Adversarial Deep Learning in EEG Biometrics

Ozdenizci, O.; Wang, Y.; Koike-Akino, T.; Erdogmus, D.

Abstract

Deep learning methods for person identification based on electroencephalographic (EEG) brain activity encounters the problem of exploiting the temporally correlated structures or recording session specific variability within EEG. Furthermore, recent methods have mostly trained and evaluated based on single session EEG data. We address this problem from an invariant representation learning perspective. We propose an adversarial inference approach to extend such deep learning models to learn session-invariant person-discriminative representations that can provide robustness in terms of longitudinal usability. Using adversarial learning within a deep convolutional network, we empirically assess and show improvements with our approach based on longitudinally collected EEG data for person identification from half-second EEG epochs.

Adversarial Deep Learning in EEG Biometrics

Ozan Özdenizci, Ye Wang, Toshiaki Koike-Akino, and Deniz Erdoğmuş

Abstract—**Deep learning methods for person identification based on electroencephalographic (EEG) brain activity encounters the problem of exploiting the temporally correlated structures or recording session specific variability within EEG. Furthermore, recent methods have mostly trained and evaluated based on single session EEG data. We address this problem from an invariant representation learning perspective. We propose an adversarial inference approach to extend such deep learning models to learn session-invariant person-discriminative representations that can provide robustness in terms of longitudinal usability. Using adversarial learning within a deep convolutional network, we empirically assess and show improvements with our approach based on longitudinally collected EEG data for person identification from half-second EEG epochs.**

Index Terms—**person identification, biometrics, EEG, adversarial learning, invariant representation, convolutional networks**

I. Introduction

Non-invasively recorded electroencephalographic (EEG) human brain activity has gained interest as an alternative person-discriminative biometric due to its continuous accessibility, privacy compliancy, and relatively harder forgeability, in comparison to today's most prevalent biometric identification approach of fingerprint recognition. A significant amount of work within the field of statistical EEG signal processing proposed novel methodologies to explicitly access person-discriminative neural sources from EEG. This problem was successfully tackled both in the context of *person identification*, where an individual is assigned to a label (class) within a specific set of people that an identification model is trained on [1], as well as for *person authentication*, where a one-to-one matching in decision making for person recognition is performed [2]. Longitudinal studies also confirm the feasibility of EEG as an alternative means of biometrics [3–5]. However, one recent study demonstrates different affective mental states potentially influencing stability of EEG as a tool for user identification [6]. As such, discriminative EEG biometric feature extractor models that can filter out specific nuisance variables are likely to enhance usability of generated invariant features for biometric identification. Similarly, this idea can extend to filtering out recording session related confounders from the feature learning process for longitudinal model robustness.

Recent progress in EEG deep learning has capabilities to tackle this problem. However importantly, current deep learn-

ing models for EEG biometric identification are vastly evaluated by within-session (i.e., within-recording) cross-validation protocols [7–10]. Due to their deep and complex nature, these models are particularly prone to capturing recording-specific variability rather than individual neural biomarkers. Going further, such models rely on the hypothesis that the deep architectures will internally learn invariant, generalizable features. This assumption is naturally constrained with the amount of available person-representative EEG data. Here, we highlight the need to extend EEG neural network models to explicitly learn invariant features against longitudinal variabilities.

In this study, we present an adversarial inference approach to extend deep learning based EEG biometric identification models, to learn session-invariant person-discriminative features (representations). We evaluated our approach based on EEG data recorded from ten healthy subjects participating in rapid serial visual presentation (RSVP) based brain-computer interface (BCI) experiments, which are identically performed on three separate sessions (i.e., days) for each participant. Empirical evaluations revealed a significant improvement with adversarial session-invariant feature learning for across-sessions person identification compared to conventional methods.

