DeepKey: A Multimodal Biometric Authentication System via Deep Decoding Gaits and Brainwaves

XIANG ZHANG, LINA YAO, and CHAORAN HUANG, University of New South Wales, Australia
TAO GU, RMIT University, Australia
ZHENG YANG, Tsinghua University, China
YUNHAO LIU, Michigan State University, USA

Biometric authentication involves various technologies to identify individuals by exploiting their unique, measurable physiological and behavioral characteristics. However, traditional biometric authentication systems (e.g., face recognition, iris, retina, voice, and fingerprint) are at increasing risks of being tricked by biometric tools such as anti-surveillance masks, contact lenses, vocoder, or fingerprint films. In this article, we design a multimodal biometric authentication system named DeepKey, which uses both Electroencephalography (EEG) and gait signals to better protect against such risk. DeepKey consists of two key components: an Invalid ID Filter Model to block unauthorized subjects, and an identification model based on attention-based Recurrent Neural Network (RNN) to identify a subject's EEG IDs and gait IDs in parallel. The subject can only be granted access while all the components produce consistent affirmations to match the user's proclaimed identity. We implement DeepKey with a live deployment in our university and conduct extensive empirical experiments to study its technical feasibility in practice. DeepKey achieves the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) of 0 and 1.0%, respectively. The preliminary results demonstrate that DeepKey is feasible, shows consistent superior performance compared to a set of methods, and has the potential to be applied to the authentication deployment in real-world settings.

CCS Concepts: • **Human-centered computing** → *Ubiquitous and mobile computing*; • **Theory of computation** → *Design and analysis of algorithms*;

Additional Key Words and Phrases: EEG (Electroencephalography), gait, biometric authentication, multimodal, deep learning

ACM Reference format:
Xiang Zhang, Lina Yao, Chaoran Huang, Tao Gu, Zheng Yang, and Yunhao Liu. 2020. DeepKey: A Multimodal Biometric Authentication System via Deep Decoding Gaits and Brainwaves. *ACM Trans. Intell. Syst. Technol.* 11, 4, Article 49 (May 2020), 24 pages.
https://doi.org/10.1145/3393619

Authors' addresses: X. Zhang, L. Yao, and C. Huang, University of New South Wales, CSE, K17, Barker St, UNSW, Sydney, NSW, 2052, Australia; emails: xiang.zhang3@student.unsw.edu.au, {lina.yao, chaoran.huang}@unsw.edu.au; T. Gu, RMIT University, 124 La Trobe St, RMIT University, Melbourne, VIC, 3001, Australia; email: tao.gu@rmit.edu.au; Z. Yang, Tsinghua University, 30 Shuangqing Rd, School of Software, Tsinghua University, Beijing, 100084, China; email: hmilyyz@gmail.com; Y. Liu, Michigan State University, 220 Trowbridge Rd, East Lansing, MI 48824, USA; email: yunhao@cse.msu.edu.

2157-6904/2020/05-ART49 $15.00
https://doi.org/10.1145/3393619

1 INTRODUCTION

Over the past decade, biometric authentication systems have gained popularity due to their reliability and adaptability. Existing biometric authentication systems generally include physiological and behavioral ones. The former is based on an individual's unique intrinsic features (e.g., face [14], iris [27], retina [38], voice [15], and fingerprint [43]) and the latter is based on an individual's behavior patterns such as gait analysis [5]. Recently, biometrics- (e.g., fingerprint, face) based authentication systems face an increasing threat of being deceived as a result of the rapid development of the manufacturing industry and technologies. For example, individuals can easily trick a fingerprint-based authentication system by using a fake fingerprint film[1] or an expensive face recognition–based authentication system by simply wearing a two-hundred-dollar anti-surveillance mask.[2] Thus, fake-resistance characteristics are becoming a more significant requirement for any authentication system. To address the aforementioned issues, EEG (Electroencephalography) signal–based cognitive biometrics and gait-based systems have been attracting increasing attention.

EEG signal–based authentication systems are an emerging approach in physiological biometrics. EEG signals measure the brain's response and record the electromagnetic, invisible, and untouchable electrical neural oscillations. Many research efforts have been made on EEG-based biometric authentication for the uniqueness and reliability. EEG data are unique for each person and almost impossible to be cloned and duplicated. Therefore, an EEG-based authentication system has the potential to uniquely identify humans and to be ingenious enough to protect against faked identities [8]. For instance, Chuang et al. [8] propose a single-channel EEG-based authentication system, which achieves an accuracy of 0.99. Keshishzadeh et al. [20] employ a statistical model for analyzing EEG signals and achieve an accuracy of 0.974. Generally, EEG signals have the following inherent advantages:

- *Fake-resistibility.* EEG data are unique for each person and almost impossible to be cloned and duplicated. EEG signals are individual-dependent. Therefore, an EEG-based authentication system has the potential to verify human identity and to be ingenious enough to protect against faked identities [8].
- *Reliability.* An EEG-based authentication system can reject the subjects under abnormal situations (e.g., dramatically spiritual fluctuating, hysterical, drunk, or under threat) since EEG signals are sensitive to human stress and mood.
- *Feasibility.* We have seen an important trend to build authentication systems based on EEG because the equipment for collecting EEG data is cheap and easy to acquire, and it is expected to be more precise, accessible, and economical in the future.

In comparison, gait-based authentication systems have been an active direction for years [37, 48]. Gait data are more generic and can be gathered easily from popular inertial sensors. Gait data are also unique because they are determined by intrinsic factors (e.g., gender, height, and limb length), temporal factors [6] (e.g., step length, walking speed, and cycle time), and kinematic factors (e.g., joint rotation of the hip, knee, and ankle, mean joint angles of the hip/knee/ankle, and thigh/trunk/foot angles). In addition, a person's gait behavior is established inherently in the long term and therefore difficult to be faked. Hoang and Choi [18] propose a gait-based authentication biometric system to analyze gait data gathered by mobile devices, adopt error correcting codes to process the variation in gait measurement, and finally achieve a False Acceptance Rate (FAR) of 3.92% and a False Rejection Rate (FRR) of 11.76%. Cola et al. [9] collect wrist signals and train

[1] http://www.instructables.com/id/How-To-Fool-a-Fingerprint-SecuritySystem-As-Easy-/.
[2] http://www.urmesurveillance.com/urme-prosthetic/.

gait patterns to detect invalid subjects (unauthenticated people). The proposed method achieves an Equal Error Rate (EER) of 2.9%.

Despite the tremendous efforts, various other challenges still remain in single EEG/gait-based authentication systems: (i) the solo EEG-based authentication system has very high fake-resistance and excellent authentication performance but is easily affected by environmental factors (e.g., noise), subjective factors (e.g., mental state), and noisy brain signals (e.g., not concentrating); (ii) the solo gait-based authentication system is more stable over different scenarios but has relatively low performance; (iii) the solo EEG/gait authentication system generally obtains a FAR (which is extremely crucial in high-confidential authentication scenarios)[3] higher than 3% [18]. It is not precise enough for highly sensitive places such as military bases, the treasuries of banks, and political offices where tiny misjudgments could provoke great economic or political catastrophes; (iv) the single authentication system may break down while under attack but no backup plan is provided.

In this article, we propose DeepKey, a novel biometric authentication system that enables dual-authentication leveraging on the advantages of both gait-based and EEG-based systems. Compared with either a gait-based or an EEG-based authentication system, a dual-authentication system offers more reliable and precise identification. Table 1 summarizes the overall comparison of DeepKey with some representative works on seven key aspects. DeepKey consists of three main components: the Invalid ID Filter Model to eliminate invalid subjects, the EEG Identification Model to identify EEG IDs, and the Gait Identification Model to identify gait IDs. An individual can be granted access only after she/he passes all the authentication components. Our main contributions are highlighted as follows:

- We present DeepKey, a dual-authentication system that exploits both EEG and gait biological traits. To the best of our knowledge, DeepKey is the first two-factor authentication system for person authentication using EEG and gaits. DeepKey is empowered with high-level fake-resistance and reliability because both EEG and gait signals are invisible and hard to be reproduced.
- We design a robust framework that includes an attention-based RNN to detect and classify multimodal sensor data, and to decode the large diversity in how people perform gaits and brain activities simultaneously. The delta band of EEG data is decomposed for its rich discriminative information.
- We validate and evaluate DeepKey on several locally collected datasets. The results show that DeepKey significantly outperforms a series of baseline models and the state-of-the-art methods, achieving FAR of 0 and FRR of 1%. Further, we design extensive experiments to investigate the impact of key elements.

The remainder of this article is organized as follows. Section 2 introduces the EEG-based, gait-based, and multimodal biometric systems briefly. Section 3 presents the methodology framework and three key models (Invalid ID Filter Model, Gait Identification Model, and EEG Identification Model) of the DeepKey authentication system in detail. Section 4 evaluates the proposed approach on the public Gait and EEG dataset and provides an analysis of the experimental results. Finally, Section 5 discusses the opening challenges of this work and highlights the future scope of this research, while Section 6 summarizes the key points of this article.

[3]In the high-confidential authentication scenario, FAR is more crucial than other metrics such as accuracy.

Table 1. Comparison of Various Biometrics

	Biometrics	Fake-resistance ↑	Universality ↑	Uniqueness ↑	Stability ↑	Accessibility ↑	Performance ↑	Computational Cost ↓
Uni–modal	Face/Vedio	Medium	Medium	Low	Low	High	Low	High
	Fingerprint/Palmprint	Low	High	High	High	Medium	High	Medium
	Iris	Medium	High	High	High	Medium	High	High
	Retina	High	Medium	High	Medium	Low	High	High
	Signature	Low	High	Low	Low	High	Low	Medium
	Voice	Low	Medium	Low	Low	Medium	Low	Low
	Gait	High	Medium	High	Medium	Medium	High	Low
	EEG	High	Low	High	Low	Low	High	Low
Mulit–modal	Fingerprint+face	Low	High	Low	Low	High	Medium	High
	Iris+pupil	Medium	Medium	High	High	Medium	High	High
	Iris+face	Medium	High	Medium	Medium	Medium	Medium	High
	ECG+fingerprint	High	Medium	Medium	High	Low	High	Medium
	EEG+gait	**High**	**Low**	**High**	**High**	**Low**	**High**	**Low**

EEG and Gait have considerable fake-resistance, which is the most significant characteristic of authentication systems. ↑ denotes the higher the better while ↓ denotes the lower the better.

