Synthetic EMG Based on Adversarial Style Transfer Can Effectively Attack Biometric-Based Personal Identification Models

Peiqi Kang, *Graduate Student Member, IEEE*, Shuo Jiang, *Member, IEEE*, and Peter B. Shull, *Member, IEEE*

***Abstract*— Biometric-based personal identification models are generally considered to be accurate and secure because biological signals are too complex and person-specific to be fabricated, and EMG signals, in particular, have been used as biological identification tokens due to their high dimension and non-linearity. We investigate the possibility of effectively attacking EMG-based identification models with adversarial biological input via a novel EMG signal individual-style transformer based on a generative adversarial network and tiny leaked data segments. Since two same EMG segments do not exist in nature; the leaked data can't be used to attack the model directly or it will be easily detected. Therefore, it is necessary to extract the style with the leaked personal signals and generate the attack signals with different contents. With our proposed method and tiny leaked personal EMG fragments, numerous EMG signals with different content can be generated in that person's style. EMG hand gesture data from eighteen subjects and three well-recognized deep EMG classifiers were used to demonstrate the effectiveness of the proposed attack methods. The proposed methods achieved an average of 99.41% success rate on confusing identification models and an average of 91.51% success rate on manipulating identification models. These results demonstrate that EMG classifiers based on deep neural networks can be vulnerable to synthetic data attacks. The proof-of-concept results reveal that synthetic EMG biological signals must be considered in biological identification system design across a vast array of relevant biometric systems to ensure personal identification security for individuals and institutions.**

***Index Terms*— EMG, synthetic biological signal, generative adversarial network, identification.**

Manuscript received 6 April 2023; revised 12 June 2023 and 19 July 2023; accepted 5 August 2023. Date of publication 7 August 2023; date of current version 18 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 52250610217 and Grant 52105033, in part by the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0100, and in part by the Chenguang Program by Shanghai Municipal Education Commission under Grant 21CGA23. *(Corresponding author: Shuo Jiang.)*

Peiqi Kang and Peter B. Shull are with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: pekkykang@sjtu.edu.cn; pshull@sjtu.edu.cn).

Shuo Jiang is with the College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China, and also with the Frontiers Science Center for Intelligent Autonomous Systems, Shanghai 200120, China (e-mail: jiangshuo@tongji.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3303316

I. Introduction

ELECTROMYOGRAPHY (EMG)-based hand gesture recognition is a representative application of human-machine interface technology [1], [2]. Due to the high information related to neural activities and high time resolution, EMG-based hand gesture recognition methods have played vital roles in prosthetic control, VR/AR interaction, user verification, and identification [3], [4].

Since the EMG signals are high-dimensional and generated from complex physiological structures, the EMG signals of the same hand gesture have significant differences between individuals. These individual differences will pose technical challenges for applications, including prosthetic control, but, in turn, hold potential to be reliable tokens for personal identification. With the help of machine learning or deep learning recognition models, EMG-based user identification systems can achieve an averaged equal error rate of 1% to 4% using multi-channel EMG bands or high-dimensional EMG systems [5], [6].

These research studies show that EMG-based user identification systems are highly functional, and since biological signals are high dimensional and complex, biometric-based identification systems are usually considered to be highly secured. However, in recent years, many attack methods have been proposed and received intensive attention; these methods usually add designed small perturbations into the input data to manipulate models' output or make models confused [7]. A brief literature review on attacking biologic recognition models will also be provided in the next section. Among these models, the most notable is the deployment of generative neural networks in image synthesis [8]. These synthetic images successfully cheated the most state-of-the-art identification models and caused huge security risks. However, for attacking EMG-based identification models using synthetic EMG signals, there are few related studies, but these attacks may cause huge security problems and potentially inevitable losses to users.

Therefore, we proposed two assumptions in this paper: first, an individual's EMG style can be learned by a neural network; second, with the learned EMG style, synthetic EMG signals with the same style can be generated and used to attack EMG-based identification models. To validate these

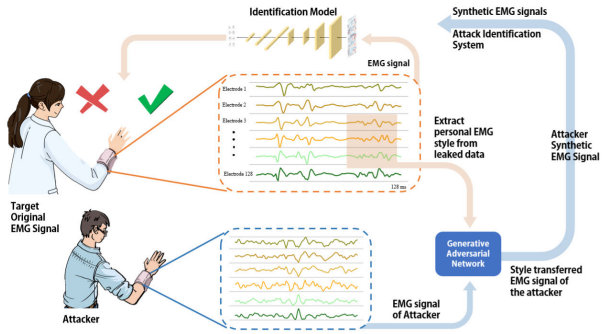

Fig. 1. Suppose that an identification model based on a deep convolutional neural network (CNN) uses EMG hand gesture signals input from a 128-channel high-density EMG armband as tokens. However, if a user's EMG signals were leaked due to various reasons, his personal EMG style can be extracted from them and used as a target style to generate synthetic EMG signals via a generative adversarial network. With these synthetic EMG signals, attackers can effectively hack the identification system and manipulate its outputs, which may cause huge risks to personal privacy and information security.

assumptions, we built an EMG signal style transformer based on the cycle generative adversarial network to learn an individual's EMG hand gesture signal style and generate other individuals' synthetic EMG hand gesture signals in the same style (Fig. 1). Experiments on three deep identification models and eight hand gestures of eighteen subjects validated that the artificial synthetic EMG signals can effectively cheat deep identification models and manipulate them to achieve the attacker's desired outcome. It is worth mentioning that, due to the high dimension and non-linearity of EMG signals, no two EMG segments should be the same; otherwise, they will be easily identified as synthetic attacks by the security system. In addition, when encountering identification systems requiring long-time sampling, few leaked data will soon run out, so attackers will be forced to use repeated data and thus be detected. Therefore, using leaked data to directly attack identification systems is not practical and deep fake data generation methods are necessary. In addition, the focus of this paper lies in exploring the possibilities of potential attacks from the biological perspective instead of from the mathematical perspective that aim at the working principle of identification models.

The contributions of this paper are as follows: (1) We proposed an attack method based on the generative strategy, which can generate style-transferred synthetic EMG hand gesture signals while keeping the content the same. (2) We demonstrated that synthetic EMG signals are capable of cheating the personal identification systems which imply that the synthetic biological signals must be taken into consideration in biological identification systems, design. Our methods hold the potential to be used to guide the design of millions of identification devices to protect people's privacy and information safety.