II. Prior Work

A. EEG-Based Biometric Identification

Pioneering neural signatures in EEG biometrics were considered to be visually evoked potentials (VEPs) [1]. Several pieces of work investigated spatio-temporal dynamics of EEG during visual stimuli perception for person identification [11–15]. Going further, since VEP-based experiment designs would require high physical and mental user attention, various studies extended this interest to different settings. Some examples include EEG recordings during silently reading texts [16], thorough EEG time-frequency domain explorations during emotion elicitation, resting-state, or motor imagery/execution tasks [17], fully task-independent designs where data are collected during various kind of auditory stimuli that does not require particular attention [18], or imagined speech [19]. An alternative approach is multitask learning in EEG biometrics, which was addressed in a study where person identification and motor task prediction was performed simultaneously through a shared representation to take advantage of latent task-specific informations [20]. Overall, these methods were mostly explored by data set specific traditional EEG processing methods that uses time-domain features [11–14], spectral decompositions [15–18] or autoregressive coefficients [19].

Motivated by its rapid progress, deep neural networks have recently gained significant interest as generic spatio-temporal EEG feature extractors. Mainly structured with convolutional

architectures, deep neural networks were introduced for P300 detection [21], steady-state visually evoked potential detection [22], rhythm perception during auditory stimuli [23], decoding of motor imagery [24], as well as recently for non-task-specific discriminative EEG feature extraction [25], [26]. Such convolutional neural networks (CNNs) were also extended to recurrent-CNNs [27], as well as deep convolutional autoencoders [28]. However, minimal progress was recently made in using these generic EEG feature extractors for biometrics. Some examples explore CNNs on resting-state [7] and motor imagery EEG [29], recurrent neural networks (RNNs) [9], as well as a combination of CNNs and RNNs for person-discriminative feature extraction under different mental states [10]. Similarly, a recent task-independent approach applies deep networks to EEG data recorded during driving [8].

Although resting-state EEG or recordings under different mental states [6] may prove a baseline for task-independence, models that can filter out any underlying cued states to learn invariant features would be useful in EEG biometrics. Furthermore, given that deep learning models can easily capture recording-specific artifacts rather than individual EEG biomarkers, feature invariance across recording sessions would be of particular interest for longitudinal usability. To this end, existing studies rely on deep capabilities of the networks to learn invariant and robust biometric EEG features when a large pool of data is used. This assumption can be constrained with the amount of person-representative EEG data. Going further, recent works mostly evaluate their methods based on within-session model learning and testing [7–10]. Hence, explicitly learning session-invariant biometric representations with deep learning remained as an open question for exploration.

B. Adversarial Invariant Representation Learning

Adversarial learning has been successfully applied in many deep learning applications to date, mainly popularized within generative model approaches for image data augmentation [30]. Adversarial training within generative models (e.g., variational autoencoder (VAE)) were also used for invariant latent representation learning, to disentangle specific attributes (e.g., nuisance variables) from the representations [31–33]. These architectures rely on training a generative model objective (e.g., evidence lower bound for VAE [34]), alongside a competing adversary with an antagonistic feedback on the overall optimization objective to enforce the invariance. Such attribute-invariant latent representations can then be used to manipulate these attributes in data augmentation [33].

In our interest, there exists significant work on learning discriminative, attribute-invariant encoder models that do not require a generative decoder counterpart. In a discriminative context, features are learned through an encoder during an adversarial training game, by maximizing label prediction certainty from learned features, while minimizing the certainty of inferring the attribute (e.g., nuisance) variables from these features [35], [36]. To date, these advancements in invariant feature learning have not been considered in EEG biometrics. In the light of these works, we hypothesize that adversarial discriminative inference can be useful in terms of session-invariant deep EEG feature learning for biometrics.

III. ADVERSARIAL CONVOLUTIONAL NETWORK

A. Adversarial Model Learning

Let $\{(\boldsymbol{X}_i, s_i, r_i)\}_{i=1}^n$ denote a model training data set, with $\boldsymbol{X}_i \in \mathbb{R}^{C \times T}$ the raw EEG data at *epoch* i recorded from C channels for T discretized time samples, $s_i \in \{1, \ldots, S\}$ the subject identification (ID) number for the person that the EEG is collected from, and $r_i \in \{1, 2, \ldots, R\}$ the recording session ID (i.e., day) that the EEG data of the subject is collected at. EEG data generation process is assumed to be jointly dependent on subject and recording session IDs (i.e., $\boldsymbol{X} \sim p(\boldsymbol{X}|s, r)$). In our decoding problem, the class labels from the discriminative perspective are the subject IDs, and we aim to learn person identification models over a specified number of subjects S in our training data set.