2 RELATED WORK

In this section, we introduce the related studies on several topics: biometric authentication technologies, EEG-based authentication, gait-based authentication, and multimodal biometric authentication.

2.1 Biometric Authentication Technologies

Since biometric features cannot be stolen or duplicated easily, biometric authentication is becoming increasingly commonplace. Currently, the most mature biometric authentication technology is fingerprint-based authentication, which has been demonstrated to have high matching accuracy and been used for decades [30]. Iris recognition is another popular approach for biometric authentication owing to its unique and stable pattern [36]. Daugman [10] proposes to use Gabor phase information and Hamming distance for iris code matching, which still is the most classic iris recognition method. Based on [10], a flurry of research [36] has emerged offering solutions to ameliorate iris authentication problems. For example, Pillai et al. [36] introduce kernel functions to represent transformations of iris biometrics. This method restrains both the intra-class and inter-class variability to solve the sensor mismatch problem. Face recognition techniques [13, 17, 52] is the most commonly used and accepted by the public for its unique features and non-invasiveness. Since face recognition systems require tackling different challenges including expression, image quality, illumination, and disguise to achieve high accuracy, Infrared Radiation (IR) [17] and 3D [13] systems have attracted much attention. According to [52], multimodal recognition combining traditional visual textual features and IR or 3D systems can achieve higher accuracy than single modal systems.

2.2 EEG-Based Authentication

Since EEG can be gathered in a safe and non-invasive way, researchers have paid great attention to exploring this kind of brain signals. For person authentication, EEG is, on the one hand, promising for being confidential and fake-resistant but, on the other hand, complex and hard to be analyzed. Marcel and Millán [33] use Gaussian Mixture Models and train client models with Maximum A Posteriori (MAP). Ashby et al. [3] extract five sets of features from EEG electrodes and inter-hemispheric data, combine them together, and process the final features with a Support Vector Machine (SVM). The study shows that EEG authentication is also feasible with less-expensive devices. Altahat et al. [2] select Power Spectral Density (PSD) as the feature instead of the widely used AutoRegressive (AR) models to achieve higher accuracy. They also conduct channel selection to determine the contributing channels among all 64 channels. Thomas and Vinod [41] take advantage of individual alpha frequency (IAF) and delta band signals to compose a specific feature vector. They also prefer PSD features but only perform the extraction merely on the gamma band.

2.3 Gait-Based Authentication

As the most basic activity in our daily lives, walking is an advanced research hotspot for activity recognition [46]. Differing from previous studies, our work focuses on human gait, a spatio-temporal biometric that measures a person's manner on walking. Existing gait recognition approaches sit in two categories: One is *model-based approach* [35], which models gait information with mathematical structures, and the other is *appearance-based approach*, which extracts features in a straightforward way irrespective of the mathematical structure. Due to its high efficiency and remarkable performance, Gait Energy Image (GEI) [31] has become one of the most popular appearance-based methods in recent years. Based on GEIs, a considerable amount of works have been proposed to explore the exterior factors and distinguish different body parts. In addition, the

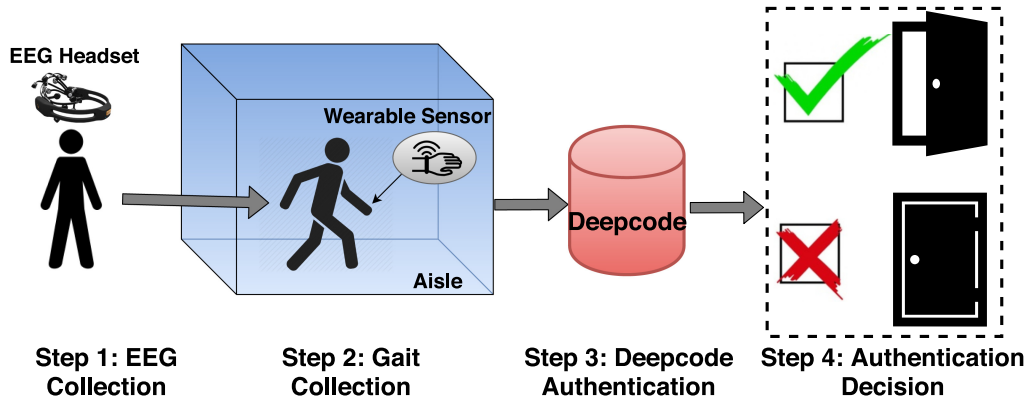


Fig. 1. Workflow of DeepKey authentication system. The data collection of EEG and gait are cascade.

cross-view variance is also a concern of gait identification [25, 26]. For example, Wu et al. [45] consider not only the cross-view variance but also deep Convolutional Neural Networks (CNNs) for robust gait identification.

2.4 Multimodal Biometric Authentication

Since traditional unimodal authentication suffers from the negative influence of loud noise, low universality, and intra-class variation, it cannot achieve higher accuracy in a wide range of applications. To address this issue, multimodal biometric authentication, which combines and uses biometric traits in different ways, is becoming popular. Taking the commonness into consideration, most works choose two biometrics from face, iris, and fingerprints and make the fusion [21, 24, 40]. In [11], an innovative combination between gait and electrocardiogram (ECG) is shown to be effective. Manjunathswamy et al. [32] combine ECG and fingerprint at the score level. To the best of our knowledge, the approach proposed in this article is the first to combine EEG and gaits for person authentication. Taking advantage of both EEG and gait signals, the combination is expected to improve the reliability of authentication systems.

This dual-authentication system is partially based on our previous work MindID [50], which is an EEG-based identification system. We emphasize several differences compared to [50]: (1) this work is an authentication system with an invalid ID filter while [50] only focuses on identification; (2) this work adopts two biometrics including EEG and gait while [50] only exploits EEG signals; and (3) this work conducts extensive real-world experiments to collect gait signals.

3 DEEPKEY AUTHENTICATION SYSTEM

In this section, we first report the workflow of DeepKey to give an overview of the authentication algorithm, and then present the technical details for each component.

3.1 DeepKey System Overview

The DeepKey system is supposed to be deployed in access to confidential locations (e.g., bank vouchers, military bases, and government confidential residences). As shown in Figure 1, the overall workflow of the DeepKey authentication system consists of the following four steps:

(1) Step 1: EEG data collection. The subject, who requests for authentication, is required to wear the EEG headset and stays in relaxation. The collection of EEG data ($\mathcal{E}$) will typically take 2 seconds.

(2) Step 2: Gait data collection. The subject takes off the EEG headset and puts on three IMUs (Inertial Measurement Units) and walks through an aisle to collect gait data $\mathcal{G}$ by IMUs.
(3) Step 3: Authentication. The gathered EEG and gait data are flattened and associated with input data $\mathcal{I} = [\mathcal{E} : \mathcal{G}]$ to be fed into the DeepKey authentication algorithm.
(4) Step 4: Decision. An *Approve* or *Deny* decision will be made according to the DeepKey authentication results.

The most crucial component among the above steps is the third step, where the DeepKey authentication system receives the associated input data $\mathcal{I}$ and accomplishes two goals: authentication and identification. For the former goal, we employ EEG signals to justify the impostor for its high fake-resistance. EEG signals are invisible and unique, making them difficult to be duplicated and hacked. For the latter goal, we adopt a deep learning model to extract the distinctive features and feed them into a non-parametric neighbor-based classifier for ID identification. In summary, the DeepKey authentication algorithm contains several key stages, namely, *Invalid ID Filter*, *Gait-based Identification*, *EEG-based Identification*, and *Decision Making*. The overall authentication contains the following several stages (Figure 3):

(1) Based on the EEG data, the Invalid ID Filter decides the subject is an impostor or a genuine. If the subject is an impostor, the request will be denied.
(2) If the individual is determined as genuine, the EEG/Gait Identification Model will identify the individual's authorized EEG/Gait ID. This model is pre-trained offline with the attention-based Long Short-Term Member (LSTM) model (Section 3.3). The output is the ID number associated with the person's detailed personal information.
(3) The final stage is to check the consistency of the EEG ID and the Gait ID. If they are identical, the system will grant an approval, otherwise it will deny the subject and take corresponding security measures.

3.2 Invalid ID Filter Model

Through our preliminary experiments, it has been found that raw EEG signals, compared to raw Gait data, have better characteristics to prevent invalid ID due to the high fake-resistance of EEG data and the richness of distinguishable features in EEG signals. Thus, in this section, we only use EEG for invalid ID filtering.

The subjects in an authentication system are categorized into two classes: *authorized* and *unauthorized*. Since the unauthorized data are not available in the training stage, an unsupervised learning algorithm is required to identify the invalid ID. In this work, we apply a one-class SVM to sort out the unauthorized subjects. Given a set of authorized subjects, $\mathcal{S} = \{S_i, i = 1, 2, \ldots, L^\circ\}$, $S_i \in R^{n_s}$, where L° denotes the number of authorized subjects and n_s denotes the number of dimensions of the input data. The input data consist of EEG data $\mathcal{E} = \{E_i, i = 1, 2, \ldots, L^\circ\}$, $E_i \in R^{n_e}$ and gait data $\mathcal{G} = \{G_i, i = 1, 2, \ldots, L^\circ\}, G_i \in R^{n_g}$. n_g and n_e denote the number of dimensions of the gait data and EEG data, respectively, and $n_s = n_g + n_e$. The notation can be found in Table 2.

For each authentication, the collected EEG data E_i includes a number of samples. Each sample is a vector with shape $[1, 14]$ where 14 denotes the number of electric-nodes in the Emotiv headset. To trade off the authentication efficiency (less collecting and waiting time) and computational performance, based on the experimental experience, we fed 200 samples ([200, 14]) into the Invalid ID Filter. 200 EEG samples are collected in 1.56 seconds, which is acceptable. The final filter result is the mean of the results on all the samples.