This paper was organized as follows: in Section II, a brief literature review on related studies is presented. In Section III, the problem formulation and structure of the proposed synthetic signal generation methods are introduced. In Section IV, experimental protocol, validation protocol, and results are introduced. Finally, in Section V, discussion on our attack methods and results is presented.

II. Background Research

A. EMG-Based User Identification

Recently, various novel identification technologies based on biometric technologies have been reported by worldwide researchers. Since the gesture-recognition-based identification technologies using EMG signals can be collected during human hand activity and are similar to the process of traditional identification methods (e.g., entering passwords by hand), they have been intensively explored by various researchers. Yamaba et al. [9] proposed the idea of utilizing EMG signals for user identification and achieved promising results using a support vector machine (SVM) classifier. Jiang et al. [10] utilized HD-sEMG signals of common daily hand gestures as identification inputs and proposed a cancelable HD-sEMG-based biometrics system to protect personal information security. Pradhan et al. [11] designed a series of experiments on the effect of different feature extraction methods and the number of channels to the EMG-based identification system and systematically investigated the performance of sixteen static wrist and hand gestures.

Compared with other biometric identification systems (e.g., gait-based [12], electrocardiograph (ECG)-based [13], and electroencephalograph (EEG)-based [14]), the EMG-based methods show advantages on high information security, high signal-to-noise ratio, high recognition accuracy, and more convenient acquisition [5].

B. Attacks on Biological Classifiers

Artificial intelligence models represented by deep learning models achieved remarkable success in various recognition tasks, but their vulnerability to interference or attacks also drew great attention [15]. There are three kinds of attack methods according to how deep the attacker can get access to the target models: white-box attacks, gray-box attacks, and black-box attacks. The black-box attacks are most practical because they only need to know the input and output of the target models. Su et al. [16] successfully cheated image classifiers by changing a single pixel of the input. Recently, the famous Open AI lab announced a simple but highly effective method called the typographic attack: simply pasting a tag with a note on the object can mislead the state-of-the-art recognition models [17]. With a similar strategy, researchers broke the state-of-the-art Face ID system with a printed sticker [18]. Attacks on the brain-computer interface (BCI) were also investigated. Zhang et al. [19] achieved effective attacks on EEG-based BCIs and proved the vulnerability of convolutional neural network (CNN) classifiers under small deliberate perturbations. Liu et al. [20] proposed a total loss minimization approach to generate universal adversarial perturbations to attack EEG-based BCIs and successfully manipulated the output of the models. Zhang et al. [21] and Bian et al. [22] conducted intensive studies on the attack on EEG-based BCI spellers, and the results showed that the BCI spellers can

be easily manipulated and may cause serious problems like medical misdiagnose.

C. Synthetic Biological Signals

The generative neural network is a kind of artificial intelligence model that can generate synthetic images or data, and the generative adversarial network (GAN) is the most representative [23]. GAN is a deep learning frame that consists of a generator and a discriminator. The generator is used to generate synthetic data, and the discriminator is used to discriminate between the synthetic data and real data, therefore helping the generator to generate more realistic synthetic data.

In addition, GAN also provides a new way to generate synthetic biological signals which were considered too complex to generate synthetically. Jiao et al. [24] utilized Wasserstein GAN generated EEG and EOG signals to expand the dataset and improve classifier performance in driver sleepiness detection. Other research also reports the contribution of Wasserstein GAN in emotion recognition [25]. Access to personal ECGs is restricted because of privacy concerns, but building automated computer-aided diagnosis systems requires vast amounts of data. Nankani et al. [26] proposed an approach for generating irregular beats (e.g., supraventricular ectopic, ventricular ectopic, and normal beats) with a conditional GAN to generate synthetic ECG data for diagnosis systems' datasets. Ding et al. [27] proposed a log-spectral matching GAN to generate PPG signals for atrial fibrillation detection.

Generated synthetic EMG data also received attention from the neuroscience community. Anicet et al. [28] utilized deep convolutional generative adversarial networks and style transfer to generate Parkinson's disease EMG signals to reduce the displeasure and pain of patients to collect lots of data. Campbell et al. [29] validated the feasibility of generating EMG signals with hand motions information using a deep generative model called sinGAN. Bird et al. [30] utilized a generative model called the generative pre-trained transformer to generate synthetic EMG signals of three gestures, including hand open, hand closed, and at rest, and demonstrated the synthetic data in a prosthetic hand control application.

To our best knowledge, there were no studies focused on EMG style transferring between individuals using generative methods or studies that tried to disclose the security risk of synthetic data in EMG-based identification systems. This work indicated that the resistance to synthetic biological signals and data protection or encryption must be taken into consideration in biological identification systems design. We hope our work can draw researchers' attention to the potential threat of synthetic biological signals to personal information security and property security.

III. Methods

A. Problem Formulation and Overview

Suppose a biometric-based identification model can accurately recognize a type of EMG hand gesture signals as subject A:

$$Y_A = ID(X_a) \tag{1}$$

TABLE I
Generator Structure

Layer	Details
Input	
ReflectionPad2d	
Conv2d	(64, (7x7), stride=1, padding=3)
InstanceNorm2d	64
ReLU	
Conv2d	(128, (3x3), stride=2, padding=1)
InstanceNorm2d	128
ReLU	
Conv2d	(128, (3x3), stride=2, padding=1)
InstanceNorm2d	128
ReLU	
Residual Blocks	$\times 4$
Upsample	scale factor=2
Conv2d	(64, (3x3), stride=1, padding=1)
InstanceNorm2d	64
ReLU	
Upsample	scale factor=2
Conv2d	(64, (3x3), stride=1, padding=1)
InstanceNorm2d	64
ReLU	
ReflectionPad2d	
Conv2d	(1, (7x7))
Tanh	
Output	

where ID is the identification model, X_a is the data input from subject A, and Y_A is the identification result of the identification model. Due to the individual differences in EMG signals between two individuals, if the same model received signals from other subjects, for instance, attacker B, then the system will recognize this subject is not A.

However, if subject A's EMG data was leaked due to various reasons, this data might be used to attack the identification model. With this leaked data, we proposed a generative adversarial EMG signal transformer to generate synthetic EMG signals to manipulate the identification model's outputs:

$$Y_A = ID(T(X_b)) \tag{2}$$

where T is the generative adversarial EMG signal transformer we proposed. and X_b is the data input from attacker B. In a real-life scenario, if the transformer model was embedded into an identification system, by transferring the input data into synthetic data, the attacker could manipulate the identification system's outputs in real time.

