 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," *J. Neural Eng.*, vol. 15, no. 3, pp. 55–110, 2018.

[67] F. Akram, S. M. Han, and T.-S. Kim, "An efficient word typing P300-BCI system using a modified t9 interface and random forest classifier," *Comput. Biol. Med.*, vol. 56, pp. 30–36, Jan. 2015.

[68] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[69] J. R. Wolpaw, "Brain–computer interfaces (BCIs) for communication and control: A mini-review," *Suppl. Clin. Neurophysiol.*, vol. 57, pp. 607–613, Feb. 2004.

[70] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, Mar. 2000.

[71] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.

[72] J. Breitenbach, H. Baumgartl, and R. Buettner, "Detection of excessive daytime sleepiness in resting-state EEG recordings: A novel machine learning approach using specific EEG sub-bands and channels," in *AMCIS, Healthcare Informat. Health Inf. Tech (SIGHealth)*, 2020, pp. 1–9, Art. no. 19.

[73] D. Flathau, J. Breitenbach, H. Baumgartl, and R. Buettner, "Early detection of alcohol use disorder based on a novel machine learning approach using EEG data," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 3897–3904.

[74] S. N. Abdulkader, A. Atia, and M.-S.-M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egyptian Informat. J.*, vol. 16, no. 2, pp. 213–230, Jul. 2015.

[75] W. Di, C. Zhihua, F. Ruifang, L. Guangyu, and L. Tian, "Notice of retraction: Study on human brain after consuming alcohol based on EEG signal," in *Proc. 3rd Int. Conf. Comput. Sci. Inf. Technol.*, vol. 5, Jul. 2010, pp. 406–409.

[76] T. K. P.Shri and N. Sriraam, "EEG based detection of alcoholics using spectral entropy with neural network classifiers," in *Proc. Int. Conf. Biomed. Eng. (ICoBE)*, Feb. 2012, pp. 89–93.

[77] L.-W. Ko, H.-C. Lee, S.-F. Tsai, T.-C. Shih, Y.-T. Chuang, H.-L. Huang, S.-Y. Ho, and C.-T. Lin, "EEG-based motion sickness classification system with genetic feature selection," in *Proc. IEEE Symp. Comput. Intell., Cognit. Algorithms, Mind, Brain (CCMB)*, Apr. 2013, pp. 158–164.

[78] V. S. Selvam and S. Shenbagadevi, "Brain tumor detection using scalp eeg with modified wavelet-ICA and multi layer feed forward neural network," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 6104–6109.

[79] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang, C. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang, "Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain–computer interface with robotic feedback," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 5549–5552.

[80] C. E. King, P. T. Wang, M. Mizuta, D. J. Reinkensmeyer, A. H. Do, S. Moromugi, and Z. Nenadic, "Noninvasive brain–computer interface driven hand orthosis," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 5786–5789.

[81] J. R. Wolpaw, "Brain–computer interfaces (BCIs) for communication and control," in *Proc. 9th Int. ACM SIGACCESS Conf. Comput. Accessibility (Assets)*. New York, NY, USA: Association for Computing Machinery, Oct. 2007, p. 12, doi: 10.1145/1296843.1296845.

[82] L. Angrisani, P. Arpaia, A. Esposito, and N. Moccaldi, "A wearable brain–computer interface instrument for augmented reality-based inspection in industry 4.0," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1530–1539, Apr. 2020.

[83] Y. Dmytriyev, A. M. A. Zaki, M. Carnevale, F. Insero, and H. Giberti, "Brain computer interface for human-cobot interaction in industrial applications," in *Proc. 3rd Int. Congr. Hum.-Comput. Interact., Optim. Robot. Appl. (HORA)*, Jun. 2021, pp. 1–6.

[84] C. J. Bell, P. Shenoy, R. Chalodhorn, and R. P. N. Rao, "Control of a humanoid robot by a noninvasive brain–computer interface in humans," *J. Neural Eng.*, vol. 5, no. 2, pp. 214–220, May 2008.

[85] 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.