Table 2. Notation

Parameters	Explanation
L°	the number of authorized subjects (genuine)
E	the set of EEG signals
G	the set of gait signals
n_s	the number of features in input data sample
G_i	the i-th gait data
n_g	the number of features in per gait sample
E_i	the i-th EEG data
n_e	the number of features in EEG data sample
X^i	the data in the i-th layer of attention RNN
N^i	the number of dimensions in X^i
K	the number of participants (genuine and impostor)
$c_j^{i'}$	the hidden state in the j-th LSTM cell
$\mathcal{T}(\cdot)$	the linear calculation of dense neural layers
$\mathcal{L}(\cdot)$	The LSTM calculation process
$\odot$	the elementwise multiplication
C_{att}	attention-based code

Table 3. Characteristics of EEG Frequency Bands

Name	Frequency (Hz)	Amplitude	Brain State	Awareness Degree	Produced Location
Delta	0.5–3.5	Higher	Deep sleep pattern	Lower	Frontally and posteriorly
Theta	4–8	High	Light sleep pattern	Low	Entorhinal cortex, hippocampus
Alpha	8–12	Medium	Closing the eyes, relax state	Medium	Posterior regions of head
Beta	12–30	Low	Active thinking, focus, high alert, anxious	High	Most evident frontally
Gamma	30–100	Lower	During cross-modal sensory processing	Higher	Somatosensory cortex

Awareness Degree denotes the degree of being aware of an external world.

3.3 EEG Identification Model

Compared to gait data, EEG data contain more noise, which is more challenging to handle. Given the complexity of EEG signals, the data pre-processing is necessary. In practical EEG data analysis, the assembled EEG signals can be divided into several different frequency patterns (delta, theta, alpha, beta, and gamma) based on the strong intra-band correlation with a distinct behavioral state. The EEG frequency patterns and the corresponding characters are listed in Table 3 [50]. Figure 2 reports the topography of EEG signals of different subjects under different frequency bands and demonstrates that the Delta wave, compared to other bands, enriches distinctive features. In detail, we calculate the inter-subject EEG signal cosine-similarity which measures the average similarity among different subjects under all the EEG bands. The results are reported as 0.1313 (full bands), 0.0722 (Delta band), 0.1672 (Theta band), 0.2819 (Alpha band), 0.0888 (Beta band), and 0.082 (Gamma band). This illustrates that the delta band with the lowest similarity contains the most

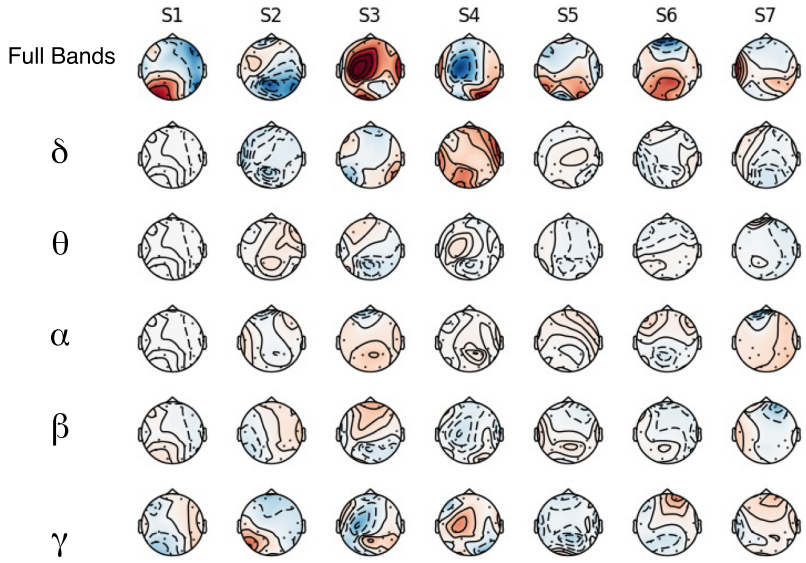


Fig. 2. EEG topography of different subjects under different frequency bands. The inter-subject EEG signal cosine-similarity is calculated under each band and the results are reported as 0.1313 (full bands), 0.0722 (Delta band), 0.1672 (Theta band), 0.2819 (Alpha band), 0.0888 (Beta band), and 0.082 (Gamma band). This illustrates that the delta band with the lowest similarity contains the most distinguishable features for person identification.

distinguishable features for person identification. Our previous work [50] has demonstrated that Delta pattern, compared to other EEG patterns, contains the most distinctive information and is the most stable pattern in different environments by qualitative analysis and empirical experiment results. Thus, in this article, we adopt a bandpass (0.5 Hz–3.5 Hz) Butterworth filter to extract Delta wave signal for further authentication. For simplicity, we denote the filtered EEG data as $\mathcal{E}$.

Since different EEG channels record different aspects of the brain signals, some of which are more representative of the individual, an approach that assumes all dimensions to be equal may not be suitable. Thus, we attempt to develop a novel model which can pay more attention to the most informative signals. In particular, the proposed approach is supposed to automatically learn the importance of the different parts of the EEG signal and focus on the valuable part. The effectiveness of attention-based RNN has been demonstrated in various domains including natural language processing [44] and speech recognition [7]. Inspired by the wide success of an attention mechanism [12], we introduce it to the Encoder-Decoder RNN model to assign varying weights to different dimensions of the EEG data. After EEG filtering, the composed Delta pattern $\mathcal{E}$ is fed into an attention-based Encoder-Decoder RNN structure [44] aiming to learn more representative features for user identification. The general Encoder-Decoder RNN framework regards all the feature dimensions of input sequence as having the same weights, no matter how important the dimension is for the output sequence. In this article, the different feature dimensions of the EEG data correspond to the different nodes of the EEG equipment. For example, the first dimension (first channel) collects the EEG data from the $AF3$[4] node located at the frontal lobe of the scalp while the seventh dimension is gathered from the $O1$ node at the occipital lobe. To assign various weights to different

[4]Both $AF3$ and $O1$ are EEG measurement positions in the International 10-20 Systems.

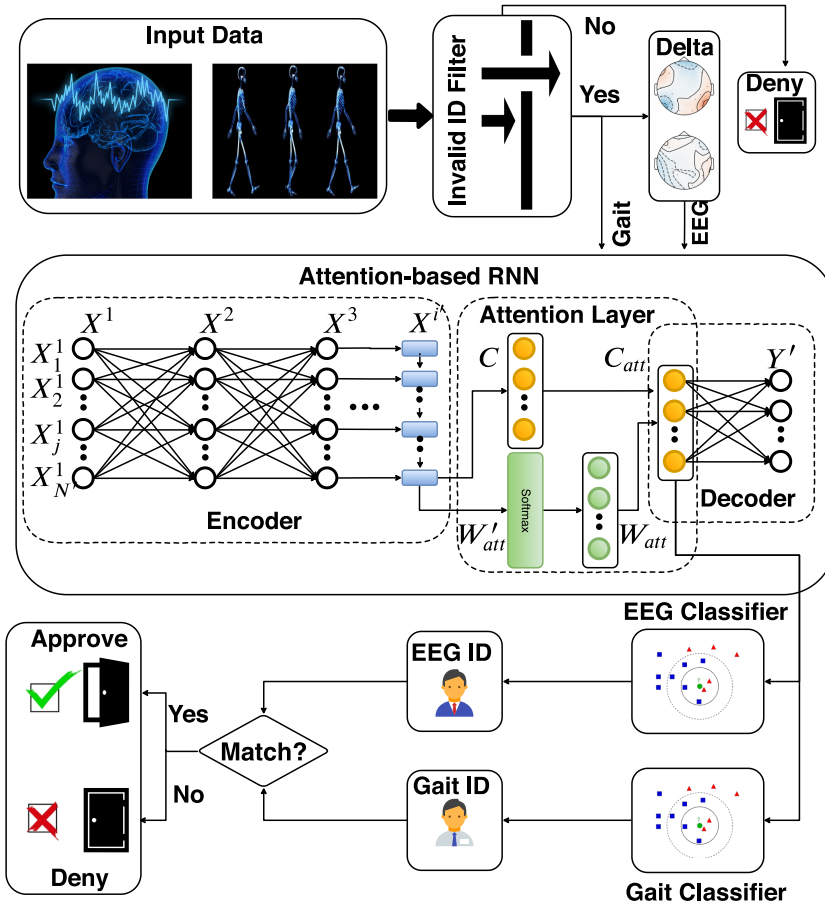


Fig. 3. Authentication workflow. If the input data cannot pass the invalid ID filter, it would directly be regarded as an impostor and deny access. If pass, the Delta pattern and gait signals are parallelly fed into an attention-based RNN structure to study the distinctive features C_{att}. The learned features are classified by the EEG and Gait classifier in order to identify the subject's EEG and Gait ID. The subject is approved only if the EEG ID is a match with Gait ID.

dimensions of $\mathcal{E}$, we introduce the attention mechanism to the Encoder-Decoder RNN model. The proposed attention-based Encoder-Decoder RNN consists of three components (Figure 3): the encoder, the attention module, and the decoder. The encoder is designed to compress the input Delta δ wave into a single intermediate code C; the attention module helps the encoder calculate a better intermediate code C_{att} by generating a sequence of the weights W_{att} of different dimensions; the decoder accepts the attention-based code C_{att} and decodes it to the output layer Y'.

Suppose the data in the i-th layer can be denoted by $X^i = (X^i_j; i \in [1, 2, \ldots, I], j \in [1, 2, \ldots, N^i])$ where j denotes the j-th dimension of X^i. I represents the number of neural network layers while N^i denotes the number of dimensions in X^i. Taking the first layer as an example, we have $X^1 = \mathcal{E}$, which indicates that the input sequence is the Delta pattern. Let the output sequence be $Y = (Y_k; k \in [1, 2, \ldots, K])$ where K denotes the number of users. In this article, the user ID is represented by the one-hot label with length K. For simplicity, we define the operation $\mathcal{T}(\cdot)$ as

$\mathcal{T}(X^i) = X^i W + b$. Furthermore, we have

$$\mathcal{T}(X_j^{i-1}, X_{j-1}^i) = X_j^{i-1} * W' + X_{j-1}^i * W'' + b',$$

where W, b, W', W'', b' denote the corresponding weight and bias parameters.