Our proposed signal transformer's network architecture (Table. I) from top to bottom is a reflection padding layer, three convolutional layers (kernel sizes are seven, three, three), four residual blocks, two convolutional layers (kernel size is three) with an up-sample function (scale factor is two), a reflection padding layer, and a convolutional layer (kernel size is seven) (Fig. 3). The input is 128 ms of EMG signal of 128 channels, and the output size is the same.

B. EMG Signal Transformer Training Process

Since two same EMG signal segments do not exist, it is necessary to extract the style with the leaked personal signals and generate the attack signals with different contents. The attack signals are generated by the transformer with attacker subject B's data. During the transformer's training process,

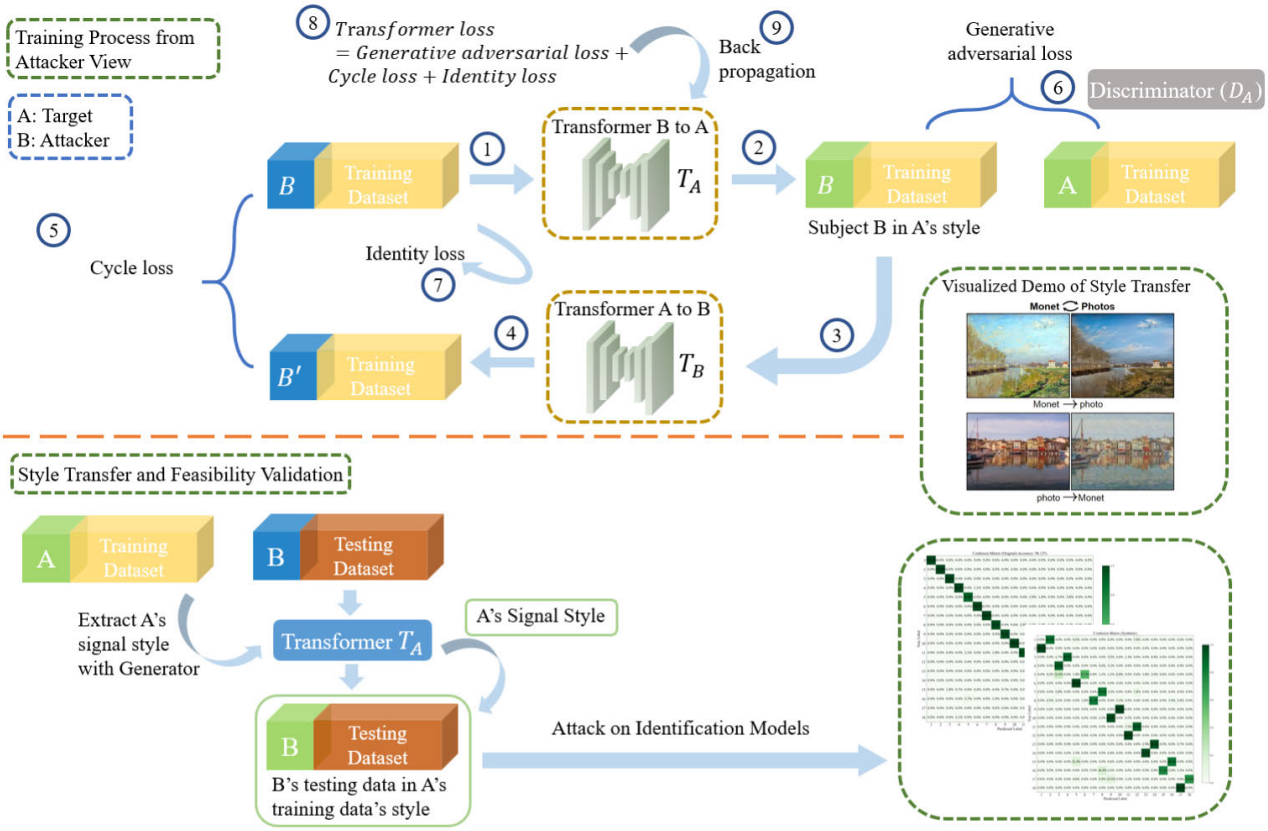

Fig. 2. The training process of the proposed signal transformer from B's (attacker) view. Suppose there are two subjects: A and B. To train a signal transformer that can generate subject B's EMG signal in subject A's style, the transformer should be optimized in the gradient descent direction of the sum of the generative adversarial loss, cycle consistency loss, and identity loss. The aim and function of the three kinds of loss are defined and introduced in the Methods section. The figure on the right shows a visualized demo of the style transfer algorithm. After the training is finished, new data from B can be transferred into A's style and be used in the attack on personal identification models.

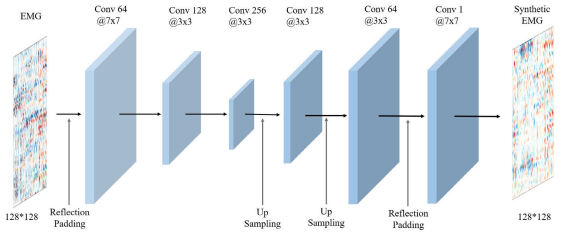

Fig. 3. The network architecture of the transformer. The input is a 128 ms of EMG signal of 128 channels, and the output is the same size synthetic EMG signal.

three kinds of loss were used to guide the optimization direction of the parameters of the transformer.

First, to improve the quality of the synthetic signal, we utilized the idea of generative adversarial networks and designed a discriminator (Table. II) for the training of the signal transformer [31]. The network architecture of the discriminator consists of four convolutional layers (kernel size is four), a zero-padding layer, and a convolutional layer (kernel size is four). During the training process, the synthetic signals generated by the transformer will be scored by the discriminator. The discriminator will give higher scores to synthetic signals that are similar to the leaked data from subject A and give lower scores to synthetic signals that are less similar to the leaked data from subject A. During the training process, the transformer and discriminator will both be trained, and therefore, can continually force the transformer to generate