[86] A. Zhang, L. Su, Y. Zhang, Y. Fu, L. Wu, and S. Liang, "EEG data augmentation for emotion recognition with a multiple generator conditional wasserstein GAN," *Complex Intell. Syst.*, vol. 8, no. 4, pp. 3059–3071, Aug. 2022.

[87] Y. Luo, L.-Z. Zhu, and B.-L. Lu, "A GAN-based data augmentation method for multimodal emotion recognition," in *Proc. Int. Symp. Neural Netw.* Moscow, Russia: Springer, 2019, pp. 141–150.

[88] G. Bouallegue and R. Djemal, "EEG data augmentation using wasserstein GAN," in *Proc. 20th Int. Conf. Sci. Techn. Autom. Control Comput. Eng. (STA)*, Dec. 2020, pp. 40–45.

[89] S. Bhat and E. Hortal, "GAN-based data augmentation for improving the classification of EEG signals," in *Proc. 14th Pervasive Technol. Rel. Assistive Environments Conf.*, Jun. 2021, pp. 453–458.

[90] Y. Jiao, Y. Deng, Y. Luo, and B.-L. Lu, "Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks," *Neurocomputing*, vol. 408, pp. 100–111, Sep. 2020.

[91] J. S. Lee and O. Lee, "CTGAN vs TGAN? Which one is more suitable for generating synthetic EEG data," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2359–2372, 2021.

[92] E. Choi, S. Biswal, B. Malin, J. Duke, W. F. Stewart, and J. Sun, "Generating multi-label discrete patient records using generative adversarial networks," in *Proc. Mach. Learn. Healthcare Conf.*, 2017, pp. 286–305.

[93] Z. Che, Y. Cheng, S. Zhai, Z. Sun, and Y. Liu, "Boosting deep learning risk prediction with generative adversarial networks for electronic health records," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 787–792.

[94] N. Park, M. Mohammadi, K. Gorde, S. Jajodia, H. Park, and Y. Kim, "Data synthesis based on generative adversarial networks," *Proc. VLDB Endowment*, vol. 11, no. 10, pp. 1071–1083, Jun. 2018, doi: 10.14778/3231751.3231757.

[95] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[96] A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, "Generative adversarial networks in EEG analysis: An overview," *J. Neuroeng. Rehabil.*, vol. 20, no. 1, pp. 1–24, Apr. 2023.

[97] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Anal.*, vol. 1, no. 1, pp. 121–143, Mar. 2006.

[98] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[99] R. H. Von Alan, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quart.*, vol. 28, no. 1, pp. 75–105, 2004.

[100] A. Jafarifarmand and M. A. Badamchizadeh, "EEG artifacts handling in a real practical brain–computer interface controlled vehicle," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1200–1208, Jun. 2019.

[101] D. Steyrl, R. Scherer, J. Faller, and G. R. Müller-Putz, "Random forests in non-invasive sensorimotor rhythm brain–computer interfaces: A practical and convenient non-linear classifier," *Biomed. Eng./Biomedizinische Technik*, vol. 61, no. 1, pp. 77–86, Feb. 2016.

[102] M. Ferdjallah and R. E. Barr, "Adaptive digital notch filter design on the unit circle for the removal of powerline noise from biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 6, pp. 529–536, Jun. 1994.

[103] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[104] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 8, 1996, pp. 145–151.

[105] J. Jin, S. Li, I. Daly, Y. Miao, C. Liu, X. Wang, and A. Cichocki, "The study of generic model set for reducing calibration time in P300-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 3–12, Jan. 2020.

[106] P. Penava, M.-L. Brozat, Y. Zimmermann, J. Breitenbach, P. Ulrich, and R. Buettner, "Subject-independent detection of yes/no decisions using EEG recordings during motor imagery tasks: A novel machine-learning approach with fine-graded EEG spectrum," in *Proc. 56th Hawaii Int. Conf. Syst. Sci.*, 2023, pp. 4057–4066.