The encoder component contains several non-recurrent fully connected layers and one recurrent LSTM layer. The non-recurrent layers are employed to construct and fit into a non-linear function to purify the input Delta pattern; the necessity is demonstrated by the preliminary experiments.[5] The dataflow in these non-recurrent layers can be calculated by

$$X^{i+1} = tanh(\mathcal{T}(X^i)),$$

where *tanh* is the activation function. We engage the *tanh* as an activation function instead of *sigmoid* for the stronger gradient [28]. The LSTM layer is adopted to compress the output of non-recurrent layers to a length-fixed sequence, which is regarded as the intermediate code C. Suppose LSTM is the i'-th layer; the code equals to the output of LSTM, which is $C = X_j^{i'}$. The $X_j^{i'}$ can be measured by

$$X_j^{i'} = \mathcal{L}(c_{j-1}^{i'}, X_j^{i-1}, X_{j-1}^{i'}), \tag{1}$$

where $c_{j-1}^{i'}$ denotes the hidden state of the $(j-1)$-th LSTM cell. The operation $\mathcal{L}(\cdot)$ denotes the calculation process of the LSTM structure, which can be inferred from the following equations.

$$X_j^{i'} = f_o \odot tanh(c_j^{i'}), c_j^{i'} = f_f \odot c_{j-1}^{i'} + f_i \odot f_m,$$

$$f_o = sigmoid(\mathcal{T}(X_j^{i'-1}, X_{j-1}^{i'})), f_f = sigmoid(\mathcal{T}(X_j^{i'-1}, X_{j-1}^{i'})),$$

$$f_i = sigmoid(\mathcal{T}(X_j^{i'-1}, X_{j-1}^{i'})), f_m = tanh(\mathcal{T}(X_j^{i'-1}, X_{j-1}^{i'})),$$

where f_o, f_f, f_i, and f_m represent the output gate, forget gate, input gate, and input modulation gate, respectively, and $\odot$ denotes the element-wise multiplication.

The attention module accepts the final hidden states as the unnormalized attention weights W'_{att}, which can be measured by the mapping operation $\mathcal{L}'(\cdot)$

$$W'_{att} = \mathcal{L}'(c_{j-1}^{i'}, X_j^{i-1}, X_{j-1}^{i'}) \tag{2}$$

and calculate the normalized attention weights W_{att}

$$W_{att} = softmax(W'_{att}).$$

The softmax function is employed to normalize the attention weights into the range of $[0, 1]$. Therefore, the weights can be explained as the probability of how the code C is relevant to the output results. Under the attention mechanism, the code C is weighted to C_{att},

$$C_{att} = C \odot W_{att}.$$

Note, C and W_{att} are trained simultaneously. The decoder receives the attention-based code C_{att} and decodes it to the output Y'. Since Y' is predicted at the output layer of the attention-based RNN model ($Y' = X^I$), we have

$$Y' = \mathcal{T}(C_{att}).$$

At last, we employ the cross-entropy cost function, and ℓ_2-norm (with parameter λ) is selected to prevent overfitting. The cost is optimized by the AdamOptimizer algorithm [22]. The iterations threshold of attention-based RNN is set as n_{iter}^E. The weighted code C_{att} has a direct linear relationship with the output layer and the predicted results. If the model is well trained with low cost, we could regard the weighted code as a high-quality representation of the user ID. We set

[5]Some optimal designs like the neural network layers are validated by the preliminary experiments, but the validation procedure will not be reported in this article due to space limitations.

ALGORITHM 1: DeepKey System

Input: EEG data $\mathcal{E}$ and Gait data $\mathcal{G}$
Output: Authentication Decision: Approve/Deny

#Invalid ID Filter:
for $\mathcal{E}, \mathcal{G}$ **do**
  Genuine/Impostor $\leftarrow \mathcal{E}$
  if Impostor **then**
    return Deny
  else if Genuine **then**
    #EEG Identification Model:
    while $iteration < n^{E}_{iter}$ **do**
      $X^{i+1} = tanh(\mathcal{T}(X^{i}))$ $\{X^{1} = \mathcal{E}\}$
      $C = X^{i'}_{j} = \mathcal{L}(c^{i'}_{j-1}, X^{i-1}_{j}, X^{i'}_{j-1})$
      $W_{att} = softmax(\mathcal{L}'(c^{i'}_{j-1}, X^{i-1}_{j}, X^{i'}_{j-1}))$
      $C_{att} = C \odot W_{att}$
      $E_{ID} \leftarrow C_{att}$
    end while
    #Gait Identification Model:
    while $iteration < n^{G}_{iter}$ **do**
      $G_{ID} \leftarrow \mathcal{G}$
    end while
    if $E_{ID} = G_{ID}$ **then**
      return Approve
    else
      return Deny
    end if
  end if
end for

the learned deep feature X_D to C_{att}, $X_D = C_{att}$, and feed it into a lightweight nearest-neighbor classifier. The EEG ID, which is denoted by E_{ID}, can be directly predicted by the classifier.

The Gait Identification Model works similarly to the EEG Identification Model except for the frequency band filtering. The iterations threshold of attention-based RNN is set as n^{G}_{iter}. The Gait Identification Model receives a subject's gait data $\mathcal{G}$ from the input data $\mathcal{I}$ and maps to the user's Gait ID G_{ID}. All the model structures, hyper-parameters, optimization, and other settings in the EEG and Gait Identification Models remain the same to keep the lower model complexity of the DeepKey system.

4 EXPERIMENTS AND RESULTS

In this section, we first outline the experimental setting including dataset, hyper-parameters settings, and evaluation metrics. Then we systematically investigate (1) the comparison with the state-of-the-art authentication systems at both system level and component level; (2) the impact of key parameters like single/multiple sessions,[6] EEG band, and data size; and (3) the authentication latency.

[6]Single session refers to the dataset collected in one session (the period from one subject putting the EEG headset on until all the experiments are finished, then taking it off). Multi-session represents the EEG data collected from different sessions, which considered the effect on EEG data quality caused by the headset position errors.

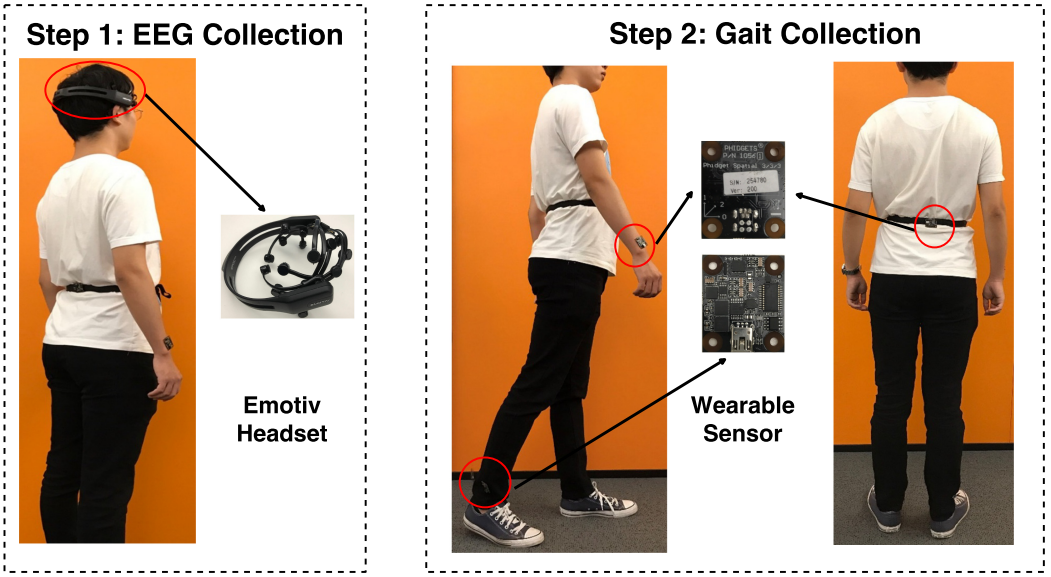


Fig. 4. Data collection. Two collection steps are cascaded to eliminate the impact on EEG data of walking. The first step collects the solo EEG signals while the second step collects the gait signals.

4.1 Experimental Settings

4.1.1 Datasets. We design real-world experiments to collect EEG data and gait data in cascade.[7] The experiments (Figure 4) are conducted by seven healthy participants aged 26 ± 2 including four males and three females. In Step 1, each participant remains standing and relaxed with eyes closed. The EEG data are collected by an EPOC+ Emotiv headset[8] which integrates 14 electrodes (corresponding to 14 EEG channels) with a sampling rate of 128 Hz. The Emotiv headset gathers brain signals in a non-invasive manner since it doesn't require surgery to insert sensors into the user's skull [51]. In Step 2, each participant walks in an aisle to generate the gait data. In the gait collection procedure, three IMUs are attached to the participants' left wrist, the middle of the back, and the left ankle, respectively. Each IMU (PhidgetSpatial 3/3/3[9]) with 80 Hz sampling rate gathering nine-dimensional motor features contains a three-axis accelerometer, three-axis gyroscope, and a three-axis magnetometer.

To investigate the impact of dataset sessions, both the EEG and gait data are collected in three sessions. In every cycle of single session, the subject puts on the equipment (headset/IMUs), gathers data, and then takes the equipment off. Therefore, in different sessions, the positions of the equipment may have a slight deviation. Table 4 reports the details of the datasets used in this article. Each EEG or gait sample contains 10 continuous instances without overlapping. The single session datasets (EID-S and GID-S) are collected in a single experiment session while the multi-session datasets (EID-M and GID-M) are gathered in three sessions. All the sessions are conducted in the same place but on three different days (each session in one day). The EEG data are easily influenced if the emotional or physical state has changed, thus, we believe the collected data are diverse because of the varying environmental factors (e.g., noise and temperature) and subjective factors (e.g., participants' mental state and fatigue state). Similarly, the gait signals could be

[7]The experiment was approved by our ethics board.
[8]https://www.emotiv.com/product/emotiv-epoc-14-channel-mobile-eeg/.
[9]https://www.phidgets.com/?&prodid=48.

Table 4. Datasets Description, Australia

Dataset	Biometric	#-D	Session	Frequency	Samples
EID-S	EEG	14	Single	128 Hz	49,000
EID-M	EEG	14	Multiple	128 Hz	147,000
GID-S	Gait	27	Single	80 Hz	140,000
GID-M	Gait	27	Multiple	80 Hz	420,000

#-D denotes the number of dimensions.

affected by lots of variables like different shoes (comfortable/uncomfortable). In this article, as an exploratory work, we focus on developng a robust discriminative deep learning model which is strong enough to prevent the corruption of the aforementioned influencing factors. The investigation of the detail effect brought by each specific factor will be left as a future research direction.