In the proposed adversarial discriminative model learning framework, we train a deterministic convolutional *encoder* $g(\boldsymbol{X}; \theta)$ with parameters θ, that will ideally output representations which are predictive of s as recovered by an *identifier* network parameterized by γ, but not predictive of r as concealed from an *adversary* network parameterized by ϕ which tries to recover r. Here, the parametric identifier models the likelihood $q_\gamma(s|g(\boldsymbol{X}; \theta))$, whereas the parametric adversary models the likelihood $q_\phi(r|g(\boldsymbol{X}; \theta))$. While the adversary is trained to maximize $q_\phi(r|g(\boldsymbol{X}; \theta))$, the encoder conceals information regarding r from the learned representations by minimizing this likelihood, as well as also retaining person-discriminative information by maximizing $q_\gamma(s|g(\boldsymbol{X}; \theta))$. This results in jointly training the networks towards the objective:

$$\min_{\theta, \gamma} \max_{\phi} \mathbb{E}[-\log q_\gamma(s|g(\boldsymbol{X}; \theta)) + \lambda \log q_\phi(r|g(\boldsymbol{X}; \theta))], \quad (1)$$

with $\lambda > 0$ denoting the adversarial loss weight to enforce stronger invariance and trading-off with identification performance. Optimization is performed by using stochastic gradient descent alternatingly for the adversary, and the encoder-identifier networks to optimize Eq. 1. An overview of the adversarial training framework is illustrated in Figure 1.

B. Convolutional Network Architecture

Our EEG feature encoder consists of four convolutional blocks (c.f. Table I), mainly structured by field-leading works [25], [26]. Sequentially, we perform temporal convolutions resembling to frequency filtering, depthwise convolutions [37] as spatial filtering of frequency-specific activity, and two more 2D convolution blocks for spatio-temporal feature aggregation. After convolutions we use batch normalizations (BatchNorm) [38], and rectified linear unit (ReLU) activations. We use ReLU activations since they typically learn faster in networks with many layers [39], and were successfully used in EEG deep learning [27]. We did not observe performance increases by using exponential linear units [26]. Deep convolution kernel sizes are mostly guided by [25], since we observed training set overfitting and overall lack of generalization with shallower layers. Beyond architectural design choices, in fact, our adversarial inference approach is applicable to any EEG neural network model by simply modifying the training objective.

Representations learned by the encoder were used by the identifier and adversary for linear classifications of subject
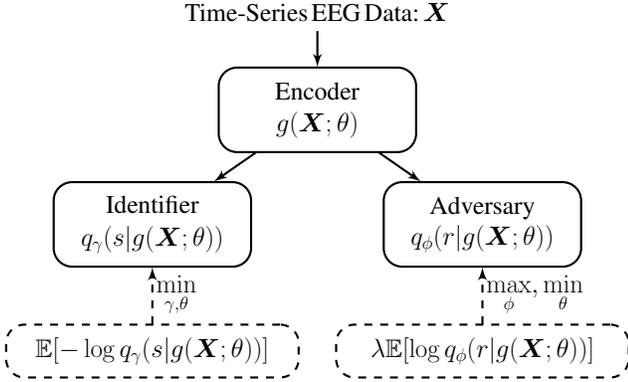
Fig. 1. Adversarial discriminative model training framework. Encoder, identifier and adversary networks are simultaneously trained towards the objective in Eq. 1, as illustrated by the loss functions in the dashed boxes.

TABLE I
CONVOLUTIONAL ENCODER NETWORK SPECIFICATIONS

Layer	Operation	Output Dim.
Encoder Input	Reshape	$(1, C, T)$
Convolutional Block 1	$20\times$ Conv $(1, T/2)$ BatchNorm	$(20, C, T/2)$ $(20, C, T/2)$
Convolutional Block 2	$20\times$ DepthwiseConv $(C, 1)$ BatchNorm + ReLU + Reshape	$(400, 1, T/2)$ $(1, 400, T/2)$
Convolutional Block 3	$200\times$ Conv $(400, T/4)$ BatchNorm + ReLU + Reshape	$(200, 1, T/4)$ $(1, 200, T/4)$
Convolutional Block 4	$100\times$ Conv $(200, T/8)$ BatchNorm + ReLU	$(100, 1, T/8)$ $(100, 1, T/8)$
Encoder Output	Flatten	$100 * T/8$

and session IDs. Both networks consisted of a fully-connected layer with S or R softmax units, respectively for the identifier and adversary, to obtain the normalized log-probabilities represented in the loss functions as shown in Figure 1.