TABLE II
DISCRIMINATOR STRUCTURE

Layer	Details
Input	
Conv2d	(64, (4x4), stride=2, padding=1)
LeakyReLU	(0.2)
Conv2d	(128, (4x4), stride=2, padding=1)
InstanceNorm2d	128
LeakyReLU	(0.2)
Conv2d	(256, (4x4), stride=2, padding=1)
InstanceNorm2d	256
LeakyReLU	(0.2)
Conv2d	(512, (4x4), padding=1)
InstanceNorm2d	512
LeakyReLU	(0.2)
ZeroPad2d	(1,0,1,0)
Conv2d	(1, (4x4), padding=1)
Output	

high-quality synthetic signals. The loss used in this process was defined as the generative adversarial loss:

$$\begin{aligned} &Loss_{GA} \\ &= L_{GA}(T_B, D_B, A, B) \\ &\quad + L_{GA}(T_A, D_A, A, B) \\ &= E_{b\sim p_{data}(b)}[log D_B(b)] \\ &\quad + E_{a\sim p_{data}(a)}[1 - log D_B T_B((a))] \\ &\quad + E_{a\sim p_{data}(a)}[log D_A(a)] \\ &\quad + E_{b\sim p_{data}(b)}[1 - log D_A T_A((b))] \end{aligned} \tag{3}$$

where A is the subject whose data was leaked, B is the attacker, a and b are data samples from A or B, T_A is the transformer that transform B's data into A's style, T_B is the transformer that transform A's data into B's style, D_A and D_B are discriminators that calculate the difference between real data and synthetic data, E is the expected value, and $a \sim p_{data}(a)$ and $b \sim p_{data}(b)$ represent data distribution.

Second, in order to transform the data style and keep the content unchanged, we adapted ideas from sentence translation and cycle generative adversarial networks to add an additional constraint condition into the training process of the signal transformer [32]. In sentence translation, the goal is to keep the original meaning but translate the sentence into a different language. To make sure the translated meaning is correct, translators always adopt a cycle strategy that translates the translated sentence back to the original language and compares its meaning with the original untranslated sentence, and tries to make the difference as small as possible. We defined this type of loss as cycle loss. The aim of the $Loss_{cycle}$ is to force the transformer to generate synthetic data only different from the real data in the style aspect and was defined by:

$$\begin{aligned} Loss_{cycle} = {} & E_{a\sim p_{data}(a)}[\|T_A T_B((a)) - a\|_1] \\ & + E_{b\sim p_{data}(b)}[\|T_B T_A((b)) - b\|_1] \end{aligned} \tag{4}$$

Third, due to the high dimensionality and nonlinearity of EMG signals, no two EMG segments should be the same; otherwise, they will easily be identified as synthetic attacks by the security system. To prevent the transformer from just cheating the discriminator by directly outputting subject A's leaked data as the transformed B's synthetic signal, in which case the transformation process would be meaningless, we utilized an identity loss as an additional constraint. The working principle of the $Loss_{identity}$ is that, if a transformer can generate data in A style and if input data is already in A style, then the synthetic data should stay the same:

$$\begin{aligned} Loss_{identity} = {} & E_{a\sim p_{data}(a)}[\|T_A(a) - a\|_1] \\ & + E_{b\sim p_{data}(b)}[\|T_B(b) - b\|_1] \end{aligned} \tag{5}$$

Therefore, the loss of the training process of the generative adversarial EMG signal transformer was given by:

$$Loss = Loss_{GA} + Loss_{cycle} + Loss_{identity} \tag{6}$$

Moreover, since the training of the transformer and discriminator is adversarial, when training the transformer, parameters of D_A and D_B should be frozen, and when training the discriminator, parameters of T_A, T_B and one of the two discriminators should be frozen. Therefore, the optimization objective of the transformer was:

$$\begin{aligned} & Min \quad Loss_{GA} \\ & = Min \quad E_{a\sim p_{data}(a)}[1 - log D_B G_B((a))] \\ & \quad + Min \quad E_{b\sim p_{data}(b)}[1 - log D_A G_A((b)) \\ & = Max \quad E_{a\sim p_{data}(a)}[log D_B G_B((a))] \\ & \quad + Max \quad E_{b\sim p_{data}(b)}[log D_A G_A((b))] \end{aligned} \tag{7}$$

This optimization objective means optimizing parameters of T_A and T_B to get higher scores from D_A and D_B and to force T_A and T_B to generate more realistic synthetic data.

In addition, the discriminator should try to prioritize real data (maximize $D_A(a)$) and try to minimize synthetic data (minimize $D_A(G_a(b))$). Therefore, the optimization objective of one discriminator (e.g., D_A) was:

$$\begin{aligned} Max \quad Loss_{GA} & = Max \quad GA(T_A, D_A, A, B) \\ & = Max \quad E_{a\sim p_{data}(a)}[log D_A(a)] \\ & \quad + Max \quad E_{b\sim p_{data}(b)}[1 - log D_A T_A((b))] \\ & = Max \quad E_{a\sim p_{data}(a)}[log D_A(a)] \\ & \quad + Min \quad E_{b\sim p_{data}(b)}[log D_A T_A((b))] \end{aligned} \tag{8}$$

After the transformer was trained, new data from attacker B can be transformed into subject A's style, and in the next section, we will validate the feasibility of attacking the identification system with the synthetic subject A's data.

IV. Experiments and Results

A. Dataset and Experimental Protocol

CapgMyo dataset A was chosen to validate our proposed methods [33]. The CapgMyo A is a public dataset consisting of eight hand gestures' HD-EMG records (8×16 electrode array) of 18 participants. Eight hand gestures were included in the dataset: 1. Thumb up (TU); 2. Extension of index and middle, flexion of the others (EIM); 3. Flexion of the ring and little finger, the extension of the others (FRL); 4. Thumb opposing base of the little finger (TO); 5. Abduction of all fingers (AA); 6. Fingers flexed together in the first (FF); 7. Pointing index (PI); 8. Adduction of extended fingers (AE). The signals were sampled at 1,000 Hz, filtered with a band-pass filter at 20-380Hz, and normalized to the [-1, 1] range, corresponding to the voltage of [-2.5 mV, 2.5 mV]. Each gesture of each subject had ten trials. To ensure the reliability of experiment results, the odd-numbered trials of each subject were used as the training dataset and the even-numbered trials of each subject were used as the testing dataset.

B. Identification Models and Baseline Benchmark

It is vital to select powerful identification models as attack targets; otherwise, the results can't prove the effectiveness of the attack methods because good attack results might be caused by weak baseline models. Therefore, for the reliability of our proposed methods, we selected three well-recognized deep EMG classification models to use as identification models and target with attacks: GengNet [33], EMGNet [34], and VGG16Net [35]. These models have been validated on multiple datasets and the exact parameters or structure of these models can be found in the above references. We followed these references and rebuild these models.

To make sure that our result is robust to different hand gestures and that individual differences are the sole variable in the experiment, for each gesture, we trained three different structured identification models. If not specifically mentioned, all the results present in the following are the mean values of all eight gestures.