4.1.2 Parameter Settings. The Invalid ID Filter attempts to recognize the unauthorized subject based on the unique EEG data. The filter chooses the RBF kernel with $nu = 0.15$. In the EEG Identification Model, the Delta band ($[0.5\ Hz, 3.5\ Hz]$) is filtered by a three-order Butterworth bandpass filter. In the attention-based RNN, in the input layer, there are 14 nodes for EEG and 27 nodes for gait signals. For both EEG and gait, we have two fully connected hidden layers (each has 64 nodes) and one LSTM layer with 64 cells; the output layer has seven nodes. The learning rate and λ are both set to 0.001. The weighted code is produced after 1,000 iterations. Here, eightfold cross-validation is used to prevent overfitting. The dataset is randomly separated into eght equal-sized subsets. One of the 10 subsets is used as a test set while the remaining subsets are used as a training set. The user ID ranges from 0 to 6 and is represented in the one-hot label.

4.1.3 Metrics. The adopted evaluation metrics are accuracy, ROC, AUC, along with FAR and FRR. DeepKey is very sensitive to the invalid subject and can acquire very high accuracy in Invalid ID Filter Model for the reason that even a tiny misjudgment may lead to catastrophic consequences. Therefore, FAR is more important than other metrics. Therefore, in DeepKey, FAR has higher priority compared to other metrics such as FRR.

4.2 Overall Comparison

4.2.1 System-Level Comparison. To evaluate the performance of DeepKey, we compare it with a set of the state-of-the-art authentication systems. DeepKey is empowered to solve both the authentication and identification problems. As shown in Table 5, DeepKey achieves a FAR of 0 and a FRR of 1%, outperforming other uni-modal and multimodal authentication systems. Specifically, our approach, compared to the listed uni-modal systems, achieves the highest EEG identification accuracy (99.96%) and Gait identification accuracy (99.61%).

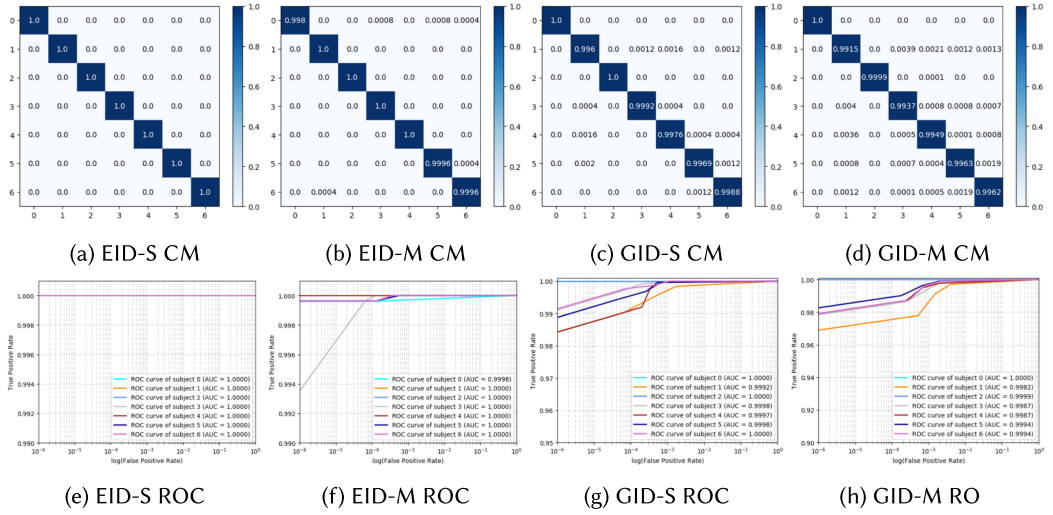
4.2.2 Component-Level Comparison. To have a closer observation, we provide the detailed performance study of each component. In the Invalid ID Filter, to enhance the accuracy and robustness of the classifier, EEG samples are separated into different segments, with each segment (without overlapping) having 200 continuous samples. Six in seven subjects are labeled as genuine while the other subject is labeled as an impostor. In the training stage, all the EEG segments are fed into the one-class SVM with RBF kernel for pattern learning. In the test stage, 1,000 genuine segments and 1,000 impostor segments are randomly selected to assess the performance. We use the leave-one-out-cross-validation training strategy and achieve a FAR of 0 and a FRR of 0.006.

In the EEG/Gait Identification Model, the proposed approach achieves an accuracy of 99.96% and 99.61% over the multi-session datasets, respectively. The detailed confusion matrix, ROC curves with AUC scores, and the classification reports (precision, recall, and F1-score) over all the datasets

Table 5. System-Level Comparison between DeepKey and Other Biometrics Authentication Systems

	Reference	Biometric	Method	#-Subject	Dataset	Accuracy	FAR	FRR
Uni-modal	[9]	Gait	semi-supervised anomaly detection+NN	15	Local	97.4		
	[34]	Gait	AVTM-PdVs	100	Public	77.72		
	[1]	Gait	MLP	60	Local	99.01		
	[39]	Gait	Artificial features + voting classifier	10	Local	98.75		
	[23]	Gait	Two SVMs	50	Local		1.0	1.0
	[42]	EEG	PSD + cross-correlation values	109	Public		1.96	1.96
	[8]	EEG	Customized Threshold	15	Local		0	2.2
	[16]	EEG	Low-pass filter+wavelets+ ANN	32	Local	90.03		
	[4]	EEG	Bandpass FIR filter +ECOC + SVM	9	Local	94.44		
	[41]	EEG	IAF + delta band EEG + Cross-correlation values	109	Public	90.21		
	[19]	EEG	CSP +LDA	12	Local	96.97		
	[50]	EEG	Attention-based RNN + XGB	8	Local	98.82		
	[20]	EEG	AR + SVM	104	Public	97.43		
Multi-modal	[29]	Fingerprint	ZM + RBF Neural Network	40	Public	92.89	7.108	7.151
		Face					11.52	13.47
		Fusion					4.95	1.12
	[47]	Iris	Gabor 2D wavelets + Gabor 2D wavelets + Hamming distance	59	Public		13.88	13.88
		Pupil					5.47	5.47
		Fusion					2.44	2.44
	[32]	ECG	wavelet decomposition	50	Public		2.37	9.52
		Fingerprint	Histogram manipulation Image Enhancement				7.77	5.55
		Fusion	Score Fusion				2.5	0
	[11]	ECG	linear time interpolation + cross correlation	30	Local		4.2	4.2
		Gait					7.5	7.5
		Fusion	Score Fusion				1.26	1.26
	Ours	EEG	Delta wave + Attention-based RNN +KNN	7	Local	99.96		
		Gait	Attention-based RNN +KNN			99.61		
		Fusion				**99.57**	**0**	**1.0**

The performance of our methods is evaluated on multi-session datasets (EID-M, GID-M).



(a) EID-S CM (b) EID-M CM (c) GID-S CM (d) GID-M CM

(e) EID-S ROC (f) EID-M ROC (g) GID-S ROC (h) GID-M RO

Fig. 5. Confusion matrix and ROC curves of the datasets. CM denotes confusion matrix. The AUC are provided on the figures.

Table 6. Classification Report of the Datasets Including Precision, Recall, and F-1 Score

Datasets	EID-S			EID-M			GID-S			GID-M		
Metrics	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
0	1.0	1.0	1.0	0.998	1.0	0.999	1.0	1.0	1.0	1.0	1.0	1.0
1	1.0	1.0	1.0	1.0	0.9996	0.9998	0.996	0.996	0.996	0.9915	0.9904	0.9909
2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9999	1.0	0.9999
3	1.0	1.0	1.0	1.0	0.9992	0.9996	0.9992	0.9988	0.999	0.9937	0.9948	0.9942
4	1.0	1.0	1.0	1.0	1.0	1.0	0.9976	0.998	0.9978	0.9949	0.996	0.9955
5	1.0	1.0	1.0	0.9996	0.9992	0.9994	0.9969	0.9984	0.9977	0.9963	0.996	0.9961
6	1.0	1.0	1.0	0.9996	0.9993	0.9994	0.9988	0.9972	0.998	0.9962	0.9953	0.9958
Average	**1.0**	**1.0**	**1.0**	**0.9996**	**0.9996**	**0.9996**	**0.9983**	**0.9983**	**0.9983**	**0.9961**	**0.9961**	**0.9961**

The proposed approach gains impressive results (higher than 99%) on all the metrics over all the seven subjects.

are presented in Figure 5 and Table 6. The above evaluation metrics demonstrate that the proposed approach achieves a performance of over 99% on all the metrics over each subject and each dataset. To make a closer observation, taking the EID-M dataset as an example, we present the training and testing curves. As shown in Figure 8, the horizontal and vertical axes denote the number of training iterations and the accuracy, respectively. We can observe that our model starts at a rather high level, converges steadily in the training stage, and does not suffer from the overfitting problem.