IV. EXPERIMENTAL STUDY

A. Study Design and Experimental Data

Ten healthy subjects participated in the experiments at three identical sessions performed on different days. The average interval between two consecutive recording sessions for the same person was 7.85 ± 14.34 days, with a minimum of 1 and a maximum of 65 days between sessions across all experiments. Before the experiments, all participants gave their informed consent in accordance with the guidelines set by the research ethics committee of Northeastern University.

During each session, EEG data were recorded from participants while they were using the RSVP Keyboard™, an EEG-based BCI speller that relies on the RSVP paradigm to visually evoke event-related brain responses for user intent detection [40]. EEG data were recorded from 16 channels (as described and also used in previous work [41], [42]), sampled at 256 Hz, using active electrodes and a g.USBamp biosignal amplifier (g.tec medical engineering GmbH, Austria).

Participants were using the RSVP Keyboard™ in offline calibration mode, attending to pre-specified letters while random sequences of visual letter stimuli are presented to them. EEG data were epoched at [0-0.5] seconds post-stimuli intervals to construct each subjects' specific session data set. We pooled all epochs within a recording session irregardless of their attended versus non-attended stimuli labels. The complete data set consisted of 41,400 epochs of 16 channel EEG data for 128 samples. Epochs were equally distributed across subjects and sessions (i.e., 1,380 epochs per subject per session).

B. Data Analysis and Implementation

We performed both within-session person identification analyses to illustrate conventional evaluation methods, as well as across-sessions analyses to demonstrate the impact of session-invariant feature learning. In within-session analyses, all subjects' data were pooled per session ID and three distinct

within-session models were trained. *Training*, *validation* and *test* data sets were randomly constructed as 70%, 10% and 20% portions of the within-session data pools. Here, we ignore the adversary network and train the encoder-identifier as a regular CNN. Across-sessions analyses were performed by a leave-one-session-out approach, where the left-out session constituted the *test* set, and the *training* and *validation* sets were constructed as 80% and 20% random splits of the other two sessions' pooled data. Here, the adversary was jointly trained to recover session IDs from the encoder outputs.

We further compare the deep CNN-based models with two baseline methods. First approach uses power spectrum features [15–18]. We concatenated channel log-bandpowers computed in θ- (4-8 Hz), α- (8-15 Hz), β- (15-20 Hz, 20-25 Hz, 25-30 Hz) and γ- (30-45 Hz, 45-60 Hz, 60-75 Hz) bands through FFT using the Welch's method and Hanning windows, into a feature vector. We used a quadratic discriminant analysis (QDA) classifier as the identifier, since it performed better than linear classifiers in our analyses. Second method uses principal component analysis (PCA) projection on the 2048 dimensional (16 times 128) vectorized EEG data, and a QDA classifier [13], [14]. PCA projection dimensionality was determined as the minimum number of components that accounted for 90% of total variance, which varied between 168 and 174.

All EEG data were normalized to have zero mean and scaled to [-1,1] range by dividing with the absolute maximum value at each epoch and channel individually. No channel selection or offline artifact removal was performed. Input dimensionality of the CNN networks is C=16 channels by T=128 samples. The identifier network has a 10-dimensional output as a S=10 class classifier across the subjects, whereas the adversary is a binary classifier across R=2 sessions. It is important to note that the binary session ID is not a shared variable across subjects, but simply indicates variability with its label in conjunction with the subject IDs. Ideally, alternating the session ID labels should not have a significant impact on the learned models.

Networks were trained with 100 epochs per batch, with 500 repeated training set passes. Early stopping was performed based on validation set loss of the identifier network. Parameters were updated once per batch with Adam [43]. We used the Chainer deep learning framework for implementations [44].