In total, we trained 24 identification models, each eight of them belonging to one model structure, and each one of them is trained and tested on one gesture category with 18 subjects'

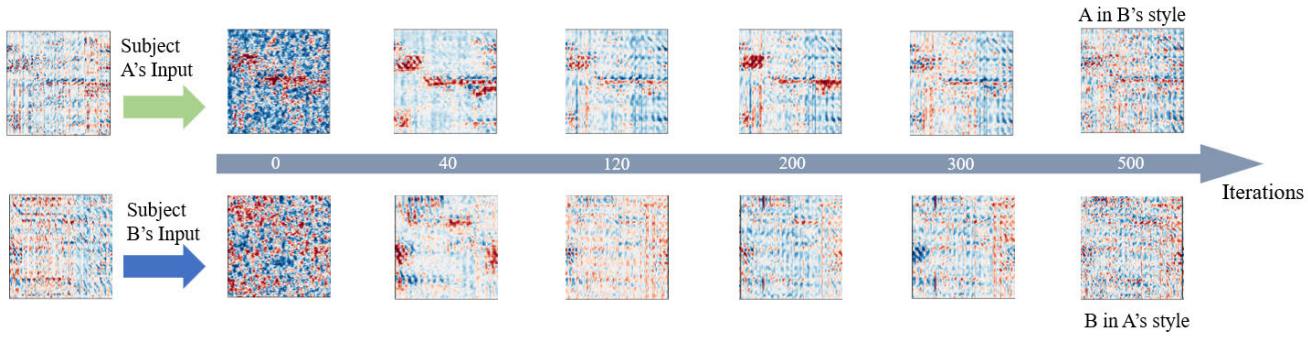

Fig. 4. Evolution of the synthetic signals during the training process. When the training started, the synthetic signals were basically irregular noises. As the training proceeded, the synthetic signals learned features from the original signals but still had major defects in some areas. After a certain number of iterations, the synthetic signals became indistinguishable from the original signals, and the value of the loss function almost stopped decreasing at which point the training could be considered finished.

data. Each model's input data consists of a specific gesture's data performed by 18 individuals, the output is the labels representing individuals (18-class classification). The training dataset is the odd-numbered trials' data of 18 individuals' one certain gesture. In the training dataset, we separated 10% of the data to act as the validation dataset. The testing dataset is the even-numbered trials' data of one certain gesture of 18 individuals and the labels are the subject numbers. This experimental paradigm is to make sure that, in each trial, individual differences are the sole variable in the experiment and prove that the proposed style transfer can learn and reproduce these differences.

To evaluate the identification performance, we utilized the rank-k identification rate as the evaluation metric and generated the cumulative match characteristic (CMC) curve. These results demonstrated that our models exhibited excellent individual recognition capability. The rank-k index is a performance metric commonly used to evaluate identification or recognition tasks. It measures the accuracy of identifying the correct label within the top-k-ranked predictions. In the context of this study, the Rank-k index assesses the ability of the model to correctly identify the intended person among the top-k predictions. As the commonly used metrics, the values of rank-1 and rank-5 were also listed, and the confusion matrixes were provided. These evaluation metrics are representative and well-recognized quantification methods for evaluating biometric-based identification models.

The test results of three identification models were shown as follows: the rank-1 and rank-5 results showed that our models have strong identification ability (Table. III), the CMC curve showed that identification systems based on these models will have reliable identification ability in a real-life scenario (Fig. 6), and the confusion matrixes showed that our identification models have good robustness among different subjects (Fig. 5). These results demonstrated that our identification models are powerful and qualified to be attack targets.

C. Synthetic Signal Generation and Attack on Identification Models

Due to the characteristics of EMG signals, two same EMG segments do not exist in nature, so it is necessary to extract the style with the leaked personal signals and generate the attack signals with different contents. The leaked data can't be used to attack the model directly, Otherwise, it will be easily detected. For better result presentation, two adjacent individuals were paired, with one acting as the data leaker and the other as the attacker, to validate the proposed EMG signal transform attack methods. Therefore, nine pairs (eighteen individuals) were tested for each experimental condition. For better understanding, odd-numbered subjects were named subject A, and even-numbered subjects were named subject B. Worth mentioning this pair setup is only for better result presentation and is not necessary for real applications.

TABLE III
RANK-1 AND RANK-5

Gesture	GengNet		EMGNet		VGG16Net	
	Rank-1	Rank-5	Rank-1	Rank-5	Rank-1	Rank-5
TU	99.21	100	86.98	98.89	93.81	100
EIM	96.03	100	88.25	98.25	94.76	98.73
FRL	98.57	100	93.34	99.83	91.27	97.78
TO	97.78	100	91.43	99.05	96.98	99.37
AA	98.41	99.84	89.37	97.78	93.02	99.21
FFF	97.30	100	93.33	100	96.03	100
PI	99.37	100	95.71	100	95.40	99.84
AE	98.25	100	92.86	99.05	96.83	99.52

During the training process, 90% of odd-numbered trials' data of subjects A and B were used as training data for the transformers, and the other 10% of odd-numbered trials' data were used as validation data to calculate the loss of the transformers. The training process repeated until the validation loss met the requirement. After the training was finished, we used subject A's and subject B's even-numbered trials' data separately to generate subject A's EMG signals in B's style and subject B's EMG signals in A's style. We also visualized the signal generation result of transformers at different training stages as shown in Fig. 4.

After the synthetic signals were generated, we tested the identification models with synthetic EMG signals (style transferred between neighboring subjects) to simulate potential attacks. Results showed that the artificial synthetic EMG signals can efficiently cheat the three kinds of identification models in all eight gestures and mislead them to make