Furthermore, the overall comparison between our model and other state-of-the-art baselines are listed in Table 7. RF denotes Random Forest, AdaB denotes Adaptive Boosting, and LDA denotes Linear Discriminant Analysis. In addition, the key parameters of the baselines are listed here: Linear SVM ($C = 1$), RF ($n = 200$), and KNN ($k = 3$). The settings of LSTM are the same as the attention-based RNN classifier, along with the GRU (Gated Recurrent Unit). The CNN contains two stacked convolutional layers (both with stride [1, 1], patch [2, 2], zero-padding, and the depths are 4 and 8, separately.) followed by one pooling layer (stride [1, 2], zero-padding) and one fully connected layer (164 nodes). Relu activation function is employed in the CNN. The methods used

Table 7. Component-Level Comparison

Baseline	Methods	EID-S				EID-M			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Non-DL Baseline	SVM	0.4588	0.5848	0.4588	0.4681	0.7796	0.7815	0.7796	0.7796
	RF	0.9875	0.9879	0.9876	0.9876	0.8124	0.8139	0.8124	0.812
	KNN	0.9897	0.9899	0.9898	0.9898	0.8211	0.8232	0.8211	0.8197
	AdaB	0.2872	0.3522	0.2871	0.2337	0.3228	0.3224	0.3228	0.2815
	LDA	0.1567	0.1347	0.1567	0.1386	0.3082	0.285	0.3082	0.2877
DL Baseline	LSTM	0.9596	0.9601	0.9596	0.9597	0.8482	0.8509	0.8483	0.8489
	GRU	0.9633	0.9636	0.99631	0.9631	0.862	0.8638	0.8626	0.8629
	CNN	0.8822	0.8912	0.8813	0.8912	0.7647	0.7731	0.7854	0.7625
State-of-the-art	[19]	0.5843	0.5726	0.5531	0.5627	0.5735	0.5721	0.5443	0.5579
	[20]	0.8254	0.8435	0.8617	0.8525	0.8029	0.7986	0.8125	0.8055
	[16]	0.8711	0.8217	0.7998	0.8106	0.8567	0.8533	0.8651	0.8592
	Att-RNN	0.9384	0.9405	0.9388	0.9391	0.9324	0.9343	0.9322	0.9326
	Ours	**1.0**	**1.0**	**1.0**	**1.0**	**0.9996**	**0.9996**	**0.9996**	**0.9996**
Baselines	**Methods**	**GID-S**				**GID-M**			
		Accuracy	**Precision**	**Recall**	**F1-score**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Non-DL Baseline	SVM	0.9981	0.9981	0.9981	0.9981	0.993	0.993	0.993	0.993
	RF	0.9878	0.9878	0.9878	0.9878	0.9954	0.9954	0.9954	0.9954
	KNN	0.9979	0.9979	0.9979	0.9979	0.9953	0.9953	0.9953	0.9953
	AdaB	0.5408	0.5689	0.5409	0.4849	0.5401	0.5135	0.542	0.4985
	LDA	0.688	0.6893	0.688	0.6855	0.6933	0.693	0.6933	0.6915
DL Baseline	LSTM	0.9951	0.9951	0.9951	0.9951	0.9935	0.9936	0.9936	0.9936
	GRU	0.9949	0.9949	0.9949	0.9949	0.9938	0.9938	0.9938	0.9938
	CNN	0.9932	0.9932	0.9932	0.9932	0.9845	0.9845	0.9845	0.9845
State-of-the-art	[9]	0.9721	0.9789	0.9745	0.9767	0.9653	0.9627	0.9669	0.9648
	[1]	0.9931	0.9934	0.9957	0.9945	0.9901	0.9931	0.9942	0.9936
	[39]	0.9917	0.9899	0.9917	0.9908	0.9875	0.9826	0.9844	0.9835
	Att-RNN	0.99	0.99	0.99	0.99	0.9894	0.9895	0.9895	0.9895
	Ours	**0.9983**	**0.9983**	**0.9983**	**0.9983**	**0.9961**	**0.9961**	**0.9961**	**0.9961**

DL denotes Deep Learning. Att-RNN denotes attention-based RNN.

for comparison (three for EEG-based authentication and three for gait-based authentication) are introduced as follows:

- Jayarathne et al. [19] feed EEG data to a bandpass filter (8 Hz − 30 Hz), extract CSP (Common Spatial Pattern) and recognize the user ID by LDA.
- Keshishzadeh et al. [20] extract autoregressive coefficients as the features and identify the subject by SVM.
- Gui et al. [16] employ low-pass filter (60 Hz) and wavelet packet decomposition to generate features and distinguish unauthorized person through deep neural network.
- Cola et al. [9] hire neural networks to analyze user gait pattern by artificial features such as kurtosis, peak-to-peak amplitude, and skewness.
- Al-Naffakh et al. [1] propose to utilize time-domain statistical features and Multilayer Perception (MLP) for person identification.

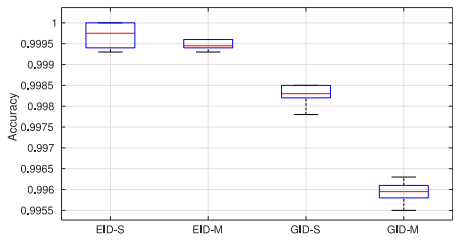


Fig. 6. Impact of session.

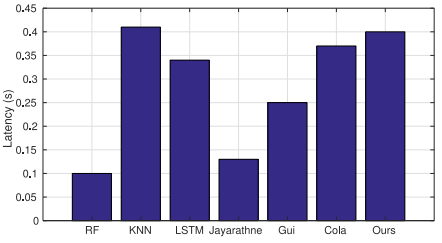


Fig. 7. Latency.

- Sun et al. [39] adopt a weighted voting classifier to process the extracted features like gait frequency, symmetry coefficient, and dynamic range.

The primary conclusions from Table 7 are summarized as follows:

- Our approach achieves the highest performance over both EEG and gait datasets under single session and multi-session settings.
- For most of the baselines, we observe lower accuracy in the EEG dataset compared to that in the gait dataset, implying the EEG-based authentication is still challenging and our approach has room to improve; nevertheless, our model outperforms others and shows superiority to both EEG-based or Gait-based methods.
- The results on single-session datasets are better than that on multi-session datasets. This is reasonable, and it demonstrates that the number of sessions does affect the authentication accuracy. This problem will be further analyzed in Section 4.3.1.
- Our model achieves better performance than Att-RNN. Since the diversity between our model and Att-RNN is that we employ an external classifier, this observation demonstrates that the external classifier is effective and efficient.

In DeepKey, the subjects passing the Invalid ID Filter are regarded as genuine only if their recognized IDs are consistent, i.e., $E_{ID} = G_{ID}$. It can be inferred easily that the FAR of DeepKey is 0 as well. However, the FRR depends on one or more of these three scenarios: the false rejection of Invalid ID Filter; the incorrect Gait identification; the incorrect EEG identification. In summary, the overall FAR is **0** and the overall FRR is calculated as $\mathbf{1}\% \approx 0.006 + 0.994 * (1 - 0.9961 * 0.9996)$.

4.3 Impact of Key Parameters

4.3.1 Impact of Sessions. In practical applications, sessions in different scenarios may result in a minor difference in equipment position, signal quality, and other factors. To investigate the impact of sessions, we conduct external experiments by comparing the performance between single-session datasets and multi-session datasets. The comprehensive evaluation metrics and the comparison over various baselines are listed in Table 6 and Table 7, while the comparison is summarized in Figure 6. The experiment results show that on the multi-session datasets, compared with the single-session datasets, we achieve a slightly lower but still highly competitive performance.

4.3.2 Impact of EEG Band. A series of comparison experiments are designed to explore the optimal EEG frequency band which contains the most discriminative features. The results presented in Table 8 illustrate the following:

- The Delta band consistently provides higher identification accuracy compared to other frequency bands for both single and multiple sessions. This observation shows that Delta pattern contains the most discriminative information for person identification.

Table 8. EEG Bands Comparison

Dataset	Baseline	Methods	EEG Bands						Best Level	Best Band
			Delta	Theta	Alpha	Beta	Gamma	Full		
EID-S	Non-DL Baseline	SVM	0.4588	0.4682	0.7484	0.5955	0.5239	0.8788	0.8788	Full
		RF	0.9875	0.8006	0.7729	0.6376	0.5469	0.8931	0.9875	Delta
		KNN	0.9897	0.8084	0.7465	0.5553	0.4606	0.8792	0.9897	Delta
		AdaB	0.2872	0.2879	0.2318	0.3069	0.2922	0.3289	0.3289	Full
		LDA	0.1567	0.1802	0.1957	0.1502	0.1306	0.4547	0.4547	Full
	DL Baseline	LSTM	0.9596	0.8126	0.8277	0.6906	0.6027	0.9273	0.9596	Delta
		GRU	0.9633	0.7996	0.8082	0.6902	**0.6985**	0.9251	0.9633	Delta
		CNN	0.8822	0.7416	0.8079	**0.6918**	0.6059	0.8985	0.8985	Full
	State-of -the-art	[19]	0.5843	0.4487	0.2918	0.3017	0.4189	0.5112	0.5843	Delta
		[20]	0.8254	0.7935	0.7019	0.6368	0.6621	0.8018	0.8254	Delta
		[16]	0.8711	0.8531	0.7556	0.6882	0.5101	0.7819	0.8711	Delta
		Att-RNN	0.9384	0.7928	0.8318	0.6854	0.6046	0.9238	0.9384	Delta
		Ours	**1.0**	**0.9285**	**0.8366**	0.5529	0.4558	0.9417	1.0	Delta
EID-M	Non-DL Baseline	SVM	0.7796	0.5424	0.5664	0.6522	0.4915	0.7477	0.7796	Delta
		RF	0.8124	0.7194	0.7351	**0.6842**	0.4765	0.8121	0.7194	Delta
		KNN	0.8211	0.7501	0.7649	0.6611	0.3821	0.8162	0.7501	Delta
		AdaB	0.3228	0.3095	0.2478	0.2548	0.2529	0.3189	0.3228	Delta
		LDA	0.3082	0.1681	0.1621	0.1824	0.1311	0.2995	0.3082	Delta
	DL Baseline	LSTM	0.8482	0.6926	0.7438	0.5726	0.5008	0.8185	0.8482	Delta
		GRU	0.862	0.6935	0.7531	0.5672	**0.5072**	0.8221	0.862	Delta
		CNN	0.7647	0.6712	0.7191	0.5588	0.4949	0.7749	0.7749	Full
	State-of- the-art	[19]	0.9721	0.7019	0.7091	0.4189	0.4089	0.8195	0.9721	Delta
		[20]	0.9931	0.6891	0.6988	0.5124	0.3397	0.7963	0.9931	Delta
		[16]	0.9917	0.7199	0.6572	0.4911	0.3977	0.8011	0.9917	Delta
		Att-RNN	0.9324	0.6847	0.6846	0.5732	0.4941	0.7976	0.9324	Delta
		Ours	**0.9996**	**0.9013**	**0.8989**	0.4428	0.3661	**0.8858**	0.9996	Delta

The full band denotes the raw EEG data with full frequency bands.

- Our method gains the best outcome on both datasets with different sessions. This validates the robustness and adaptability of the proposed approach.