TABLE II
LEAVE-ONE-SESSION-OUT PERSON IDENTIFICATION ACCURACIES (%). TEST COLUMNS SHOW THE IDENTIFIER ACCURACIES FOR THE LEFT-OUT
SESSION. FIRST TWO ROWS DENOTE THE BASELINE METHODS. NON-ADVERSARIAL MODEL DENOTES A REGULAR DEEP CNN. ADVERSARIAL MODELS
LEARN INVARIANT FEATURES ACROSS TWO TRAINING SESSIONS. PARENTHESES DENOTE THE STANDARD DEVIATIONS ACROSS 10 REPETITIONS.

| | Learning on Session 2 and 3 | | | Learning on Session 1 and 3 | | | Learning on Session 1 and 2 | | |
| | Validation Set | | Test Session 1 | Validation Set | | Test Session 2 | Validation Set | | Test Session 3 |
	Identifier	Adversary		Identifier	Adversary		Identifier	Adversary	
Spectral Powers + QDA	80.5 (.005)	–	44.9 (.001)	81.4 (.004)	–	52.1 (.001)	83.6 (.004)	–	49.5 (.002)
PCA + QDA	85.6 (.006)	–	57.6 (.002)	85.1 (.006)	–	58.9 (.001)	88.1 (.002)	–	64.1 (.002)
Non-Adversarial $\lambda = 0$	91.6 (.008)	76.0 (.02)	62.3 (.02)	97.9 (.006)	78.9 (.03)	63.2 (.02)	97.9 (.001)	76.9 (.05)	69.2 (.02)
Adversarial $\lambda = 0.005$	91.2 (.01)	65.2 (.01)	65.1 (.01)	98.4 (.005)	63.7 (.03)	69.1 (.01)	98.2 (.006)	60.4 (.03)	**71.6** (.02)
Adversarial $\lambda = 0.01$	90.7 (.01)	60.8 (.02)	**66.6** (.02)	98.4 (.002)	58.7 (.02)	**69.2** (.02)	98.4 (.003)	58.5 (.03)	71.3 (.01)
Adversarial $\lambda = 0.02$	91.1 (.01)	57.2 (.04)	65.3 (.01)	98.1 (.003)	56.0 (.03)	68.7 (.03)	98.2 (.003)	54.9 (.04)	71.2 (.01)
Adversarial $\lambda = 0.05$	91.0 (.009)	53.4 (.05)	65.4 (.02)	97.8 (.005)	53.5 (.03)	67.6 (.02)	98.2 (.003)	54.0 (.03)	71.1 (.02)
Adversarial $\lambda = 0.2$	91.8 (.005)	54.5 (.03)	64.1 (.02)	97.1 (.006)	53.1 (.04)	66.4 (.02)	97.4 (.004)	53.5 (.03)	71.1 (.03)

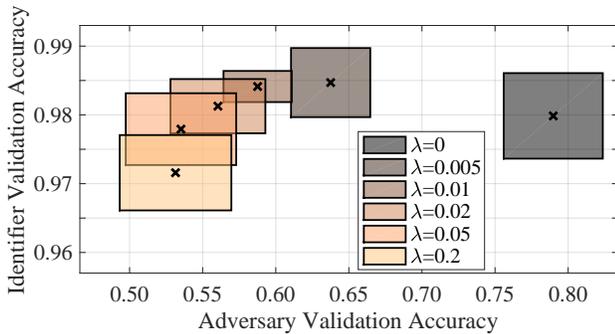
Fig. 2. Adversarial ($\lambda > 0$) and non-adversarial ($\lambda = 0$) model evaluations with identifier and adversary validation accuracies for the leave-Session 2-out learning case. Center marks denote the means across ten repetitions and widths denote ± 1 standard deviation intervals in both dimensions.

C. Within-Session Person Identification

Models were learned using the training and validation splits of single session pooled data. Our CNN model was able to discriminate 10 subjects with 98.7%±0.005, 99.3%±0.003, and 98.6%±0.006 accuracies for Sessions 1, 2, and 3 respectively, over 2,760 half-second epochs (i.e., 20% test split).