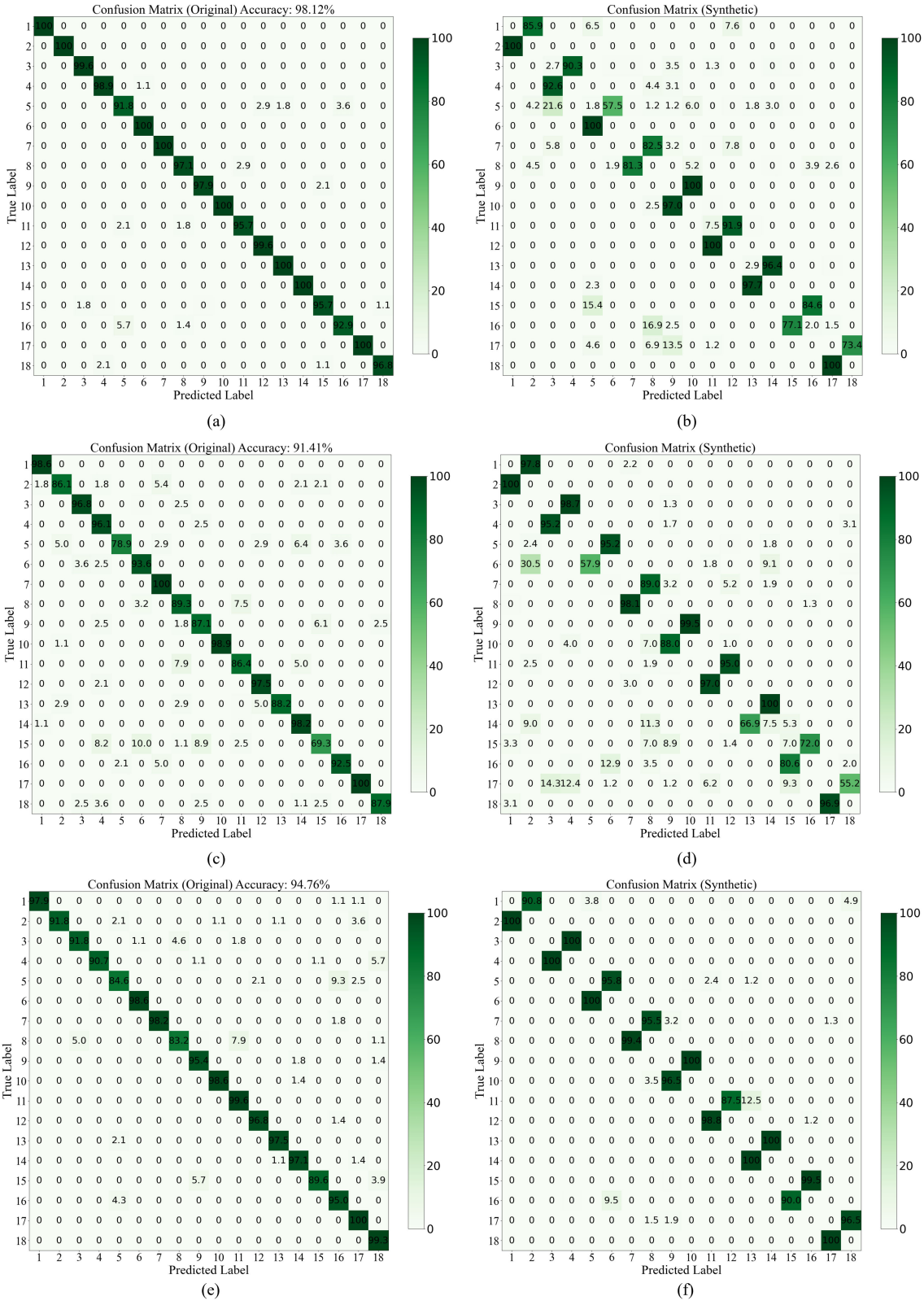

Fig. 5. The data were averaged across eight hand gestures, and the user identification accuracy of three different models was evaluated using either real signals or synthetic data as input. In the case of across-subject validation for hand gesture recognition, the average recognition accuracy of the three models was found to be 42.86%. (a) Confusion matrix for eighteen subjects' identification when testing on original data using the GengNet. (b) Confusion matrix for eighteen subjects' identification when testing on synthetic data using the GengNet. (c) Confusion matrix for eighteen subjects' identification when testing on original data using the EMGNet. (d) Confusion matrix for eighteen subjects' identification when testing on synthetic data using the EMGNet. (e) Confusion matrix for eighteen subjects' identification when testing on original data using the VGG16Net. (f) Confusion matrix for eighteen subjects' identification when testing on synthetic data using the VGG16Net. (a), (c), and (e) confusion matrices aim to validate the strong identification ability of the GengNet, EMGNet, and VGG16Net. (b), (d), and (f) confusion matrices aim to demonstrate the effectiveness of attacking GengNet, EMGNet, and VGG16Net with synthetic data.

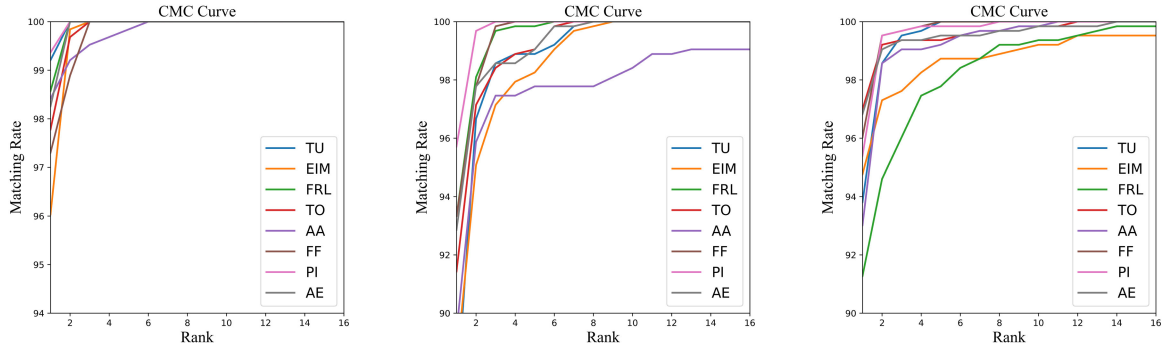

Fig. 6. The data are averaged across eighteen subjects. These figures showed that our identification models are powerful and qualified to be attack targets. (a) CMC curves for each gesture of the GengNet-based identification model. (b) CMC curves for each gesture of the EMGNet-based identification model. (c) CMC curves for each gesture of the VGG16Net-based identification model.

TABLE IV
HIT RATE AND CONFUSION RATE

Gesture	GengNet		EMGNet		VGG16Net	
	Hit	Confusion	Hit	Confusion	Hit	Confusion
TU	99.20	99.61	94.16	100	97.27	100
EIM	93.82	100	89.58	99.31	100	100
FRL	83.84	100	88.04	96.43	96.61	100
TO	98.29	100	82.66	100	98.81	100
AA	76.62	97.22	88.43	97.42	99.37	100
FFF	80.88	97.23	90.89	100	95.67	100
PI	95.66	100	84.74	98.77	95.64	100
AE	98.84	100	90.68	100	99.38	100

The data are averaged across eighteen subjects, and all the values are percentages.

judgments as we designed. To quantifiably evaluate the proposed attack methods, we defined two evaluation metrics called hit rate and confusion rate, which represent the success rate of using synthetic EMG signals on attacking identification models and the chance of identification models being confused by the synthetic EMG signals:

$$HitRate = AttackerWantedOutputs/TotalAttacks \quad (9)$$

$$ConfusionRate = WrongOutputs/TotalAttacks \quad (10)$$

Wrong outputs included outputs that were what the attacker wanted and outputs that were not what the attacker wanted but were still not correct.