Why could Delta pattern outperform other patterns since Delta wave mainly appears in deep sleep state? Here we give one possible reason. We know that the EEG patterns are associated with an individual's mental and physical states (organics and systems). For example, while the subject is under deep sleep and producing Delta pattern, the majority of physical functions of the body (such as sensing, thinking, even dreaming) are completely detached. Only the very essential life-support organs and systems (such as breathing, heart beating, and digesting) keep working, which indicates the brain areas corresponding to life-support functions are active. While the subject is awake (e.g., relax state) and producing Alpha pattern, the subject has more activated functions such as imaging, visualizing, and concentrating. Also, more brain functions like hearing, touching, and thinking are attached, which means that more physical brain areas (such as frontal lobe, temporal lobe, and parietal lobe) are activated. At this time, the life-support organs are still working. In short, only the life-support organs related brain areas are active in the first scenario (Delta pattern), while the brain areas controlling life-support and high-level functions (e.g., concentrating) are active in

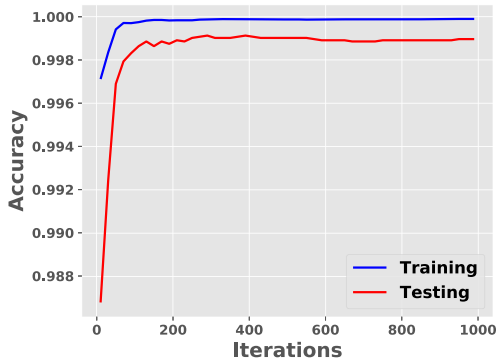


Fig. 8. Convergence curves.

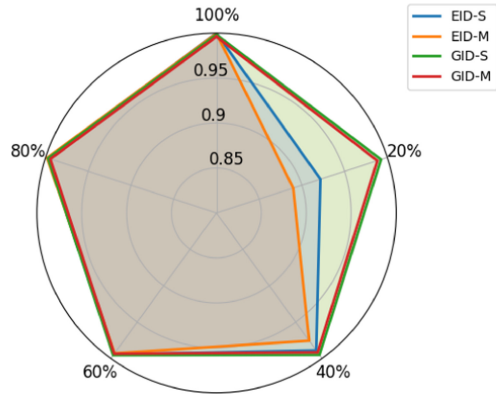


Fig. 9. Impact of datasize.

the second scenario (Alpha pattern). Thus, we infer that the Delta pattern corresponds to the life-support organs and systems, which is the most stable function in different scenarios and the most discriminative signal in inter-subject classification.

4.3.3 Impact of Datasize. Datasize is one crucial influence element in deep learning based algorithms. In this section, we conduct experiments to train the proposed method over various data sizes. As shown in the radar chart (Figure 9), four datasets are evaluated under different proportions of training datasize. It has five equi-angular spokes which represent the proportion of datasize (20%, 40%, 60%, 80%, 100%), respectively. The four concentric circles indicate the accuracy, which are 85%, 90%, 95%, and 100%, respectively. Each closed line represents a dataset and has five intersections with the five spokes. Each intersection node represents the classification accuracy of a specific dataset with a specific proportion datasize. For example, the intersection of the EID-S line and 20% spoke is about 0.92, which denotes that our approach achieves the accuracy of 92% over the EID-S dataset with 20% datasize. The radar chart infers that gait datasets (GID-S and GID-M) obtain competitive accuracy even with 20% datasize; nevertheless, EID-S and EID-M highly rely on the datasize. This phenomenon is reasonable because EEG data has a lower signal-to-noise ratio and requires more samples to learn the latent distribution pattern.

4.4 Latency Analysis

We also study the latency of DeepKey since low delay is highly desirable in real-world deployment. The latency of DeepKey is compared to the response time of several state-of-the-art authentication approaches. The comparison results are shown in Figure 7. The testing latency of our method is 0.39 s, which is competitive compared to the state-of-the-art baselines. Specifically, the reaction time of DeepKey is composed of several components, with the Invalid ID Filter taking 0.06 s and the ID Identification taking 0.33 s.

The overall system latency not only includes the computation latency but also includes the data collection latency such as the time cost when wearing the EEG headset and the IMUs. In detail, the data collection latency of our system is around 10 s, while EEG collection requires about 4 s and gait collection requires above 6 s. We can see that the data collection latency is much higher than computational latency; as a result, one major future direction to reduce system latency is to save the equipment-related time (discussed in Section 4.5).

4.5 Usability

The users have to sequentially wear two different devices to complete EEG and gait data collection, which may lead to incontinence and time-wasting. In this section, we justify the usability of the proposed DeepKey system in several aspects. First, the two equipments are non-invasive, light, and easy-to-wear, which will not cost lots of user effort. Second, as mentioned in Section 4.4, the authentication latency is very small (0.39 s); even counting the data collection latency, the whole authentication procedure will cost less than 10 s, which is tolerable for the users. Third, the proposed DeepKey aims at the highly confidential scenarios (e.g., military bases, the treasuries of banks, and political offices) which require strict precision and very high fake-resistance, in which the inconvenience is acceptable.

Even so, we agree that simplifying the authentication procedure will promote the deployment of the system. We attempt to solve this issue in the future. On the one hand, with the development of hardware- related techniques, the EEG headset is becoming more portable and affordable. For example, the developed cEEGrids,[10] flex-printed, multi-channel sensor arrays that are placed around the ear using an adhesive, are easy and comfortable to wear and dispatch. This is a good choice of EEG acquisition equipment, although the cEEGrids are currently expensive. On the other hand, a potential solution is to develop a device-free authentication system and measure gait signals by environmental sensors such as RFID (Radio Frequency IDentification) tags. The Received Signal Strength Indication (RSSI) of an RFID tag measures the received signal power which reflects the target subject's walking information [49]. Another possible solution targeting the two-factor design is to integrate gait sensors into the EEG headset. This will decrease the equipment expense and require less user effort, although this may cause a slight decrease in the authentication accuracy due to the fewer IMUs (there are three IMUs in our experiments but only one IMU after the integration with Emotiv headset). In the future, the wide deployment of the DeepKey authentication system in a real-world environment is possible.

5 DISCUSSIONS AND FUTURE WORK

In this article, we propose a biometric identification system based on both EEG and gait information. In this section, we discuss the open challenges and potential future research directions.

First, the datasets used in this article only have limited participants. Extensive evaluations over more subjects are necessary. However, compared to some existing works ([50], [4], and [39] have 8, 9, and 10 subjects in their experiments, respectively), we believe our participation scale is acceptable. Our work has already demonstrated that DeepKey can be used in settings such as small offices which are accessed by a limited group of people. In addition, evaluation can be improved by extending observations on how the system performs in different conditions. For example, considering changes in EEG signals during more trials, and longer times (hours, days, or even months) to understand if these are consistent and reliable for detection.

Second, wearable sensors like an EEG headset and wearable IMUs are required in the data collection stage of the DeepKey system. Extensive experiments are meaningful in the future to investigate how the placement position of the wearable sensor matters. The EEG headset position on the head and the IMU position on the arms may affect the authentication performance.

Additionally, a promising future work is the real-world online deployment of the proposed DeepKey system. Since we have demonstrated the effectiveness and efficiency of DeepKey in the offline situation, a future step is to build a real-time authentication environment to evaluate the online performance.

[10] http://ceegrid.com/home/.

The brain signals could be easily affected by environmental factors (e.g., noise), subjective factors (e.g., emotion and fatigue), and other brain activities (e.g., eye-opening, staring at a specific image, mental working). The proposed DeepKey is robust to objective and subjective factors due to the deep representation learning. As for other brain activities, various brain activities may lead to different authentication performance. In this work, we collect the brain signals while the user is relaxing with eyes closed because this status will produce more stable EEG signals and have less intra-subject difference. One future work is supposed to investigate how the different brain activities will impact the authentication accuracy.

Furthermore, the proposed DeepKey still faces the challenge of the "in the wild" scenario since the gathered EEG data are easily corrupted by physical actions like walking. In this work, the EEG and gait data are gathered in two separate steps. However, in outdoor environments, the user is hardly standing still and waiting for the authentication. Fortunately, our system has competitive performance in a fixed indoor environment (such as bank vouchers) which can provide a stable data collection environment and mainly concerned about high fake-resistance.

6 CONCLUSION

Taking advantage of both EEG- and gait-based systems for fake-resistance, we propose DeepKey, a multimodal biometric authentication system, to overcome the limitations of traditional unimodal biometric authentication systems. DeepKey contains three independent models: an Invalid ID Filter Model, a Delta band based EEG Identification Model, and a Gait Identification Model, to detect invalid EEG data, and recognize the EEG ID and Gait ID, respectively. The DeepKey system outperforms the state-of-the-art baselines by achieving a FAR of 0 and a FRR of 1%. In addition, the key parameters (such as sessions and EEG frequency band) and the system latency are also investigated by extensive experiments.

This work sheds light on further research on multimodal biometric authentication systems based on EEG and gait data. Our future work will focus on deploying the DeepKey system in an online real-world environment. In addition, the gait signals are currently gathered by three wearable IMUs, which may obstruct the large-scale deployment in practice. Therefore, another direction in the future is to collect gait data from non-wearable gait solutions (e.g., sensors deployed in environments).

REFERENCES

[1] Neamah Al-Naffakh, Nathan Clarke, Fudong Li, and Paul Haskell-Dowland. 2017. Unobtrusive gait recognition using smartwatches. In *2017 International Conference of the Biometrics Special Interest Group (BIOSIG'17)*. IEEE, 1–5.

[2] Salahiddin Altahat, Girija Chetty, Dat Tran, and Wanli Ma. 2015. Analysing the robust EEG channel set for person authentication. In *International Conference on Neural Information Processing*. Springer, 162–173.

[3] Corey Ashby, Amit Bhatia, Francesco Tenore, and Jacob Vogelstein. 2011. Low-cost electroencephalogram (EEG) based authentication. In *5th International IEEE/EMBS Conference on Neural Engineering (NER'11)*. IEEE, 442–445.

[4] Md Khayrul Bashar, Ishio Chiaki, and Hiroaki Yoshida. 2016. Human identification from brain EEG signals using advanced machine learning method EEG-based biometrics. In *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES'16)*. IEEE, 475–479.

[5] Nikolaos V. Boulgouris and Xiaxi Huang. 2013. Gait recognition using HMMs and dual discriminative observations for sub-dynamics analysis. *IEEE Transactions on Image Processing* 22, 9 (2013), 3636–3647.

[6] Michele L. Callisaya, Leigh Blizzard, Michael D. Schmidt, Jennifer L. McGinley, Stephen R. Lord, and Velandai K. Srikanth. 2009. A population-based study of sensorimotor factors affecting gait in older people. *Age and Ageing* 38, 3 (2009), 290–295.