D. Across-Sessions Person Identification

Since within-session learning and evaluation from EEG can overperform due to its highly correlated temporal structure, across-sessions analyses would better demonstrate model generalizability. Spectral power and time-domain representation based methods present a baseline with reasonable validation set performances, which do not generalize across-sessions well (c.f. Table II). In non-adversarial models, we observe the amount of exploited session-discriminative leakage through the adversary we train alongside the CNN, without adversarial loss feedback ($\lambda = 0$). We observed that regular CNNs exploit features that can also discriminate the two days (75–80%).

Adversarial models suppress session-variant information from the encoded features. As observed from the adversary accuracies, increasing λ censors the encoder, enforces stronger session-invariance, and converges to the 50% chance level. An intuitive way of choosing λ is by cross-validating the learning process. We train our models with varying λ, and

favor decreases in adversary performance on the validation set with increasing λ, while maintaining a similar identifier performance compared to the non-adversarial case. Figure 2 depicts (for the leave-Session 2-out case) that strong λ values can force the encoder to lose person-discriminative information. Hence, we can choose λ in a range where identifier does not start to perform poorly and adversary accuracy is low (e.g., $\lambda = 0.01$ or 0.02). When tested on an independent session, we observe up to 72% 10-class person identification accuracies based on 13,800 half-second epochs, with up to 6% gains via adversarial learning based on two sessions' invariance.

V. DISCUSSION

We propose an adversarial inference approach to extend deep learning based EEG biometric identification models, to learn session-invariant person-discriminative representations. We empirically assessed our approach based on half-second EEG epochs recorded from ten subjects during BCI experiments on three different sessions. Our results demonstrate significant contribution of adversarial learning in developing across-days EEG-based person identification models.

Within-recording deep model learning and evaluation protocols are expected to perform significantly better when temporally correlated signals (e.g., EEG) are considered. Recently expanding work in EEG biometrics are mainly evaluated by these frameworks [7–10]. Yet, one recent study that evaluates CNNs in longitudinal usability yielded significant insights to this problem [29]. We address this in a similar way, while introducing adversarial learning for deep person-discriminative models to exploit session-invariant features. Overall, our approach is applicable to any EEG neural network model.

Our model currently relies on the assumption that changing sessions and days is one specific source of variability in the data distribution. This idea could well extend to task-invariant feature learning, where subjects can ideally perform any unspecified task that representations should be invariant to (e.g., naturalistic physical movements [45]). An ideal EEG person identification model would have no possibility to be calibrated or finetuned prior to use at an arbitrary time. Hence, in the light of recent progress in deep learning, we propose adversarial inference for longitudinal model robustness.

REFERENCES

[1] P. Campisi and D. La Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 782–800, 2014.

[2] S. Marcel and J. d. R. Millán, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, 2007.

[3] R. Das, E. Maiorana, and P. Campisi, "EEG biometrics using visual stimuli: a longitudinal study," *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 341–345, 2016.

[4] E. Maiorana, D. La Rocca, and P. Campisi, "On the permanence of EEG signals for biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 163–175, 2016.

[5] E. Maiorana and P. Campisi, "Longitudinal evaluation of EEG-based biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1123–1138, 2018.

[6] P. Arnau-González, M. Arevalillo-Herráez, S. Katsigiannis, and N. Ramzan, "On the influence of affect in EEG-based subject identification," *IEEE Transactions on Affective Computing*, 2018.

[7] L. Ma, J. W. Minett, T. Blu, and W. S. Wang, "Resting state EEG-based biometrics for individual identification using convolutional neural networks," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015, pp. 2848–2851.

[8] Z. Mao, W. X. Yao, and Y. Huang, "EEG-based biometric identification with deep learning," in *8th International IEEE/EMBS Conference on Neural Engineering*, 2017, pp. 609–612.

[9] X. Zhang, L. Yao, S. S. Kanhere, Y. Liu, T. Gu, and K. Chen, "MindID: Person identification from brain waves through attention-based recurrent neural network," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, p. 149, 2018.

[10] T. Wilaiprasitporn, A. Ditthapron, K. Matcharparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective EEG-based person identification using the deep learning approach," *arXiv preprint arXiv:1807.03147*, 2018.

[11] K. Das, S. Zhang, B. Giesbrecht, and M. P. Eckstein, "Using rapid visually evoked EEG activity for person identification," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 2490–2493.