Under attacks of the synthetic EMG signals, identification models' rank-1 and rank-5 evaluation metrics dropped close to zero, which means that these identification models were disabled. The hit rate of the attack methods on three different identification models (GengNet, EMGNet, and VGG16Net) were 89.34%, 87.94%, and 97.24%, respectively. The confusion rate of the attack methods on three different identification models (GengNet, EMGNet, and VGG16Net) were 99.06%, 99.17%, and 100%, respectively. These results showed that our attack methods can effectively manipulate and confuse the state-of-the-art identification models (Table IV).

Using the same training and testing conditions, of all the models the GengNet had the strongest identification ability. Compared with the other two identification models, the GengNet has two conventional layers whose kernel size is one and much fewer channel numbers. This result shows that with a smaller kernel size identification models can obtain stronger recognition ability with fewer channels. However, smaller kernel-size models will require a significantly longer time and more computing resources to train. Therefore, if the accuracy has already met the requirement, a lightweight model might also be an option. In addition, results also showed that the anti-attack ability has no direct relationship with the identification ability. Three models with different structures show no significant difference (Tukey's test based on analysis of variance, $p < 0.05$) in confusion rate, which indicated that our attack methods have strong universality.

V. DISCUSSION

In this paper, we proposed a novel personal identification model attack method based on the EMG signal transformer. Experiments on eighteen subjects and three strong identification models proved its feasibility and reliability. Our methods achieved an average of 99.41% success rate on confusing identification models and an average of 91.51% success rate on manipulating identification models. These results demonstrate that our methods hold the potential to be used to guide the design of millions of identification devices to protect people's privacy and information safety.

Compared with other popular GAN methods including PairGAN or WGAN, our method demonstrates practical advantages. PairGAN lacks the cycle loss and identification loss employed in our method. Consequently, it fails to ensure content preservation during the signal style transfer process, making it unsuitable for attacking a biometric identification system. On the other hand, despite WGAN achieving low training losses, the generated signal quality remains inadequate. This drawback arises from the absence of the identification loss utilized in our method, which causes the generator to adopt a cheating-like strategy to deceive the discriminator, leading to a rapid decrease in loss without emphasizing the actual quality of the generated signals. Therefore, our method proves to be more suitable for this specific task.

Although there are limited research works in related EMG-attack fields, the attack on EEG systems has gained attention from researchers. Previous studies have explored methods such

as contaminating the original signals by adding attack elements (e.g., square wave signals [22] or narrow period pulses [36]) or using adversarial attacks [19]to target brain-machine interfaces. These approaches have successfully deceived the models of brain-machine interfaces in tasks like motion image classification, driver fatigue estimation, or spelling tasks, revealing the vulnerability of biometric identification models [21]. In comparison to these works, our generative approach does not require fine-tuning for different target systems or prior knowledge of the target. Therefore, it exhibits more universality. Additionally, our generated attack signals possess greater diversity, enhancing their stealthiness in attacks.

We also conducted experiments with varying channel numbers, including 8, 16, 32, 64, and 128 channels. The results revealed a decrease in model recognition accuracy as the number of channels decreased. As the recognition model itself becomes less accurate (98.12% to 77.96%), the ability of the attack signal to manipulate the recognition model also diminishes (89.34% to 48.96%). However, the attack signal still proves effective in confusing the recognition model. In addition, the phenomenon of different models exhibiting varying resistance to attacks can be attributed to their different generalization abilities. Models with high generalization ability tend to have a lower hit rate, while models with low generalization ability exhibit a higher hit rate. Therefore, we believe that enhancing the generalization ability of identification models may improve their ability to resist targeted attacks, although they may still be susceptible to confusion. These results suggest that in order to enhance the anti-attack capability of identification systems, it is necessary not only to develop more robust identification models but also to consider the identification system from a broader perspective.

Applying advanced biometric methods such as brain stripe recognition technology can be an effective way to enhance the anti-attack capability. Brain stripe recognition technology offers unique advantages, including resistance to theft, forgery, and damage. Moreover, it requires in vivo detection, making it a more secure biometric identification method for identity recognition. It can be considered as a highly secure next-generation password. In addition, implementing advanced security strategies such as cancelable algorithms [37] or user-tailored algorithms [38] can also enhance the security of biometric identification systems. Employing more complex encoding strategies or adopting sensor fusion technologies may also yield better results in terms of security. Furthermore, it is crucial to protect the original data from leakage. Some studies have proposed using encrypted data to improve data security, highlighting the importance of safeguarding sensitive information.

For style-transferred signal generating, generative models are height unconstrained and very difficult to train. During the training process, the loss will randomly uprush, and the time consumption is also larger than training a normal neural network. In addition, the signal transformer has multiple solutions, and any transformers that meet loss requirements can be used. Therefore, setting an appropriate loss threshold will accelerate the training process. However, there is no clear standard on the loss value. A recommended method is to visualize synthetic data during the training process and choose an appropriate loss threshold with multiple experiments. Additionally, the training approach of the generator network is worth discussing. There are two approaches: one involves using a single network to generate signals of different categories, while the other involves setting up multiple networks, with each network responsible for generating data of one category. The first approach is more straightforward, but the latter typically yields better generation results. Therefore, using novel generative methods and combining the advantages of both strategies (e.g., conditional diffusion models) may potentially lead to improved results.

There exist certain limitations in this paper as the field of generative models is still rapidly evolving. Firstly, the quality of the generated signals may not be optimal. However, due to the requirement of transferring signal styles while keeping the content unchanged, we did not make significant adjustments to the structure of the generative models. In the future, it would be necessary to develop higher-quality generative models that can also control the content. Additionally, generative models require substantial computational resources, and optimizing their use in real-time systems on edge devices is a challenge. Further research is needed to explore how to reduce computational requirements while ensuring high-quality generation. Moreover, as this work serves as an initial proof-of-concept study, its stability and robustness need further investigation. For example, the influence of population attributes such as gender, age, and health status on the experimental outcomes requires study. Lastly, the impact of additional security measures or specially designed models with specific training strategies to maximize user-specific information needs further exploration.