[107] E. Andonova and H. A. Taylor, "Nodding in dis/agreement: A tale of two cultures," *Cognit. Process.*, vol. 13, no. S1, pp. 79–82, Aug. 2012.

[108] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC Bioinf.*, vol. 10, no. 1, pp. 1–16, Dec. 2009.

[109] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[110] S. Bourou, A. El Saer, T.-H. Velivassaki, A. Voulkidis, and T. Zahariadis, "A review of tabular data synthesis using GANs on an IDS dataset," *Information*, vol. 12, no. 9, pp. 375–389, 2021.

[111] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. Cambridge, U.K.: Springer, 2006.

[112] J. R. Stieger, S. A. Engel, and B. He, "Continuous sensorimotor rhythm based brain computer interface learning in a large population," *Sci. Data*, vol. 8, no. 1, pp. 1–10, Apr. 2021.

[113] B. Brenninkmeijer, Y. Hille, and A. Vries, "On the generation and evaluation of tabular data using GANs," Ph.D. dissertation, Dept. Inf. Retr., Radboud Univ. Nijmegen, Nijmegen, The Netherlands, 2019.

[114] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[115] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statist. Comput.*, vol. 21, no. 2, pp. 137–146, Apr. 2011.

[116] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.

[117] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 2. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1995, pp. 1137–1143.

[118] Y. Song, L. Yang, X. Jia, and L. Xie, "Common spatial generative adversarial networks based EEG data augmentation for cross-subject brain–computer interface," Feb. 2021, pp. 1–12, *arXiv:2102.04456*.

[119] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals," 2018, pp. 1–7, *arXiv:1806.01875*.

[120] F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Semantic cosine similarity," in *Proc. 7th Int. Student Conf. Adv. Sci. Technol. ICAST*, vol. 4, no. 1, 2012, pp. 1–2.

[121] J. Lin, Y. Li, and G. Yang, "FPGAN: Face de-identification method with generative adversarial networks for social robots," *Neural Netw.*, vol. 133, pp. 132–147, Jan. 2021.

[122] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "EEG datasets for motor imagery brain–computer interface," *GigaScience*, vol. 6, no. 7, pp. 1–8, Jul. 2017.

[123] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, pp. 1–16, May 2019.

[124] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain–computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, Aug. 2007.

[125] 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, 2008.

[126] S. M. Abdelfattah, G. M. Abdelrahman, and M. Wang, "Augmenting the size of EEG datasets using generative adversarial networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–6.

[127] E. Debie, N. Moustafa, and M. T. Whitty, "A privacy-preserving generative adversarial network method for securing EEG brain signals," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[128] E. H. Livingston, "The mean and standard deviation: What does it all mean?" *J. Surgical Res.*, vol. 119, no. 2, pp. 117–123, Jun. 2004.

PASCAL PENAVA (Member, IEEE) received the B.S. degree in information systems from Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, in 2021. He is currently pursuing the M.S. degree in digitalization and entrepreneurship with the University of Bayreuth, Bayreuth, Germany. His research interests include the development of EEG based brain–computer interfaces and machine learning based analyses of time-series data.

RICARDO BUETTNER (Senior Member, IEEE) received the Dipl.-Inf. degree in computer science and the Dipl.-Wirtsch.-Ing. degree in industrial engineering and management from the Ilmenau University of Technology, Germany, the Dipl.-Kfm. degree in business administration from the University of Hagen, Germany, the Ph.D. degree in information systems from the University of Hohenheim, Germany, and the Habilitation (venia legendi) degree in information systems from the University of Trier, Germany. He is currently a Chaired Professor of information systems and data science with the University of Bayreuth, Germany. He has published over 140 peer-reviewed articles, including articles in *Electronic Markets*, *AIS Transactions on Human–Computer Interaction*, *Personality and Individual Differences*, *European Journal of Psychological Assessment*, *PLOS One*, and IEEE Access. He received 17 international best paper, the best reviewer, and the service awards and award nominations, including best paper awards by *AIS Transactions on Human–Computer Interaction*, *Electronic Markets*, and HICSS, for his work.