[12] B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin, and S. Laszlo, "Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics," *Neurocomputing*, vol. 166, pp. 59–67, 2015.

[13] H. Touyama and M. Hirose, "Non-target photo images in oddball paradigm improve EEG-based personal identification rates," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 4118–4121.

[14] T. Koike-Akino, R. Mahajan, T. K. Marks, Y. Wang, S. Watanabe, O. Tuzel, and P. Orlik, "High-accuracy user identification using EEG biometrics," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2016, pp. 854–858.

[15] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 738–742, 2007.

[16] Q. Gui, Z. Jin, and W. Xu, "Exploring EEG-based biometrics for user identification and authentication," in *Signal Processing in Medicine and Biology Symposium*, 2014, pp. 1–6.

[17] M. DelPozo-Banos, C. M. Travieso, C. T. Weidemann, and J. B. Alonso, "EEG biometric identification: a thorough exploration of the time-frequency domain," *Journal of Neural Engineering*, vol. 12, no. 5, p. 056019, 2015.

[18] D. Vinothkumar, M. G. Kumar, A. Kumar, H. Gupta, M. Saranya, M. Sur, and H. A. Murthy, "Task-independent EEG based subject identification using auditory stimulus," in *Proc. Workshop on Speech, Music and Mind*, 2018, pp. 26–30.

[19] K. Brigham and B. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–8.

[20] S. Sun, "Multitask learning for EEG-based biometrics," in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 1–4.

[21] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433–445, 2011.

[22] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PloS One*, vol. 12, no. 2, p. e0172578, 2017.

[23] S. Stober, D. J. Cameron, and J. A. Grahn, "Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings," in *Advances in Neural Information Processing Systems*, 2014, pp. 1449–1457.

[24] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2018.

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

[26] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, 2018.

[27] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in *International Conference on Learning Representations*, 2016.

[28] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep feature learning for EEG recordings," in *International Conference on Learning Representations*, 2016.

[29] R. Das, E. Maiorana, and P. Campisi, "Motor imagery for EEG biometrics using convolutional neural network," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 2062–2066.

[30] I. Goodfellow et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[31] H. Edwards and A. Storkey, "Censoring representations with an adversary," in *International Conference on Learning Representations*, 2016.

[32] C. Louizos, K. Swersky, Y. Li, M. Welling, and R. Zemel, "The variational fair autoencoder," in *International Conference on Learning Representations*, 2016.

[33] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer et al., "Fader networks: Manipulating images by sliding attributes," in *Advances in Neural Information Processing Systems*, 2017, pp. 5967–5976.

[34] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations*, 2014.

[35] Q. Xie, Z. Dai, Y. Du, E. Hovy, and G. Neubig, "Controllable invariance through adversarial feature learning," in *Advances in Neural Information Processing Systems*, 2017, pp. 585–596.

[36] G. Louppe, M. Kagan, and K. Cranmer, "Learning to pivot with adversarial networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 981–990.

[37] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Computer Vision and Pattern Recognition*, 2017.

[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015.

[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[40] U. Orhan, K. E. Hild, D. Erdoğmuş, B. Roark, B. Oken, and M. Fried-Oken, "RSVP keyboard: An EEG based typing interface," in *International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 645–648.

[41] P. Gonzalez-Navarro, M. Moghadamfalahi, M. Akcakaya, and D. Erdoğmuş, "A kronecker product structured EEG covariance estimator for a language model assisted-BCI," in *International Conference on Augmented Cognition*, 2016, pp. 35–45.

[42] P. Gonzalez-Navarro, M. Moghadamfalahi, M. Akçakaya, and D. Erdoğmuş, "Spatio-temporal EEG models for brain interfaces," *Signal Processing*, vol. 131, pp. 333–343, 2017.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[44] S. Tokui, K. Oono, S. Hido, and J. Clayton, "Chainer: a next-generation open source framework for deep learning," in *Proceedings of Workshop on Machine Learning Systems in the Twenty-ninth Annual Conference on Neural Information Processing Systems*, vol. 5, 2015, pp. 1–6.

[45] T. Nakamura, V. Goverdovsky, and D. P. Mandic, "In-ear EEG biometrics for feasible and readily collectable real-world person authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 648–661, 2018.