[7] William Chan, Navdeep Jaitly, Quoc V. Le, Oriol Vinyals, and Noam M. Shazeer. 2017. Speech recognition with attention-based recurrent neural networks. U.S. Patent No. 9,799,327 [P]. 2017-10-24.

[8] John Chuang, Hamilton Nguyen, Charles Wang, and Benjamin Johnson. 2013. I think, therefore I am: Usability and security of authentication using brainwaves. In *International Conference on Financial Cryptography and Data Security*. Springer, 1–16.

[9] Guglielmo Cola, Marco Avvenuti, Fabio Musso, and Alessio Vecchio. 2016. Gait-based authentication using a wrist-worn device. In *Mobiquitous*. ACM, 208–217.

[10] John G. Daugman. 1993. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 11 (1993), 1148–1161.

[11] Mohammad Derawi and Iurii Voitenko. 2014. Fusion of gait and ECG for biometric user authentication. In *International Conference of the Biometrics Special Interest Group (BIOSIG'14)*. IEEE, 1–4.

[12] Manqing Dong, Lina Yao, Xianzhi Wang, Boualem Benatallah, Xiang Zhang, and Quan Z. Sheng. 2019. Dual-stream self-attentive random forest for false information detection. In *2019 International Joint Conference on Neural Networks (IJCNN'19)*. IEEE, 1–8.

[13] Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi, and Rim Slama. 2013. 3D face recognition under expressions, occlusions, and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 9 (2013), 2270–2283.

[14] Geof H. Givens, J. Ross Beveridge, Yui Man Lui, David S. Bolme, Bruce A. Draper, and P. Jonathon Phillips. 2013. Biometric face recognition: From classical statistics to future challenges. *Wiley Interdisciplinary Reviews: Computational Statistics* 5, 4 (2013), 288–308.

[15] Steven Goldstein. 2016. Methods and systems for voice authentication service leveraging networking. U.S. Patent No. 9,282,096 [P]. 2016-3-8.

[16] Qiong Gui, Zhanpeng Jin, and Wenyao Xu. 2014. Exploring EEG-based biometrics for user identification and authentication. In *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB'14)*. IEEE, 1–6.

[17] Ana M. Guzman, Mohammed Goryawala, Jin Wang, Armando Barreto, Jean Andrian, Naphtali Rishe, and Malek Adjouadi. 2013. Thermal imaging as a biometrics approach to facial signature authentication. *IEEE Journal of Biomedical and Health Informatics* 17, 1 (2013), 214–222.

[18] Thang Hoang and Deokjai Choi. 2014. Secure and privacy enhanced gait authentication on smart phone. *The Scientific World Journal* 2014, 438254 (2014).

[19] Isuru Jayarathne, Michael Cohen, and Senaka Amarakeerthi. 2016. BrainID: Development of an EEG-based biometric authentication system. In *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON'16)*. IEEE, 1–6.

[20] Sarineh Keshishzadeh, Ali Fallah, and Saeid Rashidi. 2016. Improved EEG based human authentication system on large dataset. In *24th Iranian Conference on Electrical Engineering (ICEE'16)*. IEEE, 1165–1169.

[21] Nefissa Khiari-Hili, Christophe Montagne, Sylvie Lelandais, and Kamel Hamrouni. 2016. Quality dependent multimodal fusion of face and iris biometrics. In *6th International Conference on Image Processing Theory Tools and Applications (IPTA'16)*. IEEE, 1–6.

[22] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[23] Shinsuke Konno, Yoshitaka Nakamura, Yoh Shiraishi, and Osamu Takahashi. 2015. Gait-based authentication using trouser front-pocket sensors. *International Workshop on Informatics* (2015).

[24] Pavan K. Kumar, PESN Krishna Prasad, M. V. Ramakrishna, and B. D. C. N. Prasad. 2013. Feature extraction using sparse SVD for biometric fusion in multimodal authentication. *International Journal of Network Security & Its Applications* 5, 4 (2013), 83.

[25] Worapan Kusakunniran, Qiang Wu, Jian Zhang, Hongdong Li, and Liang Wang. 2014. Recognizing gaits across views through correlated motion co-clustering. *IEEE Transactions on Image Processing* 23, 2 (2014), 696–709.

[26] Worapan Kusakunniran, Qiang Wu, Jian Zhang, Yi Ma, and Hongdong Li. 2013. A new view-invariant feature for cross-view gait recognition. *IEEE Transactions on Information Forensics and Security* 8, 10 (2013), 1642–1653.

[27] Neal S. Latman and Emily Herb. 2013. A field study of the accuracy and reliability of a biometric iris recognition system. *Science & Justice* 53, 2 (2013), 98–102.

[28] Yann LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. 1998. Efficient backprop. In *Neural Networks: Tricks of the Trade*. Springer, 9–50.

[29] Tran Long, Le Thai, and Tran Hanh. 2012. Multimodal biometric person authentication using fingerprint, face features. *PRICAI 2012: Trends in Artificial Intelligence* (2012), 613–624.

[30] Dario Maio, Davide Maltoni, Raffaele Cappelli, James L. Wayman, and Anil K. Jain. 2002. FVC2002: Second fingerprint verification competition. In *Proceedings of the 16th International Conference on Pattern Recognition, 2002*, Vol. 3. IEEE, 811–814.

[31] Ju Man and Bir Bhanu. 2006. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 2 (2006), 316–322.

[32] B. E. Manjunathswamy, Appaji M. Abhishek, J. Thriveni, K. R. Venugopal, and L. M. Patnaik. 2015. Multimodal biometric authentication using ECG and fingerprint. *International Journal of Computer Applications* 111, 13 (2015).
[33] Sebastien Marcel and José del R. Millán. 2007. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 4 (2007).
[34] Daigo Muramatsu, Akira Shiraishi, Yasushi Makihara, Md Zasim Uddin, and Yasushi Yagi. 2015. Gait-based person recognition using arbitrary view transformation model. *IEEE Transactions on Image Processing* 24, 1 (2015), 140–154.
[35] C. Y. Yam and Mark Nixon. 2009. Model-based gait recognition. In *Encyclopedia of Biometrics*. Springer, 633–639.
[36] Jaishanker K. Pillai, Maria Puertas, and Rama Chellappa. 2014. Cross-sensor iris recognition through kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 1 (2014), 73–85.
[37] Kun Qian, Chenshu Wu, Zheng Yang, Zimu Zhou, Xu Wang, and Yunhao Liu. 2018. Enabling phased array signal processing for mobile WiFi devices. *IEEE Transactions on Mobile Computing* 17, 8 (2018), 1820–1833.
[38] Fahreddin Sadikoglu and Selin Uzelaltinbulat. 2016. Biometric retina identification based on neural network. *Procedia Computer Science* 102 (2016), 26–33.
[39] Bing Sun, Yang Wang, and Jacob Banda. 2014. Gait characteristic analysis and identification based on the iPhone's accelerometer and gyrometer. *Sensors* 14, 9 (2014), 17037–17054.
[40] Rupali L. Telgad, P. D. Deshmukh, and Almas M. N. Siddiqui. 2014. Combination approach to score level fusion for Multimodal Biometric system by using face and fingerprint. In *Recent Advances and Innovations in Engineering (ICRAIE), 2014*. IEEE, 1–8.
[41] Kavitha P. Thomas and A. P. Vinod. 2016. Utilizing individual alpha frequency and delta band power in EEG based biometric recognition. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC'16)*. IEEE, 004787–004791.
[42] Kavitha P. Thomas and A. P. Vinod. 2017. EEG-based biometric authentication using gamma band power during rest state. *Circuits, Systems, and Signal Processing* (2017), 1–13.
[43] J. A. Unar, Woo Chaw Seng, and Almas Abbasi. 2014. A review of biometric technology along with trends and prospects. *Pattern Recognition* 47, 8 (2014), 2673–2688.
[44] Yequan Wang, Minlie Huang, Li Zhao, et al. 2016. Attention-based LSTM for aspect-level sentiment classification. In *EMNLP*. 606–615.
[45] Zifeng Wu, Yongzhen Huang, Liang Wang, Xiaogang Wang, and Tieniu Tan. 2017. A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 2 (2017), 209–226.
[46] Weitao Xu, Guohao Lan, Qi Lin, Sara Khalifa, Mahbub Hassan, Neil Bergmann, and Wen Hu. 2018. Keh-gait: Using kinetic energy harvesting for gait-based user authentication systems. *IEEE Transactions on Mobile Computing* 18, 1 (2018), 139–152.
[47] Vitor Yano, Alessandro Zimmer, and Lee Luan Ling. 2012. Multimodal biometric authentication based on iris pattern and pupil light reflex. In *21st International Conference on Pattern Recognition (ICPR'12)*. IEEE, 2857–2860.
[48] Lina Yao, Quan Z Sheng, Xue Li, Tao Gu, Mingkui Tan, Xianzhi Wang, Sen Wang, and Wenjie Ruan. 2018. Compressive representation for device-free activity recognition with passive RFID signal strength. *IEEE Transactions on Mobile Computing* 17, 2 (2018), 293–306.
[49] Lina Yao, Quan Z. Sheng, Xue Li, Sen Wang, Tao Gu, Wenjie Ruan, and Wan Zou. 2015. Freedom: Online activity recognition via dictionary-based sparse representation of RFIID sensing data. In *2015 IEEE International Conference on Data Mining*. IEEE, 1087–1092.
[50] Xiang Zhang, Lina Yao, Salil S. Kanhere, Yunhao Liu, Tao Gu, and Kaixuan Chen. 2018. MindID: Person identification from brain waves through attention-based recurrent neural network. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 149.
[51] Xiang Zhang, Lina Yao, Xianzhi Wang, Jessica Monaghan, and David Mcalpine. 2019. A survey on deep learning based brain computer interface: Recent advances and new frontiers. *arXiv preprint arXiv:1905.04149* (2019).
[52] Hailing Zhou, Ajmal Mian, Lei Wei, Doug Creighton, Mo Hossny, and Saeid Nahavandi. 2014. Recent advances on singlemodal and multimodal face recognition: A survey. *IEEE Transactions on Human-Machine Systems* 44, 6 (2014), 701–716.

Received June 2019; revised January 2020; accepted April 2020