Our proposed EMG signal transferring and generating methods can also be applied in other areas. Since our methods can significantly reduce the individual difference between subjects, it can also be applied in hand gesture recognition or prosthetic control to improve the robustness and recognition accuracy. It can also provide a new solution to improve the capability of the human-computer interface apart from self-adaptation algorithms, transfer learning algorithms, and sensor fusion algorithms.

VI. Conclusion

This paper is the first work that focused on EMG hand gesture style transferring between individuals and the potential risk of synthetic data in personal identification systems. Rigorous experiments proved the feasibility of the cycle generative adversarial network based EMG style transformer and the vulnerability of deep EMG classifiers. The result demonstrated that synthetic biological signals must be taken into consideration in any biological identification system design, and the protection of biological signals, including EMG signals, is vital for personal privacy and security.

Acknowledgment

https://github.com/pekkykang/SyntheticEMGAttackBiometricPersonalId

REFERENCES

[1] S. Jiang, P. Kang, X. Song, B. P. L. Lo, and P. B. Shull, "Emerging wearable interfaces and algorithms for hand gesture recognition: A survey," *IEEE Rev. Biomed. Eng.*, vol. 15, pp. 85–102, 2022.

[2] P. Kang, J. Li, S. Jiang, and P. B. Shull, "Reduce system redundancy and optimize sensor disposition for EMG–IMU multimodal fusion human–machine interfaces with XAI," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023.

[3] D. Li, P. Kang, K. Zhu, J. Li, and P. B. Shull, "Feasibility of wearable PPG for simultaneous hand gesture and force level classification," *IEEE Sensors J.*, vol. 23, no. 6, pp. 6008–6017, Mar. 2023.

[4] P. Kang, J. Li, B. Fan, S. Jiang, and P. B. Shull, "Wrist-Worn hand gesture recognition while walking via transfer learning," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 952–961, Mar. 2022.

[5] J. He and N. Jiang, "Biometric from surface electromyogram (sEMG): Feasibility of user verification and identification based on gesture recognition," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 58, Feb. 2020.

[6] X. Jiang et al., "Neuromuscular password-based user authentication," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2641–2652, Apr. 2021.

[7] N. Papernot, P. McDaniel, and I. Goodfellow, "Transferability in machine learning: From phenomena to black-box attacks using adversarial samples," 2016, *arXiv:1605.07277*.

[8] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[9] H. Yamaba et al., "On applying support vector machines to a user authentication method using surface electromyogram signals," *Artif. Life Robot.*, vol. 23, no. 1, pp. 87–93, Mar. 2018.

[10] X. Jiang et al., "Cancelable HD-sEMG-based biometrics for cross-application discrepant personal identification," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1070–1079, Apr. 2021.

[11] A. Pradhan, J. He, and N. Jiang, "Performance optimization of surface electromyography based biometric sensing system for both verification and identification," *IEEE Sensors J.*, vol. 21, no. 19, pp. 21718–21729, Oct. 2021.

[12] Y. Sun and B. Lo, "An artificial neural network framework for gait-based biometrics," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 987–998, May 2019.

[13] S. Gutta and Q. Cheng, "Joint feature extraction and classifier design for ECG-based biometric recognition," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 2, pp. 460–468, Mar. 2016.

[14] D. L. Rocca et al., "Human brain distinctiveness based on EEG spectral coherence connectivity," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 9, pp. 2406–2412, Sep. 2014.

[15] C. Szegedy et al., "Intriguing properties of neural networks," 2013, *arXiv:1312.6199*.

[16] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Trans. Evol. Comput.*, vol. 23, no. 5, pp. 828–841, Oct. 2019.

[17] D. A. Noever and S. E. M. Noever, "Reading Isn't believing: Adversarial attacks on multi-modal neurons," 2021, *arXiv:2103.10480*.

[18] S. Komkov and A. Petiushko, "AdvHat: Real-world adversarial attack on ArcFace face ID system," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 819–826.

[19] X. Zhang and D. Wu, "On the vulnerability of CNN classifiers in EEG-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 814–825, May 2019.

[20] Z. Liu, L. Meng, X. Zhang, W. Fang, and D. Wu, "Universal adversarial perturbations for CNN classifiers in EEG-based BCIs," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 0460a4.

[21] X. Zhang et al., "Tiny noise, big mistakes: Adversarial perturbations induce errors in brain–computer interface spellers," *Nat. Sci. Rev.*, vol. 8, no. 4, Apr. 2021, Art. no. nwaa233.

[22] R. Bian, L. Meng, and D. Wu, "SSVEP-based brain-computer interfaces are vulnerable to square wave attacks," *Sci. China Inf. Sci.*, vol. 65, no. 4, Apr. 2022, Art. no. 140406.

[23] 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.

[24] 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.

[25] 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.

[26] D. Nankani and R. D. Baruah, "Improved diagnostic performance of arrhythmia classification using conditional GAN augmented heartbeats," in *Generative Adversarial Learning: Architectures Applications*. Cham, Switzerland: Springer, 2022, pp. 275–304.

[27] C. Ding et al., "Log-spectral matching GAN: PPG-based atrial fibrillation detection can be enhanced by GAN-based data augmentation with integration of spectral loss," 2021, *arXiv:2108.05272*.

[28] R. A. Zanini and E. Luna Colombini, "Parkinson's disease EMG data augmentation and simulation with DCGANs and style transfer," *Sensors*, vol. 20, no. 9, p. 2605, May 2020.

[29] E. Campbell, J. A. D. Cameron, and E. Scheme, "Feasibility of data-driven EMG signal generation using a deep generative model," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 3755–3758.

[30] J. J. Bird, M. Pritchard, A. Fratini, A. Ekárt, and D. R. Faria, "Synthetic biological signals machine-generated by GPT-2 improve the classification of EEG and EMG through data augmentation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 3498–3504, Apr. 2021.

[31] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," 2016, *arXiv:1605.09782*.

[32] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[33] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, Nov. 2016, Art. no. 36571.

[34] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, p. 672, Jan. 2020.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[36] L. Meng et al., "EEG-based brain–computer interfaces are vulnerable to backdoor attacks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2224–2234, 2023.

[37] X. Jiang et al., "Enhancing IoT security via cancelable HD-sEMG-based biometric authentication password, encoded by gesture," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16535–16547, Nov. 2021.

[38] L. Meng et al., "User-tailored hand gesture recognition system for wearable prosthesis and armband based on surface electromyogram," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.